Machine Learning detection of arousals from at-home sleep EEG device demonstrates high reliability compared to experts

Corey B Puryear1, Franz Fürbass1, Jack Sampiere1, Jay Pathmanathan1, Antoine Guillot1, Alexander M. Chan1
[1] Beacon Biosignals, Boston, MA, USA

An at-home sleep EEG device with an associated machine learning algorithm can detect arousals with a high epoch-level agreement with expert scorers and provides a reliable estimate of arousal index in multiple populations with disrupted sleep

Introduction

- Cortical arousals are an abrupt shift to higher EEG frequencies during sleep that may or may not be followed by an awakening.
- Psychiatric conditions such as major depressive disorders are characterized by cortical hyperarousal during sleep, and some treatments can modulate arousal, making accurate arousal detection particularly relevant in these populations.
- Scoring arousals from in-lab polysomnography (PSG) by expert readers is the gold standard, but is burdensome, time-consuming, and subject to significant inter/intra-rater variability.
- This study compares the detection of arousals from PSG to an at-home, dry-electrode EEG device (Waveband) on individuals with sleep disorders with and without concurrent antidepressant use (which elevates cortical arousability).

Methods

Purpose: To evaluate ArousalML performance on subgroups with sleep disorders with and without concurrent antidepressant use

Population: 50 total subjects with confirmed disturbed sleep: 12 with sleep disorders taking SSRI antidepressants (SDAD) and 38 not taking antidepressants (SDCTRL)

Study Design:

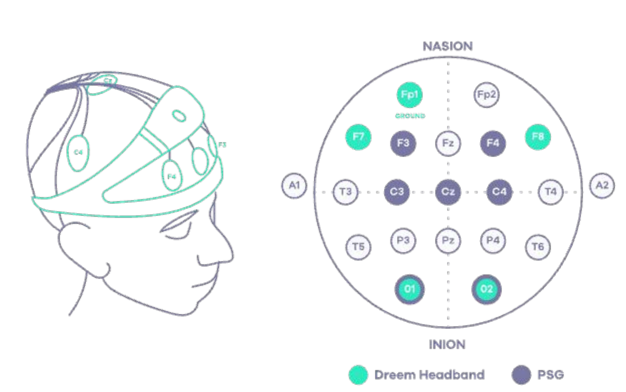
- One overnight in-lab polysomnography (PSG) while also wearing the Waveband
- Five Registered PSG Technologists (RPSGT) scored arousals in PSG only following AASM guidelines

Evaluations:

- Intraclass correlation coefficient (ICC) of the model's Arousal Index (ArI) to the scorers' average ArI
- Within-night agreement was evaluated by computing Cohen's Kappa between binarized 30-second epochs (with and without arousals) detected by the Waveband algorithm and a majority vote by human scorers (at least 3 out of 5).


Waveband

Nights were collected in-lab with polysomnography (PSG) and the at-home, dry-electrode EEG device (Waveband). Experts scored arousals on PSG. ArousalML detected arousals on Waveband. The figure shows an arousal in the signals of both devices.


NASION
INION
Dreem Headband
PSG

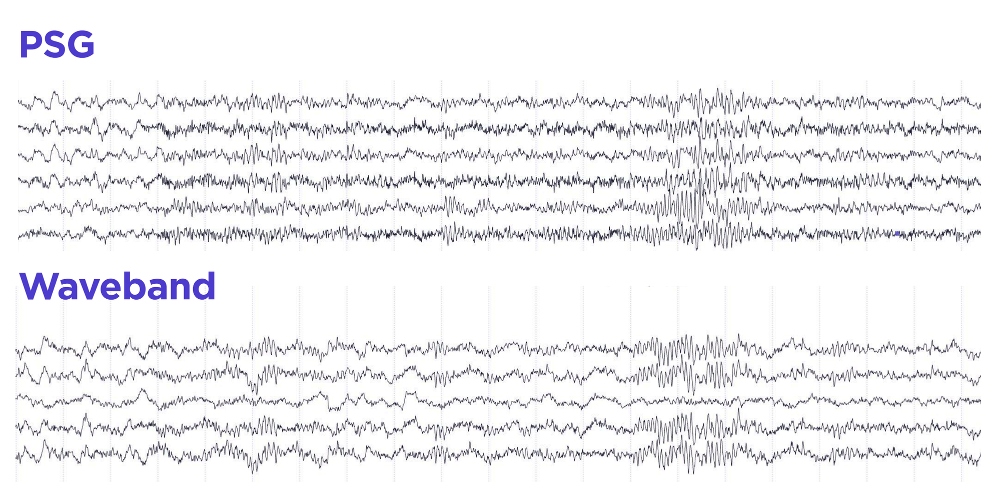
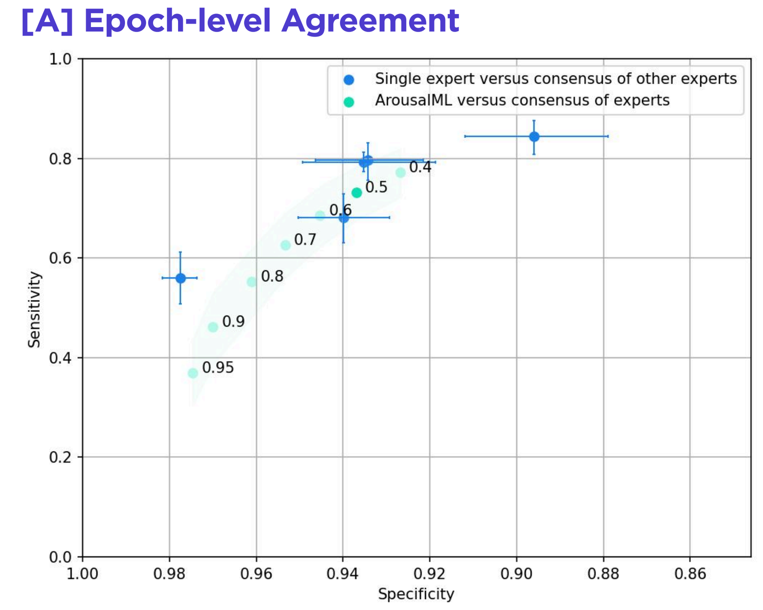

PSG
Waveband

Figure 1: Arousal on simultaneous Waveband and PSG recording

Results

[A] Epoch-level Agreement


Single expert versus consensus of other experts
ArousalML versus consensus of experts
Sensitivity
Specificity

[B] Arousal Index Agreement

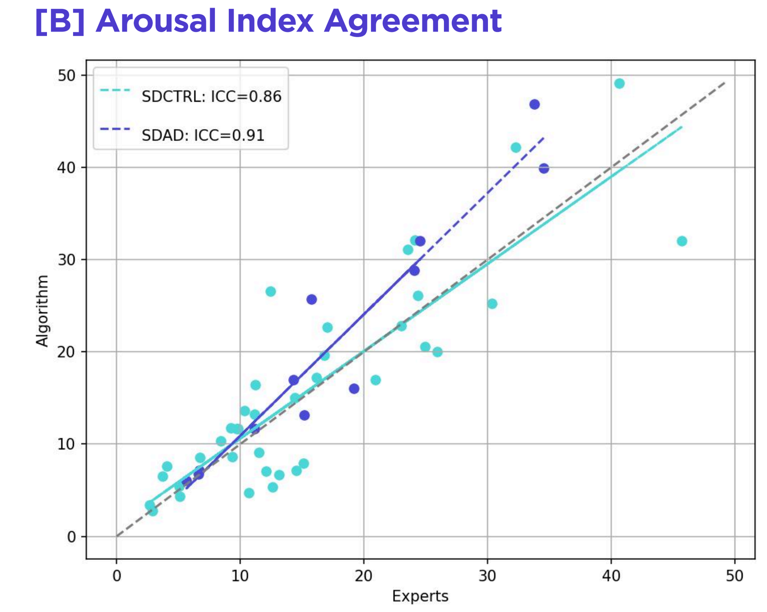
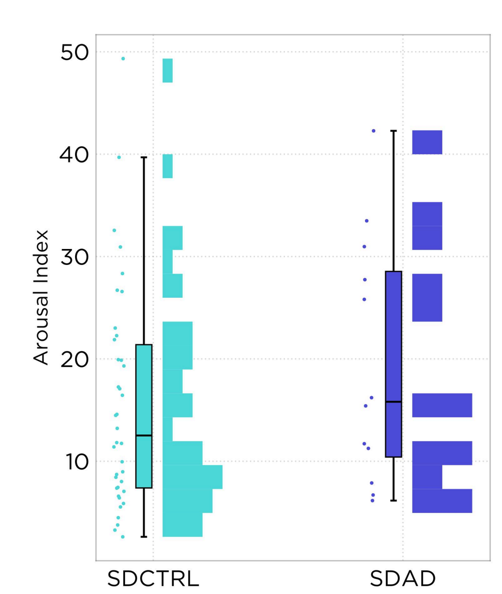

SDCTRL: ICC=0.86
SDAD: ICC=0.91
Algorithm
Experts

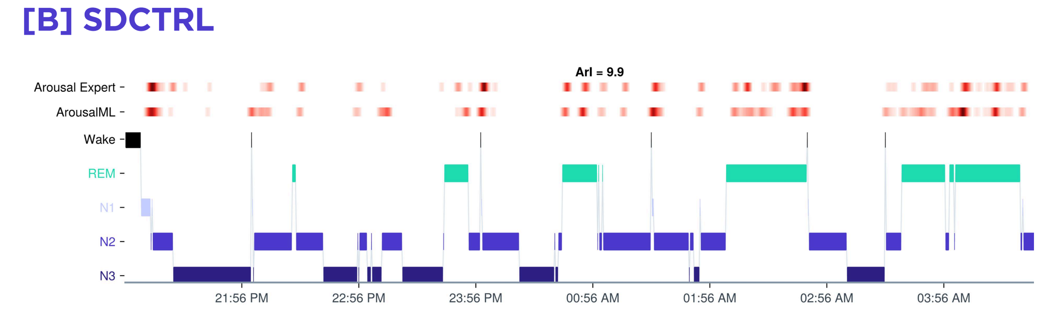
Figure 2: ArousalML showed performance in-line with human experts

[A] The at-home EEG device's epoch-level kappa of 0.57 (95% CI: 0.54-0.60) and the pairwise agreement of most individual scorers (0.45 to 0.58, evaluated relative to other scorers' majority vote).
[B] Comparing the at-home EEG device ArI to the average ArI of all scorers resulted in an ICC of 0.87 (95% CI: 0.80-0.91) overall, 0.91 (95% CI: 0.81-0.97) for SDAD and 0.86 (0.78 - 0.91) for SDCTRL.

[A] Arousal Index for groups


Arousal Index
SDCTRL
SDAD

[B] SDCTRL


ArI = 9.9
Arousal Expert
ArousalML
Wake
REM
N1
N2
N3
21:56 PM
22:56 PM
23:56 PM
00:56 AM
01:56 AM
02:56 AM
03:56 AM

[C] SDAD

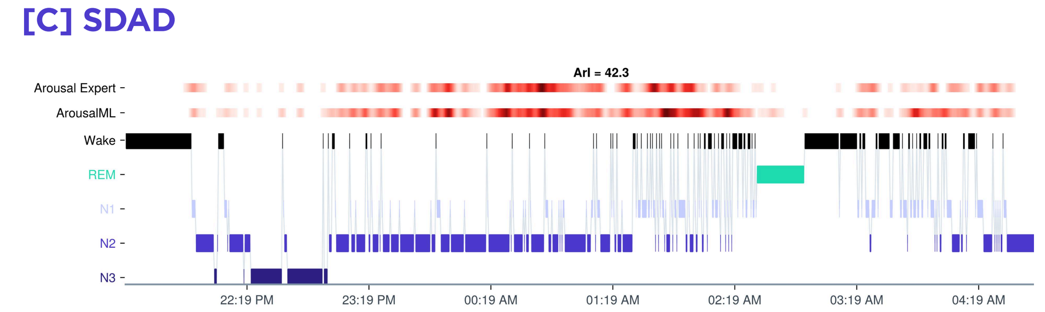

ArI = 42.3
Arousal Expert
ArousalML
Wake
REM
N1
N2
N3
22:19 PM
23:19 PM
00:19 AM
01:19 AM
02:19 AM
03:19 AM
04:19 AM

Figure 3: Waveband derived endpoints and example hypnogram with arousal interaction

ArI is higher in groups taking SSRIs (SDAD) than controls (SDCTRL) consistent with physiology and treatment interaction. [A] The distribution of arousal index in groups taking SSRIs and controls. Example hypnograms with arousals densities of the best performing expert (kappa=0.58 to other experts) and ArousalML shown in red for a night of a control participant [B] and a night of a participant taking SSRIs [C].

Summary of Results

- ArousalML demonstrated higher average agreement to human consensus than individual human experts.
 - Pairs of human scorers: 0.74 ICC (SD=0.092, range: 0.61-0.85)
 - ArI and a human scorer: 0.78 ICC (SD=0.043, range: 0.72-0.84)
 - ArI and human consensus: 0.87 ICC (95% CI: 0.80-0.91)
- ArousalML had high agreement with human scorers in sleep disorders with (SDAD) and without (SDCTRL) concurrent antidepressant use.
 - ArI and human scorer, SDCTRL: 0.77 ICC (SD=0.044, range: 0.70-0.81)
 - ArI and human scorer, SDAD: 0.83 ICC (SD=0.078, range: 0.69-0.91)
- Arousal ML had high agreement with human consensus in SDCTRL and SDAD.
 - ArI and human consensus, SDCTRL: 0.86 ICC (95% CI: 0.78 - 0.91)
 - ArI and human consensus, SDAD: 0.91 (95% CI: 0.81-0.97)
- ArousalML had better agreement with human consensus than most individual human experts.
 - ArI epoch-level kappa with human scorers: 0.57 (95% CI: 0.54-0.60)
 - Human scorers epoch-level kappa range compared to the rest of the human scorers: (0.45 to 0.58)

Conclusions

→ **An at-home sleep EEG device with an associated machine learning algorithm can detect arousals with a high epoch-level agreement with expert scorers**

→ **ArousalML-based arousal index is as good or better than individual human expert-derived arousal index**

→ **These findings demonstrate the potential for scalable, objective quantification of sleep fragmentation in real-world settings, enabling more efficient evaluation of therapies and future diagnostic applications**

Contact

Corey B. Puryear, Ph.D.
corey.puryear@beacon.bio
Director, Translational Medicine
Beacon Biosignals

Disclosure: All poster authors are employees of Beacon Biosignals.

BEACONBIOSIGNALS