Title: Deep Neural Network employing Transfer Learning to Improve Racial Bias in Pulse Oximetry Accuracy

Authors: Sungsoo Kim, MD, MS, PhD Candidate[1,2], Sohee Kwon, MD, MPH[1], Alan C. Bovik, PhD[2], Mia K. Markey, PhD[3], Maxime Cannesson, MD, PhD[1]

[1]Department of Anesthesiology and Perioperative Medicine, University of California Los Angeles, Los Angeles, California, USA; [2]Department of Electrical and Computer Engineering, The University of Texas at Austin, Texas, USA; [3]Department of Biomedical Engineering, The University of Texas at Austin, Texas, USA

Background and Aims:

Pulse oximetry has become the medical standard for continuous and noninvasive monitoring for peripheral arterial oxygen saturation (SpO2, %). However, the inaccuracy of pulse oximetry has been reported in certain racial groups. In addition, the high-fidelity dataset including race and oxygen saturation are limited. Therefore, we proposed a framework employing Deep Neural Networks (DNN) based transfer learning to reduce the inter-group bias in pulse oximeter accuracy and applyed our framework in high fidelity dataset.

Methods:

Machine Learning of Physiological Waveforms and Electronic Health Record Data to Predict, Diagnose, and Treat Hemodynamic Instability in Surgical Patients (MLORD) dataset, collected at University of California Los Angeles (UCLA), was analyzed in this study. A total of 19,525 patients underwent surgeries at UCLA between 2019 and 2022 were included. All data was collected during surgery, which increased fidelity of dataset compared to other clinical data collected outside operating room. We defined SdO2 (%), non-invasive peripheral oxygen saturation predicted by DNN, as an estimation of arterial oxygen saturation (SaO2, %). First, in Caucasian group (n=10,068), we trained the DNN. Second, using Black group (n=1,273), we applied transfer learning into the “pretrained DNN” to generate “transferred DNN”. The input of DNN was composed of non-invasive features including values of SpO2, heart rate, age, gender, body mass index, and blood pressure. The accuracy of SdO2 was evaluated by assessing the mean ± standard deviation of root mean square error (RMS) between SdO2 and SaO2, and then compared to RMS between SpO2 and SaO2 using t-test.

Results:

In each patient, we selected the first measurement of SaO2 and then selected input features in the closest time with selected SaO2. Timing difference of measurement between SaO2 and periodic features was within 5 minutes in 99.8% of observations. Each racial group split into training set and test set. DNN was trained only by the training set while the accuracy of SdO2 was evaluated in the test set. In Caucasian group, the accuracy of SdO2 (2.15 ± 2.62) using the pretrained DNN outperformed the accuracy of SpO2 (2.51 ± 3.37) with statistical significance ($p<0.001$). Similarly, in African American group, the accuracy of SdO2 (2.42±2.99) using the transferred DNN still outperformed the accuracy of SpO2 (2.81±3.84) with statistical significance ($p<0.001$).

Conclusions:

Despite a smaller number of African American group compared to Caucasian group, the accuracy of SdO2 using the transferred DNN still outperformed the accuracy of SpO2. This finding suggests that DNN employing Transfer Learning may be applicable to increase pulse oximetry accuracy in African American group.

Acknowledgements:

This research is supported by NIH research funding (R01HL144692; Machine Learning of Physiological Waveforms and Electronic Health Record Data to Predict, Diagnose, and Treat Hemodynamic Instability in Surgical Patients) and by UCLA Anesthesiology & Perioperative Medicine Seed grant (441006-2X-75014; Application of Deep Learning for real-time non-invasive continuous monitoring for enhanced peripheral oxygen saturation)