A Deterministic PPG-Based Algorithm for Obstructive Sleep Apnea Detection

A Deterministic PPG-based Algorithm for Obstructive Sleep Apnea Detection

Michele Guagnano[1,2], Sara Groppo[1,2], Andrea Romigi[3] and Massimo Violante[1]
[1]*Dept. of Control and Computer Engineering*, *Politecnico di Torino*, Turin, Italy
[2]*Sleep Advice Technologies S.R.L., Turin, Italy*
[3]*Saint Camillus International University of Health and Medical Sciences*,
Sleep Medicine Center IRCCS Neuromed Istituto Neurologico Mediterraneo Pozzilli IS, Rome, Italy
{michele.guagnano, sara.groppo, massimo.violante}@polito.it
andrea.romigi@unicamillus.org

***Abstract*—Obstructive sleep apnea (OSA) remains substantially underdiagnosed, largely due to the logistical constraints of polysomnography (PSG), the current clinical reference standard. Although consumer smartwatches provide a scalable alternative for home-based monitoring, many existing approaches rely on deep learning models that are difficult to interpret and may be challenging to deploy in resource-constrained settings. In this study, we propose a deterministic white-box pipeline for OSA screening based on heart rate (HR) and peripheral oxygen saturation (SpO_2) signals. The method is explicitly designed as a high-sensitivity screening tool and is grounded in clinically relevant physiological patterns, including cyclic variation of heart rate (CVHR) and dynamic SpO_2 desaturation behavior, together with an autonomic thresholding mechanism. Evaluation on 29 subjects showed a binary classification accuracy of 89.7%, with 100.0% sensitivity and 78.6% specificity. All OSA cases were correctly identified in this cohort, supporting the use of the proposed method as a conservative first-line screening tool. Agreement with the reference apnea–hypopnea index (AHI) classification was strong (ICC = 0.96), and severity misclassifications were limited to adjacent categories: mild overestimations in healthy subjects and minor underestimations in lower-moderate cases. These results show that a physiologically grounded, computationally efficient approach can reliably identify at-risk individuals and prioritize them for clinical PSG evaluation.**

***Index Terms*—Cyclic Variation of Heart Rate (CVHR), Deterministic Algorithm, Obstructive Sleep Apnea (OSA), OSA Screening, Photoplethysmography (PPG), Wearable Sensors**

I. Introduction

Obstructive sleep apnea (OSA) is a highly prevalent and often underdiagnosed sleep-related breathing disorder characterized by recurrent upper-airway collapse during sleep, leading to intermittent hypoxemia, sleep fragmentation, and increased cardiovascular and neurocognitive risk [1], [2]. Despite its clinical relevance, OSA is still underdiagnosed in routine practice because overnight polysomnography (PSG), the current gold standard, is resource-intensive, costly, and impractical for large-scale screening or longitudinal home monitoring [3].

In recent years, wearable technologies have emerged as a promising alternative for sleep apnea screening, enabling unobtrusive and scalable acquisition of physiological signals in real-world setting [4]. Among these signals, heart rate (HR) and peripheral oxygen saturation (SpO_2) are particularly relevant because OSA episodes produce cyclic desaturation patterns and characteristic autonomic responses. Specifically, these include cyclic variation of heart rate (CVHR), a bradycardia-tachycardia sequence, and intermittent oxygen desaturations contributing to the apnea-hypopnea index (AHI), calculated as apnea/hypopnea events per hour, the clinical standard for OSA severity [5], [6]. Smartwatches are especially attractive in this context because they combine consumer familiarity, continuous overnight monitoring, and increasing access to embedded photoplethysmography (PPG) sensing capabilities [7]. However, their clinical value lies not in replacing PSG, but in providing an accessible, high-sensitivity screening tool that may help prioritize patients for formal sleep evaluation [8]–[10].

The state of the art suggests that AI-based approaches can achieve encouraging results for sleep apnea detection [11]. A recent systematic review of 38 wearable AI studies reported pooled performance of 86.9% accuracy, 93.8% sensitivity, and 75.2% specificity across all included investigations [12]. However, many recent systems rely on deep-learning architectures such as CNNs and transfer learning [13], which often improve performance at the expense of transparency and portability across datasets [14], [15]. While these models demonstrate impressive capabilities, their clinical deployment remains constrained by limited interpretability, variable generalizability, and computational demands for continuous on-device operation [16], [17]. Consequently, there is still value in developing simpler, physiologically grounded methods that offer both acceptable accuracy and clearer clinical rationale. More specifically, smartwatch-based studies using SpO_2 and HR have shown that consumer wearables can reach high discrimination performance. For example, a recent AI-enabled smartwatch algorithm reported sensitivity, specificity, and overall accuracy of 92% for detecting moderate-to-severe OSA, while also highlighting systematic underestimation of the apnea-hypopnea index (AHI) in milder disease [18]. These findings underscore persistent challenges in generalization and calibration. Notably, many studies optimize for overall accuracy by closely following PSG scoring rules rather than prioritizing a fail-safe screening mechanism that minimizes false negatives.

Taken together, while smartwatch-based OSA detection shows increasing clinical promise, two key challenges remain: the limited interpretability of deep learning models and the difficulty of achieving PSG-equivalent accuracy across all disease severities. To address these challenges, this paper proposes a physiologically grounded, multi-stage algorithmic pipeline applicable to data acquired from wearable devices: raw HR and SpO_2 signals. In this study, the data were collected using a polysomnographic system, which was employed to obtain the reference (ground truth) labels. These signals were downsampled to match the temporal resolution of typical wearable sensors and then processed with the proposed algorithm. The primary contributions are:

- A deterministic, white-box pipeline that directly translates established OSA physiology, SpO_2 desaturation morphology and cyclic variation of heart rate (CVHR patterns), into transparent mathematical operators for full clinical interpretability.
- A high-sensitivity screening system (100.0% sensitivity, 78.6% specificity) explicitly designed as a first-line triage tool rather than a PSG replacement, prioritizing the elimination of false negatives to ensure all pathological cases receive clinical referral.
- An ultra-low computational footprint achieved through 1D morphological filters and discrete logic, enabling efficient on-device execution within resource-constrained wearables.

II. Methodology

This retrospective study included 29 subjects (13 females; mean age 39.0 ± 16.8 years), comprising healthy individuals and patients with OSA across different severity levels. Subjects were classified by AHI: healthy (AHI < 5), mild ($5 \leq$ AHI $<$ 15), moderate ($15 \leq$ AHI < 30), and severe (AHI ≥ 30). Borderline cases were included to enhance specificity assessment and dataset heterogeneity. Recordings were performed through a 32-channel polysomonography system (NOX Medical). Electrodes were positioned according to the 10–20 International System. Montage consisted of two oculographic channels, three electromyographic channels (mental and anterior tibialis muscles), eight referential EEG channels (F3, C3, F4, C4, O1, O2, A1, A2), monitoring of oro-nasal airflow, respiratory effort (thoracoabdominal belts), EKG, pulse oximetry (Nonin) and snoring. Sleep analysis was performed according to AASM criteria [6], [19]. Periodic Limb Movement Index (PLMI) and apnoea–hypopnoea index (AHI) [6] were scored under standard criteria (apnoea drop $\geq$ 90% for $\geq$ 10 s; hypopnoea drop by $\geq$ 30%, for $\geq$ 10 s in association with either $\geq$ 3% arterial oxygen desaturation or an arousal), AHI was considered an exclusion criterion if $>$ 15/h, in HCs. The scoring of the macrostructure was visually conducted by AR, a neurologist board certified in neurology, neurophysiology and sleep medicine [20]. Subjects with comorbidities known to affect heart rate or autonomic regulation, or receiving pharmacological treatment with potential chronotropic effects (including beta-blockers, antiarrhythmics, and sedatives), were excluded from the study. This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. All participants provided written informed consent prior to data collection.

The proposed algorithm (Fig. 1) implements a deterministic, four-stage pipeline operating on 1 Hz sampled SpO_2 and HR signals to automatically detect and validate apneic events [21].

A. Stage 0: Signal Pre-processing

Wearable PPG sensors are prone to motion artifacts and sensor dropouts characteristic of wrist-worn devices [22], [23]. A single-sample artifact rejection filter was therefore applied to raw SpO_2 signals, rejecting values with physiologically implausible changes while preserving underlying physiological morphology.

B. Stage 1: Desaturations Extraction

Candidate desaturation events were identified by computing the difference between the raw SpO_2 signal and a dynamically estimated morphological baseline. A continuous boolean mask ($Events$) marked temporal regions where SpO_2 fell $\geq 3\%$ below baseline. Edge detection on this mask yielded discrete onset ($starts_i$) and offset ($ends_i$) indices for each candidate event. The initial onset index marked only the 3% threshold crossing. Therefore, each candidate underwent backward extension to the preceding sample where $SpO_2 \geq baseline$, capturing the true physiological onset rather than the severity threshold crossing. Extended candidates were then validated against two criteria:

- AASM Compliance: Duration ≥ 10 seconds per American Academy of Sleep Medicine guidelines.
- Maximum drop rate: The maximum instantaneous drop rate (the absolute discrete derivative within the segment) must remain below a non-physiological upper limit, discarding abrupt signal collapses caused by sensor displacement.

Validated events were aggregated into a binary temporal vector for preliminary Oxygen Desaturation Index (ODI) computation.

C. Stage 2: CVHR Threshold

Static thresholding often fails to account for inter-subject baseline heart rate variability—such as physiological respiratory sinus arrhythmia—or the attenuated autonomic responses observed in severe OSA. To mitigate this, the algorithm employs an ODI-dependent thresholding mechanism. For non-severe cases, the algorithm computes the basal HR standard deviation (HR_{std}) during periods of stable SpO_2. The CVHR detection threshold (th_{CVHR}) is then dynamically clamped between empirically derived limits as a function of HR_{std}, ensuring that physiological arrhythmias in healthy subjects do not trigger false-positive detections. Conversely, for severe cases, a fixed, highly sensitive threshold is applied to detect the potentially attenuated CVHR signatures characteristic of advanced OSA.

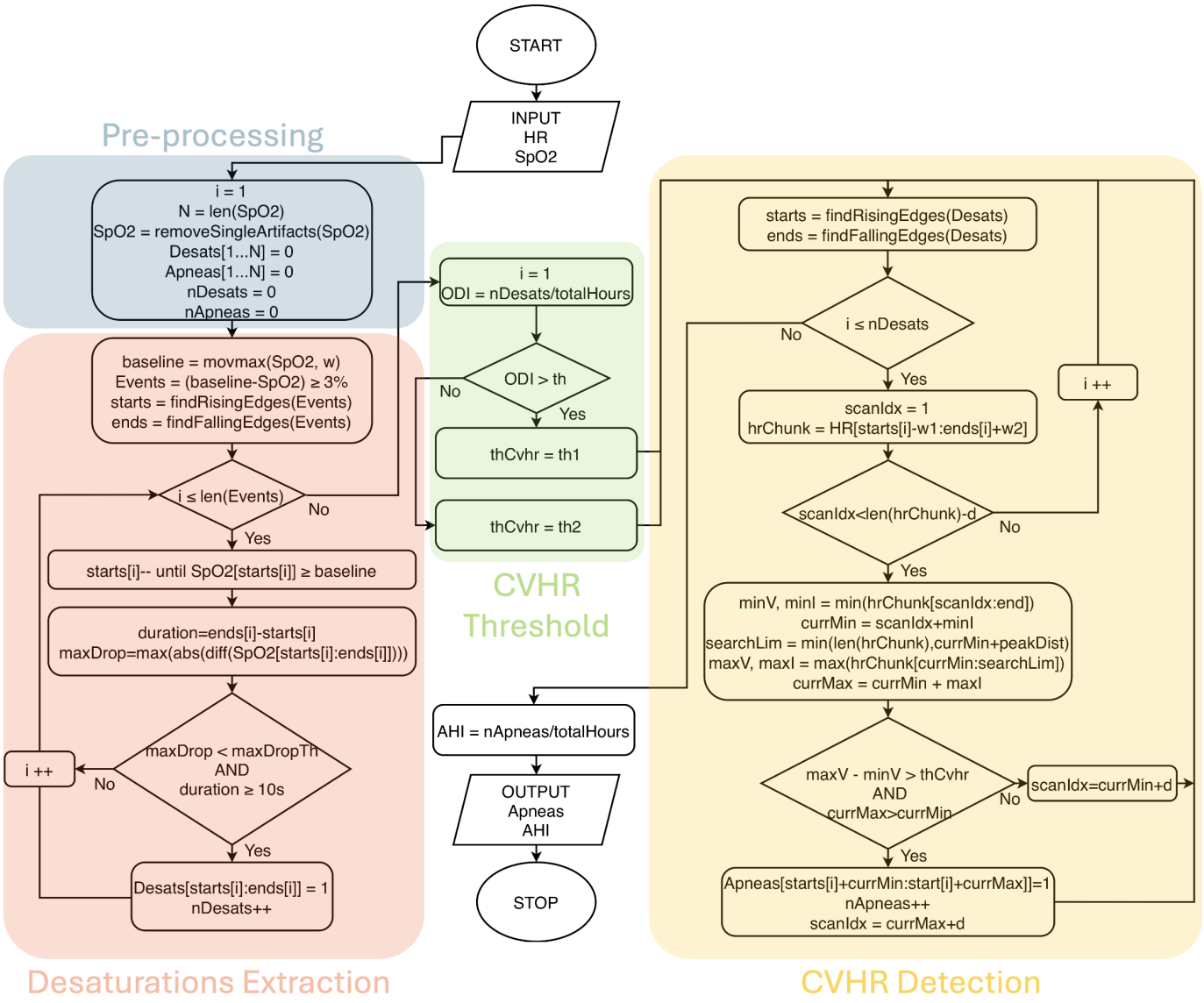

Fig. 1. Flowchart of the apneas detection algorithm. The architecture is divided into four sequential stages: Pre-processing, Desaturations Extraction, CVHR Threshold, and HR Pattern Matching for CVHR validation. The specific parameters w, $maxDropTh$, th, $th1$, $th2$, $w1$, $w2$, and d were empirically derived via grid-search optimization on an independent, proprietary training dataset to maximize baseline sensitivity

D. Stage 3: CVHR Detection (HR Pattern Matching)

In the final stage, each validated SpO_2 event initiates a localized, bounded search within the HR signal to identify the CVHR pattern—the autonomic signature of an apnea, characterized by a localized bradycardic phase followed by abrupt compensatory tachycardia [24]. A sliding window scans the HR segment surrounding the SpO_2 event onset. The algorithm identifies the local minimum (bradycardia) and performs a forward search, constrained by a predefined peak distance ($peakDist$), to locate the subsequent local maximum (tachycardia). If the peak-to-trough amplitude exceeds the subject-specific threshold (th_{CVHR}), the event is classified as an apnea. To optimize computational load and prevent redundant event counting, two temporal heuristics are applied to the sliding scanner: Upon successful validation of an apnea, the scanner advances by a d-second buffer past the tachycardic peak, allowing the heart rate to normalize and preventing secondary false detections on the descent slope. If no qualifying tachycardia is detected following a local minimum, the scanner advances by d seconds to skip prolonged non-apneic bradycardia periods and avoid redundant analysis of the same cardiac segment. Finally, the total count of validated multimodal events is aggregated to output the final AHI.

III. Results

The performance of the proposed pipeline was evaluated on the full dataset of 29 subjects and compared to the PSG ground truth. Clinical validation was performed across two diagnostic dimensions: first, the algorithm's capability to perform binary OSA screening (healthy vs. affected) using the clinical cutoff

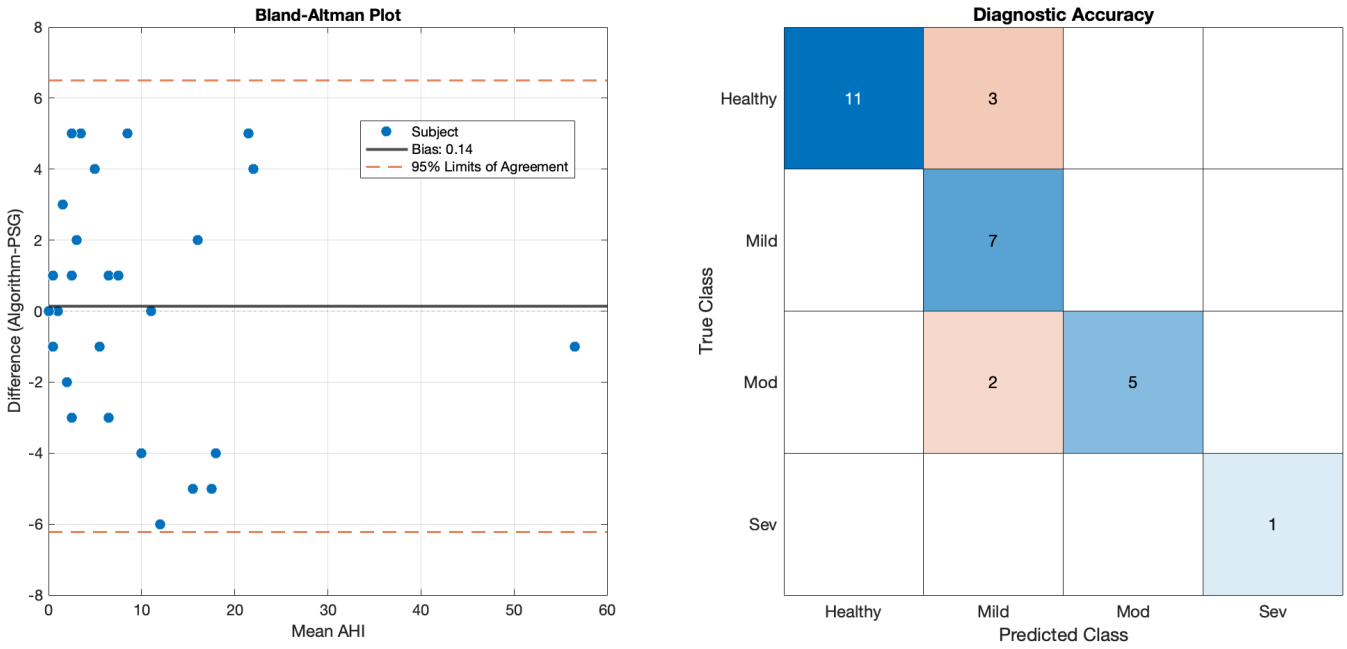

Fig. 2. Multiclass classification results

of AHI ≥ 5; and second, its accuracy in classifying disease severity according to standard AHI thresholds: healthy, mild, moderate, and severe (Fig. 2).

A. Binary Screening Performance

The system's diagnostic capacity to act as an initial screening tool was evaluated by its ability to correctly classify subjects into Healthy (AHI < 5) or OSA (AHI ≥ 5) categories based on the algorithm-derived index. Table I summarizes the patient-level classification metrics across the 29 subjects. The system achieved an overall diagnostic accuracy of 89.7%.

TABLE I
BINARY SCREENING CLASSIFICATION METRICS.

Metric	Value (%)
Accuracy	89.7
Sensitivity	100.0
Specificity	78.6
F1-Score	90.9

The most significant outcome was a sensitivity of 100.0% (0 false negatives), successfully identifying all OSA cases and ensuring that no patients requiring medical attention were overlooked. The specificity was 78.6% (3 false positives), indicating a conservative bias that occasionally misclassifies healthy participants as having mild OSA. Finally, the robust F1-Score of 90.9% confirms the reliability of the proposed pipeline as a preliminary wearable screening instrument.

B. AHI Estimation Performance

Beyond individual event detection, the clinical viability of the wearable system depends on its ability to stratify patient severity based on the AHI.

The correlation between the algorithm-derived AHI and the physician-scored PSG ground truth yielded an Intraclass Correlation Coefficient (ICC) of 0.961 (95% CI: 0.919 - 0.982), indicating a strong absolute agreement. The Mean Absolute Error (MAE) across the entire cohort was 2.55 events/hour.

To visually assess the diagnostic concordance and identify potential systemic biases, a Bland-Altman plot was performed (Fig. 2). The analysis reveals a mean bias of +0.14 events/hour, with the limits of agreement (LoA) ranging from -6.2 to +6.5. The scatter distribution indicates that the algorithm maintains robust linearity across the severity spectrum, with only minor underestimations observed in the extreme upper tail of severe OSA cases (AHI ≥ 30), likely due to the physiological fusion of rapid, consecutive apneic clusters.

IV. CONCLUSIONS

This study presented a novel, multi-stage algorithmic pipeline designed to transform consumer-grade smartwatches into reliable screening instruments for OSA. Diverging from the prevailing trend of computationally opaque and resource-intensive deep learning architectures, we propose a deterministic, white-box methodology. By anchoring the detection logic directly to the clinical physiology of CVHR and dynamic oxygen desaturation nadirs, the system ensures both high interpretability and suitability for continuous, low-power on-device execution.

The clinical validation underscores the system's exceptional viability as a fail-safe screening tool. With a patient-level sensitivity of 100.0%, the algorithm successfully eliminates false negatives, ensuring that all subjects with mild-to-severe OSA are correctly flagged for mandatory polysomnographic

evaluation. Although the deliberately conservative diagnostic bias yielded a moderate specificity of 78.6% and minor AHI underestimations during extreme, rapidly fusing apneic clusters, these outcomes represent a clinically sound trade-off. In the context of preliminary screeining, prioritizing patient safety and maximizing the true positive rate fundamentally outweighs the logistical cost of precautionary false-positive referrals.

Ultimately, this physiologically grounded approach bridges the gap between consumer wearable technology and formal clinical diagnostics. It offers a transparent, scalable, and highly accessible solution to identify at-risk populations and alleviate the escalating logistical burden on traditional sleep laboratories.

The proposed algorithm was designed to be applicable to data from consumer wearable devices. In this study, polysomnographic signals were used for the reference labels, downsampled to wearable-like sampling rates, and processed with the proposed method. Future work will collect concurrent polysomnographic and wearable data to perform a double-blind validation.

V. Limitations

Some limitations of this pivotal study should be acknowledged. First, the relatively small cohort of 29 subjects limits the statistical power of the findings and warrants caution in generalizing the results to broader clinical populations. Validation on larger, multi-center datasets is needed to confirm the robustness of the reported performance metrics, particularly the 100% sensitivity estimate, whose confidence interval remains wide at this sample size. Second, the study did not perform subgroup analyses stratified by sex or age. Given that OSA pathophysiology, autonomic responses, and AHI severity differ between males and females — with women more frequently presenting atypical or REM-predominant disease [25] — and that aging is associated with progressive attenuation of autonomic reactivity, performance may vary across these subgroups in ways that the current analysis cannot capture. Future studies should prospectively enroll larger, demographically balanced cohorts to assess potential differential performance across sex and age strata.

VI. Acknowledgment

The authors would like to sincerely thank Martina Violante for her insightful support during the data collection phase.

References

[1] T. Penzel, J. W. Kantelhardt, C.-C. Lo, K. Voigt, and C. Vogelmeier, "Dynamics of heart rate and sleep stages in normals and patients with sleep apnea," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 8, pp. 996–1004, Aug 2003.

[2] R. B. Berry, A. R. Abreu, V. Krishnan, S. F. Quan, P. J. Strollo Jr, and R. K. Malhotra, "A transition to the american academy of sleep medicine–recommended hypopnea definition in adults: initiatives of the hypopnea scoring rule task force," *Journal of clinical sleep medicine*, vol. 18, no. 5, pp. 1419–1425, 2022.

[3] T. Sofer, "Overcoming the underdiagnosis of obstructive sleep apnea to empower genetic association analyses," *Sleep*, vol. 46, no. 3, p. zsac312, 2023.

[4] S. M. A. Iqbal, I. Mahgoub, E. Du, M. A. Leavitt, and W. Asghar, "Advances in healthcare wearable devices," *npj Flex. Electron.*, vol. 5, p. 9, 2021.

[5] T. Penzel, J. W. Kantelhardt, R. P. Bartsch, M. Riedl, J. F. Kraemer, N. Wessel, C. Garcia, M. Glos, I. Fietze, and C. Schöbel, "Modulations of heart rate, ecg, and cardio-respiratory coupling observed in polysomnography," *Frontiers in physiology*, vol. 7, p. 460, 2016.

[6] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, S. Redline, K. P. Strohl, S. L. Davidson Ward, and M. M. Tangredi, "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.

[7] A. J. Goodings, K. P. Fadahunsi, D. M. Tarn, J. Lutomski, A. Chhor, F. Shiely, P. Henn, and J. O'Donoghue, "Determinants of continuous smartwatch use and data-sharing preferences with physicians, public health authorities, and private companies: cross-sectional survey of smartwatch users," *Journal of Medical Internet Research*, vol. 27, p. e67414, 2025.

[8] A. Manoni, F. Loreti, V. Radicioni, D. Pellegrino, L. Della Torre, A. Gumiero, D. Halicki, P. Palange, and F. Irrera, "A new wearable system for home sleep apnea testing, screening, and classification," *Sensors*, vol. 20, no. 24, p. 7014, 2020.

[9] M. W. Chee, M. Baumert, H. Scott, N. Cellini, C. Goldstein, K. Baron, S. A. Imtiaz, T. Penzel, C. A. Kushida *et al.*, "World sleep society recommendations for the use of wearable consumer health trackers that monitor sleep," *Sleep Medicine*, vol. 131, p. 106506, 2025.

[10] A. Romigi, "Navigating the human element in an age of wearable sleep technologies: Reflections on the world sleep society recommendations," *Sleep medicine*, vol. 133, p. 106619, 2025.

[11] S. Aziz, A. AM Ali, H. Aslam, A. A Abd-alrazaq, R. AlSaad, M. Alajlani, R. Ahmad, L. Khalil, A. Ahmed, and J. Sheikh, "Wearable artificial intelligence for sleep disorders: scoping review," *Journal of Medical Internet Research*, vol. 27, p. e65272, 2025.

[12] A. Osa-Sanchez, J. Ramos-Martinez-de Soria, A. Mendez-Zorrilla, I. O. Ruiz, and B. Garcia-Zapirain, "Wearable sensors and artificial intelligence for sleep apnea detection: A systematic review," *J. Med. Syst.*, vol. 49, no. 1, p. 66, May 2025.

[13] K. O'shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[14] A. John, K. K. Nundy, B. Cardiff, and D. John, "Somnnet: An SpO2 based deep learning network for sleep apnea detection in smartwatches," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Nov 2021, pp. 1961–1964.

[15] Z. Strumpf, W. Gu, C. W. Tsai, P. L. Chen, E. Yeh, L. Leung, C. Cheung, I. C. Wu, K. P. Strohl, T. Tsai, R. J. Folz, and A. A. Chiang, "Belun Ring (Belun Sleep System BLS-100): Deep learning-facilitated wearable enables obstructive sleep apnea detection, apnea severity categorization, and sleep stage classification in patients suspected of obstructive sleep apnea," *Sleep Health*, vol. 9, no. 4, pp. 430–440, Aug 2023.

[16] A. Vellido, "The importance of interpretability and visualization in machine learning for applications in medicine and health care," *Neural computing and applications*, vol. 32, no. 24, pp. 18 069–18 083, 2020.

[17] Q. Teng, Z. Liu, Y. Song, K. Han, and Y. Lu, "A survey on the interpretability of deep learning in medical diagnosis," *Multimedia Systems*, vol. 28, no. 6, pp. 2335–2355, 2022.

[18] D. Kim, J. Y. Han, H. Jung, D. Y. Song, C. Lee, G. Ryu, S. D. Hong, H.-Y. Kim, and Y. G. Jung, "Ai-enhanced smartwatch ahi estimation and ai-scored polysomnography for obstructive sleep apnea: Real-world validation," *Nature and Science of Sleep*, pp. 2297–2307, 2025.

[19] M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A. Keenan, M. H. Kryger, T. Penzel *et al.*, "The visual scoring of sleep in adults," *Journal of clinical sleep medicine*, vol. 3, no. 02, pp. 121–131, 2007.

[20] J. V. Rundo and R. Downey III, "Polysomnography," *Handbook of clinical neurology*, vol. 160, pp. 381–392, 2019.

[21] L. Varghese, G. Rebekah, A. Oliver, R. Kurien *et al.*, "Oxygen desaturation index as alternative parameter in screening patients with severe obstructive sleep apnea," *Sleep science*, vol. 15, no. S 01, pp. 224–228, 2022.

[22] D. Seok, S. Lee, M. Kim, J. Cho, and C. Kim, "Motion artifact removal techniques for wearable eeg and ppg sensor systems," *Frontiers in Electronics*, vol. 2, p. 685513, 2021.

[23] J. Lee, "Motion artifacts reduction from ppg using cyclic moving average filter," *Technology and Health Care*, vol. 22, no. 3, pp. 409–417, 2014.
[24] J. Hayano, E. Watanabe, Y. Saito, F. Sasaki, K. Kawai, I. Kodama, and H. Sakakibara, "Diagnosis of sleep apnea by the analysis of heart rate variation: a mini review," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2011, pp. 7731–7734.
[25] G. Insalaco, A. Braghiroli, F. Buzzi, S. Cappellano, A. Castelli, A. Castelnuovo, F. Fanfulla, L. F. Strambi, A. L. Bue, S. Marelli *et al.*, "Excessive daytime sleepiness and sex-related differences in the clinical presentation of obstructive sleep apnea in italian patients," *Sleep Medicine*, p. 106819, 2025.