Unobstrusive smartphone-based oxygen saturation measurement using a Meta-Region of interest

Firmin Leroy Kateu Demlabim, Gentian Jakllari, Emmanuel Chaput

Unobstrusive Smartphone-based Oxygen Saturation Measurement Using a Meta-Region Of Interest

Firmin Kateu[a,*], Gentian Jakllari[a], Emmanuel Chaput[a]

[a]*IRIT, Toulouse INP, CNRS, Université de Toulouse, 2 Rue Charles Camichel, Toulouse, 31000, France*

Abstract

We present SmartPhOx, a pure camera-and-flashlight smartphone-based pulse oximetry solution. We build on the ratio-of-ratios (RR) method and linear regression, an elegant approach resting on the Beer-Lambert law and landing itself to efficient smartphone implementations. However, its implementations without specialized hardware have so far proved to be unsuitable for clinical use, in particular due to the instability of the RR measurements. We use an empirical study to shed light on the reasons why and propose using the very RR measurements to filter RR measurements – a new paradigm we call the Meta-Region of interest (Meta-ROI). We design a complete-system architecture, including a novel data structure for storing and RR values in the time and space dimensions and an efficient algorithm for identifying Meta-ROI. Results from an Android implementation of SmartPhOx with the participation of 37 volunteers show that it is the first pure camera-and-flashlight solution to meet the FDA requirement for Root Mean Square Error (RMSE).

Keywords: Pulse oximetry, SpO_2, Mobile health, Ratio-of-ratios, Beer-Lambert law

1. Introduction

"*... A vast majority of Covid pneumonia patients I met had remarkably low oxygen saturations at triage — seemingly incompatible with life — but they were using their cellphones as we put them on monitors.*"[1]

What if their cellphones could have measured their oxygen saturation, how many of these patients would have seeked timely medical help and avoided intubation?

As the world is gripped by the COVID-19 pandemic, terms like oxygen saturation (SpO_2) and silent hypoxia[1] – the condition in which a patient still feels well but their SpO_2 is dangerously low [47] – have

*Corresponding author

Email addresses: firmin.kateu@toulouse-inp.fr (Firmin Kateu), jakllari@toulouse-inp.fr (Gentian Jakllari), emmanuel.chaput@toulouse-inp.fr (Emmanuel Chaput)

[1]https://www.nytimes.com/2020/04/20/opinion/sunday/coronavirus-testing-pneumonia.html

entered the mainstream. The pulse oximeter, the once-obscure fingertip device allowing home monitoring of the blood oxygen levels, has emerged as an important tool in fighting COVID-19, drawing attention to the science and technology behind it – and raising the question of whether it can be reinvented for the era of pervasive computing.

The idea of pulse oximetry – the non-invasive monitoring of oxygen saturation using the so-called ratio-of-ratios (RR) method – dates back to 1935 [33]. Scientifically it rests on the Beer-Lambert law stating that a light going through a thin body part, like a finger or earlobe, will be impacted by its thickness and concentration – the latter including oxygen saturation. Using two light beams of specific wave lengths and the fact that at different points of the cardiac cycles only oxygen saturation-related factors change, as we show in detail in Section 2, it is possible to manipulate the Beer-Lambert law through two consecutive ratios to remove non-oxygen saturation factors, like the medium thickness. The results is a relation between a *ratio-of-ratio (RR) of light measurements* and the SpO_2 – *the ratio-of-ratios (RR)* method [21]. The first pulse oximeter was developed in the '70s [44] and today a wide range of pulse oximeters can be found off-the-shelf [40, 37, 12, 25, 4]. Nevertheless, dedicated hardware adds extra burden and, as the silent hypoxia cases due to COVID-19 have revealed [47], often people are not aware their oxygen level needs monitoring.

Increasingly in people's hands and with advanced sensing, computing and communicating capabilities, the smartphone is seen as a building block of pervasive computing and key enabler of the digital healthcare [48, 8, 23, 27]. Researchers have proposed smartphone-based pulse oximetry solutions predating the COVID-19 crisis. [43] was among the first to apply the RR method for estimating SpO_2 using smartphones. A user places the finger over the flashlight – serving as the source of light – and the camera. Acquiring the photoplethysmogram (PPG) signal from processing the resulting video allows RR of light measurements and the estimation of SpO_2. However, its accuracy is below the FDA clearance threshold [3]. The fundamental reason is that it uses linear regression for implementing the RR method. Unfortunately, the PPG signal, and thus the RR measurements, can be unstable due to finger movements and pressure changes [10]. To address this issue, [11, 32] integrate into the RR measurements the camera quantum efficiency. While improving accuracy, this is information to which only manufacturers have access. PhO_2 [10] proposes attaching to the smartphone camera a custom-made device mounted with two chromatic filters, each allowing a precise wavelength to pass. The result is a system allowing SpO_2 predictions with accuracy meeting the FDA clearance threshold. Nevertheless, the custom-built hardware add-on, while manufactured with the help of 3D printing, limits its large-scale application. Recently, dedicated oxygen monitoring sensors

are being integrated in smartwatches [22], and some high-end smartphone models [1]. While accurate, such solutions leave out large sections of users with older smartphone models, particularly in developping countries.

In this paper, we introduce SmartPhOx, a smartphone-based pulse oximetry system meeting the FDA clearance threshold [3] for accuracy while relying only on the standard smartphone camera and flashlight. To achieve this, we start by first designing and conducting an empirical study aimed at shedding light on the underlying reasons behind the inaccuracy of pure camera-and-flashlight solutions. The data shows that focusing on primary factors – the quality of the PPG signal [39, 32, 11] or identifying the right region on the video [16, 43], known as the region of interest (ROI) – is misleading. We find that signals of excellent quality can still lead to unstable RR measurements. Focusing on a particular area of the video frame, such as the center, does not help either. In light of these results, we argue for a shift in approach. We propose foregoing the primary factors and instead leveraging the RR measurement values themselves for identifying stable RR measurements.

Using RR measurements to essentially filter RR measurement leads to the idea of Meta-Region of Interest – Meta-ROI, the key innovation underpinning SmartPhOx. However, transforming the Meta-ROI idea into a complete system solution running on off-the-shelf smartphones raises several challenges. First, using RR measurements to filter RR measurements requires defining what is a good RR. Second, once the good RR defined, we need an approach for automatically identifying the good RR values using camera videos as input and the processing capabilities of off-the-shelf smartphones. In short, we address these challenges by introducing a new data structure for RR measurements, we refer to as the RR Map, and two approaches representing different tradeoffs in terms of complexity and accuracy in computing SpO_2. The first approach relies on a simple linear regressor and therefore needing more stable RR measurements. The second approach use of a convolutional neural network (CNN) on the whole RR map in order to take advantage of the spatial redundancy.

Throughout this paper, we make the following contributions:

- We show that the primary factors for filtering ratio-of-ratios measurements are misleading. We shed light on the reasons why and introduce Meta-ROI – a new paradigm for identifying good RR measurements (Section 3).
- We introduce a new data structure for RR measurements, the RR Map, that enables the definition of good RR values (time-and-space consistent) (Section 5). Leveraging it, we develop an efficient algorithm for

identifying Meta-ROI (Section 6).

- We design SmartPhOx, a complete-system architecture leveraging the concept of Meta-ROI for smartphone-base pulse oximetry (Section 4).
- We introduce a convolutional neural network (CNN) architecture designed to extract relevant spatial features and regress SpO_2 from the entire RR map, using thus all the information available. We use it to design SmartPhOx v2, a complete-system architecture for measuring SpO_2.
- We implement the two versions of SmartPhOx as a standalone Android application and evaluate them with data collected from 37 volunteers. The results show that SmartPhOx is the first pure camera-and-flashlight smartphone-based solution to meet the FDA requirement for Root Mean Square Error (RMSE) [3]. Furthermore, the evaluation shows a trade-off between complexity and accuracy of the two SmartPhOx versions, with the CNN version offering better accuracy at the cost of more complexity (Section 9) .

2. Primer on the ratio-of-ratios (RR) method

In this section, we introduce the ratio-of-ratios (RR) method widely used for smartphone-based pulse oximetry [43, 11, 32, 9, 10, 16] and adopted by SmartPhOx.

2.1. Theoretical underpinning

The RR method for measuring SpO_2 rests on the law of Beer-Lambert describing the attenuation of light as a function of the traversed material. Mathematically: $I(\lambda) = I_0(\lambda)\exp^{-\epsilon(\lambda)\rho d}$, where $I_0(\lambda)$ is the incident light intensity, $\epsilon(\lambda)$, the absorptivity for the wavelength λ, ρ, the medium concentration and, d, the path length through the medium. The equation can be expressed in a form landing itself to practical systems for estimating SpO_2. Let us start by expressing it at the two extremes of the cardiac cycle : in diastole, where $d = d_{min}$, and systole, where $d = d_{max}$.Let $I_d(\lambda)$ and $I_s(\lambda)$ denote the corresponding $I(\lambda)$ values. Taking the logarithm of their ratio, we get:

$$L(\lambda) = ln(\frac{I_s(\lambda)}{I_d(\lambda)}) = (d_{min} - d_{max}).(\epsilon_{O_2}(\lambda)\rho_{O_2} + \epsilon_{H_b}(\lambda)\rho_{H_b}) \tag{1}$$

While more pratical, Eq. (1), requires measuring d_{min} and d_{max}. To relax this requirement, we can use the ratio of two values corresponding to two different wavelengths, λ_1 and λ_2:

$$RR_{\lambda_1,\lambda_2} = \frac{L(\lambda_1)}{L(\lambda_2)} \tag{2}$$

– hence the name **ratio-of-ratios**. Recognizing that $SpO2 = \frac{\rho_{O_2}}{\rho_{O_2}+\rho_{H_b}}$, where ρ_{O_2} and ρ_{H_b} denote the oxygen-saturated and oxygen-unsaturated hemoglobin, respectively, and dividing the numerator and denominator of Eq. 2 by $\rho_{O_2} + \rho_{H_b}$, we get

$$SpO2 = \frac{\epsilon_{H_b}(\lambda_1) - \epsilon_{H_b}(\lambda_2)RR_{\lambda_1,\lambda_2}}{(\epsilon_{O_2}(\lambda_2) - \epsilon_{H_b}(\lambda_2))RR_{\lambda_1,\lambda_2} + \epsilon_{H_b}(\lambda_1) - \epsilon_{O_2}(\lambda_1)}. \tag{3}$$

2.2. Ratio-of-ratios on smartphones using linear regression

Equation (3) cannot be implemented on off-the-shelf smartphones without knowledge of all the coefficients. However, studies [35, 43] have shown that it can be approximated using a linear model as follows:

$$SpO2 = A \times RR(\lambda_1, \lambda_2) + B \tag{4}$$

This equation enables the implementation of the ratio-of-ratio method on any smartphone using linear regression. RR values are measured empirically and used to train a linear regression model for estimating the coefficients A and B.

2.2.1. Measuring RR values on smartphones

The RR expression, Eq. (2), can be simplified by introducing $\delta(\lambda) = I_s(\lambda) - I_d(\lambda)$. Indeed $\frac{\delta(\lambda)}{I_d(\lambda)}$ is small – the absorbance of the blood changes lightly from systole to diastole. As a result, $RR(\lambda_1, \lambda_2)$ measurements can be made using

$$RR(\lambda_1, \lambda_2) \approx \frac{\frac{\delta(\lambda_1)}{I_d(\lambda_1)}}{\frac{\delta(\lambda_2)}{I_d(\lambda_2)}}. \tag{5}$$

This approximation is significant since measuring systolic and diastolic intensities *per se* is not necessary anymore. We measure instead a base (constant) intensity $DC_{\lambda_1} = I_d(\lambda_1)$ and variations $AC_{\lambda_1} = \delta(\lambda_1)$ over this baseline, significantly simplifying the implementation.

3. Smartphone pulse oximetry : Challenges and opportunities

Unlike dedicated pulse oximeters, smartphones use a flashlight covering a wide spectrum (400-800nm [10]) and a high resolution three-channel bitmap camera. Measuring oxygen saturation using a smartphone requires carefully applying the ratio-of-ratios method introduced in Section 2.2 on a video sequence. It involves measuring RR[2] values using Equation 5, followed by linear regression for estimating the A and B

[2]For simplicity we write RR instead of $RR(\lambda_1, \lambda_2)$.

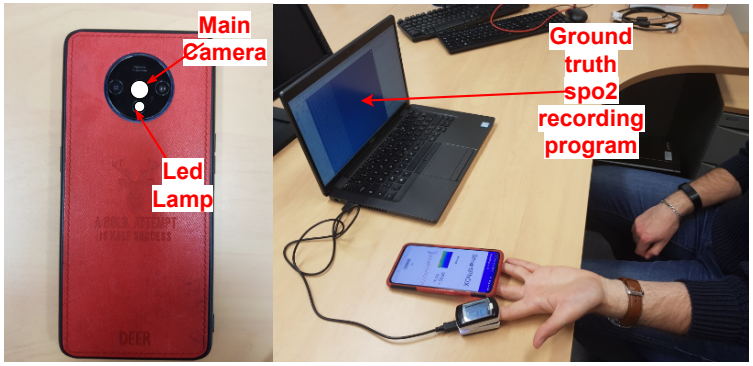


Figure 1: Experimental setup

coefficients of Equation 4. Therefore, the challenge in accurately estimating SpO_2 using a smartphone lies in how the RR are measured, both during training and inference.

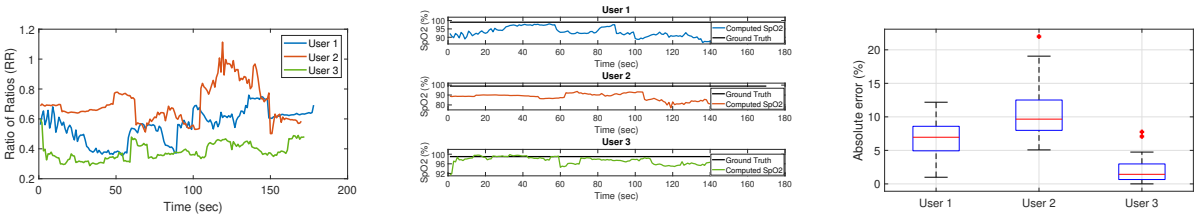
3.1. Baseline approach for measuring RR

The baseline approach for measuring RR, the basis of most works on this topic [43, 11, 10, 16], starts with a video of the subject's finger placed over the camera and flashlight. The average intensity of each channel for every video frame is collected resulting in three photoplethysmogram (PPG) signals, one for each channel. The AC/DC ratio is then calculated for each signal: taking as AC the amplitude of the oscillations of the PPG signal, and as DC the baseline of the signal. Taking as λ_1 the red channel and λ_2 the green (or blue) channel, the *RR* is finally obtained using Equation 5.

3.2. Analyzing the baseline approach

The objective of this section is not a thorough and large-scale analysis of the baseline approach for measuring RR. It is instead to introduce the simplest test case capable of shedding light on the complexities of the RR measurements on a smartphone and their underlying reasons.

Experiment: We design and conduct a controlled experiment using the setup depicted in Fig. 1 with three different users exhibiting healthy and stable SpO_2 levels (around 99%). Each user sits in a comfortable position and places their hand on a table with the palm facing up. The user's middle finger is placed on the camera of a OnePlus 7T smartphone running a custom application collecting video data, while the index finger is connected to a CMS-50E Pulse Oximeter[40] for establishing the ground truth (more details in Section 8.3). We train the linear regression model using the SpO_2 variation protocol presented in Section 8.2, and test it while SpO_2 is naturally constant.



(a) Raw RR values (b) Predicted SpO_2 (c) SpO_2 prediction errors

Figure 2: SpO_2 measurements using the ratio-of-ratio (RR) method on three different users exhibiting SpO_2 around 99%. The RR values vary significantly (Fig. 2a) even if the ground truth SpO_2 remains constant throughout the experiment – making linear regression extremely challenging. The result is a significant amount of errors in the predicted SpO_2 values (Fig. 2b, Fig. 2c).

Results: Fig. 2 plots the RR values, predicted SpO_2 and the prediction errors for all three users. Fig. 2a shows that while the SpO_2 levels are constant throughout the experiment the RR values are highly unstable. This results casts serious doubts on the feasibility of using Equation 4 for estimating SpO_2 on smartphones – no values for the A and B coefficients could associate the RR values observed in Fig. 2a to the same SpO_2 value. It is therefore no surprise that Fig. 2b and Fig. 2c show highly erroneous SpO_2 predictions.

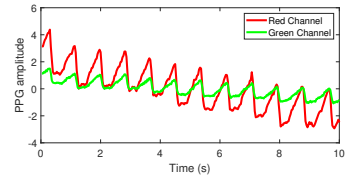
Implication or the case for consistent RR: This section's test study shows that the ratio-of-ratios method can be undermined by inconsistent RR measurements. Therefore, the smartphone-based pulse oximetry challenge reduces to the challenge of consistent RR measurements. Qualitatively, we refer to RR measurements as consistent if for a given SpO_2 level the RR measured using a smartphone are similar.

3.3. The quest for consistent RR values

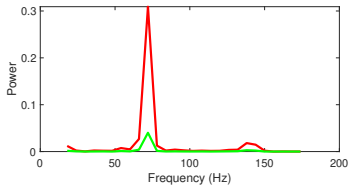
The baseline approach being highly inaccurate due to highly unstable RR values, different approaches have been proposed for acquiring better RR values. Certain approaches have proposed custom add-on hardware [10] filtering the flash light to allow only a precise wavelength. Aiming for solutions without hardware add-on, other approaches have focused on the primary factors behind the RR values. The RR being measured off the PPG signal, most focus has been on the PPG signal quality [39, 32, 11] while some focus on a particular region of the frame [16, 43]. In the following, we investigate the approaches requiring no hardware add-on.

3.3.1. The curious case of the PPG signal quality

With the RR a function of the PPG signal, a reasonable direction is to first acquire a good quality signal before applying the RR method. We investigate this approach empirically:



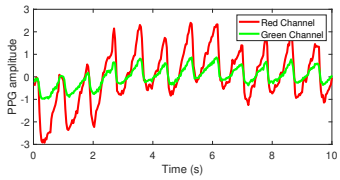
(a) PPG signals (RR = 0.7381)



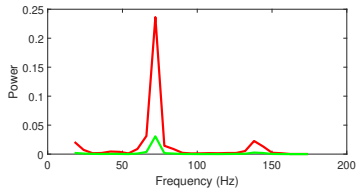
(b) PSD (RR = 0.7381)

Q_{kurt}		**Absolute error**
Red	Green	5 %
0.95	0.94	

(c) Signal quality and SpO_2 error (RR = 0.7381)



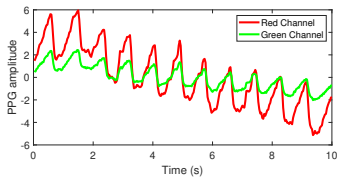
(d) PPG signals (RR = 0.7562)



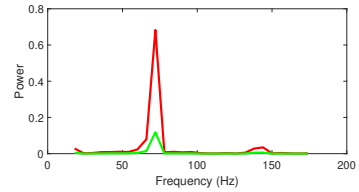
(e) PSD (RR = 0.7562)

Q_{kurt}		**Absolute error**
Red	Green	6.1 %
0.90	0.92	

(f) Signal quality and SpO_2 error (RR = 0.7562)



(g) PPG signals (RR = 0.6653)



(h) PSD (RR = 0.6653)

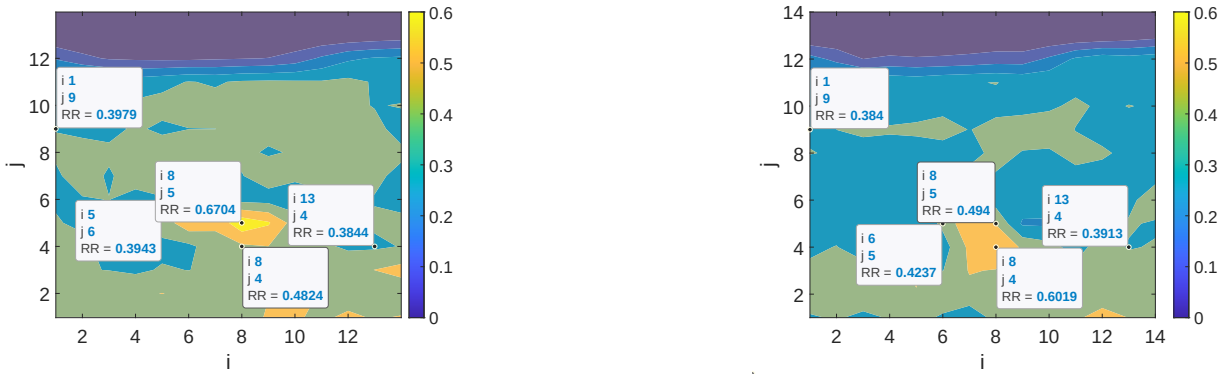
Q_{kurt}		**Absolute error**
Red	Green	0.1 %
0.94	0.94	

(i) Signal quality and SpO_2 error (RR = 0.6653)

Figure 3: PPG signals off which three RR values of Fig. 2 are computed. The signals are very similar in terms of frequency and heart rate evaluation – highest peak around the actual heart rate frequency (72 bpm), with spectrum having very similar shape and Q_{Kurt} values. Nevertheless, the respective RR values are very different, as is the quality of the SpO_2 prediction.

Methodology: To evaluate the relation between signal quality and RR consistency, we look back at the data of Fig. 2. We select three RR values – two among those leading to erroneous SpO_2 predictions and one among those leading to the accurate SpO_2– and analyze the respective PPG signals. Since the source of the PPG signal is the cardiac activity, we use Q_{kurt} in our analysis, a metric quantifying the purity of a signal related to cardiac activity [36]. Specifically, $Q_{Kurt}(s) = \frac{kurtosis(FFT(s))}{kurtosis(P_s)}$, where FFT is the Fast Fourier Transform and, P_s, the perfect sine wave with frequency corresponding to the heart rate.

Results: Fig. 3 shows that the RR values under consideration are computed off excellent PPG signals. The respective red and green channel signals exhibit their highest peaks around the ground truth heart rate (72 bpm). The Q_{kurt} values of all signals are nearly perfect. Nevertheless, the RR values are highly different. More important, two of the RR values lead to erroneous SpO_2 predictions.



(a) t_1 (b) t_2

Figure 4: RR Map at two time instances, t_1,t_2 of the same video collected while SpO_2 is constant and equal to 98%. i, and j denote the image cell indices.

Implication: While involving only three RR values, the data demonstrates that excellent PPG signals can lead to highly different RR and SpO_2 prediction – even if the ground truth SpO_2 is constant. As a result, the signal quality can be a poor proxy for consistent RR values. In Sec. 3.4, we provide an intuition as to the reasons behind this finding.

3.3.2. Location, location, location – and a new (RR) map

Instead of using PPG signal quality, an alternative approach is to simply use the central region of the image for all RR calculations and SpO_2 predictions [16, 43]. The intuition being that lighting conditions should be more uniform in this area, leading to consistent RR values.

Methodology: To evaluate the physical location-based approaches, we introduce what we refer to as the *RR Map*. The input frame is divided into *cells* and for every cell a PPG signal and an RR value are computed – the set of all the cell RR values of a particular video input constitutes its RR Map.

Results: Fig. 4 shows the RR Map at two different time instances of the data of Fig. 2. The data leads to two main observations: First, RR values from a specific region (central or not) can be highly inconsistent – they vary significantly in time and space even if the ground truth SpO_2 is constant. Focusing on cells from the central region – (6,5), (8,5) and (8,4) – shows that the respective RR values are very different. Furthermore, they vary significantly from one time instance to the other. A second observation is that RR values from physically-distant cells can be consistent. Zooming in on cells (1,9), (13,4) shows that their respective RR values are very similar and remain stable from one time instance to the other.

Implications: The RR Map values of cells (6,5), (8,5) and (8,4) underline the difficulty of reliable SpO_2 predictions using a fixed region of the frame in particular, and a physical region in general. On the other hand, the fact that the physically distant cells, (1,9), (13,4), have similar and stable RR values is leveraged in Section 3.5 for introducing a new way of selecting consistent RR values.

3.4. RR (in)consistency – the underlying reasons

The model based on Beer-Lambert's law has some limitations. It assumes that monochromatic light rays pass completely through the finger and are reflected specularly back to the camera, ignoring the complex phenomena of scattering and refraction. Light rays arriving at the camera sensor undergo an optical path that is subject to these phenomena, including intermediary reflections inside the finger, and whose impact can depend on the smoothness of the incidence region and the angle of incidence. It is as if each pixel of the camera is subject to rays following different virtual paths through the finger, producing different RR values for the same SpO_2.

Furthermore, the temporal volatility of the RRs can be explained by the fact that these virtual paths change, depending on the disturbance generated by the micro-movements of the finger. Some paths are however more stable than others: the paths for which the overall configuration of the finger surface, the camera and the flash do not change much, despite the micro movements. The Meta-Region of Interest introduced in section 3.5 is aimed at identifying the more stable paths.

3.5. A way forward : The Virtual Sensor Array

In this work, we propose a new way to deal with the camera output in the context of smartphone-based oximetry; The idea is to consider each portion of the image as the output of a small independent SpO_2 sensor, placed on top of the corresponding finger part (See Fig. 5). Thus, from a single large-area sensor, we move to an array of smaller sensors producing independent but correlated data. This proposal is motivated by the reasons behind the spatial and temporal instability of the RR across the frame, presented in section 3.4. Indeed, while the finger is not perfectly flat and never applied with the same pressure, there must be regions of the fingers for which the light pathways are similar. These regions need not be contiguous in time and space. The intuition is supported empirically by the RR Map of Fig. 4 showing cells (1,9) and (13,4), non-contiguous in space, produce similar RR values. The challenge, however, is automatically identifying the regions leading to consistent RR values. Primary factors, the focus of previous works are shown to fail: PPG signal quality is shown to be a poor proxy; the areas with consistent RR values are not

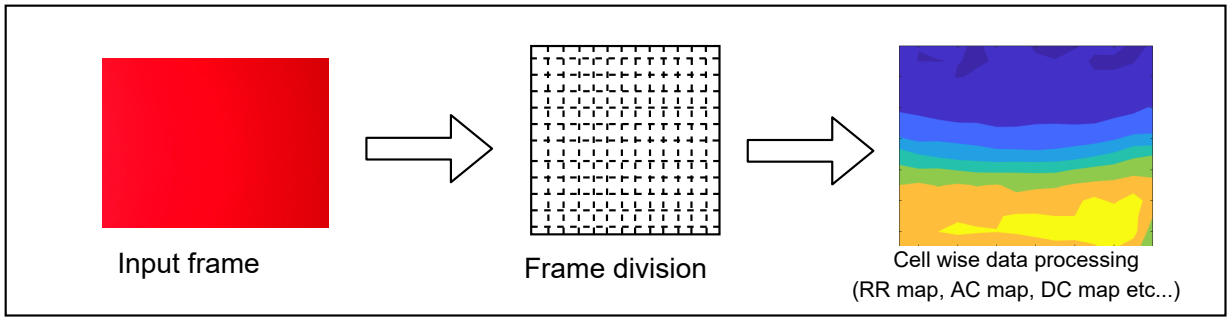


Figure 5: Virtual sensor array principle. From one single large frame to cell wise data processing, resulting in RR map, AC map, DC map etc.

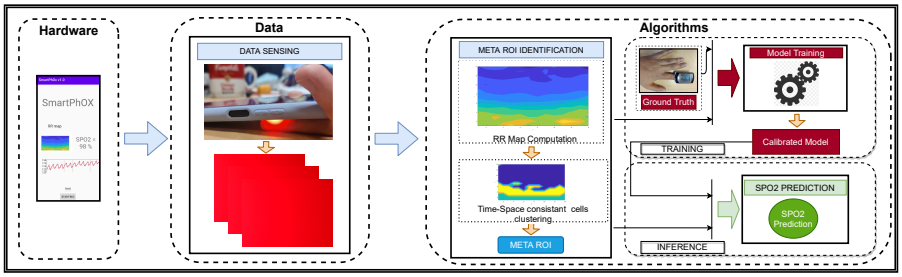


Figure 6: SmartPhOx System Architecture

necessarily contiguous, excluding an approach based on a particular physical area. We propose to forego using primary factors and to rely on the the RR values themselves to identify good RR values. Using RR values to essentially filter RR values leads to the idea of Meta-Region of Interest – Meta-ROI, the central element of SmartPhOx v1. Turning this idea into a robust smartphone-based solution raises several scientific and system challenges, which we detail and address in the following sections 4, 5 and 6.

However, such a solution implies throwing away a significant part of the RR in the RR map. This is a waste, since even if they are inconsistent with each other, they remain dependent on the SpO_2. Thus, we investigate in a second phase a way to combine efficiently all the RR of the RR map, in order to use all the available information and to reach a better accuracy. For this, we designed a convolutional neural network, trained on the RR map to deduce the SpO_2. This second version of the solution, called SmartPhOx v2, is presented in sections 7. Finally, we analyse the trade-off between the two solutions in section 9.7.

4. SmartPhOx system overview

Fig. 6 shows a high-level depiction of SmartPhOx's architecture. It comprises three modules:

1. Hardware: The SpO_2 measurement starts with the subject placing the finger on the smartphone flashlight and camera.
2. Data sensing: The smartphone camera generates a video during the measurement session. Section 5 introduces methods for customizing the video recording and transforming the data into the RR Map.
3. Algorithms: Section 6 formalizes the notion of consistent RR values and introduce an algorithm that take as input the RR Map and identifies the Meta-ROI. The latter is used for estimating SpO_2 using the ratio-of-ratio method described in Section 2.2. Section 7 introduces an alternative approach for identifying consistent information from the RR Map and regress to the SpO_2 using neural networks.

5. Data – RR Map Construction

5.1. Data sensing

The first step in SmartPhOx is recording a video session while the user places the finger on the smartphone flashlight and camera. Selecting its duration involves satisfying two constraints. It needs to be long enough to allow the calculation of several RR values for identifying time-consistent RR values. And, a single RR calculation requires a few seconds of PPG signal [43, 5]. Let w denote the PPG signal length for a single RR calculation and Z the number of consecutive RR values necessary for training and prediction. SmartPhOx calculates RR values using a sliding window of size w. Thus, the video session duration is $T = Z + w - 1$ seconds.

5.2. RR Map computation

Once a video consisting of $30 \times T$ 3-channel (Red, Green, Blue) frames is obtained, the frame surface is divided into $X \times Y$ *cells*. The choice of X and Y represents a tradeoff. Higher values translate to more cells and a finer the segmentation of the frame, enabling a more precise selection of the RR values. However, this leads to smaller individual cells with less data on their surface, making their RR more sensitive to noise induced by the camera acquisition chain. We evaluate this trade-off in Section 9. To compute the RR of a cell from the PPG signal, recall from Sec. 2.2 that $RR \approx \frac{AC_{\lambda_1}/DC_{\lambda_1}}{AC_{\lambda_2}/DC_{\lambda_2}}$. Using the green and red PPG signals as λ_1 and λ_2 of each cell, we compute its RR by selecting as AC the standard deviation of the filtered signal, and as DC the average of the raw signal.

For every cell, (x, y), $x \in \{1, 2, ..., X\}$, $y \in \{1, 2, ..., Y\}$, we obtain a vector

$$RR^{(x,y)} = \left[rr_1^{(x,y)}, rr_2^{(x,y)}, ..., rr_Z^{(x,y)} \right] \tag{6}$$

where Z is the number of RR values computed over the window T. Therefore, the RR Map can be seen as a set of $X \times Y$, Z-dimension vectors, with $X \times Y$ denoting the space dimension and Z, the time dimension. For the rest of the paper, we use the terms *cell* and *Z-dimension vector* interchangeably.

6. Meta-ROI Algorithm

The basic premise of our work, as developed in Section 3, is that accurately estimating SpO_2 requires consistent RR values. In this section, we formalize the notion of consistency and introduce an algorithm for identifying the most consistent RR values – the Meta-ROI.

6.1. Space-time consistency in the RR Map

To formalize the notion of RR consistency, first introduced qualitatively in Section 3.2, we draw on the empirical study of Section 3 and the cluster analysis. With the data showing RR values vary across frame regions and time, we define consistency in space and time. By construction, the RR Map includes the space and time dimension. Therefore, we consider RR cells to be consistent if they belong to the same cluster produced by a clustering algorithm applied on the RR Map. The clusters themselves are considered time-consistent regions. Formally:

Definition 1 (Space-time consistency). *Let $S = \{S_1, S_2, ..., S_k\}$ be a clustering of the RR Map cells. Two cells are considered consistent in space and time if they belong to the same cluster in S. The clusters $\{S_1, S_2, ..., S_k\}$ are referred to as space-time consistent regions.*

6.2. Meta-ROI algorithm

In this section, we address the challenge of identifying the *best* among the space-time consistent RR Map regions. Referred to as the the meta-region of interest (Meta-ROI), it includes the RR values SmartPhOx's linear regression model will eventually associate with a particular SpO_2 value.

A straightforward solution could be to approach this challenge as fundamentally a clustering problem and simply use an efficient heuristic for k-means. However, owing to its origins as a quantization technique [34], there is no simple way to choose the k parameter. More important, our objective is not to reduce the dimensionality of the RR Map but rather to identify the Meta-ROI.

Our solution to this two-pronged problem is a divide-and-conquer approach. We first address the challenge of identifying the best among the space-time consistent RR Map regions, assuming the k parameter is known. Subsequently, we focus on addressing the challenge of identifying the k parameter.

Algorithm 1: Meta-ROI algorithm

Input: $RR\_MAP[X][Y][Z]$

Output: The meta region of interest, $Meta\_Roi$

```
for K := 2 to K_MAX do
    Centroids[K][Z] = k_means (RR_MAP, K);
    Compute DB(Centroids[K]), using Eq.(8);
    if DB_k < minimum_DB then
        minimum_DB = DB;
        for i := 1 to K do
            Calculate cv(Centroids[i]) using Eq.(7);
        Meta_Roi = cluster with minimum cv;
return Meta_Roi;
```

To identify the best space-time consistent region, we introduce a new consistency metric. The metric needs to satisfy two requirements. It needs to quantify the consistency of a given cell cluster. Moreover, it needs to allow a meaningful comparison of the k clusters with different numerical values so as to identify the Meta-ROI. To meet these requirements, we use the coefficient of variation. It measures the dispersion of a population, allowing to quantify the consistency of a given cluster. And it is normalized, enabling a fair comparison between different clusters. Specifically:

Definition 2 (Consistency metric). *Let $S = \{S_1, S_2, ..., S_k\}$ be a clustering of the RR Map cells and $C = \{C_1, C_2, ..., C_k\}$ the respective set of the Z-dimension centroids. The time-consistency metric of a cluster, S_i, is defined as the coefficient of variation of its centroid vector, $\sigma(C_i)$:*

$$cv_i = \frac{\sigma(S_i)}{\mu_i} = \sqrt{\frac{1}{Z}\sum_{z=1}^{Z}(C_{i,z} - \mu_i)^2} \times \frac{1}{\mu_i} \tag{7}$$

$$where\ \mu_i = \frac{1}{Z}\sum_{z=1}^{Z} C_{i,z}.$$

Identifying the best value of the parameter k is a decades-old problem [13, 28] with no simple solution. The naive approach of iterating over different values of k until the consistency metric of Definition 2 is

minimized would not work as it could converge to trivial, single-cell clusters. To strike a balance between space-time consistency and region size, we couple the consistency metric with the Davies-Bouldin (DB) index [13], one of the classic validity indices for analysing clustering. Unlike its main alternative, the silhoute [28] index, which is focused on the cluster density, largely addressed by the consistency metric, the DB index rewards the creation of distinct clusters. Mathematically,

$$DB = \frac{1}{k} \sum_{i=1}^{k} max_{j \neq i} \left(\frac{d_i + d_j}{d_{ij}} \right), \tag{8}$$

with k the number of clusters, $d_i(d_j)$, the average (Euclidean) distance of all cells in cluster $i(j)$ from its centroid, and d_{ij} the distance between the centroids of clusters i and j.

The consistency metric and the DB index pave the way for our algorithm for identifying the Meta-ROI (sketched in Algorithm 1). It proceeds by making consecutive calls to a k-means algorithm with increasing values of the parameter k (lines 1, 2) up to a limit of K_MAX. Since the DB index is smallest for well-distinct clusters, the algorithms looks to minimize it (line 4). Every time a clustering with a smaller DB index is identified, the consistency metric is used for identifying the best cluster (line 5). The algorithm returns the most consistent cluster of the clustering with the smallest DB index as the Meta-ROI. A key parameter of Algorithm 1 is obviously the K_MAX. In our experiments, the smallest DB index was reached with k between 2 and 6, so we set the default value of K_MAX conservatively to 10.

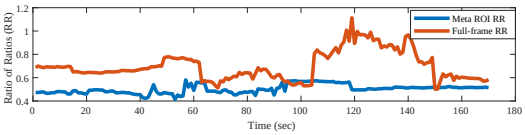
6.3. Illustration

To illustrate the impact of the Meta-ROI algorithm, we circle back to the data of Fig. 2 (Section 3.2). In the interest of clarity, we zero in on one of the users (user 2).

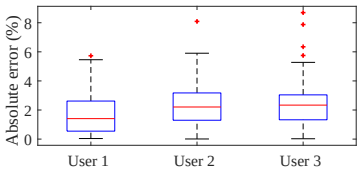
Fig. 7 plots the RR values over the experiment time and the corresponding SpO_2 prediction error against the ground truth. The data shows that the meta-ROI algorithm meets its objective of identifying consistent RR values (Fig. 7a). It leads a significant improvement in absolute prediction error when compared to the baseline approach (Fig. 7b). In Section 9, we conduct a thorough performance evaluation of our system.

7. Exploiting the spatial redundancy

The space-time consistent region obtained after the clusuring described in section 6.1 are all of them affected by the SpO_2, since their corresponding light rays goes through blood. Indeed, each consistent region exhibits a correlation with the SpO_2, as can be seeing in Fig. 8. Using the RR produced by the



(a) Baseline RR for user 2 vs Meta-ROI RR.



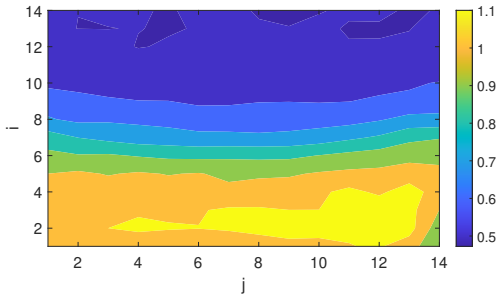
(b) SpO_2 prediction error for all users when applying SmartPhOx

Figure 7: SpO_2 Prediction with the proposed method, over data from Fig. 2. Our method lead in more stable RR, thus less prediction error.

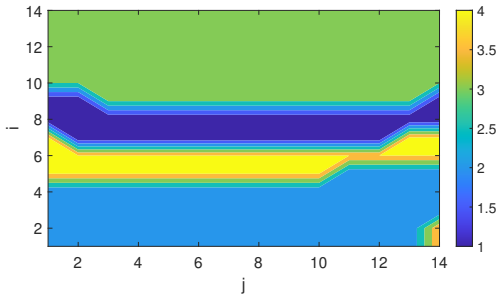
Meta-ROI alone to train the linear regressor ensures that only the best data are retained, but implies some loss of information. Indeed, we lose the RR value of some cells, as well as the relative values of more or less neighbouring cells. In this section, we examine whether it is possible to avoid such a loss. Answering this question raises two challenges. The first (a) is how to efficiently extract all the information contained in the RR map, and the second (b) is how to adapt the regression model with the new input data thus obtained. In the next section, we present how we address these two challenges, leading to a second version of our system, called SmartPhOx v2.

7.1. What regression model for SmartPhOx v2?

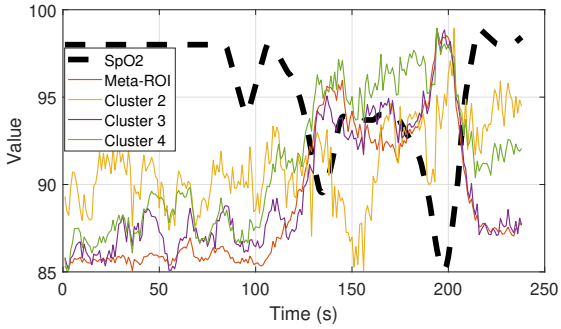
To address the dual challenge presented above, we propose to train a Convolutional Neural Network (CNN) to learn the appropriate regression function from the RR map to the SpO_2. The CNN has already been used to infer the SpO_2 [16], but directly from the raw PPG signals. This method is less efficient because PPG signals contain mostly SpO_2 independent information. They are indeed more influenced by skin colour, lighting and heart rate. The use of a CNN on the RR map has the advantage of focusing the regressor on the ratio of ratios, which are directly correlated to the SpO_2. The RR map also contains physical information in the form of the correlation between the cells. Indeed, the physical distance between the areas of the finger corresponding to each cell is closely related to the difference in path lengths of the light rays reaching these cells. The exact correlation is difficult to model but can be captured by a carefully designed neural network. Finally, the neural network can derive a more robust regression function for estimating the SpO_2 from the RR than that derived from Lambert's law, which suffers from many simplifying assumptions. The overview of SmartPhOx v2 is depicted in Fig. 9.



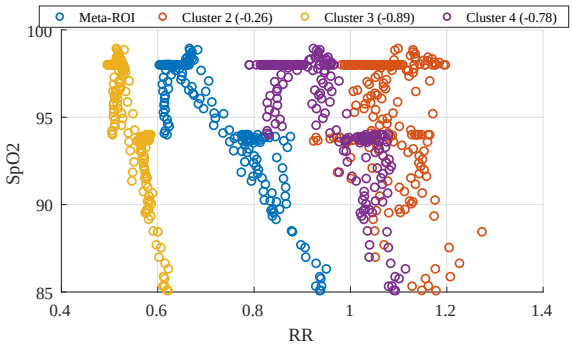
(a) Average RR map of the experiment.



(b) Clusters location with respect to the RR map.



(c) Scaled cluster centroid RR and SpO_2 in time.



(d) Scatter plot of SpO_2 and clusters centroids RR, with their pearson correlation in brackets.

Figure 8: After recording the SpO_2 and the RR map during a user's breath-holding session, we show here the evolution of the RR of the centroid of each cluster with respect to the SpO_2. The Pearson correlation between the RR and the SpO_2 is also shown in brackets. We notice that not only the Meta-ROI is correlated with the SpO_2, but also all the other cells, represented here by the centroid of the cluster to which they belong.

7.2. Neural Network Architechture for SpO_2 estimation

Unlike the previous version of SmartPhOx which used only one RR value –the Meta-ROI RR– to estimate SpO_2, the neural network of SmartPhOx v2 is designed to take as input all the RR values available in the RR map, thus encompassing as much information as possible. In addition to the RR map, we add the AC map and the element-to-element ratio of the DC map of the two colour channels. This corresponds to an input data of size $X \times Y \times 4$. The intuition behind these additions is that by observing the relative change in intensity of a cell across these 4 maps, the network would be able to identify cells more influenced by noise than by SpO_2.

The architecture of our 2D convolutional neural network is shown in Figure 10. It consists of two convolution blocks stacked together. A convolution block consists of three convolutional layers each followed by a batch normalization layer to facilitate learning convergence, and a max pooling layer to reduce in-

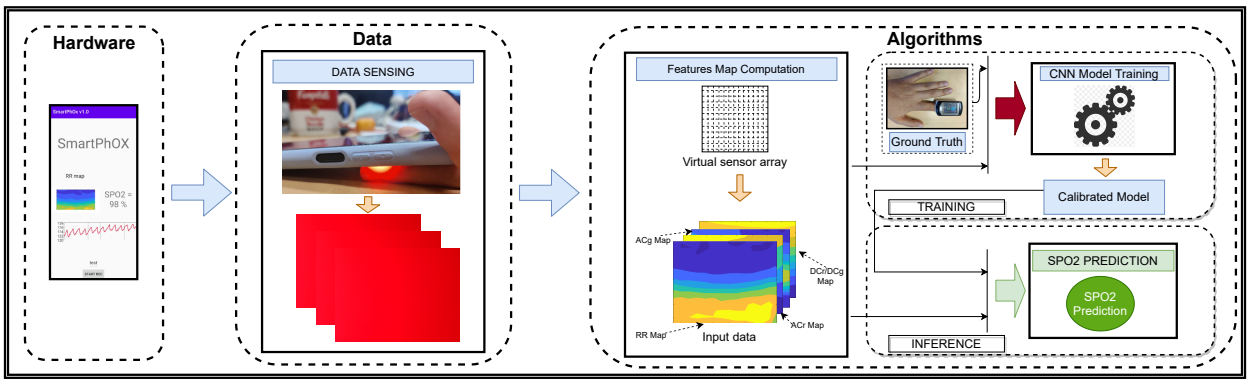


Figure 9: SmartPhOx v2 System Architecture.

Table 1: Neural network hyperparameters investigated.

Number of Conv blocks	1,2
Number of conv layers per block	1,2,3
Kernel size	3,5
Number of filters per conv	10, 16, 20
Dropout rate	0.1, 0.2, 0.3

put dimensions. The number of convolutional layers in each convolution block as well as the number of convolution blocks are studied as hyperparameters. A dropout layer is superimposed on the ensemble to avoid overfitting, with a dropout rate of 10%. The last layer is the output neuron, which corresponds to the SpO_2. The activation functions used are the exponential linear unit (ELU) for the convolutional layers, and the identity function for the output neuron. We use a cross validated grid search process to find the best combination of hyperparameters. The space of investigated hyperparameters is presented in Tab. 1. We also use early stopping and L2 norm regulariser on the convolutional layers to avoid overfitting. We optimise our model in mini-batches using adaptive momentum (ADAM)[29].

8. Implementation and dataset

8.1. Implementation

We implemented the two version of SmartPhOx as standalone Android application. The signal processing component is implemented using the IIRJ library [41]. The $k-$mean clustering is implemented in Java. The Deep learning part is implemented in TensorFlow with Keras interface and ported to the mobile as a

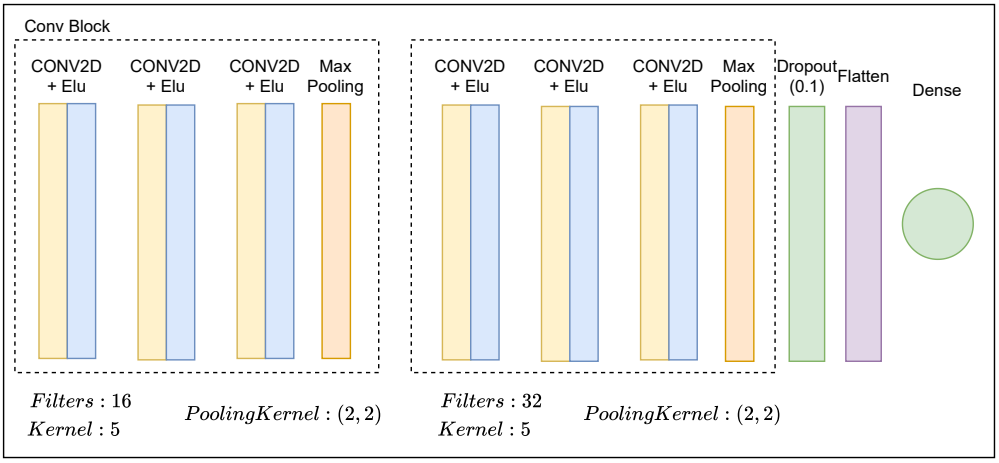


Figure 10: SmartPhOx v2 CNN architechture with best hyperparameters.

Table 2: SmartPhOx v1 implementation parameters.

w	10 s
T	15 s
RR Map Z-dimension	6
RR Map cell size	94×56 px
Video resolution	1260×720 px

tflite [46] model with float32 precision. Table 2 shows the default parameter values used in the implementation of SmartPhOx v1 (we evaluate the impact of these values on the performance of SmartPhOx v1 in Section 9.2). Table 3 shows the processing times of key processing bloc of SmartPhOx on off-the-shelf smartphones.

8.2. SpO_2 variation protocol

Ideally, we would test SpO_2 on subjects suffering from hypoxia, especially COVID-19 patients but in the current context it proved infeasible. Therefore, we have developed a protocol for inducing the oxygen level variation in healthy volunteers. The protocol starts with breathing normally for the first 30 s followed by a stop-n-go process of breathing/holding their breath, exhaling/holding their breath. The objective is to induce a gradual decrease and increase of SpO_2, thus generating a richer set of values. In particular, the volunteer is asked to take a deep breath and then hold it until starting to feel discomfort, then (b) exhale, followed by holding the breath until feeling discomfort again. At this point the SpO_2 reaches its low point, typically in

Table 3: SmartPhOx processing time on various phones. SmartPhOx v1 overall processing time is the sum of the first two columns, while SmartPhOx v2 overall processing time is the sum of the first and the last.

	RR Map computation	**Meta-ROI algorithm**	**CNN inference**	**Total (v1 / v2)**
OnePlus 8t	*27 ms*	*30 ms*	*13 ms*	***57 ms / 40 ms***
Oneplus 7T	*30 ms*	*32 ms*	*17 ms*	***62 ms / 47 ms***
Huawei P30	*57 ms*	*86 ms*	*15 ms*	***143 ms / 72 ms***

the mid-to-high 80's (%). To raise SpO_2 gradually, the volunteer is asked to take a few consecutive short breaths, each followed by holding until discomfort, returning gradually to a normal breathing pattern.

8.3. Data collection procedure

To collect the evaluation data, we followed the procedure illustrated in Fig. 1[3]. The volunteer is asked to sit in a chair with their hand resting on the table. A pulse oximeter (CMS50E) is clipped on their index finger while the back camera of the smartphone is placed on the middle finger. The person is then asked to apply the SpO_2 variation procotocol, described in Section 8.2. The average duration of each SpO_2 measurement session is 3 min.

Ground truth. To acquire the ground truth data, we use the off-the-shelf CMS50E pulse oximeter[40], which allows measurement of SpO_2 in the range of 35%-99% with a resolution of 1% [31]. It is an FDA approved device, widely used in literature for heart rate or SpO_2 monitoring [24].

As the oximeter is placed on the index finger while the smartphone on the middle finger, a question arising is whether it is valid to collect the ground truth on a different finger than the one SmartPhOx is using. To address it, we perform experiments with two oximeters placing one in each of the index and the middle finger. We then apply a T-test on the collected data to evaluate the null hypothesis that the pairwise difference between recordings of both fingers has a mean equal to zero at the 5% significance level. The test failed to reject the null hypothesis (p-value = 0.6669 > 0.05), providing support for using readings from index finger as ground truth while SmartPhOx is collecting measurements on the middle finger.

8.4. Data set

The data set is summarized in Table 4. We evaluate SmartPhOx on 37 participants and using three different smartphones, OnePlus 8T, OnePlus 7T and Huawei P30 Lite. Both Oneplus phones use a Sony

[3]Our experiments are in agreement with the ethics defined in the Helsinki Declaration [14].

Table 4: Data set summary

Age	18 - 60 Average: 30.31; Std: 12.37
Gender	Male: 27, Female: 10
Fitzpatrick phototyping scale	I:3, II:20, III: 3, IV: 2, V: 2, VI: 7
Oxygen level	85% - 99% Average: 95.8%; Std: 3%

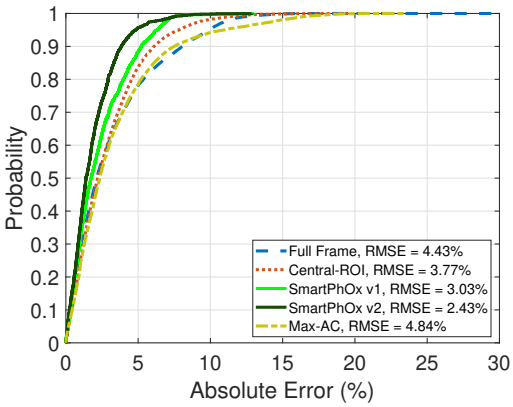


Figure 11: CDF of the absolute prediction error.

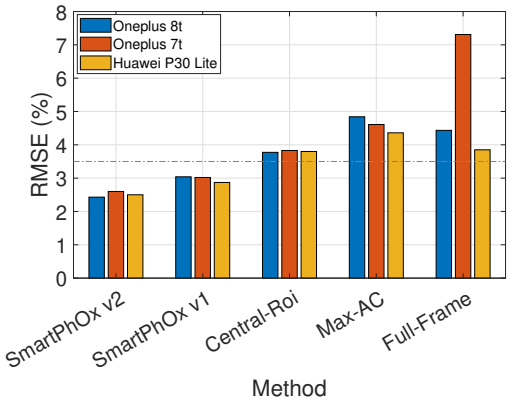


Figure 12: Root Mean Square Error (RMSE) by phone model.

IMX586 as main camera sensor, while the Huawei a Sony IMX600y. Their focal lengths are 26mm, 26mm and 28mm, and their apertures f/1.7mm, f/1.6mm and f/1.8mm, respectively. To address concerns regarding racial bias in SpO_2 measurements, especially as it regards Black patients [45], our study includes volunteers with different skin pigmentation, as classified by the Fitzpatrick phototyping scale [20].

9. Evaluation results

In this section, we perform a careful evaluation of both version of SmartPhOx, aimed at understanding their overall performance, the impact of key system parameters and experimental settings, and finally their utilization of system resources.

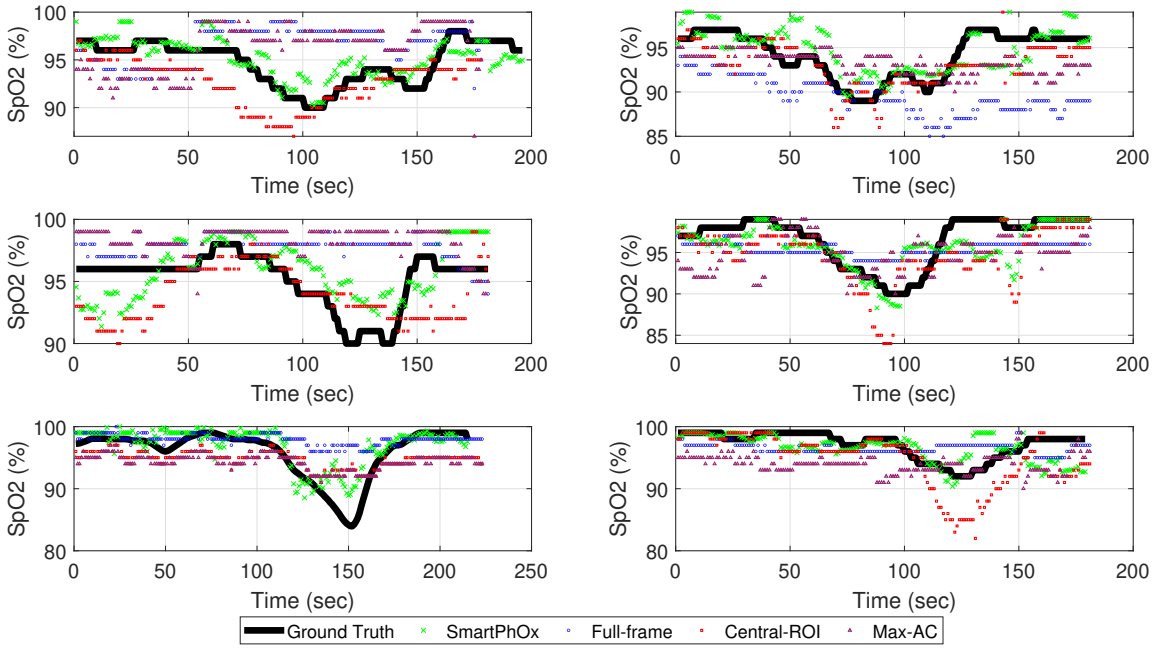


Figure 13: Raw SpO_2 results for 6 participants.

9.1. Overall SpO_2 prediction performance

Methodology: SmartPhOx is evaluated using leave-one-out cross validation, with data from 24, 12 and 1 users for train, validation and test sets, respectively. The ground truth is acquired as described in Section 8.3. We compare SmartPhOx v1's meta-ROI with the following approaches for selecting the RR values :

- **Full-frame**: Adopted by several works[39, 32, 10, 11], it uses the entire frame as the region of interest. The PPG signals are constructed by stacking in time the average value of every frame for the corresponding channel. The RR values are then computed from the resulting PPG signals.
- **Central-ROI**: It involves using the central 50x50 pixels of the frame [16, 43]. The intuition behind this approach is that the central part of the image should be least impacted by movement or ambient light, and therefore the most stable.
- **Max-AC**: It involves using the cell producing the largest value of the green channel *AC* [30]. The idea is that blood has a bigger impact on a PPG signal with a large pulse, making it, theoretically, of better quality.

For each approaches, we trained a linear regressor to estimate the SpO_2 from the RR extracted. Finally, we

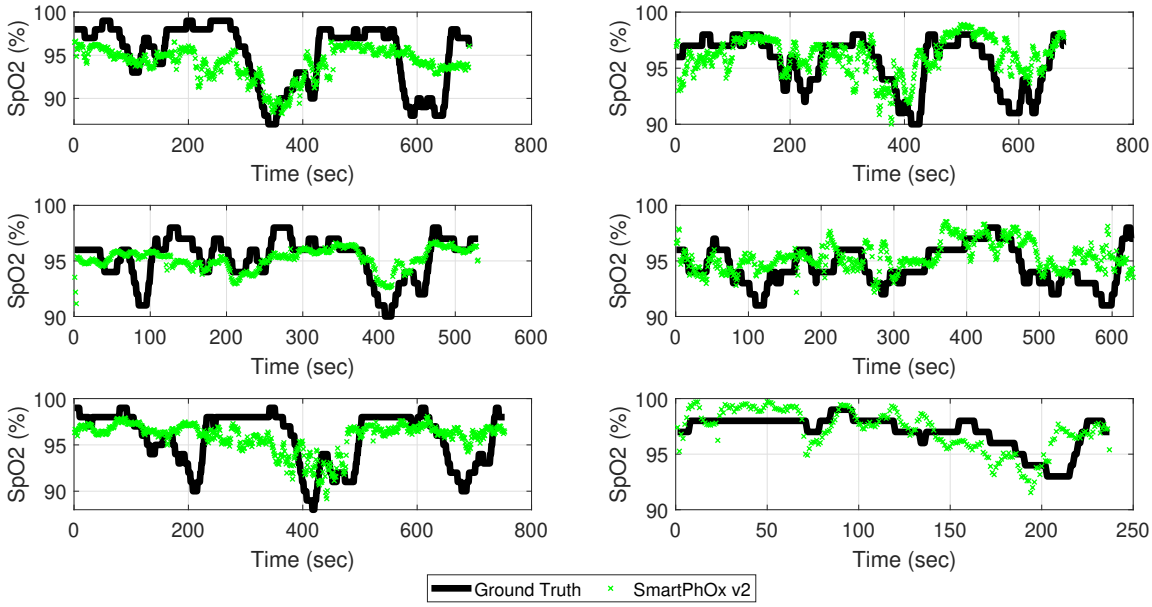


Figure 14: SmartPhOx v2 Raw SpO_2 results for 6 participants.

also add SmartPhOx v2 in the comparison.

Results: Figures 11 and 12 plots the CDF (Fig. 11) and the Root Mean Square Error (RMSE) (Fig. 12) of the SpO_2 prediction of all the considered baseline approaches. To put the results into context, Fig. 12 includes the FDA RMSE clearance threshold for pulse oximeters [3]. The data shows the two versions of SmartPhOx having the best performance. The median prediction error for SmartPhOx v1 and SmartPhOx v2 are respectively 1.75 % and 1.4 %, against 2.2% for the second-best method. The RMSE data paints a similar picture, with SmartPhOx v1 and SmartPhOx v2 delivering an RMSE of 3.04 % and 2.43 % versus 3.77, 4.84, 4.43% for Central-ROI, Max-AC and Full-Frame, respectively. Most important, both version of SmartPhOx are the only approach to meet the FDA RMSE requirement for pulse oximeters[4].

For a look into the raw data, Fig 13 shows the SpO_2 values reported by all methods during a testing session. In the interest of clarity, we show the data for 6 users. As the subjects are following the SpO_2 varying protocol, their levels drop from the healthy values of around 99% to under 90%. The data shows SmartPhOx v1 is capable of predicting the ground truth the best, which is in line with the analysis of Fig. 12.

Regarding SmartPhOx v2, in Fig. 14 we show raw SpO_2 prediction for six different participant. Interestingly, we can notice that sometimes SmartPhOx v2 is even capable of predicting a drop of SpO_2 before

[4]Obviously, this result does not imply FDA clearance, a process beyond the scope of this work.

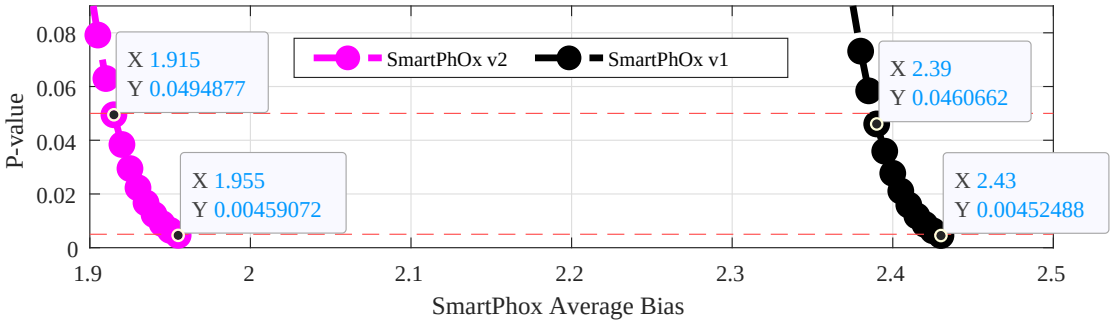


Figure 15: SmartPhOx statistical analysis.

Table 5: SmartPhOx v1 vs PhO_2. The data for PhO_2 as reported in [10] .

	PhO_2[10]	**SmartPhOx**
Number of Subjects	6	37
Skin pigmentation	Asian: 4, White: 2	Fitzpatrick I-VI
SpO2 range	81 % - 99 %	85 - 99 %
RMSE	N/A	3.04 %
Mean Absolute error, Std Absolute error	2.5 %, 1.62%	2.31%, 1.96%
Absolute error, 80^{th} percentile	3.5 %	3.83 %
Hardware add-on	Yes	No

the pulse oxymeter does.

9.2. Statistical analysis of the SmartPhOx performance

We conduct a one-tailed T-test on SmartPhOx's prediction errors observed in the experiments of Section 9.1. In particular, the statistical test is aimed at answering the question of whether SmartPhOx's prediction error is on average lower than a given value, x. Towards this, we perform a one-tailed T-test over the set of SmartPhOx's prediction errors for various values of x. Fig. 15 shows the p-value for different values of x. The data shows that the probability of SmartPhOx v1's average prediction error being above a given x drops below 0.05 for $x = 2.39$, and below 0.005 for $x = 2.43$. Regarding SmartPhOx v2, the threshold x corresponding to theses probabilities are respectively 1.91 and 1.95.

9.3. Comparison with a complete-system solution

In this section, we aim at contextualizing the performance of SmartPhOx v1 by comparing it with PhO_2 [10], a state-of-the-art system using the ratio-of-ratios (RR) method. We also compare SmartPhOx v2 with a state-of-the-art system using relying on Deep Learning.

Methodology: With a full-fledged, in-house implementation of PhO_2 being infeasible due to its using a custom-built hardware add-on, we compare with results reported in [10]. For SmartPhOx, we use the same data set and training/validation/testing protocol as in Section 9.1.

Results: Table 5 compares the performance evaluation of SmartPhOx with that of PhO_2. It shows that SmartPhOx's 80^{th} percentile of the absolute prediction error is very close to that reported for PhO_2 (no RMSE values are reported in [10]).

Implication: The results show that the meta-ROI approach of identifying regions with consistent RR values introduced by SmartPhOx can relax the requirement for custom-built hardware.

We also compared SmartPhOx v2 with the CNN proposed by [16]. We implemented their CNN and trained it using the same training strategy as of SmartPhOx v2. We want to evaluate if training a neural network to predict SpO_2 from the RR map – which is basically what SmartPhOx v2 does – is better than training the neural network using raw PPG input. Figure. 16 plots the RMSE and the median absolute error of both methods. The data shows that SmartPhOx v2 outperforms all the other methods. The RMSE and median error drop from 3.3% and 1.8%, respectively, for [16]'s CNN to 2.4% and 1.4 % when using SmartPhOx v2. In addition, it is interesting to note that SmartPhOx v2 provides a slightly better performance (3.03% and 1.75% vs 3.3% and 1.8%, respectively) when compared to applying CNN on the raw PPG data. This underscores the benefit of the virtual sensor array principle, as both SmartPhOx's algorithm take into account the difference between signals of different areas of the frame while the CNN was not capable of accurately predicting the relationship of the PPG signal from a fixed region of the frame with the SpO_2.

9.4. Sensitivity analysis

In this section, we evaluate SmartPhOx v1's performance as function of its key parameters. The RR Map being fundamental to its functionality, we focus on the RR Map cell size – defining the map's X, Y dimensions – and the RR Map Z-dimension.

9.4.1. Sensitivity to RR Map cell size

We vary the RR Map cell size from 32×18 px to 256×144 px. For our implementation using a 1260×720 px video (Table 2), this corresponds to an X, Y dimension ranging from 5×5 to 40×40.

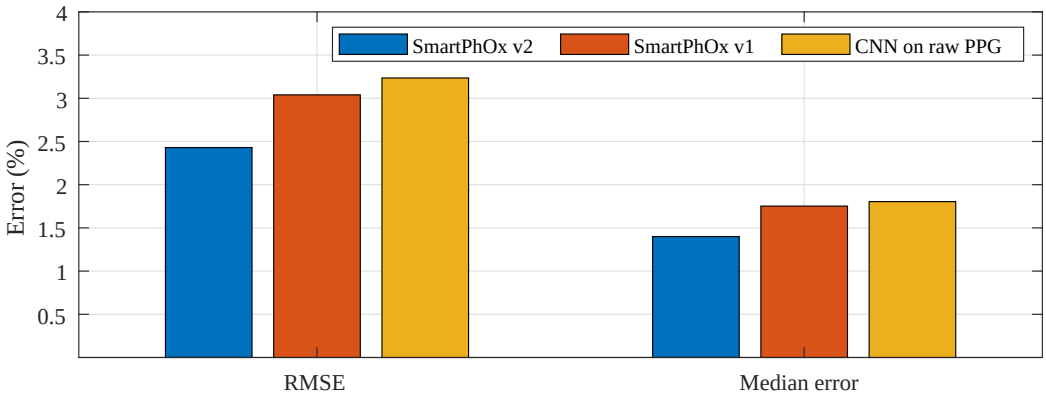


Figure 16: SmartPhOx v1 and SmartPhOx v2 vs the CNN proposed by [16]. Training the neural network on the RR map leads to better results when compared to training a CNN on the raw PPG signals.

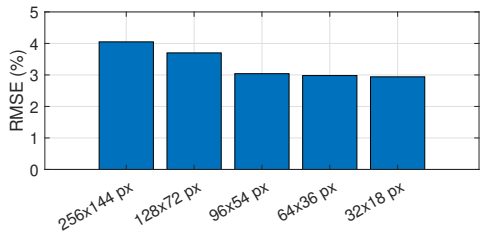


Figure 17: Impact of cell size on SmartPhOx v1 performance.

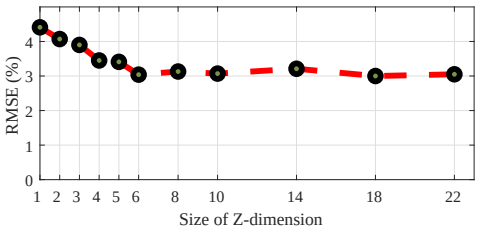
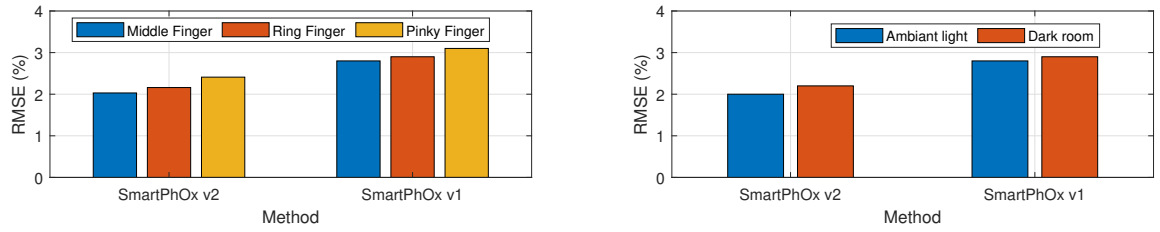


Figure 18: Impact of Z dimension on the RMSE of SmartPhOx v1.

Results: Fig. 17 reveals a binary behaviour. For large cell sizes (256×144 px, 128×72 px), the X, Y dimension of the RR Map (5×5, 10×10) is too coarse grained for SmartPhOx v1's meta-ROI algorithm to identify highly consistent meta-region of interests. However, once the cell size is 96×54 px or smaller the RMSE drops below the FDA clearance threshold. Thus, SmartPhOx v1 uses 96×54 px by default. We do not evaluate the performance of the CNN of SmartPhOx v2 as a function of cell size, since different cell size essentially means different input size and, thus, need a different neural network size. Hence, comparing the performance of different Neural Networks on different input would give no pertinent information.

9.4.2. Sensitivity to the size of Z-dimension

Fig. 18 depicts the performance of SmartPhOx v1 in terms of RMSE as function of the RR Map Z-dimension size. The data shows the importance of time consistency, embodied by the Z-dimension, in selecting the best meta-region of interest. When $Z = 1$, essentially eliminating the time dimension, the RMSE is well above the FDA clearance. As the Z-dimension increases, the performance of SmartPhOx v1



(a) SmartPhOx v1 Impact of finger

(b) SmartPhOx v1 Impact of ambient light

Figure 19: Varying experimental conditions.

improves significantly to meet the FDA requirement. Further, the data shows that once a time-consistent meta-region is identified, increasing the Z-dimension brings no additional gain. As a result, SmartPhOx v1 uses $Z = 6$ as the default value.

9.5. Varying experimental settings

In this section, we evaluate the impact of two key experimental parameters in the performance of SmartPhOx v1 and SmartPhOx v2: finger on which it measures SpO_2, and ambient lightning.

Methodology: With the help of seven of our volunteers, we run both versions of SmartPhOx with the smartphone placed successively on the middle, ring and little finger. In a second step, we successively both versions of SmartPhOx in a completely dark room, with the smartphone on the middle finger. We use two settings for the dark room conditions: in one – Dark room – we use the regressor trained with the main data set.

Results: Figure 19a shows a similar error for all fingers – RMSE of 2, 2.16, 2.41%, respectively for SmartPhOx v2 and 2.8, 2.9 3.1% for SmartPhOx v1– suggesting that both version of SmartPhOx are robust to the finger selection. In Fig. 19b We do observe a higher RMSE when using the pinky finger, which may be due to the fact that it is the smallest finger, making the light distribution over its surface more sensitive to random movements.

Fig. 19b shows that testing SmartPhOx v1 or SmartPhOx v2 in a completely dark room does not significantly alter its performance.

9.6. System resources utilization

We evaluate SmartPhOx v1' and SmartPhOx v2's utilization of CPU, memory and energy by using Android Profiler [15] and report the results in Table 6. The intrinsic multiprocessing nature of the mobile

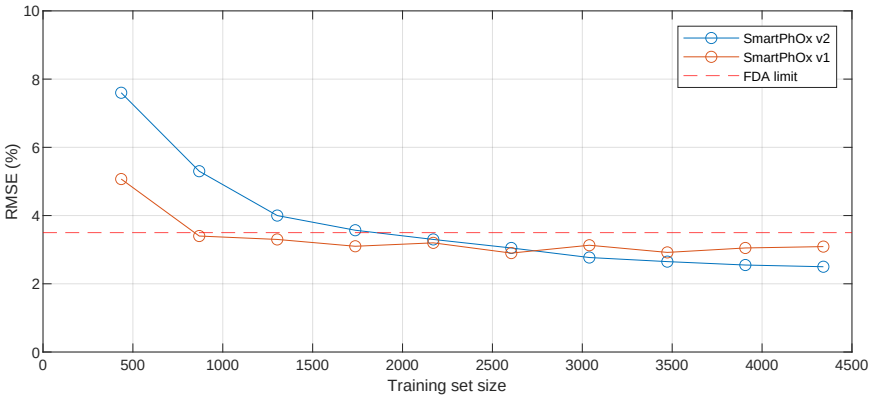


Figure 20: The performance of the two versions of SmartPhOx with respect to the training set size, for the same fixed test set. SmartPhOx v1 needs less data to achieve a Root Mean Squared Error below the FDA limit but, given enough training data, SmartPhOx v2 is more accurate.

phone operating systems makes it very challenging to measure the exact energy consumption of a given application. Thus, we show the percentage of time Android Profiler reports SmartPhOx's energy consumption as being Light (L), Medium (M) or Heavy (H). The data shows that all SmartPhOx's version utilization of resources are limited, especially in terms of energy consumption.

9.7. SmartPhOx v1 vs SmartPhOx v2: The trade off

Exploiting spatial redundancy by using CNN significantly improves the accuracy of SmartPhOx v2 compared to SmartPhOx v1 as can be seen in figure 12 and table 3. However, CNNs are more complex. For instance, training a simple linear regressor requires negligible time compared to training a CNN. In addition, with SmartPhOx v2 we need to collect much more training data to achieve reasonable accuracy compared to SmartPhOx v1. This trade-off between complexity and accuracy is illustrated in Figure 20. The data shows that SmartPhOx v1 needs half as much training data as SmartPhOx v2 to see its accuracy reach the FDA limit. It is also noticeable that the memory demand of SmartPhOx v2 is higher (table 6). This reduces its deployability, especially on memory-limited devices, such as older versions of smartphones, which are still widely used in developing countries.

10. Related work

Prior works on pulse oximetry can be grouped into two major categories: a) Works and systems using dedicated hardware dating back to the 1930s, b) More recent works built around the smartphone.

SmartPhOx v1

Device	CPU (%)	Memory (MB)	Energy (% of time) [L/M/H]
Oneplus 7t	24	176.9	90/10/0
Oneplus 8t	23	189.3	87/13/0
Huawei P30 Lite	26	298.7	95/5/0

SmartPhOx v2

Device	CPU (%)	Memory (MB)	Energy (% of time) [L/M/H]
Oneplus 7t	23	205	92/8/0
Oneplus 8t	23	198	90/10/0
Huawei P30 Lite	27	320	93/7/0

Table 6: SmartPhOx's resource utilization.

(a) Dedicated hardware: A pulse oximeter is a small portable device for noninvasive monitoring of a person's oxygen saturation in the blood. The idea dates back to 1935 [33], with the first pulse oximeter oximeter based on the ratio of red and infrared light absorption developed in the '70s [44]. Today, pulse oximetry remains an active area of research and development, leading to a plethora of devices that can be attached to the fingertip [40, 37, 12], earlobe [25, 38], forehead [4, 6, 19], trachea [7] and ring type [26] products. Despite the easy access to pulse oximeters, dedicated hardware can be impractical in everyday life, not least because, as the COVID-19 pandemic revealed, often people are not aware their oxygen level needs monitoring.

(b) Smartphone-based sensing: Recognizing smartphones as powerful sensing devices already in people's hands, researchers have proposed harnessing their capabilities for vital signs monitoring [8, 23, 27]. In the particular case of oxygen saturation, [43] was among the first to apply the ratio-of-ratios method for estimating SpO_2 using a smartphone. The RR values are computed over a 50x50px region of interest (ROI) at the centre of the frame. However, as our experiments showed, computing RR values off a particular physical frame area can lead to inaccurate SpO_2 values. To address this issue, [11, 32] integrate into the RR calculation the camera quantum efficiency, which represents the sensitivity of each channels (red, green, blue) of the image produced by the camera to the different wavelengths of the input light. While accurate, these solutions require knowledge of the camera quantum efficiency – something to which only manufacturers have access. PhO_2 [10] proposes to attach to the smartphone camera a custom-made device mounted with two chromatic filters, each allowing a precise wavelength to pass. The result is a system allowing SpO_2 predictions with very good accuracy. Nevertheless, the custom-built hardware add-on, while manufactured with the help of 3D printing, limits its large-scale application. Recently, dedicated oxygen monitoring sensors are being integrated in smartwatches [22, 42, 2], and some high-end smartphone models [18, 17, 1].

While very accurate, such solutions leave out a large section of users who have older smartphone models, particularly in developping countries. SmartPhOx, on the other hand, requires no custom hardware and can work on essentially any smartphone currently in people's hands.

11. Conclusion and discussion

We presented SmartPhOx, a smartphone-based pulse oximetry solution requiring no custom hardware. Using a carefully designed empirical study to inform our work, we identified the limitations of current approaches and introduced the notion of Meta-ROI. We transformed the Meta-ROI concept into a complete-system solution capable of running on a smartphone A carefully performance evaluation using an Android implementation of SmartPhOx and involving 37 healthy volunteers showed that it is the first smartphone-base pulse oximetry solution to meet the FDA requirement for Root Mean Square Error (RMSE) without needing custom hardware. Beyond that, we propose to consider the image acquired via the smartphone camera as the output of an array of independents sensors. This allows us to propose in a second version of the algorithm, SmartPhOx v2, a solution that exploits a deep learning architecture carefully designed to extract redundant information from the resulting array of RR, thus gaining in accuracy. Finally, we propose a comparison between the two proposed algorithms. From this, it appears that SmartPhOx v2 is recommended for recent equipment with a large memory, while SmartPhOx v1 is to be preferred for old equipment limited in this respect.

This work has its limitations. Additional cycles of engineering and testing will be necessary before it can fully meet the strict FDA requirements. In particular, SmartPhOx's evaluation needs to be extended to include non-healthy subjects. FDA requires testing in the SpO_2 range of 70% to 100%, while in healthy subjects on which we could evaluate SmartPhOx our protocol could not induce SpO_2 below the low 80s.

Acknowledgment

We thank Sandip Chakraborty (our shepherd) and the PerCom reviewers for their insightful and constructive comments. This work is supported in part by the Agence Nationale de la Recherche under the ANR JCJC CiTADEL grant. The experiments described in this work were conducted on the Grid'5000 testbed, which is powered by a group of academic institutions, including CNRS, RENATER, many universities and other groups and hosted by Inria (see https://www.grid5000.fr).

References

[1] 4, G.N., 2014. https://www.samsung.com/fr/smartphones/galaxy-note/galaxy-note-4-black-32gb-sm-n910fzkexef/. URL: https://www.samsung.com/fr/smartphones/galaxy-note/galaxy-note-4-black-32gb-sm-n910fzkexef/.

[2] 6, A.W.S., 2020. https://www.apple.com/fr/apple-watch-series-6/. URL: https://www.apple.com/fr/apple-watch-series-6/.

[3] ADMINISTRATION, U.F..D., 2013. Pulse oximeters - premarket notification submissions [510(k)s]: Guidance for industry and food and drug administration staff. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pulse-oximeters-premarket-notification-submissions-510ks-guidance-industry-and-food-and-drug.

[4] Agashe, G., Coakley, J., Mannheimer, P., 2007. Forehead pulse oximetry: Headband use helps alleviate false low readings likely related to venous pulsation artifact. Anesthesiology 105, 1111–6. doi:10.1097/00000542-200612000-00010.

[5] Bal, U., 2015. Non-contact estimation of heart rate and oxygen saturation using ambient light. Biomedical optics express 6, 86–97. doi:10.1364/BOE.6.000086.

[6] Berkenbosch, J., Tobias, J., 2006. Comparison of a new forehead reflectance pulse oximeter sensor with a conventional digit sensor in pediatric patients. Respiratory care 51, 726–31.

[7] Brimacombe, J., Keller, C., Margreiter, J., 2000. A pilot study of left tracheal pulse oximetry. Anesthesia and analgesia 91, 1003–6. doi:10.1097/00000539-200010000-00043.

[8] Bruining, N., Caiani, E., et al., 2014. Acquisition and analysis of cardiovascular signals on smartphones: potential, pitfalls and perspectives: By the task force of the e-cardiology working group of european society of cardiology. European Journal of Preventive Cardiology .

[9] Bui, N., Nguyen, A., Nguyen, P., Truong, A.H., Ashok, A., Dinh, T., Deterding, R., Vu, T., 2020. Smartphone-based spo2 measurement by exploiting wavelengths separation and chromophore compensation. ACM Transactions on Sensor Networks (TOSN) 16, 1 – 30.

[10] Bui, N., Nguyen, A., Nguyen, P., Truong, H., Ashok, A., Dinh, T., Deterding, R., Vu, T., 2017. Pho2: Smartphone based blood oxygen level measurement systems using near-ir and red wave-guided light, in: Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3131672.3131696, doi:10.1145/3131672.3131696.

[11] Carnì, D., Grimaldi, D., Sciammarella, P., Lamonaca, F., Spagnuolo, V., 2016. Setting-up of ppg scaling factors for spo2% evaluation by smartphone, pp. 1–5. doi:10.1109/MeMeA.2016.7533775.

[12] CO-Oximeter, P.P., . https://www.masimo.com/products/monitors/spot-check/pronto/.

[13] Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1, 224–227. doi:10.1109/TPAMI.1979.4766909.

[14] Declaration, H., 2013. Ethical principles for medical research involving human subjects.

[15] Developers, A., 2022. Measure app performance with android profiler. google. URL: Available:https://developer.android.com/studio/profile/cpu-profiler.

[16] Ding, X., Nassehi, D., Larson, E., 2018. Measuring oxygen saturation with smartphone cameras using convolutional neural networks. IEEE Journal of Biomedical and Health Informatics PP, 1–1. doi:10.1109/JBHI.2018.2887209.

[17] edge, S.G.S..S., 2015. https://www.samsung.com/fr/smartphones/galaxy-s6/s6/. URL: https://www.samsung.com/fr/smartphones/galaxy-s6/s6/.

[18] edge, S.G.S..S., 2016. https://www.samsung.com/fr/smartphones/galaxy-s7/overview/. URL: https://www.samsung.com/fr/smartphones/galaxy-s7/overview/.

[19] Fernandez, M., Burns, K., Calhoun, B., George, S., Martin, B., Weaver, C., 2007. Evaluation of a new pulse oximeter sensor. American journal of critical care : an official publication, American Association of Critical-Care Nurses 16, 146–52. doi:10.4037/ajcc2007.16.2.146.

[20] Fitzpatrick, T., 1988. The validity and practicality of sun-reactive skin types i through vi. Archives of dermatology 124 6, 869–71.

[21] van Gastel, M., Stuijk, S., Haan, G., 2016. New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring. Scientific Reports 6, 38609. doi:10.1038/srep38609.

[22] watch GT 2e, H., 2020. https://consumer.huawei.com/fr/wearables/watch-gt-2e/buy/. URL: https://consumer.huawei.com/fr/wearables/watch-gt-2e/buy/.

[23] Hashizume, T., Arizono, T., Yatani, K., 2018a. Auth 'n' scan: Opportunistic photoplethysmography in mobile fingerprint authentication. ACM IMWUT doi:10.1145/3161189.

[24] Hashizume, T., Arizono, T., Yatani, K., 2018b. Auth 'n' scan: Opportunistic photoplethysmography in mobile fingerprint authentication. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1. URL: https://doi.org/10.1145/3161189, doi:10.1145/3161189.

[25] Haynes, J., 2007. The ear as an alternative site for a pulse oximeter finger clip sensor. Respiratory care 52, 727–9.

[26] Huang, C.Y., Chan, M.C., Chen, C.Y., Lin, B.S., 2014. Novel wearable and wireless ring-type pulse oximeter with multi-detectors. Sensors (Basel, Switzerland) 14, 17586–17599. doi:10.3390/s140917586.

[27] Huynh, S., Balan, R.K., et al., 2019. Vitamon: Measuring heart rate variability using smartphone front camera, in: ACM SenSys'19. doi:10.1145/3356250.3360036.

[28] Kaoungku, N., Suksut, K., Chanklan, R., Kerdprasop, K., Kerdprasop, N., 2018. The silhouette width criterion for clustering and association mining to select image features. International Journal of Machine Learning and Computing 8, 69–73. doi:10.18178/ijmlc.2018.8.1.665.

[29] Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization URL: https://arxiv.org/abs/1412.6980, doi:10.48550/ARXIV.1412.6980.

[30] Kombozi, L., Park, H., 2017. A highly efficient and reliable heart rate monitoring system using smartphone cameras. Multimedia Tools and Applications 76. doi:10.1007/s11042-016-4010-1.

[31] Krzyminiewski, R., Dobosz, B., Ladzinska, M., Jemielity, M., Buczkowski, P., Urbanowicz, T., Clark, T., 2011. High signal resolution pulse wave–hope for a fast and cheap home care monitoring patients with cardiac diseases, in: Mede-Tel Conference Proceedings, pp. 252–56.

[32] Lamonaca, F., Carnì, D., Grimaldi, D., Nastro, A., Riccio, M., Spagnolo, V., 2015. Blood oxygen saturation measurement by smartphone camera. 2015 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2015 - Proceedings , 359–364doi:10.1109/MeMeA.2015.7145228.

[33] Matthes, K., 2005. Untersuchungen über die sauerstoffsättigung des menschlichen arterienblutes. Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie 179, 698–711.

[34] Max, J., 1960. Quantizing for minimum distortion. IRE Transactions on Information Theory 6, 7–12. doi:10.1109/TIT.1960.1057548.

[35] Mendelson, Y., 1992. Pulse oximetry: theory and applications for noninvasive monitoring. Clinical chemistry 38 9, 1601–7.

[36] Mohamed, R., Youssef, M., 2017. Heartsense: Ubiquitous accurate multi-modal fusion-based heart rate estimation using smartphones. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1. URL: https://doi.org/10.1145/3132028, doi:10.1145/3132028.

[37] Nellcor™ Portable SpO Patient Monitoring System, P., . https://www.medtronic.com/covidien/en-us/products/pulse-oximetry/nellcor-portable-spo2-patient-monitoring-system.html.

[38] Nonin, . Nonin 8000q2 ear clip sensor 3 foot/1 meter.

[39] Němcová, A., Jordanova, I., Varecka, M., Smisek, R., Maršánová, L., Smital, L., Vitek, M., 2020. Monitoring of heart rate, blood oxygen saturation, and blood pressure using a smartphone. Biomedical Signal Processing and Control 59, 101928. doi:10.1016/j.bspc.2020.101928.

[40] Oximeter, C.E.O.F.P., . https://www.pulseoximeter.org/cms50e.html.

[41] Porr, B., 2016. An iir filter library written in java. URL: https://www.berndporr.me.uk/iir/site/.

[42] Samsung, 2020. Samsung galaxy watch3. URL: https://www.samsung.com/fr/watches/galaxy-watch/galaxy-watch3-41mm-mystic-bronze-sm-r850nzdaeub/buy/.

[43] Scully, C., Lee, J., Meyer, J., Gorbach, A., Granquist-Fraser, D., Mendelson, Y., Chon, K., 2011. Physiological parameter monitoring from optical recordings with a mobile phone. IEEE transactions on bio-medical engineering 59, 303–6. doi:10.1109/TBME.2011.2163157.

[44] Severinghaus, J., Honda, Y., 2004. History of blood gas analysis. vii. pulse oximetry. Journal of Clinical Monitoring 3, 135–138.

[45] Sjoding, M.W., Dickson, R.P., Iwashyna, T.J., Gay, S.E., Valley, T.S., 2020. Racial bias in pulse oximetry measurement. New England Journal of Medicine 383, 2477–2478. URL: https://doi.org/10.1056/NEJMc2029240, doi:10.1056/NEJMc2029240, arXiv:https://doi.org/10.1056/NEJMc2029240. pMID: 33326721.

[46] Tensorflow, 2022. Tensorflow lite. URL: https://www.tensorflow.org/lite.

[47] Tobin, M.J., Laghi, F., Jubran, A., 2020. Why covid-19 silent hypoxemia is baffling to physicians. American journal of respiratory and critical care medicine 202, 356–360.

[48] Vagg, T., Plant, B.J., Tabirca, S., 2016. A general mhealth design pipeline, in: ACM MoMM'16, Singapore. p. 190–194.