Chapter 4

A Multi-Channel Ratio-of-Ratios Method for Noncontact Hand Video Based SpO_2 Monitoring Using Smartphone Cameras

4.1 Related Work

4.1.1 Contact-based SpO_2 measurement using smart devices

It is of significance to realize early detection of changes in SpO_2 to facilitate timely management of asymptomatic patients with clinical deterioration. Conventional SpO_2 measurement methods rely on contact-based sensing, such as pulse oximetry introduced in Chapter 1.1.4 designed by the RoR principle.

With the ubiquity of smartphones and the growing market of smart fitness devices, the RoR principle has been applied to new nonclinical settings for SpO_2 measurement. Apple Watch Series 6 has blood oxygen measurement functionality, and it requires skin contact with the watch neither too tight nor too loose for the best results [67]. The recent scientific literature also explored methods for SpO_2 estimation using a smartphone. These methods require a user to use his/her fingertip to cover an optical sensor and a nearby light source to capture the reemitted light from the illuminated tissue [17, 40, 96, 123,

140]. In this setup, an adapted ratio-of-ratios model is utilized with the red and blue (or green) channels of color videos in lieu of the traditional narrowband red and infrared wavelengths.

The aforementioned SpO_2 estimation methods based on smartphones and smartwatches are contact-based. It can present the risk of cross-contamination between individuals using the same measurement device. An additional issue with contact-based methods is that they may irritate sensitive skin or a sense of burning from the heat built up if a fingertip is in contact with the flashlight for an extended period of time. Also, pulse oximeters may not be widely available in marginalized communities and some undeveloped countries [64].

4.1.2 Noncontact SpO_2 measurement using cameras

Researchers have recently investigated measuring the saturation of blood oxygen by means of contactless techniques [57, 80, 149, 150]. These methods typically acquire a user's face video under ambient light with CCD cameras to estimate SpO_2 from pulsatile information of monochromatic wavelengths. Shao et al. [129] also use a facial video-based method to monitor SpO_2 that is implemented using a CMOS camera with a light source alternating at two different wavelengths. Tsai et al. [147] acquire hand images with CCD cameras under two monochromatic lights to analyze SpO_2 from the reflective intensity of the shallow skin tissue. These contactless methods can provide alternatives to contact-based SpO_2 measurements for individuals with finger injuries or nail polish [34, 165], for whom the traditional pulse oximeters may be inaccurate. However, the setups

used in the abovementioned studies use either high-end monochromatic cameras with selected optical filters or controlled monochromatic light sources, making it expensive and not common for daily use.

As more economical camera devices, smartphones and webcams are also applied for contactless SpO_2 estimation. Most of the SpO_2 estimation works using digital RGB cameras under ambient light [12, 22, 119, 139] adapt the conventional RoR model based on the red and infrared wavelengths directly to the use of red and blue channels of RGB videos. It is worth noting that the SpO_2 data collected in [22, 119] only covers a small dynamic range (mostly above 95%), and Tarassenko *et al.* [139] and Bal *et al.* [12] show a fitted linear relation between RoR and SpO_2 for only several minutes of data. The limitations as mentioned above can be due to: i) Signals extracted from the red and blue channels are noisier than those extracted from the green channel [152], and ii) Unlike the narrowband signals being modeled in the conventional RoR model, the RGB color channels capture a wide range of wavelengths from the ambient light. The aggregation of the broad range of wavelengths lowers the optical difference between Hb and HbO_2 and makes it less optically selective than narrowband oximeter sensors and more challenging for SpO_2 sensing. So we are motivated to disentangle the aggregation effect through a meaningful combination of the pulsatile signals from all three channels of RGB videos to distill the SpO_2 information.

4.2 Ratio-of-ratios (RoR) Model for Noncontact SpO_2 Measurement

Consider a light source with the spectral distribution $I(\lambda)$ illuminating the skin and a remote color camera with spectral responsivity $r(\lambda)$ recording an image. According to the skin-reflection model [156], the color camera will receive the specularly reflected light from the skin surface and the diffusely reflected light from the tissue-light interaction that contains the pulsatile information. Based on the assumption proposed in [57] that the specular reflection can be ignored if the color change from movement is properly treated and minimized, the camera sensor response at time t can be expressed as:

$$\mathcal{S}_c(t) = \int_{\Lambda_c} I(\lambda) \cdot e^{-\mu_d(\lambda,t)} \cdot r_c(\lambda)\, d\lambda. \tag{4.1}$$

where the λ is the wavelength. The integral range Λ_c is the sensitive response wavelength band of the c th channel of the camera, $I(\lambda)$ is the spectral intensity of the light source, $\mu_d(\lambda, t)$ is the diffusion coefficient, and $r_c(\lambda)$ is the sensor response of the c th channel of the camera.

According to Beer-Lambert's law, the diffusion coefficient $\mu_d(\lambda, t)$ can be expanded into:

$$\mu_d(\lambda, t) = \varepsilon_t(\lambda) C_t l_t + [\varepsilon_{\text{Hb}}(\lambda) C_{\text{Hb}} + \varepsilon_{\text{HbO}_2}(\lambda) C_{\text{HbO}_2}] \cdot l(t). \tag{4.2}$$

where ε_{Hb}, $\varepsilon_{\text{HbO}_2}$, and ε_t are the extinction coefficients of arterial deoxyhemoglobin, arterial oxyhemoglobin, and other tissues including the venous blood vessel, respectively. C_t, C_{Hb}, and C_{HbO_2} are the concentration of the corresponding substances. l_t is the path length that the light travels in the tissue, which is assumed to be static and invariant of

time. $l(t)$ is the path length that the light travels in the arterial blood vessels. It is modeled as time-varying because the arteries will dilate with increased blood during systole compared to diastole.

When the camera is monochromatic, incoming light is filtered by a narrowband optical filter, or the light source is a narrowband LED, the integral range Λ_c can be simplified to a single value λ_i, such that the response of the camera sensor in Eq. (4.1) can be written as:

$$\mathcal{S}_c(t) = I(\lambda_i) \cdot e^{-\varepsilon_t(\lambda_i) C_t l_t} \cdot r_c(\lambda_i) \cdot e^{-[\varepsilon_{\text{Hb}}(\lambda_i) C_{\text{Hb}} + \varepsilon_{\text{HbO}_2}(\lambda_i) C_{\text{HbO}_2}] \cdot l(t)}. \tag{4.3}$$

Let $\Delta l = l_{\text{max}} - l_{\text{min}}$ denote the difference of the light path of the pulsatile arterial blood between diastole when $l(t) = l_{\text{min}}$ and systole when $l(t) = l_{\text{max}}$. Then the ratio of the response of the c th channel of the camera sensor during diastole and systole is:

$$\text{R}(\lambda_i) = \log \left(\frac{\mathcal{S}_c|_{l=l_{\text{min}}}}{\mathcal{S}_c|_{l=l_{\text{max}}}} \right) \tag{4.4a}$$

$$= [\varepsilon_{\text{Hb}}(\lambda_i) C_{\text{Hb}} + \varepsilon_{\text{HbO}_2}(\lambda_i) C_{\text{HbO}_2}] \cdot \Delta l. \tag{4.4b}$$

The ratio-of-ratios (RoR) between two different wavelengths λ_1 and λ_2 is:

$$\text{RoR}(\lambda_1, \lambda_2) = \frac{R(\lambda_1)}{R(\lambda_2)} = \frac{\varepsilon_{\text{Hb}}(\lambda_1) C_{\text{Hb}} + \varepsilon_{\text{HbO}_2}(\lambda_1) C_{\text{HbO}_2}}{\varepsilon_{\text{Hb}}(\lambda_2) C_{\text{Hb}} + \varepsilon_{\text{HbO}_2}(\lambda_2) C_{\text{HbO}_2}}. \tag{4.5}$$

Since $\text{SpO}_2(\%) = \frac{C_{\text{HbO}_2}}{C_{\text{HbO}_2} + C_{\text{Hb}}}$, the relation between RoR and SpO_2 can be derived from

Eq. (4.5) as:

$$\mathrm{SpO_2} = \frac{\varepsilon_{\mathrm{Hb}}(\lambda_1) - \varepsilon_{\mathrm{Hb}}(\lambda_2) \cdot \mathrm{RoR}}{\varepsilon_{\mathrm{Hb}}(\lambda_1) - \varepsilon_{\mathrm{HbO_2}}(\lambda_1) + [\varepsilon_{\mathrm{HbO_2}}(\lambda_2) - \varepsilon_{\mathrm{Hb}}(\lambda_2)] \cdot \mathrm{RoR}} \tag{4.6a}$$

$$\approx \alpha \cdot \mathrm{RoR} + \beta. \tag{4.6b}$$

where the linear approximation can be obtained by Taylor expansion.

The linear RoR model in Eq. (4.6b) has been applied under different $\mathrm{SpO_2}$ measurement scenarios. For pulse oximeters, $\lambda_1 = 660$ nm and $\lambda_2 = 940$ nm are used to leverage the optical absorption difference of Hb and $\mathrm{HbO_2}$ at the two wavelengths. In some prior art using narrowband light sources or monochromatic camera sensors [80, 129] for contactless $\mathrm{SpO_2}$ monitoring, different combinations of (λ_1, λ_2) have been explored. In the prior art using consumer-grade RGB cameras [12, 22, 119, 136, 139], only two out of the three available RGB channels were used for the linear RoR model.

Among the abovementioned $\mathrm{SpO_2}$ estimation methods using consumer-grade RGB cameras, the $\mathrm{SpO_2}$ data collected in [22, 119] only cover a small dynamic range (mostly above 95%), which is not very meaningful. Bal et al. [12] and Tarassenko et al. [139] show a fitted linear relation between RoR and $\mathrm{SpO_2}$ for data that last for merely several minutes. These limitations can be attributed to that, unlike the signals captured in the narrowband setting that is modeled precisely by Eq. (4.3) and Eq. (4.4), all three RGB color channels capture a wide range of wavelengths from the ambient light, as is described in Eq. (4.1). The aggregation of the broad range of wavelengths lowers the optical difference between Hb and $\mathrm{HbO_2}$ and makes it less optically selective than narrowband sensors used in oximeters. To address this issue, we disentangle the aggregation through a careful

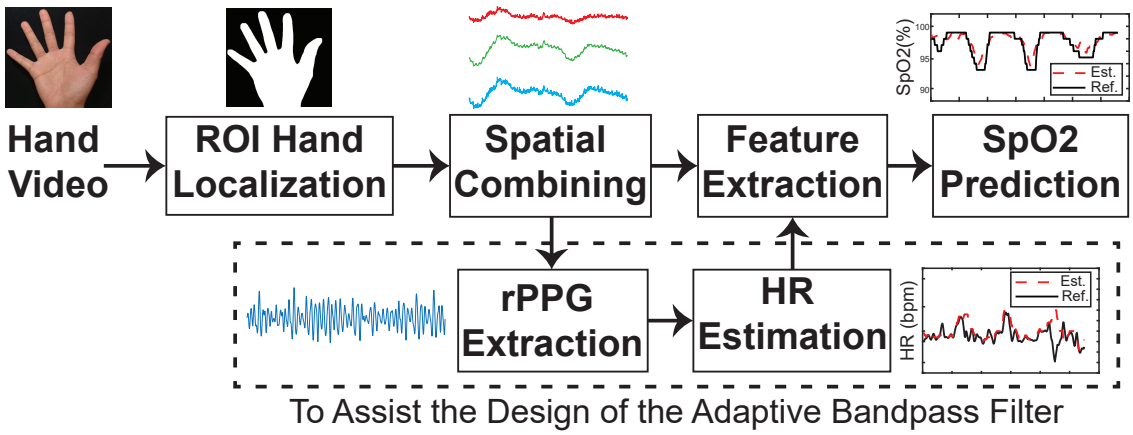


Figure 4.1: System illustration for the SpO_2 prediction using the smartphone captured hand videos. The pixels from the hand region are utilized for prediction, and an rPPG signal is extracted for heart rate (HR) estimation. Multi-channel RoR features are derived from the spatially combined RGB signals with the help of the HR-guided filters. The extracted features are then used for SpO_2 prediction.

combination of the pulsatile signals from all three channels of RGB videos to efficiently distill the SpO_2 information.

4.3 Proposed Multi-Channel RoR Method

In this work, we propose a multi-channel RoR method for noncontact SpO_2 monitoring using hand videos captured by smartphone cameras under ambient light. Fig. 4.1 illustrates the proposed procedure for the SpO_2 estimation from the smartphone captured hand videos. First, the hand is detected as the region of interest (ROI) for each frame. Second, the spatial average from the ROI is calculated to obtain three time-varying signals of RGB channels. The averaged RGB signals are extracted for two purposes: i) to estimate the heart rate (HR), and ii) to acquire the filtered cardio-related AC components using an HR-based adaptive bandpass filter. Third, the ratio between the AC and the DC components for each color channel and the pairwise ratios of the resulting three ratios are

computed as the features for a regression model where SpO_2 is treated as the label. The details of each step are provided as follows.

4.3.1 ROI Localization and Spatial Combining

First, we manually draw a rectangle to include the target hand region. This RGB region is converted to YCrCb color space, and the Cr channel is used [23] to determine a threshold that differentiates the skin pixels from the background based on the Otsu algorithm [109]. We apply an erosion and a dilation algorithm with a median filter to exclude noise pixels outside of the binary hand mask region. The final hand-shaped mask is considered as the ROI, and an example is shown in the second picture in Fig. 4.1. For all n frames in the video, we calculate the spatial average of the RGB channels in the ROI as $\mathbf{A} = [\bar{\mathbf{r}}; \bar{\mathbf{g}}; \bar{\mathbf{b}}]$, where $\bar{\mathbf{r}}, \bar{\mathbf{g}}, \bar{\mathbf{b}} \in \mathbb{R}^{1\times n}$, and $\mathbf{A} \in \mathbb{R}^{3\times n}$.

4.3.2 rPPG Extraction and HR Estimation

Typically in the RoR method, after the matrix $\mathbf{A}$ in Section 4.3.1 is calculated, the AC component for each channel of $\mathbf{A}$ is determined by either the standard deviation [123] or the peak-to-valley amplitude [129]. Since the signal-to-noise ratio (SNR) is lower for the video captured by a smartphone in a contactless manner, we propose to use an adaptive bandpass filter centered at the HR to filter the RGB channel signals and extract the AC components more precisely.

The HR can be measured contact-free by capturing the pulse-induced subtle color variations of the skin. The pulse signal, referred to as remote photoplethysmogram

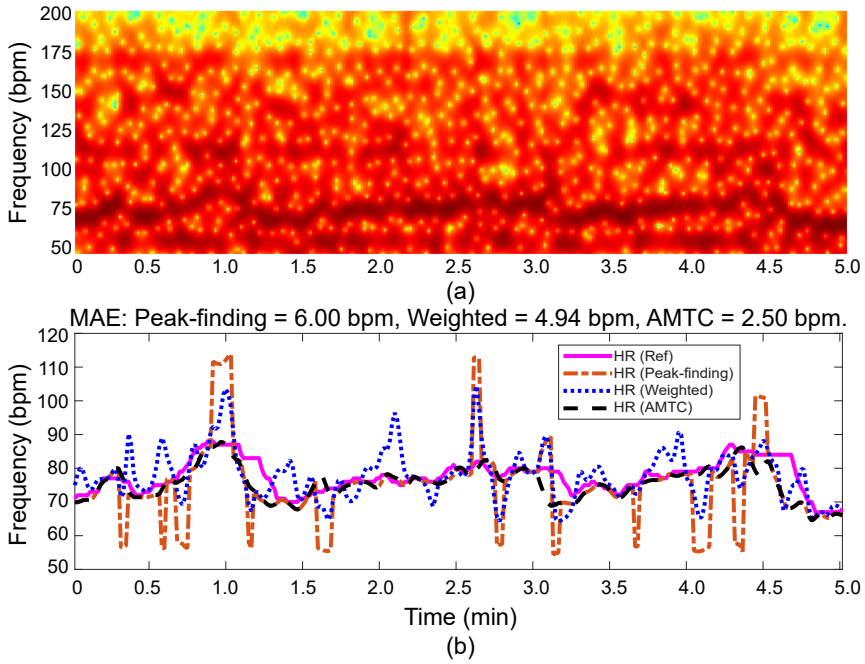


Figure 4.2: (a) Spectrogram of an rPPG signal. (b) Reference HR signals and HR signals estimated by the "naive" algorithm, the weighted energy frequency estimation algorithm, and the AMTC algorithm, respectively. The mean absolute error (MAE) of the HR estimation algorithms are 6.00 bpm, 4.94 bpm, and 2.50 bpm, respectively.

(rPPG), can be obtained by applying the plane-orthogonal-to-skin (POS) algorithm [156], which defines a plane orthogonal to the skin tone in the RGB space for robust rPPG extraction. The HR is then tracked from the rPPG signal via a state-of-the-art adaptive multi-trace carving (AMTC) [172,173] algorithm that tracks the HR from the spectrogram of rPPG by dynamic programming and adaptive trace compensation.

To study the role of accurate HR tracking for feature extraction, we also implemented a peak-finding method and a weighted energy method for frequency estimation [59] to compare with AMTC. The peak-finding method takes the peaks of the squared magnitude of the Fourier transform of rPPG as the estimated HR values, which was used in [149] and [57]. The weighted energy method finds the heart rate by weighing the frequency bins in the corresponding frame of the spectrogram of rPPG. Compared to the

peak-finding method, the weighted energy method is more robust to outliers in frequency. Fig. 4.2 illustrates an example of the HR estimation results by the peak-finding method, the weighted energy algorithm, and AMTC, respectively.

4.3.3 Feature Extraction

We use a processing window of 10 seconds with a step size of 1 second to segment the whole video into L windows. Within each window, the DC and AC components of the RGB channels are calculated to build a feature vector $\mathbf{f}$.

DC component We use a second-order lowpass Butterworth filter with a cutoff frequency of 0.1 Hz. The DC component is estimated using the median of the lowpass filtered signal of each window.

AC component The estimated heart rate values from Section 4.3.2 are used as the center frequencies for the adaptive bandpass (ABP) filters to extract the AC components of the RGB channels, which eliminates all frequency components that are unrelated to the cardiac pulse. We adopt an 8th-order Butterworth bandpass filter with ± 0.1 Hz (± 6 bpm) bandwidth, centering at the estimated HR of the current window. The magnitude of the AC component is estimated using the average peak-to-valley amplitudes of the filtered signals within the current processing window.

We define the normalized AC components at the ith window as $R(i,c) = \frac{\text{AC}(i,c)}{\text{DC}(i,c)}$, where $c \in \{r, g, b\}$ represents color channel and $i \in \{1, 2, ..., L\}$. We define the multi-channel ratio-of-ratios based feature vector of the ith window as

$\mathbf{f}_i = [R(i,r), R(i,g), R(i,b), \frac{R(i,r)}{R(i,g)}, \frac{R(i,r)}{R(i,b)}, \frac{R(i,g)}{R(i,b)}] \in \mathbb{R}^{1\times 6}$.

4.3.4 Regression and Postprocessing

As a proof-of-concept, we use linear regression and support-vector-regression (SVR) to learn the mapping between the features and the SpO_2 values.

The linear regression has limited learning capability since it captures only the linear relationship. So we use it as a baseline approach. In the objective function in Eq. (4.7), $\mathbf{y} = [y_1, ..., y_l] \in \mathbb{R}^{l\times 1}$ is the target SpO_2 value, $\omega \in \mathbb{R}^{6\times 1}$ is the predictor, and $\mathcal{F}$ is the feature matrix that serves as input. We add an L_2 regularization term in Eq. (4.7) to avoid collinearity. To select the optimal weight λ for the L_2 regularization term, we use 5-fold cross-validation.

$$\min_{\omega} \|\mathbf{y} - \mathcal{F}\omega\|_F^2 + \lambda \|\omega\|_2^2 . \tag{4.7}$$

SVR models are adopted for exploring possibly nonlinear relation between the feature vectors and the SpO_2 estimation. The Libsvm library [24] is used for training the "ϵ-SVR" in Eq. (4.8). In our implementation, we use the nonlinear Radial Basis Function (RBF) kernel for the SVR. The hyperparameters, including the penalty cost C, and the kernel parameter γ of kernel function $K(\mathbf{f}_i, \mathbf{f}_j) = \boldsymbol{\phi}(\mathbf{f}_i)^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{f}_j) = \exp\left(-\gamma \|\mathbf{f}_i - \mathbf{f}_j\|^2\right)$ are selected via grid search and a 5-fold cross-validation.

$$\begin{aligned} \min_{\omega,b,\xi,\xi^*} \frac{1}{2}\|\omega\|_2^2 + C\cdot\sum_{i=1}^{l}(\xi_i+\xi_i^*) \\ \text{s.t.} \quad \phi(f_i)\cdot\omega+b-y_i \leq \epsilon+\xi_i, \\ y_i-\phi(f_i)\cdot\omega-b \leq \epsilon+\xi_i^*, \\ \xi_i,\xi_i^* \geq 0, i=1,...,l. \end{aligned} \tag{4.8}$$

Once an estimated weight vector $\hat{\mathbf{w}}$ is learned from the linear or support vector regression, $\hat{\mathbf{w}}$ is then used to predict a preliminary SpO_2 signal. Finally, a 10-second moving average window is applied to smooth out the preliminarily predicted signal to obtain the final predicted SpO_2 signal.

4.4 Experimental Results

4.4.1 Data Collection

Fourteen volunteers, including eight females and six males, were enrolled in our study under protocol #1376735 approved by the University of Maryland Institutional Review Board (IRB), with age range between 21 and 30. Participants were asked to categorize their skin tone based on the Fitzpatrick skin types [10] shown in Fig. 4.3. There are two, eight, one, and three participants having skin types II, III, IV, and V, respectively. None of the participants had any known cardiovascular or respiratory diseases. During the data collection, participants were asked to hold their breath to induce a wide dynamic range of SpO_2 levels. The typical SpO_2 range for a healthy person is from 95% to 100%. By holding breath, the SpO_2 level can drop below 90%. Once the participant resumes

normal breathing, the SpO_2 will return to the level before the breath-holding.

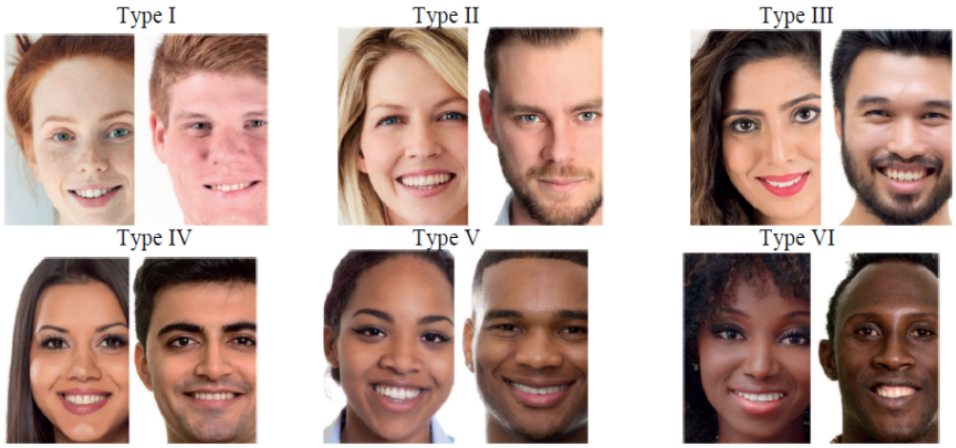


Figure 4.3: Fitzpatrick skin types [10].

Each participant was recorded for two sessions. During the recording, the participant sat comfortably in an upright position and put both hands on a clean dark foam sheet placed on a table. As shown in Fig. 4.4, the palm side of the right hand and the back side of the left hand were facing the camera. These two hand-video capturing positions are defined as *palm up (PU)* and *palm down (PD)*, respectively. The participant was asked to place his/her hands still on the table to avoid hand motion. Simultaneously, a Contec CMS-50E pulse oximeter was clipped to the left index finger to measure the participant's SpO_2 level at a sampling rate of 1Hz. As we have reviewed earlier, the oximeter is adopted clinically to be within a $\pm 2\%$ deviation from the invasive, gold standard for SpO_2 [114], so we use the oximeter measurement results as the reference in our experiments. An iPhone 7 Plus camera was fixed by a smartphone stand mounted on a tripod for video recording at a sampling rate of 30 fps. The video started 30 seconds before the oximeter started and stopped immediately after the oximeter ended to allow for proper time synchronization. The participants were asked to hold their breath for generally 30–40 seconds

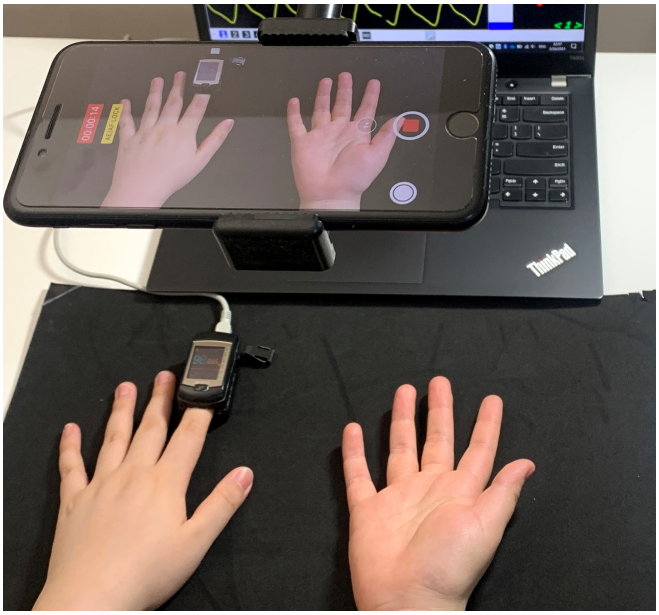


Figure 4.4: Experimental setup for data collection of hand videos and reference signals using an oximeter. The left index finger was placed in a CMS-50E pulse oximeter to record the reference HR and SpO_2 signals. The smartphone camera is recording the video of both hands.

to lower the SpO_2 level, as long as they were comfortable and able to do so. Then the participants resumed normal breathing for generally 30–40 seconds until they recovered and felt ready for the next breath-holding. The recovery period was long enough for the participants' SpO_2 to return to the levels before the breath-holding. The aforementioned process is defined as one breath-holding cycle. In each session, the breath-holding cycles were repeated three times. After the first session, the participants were asked to relax for at least 15 minutes before attending the second session for data collection. From our data collection protocol using breath-holding, we were able to obtain the SpO_2 measurements ranging from 89% to 99%.

The total length of recording time for all fourteen participants is 138.9 minutes. In terms of each participant, the minimum duration is 103 seconds and the maximum duration is 468 seconds. The average duration is 298 seconds. The current data size is

relatively small for large-scale neural network training. This is by a large part due to the restrictions for human subject related data collection imposed during the COVID-19. The available data, however, is adequate for our principled multi-channel signal based approach to SpO_2 monitoring, showing a benefit of combining signal processing and biomedical knowledge and modeling with data than the primarily data-driven approach.

Delay Estimation of Pulse Oximeter: When the CMS-50E oximeter is turned on and ready for measurement, the first reading is displayed a few seconds after the finger is inserted. This delay may be due to the oximeter's internal firmware startup and algorithmic processing. Since we need to synchronize the video and the oximeter readings using their precise starting time stamps, the delay in the oximeter can introduce misalignment errors in the reference data that we use to train the regression model. To avoid misalignment, we first estimate the delay and then subtract it from the oximeter's internal timestamp as the corrected oximeter's timestamp. To estimate the internal delay, we asked one participant to repeatedly place the left index finger, middle finger, and ring finger into the oximeter 50 times each and obtained the average delay time of 1.8s, 1.9s, and 1.7s, respectively. Because the left index finger is used for reference data collection in our setup, we take 1.8s as the delay. To further examine whether there exists any difference among the delays from the three fingers, we conducted a one-way ANOVA test. The p-value is 0.14, which shows no statistically significant different delays among the three fingers.

4.4.2 Performance Metrics

The performance of the algorithm is evaluated using the mean absolute error (MAE) and Pearson's correlation coefficient ρ given in (4.9). Note that the correlation is adopted to evaluate how well the trend of SpO_2 is tracked.

$$\mathrm{MAE}(\mathbf{y},\hat{\mathbf{y}}) = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \quad \rho(\mathbf{y},\hat{\mathbf{y}}) = \frac{(\mathbf{y}-\bar{y})^{\mathrm{T}}(\hat{\mathbf{y}}-\bar{\hat{y}})}{\|\mathbf{y}-\bar{y}\|_2 \left\|\hat{\mathbf{y}}-\bar{\hat{y}}\right\|_2}. \tag{4.9}$$

where $\mathbf{y} = [y_1, ..., y_N]^{\mathrm{T}}$, $\hat{\mathbf{y}} = [\hat{y}_1, ..., \hat{y}_N]^{\mathrm{T}}$, $\bar{y}$, and $\bar{\hat{y}}$ denote the reference SpO_2 signal, the estimated SpO_2 signal, the average values of all coordinates of vectors $\mathbf{y}$ and $\hat{\mathbf{y}}$, respectively. We adopt the correlation metric to evaluate how well the trend of the SpO_2 signal is tracked.

4.4.3 Results From Proposed Algorithm

In this subsection, we use the training data from one participant to train the regression model for the prediction of his/her testing session recorded later. We call the aforementioned training and testing procedure the *participant-specific* mode in which the models are specifically learned for each participant. We will discuss the *leave-one-out* mode of the performance of the proposed algorithm in Section 4.4.5.

Fig. 4.5 presents the learning results for all the participants using SVR for PU cases. Both training and testing sessions are shown for each participant. The SpO_2 curves in each session contain three dips that are resulted from breath-holding, except for participant #8 who had a shorter session due to limited tolerance of breath-holding. For each participant,

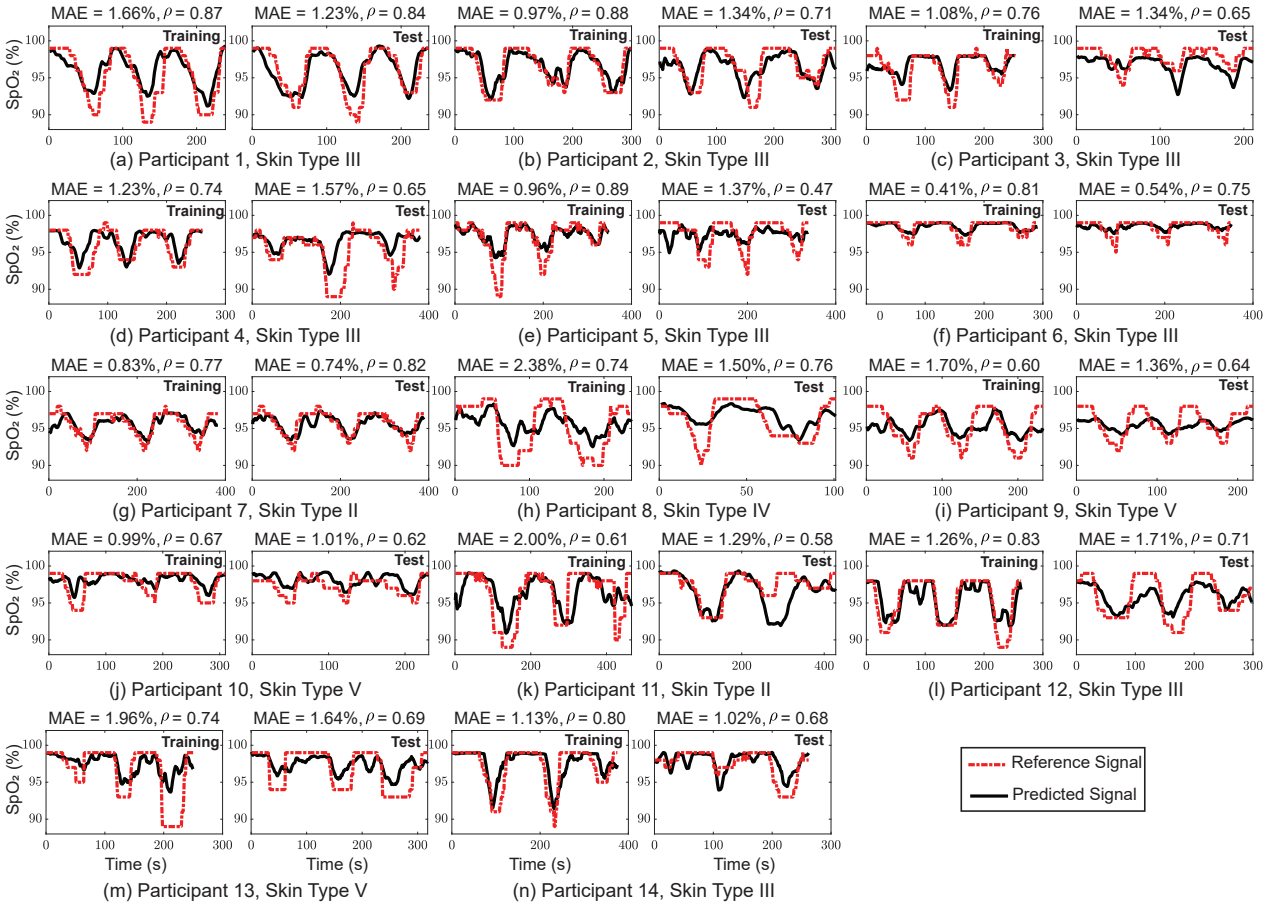


Figure 4.5: Predicted SpO_2 signals for all participants using SVR when the palm is facing the camera, i.e., the palm-up scenario. Prediction results of training and testing sessions are shown for each participant with reference SpO_2 in red dash lines and predicted SpO_2 in solid black lines. The higher the correlation ρ and the lower the MAE, the better the predicted SpO_2 captures the trend of the reference signal.

		Training		Testing	
		MAE	Correlation ρ	MAE	Correlation ρ
LR	PU	1.69% (±0.57%)	0.63 (±0.16)	1.52% (±0.54%)	0.62 (±0.11)
	PD	1.74% (±0.76%)	0.61 (±0.21)	1.53% (±0.53%)	0.56 (±0.21)
SVR	PU	**1.33%** (±0.54%)	**0.76** (±0.09)	**1.26%** (±0.33%)	**0.68** (±0.10)
	PD	1.35% (±0.45%)	0.75 (±0.09)	1.28% (±0.40%)	0.65 (±0.14)

Table 4.1: Performance of the proposed method. Results using linear regression (LR) and support vector regression (SVR) for both sides of the hand are quantified in terms of the sample mean and sample standard deviation (in parentheses).

we provide the skin-tone information in the subplot and show the accuracy indicators, MAE and ρ, for SpO_2 prediction. In all training sessions, MAE is below 2.4% and ρ is above 0.6. From this observation, we find that all the predicted SpO_2 signals in the training sessions are closely following the reference signals' trends, despite the exact value differences between the predicted and the reference signals, such as the differences around the last dip for participant #13. Furthermore, all testing MAE values are within 1.8%, suggesting that those trained models adapt well to the testing data. While there are a few cycles where the predicted signal does not fully follow the reference signal, such as the second dip for participant #4 and participant #11, the trends are consistent.

Table 4.1 summarizes the training and testing SpO_2 estimation performance of both LR and SVR based methods for both PU and PD cases. The best performance is achieved using the SVR method in the PU case. We further examine the difference between the two regression methods using boxplots in Fig. 4.6(a) that show the distributions of the correlation ρ for testing by LR and SVR, respectively. Each boxplot in Fig. 4.6(a) contains both PU and PD cases from all participants. The results are compared in terms of the

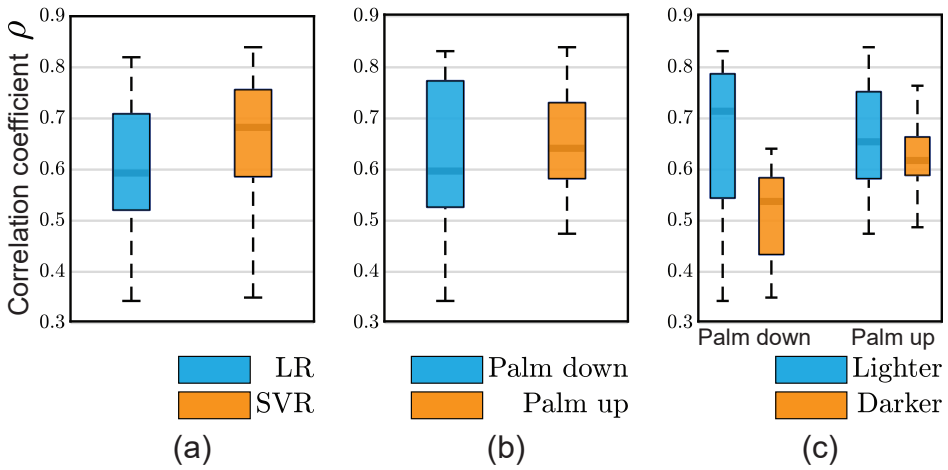


Figure 4.6: Boxplots of testing correlation coefficient ρ for all participants when grouped using different criteria. (a) Distributions contrasting linear and support vector regressions. (b) Distributions of palm-up and palm-down cases. (c) A detailed breakdown of (b) in terms of skin-tone subgroups.

median and the interquartile range (IQR). IQR quantifies the spread of the distribution by measuring the difference between the first quartile and the third quartile. The boxplots in Fig. 4.6(a) reveal that the SVR method outperforms LR with a higher median of 0.68 compared to 0.59 and with a narrower IQR of 0.17 compared to 0.19. This suggests that there may exist a nonlinear relationship between the extracted features and the SpO_2 values.

To examine the impact of the side of a hand and the skin tone on the performance of SpO_2 estimation, we analyze the following two research questions: (i) whether the side of the hand makes a difference in lighter skin (type II and III) or darker skin (type IV and V) or mixed skins (all participants), and (ii) whether the different skin tones matter in PU or PD case.

To answer question (i), we first focus on the distributions from PU and PD cases in Fig. 4.6(b) with each boxplot representing the correlation ρ in testing for all participants.

We observe that the PU case outperforms the PD case with a higher median of 0.64 compared to 0.60 and a narrower IQR of 0.15 compared to 0.25. We then zoom into each subgroup of skin tones shown in Fig. 4.6(c). For the lighter skin group, even though the median of PD case is 0.71, which is 9% better than that of PU, the IQR of PD case is 0.24, which is worse than the IQR of 0.17 of PU case. This suggests that the distributions are comparable between PU and PD cases for the lighter skin group. For the darker skin group, the PU case outperforms the PD case with a higher median of 0.62 compared to 0.54 and a narrower IQR of 0.07 compared to 0.15. In summary, there is no substantial difference between PU and PD cases in the lighter skin group, whereas, for the darker skin group and overall participants, the PU case is better than the PD case.

To answer question (ii), we first focus on the left two boxplots of Fig. 4.6(c). In the PD case, the median of the lighter skin group is significantly larger than that of the darker skin group by 31%, however, the lighter skin group also has a larger IQR. This makes it difficult to make a conclusion from the median–IQR analysis, hence we apply the t-test to complement our analysis. We note that the p-value is $0.037 < 0.05$, showing that there is a significant difference between these two groups. In the PU case shown in the right half of Fig. 4.6(c), the medians of the lighter skin group and darker skin group are 0.65 and 0.62, with IQR being 0.17 and 0.07, respectively. Thus, in our current dataset, no substantial performance difference is observed between lighter and darker skin tones in the PU case.

Method Index	Configuration		
	Multi-channel RoR features?	Narrow ABP filter?	Accurate HR tracking?
I	Two-channel RoR	✓	✓ (AMTC)
II	✓	No ABP	n/a
III	✓	Wide ABP	✓ (AMTC)
IV	✓	✓	Peak-finding
V	✓	✓	Weighted energy
Proposed	✓	✓	✓ (AMTC)

Table 4.2: Configurations for the ablation study of the proposed pipeline. The controlled experiments are conducted by replacing or removing one component at a time.

4.4.4 Ablation Study of Proposed Pipeline

In Sections 4.3.2 and 4.3.3, we have proposed three key designs in our algorithm, including a) the feature vector $\mathbf{f}$ containing pulsatile information from all RGB channels, b) the narrow ABP filter, and c) the passband of the ABP filter centered at precise HR frequency tracked by AMTC. To study the importance of each component, we conducted three controlled experiments by removing one factor at a time and the configurations of methods corresponding to the experiments are listed in Table 4.2. The results for the methods are illustrated in Fig. 4.7. The height of each bin shows the average correlation coefficient ρ or the MAE of SpO_2 estimation results from testing sessions (SVR, PU case) of all participants. Each pair of error bars corresponds to the 95% confidence interval that is calculated as $\pm 1.96\hat{\sigma}/\sqrt{N}$, where $\hat{\sigma}$ is the sample standard deviation and N is the sample size/number of participants.

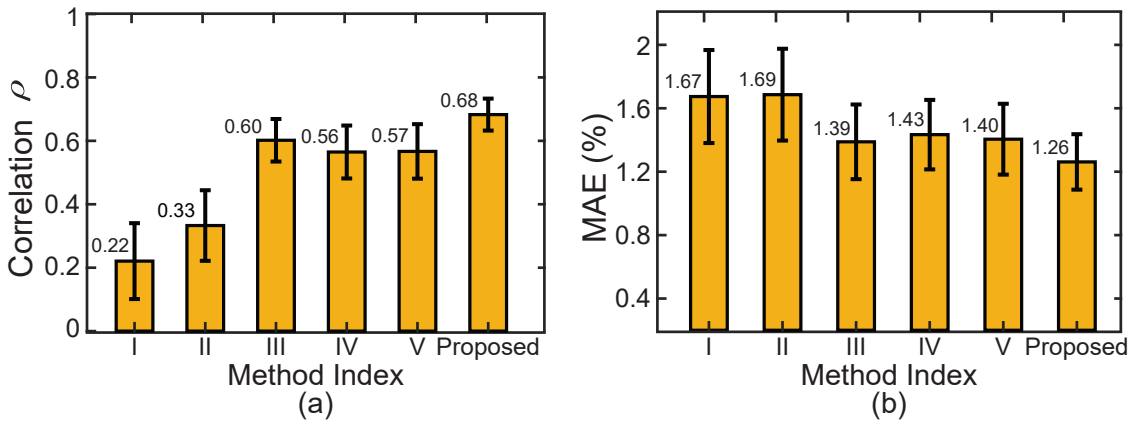


Figure 4.7: Ablation study of the proposed method. The bar plots are from testing sessions (SVR, PU case) of all participants. The error bars correspond to the 95% confidence intervals.

4.4.4.1 Advantage of The Proposed Multi-Channel RoR Over Two-Channel RoR

In this part, we compare our proposed algorithm with "**Method (I): RoR with nABP (AMTC)**." Method (I) follows the feature extraction method proposed in Section 4.3.3, including the *narrow adaptive bandpass filter (nABP)* centered at AMTC-tracked HR. The only exception is that, instead of using the feature vector $\mathbf{f}$ that contains multi-channel information, only the ratio of ratios between the red and blue channels as in traditional RoR methods is used.

Fig. 4.7 reveals that our proposed method outperforms method (I) by a big margin. More specifically, our proposed method improves the correlation coefficient from 0.22 to 0.68 and the MAE from 1.67% to 1.26%. This improvement confirms that our proposed multi-channel feature set helps with more accurate SpO_2 monitoring.

4.4.4.2 Contribution of Narrowband ABP Filter for Feature Extraction

Here we compare the following two methods to show the necessity of using a narrowband HR-guided bandpass filter:

- **Method (II): Feature vector without ABP** uses a nonadaptive, generic bandpass filter with the passband over $[1, 2]$ Hz, covering the normal range of heart rate in sedentary mode to replace the HR-based narrow ABP filter proposed in Section 4.3.3 for feature extraction.

- **Method (III): Feature vector with wide ABP (AMTC)** applies a wider ABP filter with ± 0.5 Hz bandwidth than the ± 0.1 Hz one used in our proposed method. This wider ABP filter's center frequency is provided by the AMTC tracking algorithm of the HR described in Section 4.3.2.

The bandpass filters used for methods (II) and (III) have the same bandwidth, 1 Hz. In terms of center frequency, method (II) used a fixed setting at 1.5 Hz, while method (III) is adaptively centered at the estimated HR value. Compared to method (II), method (III) has an improved testing MAE by 18%. Furthermore, compared to method (III), our proposed method with a narrow ABP filter improves the correlation coefficient ρ for testing by 13% and MAE by 9%, suggesting the contribution of the narrow HR-based ABP filter strategy for AC computation.

4.4.4.3 Importance of Accurate HR Tracking on SpO_2 Monitoring

We consider the following two methods to compare with our proposed method:

- **Method (IV): Feature vector with narrow ABP (peak-finding)** applies a narrow ABP filter of bandwidth ± 0.1 Hz for extracting the feature vector $\mathbf{f}$. The center frequency of the ABP filter is the HR estimated from the peak-finding algorithm described in Section 4.3.2.

- **Method (V): Feature vector with narrow ABP (weighted)** is similar to method (IV), except that the frequency estimation algorithm is replaced by the weighted energy in Section 4.3.2.

The averaged MAE of the HR estimation for all participants by the peak-finding algorithm, weighted frequency estimation algorithm, and AMTC algorithm are 7.11 (± 3.66) bpm, 6.42 (± 3.02) bpm, and 4.14 (± 1.72) bpm, respectively.

Fig. 4.7 shows that methods (IV) and (V) perform similarly with 0.56 vs. 0.57 for correlation ρ and 1.43% vs. 1.40% for MAE, respectively. Our proposed method guided by the AMTC tracked HR outperforms methods (IV) and (V) by 21% and 19% in correlation, and by 12% and 10% in MAE, respectively. These results suggest that the accurate HR estimation for ABP filter design improves the quality of the AC magnitude by preserving the most cardiac-related signal from RGB channels, which in turn helps with accurate SpO_2 monitoring.

4.4.5 Leave-One-Out Experiments

As a proof of concept and considering the currently limited amount of available data, we have so far discussed the SpO_2 estimation under the *participant-specific (PS)* scenario in Section 4.4 where the models are calibrated for each individual. This PS

	LOPartO		LOSessO		PS	
	MAE	ρ	MAE	ρ	MAE	ρ
PU	1.70% (±0.60%)	0.53 (±0.38)	1.59% (±0.58%)	0.55 (±0.36)	1.26% (±0.33%)	0.68 (±0.10)
PD	1.76% (±0.59%)	0.48 (±0.38)	1.70% (±0.59%)	0.50 (±0.39)	1.28% (±0.40%)	0.65 (±0.14)

Table 4.3: Testing results of leave-one-participant-out (LOPartO) and leave-one-session-out (LOSessO) experiments, measured in the sample mean and the sample standard deviation (in parentheses).

mode corresponds well to the trending "precision telehealth" that tailors the healthcare service to individuals.

In this subsection, we consider a more practical scenario where the test participant's data are never seen or only form a limited portion of the training data. In this scenario, we can develop a group-based model based on skin tone or other determinants of health, and for each subgroup, the model is "universal" and participant-independent. We will examine this group-based model through the following two modes of leave-one-out experiments:

- *Leave-one-session-out (LOSessO)*: when testing on a given participant, we include his/her training session data together with other participants' data for training.
- *Leave-one-participant-out (LOPartO)*: when testing on a given participant, we only use other people's data for training and leave out the data from this test participant.

We group the participants by skin type into lighter skin color (skin types II and III) and darker skin color (skin types IV and V) groups. We conduct LOSessO and LOPartO experiments on each subgroup and obtain the SVR generated testing results from all participants in Table 4.3. The MAE and correlation coefficient ρ improve from LOPartO to LOSessO to PS for both PU and PD cases. This result suggests that the precision

telehealth inspired PS mode is the most accurate approach to monitoring SpO_2 for an individual. Based on the overall results shown in Table 4.3, most participants demonstrate a consistent trend of the accuracy of SpO_2 estimation from LOPartO to LOSessO to PS case. The correlation ρ of participant #12 is less than -0.5 in both leave-one-out modes, suggesting that this participant may have some uncommon relation compared to others between the extracted features and SpO_2 values.

4.5 Discussions

4.5.1 Performance on Contact SpO_2 Monitoring

In addition to contact-free SpO_2 monitoring, we evaluate whether our proposed algorithm can be applied to a contact-based smartphone setup. To collect data, the left index finger covers the smartphone's illuminating flashlight and the nearby built-in camera, and the camera captures a pulse video at the fingertip. Another smartphone is used to simultaneously record a top view video of the back side of the right hand whose index finger is placed in the oximeter for SpO_2 reference data collection. One participant took part in this extended experiment where one training session with three breath-holding cycles was recorded, and three testing sessions were recorded 30 minutes after the training session.

In Table 4.4, we compare the performance of our proposed algorithm in both the contact-based and contact-free SpO_2 measurement settings. The conventional RoR models used in [123] and [96] were implemented as baseline models for contact-based SpO_2 measurement. In [123], the mean and standard deviation of each window from the red and blue channels are calculated as the DC and AC components. A linear model was built

		Training		Testing	
		MAE	ρ	MAE	ρ
Contact	RoR [123] (LR)	1.60%	0.54	1.38%	0.64
	RoR [123] (SVR)	1.14%	0.73	1.32%	0.60
	RoR [96] (LR)	1.47%	0.62	1.39%	0.63
	RoR [96] (SVR)	0.99%	0.83	1.27%	0.66
	Proposed	**0.91%**	**0.84**	**1.17%**	**0.81**
Contact-free	RoR (2-channel)	1.61%	**0.73**	1.75%	0.36
	Proposed	**1.36%**	0.62	**1.29%**	**0.68**

Table 4.4: Comparison of the proposed algorithm in both contact and contact-free SpO_2 estimation settings. The testing results are measured in the average MAE and correlation coefficient ρ.

to relate the ratio-of-ratios from the two color channels with SpO_2. In [96], the median of the pulsatile peak-to-valley amplitude is regarded as the AC component. For the two RoR methods, we implemented both LR and SVR. For contact-free SpO_2 measurement, we take the traditional two-color channel RoR method implemented in Section 4.4.4 as the baseline to compare with the proposed method.

Table 4.4 reveals that our proposed algorithm outperforms other conventional RoR models in contact-based SpO_2 monitoring. Even in the contact-free case, our algorithm presents a comparable performance to that of the contact-based cases, despite that the SNR of the fingertip video is better than the SNR from a remote hand video.

4.5.2 Resilience Against Blurring

In this subsection, we explore the robustness of our algorithm to the blurring effect on hand images. In the current setup, the hands are placed on a stable table with a cellphone camera acquiring the skin color of both hands. Ideal laboratory conditions are often not satisfied under practical scenarios, and the hand images captured by the

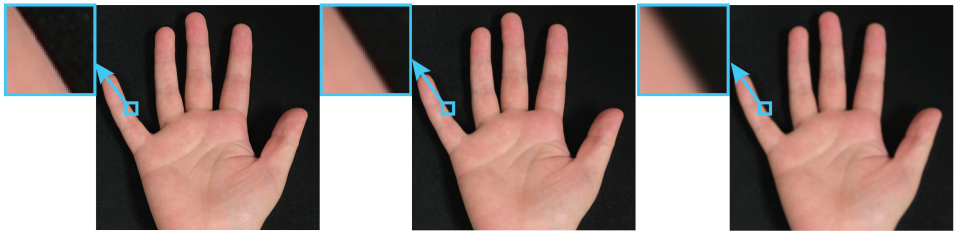


Figure 4.8: Illustration of blurring effects using different blurry levels σ on hand videos. The wider the kernel is, the blurrier the videos are.

	Training		Testing	
	MAE	ρ	MAE	ρ
$\sigma = 2.6$ blur (15×15 pixels)	1.41% ($\pm$0.50%)	0.72 ($\pm$0.11)	1.31% ($\pm$0.35%)	0.67 ($\pm$0.09)
$\sigma = 1.1$ blur (5×5 pixels)	1.42% ($\pm$0.59%)	0.70 ($\pm$0.16)	1.34% ($\pm$0.41%)	0.68 ($\pm$0.10)
No blur	1.33% ($\pm$0.54%)	0.76 ($\pm$0.09)	1.26% ($\pm$0.33%)	0.68 ($\pm$0.10)

Table 4.5: Simulation for Gaussian blurring effect on hand videos. SVR generated results for PU cases are listed for different σ and Gaussian kernel sizes. The results are quantified in terms of the sample mean and sample standard deviation (in parentheses).

cellphone cameras may be blurred due to being out of focus. The point spread function is modeled as a 2D homogeneous Gaussian kernel. The finite support of the kernel is defined manually to generate perceptually different blurry effects and then the standard deviation σ is computed based on the given support. To test different blurry effects, we experimented with two different blurry levels $\sigma = 1.1$ (5×5 pixels) and $\sigma = 2.6$ (15×15 pixels), respectively. We show the blurring effects in Fig. 4.8.

Table 4.5 presents the SVR generated results for PU cases with different σ and kernel sizes. We use the SVR, PU scenario to showcase here as it achieves the best SpO_2 prediction performance, which is verified in Section 4.4.3. From the table, we find that our algorithm is robust to the Gaussian blurring effect. After the $\sigma = 1.1$ blurring, the

testing ρ remains the same, and testing MAE is 6.3% higher than the no blurring case. After the $\sigma = 2.6$ blurring, the testing ρ is 1.5% lower and MAE is 4.0% higher than the no blurring case.

4.5.3 Limitations and Further Verification with Intermittent Hypoxia Protocols

From the recordings of our data collection protocol for voluntary breath-holding, we observed that HR and SpO_2 are correlated for many participants. That is, in one breath-holding cycle, when the participant starts to hold their breath, his/her HR increases and SpO_2 drops as the oxygen runs out. As he/she resumes normal breathing, his/her HR and SpO_2 recover to be within the normal range. Due to individuals' different physical conditions, in some participants, the peak of the HR signal and valley of the SpO_2 signal happen in such a short time interval that HR and SpO_2 are significantly negatively correlated. This observation is in line with the biological literature [56]. In the literature, breath-holding exercises were found to be able to yield significant changes in the cardiovascular system. In the central circulation, they caused significant changes in heart rate, and in the peripheral circulation, they caused significant changes in arterial blood flow and oxygen saturation.

Based on the above observation that HR is correlated with SpO_2 during breath-holding, we are curious whether our method also works for a different protocol where the instant HR change is relatively less correlated to SpO_2. An *intermittent hypoxia (IH) protocol* used in the literature shows that by receiving hypoxic air (inspired fraction of

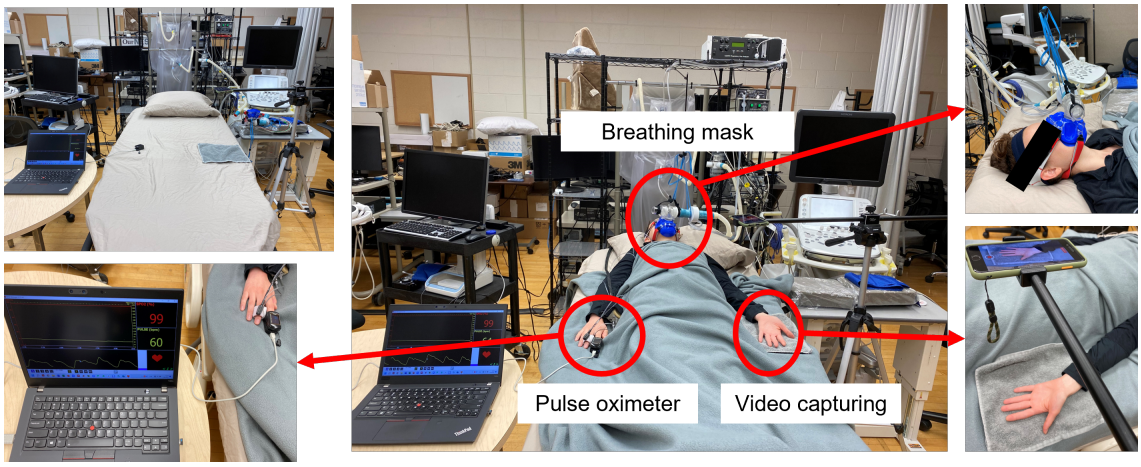


Figure 4.9: Experimental setup for the intermittent hypoxia protocol. The participant lies down on a bed with a mask controlling the breathing-in air which alternates between hypoxia and normoxia. The right index finger is clipped by the CMS-50E pulse oximeter to record the reference SpO_2 and HR signals. The palm side of the left hand (PU) is facing toward the smartphone camera during hand video recording sessions.

oxygen between 12% and 15%) intermittently with normoxic air, the participant can have a much milder HR change than breath-holding, while a significant decrease in SpO_2 can be achieved during the hypoxia [46]. The research restriction affecting human subject research in many U.S. institutions limited our ability to carry out the abovementioned hypoxia protocol before and as the restriction is eased recently, we investigate the performance of our proposed algorithm when applied to the new hypoxia protocol.

IH Protocol and Data Collection Setup:

Similar to the breath-holding protocol used in Chapter 4.4.1, the data collection setup of the IH protocol (shown in Fig. 4.9) includes a Contec CMS-50E pulse oximeter attached to the right index finger to measure the participant's SpO_2 and HR level as the reference and an iPhone 7 Plus camera mounted on a tripod for hand video recording. In lieu of holding breath to induce variation in SpO_2 values, in the IH protocol, the participant is equipped with a face mask that controls the breathing-in air. The face mask is

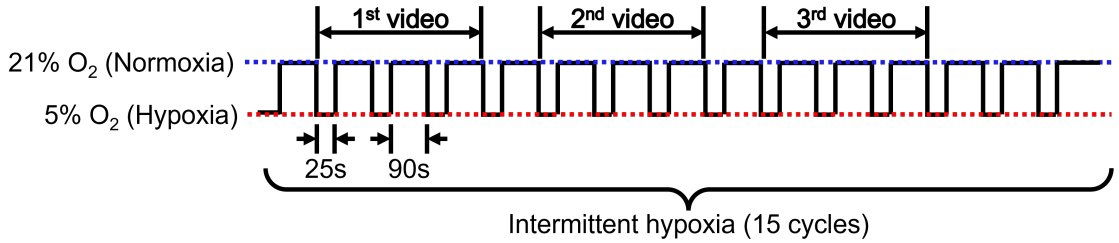
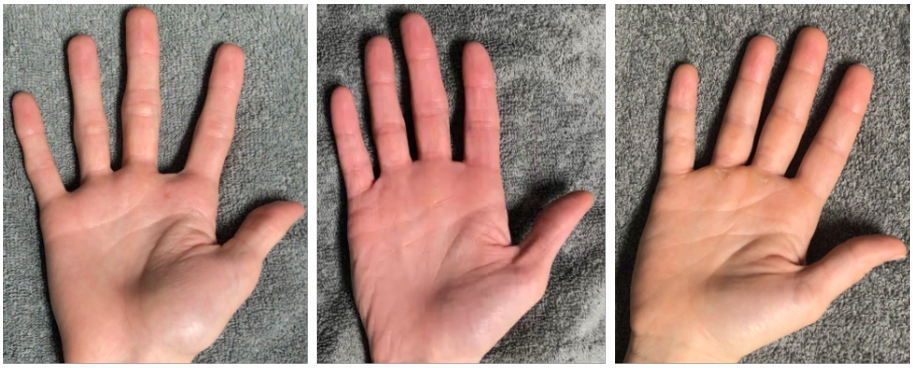


Figure 4.10: Illustration for the intermittent hypoxia (IH) protocol. The IH breathing is composed of 15 cycles of exposures to the alternating 25-second hypoxia (5% oxygen) period and the 90-second normoxia (21% oxygen) period. Three hand video sessions are recorded during the process and each video takes around 4.5 minutes. The first, second, and third videos start at the 2nd, 6th, and 10th IH cycle, respectively. In the practical data collection, the start time of the second and third video can be delayed by 1 or 2 cycles for the participants to adjust their hand positions after the previous video session.

connected to a one-way non-rebreathing valve, which is attached to a two-way switching valve. The two-way switching valve is used to control the input of either hypoxic air (5% oxygen, 3% carbon dioxide, balanced nitrogen) or room air (normoxia: 21% oxygen). Throughout the protocol, a switching valve is alternated between the acute (25-second) exposures to hypoxic air and the 90-second exposure to the normoxic medical gas for a total of 15 hypoxic events. Three hand video sessions are recorded for each participant during the process and each video takes around 4.5 minutes. The illustration for the procedure is shown in Fig. 4.10.

Overview of Participant Information and Collected Data:

Three participants, including one male and two females, were enrolled in the study under protocol #1511266 approved by the University of Maryland IRB, with one female's Fitzpatrick skin type being type I and the other two participants' being type II. One frame from the hand videos where the palm side facing the camera (PU case) of each participant is shown in Fig. 4.11. According to the IH protocol described in the previous paragraph, each participant had three hand video sessions recorded while their SpO_2 and HR were



(a) IH-1 (b) IH-2 (c) IH-3

Figure 4.11: Hand images from (a) a male participant whose Fitzpatrick skin type is II, (b) a female participant whose Fitzpatrick skin type is I, and (c) a female participant whose Fitzpatrick skin type is II.

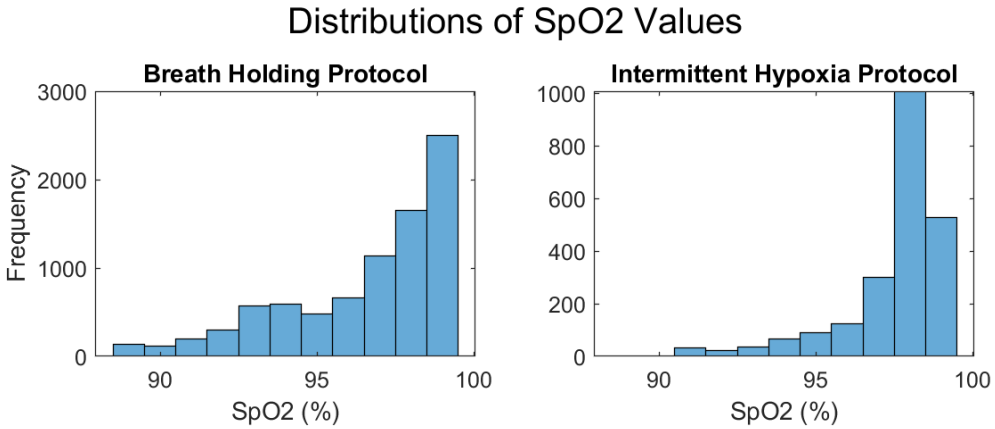


Figure 4.12: Comparison of the distributions of SpO_2 collected using the breath-holding protocol and the intermittent hypoxia protocol.

measured by the pulse oximeter during the intermittent hypoxia process. The histograms of SpO_2 values in the collected datasets using the breath holding protocol and the new IH protocol are shown in Fig. 4.12.

Recall in our previous breath holding protocol used in Chapter 4.4.1, we observed that some participants have their HR and SpO_2 correlated due to the reaction of the cardiac system during breath holding. This is manifested in the histogram shown in the left panel of Fig. 4.13, where 79% (22/28) of the participants' SpO_2 and HR have an absolute

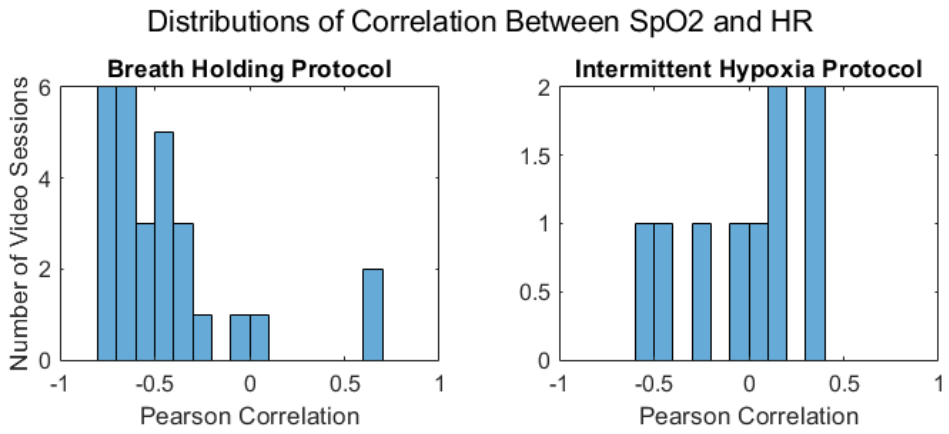


Figure 4.13: Comparison of the correlations between HR and SpO_2 from the breath-holding protocol and the intermittent hypoxia protocol.

correlation greater than a threshold of 0.4. While in the new intermittent hypoxia protocol, as shown in the left panel of Fig. 4.13, only 22% (2/9) have an absolute correlation greater than 0.4. This indicates that this new IH protocol induces less correlation between HR and SpO_2, serving as a new scenario to test the robustness of our proposed algorithm.

SpO_2 Prediction Performance:

The SpO_2 prediction is conducted in the participant-specific manner. The first video session of each participant is used for training and validation, and the second and third video sessions are used for testing. SVR is used for regression. Fig. 4.14 shows the training and testing results for the three participants. For Participant IH-1, the variation in the reference SpO_2 values is small with the lowest SpO_2 being 96% during the video sessions, resulting in no obvious dips in the SpO_2 trend. This may be due to the interpatient variability in tolerance to hypoxia. Thus, even though his predicted test SpO_2 signals do not follow the trend of reference SpO_2 well (with ρ being 0.22 and -0.26, respectively), the MAEs are less than 0.65%. Overall speaking, the test MAEs are within 1.64% while the dips of some of the SpO_2 trends are not captured well, such as Participant IH-2. From

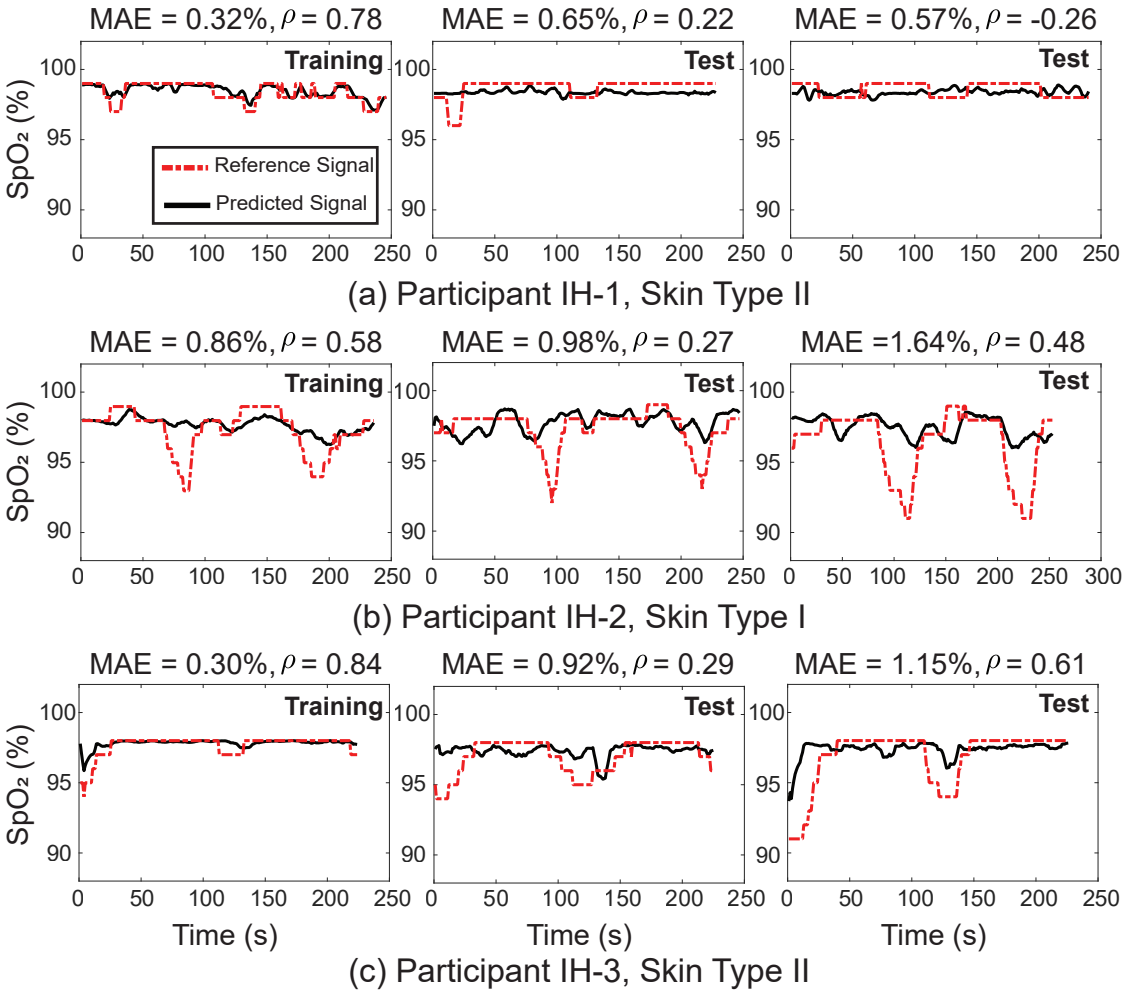


Figure 4.14: Predicted SpO_2 signals using SVR are shown for all participants from the IH protocol. The reference SpO_2 is in red dash lines and the predicted SpO_2 is in solid black lines. The higher the correlation ρ and the lower the MAE, the better the predicted SpO_2 captures the trend of the reference signal.

the limited data that we have collected so far with the IH protocol, our proposed algorithm achieved reasonable results and the results need to be verified and further improved with more data collected.

Discussions for Future Development:

From the comparison of SpO_2 distributions between the breath-holding protocol and the IH protocol shown in Fig. 4.12 and SpO_2 trends in Fig. 4.14 versus those in Fig. 4.5, we observe that the drop of SpO_2 does not get deeper and wider with the new IH protocol as described in the literature with similar IH protocols [46, 65]. The differences between our IH protocol and that in the literature mainly lie in the duration of the hypoxia period (in each episode and overall), its relative duration to the normoxia phase, and the fraction of the inspired oxygen. For example, in [65], the hypoxia environment induced by the FiO_2 (the fraction of inspired oxygen) protocol lasts for consecutively 16 minutes on average per participant to create a much wider range of SpO_2 from 61% to 100%, though the fraction of oxygen is unclear in the paper. In [46], the IH experiment is conducted 5 sessions per week throughout a 3-week duration. They found the most prominent decrease of SpO_2 was 10% on average, which happened in week 3 with 5 times of 5-minute hypoxia provoked by 12% oxygen interspersed with 3-minute normoxia intervals in each session.

With the IH protocols applied in the literature and the suggestions of a proper level of hypoxia and duration that lead to safe and positive effects and therapeutic potential of intermittent hypoxia [103], we consider the following modifications in our future design of protocol with advice and supervision from physicians to prevent adverse effects to the participants:

- having a relatively longer hypoxia period (e.g., increase from 25 seconds to several minutes); and/or
- inducing a modest fraction of inspired oxygen (e.g., 9% to 16%) [103] that can match the increased duration of the hypoxia period.

With the longer and larger decrease in SpO_2 values created by the updated protocol, we may have more meaningful training samples and better take advantage of the IH protocol.

4.6 Chapter Summary

This chapter presents a contact-free method of measuring blood oxygen saturation from hand videos captured by smartphone cameras. The whole algorithm pipeline includes 1) receiving video of the hand of a subject captured by a regular RGB camera of a smartphone; 2) extracting a region of interest of the hand video; 3) performing feature extraction of the region of interest based on spatial and temporal data analysis of more than two color channels; and 4) estimating a blood oxygen saturation level of the subject from the features. The key contributions of this chapter are mainly focused on the proposed feature engineering method, which is a synergistic combination of several key components, including the multi-channel ratio-of-ratios feature set, the narrowband filtering that adaptively centered at heart rates, and the accurately estimated heart rate. We have seen encouraging results of a mean absolute error of 1.26% with a commercial pulse oximeter as the reference, outperforming the conventional ratio-of-ratios method by 25%. We have also analyzed the impact of the sides of the hand and skin tones on the SpO_2 estimation. We have found that, given our collected dataset, the palm side performs well regardless of

the skin tone; for palm-up cases, we do not observe significant performance differences between lighter and darker skin tones.

Part of the research in this chapter was published in [142].