

Advancing sleep health equity through deep learning on large-scale nocturnal respiratory signals

Received: 7 January 2025

Accepted: 12 September 2025

Published online: 22 October 2025

 Check for updates

Zhongxu Zhuang^{1,12}, Biao Xue^{1,12}, Qiang An^{2,12}, Hui Chu¹, Yue Zhang³, Rui Chen⁴, Jing Xu⁵, Ning Ding⁶, Xiaochuan Cui⁷, E. Wang⁸, Meilin Wang⁹, Junyi Xin¹⁰, Xuan Yang¹, Yan Xu¹, Yaxian Li¹, Chang-Hong Fu¹, Xiaohua Zhu¹, Mugen Peng¹¹✉ & Hong Hong¹²✉

Sleep disorders affect billions globally, yet diagnostic access remains limited by healthcare resource constraints. Here, we develop a deep learning framework that analyzes respiratory signals for remote sleep health monitoring, trained on 15,785 nights of data across diverse populations. Our approach achieves robust performance in four-stage sleep classification (82.13% accuracy on internal validation; 79.62% on external validation) and apnea-hypopnea index estimation (intraclass correlation coefficients 0.90 and 0.94, respectively). Through transfer learning, we adapt the model to radar-derived respiratory signals, enabling contactless monitoring in home environments. The framework demonstrates consistent performance across demographic subgroups, supports real-time processing through self-supervised learning techniques, and integrates with a remote sleep health management platform for clinical deployment. This approach bridges critical gaps in sleep healthcare accessibility, supporting population-level screening and monitoring, paving the way for scalable sleep healthcare, and advancing sleep health equity.

Healthy sleep is essential for maintaining physical and mental well-being, enhancing cognitive performance, and preventing chronic diseases. However, sleep health remains underrepresented in public health agendas¹ and is often underestimated as a global health issue². Common sleep problems—such as insomnia, insufficient sleep, and sleep-disordered breathing—are linked to poor mental health³,

cardiovascular conditions like hypertension, and respiratory diseases such as chronic obstructive pulmonary disease (COPD)⁴. Despite its importance, public awareness remains low. Surveys show that a majority of individuals with sleep problems remain undiagnosed⁵ and unaware of related health risks⁶. Modern lifestyles and urban stress have also increased sleep issues among younger populations, with

¹School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China. ²Department of Biomedical Engineering, Fourth Military Medical University, Xi'an, China. ³School of Electronic and Optical Engineering, Nanjing University of Science and Technology Zijin College, Nanjing, China. ⁴Department of Respiratory and Critical Care Medicine, Sleep Center, The Second Affiliated Hospital of Soochow University, Suzhou, China. ⁵Department of Respiratory and Critical Care Medicine, The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University, Huai'an, China. ⁶Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital with Nanjing Medical University, Nanjing, China. ⁷Department of General Medicine and Sleep Medicine Center, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China. ⁸Department of Anesthesiology, Xiangya Hospital, Central South University, Changsha, China. ⁹Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, School of Public Health, Nanjing Medical University, Nanjing, China. ¹⁰Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China. ¹¹The State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. ¹²These authors contributed equally: Zhongxu Zhuang, Biao Xue, Qiang An. ✉ e-mail: pmg@bupt.edu.cn; hongnju@njnu.edu.cn

global adolescent prevalence of sleep disorders reaching nearly 10%¹. Beyond individual health, sleep-related productivity losses and long-term complications pose a significant socioeconomic burden⁷. Current sleep diagnostics still rely heavily on resource-intensive laboratory assessments, highlighting the need for scalable, accessible solutions to promote global sleep health equity.

The allocation of sleep medical resources is a key factor contributing to the aforementioned challenges. Sleep centers and diagnostic facilities are often concentrated in urban areas, leaving individuals in rural or underserved regions with limited access to essential services⁸. For instance, only 2% of primary healthcare institutions in China offer sleep-related services⁹, mainly due to shortages in specialized personnel (76%) and high equipment costs (70%). Meanwhile, the current gold standard for sleep evaluation—polysomnography (PSG)¹⁰—requires multi-channel recordings of electrophysiological and cardiorespiratory signals, making it costly, complex, and unsuitable for long-term home monitoring. Although home sleep apnea tests (HSAT) offer improved accessibility¹¹, even Type III devices require multiple contact-based sensors¹², compromising comfort and limiting user compliance. Wearable devices such as smartwatches have gained attention for home monitoring using electrocardiography (ECG)^{13,14} and photoplethysmography (PPG) signals^{14,15}. However, sensor discomfort, motion artifacts, battery limitations, and incomplete data remain significant issues. Recently, ring oximeters and actigraphy-based devices have been widely explored for sleep assessment, but challenges related to sampling rates, battery life, and limited specificity continue to restrict their clinical applicability. Moreover, although wearables may induce less psychological stress than PSG, the repeated setup required for multi-night use often leads to reduced compliance and data reliability. These challenges underscore the need for both: (1) equitable access to sleep care and (2) advanced, non-intrusive monitoring technologies that ensure scalability, usability, and data reliability.

With the growing adoption of telemedicine, digital health technologies are revolutionizing the delivery of healthcare. While telemonitoring has proven effective in chronic disease management¹⁶ (e.g., cardiovascular conditions), remote sleep health management remains underdeveloped. Integrating artificial intelligence and the internet of things (IoT) with novel sensing modalities holds great potential to enhance sleep disorder screening, personalized interventions, and long-term monitoring, extending beyond traditional clinical boundaries. Among emerging technologies, radar-based non-contact sensing offers a promising solution. These systems accurately capture thoracoabdominal movement signals during sleep without requiring physical contact, thereby eliminating the discomfort and compliance issues associated with wearable devices and PSG. Their penetrability and anti-interference capabilities enable natural sleep while maintaining signal fidelity. When integrated into remote health platforms, such radar systems can support continuous, unobtrusive sleep monitoring and real-time data transmission. This approach advances sleep healthcare delivery toward greater accessibility and equity, particularly benefiting populations with limited access to specialized facilities or long-term care services.

Respiratory activity is tightly coupled with sleep neurophysiology^{17,18}. During non-rapid eye movement (NREM) sleep, respiration slows and becomes more regular due to parasympathetic dominance, whereas rapid eye movement (REM) sleep features irregular breathing driven by sympathetic surges^{19–21}. These patterns result from the shift from wakefulness-driven to chemoreflex-driven control during sleep, reducing CO₂/O₂ responsiveness and ventilatory drive²². Recent evidence also highlights bidirectional interactions between breathing and brain oscillations^{23,24}, including hippocampal coordination of respiration-locked rhythms with memory-related sleep spindles²⁵. These distinctive respiratory signatures—arising from stage-specific autonomic and neural regulation—support the physiological

plausibility of inferring sleep stages from thoracoabdominal movement. Recent studies have demonstrated that respiratory signals correlate strongly with sleep architecture^{26–28}.

Radar-based sleep monitoring methods typically fall into two categories: (1) shallow-feature extraction followed by classical machine learning models^{29,30}, and (2) deep learning pipelines that directly infer sleep states from radar signals^{31–33}. However, most existing studies rely on limited datasets (fewer than 100 nights), and radar-specific sleep datasets remain scarce, which restricts model generalizability and clinical utility. Another limitation lies in the insufficient exploration of the physiological mapping between radar-sensed thoracoabdominal motion and sleep stages. Since many public PSG datasets³⁴ include thoracoabdominal motion signals that are broadly analogous to the respiratory motion captured by radar sensors, though signal characteristics may vary due to sensor modality and posture-related factors, leveraging these large-scale resources could help overcome radar data scarcity and enhance transferability. Furthermore, current research often treats sleep staging and sleep-disordered breathing (SDB) analysis as separate tasks, despite their physiological interdependence. A unified modeling approach would enable more comprehensive and clinically relevant monitoring, but remains underexplored in radar-based studies. In summary, challenges in data availability, signal interpretation, and task integration continue to limit the clinical translation of radar-based sleep monitoring. Bridging these gaps requires leveraging large-scale datasets, clarifying physiological mappings, and developing multi-task frameworks. Fully leveraging the diversity and breadth of public sleep datasets while integrating multi-dimensional evaluation methods provides a potential solution to address these challenges.

In this study, we present a high-precision, generalizable non-contact sleep monitoring framework leveraging large-scale thoracoabdominal respiratory signals and deep learning. Our model enables accurate sleep staging and apnea-hypopnea index (AHI) estimation, and has been validated across both public and multi-center clinical datasets, including 1103 nights from younger populations. We employ a multitask adversarial learning strategy for sleep staging and AHI estimation, and apply transfer learning to extend the framework to radar-derived signals for contactless sleep assessment. Robustness is further confirmed through subgroup analyses across age, sex, apnea severity, and comorbidities. We further introduce a self-supervised learning approach for real-time sleep staging, enabling efficient deployment without sleep labels. Finally, we integrate the model into a remote sleep health management platform, supporting accessible and scalable solutions for sleep disorder detection, management, and evaluation. The contributions of this work are reflected in three aspects: (1) a large-scale respiratory data-driven deep learning model for sleep staging and AHI estimation; (2) radar-based non-contact sleep monitoring via transfer learning; and (3) an integrated platform enabling scalable remote detection and management of sleep disorders. Together, our work demonstrates the feasibility of respiratory-based sleep monitoring, offering essential theoretical support and a practical foundation for promoting sleep health equity.

Results

Datasets and model training

In our study, four public datasets from the National Sleep Research Resource (NSRR)³⁴, namely the Sleep Heart Health Study (SHHS)³⁵, the Multi-Ethnic Study of Atherosclerosis (MESA)³⁶, the MrOS Sleep Study (MrOS)³⁷, and the Study of Osteoporotic Fractures (SOF)³⁸ were employed. Table 1 provides a summary of these datasets, including the number of nights, age groups, male proportions, AHI distribution, and disease conditions. Detailed descriptions of each dataset are presented in the Supplementary Notes. Although these public datasets encompass individuals with varying degrees of SDB severity and both sexes, they underrepresent individuals under 45. Meanwhile, recent

Table 1 | Characteristics of the datasets used in this study

Dataset	Nights	Age (%)				Male (%)	AHI (%)				Hypertension (%)	LD* (%)	PD (%)
		< 45	[45, 55]	[55, 65]	> =65		< 5	[5, 15]	[15, 30]	> =30			
SHHS1*	5736	5.9	17.6	30.6	45.9	47.8	17.3	38.3	27.3	17.1	30.7	42.7	/
SHHS2*	2638	0.1	11.6	28.9	59.4	46.2	17.4	37.1	27.0	18.5	44.2	6.7	0.5
MrOS1*	2848	0.0	0.0	0.0	100.0	100.0	9.3	33.6	32.1	25.0	49.8	5.1	1.1
MrOS2*	999	0.0	0.0	0.0	100.0	100.0	10.0	31.4	31.5	27.0	52.9	/	1.6
MESA*	2015	0.0	0.1	36.7	63.2	46.3	10.4	30.5	29.5	29.6	/	/	/
SOF*	446	0.0	0.0	0.0	100.0	0.0	18.6	38.6	28.5	14.3	59.9	12.1	/
ClinSuZhou*	458	55.5	17.5	12.9	13.8	80.3	27.1	17.2	16.8	38.9	24.7	0.4	11.1
ClinHuaiAn [#]	424	56.1	21.9	14.9	6.8	86.8	11.3	16.3	16.5	55.9	27.1	/	/
ClinRadar [#]	221	62.0	19.5	11.8	6.3	81.4	24.9	19.0	18.6	37.6	21.3	0.9	6.3

*Internal Dataset, *External Dataset.

[#]LD Lung diseases, including COPD chronic bronchitis, asthma, and emphysema. PD Parkinson's Disease.

studies have revealed a growing prevalence of SDB among younger and middle-aged populations (18–45)^{39–42}, with strong associations with cardiovascular and metabolic comorbidities^{43,44}. From a technical perspective, this age imbalance may limit model generalization. To develop a universally applicable sleep assessment model, we prospectively collected a multi-center clinical dataset covering over 1000 nights from younger adults (under 45 years), including both respiratory belt-derived respiratory signals (ClinSuZhou and ClinHuaiAn) and radar-derived respiratory signals (ClinRadar).

We developed a deep-learning model named ResSleepNet for automatic sleep staging and AHI estimation. Figure 1 illustrates the model's framework and the adversarial learning process on the internal dataset. Figure 2 offers a more detailed explanation of the model training and inference process used in this study. Phase I encompasses the pre-training process, during which we train the model on the internal dataset by minimizing the losses for sleep staging and AHI estimation while maximizing the loss for the domain discriminator. In phase II, the trained model weights are frozen, and performance evaluation is carried out on the external dataset. For details on data pre-processing and model training, validation, and testing, refer to the Methods section.

Sleep staging and AHI prediction

Table 2 presents the model's performance across different datasets, including overall accuracy in sleep staging, the sensitivity of predictions for each sleep stage, and the accuracy of AHI estimation. In the internal test datasets, the sleep staging task attained the highest average accuracy of 83.53% and Kappa of 0.73 in MrOS. Performance was somewhat attenuated yet remained acceptable in SHHS (80.34%, 0.70), MESA (83.09%, 0.73), SOF (78.57%, 0.68), and ClinSuZhou (81.72%, 0.71). In the external datasets, the ClinHuaiAn dataset achieved an average accuracy of 79.62% and Kappa of 0.67, while the radar dataset showed a decrease with an accuracy of 75.81% and Kappa of 0.62. Owing to differences in sensing methods, radar-based thoracoabdominal motion signals can slightly differ from respiratory belts, particularly during sleep posture changes. Although the frequency remains consistent, signal amplitude may vary, resulting in reduced performance when transferring a respiratory belt-based pre-trained model to radar data. Nevertheless, radar monitoring in this study still achieves comparable performance.

We systematically compared the correspondence between the sleep stages predicted by ResSleepNet and the actual sleep stages. Figure 3 and Supplementary Fig. 1 display the confusion matrices of sleep staging for each dataset, while Supplementary Table 1 further analyzes the sensitivity and specificity (as defined in Supplementary Table 2) distribution of predictions for different sleep stages across datasets. In the internal datasets, ResSleepNet performed satisfactorily

in recognizing wake, REM, and light sleep. However, its accuracy in detecting deep sleep was relatively lower, with frequent misclassifications of light sleep. The reasons are: (1) as shown in Supplementary Table 3, deep sleep makes up less than 15% of natural sleep, significantly lower than light sleep; (2) the similarity in cardiopulmonary features between deep and light sleep stages might further increase the difficulty of classification. To enhance the interpretability of the model, we visualized the intermediate decision-making process of the sleep staging prediction model in Supplementary Fig. 4. The results indicate that the model's focus differs noticeably across wake, REM, and light/deep stages, though confusion tends to occur between light and deep stages, consistent with the analysis presented earlier. Additionally, the sensitivity for detecting deep sleep varies among different datasets. For instance, it reaches 68.97% in SHHS but only 46.09% in MrOS. As shown in Supplementary Table 3, the proportion of deep sleep throughout the night is a crucial factor influencing detection performance. In MrOS, the proportion of deep sleep is merely 6.14%, significantly lower than the 11.93% in SHHS. Other datasets also exhibit this trend, where the detection performance for deep sleep correlates with its proportion in total sleep time. However, in ClinRadar, the sensitivity for the wake and REM stages decreased significantly. The confusion matrix in Fig. 3(i) reveals that 21.26% of wake and 28.23% of REM stages were misclassified as Light sleep, which is higher than in the internal test sets and ClinHuaiAn. Despite the ClinRadar dataset comprising over 200 nights of data, it remains relatively small compared to other datasets, contributing to the overall decline in accuracy.

In the AHI estimation task, the overall ICC of the internal test set was 0.90. With the exception of the SOF dataset, all other internal subsets achieved an ICC of 0.90 or higher, with ClinSuZhou attaining the highest value of 0.92. For the external datasets, the ICC values were 0.94 and 0.87, respectively. The middle two columns of Fig. 3 exhibit scatter plots and Bland-Altman plots that compare the true and predicted AHI values, demonstrating high consistency between the estimated and annotated AHI values. Supplementary Table 4 further analyzes other metrics used to evaluate AHI prediction performance and the results across different internal subsets. According to American Academy of Sleep Medicine (AASM) standards, the severity of SDB is classified based on AHI values into normal (AHI < 5), mild ($5 \leq \text{AHI} < 15$), moderate ($15 \leq \text{AHI} < 30$), and severe ($\text{AHI} \geq 30$). The four-class classification accuracies were 65.08% for the internal test set, 75.47% for ClinHuaiAn, and 63.80% for ClinRadar. The rightmost column of Fig. 3 presents the confusion matrix for the four severity categories. The results indicate that our model performs optimally in detecting severe SDB, achieving an accuracy of 91.14% in the ClinHuaiAn dataset. However, performance declines when identifying individuals with normal or moderate severity. The accuracy for detecting individuals

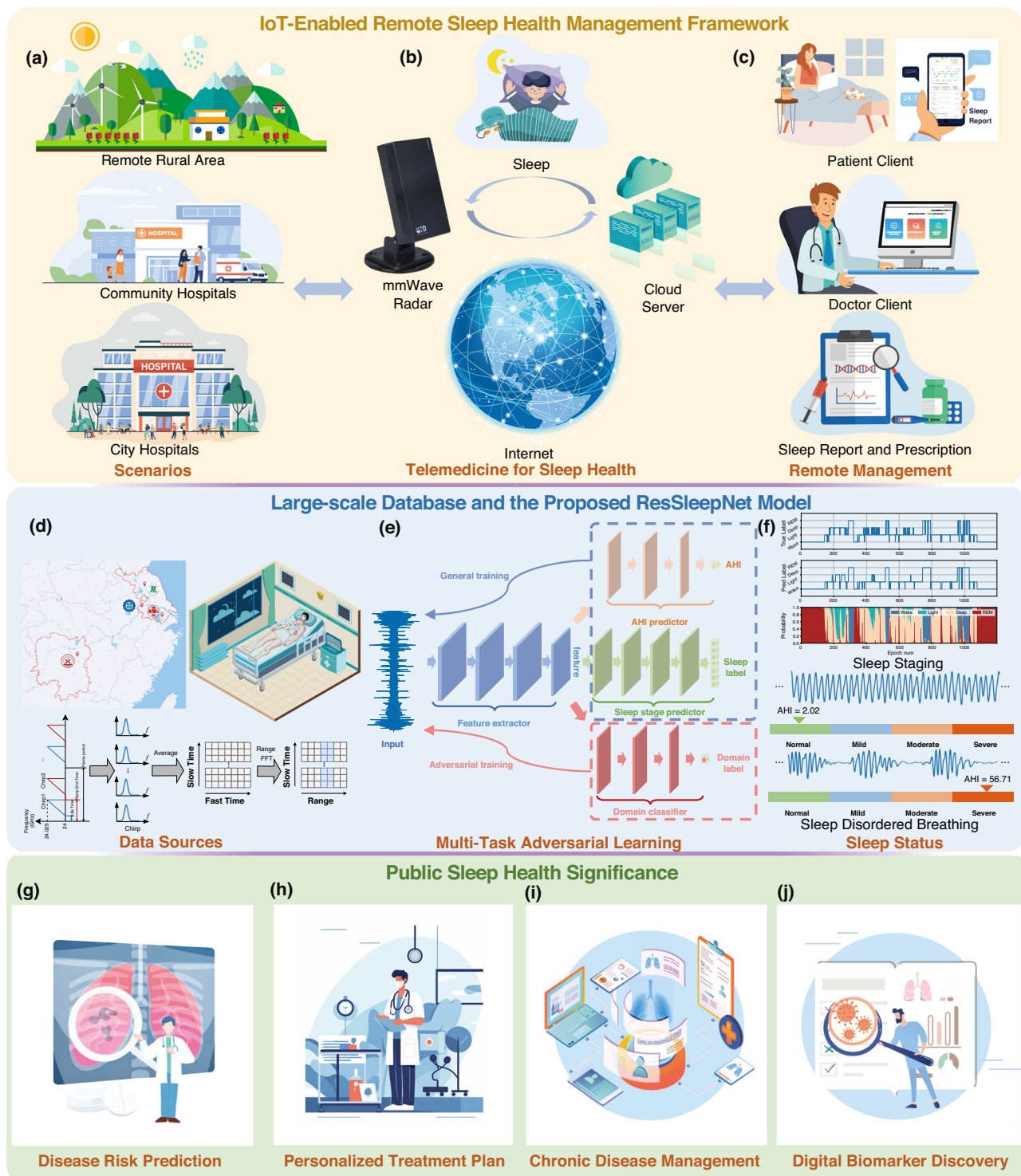


Fig. 1 | Overall framework. IoT-Enabled remote sleep health management framework. **a** Deployment scenarios, including remote rural areas, community hospitals, and city hospitals. **b** Data acquisition using mmWave radar and data transmission to the cloud server. **c** User interfaces for patients and doctors to access sleep reports and manage treatments. Large-scale Datasets and the Proposed ResSleepNet Model: **d** Data sources used for model development. **e** Multi-task adversarial learning framework for sleep staging and AHI prediction. **f** Sleep

status assessment results, including sleep staging and analysis of sleep-disordered breathing. Public Sleep Health Significance: **g** predicting disease risks associated with sleep disorders, **h** formulating personalized treatment plans tailored to individual patient needs, **i** managing chronic diseases influenced by sleep problems, and **j** discovering novel digital biomarkers to support evidence-based medicine and precision healthcare.

with mild severity is the lowest, falling below 50% and dropping to under 40% in external datasets (as shown in the rightmost column of Fig. 3). Since the SOF dataset mainly consists of individuals with mild SDB (as shown in Table 1), the ICC for AHI estimation in SOF is only

0.77. Furthermore, we evaluated the tolerance accuracy, which allows a predicted severity level to differ from the true classification by no more than one category. In the internal dataset, 98.12% of patients were either correctly classified or off by just one severity level. The

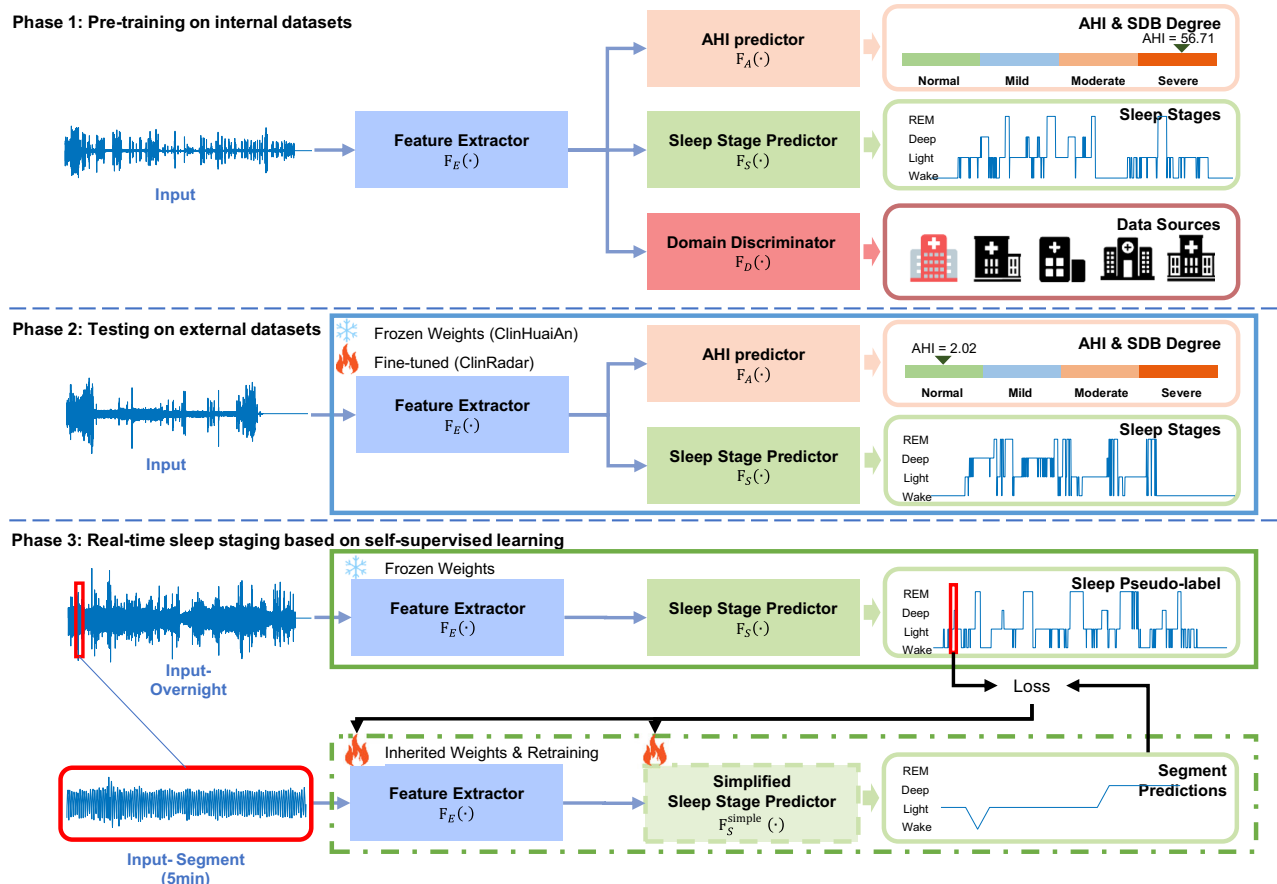


Fig. 2 | Multi-stage flowchart for sleep staging and sleep-disordered breathing prediction. In Phase 1, the model is pre-trained on internal datasets using thoracoabdominal motion signals recorded from respiratory belts. The feature extractor $F_E(\cdot)$ first processes the input sleep signals, which extracts relevant features. These features are then fed into two predictors: the AHI predictor $F_A(\cdot)$ to assess the severity of sleep apnea, and the Sleep stage predictor $F_S(\cdot)$ to predict different sleep stages. A domain discriminator $F_D(\cdot)$ is also employed to distinguish data sources, enhancing the model's generalization capability. In Phase 2, the pre-trained network (including the feature extractor, the AHI predictor, and the Sleep

stage predictor) is applied with frozen weights to respiratory belt-derived respiratory signals (ClinHuiAn) and fine-tuned for radar-derived respiratory signals (ClinRadar). In Phase 3, the model is employed for real-time sleep staging. The input signals are divided into overnight signals and 5-minute segments. The overnight signals are processed by the feature extractor and the sleep stage predictor with frozen weights to generate pseudo-labels. These pseudo-labels, together with the segment signals, are used for further fine-tuning of the feature extractor. A simplified sleep stage predictor is then used for real-time sleep stage prediction.

tolerance accuracy was 99.29% for ClinHuiAn and 92.31% for ClinRadar (Supplementary Table 4). In the 3030 nights of the internal test set, only 2 nights were misclassified between the normal and severe categories (Fig. 3d), while no such misclassifications occurred in the two external datasets (Fig. 3h, i). This single-breathing-pathway-based model maintains reliable SDB monitoring accuracy while significantly enhancing patient comfort during sleep assessment.

To contextualize our model's performance, we compare it against representative state-of-the-art (SOTA) HSAT approaches. As shown in Supplementary Table 6, our framework achieves competitive performance, benefiting from a larger dataset scale, robust model design, and enhanced portability for real-world applications.

Ablation study. To evaluate the contribution of transfer learning, we compared the performance of a model trained directly on radar-derived respiratory signals with a model pretrained on large-scale thoracoabdominal motion signals and subsequently fine-tuned on radar data. As shown in Supplementary Table 5, pretraining led to a 14.2% absolute improvement in accuracy (from 61.6% to 75.8%) and a notable increase in Cohen's Kappa (from 0.37 to 0.62) for sleep staging. For AHI estimation, the ICC improved from -0.13 to 0.87, while the MAE was reduced by more than half (from 22.97 to 8.80 events/hour).

These improvements arise because large-scale respiratory belt datasets provide richer and more diverse respiratory patterns, enabling the model to learn robust low- and mid-level temporal features that transfer effectively to radar signals. Without this pretraining, the radar-only model struggles to capture such variability due to the relatively small radar dataset size. This two-stage strategy mitigates overfitting to the smaller radar dataset and significantly enhances cross-modality generalization, particularly for challenging sleep stages such as Deep and REM, and the AHI estimation task.

Sleep parameters and clinical correlation analysis

Sleep parameters. We performed a difference analysis on key sleep parameters, including total sleep time (TST), sleep efficiency (SE), sleep onset latency (SOL), wake after sleep onset (WASO), as well as the proportions of light sleep, deep sleep, and REM sleep. Detailed definitions of these parameters are provided in Supplementary Table 2. Figure 4a–g shows the violin plots of the differences in sleep parameters calculated from the true and predicted sleep labels. Each point in these plots represents the monitoring result of one night. The text above each violin plot indicates the correlation and P-value between the true and predicted sleep parameters. For all sleep parameters, the P-values are less than 0.0001, indicating strong statistical significance. The actual and predicted values for TST, SE, SOL, WASO, and REM

Table 2 | Model performances on the internal testsets and external datasets

Internal Test Set from NSRR ³⁴ and ClinSuZhou									
Data source (n)	Sleep Staging			AHI Estimation					
	Accuracy (%)	Kappa	Wake (%)	Light (%)	Deep (%)	REM (%)	ICC	MAE	
SHHS (1675)	80.34	0.70	87.92 [87.40, 88.46]	78.19 [77.64, 78.75]	68.97 [67.66, 70.29]	82.98 [82.16, 83.82]	0.90 [0.88, 0.90]	5.49	
MirOS (770)	83.53	0.73	90.33 [89.71, 90.93]	83.80 [83.10, 84.51]	46.09 [43.66, 48.54]	80.81 [79.52, 82.19]	0.90 [0.88, 0.91]	5.65	
MESA (403)	83.09	0.73	89.08 [87.93, 90.30]	82.02 [81.06, 83.00]	53.57 [50.16, 57.00]	83.35 [81.56, 85.25]	0.91 [0.89, 0.92]	6.57	
SOF (90)	78.57	0.68	88.82 [86.69, 91.16]	76.78 [74.24, 79.30]	58.58 [52.93, 64.49]	78.22 [74.57, 82.22]	0.77 [0.67, 0.84]	6.98	
ClinSuZhou (92)	81.72	0.71	88.15 [85.69, 90.91]	81.87 [79.95, 83.90]	66.83 [61.05, 72.85]	74.30 [69.58, 79.30]	0.92 [0.88, 0.95]	6.87	

External Test Set from Respiratory Belt-derived Respiratory Signal (ClinHuiAn)									
Data source (n)	Sleep Staging			AHI Estimation					
	Accuracy (%)	Kappa	Wake (%)	Light (%)	Deep (%)	REM (%)	ICC	MAE	
ClinHuiAn (424)	79.62	0.67	88.65 [87.58, 89.77]	79.98 [79.00, 80.98]	59.61 [56.61, 62.74]	73.06 [71.09, 75.08]	0.94 [0.92, 0.95]	7.33	

External Test Set from Radar-derived Respiratory Signal (ClinRadar)									
Data source (n)	Sleep Staging			AHI Estimation					
	Accuracy (%)	Kappa	Wake (%)	Light (%)	Deep (%)	REM (%)	ICC	MAE	
ClinRadar (221)	75.81	0.62	74.91 [72.70, 77.19]	77.78 [76.42, 79.09]	71.79 [68.43, 75.36]	68.57 [65.78, 71.41]	0.87 [0.83, 0.90]	8.80	

ICC Intraclass Correlation Coefficient, MAE Mean Absolute Error, CI Confidence Interval. The sleep staging Accuracy and Kappa and AHI MAE columns represent "Mean"; the AHI ICC column represents "ICC and 95% CI"; the remaining columns represent "mean sensitivity and 95% CI".

sleep duration are highly consistent, with the corresponding error density plots concentrated around zero. In contrast, there is a slight deviation in the duration of light and deep sleep, and the error density plots have a broader distribution. The larger range in the density plots suggests misclassification between light and deep sleep stages, which leads to reduced accuracy in parameter estimation.

In addition to traditional sleep parameters, we also evaluated the model's capability to assess sleep fragmentation. Figure 4h shows the Kaplan-Meier survival curve of the actual average continuous sleep duration in the internal dataset. Different colors represent specific SDB severity levels. Each curve indicates the proportion of subjects within a category who maintain continuous sleep beyond a certain duration. The results show significant differences in survival curves across SDB severity levels, with more severe SDB leading to increased sleep fragmentation. The predicted results in Fig. 4i closely follow this trend and exhibit good consistency with the actual results in Fig. 4h. However, due to the limited data volume, the Kaplan-Meier survival curves for actual average continuous sleep duration in the two external datasets show some deviation compared to those in the internal test set.

General clinical correlation analysis. To further illustrate the adaptability of our method to different groups, we analyzed the correlation between sleep staging accuracy and various general clinical information, including sex, age, body habitus (BMI), and the severity of sleep apnea.

Sex. We analyzed the relationship between single-night sleep staging performance and sex. As shown in Supplementary Fig. 5a, d, in the internal test set, which has the largest sample size and a relatively balanced male-to-female ratio, the model's performance for males and females shows no significant difference, with similar performance distributions. In ClinHuiAn (86.8% male) and ClinRadar (82.0% male), the performance distribution between males and females remains close, with slight differences attributable to the sex imbalance. Overall, the results indicate that the model's performance does not differ between males and females.

Age. As shown in Table 1, we divided the subjects into four age groups: under 45, 45–55, 55–65, and over 65. Supplementary Fig. 6a, d show that in the internal test set, the accuracy and Kappa values are higher and more stable for participants under 45. As age increases, the median accuracy and Kappa gradually decrease, and the range of their distribution widens, indicating greater variability among older participants. A similar but more pronounced trend is observed in the external datasets (Supplementary Fig. 6b, e, c, f), with larger fluctuations, particularly in the older age groups (above 55). This suggests that the external datasets introduce additional variability, possibly due to differences in population characteristics or data collection environments. Overall, the model's performance declines in older participants, likely due to changes in sleep architecture and the presence of comorbidities in older individuals, making accurate staging more challenging.

Body habitus. To assess the model's robustness against inter-individual differences in body composition, we stratified participants by body mass index (BMI) into four conventional categories: underweight (UW, BMI < 18.5), normal weight (NW, 18.5≤BMI < 25), overweight (OW, 25≤BMI < 30), and obese (OB, BMI≥30). As shown in Supplementary Fig. 7 and Supplementary Fig. 8, the model maintained stable performance across all BMI categories and both sexes. Notably, despite potential concerns that excess adipose tissue in individuals with obesity could attenuate thoracoabdominal motion signals—especially in radar-based sensing—the staging accuracy in the obese group did not show significant degradation. These findings suggest that our preprocessing strategies (e.g., signal normalization)

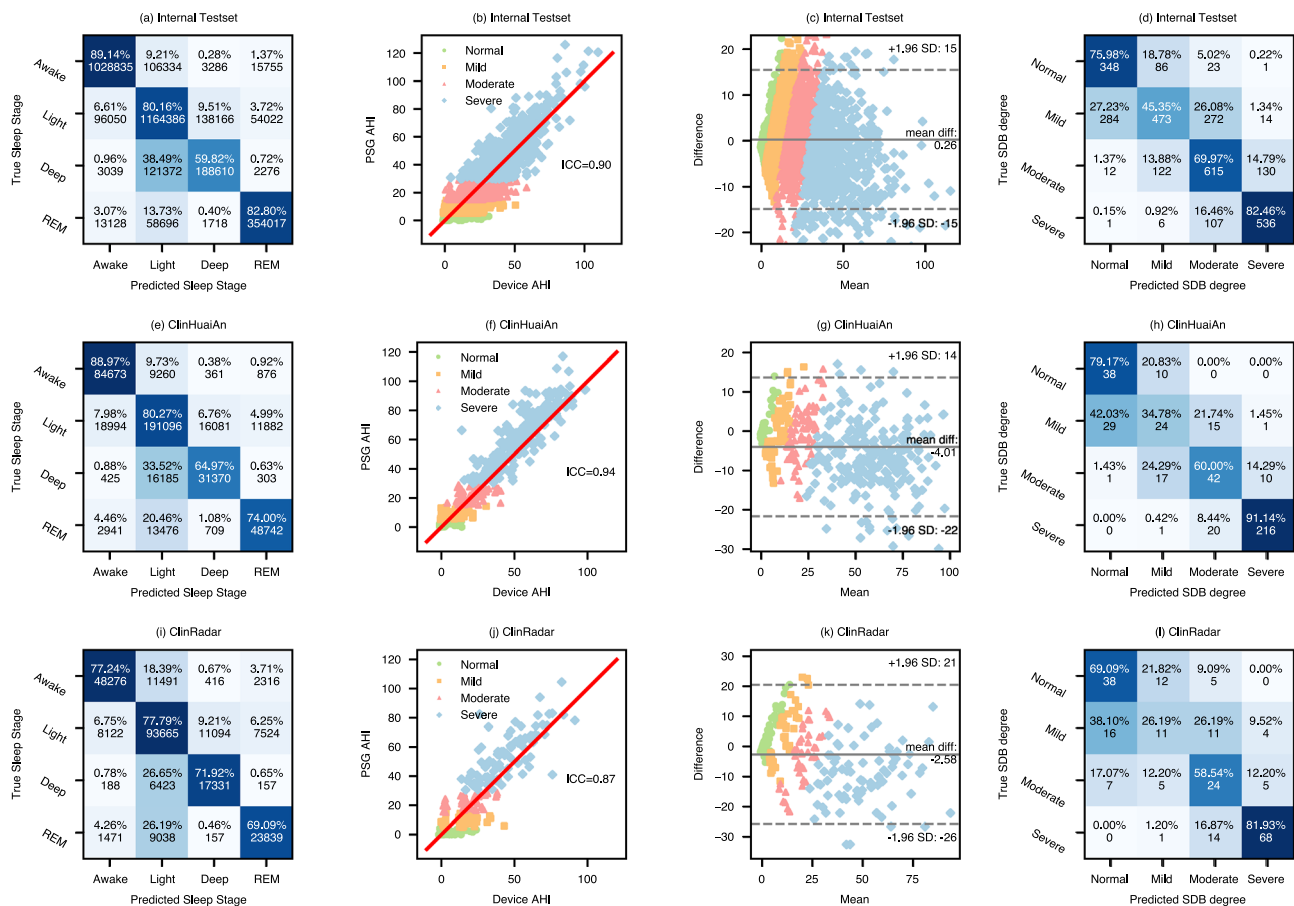


Fig. 3 | Performance evaluation of Four-class Sleep Staging, AHI estimation, and SDB severity classification using respiratory belt-derived signals (Internal Testset, ClinHuiAn) and radar-derived signals (ClinRadar). a, e, i Sleep stage confusion matrix, where the percentages indicate the proportion of correctly and incorrectly classified instances for each stage, and the numbers represent the actual

counts of these classifications. **b, f, j** AHI scatter plot with the middle line representing the equation $y = x$. **c, g, k** Bland-Altman plot for AHI with horizontal lines for the mean difference and the 95% limits of agreement. **d, h, l** SDB severity confusion matrix. Source data are provided as a Source Data file.

and model design contribute to mitigating variations introduced by body habitus, supporting reliable performance even in individuals with elevated BMI.

Sleep Apnea. We categorized the subjects into four groups based on their clinical AHI values: normal, mild, moderate, and severe SDB (as shown in Table 1). Supplementary Fig. 9 presents the distribution of sleep staging accuracy and Kappa across different datasets, indicating no significant differences in model performance between healthy subjects and patients with varying degrees of SDB. In reality, frequent abnormal events such as sleep apnea can lead to repeated micro-arousals and abnormal sleep structures. As the severity of SDB increases, sleep fragmentation worsens (as shown in Fig. 4i), making accurate sleep staging more challenging and placing higher demands on model performance. Previous studies^{5,45} have also confirmed this. However, our model's performance did not decline with increasing SDB severity. This can be attributed to the following reasons: Firstly, we used a more extensive dataset, with a larger number of individuals with SDB included in the model training compared to previous studies; Additionally, the introduction of the auxiliary task enhanced the model's ability to recognize sleep stages in patients with SDB.

Chronic comorbidities correlation analysis. Sleep is closely associated with chronic cardiovascular diseases⁴⁶, respiratory diseases, and PD. These conditions can have an impact on sleep quality and alter respiratory signals, making accurate sleep staging more challenging. In

this study, we evaluated the model's performance in the context of three representative comorbidities involving the heart (hypertension), lungs (lung diseases), and brain (PD).

The results in Supplementary Fig. 10 show that the model performs similarly in subjects with and without hypertension. While hypertension may not directly alter respiratory characteristics, recent studies suggest that respiratory effort during sleep, commonly elevated in obstructive sleep apnea (OSA), may serve as an essential predictor of prevalent hypertension, potentially offering a novel biomarker for cardiovascular risk assessment. In contrast, the model's performance is lower in subjects with lung diseases (Supplementary Fig. 11) than in healthy individuals. Lung diseases are expected to influence respiratory signals through changes in pulmonary mechanics and respiratory patterns, although the extent to which these changes reshape sleep architecture remains to be fully investigated. From another perspective, respiratory signals can serve as potential biomarkers for identifying lung diseases, suggesting the possibility of using respiratory signals for lung disease detection in future studies. Supplementary Fig. 12 illustrates the impact of PD on the model's performance. The results indicate a significant decline in performance among PD subjects. However, due to the severe imbalance in the dataset's ratio of PD to non-PD subjects, further in-depth research focusing specifically on individuals with PD is needed to accurately assess the model's performance in this population. Additionally, in the external datasets, there is an issue of missing information regarding the discussed disease categories. In the future, it will be necessary to

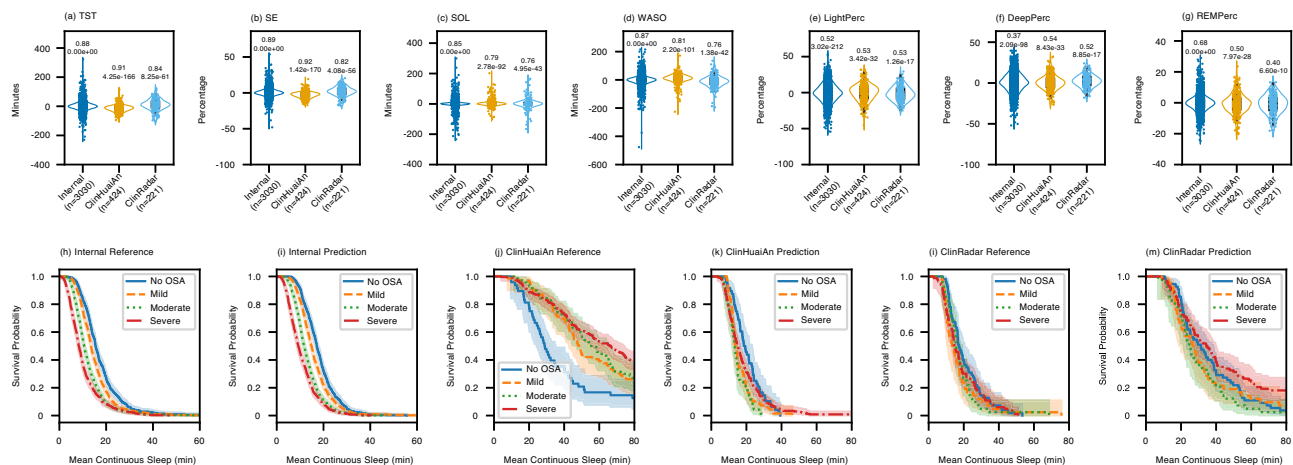


Fig. 4 | Performance evaluation of sleep parameters estimation.

a–g Comparison of sleep parameters (TST, SE, SOL, WASO, and proportions of Light sleep, Deep sleep, and REM sleep) between the actual and predicted values across the internal dataset ($n = 3030$ nights), ClinHuiAn dataset ($n = 424$ nights), and ClinRadar dataset ($n = 221$ nights). Each violin plot shows the distribution of the differences (prediction error) between the true and predicted values. The contour of each violin plot indicates the kernel density of these differences, while the horizontal line inside represents the median, the box bounds indicate the 25th and 75th percentiles, and the whiskers extend to the minimum and maximum values within 1.5 times the interquartile range (IQR). The numerical values above each plot represent the Pearson correlation coefficients between the actual and predicted

parameters, and the significance of the correlations is assessed using a two-tailed Pearson correlation test with exact p-values. All n values refer to independent nights of sleep recordings. Source data are provided as a Source Data file.

h–m Kaplan-Meier plots of the true and predicted average continuous sleep times in the internal dataset (**h, i**), the ClinHuiAn dataset (**j, k**), and the ClinRadar dataset (**l, m**). The vertical axis indicates the survival probability (proportion of subjects with an average continuous sleep time greater than the duration shown on the horizontal axis). Different colored curves represent subjects in distinct sleep apnea-hypopnea severity categories (No OSA, Mild, Moderate, Severe). The shaded areas surrounding each curve denote the 95% confidence intervals of the survival probability estimates.

Table 3 | Real-time sleep staging performance across different input segment durations

Dataset	Segment	Accuracy (%)	F1-score	Kappa	Sensitivity (%)				Inference Time* / Sleep Period	MFLOPs*	Training Time* / Night
					Wake	Light	Deep	REM			
ClinHuiAn	1 min	78.44	0.7423	0.6585	87.75	82.10	51.85	71.22	60.70ms	286.49	25.34s
	2 min	78.97	0.7475	0.6666	87.62	82.67	50.54	73.97	61.51ms	545.46	30.54s
	3 min	79.72	0.7584	0.6790	88.96	82.95	53.54	73.92	62.01ms	806.26	37.14s
	4 min	79.75	0.7586	0.6792	88.79	83.06	53.51	73.99	62.36ms	1065.23	45.85s
	5 min	79.73	0.7584	0.6789	88.72	83.06	53.51	73.99	62.46ms	1329.70	54.89s
ClinRadar ClinRadar	1 min	76.09	0.7164	0.6227	71.96	85.53	46.36	72.19	60.70ms	286.49	25.34s
	2 min	76.30	0.7198	0.6263	72.50	85.45	47.39	72.21	61.51ms	545.46	30.54s
	3 min	76.27	0.7196	0.6256	72.35	85.44	47.42	72.25	62.01ms	806.26	37.14s
	4 min	76.26	0.7194	0.6251	72.13	85.49	47.44	72.28	62.36ms	1065.23	45.85s
	5 min	76.19	0.7189	0.6240	72.09	85.34	47.49	72.29	62.46ms	1329.70	54.89s

Results are reported for respiratory belt-derived respiratory signals (ClinHuiAn) and radar-derived respiratory signals (ClinRadar).

*Inference Time: The model inference time for a single sleep epoch (30 seconds).

*MFLOPs Million Floating Point Operations, a measure of the computational complexity of a neural network.

*Training Time: Training duration of a single epoch when the model takes a whole night as input.

The model was trained on: GPU – A800-80GB, CPU – 14 vCPU Intel Xeon Gold 6348 @ 2.60 GHz.

Inference time calculations were performed on: GPU – RTX 4090, CPU – 16 vCPU Intel Xeon Gold 6430.

supplement the relevant data to further explore the relationship between sleep and comorbidities.

Real-time sleep staging

To extend the utility of our framework to online applications, we further evaluated its performance in real-time sleep staging via a self-supervised learning strategy (Fig. 2, Phase 3). In our study, “real-time” refers to the model’s ability to provide sleep stage predictions epoch-by-epoch, meaning that each 30-second segment is analyzed immediately after its completion, without access to future data. This real-time capability enables the early detection of potential sleep disorders, such as insomnia, and provides a foundation for timely and precise medical interventions.

We evaluated real-time sleep staging performance across different Historical Label Mapping Length (HLML) settings (1–5 minutes). Table 3 presents the results from the ClinHuiAn and ClinRadar datasets, displaying statistical data such as accuracy by organizing all individuals’ predictions. From the perspective of model complexity, the model’s MFLOPs exhibit an approximately linear increase with the length of the input segments, which also leads to an increase in overall training time. In terms of sleep staging accuracy, the detection accuracy for 3, 4, and 5-minute input segments in the ClinHuiAn dataset is higher than that for 1 or 2-minute segments. When the input length is extended from 1 minute to 5 minutes, the sleep staging accuracy improves by 1.29%, and the Kappa increases by 0.02. However, there is no significant performance difference between input lengths from 1 to

5 minutes, with an even smaller improvement observed in the Clin-Radar dataset. These results indicate that the model performs consistently across different input segment lengths, and this self-supervised framework can achieve accurate real-time sleep stage prediction with data as short as 1 minute.

To further illustrate the temporal performance, Supplementary Fig. 13 compares the real-time sleep staging results with different input durations, pseudo-labels generated from full-night signal input, and PSG sleep labels. All of these methods effectively reflect the overall sleep trend throughout the night. When more surrounding sleep cycles are considered, the predictions from full-night input show smoother results with fewer frequent transitions. In contrast, the real-time results exhibit more frequent transitions due to the limitations of input information, relying only on data up to the current state. Consequently, balancing model complexity and detection accuracy is essential in practical applications, requiring an appropriate input segment length based on hardware capabilities. From the perspective of showcasing optimal results, and given the sufficient hardware capabilities in this study, we opted for 5 minutes as the input length for real-time sleep staging. To demonstrate the performance of real-time sleep staging, a representative example in a clinical environment is provided in Supplementary Movie 2.

Remote sleep management

Framework overview. We developed a remote sleep management platform, which facilitates efficient data transmission and user interaction, accommodating both historical analysis and real-time monitoring. As shown in Fig. 5, the system initiates with a radar sensor equipped with a Wi-Fi module that captures respiratory signals. These data are transmitted to a Message Queuing Telemetry Transport (MQTT) broker via the MQTT protocol, which serves as an intermediary, forwarding the data to a cloud server for further processing and storage. The cloud server, equipped with computational and storage capabilities, enables remote sleep health management. To ensure flexible access, the platform supports dual communication protocols. For accessing historical sleep data and analysis results, both clinician-facing and patient-facing applications communicate with the cloud server via Hypertext Transfer Protocol (HTTP) connections, allowing users to request and retrieve stored information. For real-time monitoring, the system establishes a WebSocket connection with the cloud server, enabling continuous updates and low-latency data transmission to clinicians and patient-facing interfaces.

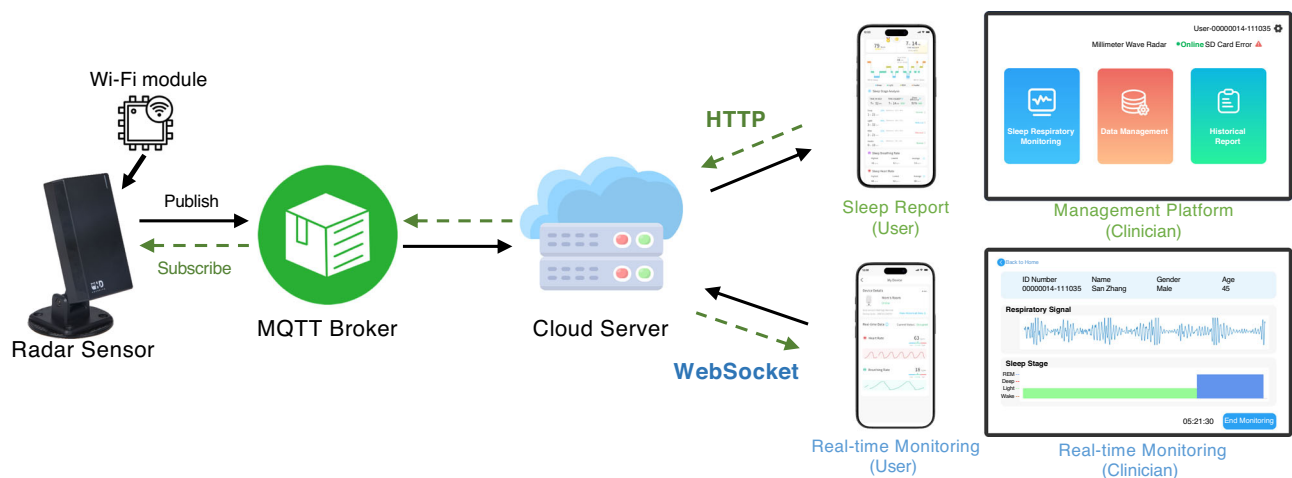


Fig. 5 | The architecture of the remote sleep management platform. The diagram illustrates the end-to-end architecture of the platform, encompassing data collection, processing, and distribution. It highlights the key components, including radar sensors for non-contact data acquisition, the MQTT broker for efficient

Model deployment. Our deployment process integrates TensorFlow-based models into the cloud server to support key analytical tasks such as sleep staging, AHI estimation, and real-time sleep staging. These models are containerized using Docker, enabling seamless deployment and scalability within the cloud environment. The deployed models subscribe to incoming data streams from the MQTT broker, process the data in real-time, and store the results in the server's database. To address the challenge of high concurrency, a load balancer is employed to distribute incoming requests across multiple server instances. This approach supports horizontal scaling, allowing the system to handle increased traffic by dynamically adding more server instances as needed, ensuring stability and efficiency during peak loads. This deployment strategy ensures efficient, accurate, and continuous operation, enabling the platform to deliver timely and reliable sleep health insights while allowing for easy updates and optimizations to the models.

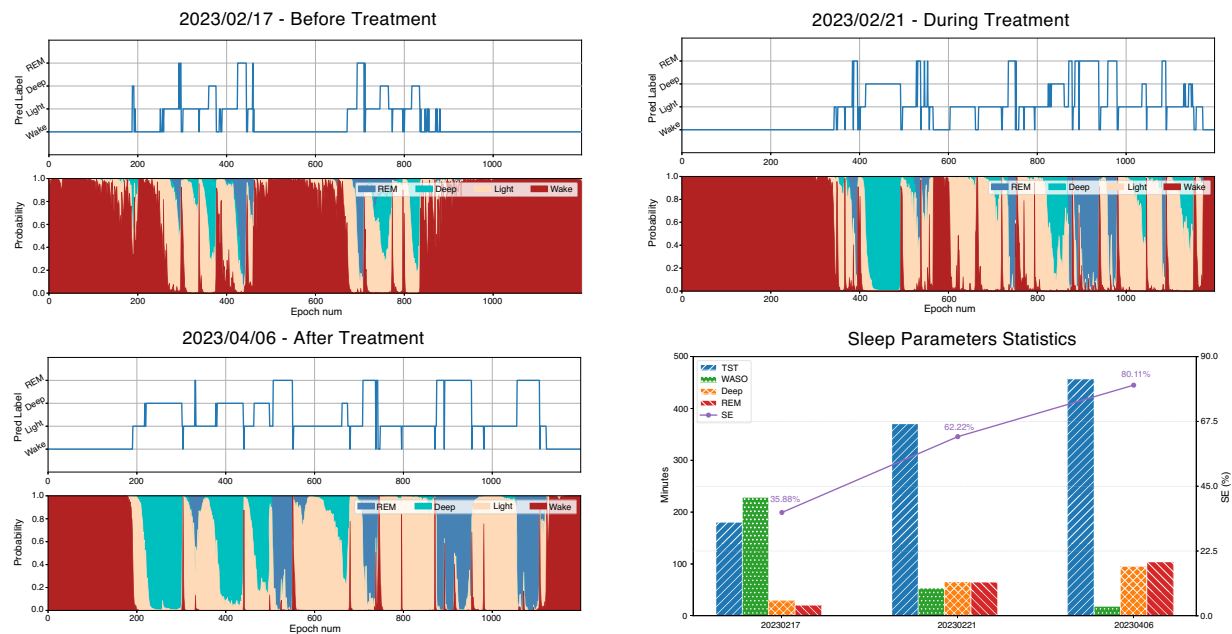
Representative case studies. In this part, we demonstrated the platform's capability to effectively detect, manage, and evaluate treatment outcomes through representative case studies.

Case of insomnia. A female patient with a primary complaint of poor sleep quality during the night is investigated in this case. Initial monitoring by our platform on February 17, 2023, (Fig. 6a) confirmed significant insomnia symptoms, characterized by fragmented sleep patterns and reduced deep sleep and REM sleep. Based on the monitoring results, the platform recommended that the patient seek medical consultation. Following her visit to the hospital, the physician prescribed medication. Subsequent monitoring on February 21, 2023, revealed noticeable improvement, with reduced wakefulness and increased deep sleep and REM stages. After continuing the treatment, monitoring on April 06, 2023, showed significant improvement, with a more consolidated sleep structure. In light of this progress, the physician recommended reducing the medication dosage. The summary chart in Fig. 6a highlights the statistical changes across these stages, confirming the effectiveness of the treatment in improving sleep quality. These results underscore the platform's ability to identify sleep issues and provide continuous feedback on treatment effectiveness, enabling personalized adjustments to the patient's treatment plan.

Case of severe sleep apnea. In another demonstration, our platform effectively identifies and manages severe sleep disorders, as

data transmission, the cloud server for computation and storage, and the user interfaces (clinician-facing and patient-facing) for accessing real-time and historical sleep health insights.

(a) A patient with insomnia



(b) A patient with severe sleep apnea

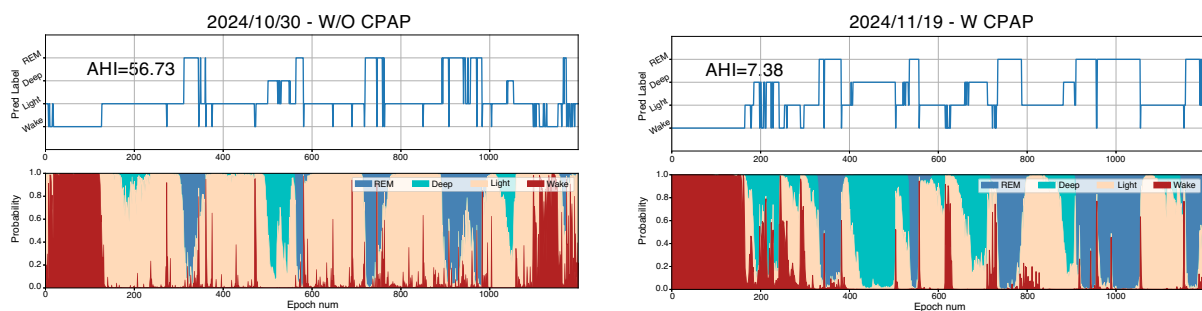


Fig. 6 | Sleep monitoring results demonstrating the platform's capabilities in detection, management, and treatment evaluation for sleep disorders. a A patient with insomnia: Sleep stage predictions and parameters (TST, SE, WASO, Deep sleep duration, and REM sleep duration) change for an insomnia patient at three time points: before treatment, during treatment, and after treatment,

showing progressive improvements. Source data are provided as a Source Data file. **b** A patient with severe sleep apnea: Sleep stage predictions and AHI changes for a severe sleep apnea patient before and during CPAP treatment, demonstrating significant improvements in sleep structure and respiratory events.

demonstrated by the case of a male patient from a remote area (Fig. 6b). Using our deployed non-contact radar device, the system detected severe sleep apnea symptoms and highly irregular sleep rhythms on October 30, 2024, with an initial AHI of 56.73, indicating a significant sleep disorder. Based on the monitoring data, the platform recommended immediate medical attention. The patient traveled to the sleep center of a Grade-A tertiary hospital in the city, where he was diagnosed with severe SDB. Physicians prescribed continuous positive airway pressure (CPAP) therapy. After CPAP treatment at home, the patient's sleep quality improved dramatically. Monitoring on November 19, 2024, revealed a markedly reduced AHI of 7.38 and significant improvement in sleep architecture, including increased deep sleep and REM stages. This case highlights the platform's ability to detect severe sleep disorders early, enabling timely intervention and supporting effective treatment outcomes through continuous monitoring.

These case studies highlight the platform's capability for precise detection, personalized management, and comprehensive evaluation of treatment outcomes. In the future, the platform aims to further expand its functionality in the management of other sleep-related chronic diseases while enhancing its real-time analysis capabilities. Ultimately, with this platform, we aim to realize improving sleep health

for the general population and reduce disparities in global sleep healthcare equity.

Discussion

This study focuses on the increasingly severe burden of sleep disorders and the pressing demand for convenient, efficient, and accurate sleep monitoring technologies in the context of limited medical resources and a shortage of specialized professionals. Current sleep medicine research primarily relies on specialized sleep centers and equipment, with diagnostic procedures that are time-consuming and labor-intensive, making it challenging to extend coverage to remote and underdeveloped areas. This limitation objectively exacerbates sleep health inequities. To address these challenges, we propose ResSleepNet, a high-precision, high-comfort, and generalizable sleep monitoring model based on large-scale thoracoabdominal motion signals. This model is designed to enable non-contact, real-time, and remotely accessible sleep management across a wider range of application scenarios.

At the data level, we collected over 1000 nights of real-world sleep data from multiple sleep centers, while integrating more than 10,000 nights of data from several public sleep datasets. Using the

thoracoabdominal motion signal from radar and the respiratory belt, we constructed a feature-rich, large-scale sleep dataset. At the technical level, we proposed an end-to-end, thoracoabdominal motion-based universal sleep monitoring framework for sleep quality assessment, encompassing full-night sleep staging and sleep parameter estimation, and AHI estimation. Notably, we developed a real-time sleep staging model based on self-supervised learning, achieving efficient and accurate real-time sleep stage prediction using thoracoabdominal motion signal segments as short as 1 minute without the need for labeled data. This approach not only overcomes the traditional reliance on post-processing of full-night data but also provides feasibility for real-time sleep monitoring across diverse scenarios. At the application level, we further enhanced the model's adaptability to radar-derived respiratory signals, providing a fully non-contact sleep monitoring solution.

Additionally, we developed a remote sleep management platform capable of identifying and tracking various sleep disorders, such as insomnia and obstructive sleep apnea, while also evaluating treatment efficacy. Compared to portable ECG devices or smartwatches, the proposed method offers enhanced comfort and enables continuous monitoring in both professional sleep centers and daily home environments, providing a scalable technological foundation for improving sleep health equity. Multidimensional evaluation results demonstrate that the proposed method exhibits robust performance across diverse populations, including individuals of different age groups, sexes, sleep apnea severity levels, and comorbidities such as hypertension. Its high concordance with expert-annotated PSG results further validates its clinical applicability.

Our method is not intended to replace in-lab PSG fully, but rather to serve as a complementary, low-cost, and scalable alternative for home-based sleep testing and remote health monitoring. It is particularly suited for initial triage, longitudinal monitoring, and settings with limited PSG access, where multi-night data and high-frequency screening are clinically valuable. Through subgroup analysis, we also demonstrated the robustness of our model across diverse populations, including individuals with obesity or chronic respiratory diseases (e.g., hypertension, COPD), where respiratory mechanics may differ. Together, these findings underscore the method's real-world utility in supporting sleep-related risk stratification and long-term sleep health management. In support of broader clinical adoption, growing evidence from randomized controlled trials (RCTs) has shown that simplified home-based approaches—such as HSAT—can match or even outperform traditional PSG in the diagnosis of OSA^{47,48}, while significantly reducing diagnostic delays and improving patient adherence⁴⁹. Moreover, early intervention based on such methods has been linked to reduced cardiovascular risks and improved quality of life^{50,51}. Complementing these findings, our method's high accuracy, non-contact nature, and scalability make it well-suited for large-scale deployment in primary care and under-resourced settings. From a health economics perspective, prior studies have demonstrated that HSAT-based screening can reduce costs by 40–60% relative to in-lab PSG^{47,48,52}. Our approach goes a step further by eliminating consumables, reducing labor and maintenance burdens, and enabling remote, continuous monitoring. These features contribute to an economically sustainable model of sleep care delivery that aligns with public health priorities and promotes equitable access to diagnostic and monitoring resources.

Our study has several limitations. First, the model is developed based solely on thoracoabdominal motion and does not incorporate EEG or EMG signals, which are essential for fine-grained sleep staging in conventional PSG. As a result, classification performance in certain transitional stages, such as deep sleep, may be suboptimal. Nonetheless, the overall performance remains competitive with, and in some cases surpasses, existing SOTA HSAT methods. Second, the model is designed for risk-oriented AHI estimation rather than direct

classification of individual respiratory event subtypes (e.g., hypopnea vs. apnea). Given the low inter-scoring agreement in manual subtype annotations, this strategy avoids label ambiguity and enhances model generalizability. Third, our validation was conducted under single-subject PSG settings, which are standard in clinical practice; however, model performance under multi-person or interference-prone environments has not been evaluated. Although prior radar-based studies have demonstrated technical feasibility for multi-subject monitoring, we emphasize that accurate and clinically applicable sleep assessment benefits from controlled, interference-free single-subject settings. This aligns with PSG practice and supports our focus on precise individual-level evaluation.

In the future, we plan to collect data from diverse regions and populations, incorporating comprehensive symptom records and comorbidity information to investigate inter-population differences and the mechanisms underlying sleep-related comorbidities in depth. Meanwhile, we plan to incorporate more non-contact sensors to enhance the robustness and generalizability of the model across diverse environments and populations. Moreover, the developed real-time sleep staging capability could enable future applications in closed-loop sleep modulation and personalized intervention, such as providing auditory or light stimulation during slow-wave sleep deficiency, or delivering early-stage feedback for abnormal sleep patterns. These potential applications could further expand the clinical and home-based utility of remote sleep monitoring systems. Long-term continuous monitoring is also a key focus for the next phase, as it will provide a solid foundation for a deeper understanding of the relationship between sleep patterns and chronic diseases, potentially paving the way for early warning systems and improved chronic disease management. To support real-world deployment, we plan to initiate prospective clinical trials and pursue regulatory approval (e.g., FDA 510(k), CE certification), with the goal of registering the proposed framework as a certified medical device. These efforts will help ensure compliance with medical standards for safety and efficacy, facilitate integration with existing hospital and home-based workflows, and broaden clinical adoption. Finally, we aim to develop personalized algorithms for “precision sleep management” tailored to individual health conditions, advancing the integrated management of sleep and overall health. This endeavor aspires to contribute to the provision of accessible and high-quality sleep healthcare services on a broader scale.

Methods

Datasets and scoring rules

Demographic information, including age and sex, was harmonized by the NSRR team to align with TOPMed and BioDataCatalyst standards, with sex information derived from self-reported data or clinical records in the original cohorts. For the clinical datasets, sex was recorded in medical records at the time of PSG acquisition.

ClinHuaian & ClinSuZhou. The ClinHuaian dataset comprises PSG records of 458 individuals aged 11 to 88. Meanwhile, the ClinSuZhou dataset includes PSG records of 424 individuals aged 12 to 78. These PSGs were recorded between 2021 and 2023 using a type I PSG. The data were collected in the sleep laboratory at The Second Affiliated Hospital of Soochow University in Jiangsu, China, and The Affiliated Huaian No.1 People's Hospital of Nanjing Medical University in Jiangsu, China. The thoracic and abdominal signals were sampled at a frequency of 32 or 128 Hz.

ClinRadar. The dataset consists of data from 221 individuals. These data were collected in the sleep laboratory at The Second Affiliated Hospital of Soochow University in Jiangsu, China, between 2022 and 2023. Wireless radar sensors were placed above the bed's headboard in patient rooms for sleep monitoring. At the same time, individuals were

monitored using a Type I PSG. The wireless radar sensors employed are frequency-modulated continuous-wave (FMCW) radar sensor BGT24MTR11 from Infineon Technologies AG. The key parameters of the FMCW radar are as follows: it has an antenna configuration of one transmitter and one receiver with a 4×2 patch, providing an antenna gain of 12 dBi. It features an opening angle of 20° by 42° , ensuring a wide field of view. The chirp signal starts at a frequency of 24.025 GHz and stops at 24.225 GHz, resulting in a bandwidth of 200 MHz. Each chirp is sampled 256 times within 1.5 ms. The radar operates with one chirp per frame and a frame interval of 25 ms.

Scoring rules. The definitions of apnea and hypopnea events and the computation of the reference AHI are homogeneous across datasets. The AHI is defined as: (Apneas with no oxygen desaturation threshold used and with or without arousal + hypopneas with $> 30\%$ flow reduction and $\geq 3\%$ oxygen desaturation or with arousal) / hour of sleep from PSG, consistent with the AASM12 recommended rule. These AHI variables are available on the NSRR website. We used the *ahi_a0h3a* variable in SHHS, MESA, and SOF, and the *poah3a* variable in MrOS. Sleep staging annotations were based on 30-second epochs. Specifically, SHHS used the Rechtschaffen and Kales (R&K) criteria⁵³, with stage labels including Wake, S1, S2, S3, S4, and REM. The MESA applied the 2007 AASM rules⁵⁴ with labels including Wake, N1, N2, N3, N4, and REM. For MrOS and SOF datasets, while the exact scoring guideline version was not explicitly specified, official scoring documents were provided, with stage labels corresponding to S1, S2, S3, S4, and REM. All self-collected datasets (ClinSuZhou, ClinHuaiAn, and ClinRadar) were scored using AASM Version 2.6 criteria⁵⁵. Given the heterogeneity in staging systems across datasets, we unified the labels into four categories for model training and evaluation: 0 = Wake, 1 = Light sleep (N1/S1 + N2/S2), 2 = Deep sleep (N3/S3 + N4/S4), and 3 = REM.

Statistics & reproducibility

This study is solely focused on thoracoabdominal motion signals across all datasets. Data from different sources exhibit varying sampling frequencies. To facilitate the neural network training process, the data from each night were standardized to the same length. Data records were excluded if they met any of the following criteria: (1) the full-night sleep labels contained undefined sleep stages; (2) sleep period time (SPT) was less than 4 hours, which is defined as the duration from the first sleep stage to the last sleep stage (see Supplementary Table 2); (3) no clear AHI reference value was available; and (4) the absence of data from either the thoracic or abdominal motion channels. Data from 15,140 nights from the SHHS, MESA, MrOS, SOF, and ClinSuZhou datasets were combined to form one internal dataset. Each subset was first split individually according to a ratio of 7:1:2 (training, validation, and testing), and the resulting splits were combined to form the overall internal training, validation, and testing sets. Data splitting was performed at the level of nights, ensuring that no data from the same night appeared in both the training and testing sets.

During model development, the internal test set was withheld and only the training and validation sets were used for model training and optimization. Additionally, the ClinHuaiAn dataset was retained as an independent external test set and was not accessed during the model development process. For the ClinRadar dataset, 4-fold cross-validation was performed, where each fold served as a held-out test set while the remaining folds were used for fine-tuning and validation, with strict isolation of test folds to prevent information leakage. To ensure unbiased evaluation, all test sets (internal test set, ClinHuaiAn, and the held-out folds of ClinRadar) remained inaccessible to the researchers until the model, hyperparameters, and thresholds were finalized. Statistical analyses, including Pearson correlation, ICC calculation, and Bland-Altman plots, were conducted using Python (SciPy and

statsmodels packages). All tests were two-sided, and exact P values, confidence intervals, and effect sizes are reported in the Supplementary Data.

Respiratory signal processing

Respiratory signal from radar. The radar-receiving antenna captures the wireless signal and is downconverted to quadrature In-phase/Quadrature (I/Q) signals that are pre-processed by the intermediate frequency (IF) amplifier before digitization. In practice, the baseband I/Q signal with direct current (DC) bias is:

$$\begin{aligned} I(t) &= \cos[\theta + 4\pi x(t)/\lambda + \Delta\varphi] + DC_I \\ Q(t) &= \sin[\theta + 4\pi x(t)/\lambda + \Delta\varphi] + DC_Q \end{aligned} \quad (1)$$

where θ is the phase shift caused by the electromagnetic signal propagation distance, $\Delta\varphi$ is the residual phase noise, $x(t)$ is the motion information of the target to be detected, λ is the wavelength of the radar carrier wave, and DC_I and DC_Q are the DC offsets.

For extracting respiratory signals using radar, Constant False Alarm Rate (CFAR) detection is performed on the product of the I/Q signals in Equation (1) to identify the positions where body movements occur. Subsequently, the stable data segments between the periods of body movement are subjected to least-squares circle fitting to remove the trend components from the I/Q signals. The detrended I/Q signals are then demodulated using the arctangent function to obtain the most primitive respiratory signal *Radar_{respi}*. To enhance signal quality, we further apply signal-level fusion using maximal-ratio combining, the principal component analysis (MRC-PCA)⁵⁶, resulting in a denoised and stable respiration waveform. The radar signal is subsequently filtered, resampled, and standardized using the same pipeline described below for the PSG-based signal. The full radar signal demodulation and pre-processing workflow is illustrated in Fig. 7(a). To illustrate the consistency between respiratory signals derived from the radar sensor and those from the reference respiration belt, a representative recording is provided in Supplementary Movie 1. Supplementary Movie 3 further demonstrates the radar's capability to capture abnormal respiratory events in a home environment.

Respiratory belt signal from PSG. In PSG sleep studies, the respiratory signal *Wave_{respi}* is obtained by directly superimposing the chest belt signal with the abdominal belt signal and then filtering and downsampling the respiratory signal. This summation follows the classical dual-compartment model of respiratory inductance plethysmography (RIP)⁵⁷, which demonstrates that the combined thoracoabdominal movements approximate respiratory effort. Uncalibrated RIPsum has been widely adopted in sleep monitoring applications^{58,59}, in accordance with AASM guidelines. Low-pass filtering is employed to remove high-frequency noise and prevent aliasing during downsampling. The filter utilized is a zero-phase 8th-order low-pass Chebyshev Type II, with a cutoff frequency of 8 Hz and a stopband attenuation of 40 decibels. The filtered respiratory signal is subsequently downsampled to approximately 34.13 Hz using linear interpolation, reducing the computational and storage demands for deep learning. We chose a sampling rate of 34.13 Hz because this results in 1024 samples per 30-second sleep window. By using a 2^n number, we can maintain full temporal alignment of data with the sleep-window during pooling operations¹⁵. The signal is cleaned by truncating values to three standard deviations and then normalized by subtracting the mean and dividing by the standard deviation to obtain the final respiratory signal. The complete respiratory signal pre-processing pipeline, including radar and PSG inputs, is summarized in Fig. 7b.

Model based on MTAL

Overall architecture. As depicted in Supplementary Fig. 3, the model first employs a feature extractor F_E , which is composed of

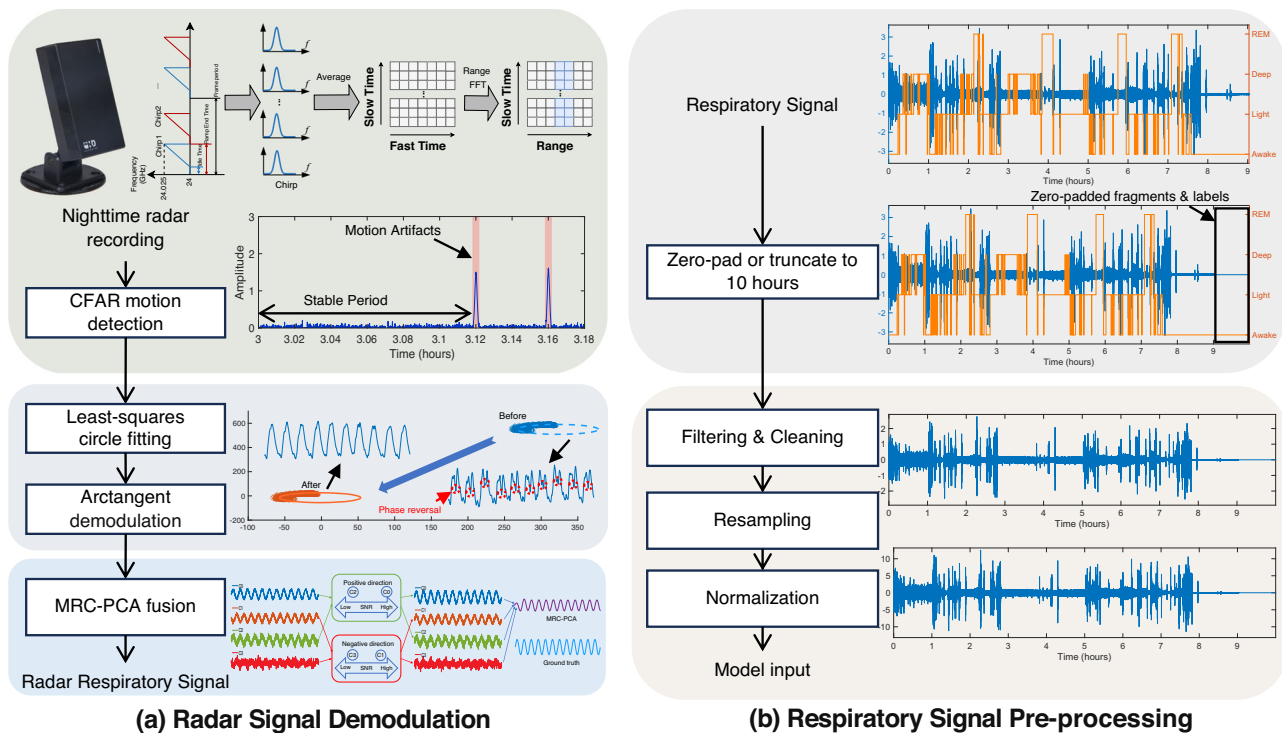


Fig. 7 | Radar signal demodulation and respiratory signal pre-processing workflow for sleep analysis. This figure illustrates the complete signal processing pipeline prior to model input. **a** Radar Signal Demodulation: Raw radar recordings undergo motion detection via CFAR to identify stable breathing segments. These are demodulated using least-squares circle fitting and arc tangent transformation, followed by signal-level enhancement through MRC-PCA fusion to yield clean

respiratory waveforms. **b** Respiratory Signal Pre-processing: Both radar- and belt-derived signals are processed using a shared pipeline that includes truncation to a fixed duration, artifact removal, low-pass filtering, uniform resampling (34.13 Hz), and z-score normalization. Sample outputs at key stages demonstrate signal refinement and alignment with annotated sleep stages for downstream analysis.

convolutional networks, to convert the overnight respiratory signal S into a continuous feature vector X . This feature vector is then fed into three parallel structures: the sleep stage predictor F_S , the AHI estimator F_A , and the domain discriminator F_D . These components respectively predict the subject's sleep stages, AHI values, and the dataset ID from which the subject's data originated.

Feature extractor. The F_E extracts information F_E^L from each individual S_i , translating the high-dimensionality inputs into a lower-dimensional space called an embedding X_i such that $F_E(S) \mapsto X$. It comprises two parallel Convolutional Neural Network (CNN) branches: a long-range branch and a short-range branch. The entire network takes respiratory signals as input, which are initially fed into the two parallel CNN branches. The long-range branch consists of multiple one-dimensional convolutional and pooling layers, with a kernel size of 9. It is primarily used to extract long-range signal features. A Leaky ReLU activation function and a max-pooling layer follow each convolutional layer. After multi-level extraction, the output features are fixed. The short-range branch first reshapes the data into different windows with a length of 30 seconds. The rest of the structure is similar to the long-range branch, but with a kernel size of 3. The output features from the two branches are fused by element-wise addition and then fed into a fully connected layer for further processing and classification. This structural design aims to combine long-range and short-range features, enhancing the model's ability to perceive features at different time scales, thereby improving the overall model performance and accuracy.

Sleep stage predictor. The sleep stage predictor is composed of two dilated convolution blocks and two Transformer encoders, which are arranged alternately. The dilated convolution block contains six one-dimensional dilated convolution layers with a kernel size of 7 and 128

channels. Each convolution layer has a different dilation rate: 1, 2, 4, 8, 16, and 32. Dilated convolutions can capture a broader temporal context without increasing computational complexity. A Leaky ReLU activation function follows each convolutional layer, and L2 regularization is applied to prevent overfitting. Finally, this module includes a Dropout layer for further regularization and uses residual connections (Add) to sum the output with the input, thereby preserving the input features. The Transformer encoder module is utilized to model long-term dependencies in the signal. Each encoder comprises a multi-head self-attention mechanism and a feed-forward neural network. Here, the input dimension of the encoder is 1200, the output dimension is 128, and it includes 4 attention heads. This module effectively captures complex temporal correlation information. After being processed through the two dilated convolution blocks and the Transformer encoder modules, the features are sent to a one-dimensional convolutional layer with a kernel size of 1 and 4 channels, corresponding to the predicted sleep stage categories. Finally, a Softmax activation function generates the probability distribution for each sleep stage. This architecture is designed to fully utilize the characteristics of dilated convolutions to extract multi-scale temporal features, while also leveraging the capabilities of the Transformer encoder to capture long-term dependencies. This combination enhances the accuracy and robustness of sleep stage predictions. Cross-entropy loss is employed to measure the difference between the predicted sleep stage labels and the true labels. The cross-entropy loss function is defined as:

$$L_{\text{sleep}} = - \sum_{i=1}^N y_i^S \log(\hat{y}_i^S) \quad (2)$$

where y_i^S is the true sleep label, \hat{y}_i^S is the predicted sleep probability, and N is the number of samples.

AHI estimator. The architecture of the AHI predictor consists of convolutional blocks and fully connected blocks. The convolutional block includes four stacked dilated convolution blocks (StackedConv), each using different dilation rates (1, 2, 3, 4) with the same convolution kernel size of 9. Each dilated convolution block is composed of three dilated convolution layers, a Batch Normalization layer, a MaxPooling layer, and a Dropout layer. These convolutional blocks progressively extract features and reduce the size of the feature maps. Finally, a Flatten layer flattens the features into a one-dimensional vector, followed by a Dense layer to extract high-level features. The fully connected block comprises three fully connected layers, each followed by a Batch Normalization layer, a Leaky ReLU activation function, and a Dropout layer. The network gradually compresses the feature dimensions through these three layers and ultimately outputs a single estimated AHI value. Mean squared error (MSE) loss is applied to measure the difference between the predicted AHI values and the true values. The mean squared error loss function is defined as:

$$L_{\text{AHI}} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i^A - y_i^A)^2 \quad (3)$$

where y_i^A is the true value, \hat{y}_i^A is the predicted value, and N is the number of samples.

Domain discriminator. To improve generalization across diverse datasets, we incorporate a domain discriminator into the model as part of an adversarial training framework. The domain discriminator is not designed to infer data source identities (e.g., device type, demographic origin) directly from raw signals, which are typically available through metadata. Instead, its purpose is to reduce the influence of domain-specific biases embedded in the physiological signal patterns themselves. This model architecture extracts deep features from the input signals through multiple convolutional and pooling layers. The initial convolution block utilizes 64 filters, with the number of filters progressively decreasing to 32, 16, and 8 in subsequent layers. This reduction in filter count helps decrease computational complexity while focusing on extracting essential information. Each convolutional layer is followed by batch normalization and pooling operations to ensure stability and reduce the size of the feature maps. After flattening, the feature vector is fed into fully connected layers, sequentially passing through layers with 128 and 64 units. The output layer uses a Softmax activation function to map the features to four domain labels. This design enables the model to perform domain classification based on the characteristics of the input data and serve as a domain classifier component in adversarial learning to minimize distribution differences between domains. Cross-entropy loss is utilized to measure the difference between the predicted domain labels and the true labels. The cross-entropy loss function is defined as:

$$L_{\text{domain}} = - \sum_{i=1}^N y_i^D \log(\hat{y}_i^D) \quad (4)$$

where y_i^D is the true domain label, \hat{y}_i^D is the predicted domain probability, and N is the number of samples.

Multi-task learning. During adversarial training, the feature extractor and the domain discriminator are optimized via an adversarial learning mechanism. The goal of the feature extractor is to generate features that are not easily distinguishable by the domain discriminator. Meanwhile, the domain discriminator attempts to distinguish features from different datasets. This adversarial training helps the feature extractor learn more generalizable features, thereby enhancing the generalization performance of the sleep stage predictor and AHI estimator across different datasets. Specifically, the optimization goal

of the model is to minimize the overall objective function V , which includes maximizing the loss of the domain discriminator, while minimizing the sleep staging loss and the AHI loss, i.e.,

$$\min_{(F_E, F_S, F_A)} \max_{(F_D)} V = L_{\text{sleep}} + \eta_1 L_{\text{AHI}} - \eta_2 L_{\text{domain}} \quad (5)$$

where $\eta_1 = 0.001$ and $\eta_2 = 1$ are the balancing factors, which have been selected empirically in our experiments. In summary, this multitask adversarial learning framework extracts features from the overnight respiratory signal. It performs normal training with the sleep stage predictor and AHI estimator. Additionally, it optimizes the domain discriminator through adversarial training to reduce inter-dataset variability. This ultimately leads to accurate sleep stage classification and AHI estimation.

Performance measures. We adopted four metrics to evaluate the performance of our method. These metrics can be defined as follows:

1. Accuracy: Calculated as the number of correctly classified samples divided by the total number of samples.

$$\text{Accuracy} = \frac{\sum_{i=1}^K \text{TP}_i}{N} \quad (6)$$

where TP_i denotes the number of correctly classified positive examples of class i , N is the number of samples, K is the number of classes.

2. Cohen's Kappa: Measures the consistency of predictions and the annotated labels to evaluate classification performance.

$$\kappa = \frac{\text{Accuracy} - P_e}{1 - P_e} \quad (7)$$

where $P_e = \sum_{n=1}^K P_{n+} P_{+n}$, P_{n+} is the proportion of category n predicted by the model, while P_{+n} is the proportion of category n in the annotated label.

3. Sensitivity: Refers to the model's ability to recognize positive examples.

$$\text{Sensitivity}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \quad (8)$$

where FN_i denotes the number of incorrectly classified negative examples of class i .

4. Specificity: Refers to the model's ability to recognize negative examples.

$$\text{Specificity}_i = \frac{\text{TN}_i}{\text{TN}_i + \text{FP}_i} \quad (9)$$

where TN_i denotes the number of correctly classified negative examples of class i , where FP_i denotes the number of incorrectly classified positive examples of class i .

5. F1-score: Measures the overall balance between precision and sensitivity, calculated as the macro-averaged F1 across all classes.

$$\text{F1} = \frac{1}{K} \sum_{i=1}^K \text{F1}_i \quad (10)$$

where K is the number of classes.

6. Class-wise F1-score: Represents the harmonic mean of precision and sensitivity for each class i .

$$\text{F1}_i = \frac{2 \cdot \text{TP}_i}{2 \cdot \text{TP}_i + \text{FP}_i + \text{FN}_i} \quad (11)$$

where TP_i , FP_i , and FN_i denote the true positive, false positive, and false negative counts for class i , respectively.

We also adopted six metrics to evaluate the performance of AHI estimation.

1. Intraclass Correlation Coefficient (ICC): Measures the reliability of measurements across observers.

$$ICC = \frac{MS_I - MS_E}{MS_I + (O - 1)MS_E + O \times \frac{MS_O - MS_E}{n}} \quad (12)$$

where O is the number of observers (two, in this case, the real and predicted AHI), MS_I is the instance mean square, MS_E is the mean square error, and MS_O is the observer mean square.

2. Confidence intervals (CI): Provide a range for the true mean of the population.

$$CI = M \pm Z \times ST \quad (13)$$

where M is the average value of samples, Z is the corresponding value found in the standard normal distribution table based on the chosen confidence level, and ST is the standard deviation of samples.

3. Mean Absolute Error (MAE): Quantifies the average difference between predicted and actual values.

$$MAE = \frac{1}{N} \sum_{i=1}^N |u_i - v_i| \quad (14)$$

where N is the number of samples, u_i is the real AHI value of the i th sample, and v_i is the predicted AHI value of the i th sample.

4. Root Mean Square Error (RMSE): Evaluates the square root of the average squared differences between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (u_i - v_i)^2} \quad (15)$$

where N is the number of samples, u_i is the real AHI value of the i th sample, and v_i is the predicted AHI value of the i th sample.

5. Pearson correlation: Assesses the linear relationship between predicted and actual values.

$$\text{Pearson correlation} = \frac{\sum_{i=1}^N (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^N (u_i - \bar{u})^2} \sqrt{\sum_{i=1}^N (v_i - \bar{v})^2}} \quad (16)$$

where N is the number of samples, u_i is the real AHI value of the i th sample, \bar{u} is the average of all real AHI values, v_i is the predicted AHI value of the i th sample, and \bar{v} is the average of all predicted AHI values.

6. R^2 : Indicates the proportion of variance in actual values explained by predicted values.

$$R^2 = 1 - \frac{\sum_{i=1}^N (u_i - v_i)^2}{\sum_{i=1}^N (u_i - \bar{u})^2} \quad (17)$$

where N is the number of samples, u_i is the real AHI value of the i th sample, and v_i is the predicted AHI value of the i th sample.

Real-time sleep staging architecture

Overview of real-time sleep staging framework. The architecture for real-time sleep staging is shown in Fig. 2 (Phase 3). During training, the full-night respiratory signals are first passed through a pretrained and fixed feature extractor and sleep stage predictor to generate pseudo-

labels. These pseudo-labels, together with short-segment respiratory inputs, are then used to fine-tune a lightweight sleep stage predictor (Supplementary Fig. 3b) for real-time inference.

Context window design. To support real-time predictions, only preceding segments are used for each epoch prediction, defined as HLML. Following previous literature^{60–63}, HLML values ranging from 1 to 5 minutes were explored to assess the trade-off between context window length and model performance. The model's computational complexity (MFLOPs) grows approximately linearly with input duration.

Training Strategy. In the training process, we minimized the cross-entropy loss between the predicted sleep stage probabilities and the generated pseudo-labels over the dataset. This strategy enabled the real-time sleep stage predictor to efficiently learn classification capabilities without relying on manually labeled data.

Ethics statement

This study complies with all relevant ethical regulations and was approved by the Institutional Review Board of Xiangya Hospital, Central South University (IRB No. 201909818). All participants provided written informed consent prior to data collection. All participants volunteered for the project and did not receive any additional compensation. Detailed information on ethical approval and consent procedures is provided in the Supplementary Notes.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The SHHS, MrOS, MESA, and SOF datasets used in this study are publicly available from the National Sleep Research Resource: SHHS, MrOS (<https://sleepdata.org/datasets/mros>), MESA, and SOF. The ClinSuZhou, ClinHuaiAn, and ClinRadar datasets contain sensitive clinical information and are available under restricted access due to institutional regulations and patient privacy considerations. To promote reproducibility, a small de-identified sample of clinical data is available in our GitHub repository (<https://github.com/zhuangzx127/ResSleepNet>) for demonstration purposes. Access to the full ClinSuZhou, ClinHuaiAn, or ClinRadar datasets can be requested for non-commercial academic research by contacting the corresponding author. Requests will be subject to institutional and ethical review, and approved users must sign a data use agreement. Access is granted to qualified researchers and is typically processed within 60 days. Source data supporting the plots and statistical analyses in the main figures are provided in the Source Data file accompanying this paper. Source data are provided with this paper.

Code availability

The source code and pre-trained model weights for ResSleepNet are publicly available at our GitHub repository: <https://github.com/zhuangzx127/ResSleepNet>. The repository includes the model architecture, inference pipeline, and testing scripts, enabling direct reproduction of the reported results. The training code and additional utilities are available for non-commercial academic use upon request, subject to a formal code usage agreement and compliance with institutional data governance policies. All released code is distributed under the MIT License.

References

1. Lim, D. C. et al. The need to promote sleep health in public health agendas across the globe. *Lancet Public Health* **8**, e820–e826 (2023).

2. The Lancet. Waking up to the importance of sleep. *Lancet* **400**, 973 (2022).
3. The Lancet Diabetes & Endocrinology Sleep: a neglected public health issue. *Lancet Diab. Endocrinol.* **12**, 365 (2024).
4. for Disease Control, C. & Prevention. Places: Local data for better health. <https://www.cdc.gov/places> (2024).
5. The Sleep Charity. New report: 14m+ undiagnosed sleep disorders damaging health, fuelling dangerous behaviour and costing billions (2023).
6. Zheng, N. S. et al. Sleep patterns and risk of chronic disease as measured by long-term monitoring with commercial wearable devices in the all of us research program. *Nat. Med.* **30**, 2648–2656 (2024).
7. Hafner, M., Stepanek, M., Taylor, J., Troxel, W. M. & Van Stolk, C. Why sleep matters-the economic costs of insufficient sleep: a cross-country comparative analysis. *Rand Health Quart.* **6**, 11 (2017).
8. Yu, F. Current status and thoughts of sleep medicine centers in China. *J. Apoplexy Nerv. Dis.* **41**, 238–240 (2024).
9. Society, C. S. R. Expert consensus on diagnosis and treatment of insomnia disorder in primary medical institutions. *Natl Med. J. China* **104**, 2296–2307 (2024).
10. Berry, R. B. et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.0*. American Academy of Sleep Medicine, Darien, Illinois <http://www.aasmnet.org/scoringmanual/> (2012).
11. Rosen, I. M. et al. Clinical use of a home sleep apnea test: an American Academy of sleep medicine position statement. *J. Clin. Sleep. Med.* **13**, 1205–1207 (2017).
12. Kapur, V. K. et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of sleep medicine clinical practice guideline. *J. Clin. Sleep. Med.* **13**, 479–504 (2017).
13. Sridhar, N. et al. Deep learning for automated sleep staging using instantaneous heart rate. *npj Digital Med.* **3**, 106 (2020).
14. Radha, M. et al. A deep transfer learning approach for wearable sleep stage classification with photoplethysmography. *npj Digital Med.* **4**, 135 (2021).
15. Kotzen, K. et al. SleepPPG-Net: A Deep Learning Algorithm for Robust Sleep Staging From Continuous Photoplethysmography. *IEEE J. Biomed. Health Inform.* **27**, 924–932 (2023).
16. Muliylil, S. Telemonitoring of patients with acute coronary syndrome. *Nature Medicine* (2024).
17. Baccelli, G., Guazzi, M., Mancina, G. & Zanchetti, A. Neural and non-neural mechanisms influencing circulation during sleep. *Nature* **223**, 184–185 (1969).
18. Pace-Schott, E. F. & Hobson, J. A. The neurobiology of sleep: genetics, cellular physiology and subcortical networks. *Nat. Rev. Neurosci.* **3**, 591–605 (2002).
19. Boudreau, P., Yeh, W.-H., Dumont, G. A. & Boivin, D. B. Circadian variation of heart rate variability across sleep stages. *Sleep* **36**, 1919–1928 (2013).
20. Fink, A. M., Bronas, U. G. & Calik, M. W. Autonomic regulation during sleep and wakefulness: a review with implications for defining the pathophysiology of neurological disorders. *Clin. Autonomic Res.* **28**, 509–518 (2018).
21. Malik, V., Smith, D. & Lee-Chiong, T. Respiratory physiology during sleep. *Sleep. Med. Clin.* **7**, 497–505 (2012).
22. Sowho, M., Amatoury, J., Kirkness, J. P. & Patil, S. P. Sleep and respiratory physiology in adults. *Clin. chest Med.* **35**, 469–481 (2014).
23. Oudiette, D. et al. Rem sleep respiratory behaviours match mental content in narcoleptic lucid dreamers. *Sci. Rep.* **8**, 2636 (2018).
24. Schreiner, T., Petzka, M., Staudigl, T. & Staresina, B. P. Respiration modulates sleep oscillations and memory reactivation in humans. *Nat. Commun.* **14**, 8351 (2023).
25. Sheriff, A. et al. Breathing orchestrates synchronization of sleep oscillations in the human hippocampus. *Proc. Natl Acad. Sci.* **121**, e2405395121 (2024).
26. Gaiduk, M. et al. Estimation of Sleep Stages Analyzing Respiratory and Movement Signals. *IEEE J. Biomed. Health Inform.* **26**, 505–514 (2022).
27. Luo, Y., Li, J., He, K. & Cheuk, W. A Hierarchical Attention-Based Method for Sleep Staging Using Movement and Cardiopulmonary Signals. *IEEE J. Biomed. Health Inform.* **27**, 1354–1363 (2023).
28. Huttunen, R. et al. A comparison of signal combinations for deep learning-based simultaneous sleep staging and respiratory event detection. *IEEE Trans. Biomed. Eng.* **70**, 1704–1714 (2022).
29. Hong, H. et al. Noncontact Sleep Stage Estimation Using a CW Doppler Radar. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **8**, 260–270 (2018).
30. Kwon, H. B. et al. Attention-Based LSTM for Non-Contact Sleep Stage Classification Using IR-UWB Radar. *IEEE J. Biomed. Health Inform.* **25**, 3844–3853 (2021).
31. Zhao, M., Yue, S., Katabi, D., Jaakkola, T. S. & Bianchi, M. T. Learning sleep stages from radio signals: A conditional adversarial architecture. In *Proceedings of the 34th International Conference on Machine Learning*, vol. 70 of *Proceedings of Machine Learning Research*, 4100–4109 (PMLR, 2017). <https://proceedings.mlr.press/v70/zhao17d.html>.
32. Toften, S., Pallesen, S., Hrozanova, M., Moen, F. & Grønli, J. Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®). *Sleep. Med.* **75**, 54–61 (2020).
33. Zhai, Q. et al. Machine Learning-Enabled Noncontact Sleep Structure Prediction. *Adv. Intell. Syst.* **4**, 2100227 (2022).
34. Zhang, G.-Q. et al. The National Sleep Research Resource: towards a sleep data commons. *J. Am. Med. Inform. Assoc.: JAMIA* **25**, 1351–1358 (2018).
35. Quan, S. F. et al. The sleep heart health study: Design, rationale, and methods. *Sleep* **20**, 1077–1085 (1997).
36. Chen, X. et al. Racial/Ethnic Differences in Sleep Disturbances: The Multi-Ethnic Study of Atherosclerosis (MESA). *Sleep* **38**, 877–888 (2015).
37. Blackwell, T. et al. Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the Osteoporotic Fractures in Men Sleep Study. *J. Am. Geriatrics Soc.* **59**, 2217–2225 (2011).
38. Spira, A. P. et al. Sleep-disordered breathing and cognition in older women. *J. Am. Geriatrics Soc.* **56**, 45–50 (2008).
39. Grote, L. The global burden of sleep apnoea. *Lancet Respiratory Med.* **7**, 645–647 (2019).
40. Benjafield, A. V. et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respiratory Med.* **7**, 687–698 (2019).
41. Zasadzińska-Stempniak, K., Zajackiewicz, H. & Kukwa, A. Prevalence of obstructive sleep apnea in the young adult population: a systematic review. *J. Clin. Med.* **13**, 1386 (2024).
42. Ohn, M. et al. Early life predictors of obstructive sleep apnoea in young adults: Insights from a longitudinal community cohort (raïne study). *Sleep. Med.* **110**, 76–81 (2023).
43. Vgontzas, A. N. et al. Age-related differences in the association of mild-to-moderate sleep apnea with incident cardiovascular and cerebrovascular diseases. *Sleep. Med.* **113**, 306–312 (2024).
44. Jordan, A. S., McSharry, D. G. & Malhotra, A. Adult obstructive sleep apnoea. *Lancet* **383**, 736–747 (2014).
45. Morokuma, S. et al. Deep learning-based sleep stage classification with cardiorespiratory and body movement activities in individuals with suspected sleep disorders. *Sci. Rep.* **13**, 17730 (2023).
46. Huynh, P. et al. Myocardial infarction augments sleep to limit cardiac inflammation and damage. *Nature* <https://www.nature.com/articles/s41586-024-08100-w> (2024).

47. Boulos, M. I. et al. SLEAP SMART (sleep apnea screening using mobile ambulatory recorders after TIA/stroke): A randomized controlled trial. *Stroke* **53**, 710–718 (2022).
48. Corral, J. et al. Conventional polysomnography is not necessary for the management of most patients with suspected obstructive sleep apnea. noninferiority, randomized controlled trial. *Am. J. Respiratory Crit. Care Med.* **196**, 1181–1190 (2017).
49. Ou, Y.-H., Tan, A. & Lee, C.-H. Management of hypertension in obstructive sleep apnea. *Am. J. Preventive Cardiol.* **13**, 100475 (2023).
50. Herth, J., Sievi, N. A., Schmidt, F. & Kohler, M. Effects of continuous positive airway pressure therapy on glucose metabolism in patients with obstructive sleep apnoea and type 2 diabetes: a systematic review and meta-analysis. *Eur. Respiratory Rev.* **32**, 230083 (2023).
51. Dharmakulaseelan, L. & Boulos, M. I. Sleep apnea and stroke. *CHEST* **166**, 857–866 (2024).
52. Henry, O. et al. A model for sleep apnea management in underserved patient populations. *J. Prim. Care Community Health* **13**, 21501319211068969 (2022).
53. Wolpert, E. A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Arch. Gen. Psychiatry* **20**, 246–247 (1969).
54. Iber, C. et al. The aasm manual for the scoring of sleep and associated events: Rules, terminology and technical specifications. 1st ed. *American Academy of Sleep Medicine* (2007).
55. Berry, R. B. et al. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.6*. American Academy of Sleep Medicine, Darien, Illinois <http://www.aasmnet.org/scoringmanual/> (2020).
56. Yu, B. et al. WiFi-Sleep: Sleep Stage Monitoring Using Commodity Wi-Fi Devices. *IEEE Internet Things J.* **8**, 13900–13913 (2021).
57. Konno, K. & Mead, J. Measurement of the separate volume changes of rib cage and abdomen during breathing. *J. Appl. Physiol.* **22**, 407–422 (1967).
58. Sabil, A. et al. Diagnosis of sleep apnea without sensors on the patient's face. *J. Clin. Sleep. Med.* **16**, 1161–1169 (2020).
59. Kogan, D., Jain, A., Kimbro, S., Gutierrez, G. & Jain, V. Respiratory inductance plethysmography improved diagnostic sensitivity and specificity of obstructive sleep apnea. *Respiratory Care* **61**, 1033–1037 (2016).
60. Rajesh, K. N., Dhuli, R. & Kumar, T. S. Obstructive sleep apnea detection using discrete wavelet transform-based statistical features. *Computers Biol. Med.* **130**, 104199 (2021).
61. Liu, H., Cui, S., Zhao, X. & Cong, F. Detection of obstructive sleep apnea from single-channel ecg signals using a cnn-transformer architecture. *Biomed. Signal Process. Control* **82**, 104581 (2023).
62. Tang, L. & Liu, G. The novel approach of temporal dependency complexity analysis of heart rate variability in obstructive sleep apnea. *Computers Biol. Med.* **135**, 104632 (2021).
63. Hu, S., Wang, Y., Liu, J. & Yang, C. Personalized transfer learning for single-lead ecg-based sleep apnea detection: exploring the label mapping length and transfer strategy using hybrid transformer model. *IEEE Trans. Instrum. Meas.* **72**, 1–15 (2023).

Acknowledgements

This work was supported by the Key Program of the National Natural Science Foundation of China (Grant No. 62431013 to H.H.), the National Key Research and Development Program of China (Grant No. 2020YFC2005300 to H.H.), the Joint Funds of the National Natural Science Foundation of China (Grant No. U24A20230 to E.W.), and the

National Natural Science Foundation of China (Grant Nos. 62401262 to B.X., 62301568 to Q.A., and 82370089 to J.X.).

Author contributions

Z. Zhuang, B.X., Q.A., and H.H. conceived the idea of remote sleep health management based on nocturnal respiratory signals and deep learning. Z. Zhuang, B.X., Q.A., and H.H. developed the deep learning models and algorithms. Z. Zhuang, H.C., Y.Z., X.Y., and H.H. developed the remote sleep health management platform and the non-contact radar acquisition system. R.C. and E.W. provided the anonymized ClinSuZhou dataset. J.X., N.D., X. Cui, and E.W. provided the anonymized ClinHuaiAn dataset. R.C., J.X., and E.W. provided the anonymized ClinRadar dataset. M. Wang and J. Xin provided support for cohort studies and data analysis. Z. Zhuang, Y.Z., Y.X., and Y. Li processed and cleaned the data. Z. Zhuang, Y.X., and Y. Li performed the experimental validation. Z. Zhuang, B.X., Q.A., C. Fu, and H.H. conducted the data analysis. Z. Zhuang and Y.X. generated the figures. Z. Zhuang, B.X., Q.A., and H.H. wrote the original manuscript. C. Fu, X. Zhu, M. Peng, and H.H. supervised the work. All authors reviewed and approved the final manuscript.

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/s41467-025-64340-y>.

Correspondence and requests for materials should be addressed to Mugen Peng or Hong Hong.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://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