UCLA
UCLA Electronic Theses and Dissertations

Title
Robust and Equitable Non-Contact Health Sensing

Permalink
https://escholarship.org/uc/item/0mb6m0cq

Author
Vilesov, Alexander

Publication Date
2023

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Robust and Equitable Non-Contact Health Sensing

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Electrical and Computer Engineering

by

Alexander Vilesov

2023

© Copyright by
Alexander Vilesov
2023

ABSTRACT OF THE THESIS

Robust and Equitable Non-Contact Health Sensing

by

Alexander Vilesov
Master of Science in Electrical and Computer Engineering
University of California, Los Angeles, 2023
Professor Achuta Kadambi, Chair

Contactless vital sensing is gaining prominence with applications in disease control, health monitoring, and medicine; airports are beginning to use infrared thermometers to screen for fevers, automobile companies are researching how cars can wirelessly detect drowsy drivers, and the medical field is exploring the benefits of how cameras can be used to remotely monitor neonates or detect diseases such as atrial fibrillation or sleep apnea. However, prior research to a large extent has not explored when remote vital sensing methods fail and if they may be disadvantageous to certain physiologies more than others such as age, weight, or gender. New methods in the field should strive to determine the impact of these variables as well as rectify inaccuracies in sensing that may occur if possible. This work explores how skin tone can adversely impact heart-rate detection with cameras and temperature evaluation with thermal cameras. Multimodal fusion and algorithmic techniques are proposed to improve skin tone equity while improving performance of contactless vital sensing methods.

The thesis of Alexander Vilesov is approved.

Jonathan Kao

Bolei Zhou

Achuta Kadambi, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

1 Introduction **1**

2 Radar and Camera Fusion for Equitable, Robust Plethysmography . . . **4**

2.1 Problem Background 4

2.1.1 Introduction 4

2.1.2 Related Work 5

2.1.3 Contributions and Scope 7

2.1.4 Problem Formulation 8

2.2 iPPG Signal Strength and Skintone Dependency 9

2.2.1 Conventional Pulse Oximetry 9

2.2.2 iPPG Signal Strength 11

2.2.3 Skin Tone Bias in iPPG 14

2.2.4 Implementation 15

2.3 Radar for Plethysmography 16

2.3.1 Radar Theory 16

2.3.2 Implementation 19

2.4 Multimodal Fusion of Camera and Radar 20

2.4.1 Resisting SNR Bias through Sensor Fusion 20

2.4.2 Overall Inferences 23

2.4.3 Implementation of Fusing RGB Camera and Radar for Plethysmography 23

2.5 MMFAIR Dataset 24

2.6 Results 26

2.6.1 Evaluation Metrics . . . 26
2.6.2 Qualitative Performance . . . 27
2.6.3 Benefit of the Skin Tone Discriminative Loss . . . 32
2.6.4 Discussion . . . 33

3 Correcting for Solar Loading in Infrared Thermometers . . . 35

3.1 Problem Background . . . 35
3.1.1 Introduction . . . 35
3.1.2 Related Work . . . 37
3.1.3 Contributions and Scope . . . 40
3.2 Thermal Light Transport Preliminaries . . . 40
3.2.1 Multipath Temperature Estimation . . . 42
3.3 Effect of Solar Loading on Temperature Estimation . . . 45
3.3.1 Core Temperature Estimation . . . 45
3.3.2 Gap in Prior Models: Solar Loading . . . 45
3.3.3 Equity of Solar Loading . . . 46
3.4 Correcting Solar Loading Induced Errors . . . 46
3.4.1 Transient Model and Correction (Solution I) . . . 46
3.4.2 Spatial Modulation and Correction (Solution II) . . . 49
3.4.3 Correcting for Solar Loading in Space-Time . . . 51
3.5 Experimental Method . . . 52
3.5.1 Protocol and Dataset . . . 52
3.5.2 Solar Loading Correction . . . 54
3.6 Results . . . 56

3.6.1 Transient Solution (Solution I) 56
3.6.2 Spatial Solution (Solution II) 57
3.6.3 Equity Analysis 59
3.7 Discussion 60

4 Conclusions 62

LIST OF FIGURES

2.1 Hb Absorption Spectras [69]. 9

2.2 **The iPPG per-pixel SNR drastically worsens with increasing skin melanin fraction.** Using biophysical skin reflectance models, we estimate the iPPG signal strength as a function of skin melanin fraction. Along with the monotonic decrease in signal strength with melanin fraction, we also note finer trend differences between the SNR for the red, green and blue channels, as well as dependence on the spectral properties of the light source. 14

2.3 **A FMCW Chirp Sequence.** The blue and red signal are the transmitted and received chirps plotted with their frequency content as a function of time. The green signal denotes the mixed signal whose phase changes while the frequency remains relatively constant. 17

2.4 **We use a 77 Ghz FMCW radar setup for non-contact radar plethysmography.** Chirp signals are bounced off the subject's chest in order to capture subtle motion. By exploiting the dependency of the phase on the distance of flight, we are able to measure this motion. 18

2.5 **The proposed approach uses a novel adversarial discriminative training-based approach for skin tone debiasing in the modality fusion module.** The fusion network operates in the frequency domain using an alternating waveform reconstruction and adversarial losses. 23

2.6 **A mobile multi-modal sensing platform was deployed to collect our remote plethysmography dataset. The parts list and reference designs may be found at https://github.com/UCLA-VMG/EquiPleth.** 25

2.7 **Qualitative analysis of estimated waveforms indicates superior overall performance for the fusion model, with reduced group-wise bias.** We highlight a randomly chosen snippet of the plethysmograph waveform to highlight qualitative differences. 29

2.8 **Plotting the heart-rate estimation error versus the ground truth heart rate (Bland-Altman plots) emphasizes performance benefits of the proposed multi-modal fusion model.** Each plot highlights the distribution of the ground truth heart rates (top), distribution of the heart-rate estimation errors (right) and the plot of the estimation errors versus the ground truth heart rates. 31

3.1 **Experimental data showing the solar loading effect. (a)** The experiment setup. **(b)** Solar loading results in elevated skin temperature measurements that are biased against dark skin toned subjects compared to light skin toned subjects. Commonly accepted fever threshold ($> 37.9°C$) [31] would misclassify solar loaded, healthy, dark skin toned subjects. 36

3.2 **Thermal image signal processing (ISP). (a)** Raw data from a thermal sensor is converted to temperature through a series of processing blocks. For thermal cameras, additional processing for temporal drift and gain control (shaded) are added. **(b)** Infrared thermometers convert sensor temperature to core temperature using a device-specific calibration. We propose an additional solar loading correction (shaded). 41

3.3 **Solution I requires transient data to extrapolate, which increases the waiting time.** 49

3.4 **Skin heat transfer.** Skin temperature is determined by the heat transfer of endogenous and exogenous factors. Penne's bio-heat equation describes these heat transfer processes. 50

3.5 **Dataset demographics.** Dataset distribution across: **(a)** Fitzpatrick scale skin tones. **(b)** Melanin Index (MI) values. 54

3.6 **Solution II enables single shot correction and does not over correct when solar loading is not present.** From left to right: RGB image, steady state thermal image (ground truth), indoor cooling after solar loading (0 min, 5 min) and the results of our correction. In the plots, we compare the solar loaded skin temperature (blue, T_{skin}), the baseline temperature (dotted, $\overline{T}_{\text{skin}}$) and our corrected skin temperature (black, T^*_{skin}) Our method successfully removes solar loading (up to 3°C) on multiple subjects. The plots show frame-by-frame inference of Solution II, showing that Solution II is not overfitting to a fixed temperature offset. 58

LIST OF TABLES

2.1 **Notation used for light transport modeling of iPPG.** The left column shows the notation used and the right column describes the notation. 12

2.2 **We use multiple performance and fairness metrics for evaluation.** Performance and fairness are not necessarily correlated properties and requires both for full comparison. 27

2.3 **Across baselines spanning the radar and camera modalities, the proposed fusion model shows performance and fairness improvements over the unimodal iPPG modality.** The performance metrics measure the average performance across the entire dataset. The pairwise difference between light and dark groups being bracketed and the sign shows direction of bias - ideally the absolute value of this bias should be low. The fairness threshold test measures the percent of the light and dark populations failing the AAMI standard. The best performing numbers are bolded between the fusion, RF, and PhysNet backbone. 28

2.4 **An adversarial network for skin tone estimation is a novel contribution that helps obtain a more equitable plethysmograph estimator across skin tone.** When compared with a fusion network trained without the adversarial network, significant improvements are noted across all performance fairness measures, at a small cost in performance measures. 33

3.1 **Today, commercial thermal IR sensors require a waiting period for solar loading to cool down.** 37

3.2 **Errors of solutions during steady state and solar loading.** 56

3.3 **Performance of commercial NCITs.** We test three NCITs which are marketed as being within the ASME and FDA acceptable error limits. However, the devices are not within their specified ranges even during steady state. 60

ACKNOWLEDGMENTS

I would like to thank the following people who have helped shape this Masters thesis and more importantly who have made my time at UCLA exciting and fun.

First, I would like to begin with thanking my advisor, Professor Achuta Kadambi. I am extremely grateful for his guidance, support, and insights that has proved vital in the development of this work and myself as a researcher.

Next, I would like to thank Dr. Laleh Jalilian. Without her, data collection, IRBs, and frankly all the projects I have worked on would be near-impossible. Thank you for your support and enthusiasm as well as for your unique insights into human physiology.

I would like to thank Professor Jonathan Kao and Professor Bolei Zhou, as part of my thesis committee, for reviewing and supporting this work.

In particular, I would like to thank some of the students I have worked most closely with: Pradyumna Chari, Adnan Armouti, and Anirudh Bindiganavale Harish. Thank you for all the arduous work, long nights, stimulating technical/non-technical conversations, and of course, the fun that made the Masters at UCLA so memorable.

I would also like to thank all my other co-authors and colleagues on this work - Ellin Zhao, Kimaya Kulkarni, Ananya Deoghare, and Shreeram Athreya - for their help and amazing technical discussions.

Lastly, I would like to thank my parents for their unwavering support and never-ending guidance.

PREVIOUS PUBLICATIONS

This thesis revises the following publications:

Vilesov A, Chari P, Armouti A, Harish AB, Kulkarni K, Deoghare A, Jalilian L, Kadambi A. Blending camera and 77 GHz radar sensing for equitable, robust plethysmography. ACM Trans. Graph.(SIGGRAPH). 2022 Jul 1;41(4):1-4. [93].

Zhao EQ, Vilesov A, Athreya S, Chari P, Merlos J, Millett K, Cyr NS, Jalilian L, Kadambi A. Making Thermal Imaging More Equitable and Accurate: Resolving Solar Loading Biases. arXiv preprint arXiv:2304.08832. 2023 Apr 18. [114]

CHAPTER 1

Introduction

Non-contact health monitoring technologies are gaining attention due to the benefits in their method of sensing over contact technologies. They are noninvasive, passive, and more flexible. In recent time, these technologies have seen interest with infrared thermometers during pandemics due to reduction of contamination [33] and radars for determining levels of drowsiness of automobile drivers through detection of their respiratory rate [76, 79]. Current research has explored more possibilities of non-contact monitoring with cameras for getting a photoplethysmograph waveform in groups like neonates whose skin is too sensitive to contact pulse oximeters [2], NIR cameras for sleep apnea detection [95], and the use of simple RGB cameras for remote-screening of cardiac diseases such as atrial fibrillation [109] or to record heart-rate and breathing-rate during a medical telehealth call [55]. The long-term goal of these technologies are to reach levels of convenience and acceptance where there is ubiquitous use of it in the medical field and everyday life. This thesis brings us closer to this goal by taking steps to making the technology both more equitable and robust.

Due to the inherent human aspect of health monitoring a major concern with all medical devices is not just their accuracy, but the consistency of their accuracy across various demographics of the population. Unfortunately, there have been many cases where medicine has indirectly favored certain demographics more than others. The form in which this takes can vary with computational, physical, and interpretation bias being proposed as some of the categories in Kadambi [43]. Computational bias may occur for a medical sensor when it fails on a certain group due to being calibrated on a different group. The same sensor

may have a physical bias such that despite being calibrated for all parts of the population, the sensor may not work well on a certain group due to unique innate physical attributes that cause the physics of the sensor to be less accurate. Lastly, inference bias occurs when even if the sensor works at the same accuracy across all people, the medical decision that a doctor makes given the sensor reading may vary based on differences in anatomy of various populations. These different errors in the operation of medical devices are not necessarily mutually exclusive. Kadambi [43] gives concrete examples of the various forms of the bias.

Remote sensing technologies used as medical devices need to be evaluated for biases just like all other medical devices. This thesis focuses on physical biases and how to solve them. Physical biases occur due to differences in anatomy causing sensor performance to be affected. These anatomy differences can be due to gender, body mass index (BMI), age, skintone, or any other physiological factor. This thesis tackles two applications of non-contact health monitoring where skintone affects sensor performance.

The first application is in image photoplethysmography (iPPG) technologies. These technologies typically detect a plethysmography waveform using a camera and ambient or active illumination. The light from illumination travels through our skin and can reach blood vessels located under the epidermis [50]. The remaining light that was not absorbed is scattered or reflected out of the skin and into the camera. The camera can then detect the plethysmography waveform by measuring subtle changes in the intensity of the light coming from the skin due to being modulated by periodic blood volume pulses at the frequency of the heart beating [91]. However, darker skintones have less light reaching these blood vessels and even less light coming back to the camera due to absorption by melanin particles. This thesis proposes a method to make the estimation of the plethysmography signal more equitable across skin tones through the use of multimodal fusion of the camera with another sensing modality, radar [93].

The second application is in the remote estimation of temperature of the human body. All remote estimation of temperature technologies rely on the infrared radiation emitting from

our body whose power is proportional to our body's temperature [67]. While the amount of infrared radiation emitted from two bodies with the same temperature but different skintones is the same [22], certain situations can cause these two bodies to have dramatically different temperature readings. In this thesis, the effect of solar loading, or exposure to sunlight, is shown to cause error in temperature estimation across all skintones but disproportionally causes larger errors in darker skintones [114]. The proposed solution attempts to remove the effect of solar loading in one shot of a thermal camera by taking advantage of the unique heat patterns caused by solar loading.

In summary, this thesis makes two specific contributions:

Contribution A: We show how skintone bias in iPPG can be reduced through multi-modal fusion on a self-collected diverse dataset. A multi-modal hardware and algorithmic approach using radar and camera is proposed to enable better performance and fairness.

Contribution B: We show how solar loading leads to erroneous temperature readings in infrared sensors that is exacerbated with darker skintones. We propose two methods for removing this effect by utilizing either temporal or spatial information.

While this thesis addresses two situations in which skintone can adversely affect medical sensors' performance. There exist many other sensor configurations and applications that may also suffer from a similar bias. Moreover, skintone is only one of the axes of fairness. BMI, age, gender, and many other factors in human physiology can physically affect the sensor. Future research should strive to increase the accuracy of sensors and concurrently remain resistant to physical biases.

CHAPTER 2

Radar and Camera Fusion for Equitable, Robust Plethysmography

2.1 Problem Background

2.1.1 Introduction

Camera-based remote plethysmography is a rapidly developing field. Most methods utilize small changes in the facial skin color as a function of dermal blood volume to capture pulse rate trends [5, 104]. Over the years, a broad range of methods have been proposed, ranging from physics-based approaches [27, 91, 98], blind source separation-based approaches [52, 68] and more recently, data-driven learning-based approaches [24, 111]. Through these advances, the heart-rate estimation performance has steadily approached levels of clinical accuracy. However, more recently, it has been established that most methods for imaging-based remote photoplethysmography (iPPG) are biased in performance against dark skin tone participants [64]. This points towards two potential problems: biases in datasets used for algorithmic evaluation, and potential fundamental biases in the physics of camera-based remote plethysmogaphy.

Another comparable modality for heart-rate estimation is the use of radio frequency (RF) devices. These devices capture the variations in chest displacement through cardiac cycles to estimate the frequency of the heart beat signal. Approaches use different types of radars and include signal processing [6, 56], as well as deep learning-based methods [106]. Ren et

al. [72] performed a comparison study between a camera and a doppler stepped-frequency continuous wave radar on one subject and showed that both modalities perform nearly equally under ideal conditions for extracting heart rates. Additionally, it is noted from our experiments on multiple subjects that while iPPG and radar are comparable, iPPG slightly outperforms radar. However, since radar systems primarily capture displacement signals [1], radar techniques do not theoretically show performance correlation with skin tone and are therefore fairer.

In this thesis, a fresh look at multi-modal fusion from the perspective of bias removal. We show that combining iPPG, a biased modality, with radar, an unbiased modality, results in a better performing algorithm compared to the unimodal methods with only small trade-offs in fairness over the unbiased method. To evaluate optimality, we establish a comprehensive set of performance and fairness metrics tailored to the task of remote plethysmography, evaluated on our novel multi-modal remote plethysmography dataset. This proposed fusion method, along with the existing iPPG and radar-based methods, constitute viable remote heart-rate detection approaches with differing performance and fairness trade-offs that an end-user may select from.

2.1.2 Related Work

Image-photoplethysmography is biased against darker skin tones. To improve fairness and performance of unimodal iPPG, we fuse it with another sensing modality (radar). In what follows, we expand on background context.

Image Photoplethysmography Heart-rate estimation using image-photoplethysmography (iPPG) has been actively studied since the early 2000s [103, 107, 108]. Typically, early methods observed and took advantage of changes in optical absorption of hemoglobin molecules at the surface of the skin during a blood volume pulse with a RGB camera. The work that followed focused on reducing error due to motion with region of interest (ROI) alignment

and clever modeling of physical properties of light reflectance [27, 98]. Remote heart-rate estimation is also achieved with Ballistocardiogram (BCG) methods which extract motion information due to the Newtonian reaction of a blood volume pulse [12]. Color analysis and BCG methods are not limited to use with RGB cameras. Near Infrared (NIR) imaging with active illumination has been employed to combat the effects of unreliable illumination in the visible spectrum [57], despite having a worse signal to noise ratio (SNR) to RGB cameras [94]. Infrared (IR) or thermal imaging has used BCG [13] and temporal temperature differentials [47]. Other work focused on visualizations of the blood volume pulse with Eulerian magnification [104] and augmented reality [38].

More recently, deep learning approaches have been utilized to attain state of the art results. [24] used an attention-based Convolutional Neural Network (CNN) to explicitly fuse skin-reflection and motion information. [111] introduced spatio-temporal CNNs to iPPG to enable temporal context-aware networks. Other work has extended these architectures [63], incorporated meta-learning [51], improved PPG waveform characteristics [81], and augmented iPPG datasets with synthetic examples [10, 100].

Radar Plethysmography Vital sensing using radar was pioneered in the 1970s for respiratory-rate detection [54]. Today, radar research has diverse applications in respiratory-rate, heart-rate, and blood-pressure detection. For heart-rate estimation, various hardware setups are used, including FMCW [6], Ultra Wide Band (UWB) Impulse [71], and Continuous Wave Doppler radars [28]. Vital sign detection is performed by observing millimeter (mm) level displacements in the chest. The average adult has chest displacements for breathing and heart pulses of 1-12 mm and 0.01-0.5 mm, respectively [29]. Through single-subject analysis, [72] note that both camera and radar-based methods perform nearly equally under ideal conditions. That is, both are potentially viable methods for remote heart-rate estimation. Since heart-rate detection is more prone to noise, applications and experimental results are often done with subjects laying down to avoid interference due to motion.

Fairness in iPPG Fairness in machine learning has been a rapidly growing area of research in the last decade. It has spanned ensuring fairness in classification [112], word embeddings [17], and computer vision [99]. Dataset bias is a common problem and it has been shown that performance is lower for minority groups in machine learning problems due to underrepresentation [18] and its effect on majority and minority performance is shown in Chari et al. [21]. In the iPPG field, fairness has been less studied. However, [10, 64] show that dataset bias as well as lower SNRs result in darker skin tones producing poorer performance than lighter skin tones. We introduce a skin tone representative dataset and propose to reduce bias across skin tones through multi-modal fusion.

Multi-modal Fusion Multi-modal fusion is the process of combining two or more modalities to achieve better performance for a given task than any singular modality on its own. In deep learning, architectures either fuse modalities in a middle latent space or at a late stage once each modality independently gives a prediction. For mid-level fusion, Restricted Boltzmann machines [83] are a common choice in popular architectures such as Deep Belief Networks [45] and Stacked Autoencoders [80]. In late-level or decision-level fusion, predictions from various modalities are simply aggregated using majority voting, weighted voting, or a meta-classifier. These architectures and formulations of multi-modal fusion have achieved great success in classification-based problems [36]. Unfortunately, they do not easily translate to a regression-based problem such as plethysmography. Nonetheless, several works have attempted fusion such as RGB+Mid-Infrared (Thermal) [61] and RGB+Near-Infrared (NIR) [58].

2.1.3 Contributions and Scope

The goal is to use camera and 77 GHz radar fusion to create a higher performing and more equitable remote plethysmograph technique. We make three specific contributions:

Contribution A: Existing unimodal remote plethysmography methods show a Pareto trade-

off between performance and fairness. We show that through carefully chosen modalities, multi-modal fusion can improve the Pareto frontier for this tradeoff, enabling improvements in both performance and fairness.

Contribution B: To the best of our knowledge, we present the first RGB and radar plethysmograph multi-modal fusion technique incorporating bias cues as part of a novel discriminative learning framework.

Contribution C: We open-source the first and largest multi-modal remote plethysmography dataset with representation across skin tones and other demographic markers.

Our code, dataset, and hardware tutorial may be accessed from https://github.com/UCLA-VMG/EquiPleth.

Scope We aim to establish the importance of multi-modal fusion towards achieving high performing and fair algorithms for vital sign sensing. We do not consider or incorporate other confounding effects, such as motion, resolution, and compression. These are relevant engineering aspects that need to be considered when looking at deployability of the technology. In this work, however, we constrain our focus on the analysis and mitigation of skin tone bias (as opposed to other kinds of biases).

2.1.4 Problem Formulation

A human subject is sensed by a non-contact sensing method, f, that processes data from M modalities (e.g. images, radar matrices, etc.). For the i-th training example, let this data be represented as an irregular list $\mathbf{x}_i = [\mathbf{v}_i^1, \mathbf{v}_i^2, ..., \mathbf{v}_i^M]$, with each $\mathbf{v}_i^m \in \mathbb{R}^{N_m}$. This subject has a ground truth and predicted plethysmograph signal, denoted as vectors $\mathbf{y}_i \in \mathbb{R}^K$ and $f(\mathbf{x}_i) = \hat{\mathbf{y}}_i \in \mathbb{R}^K$, respectively. The corresponding heart rate of the plethysmograph signal is $\mathbf{h}_i \in \mathbb{R}^+$. The sensor data, plethysmograph signal, and heart rate are drawn from distributions $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{H}$ respectively. Unique to this work's formulation (as compared to

others in iPPG literature), the subject also has a protected attribute $\mathbf{a}_i \in A$. This attribute describes skin tone categories, such that $A = \{\text{light}, \text{medium}, \text{dark}\}$. Subjects are labeled according to a modified Fitzpatrick skin tone scale [10] as light for I/II, medium for III/IV, and dark for V/VI. For brevity, the sample indexing is dropped here onward; $\mathbf{x}$, $\mathbf{y}$, $\mathbf{h}$, and $\mathbf{a}$ denote $\mathbf{x}_i$, $\mathbf{y}_i$, $\mathbf{h_i}$ and $\mathbf{a_i}$ respectively.

2.2 iPPG Signal Strength and Skintone Dependency

This section initially goes over the operation of pulse oximeters and how it relates to the signal strength in classical iPPG methods. We then show how skintone bias can manifest from the signal strength equations in iPPG. We end with our implementation of iPPG used for experiments and generating results.

2.2.1 Conventional Pulse Oximetry

The pulse oximeter is one of the primary tools doctor's use to asses two important vital signs: heart-rate (HR) and oxygen saturation (SpO2). In this work we focus on estimating heart-rate, but a brief overview of estimating oxygen saturation is also given to show the large dependence on accurate heart-rate and photoplethysmography measurements.

The classical approach to measuring heart-rate and SpO2 utilizes spectrophotometry, where light is passed through a sample and a detector measures the absorption or transmittance of light. The two main Hemoglobin (Hb) molecules of interest are oxygenated-hemoglobin and deoxygenated-hemoglobin. The two can be separated in a sample by mea-

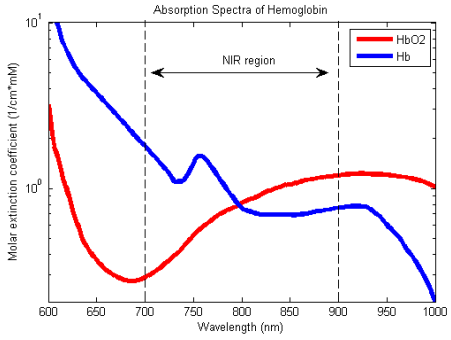

Figure 2.1: Hb Absorption Spectras [69].

suring the difference in their absorption spectras as in figure 2.1.

The origin of the pulse picked up by pulse oximeter can be best understood through the Beer-Lambert law [50]. The absorbance of light, A, is the product of molar attenuation coefficient (ϵ), optical length (l) and molar concentration (c) of the attenuating substance. The product of the molar attenuation coefficient and molar concentration is also known as the attenuation constant (μ). It is a logarithmic measure and the intensity of light that is expected is:

$$I(l) = I_o e^{-\epsilon c l} \tag{2.1}$$

In the case of human tissue, the attenuation is caused by a linear combination of absorptions:

$$I_{tissue} = I_o e^{-l_{blood}(\mu_{Hb}+\mu_{HbO_2})-l_{tissue}\mu_{Tissue}} \tag{2.2}$$

Assume that l_{blood} is the length of light travel across blood filled capillaries at rest or in between blood volume pulses (BVP). When a blood volume pulse occurs, the light travel is extended[1] slightly due to dilation of the capillaries such that:

$$I^p_{tissue} = I_o e^{-(l_{blood}+\Delta l)(\mu_{Hb}+\mu_{HbO_2})-l_{tissue}\mu_{Tissue}} \tag{2.3}$$

The largest intensity change during a blood volume pulse cycle would be when the capillaries are least and most dilated, $I^v_{tissue} - I^p_{tissue}$. In general, the blood volume pulses cause the pulse oximeter to record a distinct waveform in time, $I_{tissue}(t)$, whose AC component is called the photoplethysmogram (PPG). A simple frequency analysis or peak detector can then extract the heart-rate.

Extracting the PPG waveform is crucial for correct estimation of SpO2. SpO2 estimation relies on accurate amplitude estimates in the PPG waveform. To determine the ratio of oxygenated hemoglobin to total hemoglobin in the blood, SpO2 methods employ forms of

[1]Various sources will show that the plethysmography signal originates due to either increase in light travel due to dilation of blood vessels or due to increases in concentration of blood [50]. In this thesis, we will use the former hypothesis for the origin of the PPG signal without loss of generality.

the Ratio of Ratios method [50] which was devised to eliminate dependency on various skin thicknesses, compositions, as well as blood volume pulse characteristics through observing the PPG signal at two different wavelengths of light.

2.2.2 iPPG Signal Strength

In contrast to pulse oximeters, where light travels from an emitter on one side of the finger through to a detector on the other side, camera-based PPG or iPPG relies on reflective PPG. In this scenario, the source of light is ambient light and our detector is effectively located on the same side of skin in which light entered. The iPPG signal strength can be determined through application of previous biorealistic rendering models [7, 39]. Our signal of interest is light that transmits through the epidermis and then experiences reflection through scattering in the dermis. Table 2.1 describes and summarizes the various symbols and notations used.

The epidermal transmission is modeled through Beer-Lambert Law like in Section 2.2.1,

$$T_{epi}(\lambda) = e^{-\mu_{a,epi}(\lambda)d_{epi}}, \tag{2.4}$$

where $\mu_{a,epi}(\lambda)$ is the absorption coefficient of the epidermis and d_{epi} is the thickness of the epidermis. In general, the absorption coefficient is a convex combination between skin tissue and melanin absorption coefficients,

$$\mu_{a,epi}(\lambda) = f_{mel}\mu_{a,mel}(\lambda) + (1 - f_{mel})\mu_{a,skin}(\lambda). \tag{2.5}$$

The skin tissue absorption coefficient, $\mu_{a,skin}(\lambda)$, is a known parameter, while $\mu_{a,mel}(\lambda)$ is a another convex combination,

$$\mu_{a,mel}(\lambda) = f_{eum}\mu_{a,eum}(\lambda) + (1 - f_{eum})\mu_{a,phm}(\lambda), \tag{2.6}$$

with absorption coefficients $\mu_{a,eum}(\lambda)$ and $\mu_{a,phm}(\lambda)$ for eumelanin and pheomelanin. Equations 2.4, 2.5, and 2.6 fully define epidermal transmission light which is dominated by absorption.

Table 2.1: **Notation used for light transport modeling of iPPG.** The left column shows the notation used and the right column describes the notation.

Notation	Description
$\mu_{a,eum}(\lambda)$	Eumelanin absorption coefficient
$\mu_{a,phm}(\lambda)$	Phomelanin absorption coefficient
$\mu_{a,der}(\lambda)$	Dermal absorption coefficient
$\mu_{s,der}(\lambda)$	Dermal scattering coefficient
$\mu_{a,bld}(\lambda)$	Blood absorption coefficient
$\mu_{a,ski}(\lambda)$	Skin absorption coefficient
$\mu_{oxy}(\lambda)$	Oxygenated hemoglobin abs. coefficient
$\mu_{dox}(\lambda)$	Deoxygenated hemoglobin abs. coefficient
f_{mel}	Skin melanin fraction
f_{eum}	Epidermal eumelanin fraction
f_{bld}	Dermal blood volume fraction
f_{oxy}	Oxygenated hemoglobin fraction in blood

The light that transmits through the epidermis enters the dermis. The reflection of light from the dermal layer follows the Kubelka-Munk theory for scattering-dependent reflection that predicts the proportion of light reflected as a function of wavelength as,

$$R_d(\lambda) = \frac{(1-\beta(\lambda))^2(e^{K(\lambda)d_{der}} - e^{-K(\lambda)d_{der}})}{(1+\beta(\lambda))^2 e^{K(\lambda)d_{der}} - (1-\beta(\lambda))^2 e^{-K(\lambda)d_{der}}}. \tag{2.7}$$

d_{der} is the dermal skin depth [7], $\beta(\lambda)$ and $K(\lambda)$ are deterministically related to $\mu_{a,der}(\lambda)$ (dermal absorption coefficient) and $\mu_{s,der}(\lambda)$ (reduced dermal scattering coefficient [8]), as given in [7, 39]. The absorption coefficient of the dermis and blood are:

$$\mu_{a,der}(\lambda) = f_{bld}\mu_{a,bld}(\lambda) + (1-f_{bld})\mu_{a,ski}(\lambda), \tag{2.8}$$

$$\mu_{a,bld}(\lambda) = f_{oxy}\mu_{oxy}(\lambda) + (1-f_{oxy})\mu_{dox}(\lambda). \tag{2.9}$$

This dermal reflection takes into account reflections due to skin composition as well as oxygenated and deoxygenated blood. Now, the reflection from the skin can be modeled through transmission through the epidermis and reflection of the dermis:

$$R(\lambda) = T^2{}_{epi}.R_d(\lambda). \tag{2.10}$$

where the transmission term is squared to take into account the light going through the epidermis twice: in and out. A camera would then sense the intensity,

$$I_c = \int_\lambda E(\lambda)S_c(\lambda)R(\lambda)d\lambda, \tag{2.11}$$

where $E(\lambda)$ is the illuminant's spectrum and $S_c(\lambda)$ is the spectral response for channel c.

As mentioned in Section 2.2.1, the plethysmography signal originates due to changes in the path length through blood during blood volume pulses. The change can also be parameterized through how the blood volume fraction, f_{bld} under the skin is modulated during the blood volume pulse. The signal strength of the iPPG signal for a particular camera channel can be approximated as the largest change in the captured intensity:

$$\Sigma_c = \Delta I_c \approx \left|\frac{\partial I_c}{\partial f_{bld}}\right| \cdot \Delta f_{bld}. \tag{2.12}$$

The contribution of the change in blood volume fraction to the camera intensity can then be seen as,

$$\Sigma_c \approx \left| \int_\lambda E(\lambda)S_c(\lambda)\frac{\partial R}{\partial f_{bld}}\Big|_{\overline{f_{bld}}} d\lambda \right| \cdot \Delta f_{bld}, \tag{2.13}$$

where $\overline{f_{bld}}$ is the average blood volume fraction($f_{bld} \approx 0.05$). This approximation holds true since f_{bld} only varies by a small amount around the average blood volume fraction value, $\overline{f_{bld}}$.

This PPG signal rides on top of a DC value given by the average skin tone color:

$$\Gamma_c = \int_\lambda E(\lambda)S_c(\lambda)R(\lambda)\Big|_{\overline{f_{bl}}} d\lambda. \tag{2.14}$$

Notice that Σ_c and Γ_c are both dependent on f_{mel} due to the dermal reflection term, $R_d(\lambda)$.

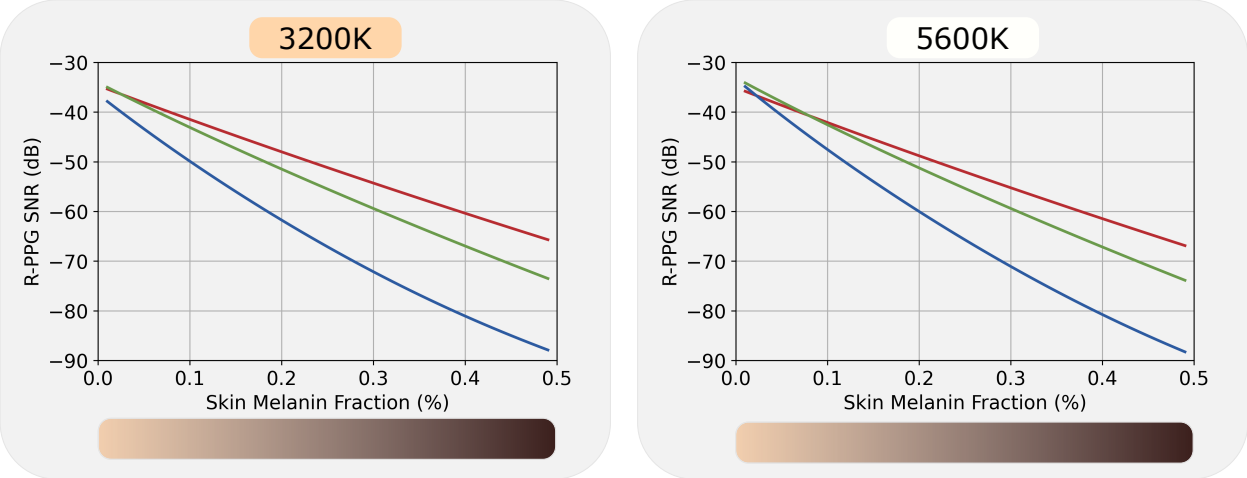

Figure 2.2: **The iPPG per-pixel SNR drastically worsens with increasing skin melanin fraction.** Using biophysical skin reflectance models, we estimate the iPPG signal strength as a function of skin melanin fraction. Along with the monotonic decrease in signal strength with melanin fraction, we also note finer trend differences between the SNR for the red, green and blue channels, as well as dependence on the spectral properties of the light source.

2.2.3 Skin Tone Bias in iPPG

Skintone bias arises due to the interplay between the signal strength of the iPPG signal as a function of melanin and imaging noise. Imaging noise occurs primarily due to the Poisson randomly distributed photon arrival processes as well as thermal noise from circuity and the quantization noise due to ADCs [37]. Assuming the pixel is not saturated past its maximum value, its noise can be written as :

$$\sigma_{pixel}^2 = \frac{\Phi t}{g^2} + \frac{\sigma_r^2}{g^2} + \sigma_q^2, \tag{2.15}$$

where Φ is the radiant power of light collected, t is the exposure time, g is the sensor gain, and σ_r and σ_q are camera noise parameters for thermal and quantization noise respectively. The same noise model that is used for calculating the signal to noise ratio (SNR) of a pixel's

value can also be used to estimate the iPPG signal's SNR as follows:

$$SNR_c = \frac{\Sigma_c t}{g\sqrt{\frac{\Gamma_c t}{g^2} + \frac{\sigma_r^2}{g^2} + \sigma_q^2}}. \tag{2.16}$$

Here, radiant power, Φ, is the average skin tone color Γ_c for a particular color channel, c.

Figure 2.2 shows the iPPG per-pixel SNR plots for the three camera color channels, across two lighting conditions (indicated by the light source 'temperature'). We use average camera response functions $\mathbf{S_c}(\boldsymbol{\lambda})$ to identify responsiveness of each of the channels to incident light, as well as exemplar camera noise parameters. Specifically, we used $\sigma_r = 140.7$, $\sigma_q = 0.08$ and $g = 1.06$. These parameters are representative and are calculated for a typical cell phone camera. Their specific values do not affect the trends and hence are not of primary importance. We note that the SNR monotonically decays with increasing skin melanin fraction. This trend is consistent across color channels and scene lighting conditions. That is, given a fixed scene and camera configuration, the underlying physical signal is poorer for dark skin tones.

2.2.4 Implementation

The RGB measurements $\mathbf{v}^1 \in [0,1]^{T\times C\times H\times W}$ are tensor-valued. Here, T is the numbers of frames, C is the number of image channels (3), and H and W are the image height and width respectively. We use the Physnet spatio-temporal network by [111] as one of the inputs for fusion. The PPG network $g_1(\cdot)$ estimates the PPG waveform $\hat{\mathbf{y}}_{RGB}$ from video inputs $\mathbf{v}^1$ with $T = 64$ frame samples as input.

The PPG network is updated using a negative Pearson loss between the estimated waveform $\hat{\mathbf{y}}$ and the ground truth waveform $\mathbf{y}$ to enforce waveform reconstruction, similar to previous work [111]. This is given by,

$$
\begin{aligned}
L_P(\mathbf{y}, \hat{\mathbf{y}}_{RGB}) =& 1 - \frac{1}{\sqrt{a_1 \times a_2}} \left(N \sum_{i=1}^{N} \mathbf{y}_i \hat{\mathbf{y}}_i - \sum_{i=1}^{N} \mathbf{y}_i \sum_{i=1}^{N} \hat{\mathbf{y}}_i \right), \\
a_1 =& \left(N \sum_{i=1}^{N} \mathbf{y}_i^2 - \left(\sum_{i=1}^{N} \mathbf{y}_i \right)^2 \right) \\
a_2 =& \left(N \sum_{i=1}^{N} (\hat{\mathbf{y}}_i)^2 - \left(\sum_{i=1}^{N} \hat{\mathbf{y}}_i \right)^2 \right),
\end{aligned}
\tag{2.17}
$$

where N is the length of $\mathbf{y}$ and $\hat{\mathbf{y}}_{RGB}$. The overall loss function used, L_{RGB} is given by,

$$
L_{RGB} = L_P(\mathbf{y}, \hat{\mathbf{y}}_{RGB}). \tag{2.18}
$$

2.3 Radar for Plethysmography

FMCW radar emits and receives (reflected) chirps, which are linearly frequency modulated electromagnetic (EM) waves, enabling the estimation of the distance travelled by the chirp before reflection. Aardal et al. [1] showed that one contributor to detecting a heartbeat is a large reflection at the air/skin interface and experimentally demonstrated that heartbeat detection is primarily reliant on physical displacements of the chest. Therefore, we assume radar is not directly affected by skin tone.

2.3.1 Radar Theory

The transmitted and received signal, $s(t)$ and $u(t)$ respectively, can be modeled as:

$$
s(t) = A_s cos(2\pi f_c t + \pi k t^2), 0 < t < T_c. \tag{2.19}
$$

$$
u(t) = A_u cos(2\pi f_c (t - t_d) + \pi k (t - t_d)^2), t_d < t < T_c. \tag{2.20}
$$

where k is the frequency slope (the rate of change of frequency of the chirp), f_c is the starting frequency of the chirp, T_c is the duration of the chirp transmission, and t_d is the time delay

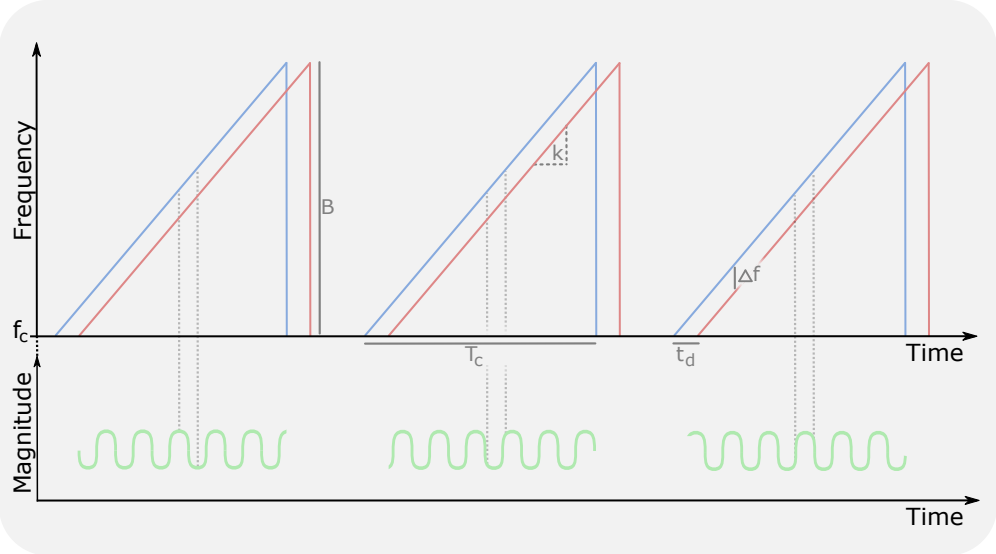

Figure 2.3: **A FMCW Chirp Sequence.** The blue and red signal are the transmitted and received chirps plotted with their frequency content as a function of time. The green signal denotes the mixed signal whose phase changes while the frequency remains relatively constant.

between the start of transmission and initial reception of the reflected wave. Then, the bandwidth of the signal, the difference between the maximum and minimum frequencies, is given by:

$$B = f_{max} - f_{min} = (f_c + kT_c) - f_c = kT_c, \tag{2.21}$$

and the time delay is proportional to the round trip distance, $t_d = \frac{2R}{c}$, where R and c are the range of the object and speed of light respectively. Figure 2.3 indicates these values on a FMCW chirp sequence.

Upon receiving a reflected chirp, the radar mixes the received chirp with the still transmitting signal. The mixed signal is proportional to $s(t) \cdot u(t)$ and contains 2 components: a beat signal component with a frequency equal to the frequency difference of $s(t)$ and $r(t)$, $\Delta f = kt_d$, and a high frequency component situated near $4\pi f_c$. The higher frequency component is filtered out by a low pass filter (LPF) to prevent aliasing, generating $m(t)$.

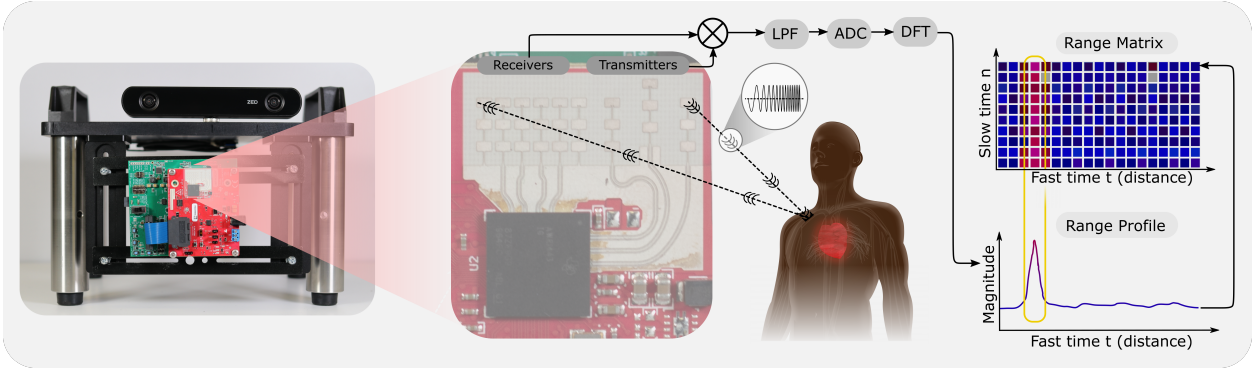

Figure 2.4: **We use a 77 Ghz FMCW radar setup for non-contact radar plethysmography.** Chirp signals are bounced off the subject's chest in order to capture subtle motion. By exploiting the dependency of the phase on the distance of flight, we are able to measure this motion.

Concretely, the radar samples in-phase and quadrature (IQ) components such that:

$$m(t) \propto \text{LPF}[s_I(t) \cdot u(t)] + j \ \text{LPF}[s_Q(t) \cdot u(t)], \tag{2.22}$$

where $s_I(t) \cdot u(t)$ and $s_Q(t) \cdot u(t)$ denote the in-phase and quadrature components respectively. The in-phase component is comprised of the transmitted signal $s(t) = s_I(t)$ multiplied with the received signal $u(t)$. The quadrature component is derived by multiplying the received signal $u(t)$ with a copy of the transmitted signal shifted by a phase of $-90°$, $s_Q(t)$. However, IQ details are excluded for brevity and the following equations are represented with just the in-phase component. The IF signal, $m(t)$, can then be written as:

$$m(t) \propto A_m cos(2\pi f_c t_d + 2\pi(kt_d)t + \pi k t_d^2), t_d < t < T_c. \tag{2.23}$$

The $\pi k t_d^2$ term is several orders of magnitude smaller than the other terms and is thus negligible. Equation 2.23 can be rewritten into a more succinct form:

$$m(t) \propto A_m cos\left(\omega t + \phi\right), t_d < t < T_c. \tag{2.24}$$

$$\omega = 4\pi \frac{kR}{c}, \quad \phi = 4\pi \frac{R}{\lambda}. \tag{2.25}$$

The phase and frequency of the resulting signal depend on the range R and can be extracted through a discrete Fourier transform (DFT) of the signal after passing it through the analog to digital converter (ADC). The frequency term, $\omega = 2\pi\Delta f$, provides the range through the following relation $R = c\frac{\Delta f}{2k}$. Therefore, the maximum unambiguous range an object can be placed from the radar follows from the Nyquist sampling theorem and ADC sampling rate f_s, to give $R_{max} = c\frac{f_s}{4k}$. The range resolution is $c\frac{f_s}{4kN}$, where N is the number of ADC samples. The phase term ϕ is inversely proportional to the wavelength of the radar, $\lambda = \frac{c}{f_c}$.

The range of an object can be parameterized as $R(t) = R_o + r(t)$, where $r(t)$ models changes due to vibrations (for example, heart beat) around the average range R_o. To extract a heart rate, $r(t)$ needs to be sampled with multiple chirps. Note that the frequency term cannot be used to extract the sub millimeter displacement of a heart beat; the frequency resolution is on the order of centimeters. Instead, we use the highly sensitive phase to determine the oscillations of $r(t)$. The reader may note that in reality the phase extracted from the digital signal would be wrapped between $[-\pi, \pi]$. This can be solved using a standard phase unwrapping algorithm.

Practically, for the transmission and processing of the nth chirp, the ADC samples, $m_n[i]$, are converted into the frequency domain or a single range profile, $M_n[f]$. To observe the periodic movements of the body due to a heart beat, we sample the range profile in time to construct a range matrix, $\mathbf{M} = [M_1[f], M_2[f], ...M_n[f]]^T$, such that a range bin is indexed by a fast time or range axis and a given chirp is indexed by the slow time axis (fast time refers to a chirp's ADC samples, while slow time refers to chirp samples). Figure 2.4 shows the processing pipeline as well as the range matrix and a range profile where the amplitude strengths determine a person's location.

2.3.2 Implementation

To extract the heart rate, a frequency analysis can be performed on the phase of the central range bin (the range bin with the maximum power occupancy)[6]. However, as [56, 115] note,

the phase is very sensitive to movement and body background reflection which can diminish the signal or cause interference due to the harmonics of the respiratory rate. To mitigate this, we were inspired by [115] work in UWB radar on extracting fine-grained respiratory signals by learning from raw IQ data. We employ a deep learning approach to estimate $g_2(\cdot)$, a mapping from $\mathbf{v}^2$ to $\hat{\mathbf{y}}_{Radar}$. Specifically, $\mathbf{v}^2$ consists of a window of range profiles around the central bin that are then processed by a series of 1D CNNs. The output is an estimate of the ground truth PPG signal even though the radar is only measuring vibrations. During training, we apply a data augmentation technique of rotating the complex components of the data [115] and a loss consisting of a Negative Pearson Loss L_P and a Signal to Noise Ratio (SNR) loss. The SNR loss is defined as follows:

$$L_{SNR}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{\int_{f_0-w}^{f_0+w} |\hat{\mathbf{Y}}(f)|^2 \, df}{\int_{-\infty}^{f_0-w} |\hat{\mathbf{Y}}(f)|^2 \, df + \int_{f_0+w}^{\infty} |\hat{\mathbf{Y}}(f)|^2 \, df}, \tag{2.26}$$

$$f_0 = \arg\max_f \mathbf{Y}(f),$$

where $\mathbf{Y}(f)$ and $\hat{\mathbf{Y}}(f)$ are the respective Fourier transforms of $\mathbf{y}$ and $\hat{\mathbf{y}}$ and w is the chosen window size.

The overall loss function used is given by,

$$L_{Radar}(\mathbf{y}, \hat{\mathbf{y}}_{Radar}) = L_P(\mathbf{y}, \hat{\mathbf{y}}_{Radar}) + \lambda_{Radar} L_{SNR}(\mathbf{y}, \hat{\mathbf{y}}_{Radar}). \tag{2.27}$$

2.4 Multimodal Fusion of Camera and Radar

2.4.1 Resisting SNR Bias through Sensor Fusion

In this section, we prove that sensor fusion, even if some of the individual sensors are biased, can lead to overall resistance of bias. Consider a temporal signal $\mathbf{s} \in \mathbb{R}^n$. Without loss of generality, we assume normalized signals such that $\|\mathbf{s}\|_2 = 1$. Our sensing setup consists

of modalities $m \in \{M_1, M_2\}$. Each modality m captures a noisy observation of the signal $\mathbf{y}_m \in \mathbb{R}^n$.

An additional property of a sample is its attribute $\mathbf{a} \in \{\text{light}, \text{dark}\}$. The observed signal is then given by $\mathbf{y}_m^{\mathbf{a}}$. Both modalities behave differently for different attributes. For example, a modality may have different expected performance for samples with different attributes.

Let the theoretical Signal to Noise Ratio (SNR) of a signal $\mathbf{y}_m^{\mathbf{a}}$ be denoted by the operator $\mathcal{S}(\mathbf{y}_m^{\mathbf{a}})$. We use the SNR to define modality-wise fundamental performance and fairness. Without loss of generality, we assume that the modality M_1 is better performing as compared to the modality M_2 (in our practical setting, M_1 would correspond to the RGB modality while M_2 would correspond to the radar modality). That is,

$$\mathop{\mathbb{E}}_{\mathbf{a},\, \mathbf{y}_{M_1}^{\mathbf{a}}} \left[\mathcal{S}(\mathbf{y}_{M_1}^{\mathbf{a}})\right] > \mathop{\mathbb{E}}_{\mathbf{a},\, \mathbf{y}_{M_2}^{\mathbf{a}}} \left[\mathcal{S}(\mathbf{y}_{M_2}^{\mathbf{a}})\right]. \tag{2.28}$$

Additionally, we note that according to our required conditions, the modality M_1 is biased in terms of attribute $\mathbf{a}$, while the modality M_2 is unbiased. That is,

$$\begin{aligned} \left| \mathop{\mathbb{E}}_{\mathbf{y}_{M_1}^{\text{Light}}} \left[\mathcal{S}(\mathbf{y}_{M_1}^{\text{Light}})\right] - \mathop{\mathbb{E}}_{\mathbf{y}_{M_1}^{\text{Dark}}} \left[\mathcal{S}(\mathbf{y}_{M_1}^{\text{Dark}})\right] \right| > \epsilon, \text{ and} \\ \left| \mathop{\mathbb{E}}_{\mathbf{y}_{M_2}^{\text{Light}}} \left[\mathcal{S}(\mathbf{y}_{M_2}^{\text{Light}})\right] - \mathop{\mathbb{E}}_{\mathbf{y}_{M_2}^{\text{Dark}}} \left[\mathcal{S}(\mathbf{y}_{M_2}^{\text{Dark}})\right] \right| < \epsilon, \end{aligned} \tag{2.29}$$

for some suitable small ϵ. We also assume without loss of generality that the 'dark' attribute is the worse performing attribute group on average. That is,

$$\mathop{\mathbb{E}}_{\mathbf{y}_m^{\text{Light}}} \left[\mathcal{S}(\mathbf{y}_m^{\text{Light}})\right] \geq \mathop{\mathbb{E}}_{\mathbf{y}_m^{\text{Dark}}} \left[\mathcal{S}(\mathbf{y}_m^{\text{Dark}})\right], \forall\, m \in \{M_1, M_2\}. \tag{2.30}$$

We wish to characterize the improvement in signal quality arising as a result of combining observations from the two modalities M_1 and M_2. That is, we wish to understand the performance and bias properties of a combined measurement signal $\mathbf{y}_{\text{comb}}^{\mathbf{a}}$ that is optimal in the SNR sense, as follows:

$$\begin{aligned} \mathcal{C}^* &= \arg\max_{\mathcal{C}} \mathbb{E}\left[\mathcal{S}(\mathcal{C}(\mathbf{y}_{M_1}^{\mathbf{a}}, \mathbf{y}_{M_2}^{\mathbf{a}}))\right], \forall\, a. \\ \mathbf{y}_{\text{comb}}^{\mathbf{a}} &= \mathcal{C}^*(\mathbf{y}_{M_1}^{\mathbf{a}}, \mathbf{y}_{M_2}^{\mathbf{a}}), \forall\, \mathbf{a}. \end{aligned} \tag{2.31}$$

Here, $\mathcal{C}(\cdot,\cdot)$ is an appropriately chosen combining operator.

We wish to quantify the benefit of multi-modal combination in terms of the gains obtained over the modality M_1. The quality gain factor $Q_{\mathbf{a}}$ is therefore given by,

$$Q_{\mathbf{a}} = \frac{\mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{\text{comb}})\right]}{\mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{M_1})\right]}. \tag{2.32}$$

We have established the required terminology for our result.

Theorem 1: Let Q_{Light} and Q_{Dark} be the quality gain factors for the light and dark attributes respectively. Then, optimally combining observations $\mathbf{y}^{\mathbf{a}}_{M_1}$ from a better performing but biased modality M_1 and $\mathbf{y}^{\mathbf{a}}_{M_2}$ from a worse performing but unbiased modality M_2, where $\mathbf{a} \in \{\text{light}, \text{dark}\}$, ensures,

$$Q_{\text{Dark}} > Q_{\text{Light}}. \tag{2.33}$$

That is, the worse performing attribute sees a greater relative gain in signal strength.

Proof: *The optimal combination method for two signals with known SNRs is the Maximal Ratio Combining [66]. The resulting expected SNR is given by,*

$$\mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{comb})\right] = \mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{M_1})\right] + \mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{M_2})\right], \forall\, \mathbf{a}. \tag{2.34}$$

Then, the quality gain factor is given by,

$$\begin{aligned} Q_{\mathbf{a}} &= \frac{\mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{M_1})\right] + \mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{M_2})\right]}{\mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{M_1})\right]} \\ &= 1 + \frac{\mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{M_2})\right]}{\mathbb{E}\left[\mathcal{S}(\mathbf{y}^{\mathbf{a}}_{M_1})\right]}. \end{aligned} \tag{2.35}$$

Then, from Equations 2.29 and 2.35, we can infer that,

$$Q_{Dark} > Q_{Light}. \tag{2.36}$$

This completes the proof. ■

We therefore establish that fusion with a worse performing but less biased modality is in fact beneficial in terms of bias mitigation.

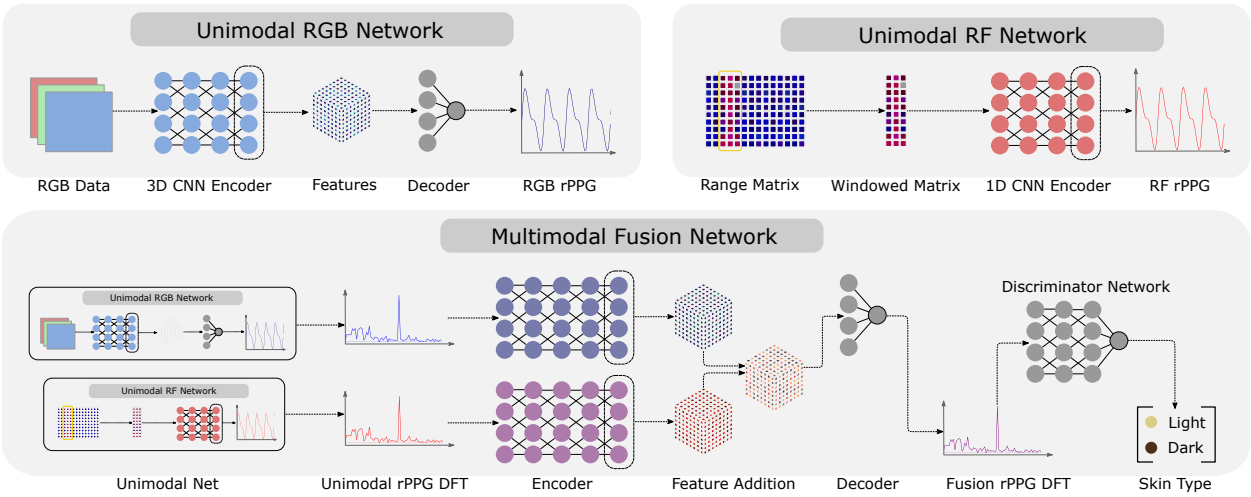

Figure 2.5: **The proposed approach uses a novel adversarial discriminative training-based approach for skin tone debiasing in the modality fusion module.** The fusion network operates in the frequency domain using an alternating waveform reconstruction and adversarial losses.

2.4.2 Overall Inferences

We now summarize the inferences from the theory section, which serve as a motivation for our multi-modal fusion hypothesis for bias alleviation.

1. The RGB modality is biased against darker skin tone samples. This arises fundamentally as a result of poor Signal to Noise Ratio.

2. Combining the RGB modality with a poorer but unbiased modality results in larger improvements for the darker skin tone samples as compared to the lighter skin tone samples.

2.4.3 Implementation of Fusing RGB Camera and Radar for Plethysmography

In the previous section, we discussed bias in the context of plethysmography. We now discuss the specific fusion of camera and 77 Ghz radar to resist bias. Referring to notation from

Section 2.1.4, let $\mathbf{v}^1$ denote an RGB modality measurement, and let $\mathbf{v}^2$ denote a FMCW radar modality measurement. Then, $\mathbf{x} = [\mathbf{v}^1, \mathbf{v}^2]$, our overall measurement. Our goal is to learn a robust functional mapping f from the measurement space to the plethysmograph space,

$$\begin{aligned} \mathbf{y} &= f(\mathbf{x}) \\ &= f([\mathbf{v}^1, \mathbf{v}^2]). \end{aligned} \tag{2.37}$$

Figure 2.5 describes our overall pipeline. We use a late fusion parameterization for the mapping f. The unimodal plethysmography signal estimates are first obtained. A fusion architecture combines these to obtain the final fused plethysmography estimate. That is,

$$\mathbf{y} = f([\mathbf{v}^1, \mathbf{v}^2]) = g_f(g_1(\mathbf{v}^1), g_2(\mathbf{v}^2)), \tag{2.38}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are the modality specific estimators for RGB and radar respectively. The function $g_f(\cdot)$ is the late fusion model. We describe each of these components in the following text. Specific implementation, training configuration and model architecture details may be found in the included supplementary code.

2.5 MMFAIR Dataset

To conduct evaluations, we recruited 91 volunteers to participate in this study. The dataset contains 28 light, 49 medium, and 14 dark skin tone volunteers. The skin tones were labeled according to the Fitzpatrick scale [75]. The volunteers are primarily from a college background, with representation between genders. 6 recordings were taken for each volunteer. A recording lasts 30 seconds and consists of a RGB video and radar IQ data. Environmental factors such as lighting variations are left in the dataset to enable more robust skin tone bias trend analysis. Figure 2.6 describes our data collection setup in detail. The data was taken using one camera from a ZED stereo camera and a Texas Instruments AWR1443 RF development board at a distance of 0.5-1m. The entire dataset consists of over 550 recordings and 18,000 unique beats of the hearts. The ground-truth plethysmograph signal was

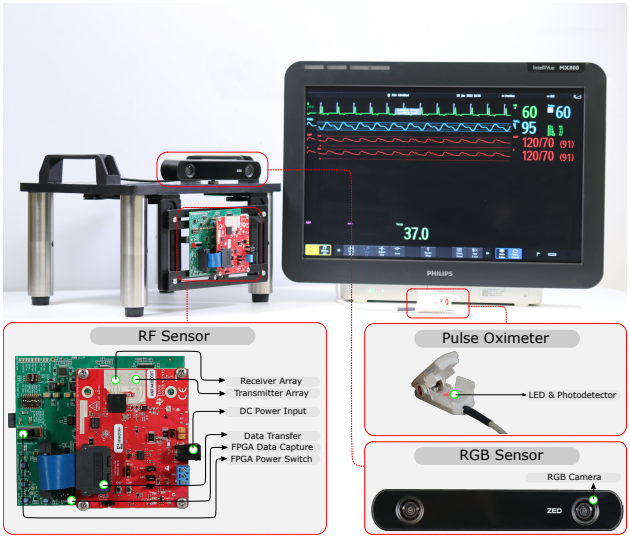

Figure 2.6: **A mobile multi-modal sensing platform was deployed to collect our remote plethysmography dataset. The parts list and reference designs may be found at https://github.com/UCLA-VMG/EquiPleth.**

acquired using an IntelliVue MX800 clinical grade patient monitor. An external computer is used to synchronize the capture and storage of the two modalities and the ground truth. The entire data collection setup is mobile, enabling large scale data capture irrespective of location.

The RGB camera was used with default factory settings at 30 fps. Videos were processed to 128x128 crops using a MTCNN [113] to locate facial regions. The FMCW radar was set to emit 120 chirps per second with a frequency slope of 60 MHz/µs, starting frequency of 77 Ghz, bandwidth of 3.720 Ghz, and sampling rate of 5 Mhz using a single transmitter-receiver pair. The sampled IQ data was processed into a range matrix and data related to regions of interest extracted within a 25 cm window.

We note that several previous works [10, 24, 111] evaluate metrics over 30 second windows of the estimated waveform. In this work, we choose to evaluate metrics over 10 second widows with a stride of 128 samples (4.27 s) instead. This provides a more realistic setting for analysis (with lower latency), in addition to better highlighting the effects related to bias

and fairness. All evaluations and metrics are evaluated over six independent data splits. For each split, we include 40 participants in the traning set, 12 participants in the validation set and all remaining participants in the test set. Note that to enable control over the participant skin tone representation in each split, we ensure that the train and validation set have equal number of participants from Fitzpatrick groups I, II, III and groups IV, V, VI.

2.6 Results

2.6.1 Evaluation Metrics

Performance To assess general performance, we use heart rate prediction metrics as in previous iPPG work [63, 111]. Table 2.2 summarizes metrics for heart-rate accuracy. Several results are reported using the absolute percent error (APE) metric due to the Association for Advancement of Medical Instruments (AAMI) defining a threshold heart-rate error as no greater than 10% relative error [59], as well as its prevalence in evaluating heart monitors and physical monitoring devices [9, 62]. We note that all performance measures are evaluated over the entire testing dataset, consisting of samples from the light, medium and dark skin tone groups.

Fairness To evaluate fairness, we adopt standard metrics from the fairness community [92]. A plethysmograph method can be considered fair if the outcome for the general population is the same as for a sub-population with a given attribute, **a**. In our case, to facilitate experimental analysis, we evaluate fairness in terms of similarity of outcome between the light and dark skin tone groups. The following definitions introduce metrics for quantifying fairness and are summarized in Table 2.2. The threshold test [14] is a notion of sufficiency of fair performance. The test shows the proportion of test samples in the light (L) and dark (D) categories that fall outside of the threshold defined by AAMI. In order to measure bias through performance metrics, we follow [19] by evaluating a performance metric, D on

	Metric	Expression
Performance	Mean Absolute Error (MAE) ↓	$\frac{1}{N}\sum_{i=1}^{N}\lvert h_i - \hat{h}_i\rvert$
	Root Mean Square Error (RMSE) ↓	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(h_i - \hat{h}_i)^2}$
	Mean Absolute Error (MAPE) ↓	$\frac{1}{N}\sum_{i=1}^{N}\frac{\lvert h_i - \hat{h}_i\rvert}{h_i}$
	Pearson Corr. Coefficient (R) ↑	$\frac{\sum_{i=1}^{N}(\hat{h}_i - \mu_{\hat{h}})(h_i - \mu_h)}{\sum_{i=1}^{N}(\hat{h}_i - \mu_{\hat{h}})^2 \sum_{i=1}^{N}(h_i - \mu_h)2}$
Fairness	Threshold Test ↓	$\mathbb{P}_{\mathbf{X}}[APE(\hat{\mathbf{H}}) > 10\% \mid \mathbf{a}]\ ,\forall\ \mathbf{a} \in \{\mathrm{L}, \mathrm{D}\}$
	Performance Bias ↓	$\lvert D_{\mathbf{a}=\mathrm{D}}(\mathbf{H}, \hat{\mathbf{H}}) - D_{\mathbf{a}=\mathrm{L}}(\mathbf{H}, \hat{\mathbf{H}})\rvert$

Table 2.2: **We use multiple performance and fairness metrics for evaluation.** Performance and fairness are not necessarily correlated properties and requires both for full comparison.

attribute groups and taking a pairwise difference.

2.6.2 Qualitative Performance

Analyzing the *generated photoplethysmograph waveforms* allows for visual inspection of the estimated waveforms. The heart-rate estimates are frequency estimates obtained out of these waveforms. Fig. 2.7 shows estimated plethysmograph waveforms for randomly chosen samples from the light and dark groups respectively. We compare the estimated waveforms from the best RGB and radar unimodal models with our fusion-based model. For the RGB only modality, a degradation in signal quality is observed from the light to the dark skin tones. This highlights the performance bias in the modality. For the radar only modality, we note across the board noisy waveforms. However, there is minimal bias across skin tones.

Table 2.3: **Across baselines spanning the radar and camera modalities, the proposed fusion model shows performance and fairness improvements over the unimodal iPPG modality.** The performance metrics measure the average performance across the entire dataset. The pairwise difference between light and dark groups being bracketed and the sign shows direction of bias - ideally the absolute value of this bias should be low. The fairness threshold test measures the percent of the light and dark populations failing the AAMI standard. The best performing numbers are bolded between the fusion, RF, and PhysNet backbone.

	Performance (Fairness)				Fairness
Method	MAE ↓ (↓)	MAPE ↓ (↓)	RMSE ↓ (↓)	r ↑ (↓)	T-Test (APE %)
Green [91]	11.61 (0.23)	15.57% (1.09%)	16.56 (-0.97)	0.23 (-0.12)	42.9,52.9
ICA [68]	8.38 (4.42)	11.65% (6.19%)	14.03 (3.15)	0.41 (-0.36)	19.9,46.9
CHROM [27]	7.45 (4.97)	10.57% (6.81%)	13.38 (4.17)	0.46 (-0.38)	14.5,42.6
BCG [12]	13.01 (-0.99)	15.03% (-1.05%)	20.66 (-1.25)	0.132 (0.05)	30.5,29.1
FFT-based RF [6]	13.51 (2.25)	1.66% (2.56%)	21.07 (2.47)	0.240 (-0.25)	39.1,44.5
PhysNet [111]	1.78 (2.22)	2.35% (2.63%)	5.26 (4.05)	0.91 (-0.25)	2.1,12.2
Our RF	2.18 (**0.51**)	3.05% (**0.69%**)	6.12 (**0.85**)	0.89 (-0.13)	5.1,8.4
Our Fusion	**1.12** (0.67)	**1.52%** (0.79%)	**3.42** (1.44)	**0.95** (**-0.10**)	**1.1**,**4.2**

Our fusion model infers the best qualitative waveforms across all three groups, while also reducing the skin tone bias that is evident in the RGB only modality.

Our second set of qualitative resources for analysis are Bland-Altman plots [35] (Figure 2.8). These plots visualize the distribution of the heart-rate estimation error versus the ground truth heart rates. The plots highlight that ground truth heart rates are distributed over a large range. In terms of heart rate estimation accuracy, we note good performance for the RGB only modality. However, visibly significant bias is present across skin tones, as visible from the error distributions. For the RF only modality, we note a larger spread in

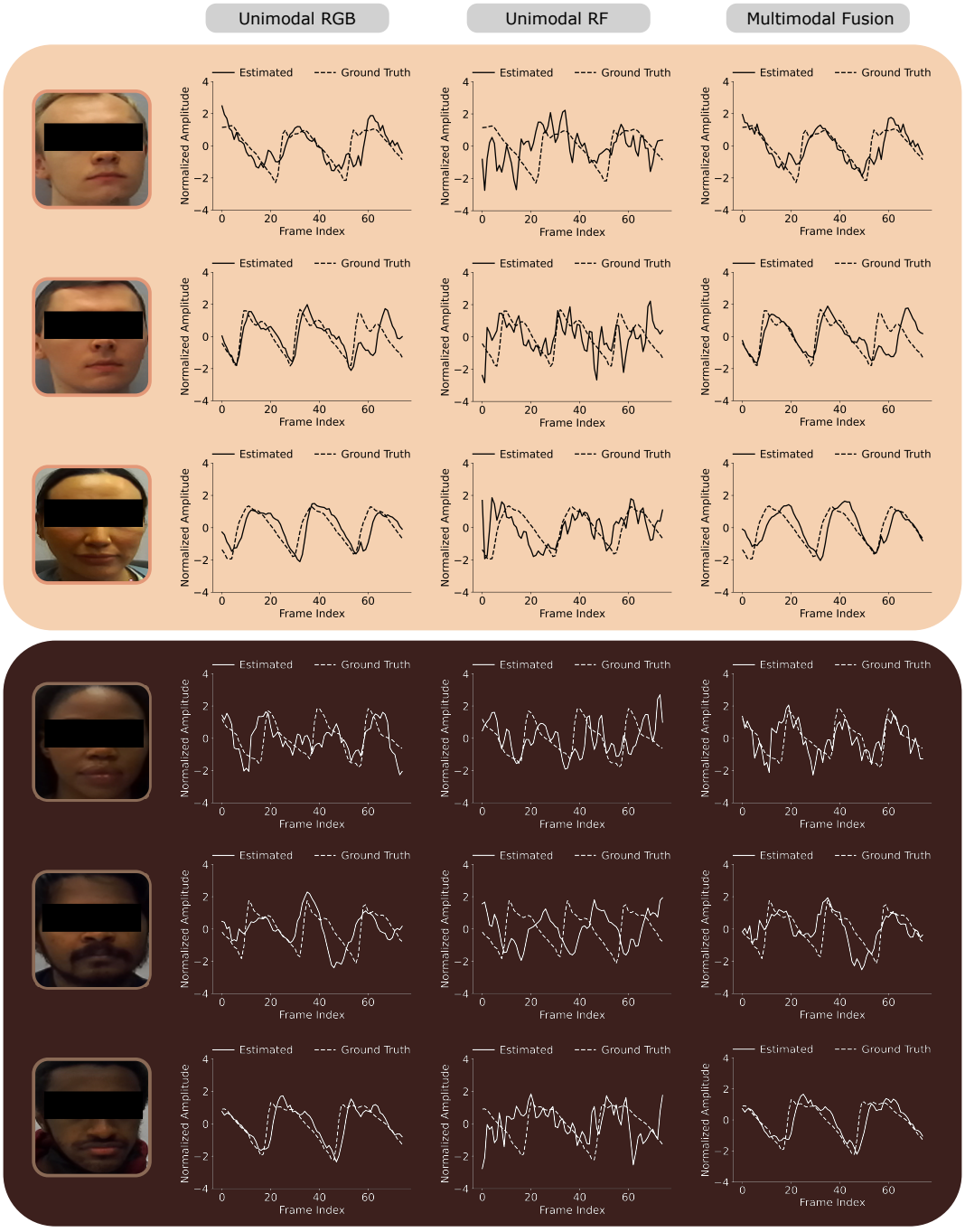

Figure 2.7: **Qualitative analysis of estimated waveforms indicates superior overall performance for the fusion model, with reduced group-wise bias.** We highlight a randomly chosen snippet of the plethysmograph waveform to highlight qualitative differences.

the distributions and a larger 1σ value. However, the skin tone bias is minimal. Again, our fusion method show significant performance and bias improvements, with lower and largely similar 1σ thresholds across skin tones.

2.6.2.1 Quantitative Performance

Table 2.3 highlights the performance measures for the various compared methods. The signal processing-based iPPG and radar-based methods reflect relatively poor performance on our dataset. We note that in general, the iPPG methods show better performance but the radar-based method is fairer.

The deep learning-based iPPG method [111] shows significant gains in performance over the signal processing methods. This highlights the benefit of data-driven nonlinear modeling. However, the performance bias between groups becomes much clearer. This reinforces the existence of fundamental bias in the iPPG modality. We also note that implementations for DeepPhys [24] and LSTM PhysNet [111] were tried on our dataset, however they did not converge during training. We believe this may be due to small misalignments between the ground-truth and video, that only the 3D-CNN PhysNet can handle.

Our deep learning-based RF method follows a similar trend. A significant improvement in performance is observed when compared to the signal processing-based RF method. However, the overall performance is lower that that of the deep learning-based iPPG method. Additionally, we note the significantly lower performance bias between the groups.

Our fusion method outperforms all previously listed methods. We see clear improvements in overall performance across all metrics. In addition, we also notice significant improvement in bias measures when compared to the RGB unimodal methods.

Notably, despite the lower performance of the radar-based method compared to the iPPG and fusion methods, the performance still remains high (with an average MAE of 2.18 beats per minute). Therefore, from the perspective of our evaluations on average performance, all

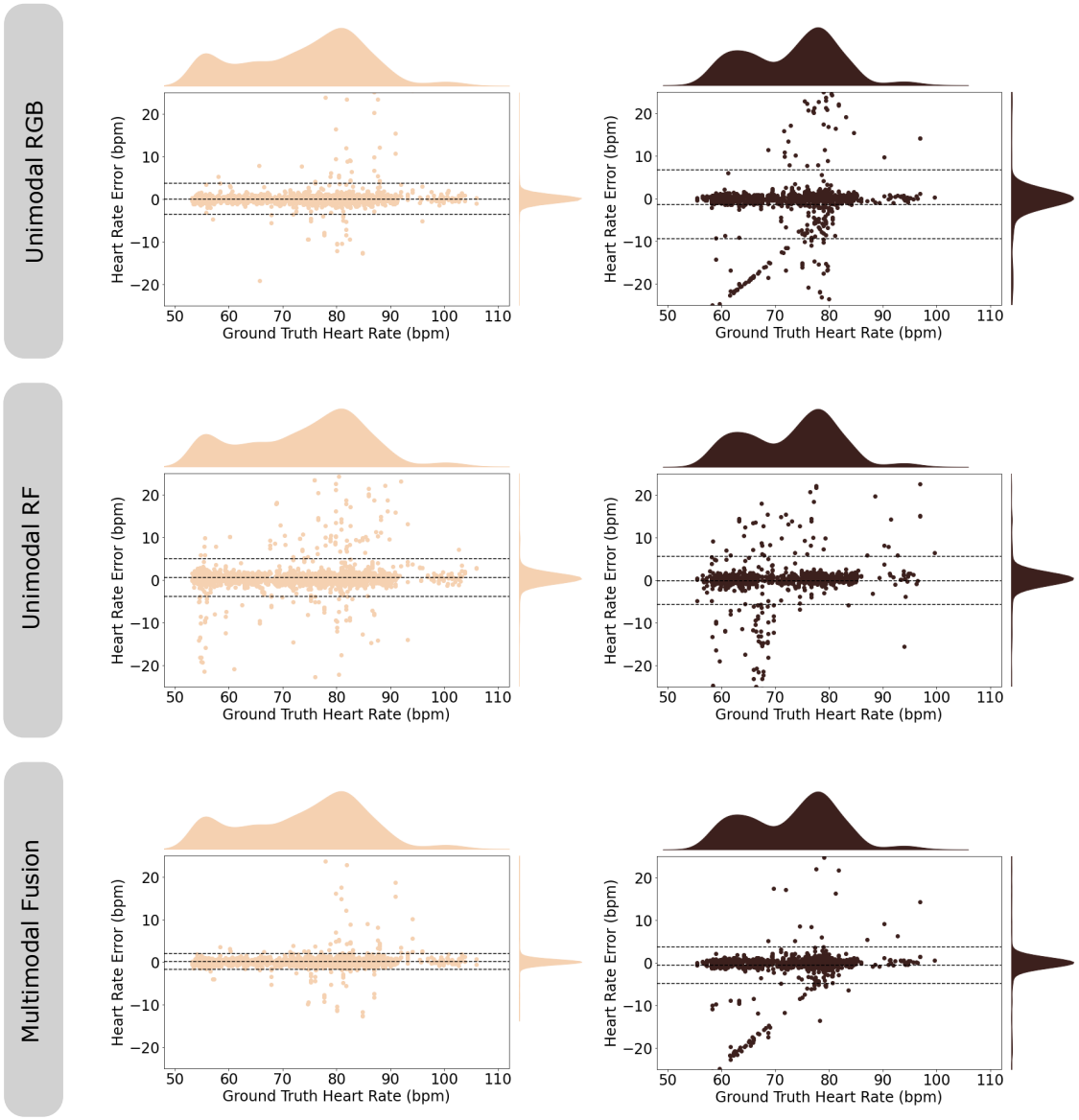

Figure 2.8: **Plotting the heart-rate estimation error versus the ground truth heart rate (Bland-Altman plots) emphasizes performance benefits of the proposed multi-modal fusion model.** Each plot highlights the distribution of the ground truth heart rates (top), distribution of the heart-rate estimation errors (right) and the plot of the estimation errors versus the ground truth heart rates.

three methods (iPPG, radar and fusion) show acceptably high average performance. Readers may note that real-world factors such as motion, distance from sensor and so on may have relevant effects on the relative performance of the three methods, that are beyond the scope of evaluation of this work. Such a detailed study, which will determine a more general conclusion on the viability of the three methods, is deferred to future work.

2.6.2.2 Fairness

Table 2.3 highlights performance bias measures for the various compared methods. We note modality specific trends for these measures. For the RGB-only modality, we note that the T-Test values for the light and dark groups, as well as the performance bias measures, show significant bias. This is consistent with our theoretical analysis and expectations. On the other hand, for the radar modality, the T-Test values show low bias. This is also noted in the performance bias measures. Finally, we note that our fusion method achieves significantly better performance bias and T-Test scores compared to the RGB-only modality. Compared to the radar modality, we note slightly worse performance bias - the fusion method shows better performance bias in terms of some metrics, while the radar-only method shows better performance bias in terms of other metrics.

The observations from the previous subsections set up a tradeoff between performance and fairness. The proposed fusion method improves on both fronts over the iPPG method, but not compared to the radar-based method. Therefore, both these methods (fusion and radar-based) are potentially deployable candidates. We discuss this in some detail in Section 2.6.4.

2.6.3 Benefit of the Skin Tone Discriminative Loss

To establish the importance of our proposed skin tone-based adversarial discriminator, we compare against a naive fusion regime, trained only with our squared Pearson loss. Table 2.4 highlights the benefit of the skin tone discriminative loss. We note that the addition of the

Table 2.4: **An adversarial network for skin tone estimation is a novel contribution that helps obtain a more equitable plethysmograph estimator across skin tone.** When compared with a fusion network trained without the adversarial network, significant improvements are noted across all performance fairness measures, at a small cost in performance measures.

	MAE ↓ (↓)	MAPE ↓ (↓)	RMSE ↓ (↓)	r ↑ (↓)
Fusion w/o AN	**1.09** (0.98)	**1.47%** (1.28%)	**3.31** (2.56)	**0.964** (-0.146)
Fusion w/ AN	1.12 (**0.67**)	1.52% (**0.79%**)	3.42 (**1.44**)	0.953 (**-0.102**)

discriminative loss significantly reduced the skin tone bias of the fusion model, at a small performance cost. The model is encouraged to minimize the distributional gap between the estimated plethysmograph waveforms for light and dark skin tone groups.

2.6.4 Discussion

In summary, we fuse data from camera streams and 77 GHz radar to create a higher performing and more equitable plethysmography technique. To encourage reproducible research, we make the dataset and reference designs available. While we have only fused camera and radar data, these two modalities were chosen for a reason. Camera-based plethysmography is generally known to be high performing but exhibits high skin tone bias. In contrast, from our experiments, we note that radar-based methods have relatively poorer performance but are more resistant to skin tone bias. This work demonstrates that in both theory and practice, sensor fusion of RGB and radar modalities can improve performance and fairness.

Comparing Fusion Results with Radar The goal of this work is to improve the fairness of camera-based plethysmography through fusion with radar measurements. Through

such fusion, we show performance and fairness gains over the camera-based modality. An alternative analysis is comparison of the fusion method with the radar modality. This sets up a tradeoff-based selection: the fused modality shows sizeable performance gains over radar, albeit with a small reduction in fairness. Notably, this means that the dark skin tone performance of the fusion method is superior to both unimodal methods. This choice between performance and fairness depends on the end-user. We present the fusion-based method to the community as a viable alternative, for potential adoption. Follow-up works can attempt to improve the fusion-model fairness.

Limitations Constrained by relatively smaller data sizes, and imperfect skin tone balance in the dataset, we use a late fusion parameterization for our model, to enable better conditioned model training. End-to-end learning for multi-modal fusion may be explored in future works for improved performance gains. In addition, while our dataset is the largest dataset for camera-radar fusion plethysmography with a focus on demographic diversity, we note the need for continued effort towards dataset collection. Our definition of bias is also specific in nature. We deal with bias in terms of the signal to noise ratio (SNR). This is one interpretation of bias, and future work can extend our tools and analysis to other definitions.

Future Work and Conclusion In follow-up work, it is possible to add further sensing modalities such as thermal, acoustic, near infrared, and polarization images to this dataset. These additional modalities have their own uses and can aid in the sensing of additional physiological information beyond the plethysmograph. We conclude by noting that this is a small step in what might be a much bigger trend for the next-generation of internet of things (IoT) devices where both performance and equity are quality metrics.

Ethical Considerations Although there is much algorithmic research on fairness, it is imperative for *devices* to also be equitable and not disadvantage segments of the population. This is particularly important for devices that may one day be used clinically.

CHAPTER 3

Correcting for Solar Loading in Infrared Thermometers

3.1 Problem Background

3.1.1 Introduction

Infrared thermometers (IRTs) offer the potential to measure human body temperature in a fast, non-invasive manner. These devices can be based on single point (hereafter, non-contact infrared thermometer (NCIT)) or image based measurements (hereafter, thermographic imaging). IRTs have been used for rapid screening of temperature to help maintain public health during the SARS, H1N1, Ebola and now, COVID-19 outbreaks. A common case is when workers enter an office building on their way to work. The lobby will have a kiosk that uses a thermal camera to screen for an elevated temperature, to assess if the person is safe to proceed into a crowded office. If she has elevated temperature, she is advised to stay home. Such screening schemes are placed at airports, hotels, and stadiums and have proven useful for minimizing spread of infectious disease.

Given the importance of measuring temperature it is critical that thermal cameras measure temperature accurately and equitably. Recent findings about thermal camera bias have not shown a clear link between skin tone and thermal bias. However, this is only true in controlled indoor scenarios. When a subject is exposed to the sun, their skin absorbs solar radiation and heats up—we term this phenomenon “solar loading". While the thermal sensor correctly measures increased skin temperature, the thermometer does not correct for the sun-induced temperature elevation.

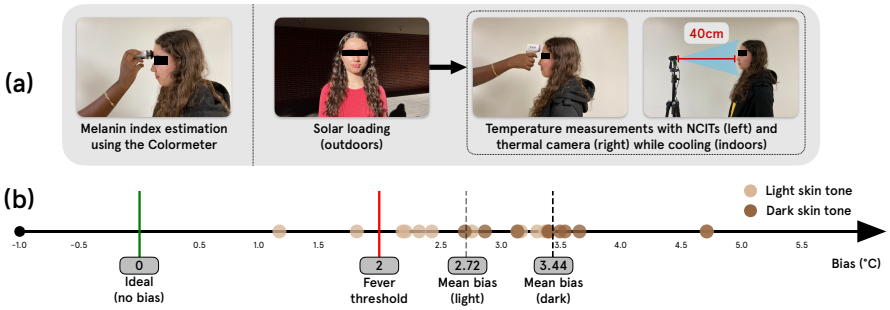

Figure 3.1: **Experimental data showing the solar loading effect. (a)** The experiment setup. **(b)** Solar loading results in elevated skin temperature measurements that are biased against dark skin toned subjects compared to light skin toned subjects. Commonly accepted fever threshold ($> 37.9°C$) [31] would misclassify solar loaded, healthy, dark skin toned subjects.

Addressing solar loading is an important problem because the effect is significant. In fact, the solar loading bias is often more than the threshold for fever. Observe the human subjects data shown in Fig. 3.1(b). The number line at bottom plots the deviation in temperature per subject with respect to a ground truth device based on contact thermography. All subjects were healthy, but most of them are falsely identified as having a fever. Moreover, there is also an issue with equity. Absorption of solar radiation in skin depends on the constituent chromophores, such as melanin. This suggests that the magnitude of solar loading scales with skin pigmentation, resulting in poorer IRT performance for select demographics. Thus, in our work we aim to not only correct solar loading, but to also study and remedy skin tone bias due to the solar loading effect.

While our focus is on solar loading, our broader vision aligns with a surge of recent scholarship in generalizing IRT operation to outdoor and less controlled environments. Dzien et al. [30], Ogawa et al. [65], Spindel et al. [82] evaluated IRTs in cold weather and found that IRT measurements were correlated with the cold. Ravi et al. [70], Tay et al. [88] tested IRTs in hot weather and found a similar correlation. It is generally accepted in the scientific

Table 3.1: **Today, commercial thermal IR sensors require a waiting period for solar loading to cool down.**

Manufacturer	**Model**	**IR Sensor**	**Waiting Period**
Welch Allyn	105801	NCIT	30 min [102]
ADC	Adtemp 429	NCIT	30 min [4]
Joytech Sejoy	DET-306	NCIT	10 min [77]
Solution I (Ours)	Generic	NCIT	**3-5 min**
Solution II (Ours)	Generic	Thermal Camera	**0 min**

community that IRTs are accurate in limited settings: a patient must be acclimated to an indoor measurement location, held at room temperature ($20 - 22°C$) for 10-30 minutes [25, 34] and this is typically the recommendation of NCIT manufacturers as can be seen in Table 3.1. In contrast to previous work, we focus our attention on the less characterized, but no less important, issue of solar loading biases.

3.1.2 Related Work

Broadly, there are two types of infrared thermometers: those that use point measurements and those that capture spatial temperature fields. The former, which refers to as non-contact infrared thermometers (NCITs), measure the forehead or temple to estimate core temperature. The latter captures larger field-of-views (FoVs) using thermal cameras and will be referred to as infrared thermography. A temporal artery thermometer (TAT) falls between these two groups: the device is manually scanned across the face to obtain multiple temperature measurements. While a TAT is highly accurate, we do not consider it in this work because (1) proper operation of the device requires training and (2) they are *contact-based* infrared thermometers. In this section, we discuss the accuracy and bias in NCITs and thermographic methods.

Non-contact infrared thermometers (NCITs): While NCITs are accurate in standard cases, in general, their performance depends on climate, measurement location and metabolic activity. Erenberk et al. [32] monitored forehead NCIT measurements in children exposed to cold weather ($0 - 4°C$). NCIT measurements underestimate core temperature, but increase and stabilize after 10 minutes indoors. Spindel et al. [82] compared NCIT performance in different measurement locations based on proximity to outdoors conditions. They found that accuracy improved as measurement locations moved further indoors. The influence of sunlight is considered but not explored. Kistemaker et al. [48] assessed the reliability of NCITs after exercise, which is expected to raise the metabolic rate. After 15 minutes of exercise, NCITs overestimated by $1.2°C$. Shajkofci [78] corrects environmental perturbations on NCIT measurements by regressing a linear relationship between skin, outdoors and ambient temperature. Additionally, they correct for diurnal variations by recording the time of day. This work, as noted by the authors, is limited in the ethnic distribution of the dataset.

Infrared Thermography: In some ways, infrared thermography is preferred over NCITs. NCITs require the device to be held $1 - 2$cm from the skin, while thermographic methods do not suffer from this limitation due to the larger FoV. Infrared cameras are however more expensive than NCITs as the hardware is more complex. In infrared thermography, an image of the subject is taken, usually of the face or neck. A region-of-interest (ROI) is selected and the aggregate skin temperature is mapped to core temperature. The inner canthus and forehead are suggested ROIs due to high perfusion from the carotid and orbital arteries respectively [34]. It is worth noting that the ROIs comprise barely 1% of the entire thermal image, meaning 99% of captured data is discarded.

IR thermography suffers from the same limitations as NCITs. Švantner et al. [85] found that measurements were influenced by ambient conditions. Wang et al. [97] use regression on multiple ROIs to achieve better accuracy then NCITs, but impact of ambient conditions are not studied.

Machine Learning for Infrared Thermometers: Machine learning has often been used with infrared thermometers in two ways: to improve the accuracy of infrared thermography, or to predict infections and illnesses using NCITs and other physiological measurements. Dagdanpurev et al. [26] used a combination of facial infrared thermography images, axillary temperature measurements and ambient temperature values to obtain optimum temperature estimates using linear regression analysis. This optimized temperature was further used for infection screening using the kNN algorithm. A skin heat transfer model for converting facial infrared thermography images into blood-perfusion maps was proposed by Wu et al. [105] to improve facial recognition model performance. The approach helped alleviate the effect of environmental conditions on infrared images. NCITs and thermographic cameras were used to detect temperatures which helped classify COVID-19 cases using recurrent neural networks (RNN) with long short term memory (LSTM) model [74]. Random forests were utilized in Li et al. [53] for the prediction of thermal comfort of healthy subjects in indoor HVAC settings using facial infrared thermography.

Bias in Infrared Thermometers: Recent works have evaluated bias in medical devices and proposed general solutions to eliminating bias, such as collection of diverse datasets [42, 116]. Other works, specifically in heart rate estimation, have demonstrated strides towards equitable technology both through hardware and software [11, 20, 93]. Recently, understanding the limitations of infrared thermometers has become important, but studies on potential biases reveal conflicting findings. Adams et al. [3] found that age and gender impacted IR thermography measurements. Regarding skin tone bias, Khan et al. [46] evaluated NCITs in subjects grouped by researcher labelled skin tone: "light" and "medium and dark". No skin tone bias was found, although the dataset was heavily skewed towards lighter skin. Strasse et al. [84] compared the performance of NCITs across body locations as well as ethnic groups (Black, White and mixed race). Again, no bias was found. Most recently, Bhavani et al. [16] compared TATs against oral thermometers. From self-reported race, performance on Black

and White patients were compared, revealing that TAT measurements underestimated temperature in Black patients ($-0.07°C$). Charlton et al. [23] found that emissivity of skin does not depend on skin pigmentation. These conflicting studies show that race, ethnicity and skin tone groupings preclude an understanding of bias in IRTs. The mechanism behind recorded biases is unclear.

3.1.3 Contributions and Scope

In this thesis, we offer two solutions to solar loading. Solution I exploits temporal modulation of solar loading to moderately reduce the waiting period. This is useful for NCITs. However, if spatial information is available with a camera, then we propose Solution II which offers single shot correction. Our approach is in the family of *computational light transport* techniques that use optical and computational principles to reveal and interpret the flow of light in our everyday world [15, 49]. The insights from light transport are practical: they inform learning-based inductive biases to correct for solar loading. A real dataset of subjects is collected and ground-truthed with contact-based thermometers. Metadata is stored with objective measures of melanin concentration, enabling an analysis of equity. The end result from this study is that solar loading is corrected. Since solar loading affects both light skinned and dark skinned people, solving for solar loading is a win-win where equity is improved as well as accuracy for everyone.

3.2 Thermal Light Transport Preliminaries

We overview the image signal processor (ISP) chain that thermal imagers use to estimate temperature of the human body depicted in Fig. 3.2. Let us begin by linking an object's temperature to its emission of light. In particular, every object hotter than absolute zero ($-273.15°C$) emits radiation proportional to its temperature—this radiation is called *thermal radiation*. The relationship between thermal radiation and temperature is wavelength

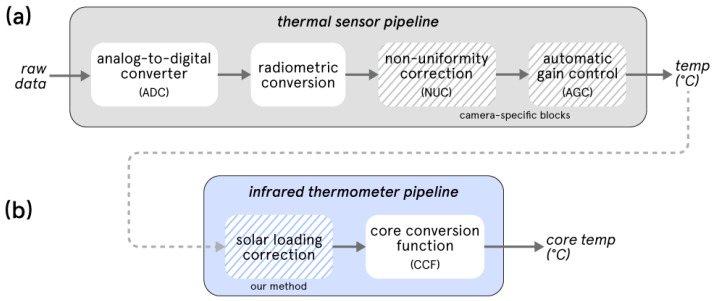

Figure 3.2: **Thermal image signal processing (ISP). (a)** Raw data from a thermal sensor is converted to temperature through a series of processing blocks. For thermal cameras, additional processing for temporal drift and gain control (shaded) are added. **(b)** Infrared thermometers convert sensor temperature to core temperature using a device-specific calibration. We propose an additional solar loading correction (shaded).

dependent. We observe this in everyday life: an iron bar heated to a very high temperature will initially glow red and then as its temperature rises it will glow orange, then yellow, and so on, until it is perceived as glowing violet. This relationship is described more precisely by Planck's law. For a given wavelength λ, an object with emissivity ε and temperature T has spectral radiant exitance $I(\lambda, T)$:

$$I(\lambda, T) = \frac{2\pi\varepsilon hc^2}{\lambda^5} \frac{1}{e^{hc/(\lambda k T)} - 1} \quad Wm^{-2}, \tag{3.1}$$

where $h = 6.63{\cdot}10^{-34} J\,s$ (Planck's constant), $k = 1.38{\cdot}10^{-23} J\,K^{-1}$ (Boltzmann constant) and $c = 3 \cdot 10^8 m\,s^{-1}$. Planck's law in Eq. (3.1) can be simplified. Assume that we have a sensor that can capture readings over the support of wavelengths considered. Then we can integrate Eq. (3.1) to obtain the total radiant flux emitted by a surface as:

$$I(T) = \int_{\lambda} M(\lambda, T)\, d\lambda, \tag{3.2}$$

$$= \varepsilon\sigma T^4. \tag{3.3}$$

This equation is known as the *Stefan-Boltzmann law* which is convenient as it relates thermal radiation to only three terms. The thermal radiation received is proportional to the product

of the fourth power of temperature multipled by emissivity and a fixed constant, $\sigma = 5.67 \cdot 10^{-8} W\, m^{-2}\, K^{-4}$, known as the *Stefan-Boltzmann constant*. Estimation of temperature can proceed by taking the fourth root of Eq. (3.3) as.

$$T = \sqrt[4]{\frac{I(T)}{\varepsilon\sigma}}. \tag{3.4}$$

We can measure $I(T)$, using for instance, a thermal camera and σ is the fixed Stefan-Boltzmann constant, leaving us with only the emissivity to plug in. Emissivity determines how efficient a surface is at emitting thermal energy and varies across objects: $\varepsilon = 1.0$ describes a perfect emitter of thermal radiation (blackbody) and $\varepsilon = 0$ describes an object that does not emit thermal radiation (e.g. polished metals). The emissivity value of human skin is known in literature as $\varepsilon = 0.98$.[1] To obtain the temperature of the human, we can take the fourth root of received radiation *vis a vis* Eq. (3.4) by setting $\varepsilon = 0.98$ to obtain temperature of the human skin surface. This assumes that the human is the only object in the scene.

3.2.1 Multipath Temperature Estimation

Real world measurement of human body temperature must account for multiple thermal emitters in a scene. In any practical setting, the human body is in a *multipath environment*. A sensor measures thermal radiation from all the objects in a scene. Consider a person sitting in a room. Their skin emits thermal radiation proportional to Eq. (3.3) and reflects radiation from the wall to the camera. This is a form of global illumination reminiscent of the Cornell Room in computer graphics, except that this setting is at thermal wavelengths. Thermal radiation arriving at the camera is compounded with radiation from the atmosphere. The object, background and atmosphere contribute to the thermal multipath.

To address this multipath problem, thermal sensor ISPs used in clinical medicine have a

[1] In the context of equitable sensing, one must be careful about using a single material constant for all humans. However, bias has not been observed in thermal cameras [73], and so this value is considered safe to use.

radiometric chain that abstracts multipath into a model with three steady state terms. A radiometric chain relates radiation intensity I_{sensor} at a single scene point as:

$$I_{\text{sensor}} = I_{\text{skin}} + I_{\text{amb}} + I_{\text{atm}}. \tag{3.5}$$

Here, I_{skin} is the intensity of thermal emission from the human body, and more precisely, the skin surface. The radiation from objects in the scene that reaches the measured spot is referred to as I_{amb} and is the multipath of objects in the background environment. The ambient term is difficult to write out explicitly as multipath reflections depend on the ambient surroundings, which are generally unknown. Finally I_{atm} describes the intensity from atmospheric air particles.

The goal of the radiometric chain in commercial devices is to rewrite Eq. (3.5) into a form that extracts $\overline{T}_{\text{skin}}$. In literature, the human body is assumed to be *opaque* to thermal radiation, such that thermal radiation does not pass through the body. The emissivity of the human body can then be described by Kirchhoff's law: $\varepsilon = \alpha$, where α is absorptivity or the fraction of incident radiation that is absorbed by the object. In analogy to visible light optics, conservation of energy on absorbed, reflected, and transmitted radiation is expressed as:

$$1 = \alpha + r + \tau, \tag{3.6}$$

$$1 = \varepsilon + r + \tau, \tag{3.7}$$

where r, τ are the fraction of reflected and transmitted radiation respectively. For opaque, human skin, $\tau = 0$ and consequently $r = 1 - \varepsilon$. When measuring a human body target, multiple radiative phenomena are recorded by a thermal sensor. Due to opacity, the human only emits radiation (Eq. (3.3)) and reflects a portion of background radiation. The emitted and reflected components then pass through an atmospheric volume. The atmosphere both attenuates the incoming radiation by τ_{atm}, the transmission factor, and emits its own radiation. Putting this together, we can use the emissivity of the human skin as well as human

temperature ($\overline{T}_{\text{skin}}$), ambient temperature ($\overline{T}_{\text{amb}}$) and atmospheric temperature ($\overline{T}_{\text{atm}}$) to write Eq. (3.5) as

$$\begin{aligned} I_{\text{sensor}} &= \tau_{\text{atm}} \left(I_{\text{skin}} + (1-\varepsilon) I_{\text{amb}}\right) + (1-\tau_{\text{atm}}) I_{\text{atm}} \\ &= \sigma \tau_{\text{atm}} \left(\varepsilon \overline{T}^4_{\text{skin}} + (1-\varepsilon)\overline{T}^4_{\text{amb}}\right) + \sigma(1-\tau_{\text{atm}})\overline{T}^4_{\text{atm}}. \end{aligned} \tag{3.8}$$

Note the explicit use of a bar on top of the temperature symbols. This emphasizes that the temperature is at steady state and not fluctuating in time. Eq. (3.8) can be algebraically rearranged to solve for the human skin temperature as

$$\overline{T}_{\text{skin}} = \sqrt[4]{\frac{\frac{I_{\text{sensor}}}{\sigma} - \tau_{\text{atm}}(1-\varepsilon)\overline{T}^4_{\text{amb}} - (1-\tau_{\text{atm}})\overline{T}^4_{\text{atm}}}{\tau_{\text{atm}}\varepsilon}}. \tag{3.9}$$

Estimation of skin temperature from Eq. (3.9) requires values for all parameters on the right hand side of the equation. This is problematic because τ_{atm} is not known. It is possible to hard-code a value as some existing thermal monitors do, by estimating the working distance and priors on air concentration. However, since our derivations will not depend on the specific value of τ_{atm}, we will without loss of generality set $\tau_{\text{atm}} = 1$. At a close range of less than 3 meters $\tau_{\text{atm}} = 0.98 - 0.99$, and this value is also a good approximation when the camera is sufficiently close to the human subject. Eq. (3.9) now simplifies to

$$\overline{T}_{\text{skin}} = \sqrt[4]{\frac{\frac{I_{\text{sensor}}}{\sigma} - (1-\varepsilon)\overline{T}^4_{\text{amb}}}{\varepsilon}}. \tag{3.10}$$

Here all terms on the right hand side are measured or estimated in commercial devices for healthcare. As discussed, I_{sensor} is the intensity measured by the instrument and $\varepsilon = 0.98$ for the human body. Clinical devices include an additional calibration step to measure $\overline{T}_{\text{amb}}$, the ambient temperature. While temperature of human skin can now be estimated in multipath conditions, there is another problem: skin temperature is not the same as clinically useful core temperature.

3.3 Effect of Solar Loading on Temperature Estimation

3.3.1 Core Temperature Estimation

We now discuss how existing devices estimate core temperature from skin temperature. The core temperature, denoted as $\overline{T}_{\text{core}}$, is the temperature of our internal organs and differs from peripheral skin temperature by a few degrees Celcius. Clinicians seek the core temperature, not the skin temperature. Existing instruments estimate $\overline{T}_{\text{core}}$ from $\overline{T}_{\text{skin}}$ by finding a map:

$$\mathcal{C} : \overline{T}_{\text{skin}} \rightarrow \overline{T}_{\text{core}}, \tag{3.11}$$

where the form of $\mathcal{C}$ varies across manufacturers, and use configurations. Existing devices use a mapping that assumes the subject is *thermoneutral*, i.e., the skin temperature is at a steady state temperature with no transient effects. Such mappings are often linear, of the form:

$$\overline{T}_{\text{core}} = b_0 + b_1 \overline{T}_{\text{skin}}, \tag{3.12}$$

where b_0 and b_1 are scalar coefficients used by a manufacturer. Unfortunately, the assumption of a steady state model is invalid. If sun is shining on a person's skin, there is a transient effect to $\overline{T}_{\text{skin}}$ but that same transient effect will not propagate to $\overline{T}_{\text{core}}$.

The core mapping, $\mathcal{C}$, is opaque to users because it is built from proprietary data collected by a manufacturer. As such, instead of directly solving for core temperature, we aim to estimate the solar loading bias in skin temperature. With a bias value, we can estimate the thermoneutral skin temperature and use existing device mappings to obtain the core temperature.

3.3.2 Gap in Prior Models: Solar Loading

The gap in the current radiometric chain in Eq. (3.5) is that heating from the sun is not modeled. As discussed from the introduction, the problem that we seek to address is *solar loading*, where the skin heats up outdoors due to sunlight. A valid indoor measurement

can then only be taken many minutes later. The focus is to correct for solar loading while minimizing the waiting period. It should also be noted that while solar loading is a source of error for all human subjects, it particularly disadvantages darker skin subjects who heat up more due to melanin absorption. We model tissue in the next section with varying melanin to underscore this disparity.

3.3.3 Equity of Solar Loading

The magnitude of the solar loading effect depends on absorption of solar radiation by the skin. Melanin absorbs solar radiation in the visible range of the EM spectrum and conservation of energy dictates that absorbed the energy must exit the skin. It does so in the form of long wave radiation, also known as thermal radiation. Eq. (3.7) dictates that absorptivity is equal to emissivity ($\alpha = \varepsilon = 0.98$), which is constant for human skin. This appears to contradict the claim that absorption of solar radiation is melanin dependent. However, α in Eq. (3.7) is defined for thermal wavelengths, while absorption of solar radiation mainly occurs for visible wavelengths.

3.4 Correcting Solar Loading Induced Errors

3.4.1 Transient Model and Correction (Solution I)

Here, we present Solution I for solar loading that uses only the transient temperature profile at a single spatial point to correct for solar loading. The advantage of Solution I (compared to Solution II) is the requirement of only one spatial location. This is useful because temperature scanners are often single pixel as they are much cheaper than a thermal camera. The downside of Solution I (compared to Solution II) is that transient information alone only reduces but does not eliminate a waiting period.

Solar loading involves transience. The existing three component model from Eq. (3.5)

needs to be generalized to include time as a parameter. Let us do so using t for time and writing

$$I_{\text{sensor}}(t) = I_{\text{skin}}(t) + I_{\text{amb}} + I_{\text{atm}}. \tag{3.13}$$

In contrast to the previous steady state model, $I_{\text{skin}}(t)$ depends on both steady state and time-varying components of the skin temperature. Solar loading skin temperature can be written as:

$$T_{\text{skin}}(t) = \overline{T}_{\text{skin}} + \overline{\beta}_{\text{solarpeak}} f(t), \tag{3.14}$$

where $\overline{\beta}_{\text{solarpeak}}$ is the maximum solar loaded bias at steady state in units of degrees. Note that the maximum solar loading is a steady state quantity, but transience arises *vis a vis* multiplication by time-varying function $f(t)$ with range from 0 to 1 representing the fraction of maximum solar loading. For example, a person who has been outdoors for a while might have a maximum temperature increase of $\overline{\beta}_{\text{solarpeak}} = 5°C$. When the person returns indoors they will start cooling down as $f(t)$ starts to decay from a maximum value of 1 (fully loaded) to 0 (no solar loading bias). It is useful to also think of a transient parametrization of solar loading bias, $\beta_{\text{solar}}(t)$. This can be written by rearranging Eq. (3.14) as

$$\beta_{\text{solar}}(t) = T_{\text{skin}}(t) - \overline{T}_{\text{skin}} = \overline{\beta}_{\text{solarpeak}} f(t). \tag{3.15}$$

Heat transfer literature studies how human body tissue heats up with radiation and offers differential equations for bio-heat. One such model, from Wang et al. [96], can be approximated such that skin temperatures approximate an exponential model during solar loading and cooling. This is visualized in Fig. 3.3. Using this principle, an estimate of temperature $T^*_{\text{skin}}(t)$ can be obtained by replacing $f(t)$ in Eq. (3.14) with an exponential. This returns two equations, one for heating/loading and another for cooling:

$$T^*_{\text{skin}}(t) \approx \begin{cases} \overline{\beta}_{\text{solarpeak}}(1 - e^{-r_h t}) + \overline{T}_{\text{skin}} & \text{heating,} \\ \overline{\beta}_{\text{solarpeak}}\, e^{-r_c t} + \overline{T}_{\text{skin}} & \text{cooling.} \end{cases} \tag{3.16}$$

where r_h, r_c are the rate of heating and cooling respectively and have non-negative values. The rates depend on factors such as skin tone, blood flow and environmental temperature.

Here, our goal is to estimate the steady state temperature $\overline{T}_{\text{skin}}$, which is at a single scene point. From Eq. (3.16) we see that skin temperature is parameterized by three values: the maximum solar loading, the level of solar loading and the steady state skin temperature. With a time series of skin temperature measurements, we can extrapolate $\overline{T}_{\text{skin}}$ by fitting to the exponential model.

Based on Eq. (3.16), only three temperature measurements are needed to fully determine $T^*_{\text{skin}}(t)$; that is, we only need to find $\overline{\beta}_{\text{solarpeak}}$, r_c and $\overline{T}_{\text{skin}}$. Our estimated steady state skin temperature $\overline{T}^*_{\text{skin}}$ can be passed through a standard IRT core estimation block to retrieve a solar loading invariant core temperature:

$$\underbrace{\overline{T}_{\text{skin}}}_{ground\,truth} \rightarrow \boxed{\begin{array}{c}Solar\\Load\end{array}} \rightarrow \underbrace{\begin{bmatrix} T_{\text{skin}}(t_{\min}) \\ \vdots \\ T_{\text{skin}}(t_{\max}) \end{bmatrix}}_{measurement} \rightarrow \boxed{\begin{array}{c}Expn.\\Fit\end{array}} \rightarrow \underbrace{\overline{T}^*_{\text{skin}}}_{estimate} \rightarrow \boxed{\begin{array}{c}Core\\Est.\end{array}} \rightarrow \underbrace{\overline{T}^*_{\text{core}}}_{estimate}, \tag{3.17}$$

where $t_{\min}$ is the time the first temperature sample was captured and $t_{\max}$ is the time the last sample was captured. The total time window observed is then $W_t = t_{\max} - t_{\min}$.

Unfortunately, thermal sensors exhibit drift and introduce a non-Gaussian temporally distributed noise that is difficult to correct, resulting in a low signal-to-noise ratio (SNR). The distribution of noise is also non-stationary. Environmental fluctuations such as wind and other factors must also be accounted for. For this reason using transient data in the real world is difficult. Solution I requires many more than three measurements and a suitably long time window of observation for the steady state extrapolation to work well. This is illustrated on real data in Fig. 3.3, where one observes that multiple minutes, between 3-5 minutes of transient data are needed to extrapolate the exponential. Nonetheless, Solution I is still an advance over previous methods which do zero solar loading correction. In particular, the waiting period is on average reduced as compared to existing solutions. However, we still seek a single shot solution, enabling mass screening without a waiting period. In Solution II, we leverage the spatial temperature field of a face to demonstrate single shot correction.

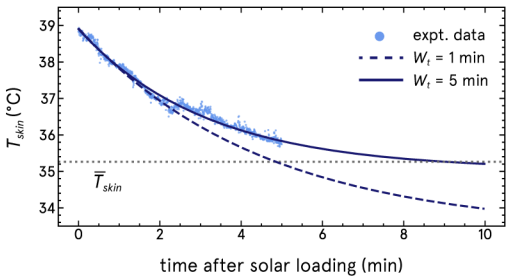

Figure 3.3: **Solution I requires transient data to extrapolate, which increases the waiting time.**

3.4.2 Spatial Modulation and Correction (Solution II)

Although solar loading seems like a solely transient effect, there is a spatial dependence as well. Solar loading is a variational quantity with respect to the spatial fields of geometry and material that characterize a human face.

The three path radiometric chain in Eq. (3.5) describes neither temporal nor spatial modulation. It can be generalized to include space as a parameter. Let us do so, using $\mathbf{x}$ for space to write:

$$I_{\text{sensor}}(\mathbf{x}) = I_{\text{skin}}(\mathbf{x}) + I_{\text{amb}} + I_{\text{atm}} \tag{3.18}$$

A few direct simplifications have been made:

- The ambient multipath and atmospheric intensity is assumed to not be spatially dependent.
- Temporal information is not involved because we are restricting scope to single shot correction.

We follow an analogous path to Equations 3.13 to 3.16 from Solution I by defining solar loading bias as a spatial functional:

$$\beta_{\text{solar}}(\mathbf{x}) = T_{\text{skin}}(\mathbf{x}) - \overline{T}_{\text{skin}}(\mathbf{x}). \tag{3.19}$$

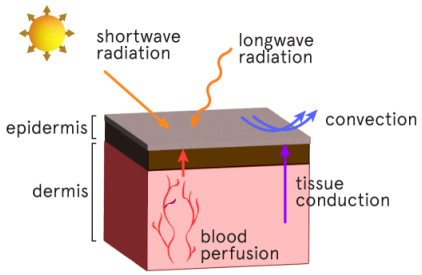

Figure 3.4: **Skin heat transfer.** Skin temperature is determined by the heat transfer of endogenous and exogenous factors. Penne's bio-heat equation describes these heat transfer processes.

We need to show that $T_{\text{skin}}(\mathbf{x})$ and $\overline{T}_{\text{skin}}(\mathbf{x})$ are both functions of space, and that they are different functions of space (so they do not cancel to a constant). In particular, the two quantities $T_{\text{skin}}(\mathbf{x})$ and $\overline{T}_{\text{skin}}(\mathbf{x})$ are defined by a partial differential equation known as Penne's Bio-heat equation whose mechanics are illustrated in Fig. 3.4. This equation is derived from the standard heat equation, where heat storage is balanced with diffusion (q_{diff}), internal heat generation and external heating sources. In our simplified setting, heat generation is from *in vivo* physiology which we abstract as (q_{blood}), external heating is from the sun (q_{rad}), and the tissue properties are defined by the constants: tissue density (ρ) and specific heat of tissue (c). For brevity, we do not include the boundary conditions and full definition of all terms. For further details, we direct the reader to Wang et al. [96]. We then have:

$$\rho c \frac{\partial T_{\text{skin}}(\mathbf{x}, t)}{\partial t} = q_{diff}(\mathbf{x}, t) + q_{blood}(\mathbf{x}, t) + q_{rad}(\mathbf{x}) \tag{3.20}$$

$$\rho c \frac{\partial \overline{T}_{\text{skin}}(\mathbf{x}, t)}{\partial t} = q_{diff}(\mathbf{x}, t) + q_{blood}(\mathbf{x}, t) \tag{3.21}$$

The primary difference between the two is q_{rad} that we define as:

$$q_{rad} = \alpha(\mathbf{x}, \mu_{mel})\, E_{sun}\, \max(0, \mathbf{l} \cdot n(\mathbf{x})) \tag{3.22}$$

where $\alpha(\mathbf{x}, \mu_{mel})$ is the proportion of incident solar energy absorbed given the melanin concentration, μ_{mel}, in the skin, E_{sun} is the incident solar radiation power, $\mathbf{l}$ is the direction of

solar radiation, and $n(\mathbf{x})$ is the surface normal vector at the location $\mathbf{x}$.

Note that $T_{\text{skin}}(\mathbf{x}, t)$ and $\overline{T}_{\text{skin}}(\mathbf{x}, t)$ are both modulated by spatial heterogeneity in biology of the skin (spatial variation of blood perfusion, melanin concentration, etc.). However, when sunlight is incident on the face, $\mathbf{l} \cdot n(\mathbf{x}) > 0$, $T_{\text{skin}}(\mathbf{x})$ and thereby $\beta_{\text{solar}}(\mathbf{x})$ are spatially dependent on the interaction of the solar rays with the geometry of the face: $\mathbf{l} \cdot n(\mathbf{x})$. However, this suggests that the bias term is jointly dependent on time and space. In the following, we show that there exists a space-time equivalence in the bias term that allows us to write it strictly as a function of space if certain conditions are met, thus enabling one-shot methods.

There exists a Space-Time Equivalence in Solar Loading Bias if $\overline{T}_{\text{skin}}(\mathbf{x}, t)$ is constrained to be spatially homogeneous and diffusion only occurs in the depth dimension. The spatially varying solar loading bias $\beta_{\text{solar}}(\mathbf{x})$ can be solved for with linear equations if $\overline{T}_{\text{skin}}(\mathbf{x}, t)$ is spatially homogeneous and in steady state such that it is written as $\overline{T}_{\text{skin}}$. Then, the solar loading bias can be solved for given at least two regions whose temperatures and surface normals are known and different from each other. The temperature of their regions can then be modeled from Eq. (3.14) at a particular time instant t_o:

$$\begin{aligned} T(x_1, t_o) &= \overline{T} + \overline{\beta}\ f(t_o)\ \left(\vec{l} \cdot \vec{n}(x_1)\right), \\ &\vdots \\ T(x_n, t_o) &= \overline{T} + \overline{\beta}\ f(t_o) \left(\vec{l} \cdot \vec{n}(x_n)\right). \end{aligned} \tag{3.23}$$

Given the temperatures of the regions, as well as the solar vector and surface normals, it is possible to solve for $\beta_{\text{solar}}(\mathbf{x}) f(t_o)$. Then, the solar bias can be determined at any location whose surface normal and solar vector is known at one point in time.

3.4.3 Correcting for Solar Loading in Space-Time

The spatial-modulation model in Eq. (3.21) describes the governing equations for solar loaded skin temperature, parameterized by $T_{\text{skin}}, T_{\text{core}}$ and melanin. Theoretically, the equations can be inverted to solve for $\overline{T}_{\text{skin}}$ given a series of observations, however, the assumed constraints

are severe and do not generalize to human physiology. Diffusion occurs across all three dimensions and occurs concurrently during solar loading and cooling; additionally, the human face is not spatially homogeneous as can be seen in Fig. 3.6. We know that skin temperature varies spatially, but the lack of an analytical solution precludes our understanding of the specific temperature patterns and gradients induced by solar loading. There is, however, no dearth of facial thermal imagery. By taking a data-driven approach, we can correct solar loading by learning a canonical pattern of heat distribution from real faces and how they change during the cooling process that allows us to relax the spatially homogeneous and depth diffusion constraint in Eq. (3.23). While spatial temperature and normal vector distributions [89] of the face can be learned through a data-driven approach, a potential drawback of the method is if the relative orientation of the surface normals and light vector vary at a high frequency during solar loading due to motion. If this occurs, than Eq. (3.23) is invalid due to the light vector being a function of time, $\mathbf{l}(t)$.

3.5 Experimental Method

3.5.1 Protocol and Dataset

We show through simulation and *in situ* experiments that the solar loading bias affects IRT measurements and is skin tone-dependent. To do so, we acquire a unique dataset of temperature and skin color information. Our imaging prototype consists of an IR camera (FLIR Lepton 3.5 LWIR camera) mounted next to a RGB webcam (ArduCam). The RGB camera is only used for better facial landmark detection when processing the dataset. In lieu of a block-body reference, the wall in the background of the images is used as a pseudo-reference to help correct for camera flat-field correction effects. Environmental conditions are recorded using a solar power meter (Tenmars Solar Meter TM-206) and a handheld anemometer (BTMeter Anemometer 866A). Throughout the experiment, the study subject is measured using multiple NCITs, an oral thermometer (Boncare Digital Thermometer MT-

601A) and our RGB-IR camera setup. Subject's skin tone information is recorded during the study; as there is no universal non-invasive skin color measurement, we discuss some existing methods of skin color measurement.

The standard for measuring skin tone in previous works has been subjective classification of skin darkness and using ethnicity as a proxy for skin tone. Classification of skin tones is done using Fitzpatrick skin tone (FST) scale, which groups skin tones into 6 categories of variable skin darkness. We record FST using the ratings of 2 or more study coordinators. The FST scale has known limitations [101] so we collect objective skin tone measurements to separate optical biases from race-related physiological biases. The FST and melanin demographics are shown in Fig. 3.5. We use the DSM III Colormeter (Cortex Technology) to measure melanin index (MI). This device has been widely used in various clinical studies for skin tone and scar color measurements [40, 90]. The colormeter measures red light reflected by the skin (I_r), to obtain:

$$MI = 100 \cdot \log\left(\frac{1}{I_r}\right). \tag{3.24}$$

In subsequent analysis, we group subjects into dark and light skin tone groups based on the MI. We select $MI = 45$ as the threshold between dark and light skin tones because it is the median value of our dataset.

We collect data from subjects under multiple conditions: thermoneutral, solar loading, and cooling. We record the subject's temperature, skin color and the experimental conditions. The setup is shown in Fig. 3.1(a). Prior to the experiment, the subject rests indoors to acclimatize to the experiment location. After resting, the subject's oral temperature, NCIT temperatures and skin color are recorded. A thermal video of their face provides steady state skin temperature. Next, the subject stands in sunlight for five minutes, after which they are immediately measured using NCITs. Back indoors, the subject cools down for five minutes and their skin temperature is recorded continuously by our thermal camera setup. Finally, NCIT temperatures are measured after five minutes of cooling.

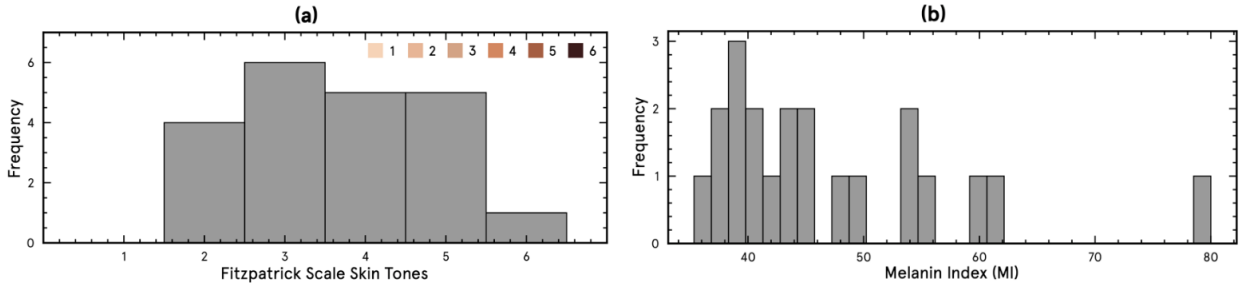

Figure 3.5: **Dataset demographics.** Dataset distribution across: **(a)** Fitzpatrick scale skin tones. **(b)** Melanin Index (MI) values.

Solar loading bias depends not only on skin tone, but also on environmental conditions, such as wind, solar azimuth, and cloudiness. To control for these covariates, we only collect data under the following conditions:

- Solar zenith close to 0 (12:00-2:00PM)
- Minimal to no clouds ($E_{sun} \geq 1000\, Wm^{-2}$)
- Minimal to no wind ($v_{wind} < 3\, ms^{-1}$)

Collecting data in a limited time of day also controls for diurnal variations. 21 subjects with melanin index (MI) values over the range $[35.34, 80]$ were selected to ensure a skin tone diverse dataset and the dataset demographics are summarized in Fig. 3.5.

3.5.2 Solar Loading Correction

The hardware required for our data acquisition is off-the-shelf and relatively inexpensive. The novelty of our method is in our computational imaging algorithms, which we discuss next.

3.5.2.1 Transient Solution

Solving for $\overline{T}_{\text{skin}}$ using Eq. (3.16) can be done in many ways. Due to the unknown steady state term, the function is not linearizable but can be solved numerically using a non-linear least squares algorithm or using a grid search as in [86, 87]. Knowledge of typical skin parameters allows us to bound the variables; in our case, we constrain skin temperature to be between $27 - 43°C$ and the solar loading bias to safely be between $0 - 20°C$. We find that fitting an exponential to the data is extremely sensitive to the camera calibration, so we additionally weight the data according to the calibration status, determined by finding temperature spikes in the background image.

3.5.2.2 Single-Shot Solution

To demonstrate the utility of spatial information, we train a lightweight convolutional neural network (CNN) to learn solar loading bias $\overline{\beta}_{\text{solarpeak}} f(t)$ from single-shot data. The network is exceedingly simple and still performs well, consisting only of two convolutional layers and three fully connected layers. We train using the ADAM optimizer for 20 epochs on cropped (50×50) facial thermal images. While we do our best to collect a diverse data from training, there will always be gaps in the acquired dataset. To combat this, we augment the dataset in multiple ways. First, we apply random horizontal flips to training data, since the level of solar loading bias in the image does not change due to flips. We simulate fever images by adding a constant offset ($1.6°C$) to the face images [110]. The linear effect of core temperature on skin temperature is shown in Wang et al. [96]. The simulated fever images help ensure that the model is not overfitting to temperature intensity, rather than spatial patterns, when correcting for solar loading. We use a statistical power analysis to determine that 14 is the minimum number of subjects needed, and we collect data from 21 subjects. The network is trained using leave-one-out cross validation; each subject is used as a test subject, 4 subjects are used for validation and remaining subjects are used for training.

Table 3.2: **Errors of solutions during steady state and solar loading.**

	Method	MAE (°C) ↓	RMSE (°C) ↓	MAPE (°C) ↓
Steady State	T_{skin}	0	0	0
	Sol. I	1.63	2.16	4.75
	Sol. II	**0.39**	**0.52**	**1.13**
Solar Loaded	T_{skin}	2.86	2.97	8.40
	Sol. I	2.48	2.79	7.23
	Sol. II	**0.72**	**0.95**	**2.12**

3.6 Results

In the following, we assess our solar loading correction methods against the thermal camera measurement of steady state skin temperature, $\overline{T}_{\text{skin}}$. We do not directly compare the camera method to the NCITs since the mapping function from T_{skin} to T_{core} is unknown as mentioned in Section 3.3.1. We compare our temporal and spatial solutions against the uncorrected thermal measurement after solar loading and cooling.

3.6.1 Transient Solution (Solution I)

Since most point sensors measure the forehead temperature, we choose this location for our transient experiments. The forehead is also the most accessible and tends to have a steady temperature, making it easy to continuously measure this location. An example forehead temperature curve during cooling is shown in Fig. 3.3, along with exponential curves fit with

variable window lengths. In both cases, the exponential fitting is successful in minimizing the residual, but an accurate $\overline{T}^*_{\text{skin}}$ does not always accompany this. In Table 3.2, the performance of this method is shown as Sol. I where 5 seconds of data are used for fitting. Assuming an exponential model causes the model to suffer on easy cases, such as the steady and cooled states, where temporal noise may present as a skin transient. However, for the solar loaded case, this model corrects a minimal amount of solar loading (MAE=2.48 compared to MAE=2.86 for the uncorrected skin temperature).

3.6.2 Spatial Solution (Solution II)

Our learned spatial solution is a data-hungry model. Since the problem of solar loading is unexplored, we collect our own dataset to test our spatial correction method. For each subject, we acquire their steady state facial temperatures as well as a continuous capture of cooling after solar loading. The steady state is used to generate $\beta_{\text{solar}}(t)$ for solar loaded frames for the model to regress on. Fig. 3.6 shows facial temperatures and results from our trained CNN on multiple subjects. We observe that the model tends to perform worse when correcting the first minute of cooling. The solar loading effect is the strongest in the early minutes of cooling and the magnitude of solar loading may be difficult to correct. Additionally, recall that skin temperature cools exponentially, so there are far less frames of strong solar loading compared to moderate solar loading. Overall, the accuracy does not suffer too much due to this effect. One failure case is shown in Fig. 3.6. The model attempts to correct solar loading but achieves a constant error in correction over time. Despite this, the model still recognizes that solar loading is present and corrects $\approx 2.5°C$ of solar loading. Quantitative results from the CNN are given in Table 3.2 as Sol. II. Sol. II significantly outperforms other methods in the case of solar loading, estimating level of solar loading with an MAE=0.97.

Note too, that Solution II is not returning false positive corrections, i.e., making a correction to temperature when solar loading does not exist. As illustrated in the curves in Fig. 3.6

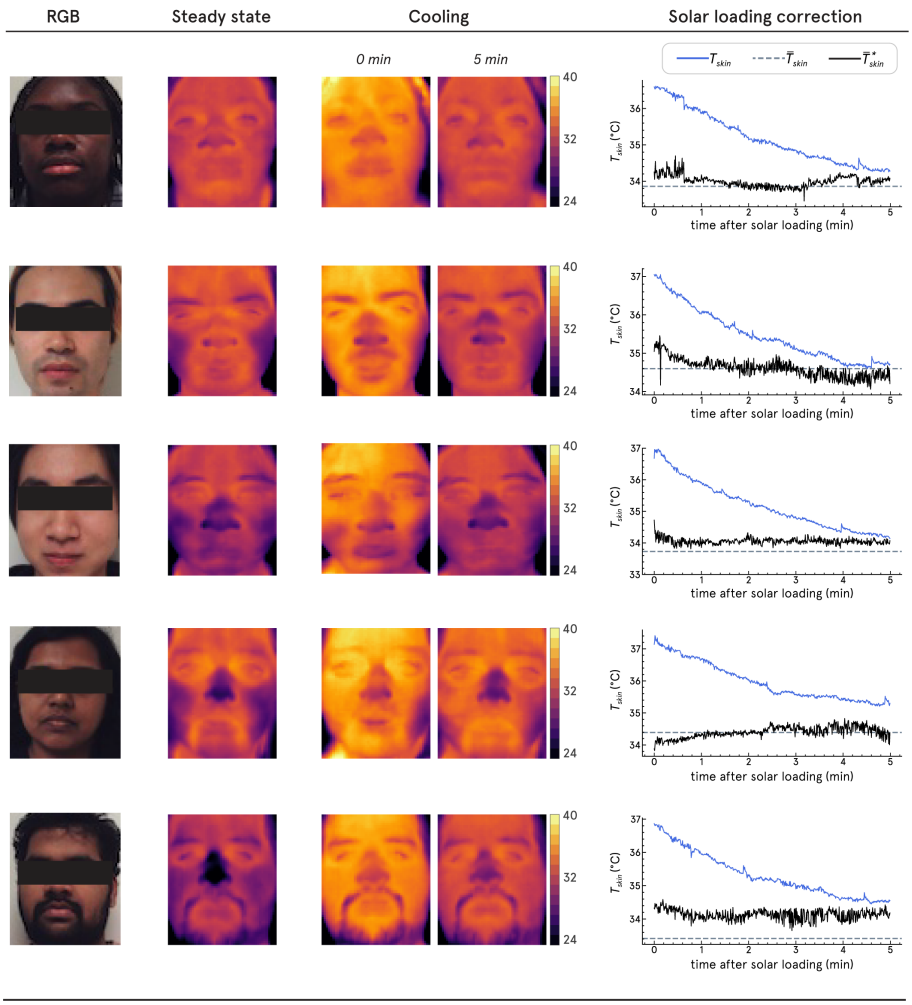

Figure 3.6: **Solution II enables single shot correction and does not over correct when solar loading is not present.** From left to right: RGB image, steady state thermal image (ground truth), indoor cooling after solar loading (0 min, 5 min) and the results of our correction. In the plots, we compare the solar loaded skin temperature (blue, T_{skin}), the baseline temperature (dotted, $\overline{T}_{\text{skin}}$) and our corrected skin temperature (black, T^*_{skin}) Our method successfully removes solar loading (up to 3°C) on multiple subjects. The plots show frame-by-frame inference of Solution II, showing that Solution II is not overfitting to a fixed temperature offset.

the temperature prediction from Solution II is relatively uncorrelated with the amount of solar loading.

3.6.3 Equity Analysis

In Section 3.3.3, we proposed melanin bias in solar loading using theory and now we further validate this finding with experimental data. The lower-level skin bias is seen in Fig. 3.1(b). Dark skin samples have, on average, a higher solar loading bias than light samples by $+0.72°C$. After our spatial solar loading correction, the absolute difference between dark and light solar loading bias is $-0.24°C$ (MAE for dark samples is $0.82°C$, MAE for light samples is $1.06°C$).

While our experimental conditions control for non-subject covariates, we can further improve our analysis by collecting data from subject pairs. Two subjects, one with dark skin and one with light skin, complete the experiment at the same time, hence undergoing the same environmental conditions. We are interested in how the NCIT error, $\epsilon = T_{\text{NCIT}} - T_{\text{oral}}$, varies between skin tones. Paired $\epsilon_{\text{dark}}, \epsilon_{\text{light}}$ measurements were compared using paired t-tests ($p = 0.05$). The subsequent analysis shows only the results for device NCIT 1; we do not report specific values for the other tested NCITs since they exhibit the same behavior.

Prior to solar loading, there is no statistical difference in NCIT error between dark and light samples. However, after five minutes of solar loading, NCIT temperatures are higher for dark subjects with a mean bias of $+1.49°C, 95\%$ CI $[0.46, 2.51]$. After five minutes of cooling, there is again no statistical difference in NCIT error between skin tones. We conclude that—under solar loading—NCITs overestimate temperatures more severely for dark skin compared to light skin subjects. NCIT performance is summarized in Section 3.6.3.

While there is no bias after 5 minutes of cooling, the NCITs are still inaccurate: NCIT values are uncorrelated with oral temperatures ($r = 0.07$). This is compared to moderate correlation for before solar loading ($r = 0.65$) and poor correlation after 5 minutes of solar

	MAE (°C)			
Method	**Steady state**	**Solar loaded**	**Cooled (5')**	**Claimed Acc.**
NCIT 1	0.39	7.10	0.65	0.1
NCIT 2	0.60	6.10	0.72	0.3
NCIT 3	0.51	7.98	0.78	0.2

Table 3.3: **Performance of commercial NCITs.** We test three NCITs which are marketed as being within the ASME and FDA acceptable error limits. However, the devices are not within their specified ranges even during steady state.

loading ($r = 0.02$). Daanen et al. [25] suggest subjects should rest for up to 30 minutes after heat exposure to ensure accurate NCIT readings. The poor accuracy of the NCIT after five minutes of cooling agrees with the time frame set forth by Daanen et al. [25].

3.7 Discussion

In summary, we present two methods of correction of solar loading. Solution I of using transient correction was a natural method to implement and relates to NCIT devices. Unfortunately, temporal errors in the measurement from today's devices means that transient correction (as we implemented it) requires a large window of sampling observation. We therefore consider our implementation of Solution II of using spatial correction to be desirable. This enables single shot elimination of solar loading with a thermal camera.

Limitations: This study characterizes light transport principles and demonstrates the ability to correct for solar loading in a single shot, but it is not a large-scale clinical study of solar loading. The scale of this study was chosen to match calculations of a power size

sample statistic. Based on the large effect size of solar loading, only 14 samples were needed to draw significance. A larger scale study, conducted in the clinic, could directly assess the impact of solar loading, not only on temperature measurement, but also clinical decision making.

Conclusion: This study has discussed solar loading and offered single shot correction (33ms) using spatial information. Before spatial correction, the solar loading bias is $3.03°C$ and post-correction the error is $0.97°C$. Before spatial correction the difference in solar loading effects between dark and light skin groups is statistically significant with test statistic of (one-tail Kolmogorov-Smirnov test, $p < 0.005$) and after correction the difference is not statistically significant ($p > 0.5$). We hope this lays a foundation for large scale efforts to improve the ability of thermal cameras to accurately and equitably sense the human body.

Ethical Considerations The equity numbers from this work were based on the sampling dataset used by the authors. The study was collected on a University campus and does not match the inclusion criteria of a societal study. Although this work analyzes skin tone, this is only one particular axis on which one seeks equitable operation. There exist several other axes of variation across humans. The study size was sufficiently large to demonstrate that solar loading is corrected (based on power analysis measures), but a large-scale clinical study that builds on the ideas here could explore a broader range of inclusion criteria, axes of demographic variation, and improvements in the performance of solar loading correction.

CHAPTER 4

Conclusions

This thesis proposed two methods for improving the accuracy of non-contact health monitoring sensors while making them more equitable. Many of the research thrusts taken in this field are often improving general performance of the devices. This metric is important, however, neglecting equity can lead to serious ramifications.

Equity, especially in the medical field, is paramount. If parts of the population feel like medical devices do not work as well on them, trust in the device's capabilities falls. Moreover, negative experiences that marginalized groups have experienced in the medical field have already led to an overall deterioration of trust in the medical field [60]. Therefore, it is extremely important that we reverse this trend by understanding devices' weaknesses and correcting them.

In order to do this, we first need to understand what anatomical factors may cause a device to perform worse. One of the easiest ways to ensure this is for future data-collection efforts to ensure as a diverse representation in the dataset as possible. A diverse representation of anatomical factors allows for an analysis of how performance varies with respect to a certain factor. For example, in Chapter 2 and Chapter 3, the acquisition of a diverse dataset allowed for a bias analysis with respect to skintone. The next step after identifying a physical bias is to solve the bias through a different approach or at least to document it so that it is well-known that the bias exists.

While the two examples shown in this thesis relate to a physical bias caused by skintone, there exist many other types of physical biases. Ideally, all sources of physical biases or axes

of fairness should be explored. These can be anatomical differences such as gender, BMI, or age, but can be more subtle such body-type, body fat/muscle composition, race, presence of a disease, or even geographical location.

Since there exist many applications of non-contact health monitoring and sources of physical bias, there is a large scope for future work. For instance, there may be a relation between the accuracy at which radar can sense heartbeats and breathing to the BMI or fat percentage of a person's body. This is even more relevant with the recent pushes towards using radar for sleep apnea detection and sleep quality measurement [41, 44]. Furthermore, while [93] improves the accuracy and equity of plethysmography estimation, it is not clear whether this waveform can be used for downstream tasks. A downstream task such as remote SpO2 detection requires accurate photoplethysmography estimation and radar and camera fusion plethysmograph signal may not be an accurate PPG surragote. Future work can try to utilize the combination of radar and camera to output a PPG signal from the camera that is somehow improved by radar data without interfering with the amplitude information of the PPG. Future work can also explore the best methodology to understand both the physical biases of a device and how to build and properly test a new one. The development of such methodologies would make research plans easier and comparisons between works straightforward. The field of non-contact health monitoring is growing and steps are being taken towards more equity. We hope to build upon this work to make sure that the future of remote health monitoring is robust and equitable.

Bibliography

[1] Øyvind Aardal, Yoann Paichard, Sverre Brovoll, Tor Berger, Tor Sverre Lande, and Svein-Erik Hamran. Physical working principles of medical radar. *IEEE Transactions on Biomedical Engineering*, 60(4):1142–1149, 2012.

[2] Lonneke AM Aarts, Vincent Jeanne, John P Cleary, C Lieber, J Stuart Nelson, Sidarto Bambang Oetomo, and Wim Verkruysse. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—a pilot study. *Early human development*, 89(12):943–948, 2013.

[3] Scott Adams, Tracey Bucknall, and Abbas Kouzani. An initial study on the agreement of body temperatures measured by infrared cameras and oral thermometry. *Scientific Reports*, 11(1):11901, June 2021. doi: 10.1038/s41598-021-91361-6. URL https://doi.org/10.1038/s41598-021-91361-6.

[4] ADC. Adtemp non-contact thermometer, model 429 user manual, jan 2017. URL https://www.adctoday.com/products/429.

[5] Md Atiqur Rahman Ahad, Upal Mahbub, and Tauhidur Rahman. *Contactless Human Activity Analysis*. Springer, 2021.

[6] Mostafa Alizadeh, George Shaker, João Carlos Martins De Almeida, Plinio Pelegrini Morita, and Safeddin Safavi-Naeini. Remote monitoring of human vital signs using mm-wave fmcw radar. *IEEE Access*, 7:54958–54968, 2019.

[7] Sarah Alotaibi and William AP Smith. A biophysical 3d morphable model of face appearance. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 824–832, 2017.

[8] R. Rox Anderson and John A. Parrish. The optics of human skin. *Journal of Investigative Dermatology*, 77(1):13–19, 1981. ISSN 0022-202X. doi: https://doi.org/

10.1111/1523-1747.ep12479191. URL https://www.sciencedirect.com/science/article/pii/S0022202X15461251.

[9] Consumer Technology Association. Physical activity monitoring for heart rate, ansi/cta-2065, 2018.

[10] Yunhao Ba, Zhen Wang, Kerim Doruk Karinca, Oyku Deniz Bozkurt, and Achuta Kadambi. Overcoming difficulty in obtaining dark-skinned subjects for remote-ppg by synthetic augmentation. *arXiv preprint arXiv:2106.06007*, 2021.

[11] Yunhao Ba, Zhen Wang, Kerim Doruk Karinca, Oyku Deniz Bozkurt, and Achuta Kadambi. Style transfer with bio-realistic appearance manipulation for skin-tone inclusive rPPG. In *2022 IEEE International Conference on Computational Photography (ICCP)*. IEEE, August 2022. doi: 10.1109/iccp54855.2022.9887649. URL https://doi.org/10.1109/iccp54855.2022.9887649.

[12] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3430–3437, 2013.

[13] Carina Barbosa Pereira, Michael Czaplik, Vladimir Blazek, Steffen Leonhardt, and Daniel Teichmann. Monitoring of cardiorespiratory signals using thermal imaging: a pilot study on healthy human subjects. *Sensors*, 18(5):1541, 2018.

[14] Solon Barocas, Moritz Hardt, and Arvind Narayanan. Fairness in machine learning. *Nips tutorial*, 1:2017, 2017.

[15] Ayush Bhandari, Achuta Kadambi, and Ramesh Raskar. *Computational Imaging*. MIT Press, 2022.

[16] Sivasubramanium V. Bhavani, Zanthia Wiley, Philip A. Verhoef, Craig M. Coopersmith, and Ighovwerha Ofotokun. Racial differences in detection of fever using tem-

poral vs oral temperature measurements in hospitalized patients. *JAMA*, 328(9):885, September 2022. doi: 10.1001/jama.2022.12290. URL https://doi.org/10.1001/jama.2022.12290.

[17] Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29:4349–4357, 2016.

[18] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

[19] Toon Calders, Asim Karim, Faisal Kamiran, Wasif Ali, and Xiangliang Zhang. Controlling attribute effect in linear regression. In *2013 IEEE 13th international conference on data mining*, pages 71–80. IEEE, 2013.

[20] Pradyumna Chari, Krish Kabra, Doruk Karinca, Soumyarup Lahiri, Diplav Srivastava, Kimaya Kulkarni, Tianyuan Chen, Maxime Cannesson, Laleh Jalilian, and A. Kadambi. Diverse r-ppg: Camera-based heart rate estimation for diverse subject skin-tones and scenes. 2020.

[21] Pradyumna Chari, Yunhao Ba, Shreeram Athreya, and Achuta Kadambi. Mime: Minority inclusion for majority group enhancement of ai performance. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII*, pages 326–343. Springer, 2022.

[22] Matthew Charlton, Sophie A Stanley, Zoë Whitman, Victoria Wenn, Timothy J Coats, Mark Sims, and Jonathan P Thompson. The effect of constitutive pigmentation on the measured emissivity of human skin. *Plos one*, 15(11):e0241843, 2020.

[23] Matthew Charlton, Sophie A. Stanley, Zoë Whitman, Victoria Wenn, Timothy J. Coats, Mark Sims, and Jonathan P. Thompson. The effect of constitutive pigmentation

on the measured emissivity of human skin. *PLOS ONE*, 15(11):e0241843, November 2020. doi: 10.1371/journal.pone.0241843. URL https://doi.org/10.1371/journal.pone.0241843.

[24] Weixuan Chen and Daniel McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 349–365, 2018.

[25] Hein Daanen, Stephan Bose-O'Reilly, Matt Brearley, D. Andreas Flouris, Nicola M. Gerrett, Maud Huynen, Hunter M. Jones, Jason Kai Wei Lee, Nathan Morris, Ian Norton, Lars Nybo, Elspeth Oppermann, Joy Shumake-Guillemot, and Peter Van den Hazel. COVID-19 and thermoregulation-related problems: Practical recommendations. *Temperature*, 8(1):1–11, August 2020. doi: 10.1080/23328940.2020.1790971. URL https://doi.org/10.1080/23328940.2020.1790971.

[26] Sumiyakhand Dagdanpurev, Guanghao Sun, Lodoiravsal Choimaa, Shigeto Abe, and Takemi Matsui. Clinical application of multiple vital signs-based infection screening system in a mongolian hospital: Optimization of facial temperature measurement by thermography at various ambient temperature conditions using linear regression analysis. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5313–5316, Honolulu, HI, July 2018. IEEE. doi: 10.1109/embc.2018.8513513. URL https://doi.org/10.1109/embc.2018.8513513.

[27] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[28] Amy D Droitcour, Olga Boric-Lubecke, Victor M Lubecke, Jenshan Lin, and Gregory TA Kovacs. Range correlation and i/q performance benefits in single-chip silicon doppler radars for noncontact cardiopulmonary monitoring. *IEEE Transactions on Microwave Theory and Techniques*, 52(3):838–848, 2004.

[29] Amy Diane Droitcour. *Non-contact measurement of heart and respiration rates with a single-chip microwave doppler radar*. Stanford University, 2006.

[30] Cornelius Dzien, Wolfgang Halder, Hannes Winner, and Monika Lechleitner. Covid-19 screening: are forehead temperature measurements during cold outdoor temperatures really helpful? *Wiener klinische Wochenschrift*, 133(7-8):331–335, October 2020. doi: 10.1007/s00508-020-01754-2. URL https://doi.org/10.1007/s00508-020-01754-2.

[31] A. Sahib El-Radhi. Fever in common infectious diseases. In *Clinical Manual of Fever in Children*, pages 85–140. Springer International Publishing, Cham, Switzerland, 2018. doi: 10.1007/978-3-319-92336-9_5. URL https://doi.org/10.1007/978-3-319-92336-9_5.

[32] Ufuk Erenberk, Emel Torun, Emin Ozkaya, Selcuk Uzuner, Aysegul Dogan Demir, and Rusen Dundaroz. Skin temperature measurement using an infrared thermometer on patients who have been exposed to cold. *Pediatrics International*, 55(6):767–770, December 2013. doi: 10.1111/ped.12188. URL https://doi.org/10.1111/ped.12188.

[33] Center for Devices and Radiological Health. Non-contact infrared thermometers, Apr 2020. URL https://www.fda.gov/medical-devices/general-hospital-devices-and-supplies/non-contact-infrared-thermometers.

[34] Josh Foster, Alex Bruce Lloyd, and George Havenith. Non-contact infrared assessment of human body temperature: The journal temperature toolbox. *Temperature*, 8(4):306–319, April 2021. doi: 10.1080/23328940.2021.1899546. URL https://doi.org/10.1080/23328940.2021.1899546.

[35] Davide Giavarina. Understanding bland altman analysis. *Biochemia medica*, 25(2): 141–151, 2015.

[36] Wei Han, Hui Chen, and Soujanya Poria. Improving multimodal fusion with hier-

archical mutual information maximization for multimodal sentiment analysis. *arXiv preprint arXiv:2109.00412*, 2021.

[37] S. W. Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 553–560, June 2010. doi: 10.1109/CVPR.2010.5540167. ISSN: 1063-6919.

[38] Christophe Hurter and Daniel McDuff. Cardiolens: remote physiological monitoring in a mixed reality environment. In *ACM siggraph 2017 emerging technologies*, pages 1–2. 2017.

[39] Takanori Igarashi, Ko Nishino, and Shree K Nayar. *The appearance of human skin: A survey.* Now Publishers Inc, 2007.

[40] MEH Jaspers, CM Stekelenburg, JM Simons, KM Brouwer, M Vlig, Eric Van den Kerckhove, E Middelkoop, and PPM van Zuijlen. Assessing blood flow, microvasculature, erythema and redness in hypertrophic scars: a cross sectional study showing different features that require precise definitions. *Burns*, 43(5):1044–1050, 2017.

[41] Abdul Q Javaid, Carlo M Noble, Russell Rosenberg, and Mary Ann Weitnauer. Towards sleep apnea screening with an under-the-mattress ir-uwb radar using machine learning. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 837–842. IEEE, 2015.

[42] Achuta Kadambi. Achieving fairness in medical devices. *Science*, 372(6537):30–31, April 2021. doi: 10.1126/science.abe9195. URL https://doi.org/10.1126/science.abe9195.

[43] Achuta Kadambi. Achieving fairness in medical devices. *Science*, 372(6537):30–31, 2021.

[44] Sun Kang, Dong-Kyu Kim, Yonggu Lee, Young-Hyo Lim, Hyun-Kyung Park, Sung Ho Cho, and Seok Hyun Cho. Non-contact diagnosis of obstructive sleep apnea using impulse-radio ultra-wideband radar. *Scientific reports*, 10(1):5261, 2020.

[45] Manjit Kaur and Dilbag Singh. Fusion of medical images using deep belief networks. *Cluster Computing*, 23(2):1439–1453, 2020.

[46] Shahrukh Khan, Bridey Saultry, Scott Adams, Abbas Z. Kouzani, Kelly Decker, Robin Digby, and Tracey Bucknall. Comparative accuracy testing of non-contact infrared thermometers and temporal artery thermometers in an adult hospital setting. *American Journal of Infection Control*, 49(5):597–602, May 2021. doi: 10.1016/j.ajic.2020.09.012. URL https://doi.org/10.1016/j.ajic.2020.09.012.

[47] Yoonkyoung Kim, Yosep Park, Jinman Kim, and Eui Chul Lee. Remote heart rate monitoring method using infrared thermal camera. *International Journal of Engineering Research and Technology*, 11(3):493–500, 2018.

[48] J. A. Kistemaker, E. A. Den Hartog, and H. A. M. Daanen. Reliability of an infrared forehead skin thermometer for core temperature measurements. *Journal of Medical Engineering & Technology*, 30(4):252–261, January 2006. doi: 10.1080/03091900600711381. URL https://doi.org/10.1080/03091900600711381.

[49] Kyros Kutulakos. Technical perspective: The dawn of computational light transport. *Communications of the ACM*, 59(9):78–78, August 2016. doi: 10.1145/2975163. URL https://doi.org/10.1145/2975163.

[50] Panicos A Kyriacou and John Allen. *Photoplethysmography: Technology, Signal Analysis and Applications*. Academic Press, 2021.

[51] Eugene Lee, Evan Chen, and Chen-Yi Lee. Meta-rppg: Remote heart rate estimation using a transductive meta-learner. In *European Conference on Computer Vision*, pages 392–409. Springer, 2020.

[52] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jędrzej Nowak. Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity. In *2011 federated conference on computer science and information systems (FedCSIS)*, pages 405–410. IEEE, 2011.

[53] Da Li, Carol C. Menassa, and Vineet R. Kamat. Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography. *Energy and Buildings*, 176:246–261, October 2018. doi: 10.1016/j.enbuild.2018.07.025. URL https://doi.org/10.1016/j.enbuild.2018.07.025.

[54] James C Lin. Noninvasive microwave measurement of respiration. *Proceedings of the IEEE*, 63(10):1530–1530, 1975.

[55] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. Multi-task temporal shift attention networks for on-device contactless vitals measurement. *Advances in Neural Information Processing Systems*, 33:19400–19411, 2020.

[56] Wenjie Lv, Wangdong He, Xipeng Lin, and Jungang Miao. Non-contact monitoring of human vital signs using fmcw millimeter wave radar in the 120 ghz band. *Sensors*, 21 (8):2732, 2021.

[57] Ewa Magdalena Nowara, Tim K Marks, Hassan Mansour, and Ashok Veeraraghavan. 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*, pages 1272–1281, 2018.

[58] Kenta Matsumura, Sogo Toda, and Yuji Kato. Rgb and near-infrared light reflectance/-transmittance photoplethysmography for measuring heart rate during motion. *IEEE Access*, 8:80233–80242, 2020.

[59] Cardiac Monitors. Heart rate meters, and alarms. *ANSI/AAMI Standard EC13*, 2002.

[60] Reem Nadeem. Black americans' views about health disparities, experiences with health care, Apr 2022. URL https://www.pewresearch.org/science/2022/04/07/black-americans-views-about-health-disparities-experiences-with-health-care/.

[61] Toshiaki Negishi, Shigeto Abe, Takemi Matsui, He Liu, Masaki Kurosawa, Tetsuo Kirimoto, and Guanghao Sun. Contactless vital signs measurement system using rgb-thermal image sensors and its clinical screening test on patients with seasonal influenza. *Sensors*, 20(8):2171, 2020.

[62] Benjamin W Nelson and Nicholas B Allen. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. *JMIR mHealth and uHealth*, 7(3):e10828, 2019.

[63] Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Transactions on Image Processing*, 29:2409–2423, 2019.

[64] Ewa M Nowara, Daniel McDuff, and Ashok Veeraraghavan. A meta-analysis of the impact of skin tone and gender on non-contact photoplethysmography measurements. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 284–285, 2020.

[65] Takuya Ogawa, Yukinori Tsukuda, Yuki Suzuki, Shigeto Hiratsuka, Ryo Inoue, and Norimasa Iwasaki. Utility of thermal image scanning in screening for febrile patients in cold climates. *Journal of Orthopaedic Science*, 27(6):1333–1337, November 2022. doi: 10.1016/j.jos.2021.08.002. URL https://doi.org/10.1016/j.jos.2021.08.002.

[66] Prashanth Pai and Mohammed Zafar Ali Khan. Comparison of sc and mrc receiver complexity for three antenna diversity systems. In *2008 24th Biennial Symposium on Communications*, pages 302–305. IEEE, 2008.

[67] Gorazd Planinsic. Infrared thermal imaging: fundamentals, research and applications, 2011.

[68] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58(1):7–11, 2010.

[69] Scott Prahl. Tabulated molar extinction coefficient for hemoglobin in water. URL https://omlc.org/spectra/hemoglobin/summary.html.

[70] Nirmal Ravi, Mathura Vithyananthan, and Aisha Saidu. Are all thermometers equal? a study of three infrared thermometers to detect fever in an african outpatient clinic. *PeerJ*, 10:e13283, June 2022. doi: 10.7717/peerj.13283. URL https://doi.org/10.7717/peerj.13283.

[71] Lingyun Ren, Haofei Wang, Krishna Naishadham, Ozlem Kilic, and Aly E Fathy. Phase-based methods for heart rate detection using uwb impulse doppler radar. *IEEE Transactions on Microwave Theory and Techniques*, 64(10):3319–3331, 2016.

[72] Lingyun Ren, Lingqin Kong, Farnaz Foroughian, Haofei Wang, Paul Theilmann, and Aly E Fathy. Comparison study of noncontact vital signs detection using a doppler stepped-frequency continuous-wave radar and camera-based imaging photoplethysmography. *IEEE Transactions on Microwave Theory and Techniques*, 65(9):3519–3529, 2017.

[73] E F J Ring and K Ammer. Infrared thermal imaging in medicine. *Physiological Measurement*, 33(3):R33–R46, February 2012. doi: 10.1088/0967-3334/33/3/r33. URL https://doi.org/10.1088/0967-3334/33/3/r33.

[74] Magesh S., Niveditha V.R., Rajakumar P.S., Radha RamMohan S., and Natrayan L. Pervasive computing in the context of COVID-19 prediction with AI-based algorithms. *International Journal of Pervasive Computing and Communications*, 16(5):

477–487, August 2020. doi: 10.1108/ijpcc-07-2020-0082. URL https://doi.org/10.1108/ijpcc-07-2020-0082.

[75] Silonie Sachdeva et al. Fitzpatrick skin typing: Applications in dermatology. *Indian journal of dermatology, venereology and leprology*, 75(1):93, 2009.

[76] Elliott Schires, Pantelis Georgiou, and Tor Sverre Lande. Vital sign monitoring through the back using an uwb impulse radar with body coupled antennas. *IEEE transactions on biomedical circuits and systems*, 12(2):292–302, 2018.

[77] Joytech Sejoy. Sejoy digital infrared forehead thermometer det-206/306, jan 2010. URL https://www.sejoygroup.com/wholesale-baby-medical-digital-infrared-forehead-thermometer.html.

[78] Adrian Shajkofci. Correction of human forehead temperature variations measured by non-contact infrared thermometer. *IEEE Sensors Journal*, 22(17):16750–16755, 2022. doi: 10.1109/JSEN.2021.3058958.

[79] Hafeez Ur Rehman Siddiqui, Adil Ali Saleem, Robert Brown, Bahattin Bademci, Ernesto Lee, Furqan Rustam, and Sandra Dudley. Non-invasive driver drowsiness detection system. *Sensors*, 21(14):4833, 2021.

[80] Vikas Singh, Nishchal K Verma, Zeeshan Ul Islam, and Yan Cui. Feature learning using stacked autoencoder for shared and multimodal fusion of medical images. In *Computational Intelligence: Theories, Applications and Future Directions-Volume I*, pages 53–66. Springer, 2019.

[81] Rencheng Song, Huan Chen, Juan Cheng, Chang Li, Yu Liu, and Xun Chen. Pulsegan: Learning to generate realistic pulse waveforms in remote photoplethysmography. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1373–1384, 2021.

[82] Jeffrey F Spindel, Stephen Pokrywa, Nathan Elder, and Clayton Smith. The environment has effects on infrared temperature screening for COVID-19 infection. *American Journal of Infection Control*, 49(11):1445–1447, November 2021. doi: 10.1016/j.ajic.2021.08.002. URL https://doi.org/10.1016/j.ajic.2021.08.002.

[83] Nitish Srivastava, Ruslan Salakhutdinov, et al. Multimodal learning with deep boltzmann machines. In *NIPS*, volume 1, page 2. Citeseer, 2012.

[84] Wally auf der Strasse, Daniel Prado Campos, Celso Júnio Aguiar Mendonça, Jamil Faissal Soni, Joaquim Mendes, and Percy Nohama. Forehead, temple and wrist temperature assessment of ethnic groups using infrared technology. *Medical Engineering & Physics*, 102:103777, April 2022. doi: 10.1016/j.medengphy.2022.103777. URL https://doi.org/10.1016/j.medengphy.2022.103777.

[85] Michal Švantner, Vladislav Lang, Tomáš Kohlschutter, Jiří Skála, Milan Honner, Lukáš Muzika, and Eliška Kosová. Study on human temperature measurement by infrared thermography. In *The 16th International Workshop on Advanced Infrared Technology & Applications*, page 1, Basel Switzerland, November 2021. MDPI. doi: 10.3390/engproc2021008004. URL https://doi.org/10.3390/engproc2021008004.

[86] Kenichiro Tanaka, Hiroyuki Kubo, Nobuhiro Ikeya, Takuya Funatomi, Tsuyoshi Takatani, and Yasuhiro Mukaigawa. Time-resolved light transport decomposition for thermal photometric stereo. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4804–4814, New York, NY, USA, June 2018. IEEE.

[87] Kenichiro Tanaka, Nobuhiro Ikeya, Tsuyoshi Takatani, Hiroyuki Kubo, Takuya Funatomi, Vijay Ravi, Achuta Kadambi, and Yasuhiro Mukaigawa. Time-resolved far infrared light transport decomposition for thermal photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):2075–2085, 2019.

[88] M.R. Tay, Y.L. Low, X. Zhao, A.R. Cook, and V.J. Lee. Comparison of infrared

thermal detection systems for mass fever screening in a tropical healthcare setting. *Public Health*, 129(11):1471–1478, November 2015. doi: 10.1016/j.puhe.2015.07.023. URL https://doi.org/10.1016/j.puhe.2015.07.023.

[89] George Trigeorgis, Patrick Snape, Iasonas Kokkinos, and Stefanos Zafeiriou. Face normals" in-the-wild" using fully convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 38–47, 2017.

[90] Martijn van der Wal, Monica Bloemen, Pauline Verhaegen, Wim Tuinebreijer, Henrica de Vet, Paul van Zuijlen, and Esther Middelkoop. Objective color measurements: clinimetric performance of three devices on normal skin and scar tissue. *Journal of Burn Care & Research*, 34(3):e187–e194, 2013.

[91] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.

[92] Sahil Verma and Julia Rubin. Fairness definitions explained. In *2018 ieee/acm international workshop on software fairness (fairware)*, pages 1–7. IEEE, 2018.

[93] Alexander Vilesov, Pradyumna Chari, Adnan Armouti, Anirudh Bindiganavale Harish, Kimaya Kulkarni, Ananya Deoghare, Laleh Jalilian, and Achuta Kadambi. Blending camera and 77 ghz radar sensing for equitable, robust plethysmography. *ACM Trans. Graph.(SIGGRAPH)*, 41(4):1–14, 2022.

[94] Vytautas Vizbara. Comparison of green, blue and infrared light in wrist and forehead photoplethysmography. *BIOMEDICAL ENGINEERING 2016*, 17(1), 2013.

[95] Tom Vogels, Mark Van Gastel, Wenjin Wang, and Gerard De Haan. Fully-automatic camera-based pulse-oximetry during sleep. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1349–1357, 2018.

[96] Hongyun Wang, Wesley A. Burgei, and Hong Zhou. Analytical solution of one-dimensional pennes' bioheat equation. *Open Physics*, 18(1):1084–1092, December 2020. doi: 10.1515/phys-2020-0197. URL https://doi.org/10.1515/phys-2020-0197.

[97] Quanzeng Wang, Yangling Zhou, Pejman Ghassemi, David McBride, Jon P. Casamento, and T. Joshua Pfefer. Infrared thermography for measuring elevated body temperature: Clinical accuracy, calibration, and evaluation. *Sensors*, 22(1):215, December 2021. doi: 10.3390/s22010215. URL https://doi.org/10.3390/s22010215.

[98] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard De Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7): 1479–1491, 2016.

[99] Zeyu Wang, Klint Qinami, Ioannis Christos Karakozis, Kyle Genova, Prem Nair, Kenji Hata, and Olga Russakovsky. Towards fairness in visual recognition: Effective strategies for bias mitigation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8919–8928, 2020.

[100] Zhen Wang, Yunhao Ba, Pradyumna Chari, Oyku Bozkurt, Gianna Brown, Parth Patwa, Niranjan Vaddi, Laleh Jalilian, and Achuta Kadambi. Synthetic generation of face videos with plethysmograph physiology. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022.

[101] Olivia R Ware, Jessica E Dawson, Michi M Shinohara, and Susan C Taylor. Racial limitations of fitzpatrick skin type. *Cutis*, 105(2):77–80, 2020.

[102] Welch. Care temp touch-free thermometer - directions for use, jan 2015. URL https://www.welchallyn.com/content/welchallyn/emeai/me/products/categories/thermometry/forehead-thermometers/caretemp-touch-free-thermometer.html.

[103] Fokko P Wieringa, Frits Mastik, and Antonius FW van der Steen. Contactless mul-

tiple wavelength photoplethysmographic imaging: A first step toward "spo 2 camera" technology. *Annals of biomedical engineering*, 33(8):1034–1041, 2005.

[104] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics (TOG)*, 31(4):1–8, 2012.

[105] Shiqian Wu, Weisi Lin, and Shoulie Xie. Skin heat transfer model of facial thermograms and its application in face recognition. *Pattern Recognition*, 41(8):2718–2729, August 2008. doi: 10.1016/j.patcog.2008.01.003. URL https://doi.org/10.1016/j.patcog.2008.01.003.

[106] Shuqiong Wu, Takuya Sakamoto, Kentaro Oishi, Toru Sato, Kenichi Inoue, Takeshi Fukuda, Kenji Mizutani, and Hiroyuki Sakai. Person-specific heart rate estimation with ultra-wideband radar using convolutional neural networks. *IEEE Access*, 7:168484–168494, 2019.

[107] T Wu. Ppgi: New development in noninvasive and contactless diagnosis of dermal perfusion using near infrared light. *J. GCPD eV*, 7(1):17–24, 2003.

[108] Ting Wu, Vladimir Blazek, and Hans Juergen Schmitt. Photoplethysmography imaging: a new noninvasive and noncontact method for mapping of the dermal perfusion changes. In *Optical Techniques and Instrumentation for the Measurement of Blood Composition, Structure, and Dynamics*, volume 4163, pages 62–70. International Society for Optics and Photonics, 2000.

[109] Bryan P Yan, William HS Lai, Christy KY Chan, Stephen Chun-Hin Chan, Lok-Hei Chan, Ka-Ming Lam, Ho-Wang Lau, Chak-Ming Ng, Lok-Yin Tai, Kin-Wai Yip, et al. Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals. *Journal of the American Heart Association*, 7(8):e008585, 2018.

[110] Mengkai Yan, Jianjun Qian, Renke Wang, Shangbing Gao, and Jian Yang. Dynamic group difference coding based on thermal infrared face image for fever screening. *IEEE Transactions on Instrumentation and Measurement*, 72:1–13, 2023. doi: 10.1109/tim.2023.3235442. URL https://doi.org/10.1109/tim.2023.3235442.

[111] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. *arXiv preprint arXiv:1905.02419*, 2019.

[112] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P Gummadi. Fairness constraints: Mechanisms for fair classification. In *Artificial Intelligence and Statistics*, pages 962–970. PMLR, 2017.

[113] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[114] Ellin Q Zhao, Alexander Vilesov, Shreeram Athreya, Pradyumna Chari, Jeanette Merlos, Kendall Millett, Nia St Cyr, Laleh Jalilian, and Achuta Kadambi. Making thermal imaging more equitable and accurate: Resolving solar loading biases. *arXiv preprint arXiv:2304.08832*, 2023.

[115] Tianyue Zheng, Zhe Chen, Shujie Zhang, Chao Cai, and Jun Luo. More-fi: Motion-robust and fine-grained respiration monitoring via deep-learning uwb radar. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 111–124, 2021.

[116] James Zou and Londa Schiebinger. Ensuring that biomedical AI benefits diverse populations. *EBioMedicine*, 67:103358, May 2021. doi: 10.1016/j.ebiom.2021.103358. URL https://doi.org/10.1016/j.ebiom.2021.103358.