AI POWERED AROUSAL SCORING FOR POLYSOMNOGRAPHY AND SELF-APPLIED-SOMNOGRAPHY SLEEP STUDIES

S.Æ. Jónsson, E. Finnsson, K. Montazeri, E. Erlingsson, H.Z. Zahawi, Þ.B. Sigmarsdóttir, E. Arnardóttir, J.S. Ágústsson

Nox Research, Nox Medical ehf, Reykjavík, Iceland,

Presented at ESRS 2022, Athens

thora@noxmedical.com

Introduction

Cortical arousals are brief changes in the EEGs recorded during a sleep study. They are defined as the EEGs having an abrupt change in frequency lasting for at least 3 seconds, preceded by at least 10 seconds of stable sleep.

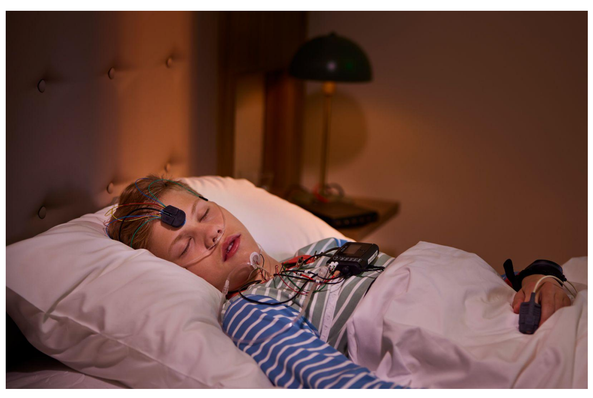
Arousal scoring is a challenge for human experts as seen by low interscorer-agreement between humans.[1, 2] The low agreement makes it difficult for AI analyses to learn arousal scoring.

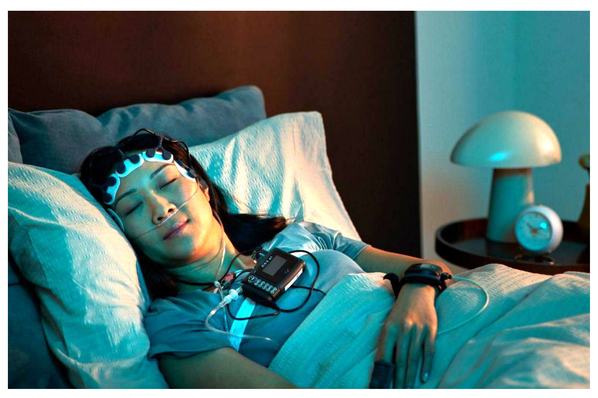
We present an end-to-end deep learning approach to robustly score arousals from PSG and Self-Applied-Somnography (SAS) sleep studies which use a reduced frontal EEG montage. The SAS is a type 2 study utilizing a frontal EEG montage and no chin EMG. The performance of the AI arousal scoring analyses is on-par with human level agreement. The AI analyses place the scored arousals correctly when the arousals occur, allowing the automatically scored arousals to be used for improving hypopnea scoring and to associate the scored arousal with preceding respiratory or periodic limb movement (PLM) events. The sensitivity of the associations are limited by the sensitivity of the arousal scoring.

Methods

The PSG and SAS analyses make predictions based on raw EEG, EOG and EMG signals, thereby avoiding manual feature extraction. This allows shorter prediction times and the analyses can learn more complex relations as the size of its training data increases. The analyses outputs are sequences of arousal probabilities which are used to generate discrete arousal events in a post-processing step.

The analyses were trained on over 1800 and 900 manually scored PSG and SAS studies, respectively. They were validated on two previously unseen datasets. A set of PSG studies (N = 151, epochs = 119,774) and SAS sleep studies (N = 88, epochs = 70,349).





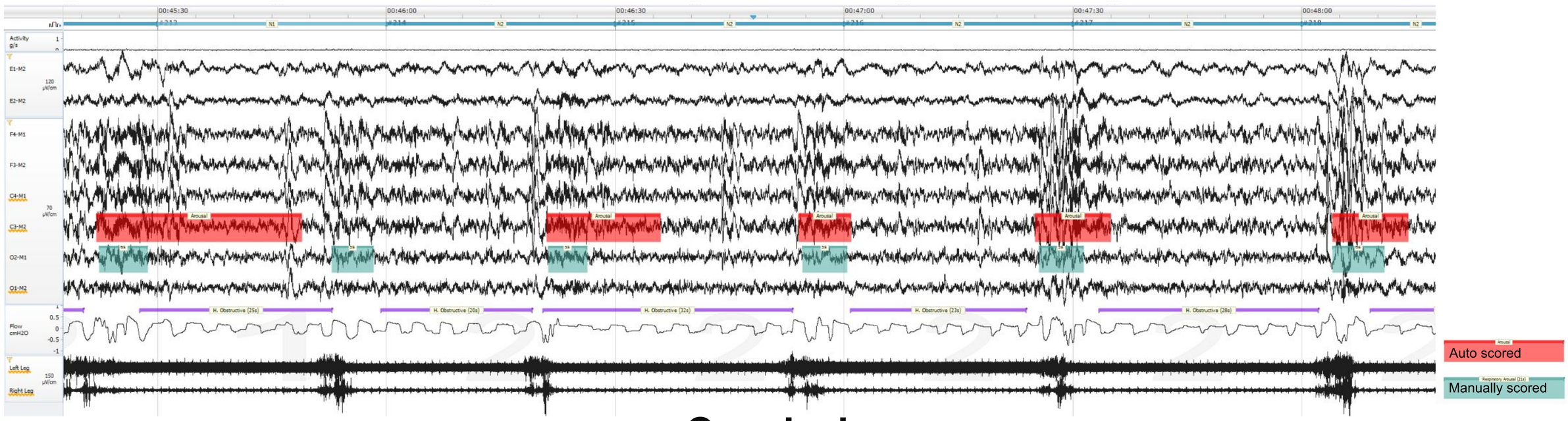
Results

The sensitivity, specificity, and accuracy was calculated on an epoch level for the PSG and SAS.

	PSG	SAS
Sensitivity % (95% CI)	67% (65% – 70%)	68% (66% – 71%)
Specificity % (95% CI)	91% (90% – 92%)	94% (93% – 95%)
Accuracy % (95% CI)	86% (85% – 87%)	91% (89% – 92%)

The downstream impact of the arousal scoring was investigated with regards to AHI classification, hypopnea detection, and event association. The performance of the downstream analyses is limited by the sensitivity of the arousal scoring.

PSG	Sensitivity (95% CI)	Specificity (95% CI)	OPA (95% CI)
AHI ≥ 5 Patients	96% (92% - 99%)	95% (86% - 100%)	96% (92% - 99%)
AHI ≥ 15 Patients	96% (90% - 100%)	96% (91% - 99%)	96% (92% - 98%)
Hypopnea events	89% (87% - 91%)	98% (97% - 98%)	96% (96% - 97%)
Respiratory Associated Arousal	65% (61% - 68%)	97% (97% - 98%)	95% (94% - 95%)
PLM Associated Arousal	65% (58% - 72%)	100% (100%-100%)	99% (99% - 100%)
Unassociated Arousal	63% (60% - 66%)	94% (93% - 95%)	90% (89% - 91%)



Conclusions

- The AI analyses perform on-par with human interscorer reliability for PSG and SAS sleep studies, with high epoch level agreement.
- The analyses place the arousal markers accurately over the EEG signals.
- The correct start of the arousal markers is important for hypopnea scoring.
- The agreement in hypopnea event scoring is high, with 89% sensitivity and 98% specificity, after the automatic arousal analyses are used, compared to manual scoring.
- The association of the scored arousals to preceding respiratory or PLM events is high and limited by the positive agreement in scoring the events.

1. C Fernandez, S Rusk, N Glattard, B Hensen, M Shokoueinejad, S Creado, J Hungerford, 0320 A Cross-Validation Approach to Inter-scorer Reliability Assessment, *Sleep*, Volume 41, Issue suppl_1, April 2018, Pages A122–A123, https://doi.org/10.1093/sleep/zsy061.319,
2. Ruehland WR, Rochford PD, Pierce RJ, Singh P, Thornton AT. External proficiency testing improves inter-scorer reliability of polysomnography scoring. Sleep & Breathing = Schlaf & Atmung. 2022 Jul. DOI: 10.1007/s11325-022-02673-4