PNAS RESEARCH ARTICLE | ENGINEERING

A skin-interfaced wireless wearable device and data analytics approach for sleep-stage and disorder detection

Yayun Du[a,1], Jianyu Gu[a,1], Shiyuan Duan[a,1], Jacob Trueb[a], Andreas Tzavelis[a,b,c], Hee-Sup Shin[a], Hany Arafa[a], Xiuyuan Li[d], Yonggang Huang[d], Andrew N. Carr[e], Charles R. Davies[f,g,2], and John A. Rogers[a,b,c,h,i,j,2]

Affiliations are included on p. 11.

Contributed by John A. Rogers; received January 16, 2025; accepted April 21, 2025; reviewed by Roozbeh Jafari and W. Hong Yeo

Accurate identification of sleep stages and disorders is crucial for maintaining health, preventing chronic conditions, and improving diagnosis and treatment. Direct respiratory measurements, as key biomarkers, are missing in traditional wrist- or finger-worn wearables, which thus limit their precision in detection of sleep stages and sleep disorders. By contrast, this work introduces a simple, multimodal, skin-integrated, energy-efficient mechanoacoustic sensor capable of synchronized cardiac and respiratory measurements. The mechanical design enhances sensitivity and durability, enabling continuous, wireless monitoring of essential vital signs (respiration rate, heart rate and corresponding variability, temperature) and various physical activities. Systematic physiology-based analytics involving explainable machine learning allows both precise sleep characterization and transparent tracking of each factor's contribution, demonstrating the dominance of respiration, as validated through a diverse range of human subjects, both healthy and with sleep disorders. This methodology enables cost-effective, clinical-quality sleep tracking with minimal user effort, suitable for home and clinical use.

wearable technologies | sleep | machine learning

Significance

This work introduces an approach to continuous sleep monitoring using low-power multimodal mechanoacoustic (LMA) sensors. Unlike traditional sleep diagnostics and existing wearable technologies, LMA directly measures respiratory biomarkers. The systematic signal processing pipeline extracts features of activities, respiratory and heart rate variability, essential for detecting sleep stages and disorders. Paired with an interpretable machine learning (ML) model, LMA-SleepNet, this approach delivers accurate, generalized, and personalized sleep analysis. Its superior performance in distinguishing sleep stages and apnea events highlights the translational potential of this technology for continuous sleep monitoring in both clinical and home settings, offering broad applicability across diverse populations.

Humans spend roughly one-third of their lives sleeping. The quality and duration of sleep are critical to physical, mental, and social well-being. In 2022, the American Heart Association introduced an updated metric called Life's Essential 8, with sleep as a key factor in assessing cardiovascular health (1). Insufficient and low-quality sleep leads to decreased labor productivity and higher mortality rates, causing an annual economic loss of $411 billion, a figure expected to rise to $467 billion by 2030 (2). Obstructive Sleep Apnea (OSA), a condition where airway obstruction leads to interruptions in breathing, poses a significant economic burden, estimated at $150 billion annually, alongside profound impacts on quality of life and life expectancy (3). Despite the clear need, challenges remain in effective detection with minimal user burden. Polysomnography (PSG) (4), the clinical standard for both monitoring sleep and detecting sleep apneas, involves sensors for electroencephalograms (EEGs), electrooculograms (EOGs), submental electromyograms (EMGs), electrocardiograms (ECGs), nasal pressure, pulse oximetry, and so on. Such partially invasive measurements cause hygiene challenges and require specialized hospital facilities with professional equipment and certified sleep technologists. Moreover, the collection of hard-wired sensors and electronics that attach across the body limits the reliability of the measurements and disrupts natural patterns of sleep. The consequences are levels of physical discomfort and mental stress that can lead to inaccurate assessments of sleep metrics (e.g., the "first night effect") (5).

The latest wearable and portable technologies support at-home sleep studies but fall short in detailed sleep-stage classification (SSC) and OSA detection due to the lack of convenient sensors for monitoring respiratory activities and the underexplored role of these activities in such applications. Wrist-worn devices like Philips ActiWatch (6), Fitbit (6), Apple Watch (6), WHOOP (6), Garmin Forerunner (6), and Polar Vantage (7) primarily use accelerometers (ACC) in inertial measurement units (IMUs) or/and heart rate (HR) metrics but lack capabilities in respiratory monitoring. Extensive research exists on assessments of sleep based on subsets of ECG, temperature, ACC, photoplethysmography, and radar measurements (8–14). Traditional microphones are unreliable due to sensitivity to environmental noise and privacy concerns (15), and two-channel in-ear microphones, while reducing noise, lack scalability for diverse populations, cause discomfort and vulnerability to movement, and fail to detect OSA (16). Other products, including ResMed S+ (wristwatch) (6), EarlySense Live (undermattress device) (6), SleepScore

Author contributions: Y.D., J.G., S.D., A.N.C., C.R.D., and J.A.R. designed research; Y.D., J.G., S.D., J.T., A.T., H.-S.S., H.A., X.L., Y.H., A.N.C., C.R.D., and J.A.R. performed research; Y.D., J.G., S.D., J.T., A.N.C., C.R.D., and J.A.R. analyzed data; Y.D., J.G., S.D., J.T., A.N.C., C.R.D., and J.A.R. wrote the paper.

Reviewers: R.J., MIT; and W.H.Y., Georgia Institute of Technology.

The authors declare no competing interest.

[1]Y.D., J.G., and S.D. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: charles.davies@carle.com or jrogers@northwestern.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2501220122/-/DCSupplemental.

Published June 6, 2025.

Max (portable) (6, 17), Oura Ring (7, 17), and Somfit (forehead-attached) (7), assess blood oxygen saturation but cannot track respiration rate (RR), RR variability (RRV), or respiratory effort, critical for detecting OSA events. The latest Apple watch only reports indirect OSA detection by sensing limb movements caused by breathing disturbances, limiting its accuracy and timeliness (18). Across most devices, loose skin contact and susceptibility to movements further reduce measurement reliability.

In sleep studies, analytics often focuses on HRV, which reflects autonomic nervous systems (ANS) activity and overall cardiovascular function. While HRV correlates with sleep stages, especially in the frequency domain (19), the development and correlation of RRV features and nonlinear HRV features (20) with sleep stages remain largely unexplored. This gap calls for physiology-supported feature development. Moreover, existing black-box ML models lack interpretability, obscuring the contribution of individual features, emphasizing the importance of explainable approaches.

This work addresses these shortcomings in sensor technologies and data analytics for characterizing sleep. Specifically, a multimodal, compact, soft, skin-interfaced, low-power mechanoacoustic (LMA) sensor (15, 21) continuously monitors key sleep-related parameters, particularly integrated cardiorespiratory activities. First, unlike commercial wearables, the LMA sensor is conformable, durable, and uniquely multimodal, capturing breathing effort, chest motion, and heart sounds-induced vibrations into a single device, enabling reliable OSA detection. This synchronized respiratory and cardiac monitoring allows deeper exploration of cardiorespiratory correlation, e.g., the effect of OSA on heart health, including HRV. Second, development of a comprehensive set of physiology-supported HRV, ACC, and RRV features from an IMU, integrated with the explainable LMA-SleepNet ML model (abbreviated as LMASN hereafter), allows superior capabilities in sleep characterization. The LMASN was evaluated on a clinical dataset from a diverse population and cross-validated against statistical methods and commonly used black-box ML algorithms. Results show that the LMASN outperforms state-of-the-art commercial products in SSC and even surpasses research platforms that rely on multiple sensors. The ML architecture tracks feature importance, revealing that ACC and RRV features dominate SSC, with RRV features outperforming the widely acknowledged HRV features in differentiating deep sleep stages. This observation underscores the significance of our contribution in using LMA to measure respiration. LMASN analyzes individual multinight sleep data, achieving human-comparable detection of rapid eye movement (REM) stage, a critical sleep stage with significant medical importance, demonstrating its effectiveness for personalized AI in sleep studies. This approach demonstrates potential for using LMA in sleep studies postsurgery or following OSA treatment, where PSG is unsuitable due to skin sensitivity. Third, its strong performance in OSA and snoring detection suggests broad applicability for continuous at-home sleep monitoring using a single, simple device.

Results

Study Participants. The results presented here correspond to data collected through 43 sleep nights from 35 adults, obtained in a clinical sleep laboratory at Carle Foundation Hospital, Urbana, IL. The study includes 26 healthy individuals and 9 with sleep apnea (apnea-hypopnea index: 5.2 to 39.6). Of the participants, 22 are women, 4 of whom are diagnosed with OSA. Ages range from 19 to 75 y, and BMIs span from 21.1 to 55.1, covering categories of healthy (18.5 to 24.9), overweight (25.0 to 29.9), and obese (>30.0), with a diversity of ethnicities. *SI Appendix*, Table S1 presents details. The participant group satisfies the three main goals for validating the sensor technology and data analysis approaches introduced here, specifically to: a) classify sleep stages (awake, NREM, REM) across a broad range of subjects; b) capture the sleep patterns of individual subjects; c) identify the effects of apnea events on sleep patterns. *SI Appendix*, Fig. S1*A* illustrates the setup for each participant, which includes a clinical PSG system and our LMA sensor. The former (2017 AASM standards) (4) captures EEG at frontal (F4), central (C4), and occipital (O2) sites, with contralateral mastoid reference electrodes (A1, A2). Additional measurements include left and right EOGs, submental EMGs, and ECGs, nasal pressure signals [pressure transducer airflow (PTAF)] and nasal/oral thermistor readings, pulse oximetry data, thoracic and abdominal excursion measurements (inductance plethysmography), and in certain cases, pressure from a CPAP machine. Labels for sleep stages (NREM N1, NREM N2, NREM N3, Wake, REM), hypopneas, and apneas follow from assignments by registered polysomnographic technologists at each 30-s epoch, following AASM criteria, assisted using commercial software (Natus SleepWorks PSG software). A sleep laboratory technician manually documents changes in body orientation. The electrical and mechanical characteristics of the LMA sensors include safe and efficient power supply systems with wireless charging, Coordinated Universal Time (UTC)-based timestamping, dynamic data rate estimation, Bluetooth Low Energy (BLE) Quality of Service (QoS)-adaptive automatic data storage switching between onboard memory and remote cloud, embedded filters and mathematical operations, and a two-phase data compression ratio reduces memory and battery power by as much as six times.

Mechanical Design for Enhanced Sensitivity, Durability, and Comfort. Fig. 1*A* depicts an LMA sensor mounted on the base of neck, covering both suprasternal notch (SN) and sternal manubrium (SM), as shown for a representative participant in *SI Appendix*, Fig. S1. This anatomical site, experimentally verified (*SI Appendix*, Fig. S1 *D* and *E*), offers a secure, comfortable location resistant to sleep movements, enabling capture of diverse cardiopulmonary signals and other parameters of interest with a high-bandwidth triaxis accelerometer. These data include core body movements, and diverse thoracic, airway, and oral functions, capturing respiration, speech, coughing, breathing, and teeth grinding (*SI Appendix*, Fig. S1 *B–E* and *Note* S1). The sensor operation on diverse participants under varying conditions, including single-night sleep sessions, exhibits no noticeable data loss or need for reapplication. A prior study (22, 23) confirms that secure attachment of adhesives is possible even during swimming, demonstrating its impermeability and robustness. The adhesive minimizes the risk of skin irritation or allergies, addressing concerns about movement, sweating, and skin impact.

Fig. 1*B* presents an exploded view schematic illustration of the device, featuring a triaxial-ACC IMU mounted on a structurally reinforced extruded island to improve sensitivity while maintaining softness and resistance to bending and peeling during use, as detailed in *SI Appendix*, Fig. S2 and *Note* S2, compared to those reported elsewhere (14, 21). The device records data, including the user's body temperature through the die sensor inside ISP1807, using a Bluetooth 5.0 modules with a built-in antenna. The data pass into an onboard memory unit for subsequent wireless transfer to a BLE-enabled device (e.g., phones, pads, and computers), and upload to a cloud server.

Autonomic neurons within the hypothalamus and brainstem (Fig. 1*C*, pink) facilitate communication between cortical and

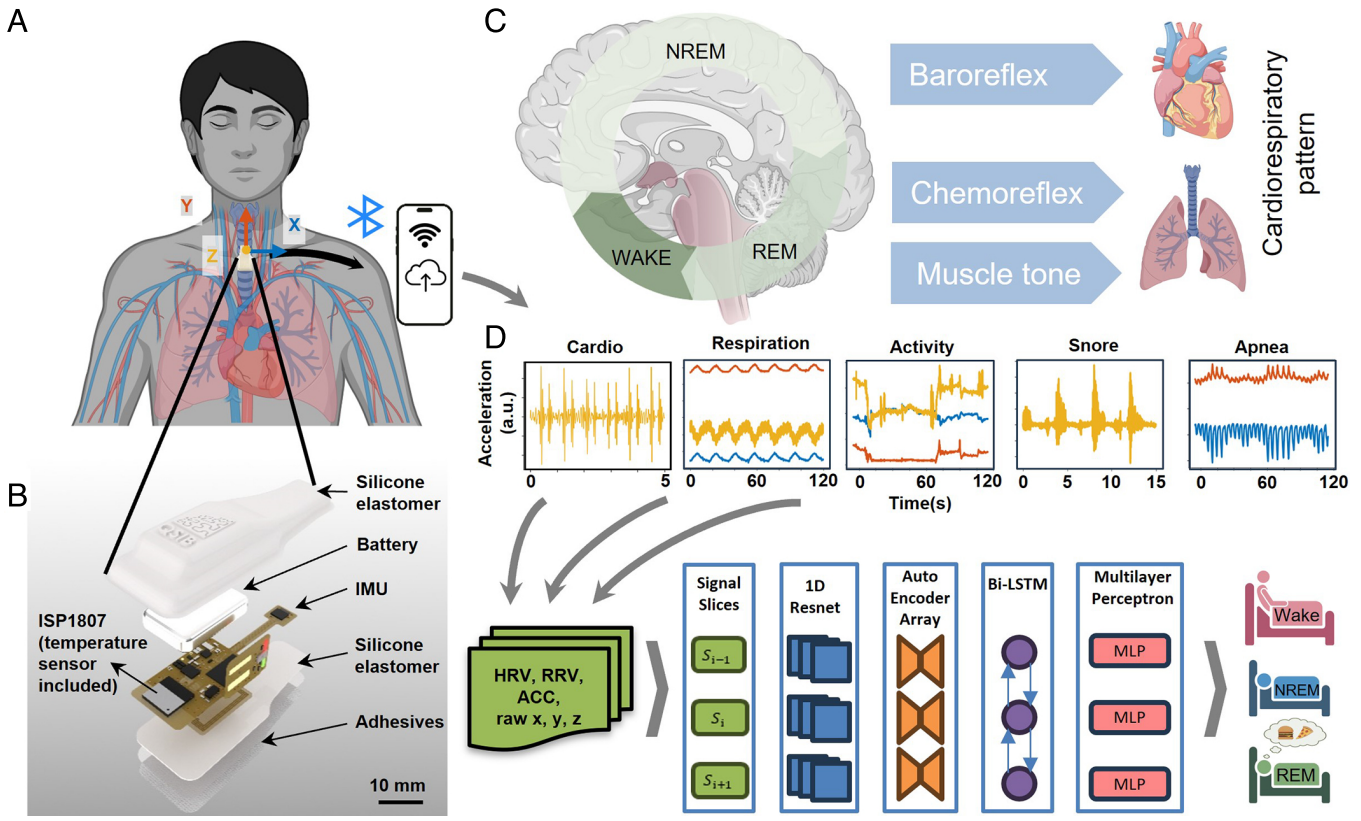


Fig. 1. Image and schematic representation of the LMA device, cloud platform, representative sleep clinical data captured at the SN, sleep stage/cardiorespiratory signal correlation, representative features derived from this correlation by our automated digital signal processing algorithms, and their use in our LMASN architecture. (*A*) Schematic illustration of the operational flow of the system, which consists of a device affixed to the base of the neck, with one end aligned with the SN and the other with the SM, user-friendly interfaces for real-time signal observation and a cloud-based database. (*B*) Exploded view schematic illustration of the main components, connection layout, and enclosure structure, with the IMU placed on an isolated island to improve sensitivity and durability. (*C*) Throughout sleep cycles, the brain modulates physiological signals via multiple pathways. (*D*) An overview of the signal processing pipeline, magnified windows with raw clinical data samples for various activities, extraction of HRV, RRV, ACC features, as partial inputs to proposed LMASN ML model for sleep-stage classification.

subcortical pathways, thereby influencing heart rate and respiration (24, 25). Sleep stages such as rapid eye movement (REM), non-REM (NREM), and wakefulness can be identified from changes in these and other activities. Although raw data from our devices include diverse patterns (Fig. 1*D*), features related to ACC, RRV, and HRV are most relevant for precise sleep-stage identification. These features serve as inputs to ML models described subsequently.

Algorithms for Capturing Cardiopulmonary and Activity Features. As shown in the flowchart in Fig. 2*A*, a deterministic digital signal processing (DSP) pipeline extracts HR, RR, and ACC from the IMU data. These results form the basis for implementing 32 HRV, 79 RRV, and 35 ACC features, across time, frequency, and nonlinear domains, as detailed in *SI Appendix*, Tables S2–S6. Frequency-domain HRV and RRV features capture the spectral characteristics of cardiac cycles and respiratory behaviors. Considering the complex nonlinear interactions of hemodynamic, electrophysiological, humoral variables, and ANS activity in regulating heart and breathing rates, nonlinear evaluations of HRV and RRV such as those based on Poincaré plots and detrended fluctuation analyses offer valuable insights into their correlation with sleep stages (20, 26–29). HRV, RRV, and ACC features reflect the activities of skeletal muscle tone (30, 31), respiratory drive (24), ANS (32), and baroreflex (33), as explained in *SI Appendix*, Table S7.

Extraction of HR begins with forward and backward bandpass filtering of the z-axis ACC data (*SI Appendix*, Fig. S3*A*), with cutoff frequencies of 10 and 30 Hz (in *SI Appendix*, Fig. S3*B*). This filtering captures the first (S1) and second (S2) heart sounds, confirmed by short-term Fourier transform spectrograms and differentiated from snoring and breathing signals across diverse subjects. The Shannon energy transformation (*SI Appendix*, Fig. S3*B*) then reduces low-amplitude noise and emphasizes rapid changes and signal features (*SI Appendix*, Fig. S3*C*). Considering the typical 100 ms separation between S1 and S2, and the average interval between cardiac cycles (34), a 1D Gaussian distribution kernel with a 100 ms window is used on the Shannon energy to generate envelopes (*SI Appendix*, Fig. S3*D*) that cover 95% of the distribution, smoothing the waveform and highlighting variations. The highest peaks generally correspond to S1 features, while the algorithm merges peaks within a 250 ms window for other cases (*SI Appendix*, Fig. S3*E*).

HRV captures cardiovascular dynamics over specific time intervals, influenced by heart–brain interactions and the ANS, with variability across sleep stages (19). As reported in previous studies (35), time-domain HRV features measure variability within 5-min heartbeat interval windows. Certain HRV spectral features are important in identifying sleep stages (19). HRV metrics follow from analysis of S1 intervals across time, frequency, and nonlinear domains. HRV is sensitive to abnormal heartbeats, where as little as 1% of ectopic beats can be significant (36). The analysis follows previous schemes to remove outlier and

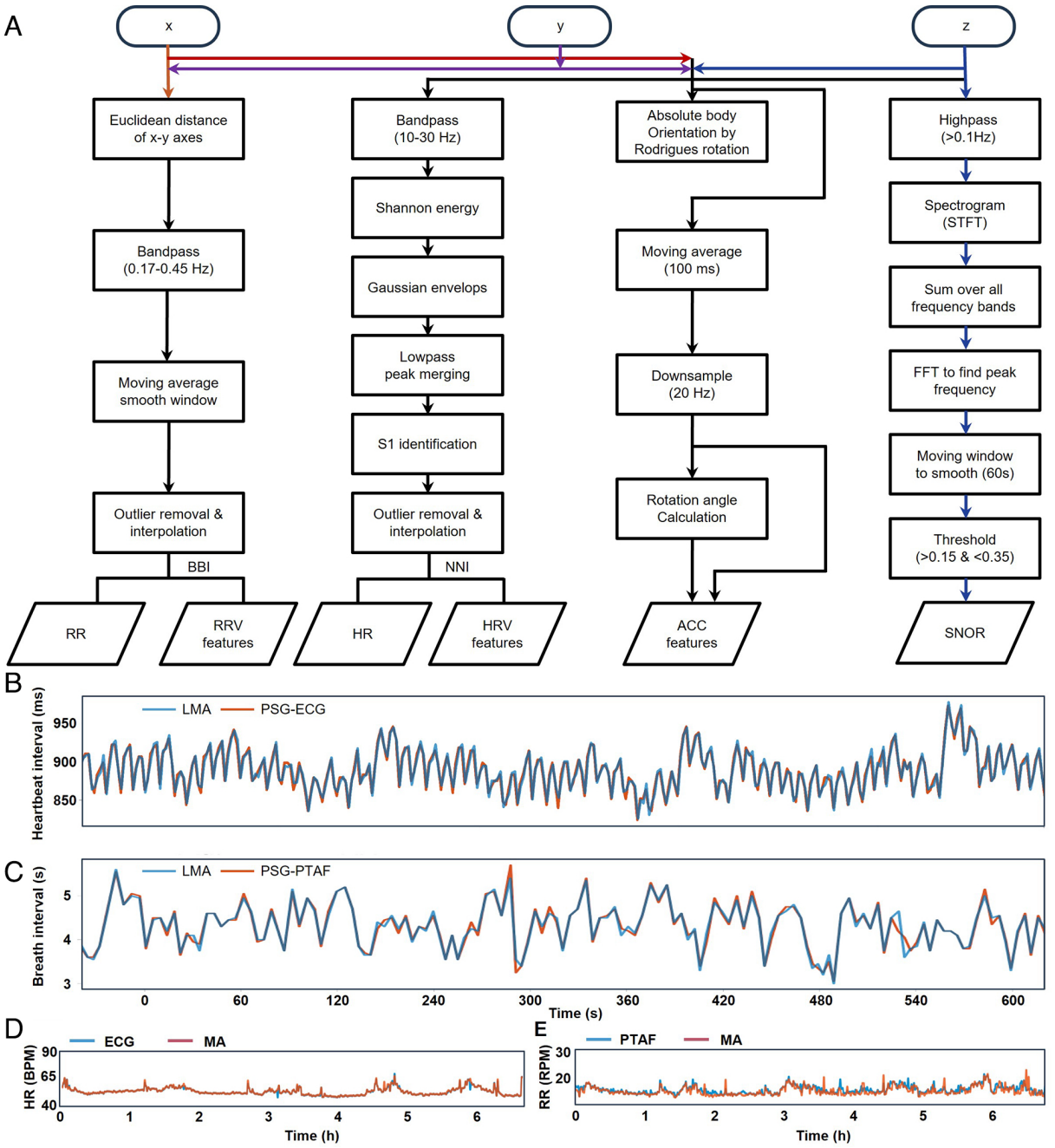


Fig. 2. Flow diagram of signal processing and corresponding results from LMA sensor and PSG from healthy subjects and subjects with sleep problems. (*A*) Block diagram of postprocessing algorithms for RR, HR, and ACC features, and SNOR detection; red, purple, blue, and black arrows indicate the flow of x-, y-, z-axis, and triaxial accelerometer data respectively. (*B*) Comparison of five-minute signals: interbeat intervals (NNI) from raw z-axis LMA vs. PSG-ECG and (*C*) Breath intervals (BBI) extracted from PSG-PTAF vs. x- and y- axis accelerometer data. Comparisons of (*D*) HR and (*E*) RR determined from one-night LMA and PSG (ECG for HR and nasal PTAF for RR) recordings.

ectopic beats by discarding intervals outside the 400 to 1,500 ms range, removing intervals that deviate beyond three SDs from the average across the sleep night, and excluding intervals that vary by 20% or more from the previous interval. A quadratic spline interpolation fills in gaps left by removal.

As a measurement of respiratory activity, the Euclidean distance between the x- and y-axis ACCs enhances the sensitivity to chest movements and minimizes interference from cardiac contributions associated with z-axis ACC. A median filter smooths the resulting signal, prior to down-sampling to 20 Hz. An upward signal corresponds to inhalation, aligning with the upward direction in PTAF readings. Breath intervals follow from measurements of peak-to-peak intervals, corresponding to troughs in the PTAF signal. RRV features include both breath effort- and rate-related metrics, as described elsewhere (31, 37). For signals reported here, effort associated with inhalation and exhalation follows from the absolute cumulative sum of these periods in the respiratory signal, as shown in *SI Appendix*, Fig. S4. Variability metrics for these features include the minimum, maximum, median absolute deviation (MAD), coefficient of variance [CV = MAD/median(x)], SD, variance, etc. In addition to RRV features calculated from every 5-min window, 15-min RRV features are also implemented to accommodate the fact that HR during sleep is 3 to 5 times higher than RR and to

capture long-range variations. Statistical and mathematical ACC features quantify both large- and small-scale movements. These include metrics like the total degree of rotation and the count of rotation events exceeding a specified threshold over time, as well as Pearson's correlation coefficients and kurtosis, which capture the distributions and linear relationships of movement patterns.

Snore Detection with Frequency Analysis and YOLO. Frequency analysis and pattern recognition serve as the basis for characterizing snoring patterns. Application of a high pass filter with a cutoff frequency of 0.1 Hz to z-axis ACC data followed by application of 60-s rolling windows with 50% overlap helps to distinguish snoring features from related signals such as coughing and talking. As shown in *SI Appendix*, Fig. S5, a Short-Time Fourier Transform (STFT) applied to each window segment (A) generates a (B) spectrogram. Evaluation of the total power in all frequencies creates a (C) time-variant power signal, $P_{MA}(t)$. Further analysis of $P_{MA}(t)$ via Fourier Transform reveals the temporal patterns of snoring in (D). The maximum power typically corresponds to breathing rhythms during snoring. In contrast, when heart sounds dominate in $P_{MA}(t)$, the maximum frequency usually matches the heart rate, thus differentiating from snoring sounds. This method is effective at identifying the onset of snoring but has difficulty predicting its duration. Open-sourced YOLOv8 algorithm (*Materials and Methods*) for object detection offers a solution with systematic metrics including accuracy [e.g., precision, recall, F1 score, and mean average precision (mAP)], speed (e.g., inference time), robustness, and versatility to detect snoring, realizing an mAP of 0.938. The mAP is a key comprehensive evaluation metric, with higher values indicating better performance. Some ground-truth signals and YOLO-predicted results are illustrated in *SI Appendix*, Fig. S6. Moreover, compared to computationally expensive STFT, YOLO effectively identifies time-domain patterns within and outside annotated bounding boxes and outputs the corresponding prediction confidence (37). Once trained, YOLO is deployable for real-time detection of snoring.

Comparisons Between Basic Extracted Features and PSG Data. Raw z-axis ACC data from the LMA sensor aligns with PSG-ECG signals, as displayed in *SI Appendix*, Fig. S7 *A–C*. Similarly, inhaling and exhaling durations from the LMA align with PSG-PTAF in *SI Appendix*, Fig. S7 *D–F*. Fig. 2 *B* and *C* demonstrates that the 10-min heartbeat and breath intervals measured by the LMA sensor match those calculated from PSG-ECG (R-R peaks) and PSG-PTAF (time difference between signal peaks). The HR and RR calculated during a representative sleep night, displayed in Fig. 2 *D* and *E*, using the methods shown in Fig. 2*A*, agree with those from PSG. *SI Appendix*, Fig. S7 *G* and *H* presents Bland–Altman plots of PSG-ECG R-R intervals (RRI) and LMA S1 intervals (*SI Appendix*, Fig. S7*G*) as well as LMA-Breath-to-Breath Interval (LMA-BBI) and PTAF-BBI signals (*SI Appendix*, Fig. S7*H*). Sensing of S1 intervals and BBI with LMA yields a mean difference of 0.13 ms and 0 s, respectively, paired with an SD of 8.00 ms and 0.18 s.

Qualitative Correlations of Features and Sleep Stages. Fig. 3*A* demonstrates the overall correlations of features RRV (e.g., RRV-mCVBB), HRV (HRV-mCSI), and ACC (ACC-GRot) features with sleep stages. Specifically, the mean coefficient of

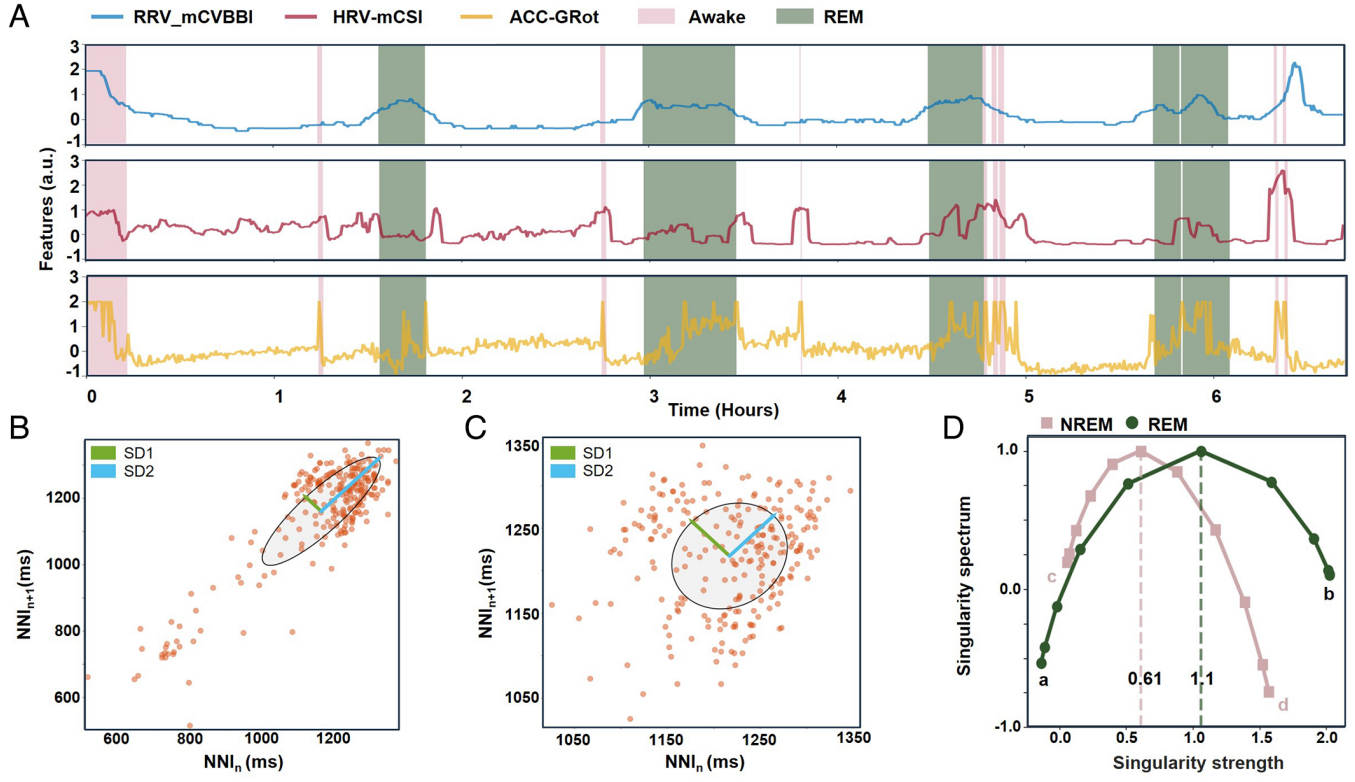


Fig. 3. Results from the LMA sensor and the PSG system from healthy and sleep-disordered subjects. (*A*) An overall result of feature correlation with sleep stages: Representative RRV feature (RRV-mCVBBI) shows a strong correlation with REM stages. The HRV feature (HRV-mCSI) and ACC feature (ACC-GRot) demonstrate high correlation with awake stages. Poincare plots (*B*) show consecutive NNI at awake stages and (*C*) NREM stages. (*D*) Singularity spectrum of RRV-MFDFA alpha1 peaks at NREM and REM stages. Dashed lines represent the MFDFA peak: singularity strength of the maximum fractal dimension. X-axis differences between points a and b, c and d are spectrum widths.

variance of the breath interval feature of RRV (RRV-mCVBB) correlates with REM sleep; modified Cardiac Sympathetic Index of HRV (HRV-mCSI) and gross amount of rotation (ACC-Grot) correlate with awake stages. Fig. 3 *B–D* provides a concise summary of the key medical findings supporting these correlations. Fig. 3 *B* and *C* displays Poincaré plots of NN-interval times series (NNI), the RRI time series with abnormal peaks removed. Physiologically, the SD1 width reflects parasympathetic activity while the SD2 length indicates sympathetic modulation (38). The Poincaré plot for awake stages (Fig. 3*B*), shows that points are more concentrated with clustering in the *Top Right* corner compared to NREM stages (Fig. 3*C*), suggesting decreased parasympathetic nervous system (PNS) and increased sympathetic nervous system (SNS) activities, consistent with medical findings (38, 39). The scattered points at the *Bottom* with smaller NNIs correspond to sharp awakenings from sleep. The narrow "comet" shape (Fig. 3*B*) represents nonrespiratory influences on HR, indicating SNS dominance. In contrast, the wide shape in Fig. 3*C* indicates respiratory components regulating HR and PNS dominance (40).

Fig. 3*D* shows one nonlinear feature, the multifractal spectrum RRV-MFDFA1 (27) during REM and NREM stages. This feature aids in understanding the distribution and variability of irregularities with the signal pattern. Both curves exhibit either left or right truncation, showing the multifractality of RRV-MFDFA1P in these stages. Distinct structures of these multifractal spectra such as peak positions and spectrum widths (x-axis differences between A and B, C, and D) highlight different long- and short-range fluctuation of RRV-MFDFA1 across sleep stages. *SI Appendix*, Fig. S8 *A–C* demonstrates that violin plots of representative RRV features, such as 15-min mCVBB, MCVRSV, and MFDFA1, are denser at lower values during NREM stages and more dispersed at higher values during REM stages. This behavior suggests increased respiratory variability during REM sleep, including breath interval and effort, as validated by technician's notes in *SI Appendix*, *Note* S3. Previous studies report that the HRV index of sympathovagal balance, LF/HF ratio, is highest during awake stages, progressively decreasing with deeper sleep stages, and reaching its lowest value during slow wave sleep (41, 42). Our data agree with these trends (*SI Appendix*, Fig. S8*D*) and further demonstrate that most RRV features (in *SI Appendix*, Fig. S8 *E–G*) more clearly follow this trend than most HRV features (*SI Appendix*, Fig. S9), better differentiating deep sleep (NREM N2 and N3) than the well-known HRV-LFHF. This observation suggests that BBI may more closely reflect the SNS/PNS balance, consistent with the understanding that respiratory pattern regulation, involving nucleus of the solitary tract (NTS), retrotrapezoid nucleus, and Pre-Bötzinger Complex (preBötC), diminishes during REM (43).

LMASN: An Explainable ML Model for Sleep-Stage Differentiation. The output from the ResNet (in *SI Appendix*, Fig. S10*A*) and DSP output features pass into an autoencoder (AE) array as shown in *SI Appendix*, Fig. S10*B*. Although the device includes a temperature sensor, temperature data are not useful in the sleep staging procedure due to its limited temporal resolution and susceptibility to environmental influences, which reduce its utility compared to cardiorespiratory features. AE is chosen by the dimension of weights its latent space set as Nf, such that each weight corresponds to the contribution/importance of each feature. The result yields an ML process that is interpretable, as detailed in *Materials and Methods*.

A weighted cross entropy (CE) loss function that aligns weights with the sample distribution in WNR (wake, NREM, REM) stages counteracts skewness in the label distribution. Fig. 4 *A–D* shows normalized confusion matrices of three-class WNR (Fig. 4 *A* and *B*) and two-class WS (wake, sleep) classifications (Fig. 4 *C* and *D*) by LMASN, using 40-fold cross-validation in a leave-one-out (LOOCV) manner. Frames A and C represent results with an unweighted CE loss, while B and D utilize weighted-CE loss. Each fold in turn selects one night as the validation set, with the rest as the training set. Notably, the results in Fig. 4*E* indicate that the weighted CE loss significantly improves precision in differentiating REM and Wake stages, with REM-Precision rising from 61.2% to 68.8%, albeit with a slight decrease in overall accuracy from 83.8% to 82.1%. This improvement is expected due to the similar and lower numbers of REM and Wake stages compared to NREM. Additionally, for WS classification, the weighted CE loss increases Wake-precision from 72.4% to 75.4%, with a roughly 3% decrease in WS-Accuracy, attributed to the significant data imbalance (Sleep: Wake $\approx$ 9:1) in our dataset. The decision to use weighted-CE loss thus depends on the specific research objective. Fig. 4*F* compares the performance of this work to commercial products (6, 7, 17), while Fig. 4*G* contrasts it with state-of-the-art multipiece wearable sensors (8, 9, 11, 12, 44), with details in *SI Appendix*, Tables S8 and S9.

This work outperforms other sleep studies using wearable devices, including multipiece sensor networks, in WNR classification. Fig. 4*H* shows results of the model using a broad range of metrics recognized in the sleep community (6, 45, 46), including Total Sleep Time (TST), Total Awake Time (TAT), Sleep Efficiency (SE), Sleep Onset Latency to the first epoch of sleep (SOL), Latency to Persistent Sleep of 10 continuous min (LPS), and Wakefulness After Sleep Onset 2 (WASO2), as detailed in the supplement note of ref. 45. The similarity of the SDs (SD) of these metrics between technician scorers (PSG) and LMASN underscores the robustness of LMASN. On an epoch-by-epoch basis, agreement averages for the same scorer with the method proposed by Chinoy et. al. (6) is 87% for REM (Fig. 4*I*). Between two different scorers (ANT1 and ANT2), these values are 87% for wakefulness (Fig. 4*J*). Awake and NEM stages show the most variability in scoring, as exemplified by detailed annotations from two well-trained scorers (ANT1–AKF and ANT2–MB in *SI Appendix*, Table S1) in Fig. 4*K*. *SI Appendix*, Fig. S11 shows boxplots of night-by-night sleep-stage agreement and kappa values for scorer pairs. The average ($\pm$SD) entire-night sleep-stage agreement for randomly chosen 9 subjects (subjects 4 to 8, 11, 15, 25, and 26) is 83% (12%) for ANT1 vs. ANT2, 88% (10%) for ANT1 vs. ANT1 and 86% (4%) for ANT2 vs. ANT2. Cohen's kappa values for interrater reliability are 0.92 (0.07), 0.94 (0.06), and 0.94 (0.02), all above 0.75, showing high reliability of our technician scores. Fig. 4*L* summarizes sleep measure from ANT1 and ANT2 with annotator 1 AKF reannotating the same recordings (ANT1s) of five subjects.

Respiration as a Major Contributor to Sleep-Stage Differentiation and Key Indicator for OSA Detection. Two methods yield consistent results for the contributions of Nf = 146 features to sleep-stage differentiation, both following the LOOCV approach. One metric of signal importance uses the gradients of LMASN, capitalizing on the gradient descent method for iterative loss function optimization. The derivatives of model predictions against each feature determine the gradients, for each of the two classification tasks, WNR and WS. Fig. 5 *A* and *B* presents the results, where higher gradients correspond to stronger influence on the predictions. For REM-NREM differentiation, RRV features (denoted by blue bars), e.g., RRV-MadBB, have prominent roles. For WS differentiation, ACC features (yellow

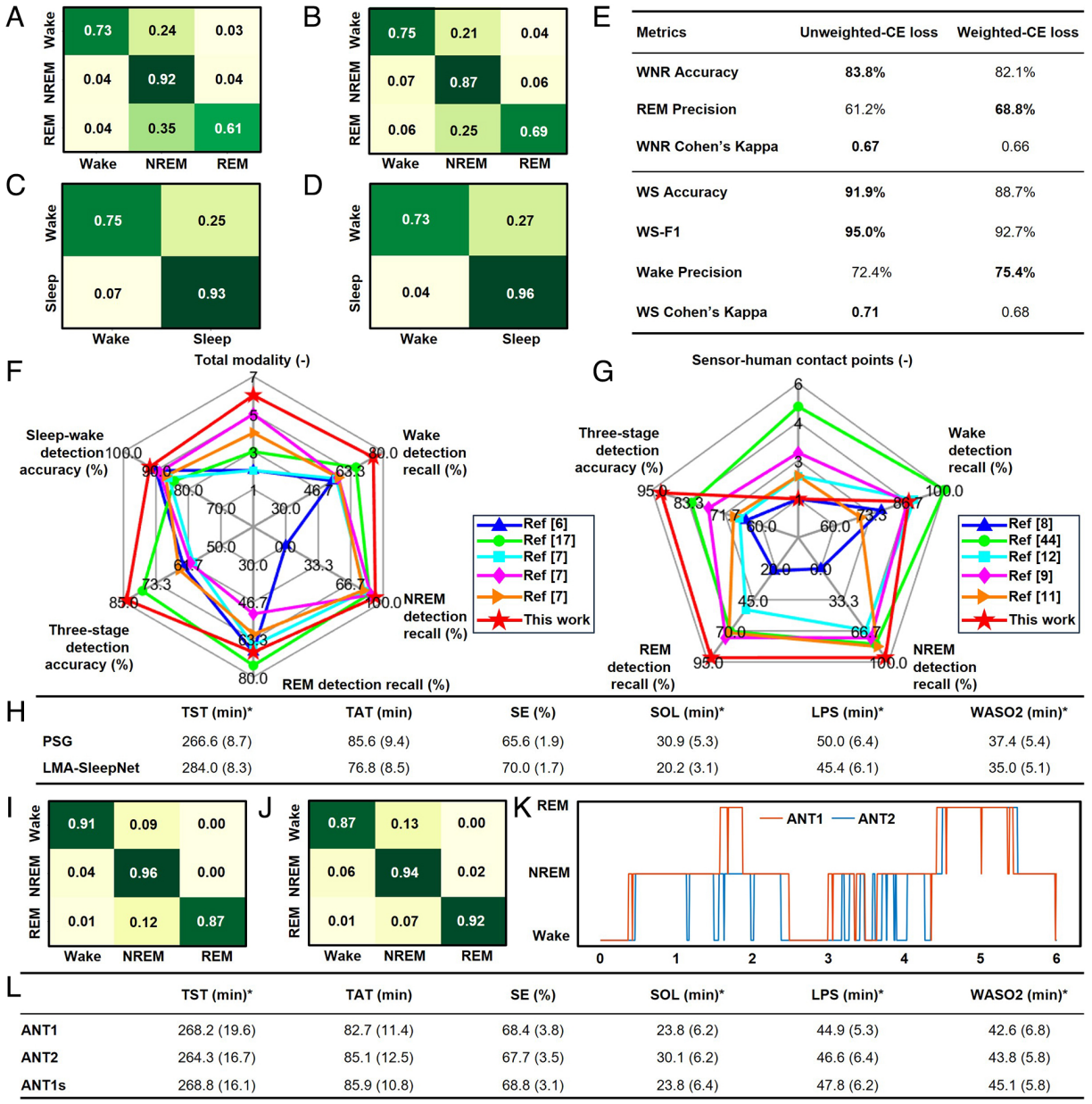


Fig. 4. Performance of technician-interpreted PSG data and LMASN for sleep-stage differentiation. Confusion matrices for (*A* and *B*) three-class WNR (wake, NREM, REM) and (*C* and *D*) two-class WS (wake, sleep) predictions by LMA-SleepNet with 40-fold cross-validation and (*A* and *C*) unweighted cross entropy (CE) loss vs. (*B* and *D*) weighted-CE loss functions. (*E*) Performance metrics of LMA-SleepNet in three-class and two-class cross-validation classification with unweighted CE and weighted-CE loss functions. Radar plots comparing performance parameters of (*F*) this work with commercial products and (*G*) this work with state-of-the-art multipiece wearables. (*H*) Summary sleep measures calculated from 40 PSG recordings and 40-fold LMA-SleepNet [mean ± SE of the mean (SEM)]. *Significant main effects by ANOVA or Kruskal–Wallis ($P < 0.05$), TST, total sleep time; TAT, total awake time; SE, sleep efficiency; SOL, sleep onset latency; LPS, latency to persistent sleep; WASO, wakefulness after sleep onset Duration of wakefulness from sleep onset (LPS) to the final epoch of sleep (stage 1, 2, 3/4, or REM). (*I* and *J*) Confusion matrices for two PSG annotators reannotate their own annotations [(*I*) ANT1 vs. ANT1s and (*J*) ANT2 vs. ANT2s]. (*K*) Representative sample highlighting annotation difference between two annotators (ANT1 vs. ANT2). (*L*) Summary sleep measure from two PSG sleep annotators' annotations (ANT1-AKF and ANT2-MB) for five patients, with Annotator 1 reannotating the same recordings (ANT1s).

bars) are most significant. This insight is important because the RRV and ACC features represent defining capabilities of the sensors used in the studies reported here, beyond traditional HRV-related characteristics (19, 41, 47). Further, Fig. 5 *A* and *B* shows that frequency features compute power across very low frequency (VLF) (0.0033 to 0.04 Hz), low frequency (LF) (0.04 to 0.15 Hz), and high frequency (HF) (0.15 to 0.40 Hz) bands. The LF/HF ratio, traditionally linked to PNS activity (31), often debated in its effectiveness for distinguishing NREM and REM sleep in recent studies (25, 41, 48), was not a distinctive feature in our research. Fig. 5 *C* and *D* summarizes the importance of RRV, HRV, and ACC features in REM-NREM differentiation (37.2%, 22.2%, and 40.6%) and WS differentiation (33.4%, 21.3%, and 45.2%). Although the total percentage contributions of ACC, RRV, and HRV features differ in REM-NREM and WS differentiations, their rank order remains consistent: 35 ACC features have the highest impact, followed by 79 RRV, and 32 HRV features. This finding demonstrates that feature importance is independent of the number of features within each category.

Violin plots, merging box and density plot features, provide additional insights as shown in Fig. 5*E* for four representative

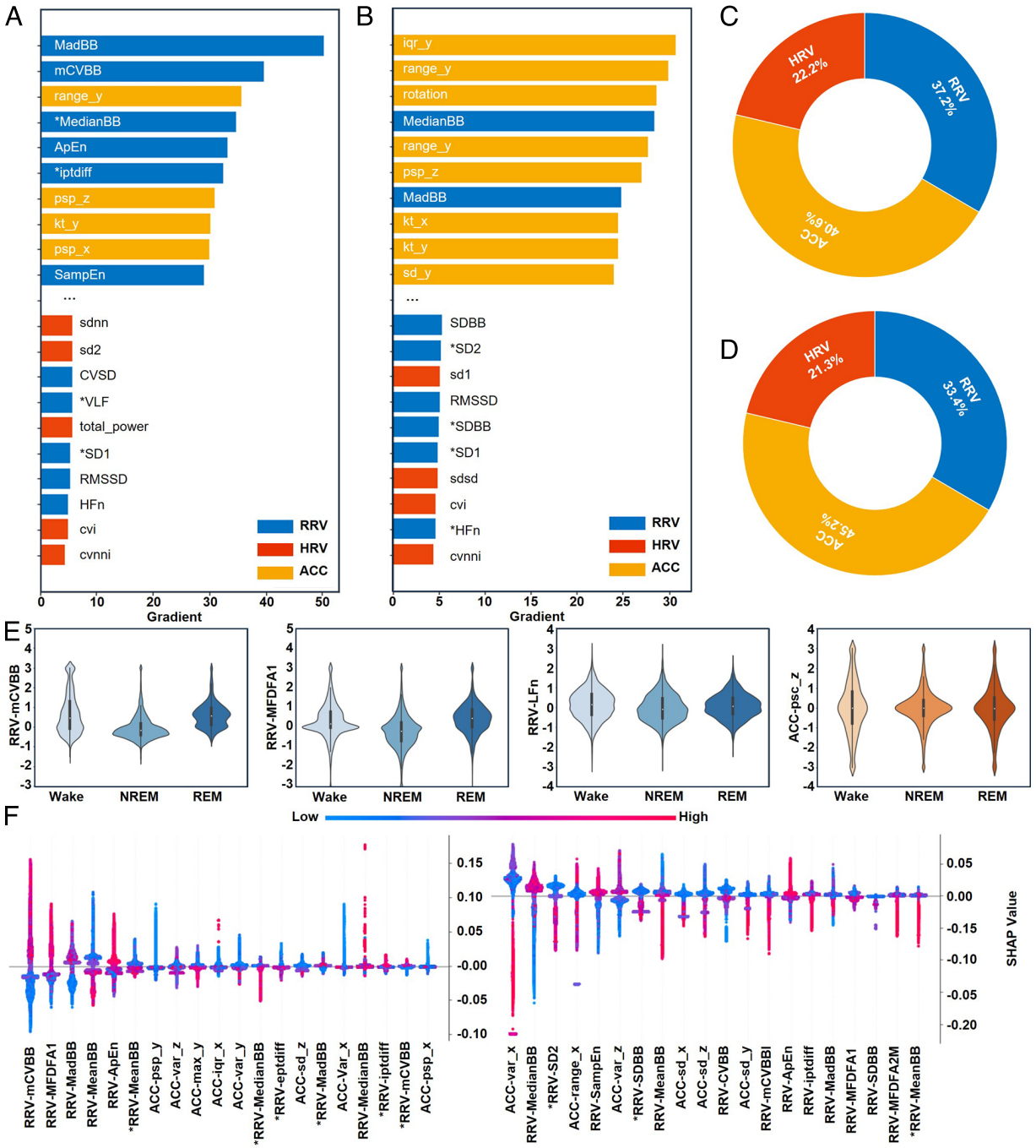
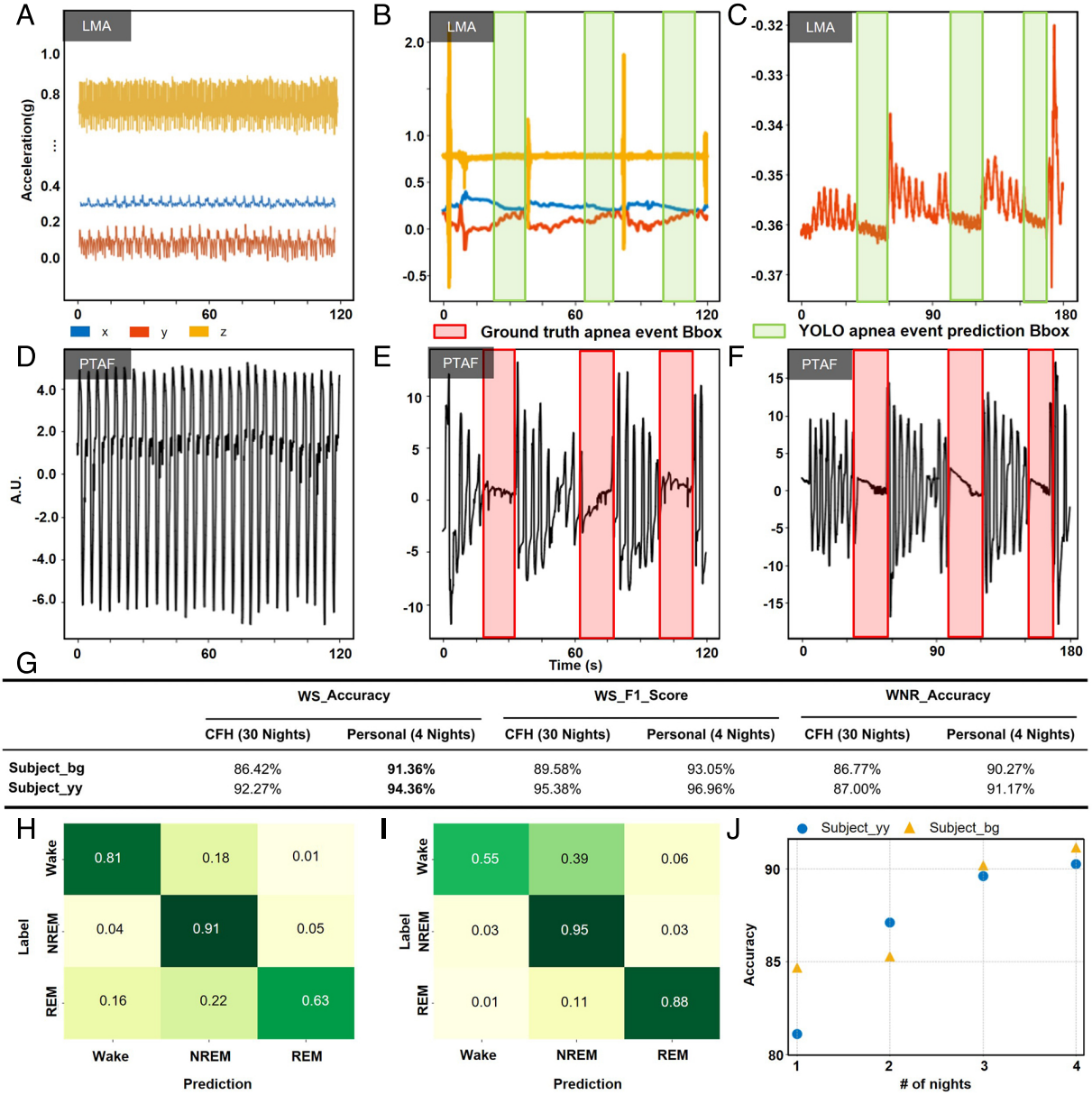


Fig. 5. Feature importance analysis with explainable ML models in sleep-stage classification. Significance of features in distinguishing between (*A*) REM-NREM and (*B*) sleep and wake stages calculated by the LMASN gradient-based method. (*C* and *D*) Donut charts quantifying the contribution of three key feature categories: RRV, HRV, and ACC for (*C*) REM-NREM and (*D*) sleep–wake stage differentiation. (*E*) Violin plots showcasing the distribution of four representative features over different sleep stages: two important features: RRV-mCVBBI and RRV-MFDFA1P from nonlinear analysis, as well as two unimportant features: *RRV − LFn* from frequency domain, and *ACC-psc_z* showing the distribution of activity features. (*F*) SHAP values of classifying WNR and WS stages using an XGBoost model. The stars in the *Upper Left* corner indicate that the features are assessed within 5-min windows, whereas other RRV features without stars are evaluated over 15-min windows.

features across WNR stages. The plots include two critical features, RRV-mCVBB and RRV-MFDFA1 from nonlinear analysis, and two least important features, RRV-LFn from the frequency domain, and ACC-psc_z. The plots display medians (white dots), interquartile ranges (thick gray bars), and the full data distribution (thin gray lines), with outliers identified by an interquartile range-based method. Kernel density estimations on each side of the gray line visualize the data distribution. These plots reveal distinct patterns, such as varied distributions of RRV-mCVBB across Wake, REM, and NREM stages, indicating its effectiveness in differentiating these stages. In contrast, RRV-LFn and ACC-psc_z, showing similar distributions and peak positions across stages, are less effective for differentiation. Frequency domain features are less critical. Moreover, *SI Appendix*, Fig. S12 shows that removing features ranked by their importance almost consistently degrades the performance of LMASN, indicating that each feature contributes unique information, thus eliminating the need for a feature selection process.

XGBoost is an ML algorithm known for its computational efficiency and ability to reduce overfitting, though often considered an unexplainable black box (49). SHAP values (50), arguably the most powerful method for explaining ML predictions, offer valuable insights into the feature-prediction correlation of XGBoost. The training process employs two binary XGBoost models with a LOOCV approach. Fig. 5*F* displays SHAP value plots for these models, where red dots denote higher feature values. In the *Left* figure, positive SHAP values indicate REM stages, while negative values indicate NREM stages. The increasing density of features from *Left* to *Right* suggests a decreasing relevance of feature changes for REM-NREM differentiation. A notable cluster of RRV-MCVBB points in blue (indicating smaller values) aligns with negative SHAP values, implying that lower RRV-MCVBB values favor the occurrence of NREM stages. ACC-psp_x points, mainly centered around 0, indicate a minimal impact on REM-NREM differentiation. In the *Right* plot, ACC-var_x is highlighted as a key feature, with higher values indicating Awake stages. Overall, the most influential features for REM-NREM and WS differentiation are consistent between LMASN and SHAP-XGBoost methods, illustrating the robustness of LMASN while retaining explainability.

Respiration is a key factor in differentiating sleep stages and a significant indicator for detecting OSA, which results from airway blockage due to relaxed throat muscles. This blockage causes noticeable changes in LMA signals around the SN area. Fig. 6 *A*–*F* compares LMA (Fig. 6 *A*–*C*) and PSG-PTAF (Fig. 6 *D*–*F*) signals across scenarios without OSA and snoring (Fig. 6 *A* and *D*), with OSA and during snoring (Fig. 6 *B* and *E*), and with OSA without snoring (Fig. 6 *C* and *F*). In comparing



	WS_Accuracy		WS_F1_Score		WNR_Accuracy	
	CFH (30 Nights)	Personal (4 Nights)	CFH (30 Nights)	Personal (4 Nights)	CFH (30 Nights)	Personal (4 Nights)
Subject_bg	86.42%	**91.36%**	89.58%	93.05%	86.77%	90.27%
Subject_yy	92.27%	**94.36%**	95.38%	96.96%	87.00%	91.17%

Fig. 6. Detection of personal sleep apnea events and sleep pattern and performance evaluation of LMASN on personal multinight data. Representative signal comparisons of personal sleep data: (*A*–*C*) LMA signals and (*D*–*F*) PSG-PTAF signals for cases (*A* and *D*) without snoring or obstructive sleep apnea (OSA), (*B* and *E*) with OSA during snoring, and (*C* and *F*) with OSA without snoring. (*G*) Performance metrics comparison of LMASN in three-class and two-class cross-validation classification, trained on data from 30 subjects at CFH, tested on single-night data from new subjects (male: Subject_gg and female: Subject_yy) vs. trained on four-night data from the same subjects with subsequent single-night testing. (*H* and *I*) Confusion matrices for LMASN using fivefold cross validation: trained on four-night data from the same subject, with subsequent single-night testing on (*H*) subject_bg and (*I*) subject_yy. (*J*) Prediction accuracy of LMASN based on training with varied numbers of nights' data and tested against one single night's data.

Fig. 6 *G* and *H* with Fig. 6 *E*, *F*, *I*, and *J*, notable differences are in larger signal amplitudes in both LMA and PSG-PTAF during snoring and clear cessation of chest movements in the *y*-axis (red data). Additionally, periodic interruptions in snoring, marked by 3 to 6 s intervals of silence (Fig. 6 *G*–*J*), and slight changes in LMA device orientation occur, linked to relaxation of tongue and throat muscles during OSA events. Similar to snoring detection, YOLOv8 is also used to detect OSA events.

Personalized Sleep Models: Total Subject Pool Performs Worse than Several Nights of Individual Sleep Data. Expanding the training sets by increasing the number and diversity of subjects can improve the generalizability and accuracy of LMASN. In certain use cases, however, personalized models might be preferred. Systems such as the Nox A1s (from Nox Medical) make a PSG sleep study feasible at home. Before head or neck surgeries, or before implanting a device to treat OSA, LMASN can be trained on LMA and PSG data collected at home. Postsurgery, when skin sensitivity may preclude the use of a full PSG device, the pretrained LMASN can be employed to predict sleep stages using only LMA, allowing for the monitoring of treatment response as stimulation parameters are adjusted. Studies to explore this possibility involve data collected across ten nights of data, five each from one female (Subject_yy) and one male (Subject_bg). Fig. 6*G* presents the averaged performance of LMASN for WNR and SW classifications based on training across 30 nights of clinical data from 30 patients at CFH and, separately, tested on one night from either Subject_yy or Subject_bg, contrasting with training across four nights of data from each individual and tested on another night from the same individual. The results indicate that a larger, more diverse training dataset does not necessarily outperform an individual dataset. For Subject_bg, there is a notable improvement in performance for WS classification, with accuracy rising from 86.42% to 91.36% and the F1 score increasing from 89.58% to 93.05%. Similarly, the accuracy for WNR classification improves from 86.77% to 90.27%. For Subject_yy, there is an approximate increase of 2.1% in WS accuracy, 4.2% in WNR accuracy, and 1.6% in the F1 score for WS classification. The reasonable performance improvement of LMASN suggests minimal overfitting and demonstrates its scalability in handling both small and large datasets, critical and challenging aspect of ML. This success is attributed to the data preprocessing and feature engineering previously described and highlights the effectiveness of personalized AI in identifying personal sleep patterns.

Fig. 6 *H* and *I* show the confusion matrices for LMASN, using a fivefold cross-validation method. Training relies on four nights of data from the same subject and then tested on a single night, for Subject_bg (Fig. 6*H*) and Subject_yy (Fig. 6*I*). For Subject_yy, the REM stage detection accuracy reaches 88%, as exemplified in *SI Appendix*, Fig. S13. This result surpasses the accuracy of individual scorers reannotating REM stage in PSG data (87% as shown in Fig. 4*I*), demonstrating that the proposed personalized approach can exceed human performance. The awake stage detection accuracy for Subject_yy is 55%. REM stage detection accuracy for Subject_bg is 63%. This lower accuracy can be partly attributed to data skewness; specifically, the proportions of awake stages in Subject_yy's total sleep stages and REM stages in Subject_bg's total sleep stages, which are approximately 6.7% and 10%, respectively. Therefore, we investigate the minimum number of nights needed for our LMASN to accurately capture personal sleep patterns and achieve a high (>70%) and human-competent accuracy in WNR classification. As depicted in Fig. 6*J*, the accuracy of WNR classification for both subjects increases with additional nights, but this rate of increase slows from 4% to below 1%, indicating a plateau in accuracy improvement. In addition, the varying nonlinear improvement in model performance across different individuals highlight individual variability.

Discussion

This paper presents a comprehensive hardware-software solution for sleep studies, featuring a low-power skin-interfaced sensor system with integrated analytics. The results include systematic signal processing approaches and the extraction of physiologically and statistically meaningful HRV, RRV, and activity features from a triaxial ACC dataset obtained from the LMA sensor. These features are inputs to a scalable, explainable ML model, enabling accurate sleep-stage detection across both large and individual multinight datasets. The platform also offers precise snoring and OSA detection using respiration-based ML. *SI Appendix*, Tables S8 and S9 show that LMASN outperforms other sleep studies using wearable devices, including sensor networks, in WNR classification. Unlike previous approaches that rely solely on IMU (ACC) or MA sensors, this work extracts key RRV and HRV features that reflect heart sympathetic innervation, baroreflex, chemoreflex, and muscle tone, all through IMU measurements. LMASN's AE-array structure enhances interpretability by tracking each feature's contribution to sleep-stage predictions. Respiratory features are more significant in SSC than traditionally emphasized HRV features.

Manual sleep-stage labeling introduces variability, as expert-labeled PSG data, despite standardized guidelines, can differ across scorers and laboratories. This subjectivity affects model performance, as ground-truth labels are not always absolute. Although our platform achieves accurate sleep staging with a minimalist sensor design, future work could explore how integrating additional modalities—when available—could further refine staging and reduce the reliance on manual scoring.

The analysis pipeline reported here can apply to any sensors that are capable of measuring or inferring HR, RR, or ACC, extending its potential beyond sleep studies to applications in precision healthcare and behavioral science. The LMA sensors automatically synchronize with UTC time, enabling synchronized sensor networks to capture multipoint physiological changes. Embedded filters and linear algebra operations support real-time processing and personalized adaptation. Moreover, integrating LMA with the YOLO ML model enables the detection of events like snoring and OSA, opening new opportunities for studying correlations between OSA and body positions or head and neck height and orientation. LMA sensors also enable continuous tracking of the impact of OSA on cardiovascular health, as reflected in HRV. Future integration of actuators may enable upper airway stimulation therapy for OSA without compromising concurrent HR and RR signal collection. While the current dataset includes nine patients with OSA, future works should expand to include more diverse populations and other sleep disorders, such as central apnea. Additionally, real-world validation in home environments-where sleep conditions and comfort differ from controlled labs-will be crucial for broader clinical translation.

Materials and Methods

Optimized Firmware for Power Efficient Operation and Data Compression. The device fabrication and encapsulation processes build on prior

techniques (14, 21); however, this work introduces significant advancements in mechanical structural design for enhanced sensitivity, durability, and firmware development for low-power, energy-efficient computation with details shown in *SI Appendix*, Table S10. Several features of the firmware enabled optimized power and onboard memory consumption with accurate timestamps. First, Kalman filtered UTC time, well aligned with stable SoC ticks, served as estimates for timestamps for each sample buffer, thereby enabling dynamic interpolation for individual sample timestamps and allowing for accurate estimates of output data rate. Second, automatic buffering and switching between writing to iOS devices and onboard memory was based on BLE QoS fluctuations. The data compression algorithm and BLE QoS-adaptive automatic data storage switching between onboard memory and cloud storage optimize battery life while ensuring remote data accessibility with minimal latency. Data were uploaded to the iOS device in batches every few minutes, with BLE data plotted within one frame on the app side ($<$16 ms). Third, signal edge computing exploited the nRF52840 SoC (in ISP1807) to handle various math operations efficiently, including filter designs like Butterworth, Kalman, Savitzky–Golay, Cascaded Biquad, and Cascaded Integrator-Comb filters, as well as fundamental linear algebra operations, FFT, and more. Fourth, data compression used a Butterworth lowpass filter and a delta-oriented integral time-series compression algorithm. A two-phase compression process enabled a $\approx$1 to 6$\times$ compression ratio. Raw data size without compression for one IMU at 1.6 kHz was 1.88 GB per day. With data compression, a 50 mAh LiPo battery enabled sustained operation for up to 10.4 d in power saving mode, 4.2 d in streaming mode, and 1.3 d for a 2GiB NAND. See details in *SI Appendix, Materials and Methods*.

Data Collection Protocols for Human Subject Studies. The research was approved by Carle Foundation Hospital's Institutional Review Board, IL, USA (IRB#18CNI1842), adhering to the Declaration of Helsinki, 1964. For data collection, informed consent was obtained from all participants and medical silicone adhesive (3M, 2477P) attached the LMA sensor to the neck of each patient (aligning IMU on the SN). To assess sensor placement, one male and one female subject wore two additional LMA sensors, one above the left breast and another on the lower ribcage, near the infrarenal aortas, for five nights of data collection. *SI Appendix*, Fig. S1 *B* and *D* suggest SN provides the best signal-to-noise ratio (SNR). The primary IMU, positioned on the isolated island aligned with the SN, achieves a higher SNR than the secondary IMU on the main sensor body, confirming the improved sensitivity of current design. Medical-grade transparent film (Tegaderm, 3M) was used to reinforce the sensor attachment. Clinical staff helped with sensor placement. Prior to each cycle of use, devices were sterilized with two cycles of wiping with 70% isopropyl alcohol and drying in air.

Extraction of Cardiopulmonary and Activity Features Guided by Physiological Considerations. REM sleep consists of phasic (sympathetically driven with rapid eye movements) and tonic (parasympathetically driven with atonia and no eye movements) phases (51, 52). Physiological patterns vary between NREM and REM stages, affecting cardiac and respiratory rhythms. Baroreflex sensitivity increases during NREM sleep but reduces in REM, to stabilize heart rate in REM (19). Skeletal muscle tone diminishes progressively from wakefulness to REM sleep, culminating in muscle paralysis (atonia) which inhibits skeletal motoneurons (53, 54), leading to reduced tidal volumes and irregular breathing (30–32). While the parasympathetic nervous system promotes stable cardiorespiratory patterns during NREM, the phasic and tonic stages during REM make the regulation of the sympathetic nervous system less distinct. The HRV, RRV, and ACC features reflect these findings, as shown in *SI Appendix*, Tables S2–S6.

Data Processing and Development of LMASN. *SI Appendix*, Fig. S14 describes the transformation of LMA raw data into inputs for the ML model. Each night-long record was first down-sampled to 200 Hz, and then standardized using a 60-s moving average with a 50% overlap. Points more than three SDs from the mean within this window were considered outliers and replaced using linear interpolation. The processed data were then segmented based on sleep stages annotated at 30-s intervals, resulting in segments that span 30 seconds before and after each label.

LMASN, abstracted in Fig. 1*D* and detailed in *SI Appendix*, Fig. S10, starts with a 1D-ResNet feature extractor (55) (*SI Appendix*, Fig. S10*A*). ResNet, a deep convolutional neural network, captures rich information that could be beneficial for SSC from raw triaxis LMA data. The autoencoder array further enriches the embedding, latent space, by explicitly integrating DSP output signals (DOS). As shown in *SI Appendix*, Fig. S10*B*, each autoencoder in the array first reduces the dimensionality of Z_{MA} to match the data stream at point *O*. The embedding at point *P* then combines with the data stream at *O* using a weighted sum and learnable weights, optimizing the significance of each DOF. Subsequently, a decoder restores the vectors to their original size such that the gradient, which involves these weights, can facilitate analysis of feature importance. The average of these outputs across all autoencoders forms a comprehensive latent vector. Bi-LSTM (56) module in *SI Appendix*, Fig. S10*C* processes the chronological progression of sleep stages. Finally, the classifier discerns subtle transitions between stages, and detects patterns that may indicate the onset of a particular stage in *SI Appendix*, Fig. S10*D*. Structural design of LMASN not only effectively extracts information from raw signals and DOS but allows determination of how each feature contributes to sleep stages.

OSA and Snoring Detection with YOLO. Unlike traditional CNN, the lightweight YOLO model enables end-to-end training and real-time object detection by reducing the number of steps, while still maintaining high average precision. For OSA detection, 113 apneic episodes from six patients, annotated by technicians with bounding boxes, serve as ground truth. Each case corresponds to 120-s intervals of LMA data. Of this complete dataset, 73 images were trained with YOLOv8. Data from each axis were normalized to center the data at zero and to fit into 640 $\times$ 640-pixel canvases. Example images were displayed to show raw LMA signals along with annotated ground-truth labeling of snoring (*SI Appendix*, Fig. S6) and OSA (*SI Appendix*, Fig. S15) events and predicted results from our dataset. The events were labeled manually using LabelMe, and the dataset was split into training, testing, and validation sets in a 73:8:32 ratio to ensure an unbiased model and to avoid overfitting. The IOU threshold is at 50% to evaluate the mean average precision of apnea events. A mean average precision (mAP) of 0.953 was achieved. YOLOv8 was also used for snoring detection trained with 120 640 $\times$ 640-pixel images, validated on 13 images, and tested on 22 images, ultimately achieving an mAP of 0.938.

Data, Materials, and Software Availability. Firmware was executed in embedded C. Signal processing and data analysis were conducted using Python 3.0 with CUDA, SciPy, PyTorch, Polars, Pandas, and scikit-learn packages. The codes are available on GitHub at: https://github.com/duyayun/MA_sleep (57). All relevant data are included in the article and *SI Appendix*. Additional supporting data are available from the corresponding authors upon request, subject to review for intellectual property or confidentiality obligations.

ACKNOWLEDGMENTS. The work was supported by the Querrey Simpson Institute for Bioelectronics at Northwestern University and the Procter & Gamble Company. We sincerely thank Dr. Steve Xu, Chief Executive Officer of Sibel Health for his valuable assistance in reviewing and providing feedback on our manuscript. Several coauthors of this submission were employees of industry while the work was conducted, but none has direct commercial interest in the work presented here.

Author affiliations: [a]Querrey Simpson Institute for Bioelectronics, Northwestern University, Evanston, IL 60208; [b]Department of Biomedical Engineering, Northwestern University, Evanston, IL 60208; [c]Medical Scientist Training Program, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611; [d]Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208; [e]Procter & Gamble Company, Cincinnati, OH 45224; [f]Carle Neuroscience Institute, Carle Health, Urbana, IL 61801; [g]Carle-Illinois College of Medicine, Urbana, IL 61801; [h]Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208; [i]Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208; and [j]Department of Neurological Surgery, Northwestern University, Evanston, IL 60611

1. D. M. Lloyd-Jones *et al*., Life's essential 8: Updating and enhancing the American heart association's construct of cardiovascular health: A presidential advisory from the American heart association. *Circulation* **146**, e18–e43 (2022).
2. S. Kwon *et al*., At-home wireless sleep monitoring patches for the clinical assessment of sleep quality and sleep apnea. *Sci. Adv.* **9**, eadg9671 (2023).
3. F. Sullivan, Hidden health crisis costing America billions (2025). https://aasm.org/resources/pdf/sleep-apnea-economic-crisis.pdf. Accessed 11 January 2025.
4. R. B. Berry *et al*., AASM scoring manual updates for 2017 (version 2.4). *J Clin Sleep Med.* **13**, 665–666 (2017).
5. A. Mayeli, S. A. Janssen, K. Sharma, F. Ferrarelli, Examining first night effect on sleep parameters with hd-EEG in healthy individuals. *Brain Sci.* **12**, 233 (2022).
6. E. D. Chinoy *et al*., Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep* **44**, zsaa291 (2021).
7. D. J. Miller, C. Sargent, G. D. Roach, A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. *Sensors* **22**, 6317 (2022).
8. R. de Goederen *et al*., Radar-based sleep stage classification in children undergoing polysomnography: A pilot-study. *Sleep Med.* **82**, 1–8 (2021).
9. N. Sridhar *et al*., Deep learning for automated sleep staging using instantaneous heart rate. *NPJ Digit. Med.* **3**, 106 (2020).
10. P. Fonseca *et al*., Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. *Sleep* **40**, zsx097 (2017).
11. Z. Beattie *et al*., Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. *Phys. Physiol.* **38**, 1968 (2017).
12. N. I. Chee *et al*., Multi-night validation of a sleep tracking ring in adolescents compared with a research actigraph and polysomnography. *Nat. Sci. Sleep* **13**, 177–190 (2021).
13. A. J. Boe *et al*., Automating sleep stage classification using wireless, wearable sensors. *NPJ Digit. Med.* **2**, 131 (2019).
14. K. Lee *et al*., Mechano-acoustic sensing of physiological processes and body motions via a soft wireless device placed at the suprasternal notch. *Nat. Biomed. Eng.* **4**, 148–158 (2020).
15. K. Chung *et al*., Noncontact sleep study by multi-modal sensor fusion. *Sensors (Basel)* **17**, 1685 (2017).
16. F. Han *et al*., Earsleep: In-ear acoustic-based physical and physiological activity recognition for sleep stage detection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **8**, 1–31 (2024).
17. T. Lee *et al*., Accuracy of 11 wearable, nearable, and airable consumer sleep trackers: Prospective multicenter validation study. *JMIR Mhealth Uhealth* **11**, e50983 (2023).
18. A Inc., Estimating breathing disturbances and sleep apnea risk from apple watch (2024). https://www.apple.com/health/pdf/sleep-apnea/Sleep_Apnea_Notifications_on_Apple_Watch_September_2024.pdf. Accessed 11 January 2025.
19. F. Shaffer, J. P. Ginsberg, An overview of heart rate variability metrics and norms. *Front. Public Health* **5**, 258 (2017).
20. U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, J. S. Suri, Heart rate variability: A review. *Med. Biol. Eng. Comput.* **44**, 1031–1051 (2006).
21. X. Ni *et al*., Automated, multiparametric monitoring of respiratory biomarkers and vital signs in clinical and home settings for Covid-19 patients. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026610118 (2021).
22. K. R. Jinkins *et al*., Thermally switchable, crystallizable oil and silicone composite adhesives for skin-interfaced wearable devices. *Sci. Adv.* **8**, eabo0537 (2022).
23. H. Jeong *et al*., Differential cardiopulmonary monitoring system for artifact-canceled physiological tracking of athletes, workers, and Covid-19 patients. *Sci. Adv.* **7**, eabg3092 (2021).
24. E. E. Benarroch, Control of the cardiovascular and respiratory systems during sleep. *Auton. Neurosci.* **218**, 54–63 (2019).
25. F. Shaffer, R. McCraty, C. L. Zerr, A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. *Front. Psychol.* **5**, 1040 (2014).
26. J. W. Kantelhardt *et al*., Multifractal detrended fluctuation analysis of nonstationary time series. *Phys. A Stat. Mech. Appl.* **316**, 87–114 (2002).
27. E. A. Ihlen, Introduction to multifractal detrended fluctuation analysis in matlab. *Front. Physiol.* **3**, 141 (2012).
28. A. Orozco-Duque, D. Novak, V. Kremen, J. Bustamante, Multifractal analysis for grading complex fractionated electrograms in atrial fibrillation. *Phys. Physiol.* **36**, 2269 (2015).
29. A. Faini, G. Parati, P. Castiglioni, Multiscale assessment of the degree of multifractality for physiological time series. *Philos. Trans. R. Soc. A* **379**, 20200254 (2021).
30. N. J. Douglas, D. P. White, C. K. Pickett, J. V. Weil, C. W. Zwillich, Respiration during sleep in normal man. *Thorax* **37**, 840–844 (1982).
31. X. Long, J. Foussier, P. Fonseca, R. Haakma, R. M. Aarts, Analyzing respiratory effort amplitude for automated sleep stage classification. *Biomed. Signal Process. Control* **14**, 197–205 (2014).
32. M. G. Miglis, Autonomic dysfunction in primary sleep disorders. *Sleep Med.* **19**, 40–49 (2016).
33. R. A. Dampney, Resetting of the baroreflex control of sympathetic vasomotor activity during natural behaviors: Description and conceptual model of central mechanisms. *Front. Neurosci.* **11**, 461 (2017).
34. V. N. Varghees, K. Ramachandran, A novel heart sound activity detection framework for automated heart sound analysis. *Biomed. Signal Process. Control* **13**, 174–188 (2014).
35. N. Storck, M. Ericson, L. Lindblad, M. Jensen-Urstad, Automatic computerized analysis of heart rate variability with digital filtering of ectopic beats. *Clin. Physiol.* **21**, 15–24 (2001).
36. M. Malik, Electrophysiology, Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Circulation* **93**, 1043–1065 (1996).
37. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Las Vegas, NV, 2016), pp. 779–788.
38. M. Brennan, M. Palaniswami, P. Kamen, Poincaré plot interpretation using a physiological model of HRV based on a network of oscillators. *Am. J. Physiol. Heart Circ. Physiol.* **283**, H1873–H1886 (2002).
39. P. W. Kamen, H. Krum, A. M. Tonkin, Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity in humans. *Clin. Sci.* **91**, 201–208 (1996).
40. J. Trinder *et al*., Autonomic activity during human sleep as a function of time and sleep stage. *J. Sleep Res.* **10**, 253–264 (2001).
41. P. Boudreau, W. H. Yeh, G. A. Dumont, D. B. Boivin, Circadian variation of heart rate variability across sleep stages. *Sleep* **36**, 1919–1928 (2013).
42. R. L. Burr, Interpretation of normalized spectral heart rate variability indices in sleep research: A critical review. *Sleep* **30**, 913–919 (2007).
43. P. G. Burke *et al*., State-dependent control of breathing by the retrotrapezoid nucleus. *J. Physiol.* **593**, 2909–2926 (2015).
44. N. Zavanelli, S. H. Lee, M. Guess, W. H. Yeo, Soft wireless sternal patch to detect systemic vasoconstriction using photoplethysmography. *Iscience* **26**, 106184 (2023).
45. J. R. Shambroom, S. E. Fábregas, J. Johnstone, Validation of an automated wireless system to monitor sleep in healthy adults. *J. Sleep Res.* **21**, 221–230 (2012).
46. K. Kim, D. Y. Park, Y. J. Song, S. Han, H. J. Kim, Consumer-grade sleep trackers are still not up to par compared to polysomnography. *Sleep Breath.* **26**, 1573–1582 (2022).
47. M. Radha *et al*., Sleep stage classification from heart-rate variability using long short-term memory neural networks. *Sci. Rep.* **9**, 14149 (2019).
48. G. E. Billman, The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance. *Front Physiol.* (2013), 10.3389/fphys.2013.00026/.
49. T. Chen, C. Guestrin, "Xgboost: A scalable tree boosting system" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, B. Krishnapuram *et al.*, Eds. (Association for Computing Machinery, 2016), pp. 785–794.
50. L. S. Shapley, "Notes on the n-person game - II: The value of an n-person game" in *International Joint Conferences on Artificial Intelligence Organization*, L. De Raedt, Ed. (IJCAI Organization, 1953).
51. M. A. Carskadon *et al*., Normal human sleep: An overview. *Princ. Pract. Sleep Med.* **4**, 13–23 (2005).
52. P. Simor, G. van der Wijk, L. Nobili, P. Peigneux, The microstructure of rem sleep: Why phasic and tonic? *Sleep Med. Rev.* **52**, 101305 (2020).
53. J. J. Fraigne, Z. A. Torontali, M. B. Snow, J. H. Peever, Rem sleep at its core-circuits, neurotransmitters, and pathophysiology. *Front. Neurol.* **6**, 123 (2015).
54. J. M. Siegel, Functional implications of sleep development. *PLoS Biol.* **3**, e178 (2005).
55. K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition" in *Proceedings of the IEEE conference on computer vision and pattern recognition* (IEEE, 2016), pp. 770–778.
56. M. F. Aslan, M. F. Unlersen, K. Sabanci, A. Durdu, CNN-based transfer learning-BiLSTM network: A novel approach for Covid-19 infection detection. *Applied Soft Computing* **98**, 106912 (2021).
57. Y. Du *et al.*, LMA_sleep: Multinight annotated PSG dataset and raw LMA signals for wearable-based sleep stage and disorder detection. GitHub. https://github.com/duyayun/LMA_sleep. Accessed 18 May 2025.