Improving Diagnostic Accuracy in Sleep Apnea Through Respiratory Based Home Sleep Tests

Þ. Sigmarsdóttir, V. Valsdóttir, S. Jónsson, E. Erlingsson, H. Hlynsson, K. Montazeri, E. Finnsson, E. Arnardóttir, D. Wilcox, G. Árnadóttir, H. Katrínardóttir, M. McPhee, R. Hugosdóttir, S. Helgason, J.S. Ágústsson

Nox Research, Nox Medical ehf, Reykjavík, Iceland

Presented at SLEEP 2025, Seattle, Washington, USA

thora@noxmedical.com

Introduction

Nox AI Scoring (FDA cleared software medical device DeepResp K241960) is a novel AI-enabled medical device designed to overcome the limitations of home sleep testing (HST) sleep recordings, which traditionally lack the signals required to score sleep stages, arousals, and thus hypopneas ending in arousals. These limitations can lead to systematic underestimation of key metrics, such as the apnea-hypopnea index (AHI), increasing the risk of missed sleep apnea diagnosis, especially in patients with poor sleep efficiency or a high proportion of hypopneas ending in arousals. Using respiratory inductance plethysmography (RIP) signals, the AI medical device estimates sleep states (Wake, REM, and NREM), arousals and respiratory events. When combined with the oximetry signal, it also enables estimation of oxygen desaturations. This novel method aims to give more accurate AHI values from HST recordings that more closely reflect the values obtained from polysomnography recordings (PSGs).

The aim of this study was to evaluate whether DeepRESP can enhance HST recordings by using RIP signals to estimate sleep stages and arousals, enabling more accurate AHI values comparable to PSGs

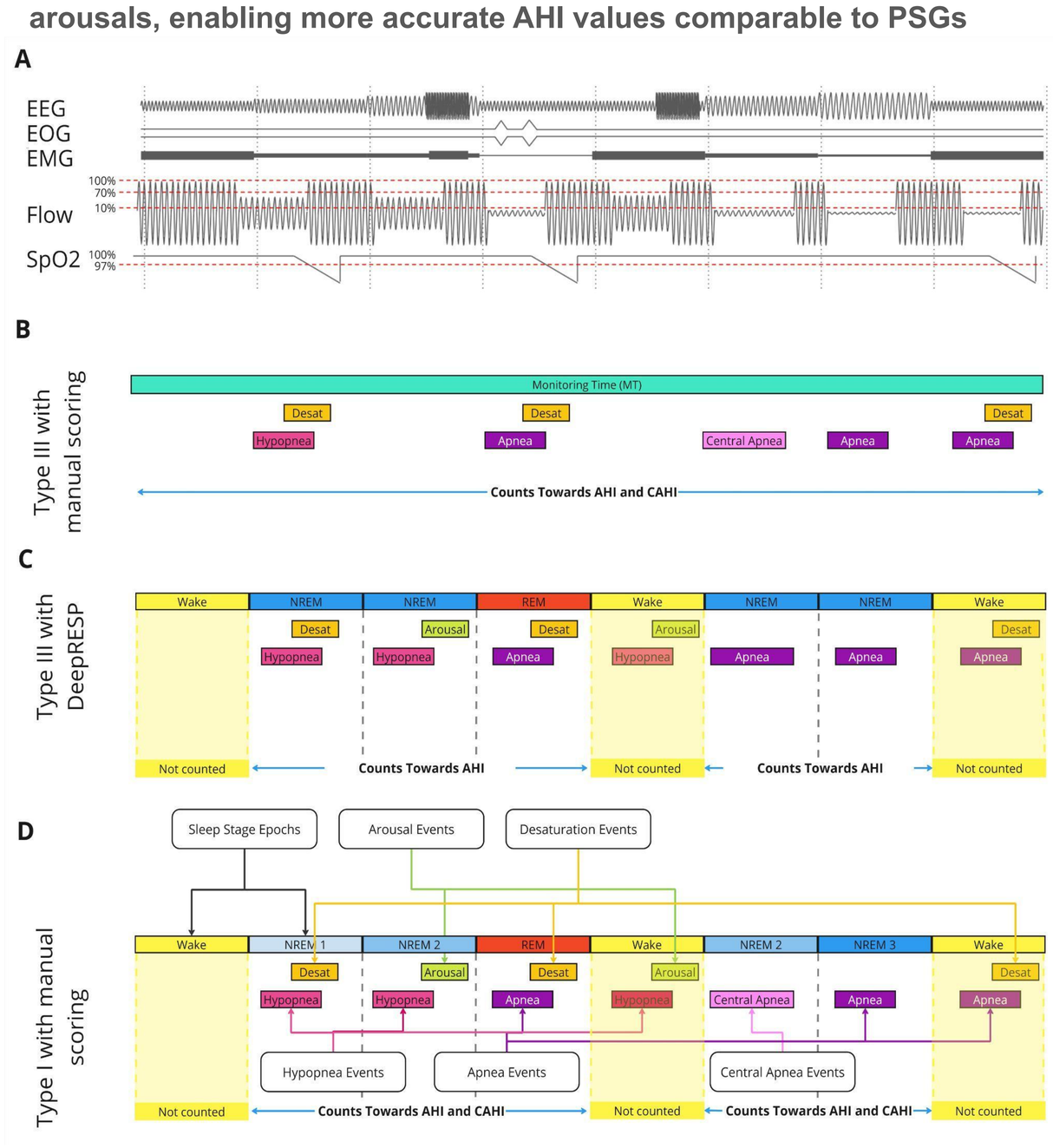

Figure 1: By using RIP signals, DeepRESP enables estimation of sleep states, arousal, and respiratory events from HST recordings. By the addition of oximetry signal it also enables estimations of oxygen desaturation.

Methods

The primary objective was to evaluate the agreement in sleep apnea severity classification between the AI-enabled medical device and manual scoring of both in-lab and ambulatory PSG recordings. Agreement was assessed at two AHI thresholds: AHI ≥ 5 and AHI ≥ 15. Secondary objectives included assessing epoch-level agreement between the AI medical device and manual scoring for Sleep states (Wake, NREM, REM), Respiratory events, and Arousals. Sensitivity, Specificity, and Accuracy were calculated using confusion matrices, with 95% confidence intervals (CI) estimated using bootstrapping.

The results were compared to the performance of HST scoring as a reference.

Results

The AI-enabled medical device demonstrated high overall agreement across critical metrics compared to manual scoring. For sleep apnea severity classifications, OPA was 92.5% for AHI ≥ 5, and 84.7% for AHI ≥ 15. Sleep state classification also showed high agreement, with an OPA of 92.7% for Wake, 89.2% for NREM, and 95.4% for REM. The OPA for arousal scoring was 81.8%, and for respiratory event scoring was 83.7%.

Table 1: Sensitivity, specificity and overall accuracy of severity classificationl agreement between DeepRESP vs Reference HST compared to manual scoring

	Sensitivity [95%CI]	Specificity [95%CI]	Accuracy [95%CI]
	AHI ≥ 5		
DeepRESP	93.1 [92.2, 93.9]	81.1 [75.1, 86.6]	92.5 [91.7, 93.3]
HST Reference	82.3	56.5	81.0
	AHI ≥ 15		
DeepRESP	82.1 [80.6, 83.5]	92.3 [90.5, 94.0]	84.7 [83.5, 85.8]
HST Reference	60.4	89.3	67.9

Table 2: Sensitivity, specificity and overall accuracy of sleep state classification agreement between DeepRESP compared to manual scoring

	Sensitivity [95%CI]	Specificity [95%CI]	Accuracy [95%CI]
	DeepRESP		
Wake	76.2 [75.5, 77.0]	96.8 [96.6, 97.0]	92.7 [92.5, 92.9]
NREM	94.5 [94.2, 94.7]	79.0 [78.4, 79.6]	89.2 [88.9, 89.4]
REM	79.1 [78.2, 79.9]	98.1 [90.0, 98.2]	95.4 [95.2, 95.5]

Table 3: Sensitivity, specificity and overall accuracy of respiratory event scoring between DeepRESP vs HST reference compared to manual scoring*,

	Sensitivity [95%CI]	Specificity [95%CI]	Accuracy [95%CI]
	DeepRESP		
Arousals	66.8 [65.8, 67.6]	86.8 [86.4, 87.1]	81.8 [81.5, 82.0]
Respiratory events	75.4 [74.6, 76.1]	87.8 [87.4, 88.1]	83.7 [83.4, 84.0]
	HST reference		
Respiratory events	58.5	95.4	81.6

**only Type I recordings were used for arousal scoring, and no comparison to the reference was made as standard HST scoring does not score arousals*

Table 4: Patient demographics.

Group	Subgroup	Total	Percentages
Sex	Female	1,138	32.6%
	Male	2,337	67%
Age group	22-35	714	20%
	36-45	740	21%
	46-55	856	25%
	56-65	677	19%
	>65	501	14%
BMI	<25	598	17%
	25-29	923	26%
	≥30	1,959	56%
AHI	Normal (<5)	175	5%
	Mild (5-14)	725	21%
	Moderate (15-29)	993	28%
	Severe (≥30)	1,595	46%

Conclusions

The results show that the novel AI-enabled medical device demonstrated high agreement with in-lab and ambulatory PSG recordings for estimating critical sleep parameters and clinical indices from HST recordings, improving the accuracy of OSA severity classification compared to reference methods. These findings suggest substantial potential to enhance diagnostic accuracy of home sleep testing.