Multimodal Sleep Signal Phenotyping for Predicting Future Cognitive Impairment

Sejuti Dhali, Tasmeya Siddique Mayesha

Department of Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh.

Overnight polysomnography (PSG) encodes rich neurophysiological, autonomic, respiratory, and oximetric signatures that may reveal cognitive decline long before clinical diagnosis. Early, objective risk stratification could enable timely intervention for Alzheimer's disease and related dementias. We address this challenge using the Human Sleep Project database, which frames future cognitive impairment prediction as a binary classification problem from baseline PSG recordings.

We audited 622 public training recordings from three acquisition sites (S0001: 484; I0006: 102; I0002: 36), revealing source imbalance and nontrivial metadata missingness (BMI missing in 232 records; follow-up interval in 311). An initial XGBoost baseline using demographic and site-level covariates achieved a validation AUROC of 0.492, confirming the need for physiological features. Our proposed architecture implements a physiology-first late-fusion framework. A primary branch extracts expert whole-night PSG phenotypes—EEG spectral and sleep-architecture descriptors, nocturnal autonomic and ECG-derived markers, respiratory burden measures and EOG/EMG indices. A secondary branch derives confidence-weighted phenotypes from CAISR algorithmic annotations (stage, arousal, respiratory, and limb channels), treating automated scores as auxiliary evidence rather than ground truth, given a measured CAISR–expert sleep-stage agreement of $\kappa = 0.52$. Branch-level predictors are integrated via calibrated stacking, with model selection governed by grouped-by-site cross-validation to control site-specific bias and ensure challenge-compliant packaging.

The baseline scaffold produced AUROC 0.492, AUPRC 0.039, accuracy 0.175, and F-measure 0.071, establishing a reproducible submission pipeline while quantifying the gap to a physiologically meaningful solution. Full multimodal feature extraction and architecture training are ongoing.

Structured dataset auditing exposed critical site-imbalance and annotation-reliability constraints that directly shaped our multimodal late-fusion design. Completing physiology-first feature extraction and calibrated stacking is expected to yield substantial gains over the demographic baseline.