Deep Learning Sleep Staging from ECG with Robust Automated Preprocessing : Toward Deployment in TBI and MRI environments

Angeliki Pollatou

Uniformed Services University · Walter Reed National Military Medical Center · The Geneva Foundation

Current Affiliation: Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc. (HJF), in support of the Military Traumatic Brain Injury Initiative (MTBI²), Uniformed Services University of the Health Sciences (USUHS), and the National Intrepid Center of Excellence

DISCLOSURE

Disclosure & conflict of interest

Presenter: **Angeliki Pollatou, Ph.D.**

The opinions and assertions expressed herein are those of the author(s) and do not reflect the official policy or position of the Uniformed Services University of the Health Sciences or the Department of War.

Neither I nor my family members have a financial interest in any commercial product, service,or organization providing financial support for this research.

Research support: ARPA-H. This talk describes investigational methods not cleared for clinical use.

LEARNING OBJECTIVES

This talk, in three objectives

01

Robust, unattended preprocessing

How automated ECG polarity detection and multi-level artifact rejection enable sleep staging without manual quality review, including in MRI environments where gradient artifacts corrupt cardiac signals.

02

Balanced sampling that generalizes

How balanced epoch sampling resolves minority-class collapse in imbalanced sleep staging datasets and outperforms standard loss-weighting approaches across independent, large-scale PSG cohorts.

03

An EEG-free sleep staging path

ECG-based automated sleep staging as an accessible alternative for TBI and EEG-intolerant military populations, and its planned deployment in MRI studies of glymphatic function across study cohorts.

WHY IT MATTERS

Why image the brain during sleep?

During sleep the brain clears metabolic waste through the **glymphatic system**, consolidates memory, and regulates emotion. These are active, stage-specific processes visible only by imaging the sleeping brain.

DISRUPTED IN THE SIGNATURE CONDITIONS OF MILITARY BRAIN HEALTH

Traumatic brain injury

Fragments sleep and blunts slow-wave activity -> poorer sleep predicts worse recovery.

PTSD

Disrupts REM sleep, with nightmares and hyperarousal -> sleep is both a symptom and a driver.

Depression

Shifts REM timing and slow-wave sleep -> among the most reliable physiological markers of mood.

Imaging brain function across sleep stages reveals the mechanisms and biomarkers behind injury and recovery.

Pollatou et al. — Mapping Human Sleep-State Brain Function with Simultaneous EEG & Neuroimaging: A Systematic Review. *NeuroImage (2026)* · *doi.org/10.1016/j.neuroimage.2026.122122*

THE HIDDEN CONFOUND

‘Resting-state’ isn’t always rest

In a typical resting-state scan, undetected sleep quietly distorts connectivity and BOLD.

~1 in 3

lose wakefulness within the first 3 minutes*

Either way, studying sleep or not, you need to know the sleep-wake state during the scan.

**Tagliazucchi & Laufs, Neuron (2014) · 1,147 resting-state fMRI datasets.*

THE GOLD STANDARD

EEG is the gold standard for sleep staging

But it is challenging to use in the MRI environment

Polysomnography with EEG is the reference for sleep staging, but bringing all of it into the scanner is punishing on every level:

Uncomfortable

A dense electrode cap makes it hard to fall asleep, compounded by relentless scanner noise in the bore.

Expensive

MR-conditional EEG hardware is costly, and readily available in very few sites.

Time consuming

Specialized training is required and lengthy setup can add real time to every MRI session.

Rigorous artifact removal required

Gradient and pulse artifacts on the EEG need correction before scoring.

After almost three decades of sleep neuroimaging and 145 EEG-fMRI simultaneous studies, study median is N = 18 and mostly healthy adults.

HOW OTHERS DO IT

Sleep stages can be estimated from many signals

Automated staging draws on signals from the brain, eyes, muscles, heart, breath, and movement:

BRAIN & EYES & MUSCLE · PSG

EEG · EOG · EMG

CARDIORESPIRATORY

ECG · Respiration · SpO_2 / PPG

MOVEMENT

Actigraphy

In the scanner, additional electrodes and sensor belts are impractical, but one signal is already there ↓

ECG — unobtrusive, already acquired in most MRI protocols, and shown to carry sleep-stage information.

(or PPG — heart rate can be estimated from both)

Sleep-staging review: Liu et al. (2024) · Automatic sleep stage classification using deep learning: signals, data representation and neural networks

OUR APPROACH

Turning the heartbeat into sleep stages

Four approaches, differing mainly in how much you hand the network:

Cardiopulmonary Coupling (CPC) · *spectral heart-rate–respiration coupling; traditional method*

Features + ML

Hand-engineered cardiac features – HRV (time, frequency, nonlinear) plus rate & respiration measures -> SVM / RF / boosting.

Deep learning · raw waveform

End-to-end from the raw ECG / PPG waveform — powerful, but heavy and bakes in device quirks.

Deep learning · heart rate (IHR)

End-to-end on the physiological signal itself, modality-agnostic (ECG or PPG), at 30-second epochs.

→ **OUR PATH**

The novelty isn't the network, it's the preprocessing and balanced training that make it work in an MRI environment even in shorter recordings, not just long overnight sessions.

ARCHITECTURE

How the network reads a full night

STAGE 1 · LOCAL

Centers a ~2-min window on every 30-s epoch and learns local heart-rhythm shapes.

STAGE 2 · LONG-RANGE

Dilated convolutions widen the view to ~90 min, a typical sleep cycle, so each epoch is read in the context of the whole night.

IN heart rate · ~10 h @ 2 Hz
OUT a sleep stage every 30 s

OUR CONTRIBUTION

What we added: robust preprocessing + balanced training

01

Automated polarity detection

Detects and corrects inverted ECG leads before R-peak detection, with no human in the loop.

02

Automated R-peak correction

Physiologically implausible and artifact-corrupted beats are filtered out before heart rate is computed.

03

Multi-level artifact rejection

Epoch- and recording-level gates tuned for MRI gradient and magnetohydrodynamic corruption.

Zero manual review *— the same pipeline ran unattended across MESA, SHHS, and our Sleep Lab, despite different hardware and sampling rates.*

AND IN TRAINING **Balanced epoch sampling** — rare stages are given equal weight during learning, so the model never gives up on them.

RESULTS

Strong agreement with expert scoring

Held-out test performance : SHHS (N = 7,985 recordings), cardiac signal only.

WAKE / SLEEP

87.7% acc

κ 0.668

recall Wake 0.75 · Sleep 0.92

WAKE / NREM / REM

82.1% acc

κ 0.683

recall Wake 0.77 · NREM 0.85 · REM 0.78

Trained separately on MESA (N = 1,870), the same balanced configuration won again.
The method ranking is preserved across two independently collected cohorts.
So the gain is not a tuning artifact of one dataset.

THE HARDEST STAGE

What about deep sleep?

Deep sleep is a small fraction of the night. The standard fix for class imbalance barely predicts it, and it stays the hardest stage to predict from the heart rate.

FOCAL LOSS

the standard remedy for class imbalance

	recall / precision
Wake	0.54 / 0.69
Light	0.85 / 0.59
Deep	**0.10 / 0.48**
REM	**0.36 / 0.51**

κ = 0.35 → survives by over-calling light sleep.

balanced sampling

BALANCED SAMPLING

ours

	recall / precision
Wake	0.78 / 0.73
Light	0.66 / 0.77
Deep	**0.60 / 0.52**
REM	**0.81 / 0.67**

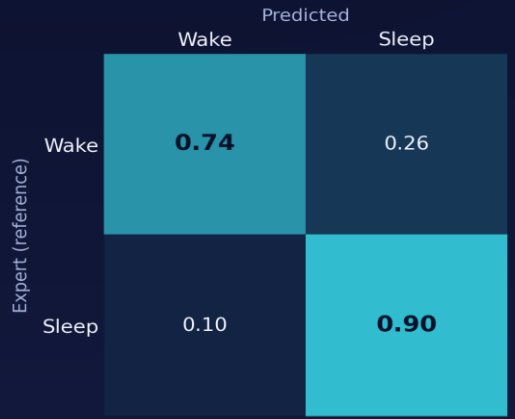
κ = 0.57 → all four stages usable.

Focal loss leans on light sleep (recall 0.85, precision 0.59) and neglects the rest; balanced sampling rebalances: Wake 0.54→0.78, Deep 0.10→0.60, REM 0.36→0.81, and light-sleep precision climbs to 0.77. ***Even so, deep sleep stays the weakest stage, so we target the two- and three-class tasks the cardiac signal handles reliably.***

INDEPENDENT VALIDATION

It transfers cleanly to an unseen cohort

56 recordings · 45,451 epochs · a new site, human scorer, and patient population never seen in training



Wake / Sleep **κ = 0.55**

Wake / NREM / REM **κ = 0.62**

3-class κ = 0.62

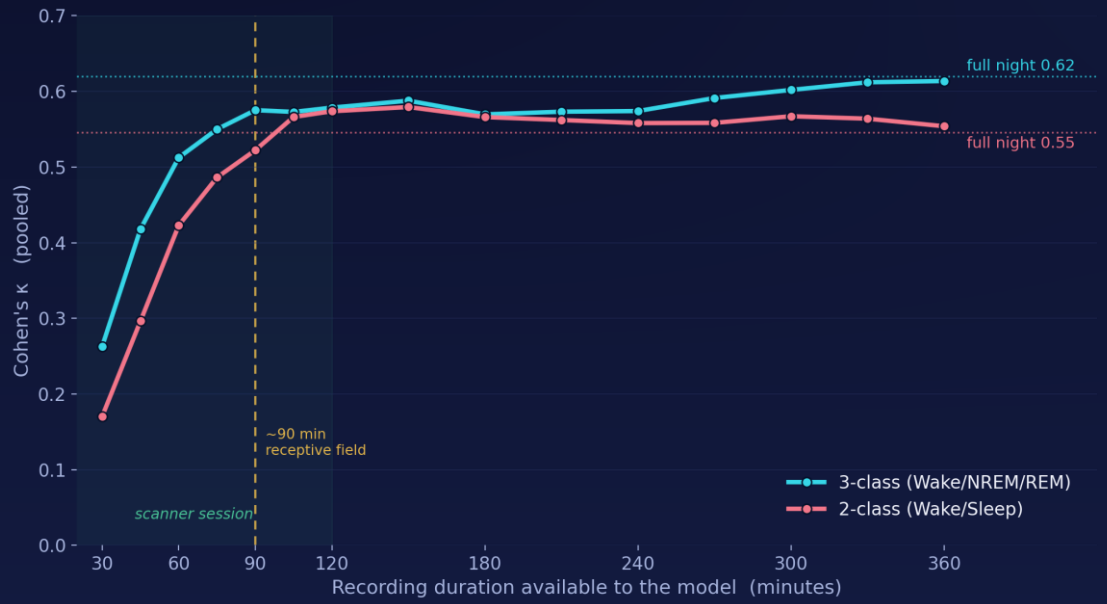
lands between the model's own MESA (0.61) and SHHS (0.68), effectively no loss on external data.

REM recall holds at **0.77**

THE SCANNER QUESTION

Nobody sleeps eight hours in a scanner

We gave the model only the first X minutes of each recording and asked how little of the night it really needs (pooled κ).



WAKE / SLEEP

2 hours is enough

κ 0.57 at 120 min, matching a full night (0.55).

FULL 3-CLASS STAGING

Nearly there at 2 h

κ 0.58 at 120 min vs 0.62 for a whole night.

A two-hour scanner session is enough for 2-classes and nearly enough for 3-classes

IN-SCANNER USE

Validated three times, then run inside the scanner

MESA

N = 1,870

Trained & tested · PSG ground truth

SHHS

N = 7,985

Trained & tested at scale · PSG ground truth

SLEEP LAB

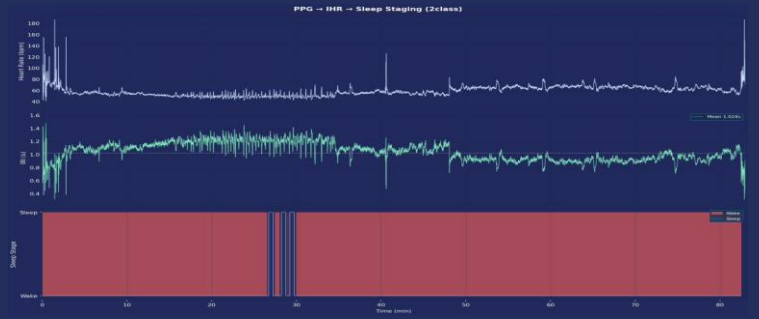
N = 56

Unseen cohort , transferred without retraining. Compared with manual PSG scoring

IN THE SCANNER

A hypnogram from inside the MRI

The same pipeline, unchanged, running end-to-end on a signal acquired concurrently with the scan predicts sleep-wake -> **no EEG, no electrode cap, no additional time to prep the participant.**



TAKEAWAYS

Sleep staging without EEG

Accurate, robust, and it works where EEG can’t

01 **It works without EEG**

The cardiac signal alone reaches strong agreement with expert scoring.

02 **It is robust and automated**

Zero manual review across three independent cohorts (MESA, SHHS, Sleep Lab), transferring without retraining.

03 **A scanner session is enough**

Two hours of recording gives full-night-quality Wake/Sleep.

ACKNOWLEDGEMENTS

Acknowledgements

Co-authors

Nastaren Abad (PI) · Isabelle Heukensfeldt Jansen · Luca Marinelli *GE HealthCare, Technology & Innovation Center*

Kent Werner *Uniformed Services University · WRNMMC · NICoE · MTBI²*

Elizabeth Metzger *The Geneva Foundation · Uniformed Services University*

Funding & collaboration

ARPA-H (award # H37810000728898) · in collaboration with GE HealthCare, Technology & Innovation Center

APPENDIX · BENCHMARK DETAIL

Held-out test performance — MESA & SHHS

Dataset · Task (best config)	Acc	κ	Per-class recall / precision
MESA · Wake / Sleep	83.1%	**0.602**	Wake 0.70 / 0.75 Sleep 0.89 / 0.86
MESA · Wake / NREM / REM	77.7%	**0.610**	Wake 0.73/0.74 NREM 0.83/0.83 REM 0.66/0.63
SHHS · Wake / Sleep	87.7%	**0.668**	Wake 0.75 / 0.75 Sleep 0.92 / 0.92
SHHS · Wake / NREM / REM	82.1%	**0.683**	Wake 0.77/0.76 NREM 0.85/0.89 REM 0.78/0.70

Sampling-rate robustness: 125 Hz (SHHS1) vs 250 Hz (SHHS2) showed no significant per-night difference for 3-class; one isolated 2-class case (88.0% vs 86.9%, p=0.015) with κ not significant — consistent with cohort effects, not sampling rate (IHR is resampled to 2 Hz upstream).

Best configuration per task (balanced sampling). Quality filtering: MESA 2,033→1,870; SHHS 8,444→7,985. MESA: AASM, 256 Hz. SHHS: R&K, 125–250 Hz. Each model trained on a single dataset.