# Robust Feature Discovery for the Prediction of Incident Dementia from Multi-Source Sleep Studies

Allan R Moser*, Sai Nyi Bhone Htut

Swarthmore College, Swarthmore, PA, USA

**Aim:** The goal of the 2026 PhysioNet Challenge is to use machine learning algorithms to predict the future diagnosis of dementia using polysomnographic (PSG) data. Evidence indicates sleep disturbances precede dementia by five to fifteen years and may even contribute to its onset via impaired glymphatic clearance. Identifying highly accurate classification features could provide actionable biomarkers for early intervention. Consequently, our motivation focuses on identifying robust, generalizable features using an optimization framework and established machine learning methods.

**Method:** In preliminary testing, our 5-fold cross-validation AUROC was 0.821 on the balanced training data provided, but dropped to 0.465 on the unofficial phase hidden validation data (5–15% positives). Since AUROC should be insensitive to class balance, this severe drop clearly demonstrates the need for a robust strategy to find features that generalize across multi-source PSG data. Team SwatBeat's four-part approach addresses this by: 1) extracting domain-agnostic features such as sleep transition probabilities, signal ratios, and sleep fragmentation indices; 2) applying an adversarial "anti-feature" filter to eliminate features statistically correlated with the source institution; 3) using time-to-diagnosis data (only available for training data) to isolate subtle prognostic signals and prioritize features correlated with disease progression; 4) using multiple draws filtered to match the original data's low positive rate, withholding instances from a single source for validation, and applying L1 regularization to obtain a small number of universally robust features.

**Conclusion:** By systematically isolating true physiological markers from institutional bias, we aim to deliver a highly generalizable and clinically viable model for the early detection of dementia.