Applied Computing and Informatics

Representation biases in remote photoplethysmography datasets – a narrative review

Shehroz S. Khan and Nour Mostafa
College of Engineering and Technology, American University of the Middle East, Egaila, Kuwait

Received 23 June 2025
Revised 24 August 2025
3 November 2025
Accepted 20 November 2025

Abstract

Purpose – Cardiovascular disease is a global public health challenge, with its rising prevalence severely impacting quality of life and leading to chronic conditions and fatalities. Monitoring heart rate (HR) is a crucial indicator of cardiac health. Remote photoplethysmography (rPPG) and HR monitoring through videos have shown promise for remote healthcare applications. Various deep learning approaches have reported state-of-the-art results on different datasets. While these results are encouraging, their applicability to real-world deployment is limited due to the presence of different types of biases present in these datasets.

Design/methodology/approach – We conducted a thorough search on various scholarly databases (e.g., Google Scholar and IEEE Xplore) with keywords "rPPG" or "remote photoplethysmography", datasets "heart rate" and similar variants. This led us to identify 32 public and private commonly used rPPG datasets.

Findings – By doing an extensive literature review on 32 commonly used rPPG datasets, we showed that most of the available datasets do not consider important demographic factors along the dimensions of age, sex, ethnicity and including patients during data collection. The most underrepresented category was patients, as only one dataset included them in the study. This could be related to the ease of data collection; however, it could lead to models that may not be generalizable to other groups, demographics or subsets of populations. Overall, our findings reveal that representation bias is present across various rPPG datasets.

Research limitations/implications – Our findings are limited to these 32 rPPG datasets, and further investigation of more datasets could reveal insights into the representation bias problem.

Practical implications – Biased rPPG datasets could lead to biased deep learning models that may not generalize in real-world conditions, leading to misdiagnosis and under- or over-estimation of HR estimation.

Social implications – Our in-depth analysis provides key insights to researchers to collect inclusive and diverse datasets and develop algorithms to mitigate biases in estimating HR and rPPG among different demographics.

Originality/value – This is one of the first literature reviews that delves deeper into common rPPG datasets and evaluates them for the presence of representation bias along the dimensions of age, sex, ethnicity and including patients.

Keywords Heart rate, Digital health, Remote monitoring, Remote photoplethysmography, Representation bias

Paper type Literature review

1. Introduction

Cardiovascular diseases (CVD) are the leading cause of death globally, taking approximately 18 million lives every year [1]. CVD comprises a variety of disorders of the heart and blood vessels [2]. Older adults are at a high risk of CVD [3]. Increased heart rate (HR) is identified as one of the major risk factors for all-cause mortality and cardiovascular events, with a stronger relationship in men than women [4]. It has been reported that an increase in HR of 10 beats per minute is associated with a 20% increase in the risk of cardiac-related deaths [4]. Continuous monitoring of HR could provide early warning for CVD risk, better management of cardiovascular conditions and prevention of potential deaths [5].

Applied Computing and Informatics
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1108/ACI-06-2025-0270

The gold standard to measure HR is an electrocardiogram (ECG) that corresponds to the electrical activity of the heart [6]. Traditionally, ECG is measured using an electrocardiograph machine with a single or multiple leads attached to the body. Due to the hardware requirement and operation by a trained technician, these devices cannot be used for continuous monitoring in a community setting. Modern-day smartwatches also provide the capability to capture ECG; however, their results remain inconclusive [7]. Moreover, wearable devices are often marred by other issues, such as frequent device recharge, ruggedness (e.g. water resistance), data accuracy, interoperability and user experience [8].

Photoplethysmogram (PPG) is a widely used technique to measure HR. It involves a light source and a photo detector to measure volumetric changes in blood vessels beneath the skin. When the light source illuminates the tissue, it detects subtle variations in the intensity of light either reflected or transmitted due to blood flow, producing the PPG signal [9] (see Figure 1). This process follows the Beer–Lambert law [10], which states that the amount of light absorbed by blood is directly proportional to the depth of light penetration in the skin and the concentration of hemoglobin in the blood [11]. During the cardiac cycle, small changes in hemoglobin concentration cause fluctuations in light absorption by blood vessels, leading to changes in skin intensity values. While PPG methods can be less accurate than ECG methods for HR measurement, they are more user-friendly and cost-effective, making them suitable for long-term use by the general population [12]. Many modern-day smartwatches provide the PPG capabilities for measuring and continuously monitoring HR [13]. However, these devices may not be appropriate for the neonatal and older population and those with delicate skin or existing skin infections [14]. Prolonged use of these devices may cause discomfort and increase the risk of skin infections and burns [15, 16]. As a result, remote photoplethysmography (rPPG) [17] methods have emerged as a promising alternative, offering a non-invasive, contactless and scalable solution for HR monitoring. The rPPG is a non-contact technique that uses a standard camera to detect subtle changes in skin color caused by blood flow variations to measure HR. By analyzing the reflected light from the skin's

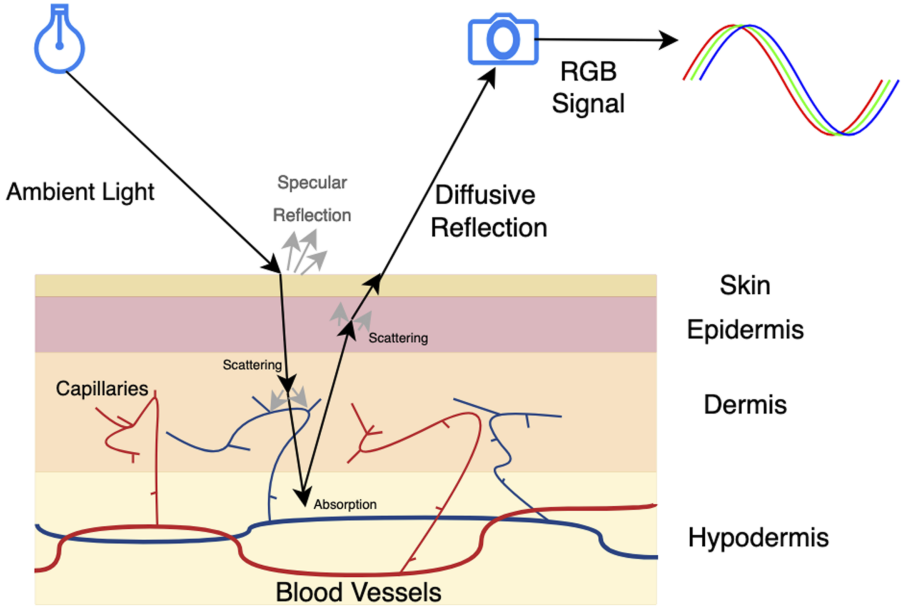

Figure 1. Principle of remote PPG based on the dichromatic reflection model

surface in a video, rPPG algorithms can estimate HR without requiring any physical contact. This method relies on the principle that heartbeat pulsations cause changes in skin blood perfusion, which alters the reflected light [17]. These approaches usually employ a digital camera to record a person's video, then apply face detection algorithms to localize the face, followed by defining a region of interest (e.g. cheek and forehead) and extracting rPPG from it [17]. HR is then computed through a series of post-processing techniques, including frequency analysis and peak detection [18].

In the past, both conventional computer vision and deep learning models have been proposed for HR estimation using rPPG signals extracted from videos of human participants [17]. While there has been significant progress in remote HR monitoring using rPPG, the specific needs of patients with heart conditions have not been thoroughly addressed in existing studies [19]. Researchers have collected and released several rPPG datasets for advancing research in this area, but they may not be representative of different demographics, such as age, sex and ethnicities. They may not contain data from cardiac patients, who may be in need for these solutions. It is not clear if models trained on a specific demographic may generalize to others. Thus, we take the position of the presence of representation bias in existing rPPG datasets. These factors could limit the value of rPPG methods to be translated into practice and improve clinical care and patient outcomes. As a result, we conducted a thorough review and reported 32 public and private rPPG datasets. These datasets clearly show the presence of representation biases along the dimensions of age, sex, ethnicity and including patients in the studies. These findings pave the way to collect diverse and inclusive datasets to benefit larger sections of society. It also points towards the development of data augmentation and synthetic data generation strategies and algorithms that could mitigate the impact of representation bias in these datasets. Our key contributions are as follows:

(1) We performed one of the first literature reviews to highlight the presence of representation bias in commonly used rPPG datasets;

(2) We identified representation bias in rPPG datasets along the dimensions of age, sex and ethnicity and not including patients, and

(3) We presented four levels of bias mitigation strategies in rPPG datasets including community, dataset, algorithm and evaluation levels.

2. Representation bias

Representation bias [20] in predictive modeling occurs when the models are trained on data that are not representative of the entire population, i.e. it may underrepresent certain parts (or demographics) of the population. Representation bias could stem from historical discrimination to selection and sampling biases in the data acquisition and preparation methods. It may also be related to the access and ease of collecting data. In the context of rPPG datasets, it is easier to collect data from young and healthy adults in comparison to patients and older populations, who may be at higher risk of CVD. Representation bias often results in skewed data distributions or omission of certain demographics' data altogether. This could lead to difficulties in modeling tail-end or unseen samples, such as in predicting abnormal HR situations among CVD patients if the model is trained only on healthy young adults with normal heart rates.

While abnormal HR may not occur frequently, absence or poor representation of those data samples could lead to unfavorable outcomes. As an example, if an rPPG dataset contains videos of the majority of young and healthy adults in a resting condition, most of their HR values may be almost similar. In these cases, the model may either learn to predict an average HR value or may not perform well when the HR values go abnormally high or low on patient data with CVD or arrhythmia conditions because the model may not have seen those variations

in the past. At best, such a model could predict the presence of "anomalous" HR [21] but may still not give an accurate prediction of HR.

Dasari *et al.* [22] mention the dominance of Caucasians (or White) and laboratory-collected rPPG datasets and that the performance of state-of-the-art deep learning models is unknown in other demographics. They collect inclusive data and evaluate state-of-the-art rPPG algorithms on participants from different demographic groups (India and Sierra Leone) in their natural environmental settings and identify the estimation bias in the measured signals. They also provide a list of rPPG datasets with and without contact-based sensors (wearable and video) with a focus on demographic diversity and laboratory environment. They did not consider patient population in their review. Our literature review is only focused on non-contact video-based rPPG datasets for HR estimation. We identified 32 rPPG datasets along the dimensions of age, sex, ethnicity and patient population. This focused research inquiry led to deeper insights into the systemic issue of representation in rPPG datasets and opened opportunities for their mitigation to develop diverse and inclusive datasets and fairer algorithms for HR estimation.

3. Remote rPPG datasets

With the advances in deep learning and interest in rPPG approaches, various researchers have collected different rPPG datasets. We conducted a thorough search on various scholarly databases (e.g., Google Scholar and IEEE Xplore) with keywords "rPPG" or "remote photoplethysmography", datasets "heart rate" and similar variants. This led us to identify 32 public and private commonly used rPPG datasets that are summarized in Table 1. This may not be a comprehensive list of all the available rPPG datasets. Xiao *et al.* [17] presented a similar (and shorter) list of rPPG datasets, where they described various attributes of these datasets alongside a discussion on the performance of conventional computer vision and supervised and unsupervised deep learning approaches on these datasets with respect to performance metrics, such as mean absolute error and root-mean-square error. Their review of datasets and predictive algorithms set the benchmark and guidelines for advancing research in this area. However, the focus of our review is not the evaluation of machine learning or deep learning algorithms or other signal processing approaches to extract rPPG or predict and/or detect HR. Interested readers are advised to refer to corresponding literature for an in-depth understanding of these topics [17–19].

We shift our focus from comparing traditional performance metrics and predictive modeling approaches to highlight the representation bias that these rPPG datasets contain. The focus of our investigation is to identify the presence of representation bias in rPPG datasets. The deductions challenge the generalizability of many state-of-the-art rPPG-based HR detection approaches across different demographics, thus diminishing their utility in real-world situations. We now discuss the presence of representation bias in existing rPPG datasets along the dimensions of age, sex, ethnicity and including patients.

3.1 Age bias

Out of the 32 datasets, 6 datasets did not provide any age information (PURE, PFF, CameraHRV, UBFC, MPSC and IMVIA-NIR), 14 datasets have data collected from younger or middle-aged adults (considering 60 years as a cut-off for older adults as per United Nations [23]) – MAHNOB-HCI, Deap-Part 2, AFRL, ECG-Fitness, LGI-PPGI-DB, OSF, VIPL-HR, Toadstool, StableSet, Moli-PPG-1, Moli-PPG-2, UBFC-Phys, iBVP and NRC-UBC. Eight papers reported data collected from both young and older individuals (MMSE-HR, OBF, MR-NIRP, BP4D+, TokyoTech-NIR, COHFACE, NRC-UBC and V4V). From the datasets with available age statistics, it is evident that older adults are generally underrepresented. For example, the V4V dataset includes only 1 older adult among its 179 participants, and the

Table 1. List of 32 rPPG datasets, - represents missing information on that field in the paper

#	Dataset	Year	Participants	Age	Sex	Ethnicity	Public
1	MAHNOB-HCI [24]	2011	27 (540 videos)	19–40 years, (26.06 ± 4.39)	11M, 16F	Different cultural backgrounds	Yes
2	Deap - Part 2 [25]	2012	22 (32 in original)	19–37 years (26.9)	16M, 16F	–	Yes
3	PURE [26]	2014	10 (60 Videos, 6 conditions)	–	8M, 2F	–	Yes
4	AFRL [27]	2014	25 (300 Videos)	18–28 years, (23.7)	17M, 8F	–	No
5	BP4D + [28, 29] (Extended from BP4D)	2014	140	18–66 years	58M, 82F	Black, White, Asian, Hispanic and others	Yes
6	MMSE-HR [29] (subset of BP4D+)	2016	40 (102 videos)	18–66 years in original	17M, 23F	Black, White, Asian, Hispanic and others	Yes
7	COHFACE [30]	2016	40 (160 Videos)	15–65 years (35.6 ± 11.5)	28M, 12F	–	Yes
8	PFF [31]	2017	13 (85 Videos)	–	–	–	Yes
9	CameraHRV [32]	2018	16	–	9M, 7F	–	Yes
10	OBF [33]	2018	106 (2,120 Videos) 6 patients	Healthy: 18–68 years (31.8) Patient: 43–81 years	61%M 39%F	White, Asian and Others	Yes
11	ECG-Fitness [34]	2018	17 (204 videos)	20–53 years	14M, 3F	–	Yes
12	LGI-PPGI-DB [35]	2018	25	25–42 years	20M, 5F	Majority White	Yes
13	OSF [36]	2018	7	23–37 years	6M, 1F	–	Yes
14	UBFC [37]	2019	-(43 Videos)	–	–	–	Yes
15	TokyoTech-NIR [38]	2019	9	20–60 years	8M, 1F	–	Yes
16	VIPL-HR [39]	2019	107 (2,378 Videos)	22–41 years	79M, 28F	Asians	Yes
17	VIPL-HR-V2 [40]	2020	500 (2,500 Videos) 2*nd* version of VIPL-HR	- (35.4 ± 18.0)	49%M 51%F	Asians	Yes
18	MR-NIRP [41]	2020	18 (180 Videos)	25–60 years	16M, 2F (5 M beard)	White, Indian and Asian	Yes
19	Toadstool [42]	2020	10 (10 Videos)	26–48 years	5M, 5F	–	Yes
20	StableSet [43]	2020	24 (24 Videos)	18–30 years (21.6)	9M, 15F	–	No
21	ViscarPPG [43]	2020	10 (20 Videos)	(29 ± 5)	7M, 3F	Skin type ranging on a Fitzpatrick scale from II to IV	Yes
22	EatingSet rPPG [43]	2020	20	(32 ± 8.6)	14M, 6F	Skin type on the Fitzpatrick scale from II to IV	No
23	MoLi-PPG-1 [44]	2020	30	18–35 years	–	–	Yes
24	MoLi-PPG-2 [44]	2020	15	18–35 years	–	–	Yes
25	UBFC-Phys [45]	2021	56 (12 eliminated)	19–38 years (21.8 ± 3.11)	10M, 46F	–	Yes

(*continued*)

Table 1. Continued

#	Dataset	Year	Participants	Age	Sex	Ethnicity	Public
26	V4V [46]	2021	179 (1,358 Videos)	18–66 years	101M, 78F	African American, White, Asian, Hispanic and Native American	Yes
27	MPSC [47]	2021	8	–	2M, 6F	Heterogeneous skin colors	Yes
28	CMU [22]	2021	140	(41.9)	(≈46% F)	40 Indians, 96 from Sierra Leone	Yes
29	BH [48]	2022	35	- (24 ± 2.31)	16M, 19F	–	Yes
30	iBVP [49]	2024	33	18–45 years, (27 ± 5.8)	10M, 23F	Asian, African, White, mixed and others	Yes
31	IMVIA-NIR [50]	2024	10	–	–	Varying ethnicities	Yes
32	NRC-UBC [51]	2024	126	Young (<40), middle age 40–59), older (>59)	69 F, 56M, 1 Other	Dark, medium and light skin tones	No

Note(s): M and F stands for male and female participants. Age is represented as range with (mean ± standard-deviation) under the range if given

COHFACE dataset features just 2 older adults out of 40 participants who are above the age of 60 years.

3.2 Sex bias

Five datasets did not report the sex of study participants (PFF, UBFC, MoLi-PPG-1, MoLi-PPG2 and IMVIA-NIR); three datasets only reported the percentage (OBF, VIPL-HR-v2 and CMU). One dataset, Deap-Part2, mentioned the sex-ratio of the full dataset including the non-rPPG dataset, while others mentioned the count of male and female participants. NRC-UBC reported an "other" gender. Only three datasets have either balanced or almost balanced distribution of sexes (Deap-Part 2 (16M, 16F), Imaging Photoplethysmography (8M, 6F) and Toadstool (5M, 5F)), while others have either proportionally large number of male participants (PURE (8M, 2F), ViscarPPG (7M, 3F), COHFACE (28M, 12F), ECG-Fitness (14M, 3F), MR-NIRP (16M, 2F), AFRL (17M, 8F), EatingSet rPPG (14M, 6F), and V4V (101M, 78F) or female participants (MAHNOB-HCI (11M, 16F), MMSE-HR (58M, 82F), StableSet (9M, 15F), NRC-UBC (69F, 56M, 1 other) and BP4D+ (58M, 82F)).

3.3 Ethnicity bias

In total, 20 datasets did not report the ethnicity distribution of their study participants (MAHNOB-HCI, Deap-Part 2, PURE, AFRL, COHFACE, PFF, Cam-eraHRV, ECG-Fitness, LGI-PPGI-DB, OSF, UBFC, TokyoTech-NIR, Toadstool, StableSet, MoLi-PPG-1, MoLi-PPG-2, UBFC-Phys, MPSC, BH and IMVIA-NIR), 3 datasets only mention the skin types and not the ethnicity (EatingSet rPPG, Vis-carPPG and NRC-UBC) and 9 datasets report participants' ethnicity (BP4D+, MMSE-HR, OBF, VIPL-HR, VIPL-HR-V2, MR-NIRP, V4V, CMU and iBVP). Of these, the VIPL-HR and VIPL-HR-V2 datasets were collected exclusively from Asian participants, while the CMU dataset was collected from Indian participants and people from Sierra Leone, while other datasets had a mix of Black, White, Asian, Hispanic, Indian, Native Americans and others. The remaining six datasets show the

ethnicity distribution (i.e. the percentage of individuals from different ethnic backgrounds) in their data (BP4D+, MMSE-HR, OBF, MR-NIRP, V4V and iBVP).

3.4 Patient bias

In our review, only the OBF dataset contains data from both healthy and CVD patients (age range from 43–81 years). OBF includes videos from 100 healthy participants and 6 patients with atrial fibrillation (A-Fib), all of whom were White. This finding that only one dataset contains patients' videos is striking with severe implications. The clinical utility of HR estimation models from rPPG datasets is to support cardiac patients who need this technology the most. However, in the absence of datasets that contain diverse videos from patients, it is very difficult to leverage the real utility of the advancement of deep learning methods to support this demographic. The dearth of patients' rPPG video data means there are limited data available to train and evaluate reliable and robust HR estimation models. Thus, the real-world performance of these models largely remains unknown. This trend is alarming and shows the difficulty in collecting data from (diverse) patients, which could be related to tedious ethical approvals, lack of access to patient population or the awareness of being inclusive of patients in these studies.

3.5 Summary

Figure 2 shows the representation bias heatmap of the 32 rPPG datasets presented in Table 1. The age bias is present in 18 datasets (56.25%), and age is not reported in 6 datasets (18.75%). The sex bias is present in 23 (71.88%) datasets, and sex is not reported in 5 datasets (15.63%). The ethnicity bias is present in 3 datasets (9.38%), and it is not reported in 16 datasets (50%).

Figure 2. Representation bias heatmap of rPPG datasets

The patients' information is only present in one data set. These findings show that all the reviewed rPPG datasets contain one or more types of representation biases. In many rPPG datasets, the age, sex and ethnicity information is not reported, with half of the studies not reporting ethnicity information.

4. Analysis and discussion

The above results show the presence of representation biases in rPPG dataset. Now, we discuss the implications of these biases.

Age bias. While the older and younger heart function the same, the HR is slightly lower in older adults [52, 53]. Therefore, if a model is trained solely on younger adults and tested on older adults, it will most likely overestimate the predicted HR because it may never have seen lower HR values during training. Therefore, this model will most likely report increased HR for older adults, which is indicative of a CVD, whereas they may be healthy. Collecting data from younger adults with lower HR than normal may be difficult to obtain and may not be a good representation of CVD patients whose HR may fluctuate abruptly. Conversely, a model trained on older adults with lower HR may underestimate the HR of younger people in the test set, indicating they may have CVD disease. Both the above decisions would be wrong and could lead to wrong diagnoses: primary care overheads and mental stress for the people. One may argue that if a dataset contains healthy adults doing exercises, then it shall record high HR values that might work to predict patients with CVD. However, studies show that both lower and higher HRs ($\leq$and $\geq$64) in CVD patients lead to all-cause mortality [54]. Moreover, during exercise, older adults' HR does not increase much compared to the younger adults [55]. Therefore, with similar reasoning as above, younger people's exercise data may not be directly applicable to the older population, and using younger adults' resting HR data to predict older adults' exercise HR will also be wrong because of other human physiological factors involved (e.g. increased breathing rate and other physiological indicators).

Sex bias. The average adult male HR (70 – 72 bpm) is slower than female HR (78 – 82 bpm) [56]. This difference is because typically the female heart's size is smaller than men's, which leads to pumping less blood with each beat and needing to beat at a faster rate to match the output of the large male heart. Female hearts also have a different intrinsic rhythmicity to the pacemaker of their hearts that leads to faster HR [57]. Therefore, models trained on one sex and tested on the other may either over- or underestimate HR due to different base line HR and could lead to erroneous conclusions. If the dataset of either sex is underrepresented during training the models, it may still lead to similar predictive issues.

Ethnicity bias. CVD and their risks are not evenly distributed across ethnicities; they may be related to many factors, including genetics [58], lifestyle [59] and health disparities [60]. In a UK biobank observational study, people of South Asian descent were found to be at more risk of CVD compared to European or Black individuals [61]. In a Canadian study, the prevalence of CVD among South Asians was found to be higher than in the Chinese population, while the Black population had a relatively low prevalence of CVD despite the least favorable overall cardiovascular risk factor profile [62]. In the US, it was reported that Mexican Americans and non-Hispanic Blacks had elevated CVD risk factors compared to non-Hispanic Whites and non-Hispanic Asians [63]. Another study on CVD projection in the US, based on the 2020 census, states that the prevalence of CVD risk factors is projected to decrease in the White population, whereas it continues to significantly increase in racial and other minorities [64]. All of these studies point to the fact that awareness of ethnic differences and CVD risks is paramount to developing ethnicity-specific cardiovascular prevention programs. From an artificial intelligence modeling perspective, it is imperative that the researchers pay attention to collecting datasets that represent a diverse cohort to build ethnicity-specific models to support their future health in a better way. Models trained and tested on one group of population may serve that population's challenges, but those models' generalization abilities across ethnicities remain unknown.

Patient bias. The patients' group is severely underrepresented in the rPPG datasets reviewed in this paper. While training and testing HR measurement models on young and healthy individuals is important to benchmark algorithms' performance, the real value of rPPG lies in helping those who are at a high risk of CVD. The exact reason to miss out on data from CVD patients in these datasets is not clear. We believe it may be related to easier access to the younger population (e.g. university students) and healthy individuals. Recruiting patients may require strict ethical approvals that could be inhibitory. However, models trained in the younger population with not much HR fluctuations or abrupt changes may not generalize well on the CVD patients' population. This further limits the validation of rPPG algorithms on cardiac population and their usefulness in clinical applications.

From the above discussion, it is clear that the majority of the current rPPG datasets suffer from different types of representation biases, including age, sex, ethnicity and patient inclusivity. We argue that these representation biases could be a hindrance in translating the real potential of deep learning and other predictive modeling algorithms to clinical and real-world settings. The reviewed datasets provide a good platform for benchmarking new or existing algorithms, yet their actual benefit to high-risk populations needs significant efforts.

4.1 Mitigating representation biases

Mitigation of representation biases in rPPG datasets can be approached by following strategies:

(1) Community level: The literature review shows that among the datasets reviewed, younger adults dominate these datasets. Most of these datasets are not diverse enough with respect to age, sex and ethnicity. Our review found only one study (OBF) containing patient videos. Ironically, heart patients should be included more in rPPG datasets, as advances in technology will benefit this population. Another finding is the lack of reporting of demographic statistics in these datasets regarding age, sex and ethnicity. These gaps in data collection allude to a clear pathway to the improve quality of rPPG datasets by including older adults and patients across sex and gender and different ethnicities and improving upon reporting the demographics in these datasets. This calls for not only data collection in laboratories but also fostering partnerships with various clinics in the communities and funding diverse studies to create inclusive rPPG data banks. Education and awareness around creating diverse and inclusive rPPG datasets for serving broader communities are paramount to this effort.

(2) Dataset level: From Table 1, we infer that only some of the datasets contain a large number of participant videos (>1,000 videos, e.g. OBF, VIPL-HR, VIPL-HR-V2 and V4V) that may be more appropriate for training deep learning algorithms. Among reviewed datasets, one of the reasons for representation biases could be related to the ease of data collection in certain demographics, such as young healthy adults from specific ethnic groups. As a result, most of the rPPG datasets are highly imbalanced with respect to age, sex, ethnicity and patient population. Data augmentation approaches [65] including synthetic data generation [66] and diffusion models [67] can help rebalancing these datasets, which can help the deep learning models to learn less biased HR estimation models.

(3) Algorithmic level: To handle imbalance in rPPG datasets, approaches that could help mitigate biases include data reweighting, loss-balancing and fine-tuning. HR estimation approaches utilizing imbalanced regression [68] and contrastive regression [69] could help predict abnormal HR or skewed data distributions. To mitigate the effect of the majority demographic group in HR estimation, approaches based on adversarial learning [70] can be helpful to debias the models. Representation bias in rPPG datasets is a conglomeration of different biases along the dimensions of age, sex, ethnicity and patient population. These different biases may require specific

mitigation approaches due to the type of bias, i.e. categorical (e.g. sex) or numerical (e.g. age). An rPPG dataset may have one or more types of biases (see Figure 2, for e.g. containing only younger healthy Asian adult males). Most of the traditional bias mitigation approaches tackle biases in datasets along one or two dimensions, as it is difficult to handle multiple dimensions of biases at the same time. Therefore, novel multi-task learning approaches [71] should be developed to handle multiple biases simultaneously and improve the fairness of HR estimation from rPPG videos.

(4) Evaluation level: The reviewed dataset papers mostly train and test new or existing deep learning models on the same dataset. Therefore, the generalization of those models across other datasets is unknown. It is highly recommended that future research in HR estimation from rPPG videos should not only benchmark on their dataset but also extend it to evaluate on other rPPG datasets to be rendered useful for practical purposes. These cross-dataset evaluations could be along the dimension of age (e.g. train models on younger adults and test on older adults), sex (e.g. train and test models across males and females), ethnicities (e.g. train on white participants and test on minority groups or vice versa) and patients (e.g. train models on healthy adults and test on patient populations). The cross-datasets evaluation approach will highlight the generalization abilities of algorithms across different datasets along the dimensions of age, sex, ethnicities and patient population. In order to improve the generalization of rPPG algorithms and make them acceptable in real-world scenarios, it is imperative to address additional challenges, such as variations in skin tones [72], illumination conditions [73], detection of multiple individuals [74] and long-distance estimations [75]. Patient-specific challenges are often related to the study settings, i.e. the reviewed datasets mostly have healthy participants who were sedentary in controlled lab conditions (low motion and good lighting). However, patients are often subject to conditions that violate these assumptions. Cardiac patients often exhibit physiological states that either weaken the rPPG signal or introduce confounding artifacts [76]. Moreover, clinical settings introduce high levels of environmental and patient-specific noise that are absent in lab datasets, e.g. motion artefacts, occlusions and obstructions and non-ideal lighting. Therefore, there is a real need to collect and evaluate rPPG datasets from participants (preferably cardiac patients) in natural or semi-naturalistic conditions with different confounding factors, including changing physical states (i.e. sitting to walking). This can further help develop frameworks to develop new algorithms, benchmark datasets, streamline data processing pipelines and perform data augmentations and systematic evaluation [77–79]. Performing fairness evaluation [80] of HR estimation models based on biased rPPG datasets could be a useful metric to identify the presence of representation biases and develop models to mitigate their impact.

To illustrate the practical application of the four-level mitigation strategies proposed above, specific datasets from Table 1 can be mapped to the technical intervention levels. While the community-level strategy is a broader call for creating new publicly available inclusive datasets (addressing the severe patient bias and lack of demographic reporting that we found), the other three levels can be directly applied to existing datasets. Table 2 shows some examples of datasets derived from Table 1 and their relation to different mitigation strategies discussed in this section.

There is a natural priority among the above-discussed mitigation strategies. We propose that addressing data collection at the community and/or data level is the foundational priority, as achieving a representative and diverse data pool is a non-negotiable requirement for achieving equitable generalization. This must be supported and validated by fixing the evaluation level through the adoption of standardized, subgroup-specific metrics. Without rigorous evaluation standards that explicitly measure performance across various

Table 2. Sample dataset candidates for applying the four-level mitigation strategies

Mitigation level	Example datasets (from Table 1)	Rationale for fit
Data level	VIPL-HR, VIPL-HR-V2 and V4V	These are the three largest datasets identified in our review. However, they suffer from significant, known demographic skews (e.g. VIPL-HR and VIPL-HR-V2 are Asian-only; VIPL-HR is sex-skewed; V4V is age-skewed. Their large scale makes them ideal candidates for re-balancing using data augmentation, synthetic data generation or diffusion models to correct for the under-representation of specific groups
Algorithm level	BP4D+, V4V, NRC-UBC and iBVP	These datasets are strong candidates because they provide explicit demographic annotations that are necessary for algorithm-level fixes. For example, BP4D+ and V4V annotate multiple ethnicities and NRC-UBC uniquely annotates both age groups and skin tones. This labeling enables the use of techniques like adversarial debiasing or multi-task learning to train models that are invariant to these specific biased attributes
Evaluation level	Example Pairings VIPL-HR and BP4D+ MAHNOB-HCI and OBF	This level requires cross-dataset generalization tests to check for bias. The pairings should be chosen to stress-test a specific bias Ethnicity: Train on the ethnically homogenous VIPL-HR (Asians) and test on the diverse BP4D+ Patient: Train on a healthy, young dataset, for, e.g. MAHNOB-HCI (19–40 years) and test on OBF, the only dataset identified that includes patient data

demographics (age and ethnicity) and patient groups, we cannot accurately confirm that data collection or complex algorithmic-level fixes are effectively reducing bias and improving fairness.

5. Conclusions

In this paper, we highlighted the presence of representation and dataset biases in various rPPG datasets. Instead of focusing on model architecture or comparing performances on standard datasets, we focused on the representation bias issue that is often overlooked and could undermine the usefulness of these models in real-world situations. By doing an extensive literature review on 32 commonly used rPPG datasets, we showed that most of the available datasets do not consider important demographic factors during data collection along the dimensions of age, sex, ethnicity and including patients. This could be related to the ease of data collection; however, it could lead to models that may not be generalizable to other groups, demographics or subsets of populations. Furthermore, the existing datasets predominantly feature young adult participants, making it essential to incorporate datasets with heart patients, newborns and older populations to assess rPPG methods effectively. Besides building biased models, an adverse consequence of excluding older adults from rPPG datasets (and other digital health applications) could constitute digital ageism [81] that could lead to discrimination and lost opportunities for older adults. The severe under-representation of diverse ethnicities and skin tones in current datasets fundamentally limits the generalizability of color-based rPPG methods, introducing a clear racial and/or ethnic bias. Furthermore, the scarcity of data from patients with complex or chronic physiological states (e.g. arrhythmias and heart failure) points to a critical form of exclusion akin to digital ableism [82], where the technology fails to serve the population most in need of continuous health monitoring. Mitigating these systemic representation biases across all demographic and physiological axes is essential for achieving equitable and robust real-world deployment of rPPG technology.

The finding that half of the reviewed studies fail to report ethnic demographic data represents one of the most significant and fundamental barriers to creating fairer rPPG algorithms. This failure to report prevents researchers from even identifying if their models perform poorly across different skin tones, thereby hindering the design of targeted algorithmic mitigation strategies and preventing the development of robust and equitable evaluation standards. Until the research community mandates and enforces comprehensive, standardized reporting of demographic data in all new dataset releases and model validations, these systemic representation gaps will persist and continue to compromise the clinical viability of rPPG.

Our findings are limited to 32 rPPG datasets, and further investigation of more datasets could reveal insights into the representation bias problem. We provided four levels of representation bias mitigation strategies comprising community, data, algorithm and evaluation. These strategies point to closer integration of researchers with the community, awareness and education around collecting and reporting inclusive and diverse datasets, along with innovations in evaluation, data augmentation and algorithms to handle multiple types of representation biases in rPPG datasets for fair estimation of HR. To summarize, there is a pressing need for comprehensive, inclusive, diverse and high-quality real-world rPPG datasets to thoroughly evaluate the generalization of new or existing methods for estimating HR from human videos.

References

1. World Health Organization Cardiovascular diseases. Available from: Https://WwwWhoInt/Health-Topics/Cardiovascular-Diseases#tab=tab_1 2021
2. Ho KJ. Cardiovascular diseases fact sheet. Nutr Asp Aging. 2018; 2: 75-100.
3. Butler J, Rodondi N, Zhu Y, Figaro K, Fazio S, Vaughan DE, Satterfield S, Newman AB, Goodpaster B, Bauer DC, Holvoet P, Harris TB, de Rekeneire N, Rubin S, Ding J, Kritchevsky SB. Metabolic syndrome and the risk of cardiovascular disease in older adults. J Am Coll Cardiol. 2006; 47(8): 1595-602. doi: 10.1016/j.jacc.2005.12.046.
4. Perret-Guillaume C, Joly L, Benetos A. Heart rate as a risk factor for cardiovascular disease. Prog Cardiovasc Dis. 2009; 52(1): 6-10. doi: 10.1016/j.pcad.2009.05.003.
5. Scalise L. Non contact heart monitoring. In: Advances in Electrocardiograms - Methods and Analysis; 2012. doi: 10.5772/22937.
6. Kaplan Berkaya S, Uysal AK, Sora Gunal E, Ergin S, Gunal S, Gulmezoglu MB. A survey on ECG analysis. Biomed Signal Process Control. 2018; 43: 216-35. doi: 10.1016/j.bspc.2018.03.003.
7. Zarak MS, Khan SA, Majeed H, Yasinzai AQK, Hamzazai W, Chung D, Koshkarian G, Fleming KS. Systematic review of validation studies for the use of wearable smartwatches in the screening of atrial fibrillation. Int J Arrhythmia. 2024; 25(1): 11. doi: 10.1186/s42444-024-00118-5.
8. Pustiek M, Beristain A, Kos A. Challenges in wearable devices based pervasive wellbeing monitoring. In: Proceedings - 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things. IIKI; 2015, 2016. doi: 10.1109/IIKI.2015.58.
9. Alian AA, Shelley KH. Photoplethysmography. Best Pract Res Clin Anaesthesiol. 2014; 28(4): 395-406. doi: 10.1016/j.bpa.2014.08.006.
10. Swinehart DF. The Beer-Lambert law. J Chem Educ. 1962; 39(7): 333. doi: 10.1021/ed039p333.
11. Kim NH, Yu SG, Kim SE, Lee EC. Non-contact oxygen saturation measurement using ycgcr color space with an rgb camera. Sensors. 2021; 21(18): 6120. doi: 10.3390/s21186120.
12. Alugubelli N, Abuissa H, Roka A. Wearable devices for remote monitoring of heart rate and heart rate variability–what we know and what is coming. Sensors. 2022; 22: 8903. doi: 10.3390/s22228903.
13. Lee J, Kim M, Park HK, Kim IY. Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths. Sensors. 2020; 20(5): 1493. doi: 10.3390/s20051493.

14. Aarts LAM, Jeanne V, Cleary JP, Lieber C, Nelson JS, Bambang OS, Verkruysse W. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - a pilot study. Early Hum Dev. 2013; 89(12): 943-8. doi: 10.1016/j.earlhumdev.2013.09.016.
15. Hawkins SD, Feldman SR. Contact urticaria caused by the Apple watch – a case report. J Dermatol Surg. 2017; 21(2): 84-6. doi: 10.1016/j.jdds.2016.08.003.
16. Xu L, Pan C, Xu B, Sun Z, Zhou K, Yan R, Weng K, Wu Y, Li H. A qualitative exploration of a user-centered model for smartwatch comfort using grounded theory. Int J Hum Comput Interact. 2024; 41(2): 1091-106. doi: 10.1080/10447318.2024.2313273.
17. Xiao H, Liu T, Sun Y, Li Y, Zhao S, Avolio A. Remote photoplethysmography for heart rate measurement: a review. Biomed Signal Process Control. 2024; 88: 105608. doi: 10.1016/j.bspc.2023.105608.
18. Lee RJ, Sivakumar S, Lim KH. Review on remote heart rate measurements using photoplethysmography. Multimed Tools Appl. 2024; 83(15): 44699-728. doi: 10.1007/s11042-023-16794-9.
19. Yu Z, Li X, Zhao G. Facial-video-based physiological signal measurement: recent advances and affective applications. IEEE Signal Process Mag. 2021; 38(6): 50-8. doi: 10.1109/MSP.2021.3106285.
20. Suresh H, Guttag J. A framework for understanding sources of harm throughout the machine learning life cycle. In: Equity and Access in Algorithms, Mechanisms, and Optimization. New York, NY: ACM; 2021. p. 1-9. doi: 10.1145/3465416.3483305.
21. Khan SS, Madden MG. One-class classification: taxonomy of study and review of techniques. Knowl Eng Rev. 2014; 29(3): 345-74. doi: 10.1017/S026988891300043X.
22. Dasari A, Prakash SKA, Jeni LA, Tucker CS. Evaluation of biases in remote photoplethysmography methods. NPJ Digit Med. 2021; 4(1): 91. doi: 10.1038/s41746-021-00462-z.
23. UNHCR Older persons. Available from: https://emergency.unhcr.org/protection/persons-risk/older-persons
24. Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. IEEE Trans Affect Comput. 2012; 3(1): 42-55. doi: 10.1109/T-AFFC.2011.25.
25. Koelstra S, Mühl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I. DEAP: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput. 2012; 3(1): 18-31. doi: 10.1109/T-AFFC.2011.15.
26. Stricker R, Muller S, Gross HM. Non-contact video-based pulse rate measurement on a mobile service robot. In: IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions; 2014. p. 1056-62. doi: 10.1109/ROMAN.2014.6926392.
27. Estepp JR, Blackford EB, Meier CM. Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. In: Conf Proc IEEE Int Conf Syst Man Cybern. 2014; 2014. January. p. 1462-69. doi: 10.1109/smc.2014.6974121.
28. Zhang X, Yin L, Cohn JF, Canavan S, Reale M, Horowitz A, Liu P, Girard JM. BP4D-spontaneous: a high-resolution spontaneous 3D dynamic facial expression database. Image Vis Comput. 2014; 32(10): 692-706. doi: 10.1016/j.imavis.2014.06.002.
29. Zhang Z, Girard JM, Wu Y, Zhang X, Liu P, Ciftci U, Canavan S, Reale M, Horowitz A, Yang H, Cohn JF, Ji Q, Yin L. Multimodal spontaneous emotion corpus for human behavior analysis. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016; 2016. December. p. 3438-46. doi: 10.1109/CVPR.2016.374.
30. Heusch, G, Anjos A, Marcel S. A reproducible study on remote heart rate measurement. 2017.
31. Hsu G-S, Ambikapathi A, Chen M-S. Deep learning with time-frequency representation for pulse estimation from facial videos. In: 2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE; 2017. p. 383-9. doi: 10.1109/BTAS.2017.8272721.

32. Pai A, Veeraraghavan A, Sabharwal A. CameraHRV: robust measurement of heart rate variability using a camera; 2018. 27. doi: 10.1117/12.2289205.
33. Li X, Alikhani I, Shi J, Seppanen T, Junttila J, Majamaa-Voltti K, *et al*. The OBF database: a large face video database for remote physiological signal measurement and atrial fibrillation detection. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE; 2018. p. 242-9. doi: 10.1109/FG.2018.00043.
34. Spetlik R, Franc V, Cech J, Matas J. Visual heart rate estimation with convolutional neural network. In: British Machine Vision Conference 2018. BMVC 2018; 2018.
35. Pilz CS, Zaunseder S, Krajewski J, Blazek V. Local group invariance for heart rate estimation from face videos in the wild. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops; 2018. 2018. June. p. 1335-13358. doi: 10.1109/CVPRW.2018.00172.
36. Kopeliovich M, Petrushan M, Shaposhnikov D. Approximation-based transformation of color signal for heart rate estimation with a webcam. Pattern Recogn Image Anal. 2018; 28(4): 646-51. doi: 10.1134/S1054661818040181.
37. Bobbia S, Macwan R, Benezeth Y, Mansouri A, Dubois J. Unsupervised skin tissue segmentation for remote photoplethysmography. Pattern Recognit Lett. 2019; 124: 82-90. doi: 10.1016/j.patrec.2017.10.017.
38. Maki Y, Monno Y, Yoshizaki K, Tanaka M, Okutomi M. Inter-beat interval estimation from facial video based on reliability of BVP signals. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBS; 2019. doi: 10.1109/EMBC.2019.8857081.
39. Niu X, Han H, Shan S, Chen X. VIPL-HR: a multi-modal database for pulse estimation from less-constrained face video. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2019; 11365(LNCS): 562-76. doi: 10.1007/978-3-030-20873-8_36.
40. Li X, Han H, Lu H, Niu X, Yu Z, Dantcheva A, Zhao G, Shan S. The 1st challenge on remote physiological signal sensing (RePSS). In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops; 2020. 2020. June. p. 1274-81. doi: 10.1109/CVPRW50498.2020.00165.
41. Nowara EM, Marks TK, Mansour H, Veeraraghavan A. Near-infrared imaging photoplethysmography during driving. IEEE Trans Intell Transp Syst. 2022; 23(4): 3589-600. doi: 10.1109/TITS.2020.3038317.
42. Svoren H, Thambawita V, Halvorsen P, Jakobsen P, Garcia-Ceja E, Noori FM, Hammer HL, Lux M, Riegler MA, Hicks SA. Toadstool: a dataset for training emotional intelligent machines playing Super Mario bros. In: MMSys 2020 - Proceedings of the 2020 Multimedia Systems Conference; 2020. p. 309-14. doi: 10.1145/3339825.3394939.
43. Gudi A, Bittner M, van Gemert J. Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation. Appl Sci. 2020; 10(23): 8630. doi: 10.3390/app10238630.
44. Perepelkina O, Artemyev M, Churikova M, Grinenko M. HeartTrack: convolutional neural network for remote video-based heart rate monitoring. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops; 2020. 2020. June. doi: 10.1109/CVPRW50498.2020.00152.
45. Sabour RM, Benezeth Y, De Oliveira P, Chappé J, Yang F. UBFC-phys: a multimodal database for psychophysiological studies of social stress. IEEE Trans Affect Comput. 2023; 14(1): 622-36. doi: 10.1109/TAFFC.2021.3056960.
46. Revanur A, Li Z, Ciftci UA, Yin L, Jeni LA. The first vision for vitals (V4V) challenge for non-contact video-based physiological estimation. In: Proceedings of the IEEE International Conference on Computer Vision; 2021. 2021. October. p. 2760-7. doi: 10.1109/ICCVW54120.2021.00310.
47. Zahid H, Sreenivasan R, Nirmalya R. MPSC-rPPG dataset. IEEE Dataport. 2021. doi: 10.21227/ddgz-tx88.

48. Yang Z, Wang H, Lu F. Assessment of deep learning-based heart rate estimation using remote photoplethysmography under different illuminations. IEEE Trans Hum Mach Syst. 2022; 52(6): 1236-46. doi: 10.1109/THMS.2022.3207755.
49. Joshi J, Cho Y. iBVP dataset: RGB-thermal rPPG dataset with high resolution signal quality labels. Electronics. 2024; 13(7): 1334. doi: 10.3390/electronics13071334.
50. Benezeth Y, Krishnamoorthy D, Botina Monsalve DJ, Nakamura K, Gomez R, Mitéran J. Video-based heart rate estimation from challenging scenarios using synthetic video generation. Biomed Signal Process Control. 2024; 96: 106598. doi: 10.1016/j.bspc.2024.106598.
51. Rivest-Hénault D, Proulx C, Ho K, Jafari N, Yakubu T, Torres S, *et al.* Influential factors in rPPG: insights from a diverse and inclusive empirical study. In: 2024 IEEE 20th International Conference on Body Sensor Networks (BSN). IEEE; 2024. p. 1-4. doi: 10.1109/BSN63547.2024.10780670.
52. Santos MAA, Sousa ACS, Reis FP, Santos TR, Lima SO, Barreto-Filho JA. Does the aging process significantly modify the mean heart rate?. Arq Bras Cardiol. 2013; 101(5): 388-98. doi: 10.5935/abc.20130188.
53. Stein PK, Kleiger RE, Rottman JN. Differing effects of age on heart rate variability in men and women. Am J Cardiol. 1997; 80(3): 302-5. doi: 10.1016/S0002-9149(97)00350-0.
54. Münzel T, Hahad O, Gori T, Hollmann S, Arnold N, Prochaska JH, Schulz A, Beutel M, Pfeiffer N, Schmidtmann I, Lackner KJ, Keaney JF Jr. Wild PS. Heart rate, mortality, and the relation with clinical and subclinical cardiovascular diseases: results from the Gutenberg health study. Clin Res Cardiol. 2019; 108(12): 1313-23. doi: 10.1007/s00392-019-01466-2.
55. Gupta JI, Shea MJ. Effects of aging on the heart and blood vessels. Available from: https://www.msdmanuals.com/home/heart-and-blood-vessel-disorders/biology-of-the-heart-and-blood-vessels/effects-of-aging-on-the-heart-and-blood-vessels
56. Prabhavathi K, Tamarai Selvi K, Poornima KN, Sarvanan A. Role of biological sex in normal cardiac function and in its disease outcome - a review. J Clin Diagn Res. 2014; 8: BE01-BE4. doi: 10.7860/JCDR/2014/9635.4771.
57. Ramaekers D, Ector H, Aubert AE, Rubens A, Van De Werf F. Heart rate variability and heart rate in healthy volunteers: is the female autonomic nervous system cardioprotective?. Eur Heart J. 1998; 19(9): 1334-41. doi: 10.1053/euhj.1998.1084.
58. Tada H, Fujino N, Hayashi K, Kawashiri M, Takamura M. Human genetics and its impact on cardiovascular disease. J Cardiol. 2022; 79(2): 233-9. doi: 10.1016/j.jjcc.2021.09.005.
59. Barbaresko J, Rienks J, Nöthlings U. Lifestyle indices and cardiovascular disease risk: a meta-analysis. Am J Prev Med. 2018; 55(4): 555-64. doi: 10.1016/j.amepre.2018.04.046.
60. Niakouei A, Tehrani M, Fulton L. Health disparities and cardiovascular disease. Healthcare. 2020; 8(1): 65. doi: 10.3390/healthcare8010065.
61. Razieh C, Zaccardi F, Miksza J, Davies MJ, Hansell AL, Khunti K, Yates T. Differences in the risk of cardiovascular disease across ethnic groups: UK biobank observational study. Nutr Metab Cardiovasc Dis. 2022; 32(11): 2594-602. doi: 10.1016/j.numecd.2022.08.002.
62. Chiu M, Austin PC, Manuel DG, Tu JV. Comparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ Can Med Assoc J. 2010; 182(8): E301-10. doi: 10.1503/cmaj.091676.
63. Lopez-Neyman SM, Davis K, Zohoori N, Broughton KS, Moore CE, Miketinas D. Racial disparities and prevalence of cardiovascular disease risk factors, cardiometabolic risk factors, and cardiovascular health metrics among US adults: NHANES 2011–2018. Sci Rep. 2022; 12(1): 19475. doi: 10.1038/s41598-022-21878-x.
64. Mohebi R, Chen C, Ibrahim NE, McCarthy CP, Gaggin HK, Singer DE, Hyle EP, Wasfy JH, Januzzi JL Jr. Cardiovascular disease projections in the United States based on the 2020 census estimates. J Am Coll Cardiol. 2022; 80(6): 565-78. doi: 10.1016/j.jacc.2022.05.033.
65. Perche S, Botina D, Benezeth Y, Nakamura K, Gomez R, Miteran J. Data-augmentation for deep learning based remote photoplethysmography methods. In: 2021 9th E-Health and Bioengineering Conference, EHB 2021; 2021. p. 1-4. doi: 10.1109/EHB52898.2021.9657650.

66. Wang Z, Ba Y, Chari P, Bozkurt OD, Brown G, Patwa P, Vaddi N, Jalilian L, Kadambi A. Synthetic generation of face videos with plethysmograph physiology. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2022. 2022. June. p. 20555-64. doi: 10.1109/CVPR52688.2022.01993.

67. Chen S, Wong K-L, Chin J-W, Chan T-T, So RHY. DiffPhys: enhancing signal-to-noise ratio in remote photoplethysmography signal using a diffusion model approach. Bioengineering. 2024; 11 (8): 743. doi: 10.3390/bioengineering11080743.

68. Yang Y, Zha K, Chen YC, Wang H, Katabi D. Delving into deep imbalanced regression. Proc Mach Learn Res. 2021; 139: 11842-11851.

69. Karagoz B, Ashraf A, Khan SS. Supervised sequential contrastive regression: improving performance on imbalanced rehabilitation exercises. Biomed Signal Process Control. 2026; 111: 108347. doi: 10.1016/j.bspc.2025.108347.

70. Khan SS, Shi T, Donato-Woodger S, Chu CH. Mitigating digital ageism in skin lesion detection with adversarial learning. Algorithms. 2025; 18(2): 55. doi: 10.3390/a18020055.

71. Yin R-N, Jia R-S, Cui Z, Sun H-M. PulseNet: a multitask learning network for remote heart rate estimation. Knowl Based Syst. 2022; 239: 108048. doi: 10.1016/j.knosys.2021.108048.

72. Nowara EM, McDuff D, Veeraraghavan A. A meta-analysis of the impact of skin tone and gender on non-contact photoplethysmography measurement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop; 2020. p. 284-5.

73. Yang Z, Wang H, Lu F. Assessment of deep learning-based heart rate estimation using remote photoplethysmography under different illuminations. IEEE Trans Hum Mach Syst. 2022; 52(6): 1236-46. doi: 10.1109/THMS.2022.3207755.

74. Khanam FTZ, Al-Naji A, Chahl J. Remote monitoring of vital signs in diverse non-clinical and clinical scenarios using computer vision systems: a review. Appl Sci. 2019; 9(20): 4474. doi: 10.3390/app9204474.

75. Blackford EB, Estepp JR. Measurements of pulse rate using long-range imaging photoplethysmography and sunlight illumination outdoors; 2017. In: Coté GL Optical Diagnostics and Sensing XVII: Toward Point-of-Care Diagnostics: 100720S. doi: 10.1117/12.2253460.

76. Hermans ANL, Isaksen JL, Gawalko M, Pluymaekers NAHA, van der Velden RMJ, Snippe H, Evens S., De Witte G, Luermans JGLM, Manninger M, Lumens J, Kanters JK, Linz D. Accuracy of continuous photoplethysmography-based 1 min mean heart rate assessment during atrial fibrillation. EP Europace. 2023; 25(3): 835-44. doi: 10.1093/europace/euad011.

77. Dawoodjee I, Ghahramani A. A flexible framework for design and validation of rPPG methods. IEEE Access. 2025; 13: 50059-76. doi: 10.1109/ACCESS.2025.3552585.

78. Boccignone G, Conte D, Cuculo V, D'Amelio A, Grossi G, Lanzarotti R. An open framework for remote-PPG methods and their assessment. IEEE Access. 2020; 8: 216083-103. doi: 10.1109/ACCESS.2020.3040936.

79. Liu X, Narayanswamy G, Paruchuri A, Zhang X, Tang J, Zhang Y, *et al.* rPPG-toolbox: deep remote PPG toolbox. In: Proceedings of the 37th International Conference on Neural Information Processing Systems, Red Hook, NY. Curran Associates; 2023.

80. Jiang Z, Seyedi S, Griner E, Abbasi A, Rad AB, Kwon H, Cotes RO, Clifford GD. Evaluating and mitigating unfairness in multimodal remote mental health assessments. PLOS Digital Health. 2024; 3(7): e0000413. doi: 10.1371/journal.pdig.0000413.

81. Chu CH, Nyrup R, Leslie K, Shi J, Bianchi A, Lyn A, McNicholl M, Khan S, Rahimi S, Grenier A. Digital ageism: challenges and opportunities in artificial intelligence for older adults. Gerontologist. 2022; 62(7): 947-55. doi: 10.1093/geront/gnab167.

82. Charitsis V, Lehtiniemi T. Data ableism: ability expectations and marginalization in automated societies. TV and New Media. 2023; 24(1): 3-18. doi: 10.1177/15274764221077660.

Corresponding author
Shehroz S. Khan can be contacted at: shehroz.khan@aum.edu.kw