Correcting Racial Bias in Measurement of Blood Oxygen Saturation

Team 28: Brooks Miner, Jaychandran Padayasi, Rohan Myers, Saad Khalid, and Woojeong Kim

GitHub: https://github.com/brooksminer/pulse-ox-correction/

Overview

Fingertip pulse oximeters are the current standard for estimating blood oxygen saturation without a blood draw, both at home and in healthcare settings. However, pulse oximeters tend to overestimate oxygen saturation, often resulting in 'hidden hypoxemia': a patient has hypoxemia (dangerously low blood oxygen saturation), but the pulse oximeter returns a healthy oxygen value. Unfortunately, oximeter overestimation of oxygen saturation is exacerbated for patients with darker skin tones due to the interaction of skin pigment with the light-based oximeter technology. This results in Black patients experiencing hidden hypoxemia at almost twice the rate of white patients.

By combining pulse oximeter readings (SpO2) with additional patient data (such as vital signs and demographic characteristics), we develop improved methods for estimating arterial blood oxygen saturation (SaO2) and identifying Hidden Hypoxemia (HH). The predictions of our machine-learning models are more accurate than pulse-oximeter readings alone, and remove the systematic racial inequity inherent in the current accepted medical practice of using oximeter readings alone.

Stakeholders: Patients, medical professionals, medical device manufacturers, regulatory agencies

KPIs for identifying Hidden Hypoxemia: Sensitivity and specificity of HH classification

KPI for predicting SaO2: Root Mean Squared Error (RMSE) between true and predicted SaO2 values

Approach

We developed two models to predict arterial blood oxygen saturation (SaO2) and HH using features that would not require a blood draw from a publicly available dataset of de-identified medical records from 50,000 unique patients at Beth Israel Deaconess Medical Center in Boston, MA, between 2008 - 2019.

- **SaO2 Regression Model:** Predicts arterial blood oxygen saturation, which requires a blood draw and is the gold standard for measuring blood oxygen saturation.
- **Hidden Hypoxemia Classification Model:** Predicts HH, which means their SpO2 measurement returns a healthy value, but their SaO2 value identifies them as hypoxic.

Results & Strategies

Using both models, we can begin to address the issue of bias in healthcare outcomes for patients of color who may encounter an underappreciation of clinical risk due to an overestimation of blood oxygen saturation.

- For our **SaO2 Regression Model**, we used an ensemble of gradient-boosted random forest regression models to predict SaO2. Both SaO2 and SpO2 distributions were highly skewed, with most values 95-100 (healthy patients) and dominating the initial regression fit. To account for this, we used ensembled undersampling of patients with healthy SaO2 values to improve model accuracy for patients with hypoxemia. For hypoxemic patients, our model predicts SaO2 with **30% greater accuracy** (based on RMSE) than pulse-oximeter (SpO2) readings alone.
- For the **HH Classification Model**, we used an ensemble of gradient boosted forest classifiers created using XGBoost. The ensemble is trained on random undersampling of the dataset in order to artificially boost the prevalence of Hidden Hypoxemia, as only 1.6% of patients in the dataset had HH. Our model correctly identifies **7 out of 10 people** (sensitivity = 0.7) as having Hidden hypoxemia, while not sacrificing specificity.

We expect implementations of these models to mitigate racial inequity in healthcare and to contribute to ongoing conversation around equitable patient care. Our models will help inform solutions such as integration of additional patient data or calibration based on skin pigmentation into the design of improved pulse oximeter devices.

Future Iterations

While race proved to be an important feature in these models, further development would ideally incorporate direct measurements of skin pigmentation, rather than patients' self-identification of race. Future implementations should also rely on patient datasets with greater demographic diversity; the dataset we used was majority white patients (over 65%).