

# Prospective Clinical Performance Validation of AI for PPG-based Sleep Staging

Chris Fernandez, MS\*, Yoav Nygate, MS\*, Sam Rusk, BS\*, Matt Sprague, MS\*, Tom Vanasse, PhD\*, Jan Wodnicki\*, Nick Glattard, MS\*, Fred Turkington\*, Kelsey Buehl\*, Shahnawaz Khan, RPSGT, CCSH\*, Andrea Ramberg, MS, RPSGT, CCSH\*, Justin Mortara, PhD\*, Nathaniel F. Watson, MD, MSc\*\*

\* EnsoData Research, Ensodata, Madison, WI, USA \*\* Department of Neurology, University of Washington (UW) School of Medicine and UW Medicine Sleep Center



## Introduction

Photoplethysmography (PPG) is the basis for both the pulse rate and oximetry during polysomnography (PSG) and Home Sleep Apnea Tests (HSAT). In recent years, with the popularization of compact and wearable health tracker devices, the PPG became an integral part of continuous measurements for the most widely adopted clinical and consumer health technologies.

High quality automated Sleep Staging analysis is crucial for both clinical and consumer wearable technologies and may assist with large scale population screening and diagnosis of sleep disorders. In this study, an interoperable AI system was designed to perform automated Sleep Staging of single-channel PPG signal data and validated on a gold standard cohort of simultaneously recorded single-channel PPG and PSG studies.

## Methodology

An AI model was trained utilizing a transfer learning-inspired approach, by applying machine learning and statistical signal processing methods, including multiple deep neural network models, to a database of over 1,000,000 diagnostic PSGs with concurrently recorded PPG.

Clinical performance validation was conducted on the AI system in an IRB approved study using a prospective, non-randomized trial design with all-comers enrollment offered to subjects undergoing a routine PSG. The study utilized FDA cleared PSG systems, Philips Respironics Sleepware G3, Natus Sandman Elite, and Polysmith Sleep System, to collect PSG studies for establishing the gold-standard comparator data. Simultaneously, PPG signals were recorded utilizing an FDA cleared single-channel PPG device, Viatom Checkme O2, to collect wearable patient data for establishing the primary validation endpoints to evaluate the AI system's performance on

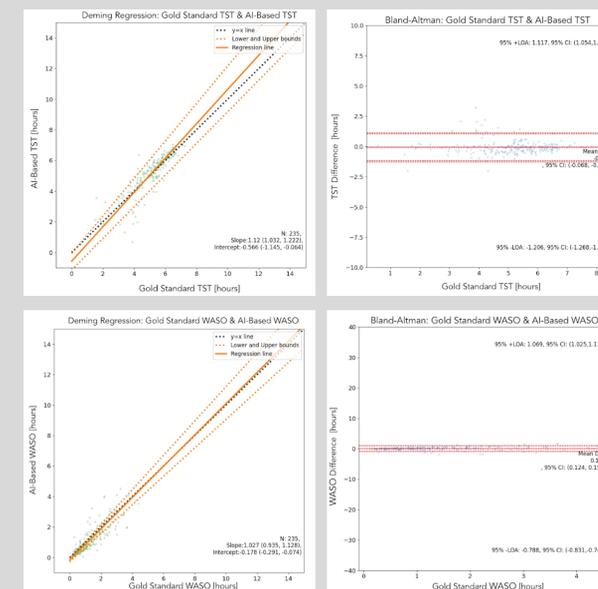
the analysis of single-channel PPG data for Sleep Staging. The study sample included N=235 subjects enrolled with informed consent, who completed PSG studies with simultaneously recorded PPG signals using wearable single-channel PPG devices, and had >4-hours of adequate data. Demographics including Age, Sex, Skin Pigmentation, BMI, ESS, confounding conditions and medications, and OSA severity were reported. The gold-standard comparative benchmark was collected by constructing a 2/3 majority scoring panel (MSP) utilizing 3 Registered Polysomnographic Technologist (RPSGT). During scoring, each of the 3 clinicians independently applied the standards of practice defined in the AASM Manual for the Scoring of Sleep and Associated Events and scored the PSG sleep test to completion. The gold-standard sleep staging was then constructed by taking the majority scored sleep stage in each 30 second epoch (i.e. the sleep stage agreed upon by at least 2/3 RPSGTs in each epoch). Taking into account inter-scorer reliability and the fidelity of the PPG signal, the five commonly used sleep stages were reduced to Wake, Light Non-REM (N1 + N2), Deep Non-REM (N3), and REM, as commonly acceptable for wearable sleep tracking devices. Performance was evaluated utilizing epoch-by-epoch sleep staging agreement. Sensitivity and Specificity were calculated for every sleep stage by aggregating the epochs and comparing each epoch's AI generated sleep stage label with its concurrent gold-standard label.

Furthermore, to evaluate the performance of the AI system's agreement with continuous sleep indices commonly included in patient sleep reports - Total Sleep Time (TST), Sleep Efficiency (SE), Sleep Latency (SL), and Wake After Sleep Onset (WASO), Deming Regression and Bland-Altman analysis was used to evaluate the level of agreement between the AI system's calculated index and the gold-standard index. Two-sided 95% median bootstrap percentile method confidence intervals (R=2,000) were calculated for all Sensitivity, Specificity, Bland-Altman, and Deming Regression performance measures.

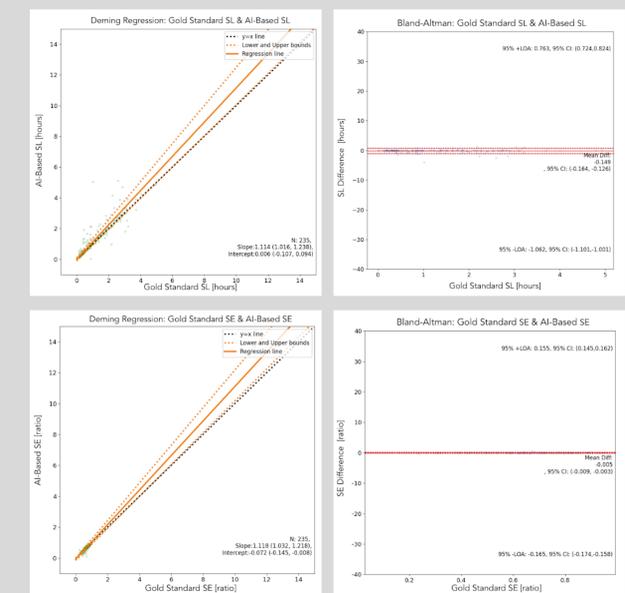
## Results

In this study, the performance of an AI-based sleep staging system using single-channel PPG was validated against gold-standard PSG data. The table below shows that PPG-based AI sleep staging demonstrated epoch-by-epoch agreement with a sensitivity (PPA) and specificity (NPA) of 84.2%/97.5% for stage REM, 80.7%/86.7% for Light Non-REM (N1/N2), 67.9%/95.5% for Deep Non-REM (N3), and 87.8%/93.7% Wake compared to PSG sleep staging. In the figures, the Bland-Altman results for PPG-based AI sleep staging demonstrated high agreement in sleep quality measures with an average difference of -2.7 minutes (95% CI: -4.08, -1.44) in TST, -0.50% (-0.90%, -0.30%) in SE, -8.94 minutes (-9.84, -7.56) in SL, and 8.46 minutes (7.44, 9.48) in WASO, along with high correlation for all sleep measures compared to gold-standard PSG as seen in the Deming Regression results.

Category	n	Sensitivity	Specificity
Wake	74,850	87.8% (87.6%, 88.0%)	93.7% (93.5%, 93.8%)
Light Non-REM	107,325	80.7% (80.5%, 80.9%)	86.7% (86.5%, 86.9%)
Deep Non-REM	14,166	67.9% (67.1%, 68.7%)	95.5% (95.5%, 95.6%)
REM	21,934	84.2% (83.7%, 84.7%)	97.5% (97.5%, 97.6%)
Total	218,275	n/a	n/a



## Results Continued



## Conclusions

The AI-based sleep staging system demonstrated high sensitivity and specificity across all sleep stages when compared to the gold-standard PSG.

The Deming regression plots indicate strong correlations for TST, WASO, SL, and SE between the AI system and the gold-standard.

Bland-Altman analyses show small mean differences and narrow limits of agreement, further validating the accuracy of the AI-based system for sleep staging.

These results suggest that the AI system can be effectively used for automated sleep staging using single-channel PPG data in both clinical and consumer health technologies, thus expanding opportunities for multi-night diagnostic testing, remote longitudinal OSA therapy monitoring, and the utility of consumer sleep technologies to promote overall sleep wellness.