Deep Learning-Based Sleep and Arousal Detection in Polysomnography Data

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

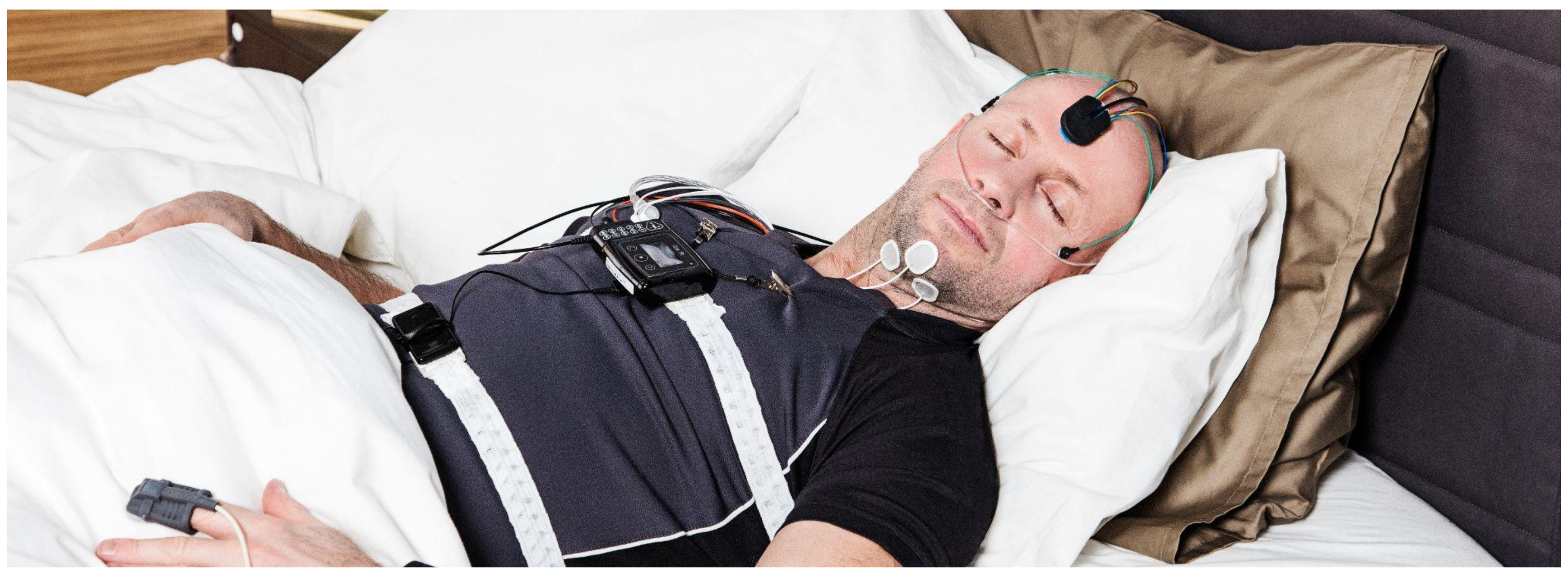
Nox Research, Nox Medical ehf, Reykjavík, Iceland

Presented at Sleep 2023, Indianapolis USA

thora@noxmedical.com

Introduction

Scoring sleep stages and arousals is crucial for accurately interpreting sleep studies and diagnosing sleep disorders. Manual scoring is time-consuming and subject to inter-scorer variability. We present an automatic polysomnography (PSG) sleep stage and arousal scorer that aims to expedite the process of sleep analysis while maintaining high classification accuracy.



Methods

We employed an end-to-end deep learning approach to classify sleep stages (Wake, REM, N1, N2, N3) and detect arousals from Nox PSG studies using EEG, EMG, and EOG signals. The models utilize residual blocks and an added temporal component, temporal convolutional network, to predict sleep stages from raw signals. Two deep-learning-based models were developed: one for sleep stage classification and another for arousal prediction. The models were trained on over 1000 manually scored PSG studies and validated on an external dataset of 820 recordings (see demographics in Table 1). This external dataset was used to evaluate the models' performance and impact on calculation of clinically relevant parameters: apnea-hypopnea index (AHI), total sleep time (TST), and arousal index (AI).

Table 1: Validation sample demographics (N = 820)

	Mean ± SD	Range		Mean ± SD	Range
Height (cm)	172.1 ± 10.6	137.2 - 205.7	**Age (years)**	49.3 ± 16.5	18.0 - 97.0
Weight (kg)	92.3 ± 25.6	38.6 - 229.5	**BMI (kg/m²)**	31.1 ± 8.2	16.0 - 70.6
Gender (N)	Female (388); Male (429); Unknown (3)				

Results

Our approach achieved robust performance. The sleep stage classification model showed good to excellent accuracy for determining all sleep stages (Table 3), and the arousal detection model showed good to excellent epoch-level agreement (Table 2). AHI classification using input from automatic sleep staging with manually scored respiratory events also demonstrated excellent accuracy (Table 2 & 4).

Table 2: Accuracy of the arousal detection model and AHI classification using automatic sleep staging and manual respiratory analysis compared to full manual analysis (Mean [95% CI])

	Sensitivity	Specificity	Class Accuracy
Arousals (Epoch-level)	**66%** [64%, 67%]	**90%** [89%, 90%]	**84%** [84%, 85%]
AHI ≥ 5 (Patients)	**97%** [96%, 98%]	**99%** [97%, 100%]	**98%** [96%, 99%]
AHI ≥ 15 (Patients)	**96%** [94%, 98%]	**100%** [99%, 100%]	**98%** [97%, 99%]

Results (cont.)

Table 3: Accuracy of sleep stage classification by our automatic sleep scorer with reference to manual scoring (Mean [95% CI])

Class	Epochs (N)	Sensitivity	Specificity	Class Accuracy
WAKE	155,317	**96%** [95%, 96%]	**94%** [94%, 95%]	**95%** [94%, 95%]
REM	79,336	**89%** [88%, 90%]	**98%** [98%, 98%]	**97%** [97%, 97%]
NREM	430,500	**91%** [91%, 92%]	**95%** [95%, 95%]	**92%** [92%, 93%]
N1	45,961	**41%** [39%, 43%]	**91%** [91%, 92%]	**88%** [87%, 88%]
N2	325,154	**76%** [75%, 77%]	**88%** [87%, 89%]	**82%** [81%, 83%]
N3	59,385	**52%** [48%, 55%]	**98%** [98%, 98%]	**94%** [93%, 94%]
Average	665,153	**71%** [70%, 72%]	**94%** [94%, 94%]	**91%** [91%, 91%]
Weighted Average	665,153	**78%** [77%, 78%]	**92%** [91%, 92%]	**88%** [88%, 89%]

Table 4: Two-way table for AHI classification using automated sleep staging and manual respiratory analysis versus full manual scoring.

		Automated Scoring			Automated Scoring	
		AHI < 5	**AHI ≥ 5**		**AHI < 15**	**AHI ≥ 15**
Manual scoring	**AHI < 5**	103	1	**AHI < 15**	357	1
	AHI ≥ 5	19	697	**AHI ≥ 15**	18	444

Intraclass correlation coefficients (ICCs) were 0.99 [0.99, 0.99] for AHI, 0.90 [0.86, 0.93] for TST, and 0.61 [0.54, 0.67] for AI (Fig. 1).

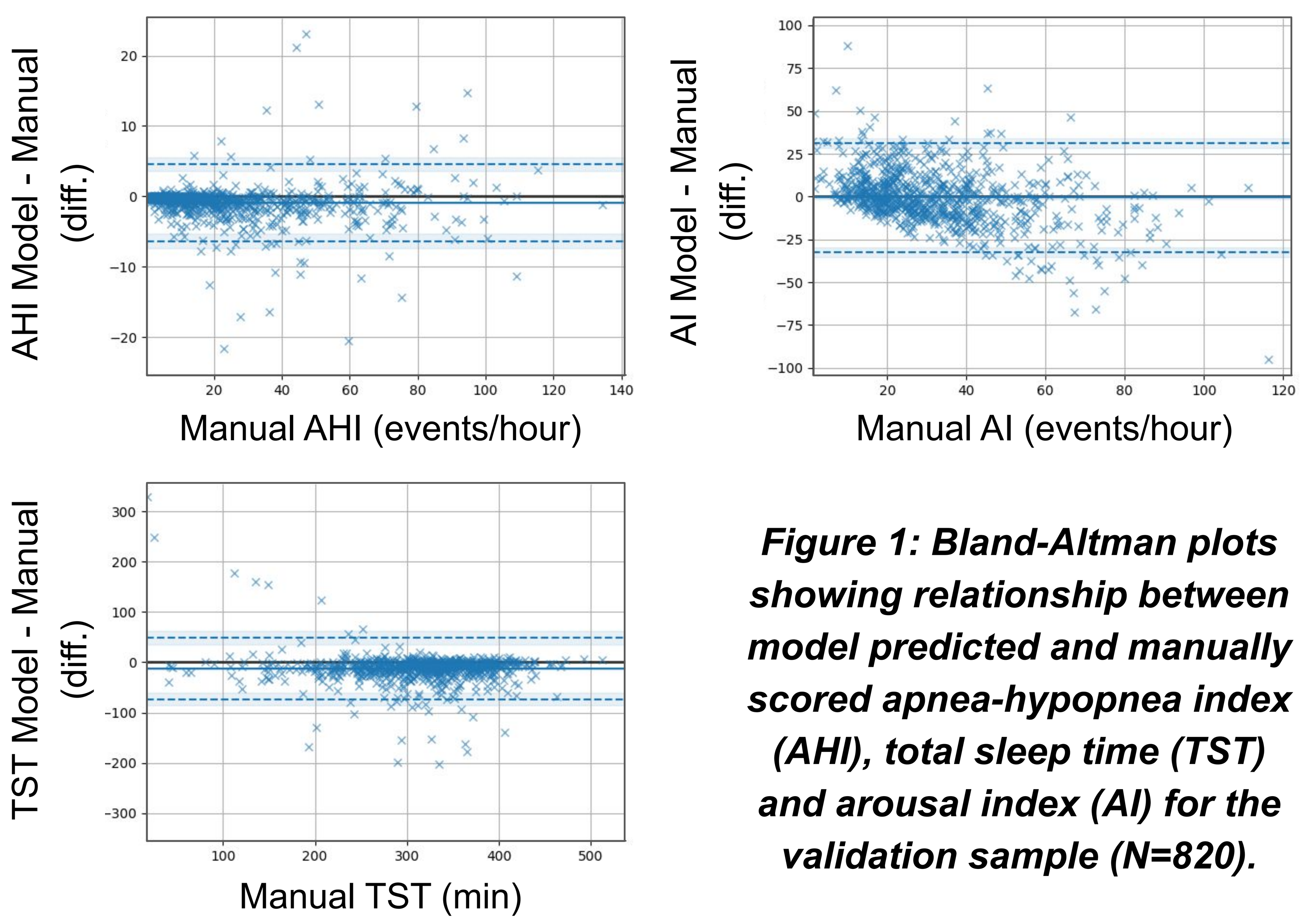


Figure 1: Bland-Altman plots showing relationship between model predicted and manually scored apnea-hypopnea index (AHI), total sleep time (TST) and arousal index (AI) for the validation sample (N=820).

Conclusions

Our deep learning-based models demonstrate promising results for automatic sleep stage classification and arousal detection from PSG data, with the potential to expedite sleep study analysis while maintaining high accuracy. This approach may facilitate more efficient and consistent sleep disorder assessment in clinical settings.