Skin Tone Calibration of Pulse Oximeter Oxygen Saturation Data

Orian Stapleton¹, Jay Luo¹, Yolanda Su¹, Sreenidhi Sankararaman¹, Esanika Mukherjee¹, Chao Cheng Chuang¹, Jiwon Woo¹, Nikita Sivakumar¹, Summer Duffy², Danielle Gottlieb-Sen², Joseph Greenstein¹*, Casey Taylor¹*

¹Johns Hopkins University, Baltimore MD, 21218
²Johns Hopkins School of Medicine, Baltimore MD, 21218

Design Day 2024

*Corresponding author

Background

Low oxygen levels cause nearly half a million deaths each year in the US. Arterial oxygen saturation (SaO₂) is the most accurate method to measure oxygen levels directly from the blood. However, pulse oximetry (SpO₂), a more common method, may be less precise, especially for those with darker skin, as melanin can interfere with its accuracy. Despite known issues with SpO₂ accuracy in darker-skinned individuals, current methods struggle to predict oxygen saturation for these patients, leading to healthcare disparities.

Clinical Need and Significance

- Pulse oximetry readings (SpO₂) are vital in healthcare, typically ranging from 95% to 100%. Values below 88% signal urgent hypoxia risks.
- Hidden hypoxemia occurs when blood oxygen (SaO₂) is below 88% but SpO₂ reads 92% to 96%, posing significant risks of organ damage, particularly for those with darker skin or diverse racial backgrounds.
- Studies indicate that black patients are nearly three times more likely to have undetected hypoxemia. Need: Algorithm that quantifies skin tone and finds pulse oximeter overestimation bias in relation to skin will give accurate oxygen saturation estimations for equitable clinical care for diverse races.

Aims and Exclusion Criteria

The aims to achieve the clinical need are as follows:
1. Establish statistical significance of pulse oximeter overestimation compared to the gold standard
2. Use computational models to better estimate O₂ saturation from physiological/demographic data and classify hidden hypoxemia
3. Develop race-independent skin tone quantification for standardization and inclusivity

Testing and Results

Figure 3. Hidden Hypoxemia (HH) Disproportionately Affects Darker Skin Tones. (A) Total HH dataset proportions (B) HH Racial Trends (C) Pulse Oximeter Overestimation

Based on Figure 3, 4% of the BOLD Dataset experience HH regardless of race (Fig. 3A). HH is more prevalent in darker skin tones (Fig. 3B) and the higher HH prevalence in darker skin tones can be attributed to oxygen saturation overestimation by pulse oximetry (Fig. 3C). HH is highly clinically relevant and is seen to disproportionately affect race due to pulse ox. overestimation.

Figure 4. Clinically Relevant Features for Reducing HH Utilizing Chi-Squared Test and Wilcoxon Rank-Sum Test (A) Chi-Squared (B) Wilcoxon Rank-Sum

Based on Figure 4, features were determined to be statistically different between HH and non-HH patients (one-hot encode categorical variables) by conducting a Wilcoxon Rank-Sum Test and Chi-Squares test. Features were ranked based on their significance and compared against critical threshold value to determine significance and the results were summarized.

Figure 5. Model and Features Workflow. (A) Feature prioritization and comparison (B) Model evaluation processing steps

Future Steps

Future steps include taking skin tone images from additional patients/students under uniform environmental conditions and utilizing colorimeters and skin tone quantification methods to make skin a quantifiable numeric variable for our analysis. Furthermore, additional models would be tested and compared against the tested regression models for model performance in predicting and accounting for HH overestimation.

Image References


Figure 1. Pulse Oximetry Bias. (A) SpO₂ darker skin bias (B) Self-reported race subjective (1,2)

Figure 2. Data Pre-Processing, BOLD Pulse Ox. Dataset (44907 patients, 142 variables, 49099 entries) found from MIMIC Dataset and criteria cleaned.

Figure 5. Model and Features Workflow. (A) Feature prioritization and comparison (B) Model evaluation processing steps

Future Steps

Future steps include taking skin tone images from additional patients/students under uniform environmental conditions and utilizing colorimeters and skin tone quantification methods to make skin a quantifiable numeric variable for our analysis. Furthermore, additional models would be tested and compared against the tested regression models for model performance in predicting and accounting for HH overestimation.