A Proposed Brazilian Dataset and Approach for Remote Blood Oxygen Saturation Measurement from Facial Videos

Êmilly Farias Bruning[1], **Gabriel M. Lunardi**[1], **Adriano Q. Oliveira**[1],
Thiago L. T. da Silveira[1], **Leonardo Ramos Emmendorfer**[1]

[1]Universidade Federal de Santa Maria - UFSM, Santa Maria, RS, Brasil

emilly.bruning@acad.ufsm.br

{gabriel.lunardi, adriano.q.oliveira}@ufsm.br

{thiago.silveira, leonardo.emmendorfer}@ufsm.br

Abstract. *Blood oxygen saturation (SpO_2) measurement is one of the most commonly used vital signs in the clinical assessment of patients. Typically, this measurement is performed using pulse oximeters, which, although accurate, present certain limitations, as they rely on direct skin contact and are sensitive to external factors. To overcome these constraints, there has been growing interest in remote monitoring of vital signs through imaging. In this work, the creation of a Brazilian dataset for SpO_2 estimation is proposed, focusing on the diversity of skin tones, ages, and medical conditions. The recorded videos will be synchronized with SpO_2 values obtained from pulse oximeters. Subsequently, the recordings will be converted into Spatial-Temporal Maps (STMaps), which compact spatial and temporal information, and a 2D convolutional neural network will be used to predict SpO_2 values from these representations.This dataset and reference model aim to support research in non-invasive remote monitoring of vital signs, providing an inclusive framework for future studies and improving the reliability of SpO_2 estimation across different populations.*

1. Introduction

Blood oxygen saturation (SpO_2) is one of the most important vital signs, and its measurement is often performed in routine medical examinations, especially in cases of pulmonary diseases or cardiovascular problems. Currently, the most common method for measuring oxygenation is through pulse oximeters, non-invasive devices that use the photoplethysmography technique [Wuyart et al. 2025]. These devices operate by being positioned on specific parts of the body, usually the fingertip or the earlobe, where two wavelengths of light are emitted and partially absorbed by the pulsatile arterial blood. The photosensitive sensor records these absorption variations, and SpO_2 is calculated from the ratio between the signals obtained at the two wavelengths [DeMeulenaere 2007].

Although accurate, pulse oximeters present limitations. Since they rely on direct skin contact, they cannot be used in patients with wounds, burns, or ulcers [Wuyart et al. 2025]. In addition, they require constant sanitization to prevent contamination. Their reliability can also be affected by external interferences, such as lighting variations, dirt on the sensor, motion artifacts, mechanical or electrical interferences, and extreme temperatures [DeMeulenaere 2007, Machado Lunardi et al. 2018].

In this context, video-based monitoring emerges as an alternative solution. Being non-invasive, it reduces the risk of contamination and enables continuous measurements without physical contact. The measurement of physiological signals through RGB cameras is based on the analysis of subtle variations in the light reflected from the skin, which capture changes in blood volume over time [Wuyart et al. 2025]. The most commonly used technique is remote photoplethysmography (rPPG), in which regions of interest, generally the face or hands, are selected in the video, and the intensity variations in the color channels (red, green, and blue) are processed to extract pulsatile signals [Wuyart et al. 2025].

Thus, this work proposes the construction of a dataset to support the measurement of vital signs, specifically SpO_2, through remote monitoring with cameras. The data collection procedure consists of recording videos and obtaining real-time oximetry values with a pulse oximeter; subsequently, the extracted data will be used to train a convolutional neural network model to predict SpO_2 values from the recordings. The main goal of this dataset is to be diverse and to provide a balanced distribution of skin tones and age groups among individuals. This proposal aims to avoid bias in the development of prediction models, especially when considering the Brazilian population, which presents diverse phenotypic characteristics [de Souza et al. 2020]. Currently available datasets have limitations in this regard, as they generally include individuals with lighter skin tones [Dasari et al. 2021]. Furthermore, while most of the literature focuses on heart rate, this work proposes to focus on the estimation of SpO_2.

2. Related Work

Currently, there are few databases that relate blood oxygen saturation to videos. The main one is the Asian dataset VIPL-HR, proposed by [Cheng et al. 2024], which includes samples of individuals with skin types III and IV on the Fitzpatrick Scale [Fitzpatrick 1988]. Considering the Brazilian context, this sample is limited, as it presents bias against individuals with darker skin tones (type VI) [Dasari et al. 2021], which compromises the generalization of remote vital signs monitoring models.

In this regard, this work distinguishes itself by creating a diverse dataset and a methodology capable of encompassing a wide spectrum of skin tones. This would reduce biases in AI training and expand the possibilities for research on remote vital sign measurement, making it a relevant tool for studies in the fields of health and technology.

3. Methodology

This section describes the complete research plan, divided into two main parts: the construction of the database and the use of this database through a deep learning approach.

3.1. Data Collection

Data collection will take place at the University Hospital of Santa Maria (HUSM), the largest public hospital complex in the countryside of Rio Grande do Sul, with extensive activity in emergency and urgent care for several municipalities in the central region. The high patient flow will allow for a substantial amount of data collection and will contribute to the main objective of the study: creating a balanced dataset that encompasses different ages, skin tones, and medical conditions.

For video acquisition, two cameras will be used: one from a smartphone and one webcam. SpO_2 will be collected simultaneously through a pulse oximeter; this measurement is relevant because it will serve as a comparative factor when the data are used for signal prediction. Data collection will follow standardized conditions regarding distance, lighting, and movement. During recording, the patient, rested and at ease, will remain in a fixed position approximately one meter from the camera, and two videos of at least 60 seconds each will be recorded, one for each of the proposed cameras, allowing for natural head movements. The videos will be recorded under moderate ambient lighting. Along with the videos and SpO_2 values, patient information such as age and sex will also be collected. All recordings will require the patient's consent, formalized through the Informed Consent Form (ICF), ensuring authorization for the use of their image in research and guaranteeing its protection [Brasil 2018, Soares et al. 2025].

3.2. Approach for SpO_2 Measurement

After data collection, the videos will undergo preprocessing for the creation of Spatial-Temporal Maps (STMaps). As illustrated in Figure 1, each video will be divided into 255 frames, and the patient's face will be detected, aligned, and centered within a 128×128 pixel area[1]. The frames will be segmented into temporal patches, and the signals extracted from each segment will be concatenated, forming a single representation that condenses spatial and temporal information suitable for input into deep learning models. This approach allows spatial and temporal information to be condensed into a single image, facilitating model learning.

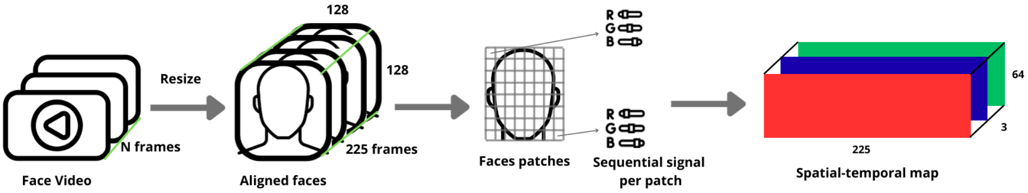

Figure 1. Process of generating a spatial-temporal map in RGB color spaces. Each video is divided into 255 frames, the patient's face is detected, aligned, and centered, and the STMaps are generated. Adapted from [Cheng et al. 2024].

As a reference model, the EfficientNet-B3 architecture is proposed, a convolutional neural network designed to optimize the trade-off between accuracy and computational complexity [Tan and Le 2019]. The choice of this model is justified by its consistent performance in computer vision tasks, combined with efficiency in terms of parameter count and computational cost.

In this study, SpO_2 estimation is formulated as a regression problem. Two-dimensional convolutional neural networks (2D CNNs) are applied to predict SpO_2 values from STMaps. The network's final layer will be replaced with a linear regression layer, allowing the prediction of continuous saturation values.

Performance will be evaluated using mean absolute error (MAE) and root mean square error (RMSE). Additionally, the results will be compared to the international tolerance standard of 4% , established for clinical pulse oximeters, ensuring adequate accuracy for remote SpO_2 monitoring [DeMeulenaere 2007].

[1] https://github.com/EmiBruning/Spo2-Estimation-from-Facial-Videos

4. Conclusion

In this work, the creation of a public and diverse dataset in the Brazilian context was proposed, with an emphasis on including different skin tones, ages, and medical conditions. Additionally, a plan was outlined to validate a deep learning approach for SpO_2 estimation from videos, using the EfficientNet-B3 architecture and STMaps to represent temporal and spatial signals.

As next steps, it is expected to complete data collection, train the proposed model, and conduct performance analyses, comparing different factors, skin tones, and recording conditions that may impact the final results. Furthermore, this work is expected to contribute to research in remote vital signs monitoring, providing reliable and inclusive tools for the non-invasive measurement of SpO_2.

Acknowledgments

The authors thank FAPERGS for the project: ARD/ARC – Process 24/2551-0000645-1.

References

Brasil (2018). Lei geral de proteção de dados pessoais (lgpd).

Cheng, C.-H., Yuen, Z., Chen, S., Wong, K.-L., Chin, J.-W., Chan, T.-T., and So, R. H. Y. (2024). Contactless blood oxygen saturation estimation from facial videos using deep learning. *Bioengineering*, 11(3).

Dasari, A., Prakash, S. K. A., Jeni, L. A., and Tucker, C. S. (2021). Evaluation of biases in remote photoplethysmography methods. *npj Digital Medicine*.

de Souza, A. M., Resende, S. S., de Sousa, T. N., and de Brito, C. F. A. (2020). A systematic scoping review of the genetic ancestry of the brazilian population. *Genetics and Molecular Biology*, 43:e20190036.

DeMeulenaere, S. (2007). Pulse oximetry: Uses and limitations. *The Journal for Nurse Practitioners*, 3(5):312–317.

Fitzpatrick, T. B. (1988). The validity and practicality of sun-reactive skin types i through vi. *Archives of Dermatology*, 124(6).

Machado Lunardi, G., Medeiros Machado, G., Al Machot, F., Maran, V., Machado, A., C. Mayr, H., A. Shekhovtsov, V., and Palazzo M. de Oliveira, J. (2018). Probabilistic ontology reasoning in ambient assistance: Predicting human actions. In *IEEE International Conference on Advanced Information Networking and Applications*.

Soares, T., Costa, R., Soares, E., Calderon, I., Lunardi, G., Valle, P., Guedes, G., and Silva, W. (2025). Machine learning-assisted tools for user experience evaluation: A systematic mapping study. In *Anais do XXI Simpósio Brasileiro de Sistemas de Informação*, pages 379–388, Porto Alegre, RS, Brasil. SBC.

Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning (ICML)*.

Wuyart, A., Abensur Vuillaume, L., Maaoui, C., and Bousefsaf, F. (2025). A survey of blood oxygen saturation assessment from video. *Biomedical Signal Processing and Control*, 110:108069.