Contactless SpO2 Monitoring using Explainable Convolutional Neural Networks from Smartphone Videos

Bo Li[1] and Rohit Sharma[1]

[1]University of Kentucky

March 17, 2023

Abstract

The estimation and monitoring of SpO2 are crucial for assessing lung function and treating chronic pulmonary diseases. The COVID-19 pandemic has highlighted the importance of early detection of changes in SpO2, particularly in asymptomatic patients with clinical deterioration. However, conventional SpO2 measurement methods rely on contact-based sensing, presenting the risk of cross-contamination and complications in patients with impaired limb perfusion. Additionally, pulse oximeters may not be available in marginalized communities and undeveloped countries.

Contactless SpO2 Monitoring using Explainable Convolutional Neural Networks from Smartphone Videos

Bo Li
Tencent, USA

Rohit Sharma
University of Kentucky, USA

Abstract—**The estimation and monitoring of SpO2 are crucial for assessing lung function and treating chronic pulmonary diseases. The COVID-19 pandemic has highlighted the importance of early detection of changes in SpO2, particularly in asymptomatic patients with clinical deterioration. However, conventional SpO2 measurement methods rely on contact-based sensing, presenting the risk of cross-contamination and complications in patients with impaired limb perfusion. Additionally, pulse oximeters may not be available in marginalized communities and undeveloped countries.**

I. Introduction and Method

The measurement and monitoring of SpO2 are critical for evaluating lung function and managing chronic pulmonary diseases. However, conventional contact-based methods present risks of cross-contamination and complications in patients with impaired limb perfusion. Additionally, pulse oximeters may not be available in marginalized communities and undeveloped countries. To address these limitations, recent studies have investigated measuring SpO2 using videos captured by cameras, including smartphones. However, this presents challenges due to weaker physiological signals and lower optical selectivity of smartphone camera sensors. Previous methods have relied on explicit feature extraction methods, but this study proposes an innovative approach using explainable convolutional neural network (CNN) models that extract features holistically from all three color channels of contactless videos captured using consumer-grade smartphone cameras. This approach aims to improve the accuracy and efficiency of SpO2 monitoring and has the potential to be adopted in health screening and telehealth, particularly in areas where conventional contact-based methods and pulse oximeters are not widely available. To overcome these limitations, recent studies have explored SpO2 measurement using videos captured without physical contact. However, measuring SpO2 using cameras in a contactless way, especially from smartphones, is challenging due to the weaker physiological signals and lower optical selectivity of smartphone camera sensors. Prior studies have relied on explicit feature extraction methods based on pulse oximeter principles. In contrast, this article proposes a novel approach to SpO2 measurement using explainable convolutional neural network (CNN) models. The method extracts features from all three color channels of contactless videos captured using consumer-grade smartphone cameras in a holistic manner.

By using CNN models, the proposed approach can address the challenges of contactless SpO2 measurement and improve the accuracy and efficiency of monitoring SpO2. This approach has the potential to be used in health screening and telehealth applications, particularly in areas where traditional contact-based methods and pulse oximeters may not be widely available.

To accurately estimate a person's SpO2 levels, the color information from their skin needs to be extracted, as the physiological information related to SpO2 is embedded in the light reflected/reemitted from the skin. This involves isolating the skin pixels from the background using a method such as Otsu's method and identifying the region of interest (ROI). The RGB values for each skin pixel are averaged over each video frame, creating R, G, and B time series, which are referred to as skin color signals. These signals are then segmented into 10-second intervals using a sliding window with a stride of 0.2 seconds to form inputs for the neural networks. Longer segments are used to improve resilience against sensing noise. Our proposed approach for SpO2 monitoring from videos using deep learning shows great promise and has the potential to be widely adopted in health screening and telehealth. This method provides a convenient and noninvasive way to continuously monitor SpO2, which can serve as an early warning sign of inadequate oxygenation and clinical deterioration. By detecting these warning signs early, medical professionals can take proactive measures to improve patient outcomes and reduce healthcare costs. While this technology is still in the early stages, it has the potential to significantly improve patient care and overall healthcare systems. As the segment length is much longer than the minimum required length for one cycle of heartbeat, a recurrent neural network structure is not necessary to capture temporal dependencies, and a fully-connected or convolutional structure is sufficient. To ensure the neural networks' predictions are interpretable, we propose the use of layer-wise relevance propagation (LRP). LRP is a widely-used technique that explains deep neural network predictions by assigning relevance scores to input features based on their contribution to the final output prediction. This allows for a better understanding of which features are most important for the prediction.

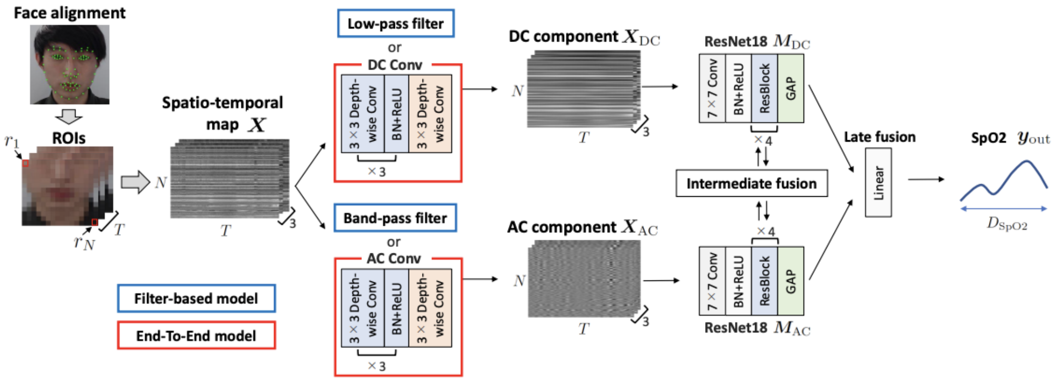

Fig. 1: System model diagram: The machine learning model (random forest) uses three types of features; CRM features, transaction features, and time-frequency features. The model is trained using previously known suspicious activity. The result of the model is a score between 0 and 1, which is converted into a decision by an optimized threshold. The threshold reflects the risk tolerance of the financial institution.

II. Observations

We have two recordings for training and testing the model for SpO2 estimation. To prevent any overlapping data between the training and validation sets, each recording is divided into three breathing cycles, with the first two cycles used for training and the third cycle used for validation. The skin color signals are extracted by averaging the R, G, and B time series from the skin pixels, which are separated from the background using Otsu's method. The skin color signals are then divided into 10-second segments using a sliding window with a stride of 0.2 seconds, which are used as input for the neural networks. As the dataset size for participant-specific experiments is small, the training and validation data are augmented by bootstrapping, where data samples are replaced during sampling. This approach ensures that the model is evaluated comprehensively by varying the random seed used for model weights initialization and random oversampling between each instance. The selected instance of the model with the highest validation RMSE is evaluated on the test recording. A 1-D convolutional neural network (CNN) model with two temporal convolutional layers, max pooling, and 25% dropout is used for feature extraction from the raw PPG signal, bandpass filtered signal, and bias signal for each color channel. The CNN model hyperparameters, including the number of convolutional layers, input window size, filters, and filter length, are optimized using a cross-validated grid search process. The implementation of the CNN model uses Python 3.5 and Tensorflow 1.2, and the training is performed on a cluster of 36 nodes with an NVIDIA P100 GPU, which takes approximately four weeks for the entire grid search process. Overall, this approach provides an accurate and efficient way to estimate SpO2 using skin color signals extracted from video recordings. After tuning the model structure and hyperparameters using training and validation data, multiple instances of the model were trained using the best-tuned hyperparameters. Random seed and random oversampling were varied between each instance to ensure comprehensive evaluation of the model. The model instance with the highest validation RMSE was chosen for evaluation on the test recording.

To obtain raw PPG signals, bandpass filtered signals, and bias signals for each color channel, various preprocessing techniques were employed. Feature extraction from these signal streams was achieved using a 1-D convolutional neural network (CNN) model with two temporal convolutional layers, max pooling, and 25% dropout. The best combination of hyperparameters, including the number of convolutional layers, input window size, filters, and filter length, was determined using a cross-validated grid search process. The CNN model was implemented using Python 3.5 and Tensorflow 1.2 and was trained on a cluster of 36 nodes with an NVIDIA P100 GPU, which took approximately 4 weeks for the entire grid search process.

References

[1] G. Indiveri and E. Chicca. A VLSI neuromorphic device for implementing spike-based neural networks. In Neural Nets WIRN11 - Proceedings of the 21st Italian Workshop on Neural Nets, pages 305–316, Jun 2011. 12

[2] K. Rajitha, et al. ”Estimation of the Cardiac Pulse from Facial Video in Realistic Conditions,” ICAART, 2019.

[3] Subramaniam, Arvind. ”A neuromorphic approach to image processing and machine vision,” 2017 Fourth International Conference on Image Information Processing (ICIIP). IEEE, 2017

[4] C. Bartolozzi and G. Indiveri. Global scaling of synaptic efficacy: Homeostasis in silicon synapses. Neurocomputing, 72(4–6):726–731, Jan 2009.

[5] Rajitha, et al. (2019, September). Spectral reflectance based heart rate measurement from facial video. In 2019 IEEE International Conference on Image Processing (ICIP) (pp. 3362-3366). IEEE.

[6] Park, H.L., Kim, M.H., Kim, M.H. and Lee, S.H., 2020. Reliable organic memristors for neuromorphic computing by predefining a localized ion-migration path in crosslinkable polymer. Nanoscale, 12(44), pp.22502-22510.