SmartPhOx: Smartphone-Based Pulse Oximetry
Using a Meta-Region Of Interest

Firmin Kateu, Gentian Jakllari and Emmanuel Chaput
IRIT, Toulouse INP, CNRS, Université de Toulouse, France – Email: firstname.lastname@toulouse-inp.fr

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).

Index Terms—Pulse oximetry, SpO$_2$, Mobile health, Ratio-of-ratios, Beer-Lambert law

I. 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."

What if their cellphones could have measured their oxygen saturation, how many of these patients would have sought 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\textsuperscript{1} – the condition in which a patient still feels well but their SpO$_2$ is dangerously low [1] – have 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 monitoring of the blood oxygen levels, has emerged as an 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 monitoring of the blood oxygen levels, has emerged as an important tool in fighting COVID-19, drawing attention to the monitoring of the blood oxygen levels, has emerged as an important tool in fighting COVID-19...

As we put them on monitors.

\textsuperscript{1}https://www.nytimes.com/2020/04/20/opinion/sunday/coronavirus-testing-pneumonia.html

and the fact that at different points of the cardiac cycles only oxygen saturation-related factors change, as we show in detail in Section II, 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 [3]. The first pulse oximeter was developed in the ’70s [4] and today a wide range of pulse oximeters can be found off-the-shelf [5]–[9]. Nevertheless, dedicated hardware adds extra burden and, as the silent hypoxia cases due to COVID-19 have revealed [1], 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 [10]–[13]. Researchers have proposed smartphone-based pulse oximetry solutions predating the COVID-19 crisis. [14] 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 [15]. 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 [16]. To address this issue, [17], [18] integrate into the RR measurements the camera quantum efficiency. While improving accuracy, this is information to which only manufacturers have access. PhO$_2$ [16] 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 [19], and some high-end smartphone models [20]. While accurate, such solutions leave out large sections of users with older smartphone models, particularly in developing countries.

In this paper, we introduce SmartPhOx, a smartphone-based pulse oximetry system meeting the FDA clearance threshold [15] for accuracy while relying only on the standard smartphone camera and flashlight. To achieve this, we start

\textsuperscript{1}https://www.nytimes.com/2020/04/20/opinion/sunday/coronavirus-testing-pneumonia.html
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 [17], [18], [21] or identifying the right region on the video [14], [22], 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 an efficient algorithm that can identify the Meta-ROI.

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 III).

• We design SmartPhOx, a complete-system architecture leveraging the concept of Meta-ROI for smartphone-base pulse oximetry (Section IV).

• 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 V). Leveraging it, we develop an efficient algorithm for identifying Meta-ROI (Section VI).

• We implement SmartPhOx as a standalone Android application and evaluate it 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) [15] (Section VIII).

II. 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 [14], [16]–[18], [22], [23] and adopted by SmartPhOx.

A. Theoretical underpinning

The RR method for measuring SpO2 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 $\lambda$, $\rho$, 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 SpO2. Let us start by expressing it at the two extremes of the cardiac cycle: in diastole, where $d = d_{\text{min}}$, and systole, where $d = d_{\text{max}}$. Let $I_d(\lambda)$ and $I_s(\lambda)$ denote the corresponding $I(\lambda)$ values. Taking the logarithm of their ratio, we get:

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

While more practical, Eq. (1), requires measuring $d_{\text{min}}$ and $d_{\text{max}}$. To relax this requirement, we can use the ratio of two values corresponding to two different wavelengths, $\lambda_1$ and $\lambda_2$:

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

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

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

B. 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 [14], [24] have shown that it can be approximated using a linear model as follows:

$$SpO2 = A \times RR(\lambda_1, \lambda_2) + B$$

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 $B$.

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

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

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.
III. SMARTPHONE PULSE OXIMETRY: CHALLENGES AND OPPORTUNITIES

Unlike dedicated pulse oximeters, smartphones use a flash-light covering a wide spectrum (400-800nm [16]) 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 II-B on a video sequence. It involves measuring RR$^2$ values using Equation 5, followed by linear regression for estimating the $A$ and $B$ 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.

A. Baseline approach for measuring RR

The baseline approach for measuring RR, the basis of most works on this topic [14], [16], [17], [22], 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 $\lambda_1$ the red channel and $\lambda_2$ the green (or blue) channel, the RR is finally obtained using Equation 5.

B. 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 [5] for establishing the ground truth (more details in Section VII-C). We train the linear regression model using the SpO$_2$ establishing the ground truth (more details in Section VII-C).

Results: Fig. 2 plots the RR values, predicted SpO$_2$ and the prediction errors for all three users. Fig. 2a shows that while

\[ RR(\lambda_1, \lambda_2) \]

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.

C. 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 [16] 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 [17], [18], [21] while some focus on a particular region of the frame [14], [22]. In the following, we investigate the approaches requiring no hardware add-on.

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:

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 $Q_{\text{kurt}}$ in our analysis, a metric quantifying the purity of a signal related to cardiac activity [25]. Specifically, $Q_{\text{kurt}}(s) = \frac{kurtosis(FFT(s))}{kurtosis(P_s)}$, 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 $Q_{\text{kurt}}$ 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.

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. III-D, we provide an intuition as to the reasons behind this finding.

Footnote: For simplicity we write RR instead of $RR(\lambda_1, \lambda_2)$. 
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 SpO₂ predictions [14], [22]. 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 SpO₂ 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 SpO₂ predictions using a fixed region of the frame in particular, and a physical region
We propose to forego using primary factors and to rely on 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. Turning this idea into a robust smartphone-based solution raises several scientific and system challenges, which we detail and address in the following sections.

IV. SmartPhOx System Overview

Fig. 5 shows a high-level depiction of SmartPhOx’s architecture. It comprises three modules:

1) Hardware: The SpO₂ 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 V introduces methods for customizing the video recording and transforming the data into the RR Map.
3) Algorithms: Section VI 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 SpO₂ using the ratio-of-ratio method described in Section II-B.

V. Data – RR Map Construction

A. 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 [14], [26]. 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.

B. 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 VIII. To compute the RR of a cell from the PPG signal, recall from Sec. II-B that \( RR \approx \frac{AC_{\lambda_1}/DC_{\lambda_1}}{AC_{\lambda_2}/DC_{\lambda_2}} \). Using the green and red PPG signals as \( \lambda_1 \) and \( \lambda_2 \) 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, \( (x, y), x \in \{1, 2, \ldots, X\}, y \in \{1, 2, \ldots, Y\} \), we obtain a vector

\[
RR(x,y) = \left[ r_{1}(x,y), r_{2}(x,y), \ldots, r_{Z}(x,y) \right]
\]  

(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.

VI. META-ROI ALGORITHM

The basic premise of our work, as developed in Section III, 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.

A. Space-time consistency in the RR Map

To formalize the notion of RR consistency, first introduced qualitatively in Section III-B, we draw on the empirical study of Section III 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.

B. 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 [27], 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.

![Algorithm 1: Meta-ROI algorithm](image)

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.

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, $\sigma(C_i)$:

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

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 [28, 29] 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 [28], one of the classic validity indices for analysing clustering. Unlike its main alternative, the silhouette [29] index, which is focused on the cluster density, largely addressed by the consistency metric, the DB index rewards the creation of distinct clusters. Mathematically,

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max\{d_{ij} - d_{ij}, 0\},$$  \hspace{1cm} (8)

with $k$ the number of clusters, $d_i(j)$ the average (Euclidean) distance of all cells in cluster $i$ 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.

VII. IMPLEMENTATION AND DATASET

A. Implementation

We implemented SmartPhOx as a standalone Android application. The signal processing component is implemented using the IIRJ library [30]. The $k$-mean clustering is implemented in Java. Table I shows the default parameter values used in the implementation (we evaluate the impact of these values on the performance of SmartPhOx in Section VIII-B). Table II shows the processing times on off-the-shelf smartphones of the SmartPhOx implementation when using the default parameter values of Table I.

B. $SpO_2$ 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 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.

C. Data collection procedure

To collect the evaluation data, we followed the procedure illustrated in Fig. 1. 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 protocol, described in Section VII-B. The average duration of each $SpO_2$ measurement session is 3 min.

a) Ground truth: To acquire the ground truth data, we use the off-the-shelf CMS50E pulse oximeter [5], which allows measurement of $SpO_2$ in the range of 35%-99% with a resolution of 1% [32]. It is an FDA approved device, widely used in literature for heart rate or $SpO_2$ monitoring [33].

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.

D. Data set

The data set is summarized in Table III. 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 IMX586 as main camera sensor, while the Huawei a Sony IMX600y. Their focal lengths are 26mm,

<table>
<thead>
<tr>
<th></th>
<th>RR Map computation</th>
<th>Meta-ROI algorithm</th>
<th>Linear Reg.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>OnePlus 8T</td>
<td>27 ms</td>
<td>30 ms</td>
<td>32 ms</td>
<td>62 ms</td>
</tr>
<tr>
<td>OnePlus 7T</td>
<td>30 ms</td>
<td>32 ms</td>
<td>30 ms</td>
<td>62 ms</td>
</tr>
<tr>
<td>Huawei P30</td>
<td>37 ms</td>
<td>36 ms</td>
<td>0.5 ms</td>
<td>745.5 ms</td>
</tr>
</tbody>
</table>

TABLE II: SmartPhOx processing time on various phones.
TABLE III: Data set summary

<table>
<thead>
<tr>
<th>Age</th>
<th>18 - 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>30.31; Std: 12.37</td>
</tr>
<tr>
<td>Gender</td>
<td>Male: 27, Female: 10</td>
</tr>
<tr>
<td>Fitzpatrick phototyping scale</td>
<td>I:3, II:20, III: 3, IV: 2, V: 2, VI: 7</td>
</tr>
<tr>
<td>Oxygen level</td>
<td>85% - 99%; Average: 95.8%; Std: 3%</td>
</tr>
</tbody>
</table>

VIII. EVALUATION RESULTS

In this section, we perform a careful evaluation of SpO2, aimed at understanding its overall performance, the impact of key system parameters and experimental settings, and finally its utilization of system resources.

A. Overall SpO2 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 VII-C. We compare SmartPhOx’s meta-ROI with the following approaches for selecting the RR values:

- **Full-frame**: Adopted by several works [16]–[18], [21], 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 [14], [22]. 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 [36]. The idea is that blood has a bigger impact on a PPG signal with a large pulse, making it, theoretically, of better quality.

Results: Fig. 6 plots the CDF (Fig. 6a) and the Root Mean Square Error (RMSE) (Fig. 6b) of the SpO2 prediction of all the considered approaches. To put the results into context, Fig. 6b includes the FDA RMSE clearance threshold for pulse oximeters [15]. The data shows SmartPhOx having the best performance. The median prediction error for SmartPhOx is 1.75% against 2.2% for the second-best method. The RMSE data paints a similar picture, with SmartPhOx delivering an RMSE of 3.04 % versus 3.77, 4.84, 4.43% for Central-ROI, Max-AC and Full-Frame, respectively. Most important, SmartPhOx is the only approach to meet the FDA RMSE requirement for pulse oximeters4.

For a look into the raw data, Fig 7 shows the SpO2 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 SpO2 varying protocol, their levels drop from the healthy values of around 99% to under 90%. The data shows SmartPhOx is capable of predicting the ground truth the best, which is in line with the analysis of Fig. 6.

B. Statistical analysis of the SmartPhOx performance

We conduct a one-tailed T-test on SmartPhOx’s prediction errors observed in the experiments of Section VIII-A. 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. 8 shows the \(p\)-value for different values of \(x\).

4Obviously, this result does not imply FDA clearance, a process beyond the scope of this work.
The data shows that the probability of SmartPhOx’s average prediction error being above a given $x$ drops below 0.05 for $x = 2.39$, and below 0.005 for $x = 2.43$.

C. Comparison with a complete-system solution

In this section, we aim at contextualizing the performance of SmartPhOx by comparing it with PhO$_2$ [16], a state-of-the-art system using the ratio-of-ratios (RR) method.

**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 [16]. For SmartPhOx, we use the same data set and training/validation/testing protocol as in Section VIII-A.

**Results:** Table IV compares the performance evaluation of SmartPhOx with that of PhO$_2$. It shows that SmartPhOx’s 80th percentile of the absolute prediction error is very close to that reported for PhO$_2$ (no RMSE values are reported in [16]).

**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.

D. Sensitivity analysis

In this section, we evaluate SmartPhOx’s performance as a 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.

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

**Results:** Fig. 9 reveals a binary behaviour. For large cell sizes ($256 \times 144$ px, $128 \times 72$ px), the $X, Y$ dimension of the RR Map ($5 \times 5, 10 \times 10$) is too coarse grained for SmartPhOx’s meta-ROI algorithm to identify highly consistent meta-region of interests. However, once the cell size is $96 \times 54$ px or smaller the RMSE drops below the FDA clearance threshold. Thus, SmartPhOx uses $96 \times 54$ px by default.

2) Sensitivity to the size of Z-dimension: Fig. 10 depicts the performance of SmartPhOx in terms of RMSE as function of the RR Map Z-dimension size. The data shows that 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 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 uses $Z = 6$ as the default value.

E. Varying experimental settings

In this section, we evaluate the impact of two key experimental parameters in the performance of SmartPhOx: finger on which it measures SpO$_2$, and ambient lightning.

**Methodology:** With the help of seven of our volunteers, we run SmartPhOx with the smartphone placed successively on the middle, ring and little finger. In a second step, we run 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 fitted on the middle finger. We run SmartPhOx with the smartphone placed successively on the middle, ring and little finger. In a second step, we run SmartPhOx in a completely dark room, with the smartphone placed successively on the middle, ring and little finger. In a second step, we run SmartPhOx with the smartphone placed successively on the middle, ring and little finger.

**Results:** Figure 11a shows a similar error distribution for all fingers – median error of 1.8, 2.15 and 2.1%, respectively – suggesting that SmartPhOx is robust to the finger selection. 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. 11b shows that testing SmartPhOx in a completely dark room does not significantly alter its performance. The data shows that training the regressor with dark-room data (Dark room*) improves slightly the performance of SmartPhOx when
compared to normal lighting conditions. This may be due to
there being less RR variability in the dark.

F. System resources utilization

We evaluate SmartPhOx’s utilization of CPU, memory and energy by using Android Profiler [37] and report the results in Table V. The intrinsic multiprocessing nature of the mobile 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 SmartPhOx’s utilization of resources is limited, especially in terms of energy consumption.

TABLE V: SmartPhOx’s resource utilization.

<table>
<thead>
<tr>
<th>Device</th>
<th>CPU (%)</th>
<th>Memory (MB)</th>
<th>Energy (% of time [L/M/H])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oneplus 7t</td>
<td>23</td>
<td>176.9</td>
<td>90/100</td>
</tr>
<tr>
<td>Oneplus 8t</td>
<td>23</td>
<td>189.3</td>
<td>87/13/00</td>
</tr>
<tr>
<td>Huawei P30 Lite</td>
<td>26</td>
<td>298.7</td>
<td>95/5/00</td>
</tr>
</tbody>
</table>

IX. 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 smartphones.

(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 [2], with the first pulse oximeter oximeter based on the ratio of red and infrared light absorption developed in the ’70s [4]. Today, pulse oximetry remains an active area of research and development, leading to a plethora of devices that can be attached to the fingertip [5]–[7], earlobe [8], [38], forehead [9], [39], [40], trachea [41] and ring type [42] 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 [11]–[13]. In the particular case of oxygen saturation, [14] was among the first to apply the ratio-of-ratios method for estimating SpO2 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 SpO2 values. To address this issue, [17], [18] 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. PhO2 [16] 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 SpO2 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 [19], [43], [44], and some high-end smartphone models [20], [45], [46]. While very accurate, such solutions leave out a large section of users who have older smartphone models, particularly in developing countries. SmartPhOx, on the other hand, requires no custom hardware and can work on essentially any smartphone currently in people’s hands.

X. 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.

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 SpO2 range of 70% to 100%, while in healthy subjects on which we could evaluate SmartPhOx our protocol could not induce SpO2 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 CIIADEL grant.

REFERENCES

