The Age is a Strong Baseline: But Is There Anything Beyond?

Petr Nejedly, Pavel Jurak and Filip Plesinger
[1] Institute of Scientific Instruments of the Czech Academy of Sciences, v. v. i.

Abstract

We describe a two-stage pipeline submitted to the George B. Moody PhysioNet Challenge 2026 by team ISIBrno-AIMT, which predicts, from a single overnight polysomnography (PSG) recording, whether a patient will receive a cognitive impairment diagnosis 1–6 years later. Stage 1 is a three-layer Transformer sleep staging model producing hypnograms. Stage 2 extracts 22 sleep stage conditioned EEG and oximetry features — spectral power ratios, spindle density, slow-oscillation–spindle coupling, and coherence — and combines them with patient age in a random forest classifier. We report the complete architecture, preprocessing, and feature definitions as implemented. Our official-phase entry (submission ID 2410) was evaluated, yielding an age-conditioned AUROC of 0.554, an unconditioned AUROC of 0.770, an AUPRC of 0.257, an accuracy of 0.911, and an F-measure of 0.259, achieving XXXth rank in the challenge.

1. Introduction

Cognitive impairment from Alzheimer's disease (AD) and related dementias is usually diagnosed after functional change is reported, by which point substantial neurodegeneration has occurred. Sleep is a candidate early-signal domain: a meta-analysis of PSG studies found reduced slow-wave and REM sleep in AD, correlating with impairment severity [1]. At the microstructure level, reduced spindle density and altered spindle characteristics have been reported in mild cognitive impairment (MCI) and early AD and linked to subsequent decline [2, 3]; a case-control study with one-year follow-up found reduced slow-wave/theta/sigma activity and altered spindles preceded early cognitive impairment. Slow-oscillation–spindle coupling, implicated in sleep-dependent memory consolidation [4], has been reported to weaken from cognitively normal to MCI to AD, and spindle/slow-oscillation measures have predicted cognition and neurodegeneration biomarkers in AD cohorts [5]. Separately, there is also a link between apnea–hypopnea causing lower nocturnal SpO2 to steeper cognitive decline, with obstructive sleep apnea associated with increased risk of cognitive impairment.

These findings motivate our solution to The George B. Moody PhysioNet Challenge 2026: "Screening for Cognitive Impairment During Sleep Studies" [6] propose a feature-based architecture utilizing a Transformer-based sleep stager coupled with feature extraction of 22 EEG/oximetry biomarkers combined with age in a random forest classifier predicting cognitive decline.

2. Methods

Our pipeline consists of two stages, i.e., sleep staging and sleep staging-conditioned feature extraction.

2.1. Stage 1: Sleep stage classifier

Per EEG channel: Welch PSD (10 s FFT segment, 50% overlap, Hann window, constant detrend, one-sided spectrum), averaged into 20 s windows with 10 s overlap; restricted to 0.5–30 Hz, column-normalized to relative power, padded/truncated to 296 frequency bins were extracted. Additionally, four scalars are appended: chin-EMG energy (filtered with 10 Hz high-pass), bipolar EOG (E1-E2) energy (filtered in range 0.5–10 Hz), and Pan–Tompkins-derived heart-rate mean/SD per 20 s. Random windows with length of 6 hours (2160 timesteps) were sampled for training of 3-layer Transformer predicting 5-stage hypnogram (Wake/N1/N2/N3/REM). The model was trained using AdamW optimizer with learning rate of 5e-4, and weight decay of 1e-2.

2.2. Stage 2: Feature extraction

Stage 2 extracts 22 stage-conditioned EEG/oximetry features plus age (23 inputs) into a class-balanced random forest. The features were selected based on an internal selection process based on a literature search of state-of-the-art features.

Prior to stage-conditioned feature extraction, each channel is passed through an institution-invariant preprocessing, applied independently of the Stage-1 sleep-staging network (which operates on the raw, uncanonicalized signal). First, every channel is resampled to a common sampling rate of 200 Hz using polyphase resampling. Second, a 4th-order Butterworth high-pass filter with a cutoff of 0.5 Hz is applied using zero-phase filtering. Third, each

channel is standardized independently via per-channel z–scoring. Oximetry channels (SpO_2/SaO_2/O_2 saturation) are resampled and z-scored in the same manner but are left unfiltered, since the high-pass step is not appropriate for a slowly varying physiological signal of this kind. The 200 Hz sampling rate was heuristically selected given variable sampling rates between recordings acquired at 200 Hz and 500 Hz. Following preprocessing, 22 stage-conditioned features (Table 1) are computed on 30-s epochs across six scalp EEG channels (F3, F4, C3, C4, O1, O2), with each feature restricted to epochs assigned the indicated sleep stage(s) by the Stage-1 hypnogram. Feature values that evaluate to `NaN` are replaced by zero.

Table 1. Stage-conditioned features (22 total).

Feature	**Sleep Stage(s)**	**Channel(s)**
Kurtosis	N2	F3
Relative theta	REM	O1
Relative theta	REM	F3
Relative theta	N3	C3
Relative theta	N1	F3
Coherence	N3	F3–C3
Coherence	N3	C3–C4
Coherence	N2	F3–F4
Hjorth complexity	N1–N3 (sleep)	O1
Hjorth complexity	N1–N3 (sleep)	C3
Spectral entropy	N1–N3 (sleep)	O2
Relative sigma	N2	C3
Relative fast-sigma	N2	C3
Relative delta	N3	F3
Delta/alpha	N3	F3
Delta/theta	N3	C3
Relative alpha	REM	O1
Spindle density	N2	C3
Spindle density	N2	F3
Spindle density (fast)	N2	C3
SO–spindle coupling	NREM	C3
Mean SpO_2	all	oximetry

Table 1 lists the 22 stage-conditioned features together with the sleep stage(s) and channel(s) over which each is computed; the definitions below describe how each feature is calculated and refer to the table for the specific stage/channel assignment in each row.

Kurtosis is computed as the epoch kurtosis of the canonicalized signal.

Relative theta power is defined as the ratio of 4–8 Hz power to total 0.5–30 Hz power.

Coherence is defined as the 0.5–4 Hz Welch coherence between a pair of channels.

Hjorth complexity is computed over 30-s sleep epochs (i.e., all epochs scored as N1, N2, N3, or REM, excluding Wake).

Spectral entropy is computed as the entropy of the 0.5–45 Hz spectrum over the same sleep-epoch pool (N1, N2, N3, REM).

Relative sigma (11–16 Hz) and relative fast-sigma (13–16 Hz) power are each expressed as a fraction of total power.

Relative delta power (0.5–4 Hz fraction of total), the delta/alpha ratio (0.5–4 Hz to 8–12 Hz power), and the delta/theta ratio (0.5–4 Hz to 4–8 Hz power) are each expressed as indicated.

Relative alpha power is defined as the 8–12 Hz fraction of total power.

Spindle density is expressed in events per minute and is detected via a 4th-order Butterworth band-pass filter, Hilbert envelope extraction, 100 ms smoothing, and thresholding at the mean plus two standard deviations, with a 0.5–3 s event-duration constraint.

Slow-oscillation–spindle coupling is quantified as the mean vector length (MVL) of phase–amplitude coupling between the slow-oscillation band (0.5–1.25 Hz) and the spindle band (11–16 Hz).

Mean SpO_2 is the recording-average oxygen saturation across all epochs, expressed as a percentage and clipped to the 75–100% range to exclude artifactual readings.

2.3. Stage 2: Cognitive Impairment Classifier

The classifier operates on 23 inputs: patient age together with the 22 stage-conditioned features. These inputs are first standardized to zero mean and unit variance, and the standardized inputs are then passed to a random forest classifier composed of 600 decision trees, with a minimum of 10 samples required at each leaf node and class weights balanced inversely to class frequency to compensate for the low prevalence of the positive (cognitive-impairment) class. This classifier is the only component trained during model development; the sleep-staging Transformer remains frozen (pretrained locally).

At inference time, the preprocessed signals are used to extract the 22 features, which are combined with age and passed through the fitted standardization transform. The classifier then outputs the predicted probability that the patient will receive a cognitive-impairment diagnosis within 1–6 years of the recording. This probability is converted to a binary prediction using a fixed threshold of 0.5.

2.4. Evaluation Metrics

Model performance is assessed on a cohort restricted to positive patients (two or more cognitive-impairment-related diagnoses, the first occurring one to six years after the sleep study, at least one week apart) and negative patients (no such diagnosis and at least six years of

follow-up); patients meeting neither definition cleanly are excluded. Four metrics are reported for reference — overall AUROC, AUPRC, accuracy, and F-measure — computed over the full cohort without any age adjustment, together with a prevalence-based reward score that weights correct and incorrect predictions against an age-matched, training-set-derived prevalence estimate; none of these determines the Challenge's official ranking.

The metric that determines the Challenge ranking is an age-conditioned AUROC, computed only over pairs of a positive and a negative patient whose ages differ by no more than two years, and defined as the probability that the model assigns a higher predicted risk to the positive patient than to the negative patient across all such age-matched pairs. Restricting every comparison to patients of approximately the same age makes the metric deliberately independent of age as a predictive signal: a model that captured only chronological age, without any additional information from the sleep study, would score at chance level here even if it achieved a high unconditioned AUROC over the full cohort, since age carries no discriminating power within a narrow age-matched pair. This construction isolates the predictive contribution of the PSG-derived signal from the contribution of age alone, which would otherwise dominate a pooled evaluation given the strong association between age and cognitive-impairment prevalence.

3. Results

Our submitted entry ID 2410 team ISIBrno-AIMT for the official phase achieved the following validation results.

Table 2. Official-phase validation results for submission ID 2410

Metric	Validation
Reward	0.034
Age-conditioned AUROC	0.554
Age-weighted AUROC	0.514
AUROC	0.770
AUPRC	0.257
Accuracy	0.911
F-measure	0.259

4. Discussion

Our unofficial-phase finding — that a simple classifier defined as "y=age/100", with no PSG-derived input, ranked first among unofficial-phase entries — highlighted a limitation in the original scoring function rather than any genuine predictive capability. Because cognitive-impairment prevalence rises with age in the training population, a metric that pools comparisons across the full age range rewards any model that recovers this age–prevalence

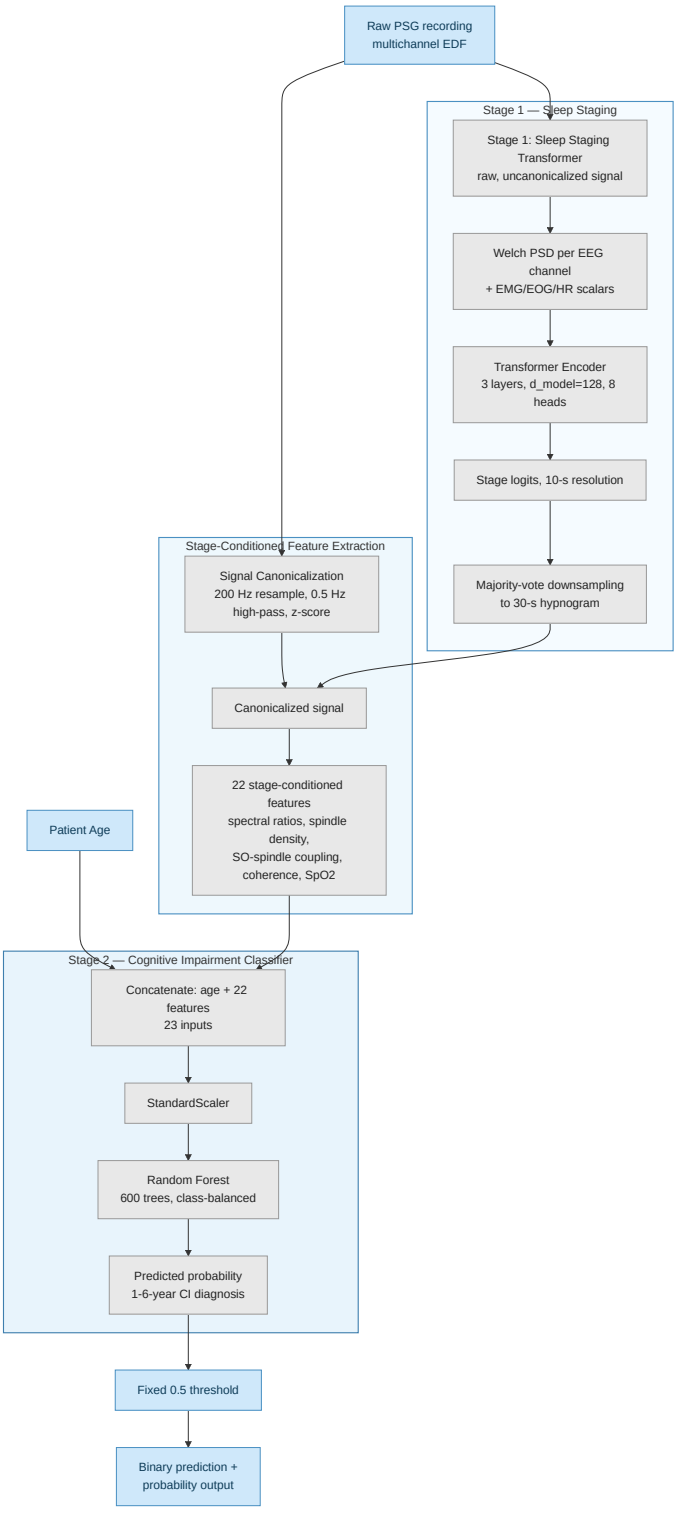

Figure 1. Proposed pipeline consisting of two stage classifiers: Stage-1 Sleep staging; Stage-2 Cognitive impairment classifier

relationship, independently of whether it extracts information from the sleep recording itself. By restricting every pairwise comparison to patients within a two-year age band, this revised metric is constructed so that a model relying on age alone cannot be distinguished from a model making uninformative predictions, since age carries no discriminating information within an age-matched pair — directly addressing the failure mode our unofficial-phase submission exposed.

Our official-phase result indicates that we did not adequately address this issue in the design of our own system. Despite many trial-and-error internal evaluations, our predictions remained substantially correlated with age rather than isolating an age-independent signal of cognitive impairment. This gap is visible directly in the discrepancy between the unconditioned AUROC (0.770) and the age-conditioned AUROC (0.554). The high accuracy (0.911) reinforces this reading rather than contradicting it: given the low prevalence of cognitive impairment in the validation cohort (5–15%, per the organizers), a classifier that defaults toward the majority (negative) class achieves high accuracy essentially independently of whether it has learned any impairment-specific signal at all. The comparatively low AUPRC (0.257) and F-measure (0.259) are more informative in this low-prevalence setting and are consistent with limited positive-class discrimination once the age confound and class imbalance are both accounted for.

We do not interpret this result as evidence that PSG-derived sleep features carry no information about cognitive-impairment risk; rather, it indicates that our presented solution did not succeed in isolating an age-independent component from the selected feature set.

5. Acknowledgments

This research was supported by the Czech Academy of Sciences (RVO68081731).

References

[1] Zhang Y, Ren R, Yang L, Zhang H, Shi Y, Okhravi HR, Vitiello MV, Sanford LD, Tang X. Sleep in Alzheimer's disease: a systematic review and meta-analysis of polysomnographic findings. Translational Psychiatry 2022;12:136.

[2] Gorgoni M, Lauri G, Truglia I, et al. Parietal fast sleep spindle density decrease in Alzheimer's disease and amnesic mild cognitive impairment. Neural Plasticity 2016; 2016:8376108.

[3] Winer JR, Mander BA, Helfrich RF, et al. Sleep as a potential biomarker of tau and β-amyloid burden in the human brain. Journal of Neuroscience 2019;39(32):6315–6324.

[4] Staresina BP, Niediek J, Borger V, Surges R, Mormann F. How coupled slow oscillations, spindles and ripples coordinate neuronal processing and communication during human sleep. Nature Neuroscience 2023;26(8):1429–1437.

[5] Páez A, Gillman SO, Dogaheh SB, Carnes A, Dakterzada F, Barbé F, et al. Sleep spindles and slow oscillations predict cognition and biomarkers of neurodegeneration in mild to moderate Alzheimer's disease. Alzheimers Dementia 2025; 21(2):e14424.

[6] PhysioNet. Screening for cognitive impairment during sleep studies: The George B. Moody PhysioNet Challenge 2026. `https://moody-challenge.physionet.org/2026/`, 2026. Accessed 2026.

Address for correspondence:

Petr Nejedly, Ph.D.

Institute of Scientific Instruments of the CAS, v. v. i.
Královopolská 147
612 00 Brno
Czech Republic

nejedly@isibrno.cz