Validating Sleep Stage Prediction: A Test of EEG-Based Models in Pediatric Sleep Analysis

Saurav Raj Pandey, PhD in Computer Science, UNC Chapel Hill
Harlin Lee, Assistant Professor, School of Data Science & Society, UNC Chapel Hill
Aaqib Saeed, Assistant Professor, Dept. of Industrial Design, Eindhoven University of Technology

We showcase the outcomes of an automated sleep scoring experiment conducted using a comprehensive pediatric sleep study dataset that was gathered during routine clinical care practices.

Extending the groundwork established in [1], our project emphasizes the niche of automated pediatric sleep scoring. This initiative is crucial in a field mainly centered on adult sleep to highlight the development in **machine learning applications for pediatric health.**

- We have developed and evaluated our **'ResNet16' model** using the Nationwide Children's Hospital (NCH) Sleep DataBank [2].
- We trained our model with **EEG data** from real-world settings, unlike previous studies that often used data from healthy adults in controlled clinical trials.
- We achieved an overall pediatric sleep scoring testing accuracy of **84.3%**, which is higher than the accuracy reported for automatic sleep scoring in [1].

Model

- Previously [1], a transformer-based model was adapted from the ViT [3] network for multi-channel time-series signals. Unlike most of previous works, our model is trained specifically for pediatric sleep scoring while **only utilizing EEG signals** and directly operating over **raw signals** instead of time-frequency images
- We use a **ResNet-16 architecture** customized for processing one-dimensional EEG signals.
- We utilize sequential residual blocks for **deep feature extraction** from EEG signals, followed by strategic **downsampling** to reduce data dimensionality and focus on relevant features.
- We then employ **global average pooling** to simplify extracted data, leading to a classifier head with linear layers for precise sleep stage classification.

Data

Dataset Utilization:
For model development and assessment, we employ the **NCH SleepBank dataset** [2], which includes around 3.6 million expertly annotated EEG samples that provide a robust foundation for our analysis.

EEG Signal Selection and Classification:
Our approach focuses on **seven EEG channels** (F4-M1, O2-M1, C4-M1, O1-M2, F3-M2, C3-M2, and CZ-01), sampled at 128Hz, to accurately categorize sleep into five stages: wakefulness, non-REM stages 1, 2, 3, and REM.

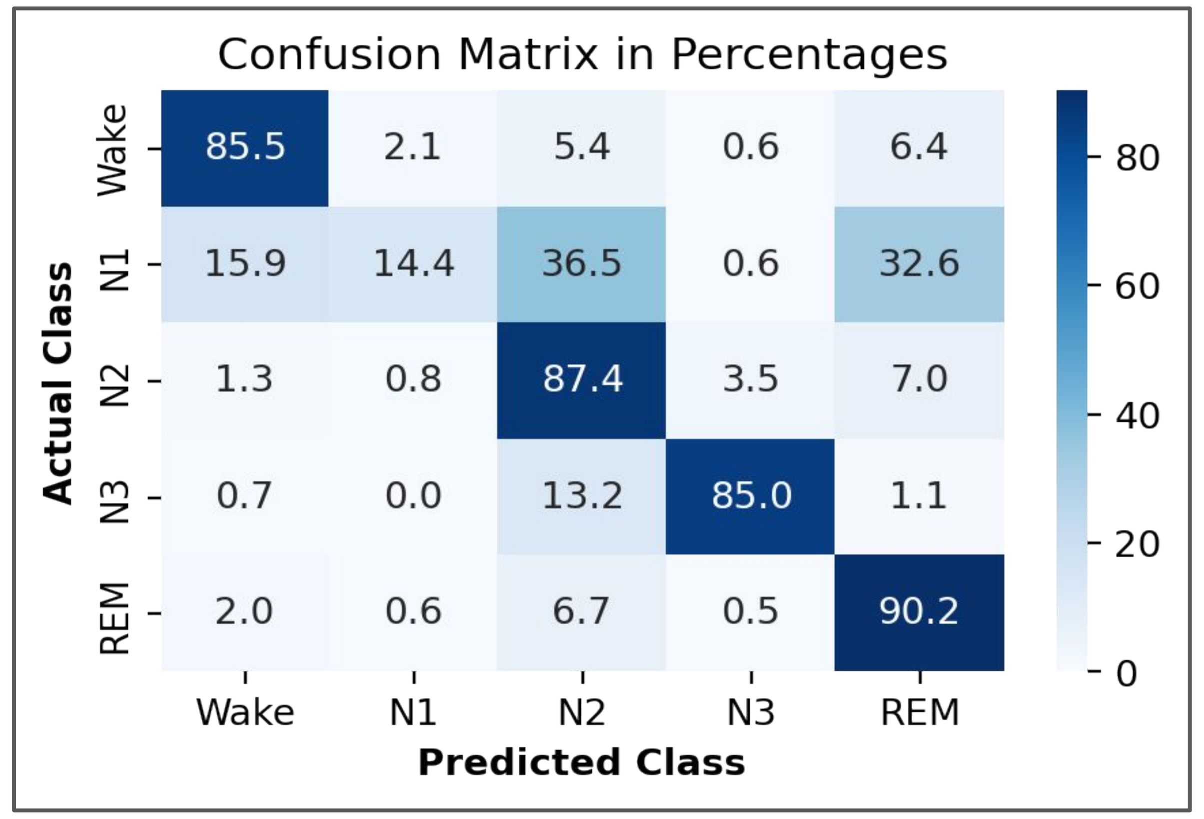
Patient Data and Model Training Split:
The model is trained and tested using **3,928 polysomnography (PSG) records from 3,631 distinct patients**. We partition these records, allocating 80% for training, 10% for validation, and 10% for testing.

Results

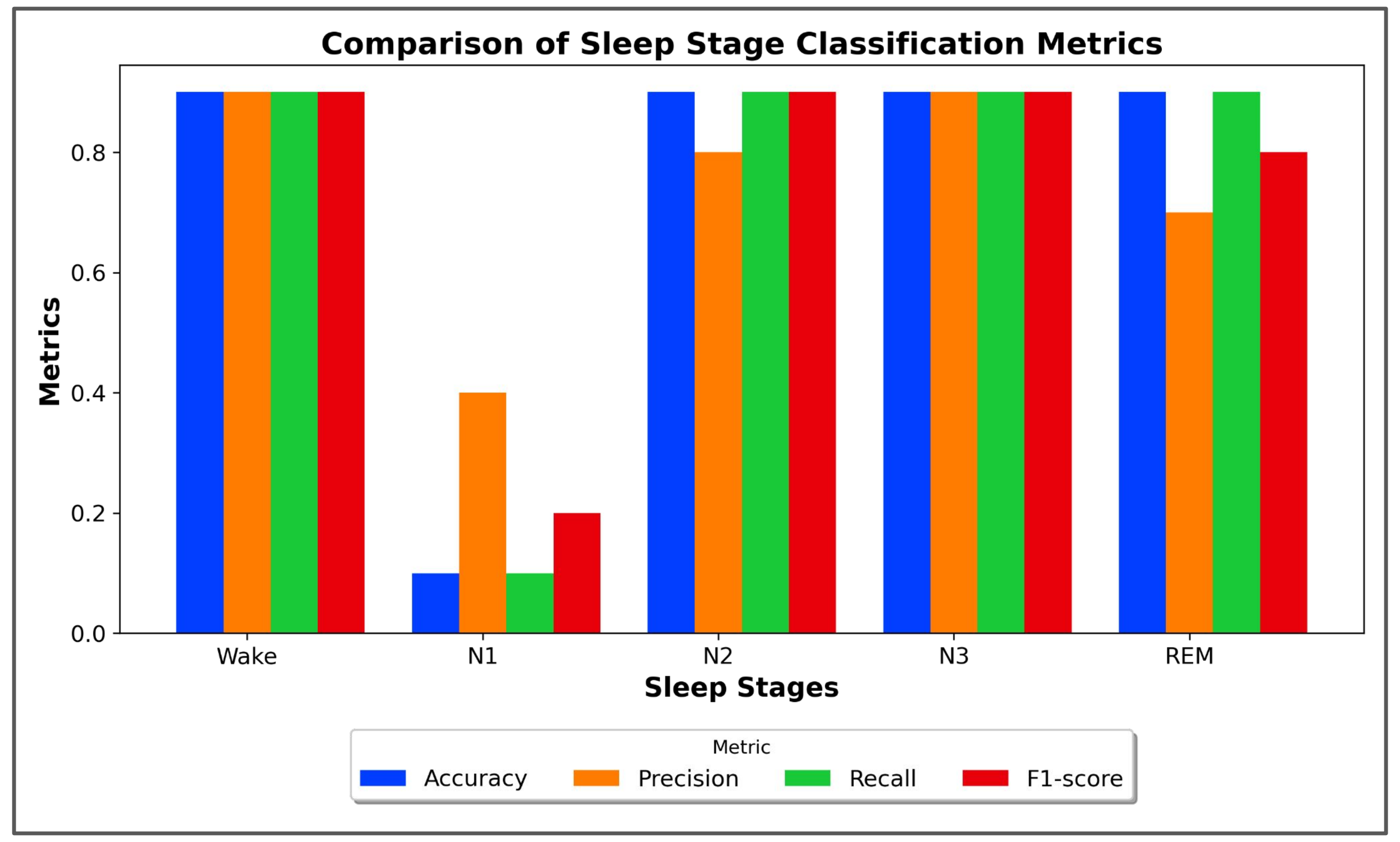
Our model demonstrates strong pediatric sleep scoring performance with an accuracy of above 85% for all but one sleep stage.

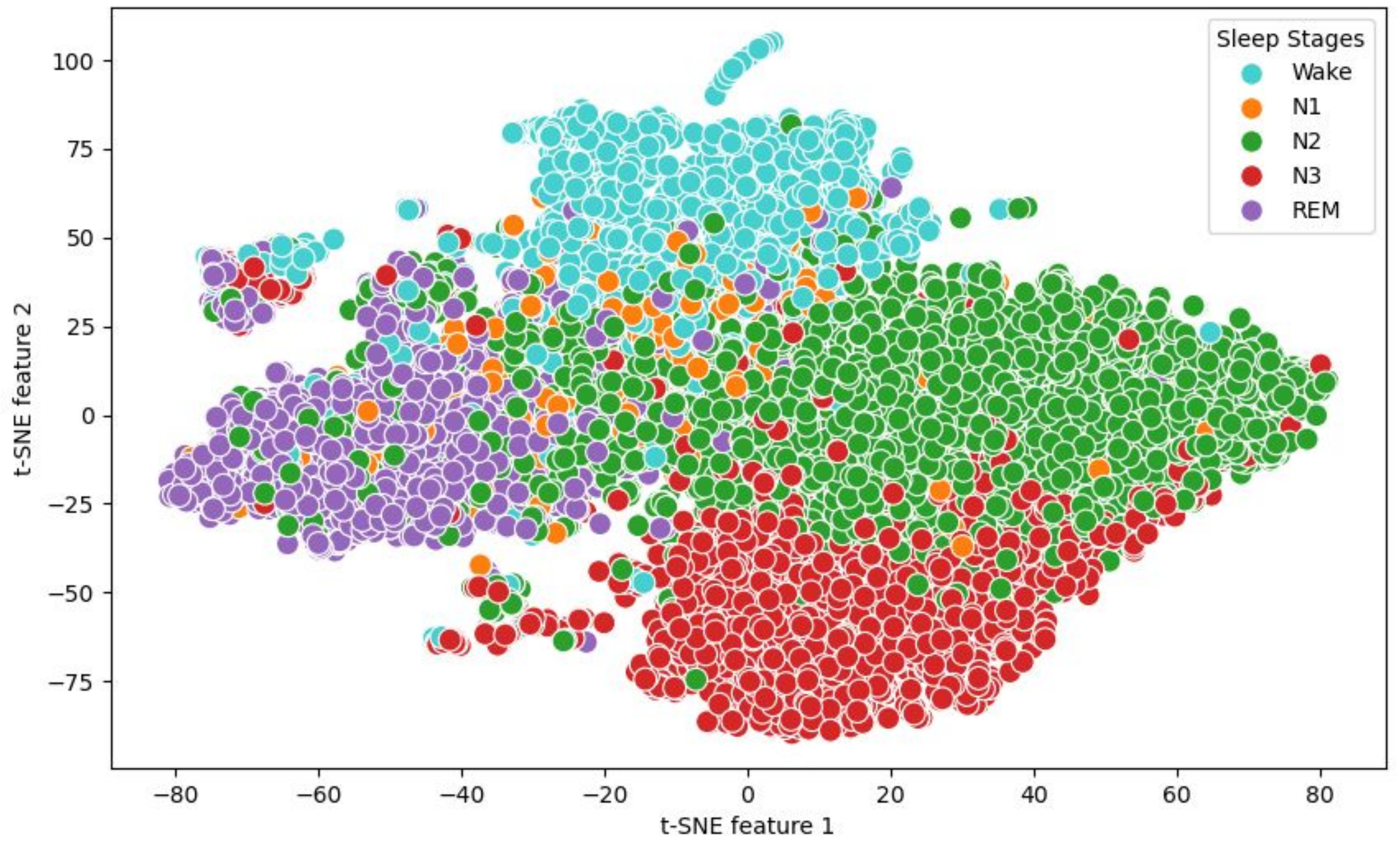
REM stage is the most accurately predicted class, with N1 having the lowest accuracy of 14.4%. **Class imbalance**, where N1 sleep stage samples are underrepresented compared to other stages, is likely to be an issue for this low accuracy in predicting N1.

To mitigate class imbalance, we can oversample class N1 or undersample the other classes to balance the class distribution. Another solution could be to assign a higher weight to class N1 during training so that the model pays more attention to correctly predicting these instances.



In addition to accuracy, other metrics were also quite high, showcasing our model's strong predictive power.





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

Future direction involves training a multi-modal self-supervised learning model for various downstream tasks, like asthma diagnosis for instance.

References

[1] Lee, H., & Saeed, A. (2022). Automatic Sleep Scoring from Large-scale Multi-channel Pediatric EEG, Neurips workshop on Workshop on Learning from Time Series for Health, 2022.

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

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