Clinical Validation of ECG-Based Obstructive Sleep Apnea Screening Using Machine Learning

Yoav Nygate, MSc*, Matt Sprague, MSc*, Sam Rusk, BSc*, Chris Fernandez, MSc*, Nathaniel F. Watson, MD, MSc**

* EnsoData Research, Ensodata, Madison, WI, USA
** Department of Neurology, University of Washington School of Medicine, Seattle, WA

UW Medicine
DEPARTMENT OF NEUROLOGY

ensodata research

Introduction

Obstructive sleep apnea (OSA) is a prevalent sleep-disordered breathing condition affecting a substantial proportion of adults worldwide. However, due to the burdensome nature of polysomnography (PSG) – the gold-standard diagnostic test – and low awareness, the majority of OSA cases remain undiagnosed.

Single-lead electrocardiogram (ECG) has emerged as a promising avenue for OSA screening. Apneic events induce characteristic fluctuations in heart rhythm and other ECG changes such as R-wave amplitude variations due to intrathoracic pressure swings. Prior studies have suggested that automated detection of such ECG signatures can serve as a useful marker for sleep apnea.

There is a substantial overlap between the population undergoing cardiovascular diagnostic testing and those with undiagnosed OSA. Many patients referred for ambulatory ECG monitoring or Holter testing due to suspected arrhythmias or unexplained fatigue may, in fact, be suffering from comorbid OSA. These patients are typically equipped with single-lead ECG devices for multi-night monitoring, creating a unique window to also screen for OSA without any additional burden.

This provides the opportunity to identify cardiovascular patients who are at risk for OSA and refer them for further testing and treatment. This approach has the potential to significantly enhance diagnostic throughput for OSA and improve care for a high-risk, underserved population.

Methodology

A Machine Learning (ML) system was developed utilizing over 100,000 diagnostic PSG studies with concurrently recorded ECG signals. The system leveraged multiple deep learning models to automatically learn respiratory event patterns, as well as sleep-stage-specific patterns from the ECG. The deep learning models takes as input sequences of the ECG signal and outputs the probability of respiratory events within a window, as well as categorical probabilities for specific sleep stages.

Methodology Continued

Clinical validation was performed on a dataset of 217 subjects from a prospective clinical study. PSG results were scored by three RPSGTs. The American Academy of Sleep Medicine (AASM) recommends utilizing the 3% hypopnea (1A) scoring rule as the current standard of care, with the 4% hypopnea (1B) scoring rule as optional for clinical reporting. Thus, respiratory events were scored using 3% AHI criteria according to AASM scoring manual specifications. For an event to be officially scored, a minimum two-thirds consensus among scorers was required. Each PSG was reviewed by a board-certified sleep physician to confirm PSG data quality, scoring accuracy, and technical adequacy.

The ML system's performance was evaluated against the gold-standard OSA diagnosis using an AHI threshold of 15 events per hour to identify positive OSA cases. To assess ECG-based sleep staging performance, sleep stages were reduced to Wake, Light Sleep (N1 + N2), Deep Sleep (N3), and REM, and agreement was evaluated utilizing an epoch-by-epoch approach. Furthermore, we benchmarked the performance against a previously developed PPG-based ML system trained to perform the same tasks.

Figure 1. Bland Altman and Deming Regression analysis for ECG-based and PPG-based AHI determination

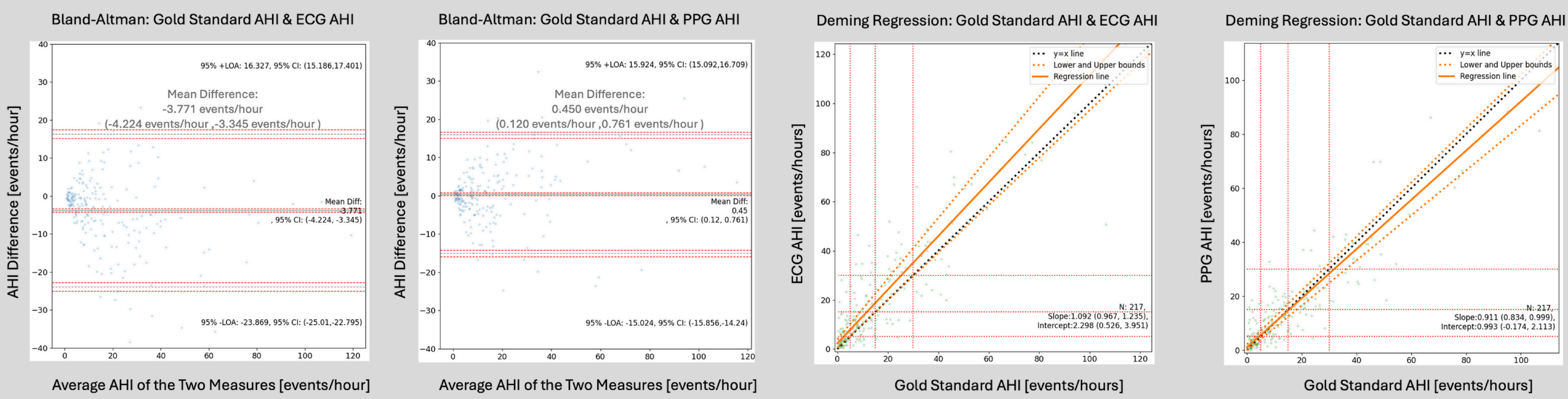
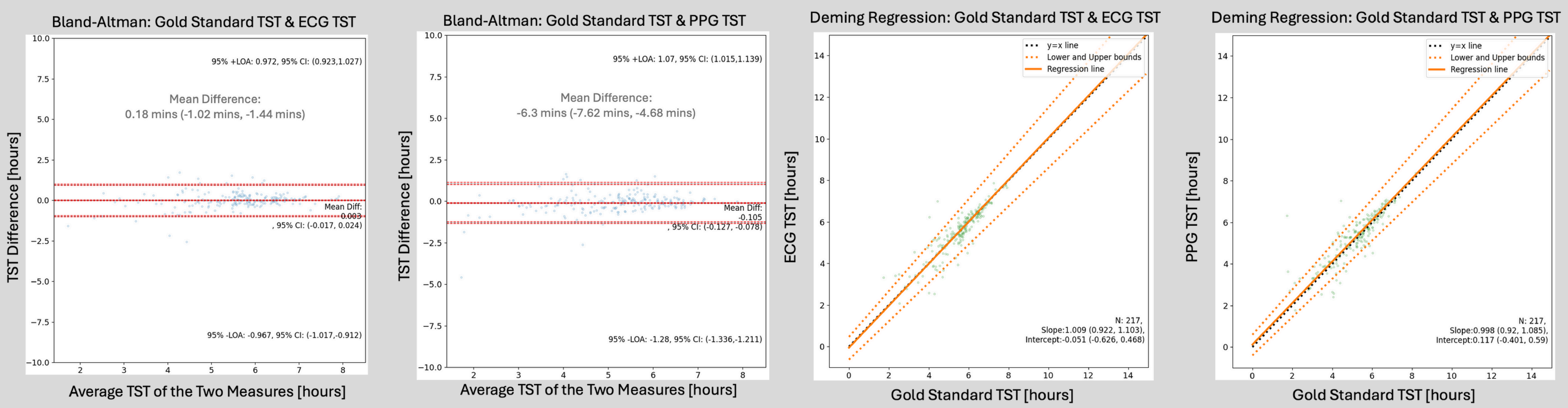


Figure 2. Bland Altman and Deming Regression analysis for ECG-based and PPG-based TST determination



Results

Table 1. Performance of ECG-based and PPG-based Moderate-Severe OSA diagnosis.

	Sample Size (N=225)		Sensitivity (%, 95% CI)		Specificity (%, 95% CI)	
	AHI ≥ 15	AHI < 15	ECG	PPG	ECG	PPG
Moderate-Severe OSA	74	143	85.1 (77.0, 93.2)	82.3 (73.0, 90.5)	80.4 (74.1, 86.7)	85.3 (79.0, 90.9)

Table 2. Performance of ECG-based and PPG-based sleep staging.

Category	N (epochs)	Sensitivity (%, 95% CI)		Specificity (%, 95% CI)	
		ECG	PPG	ECG	PPG
Wake	62,227	89.5 (89.3, 89.8)	84.5 (84.2, 84.8)	95.7 (95.6, 95.8)	94.6 (94.5, 94.7)
Light NREM	107,735	80.5 (80.2, 80.7)	80.6 (80.4, 80.8)	89.6 (89.4, 89.8)	84.2 (84.0, 84.5)
Deep NREM	14,212	81.1 (80.4, 81.8)	68.2 (67.4, 69.0)	93.9 (93.7, 94.0)	94.9 (94.8, 95.0)
REM	21,822	89.7 (89.3, 90.1)	83.1 (82.6, 83.6)	97.7 (97.6, 97.8)	97.3 (97.2, 97.4)
Total	205,996	N/A	N/A	N/A	N/A

Conclusions

This study provides clinical validation for an ECG-based ML approach to screen for OSA.

The results confirm that a single-lead ECG, analyzed with a tailored deep learning model, can reliably detect moderate-to-severe OSA with high sensitivity and specificity in comparison to the gold-standard.

This tool enables scalable screening of OSA in settings where overnight cardiovascular testing is conducted, to accurately flag OSA and refer patients to further testing and therapy.

This method has the potential to be deployed for existing devices such as Holter monitors, or as part of smart wearables and bedside monitors that continuously record single lead ECG.

By leveraging the presence of ECG in clinical practice, the approach can augment current screening efforts and reach populations that are underserved by traditional sleep medicine services and help bridge the huge diagnostic gap in OSA, potentially reducing the burden of untreated OSA.

Future work should focus on prospective trials to evaluate the impact of implementing this screening tool on patient outcomes, as well as evaluating the robustness of the approach on various overnight ECG devices (i.e different Holter monitors, as well as wearable ECG patches).