Title: Multimodal Sleep Feature Extraction and Stratified Feature Selection for Dementia Prediction

Objective: The objective is To develop a robust machine learning pipeline for predicting dementia from multimodal polysomnography data by combining detailed feature extraction with stratified feature selection.

Methods: The pipeline composed of two main components: feature extraction and feature reduction. Features are computed per channel and sleep stage from EEG, ECG, EOG, EMG, and respiratory signals using modality-specific and generic time-frequency descriptors. Cross-channel synchrony (e.g., hemispheric EEG, thoracoabdominal respiration) is also quantified. Extracted features are pivoted to a subject-session level representation. Feature reduction is performed using stratification by modality and sleep stage. Within each stratum, features are filtered based on the number of missing values to mitigate imputation bias. Dimensionality reduction is applied using principal component analysis for multimodal feature groups and recursive feature elimination otherwise. Models are trained using gradient-boosted trees, and feature importance scores are aggregated across strata, weighted by cross-validated AUC, to construct a global ranking. Final feature subsets are selected based on this ranking.

Results: An earlier submission without the full feature reduction pipeline achieved an AUC of 0.53 during the unofficial evaluation phase. The proposed extended pipeline was not yet evaluated on the official test set and thus has no official score. Cross-validated experiments on the training data yielded AUC values of 0.66 on the full feature set, while stratified subsets achieved AUC values between 0.66 and 0.70. These results are preliminary and not the official scores of evaluation of the challenge submission system.

Conclusion: Stratified feature selection combined with multimodal sleep feature extraction shows promise for dementia prediction, although a full-scale evaluation has not yet been completed.