Preliminary validation results: Improving AHI scoring accuracy using an AI model for Sleep state and arousal classification from Home Sleep Apnea Testing

A. Árnason[2*] | A.K. Sigurðarsont[2*] | E. Leonardsson[2*] | H.Þ. Hákonardóttir[2*] | H. Finnbogadóttir[2*] | L. Kristjánsdóttir[2*] | L.Ó. Andrésdóttir[2*] | M. Árnadóttir[2*] | S. Sigurðarson[2*] | T.O. Pétursson[2*] | Þ.B. Sigmarsdóttir[1] | V. Valsdóttir[1] | E. Erlingsson[1] | C. M. Texeira[1] | A. S. Islind[2] | J.S. Ágústsson[1] | E. S. Arnardóttir[2],

1. Nox Research, Nox Medical ehf, Reykjavík, Iceland

2. School of Technology, Department of Computer Science, Reykjavik University, Reykjavik, Iceland

** Equal Contribution*

Presented at **World Sleep Congress 2023**, Rio de Janeiro - Brazil

thora@noxmedical.com

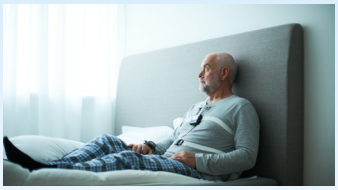
Introduction

The use of in-lab polysomnography (PSG) as a standard of care method to diagnose obstructive sleep apnea (OSA) severity can be costly, time-consuming and limited by the number of beds in each clinic. Home sleep apnea testing (HSAT) provides a lower-cost alternative to in-lab PSG, but due to its limited information on sleep stages and arousals the accuracy of the Apnea-Hypopnea-Index (AHI) is limited.

If used with HSAT sleep studies, the deep learning algorithm the Nox BodySleep 2.0, can estimate sleep states and arousals from the raw respiratory inductance plethysmography (RIP) and activity signals thus allowing for more accurate AHI estimation from HSAT studies. We present preliminary validation results for this model, in varied patient subgroups, with regards to its ability to accurately classify individuals into AHI severity groups and with regards to accurate sleep state and arousal estimation.

Methods

External performance validation of the Nox BodySleep 2.0 model was done via retrospective data analysis. The validation data, 2,477 manually scored HSAT studies with limited EEG signals (enhanced HSAT) and manually scored PSG studies, encompassed a wide demographic (including individuals with potentially interfering medical conditions or potentially using interfering medications).



Model performance was validated by comparing epoch-level agreement of sleep states and arousals when studies were manually scored vs. when estimated by the Nox BodySleep 2.0, along with comparison of the downstream parameter of AHI severity classification.

Results

With regards to AHI severity classification the Nox BodySleep 2.0 model achieved high overall agreement with manual scoring, with agreement ranging from 92%-95% across the three categories as seen in Table 1. Table 2 shows the various demographic subgroups' agreements with regards to correct classification of individuals with AHI ≥ 5. Table 3 shows epoch level agreement for sleep state and arousal estimation in males and females, compared to manually scored PSG studies.

Table 1: Accuracy of model AHI severity classification compared to manually scored enhanced HSAT and PSG studies, n = 2463

Overall	PPA%	NPA%	OPA%	F1	Cohen's Kappa
AHI≥ 5	**96** [95,97]	**91** [89,94]	**95** [94,96]	0.84	0.96
AHI≥ 15	**87** [85,89]	**97** [96,98]	**92** [91,93]	0.85	0.91
AHI≥ 30	**85** [82,88]	**99** [98,99]	**95** [95,96]	0.87	0.89

Table 2: Accuracy of model AHI severity classification for AHI ≥ 5 in various population subgroups, compared to manually scored enhanced HSAT and PSG studies

Subgroup	PPA%	NPA%	OPA%	n
Male	**97** [96,98]	**93** [90,96]	**96** [95,97]	1,264
Female	**94** [92,95]	**90** [86,93]	**93** [91,94]	1,062
18-25 YoA	**94** [89,96]	**83** [72,92]	**91** [87,94]	261
26-35 YoA	**93** [89,96]	**83** [72,92]	**91** [82,94]	267
36-45 YoA	**93** [90,96]	**89** [82,95]	**92** [89,94]	404
46-55 YoA	**97** [95,98]	**94** [89,98]	**96** [94,98]	513
56-65 YoA	**98** [97,99]	**97** [94,100]	**98** [96,99]	430
66+ YoA	**98** [97,99]	**98** [95,100]	**98** [97,99]	451
BMI < 25	**92** [90,95]	**85** [79,91]	**90** [88,93]	504
25 ≤ BMI ≤ 30	**95** [93,97]	**87** [82,92]	**93** [91,95]	682
30≤ BMI	**98** [96,98]	**97** [95,99]	**97** [96,98]	1,140
Atrial Fibrillation	**96** [91,100]	**100** [100,100]	**97** [91,100]	59
Asthma	**100** [100,100]	**100** [100,100]	**100** [100,100]	46
nGER	**95** [91,98]	**94** [87,100]	**95** [92,97]	255
Beta-Blockers	**100** [100,100]	**71** [33,100]	**96** [89,100]	46
Antidepressants*	**94** [89,99]	**89** [75,100]	**93** [88.97]	115
Benzodiazepines	**96** [86,100]	**100** [100,100]	**97** [90,100]	34

*SSRI

Table 3: Epoch level agreement of sleep state and arousal estimation between the model and manually scored PSG studies

Male (n= 1016)	PPA%	NPA%	OPA%
Arousal	**65** [64,67]	**83** [82,84]	**79** [79,80]
REM	**76** [74,77]	**98** [97,98]	**95** [95,95]
NREM	**92** [92,93]	**74** [73,76]	**85** [85,86]
Wake	**78** [77,79]	**93** [93,94]	**90** [89,91]
Female (n=941)	**PPA%**	**NPA%**	**OPA%**
Arousal	**59** [58,60]	**85** [85,86]	**80** [79,80]
REM	**80** [97,99]	**98** [97,98]	**96** [95,96]
NREM	**98** [97,99]	**79** [77,81]	**87** [87,88]
Wake	**81** [80,82]	**93** [92,94]	**90** [90,91]

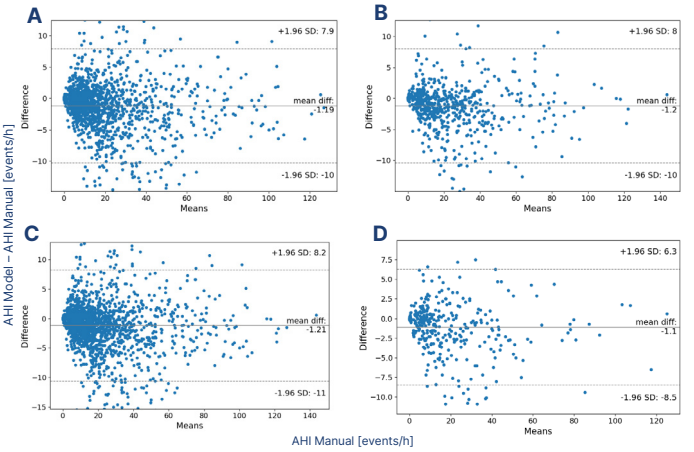


Figure 1. Bland – Altman for the model compared to manually scored enhanced HSAT and PSG studies. A: Agreement for individuals not reporting comorbid conditions B: Agreement for individuals reporting varying comorbid conditions C: Agreement for individuals not reporting any medications D: Agreement for individuals reporting various medications

Conclusions

Preliminary results indicate that when compared to PSG scoring, the Nox BodySleep 2.0 shows similar performance and shows high agreement in classifying AHI severity. No bias was found for the patient subgroups investigated.

Acknowledgment: This work was supported by Rannís (Icelandic Student Innovation fund) [Grant number: 231149-1101]. The project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement. No 965417

European Commission | Horizon 2020 European Union funding for Research & Innovation