

Performance of RIP signal-based AI in patients with PLMS: subgroup analyses from two datasets

J.S. Ágústsson¹, M. McPhee¹, R. Hugósdóttir¹, S. Jónsson¹, E. Arnardóttir¹, E. Erlingsson¹, K. Montazeri¹, G. Árnadóttir¹, D. Wilcox¹, P. Sigmarsdóttir¹, E. Finnsson¹, D. Rye², H. Riney³

- 1. Nox Research, Nox medical ehf, Reykjavík, Iceland
- 2. Emory University, Atlanta, United States
- 3. Nox Research, Nox Health Inc., Alpharetta, GA, United States.

Presented at SLEEP Europe 2026, Maastricht, The Netherlands



thora@noxmedical.com

Introduction

Periodic limb movements during sleep (PLMS) are associated with cortical and autonomic arousals and can fragment sleep. These features may mimic sleep-disordered breathing and confound AI analysis of sleep in home sleep apnea testing (HSAT) without EEG. RIP- based AI analysis may extend the diagnostic utility of HST by estimating sleep states, arousals, and apnoea–hypopnoea index (AHI). **Subgroups with low versus elevated PLMS-index (PLMSI) were assessed to determine whether PLMs impact analysis accuracy.**

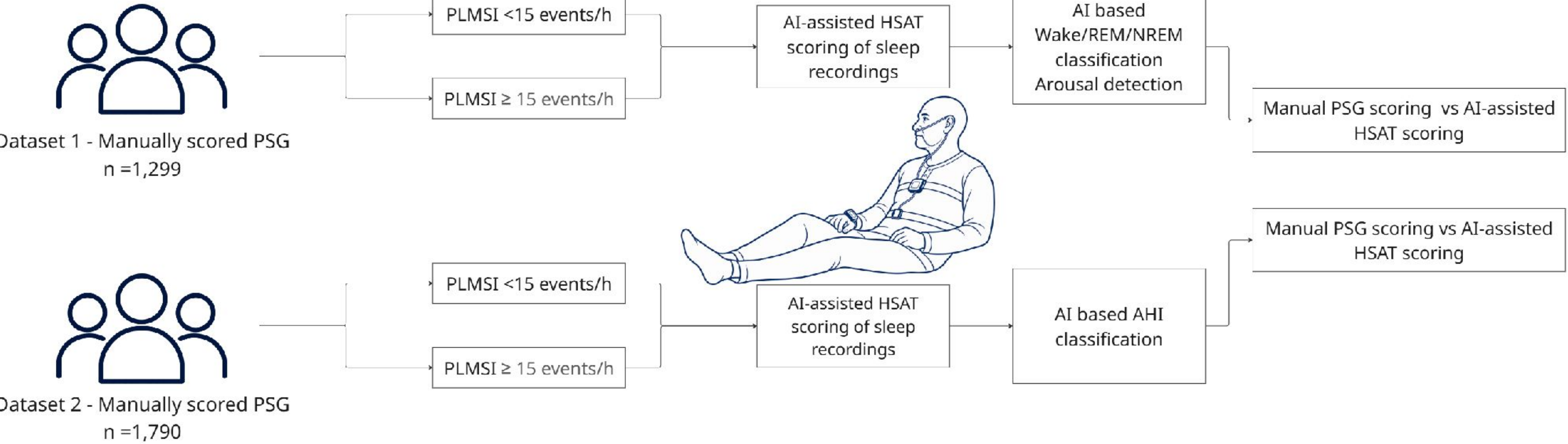


Figure 1: Study design and analysis workflow. Two retrospective datasets of treatment-seeking adults were stratified by periodic limb movement index during sleep (PLMSI <15 or ≥15 events/h). AI-derived results, obtained using the reduced channel configuration of HSAT studies, were compared with manually scored polysomnography (PSG) within each subgroup. Dataset 1 (n = 1,299) assessed epoch-level wake, REM and NREM classification and arousal detection; dataset 2 (n = 1,790) assessed apnoea–hypopnoea index (AHI) classification. AI, artificial intelligence; HSAT, home sleep apnoea testing; REM, rapid eye movement; NREM, non-rapid eye movement.

Results

Among patients with PLMSI<15 (n=1,248) in dataset-1, sensitivity, specificity and accuracy were 78.1%, 96.3%, 92.5% for wake, 93.9%, 80.7%, 89.4% for NREM, 80.9%, 98.2%, 95.9% for REM, and 66.0%, 87.0%, 81.9% for arousal detection. In patients with PLMSI≥15 (n=51), corresponding values were 73.6%, 95.4%, 90.1% for wake, 93.9%, 74.3%, 86.9% for NREM, 68.6%, 98.6%, 95.2% for REM, and 67.1%, 80.6%, 77.2% for arousal detection. In dataset-2, accuracy in sleep apnea severity classification at an AHI threshold of 5 events/hour, was 86.9% in the PLMSI<15 subgroup (n=1,176) and 91.5% in the PLMSI≥15 subgroup (n=614). For AHI≥15 events/hour, accuracy was 78.2% and 78.0%, respectively.

PLMSI < 15 events/hr (n=1248)				
	Wake	NREM	REM	# Epochs
Wake	78%	21%	1%	206,490
NREM	4%	94%	2%	658,446
REM	2%	17%	81%	131,814
PLMSI ≥ 15 events/hr (n=51)				
	Wake	NREM	REM	# Epochs
Wake	74%	24%	2%	9,688
NREM	5%	94%	1%	25,655
REM	3%	29%	68%	4,413

Table 2: Normalized epoch-level confusion matrix for algorithm-classified versus manually classified sleep states in the subset of patients with PLMSI < 15 events/hr (n=1248), and PLMSI ≥ 15 events/hr (n=51). True labels are presented in rows, predicted labels in columns, and the total number of scored epochs is displayed in the far-right column.

1A	Sensitivity % [95%CI]	Specificity % [95%CI]	Accuracy % [95%CI]
Dataset 1, PLMSI < 15			
Wake	78.1 [77.0, 79.2]	96.3 [95.9, 96.6]	92.5 [92.2, 92.8]
NREM	93.9 [93.5, 94.3]	80.7 [79.9, 81.5]	89.4 [89.1, 89.7]
REM	80.9 [79.8, 81.9]	98.2 [98.1, 98.3]	95.9 [95.7, 96.1]
Arousals	66.0 [64.8, 67.2]	87.0 [86.5, 87.4]	81.9 [81.5, 82.3]
Dataset 1, PLMSI ≥ 15			
Wake	73.6 [67.0, 79.6]	95.4 [93.4, 96.9]	90.1 [87.8, 92.1]
NREM	93.9 [91.7, 95.6]	74.3 [69.3, 78.9]	86.9 [84.7, 88.9]
REM	68.6 [58.6, 77.1]	98.6 [97.8, 99.2]	95.2 [93.8, 96.4]
Arousals	67.1 [61.6, 72.2]	80.6 [77.3, 83.5]	77.2 [75.2, 79.2]
1B	Sensitivity % [95%CI]	Specificity % [95%CI]	Accuracy % [95%CI]
Dataset 2, PLMSI < 15			
AHI ≥ 5	87.3 [85.3, 89.1]	77.8 [65.1, 89.1]	86.9 [84.9, 88.8]
AHI ≥ 15	72.0 [68.8, 74.9]	94.8 [92.1, 97.2]	78.2 [75.9, 80.5]
Dataset 2, PLMSI ≥ 15			
AHI ≥ 5	92.8 [90.7, 94.9]	67.7 [47.8, 83.9]	91.5 [89.4, 93.6]
AHI ≥ 15	73.3 [68.9, 77.6]	90.6 [85.6, 94.7]	78.0 [74.6, 81.3]

Table 1A: Sensitivity, specificity and accuracy of Nox BodySleep 2.0, vs. manual PSG scoring, with regards to classifying Wake, NREM, REM and detecting arousals in dataset 1 for PLMSI <5 (n=1,248), and PLMSI ≥15 (n=51). Table 1B: Sensitivity, specificity and accuracy of DeepRESP, vs. manual PSG scoring, with regards to OSA severity classification in dataset 2 for PLMSI <5 (n=1,176), and PLMSI ≥15 (n=614).

Conclusions

RIP-based AI showed generally high sleep staging accuracy and comparable AHI classification performance across PLMSI groups. In patients with PLMSI ≥ 15 events/hour, REM sensitivity was lower, as were NREM specificity and arousal specificity. These findings support the utility of RIP-based AI in patients with elevated PLMSI, while highlighting REM and arousal detection as areas for further evaluation. Results for REM and arousal scoring should be interpreted in light of the small PLMSI ≥ 15 events/hour subgroup in dataset 1 (n = 51) and possible differences in patient characteristics between groups.

Disclaimer

This work was conducted by Nox Health. The analyses are exploratory, post hoc subgroup analyses of previously published dataset and are not intended to represent official device performance, support promotional claims, or modify labelling.