Predicting Sleep Arousals in Home Sleep Apnea Tests Using Artificial Intelligence

E. Erlingsson, H. Zagross Zahawi, S.Æ. Jónsson, E. Arnardóttir, E. Finnsson, K. Montazeri, Þ.B. Sigmarsdóttir, H.D. Hlynsson, M.E. McPhee, J.S. Ágústsson

Nox Research, Nox Medical ehf, Reykjavík, Iceland

Presented at Sleep 2023, Indianapolis USA

thora@noxmedical.com

Introduction

As the popularity of home-sleep-apnea-tests (HSAT) grows, so too does the importance of ensuring that they provide the best information possible to facilitate patient diagnosis and treatment. One major challenge in this regard is how to estimate a patient's Apnea-Hypopnea-Index (AHI) when no electroencephalography (EEG) is available.

Until now, an EEG has been considered necessary to detect arousals, which can influence hypopnea scoring. Thus scoring hypopneas without information on arousals can lead to lower AHI for HSAT compared to polysomnography (PSG), potentially resulting in misdiagnosis. To address this issue, we have developed an effective AI model tailored to HSAT, that can predict arousals using only two non-EEG signal groups.

Methods

We developed a deep learning model to predict arousals, using only respiratory inductance plethysmography (RIP) and activity signal groups. The model performs a prediction for each second recorded and aggregates those results to score arousal events. To train and validate the model, we used 2216 manually scored PSG sleep recordings from various sleep centers in five countries.

The model's robustness and accuracy was tested, using recordings from a separate sleep center in the US that was not included in training or validation (Table 1). Additionally, we ensured that the recordings covered all categorical severities of sleep apneas; i.e., normal, mild, moderate, and severe.

Table 1: Demographic information for the US-based sample used to test the performance of the model (N = 101 patients)

	Mean	SD	Minimum	Maximum
Age (years)	37.8	21.0	4	82
Height (cm)	166.3	16.5	102.9	190.5
Weight (kg)	82.4	28.3	17.2	166.9
BMI (kg/m²)	29.3	9.1	14.4	63.5
AHI (events/hour)	21.1	21.1	0.2	106.5
Gender (N)	Male: 40; Female: 58; Unknown: 3			

Results

Compared with manual arousal scoring, using epoch-level agreement, the model exhibited good sensitivity, specificity, and accuracy (Table 2). Moreover, detecting hypopneas using model-predicted arousals gave a 24% improvement in sensitivity compared with not using arousals.

Table 2: Accuracy of arousal detection and hypopnea detection using the automatically predicted arousals

	Sensitivity	Specificity	Class Accuracy
Arousals (Epoch-level)	**62%**	**86%**	**81%**
Hypopneas (Epoch-level)	**86%**	**96%**	**94%**

We investigated differences in AHI when using the model's arousals compared to not using arousals. For AHI>=5, accuracy was excellent with arousals, but only fair to good without. Similarly, for AHI>=15, results were excellent with arousals, but only fair without (Table 3).

Table 3: Accuracy of AHI classification with or without hypopneas detected using automatically predicted arousals

	Sensitivity	Specificity	Class Accuracy
AHI scored on HSAT with automated arousal detection vs manual scoring			
AHI ≥ 5 (Patient-level)	**95%**	**100%**	**96%**
AHI ≥ 15 (Patient-level)	**95%**	**100%**	**96%**
AHI (REI) scored on HSAT without arousals vs manual scoring			
AHI ≥ 5 (Patient-level)	**68%**	**100%**	**75%**
AHI ≥ 15 (Patient-level)	**54%**	**100%**	**80%**

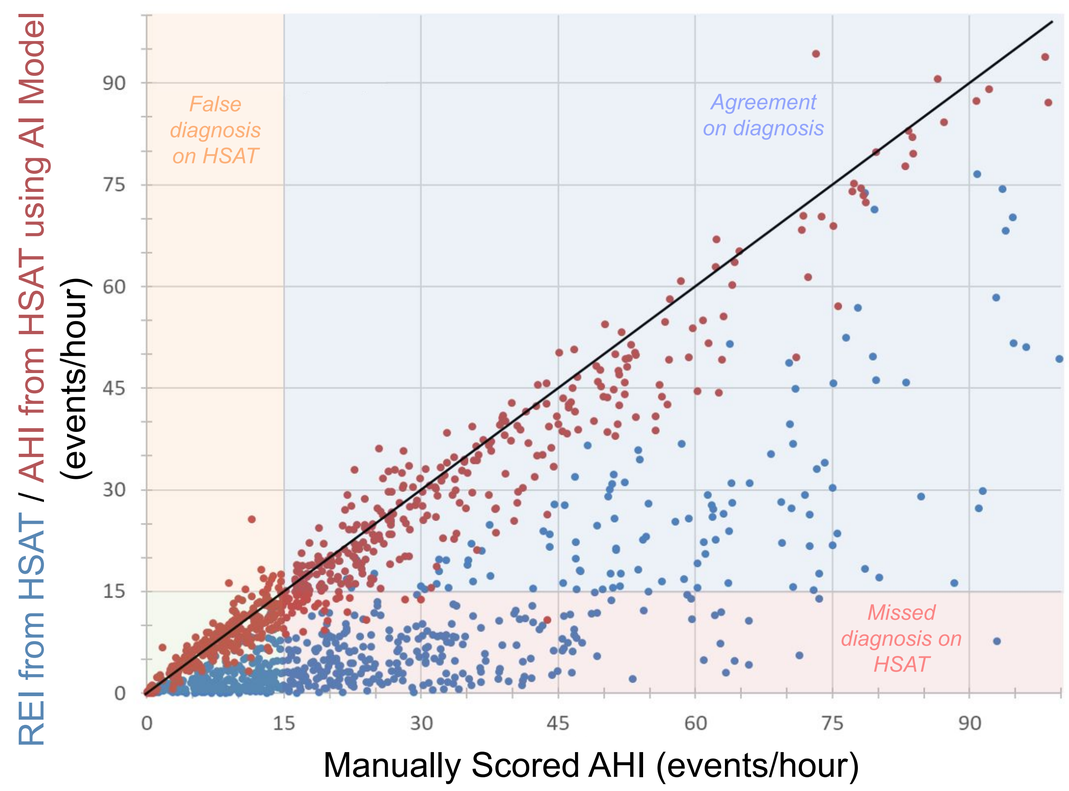

Figure 1: Visual comparison of the Respiratory Event Index (REI, blue) scored from HSAT without arousal input compared to AHI from the present AI model using automated arousal detection (red) with reference to manually scored AHI (N = 643 patients)

Conclusions

Our arousal detection model performed well, suggesting that a HSAT may be sufficient in order to predict arousals effectively. Additionally, our findings imply that using the predicted arousals can improve the scoring of hypopneas and reduce missed diagnoses based on AHI.