2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340237

Automatic monitoring of obstructive sleep apnea based on multi-modal signals by phone and smartwatch

Guangxin Zhou[1], Wenli Zhou[1], Yi Zhang[1], Zijing Zeng[1], Wei Zhao[1,2]

***Abstract*—Obstructive Sleep Apnea (OSA) is the most common sleep-related breathing disorder, with an overall population prevalence ranging from 9% to 38%, and it is associated with many cardiovascular diseases. The diagnosis of OSA requires polysomnography (PSG) testing, which is unsuitable for large-scale preliminary screening due to its high cost and discomfort to wear. Therefore, a simple and inexpensive screening method would be of great value. This study presents a novel at-home OSA screening method using a smartwatch and a smartphone to obtain several physiological signals, snoring segments, and questionnaire information during a whole night's sleep. The proposed method can distinguish four OSA risk levels based on machine learning (ML) classifications; the system was validated by conducting an in-hospital study on 350 subjects with sleep disorders. The estimated OSA risk levels are in good agreement with the OSA severity diagnosed by PSG (correlation with apnea-hypopnea index (AHI) = 0.92), and an encouraging classification performance is achieved (accuracy = 88.1%, 84.5%, 85.1%, sensitivity = 89.1%, 84.2%, 85.6% for mild, moderate and severe OSA). These findings reveal that wearable devices have the potential for large-scale OSA screening.**

I. Introduction

Sleep Apnea Hypopnea Syndrome (SAHS) is a chronic disease that significantly impacts an individual's health[1]. Obstructive Sleep Apnea is the most common type of SAHS. The main clinical manifestations of Obstructive Sleep Apnea (OSA) are loud snoring, morning headache, and daytime sleepiness. Frequent apnea, which are characterized by temporary cessation of breathing during sleep, can cause nocturnal hypoxia and hypercapnia, adding risk to complications such as hypertension, coronary heart disease, diabetes, cerebrovascular disease, and even sudden death at night[2]. As a potentially life-threatening sleep-related breathing disorder, the primary pathogenesis of OSA is upper airway stenosis, obstruction, or collapse.

The gold standard for clinical diagnosis of OSA is polysomnography (PSG) testing, which requires simultaneous monitoring of electroencephalogram (EEG), electrooculogram (EOG), electromyography (EMG), electrocardiogram (ECG), nasal pressure, airflow (AF), chest and abdomen breathing efforts, snoring, etc. Moreover, the cost of PSG testing is high, and it is complicated to wear, which will interfere with the patient's sleep and produce the first night effect. Furthermore, according to American Academy of Sleep Medicine (AASM) standards, PSG signals require manual scoring by trained technicians. However, subjective judgments can affect the results, and agreement between two scorers on the annotation of PSG events is only 71%[3]. Therefore, seeking a low-cost, convenient, and high-accuracy OSA monitoring technology is essential.

In recent years, the rise of wearable devices such as smartwatches has made it possible to monitor physiological signals continuously. Wearable devices are usually equipped with multiple sensors to monitor blood oxygen saturation, heart rate, and respiratory rate during sleep. Wearable devices provide a convenient and comfortable way to assess sleep quality and OSA risk levels.

The OPPO watch (OW) [4], a smartwatch equipped with photoplethysmography (PPG) and ACC sensors, was used in this study. It connects to the smartphone through the HeyTap Health APP. Users need to fill out a questionnaire on the APP and turn on the snoring monitoring function. The algorithm calculates the sleep stage, HR, respiration, SpO2, and snoring. After the user wakes up, the algorithm will automatically summarize the features of the whole night and calculate the respiratory event index (REI) and OSA severity through ML and deep learning (DL) models.

In this study, we developed and validated a system for estimating OSA severity using smartphones and smartwatches. The algorithm detected respiratory events and extracted OSA risk factor characteristics by combining sleep stage, SpO2, respiratory rate, heart rate variability, snoring, and questionnaire results. Then ML models estimated the REI and the OSA severity. We verified the performance by conducting an in-hospital study using both our devices and PSG.

II. Materials and Methods

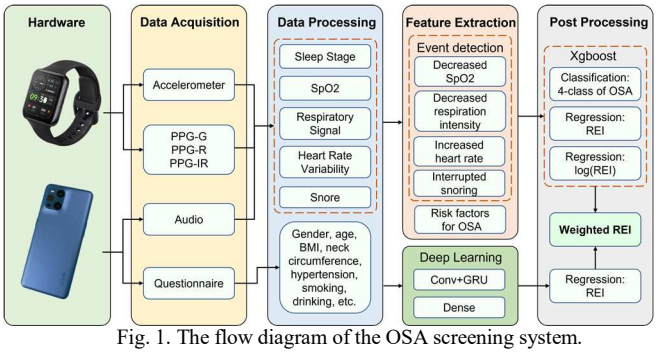

Fig. 1. The flow diagram of the OSA screening system.

Fig. 1 is the flow diagram of the OSA screening system design containing six blocks: hardware, data acquisition, data processing, feature extraction, DL, and post-processing. The system collected raw sensor signals and questionnaires through smartphones and smartwatches. The physiological signals such as sleep stage, SpO2, respiration, heart rate, and

*Research supported by OPPO.

1. Department of Health Lab, OPPO Research Institute, Guangdong OPPO MOBILE Telecommunications Corp., Ltd.

2. Corresponding author, phone number: +86 13120214327, email: zhaowei3@oppo.com

snoring were obtained by raw signal processing and analysis. We can detect respiratory events and extract OSA risk factors from the physiological signals and questionnaires. Then these factors were fed into three ML models to estimate OSA severities and REIs. Simultaneously, we estimated REIs from all the physiological signals through a DL module. All the REIs and OSA severities above were weighted to obtain the final result.

A. Clinical Dataset

The 357 participants for this study were recruited from the sleep medicine center of Shenzhen People's Hospital from June 1, 2022, to November 15, 2022. It was approved by the Ethics Committee of the Shenzhen People's Hospital as LL-KY-2022140-01. All participants signed informed consent forms.

B. Experimental Setup

Each participant was connected with PSG leads and wore an OW by the technicist during sleep. Snore monitoring is enabled through the HeyTap Health app and placed within 1 meter. The smartwatch collected infrared, red, green PPG and triaxial acceleration signal at a sampling rate of 50Hz. The phone collected audio recordings, which were collected at a sampling rate of 8kHz, and recorded the answers to a questionnaire. The PSG devices were Philips Alice 6 LDE (Philips Healthcare, Inc.), and the Sleepware G3 software was also used. Trained technicians scored PSG signals following the AASM guidelines.

Some data were excluded due to signal quality, recording duration, and wearing status problems. The detailed conditions are as follows:1) Less than 3 hours of sleep; 2) Missing audio recordings or questionnaire; 3) Poor signal quality; 4) Valid signal length is less than 3 hours; 5) Missing PSG data.

Finally, a total of 350 participants formed the dataset. 25 participants didn't have OSA (apnea-hypopnea index, AHI<5), 72 had mild OSA (5≤AHI<15), 94 had moderate OSA (15≤AHI<30), and 159 had severe OSA (AHI≥30).

C. Data Processing and Feature Extraction

1) Sleep Stages: Sleep stages are calculated from the accelerometer and green PPG. The accelerometer can detect user actions, and PPG can calculate HRV. Sleep state can be measured according to motion amplitude and HRV. PPG signals are filtered with a bandpass filter at [0.75,5] Hz. HRV features are calculated from RR intervals obtained by the peak-seeking algorithm. The algorithm uses actions and HRV features to estimate sleep states (deep sleep, light sleep, and rapid eye movement (REM)) through rules and ML methods.

2) SpO2: The OW is equipped with red and infrared PPG and accelerometers. OW can be used to estimate SpO2 by the different absorption rates of oxygenated hemoglobin (HbO2) and deoxygenated hemoglobin (Hb) [5]. After eliminating the segments with low SNR, the signal was interpolated and filtered. Then AC and DC were obtained, and r values were calculated. Finally, the SpO2 was estimated according to the r curve. The r curve is calibrated in a specially collected private dataset and independent of the dataset in this paper.

The OW calculates SpO2 per second continuously and obtains–signal quality. An oxygen desaturation event was defined as a 4% decrease in SpO2 and can be detected through the SpO2 curve. However, oxygen desaturation events did not correspond to apnea and hypopnea events one by one. Respiratory disease, sleep arousal, motion artifacts, and so on, will produce artifacts to affect the SpO2 values. The detected oxygen desaturation events are classified as respiratory-related events according to SpO2 morphology and quality. Finally, SpO2-based REI was estimated by the SpO2 algorithm.

3) HRV: Respiratory-related events are accompanied by fluctuations in instantaneous heart rate. Respiratory events can also be distinguished from the spectral characteristics of RR intervals, mainly reflected in ultra-low frequency energy. In the Sleep Stages section, we have described the method for calculating HRV by green PPG.

We calculated time domain and frequency domain HRV features using a 1-minute window. Time domain features include SDNN, RMSSD, NN50, PNN50, heart rate standard deviation, range of heart rate, etc. Frequency domain features include LF ratio ([0.04 0.15) Hz), HF ratio ([0.15 0.4) Hz), VLF ratio ([0.003 0.04) Hz), LF/HF ratio, and VLF/HF ratio[6]. Based on these features, we developed a HRV algorithm to detect respiratory events.

The HRV algorithm uses a machine learning model to calculate the probability of respiratory events as A = $\{x_1, x_2, \ldots, x_n\}$ at the interval of 10s. For the element x_i, get the num m of $x_i > 0.5$ within $i \sim i+k$ (k is a specified threshold). A respiratory event is recorded if m is greater than the specified threshold. We set minimum intervals for adjacent respiratory events to avoid double counting. Finally, the HRV-based REI was estimated by the HRV algorithm.

4) Respiration: Respiratory signals during sleep can be extracted from the PPG and ACC of the OW. OW can detect the occurrence of apnea events by the decrease in the intensity of respiratory. The end of an apnea event is usually accompanied by arousal that produces a slight movement. A decrease in respiratory intensity lasting more than 10s, accompanied by arousal movements, was defined as an apnea event. OW summarizes events and calculates REI estimated for respiratory signals.

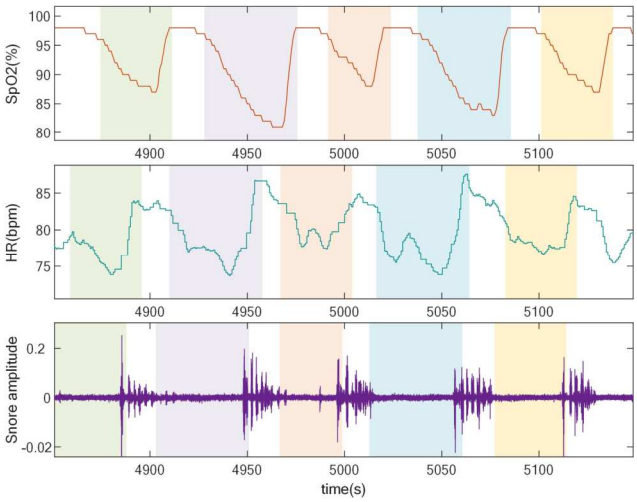

Fig. 2. Signal trends during apnea episodes.

5) Snore: We divided audio into frames (256 points per frame) and calculated short-time energy (STE) and zero-crossing rate (ZCR) on the filtered audio. Then the audio signal is detected and segmented. The Mel Frequency Cepstrum Coefficient (MFCC), ZCR, STE, and other audio signal features are calculated to classify snoring and other sounds[7].

The OW-paired HeyTap Health App detects snoring through the algorithm described above. Each snore records its start and end time, acoustic characteristics, etc. Persistent snoring is interrupted when apnea occurs. Algorithms detect snoring interruptions and determine whether the interruption is followed by awake or louder snoring, which implies respiration-related events. Finally, the snore-based REI was estimated by the snore algorithm.

6) Mutual Authenticating of Different Signals: OW detected respiratory events using SpO2, HRV, respiration, and snore signals separately. However, the effects of apnea or hypopnea on different signals are interrelated, as shown in Fig. 2, on which each color represents a single apnea event. At the onset of apnea, snoring ceases, followed by a decrease in SpO2 and respiration intensity, and an increase in heart rate. Upon apnea cessation, snoring, SpO2, respiration intensity, and heart rate return to normal levels. The hysteresis of SpO2 causes the oxygen desaturation event to be delayed by about 25 seconds compared to snoring and heart rate changes. Events detected by different signals were mutual authenticated to improve the accuracy of the final results. Respiratory-related events were confirmed if detected in two or more signals within the same time window. Then the mixed events-based REI was estimated.

7) Questionnaire: The sleep questionnaire in Table I is an example of the OSA screening questionnaire used in this paper. Most of the questions are about high-risk factors that may cause OSA, the sensitivity of some factors to OSA can reach 97.7%[8].

Questionnaire information includes subjects' personal information (neck circumference, BMI), night sleep status, daytime activity status, daily routine, and primary disease history. Nocturnal sleep conditions refer to the sleep disturbances that may occur during sleep were discovered by others: apnea, shortness of breath, wheezing, and suffocation; each event can last from several seconds to several minutes, and it occurs several times throughout the night. Sometimes there will be sounds such as suffocation or nasal sounds when rebreathing. When normal sleep is disturbed, it will cause memory loss, short temper, fatigue, and drowsiness during the day. The daily routine includes whether the subject has long-term smoking, drinking, use of sedatives, or other bad living habits. Disease history refers to diseases highly related to OSA, including 1) cardiovascular and cerebrovascular diseases, such as atrial fibrillation, coronary heart disease, stroke, and hypertension[9]; 2) respiratory narrowing diseases[10]: such as mandibular retraction, tongue hypertrophy, adenoid enlargement, long soft palate tissue, and rhinitis; 3) lung diseases: such as chronic obstructive pulmonary disease, asthma, tuberculosis.

D. Classification

We employed multi-input deep learning architectures to determine the probability of respiratory events. The deep learning model utilizes SpO2, respiratory waveforms, heart rate, and action amplitude to approximate the density of respiratory-related events within a 15-minute timeframe. Upon awakening, the REI-DL can be assessed.

We also derive statistical features from sleep stages, HR, SpO2, Resp, and snore, such as total sleep time, mean SpO2, heart rate, and snoring frequency. The XGBoost algorithm, an ensemble tree machine learning model, incorporated both REI-DL and statistical features as input, while the OSA risk level determined by polysomnography (PSG) served as a label, as shown in Fig.1. We developed classification and regression models for evaluating OSA risk levels and Apnea-Hypopnea Index (AHI) estimations, respectively.

E. Validation

To ensure the balance between different categories of data, we use Stratified 5-fold Cross Validation (S5CV, each fold maintains the same proportion of each category in the original data) to train the XGBoost model. Finally, we obtained the receiver operating characteristics (ROC) curve and the confusion matrix from the model. Meanwhile, the average accuracy, sensitivity, specificity, and area under the curve (AUC) of multi-fold validation under the AHI threshold of 5, 15, and 30 can be obtained.

III. RESULTS

Fig. 3 shows the performance of the XGBoost model, which includes: (a). The ROC curve corresponds to the AHI threshold of 15 (the thick red diagonal line represents the result of random guessing, the blue thick is the average value of multiple folds, and the thin lines of different colors represent the prediction results of each fold); (b). The confusion matrix corresponding to the S5CV algorithm, which represents the OSA severity obtained from PSG (0: normal; 1: mild; 2: moderate; 3: severe) compared with the result predicted by the XGBoost model, the value in each cell represents the number of subjects. Generally, the darker the cell color, the larger the value; the darker-colored cells on the diagonal represent a high proportion of accurately predicted objects.

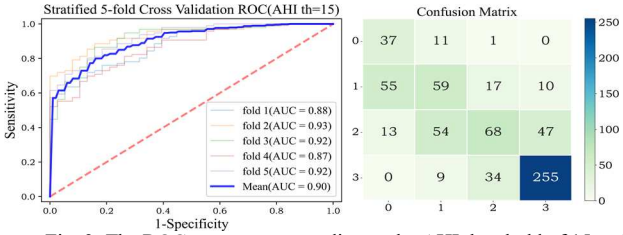

Fig. 3. The ROC curve corresponding to the AHI threshold of 15 and the confusion matrix of the XGBoost model. The performance is supported by 670 pieces of data on both hands from 350 participants.

Table I gives the statistical information of the six most related features from the questionnaire, where the Positive ratio is the proportion of patients who answered yes for the feature. Correlation r represents the correlation between the feature and the standard AHI from PSG.

In the decision module, we combine the results of ML and DL models to obtain the final OSA severity. An overview of the results can be seen in Table II. The tabular data shows that as the AHI threshold increases, the requirements for judging as positive become more stringent, and the proportion of

subjects predicted to be positive becomes smaller, so the classification sensitivity decreases and the specificity increases.

TABLE I. STATISTICAL RESULTS OF SIX FEATURES EXTRACTED FROM THE QUESTIONNAIRE.

Feature	Positive ratio	Correlation, r
Cough or snore	83.6%	0.24
Apnea or wheeze	38.2%	0.25
Shortness of breath	44.5%	0.20
Drink alcohol	11.3%	0.22
BMI > 25 kg/m^2	57.8%	0.40
Neck circ > 40 cm	24.1%	0.47

TABLE II. CLASSIFICATION PERFORMANCE UNDER THREE AHI THRESHOLDS OF 5, 15, AND 30.

Performance	Acc (%)	Sen (%)	Spec (%)
th5	88.1	89.1	75.5
th15	84.5	84.2	85.3
th30	85.1	85.6	84.7

IV. DISCUSSION

In this study, we proposed a multi-modal based method to estimate OSA severity conveniently on wearable devices and validated the result against PSG. Three types of features were used, including physiological parameters (such as HR, SpO2, respiration, and motion parameters), snoring segments, and questionnaire information are fed into our ML and DL models. The final classification results in Table II show that the system, as a non-invasive, convenient, and low-cost OSA screening tool, performed very well on the 350 subjects we collected. Furthermore, we investigated the performance of other similar technologies in literature (selecting recent studies on estimating OSA severity using wearable devices sensor data and validation with clinical datasets). The specific results are presented in Table III. Our proposed method exhibits an overall improvement in performance across different thresholds when compared to existing literature. This finding highlights the significant advantage of our approach for estimating OSA severity.

While smart wearable devices may not supplant polysomnography (PSG) for OSA screening due to their limited capacity to extract physiological signals, such as respiratory or brain waves, their ongoing development holds promise. As these devices continue to evolve, they will likely expand the scope of signal acquisition, enabling the unobtrusive and continuous monitoring of multi-dimensional sleep characteristics within the comfort of one's home. This advancement could pave the way for innovative OSA monitoring approaches and offer an affordable, objective solution for large-scale chronic disease screening.

V. CONCLUSION

This paper introduces a novel system for estimating the severity of OSA using smartphones and smartwatches. Our approach demonstrates promising results in three binary classification tasks, achieving accuracies ranging from 84.5% to 88.1%, sensitivities between 84.2% and 89.1%, and specificities from 75.5% to 85.3%. Based on these findings, we conclude that our method holds significant potential for large-scale OSA screening in the general population.

TABLE III. COMPARISON OF DIAGNOSTIC PERFORMANCE WITH THE REFERENCES.

Research paper	Equipment type	Performance	AHI threshold		
			5	15	30
Gu et al, 2020[11]	Ring	Acc (%)	86	86	86
		Sen (%)	95	85	22
		Spec (%)	29	87	100
Zhao et al, 2022[12]	Ring	Acc (%)	86	76	91
		Sen (%)	87	66	74
		Spec (%)	83	96	100
Chen et al, 2021[13]	Smartwatch	Acc (%)	81.1	88.3	88.4
		Sen (%)	67.5	88.7	85.7
		Spec (%)	94.7	87.8	91
Papini et al, 2020[14]	Smartwatch	Sen (%)	77	62	46
		Spec (%)	72	91	98
Our paper	Smartwatch	Acc (%)	88.1	84.5	85.1
		Sen (%)	89.1	84.2	85.6
		Spec (%)	75.5	85.3	84.7

REFERENCES

[1] M. H. Kryger, T. Roth, and W. C. Dement, *Principles and Practice of Sleep Medicine E-Book*: Elsevier Health Sciences, 2010.

[2] S. M. Caples, A. Garcia-Touchard, and V. K. Somers, "Sleep-disordered breathing and cardiovascular risk," *Sleep,* vol. 30, no. 3, pp. 291-303, 2007.

[3] R. G. Norman, I. Pal, C. Stewart, J. A. Walsleben, and D. M. Rapoport, "Interobserver agreement among sleep scorers from different centers in a large dataset," *Sleep,* vol. 23, no. 7, pp. 901-908, 2000.

[4] OPPO. "OPPO WATCH 3PRO," December 14, 2022; https://www.oppo.com/cn/accessories/oppo-watch-3-pro/.

[5] R. Kirszenblat, and P. Edouard, "Validation of the Withings ScanWatch as a Wrist-Worn Reflective Pulse Oximeter: Prospective Interventional Clinical Study," *Journal of Medical Internet Research,* vol. 23, no. 4, pp. e27503, 2021.

[6] A. Garde, P. Dekhordi, J. M. Ansermino, and G. A. Dumont, "Identifying individual sleep apnea/hypoapnea epochs using smartphone-based pulse oximetry." pp. 3195-3198.

[7] K. Qian, C. Janott, V. Pandit, Z. Zhang, C. Heiser, W. Hohenhorst, M. Herzog, W. Hemmert, and B. Schuller, "Classification of the excitation location of snore sounds in the upper airway by acoustic multifeature analysis," *IEEE Transactions on Biomedical Engineering,* vol. 64, no. 8, pp. 1731-1741, 2016.

[8] I. H. El-Sayed, "Comparison of four sleep questionnaires for screening obstructive sleep apnea," *egyptian Journal of Chest diseases and Tuberculosis,* vol. 61, no. 4, pp. 433-441, 2012.

[9] A. Bangash, F. Wajid, R. Poolacherla, F. K. Mim, and I. H. Rutkofsky, "Obstructive sleep apnea and hypertension: A review of the relationship and pathogenic association," *Cureus,* vol. 12, no. 5, 2020.

[10] R. Davies, N. Ali, and J. Stradling, "Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome," *Thorax,* vol. 47, no. 2, pp. 101-105, 1992.

[11] W. Gu, L. Leung, K. C. Kwok, I.-C. Wu, R. J. Folz, and A. A. Chiang, "Belun Ring Platform: a novel home sleep apnea testing system for assessment of obstructive sleep apnea," *Journal of Clinical Sleep Medicine,* vol. 16, no. 9, pp. 1611-1617, 2020.

[12] R. Zhao, J. Xue, X. Zhang, M. Peng, J. Li, B. Zhou, L. Zhao, T. Penzel, M. Kryger, and X. S. Dong, "Comparison of Ring Pulse Oximetry Using Reflective Photoplethysmography and PSG in the Detection of OSA in Chinese Adults: A Pilot Study," *Nature and Science of Sleep*, pp. 1427-1436, 2022.

[13] Y. Chen, W. Wang, Y. Guo, H. Zhang, Y. Chen, and L. Xie, "A single-center validation of the accuracy of a photoplethysmography-based smartwatch for screening obstructive sleep apnea," *Nature and Science of Sleep,* vol. 13, pp. 1533, 2021.

[14] G. B. Papini, P. Fonseca, M. M. van Gilst, J. W. Bergmans, R. Vullings, and S. Overeem, "Wearable monitoring of sleep-disordered breathing: Estimation of the apnea–hypopnea index using wrist-worn reflective photoplethysmography," *Scientific reports,* vol. 10, no. 1, pp. 1-15, 2020.