Accurate Sleep Stage Classification Using Robust End-to-End Deep Learning Approach for PSG Studies with Frontal EEG

K. Montazeri[1], S.Æ. Jonsson[1], E. Finnsson[1], E. Erlingsson[1], Þ.B. Sigmarsdóttir[1], H.D. Hlynsson[1], M.E. McPhee Christensen, E. Arnardóttir[1], J.S. Agustsson[1], A.S. Islind[2], E.S. Arnardottir[2]

1. *Nox Research, Nox medical ehf, Reykjavík, Iceland*
2. *School of Technology, Department of Computer Science, Reykjavik University, Reykjavik, Iceland*

Presented at Nordic Sleep Conference 2023, Iceland

thora@noxmedical.com

Introduction

Sleep stage classifications are crucial in diagnosing sleep disorders. However, performing a full in-lab polysomnography (PSG) recording can be time-consuming and expensive. Furthermore, manual sleep stage classification is tedious and subject to scorer variability.

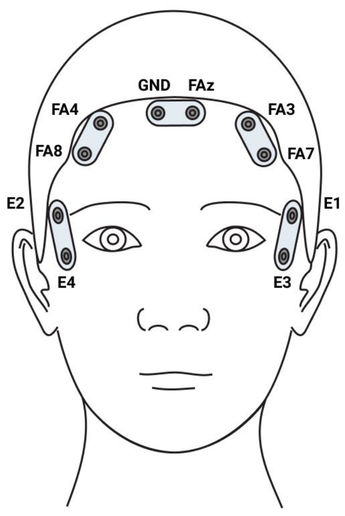


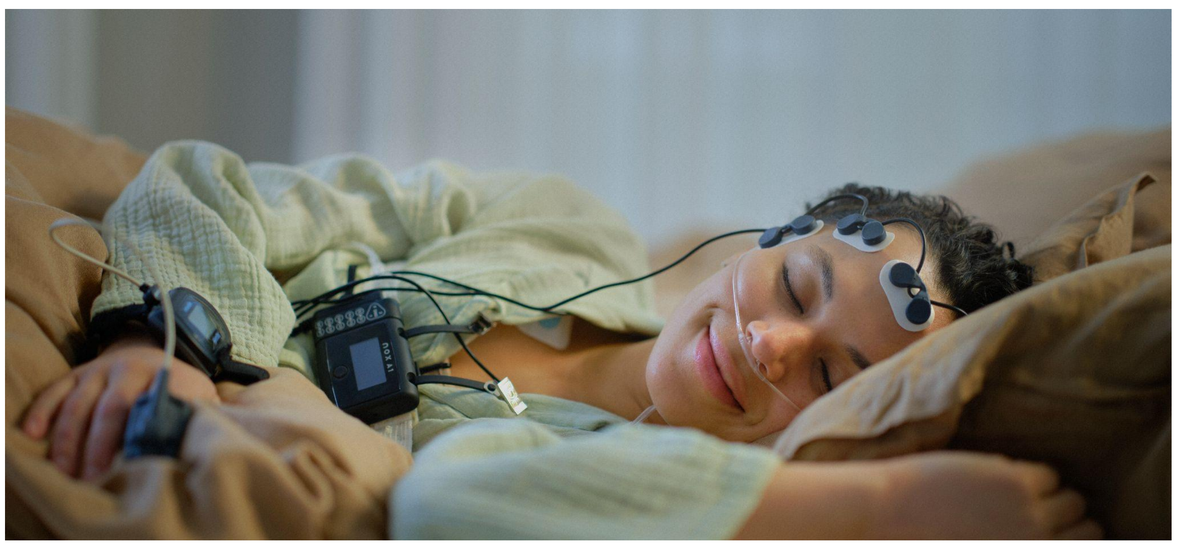
Figure 1: Nox SAS frontal montage

Here we present ResTnet, an end-to-end deep learning approach to classify sleep stages from Nox SAS-PSG studies with a reduced frontal brain (EEG), muscle (EMG) and eye (EOG) electrode montage (Fig. 1).

The model predicts sleep stages from raw signals using residual blocks with an added temporal component, creating a Temporal Convolutional Network (TCN) [1]. This reduces sleep stage prediction time and allows the model to learn more complex relations as training data volume increases.

Methods

The model was trained on 927 Nox SAS-PSG studies and validated on 204 Nox SAS-PSG studies collected at different sleep labs. Model performance was validated by comparing the sleep staging of the model, along with the downstream clinical parameters of total sleep time (TST) and apnea-hypopnea index (AHI), to manual scoring.



Results

The ResTnet model achieved high accuracy in classifying sleep stages, with an overall accuracy of 89%. Sensitivity, specificity, and accuracy for the model in classifying each sleep stage is shown in Table 1, shaded from lowest (fair, orange) to highest (excellent, blue) value. Epoch-level agreement can be seen in Table 2. The weighted F1 score, which evaluates the precision and recall of the classification model, was 0.76, and Cohen's kappa, representing the reliability of the model, was 0.68 suggesting substantial agreement.

Table 1: Accuracy of sleep staging by the ResTnet model compared to manual scoring (N = 193,211 Epochs)

Sleep Stage	Sensitivity [95% CI]	Specificity [95% CI]	Accuracy [95% CI]
WAKE	**76%** [73 - 79%]	**98%** [98 - 99%]	**96%** [96 - 96%]
REM	**76%** [74 - 78%]	**98%** [97 - 98%]	**93%** [93 - 94%]
N1	**67%** [65 - 69%]	**91%** [90 - 91%]	**87%** [86 - 88%]
N2	**87%** [86 - 89%]	**82%** [81 - 83%]	**84%** [83 - 85%]
N3	**61%** [58 - 65%]	**99%** [98 - 99%]	**92%** [92 - 93%]
Average	**74%** [72 - 75%]	**93%** [93 - 94%]	**91%** [90 - 91%]
Weighted Average	**76%** [75 - 77%]	**91%** [91 - 92%]	**89%** [88 - 90%]

Table 2: Confusion matrix of epoch-level agreement between the ResTnet model and manual scoring (N = 193,211 Epochs)

		Model Predicted Sleep Stage (% of Epochs)					
		WAKE	**REM**	**N1**	**N2**	**N3**	**Total (N)**
Manual Scoring	**WAKE**	76.2%	2.0%	18.9%	2.5%	0.4%	18,734
	REM	1.9%	76.2%	16.9%	4.9%	0.0%	39,276
	N1	6.5%	4.4%	66.7%	21.8%	0.6%	29,059
	N2	0.5%	2.8%	6.9%	87.4%	2.4%	72,298
	N3	0.2%	0.4%	0.3%	37.7%	61.3%	33,847

Agreement between the ResTnet model and manual scoring for TST and AHI scoring was excellent, with an intraclass correlation coefficient (ICC(2,1)) of 0.96 for TST and 1.00 for AHI (Fig. 2).

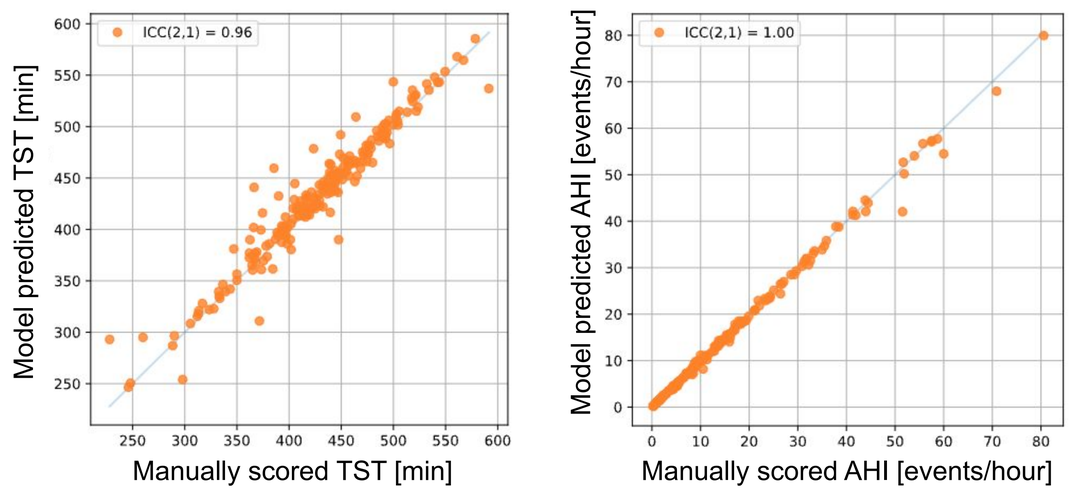


Figure 2: Agreement between the ResTnet model predicted and manually scored TST and AHI (N = 204 Patients)

Patient classification by AHI also showed excellent agreement between the ResTnet model and manual scoring (Table 3).

Table 3: Accuracy of AHI classification by the ResTnet model compared to manual scoring (N = 204 Patients)

	Number of Patients	Sensitivity [95% CI]	Specificity [95% CI]	Accuracy [95% CI]
AHI ≥ 5	137/204	**99%** [96 - 100%]	**100%** [100 - 100%]	**99%** [98 - 100%]
AHI ≥ 15	70/204	**94%** [88 - 99%]	**100%** [100 - 100%]	**98%** [96 - 100%]

Conclusions

The ResTNet model is an attractive option for scoring sleep studies. The model's performance demonstrates that it is a promising tool for accurately classifying sleep stages, estimating total sleep time and calculating apnea-hypopnea index from Nox SAS-PSG studies.

Reference: 1. Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

***Acknowledgement:** This 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