TD

Facial Video Based Physiological Variables Estimation in Dark Environments

DOCTORAL THESIS

Ankit Gupta

DOCTOR DEGREE IN INFORMATICS ENGINEERING
SPECIALIZATION: HUMAN-MACHINE INTERACTION

UNIVERSIDADE da MADEIRA
A Nossa Universidade
www.uma.pt

July | 2024

Facial Video Based Physiological Variables Estimation in Dark Environments

DOCTORAL THESIS

Ankit Gupta
DOCTOR DEGREE IN INFORMATICS ENGINEERING
SPECIALIZATION: HUMAN-MACHINE INTERACTION

SUPERVISION
Fernando Manuel Rosmaninho Morgado Ferrão Dias

CO-SUPERVISION
Antonio Gabriel Ravelo García

Facial Video based Physiological Variables Estimation in Dark Environments

Ankit Gupta

University of Madeira

JURY MEMBERS

President

Prof. Dr. Eduardo Miguel Dias Marques, University of Madeira

Committee Members

Prof. Dr. Sebastian Zaunseder, University of Augsburg

Prof. Dr. Frédéric Bouasefsaf,University of Lorraine

Dr. Sheikh Shanawaz Mostafa, Agência Regional para o Desenvolvimento da Investigação, Tecnologia e Inovação

Prof. Dr. Fernando Manuel Rosmaninho Morgado Ferrão Dias, University of Madeira

Facial Video based Physiological Variables Estimation in Dark Environments

Ankit Gupta

Supervised by:

Professor Fernando Morgado-Dias

Professor Antonio G. Ravelo-Garcia

A thesis submitted in the fulfilment of the requirements of
DOCTOR OF PHILOSOPHY
in
INFORMATICS ENGINEERING.

FACULTY OF EXACT SCIENCES AND ENGINEERING,
UNIVERSITY OF MADEIRA

July 2024

Resumo

As estimativas de parâmetros fisiológicos desempenham um papel relevante na determinação do estado de saúde de um indivíduo. Entre esses parâmetros, a frequência cardíaca e a saturação de oxigénio têm sido amplamente utilizadas para monitorização da saúde durante exames médicos, cirurgias, diagnóstico de distúrbios do sono e em unidades de cuidados intensivos. As técnicas de referência para estimar esses parâmetros são a eletrocardiografia e a fotopletismografia. Ambas são técnicas baseadas em contacto e, portanto, podem causar desconforto ao paciente em cenários como monitorização prolongada e pele sensível ou queimada. Assim, a fotopletismografia remota foi introduzida como uma variante sem contacto da fotopletismografia. Esta técnica extrai o sinal de pulso do volume sanguíneo das sequências espaço-temporais da região de interesse, seguida pela estimativa da frequência cardíaca. Por outro lado, as estimativas da saturação de oxigénio são realizadas usando o método de razão de razões usando os canais vermelho e azul. Os métodos existentes sem contacto foram projetados para condições de luz ambiente. Alguns métodos desenvolvidos para ambientes escuros usaram câmaras infravermelhas, que são caras, e os espectros resultantes têm força pulsátil inferior aos espectros visíveis. Portanto, esta tese investiga o potencial dos espectros visíveis para medidas fisiológicas em ambientes escuros (iluminância 1,0 lux). Especificamente, esta tese tem três contribuições principais: primeiro, um novo método de estimativa da frequência cardíaca baseado na análise de componentes independentes subcompleta, que foi desenvolvido e testado sob diferentes condições em tempo real, e segundo, um conjunto de dados "Dark-Video" abrangendo participantes de diferentes etnias e, finalmente, uma nova arquitetura de aprendizagem profunda para aprimoramento de imagens escuras que também foi proposta para facilitar medições fisiológicas nos ambientes escuros mencionados acima (ou seja, métodos de estimativa em cascata pelo aprimoramento de imagens). Diversas experiências foram conduzidas para a análise de desempenho usando métricas de desempenho selecionadas criticamente provaram a superioridade dos métodos desenvolvidos e também exibiram o seu potencial de serem clinicamente viáveis. A direção futura desta trabalho visa implementar esses métodos para cenários como monitorização do sono sem contacto ou monitorização durante a condução noturna.

Keywords: Separação Cegas de Fontes, · Volume Sanguíneo, · Aprendizagem Profunda, · Métodos Sem Contacto, · Estimação de Parâmetros Fisiológicos.

Abstract

Physiological parameter estimations play a significant role in determining an individual's health status. Among these parameters, heart rate and oxygen saturation have been extensively used for health monitoring during medical checkups, surgery, sleep disorders diagnosis, and intensive care units. The gold standard techniques for estimating these parameters are electrocardiography and photoplethysmography. Both are contact-based techniques and, therefore, can cause discomfort to the subject in scenarios such as prolonged monitoring and sensitive or burnt skin. Thus, remote photoplethysmography was introduced as a non-contact variant of photoplethysmography. It extracts the blood volume pulse signal from the spatiotemporal sequences of the region of interest, followed by heart rate estimation. On the other hand, oxygen saturation estimations are being performed using the ratio-of-ratios method using red and blue channels. Existing non-contact methods were designed for ambient light conditions. A few methods developed for dark environments used infrared cameras, which are expensive, and the resulting spectra have poorer pulsatile strength than visible spectra. Therefore, this thesis investigates the potential of visible spectra for physiological measurements in dark environments (illuminance $\leq 1.0\ lux$). Specifically, this thesis has three key contributions: first, a novel heart rate estimation algorithm (U-LMA) based on undercomplete independent component analysis, which was developed and tested under different real-time conditions, and second, a "Dark-Video" dataset encompassing participants of different ethnicities, and finally a novel deep learning architecture for dark image enhancement that was also proposed to facilitate physiological measurements in the above mentioned dark environments (i.e., estimation methods cascaded by image enhancement). Diverse experiments conducted for the performance analysis using critically selected performance metrics not only proved the superiority of the developed methods but also exhibited their potential of being clinically viable. The future direction of this research aims to implement these methods for scenarios such as non-contact sleep monitoring or monitoring during nighttime driving.

Keywords: Blind Source Separation,· Blood Volume Pulse,· Deep Learning,· Non-contact Approaches,· Physiological Parameters Estimation.

Acknowledgements

I wish to express my deep sense of indebtedness and sincerest gratitude to my supervisor, Prof. Dr. Fernando Morgado-Dias (Associate Professor), Faculty of Exact Sciences and Engineering, and co-supervisor Prof. Antonio G. Ravelo-García (Associate Professor), Institute for Technological Development and Innovation in Communications, Universidad de Las Palmas de Gran Canaria, for their invaluable guidance and constructive criticism throughout this project.

They have displayed unique tolerance and understanding at every step of progress and encouraged me. I deem it my privilege to have carried out my dissertation work under their supervision.

I am thankful to almighty GOD without his blessings, this work would not have been complete.

I want to thank everyone directly or indirectly involved in completing this Project.

As a Final Personal Note, I am also grateful to my family, who are inspirational in their understanding, patience, and constant encouragement.

This research could not have been possible without financial support from the following projects:

1. Acknowledgment to the project MITIExcell co-financed by Regional Development European Funds for the Operational Programme "Madeira 14-20" – EIXO PRIORITÁRIO 1, of Região Autónoma da Madeira, no. M1420-01-0145-FEDER-000002.
2. Acknowledgment to M1420-01-0247-FEDER-000033 – MTL – Marítimo Training Lab.
3. Acknowledgment to LARSyS projects (Financiamento Base e Programático, with references UIDB/50009/2020 and UIDP/50009/2020, respectively).
4. Acknowledgment to Project 761-Smart Islands Hub.

Ankit Gupta
(2118619)

Table of Contents

List of Figures

List of Tables

List of Acronyms

AASM American Association of Sleep Medicine

ACM Association of Computer Machinery

ADMM Alternating Direction Method of Multiplier

ANN Artificial Neural Network

APBV adaptive PBV method

APP Adaptive Pulsatile Plane

AR Auto-Regressive Models

B-A Bland-Altman

BCG Ballistocardiography

BP Blood Pressure

bpm beats per minute

BR Breathing Rate

BSS Blind Source Separation

BVP Blood Volume Pulse

C-MCCA Connectivity Multiset Canonical Correlation Analysis

CB Convolution Block

CC Cascaded Classifier

CDF Cumulative Density Function

CEEDMAN Combining Complementary Ensemble Empirical Mode Decomposition

CLNF Conditional Local Neural Fields

CNN Convolutional Neural Network

COVID CoronaVirus Disease

CRF Conditional Regression Forest

Cust-VJ Customized Viola Jones

CVPR Computer Vision and Pattern Recognition

CWT Continuous Wavelet Transform

DCB Deconvolution Block

DL Deep Learning

DRLSE Distance Regularized Level Set Evolution

DRMF Discriminative Response Map Fitting

DT-CWT Dual Tree-Complex Wavelet Transform

ECCV European Conference on Computer Vision

ECG Electrocardiograph/Electrocardiogram

EEMD Ensemble Empirical Mode Decomposition

EEMEFN Edge-Enhanced Multi-Exposure Fusion Network

EMD Ensemble Mode Decomposition

EVM Eulerian Video Magnification

FastICA Fast Independent Component Analysis

FFT Fast Fourier Transform

FPS Frames Per Second

GAN Generative Adversarial Network

GC Gamma Correction

GMM Gaussian Mixture Model

GRRAS Guidelines for Reporting Reliability and Agreement Studies

GRU Gated Recurrent Unit

HbO2 Oxygenated Haemoglobin

Hb Haemoglobin

HE Histogram Equalization

HR Heart Rate

HRV Heart Rate Variability

IC Independent Component

ICA Independent Component Analysis

ICCV International Conference on Computer Vision

ICU Intensive Care Unit

IEEE Institute of Electrical and Electronics Engineers

IMF Intrinsic Mode Functions

iPPG imaging PPG

IR Infrared

IVA Independent Vector Analysis

J-BSS Joint Blind Source Separation

JADE Joint Diagonalization Approximation of Matrices

KDICA Kernel Density Independent Component Analysis

KinD Kindling the Darkness

KLT Kanade-Lucas-Tomasi

LED Light Emitting Diode

LEDNet Low-Light Enhancement and Deblurring Network

LMA Levenberg-Marquardt algorithm

LPIPS Learned Perceptual Image Patch Similarity

MAE Mean Absolute Error

MasKE Mask Extractor

MAICA Multi-objective optimization using Autocorrelation and ICA

MAPE Mean Absolute Percentage Error

Mask-RCNN Mask-Region-based Convolutional Neural Networks

MATLAB Matrix Laboratory

MBLLEN Multi-Branch Low-Light Enhancement Network

MCCV Multiset Canonical Correlation Variables

ME Mean Error

MER Mean of Error-Rate percentage

MeSH Medical Subject Headings

ML Machine Learning

MRD Maximum Ratio Diversity

MSE Mean Squared Error

MS-SSIM Multi-Scale Structural Similarity Index

MTCNN Multi-Task Cascaded Convolutional Neural Networks

NeRCO Neural Representation method for Cooperative low-light image enhancement

NICU Neonatal Intensive Care Units

NIR Near Infrared

NIQE Natural image quality evaluator

NLMS Normalized Least Mean Square

NN Neural Network

NPE Naturalness Preserved Enhancement

NRN Neural Representation Network

OS Oxygen Saturation

PBV Pulse Blood Volume

PCA Principal Component Analysis

PG Plethysmography

POS Plane Orthogonal to Skin

PPG Photoplethysmography

PR Pulse Rate

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PRV Pulse Rate Variability

PSNR Peak Signal-to-Noise Ratio

ReLU Rectified Linear Unit

RF Random Forest

RFB Refinement Block

RGB Red, Green, and Blue

RMSE Root Mean Square Error

ROI Region of Interest

ROR Ratio-of-Ratios

rPPG remote PPG

RR Respiratory Rate

S3FD Single Shot Invariant Face Detector

SaO2 Atrial Oxygen Saturation

SCI Self-Calibrated Illumination

SD Standard Deviation

SICE Single Image Contrast Enhancer

SNR Signal-to-Noise Ratio

SOBI Second Order Blind Identification

SOTA state-of-the-art

SRIE Simultaneous Reflectance and Illumination Estimation

SSD Single-Shot Detector

SSIM Structured Similarity Index

SSR Spatial Subspace Rotation

SaO2 Arterial Oxygen Saturation

StO2 Tissue Oxygen saturation

SvO2 Venous oxygen saturation

SpO2 Peripheral Oxygen Saturation

TBEFN Two Branch Exposure Fusion Network

TC-DCN Task-Constrained Deep Convolutional Network

TURNIP Time Series U-Net with Recurrance for NIR imaging PPG

U-ICA Undercomplete Independent Component Analysis

U-neg Undercomplete Independent Component Analysis with Negentropy cost function

U-LMA Undercomplete ICA optimized by Levenberg Marquardt Algorithm

UBFC-rPPG Univ. Bourgogne Franche-Comté Remote PhotoPlethysmoGraphy

WoS Web of Science

Chapter 1

INTRODUCTION

Physiological parameters are quantitative measures that relate to the physiology of the human body to identify health issues. Applications of investigation using these parameters include disease diagnosis, tracking immediate or long-term effects of surgery or medicinal therapy, early identification of fatal disorders, and sleep analysis [6]. Furthermore, activities occurring inside the body, such as metabolism, respiration, and oxygen levels, can be measured using physiological parameters. The five physiological parameters also called vital signs, used for determining the individual's health status are Blood Pressure (BP), Heart Rate (HR), Arterial Oxygen Saturation (SaO2), body temperature, and Breathing Rate (BR) [7]. The HR, and SaO2 are commonly used physiological parameters by physicians [8, 9] since these have been used for health monitoring in various scenarios like Intensive Care Units (ICUs), surgery, etc. Due to the limitations of corresponding gold-standard techniques (presented in later sections) of measuring these parameters and the portability and simplicity of Photoplethysmography (PPG), it has been extensively used for these measurements. However, contact-based PPG also possesses limitations, concretely, it can cause discomfort to the subject in the scenarios such prolonged monitoring, burnt or sensitive skin. Therefore, remote PPG (rPPG) can be used as an alternative in such scenarios. Furthermore, most rPPG methods were tested in ambient light conditions, which limits their applicability in real time scenarios such as sleep monitoring or nighttime driving with insubstantial illumination conditions. This thesis aims to enhance the applicability of rPPG methods for such scenarios. To address this, this chapter presents motivation for conducting this study, followed by proposing and addressing the relevant research questions. Finally, the research objectives addressing the research questions, brief description of the thesis, and corresponding scientific contributions were presented in this chapter.

1.1 MOTIVATION

The non-contact estimation approach involves recording the subject's face video in the presence of a light source. Ambient light is predominantly used for physiological parameter estimations, while other sources could also be used, such as fluorescent, ceiling, and incandescent lights [3, 10]. Furthermore, the majority of the non-contact estimation studies used Red, Green, and Blue (RGB) color channels, whereas only a few studies have used Infrared (IR) channels [11–15]. Most estimation studies were conducted in a well-controlled laboratory environment, whereas relatively fewer studies were designed to estimate physiological parameters in clinical settings such as dim light or darkness. The existing rPPG estimation methods need an appropriate selection of Region of Interest (ROI), which might be challenging in real-time clinical conditions, resulting in inaccurate estimations. The IR spectra can be used for dealing with such scenarios, but its limitation lies in its ability to achieve relatively weaker pulsatile strength, unlike visible spectra [16]. Additionally, most studies have used self-created databases to present the effectiveness of their non-contact methods. On the contrary, publicly available databases can also be used, which provides relatively challenging conditions for parameter estimations. This study will test state-of-the-art (SOTA) methods for three publically available datasets.

However, limited estimation studies have been conducted in dark environments where darker environments, such as sleeping or nighttime driving conditions. Studies of such kind have used IR spectra for video recording besides relatively weaker PPG pulsatile strength than RGB spectra. Various research developments in the image processing domain provide tools to extract human imperceptible information from the images, which can be recorded using an RGB camera [17]. Despite this, no study has attempted physiological parameter estimations in the dark environment using visible light spectra. Therefore, it is vital to check the feasibility of RGB image color model to estimate physiological parameters in dark environments mimicking sleeping conditions or nighttime driving scenarios using the non-contact variant of PPG i.e., rPPG. Pursuing this research direction will also need the development of relevant databases, which are not available to date, to the best of the author's knowledge. Therefore, this thesis aims to develop a dataset consisting of videos captured in dark environments using RGB image color model. Subsequently, novel methods will be designed to estimate HR and Peripheral Oxygen Saturation (SpO2) in dark environments. This study will facilitate the applicability of non-contact approaches in clinical conditions because these conditions include dim or dark light conditions

instead of bright light, as used by almost all non-contact estimation studies. To be precise, the dark environment for the experiments conducted during this study is defined as the environment where the average illuminance does not exceed 1.0 lux, as measured by the lux meter (full description presented in the later sections).

1.2 RESEARCH QUESTIONS

This proposal aims to address the following research questions:

1. Is it possible to use RGB videos to estimate physiological parameters in darkness, where darkness is defined as the environment with average illuminance less than or equal to 1.0 lux?
2. Is it feasible to combine image processing methods with conventional SOTA methods to estimate HR and SpO2 in a dark environment?
3. Is it possible to achieve relatively similar performance with the RGB modality as an IR modality in a dark scenario?

1.3 OBJECTIVES

The study primarily aims at designing a non-contact RGB-based physiological parameter estimation approach in a darker environment condition mentioned in section 1.1. Specifically, a database considering the darkness condition will be created by recording facial videos with ground truth HR and SpO2 estimates. Subsequently, the estimation method will be developed in an ambient light environment and tested with publically available databases. This method will then be cascaded with a novel deep learning-based image enhancement method to test its feasibility under darker conditions. The objectives of the study can be summarized as:

1. To identify the best performing non-contact HR and SpO2 estimation methods based on a systematic analysis of the relevant SOTA methods.
2. To develop novel rPPG extraction method under normal lighting conditions and test using publically available databases.
3. To develop an image enhancement method to preserve the illuminance of the dark videos for accurate ROI extraction to facilitate the physiological parameter estimations in dark environments.

1.4 THESIS DESCRIPTION

Based on the research questions (section 1.2) and the objectives (section 1.3), this thesis aims to address the need for physiological parameters, namely, HR and SpO2 in dark environments, to scale up the ability of non-contact estimation methods for scenarios such as sleep monitoring, nighttime driver monitoring, ICU, home-based monitoring, etc. To test the feasibility of RGB spectra for physiological parameters estimations, a database was developed which includes a reference image taken in an ambient light environment, a video captured in the dark environment, and mean ground truth HR, and SpO2 values. The details about this database are explained in chapter 3.

Concurrently, a systematic analysis was conducted to identify the best SOTA HR, and SpO2 estimation methods, which resulted in a journal publication (2 of section 1.5). Additionally, it resulted in an amalgam of critical parameters required for conducting a standardized rPPG study. For instance, this analysis provides information about useful performance metrics with lower and upper thresholds to measure the efficacy of the new methods, best SOTA non-contact estimation methods, etc. Chapter 4 presents a detailed view of this systematic analysis. Based on the key findings from the systematic analysis, Independent Component Analysis (ICA) was the predominant method used for rPPG signal extraction.

After critically analyzing the ICA-based rPPG studies, it was found that ICA suffers from ordering problem, which resulted in a challenge to identify the appropriate component containing PPG information. Therefore, Undercomplete ICA optimized by Levenberg Marquardt Algorithm (U-LMA) was developed to elevate the abovementioned challenge by assuming rPPG signal extraction as an undercomplete problem (a single rPPG signal to be extracted from three mixture signals corresponding to RGB channels). U-LMA consists of three modules: 1) an entropy-based cost function to ensure the statistical independence of the cumulative density function of the Independent Component (IC); 2) Undercomplete Independent Component Analysis (U-ICA) to solve the assumed undercomplete problem; and 3) customized Levenberg-Marquardt algorithm (LMA) for entropy maximization. Since rPPG methods exhibit suboptimal performance in the presence of rigid and non-rigid motion and illumination variations, the newly developed method was tested in both scenarios. The explanations of the proposed method and respective results are presented in Chapter 5. This part of the work resulted in a journal publication (1 of section 1.5).

Furthermore, a deep learning-based image enhancement method named 2E1D-Net (named based on its architecture, i.e., two encoders and one decoder) was proposed and combined with U-LMA and Ratio-of-Ratios (ROR) methods for non-contact HR and SpO2 estimations in the assumed dark environment (average illuminance ≤ 1.0 lux), respectively. Although the combination of U-LMA with 2E1D-Net performed best, 2E1D-Net also expanded the ability of SOTA HR estimation methods in the dark environments. On the other hand, ROR, in conjunction with 2E1D-Net, also expanded the ability of ROR in dark environments. This experiment presented a proof-of-concept where RGB spectra can be used for dark environment physiological parameters estimation, unlike the conventional approach of using IR spectra without a performance compromise. Chapter 6 presents the details of the developed methods and respective results implemented on a newly proposed dataset during this thesis. This resulted in a provisional patent application, and potential journal publication (4 and, 5 of section 1.5).

Finally, chapter 7.1 concludes this thesis by presenting the key findings and limitations, followed by future research directions.

1.5 SCIENTIFIC CONTRIBUTIONS

The scientific contributions of the proposed work have been presented in three journal articles, one provisional patent application, and one book chapter, whose details are as follows:

1. Gupta, A., Ravelo-García, A. G., and Dias, F. M. (2022). A motion and illumination resistant non-contact method using undercomplete independent component analysis and Levenberg-Marquardt algorithm. IEEE Journal of Biomedical and Health Informatics, 26(10), 4837-4848.
2. Gupta, A., Ravelo-Garcia, A. G., and Dias, F. M. (2022). Availability and performance of face based non-contact methods for HR and SpO2 estimations: A systematic review. Computer Methods and Programs in Biomedicine, 219, 106771.
3. Gupta, A., Ravelo-Garcia, A. G., and Dias, F. M. (2023), Recent Advancements in Deep Learning-based Remote Photoplethysmography Methods, Data Fusion Techniques and Applications for Smart Healthcare, Elsevier.
4. Gupta, A., Dias, F. M., and Ravelo-Garcia, A. G., System and Method for Estimating At Least One Physiological Sign of a Subject in Dark Environments (PCT application number PCT/IB2024/053491).

5. Gupta, A., Ravelo-Garcia, A. G., and Dias, F. M. (2023), RGB, a Surrogate of Infrared Facial Videos for Physiological Signs Estimations in Dark, IEEE Transactions on Circuits and Systems for Video Technology (*In peer review*).

1.6 RESEARCH APPROVALS

The Ethical Commission and Data Protection Committee of the University of Madeira approved the data collection process for this study with application number 1/CEU-MAI/2021. The respective approval documents are presented in Appendix.

Chapter 2

BASIC CONCEPTS IN THE CONTEXT OF THIS THESIS

This project aims to design novel non-contact approaches for HR, and SpO2 estimations. The HR can be measured by detecting the pulse through organs such as the finger, earlobe, etc., therefore also called Pulse Rate (PR) (a surrogate of HR), while arterial oxygen saturation is called SpO2. However, HR and Pulse Rate (PR) were being used in the literature interchangeably, and this report also uses the term HR in place of PR for consistency. Since the thesis deals with measurement of above mentioned variables, this chapter first explains the importance and application of these variables along with a brief description of PPG. Subsequently, the underlying principles challenges associated with PPG and its non-contact variant, along with instruments of measuring PPG signals, are also presented in this chapter. Finally, it also sets the context and presents the research questions, objectives, and scientific contributions associated with this project.

2.1 HEART RATE MEASUREMENT

HR is the number of heart beats per minute (bpm), where each heartbeat corresponds to synchronous contractions and relaxation of the heart while blood is pumping. A healthy individual's heart beats between 60 to 100 times per minute. HR less than 60 or more than 100 bpm are considered abnormalities, termed bradycardia and Tachycardia, which must be identified by continuous HR monitoring at the earlier stage to avoid situations such as cardiac arrest or sudden death [18]. Furthermore, HR can also reveal valuable information in conditions such as high-level stress and emotion elicitation. Therefore, HR estimation has been an active area of research in biology and medicine [19]. HR can be estimated by electrocardio-

gram waveform based on R-R intervals (inverse of R-R intervals is HR). A pictorial representation of the Electrocardiograph/Electrocardiogram (ECG) waveform of a healthy individual is illustrated in Figure 1.

With every heartbeat, the blood travels from the heart to other body organs using blood vessels, which creates a pressure pulse called cardiac pulse or BVP. Detection of this pulse via body parts is known as Plethysmography (PG), in which cardiac pulse is measured by compressing the artery at the wrist, inside the elbow, knee, or ankle joint region. Furthermore, among variants of PG, the one exploiting the optical properties of skin tissues, i.e., PPG, has been widely studied due to its lesser limitations than other variants [20]. As mentioned earlier, HR measured using PPG is called PR since it is extracted from the pulse signal. HR measurement using ECG

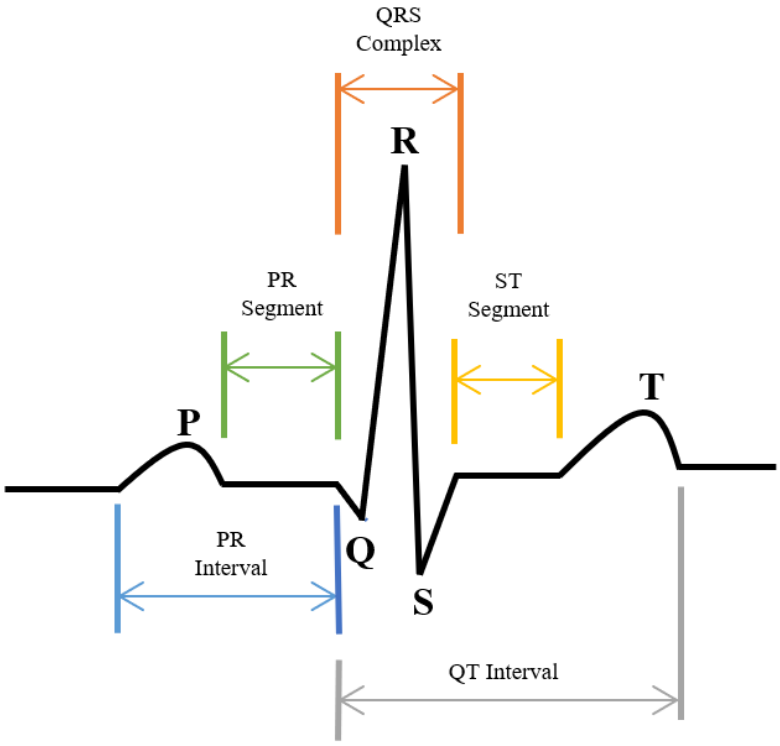


Figure. 1: ECG waveform

requires accurate placement of electrodes in contact with the skin using adhesive gels [21], whereas a spring-loaded pulse oximeter probe needs to be placed in contact with the skin for estimating PR. Therefore, the performance of both techniques depends on the correct placement of electrodes or probes, which otherwise leads to misleading results. Additionally, these techniques are unsuitable for prolonged monitoring since the continuous placement of electrodes or probes might cause discomfort or irritation in the exposed body parts. In addition, among ECG and PPG, the latter, a relatively simple and portable non-invasive technique, has been widely used in wearable and non-contact estimation and monitoring systems. Although a non-contact version of

ECG was also proposed by Fong and Chung [22], in which the sensor is placed in the chair, the proposed apparatus suffered from similar limitations as a contact-based approach, i.e., restrictive movement and accurate point of contact for non-spurious contact.

2.2 SpO2 ESTIMATION

Oxygen Saturation (OS) is a physiological variable that indicates the percentage of oxygen-bonded Haemoglobin (Hb) in the blood, which is necessary for the survival of body cells or tissues. OS value for healthy individuals ranges between 96-99% and not less than 88% during sleep. SpO2 measured using the body organs such as a finger or earlobes is called SpO2. It has been used in various scenarios, such as ICU, surgery, neonatal care, sleep disorders identification, and anesthesia. It also reflects the cardiorespiratory health of the individual. Moreover, continuous monitoring of SpO2 helps in the early detection of hypoxemia conditions. SpO2 measurement is based on PPG and is performed using a spring-loaded oximeter probe (PPG), which can be placed at the body organs such as the finger, earlobe, etc.

Typically, OS can be estimated from the pulmonary artery of the heart, tissues, and arteries, respectively. The estimation of SpO2 from the pulmonary artery, i.e., mixed Venous OS (Venous oxygen saturation (SvO2)), can be performed using fiber optic catheters. Tissue OS (Tissue Oxygen saturation (StO2)) estimation is performed using Near Infrared (NIR) spectroscopy, which also provides the assessment of tissue perfusion. Finally, OS estimation from arteries (SaO2) is performed using a blood gas analyzer via blood tests. Commonly used OS assessments use peripheral capillary, therefore called Peripheral SpO2, which is estimated using an optoelectric pulse oximeter spring-loaded probe. The pulse oximeter comprises two LEDs of different wavelengths and a photodetector to trap transmitted light post absorption by skin tissues and blood [1].

Conventionally, SpO2 estimation is computed using the ratio of Oxygenated Haemoglobin (HbO2) and deoxygenated Hb molecules of red blood cells. The SpO2 is estimated using PPG, which is based on different absorption wavelengths (red and NIR) of HbO2 and Hb. The potential reason for these wavelengths is that PPG amplitude near $600-700nm$ (red) is sensitive to the saturation level of arterial blood, while $800-950nm$ (NIR) accounts for temperature and HbO2, and deoxygenated Hb variations, respectively, as presented in Figure 2.

SpO2 is prominently estimated using the ROR method, which is the ratio of normalized PPG amplitudes of red and NIR wavelengths. In other words, the nor-

malized PPG amplitudes are calculated by taking the ratio of AC (pulsatile) and DC (slowly varying) components for red and NIR wavelengths separately. Subsequently, the ROR is finally calculated as ratios of normalized PPG amplitude of these wavelengths followed by mapping the ROR values with ground truth SpO2 value. Mathematically, it can be defined as:

$$SpO_2 = A + B \times \frac{\frac{AC_{red}}{DC_{red}}}{\frac{AC_{NIR}}{DC_{NIR}}} \times 100 \tag{1}$$

where A and B are coefficients and can be estimated using regression analysis. SpO2 estimation studies carried out in the visible spectrum used red and blue channels to compute ROR.

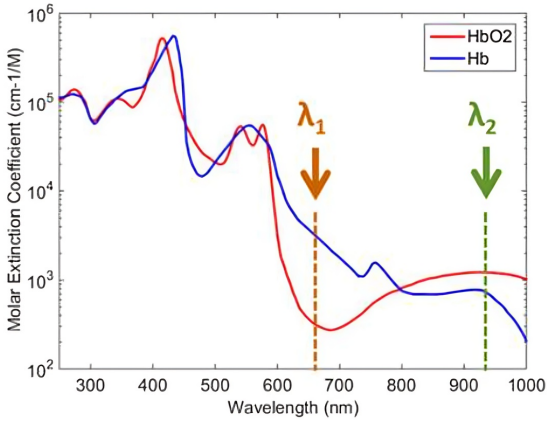


Figure. 2: Absorption spectra of oxygenated (HbO2) and deoxygenated (Hb) hemoglobin in red blood cells [1]

However, a study by Moço and Verkruysse [3] also attempted to replace blue with green channels but concluded that the ROR method using red and blue channels is more accurate without providing a potential reason. On the other hand, Gastel et al. [23] proposed a different approach by mapping BVP signals to different ranges (with an interval of 5 values each) of SpO2 values in the range of $65\% - 100\%$. Non-contact SpO2 estimation is very challenging for abnormal values, for instance, $\leq 90\%$, as it requires participants to hold their breath for a certain amount of time, which could make the participant uncomfortable and may cause casualties.

2.3 PHOTOPLETHYSMOGRAPHY

PPG is a simple, non-invasive, inexpensive technique that exploits the optical properties of the skin tissues to detect the subtle variations in the blood volume synchronized with the cardiac activity. Precisely, it measures the blood volume changes in the microvascular tissues underneath the skin, synchronized with cardiac activity,

resulting in a signal. Although PPG is non-invasive, it has been predominant in spring-loaded optoelectric pulse oximeters(3a), which has limited applicability for unobstrusive or prolonged monitoring. Therefore its non-contact variant of PPG, i.e., rPPG, was introduced which is suitable for prolonged and unobtrusive monitoring. Similar to PPG,it aims to extract a PPG, i.e., rPPG signal, containing PPG information, in proportion to the blood volume of the skin, which can be used to estimate physiological variables such as HR, SpO2, BR, BP, etc. Additionally, it also provides variables for vascular assessment, such as arterial disease, aging, stiffness, etc., and autonomic function assessment, such as Pulse Rate Variability (PRV).

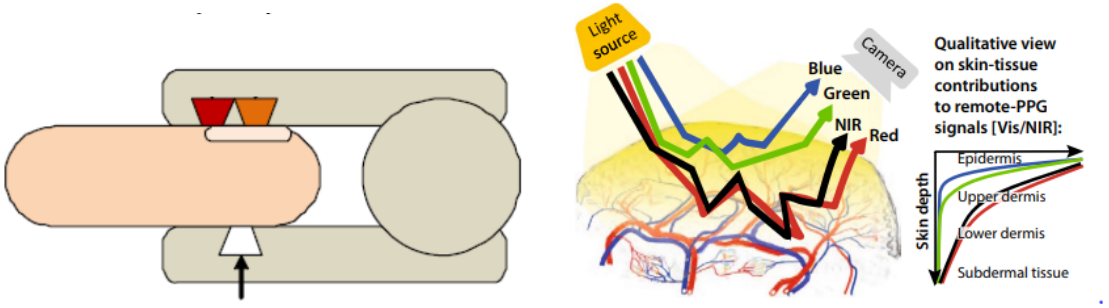
The PPG signal waveform consists of AC and DC components. The AC component corresponds to subtle variations in blood volume synchronized with the cardiac activity. Precisely, when the heart pumps blood from different organs of the body i.e., systole, the blood volume in the capillaries increases, eventually increasing the light absorption. On the other hand DC component of the PPG signal comprises reflected components resulting from the total red blood cell volume and skin, and bones. It includes information related to respiration, venous flow, the sympathetic nervous system, and thermoregulation. Accurate measurement of thePPG signal is dependent on multiple factors, such as the location and architecture of the skin region. The apparatus used for PPG analysis is shown in Figure 3a, which consists of the illumination source(s), emitting constant light radiation, and a photodetector for capturing the resultant light [24]. The detailed explanation of the underlying principle of PPG and measurement apparatus is presented is the subsequent sections.

2.3.1 Working principle and measuring instrument

The working principle of PPG is relatively simple and easy to analyze, as depicted in Figure 3b. Light, while interacting with the biological tissue, gets absorbed by tissue, bones, and blood. Additionally, most of the light is predominantly absorbed by the blood, while its few fractions are scattered due to the opaque properties of the biological tissues. The critical components affected by this light absorption are blood volume, blood vessel walls, and red blood cell orientation. Therefore, skin (in proximity to the blood vessels) is being used for PPG analysis. The skin comprises the following layers: epidermis, dermis (Upper and lower), and Subdermal (hypodermis), as shown in Figure 3b. The epidermis is the visible part of the skin, followed by the upper and lower dermis and subdermal layers constituting the vascular bed (consisting of arteries and veins). Heart pumping allows the blood to flow through multiple veins in dermis layers, which further triggers the arteries and arterioles in

the subdermal layer of the skin. The blood flow in arteries and arterioles can be used to detect the pulsatile component of the cardiac cycle [25].

The PPG signal is measured using optoelectronic oximeters whose schematic representation is shown in Figure 3a. A typical oximeter consists of two light-emitting diodes emitting red and IR wavelengths and a photodetector for capturing the transmitted light post absorption. There are three reasons to use Light Emitting Diodes (LEDs) of visible red and NIR: 1) most of the light is absorbed by the water present in the blood, but only visible red and NIR wavelength radiations can penetrate deeper into vascular bed to detect subtle variations in blood volume; 2) the 805nm wavelength is an isosbestic wavelength which allows the signal to remain unaffected from the effect of SpO2; and 3) these wavelengths can penetrate through the volume of $1cm^3$ for transmission mode PPG [25]. Different penetration depths of visible and IR light radiations are illustrated in Figure 3b. If the skin is not compressed, the change in the blood volume resulting from the skin illumination by LED reflects the cardiovascular pressure wave. Therefore, the PPG signal is often extracted from the optoelectronic pulse oximeter [24].

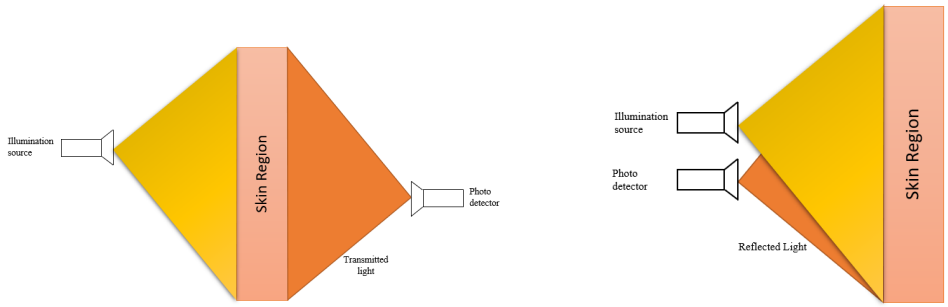


(a) (b)

Figure. 3: Pulse oximeter [2] (a) and working principle of PPG [3] (b)

2.3.2 PPG operational configurations and signal characteristics

Based on the operational configuration, PPG can be subclassed into transmission (Figure 4a) and reflection mode (Figure 4b) PPG. In transmission mode, PPG, the illumination source, i.e., LEDs and the photodetector, are placed opposite to each other. The light transmitted through the skin tissue is captured by a photodetector for PPG signal extraction. On the other hand, for reflectance mode, PPG, the LEDs, and the photodetector are placed on the same side, where the photodetector captures the reflected light from the skin tissue and is eventually used for PPG signal extraction.



(a) (b)

Figure. 4: PPG Operation configuration: Transmission mode PPG (a), reflection mode PPG (b)

The transmission mode PPG has been used by pulse oximeters using body organs such as the finger, earlobe, toe, etc. While the reflectance mode PPG has been widely used for non-contact PPG using various body organs such as the face, finger, palm, etc. The former imposes more restrictions than the latter regarding body movement and accurate placement of the oximeter probe, while overcoming the limitations due to the non-contact nature of the latter.

A PPG signal resulting from placing a finger in the pulse oximeter probe is shown in Figure 5. As mentioned earlier, it consists of a pulsatile component, also called the AC component, which is the source of cardiac information, and a slowly varying DC component due to multiple factors such as vasomotor activity, respiration, etc. Additionally, the AC component of the PPG signal consists of two peaks: the rising edge in the PPG signal, also called the anacrotic phase, which corresponds to systole; the falling edge, also called catacrotic phase corresponding to the diastole and wave reflection from the periphery. The PPG analysis eliminates the effect of slowly varying dominant DC components by high band pass filtering [13].

2.3.3 Photoplethysmography challenges

The conventional approach to the PPG signal extraction is by using optoelectronic pulse oximeters, which require the placement of an oximeter probe on the body organs such as the finger, toe, earlobe, etc. The disadvantages of using the contact-based probe are multifold: prominent prevalence of artifacts due to skin compression, restricted movement of body locations, and limited applicability in a few scenarios. If the skin is compressed while extracting the PPG signal, it may likely have the prominent prevalence of the other artifacts dominating the cardiac information. A minute deflection of the skin tissue from the actual position may result in a noisy signal. These probes can cause allergies, itching, or infection in scenarios such as prolonged monitoring, burnt or sensitive skin, and Neonatal Intensive Care Unitss

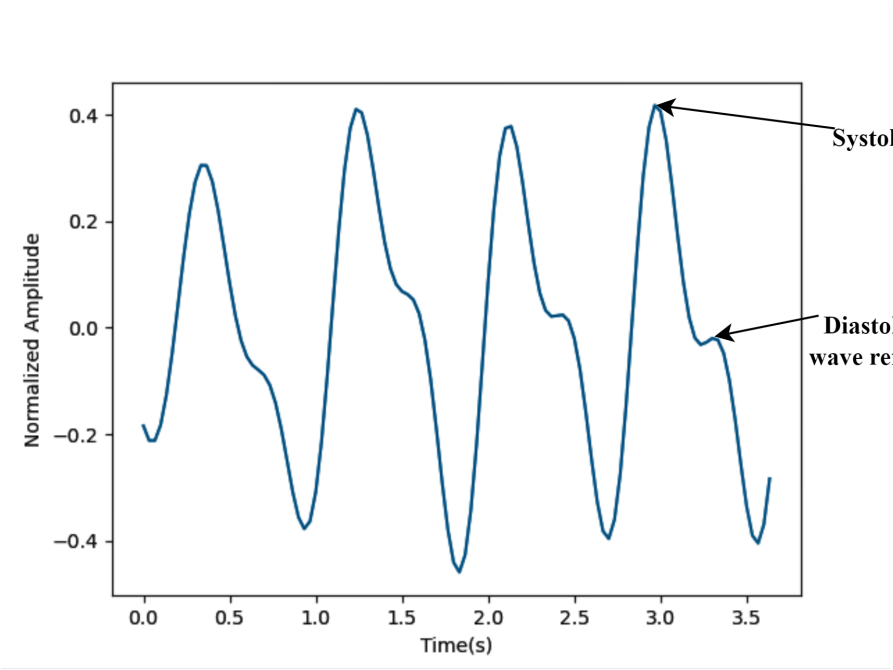


Figure. 5: A schematic representation of a PPG Signal.

(NICUs) [26,27]. Figure 6 illustrates the challenges associated with the contact-based PPG approach.

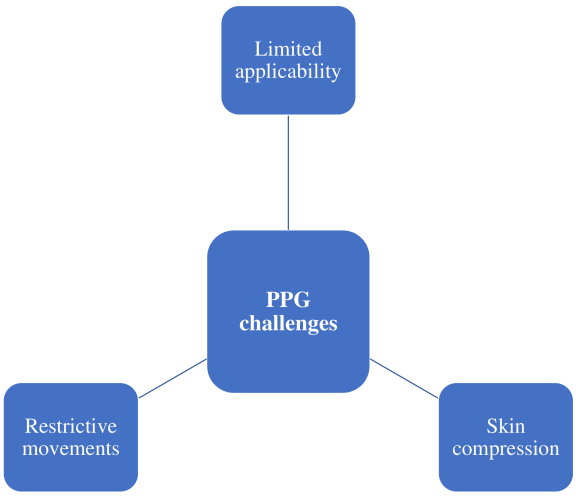


Figure. 6: A flow chart depicting challenges associated with contact-based PPG approach

2.4 NON-CONTACT PPG

Non-contact PPG, also known as rPPG [?] or imaging PPG (iPPG) [28], is based on the reflectance mode PPG, wherein the illumination source and photodetector are

placed on the same side. The first rPPG based non-contact physiological variable estimation study was conducted by Verkruysse et al. [29]; since then, numerous studies have been conducted focusing on devising non-contact methods for efficiently extracting PPG information.

These methods primarily extract the PPG information from the subject's video by extracting the subtle color variations based on the reflected light intensities from the consecutive image frames of the face. Essentially, these methods follow a three-step procedure: 1) face detection followed by skin segmentation (ROI extraction), 2) extraction of rPPG signal, and 3) frequency identification of the corresponding physiological variables to be estimated, except SpO2 estimation. Each step poses different challenges to ensure accurate extraction of rPPG signals, eventually accurate physiological variable estimations. For instance, inappropriate ROI selection results in insufficient PPG information extraction, which may lead to false estimates. A schematic representation of the rPPG methods workflow is presented in Figure 7.

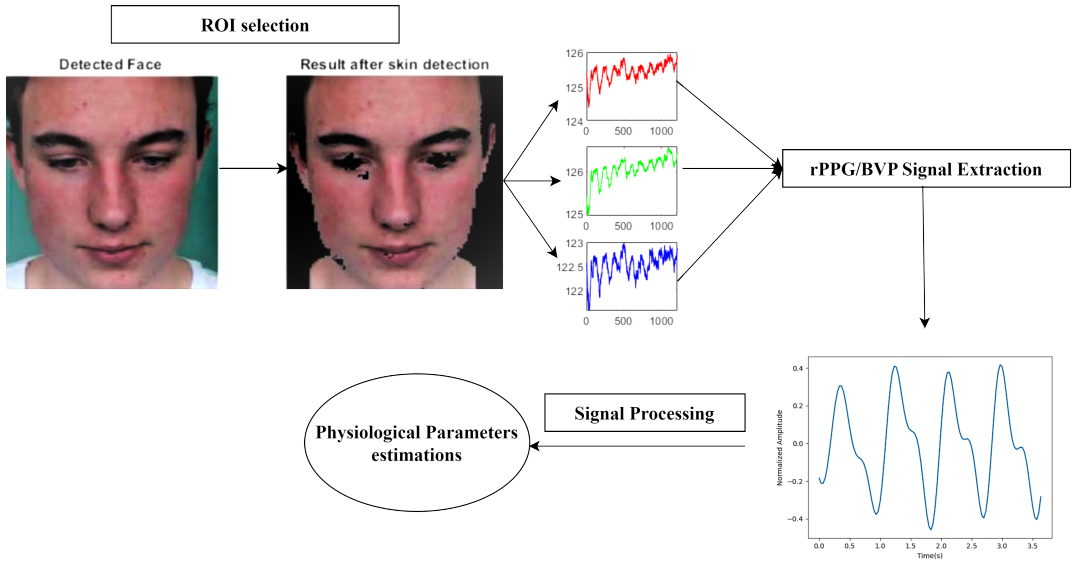


Figure. 7: A schematic representation of rPPG methods workflow.

2.4.1 Region of interest selection

Conventionally, the ROI selection for facial-based rPPG methods consists of two steps: 1) Face detection and 2) skin segmentation. However, recent methods have predominantly used attention mechanisms to select appropriate ROI, which is abundant in PPG information. Since conventional HR and SpO2 methods have used both steps, this section presents SOTA face detection and skin segmentation methods, respectively.

a) Face detection

The face detection for the rPPG signal extraction method can be broadly categorized based on multiple factors, i.e., face bounding box or facial landmarks extraction, motion constraints, and occlusion tolerance. Face bounding box methods include Viola-Jones [30] and Single Shot Invariant Face Detector (S3FD) [31]. On the other hand, the facial landmarks methods include Convolutional Neural Networks (CNNs)-based Multi-Task Cascaded Convolutional Neural Networks (MTCNN) [32], DLib (TC-DCN) [33], blazeface [34], and other Machine Learning (ML) algorithms-based methods such as Conditional Local Neural Fields (CLNF) (Openface) [35], and Conditional Regression Forest (CRF) [36]. However, RetinaFace [37] extracts both facial landmarks and the bounding box, providing substantial flexibility based on the problem. The characteristic features of face detection methods are presented in Table 1.

Viola-Jones face detection method uses a Cascaded Classifier (CC), employing Haar-like features [38]. While S3FD can detect faces at various spatial scales using anchor tiling, matching, and max-ground false labeling. The CNN-based methods, such as MTCNN, exploit the correlation between face alignment and detection by its cascaded architecture with three-stage convolution networks for face classification, bounding box, and facial landmarks detection. In contrast, the DLib face detection module uses Task-Constrained Deep Convolutional Network (TC-DCN), which is optimized for pose variations and occlusions.

A mobilenetV2 [39] type Blazeface face detection method is based on a single-shot detector with a better tie resolution strategy, primarily designed for augmented reality and smartphone devices. Other methods, such as the Openface toolkit, use the CLNF model [40], which consists of a point distribution model and patch experts for landmark shape (e.g., based on eye region orientation) and appearance variations respectively. CRF is a Random Forest (RF)-based motion robust algorithm that aims to explore the relationship between facial image patches and feature points from the set of faces, thereby learning conditional to global face properties.

On the other hand, the only face detection method providing both a bounding box and facial landmark points is RetinaFace. It is also a Single-Shot Detector (SSD) based on pixel-wise localizations on various spatial scales using joint and self-supervised multi-task learning.

Viola-Jones face detection algorithm has been predominantly used in rPPG signal extraction methods due to its robustness and faster execution. At the same time, RetinaFace is the second most used method due to its ability to detect faces under

different spatial dimensions due to efficient pixel-wise localizations. Besides, facial landmark point detection methods are often preferable for rPPG studies since these points can be tracked throughout all the consecutive image frames without any computational burden, unlike bounding box methods (bounding box extraction from each image frame).

Table 1: Characteristics of Face detection methods

Detection Methods	**Architecture**	**Occlusion**	**Motion**	**L/B**	**Facial points**
Viola-Jones	CC	No	No	B	NA
S3FD	SSD	No	No	B	No
MTCNN	CNN	Yes	Yes	L	5
Dlib (TC-DCN)	CNN	Yes	Yes	L	68
Blazeface	SSD	No	yes	L	6
OpenFace	CLNF	No	Yes	L	68
CRF	R	F No	Yes	L	10
RetinaFace	SSD	No	No	Both	5

Note: L and B stand for Landmarks and Bounding Box, respectively.

b) Skin segmentation

Skin segmentation is the process of selecting the appropriate ROI, which contains substantial PPG information. Due to the uneven distribution of PPG information throughout the face region, it is challenging to select suitable skin regions that contribute to accurate rPPG signal extraction and, eventually, HR and SpO2 estimations. Alternatively, inappropriate skin region selection leads to noisy rPPG signal, which eventually results in false estimates.

Conventional skin segmentation-based methods were based on projecting RGB values to other image models, such as YCbCr, followed by identifying the thresholds for each channel of the projections [41]. The disadvantage of this approach is its limited applicability due to fixed threshold values, which might sometimes lead to the inclusion of non-skin pixels. Consequently, Deep Learning based skin segmentation methods have been proposed, such as a study proposed by Topiwala et al. [42], which employed SOTA architectures such as Mask-Region-based Convolutional Neural Networks (Mask-RCNN) [43] and UNet [44]. Deep Learning-based methods have comparatively better skin segmentation ability due to their robustness and good generalizability independent of the conditions. However, due to the uneven distribution of the amount of rPPG information over the entire facial region, these methods also failed to achieve better implications in this field.

Therefore, current studies rely on developing soft-attention mechanisms for accurate ROI selection [12, 14, 15, 45, 46]. The benefit of using soft-attention mechanisms is twofold: first, these are very lightweight networks with faster execution time; second, they can select the regions of skin containing PPG information and discard the regions with noise due to motion and illumination artifacts as proven by the study conducted by Nowara, McDuff, and Veeraraghavan [15]. Therefore, the researchers have started designing soft-attention networks to improve the accuracy of the PPG signal. Face detection, in conjunction with the soft attention mechanism, drastically reduces the computational complexity and execution time of the rPPG signal extraction method due to the exclusion of an explicit skin segmentation module.

2.4.2 Remote Photoplethysmography signal extraction for heart rate estimation

The rPPG signal extraction methods typically suffer from motion, illumination variation artifacts, and camera quantization noise. Consequently, the proposed methods revolve around alleviating the effects of these artifacts to ensure accurate rPPG signal extraction. Existing SOTA methods primarily used visible [11, 12, 45–57] or IR cameras [11–15] since visible spectra exhibit relatively stronger pulsatile strength, while the IR spectrum is resistant to illumination variation artifacts. Furthermore, few studies have explored several other modalities, such as microwave radars [58]. Due to the expensive cost of operating instruments and their higher noise sensitivity, these studies are limited to studies utilizing visible and NIR cameras. rPPG elevates the need to place the sensor in contact with the skin, thereby avoiding the risk of skin compression, and can be used in various scenarios such as non-contact sleep monitoring, NICUs, sensitive or burnt skin, etc. Although the rPPG methods used diversified concepts, they can be broadly classified into Blind source separation, color subspace transformations, wavelet, and neural network-based methods. The details of the corresponding studies proposing these methods are presented in Table 2. Additionally, a detailed categorization of rPPG methods is presented in Chapter 4.

a) Blind Source Separation

Blind Source Separation (BSS) has been used in rPPG signal estimation since Poh et al. [4] utilized it for PPG extraction, followed by HR estimation. The advantage of using BSS-based methods is that they do not need any apriori information for

signal extraction. The underlying principle behind BSS techniques is to disentangle the PPG signal from raw RGB signals by maximizing the statistical dependence (ICA) or variance (Principal Component Analysis (PCA)) between PPG and other components. Among BSS methods, ICA has been predominantly used for rPPG signal extraction. A few studies also used PCA for rPPG signal extraction, for instance, rPPG study by Lewandows et al. [59].

It is worth mentioning that PCA is sensitive to its parameters, such as the choice of the number of eigenvectors, and is inaccurate for complex relationships among variables. Thus, it was not often used for rPPG signal extraction task, unlike ICA. However, ICA suffers from an ordering problem, i.e., there is no order of independent components based on statistical independence, which makes it challenging to select a component containing PPG information. Therefore, different variants of ICA have been proposed in the literature. For instance, to overcome the challenges associated with appropriate IC, a representative of rPPG signal, constrained ICA was proposed, which ensures statistical independence using a negentropy-based objective function to avoid local minima convergence [60]. Since constrained ICA is slow,Fast Independent Component Analysis (FastICA) has been introduced for rPPG signal extraction followed by average HR estimations from RGB-NIR fusion [16]. Exploiting the periodicity characteristics of the PPG signal, a semi-blind source separation with an objective function, a combination of autocorrelation and negentropy, was also proposed [61]. Similarly, Joint Blind Source Separation (J-BSS) [62] has also been proposed to extract cleaner rPPG signals by exploring the dependencies among video samples.

b) Color subspace transformation

The color subspace transformation methods work on projecting the motion and illumination-affected signals to orthogonal projection planes to separate pulse information and other noisy components. These methods are called color subspace transformations because they aim at eliminating the color distortions due to artifacts by transforming the color subspace to orthonormal projection planes. Most color subspace transformation methods have used two commonly used color subspace projections, CHROM [**?**] and Plane Orthogonal to Skin (POS) [63], as baseline methods. Interestingly, the limitation of CHROM, i.e., abundant information gathering due to different skin tones, was addressed, and POS was proposed. Subsequently, addressing the dimensionality of the motion artifacts, a new method was proposed, which divided the signals based on orthogonal frequency bands, resulting in signal fragments. Subsequently, the POS method was applied to these signal fragments.

Finally, the signal fragments were fused to create a rPPG signal. One of the limitations of the color subspace projections is to use the fixed projection planes, which was resolved by introducing Adaptive Pulsatile Plane (APP), which elevated the problem using fixed ROI and skin tones using fixed thresholds [64].

c) Wavelet transforms based methods

One of the predominately used wavelet processing methods in the rPPG literature is Empirical mode decomposition, which decomposes the signal into Intrinsic Mode Functionss (IMFs) of varying frequencies and amplitudes. However, it suffers from a mode aliasing problem [65]. Hence, Ensemble Empirical Mode Decomposition (EEMD) for PPG extraction was introduced [66]. The method used reflectance decomposition, which was based on Weber's law for discriminating the reflectance from illumination, followed by reducing the effect of illumination variations on the facial region. The resultant signal traces were then applied to EEMD, followed by Ensemble Mode Decomposition (EMD), for splitting it into IMF. The reason behind using EMD after EEMD is that the latter did not give real IMF due to averaging and summation of noise to each IMF. The termination criteria used for EEMD was the predefined Standard Deviation (SD) threshold, while for EMD, it was S-number, which is the number of consecutive iterations until zero crossing, and the extrema number is equal or differs by 1 [21, 66].

Although this method has been successfully utilized EEMD for BVP signal extraction, signal integrity is a primary concern while using EEMD, i.e., it may halt signal integrity. Therefore, a method consisting Combining Complementary Ensemble Empirical Mode Decomposition (CEEDMAN) in conjunction with permutation entropy was proposed [65]. Specifically, the BVP signal was denoised using the CEEDMAN algorithm with permutation entropy as an objective function with an upper and lower mode threshold of 0.3 and 0.7. Permutation entropy greater than 0.7 was further decomposed using wavelets. The problem with this method is the use of manual parameters set empirically, which may not work for all scenarios due to various factors such as skin tone.

Table 2: Summary of rPPG methods.

Baselines	**First Author**	**Physiological Variable**	**ROI Used**	**Method**	**Color channel**	**Limitations**
J-BSS	H Qi [62]	HR	Face	J-BSS, C-MCCA	RGB	The method did not address the motion and illumination artifacts.
	J Cheng [28]	HR	Face	IVA	NIR	The study did not consider motion and illumination artifacts.
BSS	R Song [67]	HR	Cheeks	KDICA	RGB	This study analyses the effect of resolution using KDICA did not propose their algorithm. Hence, the corresponding limitations could not be realized.
	C Zhang [68]	HR	Eyes Area	SOBI	RGB	The method limits its applicability to higher degrees of freedom.
	M Poh [4]	HR	Face	ICA	RGB	The method would not work well under rigid movements and different illumination conditions.
	G Tsouri [60]	PR	Face	ICA	RGB	The constrained ICA is 30 times slower than the ICA.
	R Macwan [61]	HR	Face	ICA	RGB	The proposed method uses periodicity as one of the criteria for BVP selection, which limits its applicability for estimation during periodic movements.

	S Kado [16]	HR	Cheeks	FastICA	RGB, NIR	The method's performance will degrade in case of larger HR spreads and longer intervals of motion or illumination variations.
Color Subspace Transform--ation	Q. Tran [64]	HR	Face	APP	RGB	The method works poorly if HR frequency lies in the motion spectra.
	G Haan [69]	HR	Face	CHROM	RGB	CHROM method uses skin standardization and fixed projection planes, which halts its generalizability.
	W Wang [70]	HR	Face	Subband separation	RGB	If the cardiac frequency lies in the motion spectrum and dark skin subjects, the method will not work well.
	W Wang [63]	HR	Face	POS	RGB	It does not work well with illumination variations and has fixed projection planes for pulsatile component extraction.
	C Zhao [71]	HR	Nose, Cheeks	POS	RGB	Although the method reported the lowest SD, it is sufficiently high for HR estimation. This makes the algorithm unstable.
Wavelet Based methods	J Ryu [72]	HR	Cheeks	CWT	YCbCr	The proposed method cannot work well for rigid motions.
	K Lin [21]	HR	Forehead	Wavelet EEMD, MR	Green	Extraction of the reflectance signal was performed using the assumption of uniform light on each face region, which is impractical for real-time conditions.

	D Chen [66]	HR	Forehead	Wavelet EEMD	Green	The performance may degrade for longer interval videos.
	H Yu [65]	HR	Face	SSR, CEED-MAN	RGB	The method uses many thresholds that were set empirically; hence, the method may not generalize well.
	J Cheng [28]	HR	Cheeks	Wavelet EEMD	Green	Motion artifacts were not addressed. The assumption used for the study that illumination sources are the same for face and background ROIs is impractical to consider in real-time scenarios.
	Y Zhang [68]	HR	Face	EEMD	AB	The proposed method's performance degrades for shorter video sequences.
	R Song [73]	HR	Face	EEMD	RGB	The method's performance degrades for inaccurate ROI tracking or limited canonical variables in MCCV.
	Bousefsaf	HR	Face	CWT	RGB, YCbCr (Skin segmentation)	The proposed method is dependent on the high definition image frames (using PTZ algorithm), which might limit its applicability for poor resolution videos.
NN	B Wu [74]	HR	Cheeks	NN	RGB	The method is very complex, using 31 Artificial Neural Network (ANN) models for HR prediction, which makes the method very slow.

	Y Qiu [75]	HR	Cheeks	CNN, EVM	RGB	HR distribution was not focused on this method.
	Bousefsaf [76]	HR	Forehead, Cheeks	CNN	None	The proposed network was not fine-tuned for better estimation results.
	X Niu [77]	HR	Face	CNN,GRU	YUV	RhythmNet did not work well on NIR videos.
	G Hsu [78]	HR	Face	CNN	Green	The study excluded motion artifacts.
	Z Yu [79]	HR	Face	CNN	RGB	The data augmentation strategies were dependent on the HR values, which halts the method's generalizability.
	R Song [47]	HR	Cheeks	CNN, CHROM	RGB	The method used a tiny processing window to mitigate the effect of artifacts; the persistence of artifacts for a longer duration may degrade its performance.
	M Hu [80]	HR	Face	CNN	RGB	Performance degradation was reported for rigid motions such as rotation and, fast translations and compressed videos.
	M Hu [10]	HR	Face	CNN	RGB	The proposed framework could not be able to test for diverse HR ranges due to limited labeled samples.

d) Neural Networks

Neural Networks, specifically deep learning, have shown immense potential in rPPG signal extraction and HR estimation. Unlike conventional methods, the methods employing these networks are independent of assumptions and also possess good generalization ability [62].

Literature suggested two types of deep learning approaches for HR estimation: 1)predicting mean/instantaneous HR values from facial videos by assuming facial video-HR as classification/regression task, and 2) predicting rPPG waveform from facial videos. Various studies have proposed Deep Learning (DL) architectures for predicting mean HR, such as study by Niu et al. [77, 81, 82].

The studies corresponding to the second approach of predicting rPPG waveform from facial videos, followed Shafer's skin dichromatic model [63] to propose their DL based methods for rPPG signal extraction from facial videos. Most studies employed conventional computer vision deep learning architectures utilizing transfer learning. These methods aim to extract spatial information from the regions selected by attention mechanisms and explore the temporal relationship from the consecutive image frames of the video.

Literature suggests that the rPPG signal extraction task can be performed in two approaches: simultaneously extracting spatial and temporal information [83] or spatial followed by temporal information extraction [14]. For instance, for the former approach, 3DCNN has been explored, whereas for the latter, 2D convolutions have been used for spatial feature extraction, followed by sequence models for modeling temporal relationships. Precisely, various sophisticated DL architectures, such as Generative Adversarial Networks (GANs) [51] and convolutional autoencoders [54] employing 2D [51]/3D convolutions [54], have been proposed and tested to extract clean rPPG signals. Additionally, several studies have proposed combination of conventional rPPG methods and DL architectures for signal extraction, where the latter have been used as denoisers for extracting clean signal. For instance, a study by Song et al. [84] proposed a GAN based architecture named PulseGAN, where CHROM [69] was used for initial rPPG signal extraction, followed by denoising it using PulseGAN.

2.4.3 Signal Processing

As mentioned earlier, the extracted rPPG signal is used for estimating the physiological variables using signal processing techniques. The signal processing is depen-

dent on the physiological variable to be assessed. For instance, a maximum peak estimation was performed post bandpass filtering between 0.7 to 4.0Hz for PR measurement, followed by Fast Fourier transform. On the other hand, the ratio of AC and DC components from the PPG signal was calculated for SpO2 estimation. A detailed workflow for non-contact estimation of HR and SpO2 is depicted in Figure 8.

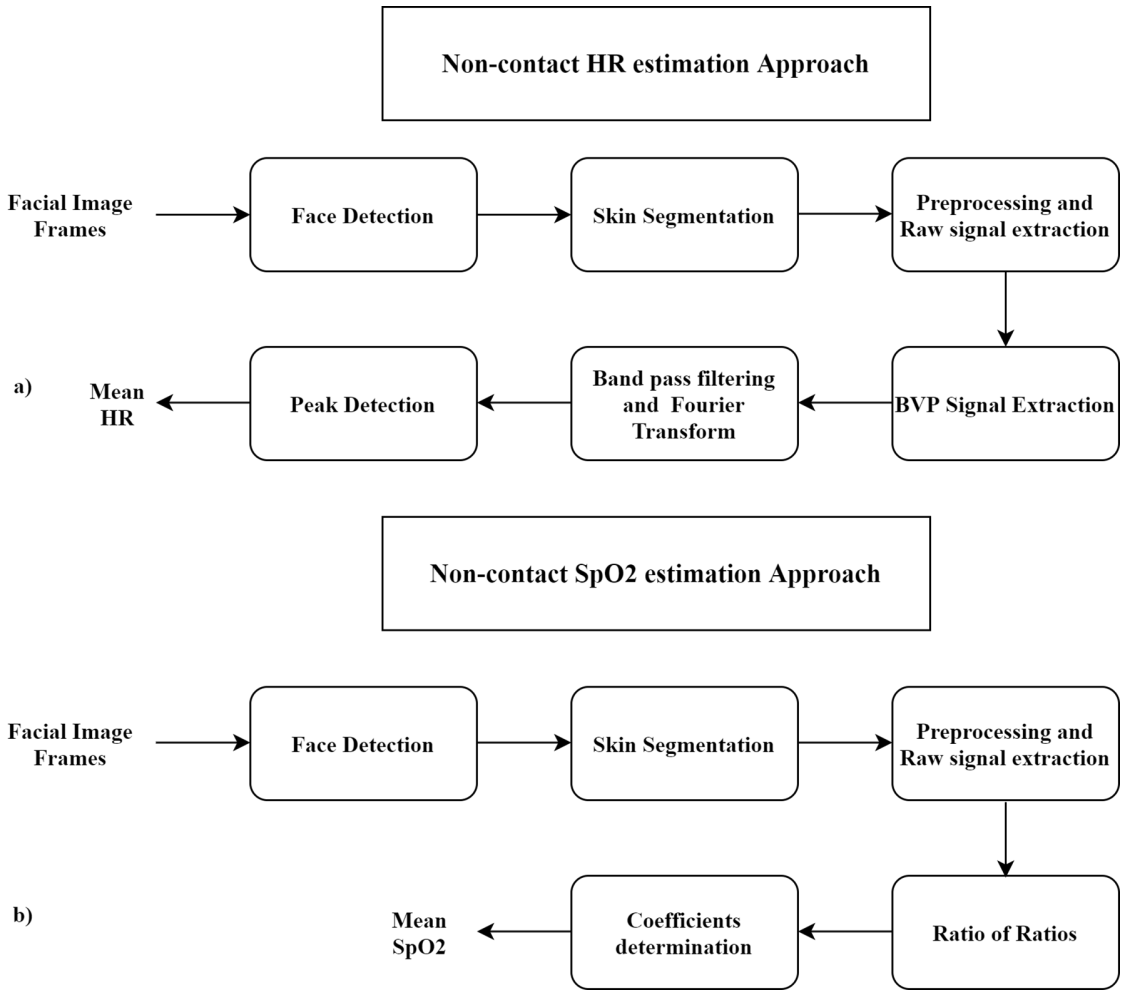


Figure. 8: Non-contact HR and SpO2 estimation approaches:HR estimation; a) SpO2 estimation (b).

2.4.4 Ratio-of-ratios method for SpO2 estimations

All the studies found in the literature have used the conventional ROR method for SpO2 calculations, except one study by Gastel, Stuijk, and Haan [23]. The details of these studies are presented in Table 3. The estimation studies differ in the selection of color channels and wavelengths. For instance, most studies [85–88] have used red and blue color channels for SpO2 estimation from the face ROI. However, Moço and Verkruysse [3] attempted to replace the blue channel with the green channel for

Table 3: A summary of SpO2 estimation studies.

First Author	**Physiological Variable**	**ROI Used**	**Method**	**Illumin. source**	**Limitations**
M Gastel [23]	SpO2	Face	APBV	IR	The APBV-based SpO2 method will not work well for periodic movements; the PPG signal is also mapped to different SpO2 levels, which can only map to those levels, not other SpO2 values.
AR Guazzi [85]	SpO2	Face	ROR	RGB	The proposed method did not address the motion, illumination variations, and effect of melanin concentrations, and the BR calculation could not be accurate due to the small processing window.
A Rosa [86]	SpO2	Forehead	ROR	R, B	The method was unable to show its applicability in low SpO2 values scenarios.
A Moço [3]	SpO2	Forehead, Cheeks	ROR	R, G, IR	The study proposed a ratio of red over green for SpO2 estimations but could not perform better than red over IR.
D Shao [89]	SpO2	Lips	ROR	Orange, NIR	The study did not address motion artifacts.
L Kong [90]	SpO2, HR	Cheeks	FFT	Mono	The method did not address motion or illumination variations.
U Bal [87]	SpO2, HR	Face	DT-CWT	RGB	The study uses few subjects to test their method; hence, analyzing its performance could have taken a relatively higher number of subjects.
L Tarassenko [88]	SpO2,HR, RR	Face	AR, ROR	RGB	Lower SpO2 values could not be measured; Motion artifacts were not addressed.

SpO2 estimations but found that the red and blue color channel combination for the ROR method is relatively more robust. Different color wavelengths, e.g., (611nm, 880nm) [89], and (520nm, 660nm) [90], have also been tested for accurate estimations for wider SpO2 range values. All of these studies are related in a way that they have used the ROR method for SpO2 estimations. However, Gastel, Stuijk, and Haan [23] extracted BVP signals using three wavelengths $(675nm, 800nm, 840nm)$ followed by mapping them to different values to SpO2 values interval. The novelty of non-contact SpO2 estimations is based on the choice of wavelengths for calculating ROR.

Chapter 3

TECHNICAL ASPECTS

This chapter presents the essential technical aspects of the elements used in this study. For instance, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines have been used to analyze and summarize existing non-contact HR and SpO2 estimation studies. The details of PRISMA guidelines for conducting a robust systematic analysis are presented in this chapter. As mentioned before, this study also aims to develop a database; the information about the database and measuring instruments for collecting data is also presented in this chapter. Additionally, the chapter presents information about the publically available databases that were used to validate new and SOTA non-contact estimation methods developed in this study. Since ICA has been used as the baseline for developing methods, this chapter presents the mathematical principles and underlying theory of ICA. Finally, a brief introduction about the performance metrics such as Root Mean Square Error (RMSE), Pearson Correlation, etc., and Bland-Altman (B-A) analysis used to validate the proposed methods is also presented in this chapter. Therefore, this chapter presents a brief introduction to the basic building blocks that contribute to providing insights into the quality and validation of different components of the study.

3.1 PREFERRED REPORTING ITEMS FOR SYSTEMATIC REVIEWS AND META-ANALYSES

The relevance of Systematic Reviews for different communities, such as healthcare providers and policymakers, can be understood by the need to keep them up to date with less time and effort, which is otherwise infeasible due to the number of research reports published on a daily basis. Therefore, the information provided

using a systematic review should be complete and transparent. For this purpose, PRISMA was published in 2009 (PRISMA, 2009) as a result of a three-day meeting of 29 participants in Ottawa, Canada [91]. The PRISMA guidelines consist of a checklist for creating complete and transparent reviews. To further improve the relevance and ensure the currency, the statement was further reviewed and updated in a 21 members' in-person meeting held in Edinburgh, Scotland (PRISMA 2020), which intends to guide the network or individual participant, data meta-analysis, systematic review of harms, diagnostic test accuracy studies and scoping reviews [92].

The PRISMA 2020 guidelines are designed to develop systematic review studies for the evaluation of health, social, and educational interventions. It supports synthesis (statistical analysis or metanalysis) and non-synthesis-based systematic reviews. Additionally, the guidelines are relevant for qualitative and quantitative analysis with the consideration related to information presentation and synthesis of qualitative methods. PRISMA guidelines can also be used to update the systematic reviews with additional considerations.

The newer version i.e., PRISMA 2020 guidelines, consists of 27 checklist items, each comprising further subsections requiring detailed information, unlike PRISMA 2009. A 27 items PRISMA checklist consists of explanations and elaborations for the following items used for this project: Title (inclusion of keyword "Systematic review"), Abstract (checklist of information inclusion presented in Table 4); Rationale; Review objectives and research questions that review addresses; Eligibility criteria (inclusion and exclusion of studies); Information sources (Registers, databases for data collection); search strategy (keywords, number limits, and fitters used), selection process (number of reviewers, studies screening criteria); data collection process; Potential Outcomes; "Risk of bias" assessment, Synthesis methods (statistical analysis, data tabulation and representations); and Reporting bias.

Additionally, the new PRISMA guidelines facilitate the assessment of the methods, which further leads to analyzing the findings of the studies. Analyzing and summarizing the characteristic features of the studies allows the readers to evaluate the applicability of the findings in their context. At the same time, the checklist also facilitates systematic review updates and replication, enabling research teams to leverage existing work. The new PRISMA guidelines are based on EQUATOR network guidance [93], which is related to designing and developing research reporting guidelines for health, but new (PRISMA) guidelines also support other interventions. A flow diagram for the studies collection to conduct a systematic review following PRISMA 2020 guidelines is also presented in Figure 9.

Table 4: Abstract Checklist

Section	**Information to be included**
Background	Research objectives questions that the review addresses.
Methods	Eligibility criteria, information sources, risk of bias (Studies quality assessment), methods for synthesis of results (statistical analysis methods).
Results	Included studies, statistical analysis, visual representations.
Discussion	Limitations of evidence, Results interpretation, and implications.
Other	Funding sources for conducting review and registration information of the protocol used for developing systematic review.

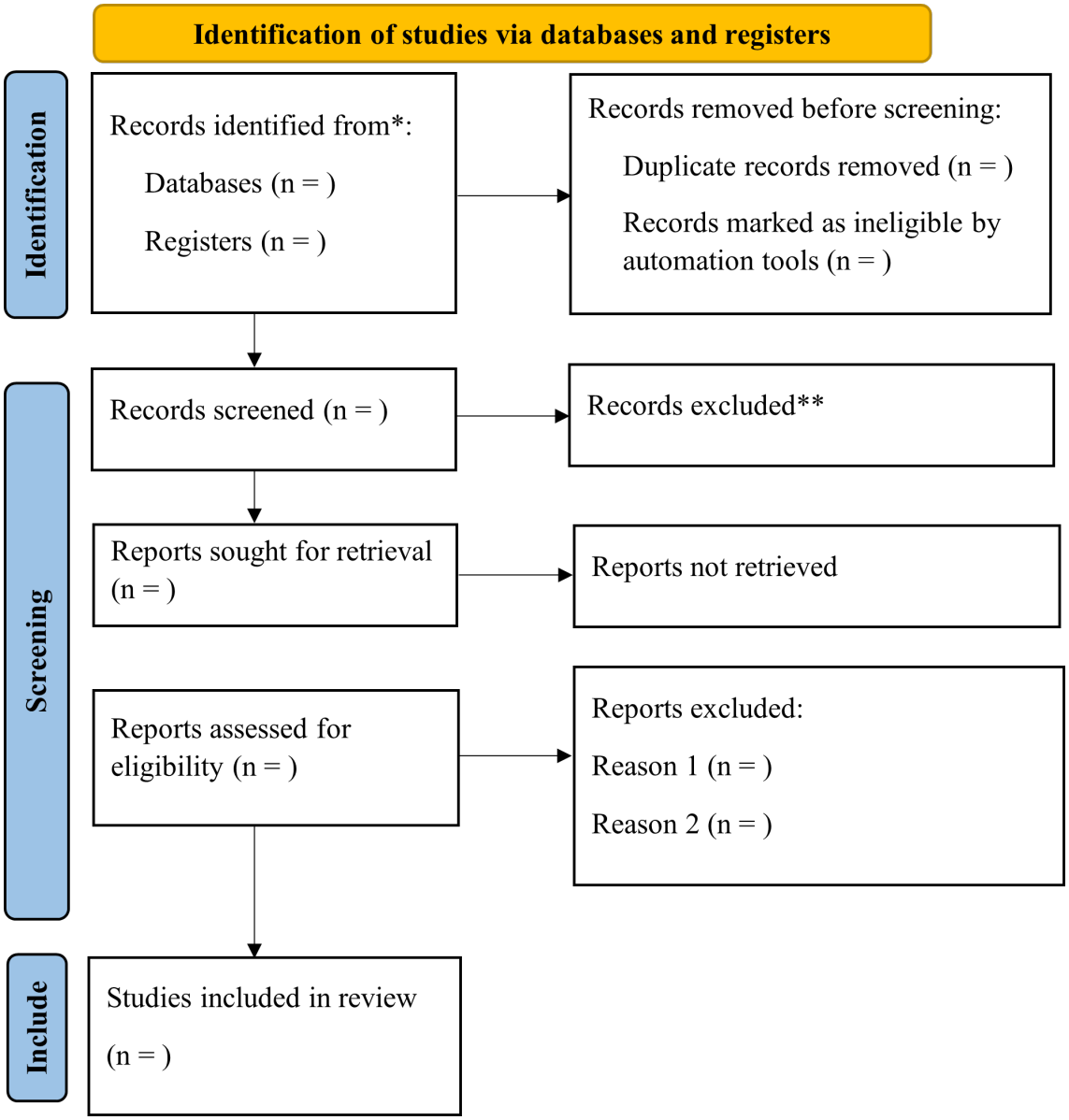


Figure. 9: A template for PRISMA flow diagram.

3.2 DATA COLLECTION

To the best of the author's knowledge, all publicly available databases have been created using single or multiple external light sources. As this study also explores the feasibility of RGB spectra for estimating physiological parameters in a darker environment, a database is developed by considering a few essential factors, such as video resolution, frame rate, and ethnicity.

Data collection at the University of Madeira requires consent from the data protection and ethical committees. Therefore, the data collection application was sent to the data protection committee, followed by the application to the ethical committee for approval, seeking consent for data collection. The data collection process was approved by both committees with the identification number $1/CEUMAI/2021$. The data collection application and approval documents from the data protection and ethical committee are presented in Appendix sections 7.3, and 7.3. Post approval from the respective committees, the data collection process was initiated in which each participant was asked to sign the informed consent forms in Portuguese or English language, followed by collecting image, video, and ground truth samples. The English and Portuguese versions of the informed consent forms approved by the data protection and ethical committee are also presented in Appendix sections 7.3,7.3, 7.3, and 7.3, respectively.

3.2.1 Database information

The proposed dataset consists of an image captured in ambient light, a video for $90 seconds$ in a dark environment using a webcam, synchronized with ground truth HR and SpO2 values using a $CMS60C$ pulse oximeter, for each subject. The dataset comprises 57 compressed RGB videos of variable framerates captured by different cameras in a dark environment with corresponding ambient light images synchronized with a ground truth HR, and SpO2 values. The distribution of HR, and SpO2 values are depicted in Figures 10, and 11, respectively.

a) Instruments Used

The video and image samples were captured using five different webcams: 720p Face Time HD camera (Apple Macbook Pro), HD webcam 720p (Asus Vivobook S15 S530F), Logitech C170-480p, Logitech hd720p. This camera covers variable framerates ranging from 7-16 Frames Per Second (FPS). Due to the unavailability of

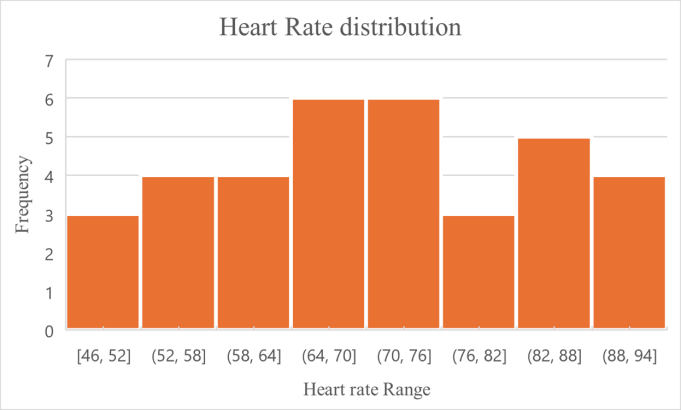


Figure. 10: Distribution of ground truth HR and SpO2 values.

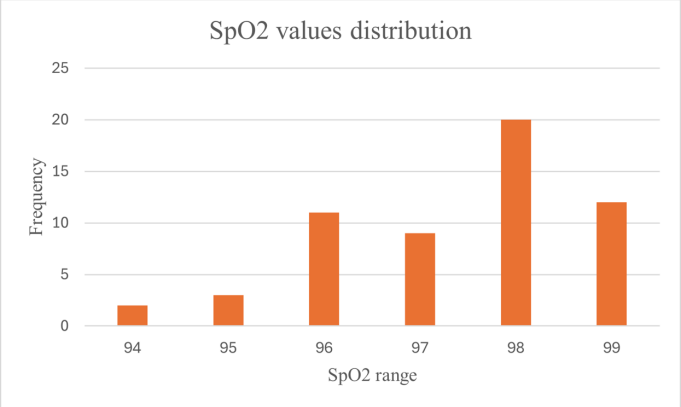


Figure. 11: Distribution of ground truth HR and SpO2 values.

information related to compression details about these cameras for commercial reasons, it was not possible to mention the compression algorithms used for each of them. Also, Apple does not reveal information about the compression algorithms of the built-in webcams used for video streaming. However, these cameras have used MJPG and H.264 compression algorithms.

Furthermore, the dark environment used for video acquisition was characterized by a luminance ≤ 1.0 lux, which was measured using a lux meter $XFUK-881F$ with a resolution of $0.1Lux/Fc$ and a $1-400,000lux$ range. The ground truth HR and SpO2 values were measured using a pulse oximeter with model number $CMS60C$. The pictorial representation of the lux meter and pulse oximeter is presented in Figure 12.

b) Data Collection Process

The data collection process is illustrated in Figure 13, which was performed using a two-stage process. First, a participant, after a substantial explanation and signing the informed consent form, was asked to sit on a chair with a ceiling light on, facing the camera for a minute, to stabilize the heart rate in the resting state.

Figure. 12: Lux meter (left) and CMS60C pulse oximeter (right) used for data collection.

The distance between the camera and the subject was kept 0.5m. Following, an image in ambient light is captured using the camera facing the subject. Finally, the ceiling light is switched off, creating a dark environment, maintaining a maximum permissible luminance of up to 1 lux, and a video of 90 seconds was captured, synchronized with ground truth HR, and SpO2 values.

The dataset consists of samples from 55 subjects with diverse ethnic regions (45 European, 5 Asian, and 5 African) and gender (41 males and 14 females), with ages between 18 and 61 years. This database was used for the performance validation of the method proposed in Chapter 6. Table 5 presents a few representative image samples (each representing a different ethnicity) of the dataset collected.

3.2.2 Other databases used

Although the main component of this study is to design and develop methods for non-contact HR and SpO2 estimations in dark environments, the first step is to develop methods in ambient light environments. Therefore, this study also used various databases publicly available on request. It used three other databases, i.e., VIPL-HR [94], Univ. Bourgogne Franche-Comté Remote PhotoPlethysmoGra-

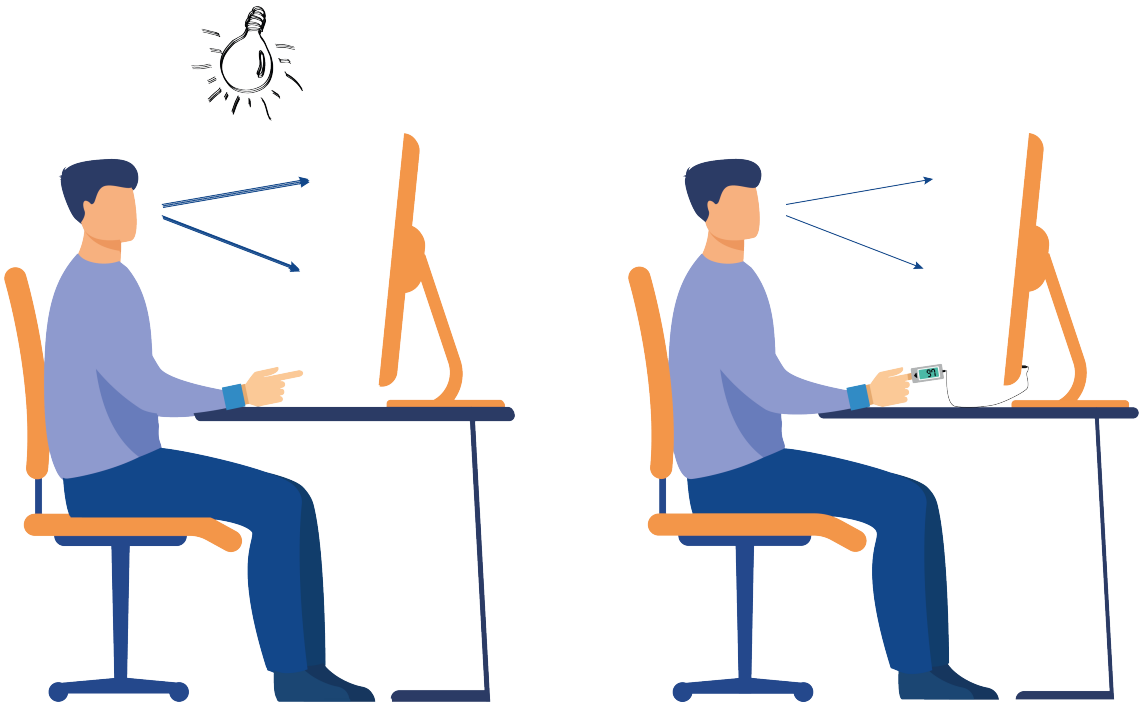
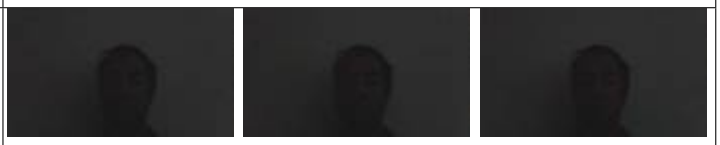


Figure. 13: Acquisition System: (a) Image acquisition consisting of a person facing the camera with illumination source for high exposure image capturing,(b) Video Acquisition system consisting of laptop camera with an internal source of light, and oximeter sensor attached to the index finger of the subject. (Video and high exposure image capturing share a common object of interest, i.e., facial region of the subject).

Table 5: Representative samples of the dataset collected.

Ethnicity gt (HR, SpO2)	**Reference image**	**Image frames from video samples**
African 67 bpm,99%		
Asian 85 bpm, 99%		
European 77 bpm,98%		

Note: gt is ground truth values. The ground truth HR and SpO2 values mentioned in the first column depict the mean values of the full sample of the subject.

phy (UBFC-rPPG) [95], and Cohface [96]. The details of these databases are presented in Table 6. These databases were used for the performance validation of the method proposed in Chapter 5. It is important to mention that the reason for using three databases is to validate the performance of the method in three scenarios i.e., constrained (VIPL-HR), rigid and non-rigid motion (UBFC-rPPG), and illumination variations (Cohface) scenarios, which corresponds to the performance evaluation of the method in the presence of motion and illumination artifacts, respectively.

a) VIPL-HR

The database consists of 2378 videos with visible light spectra and 752 videos with NIR spectra from 107 subjects (79 males and 28 females), with ages between 22 and 41 years. Nine variable scenarios were considered for sample collection. For each scenario, the samples were collected using digital cameras of different frame rates and NIR cameras. Each database sample comprises a 30 *seconds* subject video, a BVP signal, HR, and SpO2 values [94]. The publically available version of this dataset comprises compressed videos. Their compression strategy is a two-step process, which includes reducing the video resolution to two-third of the original size, and applying MJPG compression codec. This study has used the subset of these compressed videos that corresponds to a frame rate of 30 FPS with 1920×1080 pixels resolution, covering the HR range between 47 and 100 bpm. The ground truth HR was acquired using a $CMS60C$ pulse oximeter synchronized with the subject's video.

b) UBFC-rPPG

UBFC-rPPG is a publicly available database consisting of 50 video samples synchronized with a $CMS50E$ pulse oximeter with a sampling rate of 60 Hz). The videos are available in an uncompressed form with a resolution of 640×480 pixels at 30 FPS, covering the HR range between 63 and 112 bpm. Each video is 2 minutes long in which the participants were asked to sit facing the camera and play a mathematical game that causes an abrupt rise and fall in HR value promoting rigid and non-rigid movements. The database did not provide age-specific information. [95].

c) COHFACE Database

The COHFACE dataset is a collection of 160 videos with physiological recordings for the HR and the Respiratory Rate (RR) from 40 healthy subjects with a mean

age of 35.6 years. The dataset constitutes $60 seconds$ videos from 12 females and 28 males covering the HR range between 54 to 97 bpm. The videos were recorded with a resolution of 640×480 pixels at 20 FPS with the synchronized BVP measurements using the BVP model $SA9308M$, with a belt model $SA9311M$ with a sampling rate of $256Hz$ [96]. Similar to VIPL-HR, the COHFACE dataset also contains compressed videos, though the original paper did not report information about compression algorithms. However, the videos were collected using a Logitech C525 webcam, which uses H264 for video streaming. The dataset offers constrained and challenging natural conditions, especially in terms of illumination variations over the facial region. Therefore, this study tests the performance of the proposed method using natural conditions video samples.

Table 6: Publicly available databases summary used for this study.

Features	**VIPL-HR**	**UBFC-rPPG**	**COHFACE**
No. of subjects	107	50	40
Video Resolution	1920 X 1080	640 X 480	640 X 480
Frame rate	30	30	20
Video Duration	30 seconds	90 seconds	60 seconds
Ground truth sensor	CMS60C	CMS50E	SA9308, SA9311M
Subject-camera distance	1 meter	1 meter	-
HR range	47-100 bpm	63-112 bpm	54-97 bpm
Artifacts	Constrained*	Rigid and Non-rigid motion	Illumination Variation

3.3 INDEPENDENT COMPONENT ANALYSIS

This study used ICA as a baseline method for developing non-contact methods for HR and SpO2 estimations, which is a blind source separation method [97]. Therefore, it is worth presenting the details of standard ICA for better explanations of the developed methods in this study. The concept of ICA can be understood using a cocktail party problem, where all individuals are talking, which results in a mixture of voices. The task is to identify the voice of each individual by assuming that it is different for each individual. ICA aims to solve this problem by identifying the individual voices from a mixture of voices of different individuals [98].

ICA is based on the assumption that different physical processes produce unrelated signals. Mathematically, unrelated or independent signals can be defined based on uncorrelatedness and statistical independence. The common approach for

ICA is to assume that the source signals are statistically independent, i.e., one provides no information about the other. Therefore, ICA seeks to maximize the statistical independence between ICs. For further insights into ICA, suppose three temporal source signals $s_1(t), s_2(t), s_3(t)$, need to be extracted from three mixture signals $x_1(t), x_2(t), x_3(t)$, (the signals are denoted as a function of time t), where each mixture signal is a linear combination of three source signals, with a weighted contribution a_{ij} corresponding to each source signal. It can be defined as:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \tag{2}$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \tag{3}$$

$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) \tag{4}$$

It is important to mention that a_{ij} is not known since it depends on the properties of the physical system to be analyzed. The source signals are also unknown, while only mixture signals are known and available. Considering all this, the task is to extract the unknown source signals. Therefore, assuming that the A matrix (consisting of above weighted contributions) is invertible, the task is to estimate another matrix W to extract approximations of source signals as:

$$y_1(t) = w_{11}x_1(t) + w_{12}x_2(t) + w_{13}x_3(t) \tag{5}$$

$$y_2(t) = w_{21}x_1(t) + w_{22}x_2(t) + w_{23}x_3(t) \tag{6}$$

$$y_3(t) = w_{31}x_1(t) + w_{32}x_2(t) + w_{33}x_3(t) \tag{7}$$

Where $y_1(t), y_2(t), y_3(t)$, are approximations of source signals $s_1(t), s_2(t), s_3(t)$. This is called a blind source separation problem since the information about source signals, i.e. ICs that need to be extracted, is unknown. w_{ij} can be estimated by considering the statistical independence of source signals. If the signals are non-gaussian, estimating w_{ij} is straightforward. If $y_1(t), y_2(t), y_3(t)$, are statistically significant, these are equal to source signals $s_1(t), s_2(t), s_3(t)$, but this is infeasible in real-time situations. The independent components estimation can be done by decorrelating them or maximizing their non-gaussianity based on the central limit theorem, which states that the sum of non-gaussian variables is closer to Gaussian. In other words, mixture signals are assumed to be Gaussian, and independent components are extracted by increasing their non-gaussianity since the combination of non-gaussian variables (independent components) would be closer to Gaussian variables (mixture).

In context to the rPPG signal extraction task from the facial video captured in RGB space, ICA aims to approximate three statistically independent components from three mixture signals corresponding to each color channel for RGB spectra. The mixture signals are assumed to have noise and other components, such as information related to BR, parasympathetic system, etc., while one of the three IC contains the signal of interest, i.e., the rPPG signal. Therefore, for signal extraction, these independent components are investigated based on rPPG signal characteristics such as periodicity. Since ICA for biomedical signal processing requires processing using higher order statistics directly or indirectly using non-linearities, existing rPPG studies suggested the second approach for independent components estimations, i.e., maximizing the non-gaussanity of independent components using statistical measures such as Kurtosis [98]. Figure 14 shows a workflow of rPPG signal extraction using ICA.

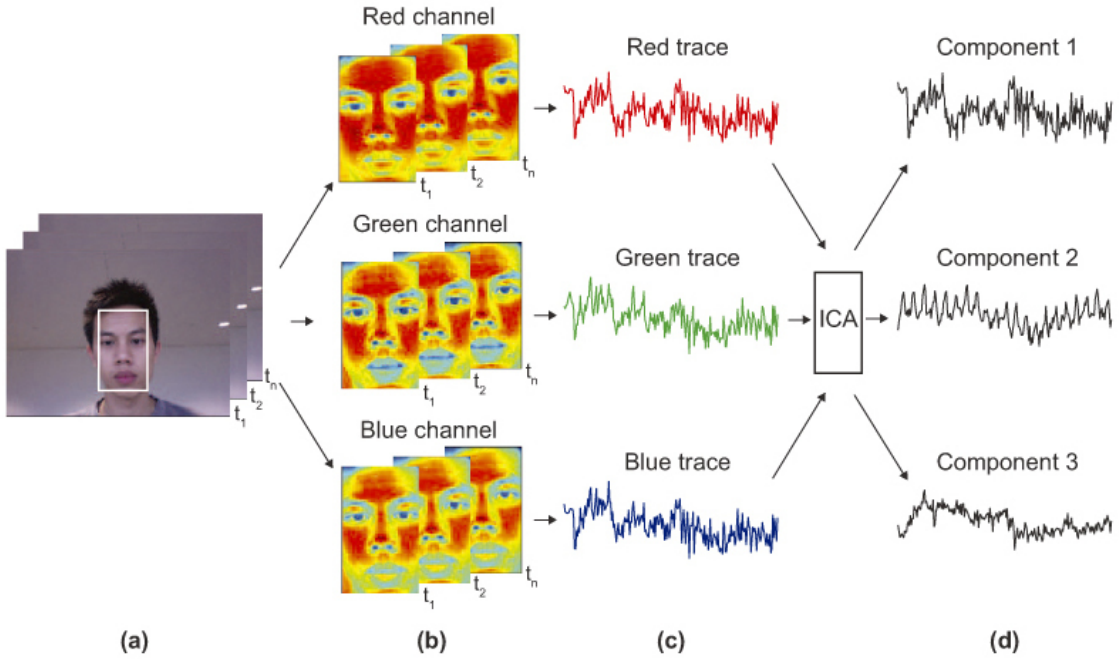


Figure. 14: A workflow of rPPG extraction task using ICA: face detection, color channels segregation, signal construction, and extracted independent components using ICA [4].

3.4 PERFORMANCE METRICS

Based on the literature review of non-contact HR and SpO2 estimation studies, it was found that the RMSE, Pearson correlation, and regression plots have been predominantly used by these estimation studies. Furthermore, a few studies have also reported Mean Error (ME) and error SD to justify the performance of the

respective methods. Besides, a measure of clinical relevance i.e. *accuracy* have also been reported in a few studies, which is defined as the percentage of samples with clinically acceptable error limits between estimated and ground truth values. The clinically accepted error limit for HR is ± 5 bpm, whereas, for SpO2, there is no clinically defined limit to the best of the knowledge. The performance analysis of all the methods developed during this study have been measured by all the parameters discussed here, i.e., ME, Mean Absolute Percentage Error (MAPE), error SD, RMSE, Pearson correlation, *accuracy*, and Coefficient of Determination (R^2).

It is worth mentioning that the consideration of these metrics is dependent on the type of physiological parameters estimated. For instance, the Accuracy metric, slightly different from the standard accuracy definition, was computed for non-contact HR estimations and not for SpO2 estimations, as this metric is based on American Association of Sleep Medicine (AASM) guidelines for HR measurement of estimation methods. On the other hand, this metric was not used for SpO2 estimations since there are no such definition guidelines available to measure the efficacy of SpO2 estimation methods. The mathematical formulations and significance of the metrics, as mentioned above, are presented in the following subsections.

3.4.1 Mean and standard deviation error

The ME is the average of the difference between ground truth and predicted values. It provides information regarding the uncertainty of measurement during estimations. The mathematical formulation for the ME, is as follows:

$$me = \frac{\Sigma_{i=1}^{N} V_{obs} - V_{est}}{N} \tag{8}$$

Where V_obs, and V_est are ground truth and estimated values, respectively, and N is the number of measurements taken from an estimation method. The limitation of this error is that negative and positive differences between ground truth and estimated value cancel each other; therefore, it needs to be used in conjunction with other performance metrics as well.

On the other hand, the error SD provides information about the distribution and variance of the error distribution. In other words, it aims at identifying the effect of outliers on the overall accuracy of the estimation method by measuring the variance. It also provides information about the stability of the method, i.e., smaller SD value means stable performance, while a higher value signifies that the method is unstable in terms of estimating values. The mathematical formulation of error SD (Average)

sd_e is as follows:

$$sd_e = \frac{\Sigma_{i=1}^{N} SD(V_{obs}, V_{est})}{N} \tag{9}$$

Where V_obs and V_est is the same as for ME, and sd_e indicates the SD for individual values of V_obs and V_est. The average value of SD is an estimate of the proximity of the individual scores from the mean.

3.4.2 Root means square error value

The RMSE value measures the average difference between ground truth and estimated values. Similar to standard error, it provides the distribution of errors with an assumption that the distribution follows Gaussian distribution. One of the limitations of RMSE is that it is non-robust for small samples [99]. The mathematical formula for RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(V_{obs} - V_{est})^2} \tag{10}$$

3.4.3 Pearson correlation

Pearson correlation measures the linear correlation between ground truth and estimated values. Its value can range between -1 to $+1$, where negative, zero, and positive values signify negative (opposing relationships with strength defined by magnitude), no (no relationships), and positive correlation (linear relationship with strength defined by magnitude), respectively. One of the limitations of Pearson correlation is that it is invariant between two sets of values, i.e., the correlation value of $\rho_{(V_{obs},V_{est})}$ is similar to $\rho_{(V_{est},V_{obs})}$.

$$\frac{\mathbb{E}\left[V_{est}, V_{obs}\right] - \mathbb{E}\left[V_{est}\right]\mathbb{E}\left[V_{obs}\right]}{\sqrt{\mathbb{E}\left[V_{obs}^2\right] - \mathbb{E}\left[V_{obs}\right]^2}\sqrt{\mathbb{E}\left[V_{est}^2\right] - \mathbb{E}\left[V_{est}\right]^2}} \tag{11}$$

where $\mathbb{E}\left[.\right]$ is the expected value and other variables have a similar meaning as above. The Pearson correlation was computed at the significance level (α) of 0.001 for this thesis.

3.4.4 Accuracy

Accuracy is defined as the percentage of samples where ground truth and estimated values are equal. However, for physiological parameters, it is very rare to have this

condition satisfied. Therefore, following AASM guidelines [100], the mathematical formula for accuracy is given as follows:

$$Acc(\%) = \frac{V_{(obs-est)\leq 5}}{N} 100 \tag{12}$$

where $V_{(obs-est\leq 5)}$ is defined as the number of samples where the error difference between ground truth and estimated values is less than or equals to five units (e.g., $\pm$ 5bpm for HR measurement). This metric was used to quantify the efficacy of non-contact HR estimation methods.

3.4.5 Coefficient of Determination

The coefficient of determination quantifies the performance of regression models. Its value ranges between 0 and 1, where 0 values signify poor performance and 1 represents good performance models. Mathematically, it is the difference between 1 and the ratio of the sum of squared errors and total residuals, where total residuals are the difference of each value of V_{obs} with its mean. It provides more insights into the model's performance than RMSE and standard error. R^2 can be mathematically defined as:

$$R^2 = 1 - \frac{\Sigma_{i=1}^{N}(V_{obs} - V_{est})^2}{\Sigma_{i=1}^{N}(V_{obs} - \mu(V_{obs}))^2} \tag{13}$$

3.5 BLAND-ALTMAN ANALYSIS

The B-A plot proposed by Eksborg in 1981 is based on quantifying the level of agreement between the two methods [5]. Precisely, it elevates the limitations of correlation and regression since both techniques are confined to assess the linear relationship between two sets of quantitative observations estimated by two methods, not the degree of agreement.

B-A analysis provides a method to measure the level of agreement by establishing the limits of agreement based on statistical measures. Specifically, the limits of agreement are computed using the μ and σ of the differences between the quantitative measurements from the two methods. B-A analysis plots a scatter plot between the difference of the paired measurements (Y-axis) and the mean of paired measurements (X-axis). It is recommended that 95% of the data points in the scatter plot should lie between the limits of agreement. The limit of agreement is set as $\mu \pm 1.96\sigma$. The scatter plots can also be plotted by taking the difference in percentage values or ratios.

Plotting the difference with the mean of measurements from the two methods also provides insights into any relationship between measurement error and actual value. If one of the methods is standard, the mean of measurements from the two methods can be replaced by the values of the standard method. Otherwise, the best estimate to consider is to take the mean of quantitative measurements from both methods.

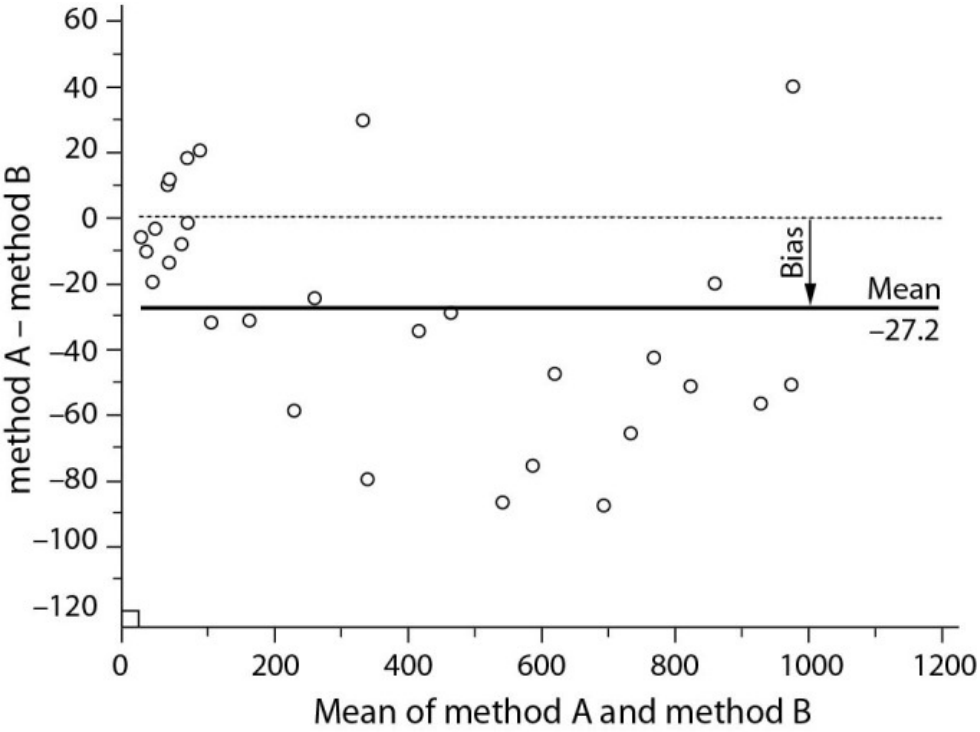


Figure. 15: A schematic representation of Bland-Altman scatter plot [5].

A schematic representation of the B-A scatter plot is depicted in Figure 15, where the Y axis represents the difference in quantitative measurements from the two compared methods, and the X axis represents the mean of both measurements. The mean bias is the mean of all paired differences, *i.e.*, -27.2, while the limits are defined by computing the SD as well, i.e., $\mu \pm 1.96\sigma$. Consequently, the lower and the upper limits of agreement from Figure 15 are 41.1 and -95.4, respectively. Also, it is worth noticing that all data points (differences) lie between the statistical limits of agreement.

Although B-A analysis provides the statistical limits of agreement between two methods, it does not provide information about the sufficiency or suitability of the levels of agreement. For instance, the mean bias is not in proximity to zero difference and, consequently, the statistical limits of agreement [5]. Therefore, acceptance of mean bias and levels of agreement is problem and domain dependent. Although the statistical limits of agreement were computed based on B-A analysis, the results were

analyzed based on guidelines by AASM for HR measurement and existing literature studies for non-contact SpO2 estimations.

Chapter 4

SYSTEMATIC ANALYSIS OF NON-CONTACT HR AND SPO2 ESTIMATION METHODS

Consumer-level cameras have provided an advantage of designing cost-effective, non-contact physiological parameters estimation approaches, which is not possible with gold standard estimation techniques. This encourages the development of non-contact estimation methods using camera technology. Therefore, this chapter presents a systematic analysis, summarizing the currently existing face-based non-contact methods along with their performance. This analysis aims to identify a set of critical factors from the existing studies, followed by using them for assessing the quality of studies based on PRISMA.

It includes all HR and SpO2 studies published in journals and a few reputed conferences, which have compared the proposed estimation methods with one or more standard reference devices. The articles for this systematic analysis were collected from the following research databases: Institute of Electrical and Electronics Engineers (IEEE), PubMed, Web of Science (WoS), Science Direct, and Association of Computer Machinery (ACM) digital library. Each study was assessed using a finite set of identified factors for reporting bias.

Out of 332 identified studies, 32 studies were selected for the final analysis. Additionally, 18 studies were included using snowballing by thoroughly checking the references of these studies. It was found that 3 out of 50 (6%) studies were performed in clinical conditions, while the remaining studies were carried out on a healthy population. Also, 42 out of 50 (84%) studies have estimated HR, while 5/50 (10%) studies have measured SpO2 only. The remaining three studies have estimated both parameters. The majority of the studies have used 1˘3 minute videos for estimation.

Among the estimation methods, DL and ICA were used in 11/42 (26.19%) and 9/42 (21.42%) studies, respectively.

According to the Bland-Altman analysis, only 8/45 (17.77%) HR studies achieved the clinically accepted error limits, whereas, for SpO2, 4/5 (80%) studies have matched the industry standards ($\pm 2 - 3\%$). DL and ICA have been predominantly used for HR estimations. Among deep learning estimation methods, different variants of neural networks, such as CNN and Transformers, have been predominantly employed due to their excellent generalization ability and independence of assumptions, unlike conventional methods.

This systematic analysis resulted in pointing out a few limitations of the existing studies, key findings, and recommendations. For instance, most non-contact HR estimation methods need significant improvements to implement them in a clinical environment. A detailed overview of this analysis is presented in the following subsections of this chapter. The work presented in this chapter corresponds to the publication of the journal, Computer Methods and Programs in Biomedicine (2 of section 1.5) [101].

4.1 BACKGROUND

Physiological parameters serve as reliable indicators of an individual's health. As mentioned before, five physiological parameters used for determining the individual's health status are BP, HR, SpO2, body temperature, and BR [7]. Estimating these parameters requires sophisticated apparatus that involves the placement of electrodes or sensors in contact with the skin using adhesive gels. This approach of estimating parameters is called the contact-based approach, which has limitations such as discomfort or allergies due to the sensors or electrode placement in contact with the subject's skin. Therefore, it is not suitable in certain scenarios, such as unobtrusive monitoring, sensitive or burnt skin, and monitoring at the NICU. This originates the need for an alternative non-contact-based estimation approach that is based on a non-contact variant of PPG known as rPPG or iPPG. It requires a camcorder with an illumination source to select the ROI appropriately, followed by a PPG signal extraction and physiological parameter estimations [4] (please see section 2.4 for details). An appropriate ROI selection is critical for accurate PPG extraction, thereby estimating correct physiological parameters.

Several studies have been performed to explore the potential ROIs for estimation. In this context, the first attempt of ROI selection was performed by Pavlidis et al. [102], which explored the potential of the face in estimating cardiac pulse, BR, and

blood flow using a thermal imaging model. The study concluded that the thin layer of tissues present in the facial region enables the estimation of the ROI accurately. Furthermore, Verkruysse et al. [29] proved the potential of the facial area for HR and RR monitoring. This was further validated in a study conducted by van der Kooij and Naber [103], which explored the potential of different body organs and found that the facial region can provide more accurate HR estimates than other body organs. Similarly, SpO2 can be accurately estimated using facial ROI [104]. Currently, four vital parameters, namely HR, SpO2, BR, and BP, can be computed using PPG, hence rPPG. Among these, HR and SpO2 are estimated in various critical scenarios, such as intensive care units, surgery, CoronaVirus Disease (COVID) diagnosis, and sleep quality analysis [86, 105]. Furthermore, BR is not the most frequently used monitoring vital signs and has relatively limited applicability [106]. Therefore, this review focuses on face-based HR and SpO2 non-contact estimation studies.

Several attempts have been made to summarize various SOTA developments using literature reviews. Several aspects have already been explored in the existing reviews in the literature. Hassan et al. [18] and Wang et al. [107] have investigated the existing HR estimation methods and compared them using benchmark datasets. Sun and Thakor [108] discussed the implications of iPPG and its technical limitations. Kranjec et al. [19] presented a review summarizing image and non-image-based methods for HR estimation. All of them are narrative review, which provides a summarization of recent developments; however, due to diversified approaches, it is challenging to select the optimum set of parameters for designing a standardized non-contact estimation study. A standardized study is a study covering as many issues as possible to a specific problem and providing a generalized solution to the problem. To mitigate this challenge, Harford et al. [109] presented a systematic review that discusses the clinical aspect of all image-based non-contact approaches using a modified Guidelines for Reporting Reliability and Agreement Studies (GRRAS) criterion.

However, no systematic analysis of non-contact estimation approaches has been presented from an engineering-specific context. Therefore, this systematic analysis aims to provide a detailed review of non-contact facial video-based physiological parameter estimation methods in clinical and non-clinical settings. To the best of the knowledge, it summarizes the technical and non-technical elements of face-based, non-contact estimation approaches. This will provide a basis for further research in this thriving domain and help improve the quality of future studies.

4.2 STUDY OBJECTIVES

This systematic analysis is focused on non-contact HR and SpO2 estimation studies utilizing face or a combination of different facial regions like cheeks, forehead, etc. as potential ROI. It aims to accomplish the following objectives:

1. This review follows the newly updated PRISMA guidelines [92] and answers all the relevant questions regarding designing a non-contact estimation study, such as selecting ROI, estimation methods, and performance metrics.
2. It summarizes and analyzes the existing HR and SpO2 non-contact estimation studies published in journals collected using five research databases. It also proposes a novel protocol for "risk of bias" analysis by identifying crucial factors from the existing studies.
3. It presents the statistical analysis of all the performance metrics used in the existing studies. This will allow a baseline to compare the newly proposed methods with existing SOTA studies' reported performance metrics.
4. It also presents the advantages and disadvantages of the various parameters, estimation methods, and data acquisition methods, enabling the identification of a suitable regime for conducting a physiological parameter estimation study.

4.3 METHODOLOGY

4.3.1 Eligibility criteria

This systematic analysis is prepared following the new PRISMA statement, 2020 [92]. The inclusion and position of its respective checklist items in this analysis is presented in Table 7. It includes all the studies related to facial video-based HR and SpO2 estimation in comparison with the appropriate reference devices. A few HR estimation studies have included other physiological parameters: BR, step count, eye blink, and Heart Rate Variability (HRV). It is important to note that no explicit searches have been performed for the estimation of these parameters. The search is limited to journal papers only. Conference papers and book chapters were not considered due to the limited novelty of the studies. However, papers published in reputed international conferences such as Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV), and European Conference on Computer Vision (ECCV) were included in this review.

4.3.2 Information Sources

The search has been performed in the following databases: IEEE, PubMed, WoS, Science Direct, and ACM digital library. The database search was completed on May 20, 2021. This analysis will be updated subsequently by including the recent relevant studies.

4.3.3 Articles search strategy

The articles search was conducted between December 2008 and May 2021 since the first attempt for HR and SpO2 estimation using the face region was published in December 2008 by Verkruysse et al. [29]. However, an updated version of this systematic review is in the pipeline. Furthermore, the studies from the engineering-specific context were considered for this review. Specifically, studies demonstrating the novel approaches for any of the components/modules (ROI/PPG signal extraction, and HR calculation) of estimation using face videos with any imaging modality were only considered for this review.

Specifically, this review includes the studies conducted using various image color models such as RGB, CIE, LAB, YCbCr, YUV, NIR, and monochrome color filters (e.g., magenta, orange). However, studies demonstrating the implications of existing non-contact methods in scenarios such as driving were not included. Related estimation studies using other human body organs, such as arms, hand palms, etc., were excluded. Moreover, the studies on fetal HR monitoring, patents related to HR or SpO2 estimation methods or devices, use of smartphone sensors by contact, PPG signals or contact-based approaches, and animal physiological parameters estimations were also excluded. The sole reason for excluding these studies is that they are beyond the scope of this systematic analysis. The details of the search strategy are presented in Table 7.

4.3.4 Data Collection

The data are collected by thoroughly reading the article's full text. Consequently, this systematic analysis aims to collect the following information: study title, estimated physiological parameter reference device(s) used, video length, frame rate (fps), camera resolution, shooting distance, databases used, number of subjects, age

[!ht]

Table 7: PRISMA Checklist

Section and Topic	**Item**	**Checklist item**	**Reported**
TITLE			
Title	1	Identify the report as a systematic review.	Y
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Y
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Y
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Y
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Y
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Y
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Y

Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and, if applicable, details of automation tools used in the process.	Y
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Y
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Y
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Y
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Y
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	NA

Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item 5)).	NA
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Y
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Y
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	NA
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Y
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	NA
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Y
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	NA
RESULTS			

Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Y/
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Y/
Study characteristics	17	Cite each included study and present its characteristics.	Y
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Y
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Y
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Y
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Y
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	NA
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	NA

Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Y
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	NA
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Y
	23b	Discuss any limitations of the evidence included in the review.	Y
	23c	Discuss any limitations of the review processes used.	Y/
	23d	Discuss implications of the results for practice, policy, and future research.	Y
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Y
	24b	Indicate where the review protocol can be accessed or state that a protocol was not prepared.	NA
	24c	Describe and explain any amendments to the information provided at registration or in the protocol.	NA
Support	25	Describe sources of financial or non-financial support for the review and the role of the funders or sponsors in the review.	Y
Competing interests	26	Declare any competing interests of review authors.	Y

Availability of data, code, and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	NA

Note: The information presented in this checklist is as per the systematic review published in the Journal (Point 2 of section 1.9).

range, gender, clinical/normal study, skin (types or color) and ethnicity information, ROI used, ROIselection method, ROI/PPG extraction method, color channels, lighting source, performance metrics, Bland-Altman analysis information, type of artifacts (Motion, Illumination or both) addressed.

4.3.5 Potential Outcomes

The potential outcome of this study is to present the available face-based non-contact HR and SpO2 estimation methods using various image modalities in comparison with the respective reference device(s). The secondary outcome is to assess their performance defined by the inclusion/exclusion of vital parameters included in every study and to analyze the corresponding reported performance metrics, their practical implications, and limitations. Furthermore, the common performance metrics reported for all studies were selected and used for study categorization using a threshold, which could mitigate the effect of low and high extreme values. Due to heterogeneity among HR and SpO2 estimation studies, different performance metrics were independently chosen for both types of studies.

4.3.6 Studies quality assessment

All the selected articles were assessed based on the identification of critical factors and a scoring scheme. The scores were given based on the inclusion of selected parameters and performance metrics for HR and SpO2 studies.

The proposed scoring protocol uses a two-step scoring process: first, scoring based on the inclusion and exclusion of the following factors: camera resolution, shooting distance, number of participants, Bland-Altman analysis, motion or illumination artifacts (HR studies), ethnicity; analyzing the performance metrics (RMSE and correlation for HR and R-squared for SpO2) and accuracy (error$\leq \pm$ bpm justifying the method's clinical relevance. For the inclusion of each parameter, a score of one is assigned and zero otherwise, except for artifacts, as apparent in Table 9.

Second, individual scores are summed up for overall quality assessment. For HR estimation studies, these scores are finally used to categorize studies into three categories: weak, fair, and strong. The studies with a score less than two will be classified as weak since in almost all the reported studies, camera characteristics (camera shooting distance, video resolution) and performance metrics (RMSE and correlation) were reported summing up to a score of 2. Studies with a score between

Table 8: Search Strategy

No.	Search strategy
1	(Contactless OR Noncontact OR non-contact OR non-invasive OR non-invasive OR remote OR contact free OR contact-free).ti
2	(Oximetry).MeSH OR (SpO2 OR "blood oxygen* OR blood saturation OR SpO2).ti
3	(HR Determination).MeSH OR (HR OR Pulse rate NOT(Variability)).ti
4	(face AND video* OR webcam OR camera OR smart-phone OR mobile camera).ti
5	(iPPG OR rPPG OR Physiological). ti OR Photoplethys-mography).MeSH
6	2 AND 3
7	3 OR 4
8	(1 OR 7) AND 6

Note: ti stands for a title search, MeSH stands for MEdical Subject Headings.

3 and 5 would be categorized as fair, whereas a score ranging from 6 to 8 would signify a strong study.

On the other hand, SpO2 studies with a score of 1 will be categorized as weak since most studies used regression, while the studies with a score of 3 and 4 would be considered fair and strong. The proposed scoring scheme for studies quality assessment is presented in Table 9 and 10, respectively. The information collected for studies quality assessment and categorization for HR and SpO2 estimation studies is presented in section 4.4.9, where scores are assigned based on the inclusion and exclusion of identified factors. For instance, the factor Camera score comprises two camera characteristics: Camera quality and subject-camera distance. If a particular study reported this information, a score of 1 is assigned, otherwise 0, followed by adding the scores for the studies quality assessment or risk of bias assessment.

4.3.7 Visual interpretation and tabulation of results

The systematic analysis uses different types of visual representations to highlight the key findings and present statistical measures to analyze new studies in the future. Precisely, it used three types of visual representations, i.e., box-whisker plots and bar and error charts. Box-whisker plots will be used to examine the distribution of performance metrics used for estimation studies, while bar charts will be used to visualize the study categorization, distributions of regions of interest, and the estimation methods for HR and SpO2 studies. Error charts will be used for analyzing

Table 9: Proposed Scoring scheme for non-contact HR/PR estimation studies.

Factors	**Score**
(Resolution AND Camera distance)	$\begin{cases} 1, \textit{reported} \\ 0, \textit{Not reported} \end{cases}$
Artifacts	$\begin{cases} 1, \textit{Motion/Illumination} \\ 2, \textit{Both} \end{cases}$
Result score (RMSE OR Correlation) (median thresholds)	$\begin{cases} 1, Resultscore \geq threshold \\ 0, Resultscore \leq threshold \end{cases}$
Accuracy (error difference<=5)	$\begin{cases} 1, \textit{reported} \\ 0, \textit{Not reported} \end{cases}$
Number of participants (median threshold)	$\begin{cases} 1, N_{Participants} \geq threshold \\ 0, N_{Participants} \leq threshold \end{cases}$
Ethnicity	$\begin{cases} 1, \textit{reported} \\ 0, \textit{Not reported} \end{cases}$
Bland and Altman Plot	$\begin{cases} 1, \textit{reported} \\ 0, \textit{Not reported} \end{cases}$

the upper and lower statistical limits of Bland-Altman plots. Additionally, tables will also be used for presenting statistical values ($\mu \pm \sigma$) of all performance metrics used in the estimation studies in the analysis and identified factors collection and scoring.

Table 10: Proposed Scoring scheme for non-contact SpO2 estimation studies.

Factors	**Score**
Camera Score (Resolution AND Camera distance)	$\begin{cases} 1, \textit{reported} \\ 0, \textit{Not reported} \end{cases}$
Result score (R^2) (median thresholds)	$\begin{cases} 1, Resultscore \geq threshold \\ 0, Resultscore \leq threshold \end{cases}$
Number of participants (median threshold)	$\begin{cases} 1, N_{Participants} \geq threshold \\ 0, N_{Participants} \leq threshold \end{cases}$
Bland and Altman Plot	$\begin{cases} 1, \textit{reported} \\ 0, \textit{Not reported} \end{cases}$

4.3.8 Heterogeneity, missing data, and subgroup analysis

Since all the studies have reported different performance metrics to support the efficacy of methods, calculating a single statistical metric would be infeasible and insufficient. Alternatively, similar numerical metrics were compared to provide a

descriptive summary. The proposed protocol penalizes for missing data, which is required for the study's quality assessment. Since the authors have collected the data from individual studies and assessed the studies' quality based on the data, they have not performed any subgroup analysis. Furthermore, this analysis aimed to provide a narrative summary based on the data collected from the included studies for HR and SpO2 estimations independently.

4.4 RESULTS

4.4.1 Study screening results

Of 332 articles retrieved from the search strategy presented in Table 7 using multiple databases, 32 were included, followed by data collection and analysis. While screening these articles, 18 more studies were included by thoroughly checking each reference list. This technique is called snowballing. Figure 16 (similar to Figure 9) depicts the PRISMA flow diagram illustrating the article's screening process. The data collected from the individual studies, according to section 4.3.4, is presented in Table 11.

Consequently, 50 articles were included in the final review, of which 38/50 (76%) studies estimated a single, 10/50 (20%) studies estimated two, and the remaining estimated three physiological parameters. It is important to note that no explicit search was performed for other parameters except HR and SpO2. However, other parameters were also calculated as a part of HR estimation studies.

However, the study did not include non-contact methods using chest, arm, palm, or finger for HR estimations since this review is constrained to face-based methods only. Furthermore, studies constituting fetal HR monitoring were not considered since HR estimation was performed using the lower abdominal area.

4.4.2 Population characteristics

a) Age and Gender

21/50 (42%) did not report the age, while the remaining 29/50 (58%) studies reported the age range. The minimum and maximum age range for all studies lies between 18 and 80. Furthermore, it is difficult to plot the distribution of age ranges due to the considerable heterogeneity. 34/50 (68%) studies have reported gender,

while 16 studies did not provide any information about age or gender. Apart from 4/50 (8%) studies [18, 65, 79, 110], all the study samples were male-dominant. However, these studies have collected data from relatively fewer participants.

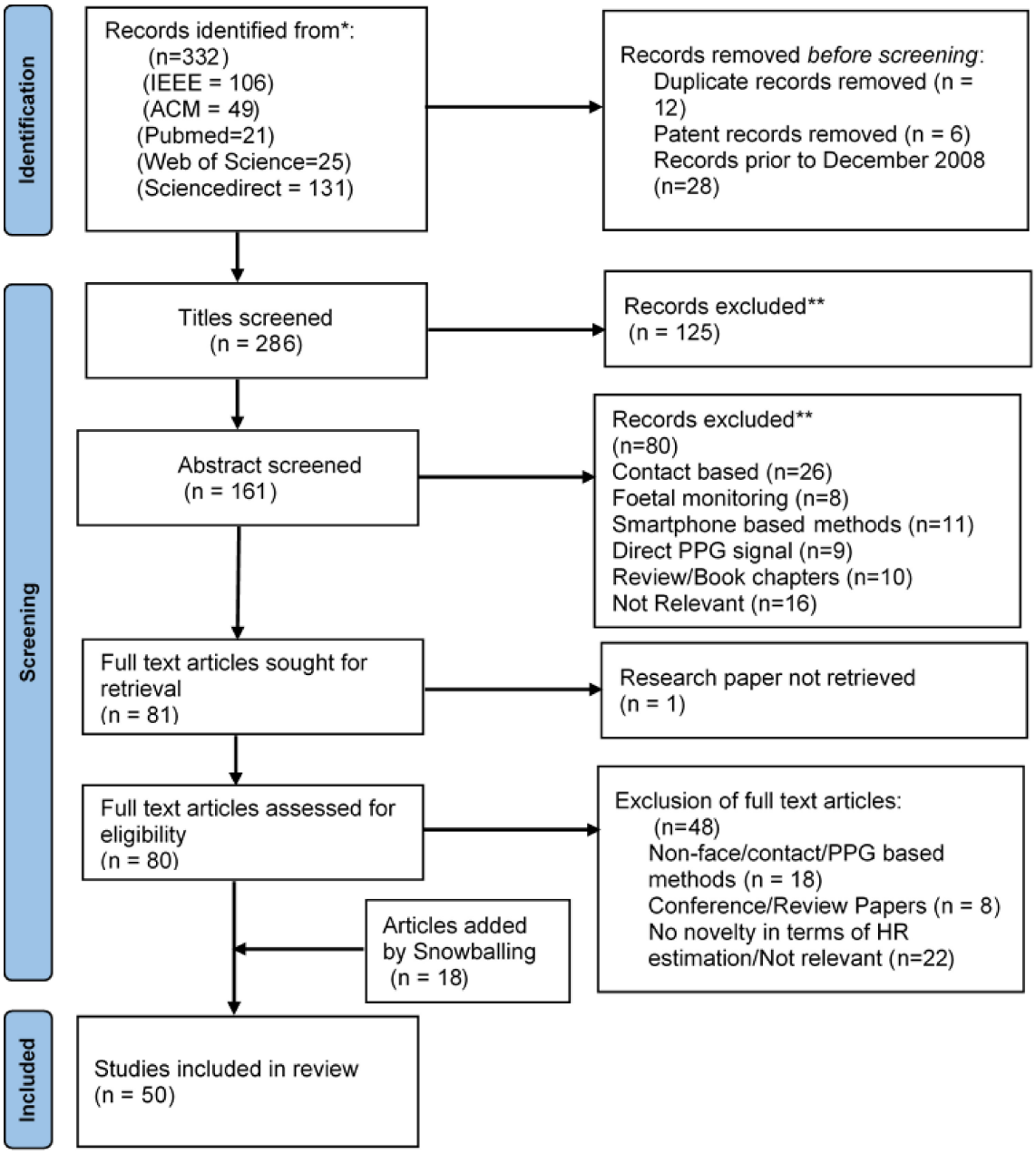


Figure. 16: PRISMA flow diagram for systematic analysis of non-contact HR and SpO2 estimation studies.

b) Ethnicity and skin color

Numerous studies proved the importance of considering the ethnicity or skin color for HR and SpO2 estimation studies since darker skin tone poses more challenges for estimation tasks than white skin. 14/50 (28%) studies reported the subjects'

Table 11: Data collection from individual non-contact HR and SpO_2 estimation studies.

First Author (year)	**Phys. Param.**	**Ref. Device**	**Quality**	**D. size**	**Age**	**Males**	**Females**	**P. Type**	**Ethnicity**	**. ROI**	**Method**	**Modality**	**Source**	**B-A plot**	**Artifacts**
J Cheng [111], (2017)	HR	ECG	0.3	14	22-30	6	8	N	Asian	Cheek	EEMD	Green	A, LED	Yes	I
J Ryu [72], (2020)	HR	ECG	0.3	63	25-28	5	8	N	NR	Cheek	CWT	YCbCr	F	Yes	M
X Yu [112], (2021)	HR, HRV	ECG	0.40	70	73-85	4#	16#	C	NR	Face	CHROM	RGB, NIR	A, C	Yes	M
H Qi [62], (2017)	HR	PPG	0.41	32	19-37	11	11	N	NR	Face	J-BSS, C-MCCA	RGB	A	No	None
W Wang [70], (2017)	HR	ECG	NR	6	NR	5	1	N	i-v	Face	Sub band separation	RGB	F, H, C	No	M
W Wang [63], (2016)	HR	ECG, PPG	0.44	11	NR	8	3	N	i-v	Face	POS	RGB	F	No	M, I

B Wu [74], (2019)	HR	HR monitor	0.3	14	20-25	11	3	N	iii-iv	Cheek	NN	RGB	A	No	M, I
F Bousefsaf [113], (2013)	HR, BR	PPG, Res. sensor	0.3	12	22-79	10	2	N	i-iv	Face	CWT	RGB, CIE	A	Yes	M
M Gastel [23], (2016)	SpO2	PPG	0.7	4	NR	NR	NR	N	ii-v	Face	APBV	IR	idt	Yes	M
AR Guazzi [85], (2015)	SpO2	PPG, ECG, Res. sensor	NR	5	NR	NR	NR	N	ii-iv	Face	ROR	RGB	LED	Yes	None
A Rosa [86], (2020)	SpO2	PPG	0.6	9	23-54	4	5	N	ii-iv	Forehead	ROR	R, B	A	Yes	None
G Haan [69], (2013)	HR	PPG	0.77	117	NR	NR	NR	N	i-vi	Face	CHROM	RGB	A	No	M
M Poh [4], (2010)	HR	PPG	0.3	12	18-31	10	2	N	Asian, African, Caussian	Face	ICA	RGB	A	Yes	M

Q. Tran [114], (2019)	HR	PPG	0.3	22	22-45	14	8	N	White negroes, Asian	Face	APP	RGB	A	No	M, I
O Gupta [115], (2016)	HR, HRV	PPG	1.25	9	25-40	NR	NR	N	Caussian, Asian	Cheek, Fore-head	Fast ICA	RGB, Magenta thermal	A	No	I
R Song [67], (2020)	HR	ECG	4.11	57	22-25	13	2	N	Asian	Cheek	KDICA	RGB	A	No	M
M Kumar [116], (2015)	PR, PRV	PPG	1.31	12	NR	7	5	N	Skin color	Face	MRD	RGB, Mono	A, F	Yes	M, I
M Poh [117], (2011)	HR, HRV, RR	PPG	0.3	12	18-31	8	4	N	NR	Face	JADE	RGB	A	No	M
L Tarassenko [88], (2014)	SpO2, HR, RR	PPG	5	46	49-80	36	10	C	NR	Face	AR, ROR	RGB	F	No	None

B Wei [118], (2017)	HR, BR	PPG, Res. Sensor	0.3	8	22-31	NR	NR	N	NR	Throat, Mouth	SOBI	RGB	A	Yes	M
W Chen [14], (2018)	HR, BR	PPG, ECG	0.32, 3.99, 0.45, 0.15	59	23-50#	53	24	N	NR	Face	CNN	RGB, NIR	A	No	M,I
Y Qiu [75], (2019)	HR	PPG	1.45	40	NR	NR	NR	N	NR	Cheek	CNN, EVM	RGB	NR	No	M
Bousefsaf [76], (2019)	HR	ECG	0.3	43	NR	NR	NR	N	NR	Fore-head Cheek	CNN	None	A, C	Yes	None
Z Yu [119], (2019)	HR, HRV	ECG, PPG	0.4, 3.99	127	NR	NR	NR	N	NR	Face	CNN	RGB	NR	No	None
G Hsu [78], (2020)	HR	ECG, PPG	0.92	157	22-41	8	2	N	NR	Face	CNN	Green	A	No	M
Z Yu [79], (2020)	HR	PPG, ECG	NR	174	22-41#	94	40	N	NR	Face	CNN	RGB	C, F	No	M, I

R Song [47], (2020)	HR	ECG, PPG	0.05	201	19-40, 20-53, 22-41#	108#	44#	N	NR	Cheek	CNN, CHROM	RGB	A	Yes	M
R Macwan [61], (2019)	HR	PPG	0.3	90	NR	NR	NR	N	NR	Face	ICA	RGB	A	No	M
J Cheng [28], (2021)	HR	PPG	0.21, 0.40	22	20-25, 20-40	9	13	N	NR	Face	IVA	NIR	None	Yes	None
M Hu [80], (2021)	HR	PPG	0.05	50	NR	36	14	N	NR	Face	CNN	RGB	A	Yes	M
M Hu [10], (2021)	HR	ECG, PPG	0.05	132	NR	NR	NR	N	NR	Face	CNN	RGB	A, C, F#	Yes	M, I
A Moço [3], (2021)	SpO2	PPG	NR	46	NR	NR	NR	N	NR	Fore-head Cheek	ROR	R, G, IR	idt	No	None
Y Lin [120], (2018)	PR	ECG	NR	10	NR	7	3	N	NR	Nose, Cheek	Adaptive filtering	RGB	A	Yes	M

C Zhao [71], (2019)	HR	PPG, ECG	0.3	18	21-35	14	4	N	NR	Nose, Cheek	POS	RGB	A	No	M, I
X Li [121], (2014)	HR	ECG	0.3, 0.45	38	24-38, 19-40#	15#	12#	N	NR	Cheek, Nose, Mouth	DRMF, DRLSE NLMS	Green	F, A	No	M, I
Y Zhang [68] (2020)	HR	PPG	0.3	30	20-28	18	12	N	NR	Face	EEMD	AB	A	Yes	M
S Kado [16], (2020)	HR	PPG	1.25	38	20-60	30	8	N	NR	Cheek	FastICA	RGB, NIR	F	No	M
X Niu [77], (2019)	HR	PPG	0.69, 0.3, 2.07	174	22-41	79	28	N	NR	Face	CNN, GRU	YUV	A, C, F	No	M
G Tsouri [60], (2012)	PR	PPG	0.07	45	18-45	NR	NR	N	NR	Face	ICA	RGB	A	Yes	None
C Zhang [122], (2017)	HR	PPG	0.92	20	NR	12	8	N	NR	Eyes Area	SOBI	RGB	A	No	M

H Yu [65] (2019)	HR	PPG	2.07	18	NR	NR	NR	N	NR	Face	SSR, CEED-MAN	RGB	A	Yes	M
D Chen [66], (2015)	HR	Sgm	NR	8	NR	7	1	N	NR	Fore-head	EEMD	Green	A	Yes	I
R Song [73], (2020)	HR	PPG	0.3	77	NR	24	13	N	NR	Face	EEMD	RGB	A	Yes	M, I
K Lin [21], (2016)	HR	Sgm	NR	9	22-31	8	1	N	NR	Fore-head	EEMD, MR	Green	A	Yes	M
U Bal [87], (2014)	SpO2, HR	ECG, PPG	0.3	18	NR	NR	NR	N,C	NR	Face	DT-CWT	RGB	A	Yes	M, I
A Woy-czyk [123] , (2021)	HR	PPG, ECG	0.13ˆ	22	22-29	12	10	N	NR	Face	GMM	RGB	A, F, C	Yes	M
D Shao [1], (2016)	SpO2	PPG	0.01	6	24-30	3	3	N	NR	Lips	ROR	Orange, NIR	NA	Yes	None

P Gupta [124], (2020)	HR	PPG	0.3	105	NR	34	31	N	NR	Face	NR	Green	A	No	M
L Kong [90], (2013)	SpO2, HR	PPG	0.07	30	18-58	NR	NR	N	NR	Cheek	FFT	Mono	A	Yes	None
J John [125], (2020)	HR	PPG	0.3	21	NR	NR	NR	N	NR	Fore head	Cust-VJ	Green	A	Yes	None

Note: #- incomplete information, NR-Not reported, Sgm-sphygmomanometer, M-Motion Artifact, I-Illumination Artifact, N-None artifacts considered, A-Ambient Light, C- Ceiling Light, Fi-Filament, Idt- incandescent, MR-Multiple Regression, Mono-Monochrome, Phys-Physiological Parameter for estimation, Quality-Camera resolution in pixels, Ref.-Reference device, D.size-Database size;B-A plot-Bland-Altman Plot.

ethnicity, skin color, or tone information for estimation studies. Furthermore, 8/14 (57.14%) [21, 23, 63, 70, 74, 85, 113] studies used the Fitzpatrick scale to define the subjects' skin tone. The study conducted by Haan and Jeanne [69] included the subjects from i-vi (all scales), while 2 studies conducted by Wang et al. [63, 70] included subjects with the i-v scale. The remaining studies included subjects with a scale ranging from $ii - iv$. On the other hand, 6/14 (42.85%) studies instead mentioned the ethnicities of the subjects: 3 studies [4, 64, 115] considered two or more, 2 studies [28, 63] considered subjects with Asian ethnicity, and one study conducted by Kumar et al. [116] has considered skin color.

4.4.3 Study design

a) Physiological variables

Out of the total included studies, 42/50 (84%) studies belong to HR estimations, while 5/50 (10%) to SpO2 estimations. Additionally, 3/50 (6%) studies estimated both physiological parameters simultaneously. Furthermore, out of 42 studies, HRV, BR, eye blink, and step counts were estimated in 4/50 (8%) [112, 115–117], 4/50 (8%) [14, 88, 117, 118], and 1/50 (2%) (each) studies, respectively.

b) Databases used

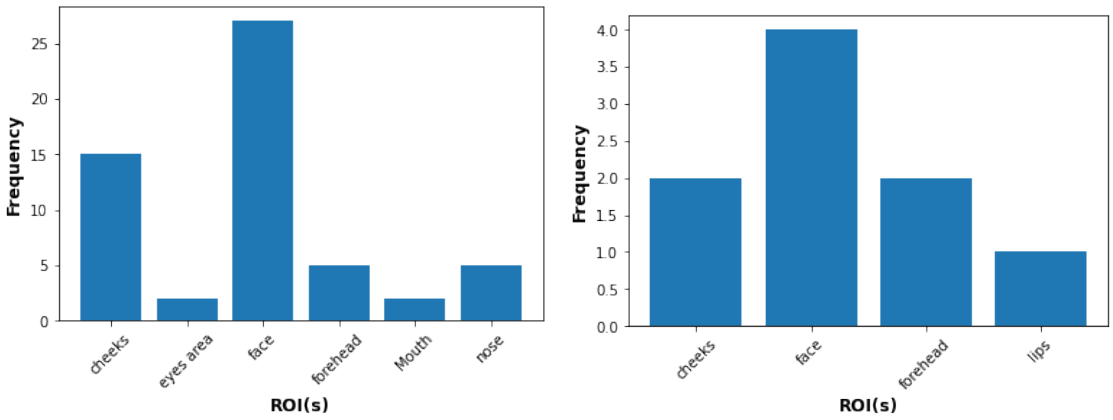
For HR studies, 34/45 (75.55%) studies used self-created databases with the number of participants ranging from $4 - 117$ with 25.82 ± 25.11 ($\mu \pm \sigma$). In contrast, the remaining studies used benchmark databases to prove the efficacy of their respective HR estimation methods. Moreover, 24/34 (35.29%) studies have only used their databases, while the rest have used self-created as well as publicly available databases. 3/11 (27.27%) studies [62, 75, 76] have used a single database, while 8/11 (72.72%) [10, 47, 61, 78–80, 119, 126] studies have employed more than one database for performance analysis of their HR estimation algorithms. All SpO2 studies created their databases with the number of participants ranging from 4 to 46 with 20.5 ± 16.78.

c) Region of interest selection

The face-based physiological parameters estimation needs an ROI, which will be used to extract the source signal. All HR estimation studies used six ROIs: face, cheeks,

nose, forehead, areas near eyes, and mouth. 7/45 (15.5%) [3,71,113,115,118,120,121] studies have used two or more ROIs from the face region, while face and cheeks were used by 27/45 (60%) and 15/45 (33.33%) studies, respectively.

On the other hand, the nose and forehead have been used by five studies each, while the remaining used the mouth and areas near the eyes. The ROI distribution for HR and SpO2 studies is shown in Figures 17a and 17b, respectively, which clearly shows the dominance of the entire face region for both estimations, despite the finding that PPG information is unevenly distributed in the face region, which resulted in the introduction of attention mechanisms to focus on the regions rich in PPG information.



(a) (b)

Figure. 17: ROI distribution for non-contact (left) and SpO2 (right) estimation studies.

d) Artifacts

Artifact removal is vital for non-contact estimation methods for ROI or rPPG signal extraction since the PPG component has a relatively weaker strength and amplitude than the artifacts. HR studies have addressed two types of artifacts, namely motion and illumination artifacts.

4/45 (8.88%) HR studies have addressed and alleviated the effect of illumination, while motion artifacts have been addressed and mitigated in 16/45 (35.55%) studies. On the other hand, 16/45 (35.55%) studies [10, 14, 61, 63, 68, 69, 72, 74, 77–79, 114, 116, 121] have addressed and proposed strategies to lessen both artifact effects. 9/45 (20%) HR and all SpO2 estimation studies did not address any artifacts.

e) Estimation Methods

As mentioned, conventional HR estimation methods extract the ROI/PPG signal from the RGB signal traces, calculate the highest frequency, and multiply it by 60 for frequency to bpm conversion. Among conventional HR estimation methods, 9/45 (20%) studies [4,16,60,61,67,115,117,118,122] used ICA, 6/45 (13.33%) studies [47,63,64,69,71,112] used color subspace transformations (CHROM/POS), and 6/45 (13.33%) studies [21,28,65,66,68,73] used EMD and its variants. However, with the advent of deep learning, several end-to-end HR estimation methods have also been proposed, which use different types of neural network architectures for estimating HR using a facial video. 11/45 (24.44%) [10,47,74,76–80,118,119,127] studies utilized neural networks and their variants for HR estimation.

Other PPG extraction methods used by HR estimation studies are wavelet transforms [72,87,113], filtering-based methods [120], autoregressive models [88], Gaussian mixture models [123], Eulerian video magnification [75], Independent Vector Analysis (IVA) [28,62,80], maximum ratio diversity [116], sub-band selection [72], multiple linear regression [21], and Second Order Blind Identification (SOBI) [118,122]. Figure 18 depicts the distribution of the estimation methods used in the literature. On the other hand, all SpO2 studies used the ROR method followed by regression, except the study conducted by Gastel, Stuijk, and De Haan [23]. This study proposed the adaptive PBV method (APBV) method for ROI signal extraction from multiple ROIs, followed by their mapping to different SpO2 levels ranging from 65 to 100% with an interval of 5%. The APBV method is based on the Pulse Blood Volume (PBV) method proposed by De Haan and Van Leest [128], which extracts the PBV vector for ROI signal extraction.

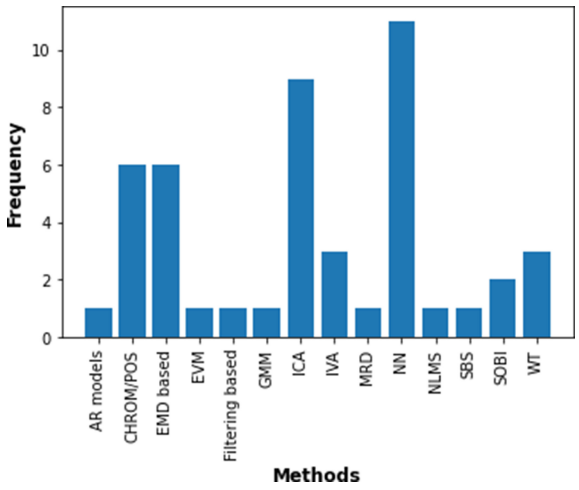


Figure. 18: Various estimation methods used for estimation of non-contact HR studies.

4.4.4 Instruments used

a) Reference Device

Four reference devices have been used for comparing the estimated values with ground truth. This analysis has only reported the use of reference devices for the self-created databases since reference device information for the benchmark databases could be easily extracted from the respective articles. As mentioned before, 34 HR estimation studies have created their databases. 8/34 (23.52%) studies have used electrocardiogram as a reference device, 22/34 (64.70%) studies have used pulse oximeters as a reference device, while one study conducted by Wang et al. [63] used both reference devices. A few HR estimation studies used an arm Band HR monitor [74] and sphygmomanometer [73,86]. On the other hand, SpO2 ground truth acquisition was carried out using a pulse oximeter only.

b) Camera Characteristics

Non-contact studies require a camera for video acquisition of ROI; hence, camera characteristics such as camera type, video resolution, frame rate, and shooting distance affect parameter estimation accuracy. However, the camera type was not included in the study's quality assessment, but it is a critical parameter since it defines the quantity and quality of PPG information for physiological parameter estimation. The camera type is determined by the color channels used for video acquisition in a non-contact study.

The color channel selection is crucial for extracting accurate PPG information. For instance, the RGB spectra have stronger pulsatile strength than the IR spectra. Therefore, this section summarizes the respective studies' color channel distribution and camera types for the included studies. 40/50 (80%) studies used RGB channels, followed by NIR used by 6/50 (10%) studies.

Other color channels used in the existing non-contact studies include monochrome color filters [89, 90, 115], YCbCr [72, 87], YUV [77], and LAB [112]. Furthermore, a few studies have used more than one spectra, e.g., Kado et al. [16] and Yu et al. [112] combined RGB and NIR spectra. The details regarding the color channels are presented in Table 11. YUV, YCbCr, and LAB color channels can be deduced from the RGB image model; therefore, 42 studies have used RGB cameras. Four studies used NIR cameras, five used customized camera setups, three used monochrome, and two used RGB-NIR spectra combination.

The distance between the subject's face and the camera is another vital parameter since a larger distance between the face and the camera deteriorates the strength of PPG signal information. Hence, it is necessary to identify a suitable shooting distance for a cleaner PPG signal. The shooting distance for HR and SpO2 estimations ranged between 0.3 and $2m$, respectively, while the widely used shooting distances were 0.5 and $1.0m$, used by 11/50 (22%) and (13/50) (26%), respectively.

However, 4/50 (8%) studies used less than 0.5 m shooting distance, while it is 1.5 m or greater for 10/50 (20%) studies. Furthermore, a few studies have acquired the videos using multiple shooting distances. Specifically, the studies conducted by Song et al. [67] and Tran et al. [64] tested the effect of video shooting distances on HR estimation. In contrast, different shooting distances for different activities were also used by Li et al. [121]. 12/50 (24%) HR estimation studies did not report camera shooting distances.

Camera resolutions also play a vital role in accurate HR estimation by providing finer details from individual image frames, which are crucial to detecting subtle color changes for extracting PPG information. A higher camera resolution provides more information but also needs intense computations. Hence, identifying a camera resolution with minimal information loss is a non-trivial task for accurate HR estimation. A diversified range of video camera resolutions has been employed to estimate accurate HR and SpO2 estimations. 19/45 (42.22%) HR estimation studies used cameras with a resolution of 640×480. In contrast, the twelve studies used 320×240, 1280×720, and 1920×1080, each employed by four HR studies.

A higher frame rate provides a larger number of contiguous images for a video, thereby providing more information to detect the ROI from raw RGB signal traces. Similar to resolution, a higher frame rate video requires more computational requirements. Hence, it is necessary to use a frame rate that ensures minimal loss and provides a noise-free PPG signal. All estimation studies, except one by Song et al. [67], have reported the frame rates for video acquisition with a range of 12˘120 FPS. 29/50 (58%) estimation studies used 30 FPS for video acquisition. Other frame rates used by estimation studies were 15 FPS [3, 4, 23, 60, 117], 20 FPS, and 25 FPS [10, 63, 69, 70, 80]. However, numerous studies have also gathered video samples at a higher sampling rate, for instance, 50 FPS [112], 60 FPS [119], 100 FPS [112], etc.

4.4.5 Clinical studies

Although most estimation studies were conducted on healthy individuals, three clinical studies [87, 88, 112] have been included using the search strategy mentioned in Table 7. Most importantly, these studies have estimated two or more physiological variables. Yu et al. [112] conducted a study on geriatric patients that aimed at estimating HR and HRV, while the study undertaken by Tarassenko et al. [88] estimated the HR, SpO2, and BR of the patients undergoing dialysis. The study conducted by Bal [87] aimed at estimating the HR and SpO2 in the pediatric ICU. It is worth mentioning that this systematic analysis suffers from lead time bias; therefore, the number of clinical studies might vary according to the date.

4.4.6 Performance metrics

a) HR Performance metrics

The performance analysis for HR estimation studies utilized five metrics, namely, ME and SD, RMSE, Mean of Error-Rate percentage (MER), Signal-to-Noise Ratio (SNR) ratio, and correlation. Among all of them, most studies used RMSE and correlation. The mean and SD of all metrics are provided in Table 12. A few studies have tested their estimation methods under different application scenarios or using multiple databases wherein average RMSE or correlation values were calculated for the analysis. Additionally, accuracy and B-A analysis were also included. 25/45

Table 12: Performance metrics statistics for HR estimation studies.

Metric	**Number of Studies**	**mean ± SD**
ME	9	0.57 ± 1.49
SD	17	4.91 ± 2.41
RMSE	28	5.15 ± 3.24
MER	8	6.03 ± 1.26
Correlation	26	0.88 ± 0.11
SNR	5	3.17 ± 1.75

(55.55%) studies reported RMSE, of which 10/25 (40%) achieved the RMSE within ± 5 bpm, whereas the remaining studies have a mean RMSE of 2.73 bpm and SD of 1.32 bpm. A box-whisker plot in Figure 19 depicts the distribution of error metrics. Higher RMSE values for the two studies correspond to testing the proposed methods under challenging conditions, such as fitness exercise and extreme temperature conditions during data acquisition.

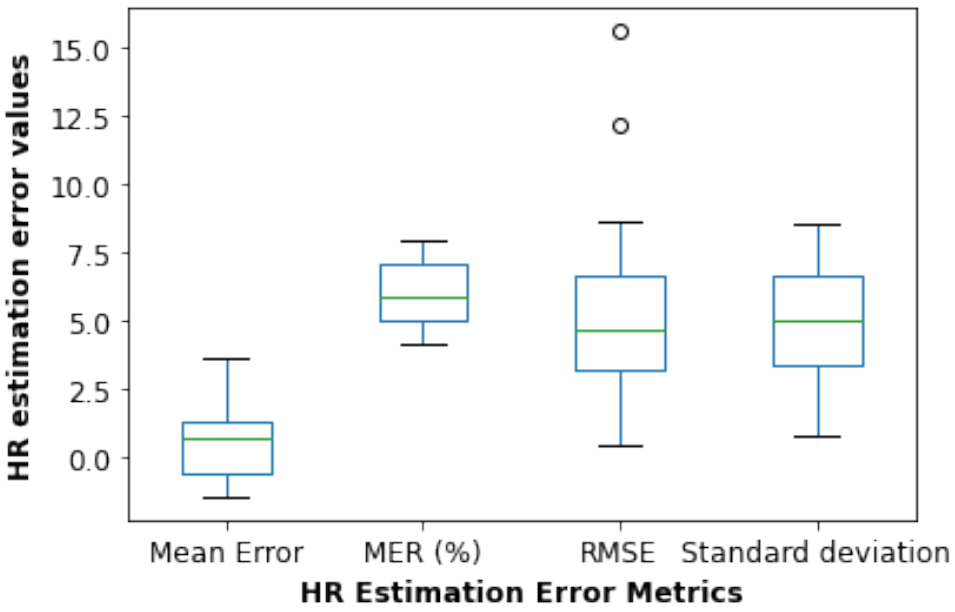


Figure. 19: Error metrics distribution for non-contact HR estimation studies.

Pearson correlation values have been reported in 26/45 (57.77%) of HR estimation studies. Among them, 15/26 (57.69%) [4,10,21,61,66–69,72,75,87,117–119,121] studies achieved a correlation value of 0.90 or more, while the average correlation value for the remaining studies is 0.77 with the SD value of 0.093. 10/45 (22.22%) [16, 62, 64, 72, 74, 112, 121, 124, 126] studies have reported accuracy, calculated as the percentage of samples with the error difference ± 5bpm between ground truth and estimated HR values. This metric is used to justify the clinical relevance since the clinically accepted error between reference device measurement and estimation is ± 5bpm [100]. The Pearson correlation values and accuracy distribution are shown in Figure 20. Additionally, Table 13 presents the reported performance metrics from the individual non-contact HR estimation studies.

Table 13: Reported performance metrics from the non-contact HR estimation studies.

First author	**ME**	**RMSE**	**MER**	**Correlation**	**Accuracy**	**SNR**	**Error SD**
Hsu [78],2020	-1.40	5.45	7.07	0.71	-	-	6.66
J Ryu [72], 2021	0.56	3.59	-	0.94	93.25	-	-
Z Yu [119], 2019	-	6.55	-	0.82	-	-	1.76
R. Song [47], 2020	-	7.45	7.97	0.75	-	-	7.31
Y. Zhang [68], 2021	-	-	-	0.96	-	-	-

P. Gupta [124], 2020	-0.54	-	-	0.80	94.5	-	5.00
R. Song [67], 2020	-	3.28	-	0.95	-	-	3.57
S. Kado [16], 2020	-	-	-	-	80.37	-	-
Y Qiu [75], 2019	-1.42	5.11	5.11	0.97	-	-	4.82
R. Macwan [61], 2019	-	-	-	0.93	-	2.98	-
B. Wu [74], 2019	-	5.79	-	-	78.25	-	-
X. Niu [77], 2020	0.73	8.14	6.71	0.76	-	-	8.11
C Zhao [71], 2020	-	15.62	-	-	-	2.16	-
Y. Lin [120], 2018	-	5.83	-	-	-	-	-
J. Cheng [28], 2016	-	-	-	0.53	-	-	8.46
H Qi [62], 2017	3.65	5.00	5.17	0.74	72.79	-	3.42
K. Lin [21], 2016	-	-	-	0.97	91.40	-	2.90
O. Gupta [115], 2016	-	-	4.90	-	-	-	-
D Chen [66], 2015	-	-	-	0.91	84.75	-	-
U. Bal [87], 2015	-	3.09	-	0.96	98.21	-	-
G. Haan [69], 2013	-	0.50	-	1.00	-	-	0.90
G. Tsouri [60], 2012	-	3.50	-	-	-	-	-
M. Poh [117], 2010	0.95	1.24	-	1.00	-	-	0.83
R. Song [73], (2021)	-	6.14	-	0.86	-	-	5.38
J. Cheng [126], 2021	-	7.26	7.145	0.78	67.97	-	5.56
M.Hu [80], 2021	-	4.37	-	-	-	-	-

Q. Tran [64], 2019	-	7.17	-	-	91.10	-	-
A. Woyczyk [123], 2021	-	12.22	-	-	-	-	-
X.Yu [112], 2021	-	0.76	-	0.88	-	-	-
W. Wang [63], 2016	-	-	-	-	-	5.16	-
F. Bousefsaf [113], 2013	-	2.15	-	0.87	-	-	-
B. Wei [118], 2017	-	1.63	-	0.95	-	-	-
X Li [121], 2014	1.30	4.45	4.20	0.90	76.5	-	3.99
W. Wang [70], 2017	-	-	-	-	-	0.56	-
F. Bousefsaf [129], 2013	1.31	8.64	-	-	-	-	-
M. Hu [10], 2021	-	4.47	-	0.90	-	-	-
Z. Hu [80], 2020	-	1.80	-	0.99	-	-	6.30
M Kumar [116], 2015	-	-	-	-	-	5.03	-
M. Poh [4], 2010	-	3.46	-	0.97	-	-	-
F. Bousefsaf [76], 2019	-	-	-	-	-	-	8.55

Note: Correlation signifies Pearson correlation at the significance level of 0.001; "-" indicates that the particular performance metric is not reported in the study. All values are reported to double-digit precision for brevity.

23/45 (51.11%) studies have included B-A plots in their analysis. Additionally, one study by Yu-Chen Lin and Yuan-Hsiang Lin [120] did not present the B-A plot but rather the level of agreement. Figure 21 depicts the mean bias and upper and lower levels of agreement for HR estimation studies in chronological order. 8/23 (34.78%) studies achieved the mean difference within the clinically accepted range, while the rest may need significant improvements in the future. Table 14 presents the mean bias upper and lower statistical limits of the B-A analysis for individual studies.

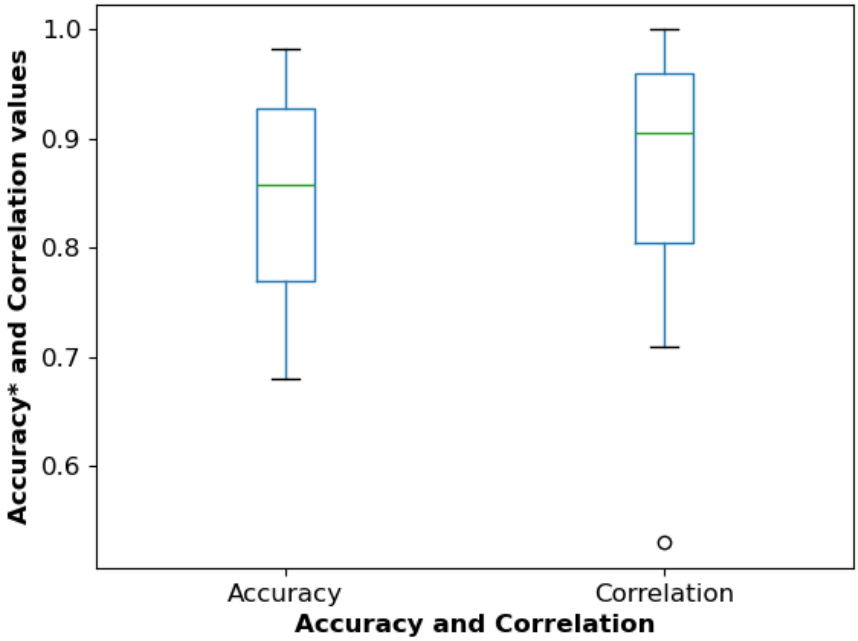


Figure. 20: Performance metrics distribution for non-contact HR estimation studies.

b) SpO2 estimation studies

As mentioned, 7/8 (87.5%) non-contact estimation studies used regression analysis for SpO2 (with range $80-100\%$) calculations, utilizing the ROR using light intensities of two wavelengths. 5/8 (62.5%) [3,4,85,89,90] studies have reported R^2 values. Furthermore, the root mean squared metric (A-rms %) was calculated for two studies [3,86], while the same number of studies [89,112] have used Pearson correlation value to test the method's performance. 3/5 (60%) studies [4,85,89] have achieved an R^2 value of 0.8 or more, while the remaining two achieved relatively lower values, i.e., 0.65 and 0.58, respectively. Overall, the mean R^2 value is 0.78, with an SD value of 0.14. Furthermore, 5/8 (62.5%) studies have used B-A plots to showcase the performance of the proposed methods, as depicted in Figure 22, and the B-A plots statistical metrics are presented in Table 15.

4.4.7 Challenges

Non-contact methods for HR estimation studies deal with three types of noises: camera quantization, motion, and illumination noise. Almost all the studies have assumed a constant light illumination incident on every region of the face, which does not comply with real-time situations. Secondly, the PPG signal's low strength

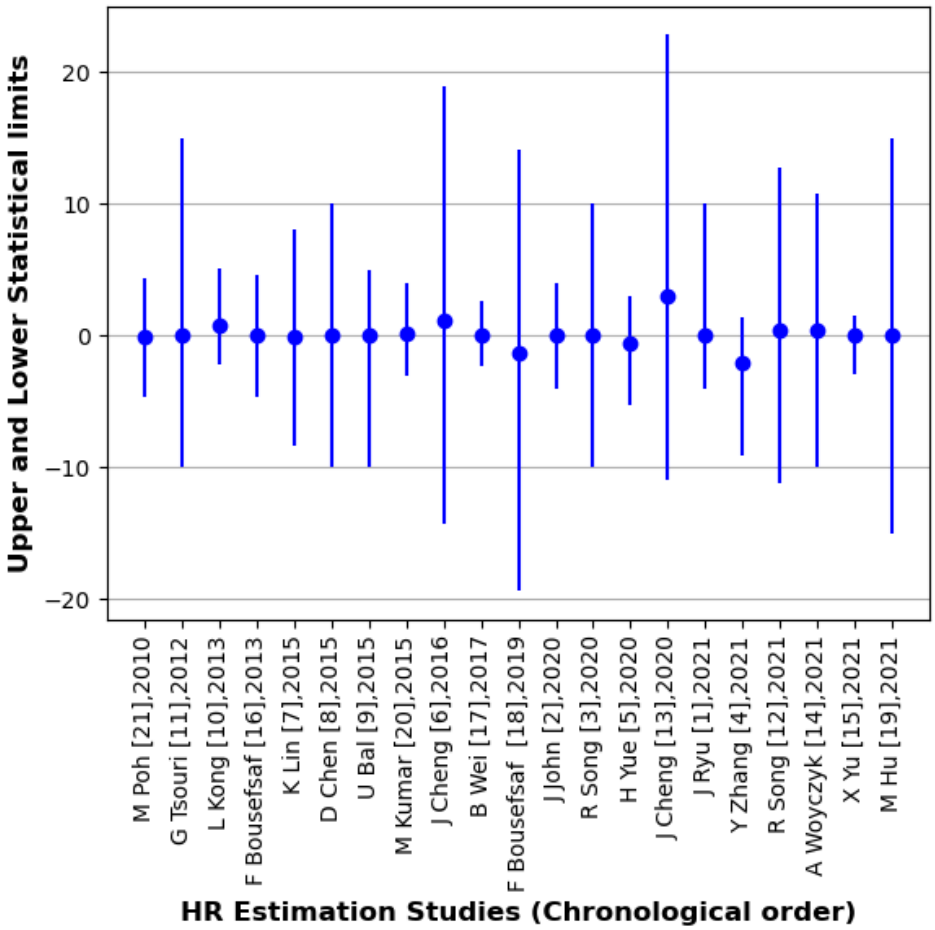


Figure. 21: Bland-Altman analysis for non-contact HR estimation studies.

compared to the noise due to motion and illumination artifacts poses a significant challenge for extracting the cleaner PPG signal. Low resolution and camera shooting distance further degrade the quality of the acquired PPG signal. Furthermore, a temperature colder than the ambient may increase the blood viscosity, reducing the blood flow and resulting in inaccuratePPG signal extraction. Therefore, camera quantization noise, motion and illumination variation artifacts, extremely colder temperature conditions, and camera characteristics such as the camera shooting distance and resolution are critical challenges that must be addressed while designing the non-contact estimation study.

4.4.8 Applications

The majority of existing non-contact methods for physiological parameters estimations are in a proof-of-concept stage. However, several studies have deployed their methods for real-time applications. 8/50 (18%) studies have deployed their proposed

Table 14: Bland-Altman metrics for non-contact HR estimation studies.

First Author (year)	**Mean Bias**	**Upper Statistical Limit**	**Lower Statistical Limit**
J Cheng [28], 2016	1.15	17.73	-15.43
J Ryu [72], 2021	0.00	10.00	-4.00
X Yu [112], 2021	0.02	-1.50	-2.95
F Bousefsaf [113], 2013	-0.01	4.575	-4.59
M Poh [4], 2010	-0.05	4.44	-4.55
M Kumar [116], 2015	0.23	3.71	-3.24
B Wei [118][110], (2017)	0.10	2.50	-2.40
F Bousefsaf [76], 2019	-1.31	15.45	-18.07
R Song [47], 2020	0.00	>10.00	< -10.00
J Cheng [126], 2020	3.00	19.90	-13.90
M Hu [10], 2021	0.00	-15.00	15.00
Y Zhang [68], 2021	-2.00	3.40	-7.14
G Tsouri [60], 2012	0.00	15.00	-10.00
H Yue [65], 2020	-0.54	3.60	-4.69
D Chen [66], 2015	0.00	10.00	-10.00
R Song [73], 2021	0.40	12.40	-11.60
K Lin [21], 2015	-0.06	8.11	-8.24
U Bal [87], 2015	0.00	5.00	-10.00
A Woyczyk [123], 2021	0.48	10.32	-10.49
L Kong [90], 2013	0.80	4.30	-3.00
J John [125], 2020	0.00	4.00	-4.00

Note: Mean bias is the mean of positive and negative differences between ground truth and estimation values, while lower and statistical limits are calculated using the formula $\mu \pm \sigma$

Table 15: Bland-Altman metrics for non-contact SpO_2 estimation studies.

First Author (year)	**Mean bias**	**Upper Statistical limit**	**Lower Statistical Limit**
M Gastel [23], 2016	0.10	4.00	3.80
A Guazzi [85], 2015	0.19	6.00	-5.81
A Rosa [86], 2019	-0.10	1.90	-2.20
D Shao [1], 2016	-0.07	2.51	-2.65
L. Kong [90], 2013	0.62	2.40	-1.60

Note: Mean bias is the mean of positive and negative differences between ground truth and estimation values, while lower and statistical limits are calculated using the formula $\mu \pm \sigma$.

non-contact estimation methods in real-time situations like fitness exercise, driving, and clinical conditions.

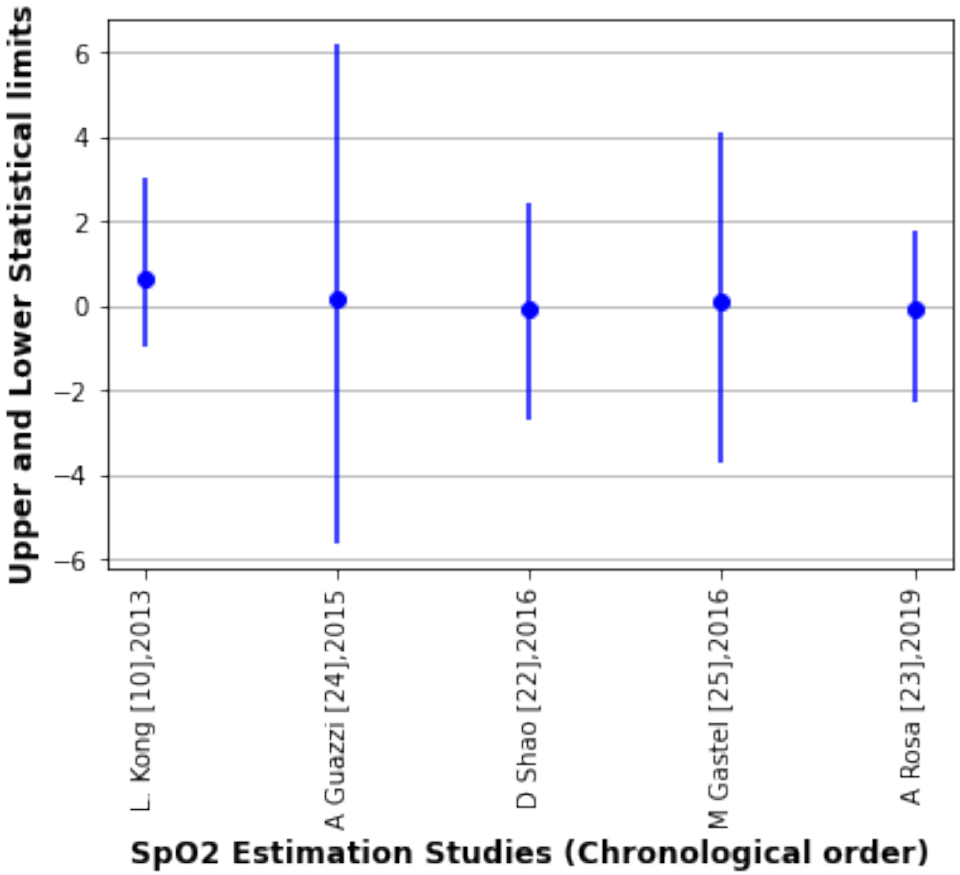


Figure. 22: Bland-Altman analysis for non-contact SpO2 estimation studies.

3/8 (33.33%) studies have tested their methods in clinical conditions. Specifically, Tarassenko et al. [88] estimated HR, SpO2, and RR of the patients undergoing dialysis, Bal [87] monitored the health status of patients in pediatric ICU by estimating HR and SpO2, and Yu et al. [112] estimated HR and HRV for Geriatric patients undergoing physiotherapy treatment. Furthermore, 4/8 (55.55%) [63, 69, 71, 80] studies have estimated heart rate during fitness exercises, while one study by Wu et al. [74] used a non-contact approach for HR monitoring of drivers. This proves the effectiveness of non-contact methods in real-time applications where it is infeasible to use conventional or gold-standard physiological parameter estimation techniques.

On the other hand, a few studies, such as [122] and [120], have also estimated parameters such as step count and eye blink. Step count can be used to monitor fitness, whereas eye blinking can be used to analyze sleep quality.

4.4.9 Study quality assessment results

This study has identified seven vital parameters for HR estimation studies, namely: camera characteristics (camera resolution and shooting distances), Bland-Altman

analysis, results score performance metrics (RMSE and correlation), artifacts, accuracy (error $\leq \pm$ 5 bpm), number of subjects used for the study, and inclusion/exclusion of ethnicity. The detailed analysis of assessing the non-contact HR estimation studies is presented in Table 16 (based on Table 9).

On the other hand, the SpO2 estimation studies quality was assessed using the following four parameters: camera characteristics, number of subjects, inclusion/exclusion of Bland-Altman analysis, and the coefficient of determination (R^2). Other parameters similar to HR studies, such as artifacts, accuracy, RMSE, correlation, and ethnicity, were not reported in most studies and, hence, not included in this analysis. The detailed analysis of the studies' quality assessment for non-contact estimation studies is presented in Table 17 (based on Table 10). The studies were categorized into three categories: "strong", "fair", and "weak", as depicted in Figure 23a. Based on the proposed protocol, 5 HR studies were identified as "strong" reporting maximum specified parameters, while the number of "fair" and "weak" studies were found to be 30 and 11, respectively. On the other hand, 3 SpO2 studies are categorized as weak, two as fair, and three as strong, as presented in Figure 23b.

4.5 DISCUSSION

This review is intended to search and summarize the currently existing facial video-based non-contact methods for estimating two widely used physiological variables, HR and SpO2, respectively. Monitoring these variables under real-time environments such as clinical conditions, driving, or fitness exercise will only be feasible using a non-contact approach since it allows higher degrees of freedom, unlike contact-based approaches. The analysis and comparison of multi-factors opted for diverse studies would enable researchers to wisely select the important parameters for designing the study and assist them in quantifying their respective studies based on the distribution of various error metrics, correlation, and accuracy using this analysis.

4.5.1 Context of evidence and limitations

The non-contact estimation approaches summarized in this analysis are at the proof-of-concept stage with a few shortcomings. These include relatively constrained video acquisition settings, smaller sample sizes, and a limited clinical context. The reference devices used for most of the studies are pulse oximeters, while some have also used ECG. While other studies have used other devices whose accuracy can be questionable compared with a standard HR monitoring device (ECG or PPG). Moreover,

Table 16: Studies quality assessment for non-contact HR estimation studies.

First Author (Year)	**Camera Score**	**Subject Score**	**Ethnicity Score**	**B-A Score**	**Artifacts Score**	**Results Score**	**Accuracy**	**Total Score (Out of 8)**	**Study Quality category**
J Cheng [16], 2016	1	0	1	1	1	0	0	4	Fair
J Ryu [72][73], 2021	1	0	0	1	2	1	1	6	Strong
X Yu [112][105], 2021	1	1	0	1	1	1	0	5	Fair
H Qiu [75], 2017	0	1	0	0	0	0	1	2	Weak
W Wang [70], 2017	0	0	1	0	1	1	0	3	Weak
W Wang [63], 2016	0	0	1	0	2	1	0	4	Fair
B Wu [74], 2019	1	0	1	0	2	1	1	6	Strong
F Bousefsaf [113], 2013	1	0	1	1	1	1	0	5	Fair
G Haan [69], 2013	0	1	1	0	2	1	0	5	Fair
M Poh [117], 2010	1	0	1	1	1	1	0	5	Fair
Q Tran [64], 2019	1	0	1	0	2	1	1	6	Fair
O Gupta [115], 2016	0	0	1	0	1	1	0	3	Fair
R Song [67], 2020	1	1	0	0	1	1	0	4	Fair
M Kumar [116], 2015	1	0	0	1	2	1	0	5	Fair
M Poh [117], 2010	1	0	0	0	0	1	0	2	Weak
L Tarassenko [88], 2014	1	1	0	0	0	1	0	3	Fair
B Wei [118], 2017	0	0	0	1	0	1	0	2	Weak
W Chen [14], 2018	0	1	0	0	2	1	0	4	Fair
Y Qiu [75], 2019	0	1	0	0	1	1	0	3	Fair

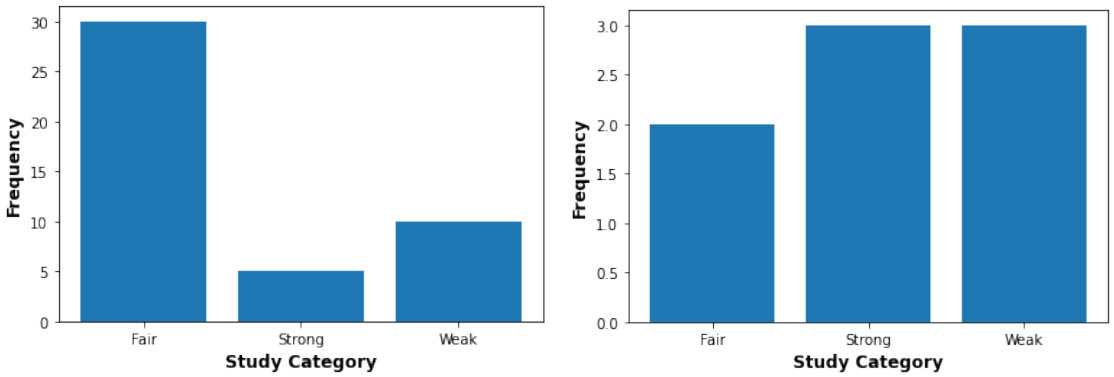
Bousefsaf [76], 2019	0	1	0	1	0	1	0	3	Fair
Z Yu [119], 2019	0	1	0		0	0 1	0	2	Weak
G Hsu [78], 2020	1	1	0	0	2	0	0	4	Fair
Z Yu [79], 2020	0	1	0	0	2		0 0	3	Fair
R Song [47], 2020	0	1	0	1	1	0	0	3	Fair
R Macwan [61], 2019	1	1	0	0	2	1	0	5	Fair
J Cheng [126], 2021	1	0	0	1	0	0	1	3	Fair
M Hu [80], 2021	0	1	0	1	1	1	0	4	Fair
M Hu [10], 2021	0	1	0	1	2	1	0	5	Fair
Y Lin [120], 2018	0	0	0	1	1	1	0	3	Fair
C Zhao [71], 2020	0	0	0	0	2	1	0	3	Fair
X Li [121], 2014	1	1	0	0	2	1	1	6	Strong
Y Zhang [68], 2021	1	1	0	1	2	1	0	6	Strong
S Kado [16], 2020	1	1	0	0	2	1	1	6	Strong
X Niu [77], 2020	1	1	0	0	2	0	0	4	Fair
G Tsouri [60], 2012	0	1	1	1	0	1	0	4	Fair
C Zhang [122], 2017	1	0	0	0	1	1	0	3	Fair
H Yue [65], 2020	1	0	0	1	1	1	0	4	Fair
D Chen [66], 2015	0	0	0	1	1	1	0	3	Fair
R Song [73], 2021	0	1	0	1	2	0	0	4	Fair
K Lin [21], 2016	0	0	0	1	1	1	1	4	Fair
U Bal [87], 2015	1	0	0	1	1	1	1	5	Fair

A Woyczyk [123], 2021	1	1	0	1	1	1	0	5	Fair
P Gupta [124], 2020	0	1	0	0	1	0	1	3	Fair
L Kong [90], 2013	1	1	0	1	0	1	0	4	Fair
J John [125], 2020	1	0	0	1	1	1	0	4	Fair

the selection of a valid reference device plays a crucial role in assessing the applicability of the proposed method in comparison to it. Additionally, a valid reference device could also address the limitations of the ECG or PPG in interpreting the results.

Table 17: Studies quality assessment for non-contact SpO2 estimation studies

First Author (Year)	**Camera Score**	**Results Score**	**B-A Score**	**Subject Score**	**Total Score (Out of 4)**	**Study Quality**
M Gastel [23], 2016	0	1	1	0	2	Fair
A Guazzi [85], 2015	0	0	1	0	1	Weak
A Rosa [86], 2019	1	0	1	1	3	Strong
L Tarassenko [88], 2014	1	0	0	1	2	Weak
A Moço [3], 2021	0	0	0	1	1	Weak
U Bal [87], 2015	1	0	0	1	2	Fair
D Shao [1], 2016	1	1	1	0	3	Strong
L Kong [90][84], 2013	1	0	1	1	3	Strong



(a) (b)

Figure. 23: Studies quality assessment results for non-contact HR (left) and SpO2 (right) estimation studies.

Most studies have used lower camera resolution, making it a cost-effective solution for real-time monitoring, such as driving, fitness exercises, and clinical monitoring. However, selecting the appropriate video resolution is challenging and also affected by the distance between the camera and the subject's face. A study conducted by Song et al. [67] attempted to find the optimal resolution and camera shooting distance. It was concluded that higher resolution enhances the quality of the PPG signal, whereas a distance of more than 1 m will deteriorate the HR estimations, which is consistent with the findings of this analysis. However, high-resolution

cameras are computationally intensive for estimating physiological variables, and a camera distance of 1 m or less limits the applicability of non-contact approaches for clinical or sleep settings. Additionally, the frame rate plays a crucial role in tracking subtle variations present in the image sequences, which ensures an accurate PPG signal.

A few attempts have been made with higher sampling rates, but the performance of HR estimation with a higher frame rate in comparison with 30 FPS did not show significant performance improvement [130]. Furthermore, a frame rate of 30 FPS worked well with most of the studies. Except for clinical and SpO2 estimation studies, most studies have acquired data for about 1 min to 3 min, which may limit the legitimacy of the methods. Moreover, a shorter interval may hinder the robustness of the proposed method under different estimation conditions.

The SpO2 estimation studies have considered relatively longer duration subject's videos. The limitation of the prolonged video acquisition is the presence of artifacts due to movement and uneven illumination variations. Besides, image quantization can also produce undesirable noise, but this effect can be mitigated by assuming constant light over the ROI. The presence of these artifacts deteriorates the estimations of the physiological variables, as the PPG signal (from the RGB color channel) is very weak and is challenging to extract from the artifacts' corrupted signals.

An obvious approach to mitigate this problem would be to convert RGB to other motion or illumination artifact-resistant color channels such as YUV [77], LAB [68], etc. For dark scenarios, an IR channel could be a better alternative, but the only problem is that the strength of the PPG signal is relatively weaker than the signal from the RGB channel. A combination of RGB with IR, which is similar to the one conducted by Kado et al. [16] or other color models, may produce promising results but at the cost of increasing the problem's complexity while also being computationally intensive.

A vast category of rPPG extraction methods has been used in the literature. Neural networks and their variants have been extensively used for HR estimation studies. The neural networks-based methods performed relatively better than other conventional rPPG estimation methods. Additionally, there has been extensive use of transfer learning for HR estimation methods. Most importantly, neural networks do not need assumptions to process the data, which was the case with the existing SOTA methods. On the other hand, SpO2 studies employed regression using the ROR method with an exception [23], which used ROI signals to map to SpO2 levels ranging from 65 to 100%.

Among the HR estimation studies, 8/45 (17.78%) [4, 65, 90, 112, 113, 116, 118, 125] have achieved clinically acceptable error differences, as depicted in Figure 21, whereas other studies might need significant improvements in the future. Furthermore, several studies have justified their method's clinical relevance by reporting the accuracy, which is calculated as the percentage of study samples having an error of less than ±5 bpm. However, it is worth noting that these studies have predominantly used their self-created databases under well-constrained laboratory conditions to test their method in the normal HR and SpO2 ranges. The performance of these methods may deteriorate for abnormal HR parameter ranges. In addition, the accuracy in this scenario will not be sufficient to justify their clinical relevance and, therefore, needs further analysis. On the other hand, the performance of SpO2 estimation under extreme conditions is challenging to test since it needs multiple breath-holding events, which is not always possible for all individuals. Consequently, developing a robust SpO2 estimation method proves to be difficult due to the need to measure the subtle changes in the saturated blood. Therefore, there are limited SpO2 estimation studies in the literature. There is a need to devise methods for estimating SpO2 values from a single PPG signal extracted from the facial ROI, similar to other physiological parameters.

Another limitation within almost all studies is that the parameters were estimated for healthy individuals, limiting the estimation methods' ability for diseased people with conditions such as hypoxemia, bradycardia, or tachycardia.

Finally, this study found some common factors applicable to all non-contact approaches to estimate both physiological parameters. Ethnicity, movement, illumination, and clinical relevance depicted by accuracy are the common factors used by all studies. Although none of the SpO2 estimation studies have reported movement and illumination artifacts, these factors are worth considering while developing non-contact approaches for physiological variable estimations.

4.5.2 Limitations of the analysis

There are certain limitations and challenges associated with synthesizing this analysis. The search strategy aimed to summarize and analyze novel methodologies for HR and SpO2 estimations using facial videos only. This excludes the estimation studies from other body parts, which may limit the findings and analysis to the face region. Being a review-based analysis, patents and commercial applications were excluded for this analysis. This review did not include unpublished studies or conference papers (except for a few discussed in“ eligibility criteria”). Therefore, this

review may have publication bias. Furthermore, reporting and lead-time bias may be possible since databases were used to collect research articles using the presented search strategy. In addition, it was difficult to compare the studies with missing information, such as an insufficient description of the population and highly diversified error metrics. These factors might affect the assessment of the "risk of bias" among studies. Additionally, some studies have presented their results using visual representation; it is challenging to extract numerical values from them for analysis. Overall, there is a high level of heterogeneity among studies, which was difficult to tackle for study quality assessment.

4.5.3 Future research and recommendations

This analysis aims to provide insight into this rapidly growing domain of developing non-contact approaches for face-based physiological variable estimations. The seven critical factors for HR estimations and four for SpO2 estimations, were identified which should be addressed while designing an estimation study. The data collection is a crucial step that might affect the efficacy of the proposed method. The study population should be described appropriately for better representation and result analysis. For instance, information such as age, gender, video resolution, and camera shooting distance should be provided to compare two or more studies. Furthermore, the selection of valid reference devices needs to be done for ground truth data collection. Traditional HR estimation is a complicated process and requires certain conventional assumptions with limited generalizability, while neural networks-based methods are less dependent on these assumptions with good generalizing ability for highly diverse study samples [47]. Hence, the applicability of neural network-based methods is recommended for physiological variable estimations. Comparing estimated values with ground truth should be done using appropriate performance metrics. Although it is challenging to have a defined set of metrics for performance analysis of the proposed methods, reporting the following metrics: RMSE, correlation, accuracy, and Bland-Altman plots is highly recommended.

Furthermore, for SpO2, all studies used the ratio of ratios method considering two color channels; an alternative approach utilizing cleaner ROI/PPG signals like in the study by Gastel, Stuijk, and Haan [23] could be beneficial in ensuring accurate estimations. Designing a SpO2 estimation study is challenging, especially under hypoxemic events or severely infected COVID patients, since it is difficult to collect data aligned with these conditions. Moreover, the conventional ROR method has certain limitations in tracking subtle variations from saturated blood. Finally,

all studies have used a shorter time frame for physiological parameter estimation in healthy individuals. Future studies should focus on methods that take less time to estimate variables using longer video sequences, handling challenges like motion and illumination artifacts, and performing under different ethnic groups in real-time scenarios. Furthermore, future estimation studies should also focus on estimating under the abnormal parameter ranges or during a cardiopulmonary or related diseased condition.

Based on the critical factors and limitations identified from this systematic analysis, and following the recommendations, a new method was proposed, which is based on ICA. This new method elevates the ordering problem and appropriate IC selection. The details about this new method and its performance analysis under different scenarios will be explained and presented in the next chapter.

Chapter 5

UNDERCOMPLETE ICA FOR HR ESTIMATION

HR estimation is of utmost importance due to its applicability in diverse fields. Conventional methods for HR estimation require skin contact and are not suitable in certain scenarios, such as sensitive skin or prolonged unobtrusive HR monitoring. Therefore, rPPG methods have become an active area of research. These methods utilize the facial videos acquired using a camera followed by extracting the BVP or rPPG signal (both terms are used interchangeably, as both terms point to the signal based on BVP) for HR calculation. The existing rPPG methods either utilized a single-color channel or weighted color differences, which has certain limitations in dealing with motion and illumination artifacts.

This chapter considers BVP extraction as an undercomplete problem and proposes a method resistant to motion and illumination variation artifacts. This method is based on an U-ICA, aiming to estimate the unmixing matrix using a non-linear Cumulative Density Function (CDF) that has been optimized using the customized LMA. Therefore, the method is named U-LMA.

The proposed method was tested under three scenarios: constrained, motion, and illumination variations scenarios. The performance of the proposed method was tested using the performance metrics explained in Chapter 3. The values achieved by the performance metrics proved the evidence of its superiority over other SOTA methods. For instance, the proposed method achieved the lowest values for error metrics and achieved higher accuracy with correlation values close to 1 between estimated and ground truth HR values. At the same time, it also has certain limitations, which will be discussed in later subsections. The work presented in this chapter is a part of the manuscript published in IEEE Journal of Biomedical Health and Informatics (1 of section 1.5) [131].

5.1 BACKGROUND

The cardiovascular disease growth rate has been increasing faster worldwide in recent years [132]. Therefore, HR is a vital physiological parameter. It reflects the physiological, physical, and emotional state of an individual. The significance and estimation approaches for HR estimation have already been presented in section 2.1. Additionally, this chapter uses rPPG due to its advantages presented in section 2.4 of chapter 2; therefore, the proposed method is based on the steps of HR estimation, as explained in section 2.4. Specifically, HR estimation using the rPPG method is a three-step process: ROI selection, BVP signal extraction, and the average HR calculation. Adhering to the common approach and associated advantages, the ROI selection includes using the facial region as ROI [18]. Capturing the facial region is predominantly done using an RGB camera because it allows less constrained conditions, unlike other methods such as NIR [16], radar, or ultrasound systems [58].

As mentioned earlier in section 2.3.2, the rPPG-based methods use reflected light acquired through a photodetector after light absorption by the skin tissues, arteries, veins, bones, and blood [24,133]. This reflected light contains blood volume variations along with various undesirable noise interferences [16] due to rigid and non-rigid motions [69] and illumination variations [66], which degrade the performance of HR estimation methods due to noisy BVP signals [121]. Furthermore, the noise due to these artifacts easily dominates the relatively weaker strength of the resultant BVP signal [66]. A few frequently used BVP extraction methods used in the literature are Wavelet transforms [113], ICA [4], and EEMD [67]. Wavelet transform requires the selection of appropriate filtering coefficients at different decomposition levels [87], whereas EEMD requires selecting amplitude and noise frequency [134].

However, ICA begins with a random initialization of unmixing matrix with just a single prerequisite of unmixing matrix dimensions, depending on the number of independent components, which is comparatively trivial compared to the other two methods. Additionally, as per the systematic analysis presented in Chapter 4, ICA is a common method for BVP signal extraction [107]. It considers BVP extraction as a BSS problem, which deals with extracting the desired signal with no or limited aprori information. Moreover, Joint Diagonalization Approximation of Matrices (JADE), which is a variant of ICA proposed by Poh et al. [117], has shown motion tolerance up to a certain extent. In addition, for the Multichannel ICA proposed by Zhang et al. [122] the experiments were conducted under low illumination as well. To the best of the author's knowledge, none of the ICA method-based studies analyzed the

impact of motion and illumination variations effect simultaneously under constrained or natural conditions.

A general assumption regarding ICA-based methods is that the number of independent signals is equal to the number of mixed signals. This assumption requires analyzing each IC as a potential candidate for the BVP signal while also requiring apriori knowledge about the BVP signal. Moreover, there is no defined criterion for selecting the BVP signal from the independent components from different color channels [135]. Conventionally, BVP signal extraction includes selecting the component with the highest periodicity, which may result in choosing the incorrect IC as a BVP signal in the case of periodic motions by the subjects [135]. Most studies selected the second IC for BVP signal extraction by discarding the 1st and 3rd IC [4, 117, 121]. This results in a loss of information [25], which may be vital for HR estimation. Color subspace transformation methods like CHROM [69] and POS [63] were proposed to overcome this information loss, which employed orthonormal vector transformations to construct a raw signal for BVP extraction. The main drawback of these methods is the fixed weights assigned to color channels, which may degrade the BVP information [63].

Considering the limitations mentioned above, this study proposed the BVP signal extraction as an undercomplete problem. In other words, given three mixture signals corresponding to RGB color channels, the task is to extract one IC that corresponds to the motion and illumination-resistant BVP signal. A novel method combining U-ICA [136] with a customized LMA [137, 138] was proposed for optimizing the unmixing matrix for BVP signal extraction without losing information from any color channel. The method is named U-LMA based on the composition of its modules i.e., U-ICA, and customized LMA. Additionally, the proposed method eradicates the need for IC selection since the output is a single IC. The mean HR calculation was performed using power spectral density analysis by using Fast Fourier Transform (FFT), post bandpass filtering.

5.2 OBJECTIVES

The U-LMA method proposed in this chapter aims to accomplish the following objectives:

1. Develop a novel non-linear optimization function constituting a CDF approximated by the hyperbolic tangent (tanh) to deal with the non-linearity due to rigid and non-rigid motions and illumination variation artifacts for BVP signal extraction.

2. Customize LMA for optimizing the entropy of the proposed non-linear least square function, ensuring the statistical independence of the resultant BVP signal.
3. Develop a novel method constituting U-ICA with customized LMA (U-LMA) for an artifacts-free BVP signal extraction, followed by its performance analysis under three scenarios (database used): Constrained (VIPL-HR [94]), motion constituting rigid and non-rigid motions (UBFC-rPPG [95]), and illumination variations (COHFACE [96]).
4. Test the performance of U-LMA with negentropy-based U-ICA (Undercomplete Independent Component Analysis with Negentropy cost function (U-neg)) and other SOTA methods to analyze the impact of a non-linear function along with optimization using LMA under the abovementioned scenarios.

5.3 RELATED WORK

The first attempt of HR estimation under normal light conditions was performed by Verkruysse et al. [29]. The study used the rPPG signal extracted using the green color channel of the ROI selected from face videos acquired by a digital camera. The PPG signal was then processed using filtering techniques and, subsequently, HR calculation. Poh et al. [4] extracted the rPPG signal using JADE (ICA) from the RGB signal traces acquired from the facial ROI captured using a webcam. Consequently, three ICs were extracted from each color channel, followed by selecting an appropriate IC as a rPPG signal for HR estimation. This study was further extended by adding a temporal filtering component, which consisted of de-trending and signal smoothing using a moving average filter for better rPPG signal extraction [117]. The above methods mainly used the green component of the ROI since it is considered to have maximum PPG information. The method given by Poh et al. [4] used kurtosis optimization, which does not have descent statistical properties to support statistical independence among components.

Gill et al. [60] addressed the problem of unsorted ICs of ICA, which proves to be challenging when selecting the appropriate independent component as a BVP signal. They proposed constrained ICA, which uses negentropy as an optimization function, avoiding local minima convergence. It is important to note that negentropy possesses better statistical properties and symmetric decorrelation than kurtosis to ensure statistical independence. Considering the periodicity of the PPG signal, Macwan et al. [61] proposed Multi-objective optimization using Autocorrelation and ICA (MAICA), which constitutes negentropy and signal autocorrelation at different

time lags, for BVP signal extraction. A Kalman filter was also utilized to address motion and illumination artifacts.

A different approach was presented by De Haan et al. [69], utilizing the chrominance features of RGB spectra. The method extracted the two chrominance vectors, orthogonal to each other, from the RGB color spectra. RGB to chrominance vector transformations were performed by empirically known coefficients. Finally, the ratio for the two vectors was used for rPPG signal extraction. Furthermore, De Haan et al. [128] further improved this method by employing the absorption spectra changes of the RGB spectra for BVP signal extraction, where the Hulsbusch noise-free spectrum model was used to develop a normalized BVP vector. Combining chrominance-based signals and ICA advantages, Song et al. [67] introduced a semi-blind source separation method named Kernal ICA, which is based on Kernel Density Independent Component Analysis (KDICA) proposed by Aiyou Chen [139]. Kernal ICA takes chrominance signals as input and extracts the rPPG signal. The kernel ICA was used to address the problem of similar magnitude among illumination variation and PPG signal. In addition, the authors have also tested the effect of different shooting distances and image resolution for rPPG signal extraction.

Realizing the need to add more channels for accurate BVP extraction, McDuff et al. [140] used a five-band lens camera to extract the orange and cyan spectra along with the three traditional color spectra. This enables monitoring the absorption of light differences between deoxygenated Hb and HbO2 by creating a bigger overlap between cyan, orange, and green spectra for accurate HR estimation, using the approach presented by Poh et al. [117]. A similar approach was proposed by Gupta et al. [115] in which a magenta color filter and thermal camera filters were utilized with an RGB camera to overcome the illumination variations effect on HR estimations. Furthermore, they concluded that the red and green channels with thermal imaging can better estimate HR. A computationally faster variant of ICA, FastICA, was used for BVP signal separation, which uses negentropy to maximize statistical independence. Alternatively, Yan et al. [141] proposed an approach of using a weighted average of RGB spectra of the selected ROI for improving the SNR, followed by denoising the signal using Wavelet transform for rPPG signal extraction. Kumar et al. [116] used a monochrome camera for extracting the green spectra followed by its weighted average using varied ROI combinations.

Pursche et al. [142] analyzed the effect of using different facial regions on HR estimation and divided the facial region into three ROIs: forehead, the area surrounding eyes and nose, and the mouth area. The HR was computed using each

ROI, applying ICA for rPPG signal extraction, followed by Fourier transform. The summary of this literature review is presented in Table 18.

The limitations of the existing SOTA methods are multifold. First, selecting the appropriate IC containing BVP information is challenging due to the unordering of independent components. Second, existing statistical dependence metrics do not consider the non-linearity associated with the PPG extraction problem. Third, adding further channels for HR estimation enhances the complexity of the problem by increasing its dimensionality with an added effect due to different types of motions and illumination variations. Fourth, color difference equations proposed in color subspace transformation methods have associated coefficients with the color channels, which may affect PPG information. Finally, the semi-blind source separation may need additional information about PPG signal statistical properties for accurate signal extraction.

The problem of unordered independent components is resolved by assuming the BVP extraction problem as undercomplete, which deals with taking raw RGB traces as input and a single BVP signal as output. To ensure the consideration of non-linearity with statistical independence, a non-linear optimization function (CDF approximated by tanh) is proposed. The presented method can deal with the associated non-linearity due to artifacts with three channels of the RGB color space. The proposed method does not assign the weights to color channels, ensuring that each color channel contributes to the BVP signal independently. Finally, U-LMA does not require apriori information to extract the BVP signal from raw signals. Hence, the proposed method manages to overcome all the limitations pointed out by the existing literature.

5.4 METHOD

The proposed method takes a face video recorded under ambient light conditions as input and estimates the mean HR. It calculates the HR via a three-step procedure: ROI selection, BVP signal extraction, and HR Estimation. This section discusses the detailed information of the proposed U-LMA for HR estimation, explaining all the constituting steps in the following subsections. A detailed flow diagram for the proposed approach is shown in Figure 24.

Table 18: Summary of existing SOTA HR estimation studies.

First Author	Physilogical Paramter	ROI Used	BVP Method	Color channel	Limitations
W Verkruysse [29]	HR, RR	face	Filtering	Green	This study was the first attempt and did not address any artifacts. The method possessed a poor signal-to-noise ratio.
M Poh [4]	HR	Face	ICA	RGB	The method did not work well under rigid movements and different illumination conditions.
G Tsouri [60]	PR	Face	ICA	RGB	The constrained ICA is 30 times slower than the ICA.
G Haan [69]	HR	Face	CHROM	RGB	CHROM method uses skin standardization and fixed projection planes, which halts its generalizability.
R Macwan [61]	HR	Face	ICA	RGB	The proposed method uses periodicity as one of the criteria for BVP selection, which limits its applicability for estimation during periodic movements.
R Song [67]	HR	Cheeks	KDICA	RGB	This study examined the influence of resolution and shooting distances; therefore, the limitations were not presented.
O Gupta [115]	HR,HRV	Cheek, Forehead	Fast ICA	RGB, Magenta thermal	The study did not consider motion artifacts.
M Kumar [116]	PR, PRV	Face	MRD	RGB,Mon	The method tracks ROI using the KLT algorithm. For rigid motion, features cannot be tracked, leading to PPG information loss.

G Haan [128]	HR	face	CHROM	RGB	The method did not work well with the stationary subjects due to unavoidable noise since the noise deviates from the BVP vector and degrades HR estimations.
D McDuff [140]	HR, BR, and HRV	face	ICA	RGB, cyan, and orange	The customized camera setup is not always feasible in real-time. Furthermore, rigid movements were not addressed in this work.
M Poh [117]	HR, RR	face	ICA	RGB	The study did not address illumination variation artifacts and rigid motions.
Y Yan [141]	HR	forehead	Wavelet Transform	RGB	The method could not work well under illumination variations.
T Pursche [142]	HR, RR, and HRV	Forehead, the area around the eyes and mouth	ICA	RGB	The study used constrained conditions for estimations and did not address any artifacts.

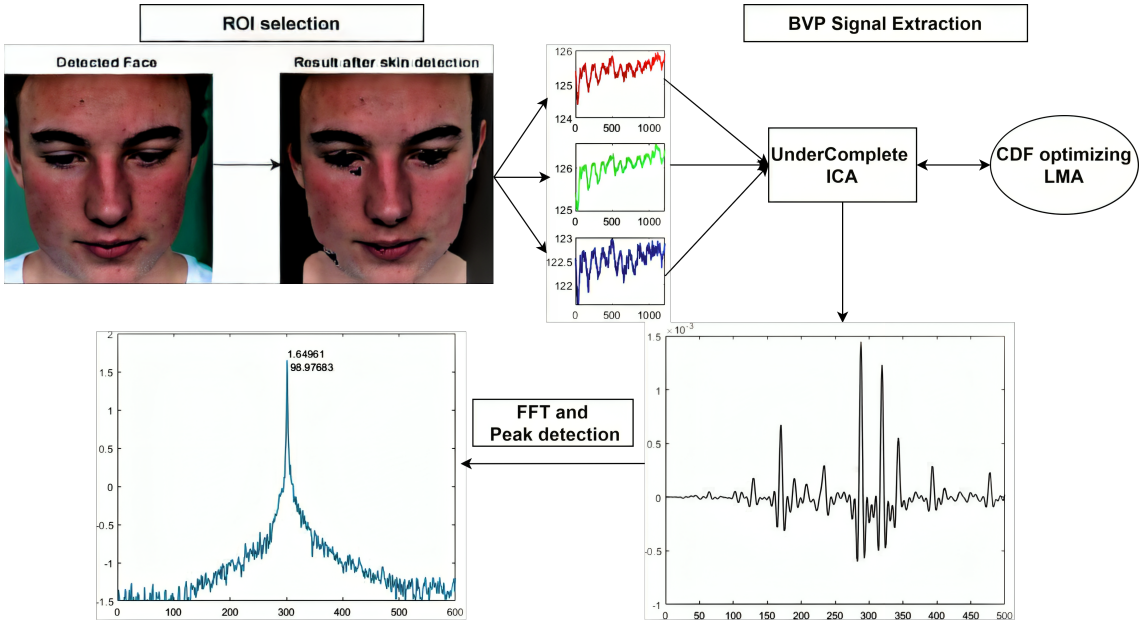


Figure. 24: The workflow of the proposed method for HR estimation.

5.4.1 ROI Selection and Signal Construction

The ROI selection involves identifying the face using the Viola-Jones face detector [30], followed by skin segmentation. The skin was segmented using the Cb and Cr components of the YCbCr color model using the parameters proposed by Mahmoud [41]. Subsequently, a spatial averaging for each channel was performed on each video image frame. A detrending process was also applied to remove slow, non-stationary drifts in the signal using the approach by Tarvainen et al. [143]. Finally, an overlapping moving window operation was applied to each channel for constructing raw signals. Figure 25 presents the steps of ROI selection and RGB raw signal trace construction.

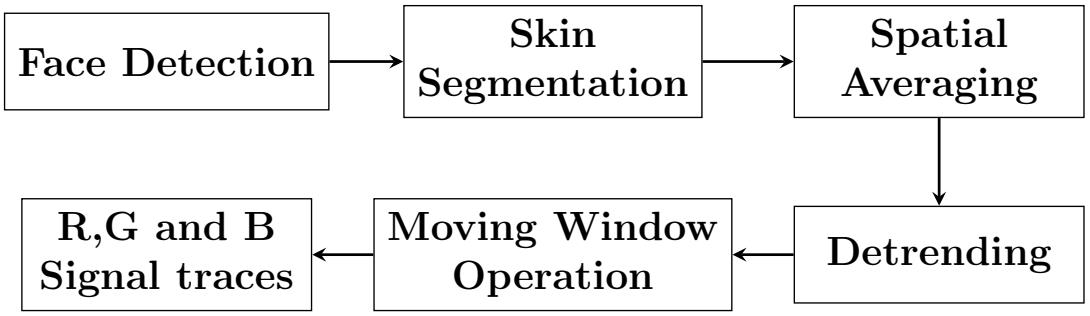


Figure. 25: ROI and raw signal construction.

5.4.2 BVP Signal Extraction

The raw signals were refined to extract the BVP signal for HR estimation. Following the standard ICA annotations, the raw signals are considered mixture signals containing BVP and other information, along with noise interferences due to motion and illumination artifacts. The goal is to extract the BVP signal as one of the ICs from them [144]. Ideally, the 2^{nd} IC is selected as a BVP signal, while other ICs are discarded, which may contain BVP information. Therefore, this study defines a BVP extraction as an undercomplete problem that takes three mixture signals and extracts a single IC consisting of BVP information from all three channels [98].

This problem is solved using the proposed U-LMA, which uses a CDF of the raw signals approximated by tanh, followed by its optimization using the customized LMA proposed in this study. The proposed approach is motivated by the work of Porrill et al. [136] for signal separation and dimensionality reduction. The difference lies in the context, optimization algorithm, and termination condition. The present work uses a customized version of LMA for optimizing the unmixing matrix W and the number of iterations as the only termination condition. The reason behind choosing the number of iterations as a termination condition is to consider the absence of BVP signal information. Since there is no reference BVP signal, it becomes impossible to check the correlation of the resultant signal with the original BVP signal, so the correlation criterion is not considered. LMA is chosen because the initial values of W due to random initialization may or may not lie near the desired solution. Both conditions need separate ways of approaching the solution. This optimization algorithm will allow the solution to efficiently converge to the desired values of W in both conditions [145]. The details of the customized version of LMA will be discussed in the following subsection.

Mathematically, $x(t) \in \mathbb{R}^{(3\times t)}$ and $y(t) \in \mathbb{R}^{(3\times t)}$ are mixed signals and IC matrix, respectively. $x(t)$ consists of three mixed signals $x_1(t)$, $x_2(t)$, and $x_3(t)$, corresponding to the color channels, whereas $y(t)$ comprises 3 ICs $y_1(t)$, $y_2(t)$, and $y_3(t)$ corresponding to three mixed signals. A standard ICA model assumes that mixed signals are linear combinations of ICs:

$$x(t) = Ay(t) \tag{14}$$

where A is the mixing matrix, which, when multiplied by ICs (signals), leads to mixed signals $x(t)$. Unfortunately, the mixing matrix and independent components are unknown; therefore, the ICs can only be extracted based on their statistical properties, as mentioned before. Furthermore, the goal is to estimate unmixing matrix

W, which will be used for IC extraction as follows:

$$y(t) = Wx(t) \tag{15}$$

From equations (14) and (15), it can be concluded that the matrix W can only be an approximation of A^{-1} for accurate IC extraction. Unlike the standard ICA, where W is a square matrix, for U-LMA, W will be a rectangular matrix since the number of ICs is less than the number of mixed signals. As the CDF of the statistically independent signals has maximum entropy [136], W can be determined by maximizing the entropy of the CDF, ensuring the statistical independence of ICs [146]. The entropy $H(y)$ of CDF for the BVP signal y is mathematically defined as:

$$H(y) = H(x) + E[\log \sigma'] \tag{16}$$

Where $H(y)$ and $H(x)$ define the CDF's entropy of the IC and multidimensional Gaussian mixed signals, respectively. Furthermore, σ' represents the derivative of CDF σ of the only statistically ICs (BVP signal). It is important to note that in equation (23), $x(t)$ and $y(t)$ are written as x and y for brevity. It is challenging to calculate $H(x)$; therefore, it can be approximated as:

$$H(x) = 0.5 * \log(c) + 0.5 * (1 + \log(2\pi)) \tag{17}$$

For $H(y)$ maximization, $H(x)$ can be reduced to $0.5 \times log(c)$ where c is written as $E[xx^T] = WSW^T$ and S is a diagonal matrix containing the covariance values of x. Considering the reduced form of $H(x)$ and approximating $\sigma = \tanh$, a new function can be deduced as:

$$h(W) = 0.5 * log|WSW^T| + E[\log(\sec^2(y))] \tag{18}$$

Equation (18) is used as a criterion for extracting the IC from the mixed signals. Differentiating equation (18) $w.r.t\ W_{ij}$ yields the ∇ vector for updating the unmixing matrix W, given by:

$$\nabla_{(w^h)} = \frac{SW^T}{WSW^T}^T - 2E[y^T x] \tag{19}$$

where $\frac{SW^T}{WSW^T}$ is the pseudo inverse of W $w.r.t$ S, a positive definite matrix. The proposed algorithm approximates the unmixing matrix W using the LMA by maximizing equation (18) and updating the matrix W using equation (19).

5.4.3 Customized Levenberg Marquardt Algorithm (LMA)

LMA is a widely used optimization algorithm to find the global minima for non-linear least-square functions with faster convergence properties [147] and dual algorithmic adaptability depending on the current solution. In other words, the LMA can be considered as a combination of gradient descent and the Gauss-newton method depending on the proximity of the current solution to the global minima.

The presented method customized the conventional LMA by introducing the entropy of CDF approximated by a tanh defined in equation (18) as an optimization function, followed by its maximization for the statistical independence of IC. The advantage of approximation using tanh lies in the fact that it introduces processing with higher-order statistics to deal with the non-linearity associated with the optimization problem [127].

The workflow of updating W using the proposed method for BVP signal extraction is presented in Figure 26. The process starts with the random initialization of W, followed by calculating the entropy of CDF and subsequently validating the convergence condition. If the convergence condition is reached, W_{curr}, the W at the current iteration is returned as an output; otherwise, the Jacobian, Hessian, and diagonal matrix is computed for updating W.

The diagonal matrix consists of the highest value of the Jacobian achieved until the last iteration performed. The cost function, i.e., entropy, is calculated before and after updating W and then compared. The damping parameter λ is increased if entropy decreases, followed by calculating the cost function again after updating W using W_{prev} from the previous iteration until the entropy is increased. If there is a rise in the entropy value after updating W, λ is decreased until the convergence condition is reached. The raw signal is multiplied by W to extract the BVP signal for HR estimation as:

$$\mathrm{BVP}(t) = W \times x(t) \tag{20}$$

5.4.4 HR estimation

The BVP signal extracted through U-LMA is processed using a bandpass filter with cut-off frequencies $0.7 - 4.0\ Hz$, respectively, corresponding to $42 - 240$ bpm. Finally, the FFT is applied for analyzing the power spectral density for maximum peak estimation, which is then used for the HR calculation by taking its log_{10} and multiplying it by 60.

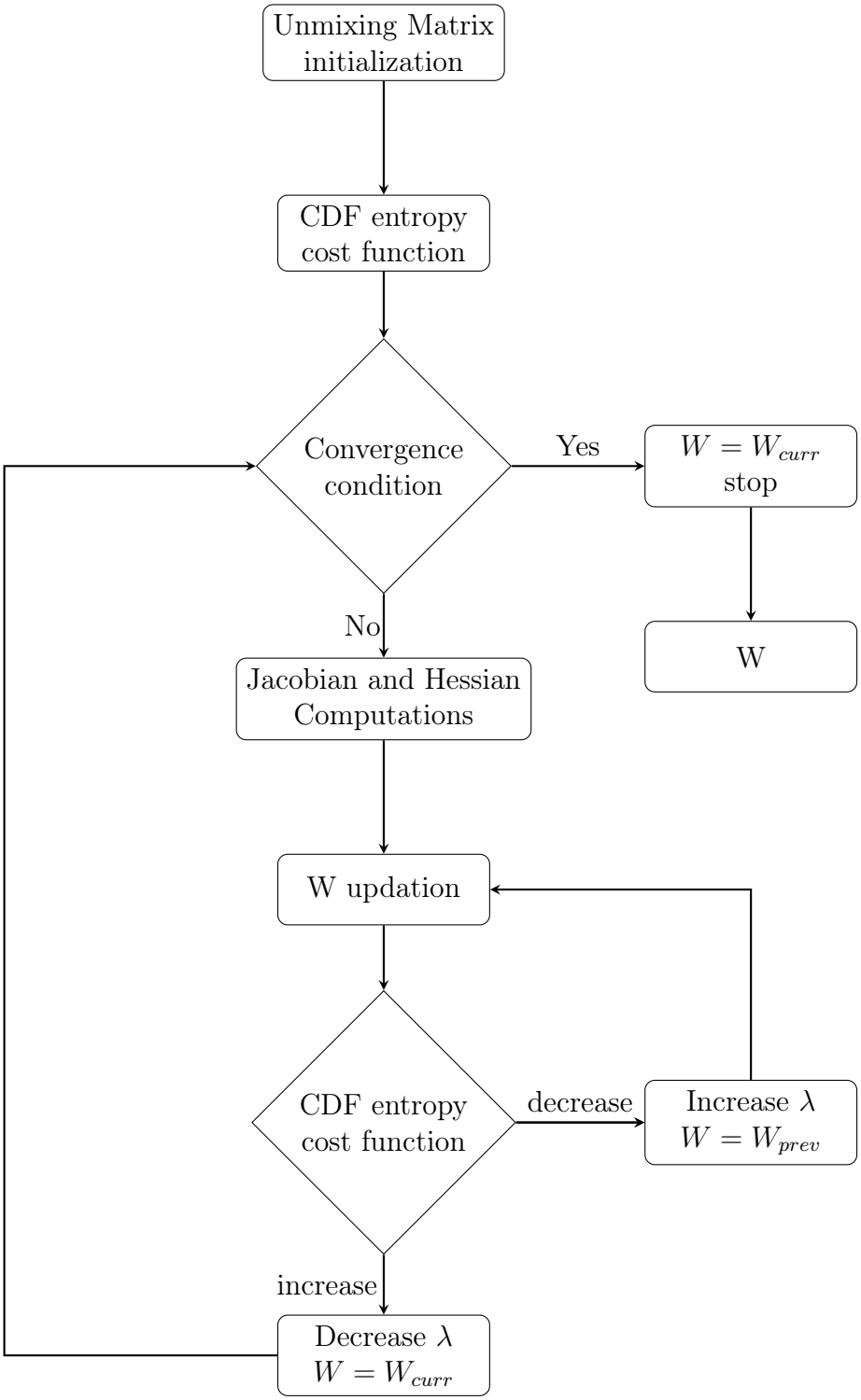


Figure. 26: Customized LMA for entropy maximization.

5.5 RESULTS

The proposed method was tested under constrained and natural conditions using three benchmark databases: VIPL-HR, UBFC-rPPG, and COHFACE. The VIPL-HR database was used for performance validation under constrained conditions,

while the UBFC-rPPG tested the method's performance for rigid and non-rigid motions, and illumination variations effect on the proposed method was tested using the COHFACE database.

It is necessary to analyze the method's performance under constrained and unconstrained conditions since testing a method under constrained conditions provides insight into its steps and their precision, whereas unconstrained conditions test its robustness. A detailed description of the databases used for this study is presented in section 3.2.2, while the performance metrics used to analyze the performance of the proposed U-LMA are also explained in section 3.4 of chapter 3.

5.5.1 ROI selection and signal construction

Before this step, the RGB image frames of the video were preprocessed to adjust the pixel intensities using gamma correction. The ROI selection deals with face detection followed by segmenting the skin in the YCbCr color space in which Y represents the luminance with pixel intensity ranges between 16 and 235, while for the chrominance blue (Cb) and chrominance red (Cr) components, the pixel values lie between 16 and 240. The thresholds used for the Cb and Cr components are between 77 to 127 and 133 to 173, respectively, with no thresholding for the luminance component [41]. Finally, the ROI is selected as 70% height and 60% width of the segmented skin region. Figure 27 depicts the results of the face detection and skin segmentation process. For detrending the temporal RGB traces, the regularization parameter was set to an empirically defined value, i.e., 10. The raw signal was constructed using a moving window operation with a 96% overlap ($1 - secincrement$) for each color channel.

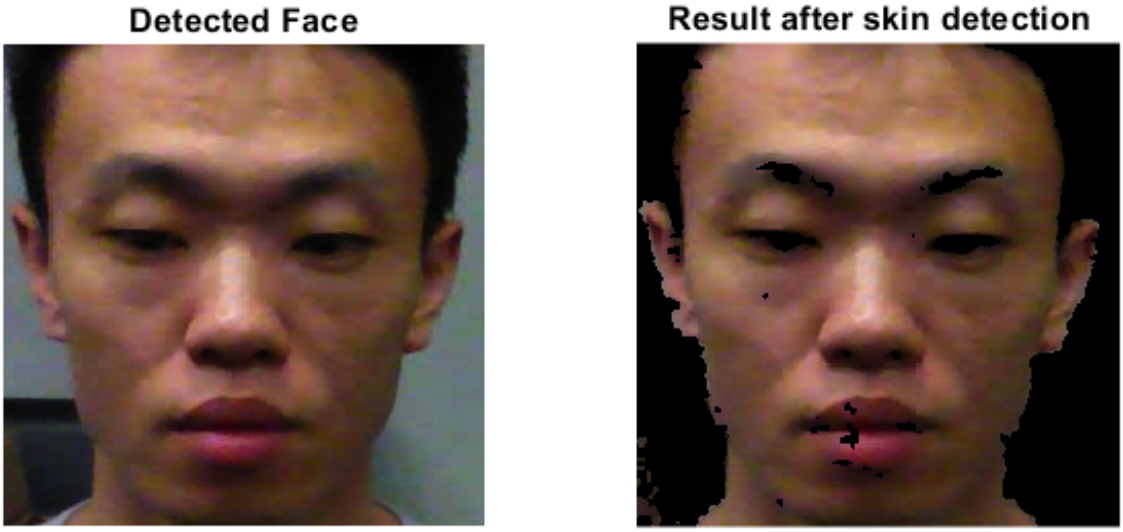


Figure. 27: Face detection and skin segmentation.

5.5.2 BVP signal extraction and HR estimation

The BVP signal extraction was performed using U-LMA. The unmixing matrix W was first initialized randomly, and the values of damping parameter λ were set empirically as 5 and 2.5, respectively, as a part of standard LMA initialization. Subsequently, the customized LMA was employed to maximize the entropy of the proposed non-linear CDF optimization function using 1000 iterative steps since none of the video samples took these many iterations for convergence to global maxima. Finally, the optimized unmixing matrix W was used to extract the BVP signal. In the last step, a FFT was applied to the resultant signal, followed by calculating the $\log_{10}$ value of peak maxima and multiplying it by 60 to obtain a mean HR estimation.

5.5.3 Performance analysis

As mentioned before, the performance of the proposed ULMA method is analyzed by considering three scenarios: constrained, rigid, and non-rigid motions and illumination variations. VIPL-HR database was used for performance testing under the constrained or stable scenario, UBFC-rPPG for testing its robustness in rigid and non-rigid motions, and COHFACE in illumination variations scenarios. For each scenario, B-A and regression plots will be presented and analyzed, taking into consideration the respective measured parameters for the plots.

a) Constrained Scenario

For the constrained (VIPL-HR database) scenario, the subjects were asked to sit in the still position at a distance of one meter away from the camera with the ceiling lamp on. The B-A and regression plots for the constrained scenario are shown in Figures 28 and 29, respectively. The mean bias for the proposed method is 0.35 bpm, which is near to a zero error difference between the ground truth and the estimated values. In other words, on average, the HR estimated by the algorithm measures 0.35 bpm less than the conventional BVP sensor used. Statistically,for mean bias, the 95% confidence intervals lie between -0.2712 to 0.9682. Additionally, for upper and lower statistical limits, 95% confidence intervals lies between -6.99 to -9.14 and 6.29 to 8.44, respectively. Since 95% intervals of mean bias, lower and upper statistical limits lie within 3 bpm, which is lower than the error standard deviation. Additionally, the Pearson correlation value denoted by r for this scenario is 0.92, thus

confirming a higher correlation between the ground truth and the estimated values. Therefore, B-A analysis and high correlation value justify the superior performance of the proposed method under the constrained scenario.

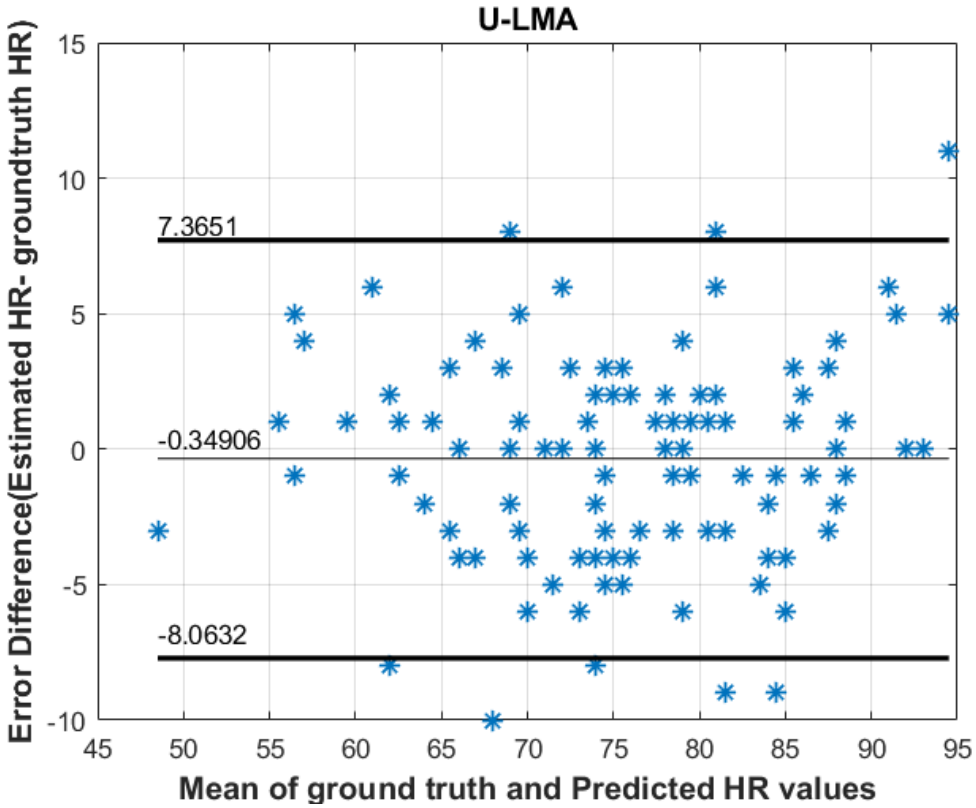


Figure. 28: Bland-Altman plot for the constrained scenario.

b) Rigid and Non-Rigid Motions Scenario

A performance analysis under this scenario was performed using all video samples of the UBFC-rPPG database. The videos were collected while subjects were playing a time-sensitive mathematical game, which causes an abrupt increase or decrease in HR values along with involuntary head movements due to the subject's action. The samples also have a certain amount of illumination variations since the video samples were collected considering natural conditions. Figures 30 and 31 depict the plots of B-A and the regression, respectively. As expected, the mean bias for this scenario is 1.84 bpm due to the presence of motion artifacts, which means the U-LMA predicts 1.84 bpm more than ground truth HR value. Statistically,for mean bias, the 95% confidence intervals lie between 0.84 to 2.84. Additionally, for upper and lower statistical limits, 95% confidence intervals lies between 8.55 to 12.01 and -8.33 to -4.87, respectively. Since 95% intervals of mean bias, lower and upper statistical limits lie within 4 bpm, which is slightly lower than the standard deviation error.

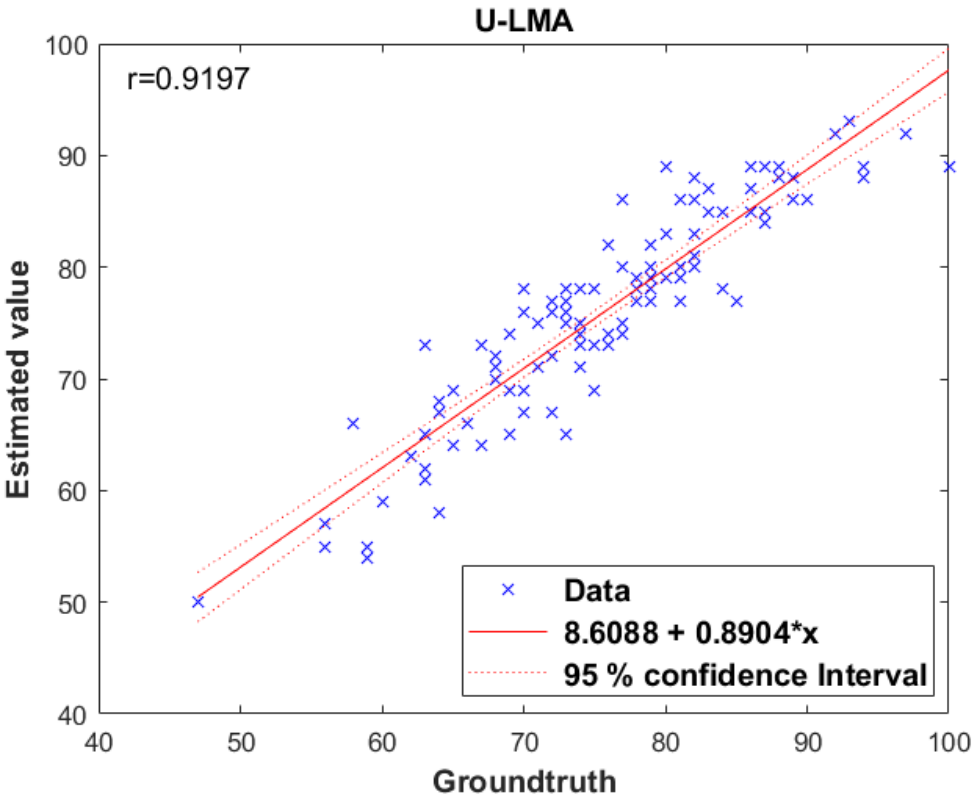


Figure. 29: Regression plot for the constrained scenario.

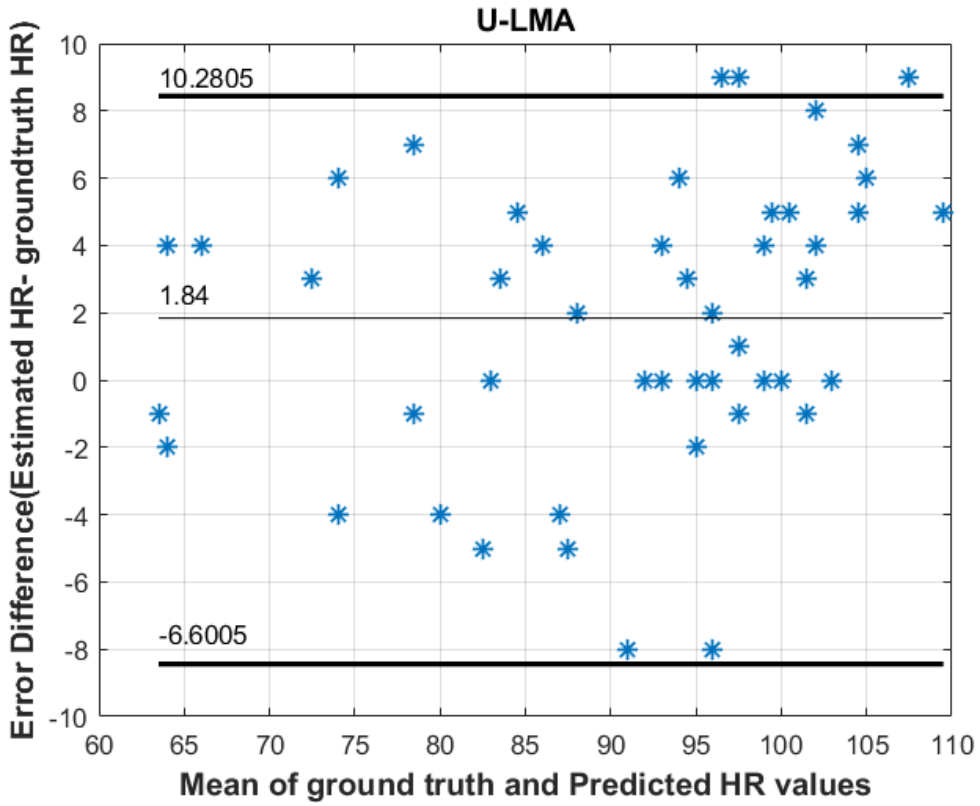


Figure. 30: Bland-Altman plot for rigid and non-rigid motion scenario.

Furthermore, the ground truth and estimated HR values demonstrated a very high correlation (0.94) despite having a higher overall mean difference. Hence, the B-A

analysis and the regression plot confirm the proposed method's effectiveness under challenging motion conditions while also handling the abrupt rise and fall of HR values when considering the mean values during the interval.

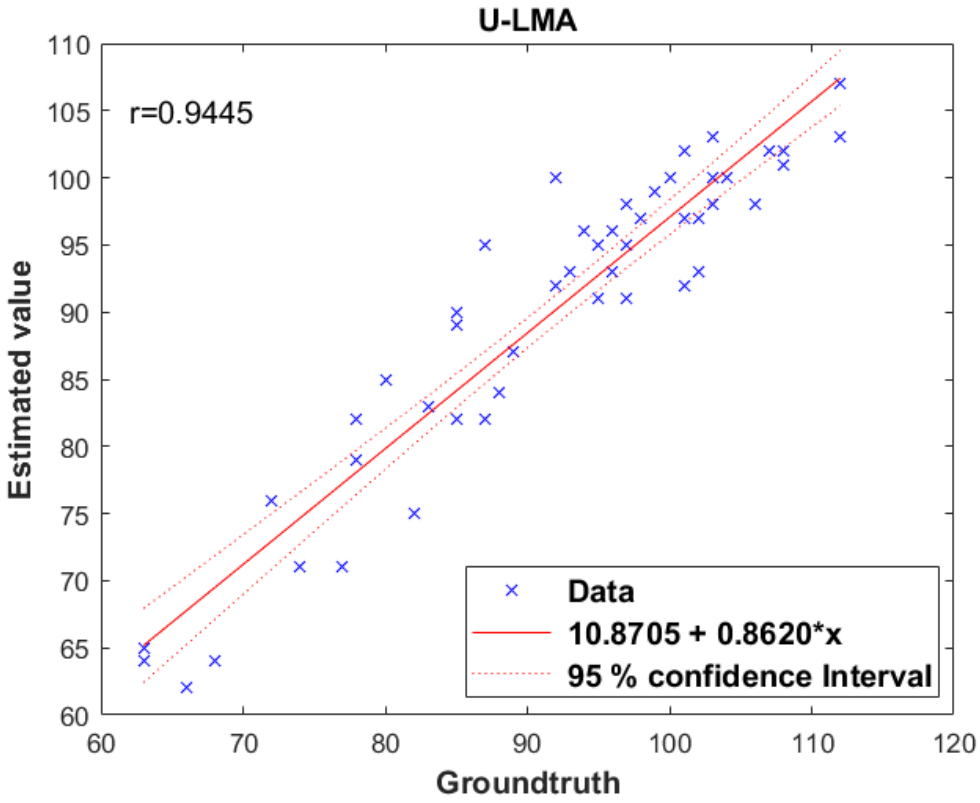


Figure. 31: Regression plot for rigid and non-rigid motion scenario

c) Illumination Variations Scenario

The COHFACE database is utilized to assess the ability of the U-LMA under different illumination scenarios. It is worth noting that the samples for the database possess motion artifacts, too, but with the predominance of uneven illumination distribution over the face due to ambient light. The performance analysis using the B-A and regression plots are presented in Figures 32 and 33, respectively. The mean bias achieved with the illumination scenario is -0.85, with lower and upper statistical limits of -9.7546 and 8.0546, respectively. Statistically,for mean bias, the 95% confidence intervals lie between -2.33 to 0.33 bpm. Additionally, for upper and lower statistical limits, 95% confidence intervals lies between 6.00 to 10.10 bpm and -11.81 to -7.70 bpm, respectively. Since 95% intervals of mean bias, lower and upper statistical limits lie closer to 4 bpm minute, which is slightly lower than the standard deviation error. Additionally, the Pearson correlation at the significance level of 0.01, achieved under this scenario, is 0.92 between ground truth and estimated HR

estimation values. Both plots and their measured parameters proved the efficiency of the U-LMA for the HR estimation using illumination variant facial videos.

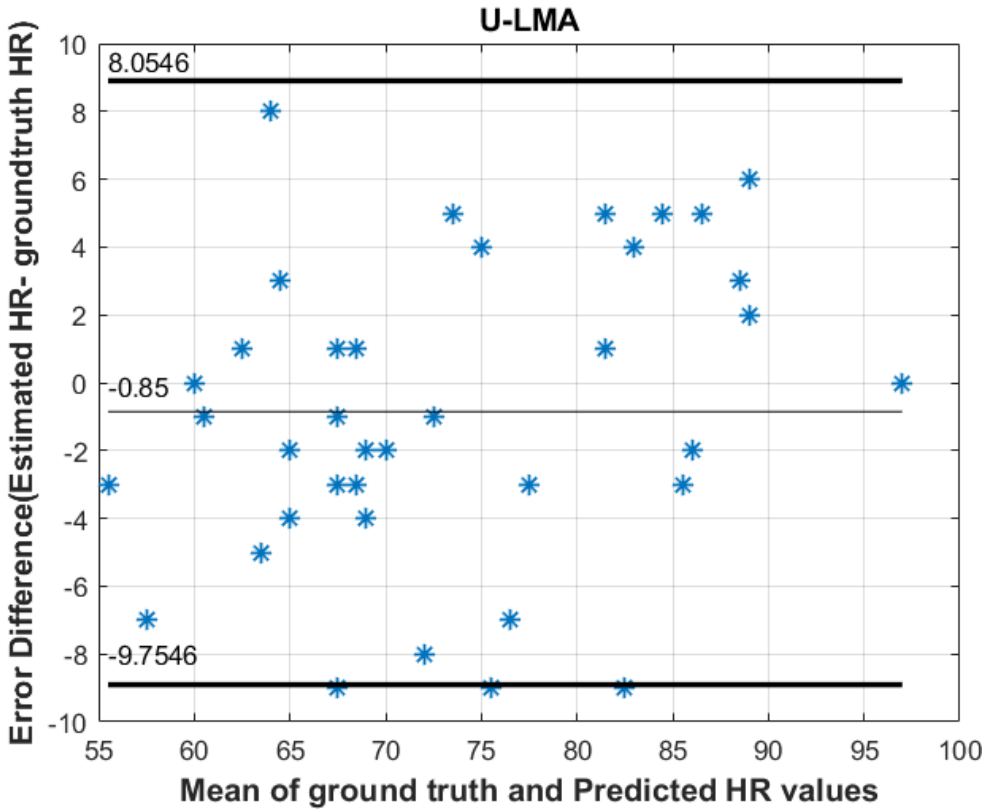


Figure. 32: Bland-Altman plot for illumination variation scenario.

5.5.4 Comparative analysis

The available related conventional rPPG methods in the literature are based on a single-color channel selection, ICA, color subspace transformations, and Wavelet-based methods. The performance of U-LMA was compared to all of these, except for Wavelet-based methods, since these methods use the time-frequency domain and empirically set coefficients, unlike other PPG methods included in the study.

The single-color channel selection method uses a filtered signal extracted from a single-color channel. Therefore, GREEN proposed by Verkruysse et al. [29] was included, which extracts the BVP signal from the green color channel of the RGB color space. Most ICA-based rPPG methods use ICA-Poh (JADE) [4, 61, 117, 140] and FastICA [16, 115, 148, 149]; hence they were included in this analysis.

JADE uses kurtosis, whereas FastICA uses a negentropy-based optimization function for an unmixing matrix estimation. The analysis also included two color subspace transformations, CHROM [69] and POS [63], due to their dependence on

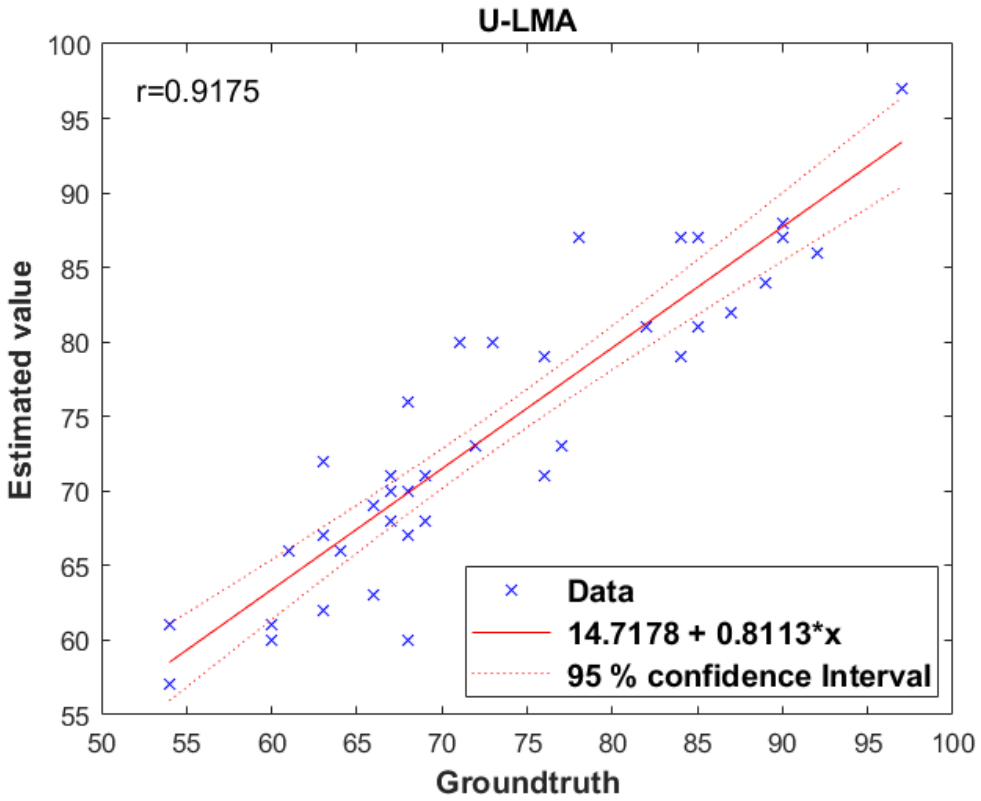


Figure. 33: Regression plot for illumination variation scenario.

optimization procedures like ICA-based methods. CHROM is a motion-intolerant algorithm, while POS performs better for uneven illumination variations.

The Ballistocardiography (BCG) method [150] was also included since it works on tracking the periodic movements of the head. As the effect of both types of motion on the proposed method is also tested in this study, it is worth including BCG as one of the SOTA methods for performance comparison. Finally, the method presented by Song et al. [67], a combination of the color subspace transformation method and KDICA proposed by Aiyou Chen [139], was also used to test the performance of the proposed method. The KDICA uses a Laplacian kernel for kernel density estimations, requiring the pulse and artifacts spectrum to be in antiphase.

Furthermore, to assess the performance of the proposed nonlinear cumulative density function, which is approximated by tanh and a customized LMA, another variant of the proposed U-ICA (U-neg) is also introduced, which utilizes the differential entropy or negentropy as an objective function. This objective function is optimized using the standard ICA procedure given by Hyvärinen et al. [144].

GREEN, ICA-Poh, CHROM, POS, and BCG were implemented using the standard implementation included in the iPhys toolbox by Mc Duff et al. [151]. Furthermore, the Matrix Laboratory (MATLAB) implemented versions of FastICA by Hyvärinen et al. [97] and KDICA by Aiyou Chen [139] were used to simulate the respective HR estimation methods as mentioned above, keeping other steps (ROI

selection, bandpass filtering, and FFT identical to the proposed U-LMA. All methods were tested under three scenarios, as explained in previous subsections. In other words, the video samples from all three databases were tested for all the methods used for comparative analysis by calculating RMSE, MAPE, ME, SD, accuracy, and Pearson correlation values under 0.01 significance level (α).

a) Constrained Scenario

For the constrained scenario, the videos from the VIPL-HR database were used for comparative analysis. The performance metrics for the comparative analysis are presented in Table 19. Among all methods, BCG was the worst performing method for this scenario despite minimal motion and illumination variation artifacts. BCG is susceptible to perform poorly in the presence of involuntary head movements, which may lead to false identification of face tracking points for estimation [150].

The CHROM, GREEN, ICA-Poh, and CHROM methods performed almost similarly. The poor performance of GREEN becomes apparent due to inappropriate method formulations and performance validations [4]. Furthermore, the original study suggested that along with green, the red and blue channels also contain complimentary PPG information [152], which is confirmed in this study. The poor performance of ICA-Poh is due to a lower frame rate than the ground truth sensor value since this leads to an inappropriate mapping of BVP peaks, which in turn leads to inaccurate interbeat intervals for HR calculation [117]. CHROM's performance depends significantly on its alpha-tuning procedure, which works better for different magnitudes of specular distortions and pulse signals. The noise due to involuntary movements is inevitable, which might have degraded the performance [67].

Like CHROM, the POS method's accuracy depends significantly on its alpha tuning procedure, which is suboptimal in the case of similar specular and pulse components' magnitude. In this case, the specular variation components projected on the two axes may not be in absolute antiphase due to complex noise distribution, which leads to false estimations of α parameter and, consequently, a poor BVP signal extraction. U-neg has performed relatively better than the SOTA methods mentioned above due to adequate information gathering from RGB color channels. However, it failed to suppress the effect of inevitable noise.

On the other hand, Kernel ICA and FastICA have performed considerably better than other methods and U-neg. However, the performance of Kernel ICA suffered due to the same reason as CHROM. However, FastICA performed better than other methods except for U-LMA, which once again proved the negentropy-based opti-

Table 19: Performance metrics for the methods under constrained scenario.

Methods	**RMSE (bpm)**	**MAPE (%)**	**SD* (bpm)**	**Mean (bpm)**	**Accuracy**	**r***
Green	21.48	24.74	12.32	17.63	1.89	0.23
ICA - Poh	19.15	22.37	11.70	15.21	2.83	0.24
CHROM	16.16	19.19	15.67	4.20	9.43	0.22
POS	18.80	22.30	11.94	14.56	0.94	0.31
BCG	24.87	27.79	12.82	21.35	7.55	-0.04
Kernel ICA	13.15	14.93	11.20	6.97	18.69	0.51
FastICA	13.10	15.37	10.38	8.06	19.62	0.60
U-neg	17.03	17.21	13.84	10.01	23.58	0.47
U-LMA	3.85	4.07	3.86	-0.35	84.91	0.92

Note: r: Pearson correlation is calculated at 0.01 significance level; Accuracy is defined as the percentage of achieving the error difference with* $\pm 5 bpm$.

mization function's effectiveness by ensuring statistical independence among independent components.

The proposed U-LMA achieved the best results, justifying its performance due to its ability to use higher-order statistics for processing non-linear signals and effective optimization procedures using LMA. Moreover, the highest accuracy with clinically accepted error difference was also achieved by U-LMA.

b) Rigid and non-rigid motions Scenario

The video samples from the UBFC-rPPG database were used to assess the effect of rigid and non-rigid motions on HR estimation. Table 20 presents the performance metrics for all the compared methods. All the methods performed well under the motion scenario due to uncompressed videos. Like the constrained scenario, BCG performed the worst for the motion scenario. BCG method's performance was suboptimal due to rigid and non-periodic head movements [150]. An improved performance of the GREEN method indicates that the method is effective for uncompressed videos and is also data-driven. ICA-Poh performed relatively well due to the accurate selection of the BVP signal since there was no loss of information from the videos. Interestingly, the statistical independence among the components suffered due to the similarity of motion and pulse spectra under the motion scenario, which led to the almost similar performance of ICA-Poh and FastICA.

Furthermore, Kernel ICA and U-neg also demonstrated a similar performance for different reasons; the former uses motion intolerant chrominance signals followed by KDICA, whereas the latter uses a negentropy-based function for the unmixing

Table 20: Performance metrics for the methods under rigid and non-rigid motion scenario.

Methods	**RMSE (bpm)**	**MAPE (%)**	**SD* (bpm)**	**Mean (bpm)**	**Accuracy**	**r***
Green	28.08	20.26	25.20	12.90	44	0.34
ICA-Poh	20.49	13.34	19.56	6.72	58	0.54
CHROM	14.08	9.00	13.16	-5.35	66	0.70
POS	14.27	9.37	13.51	-4.98	62	0.71
BCG	36.08	33.75	16.90	31.97	8	0.03
Kernel ICA	20.40	14.67	18.90	8.12	46	0.59
FastICA	20.36	14.92	19.63	6.10	46	0.56
U-neg	14.97	12.86	11.27	9.98	22	0.59
U-LMA	4.57	4.00	4.22	1.84	78	0.94

Note: r: Pearson correlation is calculated at 0.01 significance level; Accuracy is defined as the percentage of achieving the error difference with* $\pm 5 bpm$.

matrix estimation using U-ICA, combining PPG information from all color channels. Although the RMSE, MAPE, ME, and error SD of U-neg were reduced, the accuracy was degraded in the motion scenario, as expected. CHROM and POS performed relatively better than all the methods except U-LMA. This is due to their ability to perform well under motion scenarios due to the extraction of motion-resistant signals followed by the alpha tuning procedure. Nevertheless, the proposed U-LMA outperformed all the methods, reporting the minimum value of errors, highest accuracy, and Pearson correlation, justifying its exceptional performance and clinical relevance.

c) Illumination Variation Scenario

The effect of illumination variations on the methods used for this study was evaluated using the COHFACE database, as shown in Table 21. The GREEN method has shown a negative correlation for this scenario, indicating its susceptibility to uneven illumination distribution as it is susceptible to illumination variation artifacts due to varying light intensity distribution [29]. ICA-Poh did not perform well due to the low frame rate of the videos, as explained in the study conducted by Poh et al. [117]. The POS performance was suboptimal due to heterogeneous illumination conditions due to its assumption of independent intensity variations [63]. While CHROM and BCG performed better than these three methods in terms of accuracy. However, these methods could not perform adequately due to the susceptibility of BCG to illumination variations [150] and due to considerably larger differences between actual and estimated specular distortions in the video for CHROM [69].

Furthermore, the other ICA-based methods, Kernel ICA and FastICA, performed relatively better than the abovementioned methods. However, the degraded performance of the Kernel ICA is due to the same reason as the CHROM method, along with the inability of the kernel density-based ICA when performing under a higher degree of illumination distortion. Specifically, the KDICA used a laplacian Kernel, which failed to work due to illumination and pulse spectra overlapping. FastICA performed better than Kernel ICA due to the implication of a statistically better optimization function to separate specular and PPG information.

On the other hand, U-neg achieved better results than other SOTA methods, demonstrating the significance of U-ICA and negentropy. Furthermore, U-LMA achieved the lowest error and highest accuracy and correlation values, proving its superiority for HR estimation under illumination variations scenarios. Table 21 presents the performance of these methods under illumination variations scenarios.

Table 21: Performance metrics for the methods under illumination variations scenario.

Methods	**RMSE (bpm)**	**MAPE (%)**	**SD* (bpm)**	**Mean (bpm)**	**Accuracy**	**r***
Green	22.76	26.93	20.39	10.61	22.50	-0.07
ICA-Poh	24.88	28.52	17.84	17.56	35.00	0.06
CHROM	19.49	22.48	10.54	16.47	35.00	0.25
POS	26.28	30.71	12.63	23.14	22.50	0.05
BCG	25.53	27.00	15.72	20.27	2.50	0.19
Kernel ICA	14.50	12.61	11.60	8.90	50	0.36
FastICA	15.76	14.89	10.68	11.71	45	0.41
U-neg	11.13	13.03	10.76	3.33	25	0.60
U-LMA	4.48	5.16	4.45	-0.85	80	0.92

Note: r: Pearson correlation is calculated at* 0.01 *significance level; Accuracy is defined as the percentage of achieving the error difference with ±5bpm.*

d) RMSE analysis

The RMSE for HR estimation has been predominantly analyzed in most studies conducted so far. It is calculated as the square root of the averaged squared error differences among different samples, providing the overall error distribution. Figures 34 and 35 depict the box and whisker plot of RMSE for analyzing the RMSE distribution among methods based on databases and vice-versa. These RMSE plots provide a deep insight into the performance of SOTA and proposed methods for all the databases used in the study. Figure 34 shows that the UBFC-rPPG database

was challenging for all methods used in the study due to its realistic conditions considered during video acquisition, whereas the performance with VIPL-HR is the best due to constrained conditions. The COHFACE database is also challenging in terms of illumination variations throughout the samples. All methods, except BCG, performed better under the motion scenario. Specifically, ICA-Poh, CHROM, POS, and U-neg illustrated their ability to deal with different types of motions. As mentioned before, the worst performing method is BCG, as depicted in the RMSE plots in Figure 35. BCG could not adjust to the inevitable color distortions due to involuntary motions and illumination variations, as mentioned in the original study [150]. The RMSE of the GREEN method was also very high since the study did not use any formulation for BVP signal extraction. The performance of all three ICA-based SOTA methods was similar despite different objective functions used for unmixing matrix estimations. All these methods suffered from the permutation problem, making it challenging to choose the appropriate BVP component and discard other components simultaneously. Like ICA-based methods, the color subspace transfor-

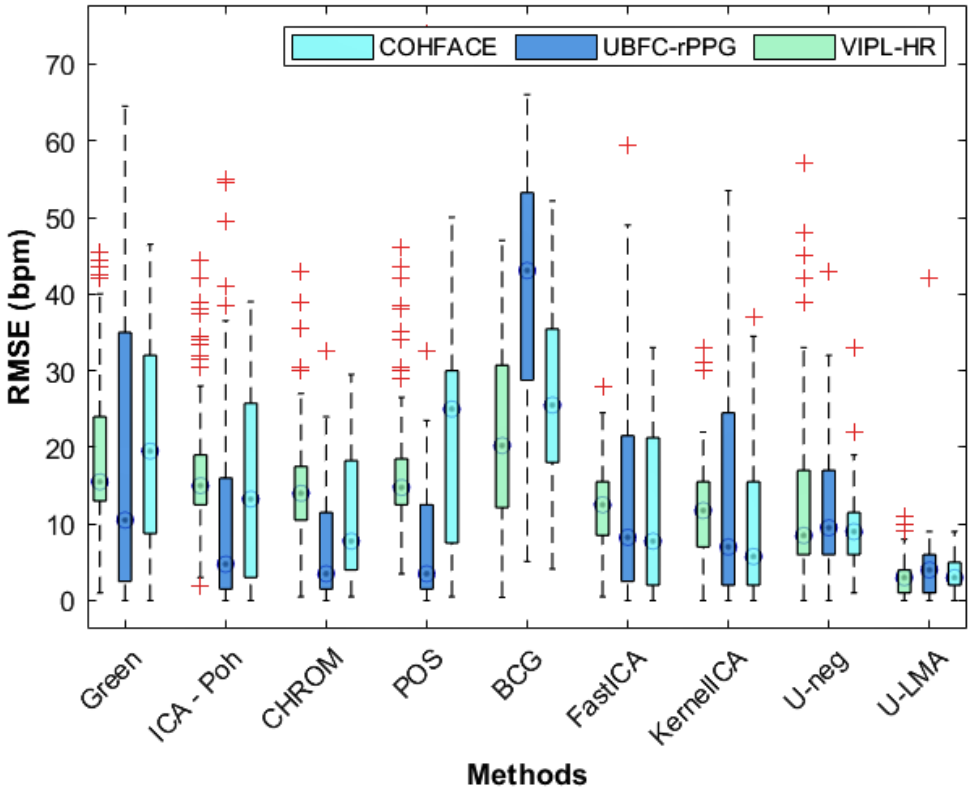


Figure. 34: RMSE Box and whisker plot for the non-contact HR estimation methods.

mation methods CHROM and POS also exhibited similar performance except for the illumination variations scenario in which the POS method failed to perform well. Since the videos were not recorded using an external light source, causing severe illumination variations on different facial regions produced an effect similar to multiple

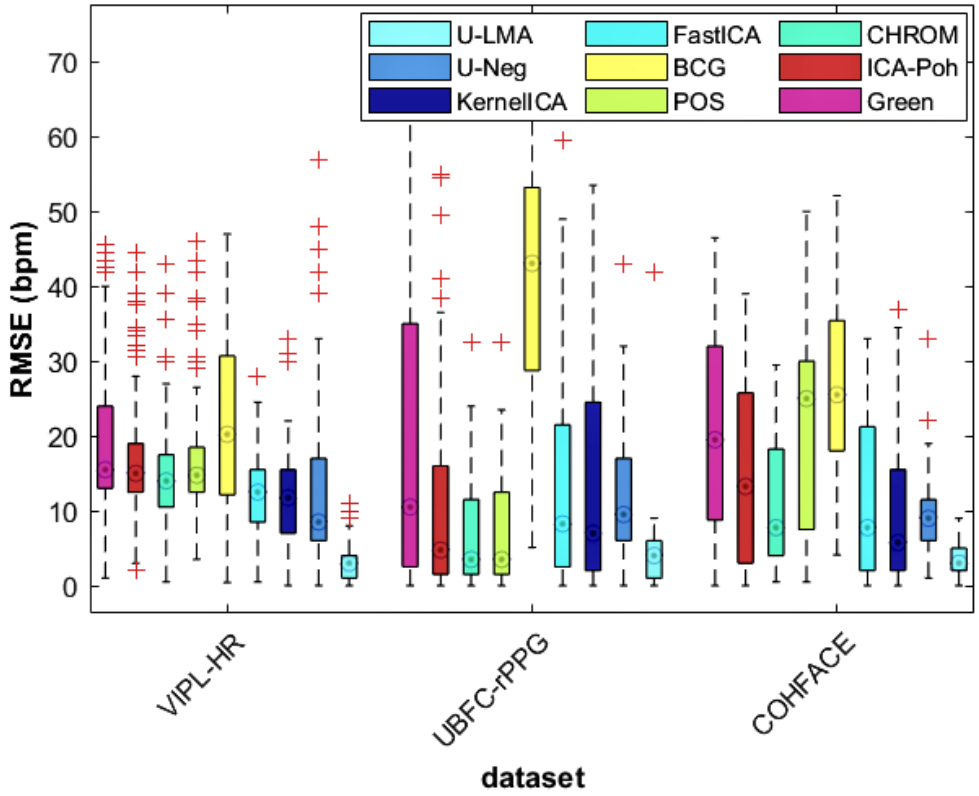


Figure. 35: RMSE Box and whisker plot of database-wise non-contact HR estimation methods.

light sources (e.g., entirely black on one side and bright on the other). U-neg performed relatively better than any SOTA method, except for the motion scenario, in which CHROM and POS performed better than U-neg. Since both methods were able to extract the BVP signal due to uncompressed videos, that ensured extracting the detailed subtle color variations, which led to an accurate BVP signal. However, the higher RMSE of the POS method for COHFACE is due to heterogeneous illumination variations, which degraded its performance [63].

Furthermore, U-LMA performed far better than U-neg and other SOTA methods, either in terms of databases or SOTA methods comparison. The proposed methods performed relatively better in all scenarios due to the proposed U-ICA, which ensures better PPG information extraction from all three channels of RGB color space. Furthermore, despite U-neg using negentropy (differential entropy) for optimizing W with a standard ICA implementation proposed by Hyvärinen et al. [144], the experiments conducted during the study revealed that the entropy of CDF approximated by tanh yielded better statistical independence than negentropy. Additionally, the lowest RMSE ranges by U-LMA when compared to U-neg in all scenarios were due to better optimization of an unmixing matrix W using the customized LMA proposed in this work.

5.6 DISCUSSION

U-LMA outperformed all other methods in the various scenarios considered for the study due to its following components: Non-linear objective function (entropy of the CDF approximated by tanh) and an efficient optimization algorithm (customized LMA). The non-linear function provided an advantage to counter the non-linearity associated with different types of motions and illumination variations. A customized LMA was able to find the global maxima for the entropy of the non-linear function in all samples with the chosen damping parameter values.

The study introduced U-neg, which included U-ICA with negentropy as an optimization function, which allowed testing the effectiveness of the U-LMA components. Kurtosis was not used because negentropy exhibits better statistical properties than kurtosis [144]. Moreover, JADE used kurtosis, whose performance was surpassed by U-neg, as shown in the previous sections. The performance comparison between both methods confirmed that processing utilizing higher-order statistics enhanced the performance of the proposed method.

On the other hand, U-ICA prevented the loss of BVP information by considering all the color channels, justifying the vitality of BVP information in red and blue channels for accurate HR estimation. Following B-A and regression analyses, the mean bias of the illumination scenario is slightly further away from a zero error difference when compared to the constrained scenario, depicting the effect of illumination artifacts on HR estimation. As expected, the mean bias for the motion scenario is comparatively larger than the illumination and constrained scenario, respectively, due to the presence of rigid and non-rigid motions of the subjects in the video samples. Furthermore, the Pearson correlation value for the UBFC-rPPG database is higher by 0.2 bpm than other databases, which could result from the utilization of uncompressed videos for HR estimation. The samples used under motion and constrained scenarios have the same sampling rate (30 FPS), whereas those tested under illumination scenarios have a relatively lower frame rate (20 FPS). It is worth mentioning that a low frame rate results in missing subtle blood volume variations, which can be captured using a higher sampling rate. The proposed U-LMA worked equally well for the videos captured with different frame rates. Hence, the loss of blood volume variation information due to compressed or low sampling rate videos did not significantly affect the proposed method for HR estimation.

Although the proposed method managed to confine most of the differences between the ground truth and the estimated HR values within the statistical limits, this did not justify its clinical relevance since none of the statistical limits matched

the clinically accepted error differences. Therefore, a new metric accuracy (error difference $\leq \pm$ 5 bpm) was defined to test the clinical relevance. The proposed U-LMA achieved sufficiently better accuracy, justifying its clinical relevance compared to other methods.

Nevertheless, U-LMA also possesses certain limitations that need to be addressed in the near future. First, although the effect of video compression and frame rates were tested, the impact of different camera-subject distances on the HR estimation was not addressed. The effect of varying shooting distances could be tested on HR estimations by creating such databases followed by analyzing them. Second, although U-LMA was tested for the rigid and non-rigid motions, scenarios with periodic motions such as walking, running, treadmill exercises, etc., or unconscious motions during sleep conditions were not addressed, which could be a future direction of the study. Third, the proposed method works well with various illumination variation conditions, but the effect of zero luminance or dark conditions was not addressed in this work due to the non-availability of reliable databases.

Most of the limitations of the proposed work are due to the limited or unavailability of sufficient benchmark databases. Finally, the proposed method only estimates HR, whereas other physiological parameter estimations like SpO2, BP, and RR using BVP signals will be worth considering in future studies.

5.7 CONCLUSION

This work addressed the BVP signal extraction as an undercomplete problem while proposing the U-LMA. Considering the non-linearity due to motion and illumination artifacts, a novel entropy-based non-linear function was proposed. The proposed non-linear function proved its effectiveness by addressing both types of artifacts. Furthermore, the non-linear function was optimized using the proposed customized LMA for entropy maximization and maximum statistical dependence due to better optimization of the unmixing matrix W. The optimization using customized LMA also aimed at reducing the effect of motion and illumination artifacts. Additionally, the proposed method eradicates the need for IC selection and preserves the maximum possible BVP information from all channels of the RGB color space. Performance analysis for U-LMA was undertaken by comparing it under three scenarios: constrained, motion (rigid and nonrigid), and illumination variations scenarios, followed by comparative analysis. The lower value of error metrics and higher correlation and accuracy value proved the proposed methods' efficacy in all scenarios.

As mentioned, the novel methods developed during this study are first tested in ambient light conditions, as presented in this chapter. Also, this chapter only focused on non-contact HR estimations, not SpO2 estimations. Adhering to this project's objectives, the next task is to modify this method in environments based on dark condition defined in chapter 6.3.4. Therefore, the next chapter presents two methods for non-contact HR and SpO2 estimations in dark environments. It is worth mentioning that U-LMA will be used for HR estimations, while SpO2 will be estimated using the popular ROR method.

Chapter 6

NON-CONTACT HR AND SPO2 IN DARK ENVIRONMENTS

Infrared-based physiological measurement methods have proven their dominance in dark environments. However, the signal extracted in these spectra exhibited poor pulsatile strength, resulting in a noise-modulated signal, hence spurious physiological estimates. Therefore, this study aims to explore the applicability of conventional RGB-based physiological variables methods in the assumed dark environment (illuminance $\leq$ 1lux). Specifically, these methods cascaded by a novel image enhancement method were analyzed to investigate the performance of state-of-the-art heart rate and oxygen saturation methods in the assumed environment. The novelty of this study is attributed to several aspects: 1) proposing a challenging illumination condition conforming to extremely dark environments such as sleep monitoring/nighttime driving, 2) proposing a robust, efficient, and unique image enhancement method complying with limitations of existing image enhancement methods, and 3) enhancing the ability of conventional RGB spectra based non-contact heart rate and oxygen saturation methods in the considered illumination condition, and 4) proving the reliability of best combinations (image enhancement method+ HR/ SpO2) as a reliable, cost-effective, and clinically viable alternative to IR-based methods.

6.1 BACKGROUND

Physiological signs are critical indicators of the physiological state of an individual. Their monitoring is vital for many applications, such as disease diagnosis, tracking immediate or long-term effects of surgery, medicinal therapy, early identification of fatal disorders, and sleep analysis [6]. Among different physiological signs estimation approaches, rPPG has been an active area of research for the last two decades due to

its advantages over conventional contact approaches. For instance, it does not require physical contact with the skin, which makes it viable for unobtrusive monitoring for prolonged periods and skin-sensitive scenarios. Secondly, unlike conventional methods, it does not constrain the subject's motion, which enhances its applicability for various scenarios such as non-contact sleep monitoring, driver health monitoring, etc. Furthermore, among different physiological signs, Heart Rate (HR) and oxygen saturation (SpO2) have been commonly monitored by physicians due to their ease of measurement and portability using PPG [8, 9].

The rPPG-based methods require a suitable ROI to extract accurate PPG information. A study by Kooji and Naber [103] concluded that the face offers better reliability for rPPG estimation. Following, the rPPG method measures subtle color variations due to blood flow across arteries, synchronous with the heartbeat. Since arteries are present in the skin's hypodermis (inner last) layer, the light radiations penetrate deeper, resulting in limited reflected radiations and ultimately leading to poor pulsatile strength of the rPPG signal. Secondly, the face topology and its inevitable head movements during video capturing lead to non-uniform light distribution, which introduces illumination variations and motion artifacts. A higher degree of motion during video capturing results in overlapping of motion and pulse spectra, while illumination variation results in an inaccurate ROI selection [69, 153], ultimately leading to false estimates.

Most real-time applications of physiological signs monitoring require dim light or dark environments with a significantly higher degree of motion, such as intensive care units and sleep monitoring labs. These conditions can be detrimental to the quality of rPPG signals, which would eventually lead to spurious physiological estimates [153–156]. For instance, insufficient lightning during the environments above may deteriorate the capturing of substantial facial details, ultimately resulting in the weaker pulsatile amplitude of the resultant rPPG signals with low signal-to-noise ratio [155]. To deal with insufficient lightning problems, the IR spectrum has proven to be the best choice since it is independent of light conditions. However, the rPPG signal extracted by IR exhibits poor pulsatile strength and is also susceptible to motion artifacts (single channel) [157], which leads to erroneous estimates. Increasing the IR wavelength channels provides promising results [158] but also increases the associated costs and complexity.

In contrast, although the RGB spectrum offers better pulsatile strength and motion robustness [158], it will also result in inaccurate estimates due to the extraction of insubstantial ROI details in dark or dim light environments. To deal with this, Odinaev et al. [159] proposed fine-tuning of camera exposure and gain. Addi-

tionally, Chen et al. [160] proposed a combination of image enhancement method Self-Calibrated Illumination (SCI) [161] with Physnet [162] using a time-shifted negative Pearson correlation loss. However, these studies have certain limitations. For instance, not all cameras are equipped with the feature of tuning camera gain and exposure time, e.g., embedded cameras in portable devices. Secondly, the selection of a suitable image enhancement method is non-trivial and challenging, as proven in the later sections of this research. Specifically, this study demonstrates that existing image enhancement methods are not generalized enough to deal with extremely low-light conditions. Additionally, existing studies were conducted using the assumed light conditions ranging from 8 to 104 lux [159, 160], which are still far from the real-time scenarios such as sleep monitoring and nighttime driving. Additionally, these studies have demonstrated their performance of HR analysis only, however, other physiological signs are also crucial for the health monitoring of an individual.

Based on the above analysis, this study highlights the following key questions when performing physiological measurements using RGB spectra in dark real-time scenarios:

1. What could be an optimal generalized light condition threshold covering most clinical and non-clinical real-time scenarios for physiological measurements?
2. Is it possible to provide a suitable enhancement method that ensures sufficient extraction of ROI details to extract the rPPG signal accurately?
3. Is it possible to estimate physiological signs other than HR using RGB spectra in dark environments?

These challenges were addressed based on the objectives presented in the subsequent section.

6.2 OBJECTIVES

Assuming an average illuminance ≤ 1 lux, this chapter aims to use a newly created database named Dark-Video dataset (the details have been presented in section 3.2.1 of Chapter 3) and accomplish the following objectives:

1. Compare existing SOTA low-light image enhancement methods to choose a suitable image enhancement method to provide a basis for the design and development of a novel deep learning-based image enhancement method for non-contact physiological variables estimations in the considered environment.

2. Develop a deep learning-based image enhancement method taking a paired input (low and slightly higher-exposure images) to enhance the low-exposure image.
3. Integrate the image enhancement method with the current SOTA non-contact HR and SpO2 estimation methods to investigate the possibility of conducting physiological measurements in extremely dark environments.
4. Conduct comprehensive experiments to demonstrate the dependability and effectiveness of the proposed approach when compared to other SOTA alternatives.

6.2.1 RELATED WORK

The physiological measurements using RGB spectra in dark environments are only possible if the darker videos are enhanced to extract substantial facial ROI details. One of the possible solutions for this is the conjuncture of image enhancement and non-contact physiological measurement methods. Therefore, this section examines state-of-the-art image enhancement and non-contact physiological variable estimation methods in context to dark environments.

6.2.2 Image Enhancement

Histogram Equalization (HE) [163] and Gamma Correction (GC) [164] are among the conventional approaches to image enhancement in poor lighting conditions. On the one hand, HE aims to widen the distribution of normalized pixel intensity values. At the same time, GC, based on the heuristically selected parameter, applies non-linear operations to individual pixels for image enhancement. A detailed overview of HE variants can be found in [165], while different variants of GC were presented in studies by Wang et al. [166], Rahman et al. [167] and Huang et al. [168]. These variants differed in terms of strategies for the gamma parameter selection. The main limitation of these techniques is their blindness to illumination, resulting in over- or under-exposed images and poor visual perceptibility. With the introduction of retinex theory by Land [169], several related methods, such as retinex methods with single/multiple scales, were proposed [170, 171]. However, these methods failed to preserve the naturalness and often generated over-exposed images. To overcome these limitations, several other retinex methods were proposed, such as Naturalness Preserved Enhancement (NPE) by Wang et al. [172], a method based on the fusion of multiple illumination maps by Ma et el. [173], and Simultaneous Reflectance and Illumination Estimation (SRIE) by Fu et al. [174]. However, these methods were prone to color distortion and high noise.

DL methods have recently been used to improve the quality of low-light images. Pioneering this approach, Lore et al. [175] first introduced a deep sparse autoencoder, LLNet, for image denoising and contrast enhancement. Subsequently, various methods such as Multi-Branch Low-Light Enhancement Network (MBLLEN) [176] and Edge-Enhanced Multi-Exposure Fusion Network (EEMEFN) [177] were proposed that considered multiscale feature extraction and fusion to improve image quality in low light conditions. Due to the similarity of the Retinex theory to the human visual system, several retinex-based deep learning methods such as LightenNet [178], KinD++ [179], PairLIE [180] and SCI [161] have also been proposed for low-light image enhancement. The novelty of these methods was based on the uniqueness of the deep-learning architectures for image decomposition, illumination enhancement, and reflectance restoration. However, the limitations of these methods are based on the assumptions on the extraction of illumination maps and reflectance components, such as piecewise smoothing of illumination maps.

Considering image enhancement, a generative modeling task, several GAN based methods such as EnlightenGAN [181] and other related methods (Meng et al. [182], Ignatov et al. [183], WESPE [184], Chen et al. [160]) have also been proposed. The main limitation of GAN-based methods is their sensitivity to hyperparameters. Unlike other supervised GAN-based methods, Enlightengan, an unsupervised method, was introduced to improve generalisability and solve the problem of overfitting in context to image enhancement tasks. Similarly, Low-Light Enhancement and Deblurring Network (LEDNet) [185] was proposed to enhance generalisability by modeling different types of degradations from low-light image datasets. Similarly, Yang et al. [186] proposed a neural representation-based method that normalizes the degradation of brightness and increase of noise by multimodal learning using an encoder-decoder framework.

Additionally, due to limited low-high pair datasets, a zero-shot learning-based approach of mapping low-light images to higher-order light enhancement curves was applied for image enhancement. Specifically, methods such as Zero-DCE++ [187], and BrightsightNet [188] (improved variants of Zero-DCE [189]) were proposed based on the approach mentioned above (i.e., image to curve mapping), which aims learn parameter maps (based on quadratic curve) for enhancement tasks.

The above methods were designed and developed for image enhancement in low light. However, the term "low-light" is not defined or reported in any of these studies, which is critical for applications such as extracting rPPG signals from face videos in dark conditions. For example, the studies by Odinaev et al. [159] and Chen et al. [160] attempted to perform rPPG signal extraction considering an environment

with illuminance levels between 8 and 104 lux. To the best of the author's knowledge, no image enhancement study has reported similar or related values, so the term low-light remained undefined. Consequently, the selection of a suitable image enhancement is a time-consuming task for applications such as the one addressed in this paper. Therefore, this study aims to perform a comparative analysis of existing low-light image enhancement methods and then select the best method for extracting the rPPG signal and subsequently estimating HR and SpO2 in the considered environment.

6.2.3 Physiological signs estimations

Numerous contactless physiological signs estimation studies exploiting the RGB and ir spectra have been proposed in the past few years [4, 63, 67, 69, 131, 190]. However, these studies used substantial illumination sources (for RGB spectra) for ROI selection and, eventually, physiological signs estimations. Additionally, limited studies have addressed the estimation problem in dark environments despite its relatively greater applicability in various fields. Among them, IR spectra have been predominantly used for HR and BR estimations in dark environments. For instance, Wang, Woster, and Brinker [190] pointed out the inability of existing IRsources to HR estimation in darkness and proposed a multichannel continuous spectrum IR illuminator based on the principle of phosphorous material. The study claimed that the proposed illuminator would be suitable for unobtrusive HR monitoring in dark environments. To address the problem of inevitable illumination variations in driving scenarios, a study by Guo et al. [191] explored the applicability of a time-of-flight camera for HR and BR estimation using cheeks and chest, respectively.Additionally, Nowara et al. [158] proposed an IR-based method, namely AutoSparsePPG, which uses NIR videos for HRestimations using sparse spectrum estimation in the driving scenario. Furthermore, Van Gastel et al. [192] depicted the applicability of IR spectra-based methods for HR estimations in NICUs in near darkness. Comas et al. [13] proposed a time series-based U-Net using IR-based videos for HR estimations. Additionally, its residual connections were modified by passing a gated recurrent unit through them to preserve temporal dependencies for rPPG signal extraction. The limitation of IR spectra-based methods is their susceptibility to motion artifacts and poor pulsatile strength of the extracted signal. There have been several attempts to overcome these limitations. For instance, motion susceptibility is addressed by using multichannel IR spectra with different wavelengths as proposed by Wang, Vosters, and Brinker [190]. Poor pulsatile strength has been addressed by using RGB-NIR fusion due to better pulsatile strength of RGB spectra [16, 193, 194]. However, these strategies increased

the complexity of rPPG signal extraction methods, which may be time-consuming, especially for real-time scenarios.

A few non-contact studies based on RGB spectra were also proposed for physiological measurements. Recently, Chen et al. [160], Odineav et al. [159], Wang and Zhang [195] have attempted estimating HR estimations under different illuminance environments ranging between from 8 and 400 lux. Although these studies have covered a few low-light real-time scenarios for non-contact HR measurements, their performance might be limited for extremely dark conditions, such as sleep monitoring and nighttime driving. Therefore, this study aims to attempt non-contact HR and SpO2 estimations in extremely dark environments, which also include these low-light scenarios as well.

On the other hand, the SpO2 estimation studies have proven their novelty by selecting different red and IR wavelength combinations for increased robustness [23]. In contrast, a few have used red and blue channels from RGB spectra of ROI for calculating ROR [196]. A different approach by Akamatsu, Onishi, and Imaoka [197] applied generative models to estimate SpO2 in ambient light conditions. Like HR, non-contact SpO2 estimations using RGB spectra have not yet been attempted in dark environments.

Three main limitations emerge from the above literature review: first, the term low-light is not substantially defined or reported in the literature; second, selecting a suitable image enhancement method for an application is challenging and time-consuming due to different specific requirements, for instance, rPPG signal extraction in the dark requires accurate preservation of color details, and third, HR estimations were mainly attempted in the dark environments, despite the importance of other physiological signs such as SpO2, BR, etc.

6.3 Proposed Methodology

The notion of RGB-based physiological measurements in dark environments represents a relatively new concept, diverging from the conventional reliance on IR spectra for estimating physiological signs [13, 158, 190–192]. Furthermore, the absence of a publicly available database addressing the specific conditions considered in this study adds to the novelty. Existing studies typically operate within an illuminance range of 8 to 400s lux, differing from the assumed low-light conditions in our study. Consequently, we propose a new database to explore the feasibility and reliability of the proposed approach.

The proposed approach for physiological measurements in the assumed environment involves two main steps: 1) enhancing the quality of facial video frames and 2) estimating physiological signs (HR and SpO2). 36 visually outlines this process. Effective enhancement of image frames necessitates an efficient enhancer capable of revealing details hidden in darkness. Subsequently, a method for estimating physiological signs should accurately extract the rPPG signal, facilitating correct HR and SpO2 estimations.

Given the limitations of existing image enhancement methods in providing substantial improvements, we introduce a new deep learning-based image enhancement method called 2E1D-Net (a two-encoder one-decoder network). This method utilizes a paired sample of images captured at different exposure levels for enhancement. The structure of 2E1D-Net is inspired by the concept of color information propagation proposed by Welsh et al. [198], with a difference that instead of image, the feature representations of the images at multiscale level are transferred for image enhancement. The trained model is then integrated with state-of-the-art non-contact estimation methods for non-contact HR and SpO2 measurements, respectively. The subsequent sections shall present the details of the proposed approach.

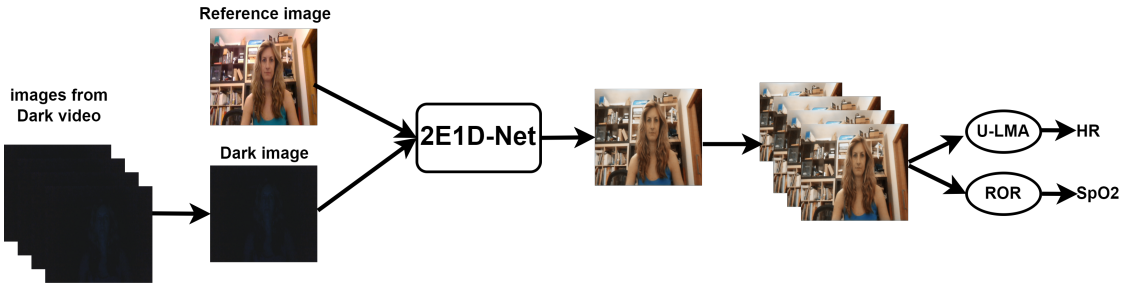


Figure. 36: A flow diagram of the proposed method for extracting HR and SpO2 estimations in dark environments.

6.3.1 Mathematical formulations of the Enhancement task

The enhancement task aims to improve the image quality by removing various distortions from the image. Alternatively, the distorted image consists of a clean image and noise due to various sources such as poor illumination, capturing devices, etc. Considering the additive nature of the noise, the distorted image can be expressed as:

$$I^d = I' + D \tag{21}$$

Where I' and D represent the clean image and noise, respectively. The enhancement task aims to approximate I' by removing D from I^d as:

$$\widetilde{I} = I^d \ \breve{} \ D \tag{22}$$

Where $\widetilde{I}$ is an approximation of I'. A few enhancement studies have assumed noise as gaussian [37,179,199], which might not hold in complex dark environments due to its dependence on illuminance and structure [200]. Therefore, the distribution of D can only be learned and minimized using a data-driven approach defined by a function G as:

$$\widetilde{I} = \min_{D} G\left(I^d\right) \tag{23}$$

By definition, G is an enhancement function that minimizes D by learning its distribution from data. Due to the non-linearity associated with the assumed dark environment (luminance $\leq$ 1 lux), it is infeasible to separate complex noise components from I^d in RGB spectra. Therefore, G projects I^d to a latent space to differentiate noise (darkness) from I', followed by decoding it back to an RGB image. Considering this, G can be seen as a composite function consisting of two functions G_{De}, and G_{En} as:

$$\widetilde{I} = \min_{D} [G_{De} \circ G_{En}](I^d) \tag{24}$$

where G_{en} aims to project the RGB spectra to a projection plane for noise removal, and G_{De} aims to decode the information of interest. Simplifying equation (24) leads to the following:

$$\widetilde{I} = \min_{D} [G_{De}(G_{En}(I^d))] \tag{25}$$

Due to unknown noise distribution, equation (25) might result in a highly distorted image (as shown in later sections). It can be solved by guiding the enhancement task with contextual information, which can be extracted from another image captured with a slightly higher exposure level, sharing common object of interest, I^a using G_{En}, to have the same projection plane. Therefore, equation (25) becomes:

$$\widetilde{I} = \min_{D} \left[G_{De}\left(G_{En}\left(I^d\right)\right) \bigoplus G_{En}\left(I^a\right)\right] \tag{26}$$

where $\bigoplus$ denotes the operation to transfer contextual information. Thus, the enhancement task aims to learn the distribution of noise D and minimize it, followed by decoding the information of interest. Adhering to equation (26), the proposed 2E1D-Net defined by G comprises two encoders following G_{En} taking a paired im-

age sample where a dark image represented by I^d is captured at a relatively lower exposure level than image I^a, sharing a decoder denoted by G_{De} for decoding the information of interest followed by element-wise aggregation operation $\bigoplus$ for aggregating feature maps post deconvolution and I^a, based on dimensionality constraints.

Post aggregation, a few artifacts still remain, which can be removed by appropriate lowpass filtering operations. Therefore, further refinement of the enhanced image is performed using a couple of convolution layers followed by Rectified Linear Unit (ReLU) activations, i.e., Refinement Block (RFB) (except for the last one, where sigmoid activation was used). It is essential to mention that from the perspective of rPPG signal extraction, only one high-exposure image is being used to enhance the whole dark video, captured for non-contact HR and SpO2 estimations in the assumed dark environment. This way, 2E1D-Net relaxes the condition of equal low-high pairwise combinations for enhancement. The proposed architecture is trained using a novel loss function:

$$\min_{G} loss(G) = \zeta L \tag{27}$$

where $\zeta = \{\alpha_i |\ 0 \leq i \leq 1\}$, denotes the coefficients associated with the different components of the loss function and $L = \{l_j |\ 1 \leq j \leq 3\}$, where i is the number of loss components. It is noteworthy that the training process of the proposed network includes minimizing (27) through subsequent iterations. The details of the proposed loss function are defined next.

6.3.2 Loss Function

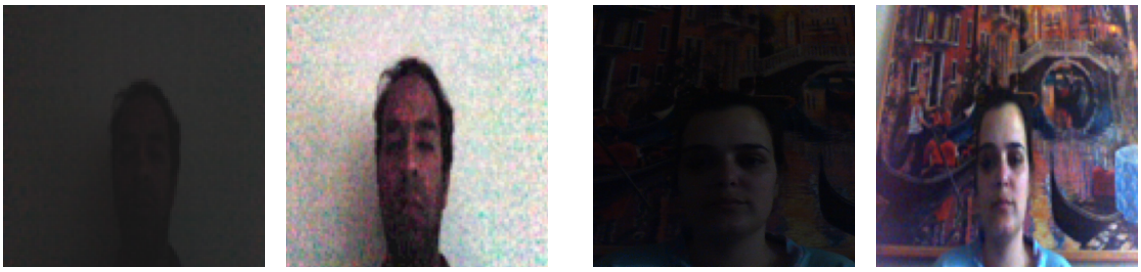
In the absence of appropriate ground truth, a loss function with well-defined constraints is key for the generalizability and robustness of unsupervised DL models. Fortunately, a paired image sample with different exposure levels with carefully designed loss components can efficiently guide the training process of the network, which must eventually result in producing enhanced images by preserving image details and improving perceptual visibility. Additionally, the enhanced image frames of the video should be able to preserve substantial ROI details, which would facilitate temporal extraction of subtle color variations for rPPG signal extraction. Therefore, the enhanced images should be able to preserve substantial image details such as color, shape, and texture.

Therefore, assuming the paired image sample [I^d,I^a] sharing a similar object of interest, the proposed loss function aims to exploit the similarity of image samples. Color information can be preserved with channel-wise pixel intensity differences

between the enhanced image, i.e., $L_1 = ||\widetilde{I} - I^a||_1$. Additionally, a higher degree of distortions in I^d resulted in shape irregularity, as shown in Figure 37 (the reflectance map is shown for better representation), which can be alleviated by restoring the edges of the object. Based on a study by Zhao et al. [201], it was found that Mean Squared Error (MSE) can efficiently preserve edges. Therefore, edge preservation can be attained by using MSE between the reconstructed and I^a as $L_2 = ||\widetilde{I} - I^a||_2^2$. Finally, the texture or the structural similarity can be improved by calculating the structural similarity index, also known as Structured Similarity Index (SSIM) [202]. Therefore the following equation $S_L = (1 - SSIM(\widetilde{I}, I^a))$ could be used to train the network. However, the flatter regions of the image cannot improved by using the conventional SSIM, as the network could not preserve the local structure, and splotchy artifacts will also be reintroduced due to substantially low standard deviation in those regions. This problem can resolved by using Multi-Scale Structural Similarity Index (MS-SSIM) [201], therefore, the above equation is modified to $(1 - MS - SSIM(\widetilde{I}, I^a))$. Based on the above analysis, the loss function can be defined as:

$$\min_G loss(G) = (\alpha_1 L_1 + \alpha_2 L_2 + \alpha_3 S_L) \tag{28}$$

where α_1, α_2, and α_3 are the balancing factors of $L1$, $L2$, and S_L. equation (28) is optimized with an Adamax algorithm [54] with a learning rate of $1e-3$, constrained to minimize the proposed loss function to enhance the images.



(a) (b) (c) (d)

Figure. 37: Original images (a,c) and corresponding reflectance map (b,d) from fine-tuned Kind++. Note that the distortions are neither visible in original images nor illumination maps, so reflectance maps are used to show the distortions in the random image samples.

6.3.3 2E1D-Net

Based on section 6.3.1, a two-encoder one-decoder network named 2E1D-Net is proposed, which takes paired image samples, each captured at different exposure levels. It is essential to mention that the image captured at a higher exposure level is not the ground truth but is used as an approximation with better perceptual visibility

than the one captured at a lower exposure level. Also, for each video, only one image as a ground truth approximation is used to enhance all image frames of the video. The schematic diagram illustrating the 2E1D-Net architecture is presented in Fig 38. The encoder-decoder framework alleviates the dark component from the dark image. On the other hand, the additional encoder improves the enhancement process by extracting and transferring the contextual information at a multiscale level from the high-exposure level image using residual learning. Additionally, the feature representations of the last convolution blocks of both encoders are fused before passing to the decoder. Subsequently, the feature maps of the additional encoder, at different scales, are fused with the deconvoluted feature map of the respective deconvolution block via—residual connections. The decoder also possessed a refinement block to avoid the effect of minute color distortions causing color overspreading and preservation of image details, followed by a sigmoid activation to produce an enhanced image. A detailed explanation of various components of the proposed 2E1D-Net is presented in the following subsections.

a) Encoder architecture

The encoder aims to encode the images to their equivalent feature representations for image enhancement tasks: 1) to efficiently segregate the noise from the information of interest due to their different distributions, and 2) to ensure appropriate contextual information propagation for improved enhancement.

The 2E1D-Net encoders share a similar architecture that consists of five Convolution Blocks (CBs), as illustrated in Fig 39. The first CB consists of four convolution layers, with the last two incorporating ReLU activations. The convolution layers extract features, while ReLU activations help to learn complex patterns. In contrast, the other CBs consists of ReLU-activated convolution layers and a Maximum pooling layer to subsample the prominent features of the feature map. All convolution layers have a kernel size of 3 with stride 1, while maximum pooling has a kernel size and stride of 2, respectively.

b) Decoder architecture

The decoder decodes the feature representations by hierarchically decoding the features from the encoders. The input to the decoder architecture, presented in Figure 40, is the aggregated set of features from both encoders. The decoder consists of three Deconvolution Blocks (DCBs), each consisting of a deconvolution followed by ReLU-activated convolution layers (40). The deconvolution layer decodes the feature

representations by maintaining the same connectivity pattern as during encoding. It is demonstrated in the respective ablation study that with a single encoder and decoder, the enhanced image cannot preserve color information and suffers from shape irregularity. Therefore, having shared the object of interest by paired image samples $[I^d, I^a]$, the respective feature maps of I^a (from the additional encoder) are fused with the deconvoluted feature maps at different scales via residual learning. Subsequently, the fused feature map is passed through a ReLU-activated convolution layer of DCB for feature extraction. Following, a RFB was also used to elevate the effect of minute color distortions causing color overspreading and preserving image details. It consists of four convolution layers with ReLU activation, except for the last one. Finally, a sigmoid activation is applied to generate the enhanced image. The contributions of these components of 2E1D-Net are demonstrated using the carefully designed ablation studies in the next section.

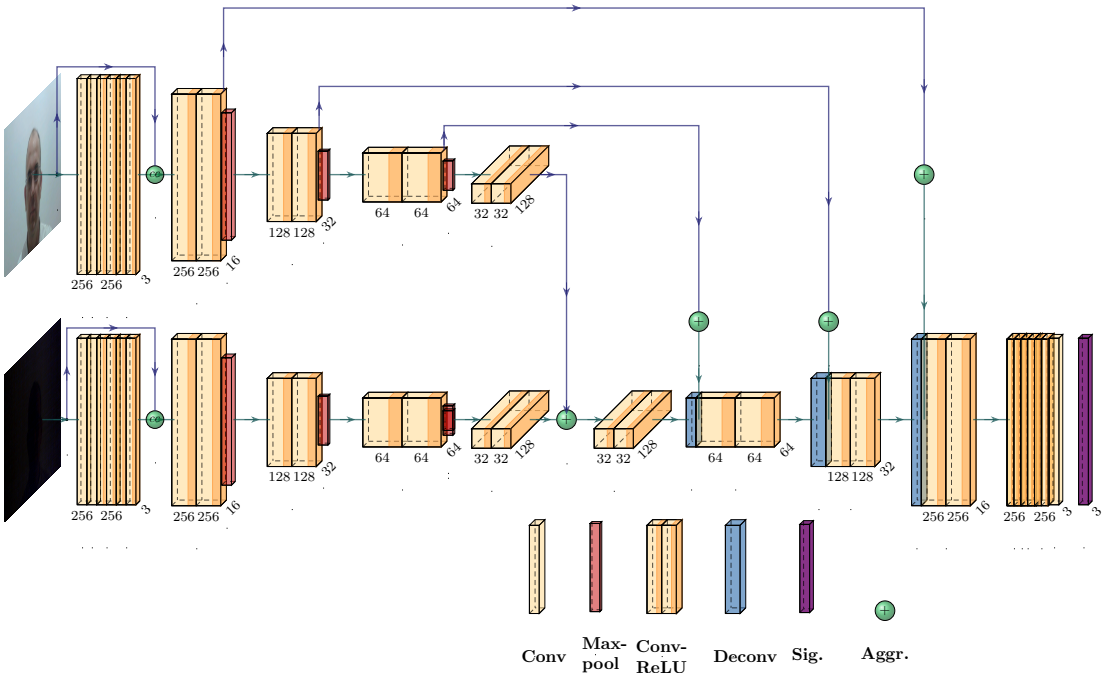


Figure. 38: Architecture of Proposed 2E1D-Net.

6.3.4 Physiological Signs Estimations

As mentioned in section , 2E1D-Net is cascaded with SOTA RGB spectra-based non-contact physiological signs estimation methods for HR and SpO2 measurement in assumed darkness condition (illuminance $\leq$ 1 lux) to attain two objectives:1) investigating the ability of 2E1D-Net to preserve substantial image details for accurate PPG information extraction, and 2) to test the conjunctions mentioned above for

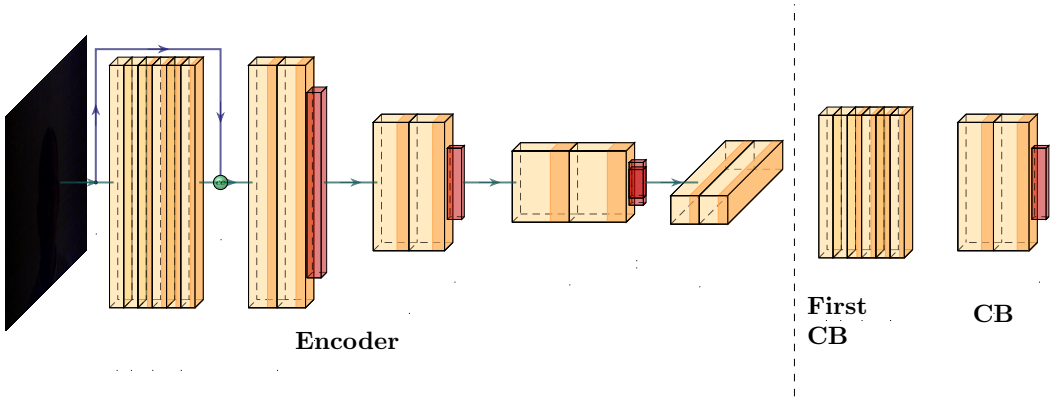


Figure. 39: Encoder architecture consists of two types of convolution blocks (1st CB) and (CB). 1st CB consists of 4 convolution layers with ReLU activations (Yellow with orange color); CB consists of a convolution layer with ReLU activation, followed by a max-pooling layer (dark orange).

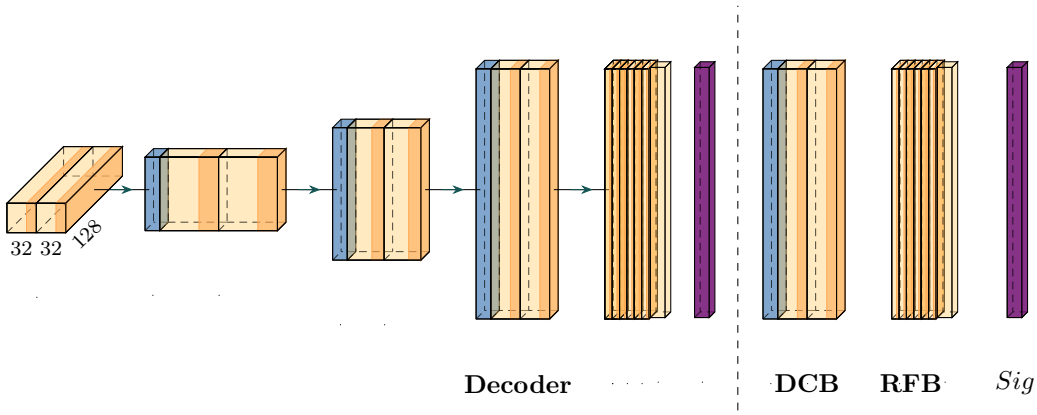


Figure. 40: Decoder architecture consists of deconvolution blocks (DCB) and refinement blocks (RFB). DCB consists of a transposed convolution layer (gray) and a ReLU activated convolution layer. RFB has the first three ReLU-activated convolution layers, followed by sigmoid activation (purple).

non-contact HR and SpO2 estimations in the assumed dark environment. This work compares nine SOTA HR estimation methods and ROR for HR and SpO2 estimations, operating in RGB color space, respectively. Finally, the best conjunctions were also compared with SOTA IR-based methods to investigate if the proposed conjunctions operating in RGB color space could provide comparable performances with IR-based methods, satisfying the darkness constraint. Since the darkness conditions were not reported in the respective IR spectra-based studies, the respective studies were selected if the proposed methods were tested in dark environments. Additionally, the default conditions for each method were maintained for better performance and fair comparisons.

6.4 Results

This section first presents the implementation details of the proposed image enhancement method, HR, and ROR methods. Following, the quantitative and qualitative comparative analysis of the proposed 2E1D-Net and conjunctions (2E1D-Net+HR/SpO2 method) were presented. Specifically, this section is divided into three parts: 1) first, the qualitative and quantitative comparative analysis of 2E1D-Net with existing state-of-the-art image enhancement methods was performed; 2) the trained model of 2E1D-Net in conjunction with SOTA RGB spectra based HR estimation, and ROR methods were analyzed to select the accurate method using suitable performance metrics; and 3) the resulting accurate HR and SpO2 conjunctions were further compared with IR based methods to evaluate their feasibility for the proposed or similar challenging dark environments.

6.4.1 Implementation details

A dataset comprising 57 videos was used to design the experiments for image enhancement and physiological signs estimations. For image enhancement, 57 videos were split into image frames, resulting in 43,534 images, which followed subject-independent and random split into training, validation, and testing datasets, in the ratio 50:25:25 (%), respectively. This resulted in a training set comprising 21,767 images, while the validation and testing dataset consisted of 10,884 and 10,883 images, respectively. The image enhancement network 2E1D-Net was trained using a batch size of 8 with image dimensions $256 \times 256 \times 3$ for 60 epochs using an *Adamax* optimizer with a learning rate of $1e-3$. An Early Stopping criterion with a patience value of five was also used as a termination criterion to stop training. For

/achr and /acspoo measurements, the default conditions reported in the respective studies were maintained for better performances and fair comparison.

6.4.2 Image Enhancement

The performance of 2E1D-Net is compared with eight SOTA image enhancement methods, which include zero-shot learning (Zero-DCE++ [187], BrightsightNet [188]), Retinex theory-based (PairLIE [180], and KinD++ [179]), and GAN based (EnlightenGAN [181]), methods using full low-light image-based methods (NerCO [186], SCI [161], LEDNet [185]), respectively. The key component of the methods included for comparative analysis is that they used the encoder-decoder framework as baselines. All methods were analyzed using the proposed Dark-Video dataset, focussing on the specific problem addressed in this work, i.e., HR and SpO2 estimations in the considered dark environment using RGB videos.

All SOTA image enhancement methods were fine-tuned based on the default conditions reported in the respective papers. Subsequently, their performance was analyzed, qualitatively and quantitatively, using visual comparisons and the following metrics: Peak Signal-to-Noise Ratio (PSNR) [55], SSIM [52], Natural image quality evaluator (NIQE) [203], and Learned Perceptual Image Patch Similarity (LPIPS) [204]. Additionally, different ablation studies were designed to investigate further the performance of constituting components of 2E1D-Net, and the visual results were presented in subsequent sub-sections.

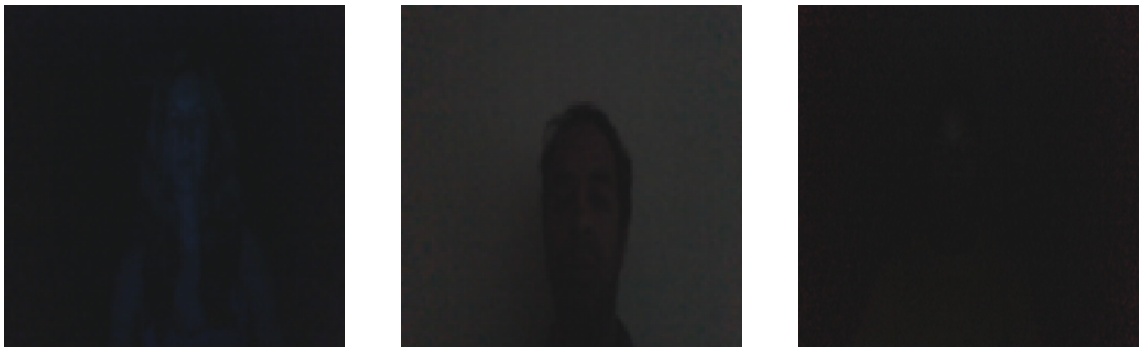
The qualitative comparisons are presented in Figure 41, while Table 22 presents the quantitative comparisons for all image enhancement methods. The sub-optimal performance of zero-shot learning methods is attributed to parameter map estimation using quadratic light curves. Due to the presence of vivid color and texture distortions in the low-light images, the estimated parameter maps, even after fine-tuning, could not provide substantial enhancement. An illustration in Figure 42a illustrates this fact. Additionally, since BrightsightNet and Zero-DCE++ are both based on the same baseline, i.e., Zero-DCE, this finding is assumed to hold for both methods. Similarly, the Retinex theory-based methods (PairLIE and KinD++) could not provide substantial enhancement due to the underlying assumptions of the Retinex theory. Specifically, Retinex theory assumes a degradation-free reflectance map, which is not always feasible in practical situations, as also pointed out by the original KinD++ study [179]. Additionally, since the low-light images were highly distorted, the mechanisms used for illumination refinement and reflectance restoration in these methods could not elevate the effect of these distortions, substan-

tially,ultimately resulting in distorted images. Figure 42b presents an illustration of the reflectance map, produced by layer decomposition Net of KinD++, fine-tuned by the proposed dataset (The illuminance component is not shown since it is a full black image).

Table 22: Comparative analysis of image enhancement methods.

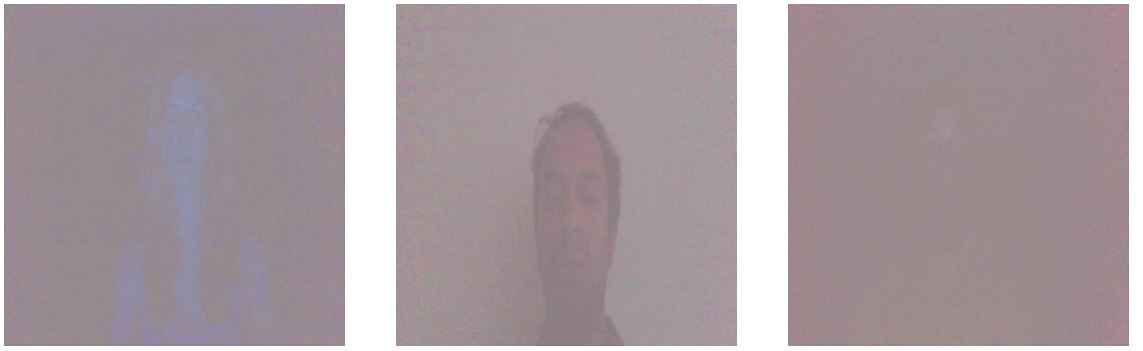
Methods	PSNR	SSIM	NIQE	LPIPS
Zero-DCE++	9.90	0.28	26.21	3.75E-05
BrightsightNet	10.30	0.34	26.20	4.29E-05
PairLIE	10.78	0.39	26.94	4.05E-05
KinD++	10.17	0.28	24.46	3.76E-05
EnlightenGAN	10.31	0.37	26.21	3.37E-05
NeRCo	8.76	0.12	**19.18**	4.43E-05
SCI	10.72	0.37	26.28	4.14E-05
LEDNet	10.48	0.40	26.35	4.04E-05
2E1D-Net	**34.19**	**0.97**	25.59	**1.00E-06**

The substandard performance of EnlightenGAN, Neural Representation method for Cooperative low-light image enhancement (NeRCO), SCI and LEDNet is also apparent from Figure 41 and Table 22, respectively. The suboptimal performance of EnlightenGAN is due to the poor performance of its self-regularized attention mechanism, which is dependent on the illuminance of the low-light image samples. Specifically, the attention mechanism could not identify the darker regions accurately (only facial and top region is identified as dark regions), which resulted in enhancement to these regions only, i.e., as shown in Figure 42c. However, the slight enhancement was due to carefully designed losses and a global-local discriminator. Upon critically analyzing the architecture of NeRCO and LEDNet, it was found that their image enhancement capability is driven by efficient and accurate feature map extraction. Additionally, feature map extractors of these methods are based on residual nets [205]. The qualitative and quantitative performances of these methods are attributed to inefficient feature extraction. For instance, NeRCO used Resnet-based encoder for feature map extraction in Mask Extractor (MasKE) and Neural Representation Network (NRN), while LEDNet used Resnet blocks, comprising Residual downsampling and upsampling blocks. To further prove this reason, an illustration of feature maps extracted from the fine-tuned versions of NeRCO (feature map from the encoders MasKE and NRN) and LEDNet (LE encoder, post feature fusion at multi scales) were presented in Figs. 42d, and 42e, respectively, which shows inefficient feature extraction (extracted features are highlighted in yellow). Therefore, replacing Resnets with powerful deblurring networks might improve the performance of these methods.



(a) Original

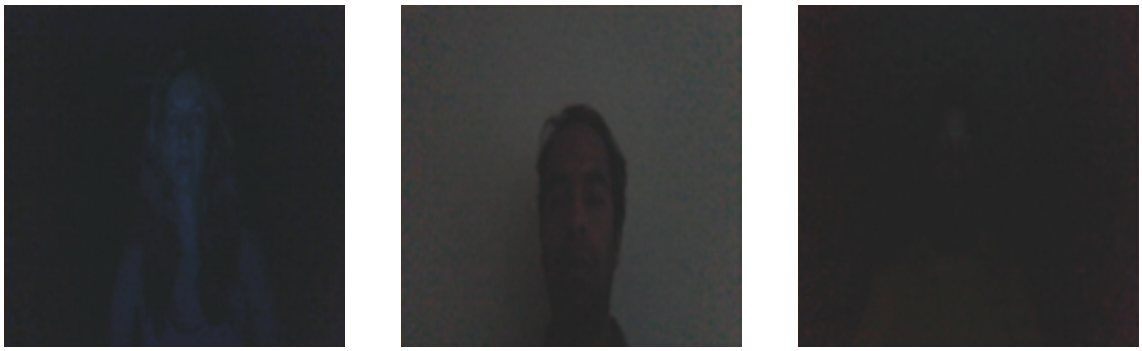
(b) Zero-DCE++



(c) BrightSightNet

(d) PairLIE

Figure. 41: Comparative analysis of SOTA enhancement methods with 2E1D-Net, depicting image samples from European (left), Asian (middle), and African (right) ethnic groups.



(e) EnlightenGAN

(f) KinD++

(g) NerCo

(h) SCI

Figure. 41: Comparative analysis of SOTA enhancement methods with 2E1D-Net, depicting image samples from European (left), Asian (middle), and African (right) ethnic groups.

(i) LEDNet

(j) 2E1D-Net (Proposed)

Figure. 41: Comparative analysis of SOTA enhancement methods with 2E1D-Net, depicting image samples from European (left), Asian (middle), and African (right) ethnic groups.

(a) (b) (c) (d) (e)

Figure. 42: Visual results of state-of-the-art methods depicting potential reasons for suboptimal performances:(high images (top), and visual results (bottom)): (a) Parameter map from ZeroDCE++ Parameter Map (b) KinD++ reflectance map,(c) EnlightenGAN attention map, (d) NerCO's encoder feature map, and (e) LEDNet feature map.

On the other hand, SCI's consideration of enhancing low-light images based on a single enhancement block did not work well, which contradicts their consideration of substantial image enhancement with only one enhancement block, only [161]. It is apparent from Figure 41h that the output looks distorted and visually unpleasing due to the presence of extreme distortions in the low-light images. Additionally, the enhanced images also consisted of haze and halo artifacts. Therefore, adding subsequent image enhancement blocks might provide better enhancement results.

Conclusively, the image samples of the proposed *Dark-video* dataset are intrinsically distorted in terms of texture, color, and shape, which contributed to the suboptimal performance of state-of-the-art image enhancement methods. This observation is consistent with the proposed 2E1D-Net, which will be explained later in the ablation studies. Therefore, to alleviate the effect of these distortions, prior knowledge is required. For instance, PairLIE assumed that a pair of low-light images could provide better insights into enhancement tasks, or KinD++ aims to restore the reflectance component guided by illumination. This supports the fact that the enhancement process needs to be guided by contextual information, as mentioned in section 6.3.1. Consequently, 2E1D-Net is designed based on equation (26). Specifically, 2E1D-Net used image pairs of different exposure levels, which facilitated the substantial (based on the proposed loss function) contextual information transfer in the form of feature maps (at different scales) through residual learning. As a result, 2E1D-Net managed to provide substantial image enhancement, unlike state-of-the-art methods, as proven by its superior quantitative metrics and visual representation from Figure 41, and 22, respectively. To further investigate the contributions of 2E1D-Net's components, two ablation studies to prove the effectiveness of architectural components and loss function components were presented in subsequent sections.

6.4.3 Ablation studies

a) Architectural Components Ablation Study

This ablation study tests the contribution of an additional encoder and the RFB of the 2E1D-Net for image enhancement. Therefore, 2E1D-Net was retrained after removing these components, one at a time, which resulted in two models: one without an additional encoder and the other without a RFB. It is apparent from Figure 43a that the additional encoder significantly contributes to the enhancement process by elevating the effect of distortion by knowledge transfer from similar high-exposure

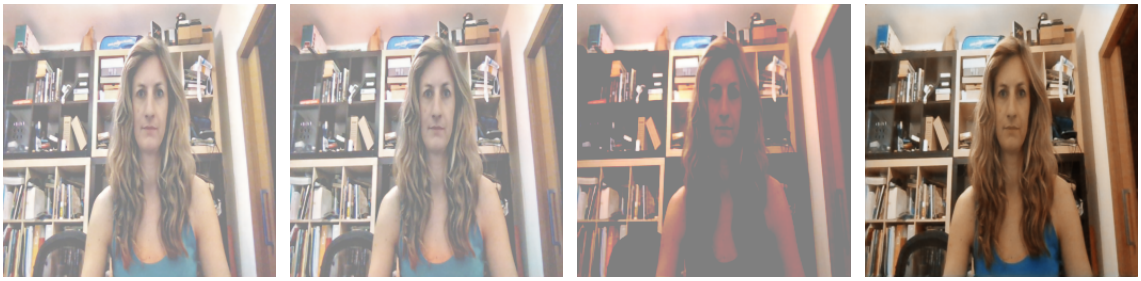
images. This observation also justifies the suboptimal performance of SOTA methods due to the absence of prior information during the enhancement process. Figure 43b presents the impact of the RFB in 2E1D-Net in elevating the effect of color distortions. As mentioned earlier, these distortions can be removed by applying lowpass filtering operations using a stack of convolution layers with non-linear activations (ReLU and Sigmoid). Therefore, the above-mentioned components contributed significantly to ensure superior performance (Figure 43c).

(a) w/o second encoder (b) w/o Refinement block (c) With all components

Figure. 43: Ablation studies analyzing the contributions of loss functions components.

b) Loss Functions Ablation Study

A similar ablation study was also designed to investigate the contribution of various components of the proposed loss function. Specifically, 2E1D-Net was trained by removing one loss component at a time, which resulted in three trained models. The visual illustrations of these models are presented in Figs. 44a, 44b, and 44c, respectively. It is apparent from these figures that the absence of L1 loss resulted in color distortions, increased haze, and halo artifacts (rooftop region of the image), while the absence of L2 loss caused illuminance degradation, resulting in poor contrast and motion blur. On the other hand, the absence of MS-SSIM caused severe color and shape distortions locally and globally with color overspreading. The combination of the above-mentioned functions ensured the superior performance of 2E1D-Net by alleviating the effect of distortions mentioned above.



(a) w/o L1 loss (b) w/o L2 loss (c) w/o MS-SSIM (d) Full loss function

Figure. 44: Ablation studies analyzing the contributions of loss functions components.

6.4.4 Physiological signs Estimation

a) HR Estimation

The performance of RGB spectra-based, non-contact SOTA HR and SpO2 estimation methods is dependent on the extraction of ROI details, which requires sufficient light conditions. Therefore, their applicability is limited in the case of low-light conditions. This work demonstrated their ability to work in extremely dark conditions when cascaded with an image enhancement method.

Specifically, seven SOTA HR estimation methods cascaded with 2E1D-Net were quantitively compared using six performance metrics: RMSE, MAPE, ME, standard deviation, accuracy, and Pearson correlation values under 0.01 significance level (α). The following HR estimation methods, ICA-Poh [4], CHROM [69], POS [63], KernelICA [67], FastICA [16], and ICA-Neg and U-LMA [131], were included in this analysis, based on the study by Gupta et al. [131].

Table 23 demonstrates the average performance metrics of the above-mentioned methods. The substandard performance of ICA is attributed to video compression and similar pulse and artifacts spectra magnitudes, which resulted in corrupted rPPG signal [4]. The same observation is also proved by the CHROM method since its alpha-tuning procedure suffered due to similar pulse-artifact spectra [69]. However, POS performed relatively better than CHROM due to different projection planes based on physiological information, unlike CHROM, where projection planes were based on specular components (challenging to estimate) [63, 69]. The performance of the aforementioned methods was notably impacted by the significantly similar magnitudes of the pulse signal and artifact spectra. It is potentially due to the prevalence of color distortions resulting from non-rigid head movement and illumination variations, potentially due to the inevitable effect of the darkness component in the original videos.

Table 23: Performance metrics for HR estimation methods.

Methods	RMSE	MAPE	SD*	μ	Accuracy	r*
ICA-Poh [4]	27.52	30.51	19.15	19.92	19.30	0.09
CHROM [69]	15.25	17.39	10.39	11.25	35.09	0.41
KernelICA [67]	12.26	12.52	9.37	8.00	56.14	0.56
POS [63]	10.84	11.46	7.02	8.30	42.10	0.74
FastICA [16]	5.87	6.29	4.09	4.25	70.17	0.86
ICA-Neg [131]	4.36	3.16	3.63	2.46	87.72	0.94
U-LMA [131]	**3.50**	**2.90**	**2.91**	**1.98**	**91.23**	**0.95**

In contrast, KernelICA performed better than the above-mentioned methods due to its kernel density-based ICA [139], which is resistant to similar pulse-artifact spectra. However, higher error metrics reported by the method were potentially due to the assumption of smoothness and continuity of independent components, which is not always possible in practical scenarios, especially in dark environments, despite substantial enhancements.

On the other hand, entropy-based ICAs performed significantly better since entropy ensured better statistical independence than kurtosis [144]. Specifically, the negentropy-based cost functions of FastICA and U-neg ensured their better performances, unlike other methods. However, U-neg performed slightly better than FastICA, especially in terms of error metrics. This proves that the entropy maximization of cumulative distributions of raw RGB signals ensures better statistical independence than the signals themselves. Finally, U-LMA outperformed other methods in all performance metrics due to the observation mentioned above, in assistance with robust optimization and faster convergence provided by the LMA [137, 138] for entropy maximization. For further insights into the superior performance of the 2E1D-Net and U-LMA combination, its performance is also analyzed using Bland-Altman analysis, as presented in Figs. 45.

The mean bias reported by the Bland-Altman plot (Figure 45) is -0.4737 beats per minute (bpm), while the upper and the lower statistical limits are defined by $\mu \pm 1.96std$ are -7.4645 and 6.5172, respectively. In other words, on average, the HR estimated by the algorithm measures 0.32 bpm less than the conventional BVP sensor used. Statistically,for mean bias, the 95% confidence intervals lie between -0.31 to 0.95 bpm. Additionally, for upper and lower statistical limits, 95% confidence intervals lies between 6.32 to 8.54 bpm and -7.91 to -5.69 bpm, respectively. Since 95% intervals of mean bias, lower and upper statistical limits lie closer to 2 bpm minute, which is lower than the standard deviation error. Additionally, the agreement between ground truth and estimated values is slightly lower, covering 91% of data points within the statistical thresholds, as per Bland-Altman Analysis. This

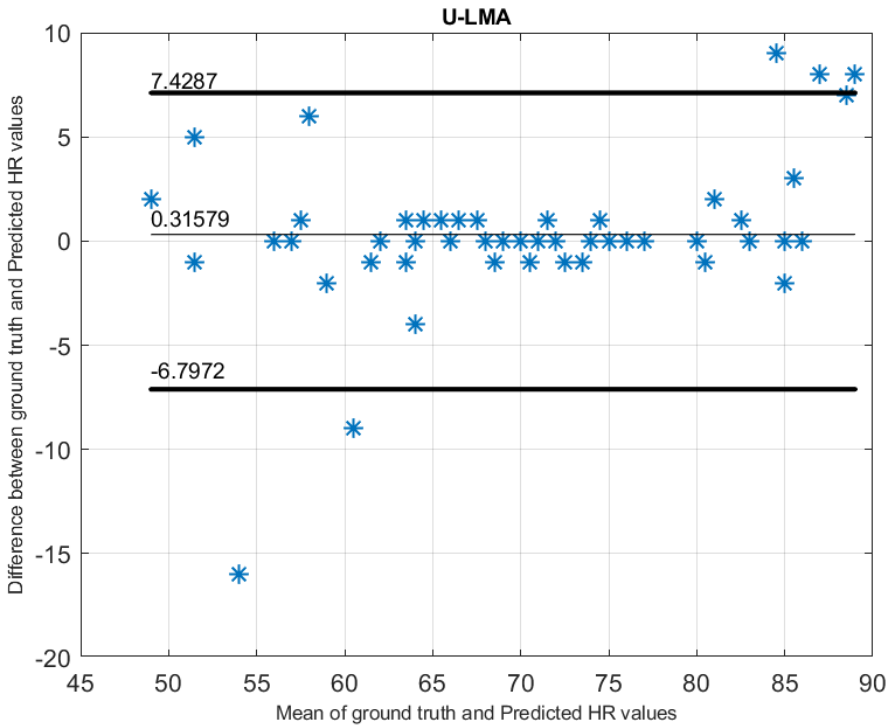


Figure. 45: Bland-Altman plot of 2E1D-Net cascaded to U-LMA.

agreement can be improved by considering more video samples for the analysis. Additionally, Pearson correlation value (r) of 0.9504, closer to 1 at 0.01% significance, justified the superior performance of U-LMA in conjunction with 2E1D-Net.

b) Comparative Analysis with IR-based Methods

Since IR spectra are robust to illumination variations, they have been conventionally used for dark environments. To prove the applicability of RGB spectra in such environments, the best combination, i.e., 2E1D-Net and U-LMA, is also compared with SOTA IR spectra-based HRestimation methods. These methods were selected on the basis of their applicability in dark conditions. Therefore, following HR estimation studies by Wang, Vosters, and Brinker [190], Guo et al. [191], Van Gastel et al. [192], Nowara et al. [158], and Comas et al. (TURNIP) [13] were included in this analysis. Specifically, Table 24 presents the performance comparison of IR-based HR estimation methods with the proposed combinations, based on reported RMSE, ME, and accuracy, respectively. The comparable performance of the proposed combination is apparent from the respective table, which suggests the applicability of RGB spectra as an alternative to IR spectra without compromising the performance.

Table 24: Comparative analysis results of 2E1D-Net-ULMA with IR spectra-based HR estimation methods.

Studies	RMSE	μ	Accuracy
Nowara et al. [158]	11.2	-	64.7
Comas et al. [13]	4.8	-	-
Wang et al. [190]	4	4.04	-
Guo et al. [191]	-	-	82
Van Gastel et al. [192]	-	1.5	87
2E1D-Net-ULMA	**3.50**	**1.98**	**91.23**

c) SpO2 Estimations

Due to the predominance of ROR method for SpO2 estimations, it is also cascaded with 2E1D-Net to demonstrate the possibility of non-contact SpO2 estimations in extremely dark environments, as considered in this study. However, this analysis is restrictive to a normal SpO2 value range,i.e., 94-99%, due to the associated complexity in acquiring abnormal SpO2 value ranges [101]. Additionally, to overcome the problem of limited samples for two SpO2 values, 94% and 95%, corresponding nine samples were also included from the VIPL-HR database [94]. Subsequently, a regression model was trained to map ROR values to ground truth SpO2 values for which the performance metrics are reported in Table 25. Additionally, the performance of the proposed combination is also demonstrated using the Bland-Altman plot, presented in Figure 46.

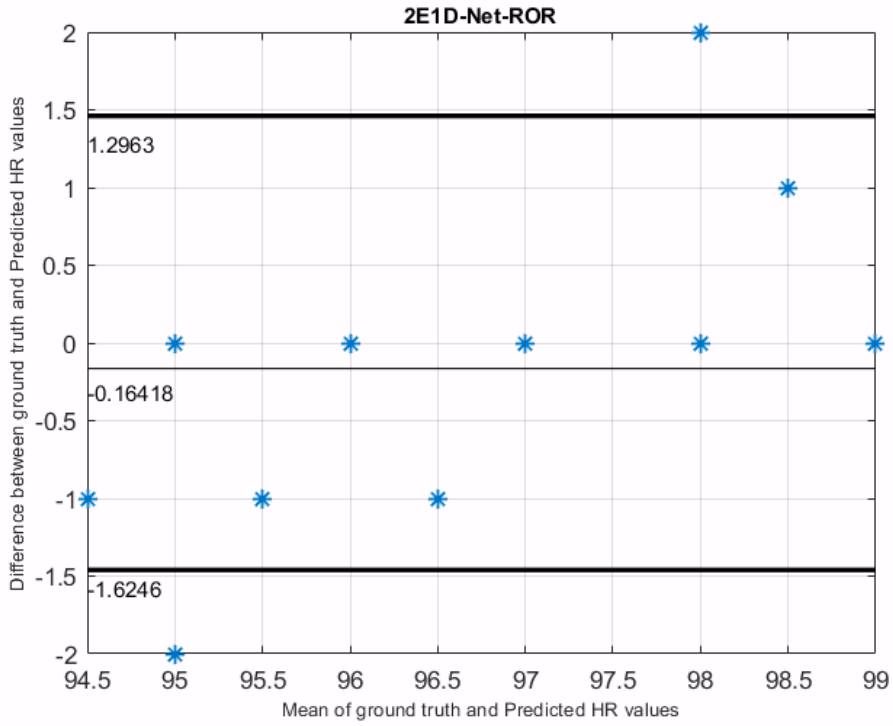


Figure. 46: Bland-Altman Plot of 2E1D-Net cascaded to ROR.

It is noteworthy that one data point in the figure corresponds to more than one video sample due to the restrictive SpO2 value range. From Figure 46, the mean bias between ground truth and estimated values is -0.1343 with upper and lower statistically limits ($\mu \pm 1.96 * std$) as 1.3791 and -1.6477, respectively. Statistically,for mean bias, the 95% confidence intervals lie between -0.31 to 0.01 bpm. Additionally, for upper and lower statistical limits, 95% confidence intervals lies between 1.05 to 1.55 bpm and -1.87 to -1.37 bpm, respectively. Since 95% intervals of mean bias, lower and upper statistical limits lie with in 0.5% minute, which is slightly lower than the standard deviation error (0.69%). Notably, the statistical thresholds are fairly smaller than the acceptable error difference, which is critical for the commercial viability of the method. Also, Pearson correlation value of 0.92 at 0.01% significance level justified a stronger correlation between ground truth and estimated SpO2 values. Hence, the proposed combination of 2E1D-Net and ROR demonstrated the potential of accurate non-contact SpO2 measurements in dark environments.

d) SpO2 comparative analysis

The proposed non-contact SpO2 measurement combination was also compared with SOTA RGB and IR-based estimation methods. Following methods, Akamatsu, Onishi, and Imaoka [197], Casalino, Castellano, and Zaza [206], and Guazzi et al. [85] were compared using MSE, Pearson correlation (r), and Mean Absolute Error (MAE). Table 25 shows that the proposed combination performed far better than other methods. However, the better performance could be due to a restrictive normal range considered in this work. Therefore, it is important to test this combination using a wider range to reach a conclusion.

Table 25: Comparative analysis of contactless SpO2 estimation methods.

Studies	Spectra	RMSE	r*	MAE
Casalino et al. [12]	RGB	1.64	0.53	1.33
Shao et al. [84]	IR	1.3	0.94	-
Akamatsu et al. [56]	RGB	0.88	0.45	0.55
Guazzi et al. [55]	RGB	-	0.81	2.08
2E1D-Net-ROR	**RGB**	**0.76**	**0.92**	**0.43**

6.4.5 Key Observations and Limitations

The extensive experiments conducted in this work proved that besides IR, RGB spectra could be a potential solution to estimate HR and SpO2 in scenarios such

as NICU and sleeping environments. However, the illumination environment, i.e., illuminance ≤ 1 lux assumed in this work, is empirically selected and does not necessarily conform with the above-mentioned environments. As mentioned earlier, several earlier attempts have been made to test the feasibility of HR estimations in dark environments [159,160,195]; however, this work has proven its novelty in terms of illumination conditions, followed by demonstrating the ineffectiveness of existing SOTA image enhancement methods for extremely dark scenarios. Additionally, this work also proved the effectiveness of conventional non-contact RGB-based HR or SpO2 estimations in conjunction with an efficient and robust image enhancement method.

However, this work has certain limitations: firstly, although 2E1D-Net was able to enhance the dark images substantially, it has a dependence on the quality of its slightly illuminated counterpart; secondly, the HR and SpO2 estimations in some scenarios require additional image processing for accurate assessments, and lastly, due to the complexity associated with SpO2 values acquisition, this study considered the normal SpO2 range, $94 - 99\%$. Future research directions of this work aim to mitigate these limitations.

6.5 Conclusion

This study proved the feasibility and reliability of RGB color space to estimate HR and SpO2 estimations in extremely dark environments (luminance ≤ 1 lux), by cascading an efficient and robust image enhancement method with conventional HR and SpO2 estimation methods. By identifying the reasons for the suboptimal performance of existing image enhancement methods in the proposed illumination condition, a two-encoder and one-decoder architecture is proposed and trained with a novel loss function (combination of L1, L2, and MS-SSIM). The encoder-decoder framework aims to alleviate the darkness component from low-light images, while the feature representations of the slightly better exposure counterpart, extracted from the additional encoder, were transferred and fused post deconvolutions (in the decoder) at a multiscale level. Subsequently, 2E1D-Net, cascaded with conventional SOTA RGB-based HR and SpO2 estimation methods, were compared and analyzed to demonstrate their applicability in extremely dark environments, with an additional Bland-Altman analysis for the best combinations. Besides, the study also proved the reliability and efficacy of the best HR and SpO2 combinations in comparison with the selected IR-based methods in dark environments.

This chapter demonstrated that RGB spectra can be used for non-contact estimations in extremely dark environments, as considered for this work. To the best of the author's knowledge, this is the first attempt to use RGB spectra for dark environments to estimate physiological parameters using a non-contact approach. This work serves as a significant contribution to this thesis. The next chapter highlights the key findings, limitations of the presented work, and future directions of this project.

Chapter 7

CONCLUSION, LIMITATIONS, AND FUTURE SCOPE

7.1 CONCLUSION

This study focuses on exploring the potential of the RGB image color model for physiological parameter estimations under dark environments with marginal illuminance inferior to 1.0 lux. This was done using a two-step approach, i.e., identifying the predominant method for rPPG signal extraction, proposing a novel method by elevating the respective limitations, and modifying the method to expand its ability for dark environments.

To accomplish the task mentioned above, an extensive systematic review of existing non-contact HR and SpO2 estimation studies was conducted to identify the research challenges associated with designing a standardized non-contact estimation study. Consequently, it was evident that ICA-based methods have been predominantly proposed for rPPG signal extraction. However, these methods suffer from the ordering problem, which resulted in the need to select appropriate IC containing PPG information. Therefore, wrong IC selection leads to inaccurate physiological parameter estimates in the scenarios, which include periodic motion. To counter the ordering problem, the BVP signal extraction was considered an undercomplete problem, and a U-ICA-based method named U-LMA was proposed, which aims to elevate it by only providing a single component consisting of PPG information.

Furthermore, it was also evident from the literature that the RGB color model possesses immense potential due to the relatively more robust pulsatile strength of the extracted rPPG signal. However, some challenges are associated with the utilization of the RGB model, i.e., susceptibility towards different types of motions and

illumination variations. Furthermore, illumination variation artifacts can be more severe in dark environments. To counter this, a deep learning-based architecture with a unique loss function was proposed to enhance dark videos (image by image) to an extent such that PPG information can be retrieved from enhanced videos. The architecture was named 2E1D-Net, which has two encoders and a decoder for enhancing a video based on a single reference image acquired in the ambient light environment.

2E1D-Net, in combination with U-LMA and ROR, facilitated the non-contact HR and SpO2 estimations in a dark environment using RGB spectra. Additionally, 2E1D-Net also expanded the ability of SOTA rPPG methods for physiological parameter estimations in a dark environment. Additionally, comprehensive experiments proved RGB spectra as a cost-effective and reliable surrogate for IR spectra without compromising the performance. Therefore, this study contributed significantly to the domain of non-contact physiological parameters monitoring in extremely dark environments, which facilitates remote health monitoring scenarios such as sleep, nighttime driving, etc.

7.2 LIMITATIONS

Although for each study, the limitations are presented in the respective chapters, this project also possessed certain limitations:

1. The project only focused on non-contact HR and SpO2 estimations but not other physiological parameters such as BP, BR, and skin temperature.
2. The developed methods were tested using various databases, but testing them for diseased individuals/patients (under abnormal ranges) requires collaboration with medical facilities and hospitals and requires intricate procedures, which is not feasible in a short span of four years
3. The deep learning-based rPPG methods were not tested during this project since combining the image enhancement with the rPPG-based method is computationally intensive, which was not possible due to limited computational resources.
4. The speculation about deep learning-based and rPPG signal extraction methods is that deep learning-based methods ensure better generalizability and are assumptions independent. This project could not test this speculation in dark environments.

7.3 FUTURE SCOPE

1. The future directions of extending the work of this project are to develop light-weight deep learning architectures that not only provide accurate rPPG signal extraction but could be easily combined with low-light enhancement methods for non-contact physiological parameter estimations in dark environments.
2. Further studies would aim to work on developing end-to-end rPPG signal extraction methods in dark environments and improving the developed methods to deploy in clinical environments such as ICUs, sleep monitoring labs, etc.
3. Additionally, a close collaboration with hospitals or other medical facilities would be encouraged to test the developed methods under real-time conditions such as non-contact sleep or home-based monitoring, nighttime driving health monitoring, etc.
4. Having explored the conventional non-contact estimation methods, future research will be focussed on exploring the DL for physiological measurements. Specifically, the future intends to deploy generating modelling to model rPPG signals by extracting PPG information from facial videos.
5. Other physiological parameter estimation, such as BP BR, will be a point of focus in future studies for expanding the ability of remote health monitoring to disease diagnostics.

References

[1] D. Shao, C. Liu, F. Tsow, Y. Yang, Z. Du, R. Iriya, H. Yu, and N. Tao. Noncontact monitoring of blood oxygen saturation using camera and dual-wavelength imaging system. *IEEE Transactions on Biomedical Engineering*, 63(6):1091–1098, 2016.

[2] B Anupama and K Ravishankar. Working mechanism and utility of pulse oximeter. *Int. J. Sport. Exerc. Heal. Res*, 2(2):111–113, 2018.

[3] Andreia Moço and Wim Verkruysse. Pulse oximetry based on photoplethysmography imaging with red and green light. *Journal of clinical monitoring and computing*, 35(1):123–133, 2021.

[4] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.

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

[6] VC Rideout and JEW Beneken. Parameter estimation applied to physiological systems. *Mathematics and Computers in Simulation*, 17(1):23–36, 1975.

[7] Wen Qi Mok, Wenru Wang, and Sok Ying Liaw. Vital signs monitoring to detect patient deterioration: An integrative literature review. *International journal of nursing practice*, 21:91–98, 2015.

[8] Karen D Fairchild, Douglas E Lake, John Kattwinkel, J Randall Moorman, David A Bateman, Philip G Grieve, Joseph R Isler, and Rakesh Sahni. Vital signs and their cross-correlation in sepsis and nec: a study of 1,065 very-low-birth-weight infants in two nicus. *Pediatric research*, 81(2):315–321, 2017.

[9] Travis J Moss, Douglas E Lake, J Forrest Calland, Kyle B Enfield, John B Delos, Karen D Fairchild, and J Randall Moorman. Signatures of subacute potentially catastrophic illness in the intensive care unit: model development and validation. *Critical care medicine*, 44(9):1639, 2016.

[10] M. Hu, F. Qian, D. Guo, X. Wang, L. He, and F. Ren. Eta-rppgnet: Effective time-domain attention network for remote heart rate measurement. *IEEE Transactions on Instrumentation and Measurement*, 70:1–12, 2021.

[11] Zhaodong Sun and Xiaobai Li. Contrast-phys: Unsupervised video-based remote physiological measurement via spatiotemporal contrast. In *European Conference on Computer Vision*, pages 492–510. Springer, 2022.

[12] Zijie Yue, Miaojing Shi, and Shuai Ding. Video-based remote physiological measurement via self-supervised learning. *arXiv preprint arXiv:2210.15401*, 2022.

[13] Armand Comas, Tim K Marks, Hassan Mansour, Suhas Lohit, Yechi Ma, and Xiaoming Liu. Turnip: Time-series u-net with recurrence for nir imaging ppg. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 309–313. IEEE, 2021.

[14] 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. Springer, 2018.

[15] Ewa M Nowara, Daniel McDuff, and Ashok Veeraraghavan. The benefit of distraction: Denoising camera-based physiological measurements using inverse attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4955–4964. IEEE, 2021.

[16] Shiika Kado, Yusuke Monno, Kazunori Yoshizaki, Masayuki Tanaka, and Masatoshi Okutomi. Spatial-spectral-temporal fusion for remote heart rate estimation. *IEEE Sensors Journal*, 20(19):11688–11697, 2020.

[17] Yen-Ting Huang, Yan-Tsung Peng, and Wen-Hung Liao. Enhancing object detection in the dark using u-net based restoration module. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–9. IEEE, 2019.

[18] Mohamed Abul Hassan, Aamir Saeed Malik, David Fofi, Naufal Saad, Babak Karasfi, Yasir Salih Ali, and Fabrice Meriaudeau. Heart rate estimation using facial video: A review. *Biomedical Signal Processing and Control*, 38:346–360, 2017.

[19] Jure Kranjec, S Beguš, G Geršak, and J Drnovšek. Non-contact heart rate and heart rate variability measurements: A review. *Biomedical signal processing and control*, 13:102–112, 2014.

[20] Clinton C Brown, Donald B Giddbon, and E Douglas Dean. Techniques of plethysmography. *Psychophysiology*, 1(3):253–266, 1965.

[21] K. Lin, D. Chen, and W. Tsai. Face-based heart rate signal decomposition and evaluation using multiple linear regression. *IEEE Sensors Journal*, 16(5):1351–1360, 2016.

[22] Ee-May Fong and Wan-Young Chung. A hygroscopic sensor electrode for fast stabilized non-contact ecg signal acquisition. *Sensors*, 15(8):19237–19250, 2015.

[23] Mark Van Gastel, Sander Stuijk, and Gerard De Haan. New principle for mea-

suring arterial blood oxygenation, enabling motion-robust remote monitoring. *Scientific reports*, 6(1):1–16, 2016.

[24] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1, 2007.

[25] Pratik Sahindrakar, Gerard de Haan, and Ihor Kirenko. Improving motion robustness of contact-less monitoring of heart rate using video analysis. *Technische Universiteit Eindhoven, Department of Mathematics and Computer Science*, 2011.

[26] Hanguang Xiao, Tianqi Liu, Yisha Sun, Yulin Li, Shiyi Zhao, and Alberto Avolio. Remote photoplethysmography for heart rate measurement: A review. *Biomedical Signal Processing and Control*, 88:105608, 2024.

[27] Ankit Gupta, Antonio G Ravelo-García, and Fernando Morgado-Dias. Recent advancements in deep learning-based remote photoplethysmography methods. *Data Fusion Techniques and Applications for Smart Healthcare*, pages 127–155, 2024.

[28] Juan Cheng, Xun Chen, Lingxi Xu, and Z Jane Wang. Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition. *IEEE journal of biomedical and health informatics*, 21(5):1422–1433, 2016.

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

[30] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

[31] Shifeng Zhang, Xiangyu Zhu, Zhen Lei, Hailin Shi, Xiaobo Wang, and Stan Z Li. S3fd: Single shot scale-invariant face detector. In *Proceedings of the IEEE international conference on computer vision*, pages 192–201. IEEE, 2017.

[32] 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.

[33] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Facial landmark detection by deep multi-task learning. In *European conference on computer vision*, pages 94–108. Springer, 2014.

[34] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, and Matthias Grundmann. Blazeface: Sub-millisecond neural face detection

on mobile gpus. *arXiv preprint arXiv:1907.05047*, 2019.

[35] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE, 2016.

[36] Matthias Dantone, Juergen Gall, Gabriele Fanelli, and Luc Van Gool. Real-time facial feature detection using conditional regression forests. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2578–2585. IEEE, 2012.

[37] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5203–5212. IEEE, 2020.

[38] Ali Sharifara, Mohd Shafry Mohd Rahim, and Yasaman Anisi. A general review of human face detection including a study of neural networks and haar feature-based cascade classifier in face detection. In *2014 International symposium on biometrics and security technologies (ISBAST)*, pages 73–78. IEEE, 2014.

[39] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520. IEEE, 2018.

[40] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency. Constrained local neural fields for robust facial landmark detection in the wild. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 354–361. IEEE, 2013.

[41] Tarek M Mahmoud. A new fast skin color detection technique. *World academy of science, engineering and technology*, 43:501–505, 2008.

[42] Anirudh Topiwala, Lidia Al-Zogbi, Thorsten Fleiter, and Axel Krieger. Adaptation and evaluation of deep learning techniques for skin segmentation on novel abdominal dataset. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 752–759. IEEE.

[43] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969. IEEE, 2017.

[44] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241.

Springer, 2015.

[45] Min Hu, Dong Guo, Mingxing Jiang, Fei Qian, Xiaohua Wang, and Fuji Ren. rppg-based heart rate estimation using spatial-temporal attention network. *IEEE Transactions on Cognitive and Developmental Systems*, 14(4):1630–1641, 2022.

[46] Xin Liu, Brian Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5008–5017. IEEE, 2023.

[47] Rencheng Song, Senle Zhang, Chang Li, Yunfei Zhang, Juan Cheng, and Xun Chen. Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks. *IEEE Transactions on Instrumentation and Measurement*, 69(10):7411–7421, 2020.

[48] Yuhang Dong, Gongping Yang, and Yilong Yin. Drnet: Decomposition and reconstruction network for remote physiological measurement. *arXiv preprint arXiv:2206.05687*, 2022.

[49] Min Hu, Dong Guo, Xiaohua Wang, Peng Ge, and Qian Chu. A novel spatial-temporal convolutional neural network for remote photoplethysmography. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6. IEEE, 2019.

[50] Si-Qi Liu and Pong C Yuen. A general remote photoplethysmography estimator with spatiotemporal convolutional network. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 481–488. IEEE, 2020.

[51] Hao Lu, Hu Han, and S Kevin Zhou. Dual-gan: Joint bvp and noise modeling for remote physiological measurement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12404–12413. IEEE, 2021.

[52] Olga Perepelkina, Mikhail Artemyev, Marina Churikova, and Mikhail Grinenko. Hearttrack: Convolutional neural network for remote video-based heart rate monitoring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 288–289. IEEE, 2020.

[53] Yuzhuo Ren, Braeden Syrnyk, and Niranjan Avadhanam. Dual attention network for heart rate and respiratory rate estimation. In *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2021.

[54] Jeremy Speth, Nathan Vance, Patrick Flynn, Kevin Bowyer, and Adam Cza-

jka. Remote pulse estimation in the presence of face masks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2086–2095. IEEE, 2021.

[55] Yun-Yun Tsou, Yi-An Lee, Chiou-Ting Hsu, and Shang-Hung Chang. Siamese-rppg network: Remote photoplethysmography signal estimation from face videos. In *Proceedings of the 35th annual ACM symposium on applied computing*, pages 2066–2073. Association for Computing Machinery, 2020.

[56] Sricharan Vijayarangan, R Vignesh, Balamurali Murugesan, SP Preejith, Jayaraj Joseph, and Mohansankar Sivaprakasam. Rpnet: A deep learning approach for robust r peak detection in noisy ecg. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 345–348. IEEE, 2020.

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

[58] Yu Rong, Panagiotis C Theofanopoulos, Georgios C Trichopoulos, and Daniel W Bliss. A new principle of pulse detection based on terahertz wave plethysmography. *Scientific reports*, 12(1):1–15, 2022.

[59] 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.

[60] G. R. Tsouri, S. Kyal, S. Dianat, and L. K. Mestha. Constrained independent component analysis approach to nonobtrusive pulse rate measurements. *J Biomed Opt*, 17(7):077011, 2012.

[61] Richard Macwan, Yannick Benezeth, and Alamin Mansouri. Heart rate estimation using remote photoplethysmography with multi-objective optimization. *Biomedical Signal Processing and Control*, 49:24–33, 2019.

[62] Huan Qi, Zhenyu Guo, Xun Chen, Zhiqi Shen, and Z. Jane Wang. Video-based human heart rate measurement using joint blind source separation. *Biomedical Signal Processing and Control*, 31:309–320, 2017.

[63] 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.

[64] Q. Tran, S. Su, W. Sun, and M. Tran. Adaptive pulsatile plane for robust noncontact heart rate monitoring. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1–13, 2019.

[65] Hongwei Yue, Xiaorong Li, Ken Cai, Huazhou Chen, Shufen Liang, Tianlei Wang, and Wenhua Huang. Non-contact heart rate detection by combining empirical mode decomposition and permutation entropy under non-cooperative face shake. *Neurocomputing*, 392:142–152, 2020.

[66] D. Chen, J. Wang, K. Lin, H. Chang, H. Wu, Y. Chen, and S. Lee. Image sensor-based heart rate evaluation from face reflectance using hilbert–huang transform. *IEEE Sensors Journal*, 15(1):618–627, 2015.

[67] Rencheng Song, Senle Zhang, Juan Cheng, Chang Li, and Xun Chen. New insights on super-high resolution for video-based heart rate estimation with a semi-blind source separation method. *Computers in Biology and Medicine*, 116:103535–103543, 2020.

[68] Yuzhong Zhang, Zhe Dong, Kezun Zhang, Shuangbao Shu, Fucheng Lu, and Jingjing Chen. Illumination variation-resistant video-based heart rate monitoring using lab color space. *Optics and Lasers in Engineering*, 136:106328, 2021.

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

[70] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. Robust heart rate from fitness videos. *Physiological measurement*, 38(6):1023, 2017.

[71] Changchen Zhao, Chun-Liang Lin, Weihai Chen, Ming-Kun Chen, and Jianhua Wang. Visual heart rate estimation and negative feedback control for fitness exercise. *Biomedical Signal Processing and Control*, 56:101680, 2020.

[72] JongSong Ryu, SunChol Hong, Shili Liang, SinIl Pak, Qingyue Chen, and Shifeng Yan. A measurement of illumination variation-resistant noncontact heart rate based on the combination of singular spectrum analysis and sub-band method. *Computer Methods and Programs in Biomedicine*, 200:105824, 2021.

[73] Rencheng Song, Jiji Li, Minda Wang, Juan Cheng, Chang Li, and Xun Chen. Remote photoplethysmography with an eemd-mcca method robust against spatially uneven illuminations. *IEEE Sensors Journal*, 21(12):13484–13494, 2021.

[74] B. F. Wu, Y. W. Chu, P. W. Huang, and M. L. Chung. Neural network based luminance variation resistant remote-photoplethysmography for driver's heart rate monitoring. *IEEE Access*, 7:57210–57225, 2019.

[75] Y. Qiu, Y. Liu, J. Arteaga-Falconi, H. Dong, and A. E. Saddik. Evm-cnn: Real-

time contactless heart rate estimation from facial video. *IEEE Transactions on Multimedia*, 21(7):1778–1787, 2019.

[76] Frédéric Bousefsaf, Alain Pruski, and Choubeila Maaoui. 3d convolutional neural networks for remote pulse rate measurement and mapping from facial video. *Applied Sciences*, 9(20):4364, 2019.

[77] X. Niu, S. Shan, H. Han, and X. Chen. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Transactions on Image Processing*, 29:2409–2423, 2020.

[78] Gee-Sern Jison Hsu, Rui-Cang Xie, ArulMurugan Ambikapathi, and Kae-Jy Chou. A deep learning framework for heart rate estimation from facial videos. *Neurocomputing*, 417:155–166, 2020.

[79] Zitong Yu, Xiaobai Li, Xuesong Niu, Jingang Shi, and Guoying Zhao. Autohr: A strong end-to-end baseline for remote heart rate measurement with neural searching. *IEEE Signal Processing Letters*, 27:1245–1249, 2020.

[80] M. Hu, F. Qian, X. Wang, L. He, D. Guo, and F. Ren. Robust heart rate estimation with spatial-temporal attention network from facial videos. *IEEE Transactions on Cognitive and Developmental Systems*, pages 1–1, 2021.

[81] Xuenan Liu, Xuezhi Yang, and Xiaobai Li. Hrunet: Assessing uncertainty in heart rates measured from facial videos. *IEEE Journal of Biomedical and Health Informatics*, 2024.

[82] Yassine Ouzar, Djamaleddine Djeldjli, Frédéric Bousefsaf, and Choubeila Maaoui. X-ippgnet: A novel one stage deep learning architecture based on depthwise separable convolutions for video-based pulse rate estimation. *Computers in Biology and Medicine*, 154:106592, 2023.

[83] Yu Zhao, Bochao Zou, Fan Yang, Lin Lu, Abdelkader Nasreddine Belkacem, and Chao Chen. Video-based physiological measurement using 3d central difference convolution attention network. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–6. IEEE, 2021.

[84] 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.

[85] Alessandro R Guazzi, Mauricio Villarroel, Joao Jorge, Jonathan Daly, Matthew C Frise, Peter A Robbins, and Lionel Tarassenko. Non-contact measurement of oxygen saturation with an rgb camera. *Biomedical optics express*, 6(9):3320, 2015.

[86] Alessandra de Fátima Galvão Rosa and Roberto Cesar Betini. Noncontact

spo 2 measurement using eulerian video magnification. *IEEE Transactions on Instrumentation and Measurement*, 69(5):2120–2130, 2019.

[87] U. Bal. Non-contact estimation of heart rate and oxygen saturation using ambient light. *Biomed Opt Express*, 6(1):86–97, 2015.

[88] Lionel Tarassenko, Mauricio Villarroel, Alessandro Guazzi, João Jorge, DA Clifton, and Chris Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological measurement*, 35(5):807, 2014.

[89] Dangdang Shao, Chenbin Liu, Francis Tsow, Yuting Yang, Zijian Du, Rafael Iriya, Hui Yu, and Nongjian Tao. Noncontact monitoring of blood oxygen saturation using camera and dual-wavelength imaging system. *IEEE Transactions on Biomedical Engineering*, 63(6):1091–1098, 2015.

[90] Lingqin Kong, Yuejin Zhao, Liquan Dong, Yiyun Jian, Xiaoli Jin, Bing Li, Yun Feng, Ming Liu, Xiaohua Liu, and Hong Wu. Non-contact detection of oxygen saturation based on visible light imaging device using ambient light. *Optics express*, 21(15):17464–17471, 2013.

[91] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and PRISMA Group*. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, 151(4):264–269, 2009.

[92] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, and Sue E Brennan. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*, 88:105906, 2021.

[93] David Moher, Kenneth F Schulz, Iveta Simera, and Douglas G Altman. Guidance for developers of health research reporting guidelines. *PLoS medicine*, 7(2):e1000217, 2010.

[94] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. Vipl-hr: A multi-modal database for pulse estimation from less-constrained face video. In *Asian conference on computer vision*, pages 562–576. Springer, 2018.

[95] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90, 2019.

[96] Guillaume Heusch, André Anjos, and Sébastien Marcel. A reproducible study on remote heart rate measurement. *arXiv preprint arXiv:1709.00962*, 2017.

[97] Aapo Hyvärinen, Jarmo Hurri, and Patrik O. Hoyer. *Independent Component*

Analysis, pages 151–175. Springer, 2009.

[98] James V Stone. Independent component analysis: an introduction. *Trends in cognitive sciences*, 6(2):59–64, 2002.

[99] Tianfeng Chai and Roland R Draxler. Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250, 2014.

[100] Association for the Advancement of Medical Instrumentation. Cardiac monitors, heart rate meters, and alarms. *American National Standard (ANSI/AAMI EC13: 2002) Arlington, VA*, pages 1–87, 2002.

[101] Ankit Gupta, Antonio G Ravelo-García, and Fernando Morgado Dias. Availability and performance of face based non-contact methods for heart rate and oxygen saturation estimations: A systematic review. *Computer Methods and Programs in Biomedicine*, page 106771, 2022.

[102] Ioannis Pavlidis, Jonathan Dowdall, Nanfei Sun, Colin Puri, Jin Fei, and Marc Garbey. Interacting with human physiology. *Computer Vision and Image Understanding*, 108(1):150–170, 2007.

[103] Koen M van der Kooij and Marnix Naber. An open-source remote heart rate imaging method with practical apparatus and algorithms. *Behavior research methods*, 51:2106–2119, 2019.

[104] Wim Verkruysse, Marek Bartula, Erik Bresch, Mukul Rocque, Mohammed Meftah, and Ihor Kirenko. Calibration of contactless pulse oximetry. *Anesthesia and analgesia*, 124(1):136, 2017.

[105] Behrouz Jafari and Vahid Mohsenin. Polysomnography. *Clinics in chest medicine*, 31(2):287–297, 2010.

[106] Andrea Nicolò, Carlo Massaroni, Emiliano Schena, and Massimo Sacchetti. The importance of respiratory rate monitoring: From healthcare to sport and exercise. *Sensors*, 20(21):6396, 2020.

[107] Chen Wang, Thierry Pun, and Guillaume Chanel. A comparative survey of methods for remote heart rate detection from frontal face videos. *Frontiers in bioengineering and biotechnology*, 6:33, 2018.

[108] Yu Sun and Nitish Thakor. Photoplethysmography revisited: from contact to noncontact, from point to imaging. *IEEE transactions on biomedical engineering*, 63(3):463–477, 2015.

[109] Mirae Harford, Jacqueline Catherall, Stephen Gerry, John Duncan Young, and P Watkinson. Availability and performance of image-based, non-contact methods of monitoring heart rate, blood pressure, respiratory rate, and oxygen

saturation: a systematic review. *Physiological measurement*, 40(6):06TR01, 2019.

[110] Yi Zhang, Zhihao Chen, and Hwan Ing Hee. Noninvasive measurement of heart rate and respiratory rate for perioperative infants. *Journal of Lightwave Technology*, 37(11):2807–2814, 2019.

[111] J. Cheng, P. Wang, R. Song, Y. Liu, C. Li, Y. Liu, and X. Chen. Remote heart rate measurement from near-infrared videos based on joint blind source separation with delay-coordinate transformation. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021.

[112] X. Yu, T. Laurentius, C. Bollheimer, S. Leonhardt, and C. H. Antink. Noncontact monitoring of heart rate and heart rate variability in geriatric patients using photoplethysmography imaging. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1781–1792, 2021.

[113] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control*, 8(6):568–574, 2013.

[114] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri. A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 6450–6459. IEEE, 2018.

[115] Otkrist Gupta, Dan McDuff, and Ramesh Raskar. Real-time physiological measurement and visualization using a synchronized multi-camera system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 46–53. IEEE, 2016.

[116] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal. Distanceppg: Robust non-contact vital signs monitoring using a camera. *Biomedical optics express*, 6(5):1565–1588, 2015.

[117] 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.

[118] Bing Wei, Xuan He, Chao Zhang, and Xiaopei Wu. Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions. *Biomedical engineering online*, 16(1):1–21, 2017.

[119] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, and Guoying Zhao. Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In *Proceedings of the*

IEEE/CVF International Conference on Computer Vision, pages 151–160. IEEE, 2019.

[120] Y. Lin and Y. Lin. Step count and pulse rate detection based on the contactless image measurement method. *IEEE Transactions on Multimedia*, 20(8):2223–2231, 2018.

[121] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen. Remote heart rate measurement from face videos under realistic situations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4264–4271. IEEE, 2019.

[122] Chao Zhang, Xiaopei Wu, Lei Zhang, Xuan He, and Zhao Lv. Simultaneous detection of blink and heart rate using multi-channel ica from smart phone videos. *Biomedical Signal Processing and Control*, 33:189–200, 2017.

[123] A. Woyczyk, V. Fleischhauer, and S. Zaunseder. Adaptive gaussian mixture model driven level set segmentation for remote pulse rate detection. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1361–1372, 2021.

[124] Puneet Gupta, Brojeshwar Bhowmick, and Arpan Pal. Mombat: Heart rate monitoring from face video using pulse modeling and bayesian tracking. *Computers in Biology and Medicine*, 121:103813, 2020.

[125] J. John, S. Krishna, and R. R. Galigekere. Automatic and adaptive signal- and background-rois with analytic-representation-based processing for robust webcam-based heart-rate estimation. *IEEE Access*, 8:34728–34736, 2020.

[126] Juan Cheng, Ping Wang, Rencheng Song, Yu Liu, Chang Li, Yong Liu, and Xun Chen. Remote heart rate measurement from near-infrared videos based on joint blind source separation with delay-coordinate transformation. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2020.

[127] Engui Fan. Extended tanh-function method and its applications to nonlinear equations. *Physics Letters A*, 277(4-5):212–218, 2000.

[128] Gerard De Haan and Arno Van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913, 2014.

[129] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. Remote assessment of the heart rate variability to detect mental stress. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 348–351. IEEE, 2013.

[130] Ethan B Blackford and Justin R Estepp. Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography. In *Medical Imaging 2015: Biomedical Applications in Molecular,*

Structural, and Functional Imaging, volume 9417, pages 639–652. SPIE, 2015.

[131] Ankit Gupta, Antonio G. Ravelo-García, and Fernando Morgado Dias. A motion and illumination resistant non-contact method using undercomplete independent component analysis and levenberg-marquardt algorithm. *IEEE Journal of Biomedical and Health Informatics*, 26(10):4837–4848, 2022.

[132] Stéphane Cook, Mario Togni, Marcus C Schaub, Peter Wenaweser, and Otto M Hess. High heart rate: a cardiovascular risk factor? *European heart journal*, 27(20):2387–2393, 2006.

[133] Tomas Ysehak Abay and Panayiotis A Kyriacou. Reflectance photoplethysmography as noninvasive monitoring of tissue blood perfusion. *IEEE Transactions on Biomedical Engineering*, 62(9):2187–2195, 2015.

[134] M Kedadouche, M Thomas, and AJMS Tahan. A comparative study between empirical wavelet transforms and empirical mode decomposition methods: Application to bearing defect diagnosis. *Mechanical Systems and Signal Processing*, 81:88–107, 2016.

[135] Iman Rahmansyah Tayibnapis, Yeon-Mo Yang, and Ki Moo Lim. Blood volume pulse extraction for non-contact heart rate measurement by digital camera using singular value decomposition and burg algorithm. *Energies*, 11(5):1076, 2018.

[136] John Porrill and James V Stone. Undercomplete independent component analysis for signal separation and dimension reduction. *report, Citeseer*, 1998.

[137] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

[138] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[139] Aiyou Chen. Fast kernel density independent component analysis. In *Independent Component Analysis and Blind Signal Separation: 6th International Conference, ICA 2006, Charleston, SC, USA, March 5-8, 2006. Proceedings 6*, pages 24–31. Springer.

[140] Daniel McDuff, Sarah Gontarek, and Rosalind W Picard. Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Transactions on Biomedical Engineering*, 61(10):2593–2601, 2014.

[141] Yonggang Yan, Xiang Ma, Lifeng Yao, and Jianfei Ouyang. Noncontact measurement of heart rate using facial video illuminated under natural light and signal weighted analysis. *Bio-medical materials and engineering*, 26(s1):S903–S909, 2015.

[142] Thomas Pursche, Jarek Krajewski, and Reinhard Moeller. Video-based heart rate measurement from human faces. In *2012 IEEE international conference on consumer electronics (ICCE)*, pages 544–545. IEEE, 2012.

[143] Mika P Tarvainen, Perttu O Ranta-Aho, and Pasi A Karjalainen. An advanced detrending method with application to hrv analysis. *IEEE transactions on biomedical engineering*, 49(2):172–175, 2002.

[144] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

[145] Mark K Transtrum and James P Sethna. Improvements to the levenberg-marquardt algorithm for nonlinear least-squares minimization. *arXiv preprint arXiv:1201.5885*, 2012.

[146] S Amarai, A Cichoki, and TP Chen. A new learning algorithm for blind source separation. *Advances in Neural Information Processing Systems*, 8:757–763, 1996.

[147] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis: proceedings of the biennial Conference held at Dundee, June 28–July 1, 1977*, pages 105–116. Springer.

[148] Ken Cai, Hongwei Yue, Bohan Li, Weitong Chen, and Wenhua Huang. Combining chrominance features and fast ica for noncontact imaging photoplethysmography. *IEEE Access*, 8:50171–50179, 2020.

[149] Kun Zheng, Kangyi Ci, Jinling Cui, Jiangping Kong, and Jing Zhou. Noncontact heart rate detection when face information is missing during online learning. *Sensors*, 20(24):7021, 2020.

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

[151] Daniel McDuff and Ethan Blackford. iphys: An open non-contact imaging-based physiological measurement toolbox. In *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 6521–6524. IEEE, 2019.

[152] Joaquim Comas, Adria Ruiz, and Federico Sukno. Efficient remote photoplethysmography with temporal derivative modules and time-shift invariant loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2182–2191. IEEE, 2022.

[153] Daniel J McDuff, Justin R Estepp, Alyssa M Piasecki, and Ethan B Blackford. A survey of remote optical photoplethysmographic imaging methods. In *2015 37th annual international conference of the IEEE engineering in medicine and*

biology society (EMBC), pages 6398–6404. IEEE, 2015.

[154] Akito Tohma, Maho Nishikawa, Takuya Hashimoto, Yoichi Yamazaki, and Guanghao Sun. Evaluation of remote photoplethysmography measurement conditions toward telemedicine applications. *Sensors*, 21(24):8357, 2021.

[155] Lin Xi, Weihai Chen, Changchen Zhao, Xingming Wu, and Jianhua Wang. Image enhancement for remote photoplethysmography in a low-light environment. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 1–7. IEEE, 2020.

[156] Sebastian Zaunseder, Andreas Heinke, Alexander Trumpp, and Hagen Malberg. Heart beat detection and analysis from videos. In *2014 IEEE 34th International Scientific Conference on Electronics and Nanotechnology (ELNANO)*, pages 286–290. IEEE, 2014.

[157] Jingjing Hu, Yunze He, Jie Liu, Min He, and Wenjin Wang. Illumination robust heart-rate extraction from single-wavelength infrared camera using spatial-channel expansion. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3896–3899. IEEE, 2019.

[158] Ewa M Nowara, Tim K Marks, Hassan Mansour, and Ashok Veeraraghavan. Near-infrared imaging photoplethysmography during driving. *IEEE Transactions on Intelligent Transportation Systems*, 23(4):3589–3600, 2020.

[159] Ismoil Odinaev, Jing Wei Chin, Kin Ho Luo, Zhang Ke, Richard HY So, and Kwan Long Wong. Optimizing camera exposure control settings for remote vital sign measurements in low-light environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6085–6092. IEEE, 2023.

[160] Shutao Chen, Sui Kei Ho, Jing Wei Chin, Kin Ho Luo, Tsz Tai Chan, Richard HY So, and Kwan Long Wong. Deep learning-based image enhancement for robust remote photoplethysmography in various illumination scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6076–6084. IEEE, 2023.

[161] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646. IEEE, 2022.

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

[163] Etta D Pisano, Shuquan Zong, Bradley M Hemminger, Marla DeLuca, R Eugene Johnston, Keith Muller, M Patricia Braeuning, and Stephen M Pizer. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, 11:193–200, 1998.

[164] Xu Guan, Su Jian, Pan Hongda, Zhang Zhiguo, and Gong Haibin. An image enhancement method based on gamma correction. In *2009 Second international symposium on computational intelligence and design*, volume 1, pages 60–63. IEEE, 2009.

[165] Krishna Gopal Dhal, Arunita Das, Swarnajit Ray, Jorge Gálvez, and Sanjoy Das. Histogram equalization variants as optimization problems: a review. *Archives of Computational Methods in Engineering*, 28:1471–1496, 2021.

[166] Zhi-Guo Wang, Zhi-Hu Liang, and Chun-Liang Liu. A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for pdp. *Displays*, 30(3):133–139, 2009.

[167] Shanto Rahman, Md Mostafijur Rahman, Mohammad Abdullah-Al-Wadud, Golam Dastegir Al-Quaderi, and Mohammad Shoyaib. An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, 2016(1):1–13, 2016.

[168] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3):1032–1041, 2012.

[169] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[170] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997.

[171] Zia-ur Rahman, Daniel J Jobson, and Glenn A Woodell. Multi-scale retinex for color image enhancement. In *Proceedings of 3rd IEEE international conference on image processing*, volume 3, pages 1003–1006. IEEE, 1996.

[172] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9):3538–3548, 2013.

[173] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015.

[174] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding.

A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2782–2790. IEEE, 2016.

[175] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017.

[176] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4. BMVC, 2018.

[177] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13106–13113. AAAI, 2020.

[178] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang. Lightennet: A convolutional neural network for weakly illuminated image enhancement. *Pattern recognition letters*, 104:15–22, 2018.

[179] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021.

[180] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22252–22261. IEEE, 2023.

[181] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021.

[182] Yingying Meng, Deqiang Kong, Zhenfeng Zhu, and Yao Zhao. From night to day: Gans based low quality image enhancement. *Neural Processing Letters*, 50:799–814, 2019.

[183] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 3277–3285. IEEE, 2017.

[184] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Wespe: weakly supervised photo enhancer for digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern*

Recognition Workshops, pages 691–700. IEEE, 2018.

[185] Shangchen Zhou, Chongyi Li, and Chen Change Loy. Lednet: Joint low-light enhancement and deblurring in the dark. In *European Conference on Computer Vision*, pages 573–589. Springer, 2022.

[186] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12918–12927. IEEE, October 2023.

[187] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4225–4238, 2021.

[188] Zhichao Chen, Jie Yang, and Chonglian Yang. Brightsightnet: A lightweight progressive low-light image enhancement network and its application in "rainbow" maglev train. *Journal of King Saud University-Computer and Information Sciences*, 35(10):101814, 2023.

[189] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1780–1789. IEEE, 2020.

[190] Wenjin Wang, Luc Vosters, and Albertus C den Brinker. Continuous-spectrum infrared illuminator for camera-ppg in darkness. *Sensors*, 20(11):3044, 2020.

[191] Kaiwen Guo, Tianqu Zhai, Manoj H Purushothama, Alexander Dobre, Shawn Meah, Elton Pashollari, Aabhaas Vaish, Carl DeWilde, and Mohammed N Islam. Contactless vital sign monitoring system for in-vehicle driver monitoring using a near-infrared time-of-flight camera. *Applied Sciences*, 12(9):4416, 2022.

[192] Mark van Gastel, Benoît Balmaekers, Sidarto Bambang Oetomo, and Wim Verkruysse. Near-continuous non-contact cardiac pulse monitoring in a neonatal intensive care unit in near darkness. In *Optical diagnostics and sensing XVIII: Toward point-of-care diagnostics*, volume 10501, pages 230–238. SPIE, 2018.

[193] Wen-Nung Lie, Dao-Quang Le, Chun-Yu Lai, and Yu-Shin Fang. Heart rate estimation from facial image sequences of a dual-modality rgb-nir camera. *Sensors*, 23(13):6079, 2023.

[194] Soyeon Park, Bo-Kyeong Kim, and Suh-Yeon Dong. Self-supervised rgb-nir fusion video vision transformer framework for rppg estimation. *IEEE Transactions on Instrumentation and Measurement*, 71:1–10, 2022.

[195] Haipeng Wang and Shuai Zhang. Non-contact human respiratory rate mea-

surement under dark environments by low-light video enhancement. *Biomedical Signal Processing and Control*, 85:104874, 2023.

[196] IR LED. A technology overview of the nellcor® oximax® pulse oximetry system.

[197] Yusuke Akamatsu, Yoshifumi Onishi, and Hitoshi Imaoka. Heart rate and oxygen saturation estimation from facial video with multimodal physiological data generation. In *ICASSP 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1111–1115. IEEE, 2022.

[198] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller. Transferring color to greyscale images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 277–280. Association for Computing Machinery, 2002.

[199] Chun-Hong Cheng, Kwan-Long Wong, Jing-Wei Chin, Tsz-Tai Chan, and Richard HY So. Deep learning methods for remote heart rate measurement: a review and future research agenda. *Sensors*, 21(18):6296, 2021.

[200] Aoxin Ni, Arian Azarang, and Nasser Kehtarnavaz. A review of deep learning-based contactless heart rate measurement methods. *Sensors*, 21(11):3719, 2021.

[201] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016.

[202] Dominique Brunet, Edward R Vrscay, and Zhou Wang. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 21(4):1488–1499, 2011.

[203] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[204] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595. IEEE, 2018.

[205] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778. IEEE, 2016.

[206] Gabriella Casalino, Giovanna Castellano, and Gianluca Zaza. Evaluating the robustness of a contact-less mhealth solution for personal and remote monitoring of blood oxygen saturation. *Journal of Ambient Intelligence and Hu-*

manized Computing, pages 1–10, 2022.

APPENDIX

DATA PROTECTION COMMITTEEE DOCUMENT

UNIVERSIDADE da MADEIRA

STATEMENT OF COMPLIANCE

As the Data Protection Officer of the University of Madeira, and after analysis by the Data Protection Team of this institution, it is our opinion that the Research Project "***Collection of Facial Image and Video Samples of People***", complies with the General Data Protection Regulation.

Funchal, May 6, 2021

Data Protection Officer

Filipa Isabel Sousa Mota

ETHICAL COMMITTEE APPROVAL DOCUMENT

UNIVERSIDADE da MADEIRA
COMISSÃO DE ÉTICA

PROPOSTA DE PARECER Nº 1/CEUMA/2021, DE 21 DE JANEIRO

Sobre o pedido de reapreciação pela Comissão de Ética da UMa do protocolo de investigação do estudo: *"Collection of Facial Image and Video Samples of People"*, no âmbito de doutoramento em Engenharia informática da Universidade da Madeira submetido pelo Professor Doutor Fernando Manuel Rosmaninho Morgado Ferrão Dias, Principal Investigador, Professor Associado da Faculdade das Ciências Exatas e da Engenharia e Madeira-Interactive Technologies Institute, Universidade da Madeira.

Na reformulação dos documentos que integram o pedido, o requerente apresenta

1. O formulário de pedido de parecer devidamente preenchido, onde indica os elementos anteriormente solicitados:
   - Membros que integram a equipa de investigação e as Instituições envolvidas;
   - Duração temporal, o local onde decorrerá o estudo e a recolha de informação;
   - Procedimentos, instrumentos de recolha dos dados fisiológicos, de imagem facial e de vídeo, descrição do *debriefing* pós-experimental;
   -Características dos participantes, explicitando a idade, a capacidade para prestar consentimento e a justificação da inclusão de migrantes;
   - Processo de recrutamento, triagem, amostragem e critérios de inclusão e de exclusão;
   - Riscos e benefícios, compensação e custos;
   - As condições de anonimização e a garantia de sigilo; que os dados recolhidos poderão serão potencialmente usados como uma base de dados de imagem e vídeo a serem compartilhados entre os investigadores e a comunidade científica que trabalha no domínio, para o propósito específico; quem tem acesso aos mesmos, quanto tempo serão guardados, a forma de destruição e as medidas de segurança;
   - Formas previstas de divulgação e disseminação de resultados.

2. O documento de Consentimento Informado, em língua portuguesa e inglesa. Consideramos que este documento deve também incluir referência explícita sobre a compensação e custos e que, na versão em português, o ponto j. necessita de melhor clarificação e adequação de linguagem para os participantes, particularmente no que concerne à autorização para a partilha dos dados registados entre os investigadores, bem como a garantia de destruição. Recomendamos revisão geral da redação em Português.

Aguarda envio de parecer da Comissão de Proteção de Dados da Universidade da Madeira, solicitado, por mail, pelo Presidente da CEUMa a 6 de janeiro de 2021.

A CEUMa deliberou emitir um parecer preliminar "favorável", condicionado à reformulação do documento de Consentimento Informado, com a inclusão dos

UNIVERSIDADE da MADEIRA
COMISSÃO DE ÉTICA

elementos supramencionados, e à entrega do parecer com aprovação da Comissão de Proteção de Dados da Universidade da Madeira.

Aprovado por unanimidade, em reunião de 21 de janeiro de 2021.

O Presidente da CEUMa
Professor Doutor Ivo da Piedade Álvares Furtado

APPROVED INFORMED CONSENT FORMS FROM THE DATA PROTECTION COMMITTEE (PORTUGUESE VERSION)

Consentimento Informado, Esclarecido e Livre para Participação em estudos de Investigação

Identificação do Investigador Prof. Fernando Manuel Rosmaninho Morgado Ferrão Dias
Professor Associado
Faculdade de Ciencias Exatas e Engenharia
Universidade da Madeira

Título do estudo: Collection of Facial Image and Video Samples of People.

a) **Enquadramento**
O estudo tem por objetivo a recolha de amostras de vídeo e fotografia, usando a câmara de um computador pessoal e a extração de informação de batimento cardíaco e a saturação de oxigénio, usando um oxímetro de pulso, para análise do sono e estudo de problemas relacionados.

b) ***Explicação do estudo*** objetivo do estudo Os objetivos de recolha das amostras para este estudo são:
   (a) Medir o batimento cardíaco de forma não invasiva
   (b) Analisar a qualidade do sono usando os parâmetros fisiológicos recolhidos.

c) ***recolha de dados de processo ou outro método***
Este estudo tem por objetivo analisar amostras de vídeo e fotografia para estimar parâmetros fisiológicos tais como o batimento cardíaco, níveis de saturação de oxigénio no sangue, que ajudam a comunidade médica a diagnosticar de forma não invasiva pacientes que sofrem de perturbações do sono.

No decurso do estudo será recolhida uma imagem sob condições normais de iluminação e amostras de vídeo sob diferentes condições de iluminação. São igualmente recolhidos através de oxímetro de pulso informações do batimento cardíaco e da saturação de oxigénio no sangue (SpO2%)) para validar os métodos de extração destas informações a partir do vídeo.

d) ***mencionar local onde o/a investigador/a se encontra com o/a participante***
O participante tem de visitar as instalações do M-ITI para participar no estudo.

e) ***quantas vezes e durante quanto tempo aproximadamente***
Considerando o esclarecimento de dúvidas, a explicação da utilização e manutenção dos dados, bem como das condições de recolha dos mesmos e a recolha propriamente dita, são necessários 10 a 15 minutos.

f) ***garantir destruição de gravações (áudio ou vídeo) num determinado prazo***
As imagens e vídeos serão guardados por tempo indefinido. Será criada uma base de dados para investigação sem dados pessoas. Os participantes serão informados que a base de dados será mantida por tempo indefinido e que pode ser usada para outros estudos e por outros investigadores e não apenas no estudo atual.

O participantes poderão pedir para os seus dados serem retirados da base de dados em qualquer altura.

Caso os dados sejam cedidos a outros investigadores, para além do estudo atual, estes terão de respeitar o definido no RGPD.

g) ***referir se há ou não pagamento de deslocações ou contrapartidas***
Não existe pagamento de deslocações nem contrapartidas por participar no estudo.

h) ***informar quem financia o estudo (o/a investigador/a ou outrem)***
Este projeto não é diretamente financiado por nenhuma fonte de financiamento.

i) ***garantir confidencialidade e uso exclusivo dos dados recolhidos para o presente estudo***
Este estudo necessita da recolha de informação pessoal dos participantes. A informação pessoal recolhida, tal como o nome e a identidade, será mantida de forma confidencial e não será partilhada com ninguém em nenhuma circunstância. Os restantes dados podem ser partilhados entre investigadores exclusivamente para propósitos de investigação. Cada participante será identificado por um código único e a informação pessoal não será partilhada.

A proteção dos dados pessoais e a livre circulação e eliminação dos dados seguirá estritamente o regulamento 2016/679 do parlamento europeu e do concelho de 27 de abril de 2016 sobre a proteção das pessoas naturais e o processamento de dados pessoais e a livre circulação destes dados, seguindo a diretiva 95/46/EC (General Data Protection Regulation (GDPR)) e a Lei 58/2019, de 8 de agosto para a implementação do GDPR.

j) ***agradecimentos e identificar o/a investigador/a – nome, profissão, local de trabalho, contacto telefónico, endereço eletrónico – e a pessoa que pede o consentimento, se for diferente***
O processo será organizado e gerido pelo investigador Ankit Gupta, cujos contac-

tos são indicados. Ele será responsável por solicitar aos participantes que assinem o documentos de consentimento informado, após o qual será feita a recolha dos dados.

k) ***Ankit Gupta***

PhD Student (Informatics Engineering)
Interactive Technologies Institute
University of Madeira
Contact: +351 926045472
Email:ankit.gupta@iti.larsys.pt

Por favor, leia com atenção esta informação. Se achar que algo está incorreto ou que não está claro, não hesite em solicitar mais informações.
Se concorda com a proposta que lhe foi feita, queira assinar este documento.

assinatura do participante: .....

APPROVED INFORMED CONSENT FORMS FROM THE ETHICAL COMMITTEE (PORTUGUESE VERSION)

Consentimento informado para a recolha de imagens faciais e amostras de vídeo

a) **Identificação do investigador**

Prof. Fernando Manuel Rosmaninho Morgado Ferrão Dias
Professor associado
Faculdade das ciências exatas e da engenharia
Universidade da Madeira

Ankit Gupta
Aluno de doutoramento (engenharia informática)
Interactive Technologies Institute
Universidade da Madeira

b) **Contacto do investigador**

Email
morgado@uma.pt
ankit.gupta@iti.larsys.pt
Contacto(s):
+351 291-705280
+351 926 045 472

c) **Âmbito do estudo**

Este estudo destina-se a analisar amostras de vídeo de faces para estimar múltiplos parâmetros clínicos como a frequência cardíaca e os níveis de oxigenação do sangue, que auxiliam os médicos a diagnosticar as pessoas que sofrem de distúrbios relacionados com o sono, de forma não invasiva. Neste estudo, uma imagem facial será coletada em condições de luminosidade normal e múltiplas amostras de vídeo serão recolhidas em diferentes condições de luminosidade. O estudo também envolve a aquisição de parâmetros fisiológicos (frequência cardíaca e nível de oxigenação do sangue (SpO2%)) de forma a validar os resultados dos algoritmos desenvolvidos. Os dados recolhidos durante o ensaio serão usados para desenvolver abordagens não invasivas para estimar a saturação de oxigénio e o ritmo cardíaco do individuo, que serão subsequentemente utilizados para a análise da qualidade do sono dos indivíduos/doentes.

d) **Objetivos** Os objetivos para a recolha das amostras usadas na realização dos estudos de investigação são:
   (a) Para medir a frequência cardíaca e SpO2 de forma não invasiva,
   (b) Analisar a qualidade do sono usando esses parâmetros fisiológicos.

e) **Seleção dos participantes**
   Os indivíduos de qualquer faixa etária entre 18-65 anos são elegíveis para participar no estudo. No entanto, se estiver sofrendo de algum distúrbio relacionado com o sono, é recomendada a menção durante a participação. A participação no estudo requer uma visita à zona da recolha de dados (Interactive Technologies Institute).

f) **Benefícios atuais e/ou futuros para o participante e/ou comunidade**
   Poderão não existir benefícios pessoais para a participação no estudo. No entanto, o conhecimento recebido poderá ser valioso para a construção de soluções não invasivas que possam estimar parâmetros médicos sem ser necessário colocar múltiplos sensores elétricos no corpo humano, evitando assim que as partes do corpo tenham efeitos colaterais devido aos sensores.

g) **Tempo requerido**
   O esclarecimento das dúvidas dos participantes relativamente ao estudo e a explicação das condições ambientais para a experiência, seguido do procedimento de recolha de dados, normalmente levaria aproximadamente 5 a 10 minutos.

h) **Riscos**
   Embora não existam riscos potenciais associados à recolha dos dados, os participantes podem sentir algum desconforto durante as recolhas das amostras de vídeo devido aos ajustes nas condições de luminosidade.

i) **Garantia de confidencialidade e anonimato**
   O estudo requer a recolha de informações pessoais dos participantes. As informações pessoais, como o nome e número do comprovativo de identidade, adquiridos durante o estudo serão mantidas em sigilo e não serão compartilhadas em nenhuma circunstância. No entanto, os dados recolhidos podem ser compartilhados entre os investigadores para o um propósito específico. A cada participante será atribuído um código exclusivo que o identifica.

   A proteção dos dados pessoais e a livre circulação desses dados, por parte dos voluntários, deve seguir estritamente o Regulamento 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no tratamento de dados pessoais e a livre circulação desses dados, revogando a Diretiva 95/46 / CE (Regulamento Geral de Proteção de Dados (RGPD)) e a Lei n.º 58/2019, de 8 de agosto, para a implementação do RGPD.

j) **Processamento, armazenamento e destruição de dados**
O processamento de dados inclui a extração da informação relativa ao pulso de volume sanguíneo das amostras de vídeo. As amostras de vídeo em ambiente escuro necessitariam das amostras de imagem para as tarefas associadas à estimação dos resultados. As amostras seriam compartilhadas após a assinatura do acordo de liberação do usuário, sendo também assinado pela parte solicitante. Os dados serão armazenados em instalações seguras, hospedadas pela instituição, como servidores seguros ou equipamentos protegidos por senha, e preservados até o limite temporal máximo permitido, sendo provavelmente destruídos posteriormente.

k) **Divulgação de resultados**
Os resultados da investigação serão divulgados por relatórios científicos (em conferências, jornais, cartas, etc.), comunicações pessoais e institucionais (como boletins informativos, páginas de onda, e-mail, etc.).

l) **Direito de recusar e respeito pelo mesmo** A participação no estudo é voluntária e pode ser interrompida a qualquer momento. Os investigadores não coagirão, em nenhuma circunstância, os participantes ou autoridades a participar neste estudo.

m) **Outros dados considerados importantes e esclarecedores** Não aplicável.

n) **Consentimento Voluntário** Ao assinar abaixo, concorda que as informações acima foram explicadas e todas as suas perguntas foram respondidas. Ao assinar este formulário, concorda em participar nesta investigação.

Nome e assinatura do participante	
Idade e sexo	
Documento de prova de identificação e número	
Data da recolha das amostras	
Nome e assinatura da pessoa que obtém o consentimento (investigador)	

INFORMED CONSENT FORMS FROM THE DATA PROTECTION (ENGLISH VERSION)

Informed consent for participation in Research Studies

Identificação do Investigador Prof. Fernando Manuel Rosmaninho Morgado Ferrão Dias
Associate Professor
Faculty of Exact Sciences and Engineering
University of Madeira

Title of the study: Collection of Facial Image and Video Sample of People

a **Context** The study aims to collect video and photo samples using the camera of a personal computer and to obtain readings on HR and SpO2 using a pulse oximeter to analyze sleep and study related problems. Explanation of the study:

b **Goal of the study**

The sampling objectives for this study are: Non-invasive measurement of HR; Analysis of the quality of sleep using the physiological parameters collected.

c **Data collection or other method**

This study aims to analyse video and photograph samples to estimate physiological parameters such as HR blood oxygen levels, which help the medical community to non-invasive diagnose patients with sleep disorders.

During the study, it will be collected an image under normal lighting conditions and video samples under different lighting conditions. HR and SpO2 level (SpO2%) information will also be collected using a pulse oximeter to validate methods for extracting this information from the video.

d **mention the place where the researcher meets the participant**

The participant has to visit the M-ITI premises to participate in the study.

e **how often and for how long approximately**

Considering the clarification of doubts, the explanation of the use and maintenance of the data, as well as the conditions of data collection and the collection itself, 10 to 15 minutes are needed.

f **Ensuring the destruction of recordings (audio or video) within a specific time period**

Images and videos will be kept indefinitely. A research database will be created without personal data. Participants will be informed that the database will be kept indefinitely and that it can be used for other studies and by other researchers and not just in the current study. Participants may ask for their data to be re-

moved from the database at any time. If data is given to researchers other than those of the current study, they will have to comply with the RGPD.

g **State whether there is any payment for travel or compensation**
There is no payment for travel or compensation for participating in the study.

h **Ensure confidentiality and exclusive use of the data collected for this study**
This study requires the collection of personal information from participants. The personal information collected, such as name and identity, will be kept confidential and will not be disclosed with anyone under any circumstances. Other data may be shared between researchers solely for research purposes. Each participant will be identified by a unique code and personal information will not be shared.
The protection of personal data and the free movement and deletion of such data will strictly follow Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation (GDPR)) and Law 58/2019, of 8 August for the implementation of GDPR.

i **acknowledgments and identify the researcher - name, profession, place of work, telephone contact, email address - and the person requesting consent, if different**
The process will be organized and managed by the researcher Ankit Gupta, whose contact details are given below. He will be responsible for asking participants to sign the informed consent document, after which the data will be collected.

j **Ankit Gupta**
PhD Student (Informatics Engineering)
Interactive Technologies Institute
University of Madeira
Contact: +351 926045472
Email: ankit.gupta@iti.larsys.pt
Please read this information carefully. If you have any questions or need further clarification, please let me know. Please sign the document if the terms meet with your approval.

INFORMED CONSENT FORMS FROM THE ETHICAL COMMITTEE (ENGLISH VERSION)

Informed consent for Face Image and Video Samples collection

a) **Researcher Identification**

Prof. Fernando Manuel Rosmaninho Morgado Ferrão Dias
Associate Professor
Faculty of Exact Sciences and Engineering
University of Madeira

Ankit Gupta
PhD Student (Informatics Engineering)
Interactive Technologies Institute
University of Madeira

b) **Researcher contact**

Email
morgado@uma.pt
ankit.gupta@iti.larsys.pt
Contact(s)
+351 291-705280
+351 926 045 472

c) **Scope of the study**

This study is intended for analyzing face video samples for estimating various clinical parameters like HR, blood oxygenation levels that assist the physicians to diagnose the people suffering from sleep-related disorders, non-invasively. Under this study, a facial image will be collected under normal light conditions and video samples will be collected under different light conditions. Further, the study involves the acquisition of physiological parameters (HR and blood saturation level (SpO2%)) to validate the results from the developed algorithms. The data collected during the experiment will be used to develop non-invasive approaches to estimate SpO2 and HR of the individual, which will further be used for sleep quality analysis of the individuals/patients.

d) **Objectives**

The objective for collecting the samples for conducting research studies are as

follows: 1. To measure the HR and SpO2 in a non-invasive way. 2. To analyze the quality of sleep using these physiological parameters.

e) **Selection of subjects**
Individuals of any age group between 18-65 years are eligible to participate in the study. However, if he/she is suffering from any sleep-related disorder/, it is recommended to mention it during participation. Participating in the study requires one visit to the data collection area (Interactive Technologies Institute).

f) **Present and / or future benefits for the subject and / or community**
There may be no personal benefit from participation in the study but the knowledge received may be of value to build noninvasive solutions in the estimation of medical parameters without intensively putting electric sensors to the human body, thereby preventing the body parts from side effects due to the sensors.

g) **Time Spent**
Clearing the doubts of individuals about the study and explaining the environmental conditions for the experiment, followed by data collection procedure would normally take approximately 5-10 minutes.

h) **Risks** Although there are no potential risks associated with the collection of the data, the participants may suffer some discomfort during video sample collections due to adjustments in the light conditions.

i) **Guarantee of confidentiality and anonymity**
The study requires the collection of personal information from the participants. The personal information like name and identity proof number acquired during the study will be kept secret and not be shared with anyone under any circumstances. However, Data collected can be shared among researchers for the purpose defined. Each participant will be assigned and identified by a unique code. The protection of personal data and the free movement of such data for the volunteers shall strictly follow Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons about the processing of personal data and the free movement of such data, repealing Directive 95/46/EC (General Data Protection Regulation (GDPR)) and Law 58/2019, of August 8 for implementation of GDPR.

j) **Processing, storage and destruction of data**
Data processing includes extracting the BVP information from the video samples. The dark video samples would need the image samples for estimation tasks. The samples would be shared after having signed the user release agreement signed by the requesting party. The data will be stored using security facilities hosted by institutions like secure servers or password protected equipment and preserved

till the maximum permissible time limit and would be likely to be destructed afterward.

k) **Dissemination of results**

The result of the research will be disclosed as scientific reports (such as conferences, journals, letters, etc.), personal and institutional commutations (such as newsletters, wave pages, emails, etc.)

l) **Right to refuse and respect for the same**

Participation in the study is voluntary and participation can be stopped at any time. The individuals will not coerce the individuals or authority to participate in this study under any circumstances.

m) **Other data considered important and clarifying**

Not Applicable

n) **Voluntary Consent**

By signing below, you agree that the above information has been explained to you and all your current questions have been answered. By signing this form, you agree to participate in this research study.

Name and Signature of the Participant	
Age and Sex	
Identity Proof Document Number	
Date of Collecting the sample	
Name and Signature of the Person obtaining the consent (Investigator)	