Sleep Stage Classification Project

Ongoing Challenges in N1 Detection

STAT 4830 Project Update

Project Goals and Evolution

Primary Objective

- Improve sleep stage classification from EEG data

- **Critical Focus**: Wake-to-N1 transition detection

- Persistent challenge in sleep medicine

Previous Work (Anphy Dataset)

- 47 spectral and time-series features

- 29 subjects

- Random forests & XGBoost

- Results:
    - N3: Strong (ROC > 0.95)

    - N2, Wake, REM: Good

    - N1: Poor (~0.06)

Transition to Sleep-EDF Dataset

Motivation

1. Well-documented labeling scheme

2. Established benchmarks

3. Public availability

4. Efficient storage

5. Validated by multiple studies

Dataset Characteristics

- Overnight polysomnography (PSG)

- 100 Hz sampling rate

- Multiple channels (EEG, EOG, EMG)

- 30-second epoch annotations

Data Preprocessing Pipeline

Channel Selection Strategy

- Primary: EEG Fpz-Cz
  - Optimal for N1 characteristics
  - Validated in literature
- Auxiliary: EOG horizontal
  - Supports transition detection
- Optional: Chin EMG

Processing Steps

1. Loading and Resampling (100 Hz)
2. Filtering (0.5–30 Hz bandpass)
3. Epoching (30-second windows)
4. Normalization (z-scoring)

Model Architecture

CNN Epoch Encoder

- Input: 30-second epochs
- Multiple convolutional layers
- Extracts time-frequency features
- 128-dimensional embeddings

Transformer Sequence Model

- Processes 20-epoch sequences
- Self-attention mechanisms
- Captures temporal dependencies
- Final classification layer

Recent Focus: Patient-Specific Split

Implementation

```python
def split_patients(data_dir, train_ratio=0.8):
    patient_ids = get_unique_patients(data_dir)
    return train_test_split(patient_ids,
                            train_size=train_ratio)
```

Training Cohort (80%)

```
['SC4032', 'SC4051', 'SC4122', 'SC4231', 'SC4252',
 'SC4352', 'SC4411', 'SC4462', 'SC4572', 'SC4761']
```

Testing Cohort (20%)

```
['SC4292', 'SC4441', 'SC4612']
```

Class Distribution Analysis

Training Set Distribution

```
Stage      Samples      Percentage
_____

Wake       38,740          72.0%
N1          2,204           4.0%
N2          8,418          16.0%
N3          1,364           2.5%
REM         3,134           5.5%
```

Impact

- N1 and N3 severely underrepresented

- Wake stage dominates dataset

- Challenges in rare class detection

Key Challenges

- N1 is still the most underrepresented class

- Critical for deep sleep assessment

- Performance: F1-score 0.2582

- Main confusion with N2 stage (299 cases)

Performance Evolution

1. Initial Results

```
Class      F1-Score
_____

Wake        0.83
N1          0.61
N2          0.90
N3          0.85
REM         0.88
```

2. After Patient-Specific Split

```
Class      F1-Score
_____

Wake        0.96
N1          0.16
N2          0.74
N3          0.26
REM         0.79
```

Current Performance Metrics

Detailed Classification Report

	Precision	Recall	F1-score	Support
W	0.9844	0.9286	0.9557	11,980
N1	0.4841	0.0990	0.1644	616
N2	0.6478	0.8623	0.7398	2,374
N3	0.2022	0.3571	0.2582	518
REM	0.7962	0.7791	0.7876	652
Accuracy			0.8628	16,140
Macro avg	0.6230	0.6052	0.5811	16,140

ROC Curve Analysis

AUC Scores

- Wake: 0.99
- N1: 0.94
- N2: 0.97
- N3: 0.94
- REM: 0.99

Confusion Matrix Deep Dive

Key Patterns

1. Strong Wake detection (11,125 correct)

2. N1 confusion with N2 (377 misclassified)

3. N2–N3 bidirectional confusion

4. REM relatively well-isolated

Critical Transition Analysis

Wake → N1 Transitions

- 118 misclassifications

- Critical for sleep onset detection

- Potential for temporal modeling

N1 → N2 Transitions

- 377 misclassifications

- Most challenging transition

- Need for better feature extraction

N2 ⬌ N3 Transitions

- Bidirectional confusion

- 299 N3→N2, 249 N2→N3

- Deep sleep transition challenges

Ongoing Challenges

1. **N1 Detection**

   - Primary challenge area
   - Patient-specific variations
   - Transition ambiguity

2. **Class Imbalance**

   - Severe underrepresentation
   - Impact on rare classes
   - Need for better balancing

3. **Patient Generalization**

   - Performance drop in new patients
   - Individual variation effects
   - Need for robust features

Next Steps

Immediate Focus

1. Improved N1 detection strategies

2. Better handling of class imbalance

3. Patient-specific adaptation

Technical Approaches

1. Dynamic sampling weights

2. Focal loss implementation

3. Transition-focused features

4. Early stopping on N1 metrics

Data Augmentation

1. Synthetic N1 samples

2. Transition period enrichment

3. Patient-specific augmentation

Performance Metrics Part 2

Classification Report (N3, REM, Overall)

	Precision	Recall	F1-score	Support
N3	0.2022	0.3571	0.2582	518
REM	0.7962	0.7791	0.7876	652
Accuracy			0.8628	16,140
Macro avg	0.6230	0.6052	0.5811	16,140

Challenges

- N3 is most underrepresented

- Imbalance affects model training

- Need for targeted solutions

Performance Metrics Part 1

Classification Report (Wake, N1, N2)

	Precision	Recall	F1-score	Support
W	0.9844	0.9286	0.9557	11,980
N1	0.4841	0.0990	0.1644	616
N2	0.6478	0.8623	0.7398	2,374

REM Stage Analysis

Distribution

Stage	Samples	Percentage
REM	3,134	5.5%

Performance Characteristics

- F1-score: 0.7876

- Precision: 0.7962

- Recall: 0.7791

- Primary confusion with N2 (137 cases)

Class Imbalance Impact

Overall Distribution Pattern

- Wake dominates (72.0%)

- N2 second most common (16.0%)

- N1, N3, REM underrepresented

Challenges for Model

- Bias toward majority classes

- Poor rare class detection

- Need for balanced training strategies