

Supplemental Materials

Tricam-rPPG: A Multimodal Multispectral Dataset for Remote Photoplethysmography

Abhijit Sarkar[✉], Surendrabikram Thapa[✉], Ishtiaque Ahmed Khan[✉],
Yogesh Deshpande[✉], and A. Lynn Abbott[✉]

Virginia Tech, Blacksburg, Virginia, USA

A.1 Introduction

This document provides additional material to complement the main paper, *Tricam-rPPG: A Multimodal Multispectral Dataset for Remote Photoplethysmography*. The goal of the supplemental material is to provide deeper insights into the dataset design, theoretical motivation, and analysis procedures that support the contributions presented in the main manuscript.

Remote photoplethysmography (rPPG) enables contactless recovery of physiological signals from subtle changes in skin appearance captured by cameras. While recent advances in computer vision and deep learning have significantly improved rPPG performance, many existing datasets rely primarily on RGB imaging and provide limited support for studying spectral variability, sensing bias, and robustness across demographic groups. The Tricam-rPPG dataset addresses these limitations by providing synchronized recordings from three co-located cameras: a high-resolution RGB camera and two near-infrared (NIR) cameras operating at 850 nm and 940 nm, together with physiological ground truth collected using a fingertip PPG sensor.

The supplemental material expands on several aspects of the dataset and experimental framework that could not be described in detail in the main paper due to space constraints. In particular, this document includes:

- additional theoretical motivation for using NIR sensing to mitigate pigmentation dependent attenuation effects in rPPG signals,
- additional discussion of the limitations of existing rPPG datasets and how the proposed dataset addresses these gaps,
- extended descriptions of the acquisition setup, instrumentation, and experimental protocol used during data collection,
- further information about data preprocessing, metadata generation, and dataset structure
- supplementary benchmarking results and analysis using standard rPPG algorithms.
- Additional results from LSTM model showing the effect of NIR data on darker skin tone.

Together, these materials provide a more comprehensive description of the Tricam-rPPG dataset and are intended to facilitate reproducibility, enable deeper analysis of multispectral rPPG signals, and support future research on robust and fairness-aware physiological sensing.

This paper mainly focus on the Dataset aspect. Hence, we do not provide a detailed literature review for all methods developed over the last two decades. Reader may refer to previous works [7, 19, 32].

A.2 NIR effectiveness for skin with high melanin concentration

A.2.1 Motivation

Most previous work on imaging-based rPPG has relied extensively on sensing in the visible band. However, studies have shown that performance suffers as melanin concentration increases. As an alternative to RGB cameras, near-infrared (NIR) sensing shows promise because melanin absorption decreases substantially as wavelengths increase into the NIR region. Inclusion of NIR-based sensing potentially reduces pigmentation-induced variability in reflectance and improves SNR parity across skin tones [12]. However, NIR alone does not guarantee fairness because exposure/ISP effects can reintroduce group-dependent distortions, motivating datasets and evaluations that disentangle algorithmic effects from acquisition physics [24]. In this section we aim to show how NIR characteristics for skin tones has potential to reduce bias.

Assuming skin to be Lambertian [13, 34], for a narrowband channel centered at wavelength λ (e.g., 850 nm or 940 nm), we write a simplified measurement model

$$I_\lambda(t) = g(E(\lambda) S(\lambda) R(\lambda, t)) + n_\lambda(t), \quad (\text{A.1})$$

where $E(\lambda)$ is the illumination spectral power, $S(\lambda)$ is the camera spectral sensitivity, $R(\lambda, t)$ is diffuse skin reflectance, $g(\cdot)$ captures camera/ISP nonlinearity (e.g., exposure control, gain, tone mapping/gamma), and $n_\lambda(t)$ denotes noise. This formulation explicitly separates *optical/tissue* factors from *system* factors, aligning with rPPG practice where exposure and ISP choices strongly influence recoverable physiology [24].

In scattering biological tissue, diffuse reflectance depends on the tissue’s absorption coefficient $\mu_a(\lambda, t)$ and reduced scattering coefficient $\mu'_s(\lambda)$ [11, 12]. We express this dependency as

$$R(\lambda, t) = f(\mu_a(\lambda, t), \mu'_s(\lambda)), \quad (\text{A.2})$$

where $f(\cdot)$ is a smooth function determined by the diffusion approximation or related radiative transfer models. The absorption coefficient $\mu_a(\lambda, t)$ represents the rate at which light energy is absorbed per unit path length. Under Beer–Lambert modeling for mixed absorbers, the absorption coefficient can be expressed as a linear combination of chromophore concentrations and their wavelength-dependent extinction spectra [10, 12]:

$$\mu_a(\lambda, t) = \sum_i c_i(t) \varepsilon_i(\lambda). \quad (\text{A.3})$$

Following standard tissue optics and modeling, the absorption coefficient is written as a sum of chromophore contributions [10, 12]:

$$\mu_a(\lambda, t) = c_{\text{mel}} \varepsilon_{\text{mel}}(\lambda) + c_{\text{HbO}_2}(t) \varepsilon_{\text{HbO}_2}(\lambda) + c_{\text{Hb}}(t) \varepsilon_{\text{Hb}}(\lambda) + \mu_{a,\text{bg}}(\lambda), \quad (\text{A.4})$$

where c_{mel} captures melanin concentration proxy of the skin tone, while $c_{\text{HbO}_2}(t)$ and $c_{\text{Hb}}(t)$ captures the concentration from oxygenated and deoxygenated hemoglobin and varies with the cardiac cycle. The terms $\varepsilon_i(\lambda)$ are corresponding extinction/absorption spectra, and $\mu_{a,\text{bg}}(\lambda)$ accounts for other approximately time-invariant absorbers.

The extinction coefficient of melanin, $\varepsilon_{\text{mel}}(\lambda)$, is a wavelength-dependent intrinsic optical property measured experimentally. Extensive spectroscopy studies report that melanin absorption decreases monotonically with increasing wavelength from the visible into the near-infrared range [2, 12]. Empirically, melanin absorption in the NIR (e.g., 800–1000 nm) is substantially lower than in the green-visible range (500–600 nm). Consequently, the sensitivity of tissue absorption to melanin concentration diminishes as wavelength increases into the NIR. These observations show that most rPPG methods might be biased by sensor design and not only by limitations of data diversity limitations.

A.2.2 From chromophore absorption to reflectance variability

Equation (A.4) describes how tissue absorption arises from a combination of chromophore concentrations and their wavelength-dependent extinction spectra. Among these components, melanin concentration c_{mel} primarily captures variations in skin pigmentation, while hemoglobin terms vary temporally with the cardiac cycle.

In camera-based rPPG, the observed signal is derived from temporal changes in reflected light intensity from the skin surface. Since diffuse skin reflectance depends on the underlying optical absorption and scattering properties of tissue [11, 12], variations in $\mu_a(\lambda, t)$ propagate to variations in the measured reflectance signal $R(\lambda, t)$ used by rPPG algorithms.

Importantly, Eq. (A.4) shows that melanin affects the absorption coefficient through the product $c_{\text{mel}}\varepsilon_{\text{mel}}(\lambda)$. Because c_{mel} varies across individuals while the extinction spectrum $\varepsilon_{\text{mel}}(\lambda)$ is fixed by physics, the magnitude of pigmentation-driven variability in optical measurements is directly modulated by the wavelength-dependent absorption spectrum of melanin.

A.2.3 Quantifying pigmentation sensitivity

To analyze how strongly reflectance measurements depend on skin pigmentation, we formulate the sensitivity of the reflectance signal to changes in melanin concentration. Specifically, we consider the log-sensitivity of reflectance to c_{mel} ,

which measures the relative change in reflectance induced by changes in melanin concentration. This quantity provides a convenient scale-invariant measure of pigmentation influence on optical measurements.

Using the chain rule, this sensitivity can be decomposed into two terms: (i) the sensitivity of reflectance to absorption and (ii) the sensitivity of absorption to melanin concentration.

Bounded melanin sensitivity. We quantify pigmentation sensitivity via the log-sensitivity of reflectance to melanin concentration:

$$\frac{\partial \log R(\lambda, t)}{\partial c_{\text{mel}}} = \frac{\partial \log R(\lambda, t)}{\partial \mu_a(\lambda, t)} \cdot \frac{\partial \mu_a(\lambda, t)}{\partial c_{\text{mel}}}. \quad (\text{A.5})$$

From Eq. (A.4), $\frac{\partial \mu_a(\lambda, t)}{\partial c_{\text{mel}}} = \varepsilon_{\text{mel}}(\lambda)$, hence

$$\frac{\partial \log R(\lambda, t)}{\partial c_{\text{mel}}} = \frac{\partial \log R(\lambda, t)}{\partial \mu_a(\lambda, t)} \cdot \varepsilon_{\text{mel}}(\lambda). \quad (\text{A.6})$$

Reflectance in biological tissue varies smoothly with the optical absorption coefficient within the physiological range encountered in practice [11, 12]. Consequently, the sensitivity of log-reflectance to absorption remains bounded. We therefore assume,

$$\left| \frac{\partial \log R(\lambda, t)}{\partial \mu_a(\lambda, t)} \right| \leq K(\lambda), \quad (\text{A.7})$$

where $K(\lambda)$ is a wavelength-dependent constant that captures the maximum reflectance sensitivity to absorption within the operating regime. We obtain the bound

$$\left| \frac{\partial \log R(\lambda, t)}{\partial c_{\text{mel}}} \right| \leq K(\lambda) \varepsilon_{\text{mel}}(\lambda). \quad (\text{A.8})$$

This last equation states that the *maximum* melanin-driven variability in log-reflectance at wavelength λ is proportional to the melanin absorption spectrum $\varepsilon_{\text{mel}}(\lambda)$. Since $\varepsilon_{\text{mel}}(\lambda)$ decreases substantially into the near-infrared wavelengths [12], NIR imaging reduces an upper bound on pigmentation-induced variability in reflectance—and, by extension, reduces one physics-driven mechanism that contributes to cross-skin-tone SNR disparities in camera-based physiological measurement.

Implication and exceptions: NIR offers improvement, but system bias can remain. While Eq. (A.8) supports the utility of NIR for reducing sensitivity to pigmentation, Eq. (A.1) suggests that NIR-based sensing alone is not sufficient to eliminate bias: $E(\lambda)$, $S(\lambda)$, and especially the nonlinearity $g(\cdot)$ (auto-exposure, gain, tone mapping, compression) can introduce group-dependent distortions and SNR shifts that are not resolved by data balancing alone [24]. Our hypothesis is that sensing from two separate bands can be used to mitigate these problems. For this reason, we have chosen to create a dataset with 2 different NIR bands as well as the traditional visible (RGB) bands. The availability of video from these bands will enable controlled studies of residual acquisition-driven bias sources.

A.3 Limitations of Existing Datasets

Table 1 of the main paper lists several publicly available datasets that include NIR videos, reinforcing our message concerning growing awareness of the efficacy of sensing in the near-infrared band for rPPG purposes. Our Tricam-rPPG dataset, which is introduced in this paper, provides several advantages compared to those previous datasets. The main difference is that our dataset provides 2 NIR videos at different wavelengths (850 nm and 940 nm), whereas all other systems in the table provide at most 1 NIR video. We have used separate cameras that are identical in spatial resolution and frame rate (60 fps). Further, our resolution of 2046×2464 pixels is the highest of all datasets in the table. Our Tricam-rPPG dataset is also the first to provide video from Aria glasses along with RGB and NIR videos of a person’s face.

For comparison, the earliest dataset in Table 1 providing NIR video is VIPL-HR [22]. This dataset contains video from a single NIR camera, at 1920×1080 and 30 fps. TokyoTech [18] also provides NIR data from a single band at a high frame rate but lower resolution of 640×480 , using a prototype camera that also produces RGB video concurrently. The IMVIA-NIR dataset [3] is a collection of synthetic videos, representing a single NIR band and not including RGB.

The first publically available dataset containing both RGB and NIR video of faces for rPPG purposes was MR-NIRP [23], providing video from a single NIR band centered at 940 nm. Spatial resolution was significantly lower than ours at 640×640 , and the frame rate is 30 fps. Another dataset that provides RGB video and NIR video from a single band (730 to 1100 nm) is DDPM [27], which uses a high frame rate of 90 fps but lower resolution of 1920×1080 .

A.4 Tri-Cam Dataset Collection

A.4.1 Instrumentation and Design

Table A.1 provides a summary of recording devices that were used to create our Tricam-rPPG dataset. As described in the main paper, this dataset was obtained using three cameras mounted side by side to collect RGB video as well as two monochrome NIR videos. All three cameras are products of Allied Vision Technologies GmbH. The RGB camera model was Alvium 1800 U-508c, and the NIR cameras were Alvium 1800 U-511m with appropriate optical filters. All three cameras produced videos at a resolution of 2064×2464 with a frame rate of 60 fps. Illumination sources were provided in the visible spectrum as well as for the NIR cameras. In addition, participants wore ARIA glasses, which provided periocular (eye-region) NIR video at 30 Hz, enabling simultaneous analysis of eye dynamics alongside face/hand recordings. Physiological ground truth was obtained using a BIOPAC MP46 fingertip system with a sampling rate of 125 Hz.

Table A.1: Details of recording devices used in the data collection for Tricam-rPPG.

Device	Configuration	Captured Signal
RGB camera	2064×2464, 60 fps	RGB video (face + hand)
NIR 850 camera	2064×2464, 60 fps	NIR video (face + hand)
NIR 940 camera	2064×2464, 60 fps	NIR video (face + hand)
BIOPAC MP46	125 Hz sampling rate	Fingertip PPG
Meta Aria glasses	Custom	NIR video (eyes), IMU,

A.4.2 Additional Task Description

Participants were guided through 2 sessions, consisting of 6 tasks and 2 tasks, respectively. The tasks are now described in more detail than was provided in the main paper. For all cases in which a video was shown to the participant, audio was also provided.

- **Session 1, Task 1** (S1_T1 - 5 minutes): The participant was asked to sit still and not talk, while facing the computer screen and cameras. The participant was asked to observe the computer screen, which was dark except for a simple fixation addition (+) symbol in the center. One purpose of this task is to provide a baseline for comparing physiological responses.
- **Session 1, Task 2** (S1_T2 - 5 minutes): The participant was asked to sit still and observe the computer screen, which presented a sequence of simple arithmetic problems and mental calculations. The participant was asked to solve the problems mentally, however, they are allowed to speak.
- **Session 1, Task 3** (S1_T3 - 5-7 minutes): The participant was asked to sit still and watch a video on the computer screen. The video was chosen to have content that is relatively calm. It was about traveling and photography in New Zealand.
- **Session 1, Task 4** (S1_T4 - 5-7 minutes): The participant was asked to sit still and watch a second video on the computer screen. The video was chosen to have content that altered their affective arousal and valence. It was a video about a family member passing away and was intended to elicit feelings of sadness and empathy.
- **Session 1, Task 5** (S1_T5 - 5 minutes): The participant was asked to turn to the side and from the camera. They were allowed to change their posture. The participant was asked to sit still in that position. This part allowed to capture the side view face of the participants.
- **Session 1, Task 6** (S1_T6 - 5 minutes): The participant was asked to sit still and not to talk, while facing the computer screen and cameras. The participant was asked to raise their hand beside their face. This shows the palm of the participant. This effort was introduced to facilitate research in studying pulse transit time between body parts.

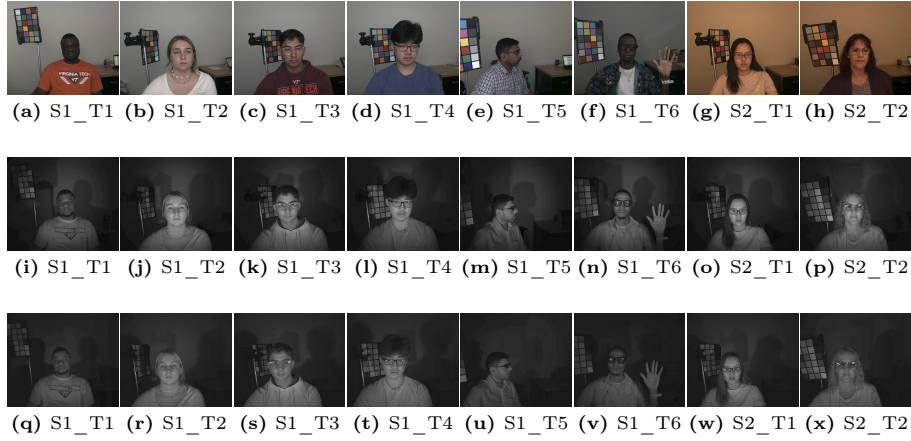


Fig. A.1: Example synchronized RGB, NIR 850nm, and NIR 940nm frames across eight participants in the Tricam-rPPG dataset. Each row corresponds to a modality; each column corresponds to a participant/session/task.

All of the tasks described above (Session 1) were performed using cool lighting conditions (8000 K). The two tasks described below (Session 2) were performed under warm lighting conditions (2700 K).

- **Session 2, Task 1** (S2_T1 - 5 minutes): The participant was asked to sit still, face the computer screen, and read silently from the computer screen. The screen showed chapter five from Moby Dick [21]. The participant read out loud. However, the audio was not recorded. This allows us to see how the rppg can be affected by a person speaking and can also help to study the effects of the face moving due to speaking.
- **Session 2, Task 2** (S2_T2 - 5-7 minutes): The participant was asked to sit still and watch a third video on the computer screen. It was a video about South Korean culture and the South Korean city of Seoul. It is a video elucidating excitement stimuli in the participant.

Figure A.1 Shows some additional example from the dataset across different participants and different sessions. The first column provides an example for one participant during Session 1, Task 1, under cool lighting conditions. The fifth column gives an example from Session 1, Task 5, with the profile view. The sixth column gives an example from Session 1, Task 6, with the raised hand. The final two columns show examples for two different participants during the two Session 2 tasks, under warm lighting.

A.5 Data Processing and Metadata

The dataset is preprocessed using standard face detection, facial keypoint estimation, and skin segmentation pipelines. Along with the raw sensor streams, we

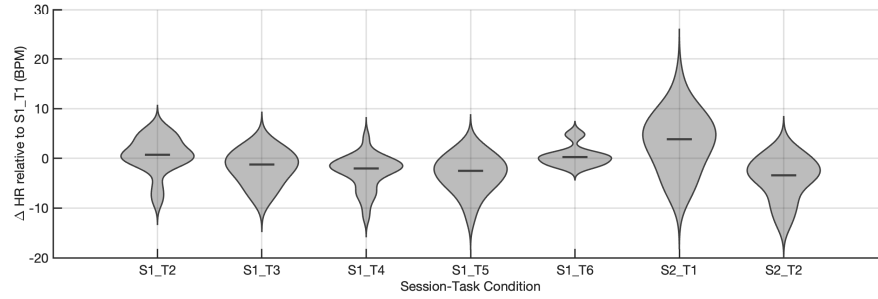


Fig. A.2: Relative changes in HR for each task with respect to the baseline task, S1_T1, which refers to Session 1, Task 1.

provide several derived attributes to facilitate rapid experimentation. These annotations are generated using widely used algorithms. Specifically, we provide: (1) **Camera Data:** synchronized recordings from three cameras stored using lossless compression (uncompressed raw streams are available upon request); (2) **Aria Data:** IMU signals, periocular NIR video; (3) **PPG Data:** raw fingertip PPG signals from a BIOPAC sensor synchronized with the video streams; (4) **PPG Landmarks:** peak locations for each pulse cycle along with a continuous signal-quality indicator; (5) **Facial Keypoints:** 468 facial landmarks for every frame in both RGB and NIR videos [17]; (6) **Skin Maps:** binary masks indicating skin pixels for each frame [33]; (7) **Patch Intensity Signals:** using facial landmarks, the face region is subdivided into 760 patches. For each patch, the mean pixel intensity is computed over time and stored as a compact temporal representation of spatial photoplethysmographic activity.

A.5.1 HR from PPG Data

The raw photoplethysmography (PPG) signal was processed to estimate beat-to-beat heart rate using a peak-based detection pipeline. The processing consisted of detrending, band-pass filtering, peak detection, and artifact correction.

All PPG signals were sampled at 125 Hz. The raw waveform was first detrended using linear detrending to remove slow baseline variations caused by sensor drift, illumination fluctuations, or gradual physiological trends unrelated to the cardiac pulse. Removing these low-frequency components helps stabilize peak detection while preserving the pulsatile characteristics of the signal.

After detrending, the signal was band-pass filtered to isolate frequency components corresponding to physiological pulse activity. A fourth-order Butterworth band-pass filter was applied with passband frequencies of 0.5–8.0 Hz. Filtering was implemented using a forward–backward filtering procedure (filtfilt) to eliminate phase distortion and preserve the temporal alignment of the signal.

Cardiac cycles were then identified by detecting systolic peaks in the filtered PPG waveform. Peaks were detected using a prominence-based local maxima detection algorithm. To avoid multiple detections within a single cardiac cycle,

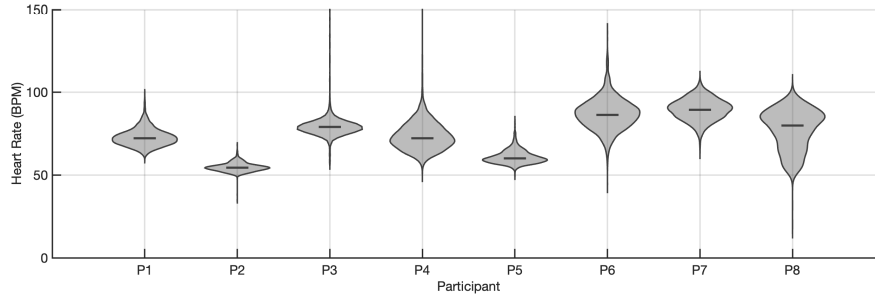


Fig. A.3: HR variations between different participants across all tasks. The plots indicate different levels of variance across participants.

a minimum separation constraint between successive peaks was enforced. This minimum distance was determined based on a maximum physiologically plausible heart rate of 220 beats per minute, ensuring that adjacent peaks correspond to distinct cardiac cycles. In addition, a peak prominence threshold proportional to the standard deviation of the filtered signal was used to suppress detections caused by noise or minor fluctuations in the waveform. Once peaks were identified, inter-beat intervals (IBIs) were computed as the time difference between consecutive peaks.

Because peak detection can occasionally produce spurious values due to motion artifacts or transient noise, an artifact correction step was applied to the resulting heart rate time series. Heart rate values outside the physiologically plausible range of 40–220 BPM were treated as invalid measurements. In addition, a Hampel filter was applied to the heart rate sequence to detect and correct outliers.

A.5.2 HR Characteristics

The instantaneous HR data in the dataset ranges between 40 BPM and 136 BPM. The average HR within sessions vary between 56 BPM and 96 BPM with std. ± 9.6 BPM. Fig. A.2, A.3, and A.4 illustrate the variability of physiological responses across tasks, individual participant’s HR variation, and the temporal dynamics of the heart rate signals.

Figure A.2 summarizes these effects at the task level by showing the distribution of heart rate deviations relative to the baseline condition (S1_T1). The boxplots reveal that different tasks produce distinct patterns of heart rate modulation across participants. Some tasks, such as S1_T6 and S2_T1, tend to produce positive deviations from baseline for multiple participants, whereas other tasks (e.g., S1_T4 and S1_T5) show more mixed responses including decreases in heart rate. The spread of the boxplots further indicates substantial variability between tasks and individuals. Fig. A.3 Shows HR distribution from eight participants. The participants shows different range of variability when they performed the tasks.

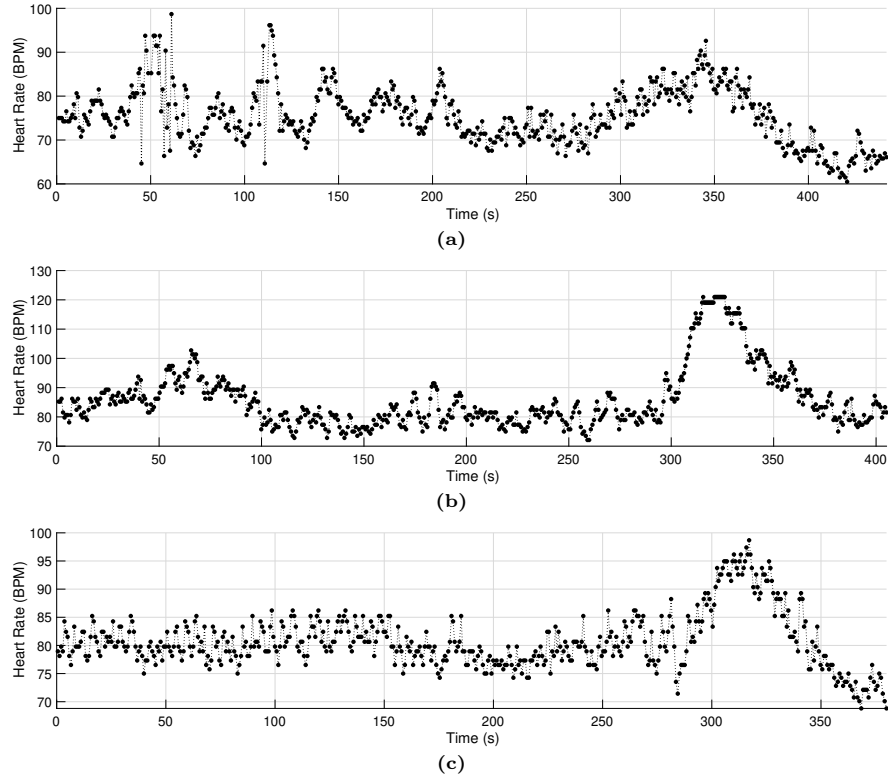


Fig. A.4: Examples of heart rate signals derived from PPG recordings. The S1_T4 results show HR dynamics around 300s mark to reflect effect of emotional stimuli.

Figure A.4 presents representative heart rate time series derived from the PPG recordings for three different participants. The signals exhibit clear beat-to-beat variability and gradual temporal changes throughout the session, reflecting natural physiological fluctuations rather than artificially smoothed trends. While baseline segments remain relatively stable, noticeable changes in heart rate appear during specific task periods. In particular, around the 300s mark corresponding to the emotional stimulus task (S1_T4), several participants show pronounced increases or shifts in heart rate, indicating a measurable physiological response to the stimulus. At the same time, the magnitude and temporal pattern of these responses vary across participants, highlighting inter-individual differences in cardiovascular dynamics.

Additionally, Scatter plots are provided (Fig A.5) to illustrate results for one light-skinned subject from our database, after training on a) PURE and b) UBFC-rPPG. Taken together, these visualizations demonstrate that the dataset captures both rich temporal dynamics and diverse physiological responses across participants and experimental conditions. This diversity is important for evalu-

ating algorithms designed to recover heart rate and related cardiovascular signals from video under realistic and heterogeneous conditions.

A.5.3 Face Data Processing

In addition to the video recordings, BVP signals, and data collected from the ARIA device, we also provide processed facial representations to facilitate downstream analysis. These include facial keypoints, skin segmentation masks, and geometry-aware facial patches derived from the facial mesh. Example visualizations of these representations, derived from RGB frames, are shown in Fig. A.6, and those from NIR (850 nm) frames are shown in Fig. A.7.

Facial Keypoints Facial landmarks were extracted using the MediaPipe Face Mesh system [1, 17]. The model estimates a dense set of 468 facial landmarks for each detected face, providing detailed geometric information about facial structures, including the forehead, cheeks, nose, lips, jawline, and eye regions. These landmarks are represented as normalized coordinates relative to the image frame and can be easily converted to pixel coordinates.

Facial Patch Representation To enable geometry-aware analysis of facial regions, the dataset also includes triangular facial patches derived from the facial mesh created using the MediaPipe Face Mesh system. Using the detected facial landmarks, we construct a triangular mesh where each triangle is defined by a triplet of landmark indices corresponding to vertices in the facial landmark graph. This patch-based representation enables region-wise analysis of facial signals, allowing researchers to investigate which facial regions contribute most to downstream tasks. For example, different patches may contain varying levels of useful information depending on factors such as illumination, motion, or physiological signal strength.

Patch Intensity Vector For each patch, we compute the average pixel intensity of the pixels contained within the triangle. Let p denote a triangular facial patch defined by the vertices (v_1, v_2, v_3) obtained from the facial landmark mesh. For frame t , we define $\Omega_p(t)$ as the set of pixels that lie within the triangle corresponding to patch p . For a given color channel $c \in \{R, G, B, 850, 940\}$, the average intensity of patch p at frame t is computed as

$$I_p^{(c)}(t) = \frac{1}{|\Omega_p(t)|} \sum_{(x,y) \in \Omega_p(t)} I^{(c)}(x, y, t) \quad (\text{A.9})$$

where $I^{(c)}(x, y, t)$ denotes the pixel intensity at spatial location (x, y) for channel c at frame t , and $|\Omega_p(t)|$ represents the number of pixels inside the triangular patch.

By computing $I_p^{(c)}(t)$ for every frame in the video sequence, each facial patch produces a temporal signal

$$\mathbf{s}_p^{(c)} = \left[I_p^{(c)}(1), I_p^{(c)}(2), \dots, I_p^{(c)}(T) \right] \quad (\text{A.10})$$

where T is the total number of frames. These patch-wise temporal signals provide localized intensity trajectories that can be used to study spatial variations in physiological signals across the face. In the context of remote photoplethysmography, such representations allow analysis of which facial regions yield stronger pulsatile signals and how these signals vary under different lighting conditions, skin tones, and motion patterns. The results for Section 6.2 in the main paper and the additional results in Section A.8.

Skin Segmentation To isolate physiologically meaningful facial regions, we also provide skin segmentation masks generated using a color-invariant skin segmentation algorithm [33]. The method formulates skin detection as a semantic segmentation task using a U-Net architecture. Unlike traditional skin detection methods that rely heavily on color cues, this approach incorporates extensive color-space augmentation during training to encourage the network to learn texture and contextual features associated with skin regions.

A.6 Dataset Use Guide

To assist researchers in understanding how different sensing modalities can support specific lines of investigation, Table A.3 summarizes several key challenges in remote photoplethysmography (rPPG) and the corresponding dataset components that enable their study. These challenges arise from a combination of physiological variability, optical interactions with skin tissue, environmental conditions, and camera processing pipelines.

One fundamental problem in camera-based physiological sensing is *spectral bias caused by skin tone*. Melanin absorption reduces the effective signal-to-noise ratio of visible-light rPPG, particularly for darker skin tones. The inclusion of synchronized RGB and dual-wavelength NIR imaging, together with demographic metadata and reference PPG, enables researchers to investigate fairness across skin tones and explore spectral fusion strategies that may mitigate such bias.

Another major challenge is *illumination variability*. Changes in lighting temperature and intensity can alter skin reflectance and interact with camera exposure and color processing. The dataset includes recordings under distinct illumination conditions as well as a ColorChecker calibration target, allowing researchers to study illumination-invariant rPPG pipelines, photometric normalization techniques, and color calibration strategies.

Motion artifacts represent another critical limitation of rPPG systems. Head movements and subtle facial dynamics can obscure the weak pulsatile signal captured by cameras. By providing facial landmarks, inertial measurements from Meta Aria glasses, the dataset supports the development of motion-aware signal extraction methods and the study of motion robustness in physiological sensing.

To the best of our knowledge, this is the first dataset that includes egoIMU data along with external camera data.

Beyond environmental factors, Tricam-rPPG also enables investigation of several *physiological signal recovery problems*. The availability of synchronized reference PPG signals allows researchers to evaluate instantaneous heart rate estimation, inter-beat interval recovery, and heart rate variability (HRV) metrics derived from video-based signals. Furthermore, the combination of spatial facial patches, multi-spectral imaging, and high-fidelity reference waveforms facilitates research on pulse waveform reconstruction and morphology recovery, which remain open challenges in camera-based cardiovascular sensing.

The dataset is first to provide RGB video data with two NIR wavelengths. This will help in studying the effect of NIR in rPPG. Especially with diverse data from participants with darker skin tone (See Table A.2).

Finally, the dataset provides structured tasks and synchronized ground-truth physiological measurements that support systematic benchmarking of rPPG algorithms. These tasks introduce controlled variations in physiological state and recording conditions, enabling consistent evaluation of algorithm robustness across motion, illumination, and spectral sensing conditions.

Table A.2: Participant distribution across skin tone and gender for Tricam-rPPG.

Skin Tone	Male	Female
Light	5	2
Medium	6	4
Dark	12	2

A.7 Results and Benchmark Using Tricam-rPPG Dataset

A.7.1 Evaluation Metrics

This section describes the heart rate (HR) metrics used in our analysis to evaluate the performance of our models. Additionally, the signal-to-noise ratio (SNR) is used to evaluate the quality of the recovered HR signal. We followed the metrics reported in the rPPG-Toolbox [16].

Mean Absolute Error (MAE) of HR Predictions The mean absolute error (MAE) measures the average absolute difference between the predicted heart rate ($\hat{H}R_i$) and the ground truth heart rate (HR_i) for N videos:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \left| \hat{H}R_i - HR_i \right| \quad (\text{A.11})$$

Mean Absolute Percentage Error (MAPE) For predicted heart rate $\hat{H}R_i$, ground truth heart rate HR_i , and for N videos:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{HR_i - \hat{H}R_i}{HR_i} \right| \quad (\text{A.12})$$

Table A.3: Research challenges in camera-based physiological sensing that can be systematically studied using the modalities and metadata provided in Tricam-rPPG.

rPPG	Research Challenge	Relevant Modalities and Metadata
	Skin tone bias and spectral attenuation	RGB + NIR (850/940 nm), demographic metadata, reference PPG
	Ambient illumination variability	RGB streams, dual illumination sessions (cool/warm), ColorChecker calibration chart
	Motion artifacts	Facial landmarks, skin masks, Meta Aria IMU, periocular NIR video
	Camera noise and ISP effects	Lossless video streams, ColorChecker reference, patch intensity signals
	Spatial modeling of physiological signals	468 facial landmarks, 760 facial patches, skin segmentation masks
	Instantaneous heart rate estimation	Reference PPG peaks, synchronized video streams
	Heart rate variability (HRV) estimation	High-fidelity PPG waveform, multi-task recordings inducing physiological variability
	Pulse waveform and morphology recovery	Reference PPG waveform, RGB + NIR signals, spatial patch intensities
	Viewpoint and geometry robustness	Multi-view face recordings, side-view videos, palm recordings
	Cross-spectral physiological sensing	RGB + dual NIR wavelengths
	Algorithm benchmarking	Structured tasks, synchronized PPG ground truth

Pearson Correlation (ρ) For predicted heart rate $\hat{H}R_i$, ground truth heart rate HR_i , N videos, and $\bar{H}R$ the average ground truth heart rate (similarly $\bar{\hat{H}}R$ for predictions):

$$\rho = \frac{\sum_{i=1}^N (HR_i - \bar{H}R) (\hat{H}R_i - \bar{\hat{H}}R)}{\sqrt{\left(\sum_{i=1}^N (HR_i - \bar{H}R)^2\right) \left(\sum_{i=1}^N (\hat{H}R_i - \bar{\hat{H}}R)^2\right)}} \quad (\text{A.13})$$

Signal-to-Noise Ratio (SNR) : Signal-to-Noise Ratio (SNR) for a predicted signal is calculated as the ratio between the area under the curve of the power spectrum around the first and second harmonic of the ground truth heart rate frequency and the area under the curve of the rest of the power spectrum. This is mathematically represented as follows:

$$SNR = 10 \log_{10} \left(\frac{\sum_{45}^{150} (U_t(f)S(f))^2}{\sum_{45}^{150} ((1 - U_t(f))S(f))^2} \right) \quad (\text{A.14})$$

Where S is the power spectrum of the estimated rPPG signal. $U_t(f)$ is equal to 1 around the first and second harmonics of the ground truth rPPG signal, while being 0 elsewhere in the power spectrum. In the context of the rPPG-Toolbox, only the power spectrum between 0.6 Hz and 3.3 Hz, or 36 beats/min and 198 beats/min, is considered. We report the mean of the SNR values calculated per video or test sample, such that:

$$MSNR = \frac{1}{N} \sum_{n=1}^N SNR \quad (\text{A.15})$$

In this paper MAE is reported in BPM and SNR is reported in dB.

A.7.2 Results Using rPPG Toolbox

This section collectively provides a comprehensive assessment of model performance using the rPPG Toolbox [16]. For post-processing and evaluation, we follow a set of commonly used procedures in the toolbox to convert the predicted waveform into heart-rate estimates. First, the predicted PPG signal is band-pass filtered using a 2nd-order Butterworth filter with cut-off frequencies of 0.75–2.5 Hz to suppress low-frequency drift and high-frequency noise outside the physiological heart-rate range. After filtering, heart rate is computed from the processed signal using either frequency-domain or time-domain analysis. In this paper, frequency-domain approach, a Fast Fourier Transform (FFT) is applied to estimate the dominant spectral peak within the physiological band. This generates one heart rate values for each whole video. The predicted and ground truth heart rate values for all the videos are then used in the equations A.11 to A.13 to calculate the metrics. SNR reflects the shape of the recovered waveform, highlighting the model’s ability to suppress noise in HR estimation.

Table A.4 shows results from traditional or physics based models when tested on Tricam-rPPG data. We have tested ICA, POS, CHROM, GREEN, LGI, PBV, and OMIT. The table also shows their performance for different skin tones. Table A.5 summarizes results for supervised methods when testing on Tricam-rPPG data. We tested five algorithms: TS-CAN, Physnet, Physformer, DeepPhys, and EfficientPhys. Each algorithm is trained with three different datasets: UBFC-rPPG [4], PURE [28], and SCAMPS [20]. Additionally, we trained these models using Tricam-rPPG data and reported cross dataset results in Table A.6.

A.7.3 Skin Tone Based Benchmarking

We have provided detailed skin tone based results for all methods reported in Table A.5. Table A.7, Table A.8, Table A.9, Table A.10, and Table A.11 shows

Table A.4: Results of running inference on Tricam-rPPG dataset using traditional methods in the rPPG Toolbox [16] (mean $\pm$ std).

Test Set: Tricam-rPPG					
Method	Skin Tone	MAE $\downarrow$	MAPE $\downarrow$	ρ $\uparrow$	SNR $\uparrow$
ICA [26]	Light	17.65 ± 2.92	26.23 ± 4.93	-0.34 ± 0.16	-6.08 ± 0.32
	Medium	14.90 ± 1.87	20.04 ± 2.75	0.15 ± 0.15	-6.54 ± 0.57
	Dark	15.81 ± 2.82	22.24 ± 4.09	0.16 ± 0.25	-6.67 ± 0.33
POS [31]	Light	18.43 ± 2.83	27.33 ± 4.82	-0.44 ± 0.15	-5.82 ± 0.39
	Medium	13.00 ± 1.66	17.70 ± 2.54	0.17 ± 0.15	-6.42 ± 0.40
	Dark	20.22 ± 3.48	29.09 ± 5.31	-0.40 ± 0.23	-7.32 ± 0.46
CHROM [8]	Light	11.99 ± 1.60	17.08 ± 2.73	0.15 ± 0.17	-5.96 ± 0.43
	Medium	13.34 ± 1.42	17.49 ± 2.00	0.19 ± 0.15	-7.17 ± 0.55
	Dark	9.67 ± 1.98	13.84 ± 3.09	0.28 ± 0.24	-6.96 ± 0.66
GREEN [30]	Light	8.41 ± 1.59	11.28 ± 2.41	0.36 ± 0.16	-4.32 ± 0.35
	Medium	11.47 ± 1.59	14.35 ± 2.02	0.31 ± 0.14	-5.51 ± 0.41
	Dark	13.36 ± 2.19	18.14 ± 3.16	0.25 ± 0.24	-6.53 ± 0.24
LGI [25]	Light	12.66 ± 2.25	18.12 ± 3.62	0.09 ± 0.17	-6.12 ± 0.33
	Medium	15.15 ± 1.91	19.61 ± 2.47	0.14 ± 0.15	-7.29 ± 0.49
	Dark	10.13 ± 2.09	14.24 ± 3.10	0.34 ± 0.24	-7.17 ± 0.50
PBV [9]	Light	9.52 ± 1.48	12.78 ± 2.06	0.45 ± 0.15	-5.52 ± 0.37
	Medium	13.41 ± 1.76	17.37 ± 2.47	0.16 ± 0.15	-6.37 ± 0.32
	Dark	8.86 ± 1.73	12.04 ± 2.36	0.57 ± 0.21	-6.63 ± 0.34
OMIT [5]	Light	12.66 ± 2.25	18.12 ± 3.62	0.09 ± 0.17	-6.12 ± 0.33
	Medium	15.15 ± 1.91	19.61 ± 2.47	0.14 ± 0.15	-7.29 ± 0.49
	Dark	10.13 ± 2.09	14.24 ± 3.10	0.34 ± 0.24	-7.17 ± 0.50

performance when testing for participants with different skin tone in the TriCam-rPPG dataset. The performance of DeepPhys has been best out of all algorithms. Fig. A.8 provides a visual demonstration of the results highlighting skin tone bias.

UBFC-rPPG, and PURE has representation bias in the data as they mainly contain participants from Caucasian group with light skin tone. SCAMP has better diversity, but is synthetic. Hence, they may not reflect the Lambertian characteristics well. So, in order to see if there is more than dataset bias, we tested with MMPD dataset that represents mainly medium and dark skin participants. The results are shown in Table A.12. Algorithm like DeepPhys perform significantly worse on the dataset when compared to Table A.5. Figure 1 in the main paper shows some visual illustration of results from this section and Tables.

Table A.5: Results of running inference on Tricam-rPPG dataset using supervised methods in the rPPG Toolbox [16] (mean $\pm$ std). This is also reported in the Main Paper, Table 5.

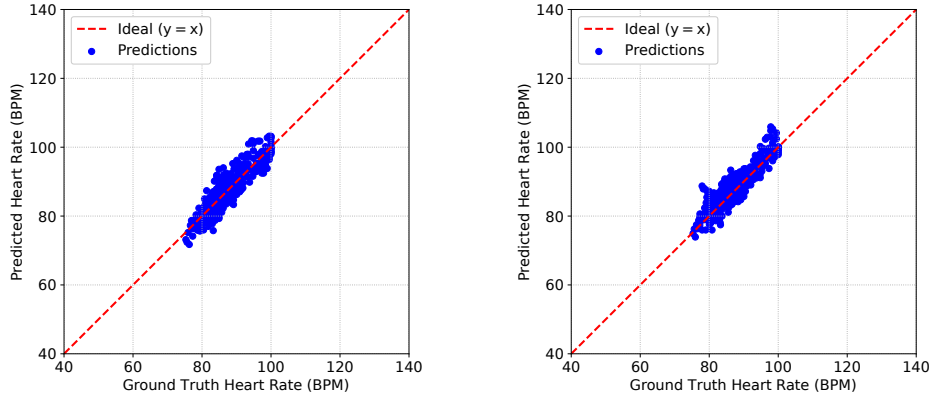
Test Set: Tricam-rPPG					
Method	Train Set	MAE $\downarrow$	MAPE $\downarrow$	ρ $\uparrow$	SNR $\uparrow$
TS-CAN [14]	UBFC-rPPG	6.16 ± 1.00	7.63 ± 1.15	0.62 ± 0.07	-2.33 ± 0.38
	PURE	16.74 ± 2.89	21.09 ± 3.67	0.33 ± 0.09	-3.62 ± 0.31
	SCAMPS	37.31 ± 4.37	49.07 ± 5.96	0.09 ± 0.10	-4.90 ± 0.22
PHYSNET [35]	UBFC-rPPG	19.09 ± 2.02	26.75 ± 3.01	-0.14 ± 0.09	-5.70 ± 0.17
	PURE	17.30 ± 1.76	23.76 ± 2.57	0.10 ± 0.09	-5.23 ± 0.16
	SCAMPS	16.38 ± 1.40	22.45 ± 2.05	0.10 ± 0.09	-6.82 ± 0.08
PHYSFORMER [36]	UBFC-rPPG	31.23 ± 2.77	43.88 ± 4.21	-0.29 ± 0.09	-5.46 ± 0.18
	PURE	12.81 ± 1.14	17.30 ± 1.69	0.23 ± 0.09	-5.01 ± 0.20
	SCAMPS	24.71 ± 2.01	34.03 ± 3.01	-0.04 ± 0.10	-6.54 ± 0.10
DEEPPHYS [29]	UBFC-rPPG	3.85 ± 0.56	4.85 ± 0.71	0.77 ± 0.06	-1.28 ± 0.39
	PURE	5.21 ± 0.67	6.61 ± 0.85	0.67 ± 0.07	-1.89 ± 0.40
	SCAMPS	6.90 ± 0.77	8.82 ± 1.03	0.51 ± 0.09	-2.95 ± 0.32
EFF.PHYS-C [15]	UBFC-rPPG	6.30 ± 1.24	8.20 ± 1.70	0.40 ± 0.09	-2.47 ± 0.34
	PURE	22.48 ± 3.46	29.26 ± 4.60	0.15 ± 0.09	-3.92 ± 0.27
	SCAMPS	69.13 ± 4.24	92.14 ± 6.18	-0.00 ± 0.10	-5.80 ± 0.15

A.8 Effectiveness of Using NIR Data

This section shows additional results to support Section 6.2 in the main paper. Tables A.13, A.14, and A.15 shows results from the custom algorithm that uses different combinations of NIR and RGB data. The results shows prominently that use of NIR significantly increases the performance of the subjects who has darker skin.

Table A.6: Results of running inference on other datasets using supervised models trained on Tricam-rPPG (mean $\pm$ std).

Train Set: Tricam-rPPG					
Method	Test Set	MAE $\downarrow$	MAPE $\downarrow$	ρ $\uparrow$	SNR $\uparrow$
TS-CAN [14]	UBFC-rPPG	25.78 ± 3.59	23.70 ± 3.04	0.11 ± 0.16	-8.29 ± 0.61
	PURE	14.93 ± 2.70	18.13 ± 2.47	0.12 ± 0.13	-4.93 ± 0.55
	UBFC-Phys	12.69 ± 1.17	15.46 ± 1.46	0.22 ± 0.10	-5.48 ± 0.23
PHYSNET [35]	UBFC-rPPG	31.43 ± 3.50	28.77 ± 2.94	-0.09 ± 0.16	-10.23 ± 0.82
	PURE	15.66 ± 2.71	19.89 ± 2.82	-0.00 ± 0.13	-5.43 ± 0.73
	UBFC-Phys	12.20 ± 1.06	14.18 ± 1.11	0.22 ± 0.10	-5.86 ± 0.24
PHYSFORMER [36]	UBFC-rPPG	38.53 ± 3.13	36.44 ± 2.56	-0.25 ± 0.15	-10.94 ± 0.49
	PURE	19.25 ± 2.40	26.25 ± 2.66	-0.00 ± 0.13	-6.79 ± 0.49
	UBFC-Phys	17.32 ± 1.15	19.93 ± 1.16	0.06 ± 0.10	-7.24 ± 0.20
DEEPHYS [6]	UBFC-rPPG	34.30 ± 3.12	32.15 ± 2.56	-0.05 ± 0.16	-8.75 ± 0.52
	PURE	17.01 ± 2.77	21.46 ± 2.64	0.12 ± 0.13	-5.90 ± 0.55
	UBFC-Phys	14.09 ± 1.25	17.05 ± 1.54	0.02 ± 0.10	-6.22 ± 0.22
EFF.PHYS-C [15]	UBFC-rPPG	28.71 ± 3.75	26.02 ± 3.19	-0.18 ± 0.16	-7.95 ± 0.57
	PURE	13.54 ± 2.57	16.04 ± 2.22	0.32 ± 0.13	-4.73 ± 0.56
	UBFC-Phys	11.13 ± 1.15	13.36 ± 1.30	0.31 ± 0.10	-4.95 ± 0.27

**Fig. A.5:** Test: Tricam-rPPG. Train: a) PURE b) UBFC-rPPG

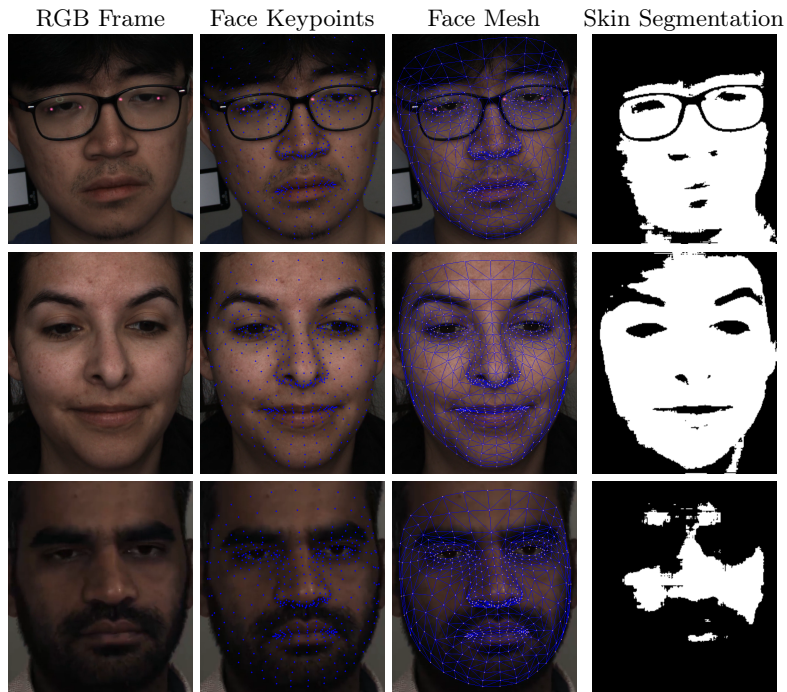


Fig. A.6: Example visualizations of facial representations derived from RGB frames in the dataset.

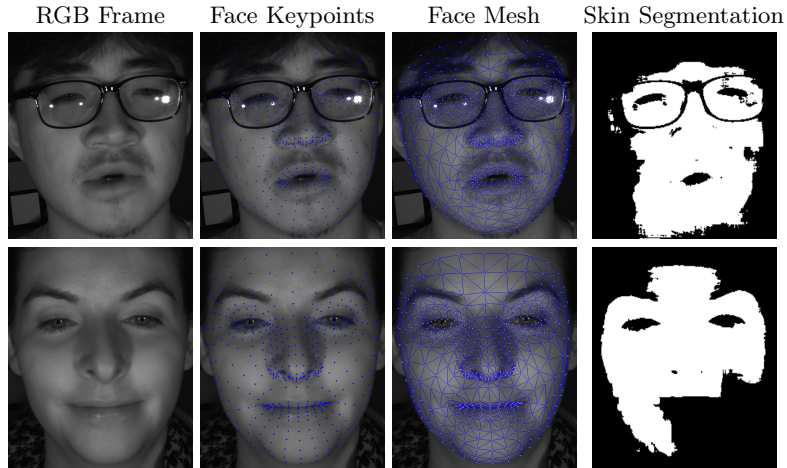


Fig. A.7: Example visualizations of facial representations derived from NIR (850nm) frames in the dataset.

Table A.7: Results across different skin tone ranges for DEEPHYS [6] (mean $\pm$ std).

Test Set: Tricam-rPPG					
Train Set	Skin Tone	MAE $\downarrow$	MAPE $\downarrow$	ρ $\uparrow$	SNR $\uparrow$
UBFC-rPPG [4]	Light	1.72 ± 0.51	2.21 ± 0.63	0.94 ± 0.06	0.62 ± 0.63
	Medium	4.42 ± 0.90	5.12 ± 0.99	0.78 ± 0.09	-1.39 ± 0.57
	Dark	6.60 ± 1.58	9.43 ± 2.40	0.45 ± 0.22	-4.78 ± 0.20
PURE [28]	Light	1.50 ± 0.47	1.98 ± 0.63	0.95 ± 0.05	0.15 ± 0.65
	Medium	4.92 ± 0.89	5.82 ± 1.03	0.76 ± 0.10	-1.95 ± 0.55
	Dark	13.43 ± 1.65	17.99 ± 2.15	-0.01 ± 0.25	-5.79 ± 0.26
SCAMPS [20]	Light	3.26 ± 0.83	4.12 ± 1.07	0.82 ± 0.10	-1.61 ± 0.51
	Medium	6.45 ± 0.95	7.67 ± 1.11	0.67 ± 0.11	-2.90 ± 0.49
	Dark	15.38 ± 2.06	21.27 ± 3.06	-0.42 ± 0.23	-5.73 ± 0.16

Table A.8: EfficientPhys performance across skin tones on the Tricam-rPPG test set.

Test Set: Tricam-rPPG					
Train Set	Skin Tone	MAE $\downarrow$	MAPE $\downarrow$	ρ $\uparrow$	SNR $\uparrow$
UBFC-rPPG [4]	Light	2.39 ± 0.77	3.07 ± 0.91	0.87 ± 0.08	-0.81 ± 0.53
	Medium	4.51 ± 0.82	5.27 ± 0.90	0.77 ± 0.09	-2.58 ± 0.49
	Dark	18.51 ± 5.55	25.79 ± 7.79	-0.18 ± 0.25	-5.71 ± 0.14
PURE [28]	Light	24.57 ± 6.55	32.68 ± 8.81	0.20 ± 0.17	-2.83 ± 0.42
	Medium	13.48 ± 3.28	15.69 ± 3.76	0.42 ± 0.13	-3.99 ± 0.42
	Dark	41.51 ± 9.83	57.53 ± 13.61	-0.10 ± 0.25	-6.08 ± 0.20
SCAMPS [20]	Light	74.24 ± 6.99	102.50 ± 10.68	-0.23 ± 0.17	-5.79 ± 0.23
	Medium	55.39 ± 6.26	70.14 ± 8.50	0.21 ± 0.14	-5.48 ± 0.23
	Dark	100.16 ± 2.35	137.34 ± 6.48	-0.23 ± 0.24	-6.76 ± 0.34

Table A.9: PhysFormer performance across skin tones on the Tricam-rPPG test set.

Test Set: Tricam-rPPG					
Train Set	Skin Tone	MAE $\downarrow$	MAPE $\downarrow$	ρ $\uparrow$	SNR $\uparrow$
UBFC-rPPG [4]	Light	29.62 ± 4.54	42.55 ± 6.93	-0.38 ± 0.16	-4.70 ± 0.37
	Medium	34.18 ± 4.32	47.13 ± 6.53	-0.37 ± 0.14	-5.67 ± 0.23
	Dark	26.64 ± 5.39	37.89 ± 8.42	-0.20 ± 0.24	-6.43 ± 0.29
PURE [28]	Light	12.48 ± 1.98	17.00 ± 2.90	0.37 ± 0.16	-4.17 ± 0.35
	Medium	11.89 ± 1.29	15.78 ± 1.94	0.19 ± 0.14	-5.25 ± 0.29
	Dark	15.96 ± 3.75	21.98 ± 5.47	0.14 ± 0.25	-6.09 ± 0.30
SCAMPS [20]	Light	25.67 ± 3.86	36.01 ± 5.89	0.03 ± 0.17	-6.43 ± 0.18
	Medium	24.01 ± 2.38	31.71 ± 3.34	0.06 ± 0.15	-6.61 ± 0.13
	Dark	24.68 ± 5.48	36.28 ± 8.49	-0.64 ± 0.19	-6.59 ± 0.29

Table A.10: PhysNet performance across skin tones on the Tricam-rPPG test set.

Test Set: Tricam-rPPG					
Train Set	Skin Tone	MAE ↓	MAPE ↓	ρ ↑	SNR ↑
UBFC-rPPG [4]	Light	19.14 ± 3.50	27.75 ± 5.39	-0.21 ± 0.17	-5.01 ± 0.31
	Medium	19.44 ± 2.90	26.31 ± 4.21	-0.12 ± 0.15	-5.89 ± 0.23
	Dark	18.06 ± 4.77	25.95 ± 7.04	-0.16 ± 0.25	-6.62 ± 0.33
PURE [28]	Light	20.35 ± 3.59	29.14 ± 5.39	-0.10 ± 0.17	-4.65 ± 0.21
	Medium	17.80 ± 2.36	23.30 ± 3.36	0.16 ± 0.15	-5.34 ± 0.27
	Dark	10.67 ± 2.04	15.14 ± 2.92	0.34 ± 0.23	-6.10 ± 0.30
SCAMPS [20]	Light	16.56 ± 2.73	23.50 ± 4.00	0.05 ± 0.17	-6.62 ± 0.11
	Medium	14.87 ± 1.60	19.71 ± 2.41	0.19 ± 0.14	-6.88 ± 0.12
	Dark	20.09 ± 3.77	27.65 ± 5.29	0.05 ± 0.25	-7.11 ± 0.24

Table A.11: TS-CAN performance across skin tones on the Tricam-rPPG test set.

Test Set: Tricam-rPPG					
Train Set	Skin Tone	MAE ↓	MAPE ↓	ρ ↑	SNR ↑
UBFC-rPPG [4]	Light	2.61 ± 0.89	3.34 ± 1.06	0.85 ± 0.09	-0.43 ± 0.61
	Medium	6.21 ± 1.73	7.09 ± 1.83	0.65 ± 0.11	-2.50 ± 0.55
	Dark	13.14 ± 1.91	17.70 ± 2.60	0.20 ± 0.24	-5.68 ± 0.20
PURE [28]	Light	16.03 ± 5.62	20.36 ± 7.14	0.40 ± 0.16	-2.09 ± 0.51
	Medium	14.08 ± 3.59	16.34 ± 4.08	0.50 ± 0.13	-3.91 ± 0.45
	Dark	25.27 ± 6.85	35.23 ± 9.84	-0.23 ± 0.24	-5.95 ± 0.24
SCAMPS [20]	Light	40.87 ± 8.51	56.51 ± 12.12	-0.04 ± 0.17	-4.23 ± 0.36
	Medium	33.20 ± 5.59	40.91 ± 7.07	0.27 ± 0.14	-4.84 ± 0.34
	Dark	41.16 ± 9.86	55.98 ± 13.41	-0.02 ± 0.25	-6.43 ± 0.22

Table A.12: Cross-dataset evaluation on the Tricam-rPPG using models trained on different subsets of the MMPD dataset. Performance is reported as mean $\pm$ standard deviation. Training is done on three subset of MMPD dataset with different combination of skin tones (34 $\rightarrow$ medium, 56 $\rightarrow$ Dark)

Method	Train Set	MAE ↓	MAPE ↓	ρ ↑	SNR ↑
TS-CAN [14]	MMPD-34	11.12 ± 2.18	14.35 ± 2.74	0.43 ± 0.18	-4.35 ± 0.37
	MMPD-56	7.66 ± 1.92	10.12 ± 2.11	0.48 ± 0.18	-4.87 ± 0.23
	MMPD-All	10.66 ± 2.04	13.82 ± 2.56	0.40 ± 0.17	-4.13 ± 0.32
DeepPhys [6]	MMPD-34	8.73 ± 1.71	11.22 ± 2.36	0.46 ± 0.17	-4.99 ± 0.26
	MMPD-56	7.94 ± 1.65	10.81 ± 2.14	0.49 ± 0.18	-5.08 ± 0.21
	MMPD-All	9.67 ± 1.88	12.93 ± 2.49	0.41 ± 0.18	-5.71 ± 0.29
EfficientPhys [15]	MMPD-34	10.59 ± 2.11	13.56 ± 2.80	0.38 ± 0.19	-6.33 ± 0.07
	MMPD-56	11.18 ± 2.24	14.08 ± 2.91	0.36 ± 0.19	-6.43 ± 0.12
	MMPD-All	8.13 ± 1.64	10.92 ± 2.05	0.47 ± 0.19	-6.33 ± 0.05

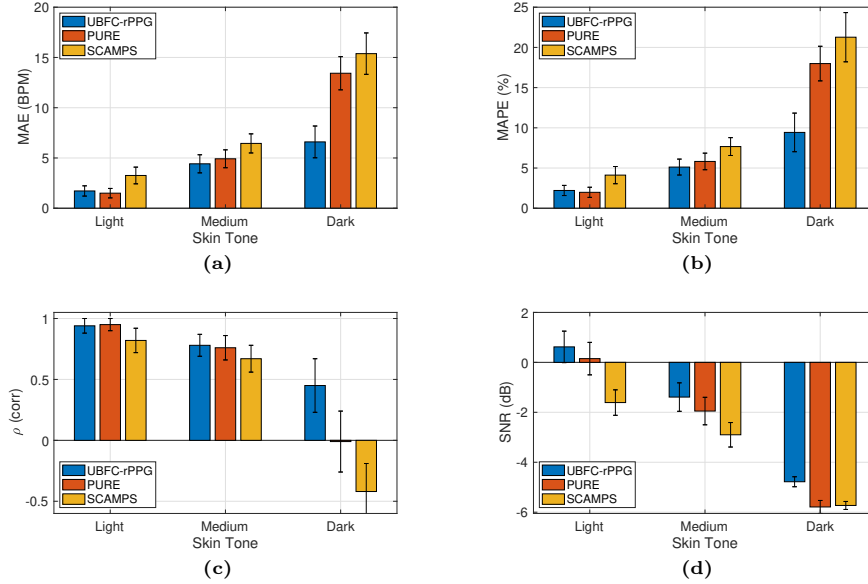


Fig. A.8: Qualitative comparison across skin tones and datasets while training using DEEPHYS. It is a visualization of Table A.7

Table A.13: Performance of LSTM models for Light skin tone subjects.

Metrics	RGB	RG850	RG940	RIR	GIR	BIR	RGBIR
$\rho \uparrow$	0.82	0.86	0.81	0.78	0.77	0.79	0.89
MAE (instant) $\downarrow$	7.37	5.85	6.56	7.47	8.53	7.41	6.07
RMSE (instant) $\downarrow$	11.96	9.28	10.42	13.36	14.95	11.45	10.20
MAE (full clips) $\downarrow$	4.49	3.70	3.77	5.14	5.79	4.82	3.48
RMSE (full clips) $\downarrow$	6.91	6.97	6.09	9.14	9.51	7.59	5.79
SNR (dB) $\uparrow$	4.85	7.27	5.44	5.58	4.36	3.91	7.30

Table A.14: Performance of LSTM models for Medium-Light skin tone subjects.

Metrics	RGB	RG850	RG940	RIR	GIR	BIR	RGBIR
$\rho \uparrow$	0.82	0.88	0.84	0.83	0.84	0.82	0.92
MAE (instant) $\downarrow$	12.60	8.30	11.15	12.30	9.56	11.88	7.02
RMSE (instant) $\downarrow$	29.83	14.69	25.20	30.13	16.07	18.10	10.84
MAE (full clips) $\downarrow$	8.52	4.66	7.00	7.05	5.84	8.05	4.17
RMSE (full clips) $\downarrow$	17.99	8.47	15.76	13.18	9.98	12.21	6.49
SNR (dB) $\uparrow$	2.95	5.20	5.88	4.46	2.75	-0.53	5.77

Table A.15: Performance of LSTM models for Medium-Dark skin tone subjects.

Metrics	RGB	RG850	RG940	RIR	GIR	BIR	RGBIR
$\rho \uparrow$	0.71	0.77	0.74	0.73	0.72	0.73	0.87
MAE (instant) $\downarrow$	12.50	9.42	10.97	11.69	12.67	11.89	8.96
RMSE (instant) $\downarrow$	17.12	12.87	16.69	17.88	21.73	17.24	12.29
MAE (full clips) $\downarrow$	7.26	5.04	6.21	7.19	7.82	7.18	5.60
RMSE (full clips) $\downarrow$	9.46	6.29	7.80	8.38	10.01	8.50	7.60
SNR (dB) $\uparrow$	-4.68	-1.59	-0.76	-2.83	-1.20	-2.72	0.06

References

1. Mediapipe face landmarker. Google AI Edge Documentation (2026), https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker, accessed: 2026-03-12
2. Bashkatov, A.N., Genina, E.A., Kochubey, V.I., Tuchin, V.: Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *Journal of Physics D: Applied Physics* **38**(15), 2543–2555 (2005)
3. Benezeth, Y., Krishnamoorthy, D., Monsalve, D.J.B., Nakamura, K., Gomez, R., Mitéran, J.: Video-based heart rate estimation from challenging scenarios using synthetic video generation. *Biomedical Signal Processing and Control* **96**, 106598 (2024)
4. Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., Dubois, J.: Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters* **124**, 82–90 (2019)
5. Casado, C.A., López, M.B.: Face2ppg: An unsupervised pipeline for blood volume pulse extraction from faces. *IEEE Journal of Biomedical and Health Informatics* **27**(11), 5530–5541 (2023)
6. Chen, W., McDuff, D.: DeepPhys: Video-based physiological measurement using convolutional attention networks. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 349–365 (2018)
7. Curran, T., McDuff, D., Liu, X., Narayanswamy, G., Ma, C., Patel, S., Yang, E.: Camera-based remote photoplethysmography for blood pressure measurement: current evidence, clinical perspectives, and future applications. *Connected Health And Telemedicine* **2**(2), N–A (2023)
8. De Haan, G., Jeanne, V.: Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering* **60**(10), 2878–2886 (2013)
9. De Haan, G., Van Leest, A.: Improved motion robustness of remote-PPG by using the blood volume pulse signature. *Physiological Measurement* **35**(9), 1913 (2014)
10. Delpy, D.T., Cope, M., van der Zee, P., Arridge, S., Wray, S., Wyatt, J.: Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology* **33**(12), 1433–1442 (1988). <https://doi.org/10.1088/0031-9155/33/12/008>
11. Farrell, T.J., Patterson, M.S., Wilson, B.C.: A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *Medical Physics* **19**(4), 879–888 (1992). <https://doi.org/10.1118/1.596777>
12. Jacques, S.L.: Optical properties of biological tissues: a review. *Physics in Medicine and Biology* **58**(11), R37–R61 (2013). <https://doi.org/10.1088/0031-9155/58/11/R37>
13. Lam, A., Kuno, Y.: Robust heart rate measurement from video using select random patches. In: *Proceedings of the IEEE international conference on computer vision*. pp. 3640–3648 (2015)
14. Liu, X., Fromm, J., Patel, S., McDuff, D.: Multi-task temporal shift attention networks for on-device contactless vitals measurement. *Advances in Neural Information Processing Systems* **33**, 19400–19411 (2020)
15. Liu, X., Hill, B., Jiang, Z., Patel, S., McDuff, D.: Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 5008–5017 (2023)

16. Liu, X., Narayanswamy, G., Paruchuri, A., Zhang, X., Tang, J., Zhang, Y., Sengupta, R., Patel, S., Wang, Y., McDuff, D.: rppg-toolbox: Deep remote ppg toolbox. *Advances in Neural Information Processing Systems* **36**, 68485–68510 (2023)
17. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.L., Yong, M.G., Lee, J., et al.: Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019)
18. Maki, Y., Monno, Y., Yoshizaki, K., Tanaka, M., Okutomi, M.: Inter-beat interval estimation from facial video based on reliability of bvp signals. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. pp. 6525–6528. IEEE (2019)
19. McDuff, D.: Camera measurement of physiological vital signs. *ACM Computing Surveys* **55**(9), 1–40 (2023)
20. McDuff, D., Wander, M., Liu, X., Hill, B., Hernandez, J., Lester, J., Baltrusaitis, T.: Scamps: Synthetics for camera measurement of physiological signals. *Advances in Neural Information Processing Systems* **35**, 3744–3757 (2022)
21. Melville, H.: *Moby-Dick; or, The Whale*. New York: Harper & Brothers (1851)
22. Niu, X., Han, H., Shan, S., Chen, X.: Vipl-hr: A multi-modal database for pulse estimation from less-constrained face video. In: *Asian conference on computer vision*. pp. 562–576. Springer (2018)
23. Nowara, E.M., Marks, T.K., Mansour, H., Veeraraghavan, A.: Near-infrared imaging photoplethysmography during driving. *IEEE transactions on intelligent transportation systems* **23**(4), 3589–3600 (2020)
24. Odinaev, I., Chin, J.W., Luo, K.H., Ke, Z., So, R.H., Wong, K.L.: Optimizing camera exposure control settings for remote vital sign measurements in low-light environments. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 6086–6093 (2023)
25. Pilz, C.S., Zauneder, S., Krajewski, J., Blazek, V.: Local group invariance for heart rate estimation from face videos in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. pp. 1254–1262 (2018)
26. Poh, M.Z., McDuff, D.J., Picard, R.W.: Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering* **58**(1), 7–11 (2010)
27. Speth, J., Vance, N., Czajka, A., Bowyer, K.W., Wright, D., Flynn, P.: Deception detection and remote physiological monitoring: A dataset and baseline experimental results. In: *International Joint Conference on Biometrics (IJCB)*. pp. 1–8. IEEE (2021)
28. Stricker, R., Müller, S., Gross, H.M.: Non-contact video-based pulse rate measurement on a mobile service robot. In: *The 23rd IEEE international symposium on robot and human interactive communication*. pp. 1056–1062. IEEE (2014)
29. Terbe, D.: *Camera-Based Pulse Monitoring Using Deep Learning Tools*. Ph.D. thesis, Pázmány Péter Catholic University (PPKE) (2020)
30. Verkruysse, W., Svaasand, L.O., Nelson, J.S.: Remote plethysmographic imaging using ambient light. *Optics express* **16**(26), 21434–21445 (2008)
31. Wang, W., Den Brinker, A.C., Stuijk, S., De Haan, G.: Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering* **64**(7), 1479–1491 (2016)
32. Wang, W., Leonhardt, S., Tarassenko, L., Shan, C., McDuff, D.: Guest editorial: Camera-based monitoring for pervasive healthcare informatics. *IEEE Journal of Biomedical and Health Informatics* **25**(5), 1358–1360 (2021)

33. Xu, H., Sarkar, A., Abbott, A.L.: Color invariant skin segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2906–2915 (2022)
34. Xu, S., Sun, L., Rohde, G.K.: Robust efficient estimation of heart rate pulse from video. *Biomedical optics express* **5**(4), 1124–1135 (2014)
35. Yu, Z., Li, X., Zhao, G.: Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. In: 30th British Machine Vision Conference: BMVC 2019. 9th-12th September 2019, Cardiff, UK. The British Machine Vision Conference (BMVC) (2019)
36. Yu, Z., Shen, Y., Shi, J., Zhao, H., Torr, P.H., Zhao, G.: Physformer: Facial video-based physiological measurement with temporal difference transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4186–4196 (2022)