Idoia Badiola Aguirregomezcorta*, Vladimir Blazek, Steffen Leonhardt, Christoph Hoog Antink

Learning about reflective PPG for SpO_2 determination using Machine Learning

Abstract: Reflective Photoplethysmography (PPG) sensors are less obtrusive than transmissive sensors, but they present patient-dependent variations in the so-called "Ratio of Modulation" (R). Thus, the conventionally employed calibration curves for determining peripheral oxygen saturation ($S_{\mathrm{p}}O_2$) may report inaccurate values. In this paper, we study the possibility of overcoming these limitations through Machine Learning (ML). For that, we show the results of applying several algorithms and feature combinations to PPG data from a human hypoxia study. The study was performed on ten healthy subjects. Their target oxygen saturation was reduced in five steps from 98-100% to 70-77% through an oral mask. Blood Gas Analysis (BGA) was performed five times for each saturation level to measure the arterial oxygen saturation. PPG data were acquired from a reflective pulse oximeter placed in the subjects' ear canals. PPG signals were pre-processed, and several features in the frequency and temporal domain were calculated. For the ML algorithms' input, we explored different combinations of the features. We trained and validated the algorithms with the data from seven patients, and we tested them on three. Finally, we performed leave-one-out cross-validation to ensure the universality of the methods. The results show a good agreement of the predictions with the BGA values for Linear Regression, k-Nearest Neighbors, Stochastic Gradient Descent, and Neural Network for all input feature combinations with an average RMSE in the range of 3%. However, the performance of the Linear Regression was not beaten by the Neural Network, even for overfitting with 2000 hidden layers. The combination of R calculated with a Fast-Fourier Transform and $AC_{\mathrm{RMS,red}}/AC_{\mathrm{RMS,ir}}$ significantly improved the results, reducing the RMSE by 25%. This work demonstrates that a straight-forward Linear Regression is capable of determining $S_{\mathrm{p}}O_2$ with reflective PPG independently of the subject if the ratio $AC_{\mathrm{RMS,red}}/AC_{\mathrm{RMS,ir}}$ is considered simultaneously with the Ratio of Modulation.

Keywords: reflective PPG, SpO_2, arterial oxygen saturation, Machine Learning, hypoxia, Ratio of Modulation.

https://doi.org/10.1515/cdbme-2021-2009

***Corresponding author: Idoia Badiola Aguirregomezcorta:** Lehrstuhl für Medizinische Informationstechnik (MedIT), RWTH Aachen University, Pauwelsstraße 20, Aachen, Germany, e-mail: badiola@hia.rwth-aachen.de
Vladimir Blazek, Steffen Leonhardt, Christoph Hoog Antink: Lehrstuhl für Medizinische Informationstechnik (MedIT), RWTH Aachen University, Aachen, Germany

1 Introduction

Photoplethysmography (PPG) is an optical technique that detects blood volume changes in the tissue's microvascular bed under the skin surface [1]. Its clinical applications range from monitoring blood oxygen saturation to determining heart rate and respiratory rate. PPG sensors typically consist of two LEDs of different wavelengths of light (red and infrared) and a photodetector. Oxyhemoglobin (HbO_2) and deoxyhemoglobin (Hb) in blood absorb light differently depending upon the wavelength. For measuring $S_{\mathrm{p}}O_2$, the amount of light absorbed by the HbO_2 is calibrated against the total amount of light received by the photodetector. Typical $S_{\mathrm{p}}O_2$ values range from 95 % to 100 %, and lower values reflect insufficient oxygen levels or hypoxia.

In practice, we calculate $S_{\mathrm{p}}O_2$ with the aid of a figure of merit known as Ratio of Modulation or Ratio of Ratios (R), defined in eq 1.

$$\boldsymbol{R} = \left(\frac{AC}{DC}\right)_{\mathbf{RED}} / \left(\frac{AC}{DC}\right)_{\mathbf{IR}} \tag{1}$$

where $AC_{\mathrm{RED}}, DC_{\mathrm{RED}}, AC_{\mathrm{IR}}$ and DC_{IR} are the AC and DC components of the PPG signal for the red and the infrared wavelengths, respectively. After R is computed, a calibration curve is usually employed for determining oxygen saturation. Theoretically, Beer-Lambert's Law can be modified to relate the saturation to R [2]. However, today manufacturers calibrate each device empirically by performing a hypoxia study [3], as the one we employed for this work. For that, they collect reference arterial oxygen saturation ($S_{\mathrm{a}}O_2$) values invasively with Blood Gas Analysis (BGA), which is considered the gold standard. In this technique, a needle or a catheter is inserted into the artery, and blood is extracted.

Then, S_aO_2 can be calculated accurately from the concentration of HbO_2 and Hb in blood.

In a previous study from 2012, Venema et al. collected reflective PPG signals from an ear pulse oximeter and BGA values from ten healthy individuals during a human hypoxia study [4]. In that study, they showed that a general calibration for all patients was impossible without a previous individual normalization of R at a starting value of 0.7. Other authors such as Arsath et al. have also shown that for reflective PPG, "R-value is subject-specific and is also heavily dependent on the measurement site" [5], and Guo et al. also proposed an individual calibration curve for each participant [6].

In a recent publication from Vencat et al. [7], the authors suggested using ML to overcome those limitations on reflective pulse oximetry on the finger. Their model implemented k-Nearest Neighbor (kNN), Quadratic Discriminant Analysis, and Bagged Trees, and they used a commercial pulse oximeter as a reference. However, they observed a significantly lower performance at oxygen saturations below 90%, and they did not reach any SpO_2 below 81%.

In this work, we study the possibility of employing further ML algorithms for the determination of SpO_2 on reflective PPG measured on the ear even for lower values of S_aO_2. Haynes [8] concluded that a transmissive PPG sensor on the ear shows a much lower agreement between S_aO_2 and conventionally determined S_pO_2 than a sensor on the finger. Therefore, we determined S_pO_2 with the aid of ML algorithms, which might lead to a higher agreement between both measures. For that, PPG and BGA datasets from a hypoxia human study [4] were used with induced oxygen saturations ranging from 70-76% to close to 100%.

2 Methods

The hypoxia study was performed on ten healthy patients [4]. The patients lay on a bed, and oxygen input for the patients was reduced in five steps of 150 to 200 seconds from 98-100% to 70-77% through an oral mask.

The saturation was then measured using several methods. On the one hand, arterial blood was extracted five times for each level of saturation, and BGA was performed to determine the S_aO_2. That occurred every 20 seconds after at least 30 seconds of stabilization of each saturation stage. On the other hand, two "LAVIMO" [4] reflective sensors were placed on one ear of the patients: a sensor with a universal fit for all patients and another with an individually designed fit. For this paper, we only considered the data from the universal-fit sensor. Finally, reference values of the saturation were also extracted from commercial pulse oximeters connected to the patients' index finger.

The study dataset consists of PPG amplitudes for red and infrared wavelengths at a sampling frequency of 200 Hz and BGA values with a specific timestamp. The data were processed and analyzed in MATLAB 2019a. For pre-processing, we aligned all the values with respect to time. Afterward, we segmented the PPG data into 25 levels for each patient: five segments for each of the five stages of saturation. Each segment starts 10 seconds before the corresponding BGA's blood extraction time and ends 10 seconds after.

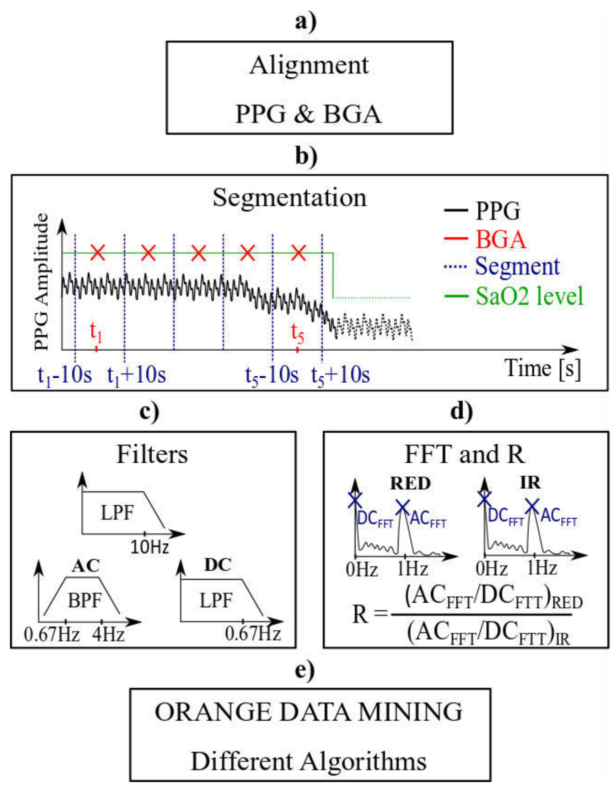

Figure 1: Flow diagram of the pre-processing of the data.

The next step was to filter the PPG signals for red and infrared wavelengths at each segment (see also Figure 1 c). First, a Butterworth low-pass filter with a cutoff frequency of 10 Hz, attenuation of 50 dB and 170th order removes the higher frequencies from the signals. Because the AC part relates to the heart activity, a 4th order Butterworth band-pass filter with cutoff frequencies 0.67 Hz to 4.5 Hz extracts the AC part of the PPG. Finally, a 6th order Butterworth low-pass filter isolates the DC part of the PPG signals below 0.67 Hz.

We selected three features from the PPG signal segments for each BGA value for training the ML algorithms: the ratios $AC_{\mathrm{RMS}_{\mathrm{RED}}}/AC_{\mathrm{RMS}_{\mathrm{IR}}}$ and $DC_{\mathrm{RMS}_{\mathrm{RED}}}/DC_{\mathrm{RMS}_{\mathrm{IR}}}$ and R. For calculating the latter, a Fast Fourier Transform (FFT) was applied to the segments, and the maximum values from the

FFT around 0 Hz and 1 Hz were considered as DC_{FFT} and AC_{FFT}, respectively (see also Figure 1 d). For the implementation of the ML algorithms, their training, and the results' visualization, we employed Orange Data Mining, which is based on Python scripting and a C++ backend [9].

The software allows for importing features and related targets in different formats. Among the available algorithmic models are Neural Networks (NN), Random Forest, kNN, and Linear Regression. Several parameters can be adjusted or selected within the computational block for these algorithms for optimal performance on the training data.

For the optimization of the solutions, it analyzes the convergence of the algorithms depending on several values: the Mean Absolute Error (MAE), the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE), and the Coefficient of Determination (R2). Finally, the predictions for the test data can be exported in many formats.

In a previous analysis of the same datasets, Venema et al. applied the conventional method for determining $S_{\mathrm{p}}O_2$: they calculated R from the data, and they tried to find a calibration curve for the saturation. However, they showed that, although individual calibration polynomials for each patient presented highly accurate results, it was impossible to find a global calibration curve for all of them. The reason was that the R-values for each patient showed a parallel shift, so they needed an individual normalization of the initial value at 0.7.

For this work, we collected data from 10 patients and 25 levels for each patient, ranging oxygen saturations from 100% down to 70%; therefore, the number of samples available for the study was 250. To test the ML algorithms' effectiveness, we trained them with 60% of the data from seven patients and tested them with the reamining 40%. Afterward, we tested the algorithms on all the remaining three patients' data.

The training inputs consisted of different combinations of the features from the PPG signals and the BGA values as targets. Those combinations were the following: only R; R, $AC_{\mathrm{RMS,red}}/AC_{\mathrm{RMS,ir}}$ and $DC_{\mathrm{RMS,red}}/DC_{\mathrm{RMS,ir}}$; R and $AC_{\mathrm{RMS,red}}/AC_{\mathrm{RMS,ir}}$; and $AC_{\mathrm{RMS,red}}/AC_{\mathrm{RMS,ir}}$ and $DC_{\mathrm{RMS,red}}/DC_{\mathrm{RMS,ir}}$.We implemented and tested the following algorithms and optimizers: AdaBoost, Linear Regression, Neural Network, Random Forest, Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Decision Tree, and kNN. In Orange Data Mining, we selected each algorithm's parameters for optimal performance. For instance, in the case of the Neural Network, increasing numbers of hidden layers were implemented, ranging from 1 up to the overfitting case of 2000.

3 Results and discussion

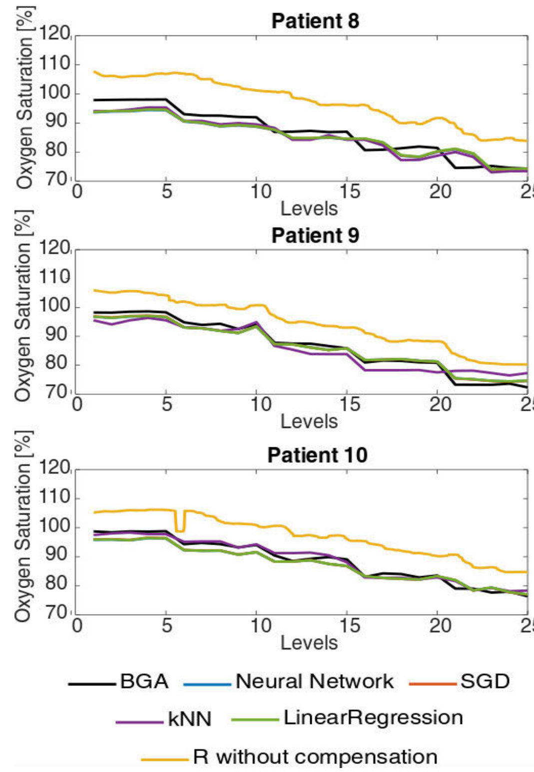

Figure 2: Evolution of predicted values for the tested 3 patients.

Figure 2 shows the predictions for the best four algorithms vs. the actual BGA values. The best results for all four ML algorithms were obtained from the input parameter combination of R and $AC_{\mathrm{RMS,red}}/AC_{\mathrm{RMS,ir}}$, which was considered in Figure 2. We can observe that Linear Regression, Neural Network, and SGD present an overlapping course of the predictions, whereas kNN substantially differs for Patient 9, especially for lower $S_{\mathrm{a}}O_2$. However, the predictions for levels 9, 10 and 11 in the x-axis match the BGA values, as they follow the original signal form. It is also observable that the predictions for the last levels in Patient 8 are shifted in time compared to $S_{\mathrm{a}}O_2$, which also occurs for the calculations with the conventional method. Strong motion artifacts might cause this behavior due to the relocation of a sensor in the ear, which also affected the PPG signal's DC component.

Figure 3 represents the leave-one-out cross-validation for all combinations of input data for the tested algorithms. It is observable that the average RMSE is approximately 3% for all the patients. From the upper two graphs, we conclude that the combination of R calculated with a Fast-Fourier Transform and $AC_{\mathrm{RMS,red}}/AC_{\mathrm{RMS,ir}}$ improved the results, reducing the sum of the RMSE of each patient from 40 to 30. Besides, not even a NN with 2000 hidden layers performs better than the

Linear Regression with that combination. This might be a consequence of the small dataset from the 250 samples from the study that was available for training. This will be further investigated with techniques of data augmentation.

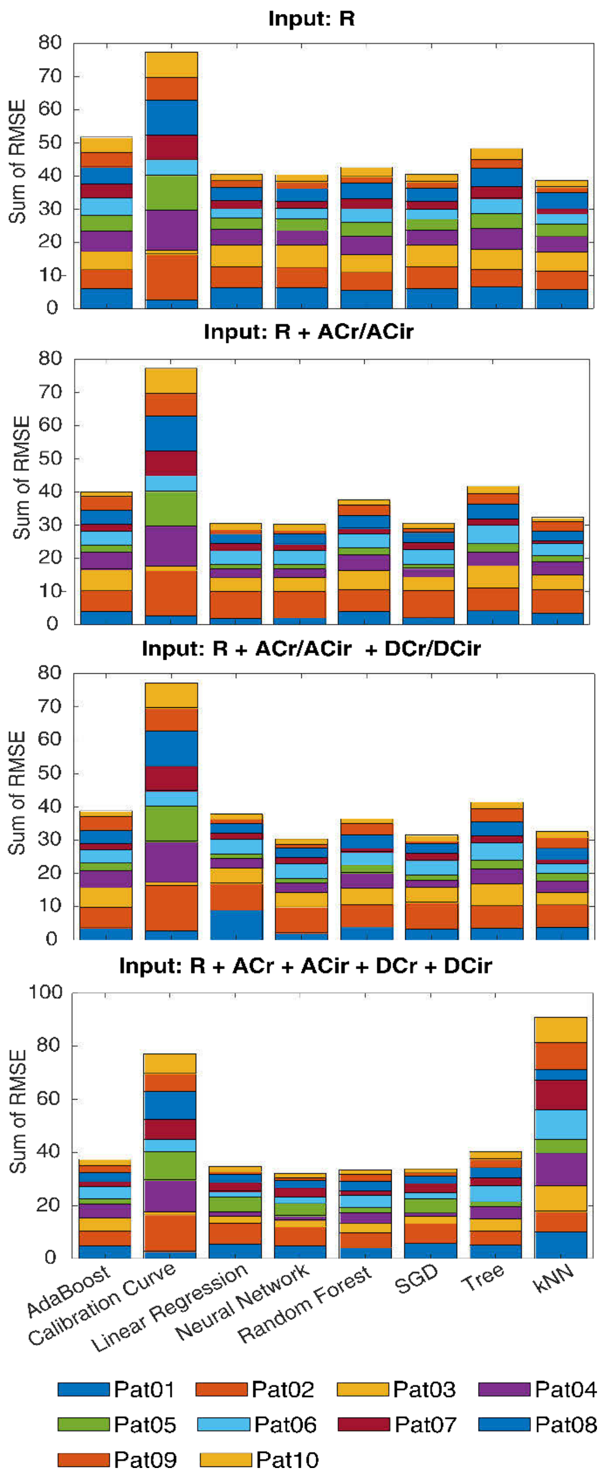

Figure 3: Leave-one-out cross-validation for all algorithms and input feature combinations.

4 Conclusions

In this study, we tested several ML algorithms for the determination of S_pO_2 from reflective PPG in the ear. The PPG signals were preprocessed to extract input features for the ML models. The cross-validation demonstrates that a straight-forward Linear Regression is capable of determining S_pO_2 with reflective PPG independently of the subject. For that, the ratio $AC_{\mathrm{RMS,red}}/AC_{\mathrm{RMS,ir}}$ shall be considered simultaneously with the Ratio of Modulation R.

Author Statement

Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] J. Allen. Photoplethysmography and its application in clinical physiological measurement. Physiological measurement, vol. 28, no. 3, R1-39, 2007.

[2] J. G. Webster, *Design of pulse oximeters*: CRC Press, 1997.

[3] S. Chatterjee. Monte Carlo investigation of light-tissue interaction in photoplethysmography: City, University of London, 2018.

[4] B. Venema, N. Blanik, V. Blazek, H. Gehring, A. Opp, and S. Leonhardt. Advances in Reflective Oxygen Saturation Monitoring With a Novel In-Ear Sensor System: Results of a Human Hypoxia Study, IEEE Transactions on Biomedical Engineering, vol. 59, no. 7, pp. 2003–2010, 2012.

[5] A. P. S. Mohamed Tanveejul et al., A Study on the Subject and Location Specificity in Reflectance based SpO2 Estimation using R-value based Calibration Curve, in 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2020, pp. 1–6.

[6] T. Guo, Z. Cao, Z. Zhang, D. Li, and M. Yu, Reflective oxygen saturation monitoring at hypothenar and its validation by human hypoxia experiment. BioMedical Engineering OnLine, vol. 14, no. 1, p. 76, 2015.

[7] S. Venkat et al., Machine Learning based SpO2 Computation Using Reflectance Pulse Oximetry. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 482–485.

[8] J. M. Haynes, The ear as an alternative site for a pulse oximeter finger clip sensor. Respiratory care, vol. 52, no. 6, pp. 727–729, 2007.

[9] J. Demšar et al., Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research, vol. 14, pp. 2349–2353, 2013.