Predicting Cognitive Impairment From Polysomnography Using Combined Sleep Depth and Fragmentation Features

Richard Hohmuth[1], Jakob J. Mueller[1], Hagen Malberg[1], Martin Schmidt[1]

[1] Institute of Biomedical Engineering, TU Dresden, Dresden, Germany

Abstract

Changes in sleep dynamics are increasingly recognized as early biomarkers of neurodegenerative diseases. We present a machine learning framework for predicting cognitive impairment (CI) from nocturnal polysomnographic (PSG) recordings by jointly modeling two complementary dimensions: sleep depth and sleep fragmentation. Our framework extracts a multimodal feature space across 6,600 multi-site PSG recordings. Sleep depth is quantified via slow-wave activity, odds ratio product, and delta power entropy, while sleep fragmentation is assessed through cyclic alternating pattern, arousal temporal distributions, and respiratory event indices. Patient-level features are filtered using combined ANOVA and mutual information ranking, and gradient-boosted trees (XGBoost) are trained with class-imbalance weighting. Model interpretability is evaluated using Shapley additive explanations (SHAP). In official evaluation, our approach achieved an age-conditioned AUROC of 0.661 ± 0.037 *in 5-fold cross-validation on the training set and* 0.636 *on the official hidden validation set. SHAP analysis highlighted the predominant importance of cycle-dependent frontal slow-oscillation dynamics, non-REM stage 2 (N2)* θ *variability, cyclic alternating pattern duration, and sleep spindle power, indicating that sleep depth and fragmentation features jointly inform CI prediction.*

1. Introduction

Sleep is a fundamental physiological process that serves restorative, metabolic, and cognitive functions. Sufficient sleep is essential for memory consolidation, synaptic adaptation, and the clearance of amyloid-beta peptides via the glymphatic system, which, alongside tau proteins, are primary biomarkers of Alzheimer's disease and dementia [1, 2]. Conversely, these biomarkers exhibit pathological accumulation years prior to clinical cognitive decline and are associated with alterations in sleep efficiency and sleep quality [3, 4]. Changes in sleep architecture, such as diminished slow-wave sleep [5], a reduced number of fast sleep spindles [6], and increased sleep fragmentation [7], are associated with higher levels of amyloid-beta and/or tau protein pathology. These findings underline the utility of polysomnography (PSG) as a non-invasive, population-level tool for assessing the risk of cognitive impairment (CI).

The George B. Moody PhysioNet Challenge 2026 invites teams to develop automated algorithms that use PSG recordings to screen for CI [8]. Our team, Biosignal Pilots, addresses this challenge by generating a comprehensive feature space structured around two complementary dimensions of sleep quality, supplemented with cardiorespiratory measures. We hypothesize that sleep quality is a joint product of both sleep depth and sleep fragmentation and that the combined analysis of these two dimensions reveals early stages of CI before patients exhibit overt clinical symptoms. Our pipeline integrates multimodal PSG signals, such as electroencephalogram (EEG), electrocardiogram (ECG), respiratory signals, and expert sleep stage annotations, into a unified feature space and trains regularized gradient-boosted classifiers to predict CI status.

2. Methods

2.1. Data

We used the large training dataset provided by the PhysioNet Challenge 2026, comprising 6,600 multi-site PSG recordings with corresponding sleep stage and respiratory event annotations, demographic information, and cognitive-impairment labels [8]. Each recording contained at least one ECG channel and one or more EEG channels (e.g., F3, F4, C3, C4), alongside respiratory effort and airflow signals. Sleep stages (30-second epochs), respiratory events, and sleep arousals were provided via expert-human and algorithmic annotations.

2.2. ECG Preprocessing

The raw ECG signal was high-pass filtered with a cutoff frequency of 0.5 Hz using the NeuroKit2 *ecg_clean* function. R peaks were identified in chunked 5-minute seg-

ments using *ecg_peaks*, and redundant detections at segment boundaries were removed by enforcing a minimum inter-peak distance of 300 ms. RR intervals were computed, followed by median-based artifact rejection (deviations > 20 % from the moving median were discarded). A per-segment signal quality index (SQI) was computed as a weighted combination of QRS template-correlation SQI (40 %), RR interval regularity SQI (30 %), and signal-to-noise ratio SQI (30 %). Segments with low SQI (< 0.5) were excluded from heart rate variability (HRV) feature extraction.

2.3. EEG Preprocessing

Each available EEG channel was bandpass-filtered using a 4th-order Butterworth filter between 0.3 and 35 Hz, and notch-filtered at the powerline frequency (60 Hz). Re-referencing of EEG leads to bipolar reference was conducted if necessary. Global bad-channel detection was performed across all channels using NeuroKit2's *eeg_badchannels* function. Channels identified as bad quality received an SQI of zero for all segments. For remaining channels, a segment-wise SQI combined three sub-metrics: amplitude SQI (fraction of samples within physiological bounds, weight 0.35), flatliner SQI (fraction of 1-second windows with variance within specified bounds, weight 0.30), and spectral SQI (ratio of power within 0.5–30 Hz to total power, combined with normalized entropy of the band-power distribution, weight 0.35).

2.4. Annotation Preprocessing

Sleep stage annotations (wake (W), non-rapid eye movement stages 1–3 (NREM, N1, N2, N3), and rapid eye movement (REM)), respiratory events, and sleep arousals were parsed from numeric encodings into standard labels. Both sleep stages and sleep events were mapped onto analysis segments using an overlap-weighted assignment approach. For sleep stages, the dominant stage within each segment was defined as the stage occupying the largest proportion of the segment duration. For sleep events, segment-wise metrics were computed, including event counts, cumulative durations, and bidirectional distances to the nearest event of each type.

2.5. Segmentation

All signals and annotations were segmented into uniform, non-overlapping time windows with a length of 300 s, corresponding to the minimum gold-standard window length for HRV parameter calculation [9]. Segment boundaries were computed once and applied consistently across ECG, EEG, and respiratory channels. Segments shorter than 50 % of the target length, occurring only at the end of a recording, were removed from analysis. Per-segment metadata, including SQI values for each signal type, dominant sleep stage, and event flags, was compiled into a segment-metadata table which was further used for feature extraction.

2.6. Feature Extraction

2.6.1. HRV Features

For each segment passing the ECG SQI threshold, a set of 33 HRV features was extracted across five groups: (1) *Time-domain*: standard deviation of normal RR intervals (SDNN), root mean square of successive differences (RMSSD), and related statistics. (2) *Frequency-domain*: low frequency (LF), high frequency (HF), and spectral power measures via Welch's method [10]. (3) *Nonlinear*: Poincaré plot parameters (SD_1, SD_2), sample entropy, and complexity metrics. (4) *Cardiorespiratory coupling*: respiratory sinus arrhythmia (RSA) metrics and cross-correlation parameters. (5) *Heart rate statistics*: mean HR, standard deviation, artifact percentage, and longest valid/artifact run lengths.

2.6.2. Sleep Quality Features

Sleep quality features were calculated for each EEG channel and segment passing quality thresholds alongside annotation metrics.

Sleep Depth Features: A total of 73 EEG-derived and 18 annotation-derived features were extracted: (1) *odds ratio product (ORP)*: continuous sleep depth metric computed following Younes et al. [11] using lookup tables derived from $>$ 3,000 PSG recordings. (2) *spectral and microstructural markers*: δ power (0.5–4 Hz) [10], δ power entropy [12], sleep spindle features (density, duration, amplitude) from the Luna framework [13], slow-oscillation (SO) characteristics, σ power (12–16 Hz), and cross-frequency power ratios. (3) *PSG annotations*: dominant sleep stages, stage fractions, and NREM – REM cycle progression parameters.

Sleep Fragmentation Features: A set of 31 EEG-derived and 33 annotation-derived features were computed: (1) *cyclic alternating pattern (CAP)*: A-phase detection was performed using a deep neural network framework [14], extracting 31 CAP parameters including A-phase subclass fractions (A1, A2, A3) and cycle durations. (2) *sleep disruption metrics*: statistics for sleep arousals, obstructive/central/mixed apneas, hypopneas, stage transitions, wake fraction, and WASO. Additional spectral features, nonlinear metrics, and global architecture sum-

maries (stage transitions, event indices) were also computed.

2.7. Feature Aggregation

Segment-level features were aggregated to patient-level summaries by computing the mean, standard deviation, median, 25^{th} and 75^{th} percentiles, and interquartile range across all valid segments. Sleep stage-conditioned combinations (e.g., mean SDNN during N3, mean δ power during NREM) were also computed where sufficient segments existed. Global sleep-architecture summaries and demographic variables (age, sex, BMI) were appended, yielding an initial candidate pool of 8,193 features (HRV: 440, sleep depth: 3,511, sleep fragmentation: 524, demographics/global architecture: 3,718).

2.8. Feature Selection

A combined two-stage feature selection strategy was applied. Features with > 50 % missing values or zero variance were removed prior to selection. For remaining features, min-max normalized ANOVA F-statistic scores and mutual information scores with respect to the binary CI label were averaged with equal weights. The top $k = \min(100,\ 3\sqrt{N}) \leq 100$ features were retained (selecting 74–83 features across folds), compressing the feature space by > 99 % to constrain model complexity.

2.9. Classification and Model Explainability

Missing feature values were imputed with the column median, and all features were standardized to zero mean and unit variance. Gradient-boosted classifiers (XGBoost, LightGBM) and Random Forest were trained. Class imbalance was handled via class-weight scaling. Hyperparameters were tuned using the Optuna framework [15]. Performance was evaluated using stratified 5-fold cross-validation with age-conditioned AUROC, AUROC, AUPRC, accuracy, and F1 score as primary metrics. Shapley additive explanations (SHAP) [16] was performed to quantify feature contributions.

To evaluate whether the model learned site-invariant biomarkers rather than acquisition artifacts, SHAP distributions of the top features were tested across clinical recording sites using Kruskal-Wallis H tests, followed by Bonferroni-corrected Mann-Whitney U tests.

3. Results

The performance of the proposed XGBoost model across 5-fold cross-validation on the training set ($N = 6{,}600$) and on the official hidden validation set is summarized in Table 1. In 5-fold cross-validation, the model achieved a mean age-conditioned AUROC of 0.661 ± 0.037, an AUROC of 0.810 ± 0.021, and an AUPRC of 0.343 ± 0.056. On the official hidden validation set, the model demonstrated good discrimination, achieving an age-conditioned AUROC of 0.636, an AUROC of 0.760, an AUPRC of 0.320, an accuracy of 0.065, an F1 score of 0.122, and a challenge reward of -0.299.

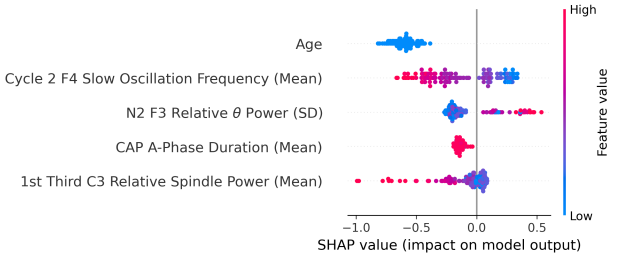

Figure 1. SHAP beeswarm plot of top features for cognitive impairment prediction. SO: slow oscillation, N2: non-REM stage 2, CAP: cyclic alternating pattern, SD: standard deviation.

Fig. 1 illustrates the SHAP beeswarm summary for the top predictors. Demographic age emerged as the primary global baseline predictor (mean absolute SHAP: 0.595). Among electrophysiological markers, the frontal SO mean frequency in sleep cycle 2 (channel F4) was the top sleep predictor (mean absolute SHAP: 0.276), where reduced SO frequency was associated with increased CI risk. Frontal relative θ power variability in N2 sleep (F3 channel standard deviation: 0.206; interquartile range: 0.100) followed, indicating disrupted spectral stability during non-REM stages.

Among sleep fragmentation markers, the mean duration of CAP A-phases ranked fourth overall (mean absolute SHAP: 0.141), followed by C3 relative spindle power during the first sleep third (0.136). Spindle power and density stability in subsequent sleep thirds and N2 sleep (third-third C3 spindle power: 0.102; C3 spindle density SD during N2: 0.095; F4 SO slope during N2: 0.081; CAP cycle duration: 0.071) further contributed to CI risk prediction.

In the cross-site robustness analysis across clinical cohorts, four of the top five features exhibited no statistically significant differences in SHAP distributions across recording sites: age ($H = 0.048, p = 0.83$), cycle 2 F4 SO frequency ($H = 2.32, p = 0.13$), F3 relative θ power standard deviation during N2 ($H = 2.96, p = 0.085$), and first-third C3 relative spindle power ($H = 2.09, p = 0.15$). Only mean CAP A-phase duration showed moderate site sensitivity ($H = 6.53, p = 0.011$, effect size $r = -0.26$).

Table 1. Model performance on 5-fold cross-validation and the official hidden validation set.

Evaluation set	Age-conditioned AUROC	AUROC	AUPRC	Accuracy	F1 score	Reward
5-fold CV (Training set, $N = 6{,}600$)	0.661 ± 0.037	0.810 ± 0.021	0.343 ± 0.056	0.380 ± 0.026	0.184 ± 0.005	$+0.143 \pm 0.184$
Hidden validation set	0.636	0.760	0.320	0.065	0.122	-0.299

4. Discussion

The prominence of cycle-specific SO and spindle metrics in our SHAP analysis aligns with established neurophysiological mechanisms of neurodegeneration. SO (< 1 Hz) and sleep spindles (12–16 Hz) during NREM sleep orchestrate memory consolidation, synaptic plasticity, and glymphatic clearance [2, 17]. Our finding that altered SO frequencies and reduced spindle power across sleep thirds strongly predict CI agrees with clinical observations linking slow-wave degradation and spindle loss to early tau and amyloid pathology [5, 6].

Furthermore, the high importance of CAP metrics (A-phase duration and cycle duration) underscores the pathological relevance of sleep fragmentation. While standard macro-architectural sleep stages provide coarse summaries, fine-grained micro-structural oscillations provide critical physiological insights for early cognitive decline screening [4].

To ensure generalizability, our two-stage feature selection and gradient-boosted tree regularization effectively constrain model complexity, as evidenced by consistent discrimination from 5-fold cross-validation (age-conditioned AUROC: 0.661) to the hidden validation set (age-conditioned AUROC: 0.636). Cross-site analysis confirmed robust cohort-invariance for four top features, with minor CAP variations reflecting scored time differences.

5. Conclusion

Our results demonstrate that combining sleep depth and sleep fragmentation features extracted from PSG indicates the potential for early screening for neurodegenerative diseases such as CI, achieving an age-conditioned AUROC of 0.636 on the official hidden validation set. These findings emphasize the clinical potential of non-invasive, PSG-derived digital biomarkers for early CI risk stratification.

References

[1] Lulu X, et al. Sleep drives metabolite clearance from the adult brain. Science 2013;342.

[2] Rasch B, Born J. About sleep's role in memory. Physiol Rev 2013;93.

[3] Ju YES, Lucey BP, Holtzman DM. Sleep and alzheimer disease pathology - a bidirectional relationship. Nat Rev Neurol 2014;10.

[4] Zhang Y, et al. Sleep in alzheimers disease: a systematic review and meta-analysis of polysomnographic findings. Transl Psychiatry 2022;12.

[5] Lucey BP, et al. Reduced non-rapid eye movement sleep is associated with tau pathology in early alzheimer's disease. Sci Transl Med 2019;11.

[6] Rauchs G, et al. Is there a link between sleep changes and memory in Alzheimer's disease? Neuroreport 2008;19.

[7] Sanchez-Espinosa MP, Atienza M, Cantero JL. Sleep deficits in mild cognitive impairment are related to increased levels of plasma amyloid-beta and cortical thinning. Neuroimage 2014;98.

[8] PyhsioNet. George B. Moody physionet challenge 2026: Screening for cognitive impairment during sleep studies, 2025.

[9] Heart Rate Variability. Task Force of the ESC and the NASPE. Circulation 1996;93(5).

[10] Welch P. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Trans On Aud And Electorac 1967;15.

[11] Younes M, et al. Odds ratio product of sleep EEG as a continuous measure of sleep state. Sleep 2015;38.

[12] Lechat B, et al. A novel electroencephalogram-derived measure of disrupted delta wave activity during sleep predicts all-cause mortality risk. Ann Am Thorac Soc 2022; 19.

[13] Wamsley EJ, et al. Reduced sleep spindles and spindle coherence in schizophrenia: Mechanisms of impaired memory consolidation? Biol Psychiatry 2012;71.

[14] Hartmann S, Baumert M. Automatic a-phase detection of cyclic alternating patterns in sleep using dynamic temporal information. IEEE Trans Neur Syst Rehabil Eng 2019;27.

[15] Akiba T, et al. Optuna: A next-generation hyperparameter optimization framework. KDD 2019;.

[16] Lundberg SM, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018;2.

[17] Fernandez LMJ, Lüthi A. Sleep spindles: Mechanisms and functions. Physiol Rev 2020;100(2):805–868.

Address for correspondence:

Richard Hohmuth
Fetscherstr. 29, 01307 Dresden, Germany
richard.hohmuth@tu-dresden.de