Does Skin Tone Affect Machine Learning Classification Accuracy Applied to Photoplethysmography Signals?

Philip J Aston

National Physical Laboratory/University of Surrey
Teddington/Guildford, UK

Introduction: A recent UK report highlighted that pulse oximeters are sensitive to the optical properties of the skin and so can give inaccurate readings of blood oxygenation for patients with darker skin tones. Thus, these devices have the potential for ethnic and racial bias, resulting in some ethnic groups receiving suboptimal treatment. Photoplethysmography (PPG) signals from pulse oximeters and a wide range of wearable devices are now widely used with machine learning to predict other physiological parameters such as blood pressure (BP). We investigated whether skin tone had any effect on the performance of machine learning classifiers trained on PPG signals.

Methods: We considered the machine learning problem of a binary classification of BP as high ($\geq 140/90$ mmHg) or not high ($< 140/90$ mmHg) and restricted attention to systolic BP. We used the Aurora BP dataset, which contains PPG signals, BP readings and Fitzpatrick Skin Tone for many records. Symmetric Projection Attractor Reconstruction (SPAR) was used to extract features from the PPG signals. Ten-fold cross validation was performed using signals with Fitzpatrick Skin Tone 1 (lightest skin tone) in order to give a benchmark classification performance. A single model was then trained using all of the signals with Fitzpatrick Skin Tone 1 and tested using data for all other skin tones to see the extent to which skin tone affects model performance.

Results: The classification accuracy for each of the skin tone classes is shown in Fig. 1, which demonstrates a significant decrease in accuracy for all other skin tones compared to class 1 except, surprisingly, for class 5.

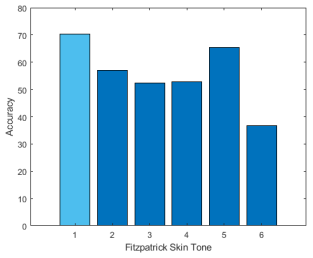

Figure 1: Accuracy of a machine learning model trained on Skin Tone 1 data and tested on other skin tone classes.

Conclusions: While it is known that PPG signals are sensitive to skin tone, it was not known whether this translates into poorer performance for machine learning models trained only on lighter skin tones. We have shown that there is a significant decrease in accuracy for most darker skin tones for a binary classification problem related to BP.