Learning from Time Series for Health Workshop at NeurIPS 2022

Automatic Sleep Scoring from Large-scale Multi-channel Pediatric EEG

Harlin Lee, UCLA Mathematics
Aaqib Saeed, Philips Research, The Netherlands

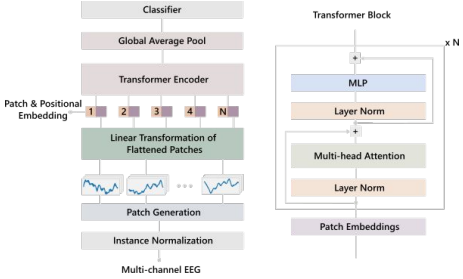
We present the first automated sleep scoring results on a recent large-scale pediatric sleep study dataset that was collected during standard clinical care

This work brings attention to the **growing need for machine learning research on pediatric sleep** by focusing on automated *pediatric* sleep scoring, which has been overlooked in favor of automated *adult* sleep scoring by the community.

- We develop and demonstrate our transformer-based model on the new Nationwide Children's Hospital (NCH) Sleep DataBank [1], which has not been explored in sleep scoring literature before.
- Our model is trained with EEGs that are closest to what it will see in future deployment, which is unlike prior work on sleep scoring that learn from mostly healthy adults in a clinical trial.
- Achieves an overall pediatric sleep scoring accuracy of **78.2%**, and our analysis reveals that the accuracy is above 80% for 6-15 year old patients.

Model

Our **transformer-based model** is inspired by the ViT [2] network, which we adapt here to multi-channel time-series signals.
It is different from previous works as: 1) it is trained specifically for pediatric sleep scoring; 2) it does not utilize any other modalities except EEG signals; and 3) it directly operates over raw signals as opposed to time- frequency images.



Data

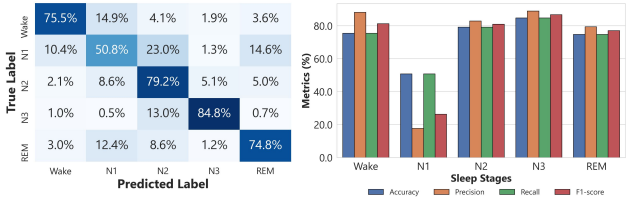
We utilize the **NCH SleepBank dataset** [1], which comprises approximately 3.6 million fully-annotated EEG examples by domain experts, for training and evaluating models.

Only **seven-channel EEG signals** (F4-M1, O2-M1, C4-M1, O1-M2, F3-M2, C3-M2, and CZ-01) at 128Hz are used to classify instances into 5 sleep stages (i.e., wakefulness, non-REM stages 1, 2, 3, REM).

We use **3,928 polysomnography (PSG) from 3, 631 unique patients** for model training and evaluation. We split the patients into 70%, 10%, and 20% for training, validation, and testing, respectively, so that the three splits have no overlap in patients.

Results

Model demonstrates strong pediatric sleep scoring performance.

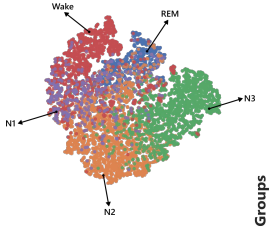


Model sleep scores better on 6 to 15 year olds and children of Asian, Others and Unknown race with over 80% accuracy.

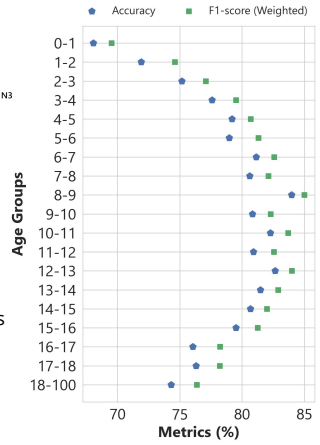
	Accuracy (%)	F1-score (%)
Race		
White	78.6	80.3
Black or African American	76.4	78.3
Multiple Races	78.0	79.6
Asian	78.7	80.3
Others and Unknown	80.6	82.6
Sex		
Male	77.9	79.6
Female	78.5	80.4

Predictive power is not from a single EEG channel.

Channel	Sleep Stages					
	Wake	N1	N2	N3	REM	All
F4-M1	69.1	39.8	75.0	83.1	**78.0**	75.1
O2-M1	70.1	31.5	74.1	81.4	60.7	71.4
C4-M1	68.2	39.9	77.0	83.5	70.8	74.6
O1-M2	67.7	35.0	73.8	78.7	67.4	71.4
F3-M2	**72.4**	**41.8**	75.5	**84.2**	71.1	75.1
C3-M2	72.2	34.7	**78.3**	83.8	70.8	75.7
CZ-O1	69.1	34.6	76.5	80.1	66.5	72.9



t-SNE embeddings of the features from penultimate layer of randomly selected test set samples. The colors are added during post-hoc analysis for better interpretability.



References

[1] Lee, Harlin, et al. "A large collection of real-world pediatric sleep studies." Scientific Data 9.1 (2022): 1-12.

[2] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).