Tricam-rPPG: A Multimodal Multispectral Dataset for Remote Photoplethysmography

Abhijit Sarkar[1], Surendrabikram Thapa[2,3], Ishtiaque Ahmed Khan[2,3], Yogesh Deshpande[2,3], and A. Lynn Abbott[3]

Virginia Tech, Blacksburg, VA USA

Abstract. This paper introduces Tricam-rPPG, a multimodal dataset designed to support systematic studies of remote photoplethysmography (rPPG) using multispectral imaging. Remote (noncontact) monitoring offers the potential for unobtrusive measurement of physiological signals related to health, cognitive load, and affect. However, most existing rPPG datasets rely primarily on RGB imaging, limiting the study of spectral effects, illumination variability, and sensing biases associated with differences in optical properties of the skin. Tricam-rPPG provides synchronized recordings of 31 human subjects from three co-located cameras: a standard RGB camera and two near-infrared (NIR) cameras operating at 850 nm and 940 nm, all captured under controlled illumination conditions. For fourteen participants, simultaneous recordings from Meta Aria glasses are also included. All cases are supplemented by reference waveforms of blood volume pulses (BVP) measured using a fingertip PPG sensor. In addition to presenting the dataset, we establish baseline benchmarks with widely used rPPG algorithms to evaluate heart-rate estimation performance across different combinations of spectral channels. By providing synchronized RGB and NIR video together with physiological ground truth, Tricam-rPPG enables new research directions in multispectral physiological sensing, fairness-aware rPPG algorithms, and robust remote cardiovascular monitoring. The dataset is available under a data use license and mutual agreement (check https://tricam-rppg.github.io).

Keywords: Cardiovascular Monitoring · Photoplethysmography · Remote PPG · Multispectral Imaging · Dataset

1 Introduction

Cardiovascular diseases (CVD) remain the dominant global health burden. The World Health Organization estimates that 19.8 million deaths from CVD occurred in 2022, accounting for roughly 32% of all deaths globally, with over three-quarters of CVD deaths occurring in low- and middle-income countries (LMICs) [46]. Early detection may have prevented 80% of those deaths. Along with physical health, global mental health situation including stress and depression, is declining post COVID [32]. Early signs of CVD and stress are reflected

in physiological signals. Most prominently, heart rate (HR) and heart-rate variability (HRV) are two key measures which act as "system indicators" of cardiometabolic health and autonomic stress. Yet, continuous monitoring at population scale remains structurally difficult; clinic-based measurement is episodic, wearable adoption is uneven, and many communities lack the infrastructure for sustained follow-up, especially across rural regions and LMIC contexts [46]. With increasing medical costs and shrinking medical facilities, new cost-effective cardiovascular monitoring methods are needed. This paper concerns advancements of one such method: remote photoplethysmography (rPPG) [1].

Photoplethysmography (PPG) refers to the use of illumination as a means of measuring volumetric change, and rPPG typically refers to noncontact, camera-based measurements associated with blood volume pulses (BVP). With every heartbeat, the cardiovascular system pumps blood throughout the body, producing subtle changes in skin color that are invisible to the naked eye. Over the past two decades, researchers have shown that these minute variations can be recovered using computational methods and commercially available cameras [7, 11, 19, 27, 33, 37, 45]. With widespread availability of cameras in mobile phones, laptops, and other consumer devices, there is increasing potential to democratize remote measurements for cardiovascular health monitoring.

Because rPPG enables contactless measurement of physiological signals using standard imaging devices, these methods have been explored in domains such as telemedicine, affective computing, driver monitoring, and human–computer interaction [8, 27, 33, 36]. In healthcare settings, rPPG has been investigated for unobtrusive monitoring of cardiovascular activity and respiratory dynamics, supporting remote patient monitoring, telehealth applications, and hospital patient monitoring [23, 27]. These approaches have also been applied to specialized clinical scenarios such as neonatal monitoring, and post surgery recovery where contactless sensing is desirable for fragile patients [35]. Looking beyond estimation of heart rate, camera-based photoplethysmography has been used to study spatial blood perfusion, oxygen saturation, respiration rate, blood pressure and hemodynamic changes across the skin [18, 30, 34, 38]. These diverse applications highlight the growing interest in camera-based physiological sensing as a scalable and non-invasive alternative to traditional contact sensors.

In spite of widespread interest, current systems are not yet capable of providing a scalable, equitable, and commercial solution that can monitor vital signs continuously and unobtrusively. In practice, rPPG systems often struggle under different real-world conditions [27, 45]: (i) non-uniform illumination, (ii) subject motion, (iii) camera auto-exposure/auto-gain and ISP nonlinearities, (iv) compression artifacts and sensor noise in consumer cameras, (v) skin-tone-dependent optical attenuation that reduces effective signal-to-noise ratio (SNR) in visible bands, (vi) partial occlusions from hair, facial accessories, or hand movements, (vii) geometric variations such as head pose and distance changes that alter pixelwise sampling of skin regions, and (viii) background interference and ambient lighting fluctuations that introduce temporally correlated noise.

[1] In some literature, rPPG is called imaging photoplethysmography (iPPG).

Different datasets have been proposed over the years to address these problems. However, we revisit rPPG-related developments from the past two decades to evaluate the feasibility of deployment at global scale, and therefore across a diverse population. Our assessment indicates that many recent methods and data collection systems lack key components necessary to ensure reliable performance across different skin tones and ambient illumination variations. One clear gap that emerges is extensive research in multi-spectral rPPG development, especially the use of near-infrared (NIR) wavelengths in conjunction with RGB channels. To address this limitation and to support further progress in the field, this paper introduces the Tricam-rPPG dataset, which we have developed with emphasis on advancing research in the field related to rPPG. The main contributions of this work are as follows:

- The Tricam-rPPG dataset is among the first to provide *multispectral video recordings* for rPPG research, combining standard RGB video with two NIR streams at 850 nm and 940 nm. Videos are provided for 31 subjects, and recordings from Meta Aria glasses are included for a subset of those participants (14 participants). Reference physiological signals are provided by a fingertip PPG sensor.
- We provide rich *metadata and preprocessing annotations* (including facial landmarks and skin-region information), together with baseline benchmark results to support reproducible evaluation of rPPG algorithms.
- We demonstrate the benefit of multispectral sensing for rPPG through *baseline experiments* showing that incorporating NIR channels improves SNR and reduces MAE, particularly as skin tone gets darker, highlighting the potential of longer wavelengths to mitigate sensing limitations of pipelines using visible wavelengths only.

2 Motivation for a Multispectral Dataset

2.1 Systematic Bias against Darker Skin Tones

We evaluated different pretrained rPPG models on a subset of Tricam-rPPG data with the widely used rPPG-Toolbox [22]. Fig. 1 shows that algorithms such as DeepPhys [7], TSCAN [15], and EfficientPhys [21] exhibit low MAE for lighter skin tones. However, they degrade monotonically as the skin tones become darker. These findings are consistent with those presented by Bondarenko et al. [5]. We further trained TSCAN on the MMPD [42] dataset, which contains data from darker skin tones. However, while even if the dataset contained no data from light skin, it performed best for light skin tones, and the performance for darker skin tones continued to degrade (See Table 1, and Table A.7 - A.9 in Supp material for more details). When performance gaps persist even after balancing training data across skin tones, this indicates that bias is not purely *data bias* (e.g., representation imbalance) but also *system bias* arising from the sensing and image-formation pipeline: illumination spectra, camera spectral sensitivity, exposure/gain control, tone mapping, compression, and noise floors/ceilings.

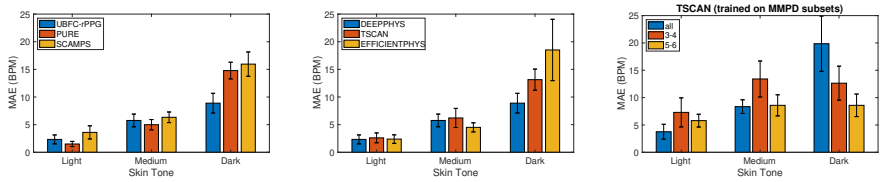

Fig. 1: Systematic bias and representative bias in rPPG methods. (a) Performance by DeepPhys [7] across different skin tones in Tricam-rPPG, when the model was trained separately on 3 datasets that have participants with lighter skin tones. (b) Three different deep-learning algorithms show similar degradation characteristics for darker skin tones when trained with UBFC-rPPG and tested on Tricam-rPPG. (c) Performance of TSCAN [15] degrades even when the model is trained using subsets of MMPD [42], a dataset that includes participants with darker skin tones.

Table 1: Results across different skin tone ranges for DEEPPHYS [7] (mean ± std).

Test Set: Tricam-rPPG					
Train Set	Skin Tone	MAE ↓	MAPE ↓	ρ ↑	SNR ↑
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 [41]	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 [?]	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

2.2 Photometric Characteristics of NIR for rPPG

This paper mainly concerns the effect of multi-spectral imaging, especially use of NIR. NIR sensing shows promise because melanin absorption decreases substantially with increasing wavelength into the NIR, potentially reducing pigmentation-induced variability in reflectance and improving SNR parity across skin tones [14]. 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 [31].

Empirically, melanin absorption in the NIR (e.g., 850–940 nm) is substantially lower than in the green-visible range (500–600 nm) [2, 14]. Consequently, the sensitivity of tissue absorption to melanin concentration diminishes as wavelength increases into the NIR. These shows that most rPPG methods might be biased by sensor design and not only by limitations of data diversity limitations. (See supplemental material for more details.)

2.3 Scope and Challenges in Using NIR for rPPG

NIR sensing has been explored in remote measurement settings [44], including in driving environments and other low-light scenarios [25, 29]. Prior studies have demonstrated that physiological signals are observable in the NIR spectrum and that NIR-based measurements can exhibit stronger rPPG responses under certain conditions compared to visible-light imaging. However, the optical characteristics of NIR introduce additional challenges. NIR light penetrates deeper into biological tissue, which can lead to increased scattering and diffusion of the reflected signal. While deeper penetration may capture information from richer vascular structures, it may also introduce spatial mixing that can degrade signal quality [1]. As a result, an important question remains insufficiently explored: how these diffusion effects influence different stages of rPPG analysis. In particular, it is unclear whether deeper tissue diffusion primarily affects pulse peak detection algorithms or whether it also alters waveform morphology used by shape-based estimation methods.

From a deployment perspective, NIR sensing opens additional opportunities for rPPG in real-world environments. In low-illumination scenarios, RGB-based approaches often suffer from severe signal degradation due to insufficient visible light. In contrast, NIR imaging can operate with dedicated infrared illumination, enabling physiological signal recovery even under dim or nighttime conditions. This capability has already motivated the adoption of NIR cameras in applications such as driver monitoring systems, wearable devices and other human-sensing platforms operating in uncontrolled environments [10, 29].

Furthermore, modern indoor lighting systems are designed to emit energy primarily in the visible spectrum [16] for efficiency. As a result, visible-light rPPG measurements can be strongly influenced by changes in ambient illumination, including flicker and intensity variations. NIR imaging, particularly when combined with controlled illumination source, can reduce this dependency on environmental lighting and provide a stable sensing modality. These properties make NIR an attractive complementary sensing channel for robust remote physiological measurement.

Despite these advantages, datasets that simultaneously capture synchronized RGB and NIR signals under controlled experimental conditions remain extremely limited. To the best of our knowledge, there is currently no publicly available dataset designed to systematically study rPPG performance across both RGB and NIR modalities within a structured experimental protocol.

3 Background: Existing Datasets

The past 15 years have shown considerable interest by the research community in producing video datasets of human faces. This section describes publicly available datasets that have been used to support research involving rPPG. All of these provide video recordings of human faces, typically using RGB cameras in indoor environments. To provide ground-truth references, most of these datasets

include recordings of BVP signals, PPG, ECG, and HR. Tables 2 and 3 provide a summary of datasets in chronological order.

MAHNOB-HCI [39] was the first relevant dataset to gain popularity in rPPG research, although its primary purpose was to support studies of emotion. Along with RGB recordings of the face, physiological reference data was recorded (ECG, GSR, respiration, and skin temperature) using sensors mounted on the body, including a cap containing EEG electrodes. The DEAP dataset [17] was also concerned with emotion analysis. DEAP similarly provides EEG signals along with videos of faces. While wearing a cap with EEG electrodes, each participant was recorded while watching music videos on a monitor. It is interesting to note that the music videos served as secondary illumination sources that subtly affected the appearance of the participants' faces. Further interest in analyzing emotions and facial expressions led to the development of the large multimodal dataset BP4D+ [50]. In addition to an RGB camera and a thermal camera (spectral range 7.5-14.0 μm), data was recorded using a stereo-based depth sensor. MMSE-HR [43] also contains videos showing elicited emotion.

PURE [41] was the first publicly available dataset to be created specifically for the purpose of investigating camera-based rPPG. Whereas the participants for MAHNOB-HCI and DEAP remained stationary, the participants in PURE exhibit head motion and facial movements induced by talking. COHFACE [13] is another early dataset that was created for investigating rPPG. It consists of recordings of 40 participants, with RGB video obtained using a commodity webcam (Logitech C525) under two lighting conditions. VIPL-HR [28] is a relatively large dataset for rPPG that was the first intended to support deep-learning techniques. Recordings are provided from several RGB video cameras, an RGB-D sensor, and a fingertip device.

TokyoTech [26] is a dataset that emphasized the utility of a prototype camera that simultaneously captured the standard RGB color bands along with an NIR band. A high frame rate (300 fps) was used to capture recordings of 9 participants. This work was among the first to emphasize the use of video data for estimating inter-beat intervals (IBI) and HRV. Nowara et al. introduced the MR-NIRP dataset [29], which was the first publicly available dataset containing both RGB and NIR video of faces for rPPG purposes. The videos were obtained of passengers in automobiles, for both parked and motion situations. Another dataset that provided both RGB and NIR video is DDPM [40], which was developed to aid in the study of deception based on facial appearance. This dataset provides recordings of 70 subjects in several modalities, including RGB and NIR video, both at the high frame rate of 90 fps. DDPM also provides long-wave IR video at lower resolution and and slower frame rate. Benezeth et al. describe the creation of a *synthetic* video NIR dataset of 10 participants called IMVIA-NIR [3].

UBFC-rPPG [4] and, later, UBFC-Phys [36] provide RGB videos of 43 and 140 participants, respectively. The former emphasized cognitive load, as the participants concentrated on a time-sensitive mathematical game, whereas the latter was intended to analyze the effect of social stress and anxiety. The MMPD [42]

Table 2: Comparison of existing rPPG datasets and the proposed Tricam-rPPG dataset in terms of acquisition properties and task diversity. ("IMU" refers to inertial measure unit, as provided on Aria glasses. "N.A.": Not Available.)

Dataset	Resolution	FPS	NIR	Multi-Spectral	Ambient Light	Tasks	IMU
MAHNOB-HCI [39]	780×580	60	–	–	1	2	–
DEAP [17]	720×576	50	–	–	1	40	–
PURE [41]	640×480	30	–	–	1	2	–
BP4D+ [50]	1040×1392	25	–	✓	1	10	–
MMSE-HR [43]	1040×1392	25	–	✓	1	Multiple	–
COHFACE [13]	640×480	20	–	–	2	2	–
VIPL-HR [28]	1920×1080	30	✓	–	3	4	–
TokyoTech [26]	640×480	300	✓	✓	1	Multiple	–
UBFC-rPPG [4]	640×480	30	–	–	2	2	–
MR-NIRP [29]	640×640	30	✓	✓	Multiple	2	–
UBFC-Phys [36]	1024×1024	35	–	–	1	3	–
DDPM [40]	1920×1080	90	✓	✓	1	Multiple	–
MMPD [42]	320×240	30	–	–	3 or 4	Multiple	–
IMVIA-NIR [3]	1280×1024	20	✓	–	1	1	–
egoPPG-DB [6]	320×240	30	✓	–	Multiple	Multiple	✓
Tricam-rPPG (Ours)	2046×2464	60	✓	✓	2	8	✓

dataset contains videos obtained from mobile phones. There is a growing level of interest monitoring regions of the face near the eyes using eyeglasses. The egoPPG-DB dataset [6] provides video from Aria glasses. These NIR videos provide close-up views of the eyes, and are intended for tracking of gaze direction.

4 Tricam-rPPG Dataset Collection

4.1 Instrumentation and Design

We designed a controlled multimodal acquisition platform to collect synchronized RGB, NIR, and physiological reference signals. The experimental setup (see supp material) consists of three co-located cameras mounted on rigid supports to ensure consistent viewpoint and minimal relative motion across modalities. An RGB camera (2046×2464 resolution at 60 fps) captures high-resolution color videos of the face and hand. Two monochrome NIR cameras operating at 850 nm and 940 nm (2046×2464 resolution at 60 fps) simultaneously record facial and hand videos under active NIR illumination. All cameras had fixed gain and intrinsic settings to reduce any unwanted variabilities. Physiological ground truth is obtained using a BIOPAC MP46 system sampling finger photoplethysmography (PPG) at 125 Hz. 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. To ensure photometric consistency

Table 3: Demographic and recording diversity across rPPG datasets. For body parts: FF → Full Face, PF → Profile Face, P → Palm, PO → Periocular.

Dataset	SkinTone Diversity	Age Range	Body Parts	Gaze Point	No. of Participants	Total Time(min.)
MAHNOB-HCI [39]	Medium	19-40	FF	✓	27	264
DEAP [17]	Low	19-37	FF	–	22	880
PURE [41]	Low	N.A.	FF	–	10	60
BP4D+ [50]	High	18-66	FF	–	140	N.A.
MMSE-HR [43]	High	N.A.	FF	–	40	4500
COHFACE [13]	Low	N.A.	FF	–	40	160
VIPL-HR [28]	Low	22-41	FF	–	107	1150[1]
TokyoTech [26]	Low	20s–60s	FF	–	9	27
UBFC-rPPG [4]	Low	N.A.	FF	–	43	70
MR-NIRP [29]	High	20–60	FF	–	18	115
UBFC-Phys [36]	Low	19–38	FF	–	140	170
DDPM [40]	Low	N.A.	FF	–	70	776
MMPD [42]	High	N.A.	FF	–	33	660
IMVIA-NIR [3]	Low	N.A.	FF	–	10	140
egoPPG-DB [6]	Medium	19-32	PO	✓	25	780
Tricam-rPPG (Ours)	High	21–66	FF, PF, P, PO	✓	31 (14)	1130

[1]This duration is for RGB videos only. For NIR, the total time is 380 min.

and enable post-hoc color calibration across sessions, a standardized color calibration palette is included within the RGB camera's field of view during acquisition. All streams were temporally aligned to support frame-accurate multimodal analysis. All data streams were synchronized post-hoc using UTC time stamps. This configuration enables high-fidelity multimodal analysis under controlled yet realistic recording conditions.

4.2 Experimental Protocol

The dataset is organized into two structured recording sessions conducted under two illumination conditions: a cool setting (8000 K) and a warm setting (2700 K). Across both sessions, participants perform eight tasks designed to elicit diverse physiological responses spanning resting, cognitive, emotional, and pose-dependent conditions. Session S1 includes a resting baseline with visual fixation, cognitive load induction through mental arithmetic, low-arousal video viewing, sadness-inducing affective stimulation, and a side-view posture to enable profile-based facial analysis. S1 also includes a resting condition with a raised hand to capture pulse signals from multiple anatomical regions, enabling investigation of spatial pulse propagation and potential pulse transit time estimation. Session S2 complements these tasks with a reading activity, excitement-inducing affective stimulation. Together, these sessions capture physiological responses across baseline, affective, cognitive, and non-frontal configurations.

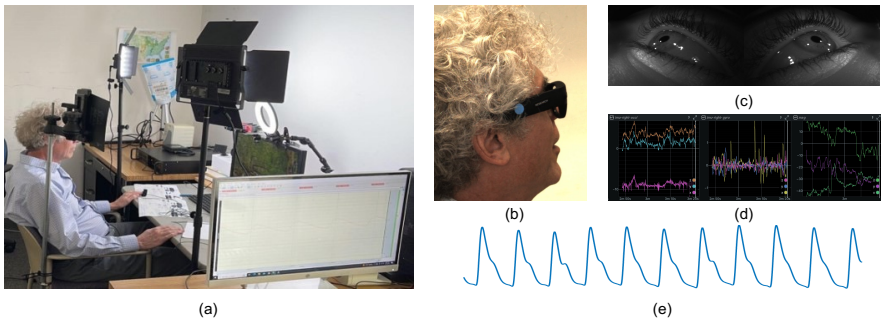

Fig. 2: (a) Overview of recording studio. (b) Volunteer wearing Meta Aria glasses. (c) Periocular video from Aria glasses. (d) Accelerometer and gyro data. (e) Example PPG signal from fingertip sensor.

Each recording segment lasts at least five minutes to enable reliable estimation of HRV. Prior work highlights that shorter durations may fail to capture slower physiological oscillations associated with autonomic regulation [12]. Table 4 summarizes the full protocol. Fig. 2 and Fig. 3 show representative samples from the dataset.

Table 4: Overview of sessions and tasks in the Tricam-rPPG dataset. Session 1 (S1) indicates tasks that were conducted using "cool" lighting conditions. Session 2 (S2) indicates tasks that were conducted using "warm" illumination settings.

Session	Task	Description
S1	T1	Resting baseline; static stare at central cross (+) without moving.
S1	T2	Mental arithmetic; cognitive load induction.
S1	T3	Video viewing (calm state) to capture low-arousal responses.
S1	T4	Video viewing to elicit altered emotional valence and arousal (sadness).
S1	T5	Static side-view posture to capture profile.
S1	T6	Resting baseline; participant held a raised hand at the side of the face.
S2	T1	Reading task; participant read portion of Chapter 5 of *Moby Dick*.
S2	T2	Video viewing to elicit emotional valence and arousal (excitement).

4.3 IRB Approval and Participant Recruitment

We submitted the protocol to our Institutional Review Board (IRB) and received approval before collection of the data. Advertisements were published for participant recruitment. The participants were first screened by a phone call. The participants were compensated for their time. The participants also signed an agreement to allow release of this dataset to other researchers. The dataset contains a total of 31 participants. Participants were asked to self-report their skin

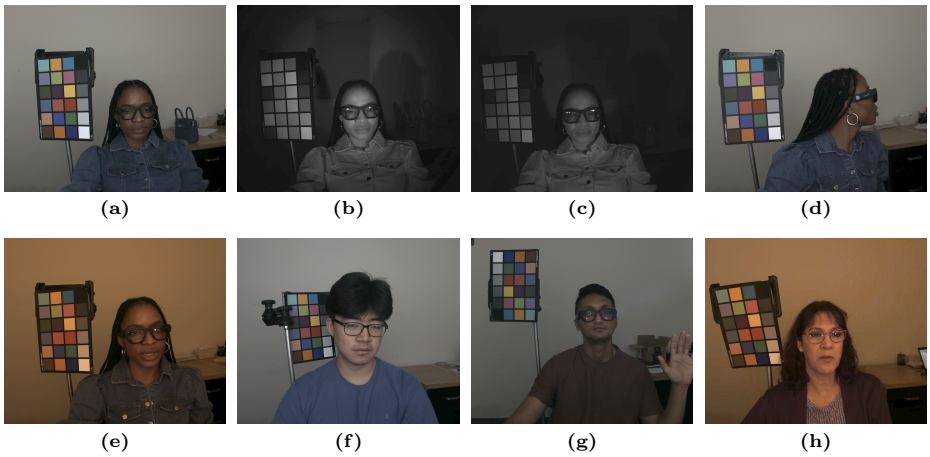

Fig. 3: Example frames from Tricam-rPPG dataset. (a) RGB. (b) NIR 850 nm. (c) NIR 940 nm. (d) Profile view. (e-f) Warm and cool illumination settings, respectively. (g) Showing palm (session S1, task T6). (h) Reading task (session S2, task T1).

tones according to a I - VI scale that was inspired by an NHS Fitz- patrick Skin-Type Chart. The dataset includes three skin-tone groups: Light (equivalent to Fitzpatrick scale I-II), Medium (equivalent to Fitzpatrick scale III-IV), and Dark (equivalent to Fitzpatrick scale V-VI). The Light group contains 7 participants, with 5 males and 2 females. The Medium group includes 10 participants, with 6 males and 4 females. The Dark group includes 14 participants, with 12 males and 2 females. Overall, the Table 5 shows that the dataset contains participants from a range of skin tones, although the gender distribution is male-dominant, particularly in the Dark skin-tone group. Notably, the dataset contains stronger representation in the dark skin tone group, supporting evaluation.

Table 5: Participant distribution across skin tone and gender for Tricam-rPPG.

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

5 Tricam-rPPG Dataset Characteristics

An rPPG dataset will benefit from the following properties (1) synchronized RGB–NIR recordings, (2) high-fidelity physiological ground truth, (3) viewpoint and body-region diversity, (4) demographic diversity, (5) head motion reference,

(6) photometric reference, and (7) HRV variations. Tricam-rPPG is designed to address these needs.

RGB+NIR (850/940 nm): In addition to high-resolution RGB, we provide two co-registered NIR streams to enable cross-spectral analyses and to investigate whether longer wavelengths reduce melanin-related attenuation and improve SNR parity across skin tones.

Viewpoint and body region: Beyond frontal face videos, the dataset includes lateral facial views and right-hand (palm) recordings, supporting pose-robust rPPG, cross-region pulse recovery, and geometry-aware physiological modeling.

Demographic coverage: Participants span light/medium/dark self-reported skin tones, enabling stratified evaluation and fairness analysis across spectra and tasks. (Skin-tone designations are self-reported by the participants.)

Controlled photometric variation: Recordings are collected under two illumination settings (cool and warm). Additionally, video stimuli introduce structured, time-varying illumination; the source videos are shared to allow careful analysis of illumination changes.

Physiological variability across tasks: Multiple sessions include baseline, cognitive stimulation (e.g., mental math), and affective video viewing to induce variability in HR/HRV and support task-level benchmarking.

Reference physiology: We provide synchronized fingertip PPG ground truth (BIOPAC) for physically grounded evaluation of HR and waveform fidelity.

Color calibration: Each recording includes a ColorChecker chart to support photometric calibration, illumination estimation, and analysis of ISP/spectral bias that can confound rPPG.

Aria subset (motion-aware sensing): A subset of participants wear Meta Aria glasses providing periocular NIR video and synchronized IMU, enabling quantitative studies of motion-induced degradation and motion compensation.

Collectively, Tricam-rPPG supports benchmarking beyond intensity-based frontal analysis toward multi-view, cross-spectral, motion-aware, and waveform-consistent physiological reconstruction.

5.1 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 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 [24]; (6) **Skin Maps**: binary masks indicating skin pixels for each frame [47]; (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.

6 Results and Baseline

6.1 Benchmark with Standard Methods

To facilitate reproducible evaluation and enable meaningful comparison across future methods, we establish a comprehensive benchmark using several widely adopted rPPG algorithms implemented in the rPPG Toolbox [22]. Specifically, we evaluate representative classical and learning-based approaches under a unified pipeline and report performance using the standard metrics commonly adopted in the rPPG literature, including mean absolute error (MAE) in BPM, mean absolute percentage error (MAPE) in BPM, correlation coefficient (ρ), and signal-to-noise ratio (SNR) in dB. These metrics allow us to characterize both HR estimation accuracy and signal quality in a consistent manner across methods. All experiments follow the default preprocessing and evaluation protocols provided by the toolbox to ensure reproducible by the community.

Beyond reporting aggregated performance, we provide a detailed analysis to highlight the strengths and limitations of current rPPG methods under different conditions. First, we report overall benchmark performance across the full dataset to establish reference baselines. Second, recognizing the growing concern regarding fairness and sensing bias in camera-based physiological measurement, we further report results stratified by self- reported skin tone groups. Third, to evaluate cross-dataset generalization, we conduct experiments where models are trained on the proposed Tricam-rPPG dataset and tested on several widely used public datasets, as well as the reverse setting where models trained on external datasets are evaluated on Tricam-rPPG. Finally, we present task-level granular analyses to provide deeper insight into how existing methods perform under different experimental conditions. Together, these evaluations establish a transparent and extensible benchmark that enables future work to systematically compare algorithms and better understand the challenges of robust remote physiological sensing. Due to space constraint only Table 6 is shown in the main document. Other benchmarks are reported in the supplementary document.

6.2 NIR Effectiveness

In order to test the effectiveness of the NIR on darker skin tone, we tested model inspired by Li et al [19]. It uses temporal data from face patches and train a Demucs model [9]. We adopted the structure to train different combination of five available channels from the dataset: red (R), green (G), blue (B), NIR (850 nm), and NIR (940 nm). We trained a total of five three channel combinations R+IR, G+IR, B+IR, RG+850, and RG+940 (Fig. 4b). We specifically prioritize G channel as research shows higher response of BVP in G. *IR* denotes that both 850 and 940 channels were included. Additionally, we train a model that can take all five channels RGBIR (Fig. 4a).

Table 6: HR benchmark performance on Tricam-rPPG dataset using supervised methods in the rPPG Toolbox [22] (mean ± standard deviation).

Test Set: Tricam-rPPG					
Method	Train Set	MAE ↓	MAPE ↓	ρ ↑	SNR ↑
TS-CAN [20]	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 [48]	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 [49]	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 [7]	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 [21]	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

Data from eight subjects were tested for verification. The inclusion of NIR channels leads to clear improvements in both signal quality and HR estimation accuracy, with the gains becoming more pronounced as skin tone increases. As shown in Fig. 5, the effect of NIR channels on lighter skin tones (equivalent to Fitzpatrick scale II) is relatively modest, while the improvements become substantially larger for light–medium (equivalent to Fitzpatrick scale III) and medium–dark ((equivalent to Fitzpatrick scale IV) skin tones. For example, the RG850 configuration increases SNR by 50% for light skin, but the improvement grows to 80% and 70% for light–medium and medium–dark skin tones, respectively. A similar pattern is observed for RG940, where the SNR gain is limited to 10% for light skin but increases dramatically to 100% and 85% for light–medium and medium–dark skin tones. The multispectral RGBIR configuration further reinforces this trend, increasing SNR by 50%, 95%, and 100% for the three groups, respectively. These results indicate that the primary benefit of introducing NIR sensing emerges for higher skin pigmentation levels, while the performance for lighter skin tones remains largely stable.

A comparable pattern appears in the HR estimation error shown in Fig.5b. While the inclusion of NIR channels produces moderate reductions in MAE for lighter skin tones, the reduction becomes substantially larger for light–medium and medium–dark skin tones. For instance, the RG850 configuration reduces MAE by 20% for light skin, while the reduction reaches 35% and 25% for

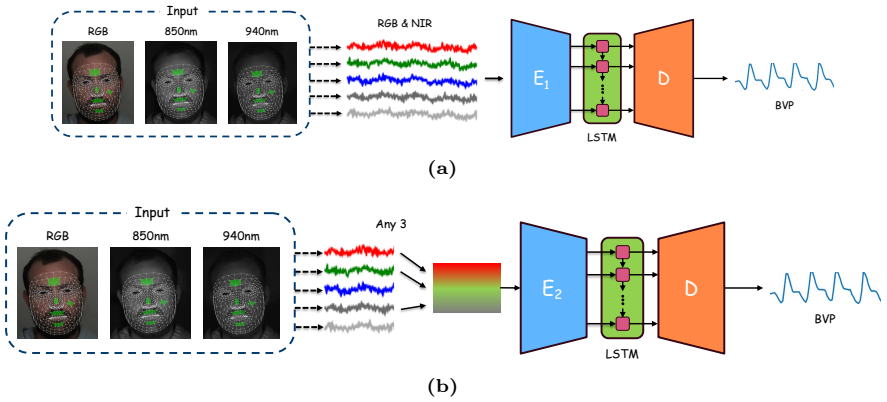

Fig. 4: Encoder-decorder model to test different combinations of R, G, B, IR(850nm), and IR(940nm) channels for rPPG information retrieval. Two models were developed: (a) takes all five channels as input simultaneously; and (b) takes any three channels.

light–medium and medium–dark tones while keeping an MAE value of 4.66 BPM and 5.04 BPM respectively. The RGBIR configuration yields the largest improvement, reducing MAE by 15% for light skin but by 45% and 30% for light–medium and medium–dark skin tones, respectively while the absolute value remains moderately low at 4.17 BPM and 5.60 BPM respectively. Importantly, these improvements are achieved without degrading performance for lighter skin tones, indicating that the addition of NIR channels primarily addresses limitations that arise under higher melanin concentrations.

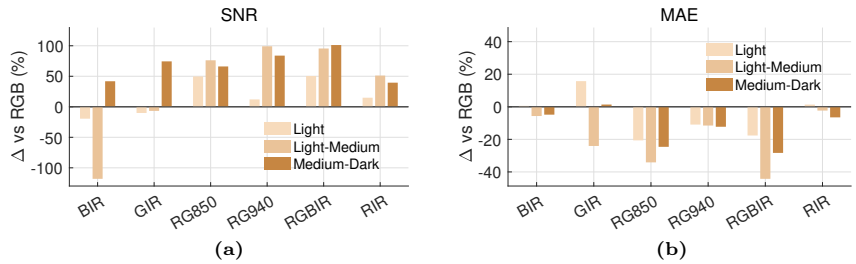

Fig. 5: Change in SNR↑ and MAE↓ when NIR data is introduced, relative to RGB-only data. Percentage changes are shown for participants who self-reported their skin tones as light, light-medium, and medium-dark. (a) SNR↑. (b) MAE↓.

.

7 Conclusion

This paper has introduced Tricam-rPPG, a multimodal dataset designed to support systematic studies of remote photoplethysmography under controlled yet realistic sensing conditions. The dataset provides synchronized recordings from RGB and two NIR cameras (850 nm and 940 nm) together with reference physiological signals, enabling analysis of both HR estimation and BVP waveform recovery tasks. In addition to presenting the dataset, we established a reproducible benchmark using widely adopted rPPG algorithms implemented in the rPPG Toolbox [22] and evaluated them using standard metrics across multiple experimental settings. Tricam-rPPG provides a valuable resource for the research community. This dataset will enable reproducible evaluation of existing methods while opening new directions for research in multimodal physiological sensing. In particular, the dataset supports future work on robust rPPG algorithms, fairness-aware physiological sensing, multimodal fusion across visible and NIR channels, and improved recovery of physiological waveforms beyond HR estimation. By making the dataset and evaluation protocols publicly available, we hope to encourage further progress toward reliable and equitable remote cardiovascular monitoring technologies. The team will host the data on their own internal server, and will distribute it to interested researchers. The data will be distributed after the interested party signs a data user license in order to honor IRB agreements.

8 Data Access

The dataset will be made available to researchers for non-commercial research purposes. Interested researchers must submit a data access request and execute a Data Use License Agreement before obtaining access to the dataset. Additional institutional or project-specific agreements may be required on a case-by-case basis. Further details regarding the access procedure are provided on the project website: https://tricam-rppg.github.io.

9 Acknowledgments

The team would like to thank Fulan Li, Zeeshan Karamat, Gayatri Bhatambarekar, Jin Woo Baik, Joe Bekiranov, Ayush Sadekar, Calvin Winkowski, Ryan Mowry, and Ryan Talbot for their support during the data collection effort. This work was partially funded by National Science Foundation (NSF) award 2136915, and by Virginia's Commonwealth Cyber Initiative (CCI) award HS-4Q24-060.

References

1. Allen, J.: Photoplethysmography and its application in clinical physiological measurement. Physiological measurement **28**(3), R1–R39 (2007)
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. Bondarenko, M., Menon, C., Elgendi, M.: Demographic bias in public remote photoplethysmography datasets. NPJ Digital Medicine **8**(1), 593 (2025)
6. Braun, B., Armani, R., Meier, M., Moebus, M., Holz, C.: egoppg: Heart rate estimation from eye-tracking cameras in egocentric systems to benefit downstream vision tasks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5579–5590 (2025)
7. 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)
8. Cheng, C.H., Wong, K.L., Chin, J.W., Chan, T.T., So, R.H.: Deep learning methods for remote heart rate measurement: A review and future research agenda. sensors **21**(18), 6296 (2021)
9. Défossez, A., Usunier, N., Bottou, L., Bach, F.: Demucs: Deep extractor for music sources with extra unlabeled data remixed. arXiv preprint arXiv:1909.01174 (2019)
10. Engel, J., Somasundaram, K., Goesele, M., Sun, A., Gamino, A., Turner, A., Talattof, A., Yuan, A., Souti, B., Meredith, B., et al.: Project aria: A new tool for egocentric multi-modal ai research. arXiv preprint arXiv:2308.13561 (2023)
11. de Haan, G., Jeanne, V.: Robust pulse rate from chrominance-based rppg. IEEE Transactions on Biomedical Engineering **60**(10), 2878–2886 (2013). https://doi.org/10.1109/TBME.2013.2266196
12. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on intelligent transportation systems **6**(2), 156–166 (2005)
13. Heusch, G., Anjos, A., Marcel, S.: A reproducible study on remote heart rate measurement. arXiv preprint arXiv:1709.00962 (2017)
14. 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
15. Kim, D.Y., Goh, E., Lee, K., Chae, J., Mun, J., Na, J., Sohn, C.b., Kim, D.Y.: Remote bio-sensing: Open source benchmark framework for fair evaluation of rppg. arXiv preprint arXiv:2307.12644 (2023)
16. Kim, S., Jahandar, M., Jeong, J.H., Lim, D.C.: Recent progress in solar cell technology for low-light indoor applications. Curr. Altern. Energy **3**(1), 3–17 (2019)
17. Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: DEAP: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing **3**(1), 18–31 (2011)

18. Lai, H., Huang, J., Tian, X., Yu, D., Ye, Y., Zeng, Y., Huang, Y., Wang, G., Pan, L., Sun, J., et al.: A ubiquitous platform for camera-based multi-parameter vital signs monitoring in hospital icus: A double-center clinical study. IEEE Journal of Biomedical and Health Informatics **29**(12), 8671–8678 (2025)
19. Li, F., Thapa, S., Bhat, S., Sarkar, A., Abbott, A.L.: A temporal encoder-decoder approach to extracting blood volume pulse signal morphology from face videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 5965–5974 (June 2023)
20. 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)
21. 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)
22. 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)
23. Liu, Z., Huang, B., Lin, C.L., Wu, C.L., Zhao, C., Chao, W.C., Wu, Y.C., Zheng, Y., Wang, Z.: Contactless respiratory rate monitoring for icu patients based on unsupervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6005–6014 (2023)
24. 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)
25. Magdalena Nowara, E., Marks, T.K., Mansour, H., Veeraraghavan, A.: Sparseppg: Towards driver monitoring using camera-based vital signs estimation in near-infrared. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 1272–1281 (2018)
26. 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)
27. McDuff, D.: Camera measurement of physiological vital signs. ACM Computing Surveys **55**(9), 1–40 (2023)
28. 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)
29. 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)
30. Nwibor, C., Haxha, S., Ali, M.M., Sakel, M., Haxha, A.R., Saunders, K., Nabakooza, S.: Remote health monitoring system for the estimation of blood pressure, heart rate, and blood oxygen saturation level. IEEE Sensors Journal **23**(5), 5401–5411 (2023)
31. Odinaev, I., Chin, J.W., Luo, K.H., Ke, Z., So, R.H.Y., Wong, K.L.: Optimizing camera exposure control settings for remote vital sign measurements in low-light environments. In: CVPR Workshops (CVPM) (2023)
32. Organization, W.H.: World mental health report: Transforming mental health for all. World Health Organization (2022)

33. Poh, M.Z., McDuff, D.J., Picard, R.W.: Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Transactions on Biomedicaal Engineering **58**(1), 7–11 (2010)
34. Rubins, U., Miscuks, A., Qawqzeh, Y., Marcinkevics, Z., Grabovskis, A.: Photoplethysmography imaging algorithm for real-time monitoring of skin perfusion maps. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5950–5956 (2023)
35. Ruhrberg Estévez, S., Grafton, A., Thomson, L., Warnecke, J., Beardsall, K., Lasenby, J.: Continuous non-contact vital sign monitoring of neonates in intensive care units using rgb-d cameras. Scientific Reports **15**(1), 16863 (2025)
36. Sabour, R.M., Benezeth, Y., De Oliveira, P., Chappe, J., Yang, F.: Ubfc-phys: A multimodal database for psychophysiological studies of social stress. IEEE Transactions on Affective Computing **14**(1), 622–636 (2021)
37. Sarkar, A.: Cardiac signals: remote measurement and applications. Ph.D. thesis, Virginia Tech (2017)
38. Schrumpf, F., Frenzel, P., Aust, C., Osterhoff, G., Fuchs, M.: Assessment of deep learning based blood pressure prediction from ppg and rppg signals. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3820–3830 (2021)
39. Soleymani, M., Lichtenauer, J., Pun, T., Pantic, M.: A multimodal database for affect recognition and implicit tagging. IEEE transactions on affective computing **3**(1), 42–55 (2011)
40. 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)
41. 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)
42. Tang, J., Chen, K., Wang, Y., Shi, Y., Patel, S., McDuff, D., Liu, X.: Mmpd: multi-domain mobile video physiology dataset. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 1–5. IEEE (2023)
43. Tulyakov, S., Alameda-Pineda, X., Ricci, E., Yin, L., Cohn, J.F., Sebe, N.: Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2396–2404 (2016)
44. Van Gastel, M., Stuijk, S., de Haan, G.: Motion robust remote-ppg in infrared. IEEE Transactions on Biomedical Engineering **62**(5), 1425–1433 (2015)
45. 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)
46. World Health Organization: Cardiovascular diseases (cvds): Fact sheet. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (2025), accessed 2026-02-28
47. 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)
48. Yu, Z., Li, X., Zhao, G.: Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. In: 30th British Machine Visison Con-

ference: BMVC 2019. 9th-12th September 2019, Cardiff, UK. The British Machine Vision Conference (BMVC) (2019)
49. 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)
50. Zhang, Z., Girard, J.M., Wu, Y., Zhang, X., Liu, P., Ciftci, U., Canavan, S., Reale, M., Horowitz, A., Yang, H., et al.: Multimodal spontaneous emotion corpus for human behavior analysis. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3438–3446 (2016)