AFML Project One Pager

T3N5OR Team Members:

Adarsh R Menon (PES1UG23AM016)

Akshay P Shetti (PES1UG23AM039)

Tarun S (PES1UG23AM919)

Adityaa Kumar H (PES1UG23AM025)

Project Title: Robust Sleep Stage Detection from Scarce & Imbalanced EEG Data using Generative Modelling and Semi-Supervised Learning

Dataset: Sleep-EDF (Original Subset)

The Sleep-EDF (Sleep Cassette) dataset contains EEG recordings from 20 healthy adults (aged 25–34 years) over 39 nights. Signals were recorded using Fpz–Cz and Pz–Oz EEG channels at 100 Hz. Each 30-second segment is annotated with expert-scored sleep stages (W, N1, N2, N3, REM). It is inherently imbalanced (N1/N3 underrepresented), making it ideal for studying robust classification and data augmentation techniques.

Project Pipeline (1-Month Timeline):

Week 1 – Data Preparation: Load Sleep-EDF subset, filter and segment into 30 s epochs, normalize, and visualize imbalance.

Week 2 – Baseline Model: Train a 1D-CNN classifier with class-weighted loss and focal loss; evaluate macro-F1 and per-class recall.

Week 3 – Generative Modelling: Implement **Variational Autoencoder (VAE)** to synthesize minority-class (N1/N3) EEG data and retrain classifier.

Week 4 – Semi-Supervised Learning: Apply SSL algorithms *(CisTran/SimCLR/CPC/TS-TCC)* to leverage unlabeled data; compare against baselines using macro-F1, κ, and confusion matrices.

Model Selection Rationale:

Model	Reason for Selection
1D-CNN	Captures local temporal EEG features efficiently; serves as strong baseline.
VAE	Generates realistic synthetic EEG for minority classes; stable on small datasets.
SSL	Leverages unlabeled EEG data to improve generalization and label efficiency.

Reference Papers:

1. L. Huang, L. Yao, X. Li and B. Dong, "Staging study of single-channel sleep EEG signals based on data augmentation," Frontiers in Public Health, vol. 10, 2022

2. Self-supervised Learning for Label-Efficient Sleep Stage Classification: A Comprehensive Evaluation

3. Sleep-EDF Dataset: https://physionet.org/content/sleep-edf/1.0.0/