Detection and Prediction of Sleep Disorders by Covert Bed-Integrated RF Sensors

Zijing Zhang*, Thomas B. Conroy, Ana C. Krieger and Edwin C. Kan, *Senior Member, IEEE*

***Abstract*— *Objective:* Respiratory disturbances during sleep are a prevalent health condition that affects a large adult population. The gold standard to evaluate sleep disorders including apnea is overnight polysomnography, which requires a trained technician for live monitoring and post-processing scoring. Currently, the disorder events can hardly be predicted using the respiratory waveforms preceding the events. The objective of this paper is to develop an autonomous system to detect and predict respiratory events reliably based on real-time covert sensing. *Methods*: A bed-integrated radio-frequency (RF) sensor by near-field coherent sensing (NCS) was employed to retrieve continuous respiratory waveforms without user's awareness. Overnight recordings were collected from 27 patients in the Weill Cornell Center for Sleep Medicine. We extracted respiratory features to feed into the random-forest machine learning model for disorder detection and prediction. The technician annotation, derived from observation by polysomnography, was used as the ground truth during the supervised learning. *Results*: Apneic event detection achieved a sensitivity and specificity up to 88.6% and 89.0% for *k*-fold validation, and 83.1% and 91.6% for subject-independent validation. Prediction of forthcoming apneic events could be made up to 90 s in advance. Apneic event prediction achieved a sensitivity and specificity up to 81.3% and 82.1% for *k*-fold validation, and 80.5% and 82.4% for subject-independent validation. The most important features for event detection and prediction can be assessed in the learning model. *Conclusion*: A bed-integrated RF sensor can covertly and reliably detect and predict apneic events. *Significance*: Predictive warning of the sleep disorders in advance can intervene serious apnea, especially for infants, servicemen, and patients with chronic conditions.**

***Index Terms*— Clinical diagnosis; respiration sensors; sleep apnea.**

I. Introduction

SLEEP disorders are a major public health problem, and 50 to 70 million Americans chronically suffer from the consequences from sleep disorders [1]. Obstructive sleep apnea (OSA) is the most common sleep-related breathing disorder [2][3], with a prevalence in the adult population ranging from 6% to 17%, explicated by the apnea-hypopnea index (AHI) greater than 15 events per hour. OSA can be as high as 49% in geriatrics [4], and can still be under-diagnosed due to the inconvenience of the present monitoring setup [5]. OSA is characterized by repeated episodes of partial or complete obstruction of the respiratory passages during sleep, and may result in sleep fragmentation and non-restorative sleep. The consequences of OSA include excessive day-time sleepiness, insomnia, and increased risks of stroke, obesity, pulmonary hypertension and heart attack [6]. Missed identification of sleep disorders can be especially serious for young children with concerns of sudden infant death [7], and servicemen whose circadian rhythm are difficult to maintain but continuous vigilance is frequently required [8].

Currently, the gold standard of sleep disorder diagnosis is an overnight sleep study by polysomnography (PSG) [9], which records the breath airflow, respiratory torso movement, oxygen saturation (SpO2), body motion, electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG) by a plethora of sensors. Sleep disorders are most often scored by trained sleep technicians considering various PSG waveforms. Despite of high accuracy by PSG, disadvantages include the high cost and scarce availability in clinics, and the uncomfortable experience for users. Several methods have been put forward as PSG alternatives for sleep disordered breathing (SDB) detection, such as the at-home systems using portable devices [10] which often has inadequate accuracy [11].

Electrocardiogram (ECG) is one of the most extensively scrutinized signals for sleep study [12][13]. Together with the chest belts for respiratory patterns, SDB can be reasonably detected. Apnea alarm systems also usually employ SpO2 from the pulse oximeter to provide warning when SpO2 falls below a predefined threshold [14]. However, the high rate of false alarms can be triggered by motion artifacts and poor sensor contact [15].

Many methods and algorithms have been proposed to detect sleep disorders autonomously [9][16]-[18], including various machine-learning (ML) models of neural networks, regression, and ensemble learning. Nevertheless, in the past decades, fewer studies have explored the prediction capability [19][20]. Predictive warning of the SDB events in advance can potentially improve the effectiveness of therapy. Currently, the gold-standard treatment for OSA is the continuous positive airway pressure (CPAP) [21][22] by blowing air into the nose. Though effective, the use of a single pressure and cumbersome

This work is supported by Department of Defense of the United States through Office of the Congressionally Directed Medical Research Programs (CDMRP) Discovery Award PR-182496, and by National Institute of Health (NIH) R21 DA049566-01A1.

Z. Zhang, T. B. Conroy, and E. C. Kan are with School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA. A. C. Krieger is with Center for Sleep Medicine at Weill Cornell Medicine, New York, NY 10065, USA. E-mails: {zz58, tbc38, eck5}@cornell.edu; ack2003@med.cornell.edu).

equipment could cause pressure intolerance and reduce long-term acceptance. To improve user comfort, auto-titrating continuous positive airway pressure (APAP) [23] was later developed [24]. Recently the COVID-19 pandemic has cast a spotlight on the ventilators [25] as life-support machines providing intensive ventilatory support. Other innovative methods have also been proposed to intervene sleep disorders after detecting abnormal events [26]. However, interventions can be best applied with apneic prediction, as detection may be too late for intervention after 30s. Predictive warning of SDB for an advance of 30 – 90 s might be critical to improve therapeutic outcomes and reduce the impact on oxygen levels and sleep structure.

Motivated by the unmet needs of reliable prediction of sleep disorders with minimal user disturbance, we developed a bed-integrated system with predicative SDB warning up to 90 s. Our system is based on the near-field coherent sensing (NCS) of ultra-high frequency (UHF) electromagnetic (EM) waves to monitor the dielectric boundary movement of internal organs and body parts [26]-[28]. It can be invisible to users and requires no personal setup time, especially considering occasional leave from beds such as restroom visits. Comfort and convenience are of critical importance for overnight sleep monitoring because apnea is a sparse event in the long recording of sleep in different stages. If users decide to take the monitor off and do not take the trouble of re-installation, serious apneic incidences can be missed.

We have further developed a learning-based algorithm for detection and prediction. We recruited 27 patients with sleep disorders from the Weill Cornell Center for Sleep Medicine for overnight recording. The demographic properties and final AHI scores were shown in Supplementary Table I. The data from the NCS bed sensor was processed for feature extraction to feed into the random-forest ML model, which gave accurate SDB detection. Apneic events can be further predicted up to 90 s in advance based on the present respiratory features. We could also determine the correlation between respiratory features and SDB to identify the most critical physiological factors for detection and prediction of SDB episodes.

II. Experimental Setup and Protocol

A. Experimental setup

As shown in Fig. 1(a), the bed-integrated sensor consisted of a notched miniature coaxial radio-frequency (RF) cable, where the metal shield of the middle 5 inches was removed to allow a small amount of EM energy leaking into the nearby user body [28]. The notch length was designed to accommodate positional variations for different patients. In the near-filed region, the dielectric boundary movement by lungs and associated muscles would couple to the leaked EM energy, and hence affect the signals between the transmitter (Tx) and the receiver (Rx) [29]. Two notched-cable sensors underneath the approximate thorax and abdomen positions were adopted to capture the motion in separate regions during breathing [27]. Software-defined radio (SDR) was used to connect the notched sensors and then interface with the host computer through USB (Universal Serial Bus). As shown in Fig. 1(c), SDR was implemented by the National Instrument Ettus B210 with two Tx/Rx ports. The notched cables were sewn on the bottom side of the mattress pad, as shown in Fig. 1(d), which were placed under an incontinence protector and a fitted bedsheet for comfort and aesthetic. No apparent performance degradation was observed for the additional layers, as evident from our prior studies for sensing over several layers of fabrics [30]. The schematic of the experimental setup was summarized in Fig. 1(e). In our previous work, vital-sign monitoring by this setup has been benchmarked against the strain-based chest belts and ECG with various sleep postures and large position variation [28].

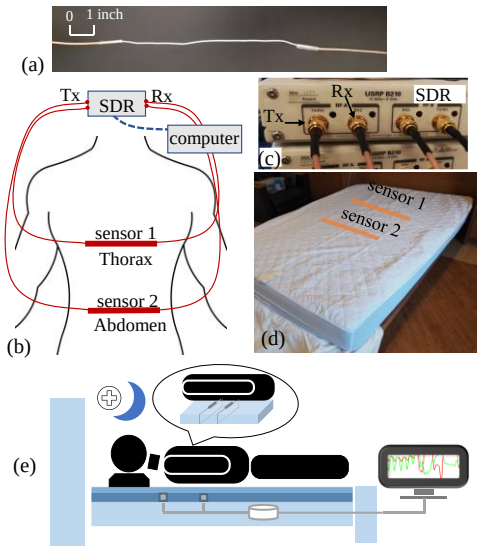

Fig. 1. The RF sensor setup: (a) A photo of the notched transmission line sensor; (b) NCS sensor positions in relation to the body; (c) The photo of the SDR transceiver connected to two NCS sensors; (d) The position of the notched sensors under the mattress pad; (e) Schematics of the overall experimental setup (PSG not shown).

For the SDR Tx signal chain, the digital baseband went through the digital-to-analog converter (DAC) and was then mixed with the carrier frequency f_{RF}. The RF power is less than −10 dBm or 0.1 mW, well under the safety limits set by occupational safety and health administration (OSHA) in the UHF band. The RF signal leaked from the notched structure is coupled into internal organ motion, received by Rx, and then demodulated and sampled by the analog-to-digital converter (ADC) to retrieve the baseband. We employed the quadrature scheme as the baseband tone f_{BB}, and the NCS signal can be represented by the magnitude and phase modulation on the quadrature signal as

$$NCS_{amp}(t) = \sqrt{I_{Rx}(t)^2 + Q_{Rx}(t)^2} \quad (1)$$

$$NCS_{ph}(t) = unwrap(\tan^{-1}\frac{Q_{Rx}(t)}{I_{Rx}(t)} - 2\pi f_{BB} - \theta_0) \quad (2)$$

$$I_{Rx}(t) = A(t)\cos(2\pi f_{BB}t + \theta_0) \quad (3)$$

$$Q_{Rx}(t) = A(t)\sin(2\pi f_{BB}t + \theta_0) \quad (4)$$

where θ_0 is the phase offset accumulated from the Tx-Rx signal chains. f_{RF} was selected at 1 GHz, and any choice between 0.9 and 2.4 GHz should have similar performance [27]. Two RF notched sensors, one approximately at the thorax and the other at the abdomen, were operated at two distinctive basebands of

f_{BB1} =355 kHz and f_{BB2} = 440 kHz, which were implemented by two Tx-Rx channels synchronized in one SDR to reduce cross interference. Both channels were sampled at 10^6 samples per second (Sps), and were further down-sampled to 500 Sps after demodulation.

B. Subjects and Data Collection

Overnight PSG was performed at the Weill Cornell Center for Sleep Medicine at approximately the participants' regular sleep time, and included recording of EEG, EOG, submental and anterior tibialis EMG, two-lead ECG, chest and abdominal movement by inductive plethysmography, body position, SpO2 pulse oximetry, and nasal pressure respiratory flow monitoring. Scoring of SDB and sleep stages was performed by a registered PSG technician licensed in the State of New York. All events were scored according to the recommended rules by the American Academy of Sleep Medicine [31]. The human study was performed under the approved protocol of Weill Cornell Medical Center IRB# 19-12021223.

III. Data Processing

After gathering overnight recording of patients, we first processed our NCS data along with PSG respiratory data, then extracted the respiratory features, and finally fed the data into the ML classifier for SDB detection and prediction. Notice that NCS and PSG respiratory channels had the same signal processing procedure for fair comparison. We used MATLAB for signal processing and feature extraction.

A. Signal processing

We included the full overnight NCS recording of 27 patients with the duration of 7 – 8 hours. PSG was mainly used for feature validation and model comparison, and had 3 respiratory channels of the airflow, thorax belt, and abdomen belt. SpO2 from PSG was also collected as an additional input apart from the respiratory motion waveforms. Two NCS sensors from the thorax and abdomen positions produced respective magnitude and phase as four individual inputs.

The feature extraction contained 8 steps:

1. Down sample NCS and PSG to 25 Hz.
2. Synchronize NCS and PSG signals (precision to 1s).
3. Perform bandpass filtering and smoothing (0.05 Hz to 2 Hz).
4. Segment waveform into epochs of T_{epoch} = 40 s and a sliding window of T_{slide} = 15 s.
5. Label operator annotation in epoch.
6. Normalize waveform and extract features in epoch.
7. Select epochs by signal quality for NCS and PSG.
8. Output features and annotation to the ML model.

The bandpass filter in Step 3 was implemented in MATLAB by the digital infinite impulse response (IIR). Another Savitzky-Golay finite impulse response (FIR) smoothing filter [32] with 4^{th} polynomial order was further employed to rid of high-frequency noises. The operator annotations were adapted to give epoch-based references. In each epoch, if any annotation has a time duration > 40% × T_{epoch} = 16 s, the current epoch will be labelled accordingly. The choice of T_{epoch} between 10 – 20 s will not affect the end result significantly. Epoch labels include normal, snore, hypopnea, OSA, mixed apnea, and CSA (central sleep apnea). If the annotated disorder event lasts < 16 s, the epoch will be labeled as normal. If the time duration with no annotation > 20 s, current epoch will be removed from dataset. This is often due to insufficient evidence of proper PSG monitoring, such as patients going to restroom or taking some sensors off.

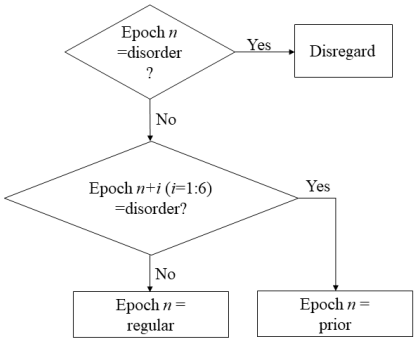

Fig. 2. The prediction labelling criterion.

The annotation was directly used as the ground truth for disorder detection training and validation. For prediction, labelling criterion should be modified. In each epoch, if the current label was a disorder event, we would remove it from the dataset because prediction was based on the normal period before the disorder events. Aiming to predict disorders 0 – 90s in advance, we labelled the epochs containing six sliding windows before the forthcoming disorder event as the disorder precedence "prior", and the other epochs as "regular". A simple flow chart of prediction labelling is shown in Fig. 2.

B. Feature extraction

After epoch segmentation, waveforms were normalized to [–1,1] in each epoch. To extract respiratory features, we first implemented the peak detection algorithm [33]. A moving-average curve was first calculated at each time point in a given period, which was around one respiration cycle and then constantly updated. The points when the moving-average curve crossed the original signal were marked as up-crossing points for positive slopes in the original signal or down-crossing points for negative. Local maximum was labelled as the maximal point between two up-down crossing points, and local minimum as the minimal point between two down-up crossing points. Fig. 3 presents sample epochs of respiratory waveforms, annotated with different labels of (a) normal, (b) OSA, and (c) hypopnea. The red waveform was derived from NCS and the green ones from PSG. Solid magenta upward-pointing triangles marked the maximum peaks detected by the algorithm and blue downward-pointing triangles mark the minimum peaks. We could observe distinctive patterns in different events. Waveforms showed a more regular respiratory pattern in normal epochs, while irregular patterns more frequently indicated disorder epochs.

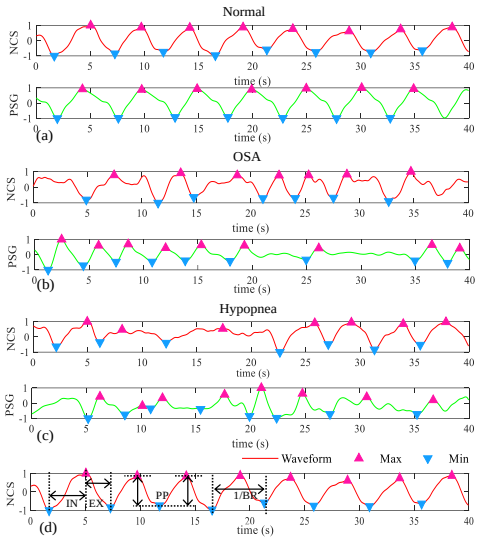

Fig. 3. Waveform examples from NCS and PSG in the epochs labelled as (a) normal; (b) OSA; (c) hypopnea. NCS channels are from (a): thorax phase; (b): thorax phase; (c): abdomen amplitude. PSG channels are from (a): chest; (b): airflow; (c): airflow. (d) Examples of the peak detection and feature extraction process. The red line is the respiratory waveform from (a). The magenta and blue triangles are the detected maximum and minimum peaks. Features of IN, EX, PP and BR are marked.

TABLE I. RESPIRATORY FEATURES I (16).

μ_{BR}	μ_{PP}	μ_{IN}	μ_{EX}
σ_{BR}	σ_{PP}	σ_{IN}	σ_{EX}
CoV_{BR}	CoV_{PP}	*skew*	*kurt*
n	*entr*	η_{p0}	T_{hold}

TABLE II. RESPIRATORY FEATURES II (21).

η_{p1}	η_{p2}	η_{p3}	η_{p4}
$f1_{max}$	$f2_{max}$	$f3_{max}$	$f4_{max}$
Cor_{BR}	Cor_{PP}	Cor_{IN}	Cor_{EX}
SD_{BR}	SD_{PP}	SD_{IN}	SD_{EX}
max_{IN}	max_{EX}	max_{BR}	
min_{BR}	min_{PP}		

TABLE III. SPO2 FEATURES (4).

μ_{SpO2}	σ_{SpO2}	η_{SpO2}	min_{SpO2}

After identifying respiratory cycles using peak detection, we could first extract the 4 respiratory parameters in each breath cycle to represent the instantaneous respiratory characteristics, including *BR* (breath rate in BPM), *PP* (peak-to-peak in arbitrary units as an estimate of the lung volume), *IN* (inhalation interval in s), and *EX* (exhalation interval in s). Several examples are shown in Fig. 3(d) for the extraction process of respiratory parameters.

After gathering respiratory cycles and parameters, we extracted 16 respiratory features which would function as the epoch features fed into the ML classifier for detection as listed in Table I. The first 8 features were the mean (μ) and the standard deviation (σ) of the above 4 respiratory parameters. Because BR and PP were two significant factors representing the respiratory pattern, we added 2 more features as the coefficient of variation (*CoV*) of BR and PP,

$$CoV = \left(\frac{\sigma}{\mu}\right)^2 \quad (4)$$

CoV showed the extent of variability in relation to the mean. Additionally, *Skew* and *kurt* measured the tailedness and asymmetry of each respiratory cycle, and were averaged over all cycles within the epoch. Apart from features derived from respiratory parameters, we appended four supplemental features including 1) the total number of detected respiratory cycles *n*; 2) the total randomness or entropy of the waveform *entr*; 3) the power in the lower frequency band ((0.05,0.5) Hz) divided by the total power in all frequencies η_{p0}; 4) the time duration when no peak was detected within the epoch T_{hold}.

Other than 16 respiratory features in Table I, we added 21 respiratory features in Table II for the prediction classifier, which had 37 respiratory features in total. Augmentation of features can enhance the performance of the ML model before overfitting becomes dominant. η_{pi} and fi_{max} (i = 1~4) represented the power in specific bandwidth divided by the total power in all frequencies and the frequency with the maximum power density within the bandwidth, respectively. The four chosen bandwidths were f_1 = (1, 2) Hz; f_2 = (2, 5) Hz; f_3 = (5, 8) Hz; f_4 = (8, 12.5) Hz. *Cor* was the autocorrelation in a time lag of one respiratory cycle to measure the successive similarity of a given respiratory parameter. *SD* representing the successive difference was defined as the mean absolute difference between adjacent cycles. At last, we added the maximum of IN, EX and BR within the epoch and the minimum of BR and PP. The choice of these features is based on the physiological reasoning that in the events of disorder, or in the anticipation of the events, there would be larger variation in PP and BR within the epoch. η_{pi} and fi_{maxc} contained the regularity of IN and EX, and possibly some tissue vibration characteristics during the disorder events below the audible range.

Beyond respiratory features derived from the NCS and PSG waveforms, we also added features representing oxygen saturation as listed in Table III: 1) μ_{SpO2}: the mean SpO2 level; 2) σ_{SpO2}: the standard deviation of SpO2 level; 3) η_{SpO2}: the percentage of time when SpO2 < threshold (92%); 4) min_{SpO2}: the minimum level of SpO2.

After segmentation and feature extraction, we added an extra step for NCS epoch selection. Signal quality cannot be guaranteed during the entire course of overnight recording because patients may have random motion lying on the bed or leave the bed for restroom visits. Various other factors such as ambient movement might bring about noises to cause SNR (signal-to-noise ratio) degradation. Therefore, we opted to remove the epochs with very low SNR by pre-determined thresholds, i.e., epochs with $\eta_{p0} < Th_{\eta p0}$ = 70 % and $\sigma_{PP} > Th_{\sigma PP}$ = 0.3 will be removed from the dataset.

IV. MACHINE LEARNING MODELS

A. Data composition

For output datasets, Table IV shows NCS and PSG dataset composition for detection and prediction, respectively. Labels for epochs were divided into 7 classes, namely, normal, snore, arousal, hypopnea, OSA, mixed apnea, and CSA. Labels of normal, snore, and arousal were further grouped into the binary classification of "normal", while labels of hypopnea, OSA, mixed apnea, and CSA into "disorder". We studied the

performances of our ML model by both 7 classes and 2 classes, although the main focus was on the binary classes of normal and disorder. The first 7 rows in Table IV presented the total number of epochs annotated with these labels in NCS and PSG. For the last 2 rows, the disorder ratio was the proportion of disorder epochs within all epochs, and the epoch selection ratio was the ratio between the total duration of selected epochs and overall recording time. Because prediction only included normal epochs for disorder precedence, NCS prediction dataset has a relatively smaller ratio than detection. As for the PSG dataset, we used the epochs derived from the same time periods as selected from NCS for fair comparison. Note that though PSG and NCS shared the same recording time, PSG only utilized one optimal channel out of three respiratory channels for each epoch, while NCS may include more than one channel with acceptable signal quality within each epoch.

TABLE IV. NCS DATASET COMPOSITION OF EVENTS AND PRECEDENCIES.

	NCS Detection	PSG Detection	NCS Prediction	PSG Prediction
Normal	23574	15334	13350	9538
Snore	621	244	265	126
Arousal	548	274	578	334
Hypopnea	7674	3483	3221	1689
OSA	1902	852	1452	560
Mixed Apnea	63	34	7	4
CSA	401	188	405	157
Disorder ratio	0.289	0.223	0.307	0.231
Epoch selection ratio	0.413	0.413	0.254	0.254

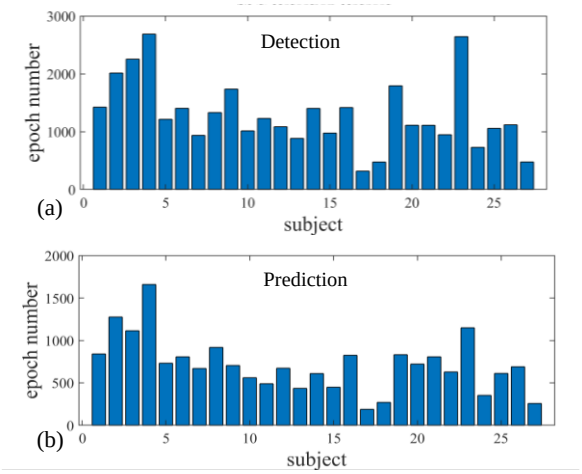

Fig. 4. The numbers of the selected NCS and PSG epochs from each patient in the dataset of (a) detection and (b) prediction.

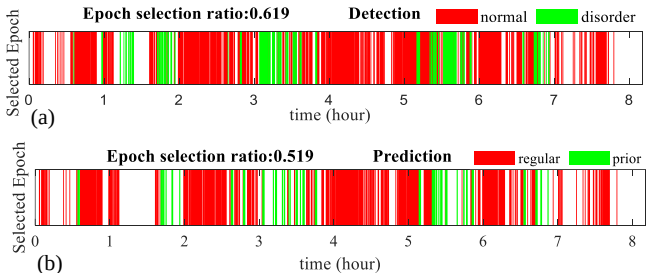

Fig. 5. An example of the selected NCS epoch distribution within the whole overnight recording in the (a) detection and (b) prediction datasets.

TABLE V. COMPARISON OF THE MEAN AND STANDARD DEVIATION OF NORMAL AND DISORDER EPOCHS OF SELECTED FEATURES.

Avg± Dev	μ_{BR}	σ_{BR}	μ_{PP}	σ_{PP}	η_{p0}
NCS: Normal	15.6± 3.12	2.62± 1.65	0.992± 0.206	0.147± 0.050	87.2± 7.44
NCS: Disorder	17.8± 3.64	4.61± 2.91	0.859± 0.220	0.206± 0.079	86.1± 7.37
PSG: Normal	15.6± 3.23	1.97± 1.67	1.171± 0.208	0.138± 0.085	91.7± 8.52
PSG: Disorder	16.7± 3.47	2.73± 1.88	1.032± 0.209	0.213± 0.096	89.6± 8.98
Avg± Dev	CoV_{BR}	CoV_{PP}	μ_{SpO2}	σ_{SpO2}	
NCS: Normal	0.164± 0.094	0.159± 0.074	93.5± 5.52	0.620± 1.85	
NCS: Disorder	0.252± 0.141	0.266± 0.145	91.6± 3.59	1.70± 1.34	
PSG: Normal	0.122± 0.089	0.135± 0.115	93.5± 5.67	0.622± 1.82	
PSG: Disorder	0.162± 0.101	0.228± 0.137	91.7± 3.61	1.68± 1.29	

Fig. 4 showed the selected NCS epoch number for each patient in the detection and prediction datasets. The NCS epoch selection had a large variation mainly due to subject variation. Similar results for PSG dataset were presented in supplementary Fig. 1. An example of the NCS epoch selection during the whole overnight recording was shown in Fig. 5 for detection and prediction datasets from one representative patient. For detection in Fig. 5(a), the red bars represented the selected normal epochs and the green bars represented the selected disorder epochs. For prediction in Fig. 5(b), the normal epochs in the detection dataset are further divided into regular and prior according to whether an abnormal event will happen in the forthcoming 90 s. Note that the PSG dataset had the same epoch time distribution with NCS. Table V presented the comparison of the mean and standard deviation of the selected features in normal and disorder epochs in the NCS and PSG detection datasets which had dominant significance in the ML model in the following section. Disorder epochs had higher standard deviation for all respiratory features, indicating disorder epochs were less stable and tend to fluctuate more. Meanwhile, disorder epochs also have distinctively higher CoV_{BR}, and CoV_{PP} in comparison with normal ones, which were important factors to distinguish the two classes as well. For oximetry, disorder epochs usually had smaller μ_{SpO2} and higher σ_{SpO2}. Similar analyses for prediction dataset were presented in supplementary Table II.

B. The Random-forest Model

We chose the random-forest classifier [34] as the ML model, which was an ensemble learning method for classification that constructed a multitude of decision trees during training, and then output the class selected by most trees. There are two advantages in the tree-based ML models: 1) Straightforward to interpret as a white-box model, which can help us understand the intrinsic relationship between respiratory features and sleep disorders; 2) Non-parametric without assumption on the data distribution or linearity. Random forest can reduce overfitting in a single decision tree but keep the advantages of the decision tree.

However, we faced the problem of class imbalance as shown in Table IV, where the number of normal epochs was much larger than those of the disorder epochs. In other words, a bias or skewness would shift towards the normal event present in the dataset. For remediation, we added class weights for statistical amplification, which assigned different weights to the normal and disorder labels (normal:disorder = 1:3). The model thus penalized the misclassification made on the minority class of disorder. This practice achieved higher sensitivity to disorder detection effectively.

Class weighting can improve sensitivity, while outlier removal can improve specificity. Before constructing the ML model, we first cleaned the dataset by removing the normal epochs that were distinctively deviant from the majority. In our study, we assumed that the normal epochs in the dataset should have a relatively regular and similar respiratory pattern and thus data should form a dense cluster. Abnormal observations that are far from the majority ones within normal epochs were removed as outliers that were most likely due to noisy or wrong data. Here, we used the isolation forest algorithm [35], an unsupervised anomaly detection method based on random forests, as the outlier detection method. The outlier removal ratio was set at 0.2, which efficiently promoted specificity by eliminating noisy data in normal epochs.

C. *Results for disorder detection*

The *k*-fold and leave-one-participant-out cross-validations (CV) were employed as model verification, where *k*-fold CV tested the skill of the model on new data, and leave-one-participant-out CV tested the robustness to unseen patients. For *k*-fold CV of $k = 5$, we divided the whole dataset (N cases) into separate training ($0.8N$ cases) and testing ($0.2N$ cases), and the process was repeated 5 times until all cases had been tested as unseen data. For leave-one-participant-out CV, the model was trained on the data sets from all patients except one, whose data were then used as testing. The CV process was reinitialized and repeated for each patient as the testing case.

Fig. 6 shows the overall confusion matrices for detection using *k*-fold random forest, while Table VI further presents the statistics for *k*-fold. The binary class of normal and disorder achieved better accuracy than the full seven classes, whose confusion matrices are shown in supplementary Fig. 2. NCS + SpO2 resulted in the best performance for disorder detection with 88.9% accuracy, 88.6% sensitivity and 89.0% specificity. The top three important features were SpO2 deviation σ_{SpO2}, peak-to-peak deviation σ_{PP}, and breath rate deviation σ_{BR}. When only NCS datasets were used, the sensitivity has significantly degraded to 63.6%, indicating SpO2 was an important factor for apnea identification apart from respiratory patterns. The top three important features became σ_{PP}, σ_{BR}, and μ_{BR}.

Note that the PSG detection here only utilized one optimal channel out of three respiratory channels for each epoch. In supplementary Fig. 3, we also presented the detection performance using only airflow or respiratory-belt channels from PSG. Overall, the accuracy using different PSG respiratory channels was similar.

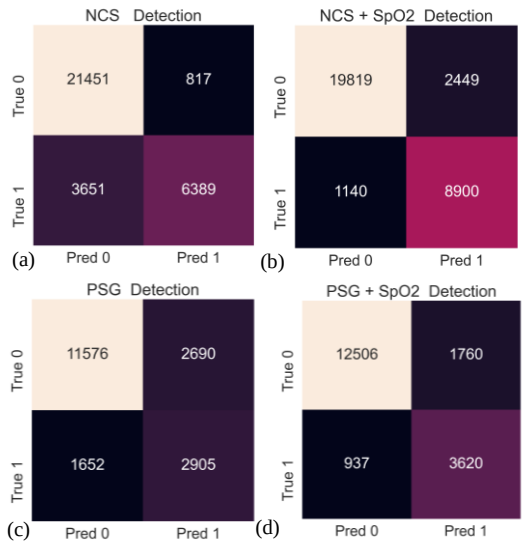

Fig. 6. The confusion matrices showing Normal (0) and Disorders (1) detection by the random forest model using the features from (a) NCS; (b) NCS +SpO2; (c) PSG; (d) PSG +SpO2. The cells list the number of epochs in each category. A 5-fold CV was tested on the entire data.

TABLE VI. COMPARISON OF THE DETECTION DEVICES BY 5-FOLD CV.

Data set	NCS	NCS + SpO2	PSG	PSG + SpO2
Accuracy	86.2%	88.9%	76.9%	85.2%
Sensitivity	63.6%	88.6%	63.7%	78.3%
Specificity	96.3%	89.0%	81.1%	87.4%
Feature importance	σ_{PP} (0.50) σ_{BR} (0.27) μ_{BR} (0.05) CoV_{PP} (0.03)	σ_{SpO2} (0.58) σ_{PP} (0.14) σ_{BR} (0.08) CoV_{PP} (0.04)	CoV_{PP} (0.42) T_{hold} (0.10) μ_{PP} (0.05) μ_{EX} (0.05)	σ_{SpO2} (0.36) η_{SpO2} (0.05) μ_{BR} (0.05) σ_{PP} (0.05)

We further compared multiple classifiers including k-nearest neighbor (kNN), support vector machine (SVM), decision tree, hybrid model and random forest, as presented in Table VII. The hybrid model consisted of the voting classifier ensembled from SVM, kNN and decision tree altogether. kNN had the lowest sensitivity to disorder detection, although the specificity was very high. Random forest resulted in highest accuracy for NCS + SpO2 dataset, and also achieved high sensitivity and specificity. The overall difference among SVM, decision tree and random forest was relatively small.

TABLE VII. ALGORITHM COMPARISON FOR DETECTION BY 5-FOLD CV.

Algorithm	CV Accuracy (%)		Sensitivity (%)		Specificity (%)	
	NCS	NCS + SpO2	NCS	NCS + SpO2	NCS	NCS + SpO2
kNN	82.9	87.0	47.1	62.0	99.1	98.3
SVM	71.8	87.1	80.3	90.3	68.0	85.7
Decision Tree	77.3	87.5	72.1	88.3	79.7	87.2
Hybrid*	81.5	89.7	69.6	86.8	86.9	91.0
Random Forest	86.2	88.9	63.6	88.6	96.3	89.0

Hybrid* is the voting classifier ensembled from SVM, kNN and decision tree.

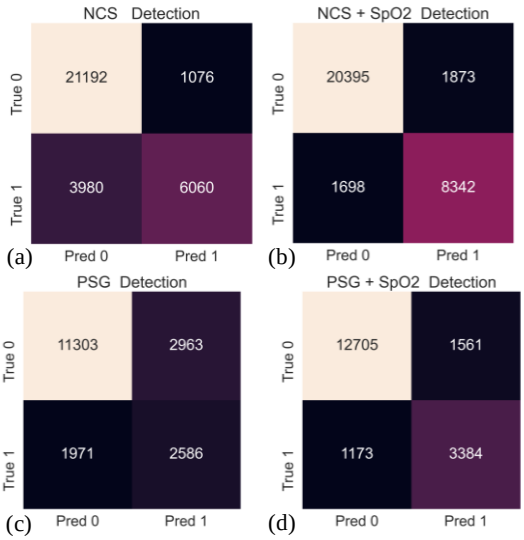

Fig. 7. The confusion matrices showing Normal (0) and Disorders (1) detection by the random forest model using the features from (a) NCS; (b) NCS +SpO2 (c) PSG; (d) PSG +SpO2 by the leave-one-participant-out CV on the entire data.

TABLE VIII. COMPARISON OF THE DETECTION DEVICES BY LEAVE-ONE-PARTICIPANT-OUT CV.

Data set	NCS	NCS + SpO2	PSG	PSG + SpO2
Accuracy	84.4%	88.9%	73.8%	85.5%
Sensitivity	60.4%	83.1%	56.7%	74.3%
Specificity	95.2%	91.6%	79.2%	89.1%
Feature importance	σ_{PP} (0.50) σ_{BR} (0.28) μ_{BR} (0.05) CoV_{PP} (0.02)	σ_{SpO2} (0.62) σ_{PP} (0.15) σ_{BR} (0.08) CoV_{PP} (0.03)	CoV_{PP} (0.43) T_{hold} (0.10) μ_{EX} (0.05) μ_{BR} (0.05)	σ_{SpO2} (0.70) CoV_{PP} (0.03) η_{SpO2} (0.03) *kurt* (0.02)

The results above were tested from *k*-fold CV, and the leave-one-participant-out CV for unseen patients was shown in Fig. 7. High performance of 88.9% accuracy, 83.1% sensitivity and 91.6% specificity was maintained using NCS + SpO2 features. As shown in Figs. 7 (c)(d), for PSG dataset, sensitivity to disorder events remain slightly lower than NCS. The accuracy and feature importance using *k*-fold and leave-one-participant-out CV were similar, as shown in Table VIII.

The misalignment in timing between the thorax and abdomen waveforms in the epoch can be a potential feature for OSA detection from the paradoxical breathing patterns [29]. In supplementary Fig. 4, we presented the additional comparison results with/without adding the time lag feature T_{lag} indicating the misalignment in NCS. T_{lag} was calculated by the shifted time lag of the abdomen channel that gave the highest cross-correlation between the thorax and shifted abdomen channels. Phase and amplitude channels were separately compared. However, the accuracy results after adding the T_{lag} feature were not much improved as shown in supplementary Fig. 4, probably because the OSA event was already represented in other respiratory features. Therefore, we did not include T_{lag} as a feature in the other benchmarks.

D. Results for disorder prediction

In this section, we presented the accuracy statistics for SDB prediction using the waveforms in the normal epochs preceding the disorder epoch by 0 – 90 s. Similar CV tests were performed on prediction datasets as in detection.

Fig. 8 shows the overall confusion matrices using the *k*-fold random forest model, and Table VIII further presents the statistics. Results for the full 7 classes are shown in Supplementary Fig. 5. Similar to detection, prediction of individual events from 7 classes had lower accuracy in comparison with the binary class. In contrast to the detection results when NCS + SpO2 was better than NCS alone, we can find that the performance of using only NCS was comparable to NCS + SpO2, meaning that NCS respiratory information alone can function as a predictor for disorders. Physiologically speaking, low SpO2 was the result of the apneic event, and therefore was useful in detection, but not in prediction. Sensitivity to disorder precedence epochs (81.3%) was relatively lower than those of detection (88.6%), which was also understandable because disorder precedence has less evident changes in respiratory features than the actual disorder events.

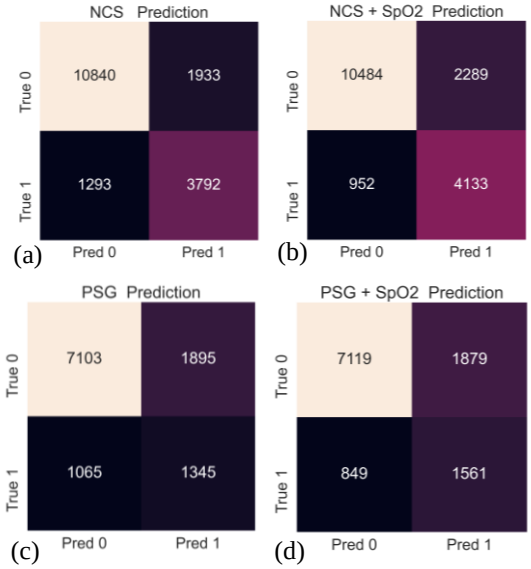

Fig. 8. The confusion matrices showing Regular (0) and Prior (1) prediction by the random forest model using the features from (a) NCS; (b) NCS +SpO2;(c) PSG; (d) PSG +SpO2. The cells list the number of epochs in each category. A 5-fold CV was tested on the entire data.

TABLE IX. COMPARISON OF THE PREDICTION DEVICES BY 5-FOLD CV.

In comparison with PSG, NCS had distinctively higher accuracy and sensitivity for SDB prediction. This was likely because the important feature η_{p0}, representing the signal power dominance within the bandwidth of (0.05, 0.5) Hz, was not well represented in PSG, as NCS had a unique capability to extract motion characteristics of a broad bandwidth [38][39]. The important features for prediction in NCS included η_{p0}, σ_{PP} CoV_{PP}, and σ_{BR} according to Table VIII. To identify forthcoming abnormality, the feature η_{p0} captured whether the waveform was monotonic in the fundamental BR or contained more high-frequency attributes. The features CoV_{PP} and σ_{PP}

represented the peak-to-peak variations, which corresponded to the lung volume. σ_{BR}, representing the breath rate variation, can be important too.

TABLE IX. COMPARISON OF THE PREDICTION DEVICES BY 5-FOLD CV.

Data set	NCS	NCS + SpO2	PSG	PSG + SpO2
Accuracy	81.9%	81.9%	74.1%	76.1%
Sensitivity	74.6%	81.3%	55.8%	64.8%
Specificity	84.9%	82.1%	78.9%	79.1%
Feature importance	η_{p0} (0.31) σ_{PP} (0.30) CoV_{PP} (0.19) σ_{BR} (0.13)	σ_{PP} (0.35) η_{p0} (0.28) σ_{SpO2} (0.18) σ_{BR} (0.09)	σ_{PP} (0.21) min_{PP} (0.17) CoV_{PP} (0.10) μ_{PP} (0.09)	σ_{SpO2} (0.49) η_{SpO2} (0.05) μ_{PP} (0.05) μ_{PP} (0.05)

TABLE X. ALGORITHM COMPARISON FOR PREDICTION BY 5-FOLD CV.

Algorithm	CV Accuracy (%)		Sensitivity (%)		Specificity (%)	
	NCS	NCS + SpO2	NCS	NCS + SpO2	NCS	NCS + SpO2
SVM	77.4	80.4	76.7	80.1	77.8	80.5
kNN	83.0	83.4	46.3	47.6	99.6	99.5
Decision tree	79.7	79.7	81.0	81.0	79.2	79.2
Hybrid*	79.3	82.7	73.9	77.7	81.5	84.7
Random forest	81.9	81.9	74.6	81.3	84.9	82.1

Hybrid* is the voting classifier ensembled from SVM, kNN and decision tree.

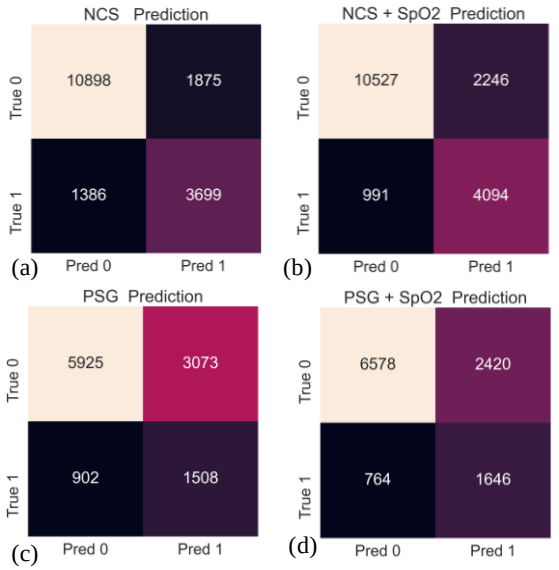

Fig. 9. The confusion matrices showing Regular (0) and Prior (1) prediction by the random forest model using the features from (a) NCS; (b) NCS +SpO2 (c) PSG; (d) PSG +SpO2 by the leave-one-participant-out CV.

TABLE XI. COMPARISON OF THE PREDICTION DEVICES BY LEAVE-ONE-PARTICIPANT-OUT CV.

Data set	NCS	NCS + SpO2	PSG	PSG + SpO2
Accuracy	81.7%	81.9%	65.2%	72.1%
Sensitivity	72.7%	80.5%	62.6%	68.3%
Specificity	85.3%	82.4%	65.8%	73.1%
Feature importance	η_{p0} (0.32) σ_{PP} (0.27) CoV_{PP} (0.21) σ_{BR} (0.14)	σ_{PP} (0.36) η_{p0} (0.28) σ_{SpO2} (0.18) σ_{BR} (0.10)	min_{PP} (0.22) σ_{PP} (0.15) μ_{PP} (0.11) CoV_{PP} (0.08)	σ_{SpO2} (0.51) μ_{PP} (0.06) min_{SpO2} (0.05) μ_{SpO2} (0.05)

Different algorithms for prediction were compared regarding classification performance in Table IX. kNN lacks a reasonable sensitivity to disorder events. Random forest has a good performance on specificity and achieves reasonably high sensitivity and overall accuracy. SVM and decision tree can also generate results with nearly similar performance.

In addition to *k*-fold CV, leave-one-participant-out CVs were performed to validate the robustness of the prediction system on unseen patients and the results are shown in Fig. 9. The result reached 81.7% accuracy, 72.7% sensitivity and 85.3% specificity using only the NCS features. The accuracy remained high for unseen patients and showed a good match with *k*-fold CV results. Similar results from the PSG dataset were presented in Figs. 9 (c)(d). The accuracy and feature importance using leave-one-participant-out CV were shown in Table XI.

We also compared the results for prediction in advance of different time length ranging from 30s to 120s as shown in Fig. 10. Accuracy decreased when the prediction time length increased, and sensitivity degraded significantly when the prediction time exceeded 90 s. We selected 90 s as the final choice to obtain a reasonably high accuracy as well as a longest feasible warning time.

In Fig. 10(b), we presented additional comparison for the different combinations of epoch duration T_{epoch} and sliding window T_{slide} using (upper) NCS and (lower) NCS +SpO2 features. Accuracy was hardly affected by the choices of T_{epoch} and T_{slide} . We chose T_{epoch} = 40 s and T_{slide} = 15 s in the main analysis mostly for convenience and a relatively large number of epochs.

The above results of our SDB detection and prediction were based on the feature extraction followed by the ML model. We also experimented on convolutional neural network (CNN) as the ML model, eliminating the feature extraction process. For CNN, we used the waveform from NCS as the direct input and constructed the network consisting of 5 convolution layers and 3 linear layers. Dataset was divided into training (80%) and testing (20%) parts and the accuracy for the unseen testing data was estimated. Results were presented in Supplementary Fig. 6 for detection (accuracy = 0.650) and Supplementary Fig. 7 for prediction (accuracy = 0.653) using the NCS dataset. In this study, CNN had inferior performance to the approaches using feature extraction followed by the classic ML models. This was

likely due to the insufficient SNR in the raw waveforms, where the feature extraction in the epoch duration could provide some data smoothing and selection effects.

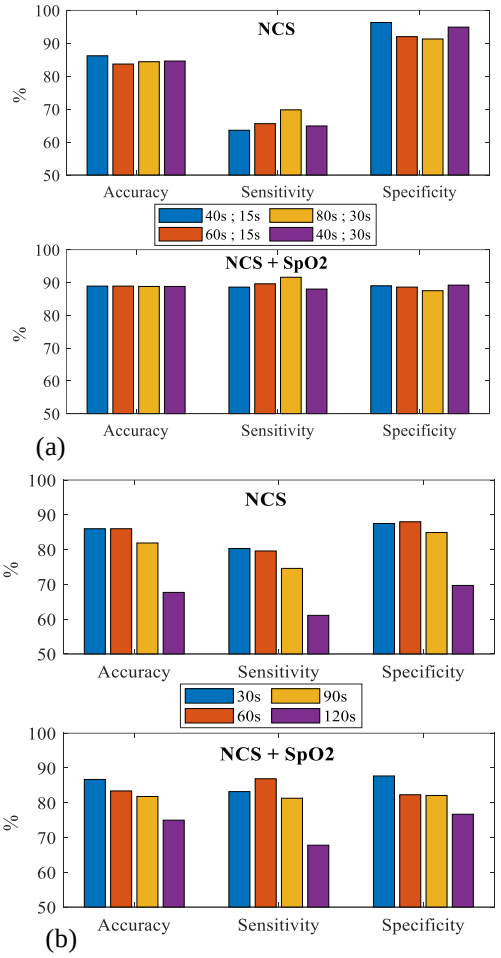

Fig. 10. (a) Comparison of NCS prediction accuracy of the different time length ranging from 30s to 120 s using features from (upper) NCS and (lower) NCS +SpO2. (b) Comparison of detection accuracy using different combinations of epoch duration T_{epoch} and sliding window T_{slide} in the (upper) NCS and (lower) NCS +SpO2 datasets.

E. Results for AHI classification

AHI, calculated by the number of apnea and hypopnea events per hour of sleep, is an important feedback to the patient to indicate the severity of sleep apnea [36]. Individual AHI information for each participant is shown in Supplementary Table 1 according to the technician annotation derived from full PSG observation.

In addition to epoch-based apnea detection and prediction, we also evaluated the performance for AHI classification from the NCS inputs only. We first divided all participants into binary classes of AHI $\leqslant$ 5 as "Normal" and AHI > 5 as "OSA Present" [37]. Using our NCS detection results for normal and disorder, we extracted overall features for each participant including the total epoch numbers of normal and disorder and the NCS F selection rate. We then adopted a simple random forest model by these features from each participant as the input, and estimated the AHI class. The resulting confusion matrix of the two-class AHI classification between annotation and NCS output is shown in Fig. 11 (a). In Fig. 11 (b), we also presented the three-class confusion matrix from AHI $\leqslant$ 5 as "Normal"; 5 < AHI $\leqslant$ 15 as "Mild OSA"; 15 < AHI as "Moderate OSA" [37]. Our NCS estimation achieved accuracy of 0.93 for binary AHI classification, and 0.70 for three classes. Our AHI classification performance is limited due to the present small sample size. Alternatively a random-forest regressor ML model can produce a continuous AHI score [27], which is described in Supplementary Fig. 8. Our AHI accuracy can likely make further improvement in future studies when data from more patients with broader distribution become available.

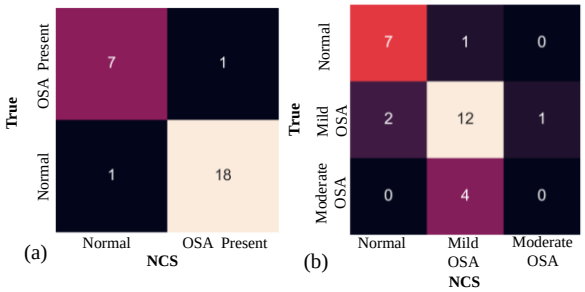

Fig. 11. The Confusion matrix of the (a) two-class and (b) three-class AHI classification between annotation and NCS output from 27 patients. In (a), the two AHI classes are: AHI $\leqslant$ 5: Normal; 5 < AHI : OSA Present, with an accuracy = 0.93; sensitivity = 0.95; specificity = 0.88. In (b), the three AHI classes are: AHI $\leqslant$ 5: Normal; 5 < AHI $\leqslant$ 15: Mild OSA; 15 < AHI < 30: Moderate OSA, with an accuracy = 0.70.

V. Discussion

A. Remaining Challenges

Challenges to construct a clinically acceptable sleep apnea detection and prediction platform still remained:

1) Sensitivity was only above 70% for prediction at the current stage. On the brighter side, the NCS deployment can be invisible to patients throughout the monitoring, and was hence ultimately comfortable and convenient. Our present system can still be a reasonable complement to guide interventions [21]-[26].
2) Our current system cannot classify different respiratory events with sufficient accuracy. Though the binary-class (normal/disorder) results had a relatively high accuracy, our system did not perform well for identifying OSA, CSA, and hypopnea individually. This may be due to the much more hypopnea events than OSA and CSA ones in the dataset. The system thus had limited learning for OSA and CSA, and tend to classify new disorder events as hypopnea.
3) SNR of the waveform needed further improvements. In this study, we relied on NCS epoch selection to eliminate noisy episodes. Sensor improvement for higher SNR and higher tolerance to subject variation and motion interference should be investigated.
4) The snoring event needed features from higher frequency. In this study, the snoring event was not included in the disorder class. However, snoring was an important sleep abnormality in need of more comprehensive investigation. Our NCS sensing technology can couple to low-frequency motion like

respiration [27] as well as high-frequency motion of internal tissues [38]. Snoring can be detected with minimal ambient interference by an additional NCS probe on the jugular or submental area, similar to cough sensing [39].

B. *Future Improvements from Expanded Scope*

The potential extension in future research studies includes:

1) We will expand future studies by adding more severe cases with AHI $\geqslant$ 30. During the execution period of our study protocol in the Weill Cornell Center for Sleep Medicine, we recruited participants with suspected sleep apnea that turned out to have mild and moderate conditions, while others were normal. Future study including patients with wider range of AHI can also help build a more comprehensive and mature learning model.
2) Broadening the demographic groups to include high-risk patients for SDB, including opioid addicts, COPD, or infants at risk for severe apneic events and respiratory arrest, especially those who are born preterm [7][40]. Reliable sleep apnea detection and prediction in these high-risk patients would help improve outcome and prevent fatality.
3) Developing a system that uses SDB prediction to guide real-time intervention including user warning and ambient stimulation, followed by overall effectiveness assessment. The benefit of intervention will likely depend on the prediction accuracy and reliability. By integrating detection, prediction and intervention, we would hopefully improve diagnosis, prognosis and therapy for SDB.
4) Extending the study to include patients with more severe cases of OSA and the associated risk in comorbidities by examining the possible correlation [41]. We will also try to include patients with CSA to further explore the clinical utility of the proposed technology, although the number of patients needs to be much larger due to the infrequent occurrence of CSA. The detection model will also need broader dimension of pathological features in order to achieve a higher confidence level.
5) Improving the learning model to achieve higher accuracy and reliability. Improvements can include additional preprocessing for feature extraction, better noise reduction algorithms, and more complex ML models such as gain-adversarial network (GAN) [42].

VI. Conclusion

In this paper, we reported a hardware-software co-designed system that can detect and predict SDB. This system was based on a covert bed-integrated RF sensor by NCS, which can be non-invasive and invisible to user. SDB detection for considering apneas and hypopneas together achieved a sensitivity and specificity up to 88.6% and 89.0% for *k*-fold validation, and 83.1% and 91.6% for subject-independent validation, respectively. Subsequent apneic events can be predicted up to 90 s in advance based on the present respiratory features. Disorder prediction achieved a sensitivity and specificity up to 81.3% and 82.1% for *k*-fold validation, and 80.5.0% and 82.4% for subject-independent validation, respectively. By the random forest ML model, the most significant physiological symptoms before and during the SDB episodes can also be revealed.

The current sleep apnea diagnosis platform was mostly based on PSG, which remained expensive in terms of hardware and operators, uncomfortable from body electrodes, and time-consuming for deployment. The ability to predict upcoming SDB events by PSG was also limited. In the future, our covert detection and prediction system could expedite intervention, and improve diagnosis and therapy for respiratory disturbance during sleep.

References

[1] H. R. Colten and B. M. Altevogt, Eds., *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press, 2006.
[2] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, "The occurrence of sleep-disordered breathing among middle-aged adults," *New Engl. J. Med.*, vol. 328, no. 17, pp. 1230-1235, 1993.
[3] T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women," *Sleep*, vol. 20, no. 9, pp. 705-706, 1997.
[4] C. Senaratna et al., "Prevalence of obstructive sleep apnea in the general population: A systematic review," *Sleep Med. Rev.*, vol. 34, no. 1, pp. 70–81, Aug. 2017.
[5] A. Roebuck et al., "A review of signals used in sleep analysis," *Physiol. Meas.*, vol. 27, no. 3, pp. 320–331, 2015.
[6] E. A. Phillipson, "Sleep apnea--a major public health problem," *New Engl. J. Med.*, vol. 328, no. 17, pp. 1271-1273, 1993.
[7] E. S. Katz, R. B. Mitchell, and C.M. D'Ambrosio, "Obstructive sleep apnea in infants," *Am. J. Respir. Crit.*, vol. 185, no. 8, pp. 805-816, Apr. 2012.
[8] V. Mysliwiec et al., "Sleep disorders in US military personnel: a high rate of comorbid insomnia and obstructive sleep apnea," *Chest*, vol. 144, no. 2, pp. 549-557, Aug. 2013.
[9] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias and T. Penzel, "A review of obstructive sleep apnea detection approaches," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 825-837, Mar. 2019.
[10] W. W. Flemons et al., "Home diagnosis of sleep apnea: A systematic review of the literature," *Chest*, vol. 124, no. 4, pp. 1543-1579, Oct. 2003.
[11] O. C. Ioachimescu et al., "Performance of peripheral arterial tonometry–based testing for the diagnosis of obstructive sleep apnea in a large sleep clinic cohort," *J Clin. Sleep Med.*, vol. 16, no. 10, pp. 1663-1674, Oct. 2020.
[12] O. Faust, U. R. Acharya, E. Ng, and H. Fujita, "A review of ECG-based diagnosis support systems for obstructive sleep apnea," *J. Mech. Med. Biol.*, vol. 16, no. 01, p. 1640004, Feb. 2016.
[13] G. D. Clifford, F. Azuaje, and P. McSharry, *Advanced Methods and Tools for ECG Data Analysis*. Artech house Boston, 2006.
[14] U. J. Magalang et al., "Prediction of the apnea-hypopnea index from overnight pulse oximetry," *Chest*, vol. 124, no. 5, pp. 1694-1701, Nov. 2003.
[15] V. Monasterio, F. Burgess, and G. D. Clifford, "Robust classification of neonatal apnoea-related desaturations," *Physiol. Meas.*, vol. 33, no. 9, p. 1503, Aug. 2012.
[16] N. Pombo, N. Garcia, and K. Bousson, "Classification techniques on computerized systems to predict and/or to detect Apnea: A systematic review," *Comput. Methods Programs Biomed.*, vol. 140, pp. 265-274, Mar. 2017.
[17] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 469-477, May 2012.
[18] A. Burgos, A. Goni, A. Illarramendi, and J. Bermudez, "Real-time detection of apneas on a PDA,", *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 4, pp. 995-1002, Nov. 2009.
[19] J. Bock and D. A. Gough, "Toward prediction of physiological state signals in sleep apnea," *IEEE. Trans. Biomed. Eng.*, vol. 45, no. 11, pp. 1332-1341, Nov. 1998.
[20] J. A. Waxman, D. Graupe, and D. W. J. A. j. o. r. Carley, "Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network," *Am. J. Respir. Crit.*, vol. 181, no. 7, pp. 727-733, Apr. 2010.

[21] N. B. Kribbs et al., "Objective measurement of patterns of nasal CPAP use by patients with obstructive sleep apnea," *Am. Rev. Respir. Dis.*, vol. 147, no. 4, pp. 887-895, Apr. 1993.
[22] R. G. Andrade et al., "Nasal vs. oronasal CPAP for OSA treatment: A meta-analysis," *Chest*, vol. 153, no. 3, pp. 665-674, Mar. 2018.
[23] R. B. Berry, J. M. Parish, and K. M. Hartse, "The use of auto-titrating continuous positive airway pressure for treatment of adult obstructive sleep apnea," *Sleep*, vol. 25, no. 2, pp. 148-173, Jan. 2002.
[24] V. Patruno et al., "Fixed and auto-adjusting continuous positive airway pressure treatments are not similar in reducing cardiovascular risk factors in patients with obstructive sleep apnea," *Chest*, vol. 131, no. 5, pp. 1393-1399, May 2007.
[25] A. Esteban et al., "How is mechanical ventilation employed in the intensive care unit? An international utilization review," *Am. J. Respir. Crit. Care Med.*, vol. 161, no. 5, pp. 1450-1458, May 2000.
[26] "Sleep smart, smart pillow," ZEREMA. [Online]. Available: https://www.zerema.co/. [Accessed: 06-Oct-2021].
[27] Z. Zhang, P. Sharma, T. B. Conroy, V. Phongtankuel, and E. C. Kan, "Objective scoring of physiologically induced dyspnea by non-invasive RF sensors," *IEEE. Trans. Biomed. Eng.*, July 2021, doi: 10.1109/TBME.2021.3096462.
[28] Z. Zhang, P. Sharma, J. Zhou, X. Hui, and E. C. Kan, "Furniture-integrated respiration sensors by notched transmission lines," *IEEE Sens. J.*, vol. 21, no. 4, pp. 5303-5311, Feb. 2021.
[29] P. Sharma, X. Hui, J. Zhou, T. B. Conroy, and E. C. Kan, "Wearable radio-frequency sensing of respiratory rate, respiratory volume, and heart rate," *NPJ Digit. Med*., vol. 3, p. 98, July 2020.
[30] X. Hui, P. Sharma, and E. C. Kan, "Microwave stethoscope for heart sound by near-field coherent sensing," *Proc. IEEE MTT-S International Microwave Symposium (IMS)*, 2019, pp. 365-368.
[31] R. B. Berry et al., "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events: Deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine," *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597-619, Oct. 2012.
[32] R. W. Schafer, "What is a Savitzky-Golay filter? [lecture notes]," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111-117, July 2011.
[33] W. Lu et al., "A semi-automatic method for peak and valley detection in free-breathing respiratory waveforms," *Med. Phys.*, vol. 33, no. 10, pp. 3634-6, Oct. 2006, doi: 10.1118/1.2348764.
[34] T. K. Ho, "The random subspace method for constructing decision forests *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832-844, Aug. 1998.
[35] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discov. Data (TKDD)*, vol. 6, no. 1, pp. 1-39, Mar. 2012.
[36] V. Hoffstein and J. Szalai, "Predictive value of clinical features in diagnosing obstructive sleep apnea," *Sleep*, vol. 16, no. 2, pp. 118-122, Mar. 1993.
[37] W. R. Ruehland, P. D. Rochford, F. J. O'Donoghue, R. J. Pierce, P. Singh, and A. T. Thornton, "The new AASM criteria for scoring hypopneas: impact on the apnea hypopnea index," *Sleep*, vol. 32, no. 2, pp. 150-157, Feb. 2009.
[38] X. Hui, T. B. Conroy, and E. C. Kan, "Near-field coherent sensing of vibration with harmonic analysis and balance signal injection," *IEEE Trans. Micro. Theory Tech.*, vol. 69, no. 3, pp. 1906-1916, May 2021.
[39] X. Hui, J. Zhou, P. Sharma, T. B. Conroy, Z. Zhang and E. C. Kan, "Wearable RF near-field cough monitoring by frequency-time deep learning", *IEEE Trans. Biomed. Circuits & Sys*, vol. 15, no. 4, pp. 756 – 764, Aug. 2021, doi: 10.1109/TBCAS.2021.3099865.
[40] D. R. Halloran and G. R. Alexander, "Preterm delivery and age of SIDS death," *Ann Epidemiol.*, vol. 16, no. 8, pp. 600 – 606, Aug. 2006, doi: 10.1016/j.annepidem.2005.11.007.
[41] J. A. Pinto, D. K. Ribeiro, A. F. d. S. Cavallini, C. Duarte, and G. S. Freitas, "Comorbidities associated with obstructive sleep apnea: A retrospective study," *Int. Arch. Otorhinolaryngol.*, vol. 20, pp. 145-150, 2016.
[42] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowledge-Based Systems*, vol. 187, p. 104837, Oct. 2020.