



OPEN Apnea detection using wrist actigraphy in patients with heterogeneous sleep disorders

Xiaoman Xing^{1,2,7}, Sizhi Ai^{3,7}, Jihui Zhang³, Rui Huang^{2,5}, Yaping Liu³, Dongming Quan⁴, Jiacheng Ma³, Guoli Wu³, Jiangen Xu², Yuan Zhang⁶, Hongliang Feng³✉ & Wen-fei Dong²✉

Obstructive sleep apnea (OSA) and related hypoxia are well-established cardiovascular and neurocognitive risk factors. Current multi-sensor diagnostic approaches are intrusive and prone to misdiagnosis when simplified. This study introduces an enhanced single-sensor-based OSA screening method, leveraging novel signal processing and machine learning to ensure accurate detection across diverse populations. Wrist actigraphy is a widely-used and energy-efficient tool for respiratory rate estimation. The main challenge in OSA pattern recognition is handling various disturbances in real-world applications. We developed a novel approach combining apex-centric tokenization with a Multi-Head Causal Attention (MHCA) mechanism. Apex-centric tokenization enhances sensitivity to OSA events, while MHCA refines predictions and increases specificity in detecting oxygen desaturation. Our study involved 58 participants, with overnight bilateral wrist actigraphy and concurrent polysomnography used as a reference for thorough analysis. By focusing on the physiological causal relationship of the events, the algorithm excelled in detecting moderate to severe oxygen desaturation, achieving a sensitivity of 85.7% and a specificity of 98.1%, even in the presence of disturbances such as restless leg movements and snoring. The estimated oxygen desaturation index correlated strongly with clinical standards ($r = 0.89$), and the correlation with the apnea-hypopnea index was 0.87. Both apex-centric tokenization and MHCA were crucial for this performance. Our approach shows potential for analyzing apnea patterns and related oxygen desaturation in a broader population using only wrist actigraphy, reducing measurement burdens and improving understanding of complex sleep disorders.

Keywords Wrist actigraphy, Apnea, Oxygen desaturation, Tokenization, Multi-head causal attention

Obstructive sleep apnea (OSA) is the most common sleep-related respiratory disorder, characterized by airway obstruction, decreased blood oxygen, and interrupted sleep cycles¹. Affecting nearly one billion people globally, OSA is linked to serious health risks, including cardiovascular diseases, cognitive impairments, and traffic accidents². In 2015, the cost of diagnosing and treating OSA in the USA was approximately US\$12.4 billion³. Traditional diagnostic methods for OSA rely on multiple intrusive sensors, such as respiratory inductive plethysmography (RIP), electromyography (EMG), nasal/oral airflow sensors, and pulse oximeters, which can be uncomfortable and disruptive to patients⁴. This complexity contributes to underdiagnosis in general population^{5,6}. Simplifying the diagnostic process while maintaining accuracy remains a challenge, especially in patients with comorbidities.

Wrist actigraphy, an energy-efficient and non-invasive method for monitoring respiratory rate (RR), has shown promise for OSA detection⁷. For healthy and OSA individuals with no comorbidities, wrist actigraphy demonstrates comparable accuracy to traditional chest-worn sensors. Hernandez et al. reported a mean absolute error (MAE) of 0.97 ± 2.20 breaths per minute (bpm) in RR estimation⁸. Xiao Sun et al. integrated a Kalman filter, achieving MAEs of 0.72 and 1.08 bpm for healthy and simple OSA patients, respectively⁹. Leube et al.

¹Division of Life Sciences and Medicine, School of Biomedical Engineering (Suzhou), University of Science and Technology of China, Suzhou, Jiangsu, China. ²Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, Jiangsu, China. ³Center for Sleep and Circadian Medicine, The Affiliated Brain Hospital, Guangzhou Medical University, Guangzhou, Guangdong, China. ⁴Guangdong Mental Health Center, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, China. ⁵Academy for Engineering and Technology, Fudan University, Shanghai, China. ⁶Department of Respiratory Medicine, Xiangya Hospital, Central South University, Changsha, China. ⁷These authors contributed equally: Xiaoman Xing and Sizhi Ai. ✉email: hlfeng@link.cuhk.edu.hk; wenfeidong@126.com

demonstrated the potential of reconstructing respiratory waveforms using electrocardiogram and accelerometer data¹⁰. Chen et al. have tried a random forest algorithm to detect OSA without comorbidities, achieving an F1 score of 0.96 and an Apnea-Hypopnea Index (AHI) intraclass correlation of 0.88¹¹. Zhao et al. suggested nonlinear Poincaré analysis to enhance respiratory pattern recognition¹².

Despite these advancements, real-world applications of respiratory monitoring face additional challenges. Comorbidities such as snoring and restless leg syndrome (RLS) can significantly disrupt measurements, reducing their accuracy. Schipper et al. assessed respiratory effort using a chest-worn accelerometer, observing a fourfold increase in mean squared errors during leg movements and a tenfold increase during wakefulness compared to non-rapid-eye-movement (NREM) sleep¹³. D. Liaqat et al. used Convolutional Neural Networks (CNN) to extract features from accelerometer and gyroscope data, by discarding “unsuitable” data, they achieved a MAE of 2.05 bpm for ambulatory “real-life” RR estimation, with varying accuracies for healthy individuals and chronic obstructive pulmonary disease (COPD) patients¹⁴. Papini et al. evaluated OSA detection using even-spaced 30-second window, achieving an AHI correlation of 0.72 and an error of 0.56 ± 14.74 events/h. However, this short window may miss prolonged OSA events, possibly reducing positive predictive value for severe cases¹⁵. Although these studies were conducted in real-world settings, the influence of comorbidities on OSA detection was not thoroughly analyzed, and there is a need for further improvement in performance.

To address these challenges, recent research has increasingly turned to advanced machine learning architectures, such as Fully Convolutional Networks (FCN), CNN, Long Short-Term Memory networks (LSTM), and U-Nets^{14,16–19}. The integration of transformer architectures, typically seen in token-based Large Language Models (LLMs)²⁰, signifies a new exploration in this domain. Nonetheless, these methodologies face limitations: they either adhere to a predetermined window size without refining the token definition, thereby risking misclassification due to incomplete capture of events, or they employ pixel-to-pixel sequence classification approaches that demand meticulous labeling and may prove unfeasible for extensive datasets with unavoidable temporal misalignments.

In this work, we made several key contributions to address these problems:

- (1) Infer oxygen levels from motion patterns: By leveraging physiological principles, we demonstrated that a single energy-efficient motion sensor can reliably detect apnea events and oxygen desaturation episodes - even in the presence of RLS and snoring^{21–23}.
- (2) Apex-Centric Flexible Windowing: We introduce an apex-centric flexible windowing approach to enable comprehensive and precise event detection.
- (3) Multi-Head Causal Attention (MHCA): We leverage MHCA to focus on physiologically plausible event sequences, enhancing the accurate detection and classification of diverse OSA-related desaturation patterns^{24,25}.

By distinguishing these events, our method enables more informative, accurate and low-cost actigraphy-based monitoring, enhancing its clinical utility for large-scale retrospective studies.

Methods

This section details our experimental protocol, data processing pipeline, and analytical framework. First, we describe the study design and data acquisition procedures. Then, we present our standardized preprocessing workflow, comprising signal filtering, data alignment, and a novel flexible apex-centered windowing approach. Lastly, we introduce our two-stage machine learning architecture, emphasizing its physiologically-grounded design principles alongside the comprehensive validation process.

Data collection

Our study involved 58 participants who underwent full-night digital polysomnography (PSG, Grael, Compumedics, Australia) and concurrent bilateral wrist actigraphy (Geneactiv, Activinsights Ltd, UK; Axivity, Axivity Ltd, UK) at the Department of Sleep Medicine, Guangdong Mental Health Center, Guangdong Province People's Hospital. The participants included 30 males, with ages ranging from 12 to 73 years. The study was approved by the Ethics Committee (IRB KY-Q-2022-171-01) and followed the Declaration of Helsinki. Informed consents have been obtained from all subjects. The two actigraphy brands were selected due to their widespread use in retrospective studies and their demonstrated equivalent performance^{26,27}.

Figure 1 illustrates the experimental setup. Actigraphy data were collected at a 100 Hz sampling rate. A registered technologist annotated PSG data according to the American Academy of Sleep Medicine (AASM) criteria²⁴. OSA is characterized by episodes of upper airway obstruction during sleep, including apneas (complete cessation of airflow for ≥ 10 s) and hypopneas ($\geq 30\%$ reduction in airflow for ≥ 10 s with $\geq 3\text{--}4\%$ oxygen desaturation or arousal). Limb movements were measured by EMG and were not scored within 0.5s of a respiratory event. Electroencephalogram (EEG) was used to annotate sleep stages, although this is not shown in the illustration.

Data processing

Preprocessing

Actigraphy data contains rich spectral information. Gravitational and seismic vibrations were separated using different filters. A Butterworth bandpass filter (0.05–0.6 Hz) was used to obtain respiratory information. A low-pass filter of 0.01 Hz was used to obtain posture-related information. Principal Component Analysis (PCA) was then applied to the three-axis signal, and the first principal component was used.

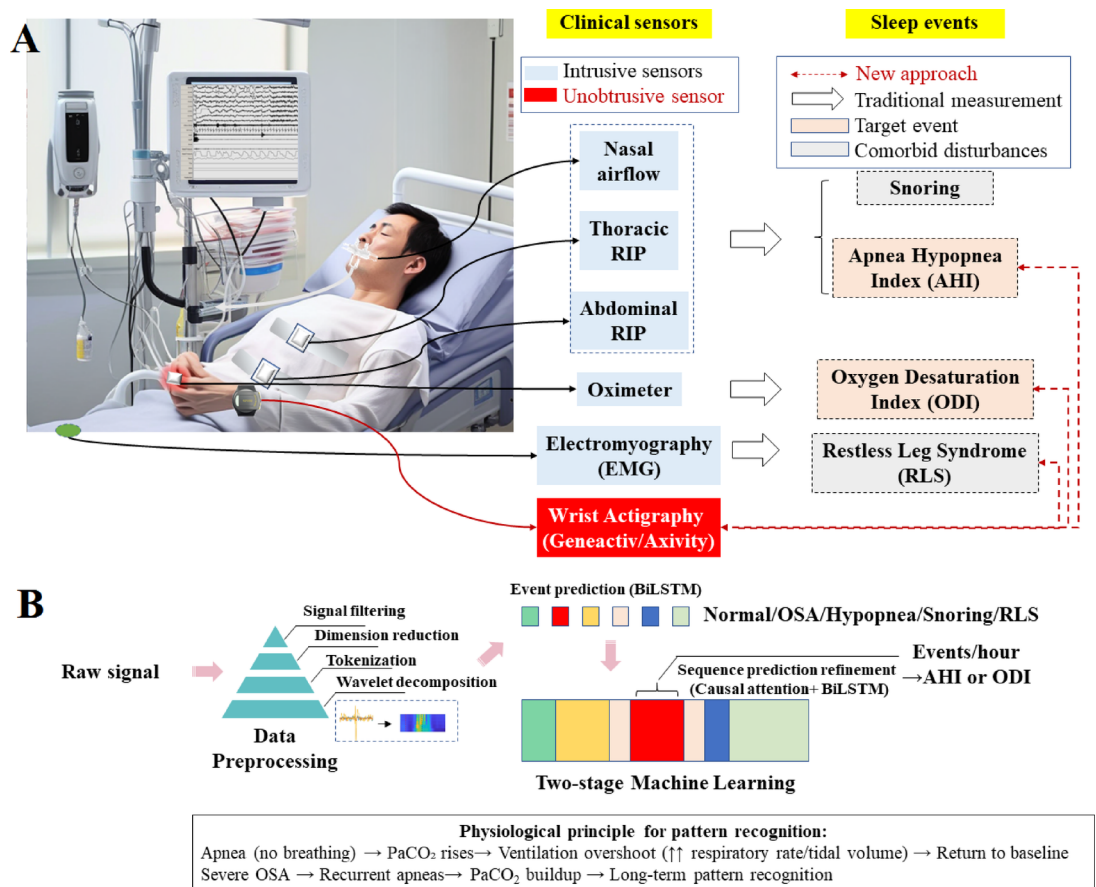


Fig. 1. (A) Illustration of experimental setup. (B) Detailed data processing flowchart for OSA detection. PaCO_2 : arterial partial pressure of carbon dioxide; BiLSTM: Bi-directional Long and Short-Term Memory.

Complete ensemble empirical mode decomposition (CEEMD)

To extract physiologically meaningful motions, the CEEMD method was used to decompose the data into intrinsic mode functions (IMFs)^{28,29}. Unlike traditional EMD, CEEMD reduces mode mixing, enhances robustness, and improves signal separation by adding white noise multiple times and averaging the results, as shown in the Appendix. Using a noise standard deviation of 0.1 and maximum 100 iterations, we decomposed each epoch into ten IMFs. The fourth IMF was selected for its strong respiratory pattern correlation.

The fourier synchrosqueezed transform (FSST) and RR Estimation

FSST is a time-frequency analysis technique that combines ideas from the Continuous Wavelet Transform and the Fourier Transform, as shown in the Appendix³⁰. It provides a sharper localization of signal components in both time and frequency domains simultaneously. RR was estimated as the highest amplitude peak in each epoch's spectrum. In our study, RR was used as an auxiliary feature for apnea detection. Therefore, we did not employ more advanced techniques to refine RR estimation.

Signal alignment

Due to the offline nature of PSG clocks, PSG and actigraphy data can become misaligned. To correct this, we implemented a two-step realignment:

Posture Alignment: We aligned the sleep posture changes obtained from actigraphy data with the PSG annotations.

Respiratory Synchronization: We then calculated the lag for maximum cross-correlation between the thoracic RIP signal and preprocessed respiratory band signal from actigraphy, as described in 2.2.1.

This process ensured accurate synchronization between the PSG and actigraphy data. However, 16 out of 58 subjects had noisy posture-change peaks and even noisier respiratory signals due to restless motion disturbances, leading to their exclusion from the training set. Since “per-subject” AHI and ODI calculations do not require precise alignment, these 16 subjects were used as testing data, providing a rigorous test of the algorithm's performance on noisy, unseen data.

Apex-centric tokenization

Tokenization, a concept borrowed from natural language processing (NLP), is a critical step in machine learning that significantly influences model outcomes. Respiratory events like OSA and hypopnea last over 20 s and are

often followed by oxygen desaturation, while RLS and snoring are much shorter³¹. To capture these events, we used 100-second epochs centered on the potential apex of apnea episodes, labeling each epoch with the dominant event to minimize secondary interferences.

The signal was segmented using a sliding 3-minute window with a 1-minute step. Potential events (PEs) were identified as points exceeding two standard deviations from the median of each window. Only the apex of successive PEs was used as the center of respiratory events, defined by Eqs. (1–2). If PEs were within 0.05s time windows, they were included in the same cluster. Here ACC_{3min} refers to the 3-minute window used to search for PEs, and $PE_{cluster}$ refers to the group of successive PE points, as shown in Fig. 2.

$$PE = |ACC_{3min} - median(ACC_{3min})| > 2 \times std(ACC_{3min}) \quad (1)$$

$$Apex = \max(|PE_{cluster}|) \quad (2)$$

A full-night scan isolates these potential events, while the remaining signal is deemed largely devoid of events. For each apex, a symmetrical 100-second window is established, centered on the apex, encapsulating potential OSA events. PEs caused by noise are quite different from the OSA patterns, as shown in Supplementary Fig. 1. Segments without any apex are deemed ‘clean’ and were partitioned into even-spaced non-overlapping 100-second intervals. If the edges of these ‘clean’ segments fell within 1 min of the edge of a PE window, the corresponding segments were discarded to prevent potential artifact.

Given the uneven distribution of apex positions, the inter-epoch intervals vary, with a minimum interval of 0.06 s. The median length of a 30-epoch sequence is 430.7 ± 127.5 s, while a “clean” 30-epoch sequence can have a maximum length of 3000 s. This tokenization method ensures the meticulous capture of potential events, while the strategic sampling in highly unstable regions forces the subsequent MHCA model to focus on detailed information about apnea, enhancing discrimination.

BiLSTM-MHCA algorithm

Algorithm design

To enhance specificity in the presence of motion artifacts, such as limb movements or respiratory asynchrony between the chest and abdomen, we employed a BiLSTM-MHCA architecture¹⁵. This model leverages the principle of respiratory control: apneic pauses elevate $PaCO_2$ levels, triggering compensatory increases in respiratory rate and tidal volume until normocapnia is restored³². Our model utilizes BiLSTM to compare the post-apneic ventilation patterns against expected $PaCO_2$ responses to differentiate true OSA events.

The BiLSTM detects epoch-wise OSA events but may struggle with severe OSA, where recurrent apneas interrupt $PaCO_2$ recovery. We therefore added a second-stage MHCA to model longer-term respiratory patterns across multiple cycles.

The detailed procedures are as follows:

The BiLSTM first classifies respiratory epochs into five categories, generating similarity scores. The scores are generated by encoding input X into hidden states h , then transform h to class logits $z = W^T h + b$. Here, W and b are learned parameters, and y is the ground truth label. Class probabilities are computed by Eq. (3).

$$p(y = c|X) = \frac{e^{z_c}}{\sum_k e^{z_k}}, \quad c, k \in \{1, \dots, 5\}. \quad (3)$$

These scores are combined with entropy, amplitude, posture, and respiratory rate into 30-epoch non-overlapping sequences, which are subsequently fed into the MHCA for refined predictions. The workflow is illustrated in Fig. 3.

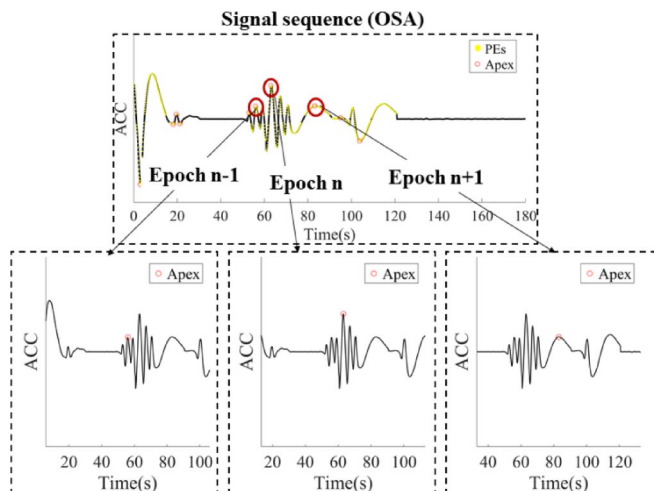


Fig. 2. Apex-centric tokenization procedure for an OSA event.

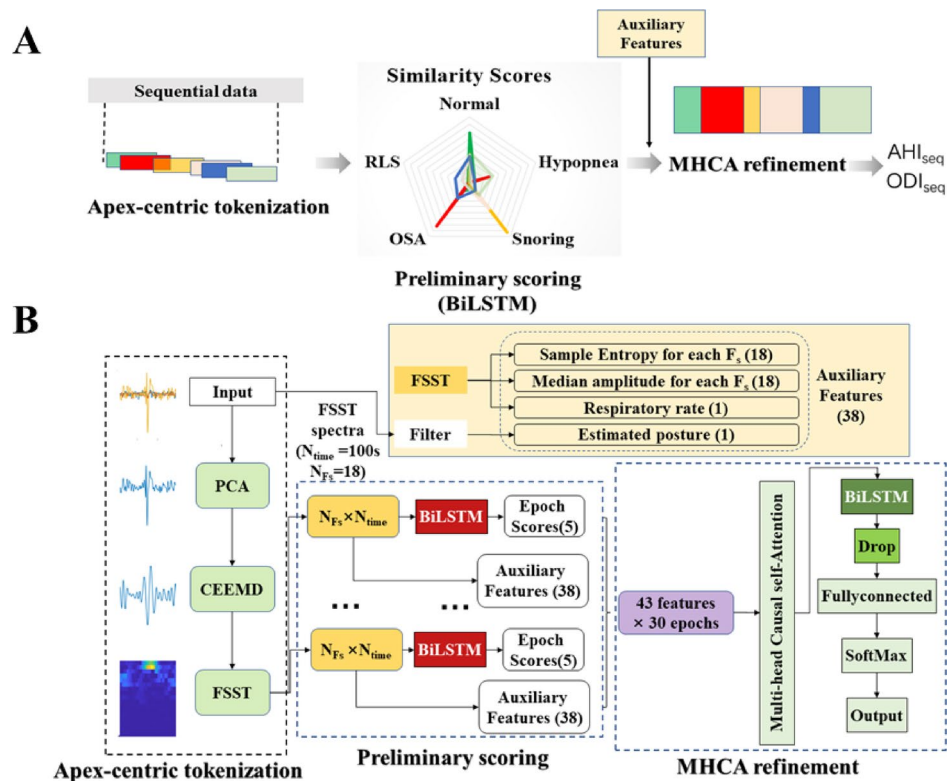


Fig. 3. (A) Illustration of the multi-stage data flow. (B) Detailed design of each module. AHI_{seq} and ODI_{seq} were defined in “Target labelling and recovering” section.

The key component of this structure is causal self-attention, also known as masked self-attention. It ensures that the model aligns with the natural progression of events, and refines the epoch-wise predictions, reducing erratic epoch-wise prediction and enhancing the model’s sensitivity to specific occurrences.

To enforce the causal (or autoregressive) property, a mask is applied to the attention scores. The attention scores are computed as the dot product between the Query and Key vectors, scaled by the square root of the key dimension d_k . In the attention mechanism, Query (Q), Key (K), and Value (V) matrices are standard. Q queries information from the Keys, K determines the relevance of the corresponding Values, and V contains the actual content to be extracted.

The mask M sets the future contribution to negative infinity, ensuring that they do not influence the attention weights.

$$Attention(K, Q, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V \quad (4)$$

To demonstrate the advantages of the apex-centric tokenization and MHCA, we conducted an ablation study. We compared the performance of our approach with variations that used an even-spaced window size and those without the MHCA module. To ensure comparable sequence lengths, we used a 100-second epoch length with a 15-second sliding step to create 450-second sequences, which is similar to the median sequence length of our flexible apex-centric tokenization method, as described in “Apex-centric tokenization” section.

Target labelling and recovering

AHI is a measure used in the diagnosis and assessment of sleep-disordered breathing, particularly in cases of OSA and other sleep-related breathing disorders³³. The definition of AHI used in this study is shown in Eqs. 5–6. It represents the average number of OSA (complete pauses in breathing) and hypopnea (shallow breaths or reduced airflow) per hour of sleep³⁴, either using the whole-night data (AHI_{sub}) or using a 30-epoch sequence (AHI_{seq}). Our study uses five labels for each epoch: normal, OSA, hypopnea, RLS, and snoring.

For each subject, we use AHI_{sub} to stratify the OSA severity.

$$AHI_{sub} = \frac{Apnea\ counts + Hypopnea\ counts}{Total\ sleep\ time} \quad (5)$$

$$AHI_{seq} = \frac{Apnea\ counts + Hypopnea\ counts}{Sequence\ duration} \quad (6)$$

$$ODI_{sub} = \frac{\text{Oxygen desaturation counts}}{\text{Total sleep time}} \quad (7)$$

$$ODI_{seq} = \frac{\text{Oxygen desaturation counts}}{\text{Sequence duration}} \quad (8)$$

ODI measures the frequency of reduced oxygen levels³⁵, which is used to stratify hypoxia severity, as in Eqs. (7–8). ODI_{sub} and AHI_{sub} are positively correlated but not interchangeable, as shown by the red arrow in Fig. 4. RLS_{sub} and $Snoring_{sub}$ were defined similarly.

Since we adopted an apex-centric tokenization method, which may have significant overlap between epochs, an event counting procedure was necessary to derive AHI_{sub} or AHI_{seq} . For a sequence with 30 estimated labels, we identified the epochs labeled as OSA and hypopnea. We then evaluated the distances between the apex of these events and defined a 60-second ‘dead-response’ time. An apex fell within this ‘dead-response’ zone was absorbed by the prior apex. This process is repeated until all the intervals are higher than 60s, as illustrated in Supplementary Fig. 2. The threshold was chosen by optimizing the consistency with the reference counts.

Data partitioning

Due to misalignment in PSG and actigraphy data, 16 challenging-to-align subjects were excluded from training and used as testing data. For the remaining 42 subjects:

First-Tier Training: To prevent data leakage from overlapping epochs, each training iteration randomly selected a single epoch from each overlapping set. Adjacent epochs were excluded to prevent data leakage. This produced 592 OSA, 638 hypopnea, 804 RLS, and 1518 snoring events per iteration. We balanced categories by resampling to match the least frequent class and split data 80/10/10 for training, validation, and testing. By randomly selecting a single epoch from each cluster of overlapping epochs during each training iteration, we enhanced training diversity while effectively controlling overfitting.

Second-Tier Training: Data was sorted into three AHI_{seq} categories: <10, 10–30, and >30 events/h. An equal number of sequences was selected from each category, with the same 80/10/10 split for training, validation and testing. The remaining sequences were allocated to the testing set. Since the sequences were non-overlapping to begin with, no additional screening was required. This division helped maintain a large training pool while minimizing data contamination.

The training process was repeated 20 times to ensure robust evaluation.

Performance evaluation

Median absolute error (MAE) and correlation evaluation MAE was used to evaluate AHI_{seq} and ODI_{seq} estimation performance. Pearson correlation coefficients assessed the similarity between reference and estimated values.

Classification evaluation We used clinical standards to classify the subjects into normal, mild, moderate, and severe OSA patients. For our analysis, we also employed a simplified classification, dividing the subjects into two groups: normal/mild and moderate/severe. The first group (normal/mild) does not require clinical intervention, while the second group (moderate/severe) does.

To evaluate the performance of our classification model, we employed sensitivity and specificity, as defined in Eqs. 9–10. For event-level or subject-level classification, TP (True Positives) is the number of correct identifications of an event or subject-level OSA severity. FN (False Negatives) is the number of missed detections of non-event or opposite labels. FP (False Positives) is the number of incorrect detections of an event or label, and TN (True Negatives) is the number of correct non-event or opposite label detections.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (10)$$

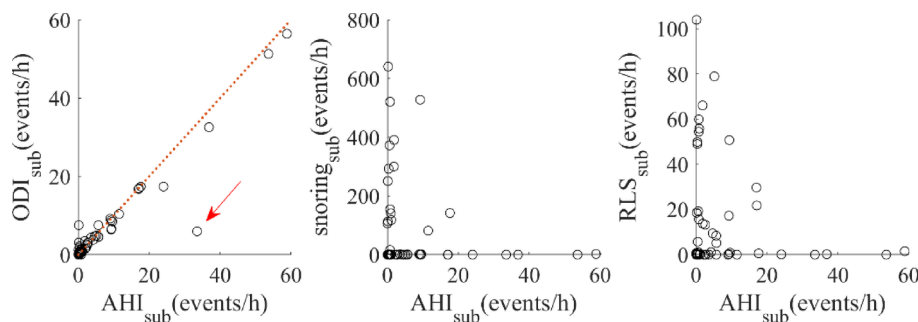


Fig. 4. Correlation of AHI_{sub} with (A) ODI_{sub} , (B) $snoring_{sub}$ and (C) RLS_{sub} .

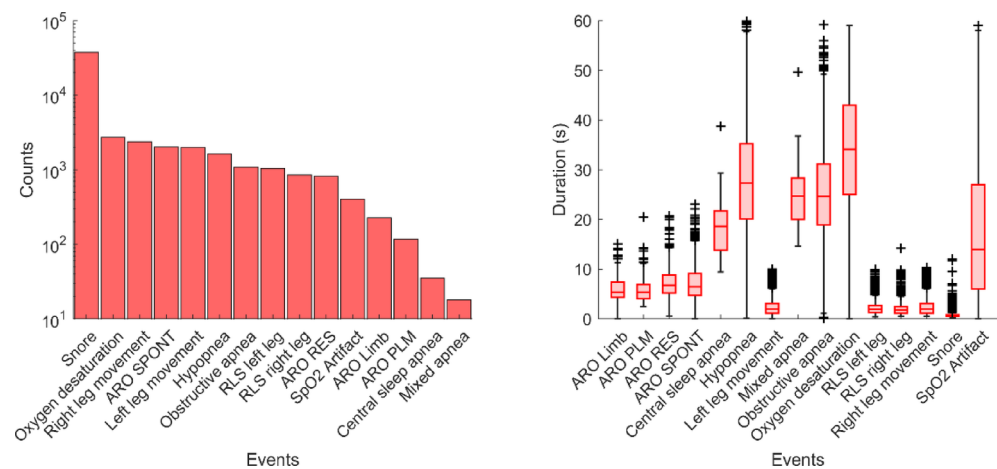


Fig. 5. Event distribution and durations.

	Group I	Group II
Device	Geneactiv	Axivity
Age (years)	39.7 ± 14.2	28.3 ± 17.6
Sex (M/F)	15/13	15/15
Total sleep time (hours)	6.6 ± 2.7	8.4 ± 1.9
Recording time (hours)	7.9 ± 2.4	8.8 ± 1.1
AHI _{sub} (events/h)	8.3 ± 15.0	4.5 ± 8.6
ODI _{sub} (events/h)	6.7 ± 13.2	4.6 ± 7.4
Snoring _{sub} (events/h)	3.1 ± 14.6	131.9 ± 187.4
RLS _{sub} (events/h)	11.1 ± 25.1	13.8 ± 21.4

Table 1. Subject characteristics and sleep-related parameters (mean ± SD).

	OSA	Hypopnea	Oxygen desaturation	RLS	Snoring
Event number	1077	1630	2730	1894	37,327
Duration (s)	25.5 ± 9.1	28.2 ± 11.0	33.5 ± 13.4	2.2 ± 1.5	0.7 ± 0.4

Table 2. Event counts and durations (mean ± SD).

For multi-class classification, the overall sensitivity and specificity are the weighted sum of the class-wise evaluations. We also used Cohen's kappa for consistency test.

These metrics were analyzed on a per-event, per-measurement, and per-subject basis to address their different application scenarios. For decisions regarding medical intervention, a per-subject analysis is sufficient. For a more detailed analysis of event distribution, a per-event analysis is necessary. The per-measurement analysis can compare the performance of the dominant and non-dominant hands and help in designing a more effective protocol.

Influence of handedness, co-existing diseases, sleep stages and postures In many clinical studies, the wrist watches were worn on the dominant hand. To evaluate the effect of handedness, we compared the AHI_{sub} and ODI_{sub} estimation accuracy between the dominant and non-dominant hands.

We also analyzed the impact of co-existing conditions such as RLS and snoring. Additionally, we examined variations in sleep stages (N1, N2, N3, and REM) and body postures (supine, left-side, and right-side), which account for 93.8% of all observed postures. These factors were essential for a comprehensive understanding of the data.

Results and discussion

Patient characterization

The cohort's characteristics, sleep durations, and event distribution are shown in Fig. 5; Tables 1 and 2. Sleep patterns are heterogeneous, with various events including spontaneous awakenings (ARO SPONT), respiratory-related awakenings (ARO RES), limb movement-related awakenings (ARO Limb), periodic leg movement-

related awakenings (ARO PLM), blood oxygenation (SpO_2) artifacts, central sleep apnea (CSA), and mixed apnea. CSA occurs when the brain fails to send proper signals to breathing muscles, while mixed apnea combines elements of OSA and CSA.

Snoring happens most frequently, followed by oxygen desaturation. OSA incidents last an average of 20 s, while oxygen desaturation events span about 30 s. This distribution allowed our apex-centered tokenization to achieve complete or near-complete encapsulation of OSA events. Among the study cohort, 41 subjects were normal ($\text{AHI}_{\text{sub}} < 5$), with 9 subjects classified as having mild OSA ($5 \leq \text{AHI}_{\text{sub}} < 15$), four subjects had moderate OSA ($15 \leq \text{AHI}_{\text{sub}} < 30$), and four subjects had severe OSA ($\text{AHI}_{\text{sub}} \geq 30$).

Comparison of RIP, nasal flow and processed actigraphy data

Our investigation revealed that transforming raw triaxial accelerometer data through PCA to IMFs effectively retained the subtle vibrations induced by respiratory activity. Simultaneous examination of signals during OSA episodes demonstrated that the chosen IMF closely aligns with the thoracic movements, as shown in Fig. 6. The signal comparison during RLS, snoring and hypopnea are shown in the Supplementary Fig. 3.

Apnea severity estimation

The distribution of AHI_{seq} and AHI_{sub} is shown in Supplementary Fig. 4, with AHI_{sub} being the weighted sum of AHI_{seq} for each subject.

Oxygen desaturation can result from both hypopneas and OSA. Our event classification system effectively identifies OSA-related oxygen desaturation events, achieving a per-event sensitivity of 67.4% and specificity of 92.5%.

On a per-measurement level, the Pearson correlation coefficient between the estimated and reference AHI_{sub} is 0.87, and for ODI_{sub} , it is 0.89. The red circle with dashed line in Fig. 7A and B represents measurements when subjects have high AHI but low ODI, as illustrated in Fig. 4. These results suggest that the system is less sensitive to hypopneas without accompanying oxygen desaturation. The comparison of reference and estimated AHI_{sub} , ODI_{sub} , and RLS_{sub} is shown in Fig. 7. However, snoring detection showed a lack of correlation with reference data, which is presented in Supplementary Fig. 5.

On a per-subject level, we used left-hand data due to its better performance. Given that AHI_{sub} was underestimated at higher values (Fig. 7A), we stratified $\text{AHI}_{\text{sub-est}}$ using thresholds of 5, 10, and 15, similar to¹⁵. For the standard four-class apnea severity (normal, mild, moderate, severe), the Cohen's kappa for AHI_{sub} and ODI_{sub} on the left hand was 0.68 and 0.54, respectively (Table 3). When simplified into two categories (normal/mild and moderate/severe), Cohen's kappa improved significantly. For ODI_{sub} , it increased to 0.84. The sensitivity and specificity for moderate/severe OSA detection were 85.7% and 98.1%, respectively.

This result is reflected in the confusion matrix (Fig. 8), where the algorithm nearly perfectly distinguishes between conditions requiring intervention (moderate/severe) and those that do not. Notably, the algorithm effectively distinguishes cases with high AHI but relatively lower ODI, highlighting its ability to detect subtle clinical nuances of diagnostic information. Only one subject with moderate OSA was misclassified as normal, who had 58 arousals and 7 central apneas during sleep, indicating potential signal degradation. Other comorbidities, such as snoring or RLS, slightly reduce quantification accuracy but do not significantly impact diagnostic value, as illustrated in Fig. 9.

Notably, while right-hand data performed less optimally for four-class classification, it achieved similar results when using the two-class system.

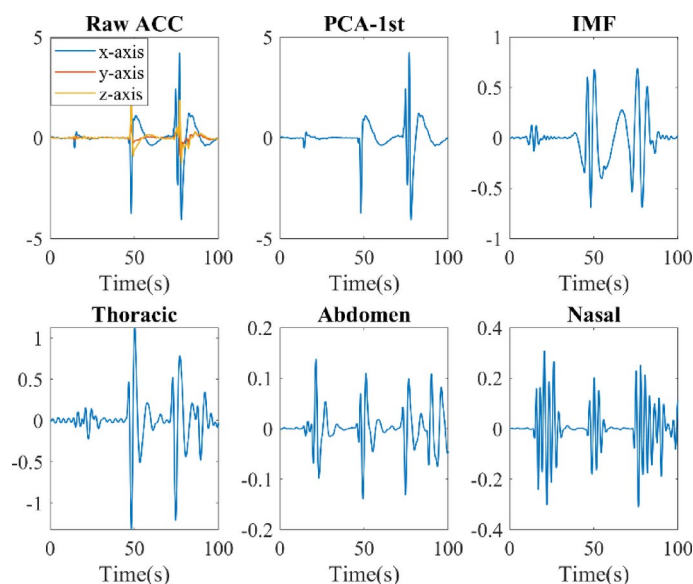


Fig. 6. Concurrent signals and processed data during OSA events. First row: the raw data, the first principal component and the used IMF; Second row: experimental thoracic, abdomen and nasal flow signal.

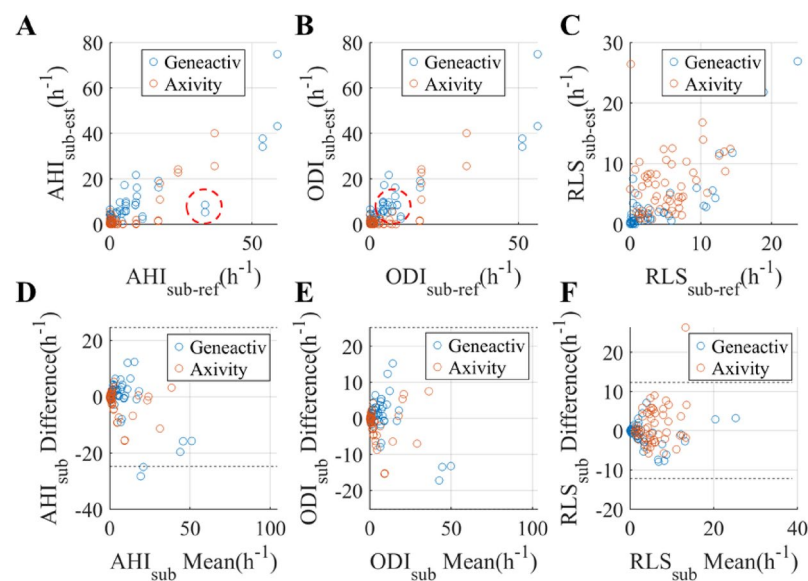


Fig. 7. (A–C) correlation between reference AHI_{sub} , ODI_{sub} and RLS_{sub} versus their estimations; (D–F) Corresponding Bland–Altman plot. The limits of agreement were drawn as dotted line (1.96 SD).

	Four classes		Two classes	
	AHI_{sub}	ODI_{sub}	AHI_{sub}	ODI_{sub}
Left-hand	0.68	0.54	0.77	0.84
Right-hand	0.59	0.50	0.77	0.84

Table 3. Cohen’s kappa between estimated and expected AHI_{sub} and ODI_{sub} .

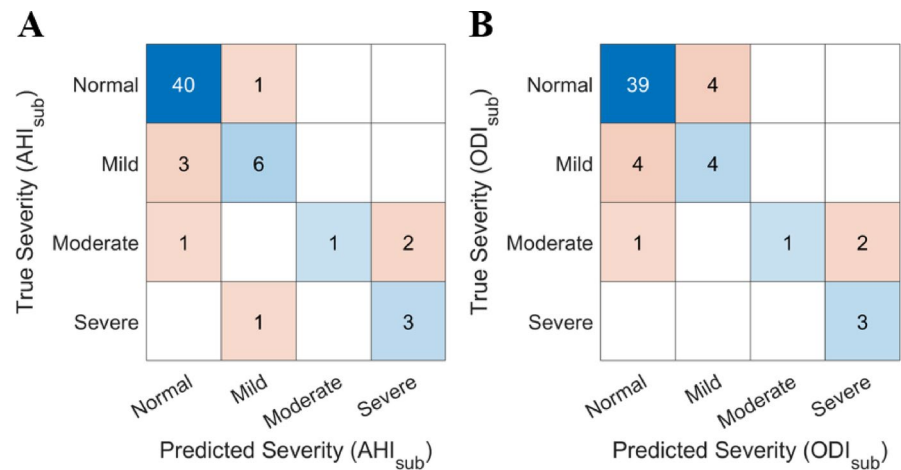


Fig. 8. OSA severity comparison, if the classification is made based on (A) AHI_{sub} (B) ODI_{sub} .

Among the 16 subjects reserved for testing, their performance is shown in Supplementary Fig. 6. Using four AHI_{sub} classes, 6 out of 16 were misclassified, primarily normal and mild OSA patients. When using two-category system, only 1 subject was misclassified.

Influencing factors

Handedness and posture

OSA detection performance was enhanced when using left-hand data or when the individual adopted a left-side sleeping posture, as shown in Table 3; Fig. 9. Handedness did not significantly influence outcomes, as both left-handed and right-handed individuals achieved optimal results with their left hand. One possible explanation is that, left-side positioning was associated with a ~50% higher average AHI_{seq} compared to right-side sleeping,

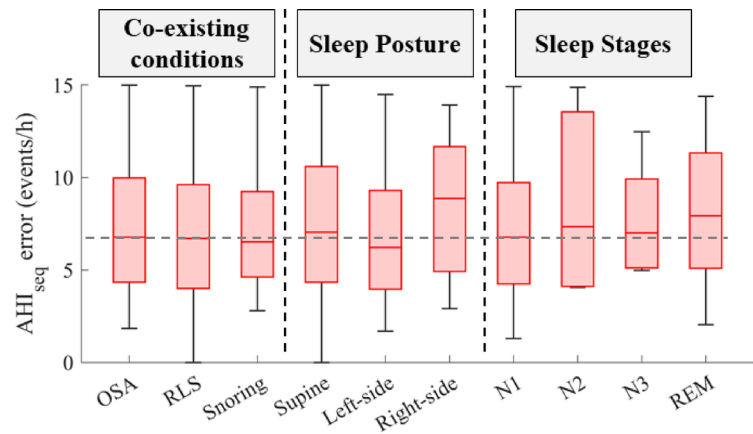


Fig. 9. For sequences with OSA events, AHI_{seq} estimation error under various influences.

	Sensitivity (%)	Specificity (%)	Correlation with ODI_{sub} (Cohen's kappa)
Even-spaced Tokenization (BiLSTM)	38.2	74.0	0.16(0.10)
Even-spaced Tokenization (BiLSTM-MCHA)	13.0	97.2	0.02(0.01)
Apex-centric tokenization (BiLSTM)	52.0	73.6	0.21(0.20)
Apex-centric tokenization (BiLSTM-MCHA)	67.4	92.5	0.74(0.67)

Table 4. Performance enhancement by proper tokenization and multi-stage structure.

consistent with reports that left decubitus may promote airway collapse in OSA patients³⁶. We speculate that left-side episodes generate more characteristic signals, potentially reflecting more severe airway obstructions.

Influence of sleep stages and co-existing conditions on OSA detection

In the absence of RLS or snoring, the MAE of AHI_{seq} for sequences with OSA ($AHI_{seq} > 0$) was 7.37 events/h, with a standard deviation of 11.36 events/h, lower than the 14.74 events/h reported standard deviation in a previous study¹⁵. The presence of RLS and snoring did not negatively impact AHI_{seq} accuracy (Fig. 9).

AHI_{seq} estimation shows stage-dependent variability: In REM sleep, accuracy decreases, probably due to motion artifacts from wrist movements and irregular breathing patterns, compounded by reduced respiratory muscle activity that weakens motion signals^{37,38}. During N2 sleep, diminished $PaCO_2$ chemosensitivity results in mild hypoventilation and less distinct post-apneic ventilation patterns, increasing AHI_{seq} variability³⁹. Apnea rates rose moderately in N3 (30% increase) and substantially in REM (66% increase), with ODI_{seq} following similar trends.

It is worth noting that the for sequences without OSA events ($AHI_{seq} = 0$), the mean AHI_{seq} error is 1.67 events/h, with a median of 0.

Influence of device brands

In this study, subjects were matched across groups by sample size and gender. Preliminary interviews also indicated comparable self-reported sleep apnea risk levels in daily life. However, during actual testing, we observed that subjects wearing Axivity devices exhibited significantly more snoring events, indicating that before objective testing, individuals may misinterpret severe snoring as OSA.

Both devices demonstrated comparable performance in AHI_{seq} estimation, as confirmed by a two-sample t-test (Geneactiv: 2.64 events/hour vs. Axivity: 2.42 events/hour; $p > 0.05$), indicating that both are equally suitable for OSA detection and AHI/ODI estimation.

Ablation study

We found that the apex-centric tokenization method enhanced OSA event detection sensitivity while maintaining specificity. The MHCA module significantly improved both sensitivity and specificity by optimizing physiological sequential predictions, but only when used with apex-centric tokenization.

Interestingly, adding the MHCA mechanism decreased the overall performance when even-spaced tokenization was used, likely due to low sensitivity to OSA events (Table 4). In contrast, apex-centric tokenization retained most potential events, resulting in higher sensitivity. The sensitivity and specificity comparison in Table 4 was based on a per-event evaluation.

Algorithm	Sensor type	Evaluation	Sensitivity (%)	Specificity (%)	Correlation with AHI/ODI (Cohen's kappa)
CNN-LSTM ¹⁸	Electrocardiogram	Per event	95	82	0.74, AHI _{sub} (NA)
Discriminant analysis ¹⁵	Electrocardiogram	Per event (heterogeneous)	91.9	52.5	0.72, AHI _{sub} (0.53)
Decision tree ⁴⁰	Pressure sensor	Per event	74	74	NA
U-Net ¹⁹	Pressure sensor	Per event	76.4	98.5	0.89, AHI _{sub} (NA)
Random forest ¹¹	Accelerometer	Per event	96.3	96.7	0.88, AHI _{sub} (NA)
BiLSTM-MCHA	Accelerometer	Per event (heterogeneous)	67.4	92.5	0.74, ODI _{sub} (0.67)
BiLSTM-MCHA	Accelerometer	Per subject (heterogeneous)	85.7	98.1	0.89, ODI _{sub} (0.84)

Table 5. Comparison with previous studies with advanced machine-learning with/without heterogeneous population.

Comparison with previous studies

Few studies have used wrist actigraphy for apnea detection, so we compared our work with similar sensor types (Table 5). Our multi-stage approach outperformed previous studies in a ‘real-world’ population¹⁵. While our per-event classification in a heterogeneous population had slightly lower performance than studies using simpler cohorts, our per-subject evaluation showed better or comparable sensitivity and specificity for moderate to severe OSA.

Traditional actigraphy devices like Geneactiv and Axivity have lower resolution (3.9 mg) compared to newer sensors (<0.1 mg), making respiratory detection challenging. Our algorithm was developed to work with these devices due to their retrospective research value. There is a wealth of data available from these actigraphy, such as the UK Biobank with over 100,000 participants who had 7-day actigraphy data⁴¹, which can provide valuable clinical insights if well-utilized.

Novelties and implications

Several key technological innovations distinguished this study:

- (1) ODI estimation using only motion sensors: AHI is commonly used for OSA diagnosis, but ODI may be more relevant for long-term complications and prognosis. Wrist actigraphy lacks optical sensors for direct oxygen saturation measurement. However, we successfully inferred ODI using accelerometer data, advancing the use of this low-cost, power-efficient technology.
- (2) Apex-centric tokenization Method: We designed a novel apex-centric tokenization method based on apex detection and clustering. This method addressed the issue of incomplete event coverage and enhanced the sensitivity to potential events. By strategically sampling unstable region and use physiological reasoning, we forced the MHCA to focus on meaningful temporal patterns. This resulted in better specificity, while maintaining good sensitivity.
- (3) Multi-stage approach with causal attention mechanism: This design integrates the strengths of BiLSTM and MHCA. The causal attention emphasizes physical plausibility at different levels of granularity, significantly reducing the impact of noise and motion artifacts caused by disturbances such as RLS and snoring.
- (4) Comprehensive evaluation in practical application: Using novel bilateral experimental design, we demonstrated that wearing the actigraphy device on the left hand led to improved performance. RLS and snoring was shown to have minimal impact on OSA detection. However, the accuracy of OSA detection varies with OSA severity, which correlates with different sleep stages and postures. These factors should be considered in clinical practice.

The findings of this study showed that actigraphy could be a practical alternative to PSG for OSA screening, which is robust against disturbances from RLS and snoring.

Although our algorithm produced promising results, several limitations need to be addressed. The low resolution of the accelerometer sensors limits the reliable quantification of mild OSA. In addition, the duration of each event cannot be accurately determined. Our method is more effective for detecting apnea with oxygen desaturation than pure apnea. However, most subjects had mental comorbidities, and since the algorithm does not rely on biometric information, this limitation does not significantly affect the results.

Conclusion

In conclusion, our innovative combination of apex-centric tokenization and MHCA demonstrated the feasibility of using low-resolution wrist actigraphy to detect moderate to severe OSA with high accuracy, even in the presence of snoring and RLS. This method shows significant potential for large-scale public health studies.

Data availability

De-identified data will be provided upon reasonable request. Please send the request to xingxm@sibet.ac.cn.

Received: 24 October 2024; Accepted: 6 May 2025

Published online: 22 May 2025

References

1. Park, J. G., Ramar, K. & Olson, E. J. Updates on definition, consequences, and management of obstructive sleep apnea. *Mayo Clin. Proc.* **86**(6), 549–554 (2011).
2. Lechat, B. et al. Multinight prevalence, variability, and diagnostic misclassification of obstructive sleep apnea. *Am. J. Respir. Crit. Care Med.* **205**(5), 563–569 (2022).
3. Knauer, M. et al. Clinical consequences and economic costs of untreated obstructive sleep apnea syndrome. *World J. Otorhinolaryngol. Head Neck Surg.* **1**(1), 17–27 (2015).
4. Suzuki, M. et al. Comparison of diagnostic reliability of out-of-center sleep tests for obstructive sleep apnea between adults and children. *Int. J. Pediatr. Otorhinolaryngol.* **94**, 54–58 (2017).
5. Young, T. et al. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep* **20**(9), 705–706 (1997).
6. Balagny, P. et al. Prevalence, treatment and determinants of obstructive sleep Apnoea and its symptoms in a population-based French cohort. *ERJ Open. Res.* **9**(3) (2023).
7. Bin, K. J. et al. Digital platform for continuous monitoring of patients using a Smartwatch: longitudinal prospective cohort study. *JMIR Form. Res.* **7**, e47388 (2023).
8. Hernandez, J., McDuff, D. & Picard, R. W. *Biowatch: Estimation of heart and breathing rates from wrist motions*. in *9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)* (2015).
9. Sun, X. et al. SleepMonitor: monitoring respiratory rate and body position during sleep using smartwatch. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **1**(3), 104 (2017).
10. Leube, J. et al. Reconstruction of the respiratory signal through ECG and wrist accelerometer data. *Sci. Rep.* **10**(1), 14530 (2020).
11. Chen, X. et al. ApneaDetector: detecting sleep apnea with smartwatches. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **5**(2), 59 (2021).
12. Zhao, L. et al. Robust respiratory rate monitoring using smartwatch photoplethysmography. *IEEE Internet Things J.* **10**(6), 4830–4844 (2023).
13. Schipper, F. et al. A deep-learning approach to assess respiratory effort with a chest-worn accelerometer during sleep. *Biomed. Signal Process. Control* **83**, 104726 (2023).
14. Liaqat, D. et al. WearBreathing: real world respiratory rate monitoring using smartwatches. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **3**(2), 56 (2019).
15. Papini, G. B. et al. Estimation of the apnea-hypopnea index in a heterogeneous sleep-disordered population using optimised cardiovascular features. *Sci. Rep.* **9**(1), 17448 (2019).
16. Ayashm, S., Chehel Amirani, M. & Valizadeh, M. Analysis of ECG signal by using an FCN network for automatic diagnosis of obstructive sleep apnea. *Syst. Signal. Process.* **41**(11), 6411–6426 (2022).
17. Li, R. et al. Convolutional neural network for screening of obstructive sleep apnea using snoring sounds. *Biomed. Signal Process. Control* **86**, 104966 (2023).
18. Srivastava, G. et al. ApneaNet: A hybrid 1DCNN-LSTM architecture for detection of obstructive sleep apnea using digitized ECG signals. *Biomed. Signal Process. Control* **84**, 104754 (2023).
19. Chen, Y. et al. Contactless screening for sleep apnea with breathing vibration signals based on modified U-Net. *Sleep Med.* **107**, 187–195 (2023).
20. Almarshad, M. A. et al. Adoption of transformer neural network to improve the diagnostic performance of oximetry for obstructive sleep apnea. *Sensors (Basel)* **23**(18) (2023).
21. Mehra, R. et al. Prevalence and correlates of sleep-disordered breathing in older men: osteoporotic fractures in men sleep study. *J. Am. Geriatr. Soc.* **55**(9), 1356–1364 (2007).
22. Ohayon, M. M., O'Hara, R. & Vitiello, M. V. Epidemiology of restless legs syndrome: A synthesis of the literature. *Sleep Med. Rev.* **16**(4), 283–295 (2012).
23. Fry, J. M., DiPhillipo, M. A. & Pressman, M. R. Periodic leg movements in sleep following treatment of obstructive sleep apnea with nasal continuous positive airway pressure. *Chest* **96**(1), 89–91 (1989).
24. Athavale, Y. et al. Advanced signal analysis for the detection of periodic limb movements from bilateral ankle actigraphy. *J. Sleep. Res.* **26**(1), 14–20 (2017).
25. Gschliesser, V. et al. PLM detection by actigraphy compared to polysomnography: A validation and comparison of two actigraphs. *Sleep. Med.* **10**(3), 306–311 (2009).
26. Plekhanova, T. et al. Equivalency of sleep estimates: Comparison of three research-grade accelerometers. *J. Meas. Phys. Behav.* **3**(4), 294–303 (2020).
27. Sudlow, C. et al. UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**(3), e1001779 (2015).
28. Muñoz-Gutiérrez, P. A. et al. Localization of active brain sources from EEG signals using empirical mode decomposition: A comparative study. *Front. Integr. Neurosci.* **12** (2018).
29. Chen, Y. C. & Hsiao, T. C. *Towards Estimation of Respiratory Muscle Effort with Respiratory Inductance Plethysmography Signals and Complementary Ensemble Empirical Mode Decomposition* vol. 56, 1293–1303 (Medical & Biological Engineering & Computing, 2018).
30. Auger, F. et al. Time-frequency reassignment and synchrosqueezing: An overview. *IEEE. Signal. Process. Mag.* **30**(6), 32–41 (2013).
31. Huang, C. Y. & Yu, C. C. Different diagnostic criteria for periodic leg movements in patients with obstructive sleep apnea after continuous positive airway pressure Titration. *Neuropsychiatr Dis. Treat.* **15**, 2129–2136 (2019).
32. Albanese, A. et al. An integrated mathematical model of the human cardiopulmonary system: Model development. *Am. J. Physiol. Heart Circ. Physiol.* **310**(7), H899–H921 (2015).
33. Veasey Sigrid, C. & Ilene, M. R. Obstructive sleep apnea in adults. *N. Engl. J. Med.* **380**(15), 1442–1449 (2019).
34. Malhotra, A. et al. Metrics of sleep apnea severity: Beyond the apnea-hypopnea index. *Sleep* **44**(7) (2021).
35. Rashid, N. H. et al. The value of oxygen desaturation index for diagnosing obstructive sleep apnea: A systematic review. *Laryngoscope* **131**(2), 440–447 (2021).
36. Ozeke, O. et al. Influence of the right- versus left-sided sleeping position on the apnea-hypopnea index in patients with sleep apnea. *Sleep. Breath.* **16**(3), 617–620 (2012).
37. Morrell, M. J., Browne, H. A. K. & Adams, L. Respiratory response inspiratory resistive load. *Dur. Rapid Eye Mov. Sleep. Hum.* **526**(1), 195–202 (2000).
38. Gutierrez, G. et al. Respiratory rate variability in sleeping adults without obstructive sleep apnea. *Physiol. Rep.* **4** (17), e12949 (2016).
39. Sowho, M. et al. Sleep and respiratory physiology in adults. *Clin. Chest. Med.* **35**(3), 469–481 (2014).
40. Weidong, G. et al. Obstructive sleep apnea syndrome detection based on ballistocardiogram via machine learning approach. *Math. Biosci. Eng.* **16**(5), 5672–5686 (2019).

41. Doherty, A. et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. *PLoS One* **12**(2), e0169649 (2017).

Author contributions

X.X. wrote the main manuscript text. S.A. organized the clinical study. J.Z., R.H., Y.L., and D.Q. assisted with data collection, deposition, and analysis. J.M. collected the PSG data. G.W., J.X., and Y.Z. contributed to revising the manuscript. H.F. and W.D. conceptualized the study and secured funding for the project. All authors reviewed the manuscript.

Funding

This work was supported by the National Key R&D Program of China (2022YFC3601000), which provided funding for data collection. This work was also supported by the Youth Innovation Promotion Association CAS (2021323), which helped analysis and interpretation of data.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-01430-3>.

Correspondence and requests for materials should be addressed to H.F. or W.-f.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025