Optimizing EEG-Based Models for Sleep Onset Detection

STAT 4830

Joshua George, Kimberly Liang, Tereza Okalova, Stefan Zaharia

Introduction & Motivation

- **Problem:** Sleep stage classification (especially N1) is critical for sleep disorder diagnosis but remains challenging.
- **Focus:** Improve detection of **N1 (transitional sleep)** using EEG data.
- **Limitations of Prior Work:**
    - Manual feature extraction struggled with N1 (F1 ~0.06).
    - Over-reliance on multi-channel setups.
- **Solution:**
    - Shift to Sleep-EDF dataset for benchmarking.
    - End-to-end deep learning with **CNN-Transformer hybrid model**.

Jargon Preliminaries

- Power Spectral Density (PSD)
  - Distribution of power into frequency components of which the signal is composed.
  - Canonical frequency bands:
    - **Delta**: 1–4 Hz
    - **Theta**: 4–8 Hz
    - **Alpha/mu**: 8–13 Hz
    - **Beta**: 13–30 Hz
    - **Gamma**: 30–80 Hz
    - **High gamma**: 80–150 Hz
- Electroencephalogram (EEG)
- Polysomnography (PSG)

I. Sleep Neuroscience Primer

When we all fall asleep where do we go

- N1 (Light Sleep)
  - Transition from wakefulness.
  - EEG: Mixed alpha (8–13 Hz) and theta (4–7 Hz) activity; vertex sharp waves may occur.
  - Physiology: Slower heart rate, reduced muscle tone, lower core body temperature.
  - Often, individuals awakened during N1 may not realize they slept.
- N2 (Intermediate Sleep)
  - EEG: Predominant theta (4–7 Hz) with sleep spindles (12–14 Hz) and K-complexes.
- N3 (Deep/Slow-Wave Sleep)
  - EEG: Dominated by delta waves (0.5–2 Hz); occasional alpha intrusions (alpha-delta phenomenon).
- REM Sleep
  - EEG: Low-amplitude, mixed-frequency activity with sawtooth waves, resembling wakefulness.
  - Physiology: Rapid eye movements and near-complete muscle atonia

N1 as a fundamentally transitional state

- By definition, N1 begins once more than 50% of the alpha rhythm is replaced by low-amplitude mixed-frequency (theta) activity.
- It typically lasts only 1–5 minutes (~5% of total sleep) and is very easy to awaken from
- Physiologically, N1 shows features overlapping with relaxed wakefulness (e.g. residual alpha waves, muscle tone still present) and early stage N2 (theta activity but *without* the definitive sleep spindles or K-complexes of N2)
- Human scorers exhibit only “fair” inter-rater agreement on N1, considerably lower than for other stages. Disagreements commonly occur at transitions *into* N1 from wake and *out of* N1 into N2; they often struggle to decide exactly when a drowsy wake epoch becomes N1 or when N1 has progressed to N2.

II. Overview of Machine Learning Approaches for Sleep Staging

SOTA Architectures

Any state-of-the-art model’s merit is partly judged by how well it handles N1 without sacrificing overall accuracy. In comparative evaluations, models like DeepSleepNet, SeqSleepNet, XSleepNet, U-Sleep, and various CNN-LSTM hybrids are often compared on common benchmarks (Sleep-EDF, MASS, etc.). These comparisons show a gradual improvement over time – e.g., XSleepNet (a multimodal RNN+CNN) and U-Sleep (a UNet-like CNN) both pushed sleep staging performance into the low 80s F1-score on Sleep-EDF.

Benchmarks and Datasets

The *National Sleep Research Resource (NSRR)* portal (sleepdata.org) hosts many PSG collections, including the Sleep Heart Health Study (SHHS, ~6,400 overnight recordings) and the Multi-Ethnic Study of Atherosclerosis (MESA, ~2,000 recordings)

For example, Zhang *et al.* (2023) first developed their model on SHHS and then evaluated it on four other cohorts (including MrOS, CCSHS, and SOF) to show robustness.

The Sleep-EDF dataset is a popular **benchmark** for academic papers – it's relatively small (circa 20 healthy subjects recorded in 1980s, plus 8 in a 2013 study), but many papers report results on it, making it useful for baseline comparisons. However, models solely tuned on Sleep-EDF can overfit its peculiarities; using a diverse dataset like SHHS or MASS (Montreal Archive of Sleep Studies) for training will yield a model that likely performs better universally. Other notable datasets include **MASS** (recordings from 200 subjects with multiple nights, used in many GNN papers), **ISRUC-Sleep** (100 subjects from a Portuguese lab, often used for testing novel methods.

Other datasets include **PhysioNet Challenge 2018 data** (a subset of MASS), and clinical datasets like **UCDatabase** or **CAP Sleep database** (for more specialized analysis like cyclic alternating pattern).

Inter-dataset differences need to be accounted for – e.g., Sleep-EDF uses older R&K scoring criteria for some recordings, SHHS and others use AASM; sampling rates and channel setups vary. **Fine-tuning** or at least adjusting the model for these differences (for instance, mapping R&K stages to AASM) is necessary for optimal performance.

III. Data

Data Preprocessing Pipeline

1. **Sleep-EDF**
   - Standardized 30s epochs with expert annotations (W, N1, N2, N3, REM).
   - Publicly available, widely used for benchmarking (e.g., DeepSleepNet, U-Sleep).
   - 100 Hz sampling rate

2. **Channel Selection:**
   - Primary: EEG Fpz-Cz (captures N1 transitions).
   - Optional: multi-channel (EEG+EOG)
3. **Steps:**
   - Resample to 100 Hz.
   - Bandpass filter (0.5–30 Hz).
   - Segment into 30s epochs (3000 samples).
   - Z-score normalization.
4. **Output:** .npz files (epochs + labels) for PyTorch/TensorFlow.

Preprocessed Signal Segments

Epoch-level Files (_epochs.npz):

- **Data Array ("data"):** A NumPy array of shape (n_epochs, n_channels, 3000).
  - n_epochs: The total number of 30-second segments from the recording.
  - n_channels: The number of channels loaded (typically 1 for just EEG Fpz-Cz, or 2–3 if EOG and/or EMG are also included).
  - 3000: Number of samples per epoch at 100 Hz (30 seconds × 100 Hz).
- **Labels Array ("labels"):** A NumPy array of shape (n_epochs,) containing the numeric sleep stage label for each epoch (0 for Wake, 1 for N1, 2 for N2, 3 for N3, 4 for REM).

Sequence-level Files (_sequences.npz):

- **Sequences Array ("sequences"):** A NumPy array of shape (n_sequences, seq_length, n_channels, 3000), where seq_length is the fixed number of consecutive epochs grouped together (e.g., 20). This grouping enables the model to capture temporal context across epochs.
- **Sequence Labels ("seq_labels"):** A NumPy array of shape (n_sequences, seq_length) containing the labels for each epoch within the sequence.

For each PSG file we locate the corresponding hypnogram EDF file, then use MNE's annotation functions to read and set the sleep stage labels. Specifically, we call mne.read_annotations on the hypnogram file and then use mne.events_from_annotations to generate events for each 30-second epoch. These events are then used to create mne.Epochs, from which we extract the labels (mapped to numeric classes) for each epoch.

IV. Model

Model Architecture Overview

- **Hybrid Design:** Combines local feature extraction (CNN) and temporal context (Transformer).
- **Two-Stage Processing:**
  1. **CNN Epoch Encoder:** Converts raw EEG into 128D embeddings.
  2. **Transformer Sequence Model:** Captures dependencies across 20-epoch sequences.

CNN Epoch Encoder

Input: A 30-second epoch with shape (n_channels, 3000 samples); for single-channel use EEG only, while multi-channel includes EEG+EOG (and optionally EMG).

Convolutional Layers: Sequential 1D convolutions capture local time-frequency patterns, filtering the raw signal into meaningful features.

- Conv1D (16 filters, kernel=50, stride=6) → ReLU → MaxPool (kernel=8).
- Conv1D (32 filters, kernel=8, stride=2) → ReLU → MaxPool (kernel=4).
- Flatten → FC layer (128D embedding).

Pooling Operations: Intermediate pooling layers reduce the temporal resolution, highlighting dominant spectral features while lowering computational load.

Feature Embedding: The final CNN layer produces a 128-dimensional embedding, representing the epoch's spectral characteristics for further processing.

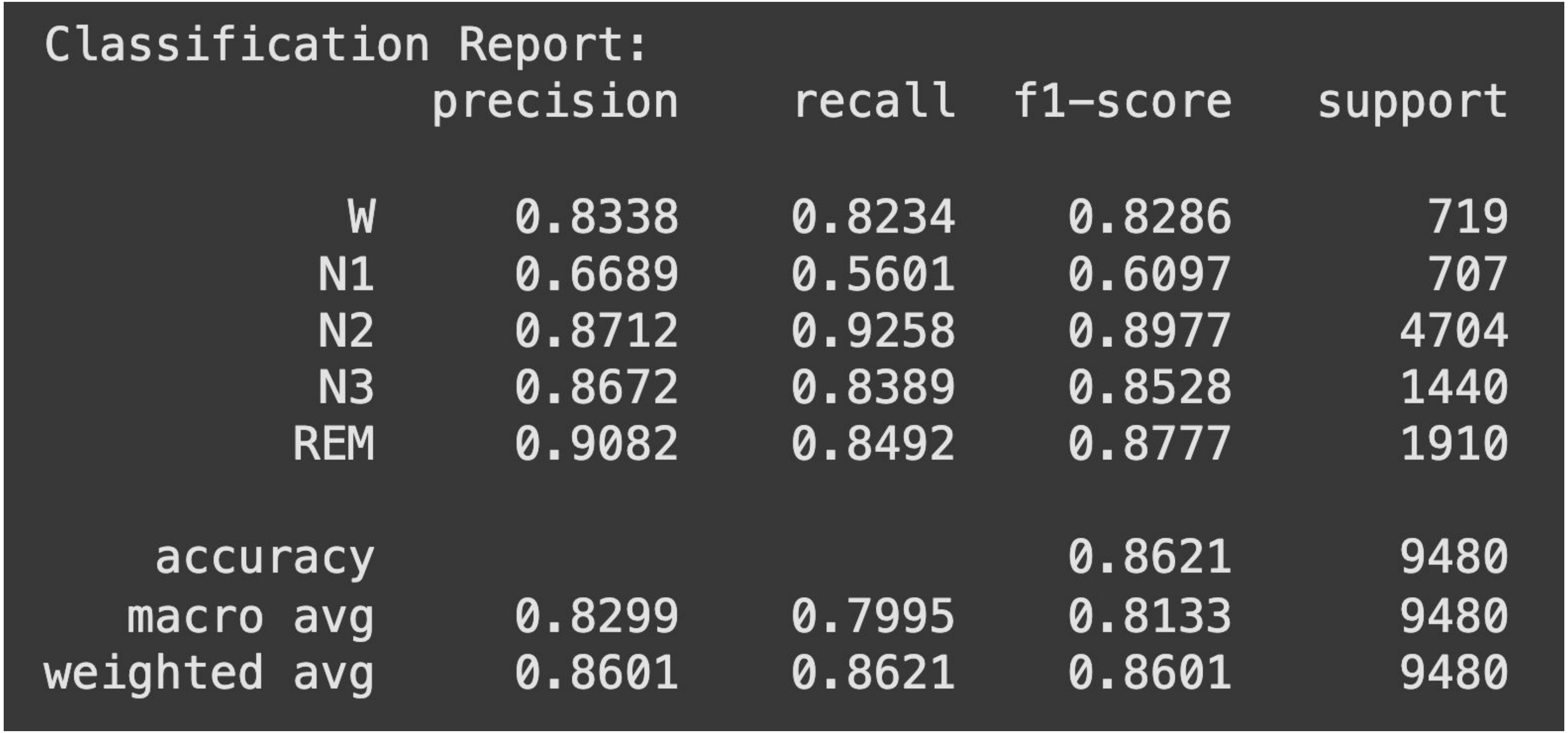
Transformer Sequence Model

- **Input:** Sequence of 20 CNN embeddings.
- **Transformer Configuration:**
  - 2 encoder layers.
  - 4 attention heads.
  - Dropout = 0.1.
- **Output:** Class probabilities per epoch.
- **Key Advantage:** Self-attention helps capture transitions (e.g., N1 ↔ Wake/N2).

Training Loop

- **Loss Function:** Weighted cross-entropy (emphasis on N1).
- **Optimizer:** Adam (LR = 1e-3).
- **Training:** 20 epochs, batch size=16 (subject-level split).
- **Hardware:** M1 MacBook (8 GB RAM; no GPU).

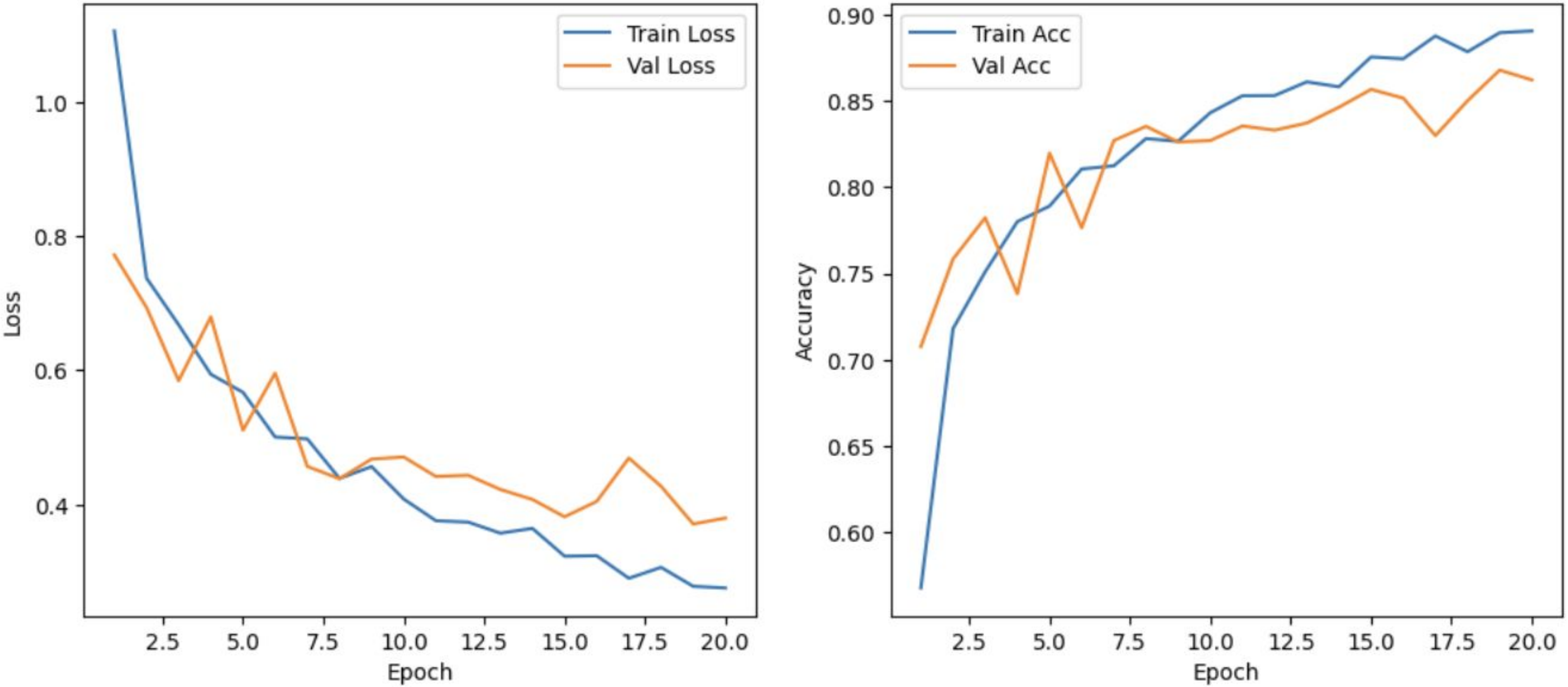
VI. Results



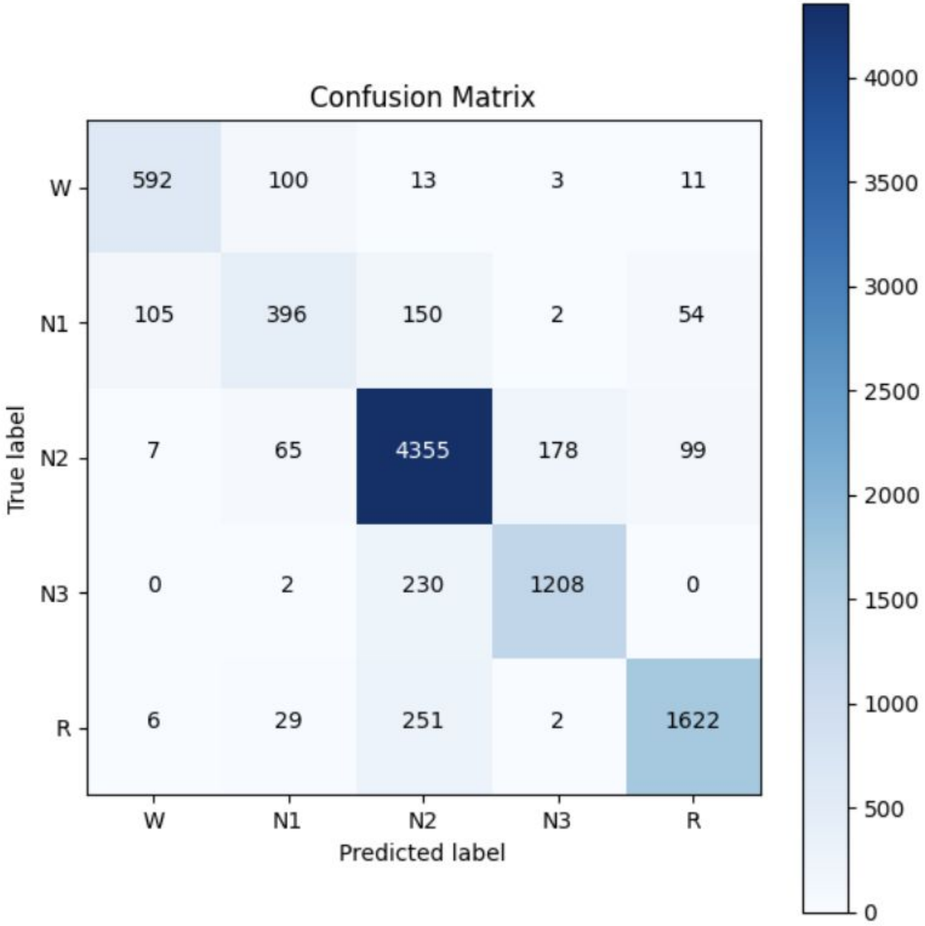
```
Classification Report:
              precision    recall  f1-score   support

           W     0.8338    0.8234    0.8286       719
          N1     0.6689    0.5601    0.6097       707
          N2     0.8712    0.9258    0.8977      4704
          N3     0.8672    0.8389    0.8528      1440
         REM     0.9082    0.8492    0.8777      1910

    accuracy                         0.8621      9480
   macro avg     0.8299    0.7995    0.8133      9480
weighted avg     0.8601    0.8621    0.8601      9480
```

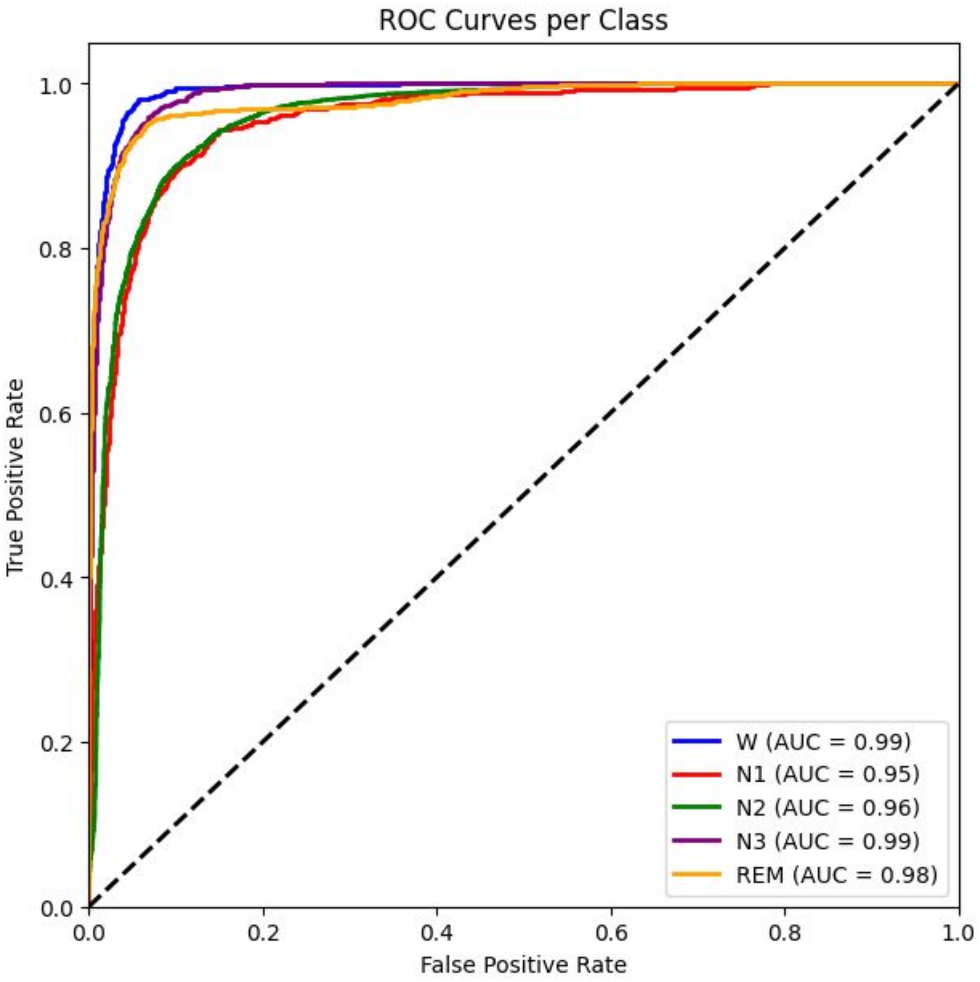


Train Loss
Val Loss
Loss
1.0
0.8
0.6
0.4
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epoch
Train Acc
Val Acc
Accuracy
0.90
0.85
0.80
0.75
0.70
0.65
0.60
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epoch



Confusion Matrix

True label \ Predicted label	W	N1	N2	N3	R
W	592	100	13	3	11
N1	105	396	150	2	54
N2	7	65	4355	178	99
N3	0	2	230	1208	0
R	6	29	251	2	1622


ROC Curves per Class
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
W (AUC = 0.99)
N1 (AUC = 0.95)
N2 (AUC = 0.96)
N3 (AUC = 0.99)
REM (AUC = 0.98)

Post-Processing Work

To improve the biological plausibility of our sleep stage predictions, we will apply rule-based post-processing to smooth and correct the model outputs. Since stage N1 typically occurs as a transitional stage between Wake and N2 (and rarely appears directly adjacent to N3 or REM) we will enforce known sleep architecture constraints by correcting implausible transitions (e.g., Wake→N3 or N1→REM) and relabeling short, isolated stage epochs that are likely misclassified.

Specifically, we plan to:

1. Merge single-epoch N1 or REM labels into surrounding N2 epochs when flanked by consistent stages
2. Apply a 3-epoch median filter to reduce noise from brief outlier predictions

We also may do a Hidden Markov Model so that we are able to classify the transitions in a more algorithmic way

Future Work

1. **Transition Constraints:** Add Markov/HMM layers to penalize unlikely stage jumps.
2. **Self-Supervised Pretraining:** Improve feature learning on unlabeled data.
3. **Channel Ablation Study:** Validate minimal electrode setup.
4. **Cross-Dataset Validation:** Test on MASS or SHHS datasets.

References

1. https://pmc.ncbi.nlm.nih.gov/articles/PMC10317531/#:~:text=is%20lower%20when%20considering%20separately,that%20scorers%20had%20difficulty%20discriminating
2. https://elifesciences.org/articles/70092#:~:text=median%20F1,These%20algorithm%20performance%20results%20were
3. https://www.mdpi.com/2076-3417/13/24/13280#:~:text=signal,pooling%20layers
4. https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2023.1162998/full#:~:text=We%20adopted%20an%20architecture%20comprised,in%20the%20preadolescent%20age%20group
5. https://pmc.ncbi.nlm.nih.gov/articles/PMC9735886/#:~:text=Graph%20neural%20networks%20have%20been,head%20spatial
6. https://pubmed.ncbi.nlm.nih.gov/39243591/#:~:text=Methods%3A%20%20In%20this%20paper%2C,The%20STRL%20module%20employs%20a
7. https://pubmed.ncbi.nlm.nih.gov/31946665/#:~:text=patterns%20themselves%20being%20highly%20sequential,Simulations%20on%20publicly%20available
8. https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2021.653659/full#:~:text=the%20approaches%20developed%20for%20,set%20that%20also%20includes%20incorrect
9. https://ai.jmir.org/2023/1/e46769#:~:text=Self,3%20large%20EEG%20data%20sets
10. https://arxiv.org/html/2412.01929v1#:~:text=al,class%20classification%20models%2C%20respectively%5B%2029
11. https://www.mdpi.com/2227-9067/10/10/1702#:~:text=,switch%20between%20the%20two%20states
12. https://www.researchgate.net/figure/Steps-in-analysis-to-formalise-critical-transitions-in-sleep-NREM-non-rapid-eye_fig2_340086449#:~:text=NREM%20www.researchgate.net%20%20In%20,system%20to%20develop%20new
13. https://pmc.ncbi.nlm.nih.gov/articles/PMC9741833/#:~:text=Dynamics%20of%20sleep%20stage%20transitions,be%20promising%20for%20future%20research
14. https://pubmed.ncbi.nlm.nih.gov/39243591/#:~:text=enhancement%20block%20to%20obtain%20the,labels%20of%20the%20input%20signals