Detection and Classification of Sleep Apnea and Hypopnea using PPG and SpO$_2$ signals

Remo Lazazzera [iD], Margot Deviaene [iD], Carolina Varon [iD], Bertien Buyse, Dries Testelmans [iD], Pablo Laguna [iD], Eduardo Gil [iD], and Guy Carrault [iD].

Abstract—In this work, a detection and classification method for sleep apnea and hypopnea, using photopletysmography (PPG) and peripheral oxygen saturation (SpO$_2$) signals, is proposed. The detector consists of two parts: one that detects reductions in amplitude fluctuation of PPG (DAP) and one that detects oxygen desaturations. To further differentiate among sleep disordered breathing events (SDBE), the pulse rate variability (PRV) was extracted from the PPG signal, and then used to extract features that enhance the sympatho-vagal arousals during apneas and hypopneas. A classification was performed to discriminate between central and obstructive events, apneas and hypopneas. The algorithms were tested on 96 overnight signals recorded at the UZ Leuven hospital, annotated by clinical experts, and from patients without any kind of co-morbidity. An accuracy of 75.1% for the detection of apneas and hypopneas, in one-minute segments, was reached. The classification of the detected events showed 92.6% accuracy in separating central from obstructive apnea, 83.7% for central apnea and central hypopnea and 82.7% for obstructive apnea and obstructive hypopnea. The low implementation cost showed a potential for the proposed method of being used as screening device, in ambulatory scenarios.

Index Terms—Apnea detection, apnea classification, PPG, SpO$_2$, PRV, DAP.

I. Introduction

SLEEP disordered breathing events (SDBE) detection and classification are presented in this work, focusing on apnea and hypopneas. Sleep apnea is the absence of airflow during sleep, while hypopnea is a sleep airflow reduction [1]. Both can be obstructive or central: obstructive if the breathing effort continues, while there is a mechanical obstruction of the airways, resulting in interruptions of the airflow; central if no breathing effort is present. This disorder disrupts the normal sleep pattern and is associated with daytime sleepiness and fatigue [2]. The stress that the heart and brain undergo with each apnea/hypopnea, can lead to long-term complications: high blood pressure, cardiovascular disease, stroke and

R. Lazazzera and G. Carrault are with Laboratoire Traitement du Signal et de l'Image (LTSI-Inserm UMR 1099), Université de Rennes 1, Rennes 35000, France (e-mail: guy.carrault@univ-rennes1.fr; remo.lazazzera@gmail.com).

E. Gil and P. Laguna with the Biomedical Signal Interpretation and Computational Simulation (BSICoS) Group, I3A, IIS Aragón, University of Zaragoza, and with the CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza 50018, Spain (e-mail: edugilh@unizar.es; laguna@unizar.es).

M. Deviaene is with the Department of Electrical Engineering-ESAT, STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven B-3001, Belgium; C. Varon is with the Circuits and Systems (CAS) group, Delft University of Technology, the Netherlands (e-mail: margot.deviaene@esat.kuleuven.be; j.c.varon@tudelft.nl).

D. Testelmans and B. Buyse are with the Department of Pneumology, UZ Leuven, Leuven 3000, Belgium (e-mail: dries.testelmans@uzleuven.be; bertien.buyse@uzleuven.be).

diabetes [3], [4]. For this reason its detection, diagnosis and treatment are important.

Benjafield et al. in 2019, published the first study to report global prevalence of obstructive sleep apnea, showing that almost 1 billion people are globally affected, with prevalence exceeding 50% in some countries. The number of affected individuals was highest in China, followed by the USA, Brazil, and India. The study estimated that, globally, 936 million adults aged 30-69 years (men and women) have mild to severe obstructive sleep apnea and 425 million adults aged 30-69 years have moderate to severe obstructive sleep apnea [5].

The diagnosis of sleep apnea and hypopnea is done using either polysomnography or polygraphy. Polysomnography (PSG) is the gold standard procedure for its diagnosis. It consists of an overnight recording of different electrophysiological signals such as electroencephalogram, electromyogram, electrooculogram, electrocardiogram, airflow, peripheral oxygen saturation (SpO$_2$) and photoplethysmogram (PPG), chest and abdominal movements. The acquisition and analysis of these signals require human expertise and specialized equipment, being a very uncomfortable and costly procedure. This is why sleep disorders are often under-diagnosed. Polygraphy is a less restrictive examination only measuring (cardio-)respiratory parameters usually including airflow, respiratory movement and oxygen saturation. This technique can more easily be performed at home and requires the use of a nasal cannula, which is cumbersome and could interfere with the natural sleep. So different techniques for home sleep apnea/hypopnea monitoring were extensively developed [6], minimizing the equipment used. It motivated multiple studies focused on developing low cost and unobtrusive systems for sleep apnea/hypopnea detection, based on few physiological signals. Pulse oximetry, which exploits the PPG and SpO$_2$ signals recorded at the level of peripheral parts of the body, would be less invasive and more convenient.

The PPG signal is obtained by means of a pulse oximeter, which illuminates the skin, and measures changes in light absorption of the blood volume [7]. The SpO$_2$ signal, furthermore, is computed by combining the information of two PPG signals, acquired at different wavelengths: infrared and red light, usually at the level of the index finger [8].

From the PPG signal, reductions in blood flow can be observed. These fluctuations in the amplitude are markers of sympathetic discharge [9], [10] that produces vasoconstriction [11]–[13], possibly related to transient arousals. Movements and deep inspiratory gasps produce sympathetic activation and besides them, the same effect is also caused by apneas

and hypopneas [14]. This activation occurs because after the manifestation of SDBE, the ANS tries to restart the normal breathing with a discharge of the sympathetic activity. In addition, hypoxia plays a key role during apneas and hypopneas [15], [16], and it can be quantified using the SpO_2 signal. Therefore, these effects can be detected using pulse oximeter systems, by combining the information of PPG and SpO_2 signals [9], [10], [17]–[19].

Heart rate variability (HRV) represents fluctuations in the heart rate, related to autonomic nervous system (ANS) control. High frequency (HF) components between 0.15 and 0.4 Hz represent the vagal tone, while frequencies from 0.04 to 0.15 Hz manifest the activation of both parasympathetic and sympathetic nervous systems, and these are labeled low-frequency (LF) components. The ratio between LF and HF is defined as the sympatho-vagal balance [20]. The pulse rate time series, or so-called pulse rate variability (PRV), can be derived from the PPG and is a surrogate of the heart rate time series. Its usefulness was for instance demonstrated in [21], during tilt table tests, when the sympathetic activation significantly influences the heart rate modulation.

Multiple studies investigated the diagnosis of obstructive sleep apnea syndrome (OSAS) based on the detection of vasoconstriction, using peripheral arterial tonometry [13], [22], [23]. The relationship between autonomic nervous system and PPG was also studied in [12], [24]. Detection of sleep apnea and hypopnea, from PPG [25] and HRV were already explored in literature [26]–[29]. The decreases in the amplitude fluctuation of PPG (DAP) detector was tested for detecting obstructive hypopnea events in [25] and combined with HRV analysis in [30], [31], thereby demonstrating the usefulness of DAP to diagnose OSAS in children.

Since studies [9], [10], [14], [17]–[19] already showed that DAP and oxygen desaturation information are linked to apneas/hypopneas, the aim of this work is to test the DAP detector on adults, combining it with oxygen desaturation events for detecting SDBE. The performances of the proposed method are then compared with other methods, tested on the same database. In literature several studies also explored the possibility of detecting apneas and hypopneas by investigating the oxygen desaturation signal. The most promising techniques used a machine learning approach [32], [33]. In [32], Deviaene et al. reached an averaged desaturation classification accuracy of 82.8% using a random forest classifier, over different test sets of the Sleep Heart Health Study containing 8052 subjects in total. Mostafa et al. in [33] implemented a Deep Belief Network and used a 10-fold cross validation on two public databases: one with 8 subjects and other with 25 subjects. They achieved, respectively, an accuracy of 85.26% and of 97.64%.

Finally, another hypothesis tested in the present paper, is whether the use of sympatho-vagal balance and oxygen desaturation information could be significant to classify the different types of apneas and hypopneas. Besides the detection, a respiratory event classification was performed per apneic/hypopneic SDBE detected, comparing different classifiers. Classification in central or obstructive apnea is as relevant as the distinction between apnea and hypopnea, to discover the nature of the event.

The aim of this study is to detect and later classify SDBEs from a PPG and a SpO_2 signal, both acquired at the level of the index finger. Events of Central Apnea (CA), Central Hypopnea (CH), Obstructive Apnea (OA), Obstructive Hypopnea (OH) and Mixed Apnea (MA) were all initially considered as Sleep Disruptive Breathing Events (SDBE). First, SDBE detection is performed without identifying the SDBE nature. Next, a classification is performed by including features derived from PRV, for separating Central (C) from Obstructive (O) SDBEs and Apneic from Hypopneic SDBEs.

The system proposed in this work is designed for feasible ambulatory monitoring, based on oximetry devices providing both SpO_2 and PPG signals. It would be of great interest for an ambulatory sleep monitoring system, not intended as an alternative to the PSG but as a first screening step.

This paper is organized as follows: at first, the DAP detector and the oxygen-desaturation detector, used to discriminate between apneic and non-apneic SDBEs, are described; next, the classification of the SDBEs as apneic or hypopneic, central or obstructive is presented. This classification was performed by introducing indexes of the PRV spectral-analysis that give relevant information about the ANS activity. Both detection and classification results are then introduced. In the discussion section, a comparison of the proposed algorithms, with those present in literature and those implemented on the same database, is presented; the final conclusions and further improvements close the paper.

II. MATERIALS AND METHODS

A. Data

A database consisting of 96 overnight recordings of patients suspected to suffer from sleep apnea-hypopnea syndrome and without any cardiovascular co-morbidity, was provided by the Sleep Laboratory of the University Hospitals Leuven (UZ Leuven, Belgium). Each record contains a PPG and an SpO_2 signal sampled at 500 Hz, and the apnea-hypopnea index (AHI) was calculated as the amount of respiratory events per hour of sleep, scored according to the AASM 2012 rules [34]. The average AHI in the dataset is 31.3 and 39% of the subjects had an AHI larger than 30; 53% of them had an AHI between 5 and 30 and the remaining 8%, an AHI less or equal to 5. The annotations contain the beginning and duration of CA, CH, OA, OH and MA. Table I shows the total number of annotations per SDBE category, in the database.

TABLE I
TOTAL NUMBER OF SLEEP DISORDERED BREATHING EVENTS, PER CATEGORY, IN THE DATABASE

	CA	CH	OA	OH	MA
Number of events	765	689	4984	14140	750
Percentage of events	3.6%	3.2%	23.4%	66.3%	3.5%

B. Apnea and hypopnea detection

The algorithm for sleep apnea and hypopnea detection was based on the detection of DAP and desaturation events.

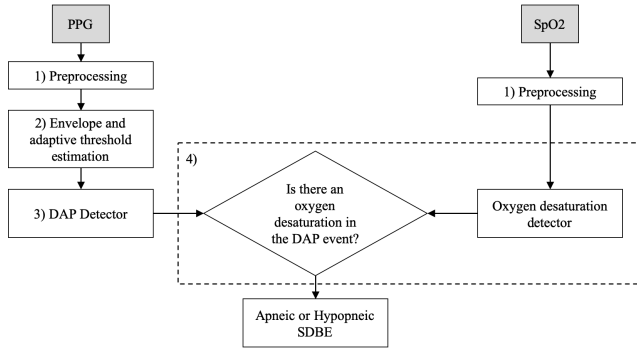This whole detection process is summarized in the flowchart depicted in Fig.1.

Fig. 1. Apneic and hypopneic events detection flowchart.

1) Pre-processing: Both the PPG ($x(n)$) and SpO$_2$ ($x_{SpO_2}(n)$) signals, acquired at 500 Hz, were down-sampled at 100 Hz. After that, the SpO$_2$ samples were rounded to the nearest integer. These two operations aim at simulate signals acquired with low time resolution acquisition devices: this makes it possible to test the usability of the proposed algorithm, to a wider range of acquisition hardware. The PPG signal $x(n)$ was then detrended by removing the time-varying mean value, obtained with a moving average filter: the resulting detrended signal was referred to as $x_d(n)$. Then, the mean PPG cardiac cycle length, T, was estimated using a zero-crossing detector applied to $x_d(n)$. An artifact detector based on Hjorth parameters [35] was implemented and the artifact signal segments were rejected as in [25].

2) Envelope and adaptive threshold estimation: In order to follow the amplitude fluctuations of the PPG, its envelope was analyzed. It was computed from $x_d(n)$, by using the root mean square series method as in [25]:

$$x_e(n) = \sqrt{\frac{1}{N_p} \sum_{k=n-(N_p-1)}^{n} x_d^2(k)} \ , \tag{1}$$

where N_p is the number of samples in two cardiac cycles.

A DAP event was identified when $x_e(n)$ was lower than an adaptive threshold $\zeta(n)$, [25], computed as a percentage (U_p) of the mean of the last L_p non-artifact samples of the envelope:

$$\zeta(n) = \begin{cases} \dfrac{U_p}{100 L_p} \displaystyle\sum_{k=n-(L_p-1)-T_{L_p,n}}^{n} x_e(k) & n, k \in \{n_a\}_\zeta \\ \zeta(n-1) & n \in \{n_c\}_\zeta \end{cases} \tag{2}$$

where $\{n_a\}_\zeta$ is the sample set eligible for the computation of the adaptive threshold and L_p is the number of samples in $\{n_a\}_\zeta$. Then, $\{n_c\}_\zeta$ is the sample set not eligible for the computation and $T_{L_p,n}$ is the number of samples in $\{n_c\}_\zeta$.

The ineligibility condition of a sample ($n \in \{n_a\}_\zeta$), for the adaptive threshold computation, keeping constant $\zeta(n)$, is any of the following:

- the sample belongs to a DAP event, $x_e(k) < \zeta(n-1)$
- the sample belongs to an artifact according to the Hjorth parameters artifact detector
- the sample belongs to an abrupt change in the $x_e(n)$ signal. A change was considered abrupt when

$$|x_e(n) - x_e(n-1)| > \frac{\alpha}{f_s} A_e \tag{3}$$

where A_e is half of the mean oscillation amplitude range of $x_d(n)$ in the recording, f_s is the sample frequency and α was experimentally chosen equal to 5 on accuracy basis.

3) DAP detector: A DAP event was identified when the PPG envelope was lower than the predefined adaptive threshold $\zeta(n)$, for a minimum time duration (Δn_{DAP}), set a priori. Fig.2 shows an example of DAP detection from a PPG signal.
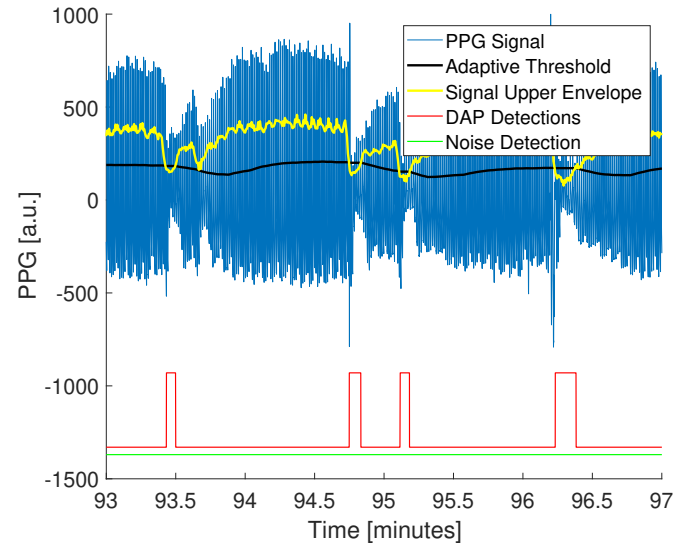

Fig. 2. The DAP detection method applied on a PPG signal. The threshold was computed with $U_p = 70\%$.

4) Oxygen-desaturation detector: At first, an artifact detector was applied to the whole SpO$_2$ signal. A signal segment was annotated as artifact when $x_{SpO_2}(n) < 50\%$: if so, the segment was excluded from the analysis. SpO$_2$ values lower than 50% are usually indication of PPG signal degradation. Afterwards, the oxygen-desaturation detector receives as input the time instants, at which the *i-th* DAP event starts n_o^i (named the onset), and ends n_e^i. Then, an enlarged window around the DAP event was used, by considering $\Delta n_o = 5 * f_s$ samples before the DAP onset and $\Delta n_e = 15 * f_s$ after the DAP ends. The parameters Δn_o and Δn_e were defined a priori, taking in consideration the delay the SpO$_2$ signal has over the PPG [8]. In these new enlarged DAP temporal windows, the oxygen-desaturation detector analyzes if a drop in SpO$_2$, $\Delta SpO_2\%$ takes place. A SDBE was detected, when to the *i-th* DAP event was associated a drop in amplitude between the maximum value of the SpO$_2$ (within the enlarged window) and its minimum, equal or higher than different ΔSpO_2 tested levels (from 1% to 3%):

$$\max[x_{SpO_2}(n)] - \min[x_{SpO_2}(n)] \geq \Delta SpO_2 \quad (4)$$
$$\text{with} \quad n \in [n_o^i - \Delta n_o, n_e^i + \Delta n_e]$$

This computation does not consider a deviation from a SpO$_2$ predefined reference baseline, since the baseline oxygenation level can vary throughout the night. Moreover, the first part of each SpO$_2$ signal segment could not either be used as reference since, in case of consecutive SDBE, the normal oxygenation values can be present at the end of the DAP instead of the beginning.

Fig.3 presents a detection example of obstructive-hypopnea: the three plots show respectively the nasal pressure signal, DAP detections and oxygen-desaturation detections.
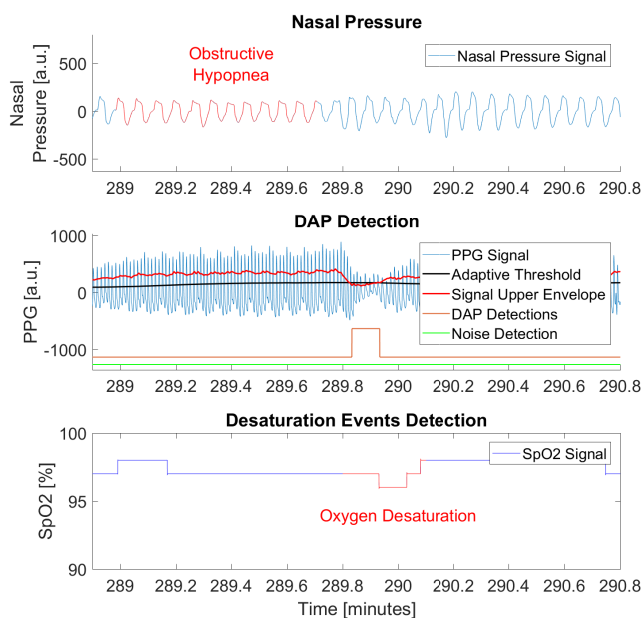

Fig. 3. Apneic and hypopneic events detection. Starting from the top: the first figure shows the nasal pressure signal, where the red portion highlights an annotated obstructive hypopnea event; the second plot visualizes the PPG signal with DAP detections; in the last figure, an oxygen-desaturation was detected in the SpO$_2$ signal.

5) Performance evaluation: For the overnight sleep apnea/hypopnea detection, each recording was divided in segments of one-minute. This segmentation represents the field standard in literature and was used, in the present work, for state-of-the-art comparison. In addition, splitting data into 60s epochs, gave the possibility to take into account the delay between the SDBE occurrence in the airflow signal (annotated manually by experts [34]) and its manifestation in the PPG and SpO$_2$ signals. This delay depends on the subject physiological characteristics and lasts approximately 30s [36].

The DAP detector is event-based and a SDBE was defined apneic/hypopneic if a desaturation event was associated to the corresponding DAP. Each segment was labeled as true positive (TP) if it contained both: an apneic/hypopneic reference annotation and a detected apneic/hypopneic SDBE. Otherwise, if the segment contained none of them, the segment was labeled

as true negative (TN). In case the segments contained only an apneic/hypopneic SDBE, then it was assigned the label false positive (FP). On the contrary, if only the apneic/hypopneic reference annotation was present, it was labeled as false negative (FN). No overlap was considered neither between one minute windows, nor between the annotations and the one minute windows. However, if an apneic/hypopneic SDBE was crossing the border of two one-minute windows, it was considered present in both windows. A receiver operating characteristic (ROC) analysis was performed and the sensitivity (Se), specificity (Sp) and accuracy (Acc) indexes were computed. To get the best classification, it was necessary to maximize Se and Sp, by varying the parameters of the proposed detectors. Finally, in the ROC curve, the combination of parameters that gave the pair (1-Sp, Se) closer to the point (0, 1), was selected.

To get the best parameters for the DAP and oxygen-desaturation detectors, a 3-fold cross-validation (CV) [37] was used. At each fold, 2/3 of the patients present in database were used to train the detectors and 1/3, to test the trained detectors. The 3-fold CV was performed per patient without re-substitution. Since PPG features can be patient specific, it was taken into account that each patient should be either in the detection test set or in the detection training set, in order to ensure a good generalization on new subjects. Data balancing was not performed for SDBE detection because at this stage, no separation of SDBE was considered.

The parameters modified during the Se and Sp maximization procedure were, for the DAP detector, the percentage of the adaptive threshold U_p (from 30% to 80% of the PPG envelope) and DAP minimum duration Δn_{DAP} (from 0 to 3 s). In the oxygen-desaturation detector, the desaturation threshold ΔSpO_2 was varied from 1% to 3%. The detector sensitivity lowered as the desaturation threshold increased: for this reason, results for a desaturation threshold larger than 3% were not taken into account.

C. Sleep disordered breathing events classification

The method used to classify SDBEs involved the implementation of three binary classifications (C-O, CA-CH, OA-OH), instead of one classification with five predictors (CA, CH, OA, OH, MA). Using 5 predictors would result in separating SDBEs that are not completely independent, with less performant classifiers (e.g. CA and CH have both a central origin). SDBEs classification was accomplished in a two steps process. First central versus obstructive SDBEs were classified, as they have independent causes. Then apnea and hypopnea were independently discriminated for central and obstructive SDBEs. From the original database, a subset was selected by extracting all the true positive apneic/hypopneic SDBEs whose duration was falling inside any SDBE annotation. In this way all normal activity was removed and all FP DAP, whose link to the SDBE was not directly visible, were discarded. At first, central SDBEs were separated from the obstructive ones and then, within each category, apneic DAP were separated from hypopneic. Finally, three binary DAP identification problems were defined for classification: central versus obstructive apneic DAP, central apneic DAP versus central hypopneic DAP,

obstructive apneic DAP versus obstructive hypopneic DAP. The proposed classification is presented in Fig.4. Each SDBE was labeled as central or obstructive, apneic or hypopneic, in correspondence to the respiratory event that occurred during the DAP event. According to the AASM 2012 rules, an event labeled as apneic should last at least 10s [34]; for this reason, in the database it never occurred that, during a DAP, different SDBE were present. The classification was performed using right-shifted windows around the DAP events, and not per one minute windows: this because, within one minute, there could be multiple DAP events corresponding to different irregular respiratory events.
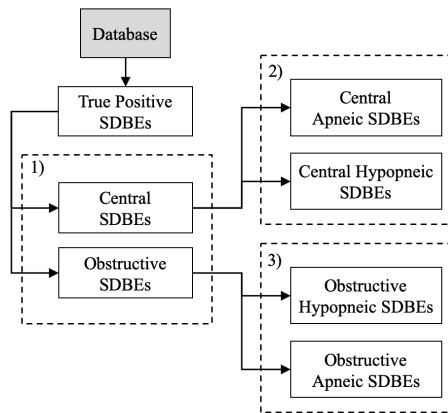

Fig. 4. For classification, three binary DAP identification problems were defined: 1) central versus obstructive apneic SDBEs; 2) central apneic SDBEs versus central hypopneic SDBEs; 3) obstructive apneic SDBEs versus obstructive hypopneic SDBEs.

A set of features was extracted from the PPG and SpO$_2$ signals for each apneic/hypopneic SDBE. A time-frequency analysis, on the PRV signal, was used to quantify the sympathovagal response due to a respiratory event. From the ensemble of the features, only those that minimized the misclassification error (MCE) were selected for the classification. The MCE is defined as the number of misclassified observations divided by the number of observations. Finally, a 10-fold CV classification was performed.

In order to have more generalization power, the apneic/hypopneic SDBE classification was not performed per patient. In fact, in some patients, the number of samples for a certain SDBE sub-type was too low and it would cause a classifier overfitting [37]. The underlying conception was based on the idea of letting the detection task to recognize apneic/hypopneic SDBEs and the classifier to label them.

1) Features extraction and features set: A total of 37 features were extracted from both the PPG and the SpO$_2$ signals. These include PPG features, SpO$_2$ features, pulse rate time domain features and pulse rate frequency domain features; the latter computed by using the smooth pseudo Wigner-Ville distribution (SPWV) [38] and the Lomb periodogram [39], [40].

All features were extracted from an enlarged temporal analysis window right-shifted around the apneic/hypopneic SDBE: the temporal DAP windows were enlarged to 15s before the apneic/hypopneic DAP starts and 50s after the apneic/hypopneic DAP finishes. The duration of the temporal analysis window was at least 65 s, limiting the minimum computed spectral power frequency to 0.03 Hz.

In this temporal DAP window, the PPG signal $x_d(n)$ was considered. Each peak of the PPG signal represents the pulse occurrence time. To detect the pulse, a search rule was implemented by using the MATLAB® Signal Processing Toolbox [41]: it detects peaks with a minimum prominence, fixed by empirical analysis on the database. The peak prominence is an adimensional parameter that measures how much the peak stands out, due to its intrinsic height and its location, relative to other peaks. This strategy was chosen to discriminate among pulses, discarding some pronounced dicrotic notches of the PPG signal that, otherwise, could be detected as single pulses. A further analysis to discover and handle ectopic pulses was implemented following the works in [42], [43], under the assumption that the pulse rate time series is a surrogate of the heart rate time series.

By analyzing the pulse-to-pulse time instants, the same indexes, that usually are exploited in the RR (or NN if normal beats) intervals of the electrocardiogram, were computed. These indexes belong to the time domain features.

For the time-frequency analysis, the inverse interval function $d_{iif}(t_i)$ [44], [45], resampled at 2 Hz by cubic spline interpolation, was used. The resulting time series was then centered by subtracting the mean value. The spectral information of the pulse rate time series was then obtained, by both applying the SPWV [46], [47] and the Lomb periodogram [48]. The spectral analysis was performed by using two mathematical functions, in order to test if the Lomb periodogram would be enough for a low cost computation of the spectral information, or a higher time-frequency resolution was needed. Before applying the SPWV distribution, the Hilbert transform was applied to the $d_{iif}(t_i)$ in order to get its analytic function. The parameters of the quadratic time-frequency distribution were then selected on the basis of recommendations and experimental results reported in the studies [30], [49], [50]. For smoothing in time, a Hamming window of 10.5s was selected, whereas for smoothing in frequency, a Hamming window of 64.5s was used.

Using the spectral analysis, the high frequencies power (P_{HF}) was extracted, as well as the low frequencies (P_{LF}) and the very low frequencies power (P_{VLF}). The sympathovagal information is quantified by the power ratio $R_{LF/HF} = P_{LF}/P_{HF}$: a value of $R_{LF/HF}$ higher than the unity reveals a sympathetic activation. This activation, caused by an obstructive hypopnea is presented in Figure 5. Once the main LF/HF peak was detected in the DAP enlarged window, thanks to the quadratic energy distribution, it was possible to compute the VLF, LF and HF spectral power in correspondence to the time instant of the LF/HF main peak.

The SPWV high resolution in the frequency domain, allows also to determine which frequency was carrying the most energy, per frequency band (e.g. LF, HF) in the PRV signal. For each of the frequencies present in the selected band, the total spectral power was computed and compared to the spectral power evaluated for the adjacent frequencies. Finally the frequency carrying the most power was selected.
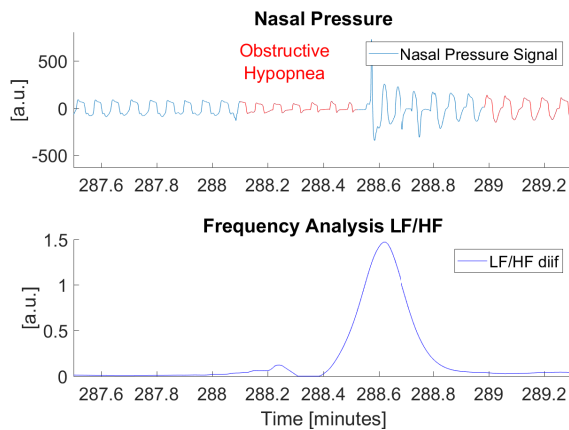
Fig. 5. This figure shows a sympatho-vagal arousal in the LF/HF, linked to an obstructive hypopnea occurrence.

The list of the extracted feature is reported in Table II. For a detailed description, please refer to the original papers.

TABLE II
CLASSIFICATION FEATURES

PPG and SpO_2 features
DAP duration, DAP area (PPG area between the threshold and the envelope, during the DAP event), DAP amplitude (distance between the DAP minimum and the corresponding threshold value) and desaturation loss (ΔSpO_2 desaturation drop in the SpO_2 signal during the DAP event).
Pulse rate, time and frequency domain features [51], [52]
NNmean, NNmode, NNmedian, NNvariance, NNskew, NNkurt, NNiqr, SDNN, RMSSD, pnn50, PRSA AC, PRSA DC, SDANN, SDNNI, SD1, SD2, SD2/SD1; using the Lomb periodogram: vlf (power in very low frequency range), lf (power in low frequency range), hf (power in high frequency range), lfhf ratio and ttlpwr (total spectral power, approximately < 0.4 Hz).
Pulse rate, SPWV frequency domain features [53]
signal main HF frequency (frequency carrying the most energy in HF frequency band), signal main frequency (frequency carrying the most energy among the VLF, LF and HF bands), maximum value of the LF/HF power signal in the enlarged DAP time window, LF/HF main peak prominence, the VLF, LF and HF power corresponding to the LF/HF main peak, the LF/HF peak total energy (sum of the spectral power of VLF, LF, HF), the VLFn (ratio between VLF and LF/HF peak total energy), the LFn (ratio between LF and LF/HF peak total energy), the HFn (ratio between HF and LF/HF peak total energy).

2) Data balancing: For the binary classifiers, it was necessary that both classes were well represented by the same amount of data. The imbalanced data problem was solved by both generating additional data for the least represented class and by removing data from the most represented one. As first step, the ADASYN algorithm proposed in [54] was used. The ADASYN method allows to improve class balance by synthetically creating new samples, via linear interpolation between existing minority class samples. ADASYN is an extension of SMOTE method [55], creating more samples in the vicinity of the boundary between the two classes,

than in the interior of the minority class. After applying this oversampling method, if still a class disparity exists, samples from the more represented class were randomly removed, until the equality was reached.

3) Features selection: The objective of features selection was to reduce the data dimension by finding a small subset of important features which could result in a good classification performance. To help interpretability, two filter and a wrapper method were used [56].

Specifically, at this point, the database was split for a 10% hold-out validation. The 90% of the database was used to discover the best features subset (classification training set) and the remaining 10% of it (classification test set) was used to test the classifier performance with the best subset of features. On the classification training set, the 10-fold CV was then used, for sequential feature selection. The overall procedure for feature selection is shown in Fig.6.
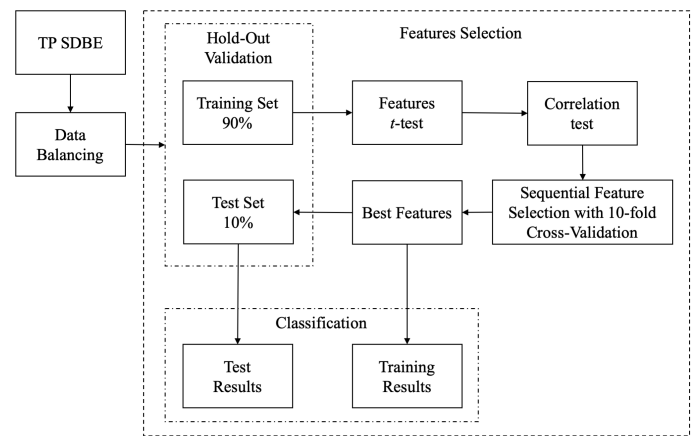

Fig. 6. Feature selection schematics: from the hold-out validation, the classification training set was used to perform the 10-fold CV for sequential features selection, and the classification test set was used to evaluate the performance of the best subgroup of features.

A t-test was computed on all the features of the classification training set, to discard any interaction among them. So the p-value of each feature was compared, as a measure of how effective the feature was at separating groups, and only those having p-values smaller than 0.05 were kept.

Features selected from the list, based on their individual ranking, may also contain redundant information. Following this reasoning, as second filter method, the correlation matrix among all the features was calculated. For each couple of features, if the absolute value of the linear correlation coefficient was equal to or higher than 0.9, only the feature that was the least correlated with others was kept. In case none of the two had other correlations, one of them was randomly discarded.

To further reduce the number of features, a backwards wrapper method was used. Different classifier methods were tested: decision trees, discriminant analysis, logistic regression classifiers, naive Bayes classifiers, support vector machines, nearest neighbor classifiers and ensemble classifiers, with all features included. Finally, the k-Nearest Neighbor (k-NN) with euclidean distance metric was chosen, because it showed the best balanced-class accuracy. To decide the number of features,

the MCE was computed implementing the 10-fold CV in the backwards wrapper method, on the classification training set, as a function of the number of features used to fit the model. As first step, all features were taken into consideration for 10-fold CV, and the MCE was computed; in the second iteration one feature was discarded from the group and the remaining ones were used to compute the MCE in the 10-fold CV. This process was iterated until a minimum was reached. Finally, the selected best subgroup of features for classification, was the one that corresponded to the first local minimum of the MCE.

When the best subgroup of features was obtained, these features were used to compute the MCE on the classification test set, from the hold-out validation.

4) Model selection and evaluation: At this point, the best subgroup of extracted features was obtained and it was possible to find the best binary classifiers for the dataset. In this step, the entire original dataset for classification, with balanced data, was partitioned for a 10-fold CV. All the classifiers present in Table VII were then trained and evaluated. The evaluation of each of these classifiers was performed by computing the Se, Sp and Acc indexes, using MATLAB® Classification Lerner.

III. RESULTS

A. Apnea and hypopnea detection results

ROC curves in Fig.7 show the performance of the proposed algorithm for the detection training set of each of the three CV. The best results were highlighted with a circle. Se, Sp and Acc results for the detection training and test set of each CV fold, using the best detection parameters, are shown in Table III. For all the 3-fold CV, the best parameters resulted to be: $U_p = 70\%$, $\Delta n_{DAP} = 0s$ and $\Delta SpO_2 = 2\%$. The same identical values, obtained for the tested parameters in the validation, show a robust generalization of the method.

TABLE III
APNEA AND HYPOPNEA DETECTION RESULTS

	CV fold	Dataset	Se [%]	Sp [%]	Acc [%]
All SDBE	1	Training	75.1	73.6	74.4
		Test	80.1	72.6	76.4
	2	Training	76.7	72.5	74.6
		Test	77.4	74.6	76.0
	3	Training	78.9	73.7	76.3
		Test	73.1	72.5	72.8
	All data		**76.9**	**73.2**	**75.1**

Event Type	CA	CH	MA	OA	OH
Total Se [%]	86.6	73.3	88.2	86.4	76.2

An overall 75.1% accuracy was reached in detecting all sleep apneas and hypopneas from a PPG and a SpO_2 signal on the UZ Leuven database. In Table III are reported the CV results for apnea/hypopnea detection. In addition, the results obtained by testing the detector on the whole detection database, using the best subset of parameters, are expressed in the row Total of the table. Finally, the last row contains the total Se computed in the CV, by grouping the SDBE in apnea or hypopnea, central or obstructive: it is possible to notice, that hypopnea was the most difficult to detect.

To test the detector performances on those patients with low AHI index, the best detection parameters earlier obtained, were used for apnea/hypopnea detection on patients with AHI≤5. These results are presented in Table IV. In these results, considering for example the central apnea detection, a FP event was considered when an apneic/hypopneic SDBE did not correspond to a central apnea annotation, even if other apnea types were present.

TABLE IV
APNEA/HYPOPNEA DETECTION RESULTS ON PATIENTS WITH AHI≤5

Event Type	Se [%]	Sp [%]	Acc [%]
All	79.9	75.8	76.1
Central Apnea	100	72.0	72.2
Central Hypopnea	83.6	72.4	72.6
Mixed	-	-	-
Obstructive Apnea	88.9	63.6	63.8
Obstructive Hypopnea	76.8	74.2	74.3

B. Classification results of sleep disordered breathing events

Fig.8 shows three plots (one per binary classification), visualizing the MCE versus the number of features used for the 10-fold CV on the classification training set. In the plots is visible that after an initial phase in which the MCE decreases as the number of features increases, follows a phase in which the MCE starts increasing. This can be explained by the fact that an overfitting was reached. To avoid overfitting and generalize the model, a subgroup of features was used. From the overall 37 features used for the classification, the first 7 ones were selected for C versus O classification: the selection was arrested before entering the MCE plateau. Also, for CA - CH classification, 7 features were selected, by having the MCE lower than 0.2. Instead, for the OA - OHA classification, 8 features were selected in order to get closer to the 0.2 value of MCE and also by not entering the following MCE plateau. All the selected features are reported in Table V. Finally, Table VI details the MCE, corresponding to the classifications performed by the best subset of features, computed on the classification test and training set using the 10-fold CV.

TABLE V
BEST CLASSIFICATION FEATURES

Best subset of features for C-O classification
DAP amplitude, DAP duration, NNmedian, lfhf, signal main frequency, signal main HF frequency, LFn.

Best subset of features for CA-CH classification
DAP amplitude, DAP duration, NNmedian, lf, lfhf, ttlpwr, LFn.

Best subset of features for OA-OH classification
DAP amplitude, Desaturation loss (ΔSpO_2), NNmedian, lf, ttlpwr, lfhf, signal main HF frequency, LFn.

After the selection and evaluation performance of the best subgroup of features, this subset was used to perform a 10-fold CV on the whole classification dataset, using different classifiers for each of the three binary classifications. The
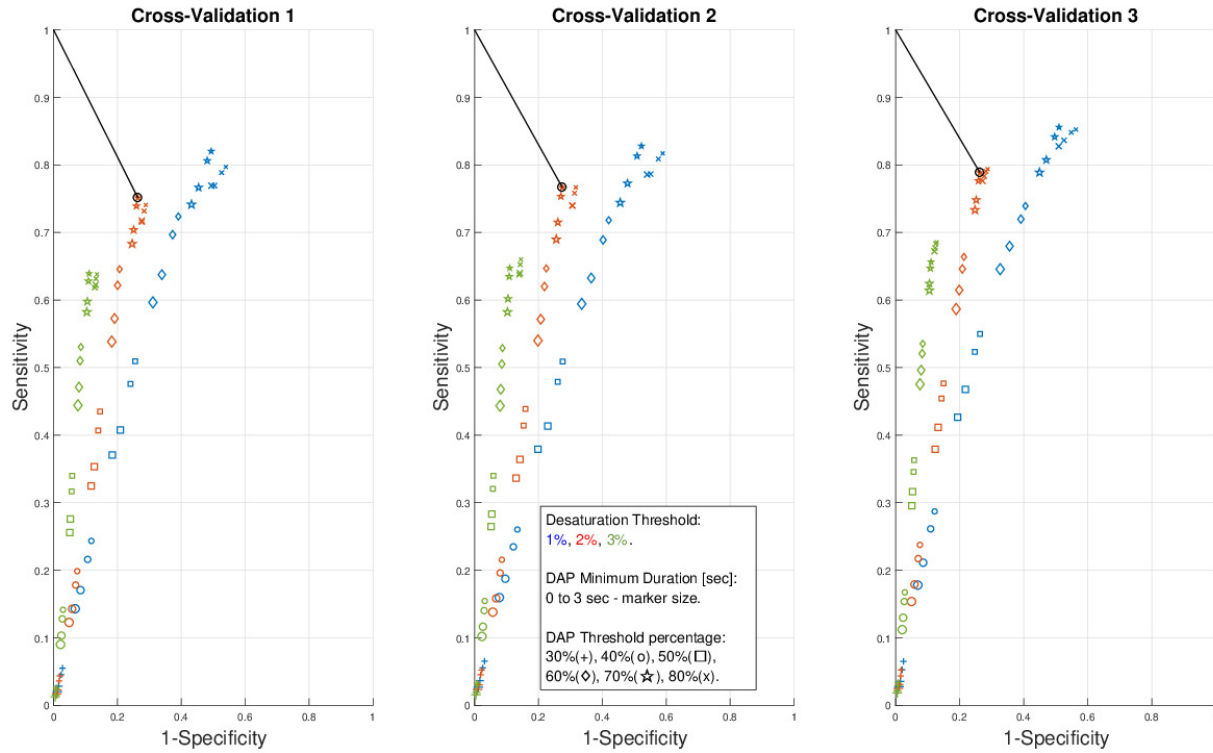
Fig. 7. ROC curves for sleep apneas/hypopneas detection. The curves were built by varying the desaturation threshold from 1% (blue) to 2% (red) until 3% (green). The DAP minimum duration scales from 0 to 3s and is represented by the size of the marker. The DAP threshold percentage, instead, was varied of 30% (+), 40% (o), 50% (square), 60% (diamond), 70% (star), 80% (x).
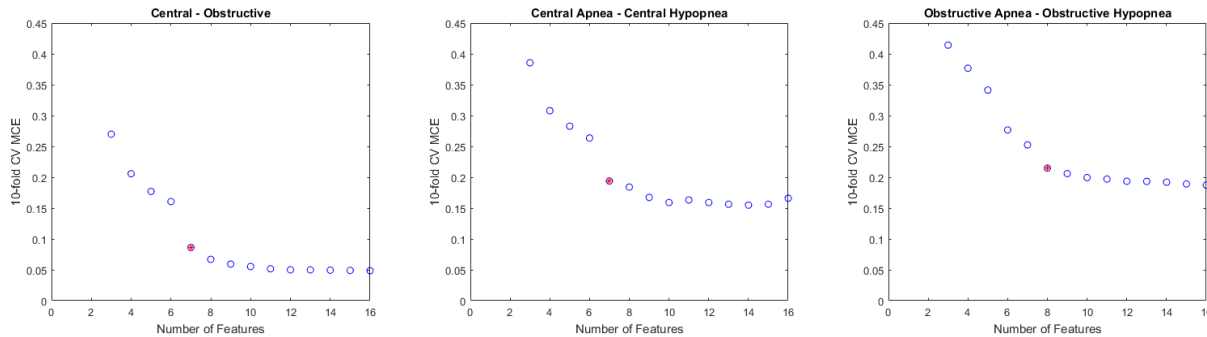

Fig. 8. Sequential Feature Selection with 10-fold Cross-Validation. The MCE was minimized by 7 features for central and obstructive classification, 7 for central apnea and central hypopnea, 8 for obstructive apnea and obstructive hypopnea.

TABLE VI
FEATURES SELECTION MCE

	MCE [%]		
Classification	C-O	CA-CH	OA-OH
10-fold CV Training Set	8.6	19.4	21.5
Test Set	7.4	13.9	19.6

for central and obstructive apnea, 83.7% for central apnea and central hypopnea and 82.7% for obstructive apnea and obstructive hypopnea. The True Positive rate (TPr=Se) and False Positive rate (FPr=1-Sp), as well as the Acc and ROC area under the curve (AUC) are shown in Table VIII. For comparison, on the same database, no other works were performed, in classifying the SDBE types.

performances of each of these classifiers were reported in Table VII. The best classifier for the three classifications, revealed to be the Fine Gaussian Support Vector Machines (SVM) classifier with Gaussian kernel with 0.56 scale. The overall accuracy reached with the classification was 92.6%

In conclusion to the results section, it is possible to assert that the detection and classification methods showed to be consistent for being implemented in a smart device, for sleep monitoring. The detection task recognizes apneic/hypopneic SDBEs and the classifier labels the events.

TABLE VII
CLASSIFIERS ACCURACY PERFORMANCES

Classifier	Classification Accuracy [%]		
	C-O	CA-CH	OA-OH
Decision Trees			
Fine Tree	79.3	74.4	76.4
Medium Tree	74.8	70.6	75.4
Coarse Tree	69.8	71.4	70.2
Discriminant Analysis			
Linear Discriminant	66.2	69.0	67.7
Quadratic Discriminant	67.0	67.6	67.4
Logistic Regression Classifiers			
Logistic Regression	66.4	69.6	67.8
Naive Bayes Classifiers			
Gaussian Naive Bayes	67.4	67.5	65.4
Kernel Naive Bayes	71.6	70.7	69.8
Support Vector Machines			
Linear SVM	67.1	69.5	67.8
Quadratic SVM	78.8	74.7	72.5
Cubic SVM	85.2	77.8	82.9
Fine Gaussian SVM	**92.6**	**83.7**	**82.7**
Medium Gaussian SVM	78.1	75.8	73.7
Coarse Gaussian SVM	69.3	69.3	69.7
K-Nearest Neighbor Classifiers			
Fine KNN	92.0	82.8	79.5
Medium KNN	84.8	74.4	75.5
Coarse KNN	76.0	70.7	72.0
Cosine KNN	84.2	75.4	74.2
Cubic KNN	84.2	73.7	75.0
Weighted KNN	89.6	81.3	78.3
Ensemble Classifiers			
Boosted Trees	77.7	77.4	76.3
Bagged Trees	92.8	79.0	80.4
Subspace Discriminant	66.1	68.5	67.1
Subspace KNN	85.9	68.0	64.7
RUSBoosted Trees	74.8	71.1	75.4

TABLE VIII
RESPIRATORY EVENTS CLASSIFICATION PERFORMANCES FOR FINE
GAUSSIAN SVM

		TPr [%]	FPr [%]	Acc [%]	AUC
C-O	C	95	10	93	0.97
	O	90	5		
CA-CH	CA	86	19	84	0.91
	CH	82	14		
OA-OHA	OA	85	20	83	0.89
	OHA	81	14		

IV. DISCUSSION

A. Apnea and hypopnea detection

The DAP detector was designed to highlight PPG signal shape variations and then those detections were discriminated, by verifying if a desaturation occurred in those time instances. The desaturation detector was then essential, in order to lower the FP detections performed by the DAP. In this case, the SDBE that were left out corresponded to those that were not accompanied by desaturation but that were annotated as apneas/hypopneas because they were characterized by an arousal.

The 3-fold CV, for SDBE detection, was performed per patient, without re-substitution or data balancing. Similar detection results were obtained for all the the 3-fold CV. The DAP signal processing technique for apnea and hypopnea detection improved the interpretability of the recognition algorithm. However, the performance did not outperform the detection method, recently presented in [57]. The proposed detection method used the same DAP detector already applied in [25], [30], [31], for detecting apneas in 26 children with OSAS. In this work, the DAP detector was adapted to detect different kinds of SDBE in adults. Finally, results between this work and the others implementing the DAP detector, are not directly comparable because no apneic events annotations existed in children database and the reference was based on airflow signal and the final clinical diagnosis.

A results comparison can be performed with the work of Lazaro et al. [58], where a similar database collected in UZ Leuven was analyzed. An accuracy of 72.66% in detecting obstructive sleep apneas using PPG signals, in one minute segments of 26 polysomnographic recordings, was reported. This study used a least squares SVM classifier with an RBF kernel to classify apneic and not apneic one minute recordings by extracting features of amplitude and width variability, pulse up-slopes and slope transit time. A comparison with this work, reveled that the DAP detector performances were better than those in [58].

Other SDBE detection algorithms present in literature, that were tested on the same database than the one used for this paper, also including other patients with cardiovascular comorbidities, are those in [32], [57], [58]. The best performances were reached by Deviaene et al. in [57] using the classifier implemented in [58], on PPG and SpO$_2$ signals. In this case, the database consisted of 102 patients and the results showed an accuracy of 83.4%, with 73.7% Se and 86.6% Sp in detecting sleep apnea/hypopnea minutes; while only using SpO$_2$, already an accuracy of 82% was obtained. These results confirmed the importance of using both PPG and SpO$_2$ information to increases the overall accuracy. The Acc reached in the work [32], [57] outperformed the Acc obtained in the present work, indicating that a classification method for apnea/hypopnea detection performs better than the DAP signal processing technique.

The database used in this work, presented also patients with AHI$\leq$5 (8% of the database subjects) and these detections can validate the detector performances in subjects whose sleep health condition was almost normal. The performance on this group revealed to be consistent with the results obtained on the overall database. In particular, the lowest Se was obtained for obstructive hypopnea while the highest (Se=100%) was obtained for central apnea. The lowest Sp was computed for obstructive apnea, indicating a higher number of FP. From these evidences we can conclude that the detector revealed to be robust also for persons with low AHI.

A point of reflection would be the use of the time-frequency analysis of the PRV in the SDBE detection algorithms. For this purpose, a time-frequency algorithm was implemented to verify if, in a temporal windows of 60s right-shifted around the DAP event, the peak of LF/HF signal overpass the unity. If

so, the SDBE was labeled as apneic/hypopneic. Adding this processing step to the presented detector and re-optimizing the different hyper-parameters of the detector, the best results for all SDBE, gave an accuracy of 68.9%. Replacing, instead, the desaturation detector by the time-frequency algorithm, the results were not better: they reached a 61.3% in Acc. If the DAP detector was used alone, the best hyper-parameter value in the ROC curve gave an accuracy of 64.6%. The desaturation detector instead, if not combined with the others, showed a result of 67.1% in Acc and the LF/HF detector a 57.4%.

Finally, it can be concluded that the presented detection tool, exploiting PPG and SpO_2 signals, can be used as first screening for apnea/hypopnea monitoring. This can validate the possibility to implement the method in a non-invasive, home sleep monitoring device.

B. Sleep disordered breathing events classification

For each apneic/hypopneic SDBE, a classification was performed by extracting the features connected to the DAP. This reached 92.6% accuracy in classifying central and obstructive apnea, 83.7% for central apnea and central hypopnea and 82.7% for obstructive apnea and obstructive hypopnea. The Fine Gaussian SVM was chosen, respect to other classifiers, for its better performances in SDBE classification.

The best set of selected features, confirms the importance of PRV time-frequency analysis and highlights the role of the heart rate modulation and ANS activity, during SDBE. It can be said that DAP amplitude, NNmedian, lfhf and LFn were the most relevant features in all classification problems. Additionally, the frequency analysis performed with the SPWV distribution added an important contribution, along with the Lomb periodogram, in discriminating the SDBE classes.

For the classification, all the SDBE events, from all the patients, were collected and then separated for CV. Differently from what performed with SDBE detection, during classification, patients were not separated for validation. This choice was motivated by the fact that the distribution of SDBE, over different patients, was unbalanced. In fact, to CV on patients, it is necessary to balance data at every fold, for training and test data-sets; hence removing data at each fold, would result in a model with less variance, but higher bias.

In the paper of Lazaro et al. [58], no classification in different SDBE types was performed, but a classifier was used for obstructive apnea detection. In the present work, features extracted from the pulse rate time series were used more than those extracted from the PPG morphology as in [58]. This was motivated by the fact that the pulse rate was easier to detect than the morphological features, because the last ones were more affected by noise.

Others literature works worth mentioning are those presented in [59] and [60]. The first one uses ECG signals for apnea classification, while the second one PPG and SpO_2, but not a direct comparison can be performed due to the different dataset used. In [59], Gubbi et al. differentiate central sleep apnea and obstructive sleep apneas using wavelet packet analysis and SVM applied to ECG signals. An advantage of ECG signals is that they are are less affected by noise than the PPG.

The algorithm showed an overall classification Acc of 91.08%, Se of 91.02% and Sp of 91.09%, respectively using wavelet packet analysis on the test set. The study proposed in [60], explains a method for automatically classifying sleep apnea and hypopnea events using PPG and SpO_2 signals, acquired from a pulse oximeter. The PPG was used to classify sleep state, while SpO_2 to classify the sleep-disordered breathing events exploiting a SVM. The classification results showed sensitivity performances and positive predictive values of 74.2% and 87.5% for apnea, 87.5% and 63.4% for hypopnea.

As suggested in [60], it would be interesting to extend the presented SDBE analysis, to the classification of sleep staging from PRV and PPG signals. This might lead to an improvement in the results and enhance the possibility of ambulatory information extraction.

V. CONCLUSION

The aim of this study was to validate the hypothesis to detect and classify sleep disordered breathing events from a PPG and a SpO_2 signal, both acquired at the level of the index finger. The detection and classification results show this method to be suitable for a first sleep screening. Because an oximetry device implements two PPG sensors for computing the oxygen desaturation, it would be enough to both acquire PPG and SpO_2 signals. In this context, developing a new device that implements this sensor, would be of great interest for a home sleep monitoring system. However, the present study showed that it would not replace clinical devices (like those based on the nasal pressure signal), due to the noisy nature of the PPG signal. Nevertheless, the results suggest that such approach can be used as an initial report to better select patients at home, before clinical sleep center observation. This perspective, lights the way for an embedded system for sleep apnea/hypopnea home monitoring, in a Medical of Things (MoT) device.

REFERENCES

[1] A. A. of Sleep Medicine Task Force *et al.*, "Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research," *Sleep*, vol. 22, pp. 667–689, 1999.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2020.3028041, IEEE Transactions on Biomedical Engineering

11

[2] U. D. of Health, H. Services *et al.*, "What are the signs and symptoms of sleep apnea? nhlbi: Health information for the public," 2012.

[3] T. Young, P. E. Peppard, and D. J. Gottlieb, "Epidemiology of obstructive sleep apnea," *American Journal of Respiratory and Critical Care Medicine*, vol. 165, no. 9, pp. 1217–1239, may 2002.

[4] F. J. Nieto, "Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study," *JAMA*, vol. 283, no. 14, p. 1829, apr 2000.

[5] A. V. Benjafield *et al.*, "Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis," *The Lancet Respiratory Medicine*, vol. 7, no. 8, pp. 687–698, 2019.

[6] W. W. Flemons *et al.*, "Home diagnosis of sleep apnea: A systematic review of the literature," *Chest*, vol. 124, no. 4, pp. 1543–1579, oct 2003.

[7] K. H. Shelley, "Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate," pp. S31–S36, 2007.

[8] K. Matthes, "Studies on the oxygen saturation of arterial human blood," *Naunyn-Schmiedeberg's Archives of Pharmacology*, vol. 179, no. 6, pp. 698–711, 1935.

[9] V. K. Somers *et al.*, "Sympathetic neural mechanisms in obstructive sleep apnea." *Journal of Clinical Investigation*, vol. 96, no. 4, pp. 1897–1904, oct 1995.

[10] V. A. Imadojemu *et al.*, "Obstructive apnea during sleep is associated with peripheral vasoconstriction," *American journal of respiratory and critical care medicine*, vol. 165, no. 1, pp. 61–66, 2002.

[11] Y. Mendelson, "Pulse oximetry: theory and applications for noninvasive monitoring." *Clinical chemistry*, vol. 38, no. 9, pp. 1601–1607, 1992.

[12] M. Nitzan *et al.*, "The variability of the photoplethysmographic signal - a potential method for the evaluation of the autonomic nervous system," *Physiological Measurement*, vol. 19, no. 1, pp. 93–102, feb 1998.

[13] R. P. Schnall *et al.*, "Periodic, profound peripheral vasoconstriction—a new marker of obstructive sleep apnea," *Sleep*, oct 1999.

[14] R. Smith *et al.*, "Autonomic nervous system and sleep apnea syndromes," *Revue des maladies respiratoires*, vol. 17, pp. S51–67, 2000.

[15] J. C. Hardy *et al.*, "Sympathetic and blood pressure responses to voluntary apnea are augmented by hypoxemia," *Journal of Applied Physiology*, vol. 77, no. 5, pp. 2360–2365, nov 1994.

[16] U. Leuenberger *et al.*, "Surges of muscle sympathetic nerve activity during obstructive apnea are linked to hypoxemia," *Journal of Applied Physiology*, vol. 79, no. 2, pp. 581–588, aug 1995.

[17] H. Schneider *et al.*, "Neural and local effects of hypoxia on cardiovascular responses to obstructive apnea," *Journal of Applied Physiology*, vol. 88, no. 3, pp. 1093–1102, mar 2000.

[18] U. A. Leuenberger *et al.*, "Hypoxia augments apnea-induced peripheral vasoconstriction in humans," *Journal of Applied Physiology*, vol. 90, no. 4, pp. 1516–1522, 2001.

[19] A. Anand *et al.*, "Peripheral vascular resistance increases after termination of obstructive apneas," *Journal of Applied Physiology*, vol. 91, no. 5, pp. 2359–2365, nov 2001.

[20] A. Malliani, "The pattern of sympathovagal balance explored in the frequency domain," *Physiology*, vol. 14, no. 3, pp. 111–117, jun 1999.

[21] E. Gil *et al.*, "Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions," *Physiological Measurement*, vol. 31, no. 9, pp. 1271–1290, aug 2010.

[22] A. Bar *et al.*, "Evaluation of a portable device based on peripheral arterial tone for unattended home sleep studiesa," *Chest*, vol. 123, no. 3, pp. 695–703, mar 2003.

[23] C. P. O. Donnell *et al.*, "The effect of upper airway obstruction and arousal on peripheral arterial tonometry in obstructive sleep apnea," *American Journal of Respiratory and Critical Care Medicine*, vol. 166, no. 7, pp. 965–971, oct 2002.

[24] M. Nitzan *et al.*, "Assessment of changes in arterial compliance by photoplethysmography," in *21st IEEE Convention of the Electrical and Electronic Engineers in Israel. Proceedings (Cat. No.00EX377)*, IEEE. IEEE, 2000, pp. 351–354.

[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," *Biomedical Signal Processing and Control*, vol. 3, no. 3, pp. 267–277, jul 2008.

[26] T. Penzel *et al.*, "Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings," *Medical and Biological Engineering and Computing*, vol. 40, no. 4, pp. 402–407, jul 2002.

[27] K. Dingli *et al.*, "Spectral oscillations of RR intervals in sleep apnoea/hypopnoea syndrome patients," *European Respiratory Journal*, vol. 22, no. 6, pp. 943–950, dec 2003.

[28] F. Roche *et al.*, "Screening of obstructive sleep apnea syndrome by heart rate variability analysis," *Circulation*, vol. 100, no. 13, pp. 1411–1415, sep 1999.

[29] ——, "Predicting sleep apnoea syndrome from heart period: a time-frequency wavelet analysis," *European Respiratory Journal*, vol. 22, no. 6, pp. 937–942, dec 2003.

[30] E. Gil *et al.*, "Discrimination of sleep-apnea-related decreases in the amplitude fluctuations of PPG signal in children by HRV analysis," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1005–1014, apr 2009.

[31] J. Lázaro *et al.*, "Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children," *IEEE journal of biomedical and health informatics*, vol. 18, no. 1, pp. 240–246, 2013.

[32] M. Deviaene *et al.*, "Automatic screening of sleep apnea patients based on the spo 2 signal," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 607–617, 2018.

[33] S. S. Mostafa *et al.*, "Spo2 based sleep apnea detection using deep learning," in *2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES)*. IEEE, 2017, pp. 000 091–000 096.

[34] 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," *Journal of Clinical Sleep Medicine*, oct 2012.

[35] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography and Clinical Neurophysiology*, vol. 29, no. 3, pp. 306–310, sep 1970.

[36] V. Moret-Bonillo *et al.*, "Intelligent approach for analysis of respiratory signals and oxygen saturation in the sleep apnea/hypopnea syndrome," *The open medical informatics journal*, vol. 8, p. 1, 2014.

[37] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *Journal of Machine Learning Research*, vol. 11, no. Jul, pp. 2079–2107, 2010.

[38] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, 1989.

[39] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and Space Science*, vol. 39, no. 2, pp. 447–462, feb 1976.

[40] J. D. Scargle, "Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, pp. 835–853, 1982.

[41] U. The MathWorks, MA, "Signal processing toolbox™," 2019.

[42] J. Mateo and P. Laguna, "Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 334–343, mar 2003.

[43] K. Solem, P. Laguna, and L. Sornmo, "An efficient method for handling ectopic beats using the heart timing signal," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 1, pp. 13–20, jan 2006.

[44] P. Laguna, G. B. Moody, and R. G. Mark, "Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 6, pp. 698–715, 1998.

[45] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press, 2005, vol. 8.

[46] W. Martin and P. Flandrin, "Wigner-ville spectral analysis of nonstationary processes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 6, pp. 1461–1470, dec 1985.

[47] M. O. Mendez *et al.*, "On arousal from sleep: time-frequency analysis," *Medical & Biological Engineering & Computing*, vol. 46, no. 4, pp. 341–351, feb 2008.

[48] W. H. Press *et al.*, *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[49] S. Pola *et al.*, "Estimation of the power spectral density in nonstationary cardiovascular time series: assessing the role of the time-frequency representations (TFR)," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 1, p. 46, jan 1996.

[50] S. Cerutti, L. T. Mainardi, and A. M. Bianchi, "Time-frequency and time-varying analysis for assessing the dynamic responses of cardiovascular control," *Critical Reviews in Biomedical Engineering*, vol. 30, no. 1-3, pp. 175–217, 2002.

[51] A. N. Vest *et al.*, "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiological Measurement*, vol. 39, no. 10, p. 105004, oct 2018.

[52] A. L. Goldberger *et al.*, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[53] F. Auger *et al.*, "Time-frequency toolbox," *CNRS France-Rice University*, vol. 46, 1996.

[54] H. He *et al.*, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. IEEE, 2008, pp. 1322–1328.

[55] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, sep 2009.

[56] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, dec 1997.

[57] M. Deviaene *et al.*, "Sleep apnea detection using pulse photoplethysmography," in *2018 Computing in Cardiology Conference (CinC)*, vol. 45. IEEE, 2018, pp. 1–4.

[58] J. Lázaro *et al.*, "Pulse photoplethysmography derived respiration for obstructive sleep apnea detection," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[59] J. Gubbi, A. Khandoker, and M. Palaniswami, "Classification of sleep apnea types using wavelet packet analysis of short-term ECG signals," *Journal of Clinical Monitoring and Computing*, vol. 26, no. 1, pp. 1–11, dec 2011.

[60] J.-U. Park *et al.*, "Automatic classification of apnea/hypopnea events through sleep/wake states and severity of SDB from a pulse oximeter," *Physiological Measurement*, vol. 36, no. 9, pp. 2009–2025, aug 2015.