SLEEP BREATHING PHYSIOLOGY AND DISORDERS · ORIGINAL ARTICLE

Detecting arousals and sleep from respiratory inductance plethysmography

Eysteinn Finnsson[1,2] · Ernir Erlingsson[1,3] · Hlynur D. Hlynsson[1] · Vaka Valsdóttir[1] · Thora B. Sigmarsdottir[1] · Eydís Arnardóttir[1] · Scott A. Sands[4] · Sigurður Æ Jónsson[1] · Anna S. Islind[2] · Jón S. Ágústsson[1]

Abstract

Purpose Accurately identifying sleep states (REM, NREM, and Wake) and brief awakenings (arousals) is essential for diagnosing sleep disorders. Polysomnography (PSG) is the gold standard for such assessments but is costly and requires overnight monitoring in a lab. Home sleep testing (HST) offers a more accessible alternative, relying primarily on breathing measurements but lacks electroencephalography, limiting its ability to evaluate sleep and arousals directly. This study evaluates a deep learning algorithm which determines sleep states and arousals from breathing signals.

Methods A novel deep learning algorithm was developed to classify sleep states and detect arousals from respiratory inductance plethysmography signals. Sleep states were predicted for 30-s intervals (one sleep epoch), while arousal probabilities were calculated at 1-s resolution. Validation was conducted on a clinical dataset of 1,299 adults with suspected sleep disorders. Performance was assessed at the epoch level for sensitivity and specificity, with agreement analyses for arousal index (ArI) and total sleep time (TST).

Results The algorithm achieved sensitivity and specificity of 77.9% and 96.2% for Wake, 93.9% and 80.4% for NREM, 80.5% and 98.2% for REM, and 66.1% and 86.7% for arousals. Bland–Altman analysis showed ArI limits of agreement ranging from - 32 to 24 events/hour (bias: - 4.4) and TST limits from - 47 to 64 min (bias: 8.0). Intraclass correlation was 0.74 for ArI and 0.91 for TST.

Conclusion The algorithm identifies sleep states and arousals from breathing signals with agreement comparable to established variability in manual scoring. These results highlight its potential to advance HST capabilities, enabling more accessible, cost-effective and reliable sleep diagnostics.

Keywords Machine learning · Home sleep testing · Sleep disorders · Sleep stage · Polysomnography

Introduction

During sleep, the brain transitions between different states: rapid eye movement (REM) sleep, non-REM (NREM) sleep (which encompasses N1, N2, and N3), and wakefulness. Monitoring these states is crucial for diagnosing and assessing various sleep disorders which may be characterized by disruption of the usual patterns and duration of NREM and REM sleep, or an unusual number of awakenings or arousals [1, 2]. This is typically done using polysomnography (PSG) equipped with electroencephalography (EEG),

electrooculography (EOG), electromyography (EMG), electrocardiography (ECG), respiratory inductance plethysmography (RIP), nasal cannula, and pulse oximetry. However, PSG studies are expensive and often inconvenient for patients, leading to the adoption of home sleep testing (HST) as a more accessible alternative that focuses primarily on breathing and oxygenation signals. While HST is more affordable and easier to administer it has certain limitations. Notably, it cannot accurately assess sleep states and arousals due to the absence of EEG, EOG, and EMG measurements. These limitations reduce the effectiveness of HST in diagnosing sleep disorders other than sleep apnea. Moreover, it may result in underdiagnosis of sleep apnea in patients with reduced sleep time or those with a high proportion of hypopneas associated with arousals rather than desaturations [3, 4].

Eysteinn Finnsson and Ernir Erlingsson contributed equally to this work.

Extended author information available on the last page of the article

There are indications that sleep states and arousals can be inferred from breathing patterns. For instance, breathing during wakefulness is characterized by volitional overwrite of breathing control, including swallows, sniffs, breath-holds, speech, sighs, coughs, and body movements. In contrast, breathing is remarkably monotonous during NREM sleep, controlled primarily by metabolic demand [5, 6]. During REM sleep, breathing becomes more variable, both in terms of breath amplitude and timing. Additionally, REM sleep is characterized by a significant reduction in skeletal muscle tone, with breathing predominantly driven by the diaphragm [7, 8]. This loss of muscle tone affects the intercostal muscles, whose contraction expands the chest wall, leading to a distinctive breathing pattern during REM where the chest wall caves in during inspiration [6]. These variations in breathing dynamics during sleep reveal a strong coupling between the sleep states and physiological changes that are reflected in the breathing patterns. Indeed, previous work has indicated that there is information in the respiratory signal for discriminating between REM, NREM and Wake [9–12].

Likewise, there are reasons to think that arousals can also be detected by analyzing breathing patterns. Arousals are brief awakenings from sleep, often triggered by stimuli such as loud sounds or breathing obstructions. These arousals are associated with sympathetic activation within the autonomic nervous system, leading to increases in heart rate and blood pressure [13]. Additionally, ventilation dynamics are altered following arousal through both chemical and state-related mechanisms [6]. Accompanying arousals is an increased activation of skeletal muscles, including the engagement of the upper airway dilator muscles [6, 14]. This muscle activation plays a crucial role in resolving upper airway obstructions, particularly in the context of sleep apnea [15]. Arousals have further been shown to cause a transient increase in ventilation, a gasp-like reflex, that is called the ventilatory response to arousal (VRA). This reflex is independent of metabolic demand and occurs even in the absence of upper airway obstruction (e.g., apnea and hypopnea) [16–20]. Research on auditory-induced arousals demonstrated that in addition to an increase in ventilation, the VRA is accompanied by a decrease in inspiratory duration and remains consistent across different sleep states [16].

To measure breathing movements of the chest wall and abdomen, both PSG and HST rely on RIP technology. The American Academy of Sleep Medicine (AASM) recommends using RIP belts for detecting the presence of respiratory effort and as a secondary flow sensor to the nasal cannula [2]. High-quality RIP belts have been shown to provide a reliable measure of airflow, with strong correlation to nasal cannula readings [21]. The ability of RIP to simultaneously measure the dynamics of chest wall movements and airflow means that it captures all the relevant physiology required to assess sleep states and arousals from breathing.

In this paper, we apply artificial intelligence (AI) to infer the relationship between sleep and breathing, developing an algorithm that detects sleep states and arousals solely from breathing. This enables subsequent calculation of total sleep time (TST) as well as arousal index (ArI)—an indicator of sleep fragmentation—from HST, significantly expanding its scope and efficacy. This innovative approach has the potential to transform HST by providing a comprehensive assessment of sleep, while simultaneously paving the way towards more accurate and accessible diagnosis of sleep disorders.

Materials and methods

Nox BodySleep 2.0 (NBS2) is an AI algorithm designed to determine sleep states and detect arousals using abdominal and thoracic RIP signals. The input signals are processed by a neural network, which assigns probabilities to each sleep state as well as to the likelihood of an arousal. An extensive dataset was used to assess the algorithm's efficacy, with careful attention to ensure that the subjects and data reflect the algorithm's intended application in standard clinical settings.

NBS2 is designed to aid in the scoring of home sleep tests that include no electrophysiological signals (e.g., EEG, EOG, and EMG). However, to rigorously evaluate NBS2, it was necessary to compare it with "gold standard" sleep and arousal scoring, which requires in-laboratory PSG recordings with high-quality electrophysiological signals. Accordingly, in the current study, PSG recordings were manually scored using all available signals while only RIP data were input to NBS2 to provide AI scoring for comparison, see Table 1. Of note, since our goal was to discriminate between three distinct states, namely Wake, REM and NREM, the gold standard labels for N1, N2, and N3 were combined into a single NREM state.

Algorithm development

The NBS2 algorithm was developed using a dataset of PSG recordings from clinical and research settings across seven sites in Europe, Asia, and the United States. In total, 3,185 recordings were used: 2,216 for training, 482 for tuning, and 487 for testing. Note that the results presented in this study are based on an entirely separate dataset, unrelated to those used for algorithm training and development.

Training was performed using cross-entropy loss, optimized with the AdamW method as implemented in TensorFlow. Overfitting was monitored and the following regularization techniques were applied: spatial dropout and batch normalization. Data augmentation was employed which involved scaling the input signals and introducing random

Table 1 Comparison of signals recorded during in-laboratory polysomnography (PSG) and home sleep tests, along with the required inputs for the current algorithm under evaluation

	EEG	EOG	EMG	ECG	Flow	RIP	SpO₂	Actigraphy
In-lab PSG	✓	✓	✓	✓	✓	✓	✓	✓
Home sleep test					✓	✓	✓	✓
Nox BodySleep 2.0						✓		

* Electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), electrocardiography (ECG), airflow (Flow), respiratory inductance plethysmography (RIP), oxygen saturation (SpO₂)

temporal shifts. Training samples were selected with a 50% probability of containing a scored arousal.

Participants

A large set of in-lab PSG recordings were used to comparatively evaluate the algorithm's performance. The PSG studies were drawn from dataset provided by FusionSleep clinics in Atlanta, Georgia, USA, between 2020 and 2023 (inclusive). This dataset is representative of the algorithm's intended use in standard clinical practice in sleep clinics and hospitals, ensuring that the comparative results accurately reflect its real-world performance. Data was recorded using the Nox A1 sleep recorder (Nox Medical, Iceland). All recordings underwent standardized quality control and were scored by qualified sleep technologists according to the AASM manual for scoring sleep and associated events [2]. Finally, the scoring was reviewed and approved by a medical doctor.

The PSG recordings used for evaluation were selected based on predefined inclusion criteria, requiring that each study be manually scored, conducted during nighttime, and include an adult participant (≥ 18 years old). Additionally, both RIP belts had to be connected for at least half of the recorded sleep time. Studies not meeting these criteria were excluded from the analysis. No information was available on subject's medication use or preexisting health conditions for this study.

Algorithm description

NBS2 requires only the abdomen and thorax RIP belt inductance signals as input. These signals are sampled at 25 Hz with a bandwidth of 0–12.5 Hz, measuring the loop inductance of the wire in each belt encircling the patient. A non-linear adaptive filter is used to remove the baseline, gradually suppressing frequencies below 0.1 Hz while adapting to larger changes, such as patient movement. Nine consecutive signal epochs (where each epoch is a 30 s temporal window) are employed to process each target epoch, with four epochs preceding and four epochs following the target, as depicted in Fig. 1. For each of these epochs, there are two
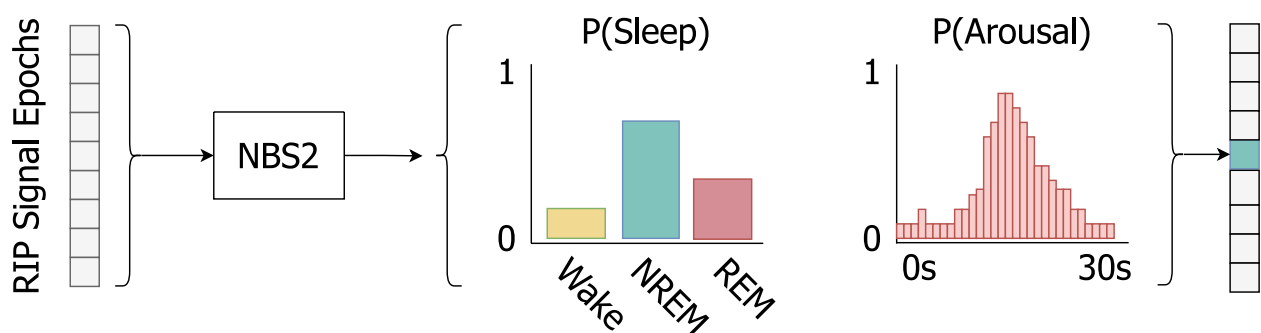
Fig. 1 Overview of the Nox BodySleep 2.0 (NBS2) algorithm. The input consists of nine consecutive 30 s respiratory inductance plethysmography (RIP) signal epochs. NBS2 predicts the probability of sleep states (Wake, NREM, REM) and arousal (1 s resolution) for the center epoch. Sleep state is scored by selecting the state with the highest probability. An arousal is scored using a rule-based post-processing step, which requires that the arousal probability exceeds a specific threshold for at least 3 s. The output probabilities shown in the figure are illustrative and do not represent actual algorithm outputs

target outputs: 1) The sleep state in each epoch is derived according to the highest probability, followed by a holistic regression analysis that refines it, i.e., smooths consecutive sleep state outputs, which is a common regularization technique for time-series data [22]; 2) The algorithm computes the probability of an arousal event for each second within the target epoch. An arousal is subsequently identified if at least three consecutive seconds exceed a predefined threshold, in which case an arousal is scored along with its duration, see Fig. 2.

The neural network architecture is based on temporal convolutional networks (TCN), following the suggestions made by S. Bai et al. [23] to improve the performance of sequence modelling tasks. The architecture employs residual connections, regularization, and activation functions, to form a chain of residual blocks [24–26]. A schematic of model architecture is available in the supplementary materials (Figures S1-S2).

TCNs possess several features well suited to the task at hand. They efficiently handle long-range dependencies using dilated convolutions, effectively capturing the necessary context surrounding the window of interest. Furthermore, their ability to extract time-invariant features is particularly important for precisely localizing sleep-related events like arousals. Using a multi-task learning approach, we train a single model to handle both arousal detection and sleep state classification, enabling shared feature learning with separate outputs for each task (see Figure S2. in the supplementary materials).

Statistical analysis

To evaluate the NBS2 algorithm's performance, we use both epoch-level agreement metrics and patient-level measures. Sensitivity, specificity, accuracy, and F1 score, are calculated to assess epoch-level agreement between NBS2 outputs and manual scoring. Additionally, Cohen's Kappa is computed to quantify agreement, corrected for chance. Epochs containing manually scored signal artifacts are excluded from analysis. For patient-level measures, specifically TST and ArI, agreement are evaluated using Bland–Altman analysis and intraclass correlation coefficients (ICC).

Epoch-level sensitivity, specificity, accuracy, and F1 score are defined as follows: sensitivity $= TP/(TP + FN)$, specificity $= TN/(TN + FP)$, accuracy $= (TP + TN)/(TP + TN + FP + FN)$, and F1 score $= 2\,TP/(2\,TP + FP + FN)$ where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Sensitivity, specificity, accuracy, and F1 score, are calculated with 95% confidence intervals. The sleep state metrics are calculated for each class using a one-vs-rest method, while arousal epochs are identified based on their epoch presence or absence in either NBS2 or the manual scoring. Arousal detection performance is also analyzed by REM and NREM sleep states. Epoch-level measures of arousal detection are selected for comparison of medical devices through their FDA 510(k) summaries [27–29].

For patient-level measures, Bland–Altman analysis is used to calculate mean difference (bias) and limits of
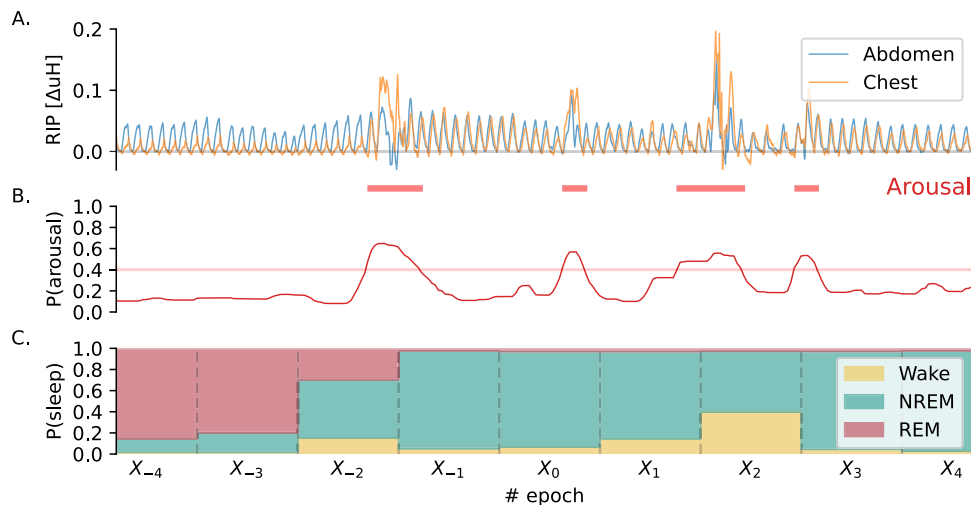
Fig. 2 The algorithm analyses 4.5 min data segments at a time. Subplot A shows breathing movements recorded by two respiratory inductance plethysmography (RIP) belts placed around the abdomen and chest, which serve as inputs to the algorithm. Subplots B and C display the neural-network outputs: arousal probabilities and sleep state probabilities. Arousals are marked with red squares when the arousal probability exceeds a threshold for at least 3 s. The sleep state prediction in subplot C is presented as a hypnodensity graph [53], predicting the likelihood of each sleep state—Wake, NREM, or REM—across epochs. The depicted window of data captures a transition from REM sleep, characterized by reduced chest wall movement, to NREM sleep, where chest movement increases. This transition is associated with arousals and a higher probability of wakefulness

Table 2 Subject characteristics

Characteristic	Value	Unit
Age (Mean $\pm$ SD)	48.6 ± 17.0	years
Sex, N (M:F:unknown)	616:673:10	count
BMI (Mean $\pm$ SD)	29.5 ± 7.8	kg/m^2
AHI (Mean $\pm$ SD)	26.9 ± 24.9	events/hour
AHI < 5	118	count
$5 \leq$ AHI < 15	416	count
$15 \leq$ AHI < 30	344	count
AHI ≥ 30	421	count
PLMSI (Mean $\pm$ SD)	2.2 ± 9.2	events/hour

* Body Mass Index (BMI), Apnea Hypopnea Index (AHI), Periodic Limb Movement of Sleep Index (PLMSI)

agreement (LoA), both with 95% confidence intervals. Similarly, ICC is computed, with corresponding 95% confidence intervals. Additionally, the algorithm's arousal scoring performance is evaluated across obstructive sleep apnea (OSA) severity levels, specifically examining ArI bias, LoA and ICC within defined apnea–hypopnea index (AHI) groups: normal (AHI < 5), mild ($5 \leq$ AHI < 15), moderate ($15 \leq$ AHI < 30), and severe (AHI ≥ 30).

In a supplementary analysis, all epoch- and patient-level measures are provided for each OSA severity group, as well as for periodic limb movement of sleep index (PLMSI) subgroups (PLMSI < 15 and PLMSI ≥ 15).

All confidence intervals are determined via bootstrapping with 10,000 iterations.

Results

After applying the exclusion criteria to the initial pool of 1829 studies, the final validation dataset comprised 1,299 sleep studies, including 616 male subjects, 673 female

Table 4 The normalized epoch-level confusion matrix for algorithm-classified versus manually classified sleep states, with the total number of scored epochs in the far-right column

		Algorithm prediction			
		Wake	NREM	REM	# Epochs
Manual scoring	Wake	78%	21%	1%	216,260
	NREM	4%	94%	2%	684,101
	REM	2%	18%	80%	136,239

Table 5 Normalized epoch-level confusion matrix for algorithm-classified versus manually classified arousal scoring, with the total number of scored epochs in the far-right column

|		Algorithm prediction		
		Positive	Negative	# Epochs
Manual scoring	Positive	66%	34%	250,773
	Negative	13%	87%	786,033

subjects, and 10 subjects with unspecified gender (Table 2 provides additional anthropometric data). Overall, the RIP signal connectivity was good, with 98% of sleep studies exceeding 95% connectivity throughout the recording period, including initial sensor setup and end-of-study removal.

Table 3 presents epoch-level sensitivity, specificity, accuracy, and F1 score, along with their respective 95% confidence intervals. Additionally, Table S1 in the supplementary materials shows the distribution of performance across the study population. Detailed confusion matrices are provided in Tables 4 and 5. In Table 4, the classification of manually scored epochs by the NBS2 algorithm is displayed. The results indicate that NBS2 accurately classifies 78% of Wake epochs, though it misclassifies 21% of these as NREM sleep. NBS2 correctly scores 94% of NREM epochs and 80%

Table 3 Epoch-level performance of the algorithm under investigation compared to manual scoring, presented in terms of sensitivity, specificity, accuracy, and F1 score. Arousal detection performance is further stratified by NREM and REM sleep states

	Sensitivity % [95% CI]	Specificity % [95% CI]	Accuracy % [95% CI]	F1 Score % [95% CI]
Sleep State Classification				
Wake	77.9 [76.8, 78.9]	96.2 [95.9, 96.5]	92.4 [92.1, 92.7]	81.0 [80.2, 81.8]
NREM	93.9 [93.5, 94.3]	80.4 [79.7, 81.2]	89.3 [89.0, 89.6]	92.1 [91.8, 92.3]
REM	80.5 [79.3, 81.6]	98.2 [98.1, 98.3]	95.9 [95.7, 96.1]	83.7 [83.0, 84.4]
Arousal Classification				
Arousal	66.1 [64.9, 67.2]	86.7 [86.2, 87.2]	81.7 [81.3, 82.1]	63.6 [62.7, 64.5]
Arousal Classification by sleep state				
Arousal | NREM	64.0 [62.7, 65.3]	90.9 [90.4, 91.4]	83.0 [82.6, 83.5]	68.8 [67.8, 69.8]
Arousal | REM	66.8 [65.3, 68.3]	86.9 [86.2, 87.6]	82.3 [81.7, 82.8]	63.5 [62.3, 64.5]
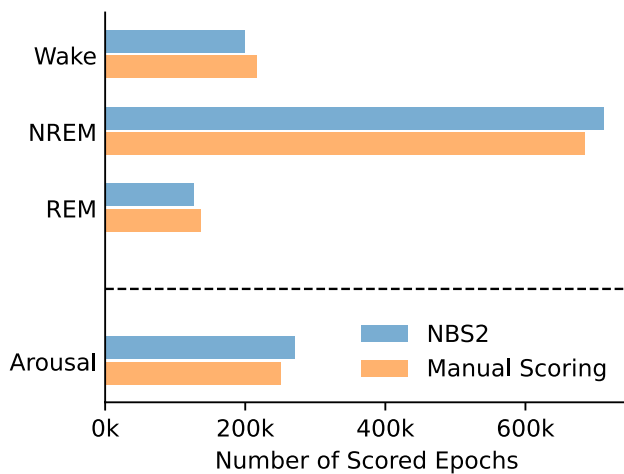
Fig. 3 Comparison of the total number of epochs scored as Wake, NREM, REM, and those with arousals, between Nox BodySleep 2.0 (NBS2) and manual polysomnography scoring

of the REM epochs. Of the REM epochs that were manually scored as REM, NBS2 misclassified 18% as NREM. Cohen's Kappa value for overall sleep-state agreement was

0.77. Table 5 shows that NBS2 correctly identifies arousals in 66% of the epochs with a manually scored arousal and appropriately avoids scoring arousals in 87% of the epochs where no arousal was manually scored. Cohen's Kappa value for overall arousal scoring agreement was 0.51.

Figure 3 compares the distribution of sleep states and arousal events between the NBS2 algorithm and manual PSG scoring, showing that NBS2 produces similar numbers of Wake, NREM, and REM sleep states, as well as a comparable number of arousal events, across the entire dataset. On average, arousals scored by NBS2 were 1.9 s longer than those scored manually. However, the total number of arousals identified by each approach was nearly identical, with NBS2 detecting 241,817 arousals and manual scoring identifying 241,377 see Figure S3 in the supplementary materials.

The agreement between NBS2 and manual scoring for both ArI and TST is assessed using Bland–Altman analysis of bias and LoA. For ArI, the analysis showed a bias of -4.4 (95% CI: -5.2, -3.6) events/hour. The 95% LoA ranged from -32 (95% CI: -35, -29) to 24 (95% CI: 21, 25) events/hour. Proportionality bias with slope -0.127 ($p < 0.05$) was found. Log transformed Bland–Altman plot is available in Figure S4 in supplementary materials. For

Fig. 4 Bland–Altman plot showing the agreement between polysomnography scoring and Nox BodySleep 2.0 (NBS2) for Arousal Index (ArI). The plot includes the mean difference (bias, dot-dash line) and 95% limits of agreement (dashed blue lines). Each data point represents a single sleep study. The 95% confidence intervals for the bias and limits of agreement are shown as shaded regions
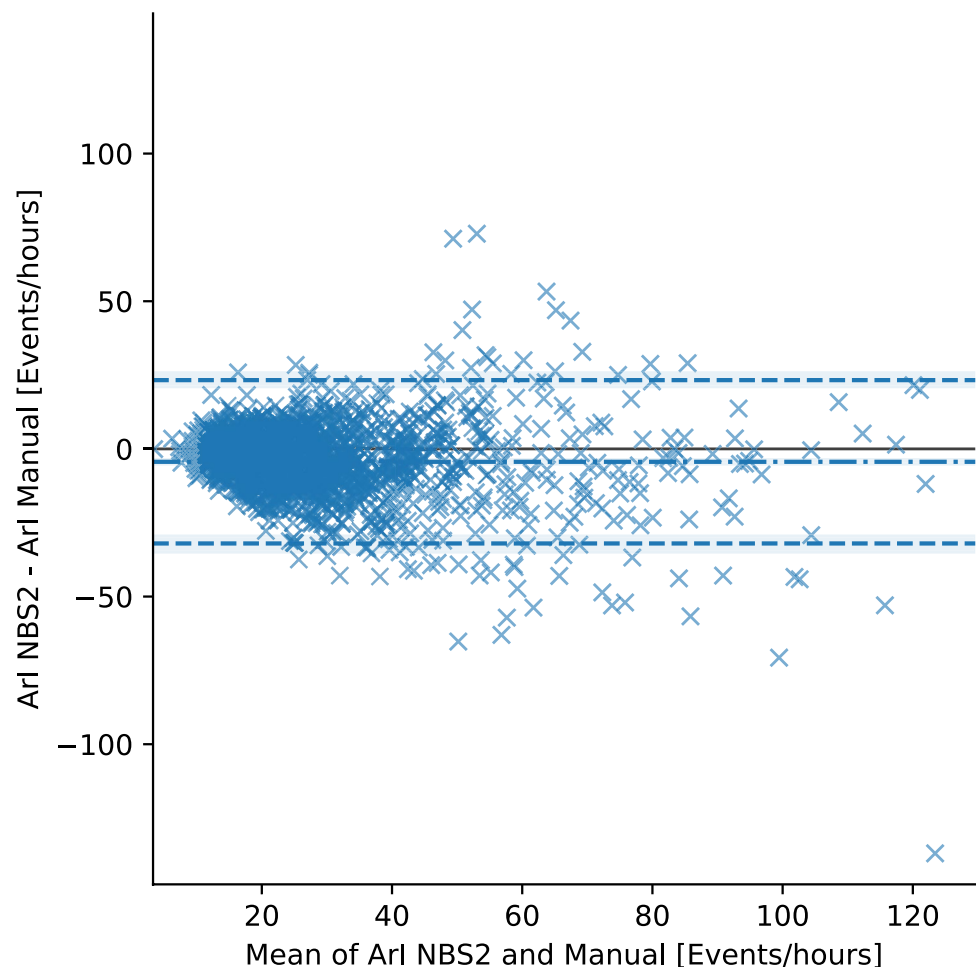
Fig. 5 Bland–Altman plot showing the agreement between manual polysomnography scoring and Nox BodySleep 2.0 (NBS2) for total sleep time (TST). The plot includes the mean difference (bias, dot-dash line) and 95% limits of agreement (dashed blue lines). Each data point represents a single sleep study. The 95% confidence intervals for the bias and limits of agreement are shown as shaded regions
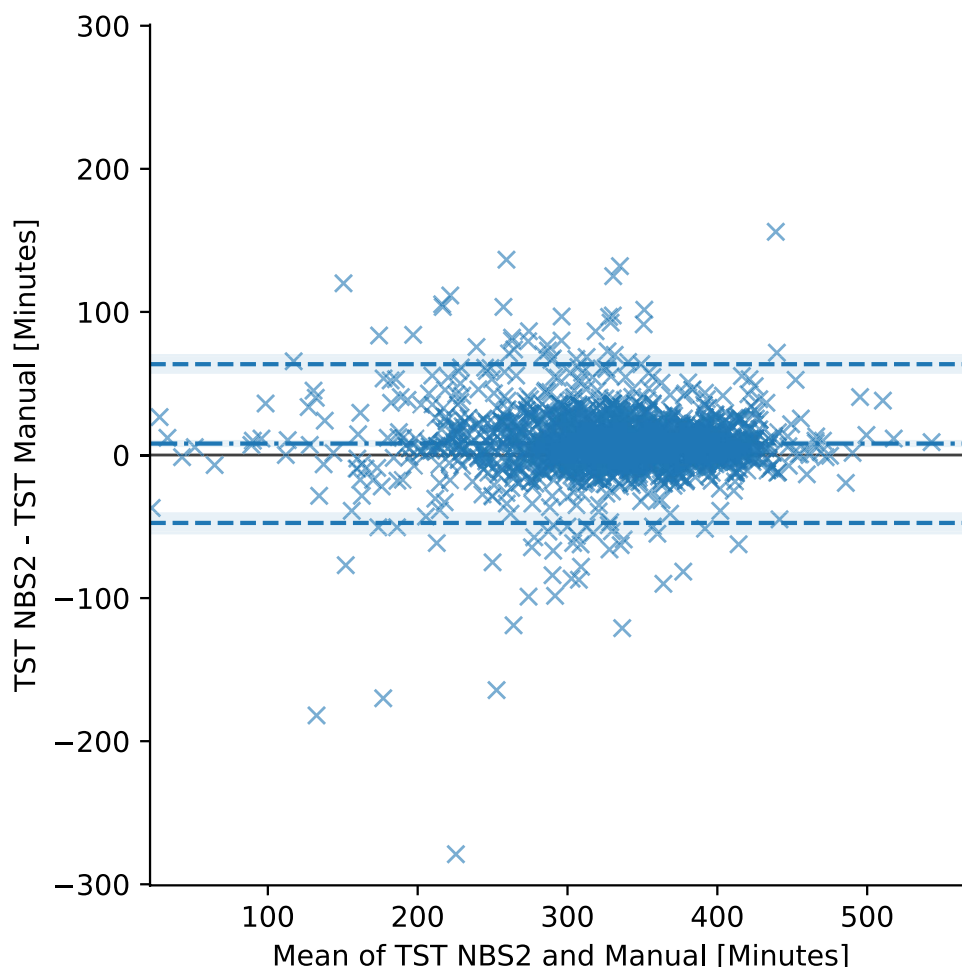
TST, the bias is 8.0 (95% CI: 6.5, 9.5) minutes, with 95% LoA spanning from − 47 (95% CI: − 54, − 41) to 64 (95% CI: 57, 69) minutes. See Bland–Altman plots in Figs. 4 and 5. ICC for ArI is 0.74 (95% CI: 0.69, 0.77) and for TST, it is 0.91 (95% CI: 0.89, 0.93).

In evaluating NBS2's performance in scoring arousals across OSA severity levels we observed consistent results. Here, LoA is shown as 1.96 times the standard deviation of the error. For patients without OSA (AHI < 5; $n = 118$), the arousal index showed a bias of − 3.63 events/hour, a LoA of 20.22 events/hour, and an ICC of 0.68 (95% CI: 0.57, 0.78). In mild OSA ($5 \leq$ AHI < 15; $n = 416$), the bias was − 3.08 events/hour with a LoA of 19.83 events/hour and an ICC of 0.67 (95% CI: 0.57, 0.76). Moderate OSA ($15 \leq$ AHI < 30; $n = 344$) had a bias of − 3.61 events/hour, a LoA of 21.34 events/hour, and an ICC of 0.54 (95% CI: 0.42, 0.66). In severe OSA (AHI $\geq$ 30; $n = 421$), the bias was higher at − 6.46 events/hour, with a LoA of 37.80 events/hour and an ICC of 0.62 (95% CI: 0.53, 0.69). A more detailed stratification of results across OSA severity levels as well as PLMSI subgroups (PLMSI ≥ 15, $n = 51$) can be found in Tables S2-S25 in the supplementary materials.

Additionally, each arousal scored by NBS2 was classified as either respiratory-related or not. An arousal was considered respiratory-related if a manually scored respiratory event (apnea or hypopnea) occurred within the 15 s preceding it. The proportion of respiratory-related arousals varied widely across the study population, ranging from 10 to 90% (median: 49%, central 95% range). This broad distribution underscores the algorithm's ability to detect arousals beyond those linked to respiratory events, effectively capturing patients with predominantly non-respiratory arousals as well as those with predominantly respiratory-related arousals.

Discussion

This study introduces an innovative algorithm that utilizes RIP signals, analyzed through a novel AI architecture, to classify sleep states and arousals. We have demonstrated that both sleep states and arousals can be derived from RIP signals, achieving performance comparable to that of human EEG-based scoring. ICC is commonly used to quantify variability in expert human scoring; for TST, reported ICC values are

0.82 [30] and 0.87 [31], and for ArI, values range from 0.41 [32] to 0.68 [33]. The current algorithm achieved an ICC of 0.91 for TST and 0.74 for ArI, indicating strong alignment with expert EEG-based scoring. Detecting sleep states and arousal events without relying on EEG represents a substantial advancement, with potential to improve the accuracy and accessibility of sleep disorder diagnoses via HST.

Previous studies on deriving sleep states from non-EEG signals have used various combinations of physiological data, including respiratory and cardiac signals [34], respiratory, cardiac, and movement signals [35], and respiratory and movement signals [11, 12]. Much research has also focused on peripheral cardiac function using photoplethysmography (PPG), often combined with peripheral arterial tonometry (PAT) and actigraphy [36–41]. RIP signals alone have also been successfully used to determine sleep states in infants [9, 10]. Our group previously utilized RIP and actigraphy to estimate sleep states [11]. Notably, NBS2 improves upon the earlier method while relying solely on RIP signals, achieving higher sensitivity and specificity for all three sleep states as well as reducing TST overestimation from 14 to 8 min.

Additionally, achieving automatic arousal detection from breathing signals represents a significant milestone, given the longstanding challenge of detecting arousal events from any PSG signal. However, recent advancements in deep learning methods have shown promise in overcoming this challenge [42]. Attempts have been made at arousal detection from non-EEG signals using ECG [43, 44], PPG combined with respiratory flow [45] and PAT [46]. A downside of using cardiogenic signals for sleep state and arousal detection is the potential for confounding effects due to cardiac conditions and medications [47, 48], which are common among patients with sleep disorders [49–51]. By focusing on RIP signals, our approach avoids some of these confounding factors, potentially leading to more accurate and reliable detection of sleep state and arousal events. This is particularly important for patient populations prone to sleep fragmentation, such as the elderly [52] and those with sleep-disordered breathing characterized by arousals rather than desaturations [3]. Future studies should investigate the broader implications of RIP-based sleep and arousal detection in these groups.

Nevertheless, this study is not without limitations. The design and quality of RIP belts is expected to impact the generalizability of our results [21]. Similarly, excessively filtered low bandwidth signals are likely to impact performance of the algorithm, potentially compromising the accuracy of sleep staging and arousal detection. Thus, the choice of hardware and the preprocessing of signals are critical considerations for the successful application of this method. Another limitation of the study is that the validation cohort might not fully express the diversity of patients with sleep disorders and various comorbidities.

Future studies will need to validate the algorithm's performance across more diverse populations to ensure its effectiveness and reliability across a wide range of patient groups. In addition, the current study did not assess the impact of NBS2 on diagnostic measures such as the AHI, which incorporates sleep state scoring through the TST and relies on precise arousal detection for scoring hypopneas that do not meet desaturation criteria.

In conclusion, this study makes a two-fold contribution. First, it demonstrates that information about sleep state and arousals can be accurately extracted from breathing signals. Second, it introduces a novel AI-driven algorithm for sleep state scoring and arousal detection using respiratory signals, which could greatly enhance the accuracy and accessibility of sleep disorder diagnoses through HST. Combined, these contributions represent a promising advancement in our understanding and evaluation of sleep disorders.

Declarations

Institutional review board statement The study was deemed exempt from institutional review board (IRB) review by the WCG IRB under 45 CFR § 46.104(d)(4), as the data involved did not allow for the direct identification or re-identification of human subjects.

Declaration of generative AI and AI-assisted technologies in the writing process During the preparation of this work the author(s) used OpenAI's ChatGPT 4o to suggest improvements in language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Competing interest Eysteinn Finnsson, Ernir Erlingsson, Hlynur D. Hlynsson, Vaka Valsdóttir, Thora B. Sigmarsdottir, Eydís Arnardóttir, Sigurður Æ. Jónsson, and Jón S. Ágústsson are all employees of Nox Medical.

References

1. American Academy of Sleep Medicine (2014) International classification of sleep disorders, 3. ed. American Acad. of Sleep Medicine, Darien, Ill

2. Berry RB, Quan SF, Abreu AR et al (2020) The AASM manual for the scoring of sleep and associated events; Rules, terminology and technical specifications, version 2.6. Am Acad Sleep Med

3. Malhotra RK (2021) Arousal-based scoring of obstructive hypopneas. Curr Opin Pulm Med 27:491–495. https://doi.org/10.1097/MCP.0000000000000820

4. Berry RB, Abreu AR, Krishnan V et al (2022) A transition to the American Academy of Sleep Medicine–recommended hypopnea definition in adults: initiatives of the Hypopnea Scoring Rule Task Force. J Clin Sleep Med 18:1419–1425. https://doi.org/10.5664/jcsm.9952

5. Phillipson EA (1978) Control of breathing during sleep. Am Rev Respir Dis 118:909–939. https://doi.org/10.1164/arrd.1978.118.5.909

6. Malik V, Smith D, Lee-Chiong T (2012) Respiratory physiology during sleep. Sleep Med Clin 7:497–505. https://doi.org/10.1016/j.jsmc.2012.06.011

7. Tusiewicz K, Moldofsky H, Bryan AC, Bryan MH (1977) Mechanics of the rib cage and diaphragm during sleep. J Appl Physiol Respir Environ Exerc Physiol 43:600–602. https://doi.org/10.1152/jappl.1977.43.4.600

8. Tabachnik E, Muller NL, Bryan AC, Levison H (1981) Changes in ventilation and chest wall mechanics during sleep in normal adolescents. J Appl Physiol Respir Environ Exerc Physiol 51:557–564. https://doi.org/10.1152/jappl.1981.51.3.557

9. Terrill PI, Wilson SJ, Suresh S et al (2010) Attractor structure discriminates sleep states: Recurrence plot analysis applied to infant breathing patterns. IEEE Trans Biomed Eng 57:1108–1116. https://doi.org/10.1109/TBME.2009.2038362

10. Terrill PI, Wilson SJ, Suresh S et al (2012) Application of recurrence quantification analysis to automatically estimate infant sleep states using a single channel of respiratory data. Med Biol Eng Comput 50:851–865. https://doi.org/10.1007/s11517-012-0918-4

11. Dietz-Terjung S, Martin AR, Finnsson E et al (2021) Proof of principle study: diagnostic accuracy of a novel algorithm for the estimation of sleep stages and disease severity in patients with sleep-disordered breathing based on actigraphy and respiratory inductance plethysmography. Sleep Breath 25:1945–1952. https://doi.org/10.1007/s11325-021-02316-0

12. Gaiduk M, Rodríguez JJP, Seepold R et al (2022) Estimation of sleep stages analyzing respiratory and movement signals. IEEE J Biomed Health Inform 26:505–514. https://doi.org/10.1109/JBHI.2021.3099295

13. Azarbarzin A, Ostrowski M, Hanly P, Younes M (2014) Relationship between arousal intensity and heart rate response to arousal. Sleep 37:645–653. https://doi.org/10.5665/sleep.3560

14. Amatoury J, Jordan AS, Toson B et al (2018) New insights into the timing and potential mechanisms of respiratory-induced cortical arousals in obstructive sleep apnea. Sleep 41:zsy160. https://doi.org/10.1093/sleep/zsy160

15. Eckert DJ, Malhotra A (2008) Pathophysiology of adult obstructive sleep apnea. Proc Am Thorac Soc 5:144–153. https://doi.org/10.1513/pats.200707-114MG

16. Carley DW, Applebaum R, Basner RC et al (1997) Respiratory and arousal responses to acoustic stimulation. Chest 112:1567–1571. https://doi.org/10.1378/chest.112.6.1567

17. Khoo MC, Shin JJ, Asyali MH et al (1998) Ventilatory dynamics of transient arousal in patients with obstructive sleep apnea. Respir Physiol 112:291–303. https://doi.org/10.1016/s0034-5687(98)00041-3

18. Horner RL, Rivera MP, Kozar LF, Phillipson EA (2001) The ventilatory response to arousal from sleep is not fully explained by differences in CO2 levels between sleep and wakefulness. J Physiol 534:881–890. https://doi.org/10.1111/j.1469-7793.2001.00881.x

19. Edwards BA, Connolly JG, Campana LM et al (2013) Acetazolamide attenuates the ventilatory response to arousal in patients with obstructive sleep apnea. Sleep 36:281–285. https://doi.org/10.5665/sleep.2390

20. Terrill PI, Edwards BA, Nemati S et al (2015) Quantifying the ventilatory control contribution to sleep apnoea using polysomnography. Eur Respir J 45:408–418. https://doi.org/10.1183/09031936.00062914

21. Montazeri K, Jonsson SA, Agustsson JS et al (2021) The design of RIP belts impacts the reliability and quality of the measured respiratory signals. Sleep Breath 25:1535–1541. https://doi.org/10.1007/s11325-020-02268-x

22. Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2016) Time series analysis: forecasting and control, 5th edn. John Wiley & Sons Inc, Hoboken, New Jersey

23. Bai S, Kolter JZ, Koltun V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling

24. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks

25. He T, Zhang Z, Zhang H et al (2019) Bag of tricks for image classification with convolutional neural networks

26. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition

27. U.S. Food and Drug Administration (2011) 510(k) Summary for MICHELE Sleep Scoring Systems (K112102)

28. U.S. Food and Drug Administration (2017) 510(k) Summary for EnsoSleep (K162627)

29. U.S. Food and Drug Administration (2021) 510(k) Summary for EnsoSleep (K210034)

30. Van Pee B, Massie F, Vits S et al (2022) A multicentric validation study of a novel home sleep apnea test based on peripheral arterial tonometry. Sleep 45:zsac028. https://doi.org/10.1093/sleep/zsac028

31. Kuna ST, Benca R, Kushida CA et al (2013) Agreement in computer-assisted manual scoring of polysomnograms across sleep centers. Sleep 36:583–589. https://doi.org/10.5665/sleep.2550

32. Pitkänen H, Nikkonen S, Rissanen M et al (2024) Multi-centre arousal scoring agreement in the sleep revolution. J Sleep Res 33:e14127. https://doi.org/10.1111/jsr.14127

33. Magalang UJ, Chen N-H, Cistulli PA et al (2013) Agreement in the scoring of respiratory events and sleep among international sleep centers. Sleep 36:591–596. https://doi.org/10.5665/sleep.2552

34. Bakker JP, Ross M, Vasko R et al (2021) Estimating sleep stages using cardiorespiratory signals: validation of a novel algorithm across a wide range of sleep-disordered breathing severity. J Clin Sleep Med 17:1343–1354. https://doi.org/10.5664/jcsm.9192

35. Gaiduk M, Penzel T, Ortega JA, Seepold R (2018) Automatic sleep stages classification using respiratory, heart rate and movement signals. Physiol Meas 39:124008. https://doi.org/10.1088/1361-6579/aaf5d4

36. Fonseca P, Weysen T, Goelema MS et al (2017) Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. Sleep 40:zsx097. https://doi.org/10.1093/sleep/zsx097

37. Ioachimescu OC, Allam JS, Samarghandi A et al (2020) Performance of peripheral arterial tonometry–based testing for the diagnosis of obstructive sleep apnea in a large sleep clinic cohort. J Clin Sleep Med 16:1663–1674. https://doi.org/10.5664/jcsm.8620

38. Walch O, Huang Y, Forger D, Goldstein C (2019) Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. Sleep 42:zsz180. https://doi.org/10.1093/sleep/zsz180

39. Wulterkens BM, Fonseca P, Hermans LW et al (2021) It is all in the wrist: Wearable sleep staging in a clinical population versus reference polysomnography. NSS 13:885–897. https://doi.org/10.2147/NSS.S306808

40. Li Q, Li Q, Cakmak AS et al (2021) Transfer learning from ECG to PPG for improved sleep staging from wrist-worn wearables. Physiol Meas 42:044004. https://doi.org/10.1088/1361-6579/abf1b0

41. Radha M, Fonseca P, Moreau A et al (2021) A deep transfer learning approach for wearable sleep stage classification with photoplethysmography. npj Digit Med 4:1–11. https://doi.org/10.1038/s41746-021-00510-8

42. Qian X, Qiu Y, He Q et al (2021) A review of methods for sleep arousal detection using polysomnographic signals. Brain Sci 11:1274. https://doi.org/10.3390/brainsci11101274

43. Olsen M, Schneider LD, Cheung J et al (2018) Automatic, electrocardiographic-based detection of autonomic arousals and their association with cortical arousals, leg movements, and respiratory events in sleep. Sleep 41:zsy006. https://doi.org/10.1093/sleep/zsy006

44. Li A, Chen S, Quan SF et al (2020) A deep learning-based algorithm for detection of cortical arousal during sleep. Sleep 43:zsaa120. https://doi.org/10.1093/sleep/zsaa120

45. Ross M, Fonseca P, Overeem S et al (2023) Autonomic arousal detection and cardio-respiratory sleep staging improve the accuracy of home sleep apnea tests. Front Physiol 14:. https://doi.org/10.3389/fphys.2023.1254679

46. Pillar G, Bar A, Shlitner A et al (2002) Autonomic arousal index: an automated detection based on peripheral arterial tonometry. Sleep 25:543–549. https://doi.org/10.1093/sleep/25.5.541

47. Kinoshita T, Yahaba M, Terada J et al (2018) Impact of arterial stiffness on WatchPAT variables in patients with obstructive sleep apnea. J Clin Sleep Med 14:319–325. https://doi.org/10.5664/jcsm.6970

48. Kinoshita T, Terada J, Koshikawa K et al (2019) Impact of antihypertensive drugs on Watch-PAT variables in patients with obstructive sleep apnea. Eur Respir J 54:. https://doi.org/10.1183/13993003.congress-2019.PA817

49. Yeghiazarians Y, Jneid H, Tietjens JR et al (2021) Obstructive sleep apnea and cardiovascular disease: A scientific statement from the American heart association. Circulation 144:e56–e67. https://doi.org/10.1161/CIR.0000000000000988

50. Grandner MA, Jackson NJ, Pak VM, Gehrman PR (2012) Sleep disturbance is associated with cardiovascular and metabolic disorders. J Sleep Res 21:427–433. https://doi.org/10.1111/j.1365-2869.2011.00990.x

51. Javaheri S, Redline S (2017) Insomnia and risk of cardiovascular disease. Chest 152:435–444. https://doi.org/10.1016/j.chest.2017.01.026

52. Bonnet MH, Arand DL (2007) EEG arousal norms by age. J Clin Sleep Med 3:271–274. https://doi.org/10.5664/jcsm.26796

53. Stephansen JB, Olesen AN, Olsen M et al (2018) Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. Nat Commun 9:5229. https://doi.org/10.1038/s41467-018-07229-3

Authors and Affiliations

Eysteinn Finnsson[1,2] · **Ernir Erlingsson[1,3]** · **Hlynur D. Hlynsson[1]** · **Vaka Valsdóttir[1]** · **Thora B. Sigmarsdottir[1]** · **Eydís Arnardóttir[1]** · **Scott A. Sands[4]** · **Sigurður Æ Jónsson[1]** · **Anna S. Islind[2]** · **Jón S. Ágústsson[1]**

✉ Eysteinn Finnsson
eysteinnf@noxmedical.com; eysteinn23@ru.is

Ernir Erlingsson
ernir@noxmedical.com; ere29@hi.is

Hlynur D. Hlynsson
hlynurd@gmail.com

Vaka Valsdóttir
vaka@noxmedical.com

Thora B. Sigmarsdottir
thora@noxmedical.com

Eydís Arnardóttir
eydis@noxmedical.com

Scott A. Sands
sasands@bwh.harvard.edu

Sigurður Æ Jónsson
sigurdur@noxmedical.com

Anna S. Islind
annasi@ru.is

Jón S. Ágústsson
jons@noxmedical.com

1   Nox Research, Nox Medical, Katrínartún 2, 105 Reykjavík, Iceland

2   Department of Computer Science, Reykjavik University, Reykjavik, Iceland

3   Department of Computer Science, University of Iceland, Reykjavik, Iceland

4   Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA