Prospective Clinical Validation of AI for PPG-based OSA detection utilizing Standardized Skin Color Assessments

UW Medicine DEPARTMENT OF NEUROLOGY

ensodata research

Chris Fernandez, MS*, Yoav Nygate, MS*, Sam Rusk*, 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. However, PPG technology depends on the transmission, absorption, and reflectance of light. As such, skin pigmentation influences pulse oximeter accuracy, as well as the detection of respiratory events and sleep stages in PPG-based wearable sleep devices. Even Though skin pigmentation plays a crucial part in the fidelity of PPG, standardized assessment is limited regarding skin pigmentation measures in PSG, HSAT, and PPG-based wearable sleep tracker performance validation. HSAT, PPG-based wearable sleep tracking, and even PSG performance rely crucially on pulse oximetry and PPG signal accuracy. A clinical performance validation study was conducted on a PPG-based AI system for OSA detection and Sleep Staging, utilizing a prospective, non-randomized, IRB approved trial design with all-comers participation offered to subjects referred for PSGs. This study included standardized skin pigmentation assessments to enable bias analyses.

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 OSA detection and Sleep Staging. Furthermore, during trial enrollment, each patient's skin pigmentation was collected by the administrating clinician utilizing standardized Munsell 7.5YR Soil-Color Chart where both the Skin Chroma and Skin Color Value were recorded for each patient. A sample of the color chart can be seen in Figure 1 below.

The study sample included N=215 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 scoring was then constructed by taking the majority scoring of sleep stages and all associated events by assigning each event to a 30 second epoch and assigning the majority label scored by the MSP (i.e. the sleep stage/event agreed upon by at least 2/3 RPSGTs in each epoch). OSA diagnostic performance measures including AHI and TST were evaluated in relation to standardized skin color measures utilizing Ordinary Least Squares (OLS) analysis to evaluate for any systematic, directional, or significant differences in AHI or TST performance as a function of the two continuous measures of skin pigmentation (i.e. Skin Chroma and Skin Color-Value). To evaluate the presence of bias based on skin chroma and color value, the absolute difference between the gold-standard MSP calculated index and the AI-calculated index was used for both AHI and TST.

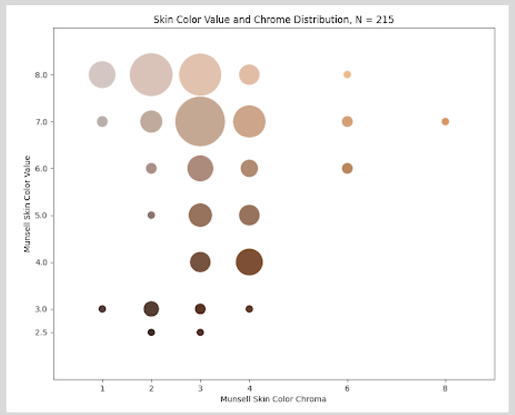
Results

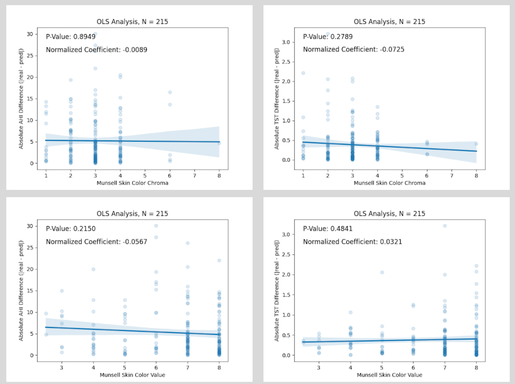
The study analyzed the performance of a PPG-based AI system for OSA detection and sleep staging in relation to standardized skin color measures, specifically skin color value and chroma, using data from 215 subjects. As seen in Table 1, the descriptive statistics show that the mean AHI was 15.829 with a median of 9.205, while the mean TST was 5.116 with a median of 5.300. Skin color value had a mean of 6.627 and a median of 7.000, whereas skin color chroma had a mean of 3.004 and a median of 3.000.

	Mean	Median	STD	Minimum	Maximum
AHI	15.829	9.205	18.278	0.135	110.500
TST	5.116	5.300	1.259	1.483	8.183
Skin Color Value	6.627	7.000	1.534	2.500	8.000
Skin Color Chroma	3.004	3.000	1.095	1.000	8.000

Furthermore, as seen in Figure 2, the bubble chart demonstrates that most subjects had higher skin color values (around 7.0) and moderate skin color chroma (around 3.0) which demonstrates a slight tilt towards lighter skin color. Figures 3-6 illustrate the OLS analysis results which evaluated the relationship between skin color measures and the absolute differences in AI-calculated AHI and TST compared to the gold-standard scoring by the MSP. The results showed no significant correlation between skin color chroma and AHI (P-Value: 0.8949) or TST (P-Value: 0.2789). Similarly, no significant correlation was found between skin color value and AHI (P-Value: 0.2150) or TST (P-Value: 0.4841).

Results Continued





Conclusions

The prospective trial findings showed that the AI system's performance in calculating AHI and TST is not significantly influenced by the subjects' skin pigmentation, indicating no systematic bias based on skin color in the AI system's OSA detection and sleep staging, 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, while ensuring reliability across a wide spectrum of skin pigmentations.