CONTACT-FREE PHYSIOLOGICAL MONITORING OF CARDIORESPIRATORY STATES USING RADAR AND OPTICAL SENSORS

Gašper Slapničar

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Dr. Mitja Luštrek, Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia
Co-Supervisor: Assoc. Prof. Dr. Wenjin Wang, Biomedical Engineering Department, Southern University of Science and Technology, Shenzhen, China; Department of Electrical Engineering, Eindhoven University of Technology, the Netherlands

Evaluation Board:
Asst. Prof. Dr. Aleksandra Rashkovska Koceva, Chair, Department of Communication Systems, Jožef Stefan Institute, Ljubljana, Slovenia
Prof. Dr. Matej Kristan, Member, Visual Cognitive Systems Lab, Faculty of Computer and Information Science, Ljubljana, Slovenia
Asst. Prof. Dr. Gennaro Tartarisco, Member, National Research Council of Italy, Rome, Italy

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Gašper Slapničar

CONTACT-FREE PHYSIOLOGICAL MONITORING OF CARDIORESPIRATORY STATES USING RADAR AND OPTICAL SENSORS

Doctoral Dissertation

BREZSTIČNO FIZIOLOŠKO SPREMLJANJE KARDIORESPIRATORNIH STANJ Z RADARSKIMI IN OPTIČNIMI SENZORJI

Doktorska disertacija

Supervisor: Prof. Dr. Mitja Luštrek

Co-Supervisor: Assoc. Prof. Dr. Wenjin Wang

Ljubljana, Slovenia, March 2024

First to my parents and family who have shown unwavering support and instilled values in me that got me here.
Then to my wife Jasmina who is the best thing that has ever happened to me.
To my supervisors, Mitja and Wenjin, who offered discussions rather than lectures, and without whom this would not have been possible. They were always there and made problems seem smaller than they were.
To my colleagues and coworkers at the Department of Intelligent Systems who made this journey fun.
To Peter, for the content and talks that made me laugh.
To everyone that were a part of this journey in one way or another.
Finally, to myself.

Acknowledgments

First, I would like to thank and acknowledge my supervisors, Prof. Dr. Mitja Luštrek and Assoc. Prof. Dr. Wenjin Wang – they offered great insights and guidance throughout the creation of this dissertation, from high-level conceptual discussions on study designs, to micro-level internal reviews of several papers and the dissertation itself.

I would also like to thank my committee members – Asst. Prof. Dr. Aleksandra Rashkovska Koceva, Prof. Dr. Matej Kristan and Asst. Prof. Dr. Gennaro Tartarisco – for taking the time to analyze, understand and review my dissertation, giving valuable suggestions and providing constructive feedback.

Special thanks to Prof. Dr. Matjaž Gams for his ideas, motivation, knowledge, and opportunities provided.

Furthermore, I would like to acknowledge the support and contribution of the whole Department of Intelligent Systems, Jožef Stefan Institute, including current and former employees, as well as student collaborators – there have been many discussions throughout the years with many people that helped improve the developed methodology. Many have also participated in the data-recording experiments that made part of this research possible.

Special thanks goes to student collaborators Kitoshi Kawai (University of Tokyo), Nejc Mlakar (Faculty of Computer and Information Science, University of Ljubljana) and Rok Hudobivnik (Faculty of Computer and Information Science, University of Ljubljana) who all helped write some parts of code for some subsets of experiments throughout the years. Kitoshi was also instrumental in the data collection procedure.

Additionally, I would like to express sincere gratitude to my colleague Junoš Lukan who explained and helped me understand many aspects of statistical analysis, which was instrumental in validating the results of this work.

Penultimately, a special thank you to the Faculty of Computer and Information Science, University of Ljubljana, and the Jožef Stefan International Postgraduate School, for providing the building blocks of my knowledge that led to this dissertation.

Ultimately I would like to acknowledge the financial support of the Slovenian Research and Innovation Agency (ARIS) and Jožef Stefan Institute for my position as a young researcher under grant agreement PR-09783.

Abstract

Omnipresence of sensor-equipped devices spurred rapid development of e-health and m-health applications in the past decades. Despite their wide-spread adoption in the form of wearables, such devices are ultimately not a universal or ideal solution for regular health monitoring due to their reliance on battery, requiring skin contact and general obtrusive nature. Ideally, the need for direct user-device interaction should be completely removed in the paradigm of ubiquitous and pervasive computing, which can be achieved using contact-free sensors such as radars and cameras that monitor different parts of electromagnetic (EM) spectrum. These devices can be used to monitor different physiological parameters in an unobtrusive manner, making them feasible for subjects who cannot use wearables (e.g., neonates, burn victims, elderly with dementia).

We initially explored the potential of radio-frequency part of the EM spectrum, measured by radars, for detection of complex hemodynamic states. These are expressed via several physiological parameters, including respiration. Radars allow for measurement of periodic thoracic expansion and contraction even in challenging conditions, such as night time and occlusion, making them ideal for sleep monitoring. We proposed a novel branched neural network architecture that can take a different number and type of input signals. We showed that we can detect five different hemodynamic states available in a public dataset (including apnea) with up to 0.83 accuracy and F1 score when using only radar signals as input. These results were only 4-5% behind traditional contact sensors, confirming feasibility of radar-based physiological monitoring.

In the second part, we investigated the feasibility of using the visible part of the EM spectrum, specifically the feasibility of a modified consumer RGB camera for multi-wavelength (MW) pulse transit time (PTT) measurement between different skin layers. Different wavelengths penetrate to different depths and allow for depth-specific photoplethysmogram (PPG) reconstruction. These can be used for MW PTT computation and subsequent blood pressure (BP) estimation. We found that algorithmic channel separation of mentioned PPGs is mandatory due to the imperfect nature of image sensor design, which causes spectral overlap between PPGs from different depths. We thus developed several algorithms that allow for data-driven camera-independent channel separation, which in turn allows for precise measurement of MW PTTs. Finally we confirmed on an in-house dataset that such MW PTTs are well-correlated with BP and that a personalized regression model can be trained to predict both systolic and diastolic BP with errors within clinical standards.

Overall we showed in this dissertation that contact-free sensors leveraging the information from the EM spectrum are an affordable unobtrusive alternative to wearables, and can achieve similar performance in monitoring of important physiological parameters and states. While some limitations and challenges remain, such as difficult uncontrolled conditions and privacy concerns, there is potential for implementing the proposed methods in a single device, which could immensely improve the speed, cost and comfort of physiological monitoring both at home and in hospitals.

Povzetek

Vseprisotnost naprav s senzorji je v preteklih desetletjih spodbudila pospešen razvoj aplikacij za elektronsko in mobilno zdravje. Navkljub razširjeni uporabi nosljivih naprav s senzorji le-te niso splošna in idealna rešitev za redno spremljanje zdravstvenega stanja, saj so odvisne od baterije, terjajo stik s kožo in so lahko moteče. V idealnem primeru bi v kontekstu vseprisotnega računalništva povsem odpravili potrebo po neposredni interakciji med uporabnikom in senzorjem, kar je dosegljivo z uporabo brezstičnih senzorjev, kot so radarji in kamere, ki spremljajo različne dele elektromagnetnega (EM) spektra. Te naprave omogočajo spremljanje različnih fizioloških signalov na nemoteč način, kar jih naredi primerne tudi za uporabnike, ki ne morejo nositi nosljivih naprav (npr. nedonošenčki, žrtve opeklin, starejši z demenco).

Sprva smo raziskali potencial radiofrekvenčnega dela EM spektra, ki ga merijo radarji, za zaznavo kompleksnih hemodinamičnih stanj. Ta se odražajo v več fizioloških signalih, vključno z dihanjem. Radar omogoča merjenje periodičnega širjenja in krčenja prsnega koša tudi v zahtevnih pogojih, kot so tema in zakritost opazovanca, kar ga naredi zelo primernega za spremljanje spanja. Predlagali smo novo razvejano arhitekturo nevronske mreže, ki lahko sprejme različno število in tip vhodnih signalov. Pokazali smo, da lahko zaznamo pet različnih hemodinamičnih stanj, ki so bila na voljo v javni bazi podatkov, s točnostjo in mero F1 0.83, če kot vhode uporabimo le radarske signale. Ti rezultati zaostajajo le 4-5% za tistimi z uporabo tradicionalnih motečih nosljivih senzorjev, kar potrjuje izvedljivost radarskega spremljanja fizioloških signalov.

V drugem delu smo raziskali izvedljivost izrabe vidnega dela EM spektra, konkretno izvedljivost uporabe modificirane splošnonamenske RGB kamere za merjenje večvalovnega (ang. multi-wavelength) časa prenosa pulza (ang. pulse transit time, PTT) med različnimi plastmi kože. Različne valovne dolžine prodrejo različno globoko, kar omogoča rekonstrukcijo fotopletizmogramov (ang. photoplethysmogram, PPG) iz vsake globine. Iz teh se lahko izračunajo PTT-ji, ki se nato uporabijo za ocenjevanje krvnega tlaka. Ugotovili smo, da je obvezen korak algoritmično ločevanje omenjenih PPG-jev zaradi nepopolne implementacije slikovnega senzorja, saj slednja povzroči spektralni presek med PPG-ji. Posledično smo razvili več algoritmov, ki omogočajo ločevanje kanalov in natančno merjenje večvalovnih PTT-jev samo na osnovi podatkov, neodvisno od specifične kamere. Nazadnje smo potrdili, da so takšni večvalovni PTT-ji dobro korelirani s krvnim tlakom in da je možno naučiti personaliziran regresijski model za napovedovanje sistoličnega in diastoličnega krvnega tlaka, ki dosega napake znotraj mej kliničnih standardov.

Celokupno smo v tej disertaciji pokazali, da so brezstični senzorji, ki izrabljajo informacijo iz EM spektra, cenovno dostopna in nemoteča alternativa nosljivim napravam ter da lahko dosežejo podobne rezultate pri spremljanju pomembnih fizioloških parametrov in stanj. Čeprav omejitve in izzivi ostajajo, kot so naprimer zahtevni nenadzorovani pogoji in ohranjanje zasebnosti, obstaja potencial za implementacijo predlaganih metod v eni sami napravi, ki bi lahko znatno izboljšala hitrost, ceno in udobje fizioloških meritev tako doma kot v bolnišnicah.

Contents

List of Figures

List of Tables

List of Algorithms

Abbreviations

(A)NN	... (artificial) neural network
(N)IR	... (near) infrared
ABP	... arterial blood pressure
AC	... pulsatile signal amplitude
AF	... atrial fibrillation
AHI	... apnea-hypopnea index
ANOVA	... analysis of variance
ANS	... autonomic nervous system
BCG	... ballistocardiograph
BMI	... body mass index
BP	... blood pressure
bpm	... beats-per-minute
BSS	... blind source separation
CHROM	... chrominance (algorithm)
CMOS	... complementary metal-oxide-semiconductor
CNAP	... continuous noninvasive arterial pressure
CNN	... convolutional neural network
CO	... cardiac output
COVID-19	... SARS-CoV-2 virus disease
CV	... cross validation
CVD	... cardiovascular disease
DBP	... diastolic blood pressure
DC	... non-pulsatile baseline variations or wandering
DL	... deep learning
DNA	... deoxyribonucleic acid
ECG	... electrocardiogram
EM	... electromagnetic
FFT	... fast Fourier transform
FIR	... final impulse response
FMCW	... frequency modulated continuous wave
GA	... genetic algorithm
GDDR5	... graphics double data rate 5
GPU	... graphics processing unit
HR	... heart rate
HRV	... heart rate variability
I	... In-phase (component)
IBI	... interbeat intervals
ICA	... independent component analysis
ICG	... impedance cardiogram
JSI	... Jožef Stefan Institute
LED	... light emitting diode

LIDAR	...	light detection and ranging
LOSO	...	leave one subject out
LSTM	...	long-short-term-memory
MAE	...	mean absolute error
MDI	...	mean decrease in impurity
ME	...	mean error
ML	...	machine learning
MW	...	multi-wavelength
PAT	...	pulse arrival time
PCA	...	principal component analysis
PI	...	perfusion index
PIMD	...	profound intellectual and multiple disabilities
POS	...	plane orthogonal to skin (algorithm)
PPG	...	photoplethysmogram
PTT	...	pulse transit time
PWV	...	pulse wave velocity
Q	...	Quadrature (component)
RAM	...	random access memory
RE	...	relative error
RF	...	radio frequency
RGB	...	red, green, blue
RMSE	...	root mean squared error
RMSSD	...	root mean square of successive differences
ROI	...	region of interest
rPPG	...	remote photoplethysmography
RR	...	respiratory rate
SBP	...	systolic blood pressure
SDK	...	software development kit
SDNN	...	standard deviation of N-N intervals
SMOTE	...	synthetic minority oversampling technique
SNR	...	signal to noise ratio
SOTA	...	state of the art
SpO2	...	blood oxygen saturation
SV	...	stroke volume
SVM	...	support vector machines
SVR	...	systemic vascular resistance
TFM	...	task force monitor
Tr	...	receiver antenna
Tx	...	transmission antenna
UV	...	ultra-violet
VR	...	virtual reality
vRAM	...	video RAM

Chapter 1

Introduction

Sensor-equipped devices have become omnipresent over the last decade [1], [2] and are commonly equipped with a variety of sensors, which can be used for different purposes, such as obtaining physiological data about the user. This spurred rapid development of e-health and m-health applications [3], which manifested in great improvements of continuous patient monitoring and wide-spread adoption of telemedicine [4]. Importance of such systems was further reinforced post 2019 with the emergence and rise of the SARS-CoV-19 (COVID-19) pandemic, which made traditional medical treatment difficult, both in terms of availability, as well as safety, especially for high-risk groups like chronic patients and elderly people [5]–[7].

Despite the increasing prevalence and wide-spread use of sensors, mostly in the form of wearables that require contact with the user (e.g., smart watches, smartphones, etc.), the ultimate goal of pervasive and ubiquitous computing is to completely remove the need for the user–device interaction [8], [9]. In practice, this can be naturally achieved using contact-free sensors, such as cameras. These can be present in an environment without limiting or bothering the user in any way, while allowing for continuous non-invasive monitoring of important physiological parameters and related medical conditions of almost any subject.

1.1 Motivation

The potential of contact-free sensing becomes apparent in the context of current gold-standard devices used for monitoring different important aspects of user health.

Electrical cardiac activity is recorded via electrocardiogram (ECG), which is obtained using electrodes placed at correct position on the limbs and thorax [10]. These electrodes must be surrounded by a conductive gel and adhesive substance to keep them in the correct place. They are equipped with an analog-to-digital converter connected to a central unit, most commonly with cables. Such a traditional setup can be often found in a doctor's office and laboratories for continuous patient monitoring.

Respiratory monitoring is another important aspect of healthcare, which is of interest both for relatively common populations, such as long-term smokers and elderly people, as well as more specific high-risk populations, such as sleep apnea patients [11]. Such monitoring is exceptionally obtrusive, where the golden standard remains a complex device including a mask or semi-invasive nasal cannulas placed on the subjects face, covering the mouth and / or nasal area. These are commonly accompanied by a respiration belt around the thorax. The mask or cannulas are again connected with a tube to a central device where the airflow and respiratory rate (RR) is recorded.

Another parameter that is extremely commonly measured and monitored, both in office and at home, is blood pressure (BP). Digital sphygmomanometer is the de-facto standard

for general measurement, although the perfect ground truth of continuous arterial blood pressure (ABP) can only be obtained using invasive intravenous catheters. Sphygmomanometer once again includes a tube connecting the cuff and the central device, which is used to inflate the cuff and allow for a measurement to be made in a specific body position.

More physiological details contributing to the requirement for such devices will be discussed in later sections, however, one can already discern the common denominator of these methods – relatively complex devices and protocols for monitoring important cardiovascular and respiratory activity, which are quite obtrusive and can be frustrating for the subjects. This results in many undesired effects, such as problems in adherence to regular measurement and potentially even incorrect measurements due to the procedure itself influencing the parameters. A good example of this is the white-coat syndrome, a common problem with in-office BP measurement, where the person following the protocol becomes anxious and worried about the result, subsequently causing changes in hemodynamics that are reflected in the measured values. The latter thus lose some of the important information about the default state of the subject and cannot be completely trusted [12].

Examples of the discussed devices and the required sensor positions and / or measurement protocols are shown in Figure 1.1.

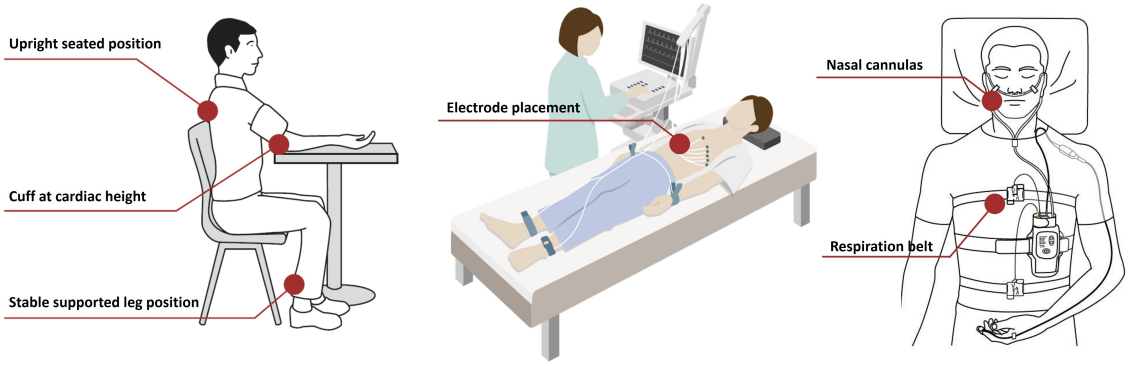

Figure 1.1: Examples of golden standard devices and measurement protocols for monitoring important physiological parameters, including blood pressure with a sphygmomanometer, cardiac activity with electrodes, and respiratory activity with nasal cannulas.

The nature of the described wearable sensor systems is not only obtrusive and cumbersome, but fundamentally infeasible for use with specific high-risk subject groups. A prime example of this are (preborn) neonates [13], which have highly sensitive skin and fragile body, and are kept and monitored inside incubators. Another example are burn victims or people with severe allergies, where wearable sensors requiring skin contact cannot be used. Finally, limitations of high-risk groups are not only physical, but also cognitive. People with profound intellectual and multiple disabilities (PIMD) [14], [15] and elderly people with cognitive impairment (e.g., dementia) can neither understand nor remember how and when to use devices like a digital sphygmomanometer. Even after precise instructions, it cannot be expected for such subjects to adhere to a rigorous measurement routine [16].

There is a clear need to resolve the described issues, as physiological parameters often serve as key indicators of many high-risk diseases. For instance, cardiovascular diseases (CVDs) – including coronary heart disease, hypertensive heart disease, atrial fibrillation, etc. – are the leading cause of death globally [17], [18]. Similarly, respiratory disorders – including sleep apnea, asthma, bronchitis, etc. – also have major negative effects on large parts of population [19]. These facts serve as a motivation for our research into contact-free monitoring of important respiratory and cardiovascular physiological parameters and

detection of related conditions.

1.2 Problem Definition

Our problem directly follows the motivation and reasoning from the previous section. We have to propose, develop and validate contact-free methods with suitable sensors that enable contact-free monitoring of important physiological parameters and conditions, as such parameters are important indicators of highly impactful and dangerous diseases mentioned previously.

More precisely, we first have to consider possible alternatives to existing sensors that are suitable for capturing the information related to cardiovascular and respiratory activity. These physiological modalities are chosen due to their medical importance and feasible contact-free alternatives to traditional sensors. We then have to analyze their pros and cons, gather data, propose methods for physiological parameter extraction, investigate their feasibility and finally validate the setup using robust experiments.

The idea of contact-free sensing itself is not novel [20], however, we are interested in more than only basic parameters such as heart rate (HR) and RR, which are the most commonly measured vitals using such approaches. The aim of this dissertation is to investigate the feasibility of measuring and detecting more complex conditions and parameters that are important indicators of the subject's health and well-being. In the respiration modality, a good example of this is sleep apnea, which is a widely-investigated and deadly condition [11]. In cardiovascular hemodynamics, an important parameter that influences a severe portion of elderly population is BP, and, directly following from it, potential hypo- and hyper-tension [21]. Respiration patterns and BP are universally important, as they were shown to be connected to general well-being, stress and other complex states [22]. Explicitly, the following problems are of interest:

- Wearable sensors are predominantly used for home monitoring of physiological parameters, which then drives telemedicine. However, such sensors have several downsides, such as battery or cable dependence, discomfort, specific placements and ultimately even the requirement to be worn, which is not trivial, especially for populations like elderly with dementia. Subsequently, contact-free unobtrusive alternatives should be considered for monitoring as many parameters as possible.

- The current standard for sleep monitoring and apnea detection are sleep laboratories, where subjects are equipped with masks and chest straps to measure respiration. Such sleep monitoring setups substantially decrease sleep quality and subject comfort.

- A large amount of the elderly population is suffering from hypertension and related diseases, which will become increasingly prevalent with aging populations globally. BP is the de-facto indicator of such diseases, but current standard cuff-based monitoring is cumbersome, requires knowledge of the measurement protocol and can cause incorrect measurements.

In the following section, we will focus on the mechanisms influencing physiological parameters and governing the complex hemodynamics in the human body. Our aim is to first thoroughly present and understand the foundation on which computer-assisted contact-free health monitoring methods can be feasibly built, with good understanding of the fundamentals governing the monitored phenomena and causing the measurable effects. This, alongside the analysis of existing state-of-the-art (SOTA) related work, will allow us to propose hypotheses for our research, which originate from the problems listed here and

the solutions already existing in related work. We will then propose and investigate both hardware and algorithmic solutions. The interpretation of obtained results will allow us to ultimately accept or reject the hypotheses proposed in Section 3.3.

Chapter 2

Physiological Background

A detailed understanding of the physiological phenomena, governing the physiological parameters of interest, is of utmost importance, especially when designing novel methods of measurement. In this section, we give a detailed overview of the human circulatory system and hemodynamics, starting from a high-level overview and then narrowing on more detailed principles of blood flow at organ level, focusing on skin, as it is the organ that is directly observable and most easily measurable.

2.1 Circulatory System

Human cardiovascular and respiratory system are closely connected. The former consists of the heart and the vascular network (including arteries, veins and capillaries) and the latter includes lungs and the airways (nose, mouth, windpipe) that deliver air rich with oxygen [23]. They work in tandem to provide continuous oxygen supply to the blood cells and in turn maintain cell-level metabolism [24]. The hemoglobin-rich oxygenated blood carries oxygen and nutrients to the organs, tissues and muscles away from the heart via arteries. Deoxygenated blood returns towards the heart and lungs via veins. Two notable naming exceptions are the pulmonary artery and vein – the former is the only artery carrying deoxygenated blood (away from the heart, towards the lungs), while the latter is the only vein carrying oxygenated blood (from the lungs towards the heart). Together, this is known as the circulatory system and is schematically shown in Figure 2.1. Note that the separation between the respiratory and cardiovascular system is not clear-cut, as both are closely intertwined.

Following the structure of the circulatory system, we can determine that the respiratory and cardiovascular activity are crucial for normal functioning of the body. Should they deviate from expected functioning, such conditions can have serious consequences. While cardiovascular and respiratory activity in themselves are well-known and subject of extensive research [23], there is another important aspect connecting them and ensuring correct functioning – hemodynamics. These are the dynamics of the blood flow through the circulatory system. More precisely, they describe the physical laws that govern the flow of blood in the circulatory system or more specifically in the vascular network [26].

2.1.1 Respiratory System Structure and Functions

Normal respiratory activity is an obvious requirement for the circulatory system to successfully deliver oxygen to the body. This means that continuous delivery of fresh air (rich with oxygen) from the outside environment to the lungs, and subsequently to the cells within

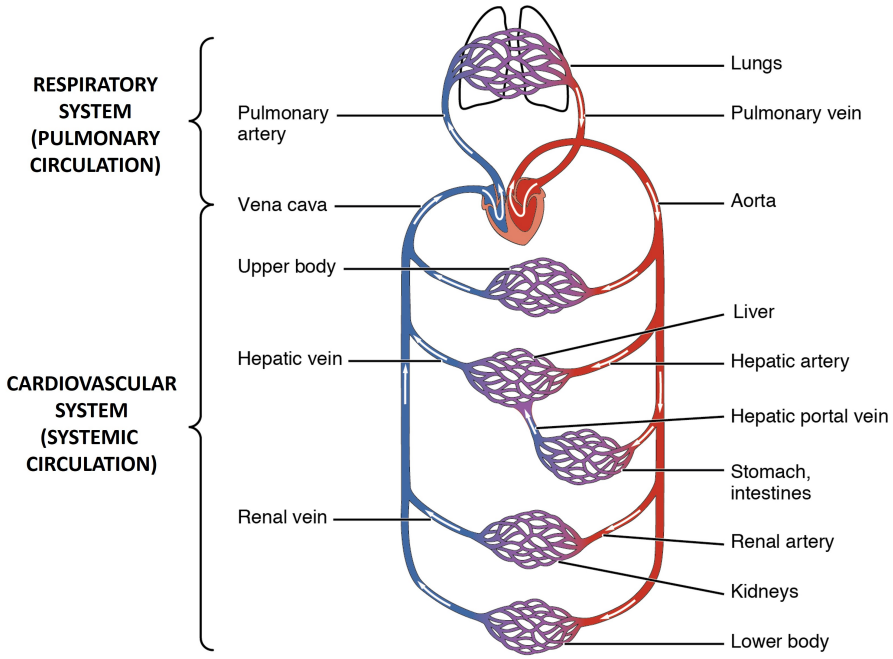

Figure 2.1: Human circulatory system [25], comprising the lungs, the heart, and the vessel network connecting to other organs. This system ensures oxygen and nutrient delivery to cells throughout the body.

tissues, must be ensured. Following this, removal of the carbon dioxide is also mandatory, as it becomes toxic in higher concentrations [27].

The gas exchange described above occurs between the lung and the external environment by the process of ventilation. This involves inhalations and exhalations within a respiratory cycle. Inhalation is the active movement of bringing the air into the lungs with the help of the diaphragm muscle, which causes the lungs and chest to expand substantially. The gas exchange then takes place with the cardiovascular system via the blood-air barrier within structures called pulmonary alveoli, shown in Figure 2.2 [28]. In contrast, exhalation is usually a passive process, which is the consequence of the elastic recoil of the lungs and the chest wall, and causes the lungs and chest area to contract [29].

The changes in the pulmonary volume cause the frontal wall of the thorax to displace periodically. This displacement is minor, measured in millimeters (mm) to centimeters (cm) at most [31], but very highly correlated with respiratory rate [32], making it a useful indicator for respiration monitoring and detection of dangerous conditions, such as sleep apnea. Respiratory rate must be in an expected range and uninterrupted for periods longer than a minute, as failure to provide oxygen can swiftly have severe life-threatening effects on the brain [33].

2.1.2 Cardiovascular System Structure and Functions

The cardiovascular system includes the heart and vascular network, as shown in Figure 2.3.

The heart is a muscular organ that pumps blood through the vessels and is located between the lung wings in the middle of the thorax area. It is divided into four chambers, two atria and two ventricles. Similarly to respiration, cardiac activity is continuous and periodic, but with higher frequency compared to respiration. A single cardiac cycle comprises atrial and ventricular systole, followed by atrial and ventricular diastole. The atrial

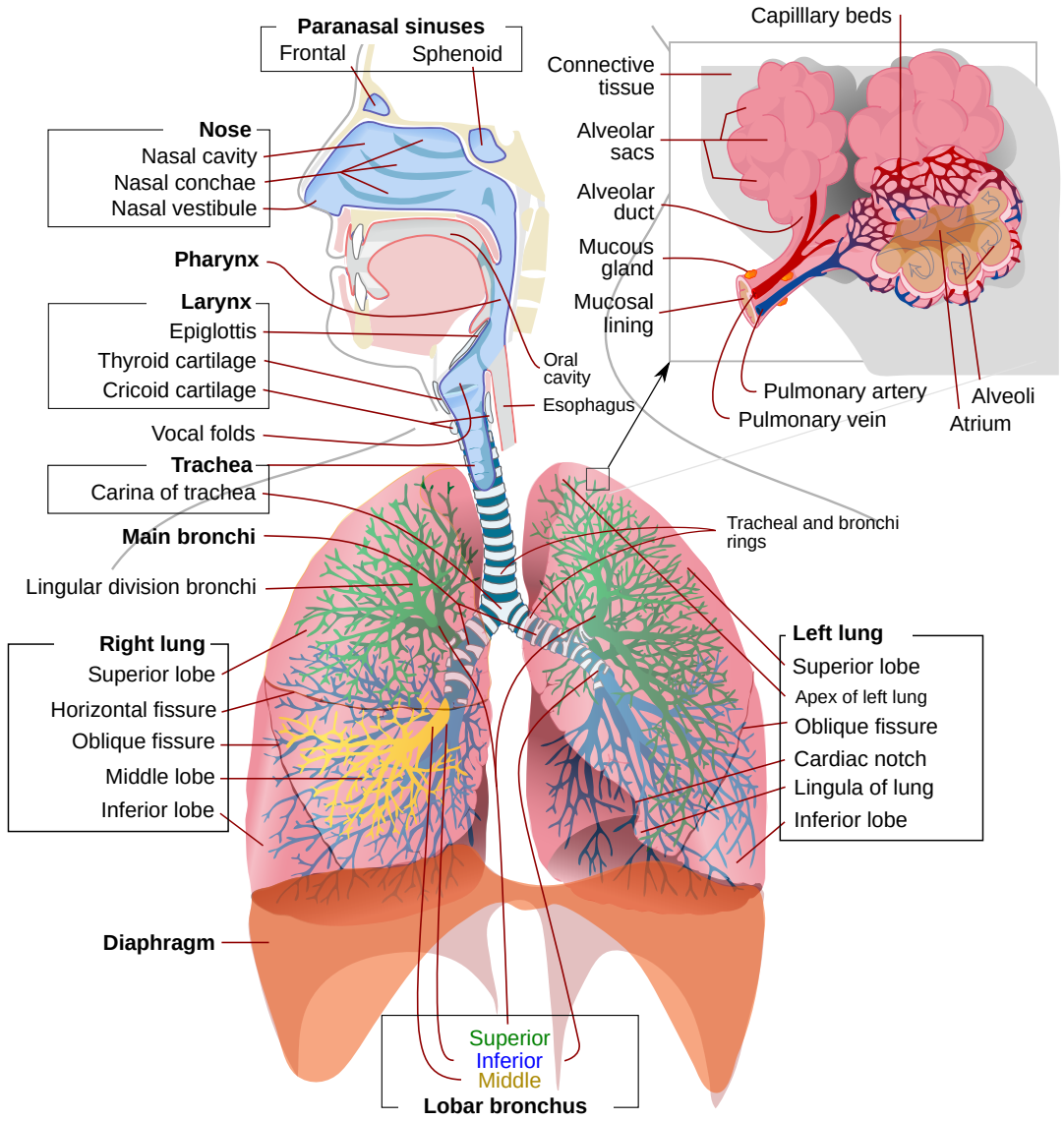

Figure 2.2: Human respiratory system [30] that allows for oxygen transfer between the outside environment and the blood.

systole is the period in which the atria contract, filling the ventricles with blood. Ventricular systole is the period of contraction of the ventricles, forcing blood into the aorta and pulmonary artery. Atrial diastole is the period during which atria relax and fill with blood coming from the lungs. Similarly during ventricular diastole the ventricles relax and refill with oxygenated blood from the atria. This procedure is coordinated between the atria and ventricles, meaning that atria relax when ventricles contract and vice-versa [35]. This procedure is reflected in the ECG in the form of the well-known QRS complex, with the additional P and T waves before and after QRS, respectively. In brief, the cycle starts with ventricular filling in the late parts of the ventricular diastole, immediately followed by atrial contraction in the atrial systole. These are reflected in the ECG as the P wave and the Q deflection. This is followed by the isovolumetric ventricular contraction and blood ejection during the ventricular systole, which is seen in the ECG as the sharp R

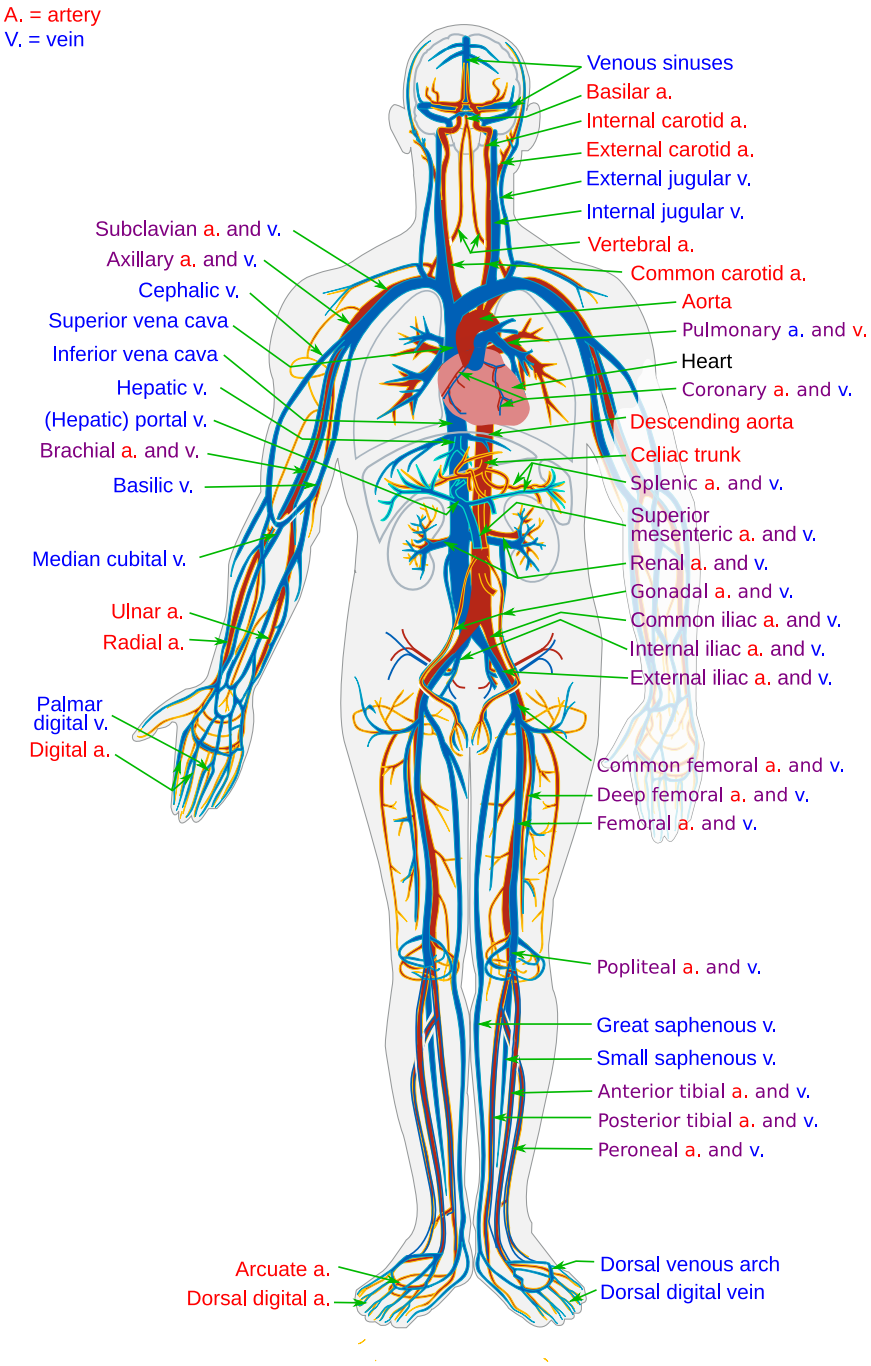

Figure 2.3: Human cardiovascular system [34]. Red denotes arteries carrying oxygenated blood, blue denotes veins carrying deoxygenated blood. Purple denotes vasculature including both arteries and veins.

wave. Finally, the isovolumetric relaxation of the ventricles and another refilling follow, which are reflected in the ECG as the S and T wave, respectively [36]. The procedure is illustrated in Figure 2.4.

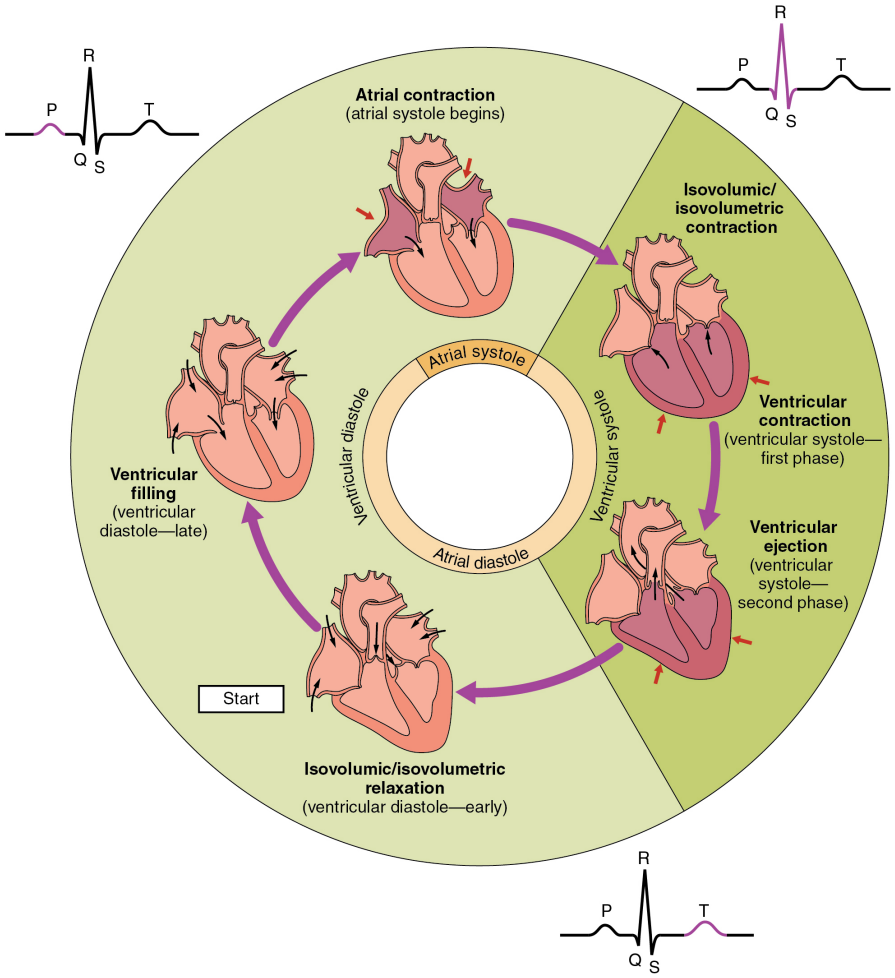

Figure 2.4: Illustration of the cardiac cycle and its reflection in the QRS complex [37].

Note that despite Figure 2.4 showing a very sequential procedure, it is in reality performed in parallel, as mentioned previously and highlighted in Figure 2.5.

Despite precise separation into two pairs of systoles and diastoles, phrasing in literature commonly refers to a single cardiac systole and diastole, not separating between atrial and ventricular. The diastole generally refers to the period during which the heart muscle relaxes and refills with blood, while the systole refers to the period of contraction and pumping of blood. In practice, this naming is more closely associated with ventricular systole and diastole, since it is the precise period during which blood is actually ejected from the heart.

2.1.3 Hemodynamics

Once the blood is ejected from the heart in the ventricular systole, it immediately enters the main arteries, which in turn split into smaller arteries, arterioles and finally capillaries throughout the body. Based on the structure (elasticity and related vascular resistance) of the cardiac and vascular walls, the amount of blood being ejected – known as the

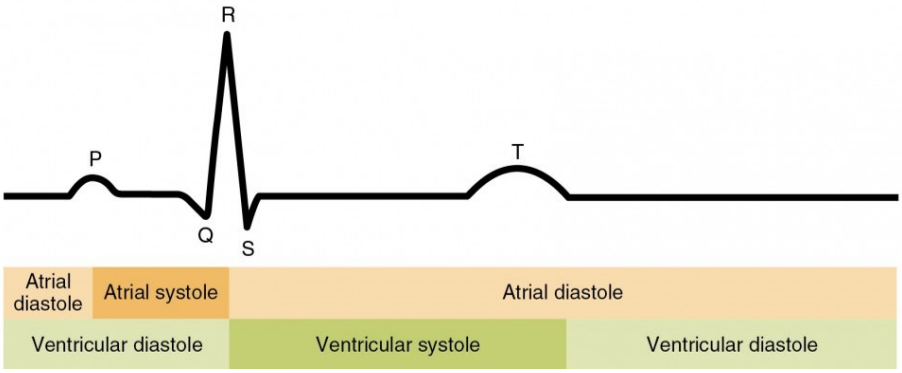

Figure 2.5: Systoles and diastoles of cardiac cycle happening in parallel between atria and ventricles.

cardiac output (CO) – and the relative chemical structure alongside corresponding physical properties of the blood, there is different pressure exerted on the vascular wall [38]. Optimal performance of the previously described circulatory system depends on maintaining precise value of such ABP. It is of vital importance that it remains within some bounds, which allows for normal blood flow (pressure must not be too low), while not exerting too much pressure on the vascular wall (pressure must not be too high). Stable ABP in expected range ensures optimal perfusion of critical organs (e.g., brain) and other parts of the body [39].

The flow of blood is orchestrated by the previously described cardiac cycle and the arterial pressure is continuously controlled by several mechanisms that ensure biological homeostasis:

- **Baroreflex.** Responsible for continuous short-term regulation of BP. It relies on specialized neurons (baroreceptors) that relay changes in BP detected via the tension / stretching of the arterial wall. The detected stretch changes are mediated by the autonomic nervous system (ANS) and the HR is adjusted according to the change. In case of lowered BP, the HR is increased, and vice-versa [40].

- **Chemical regulation via hormones.** Responsible for long-term maintenance of normal functioning BP, the heart releases atrial natriuretic peptide when BP is too high, and the kidneys detect and correct low BP with the renin-angiotensin system [41]. Details of these mechanisms are the subject of active research and out of the scope of this dissertation.

As briefly mentioned before, the ABP measured value is determined mainly by three subject-specific characteristics. First is the CO, or the amount of blood ejected by the heart into the aorta, second is the systemic vascular resistance (SVR), or the fluidic resistance of the vessels opposing the blood flow, and third is the chemical structure of the blood that gives it specific physical properties [40]. While the structure of blood can change (e.g., medication like blood thinners), we assume it to be constant within an average individual that is not treated for some disease. Thus we can define ABP as:

$$ABP = CO \cdot SVR \tag{2.1}$$

where ABP is the arterial blood pressure in millimeters of mercury (mmHg), CO is the cardiac output volume in L/min and SVR is the systemic vascular resistance in

mmHg/(L/min). The specific CO further depends on the HR and the stroke volume (SV), which is the amount of blood ejected by the heart at each contraction [40]:

$$CO = HR \cdot SV \tag{2.2}$$

where CO is the cardiac output in L/min, HR is the heart rate in beats-per-minute (bpm) or min^{-1}, and SV is the stroke volume in L. Physiologically, HR is determined by the sinoatrial node made up of pacemaker cells, which serves as a natural pacemaker of the heart. A certain SV is achieved by the signalled contraction of the ventricular nodes. Both are governed by the autonomic nervous system (ANS), which can achieve quick changes in ABP via the previously described baroreflex regulation mechanism.

2.1.4 Systolic and Diastolic Blood Pressure

Due to continuous periodic nature of the cardiac activity, ABP is not a single value but a continuous time series changing with the blood flow and pulse wave propagation. When a pulse is ejected and propagated during the cardiac systole, the pressure is increasing, while the refilling of the heart during the diastole causes it to decrease. In standard clinical practice however – meaning outside of invasive procedures such as surgery – there are two main indicators commonly provided and monitored [40]:

- **Systolic blood pressure (SBP).** Defined as the maximum value of the ABP trace in a single cardiac cycle or defined period.

- **Diastolic blood pressure (DBP).** Defined as the minimum value of the ABP trace in a single cardiac cycle or defined period.

The clinical significance and diagnostic value of SBP and DBP are immense, and they are an essential part of any medical examination [42]. As a direct indicator of hypo- and hyper-tension, two of the most wide-spread and dangerous CVDs, their importance cannot be overstated [17], [18].

They are commonly measured in mmHg due to historic reasons, as the first devices for measurement used a mercury column. The measurement protocol should be followed precisely due to notable influence of any deviation on the measured BP values. As shown in Figure 1.1, the subject must sit upright, with supported back and legs uncrossed, as such position ensures that the vascular system is not influenced by factors not relating to hemodynamic state (such as compression of the arteries when legs are crossed) [42]. Furthermore, arms should never be raised, but should be positioned naturally with the upper arm resting at heart level. Should the arm be raised, this requires the pulse wave to additionally propagate against gravity, increasing the required pressure at the aorta and influencing the measured values [42].

2.2 Pulse Wave Propagation and Skin

The ejected blood at each cardiac cycle traverses the vascular system in waves, as the blood volume increases during each systole. Physical characteristics of these waves are reflecting the cardiovascular state via several metrics, and they change based on hemodynamic state as well as tissue specifics.

2.2.1 Pulse Wave Velocity and Pulse Transit Time

We have shown in Section 2.1.3 that systemic vascular resistance plays a crucial role with BP [43]. Moreover, SVR is related to the speed of the blood flow – the pulse wave velocity (PWV) – via the Moens-Korteweg equation from biomechanics [44]:

$$PWV = \frac{L}{PTT} = \sqrt{\frac{E_{inc} \cdot h}{2r\rho}} \tag{2.3}$$

where PWV is the pulse wave velocity at which a pulse wave propagates a certain distance L in time PTT, E_{inc} is the Young's elastic modulus of the vessel wall, h is the vascular wall thickness, r is the vessel radius, and ρ is the blood density. SVR is related to and reflected in the elastic modulus on the right-hand side of the equation [40].

When the elasticity of the arterial wall is high (E_{inc} is low), the resistance to blood flow is also high, causing the blood flow to slow down, and the time it takes for a pulse wave to propagate a certain distance to increase. This time is known in literature as the pulse transit time (PTT) [44], and denoted as such in Equation 2.3. The opposite holds true when the SVR is low (the pulse wave propagates faster and the PTT decreases). In terms of pressure, when the SVR is lower, the vascular walls are less prone to deformation and generally stiffer, which causes the pressure to be larger compared to a more elastic tube [43].

The speeds and times of pulse propagation change based on the type of vessels the blood traverses, and on the distance from the heart. We report typical ranges of these parameters, alongside other relevant parameter ranges, in Section 2.4.

2.2.2 Skin Structure and Properties

The same principles described so far at a macro body level apply on the micro organ level as well. In light of our aim to develop contact-free sensing methods for parameters of the circulatory system, we focused on skin, as skin is the organ that is most easily observed directly, while being highly perfused, thus reflecting cardiovascular activity and hemodynamics [45]. The layered structure of skin tissue is shown in Figure 2.6.

The skin comprises three main layers, the top-most epidermis, the middle dermis and the bottom-most hypodermis [47]:

- **Epidermis** is the outermost layer of the skin that forms the protective layer of the body surface. It contains no blood vessels and is nourished by diffusion from the underlying dermis.
- **Dermis** is the layer beneath the epidermis and it consists of connective tissue that absorbs stress. Importantly, among other structures, it also contains plentiful small blood vessels that provide nourishment and waste removal.
- **Hypodermis** is sometimes not considered as part of the skin, however it is commonly included in research dealing with skin [47]. It attaches the skin to underlying muscles and hosts slightly larger vessels that branch into dermis.

An important property of the skin is its melanin content, which gives it its pigment. Naturally it is brown in color and the amount varies between different sites of an individual, as well as between individuals. Its purpose is to protect the cells in the lower layers against ultra-violet (UV) light, which can cause genetic damage to the deoxyribonucleic acid (DNA) [48]. A well-known numerical classification schema for human skin color based on melanin content is the Fitzpatrick scale, which classifies skin into 6 types, based on

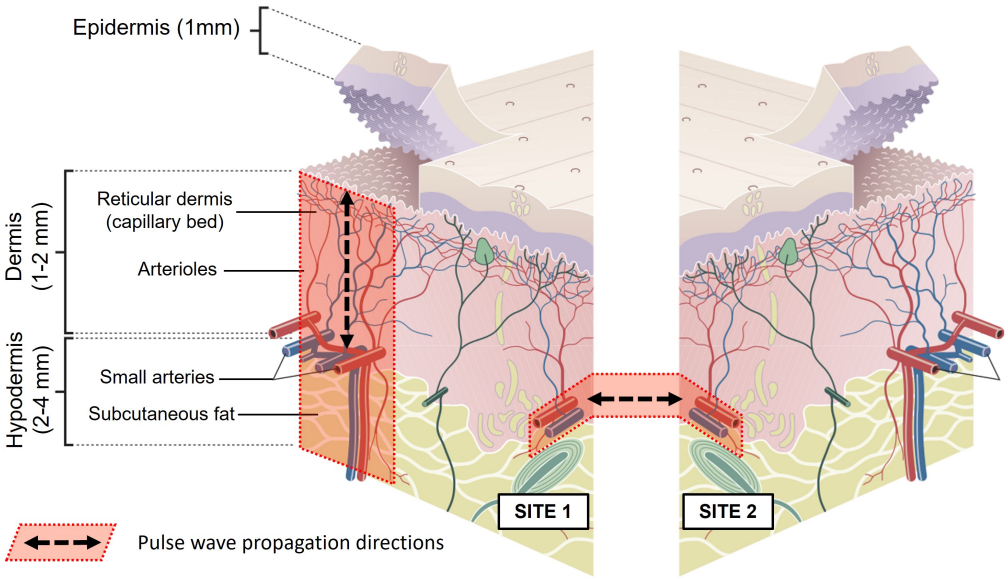

Figure 2.6: Layered structure of human skin [46], showing vascular presence and important structures. Pulse waves propagate through the vessels between different sites (horizontal arrow) and between different layers (vertical arrow).

response to UV and related melanin content [49]. Based on melanin content, the optical properties differ, which makes some skin types behave differently in regards to electromagnetic radiation and potential measurement applications thus become skin-tone-specific [50], [51]. Despite this, there is a consistently lower amount of melanin present in certain parts of the body, such as palms and soles of the feet. This is due to evolutionary reasons, as these parts were less exposed to radiation (e.g., the sun) while being frequently used. This means that less melanin was required and instead another translucent protective layer of dead skin cells called stratum lucidum developed [52].

2.2.3 Skin Perfusion

As described in the previous section, the skin, especially its dermis layer, is dense with vessels, specifically capillary loops. These are the thin final microvessels in the vascular network originating from the heart. It starts with the aorta, which then splits in the aortic arch into smaller arteries leading towards the brain and towards limbs. Arteries split further along the network into yet smaller arterioles, and finally capillaries. The latter form capillary loops, where the input is the arterial pathway described previously, which delivers oxygenated blood and nutrients, and the output are called venules, which then widen and become veins, returning blood towards the heart [53].

A network of many capillary loops is called the capillary bed. Presence and density of the capillary bed differs throughout the body, thus causing different perfusion of different tissues. The hand is an example of a relatively well-perfused part of the body, with two large arteries – the ulnar and radial artery – delivering blood in the palmar arch, as seen in Figure 2.7. From here, many arterioles split towards the fingers and general presence of capillary loops is quite high [55].

While the perfusion in the hand is often depicted in two dimensions, as shown in Figure 2.7, it actually takes place in the third dimension as well, as highlighted in Figure 2.6.

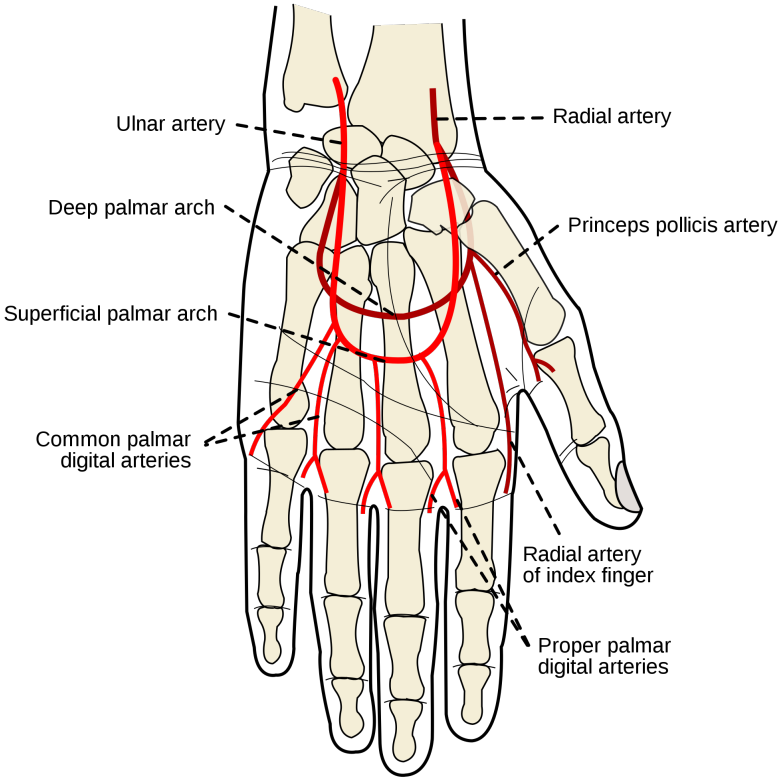

Figure 2.7: Arteries delivering blood, which allows for good perfusion of the human hand [54].

This means that pulse waves propagate from the deeper layers with larger arterioles towards the upper layers with the capillary bed, before returning through the venous network. Subsequently, the traversal of a pulse wave takes place between two locations on the arterial path, as well as between skin layers [56].

2.3 Observable and Measurable Effects

Having a good understanding of the underlying phenomena causing physiological changes in the body, we can identify the effects that are observable and measurable in a non-invasive contact-free manner. In this section, we will briefly highlight the relevant measurable effects, while a detailed and extensive overview of existing work describing the methods and results of such measurement will be given in the following chapter.

2.3.1 Respiratory Chest Displacement

The previously described pulmonary volume changes are reflected in the movement of the chest wall [57]. This movement manifests in all directions, but is most prominent in the direction perpendicular to the spine, meaning directly forward. Lateral displacement is also present, but is a lot less obvious [58].

This displacement can be measured with devices allowing for distance measurement with precision on the order of millimeters. However, as the displacement is relatively small, it can be sensitive to other larger movement artefacts that are omnipresent during the daily routine of individuals.

2.3.2 Light Absorption Changes

The previously described perfusion is periodic in sync with cardiac contractions. As each pulse wave traverses the network, it fills the vessels and tissue with increased volume of oxygenated hemoglobin-rich blood. Subsequently, after the gas and nutrient exchange, deoxygenated blood with low hemoglobin returns towards the heart and lungs, lowering the volume of blood in the tissue. This phenomenon of changes in blood volume is reflected in color changes of the skin obtained by measuring the light reflected from or transmitted through the tissue. It can be measured relatively easily with a consistent light source and a photodiode. The obtained periodic signal is called photoplethysmogram (PPG) and is commonly used to measure blood oxygen saturation (SpO2), although its range of applications and its usefulness are extensive, as we will show in the next chapter [59].

Such changes in color can also be captured remotely with devices such as RGB cameras [45], however, both temporal resolution and color sensitivity must be sufficient to capture them. We discuss the details of different options for such measurement in the following chapter.

2.3.3 Ballistocardiographic Effects

Another subtle effect of cardiac activity is the movement that can be observed as the downward movement of blood through the descending aorta produces an upward recoil, moving the body upward with each heartbeat [60]. This is known as the ballistocardiographic (BCG) effect.

While this causes movement, which can be visually observed in the wrist or neck area, and the tactile effects felt with the fingertips, the movement is very subtle and sensitive. So instead, it is usually measured in terms of changes of force exerted by the body, meaning contact-free sensors are not as feasible [61].

2.4 Typical Ranges of Measured Parameters

In the previous sections we discussed several parameters relating to the performance of the circulatory system and hemodynamics, such as the CO, SVR and the related ABP. Further, we described the pulse wave propagation and discussed PWV and PTT. Finally, we also briefly reviewed the non-invasively measurable effects of the physiological phenomena. In Table 2.1, we give an overview of typical values for some of these parameters, as reported in related work [40].

Table 2.1: Typical ranges of relevant pulmonary, cardiovascular, hemodynamic and tissue parameters by groups [40].

Group	**Parameter**	**Typical value and unit**
Cardiac timing	Heart rate	50–100 bpm
	Isovolumetric contraction time	50 ms
	Ventricular ejection time	300 ms
Blood volume	Total blood volume	5 L
	Arterial / Venous volume	2 L / 3 L
	Stroke volume (at rest)	70–80 cm^3
	Stroke volume (maximum)	110–120 cm^3
Blood flow	Cardiac output (at rest)	4–7 L/min
	Cardiac output (maximum)	20–35 L/min
Blood pressure	Systemic BP (SBP / DBP)	120 / 80 mmHg
	Pulmonary BP (SBP / DBP)	25 / 10 mmHg
Vascular resistance	Systemic vascular resistance	20 mmHg/(L/min)
	Pulmonary vascular resistance	3 mmHg/(L/min)
Pulse wave velocities	Aortic PWV (young $<$ 30 y)	4–5 m/s
	Aortic PWV (elderly $>$ 60 y)	10–15 m/s
	Capillary PWV	6–17 mm/s
Skin thickness (palm)	Epidermis	1 mm
	Dermis	1–2 mm
	Hypodermis	2–4 mm
Pulmonary volume	Total lung capacity	5 – 7 L
	Tidal volume (air moving through lungs in a single breath)	400–500 mL
Displacements	Chest wall displacement	3–5 mm
	Ballistocardiographic displacement	$<$ 1 mm

It is important to mention that the reported values in Table 2.1 are subject to some disagreement in literature, as the reported ranges of PWV span several orders of magnitude [40], [62]. These differences are the result of different vessels at which the PWV was measured, as the blood propagates notably faster in the thick large vessels close to the heart, while slowing down in the peripheral capillaries.

Chapter 3

Related Work

In the previous chapter, we briefly touched on the basic common ideas for contact-free measurement of the effects produced by the physiological mechanisms of the circulatory system. Such monitoring is not novel, but a well-developed and highly-researched field [20]. In this chapter, we will investigate existing methods and applications of such measurement in depth, with the aim of both understanding the state of the art work, as well as identifying the remaining challenges and gaps in knowledge that offer room for improvement beyond the state of the art.

3.1 Remote Monitoring of Fundamental Cardiorespiratory Parameters

A wide range of research on remote cardiorespiratory monitoring systems has been proposed using different methods, including methods based on Doppler effect, video camera imaging and thermal imaging. In this section, we review some of the state of the art methods for physiological parameter estimation [20], which allow for potential monitoring of specific fundamental parameters (e.g., HR and RR). Related work suggests these can be relatively robustly estimated from radar signals and the previously described (r)PPG signal, both of which can be obtained remotely with different sensors [63].

3.1.1 Radio-frequency Methods

It is not uncommon in practical applications (e.g., sleep monitoring) for some type of occlusion to be present (e.g., blanket, clothes, hair, makeup etc.) between the subject and the sensor. In such cases it makes sense to primarily use the part of the EM spectrum with deeper penetration and longer wavelengths. Some research deals with the IR part of the spectrum [64], going beyond visible light towards longer wavelengths. However, a large body of work goes even further, and instead investigates the microwave and radio-frequency signals for contact-free monitoring [20].

3.1.1.1 Hardware

- **Radar.** Such work commonly relies on the Doppler effect, which is an active measurement technique capable of detecting previously described pulmonary expansion and contraction, which results in changes of distance between the sensor and the observed chest wall [65]. It works by emitting a continuous wave towards the object from the transmission (Tx) antenna and then recording the reflected wave via the receiver (Tr) antenna. The reflected wave is either frequency or phase modulated

due to movement of the object. Comparing the emitted and received signal allows for computation of change in frequency and phase [66]. Subsequently, radar systems, specifically frequency modulated continuous wave (FMCW) radars, are most often employed to measure the movement of the chest wall. Although the used frequency ranges vary substantially, it was shown that the use of high-frequency radars ($>$ 10 GHz) increases the SNR of the obtained signal [67].

- **LIDAR.** An alternative to the above is to instead use laser beams to measure changes in distance, for which light detection and ranging (LIDAR) systems can be used. The latter have a shorter range compared to radars, but can provide more precise measurements at close range [68]. LIDARs are a lot less mature compared to radars and can be considered an emerging trend.

3.1.1.2 Measurable Parameters

- **Heart rate.** There are several reported studies that estimate HR from the cardiac vibrations of the chest wall [65]. Ling et al. [69] used mmWave FMCW radar to record chest movement by first removing the body motion and respiration artefacts. Following this, they proposed a heart beat extraction algorithm based on filtering and wavelet transform. It was evaluated on 20 subjects and HR was estimated with a MAE of around 4 bpm. Jung et al. [70] also used a FMCW 60 GHz radar to extract HR in short windows and compared with an ECG ground truth. They obtained similar errors of around 4 bpm. Most other work used similar hardware and idea, simply using different signal processing and separation algorithms, all achieving comparable results in terms of MAE.

- **Heart rate variability.** HRV was also thoroughly investigated. Nosrati et al. [71] used a 2.4 GHz radar placed 1.5 m from the subjects to reconstruct the cardiac signal. They trained what they call a Gaussian pulse model and a custom autocorrelation and frequency-phase regression technique to obtain a clean signal. Interbeat intervals (IBIs) were compared with a reference PPG and they reported a relative error (RE) of 1.97%. Shi et al. [72] recently proposed an end-to-end approach using bi-directional LSTM network to estimate IBIs. On 25 test subjects they achieved RE of around 5%. Again, other work used the same idea, albeit with different signal processing methods, with the aim of obtaining a clean cardiac signal that can be used to compute HRV features [73].

- **Respiratory rate.** In terms of radars, respiration monitoring is an especially mature research field, as it is arguably easier to measure compared to cardiac activity due to larger consistent displacement of the chest. An early method was proposed by Xiao et al. [74], who used a 5 GHz radar at a distance between 0.5 and 2.8 m, and achieved accuracy of 81% for HR and RR detection at maximum distance. Accuracy was calculated as the percentage of time the calculated rate (HR and RR) was within 2% of the reference rate. However, this metric is less intuitive compared to commonly reported errors in terms of bpm. More recently Yang et al. [75] proposed a more elaborate method which uses image representations of the obtained data (similar to spectrograms) and then first detects the number of present people (periodic patterns) and finally estimates RRs based on edge detection in the image. They reported MAEs of around 0.5 breaths per minute, even for scenarios where multiple people are present.

  A novel approach that enables LIDAR monitoring of respiration was proposed by Zhang et al. [68]. They initially proposed a photonic radar, which is an alternative

technique to traditional radars, using lasers to create and analyze signals in the RF frequencies. In this sense, it is similar to a LIDAR, which they also demonstrated. They analyzed the obtained respiratory signals of a cane toad (which has a high relative pulmonary expansion in relation to their body size) in terms of cross-correlation between the photonic radar / LIDAR and reference camera-extracted data, achieving a correlation coefficient of 0.72.

- **Blood pressure.** In recent years, researchers also estimated BP using radars, most commonly in combination with other sensors via PTT or pulse arrival time (PAT). The former is the delay between two pulses measured at two locations (could be using two PPG sensors or other alternatives) and the latter is the time it takes for a pulse to travel from the heart (ECG) to some location (commonly PPG). Buxi et al. [76] used a radar to acquire arterial pulsation at the aortic arch and then measured the delays between the corresponding cycles and their counterparts in the ECG. They measured data from 6 subjects and got Pearson's correlation coefficients of -0.66 and -0.48 for SBP, depending on the specifics of the measured PTT or PAT. Zhao et al. [77] proposed another BP estimation approach, again based on PTT obtained between the aortic and radial pulse waves measured with a radar. They evaluated on 3 subjects in stable conditions (stationary, holding breath, at rest) using a linear regression, and they reported errors within 3 mmHg, although details are not clear – the precise definition of the error metric is not specified and the reported values are only integers.

3.1.2 Optical Methods

The most common idea behind contact-free physiological measurement is to use a camera to capture the changes in skin color as a consequence of blood volume changes [78]. The principal idea behind camera-based physiological monitoring is to use the information from the non-ionizing part of the electromagnetic (EM) spectrum, namely visible and (near) infrared light [63]. This part of the spectrum has different penetration depths in accordance with wavelength, which makes it suitable to probe different depths of tissue – usually skin – as shown in Figure 3.1.

Alternatively, cameras can also be used to detect subtle movement relating to the BCG [79], however, this is extremely sensitive to other movement noise and has to be separated precisely. This means that practical application of camera-based BCG monitoring is more difficult and less feasible compared to the color-based PPG approach.

3.1.2.1 Hardware

- **RGB Cameras.** RGB cameras by default are sensitive to three distinct bands of visible light around 475 nm (blue), 550 nm (green) and 650 nm (red). This type of camera is by far the most wide-spread, as it is available in laptops, phones and other smart devices like mirrors, TVs, virtual reality (VR) headsets, etc. [63]. Moreover, this type of camera captures the information from the spectral band at which the signal-to-noise ratio (SNR) for PPG is the greatest, which is around 570 nm [80]. In terms of skin layer reached by light with such wavelength, it mostly corresponds to the capillary bed in the skin dermis (depends in part on measuring site and skin thickness). Hence, research dealing with physiological parameters obtained from RGB cameras is most mature [81]–[83]. The illumination source for such cameras can be ambient lighting, which is convenient, however, if a dedicated light source is needed, it interferes with the human eyes and can be unpleasant.

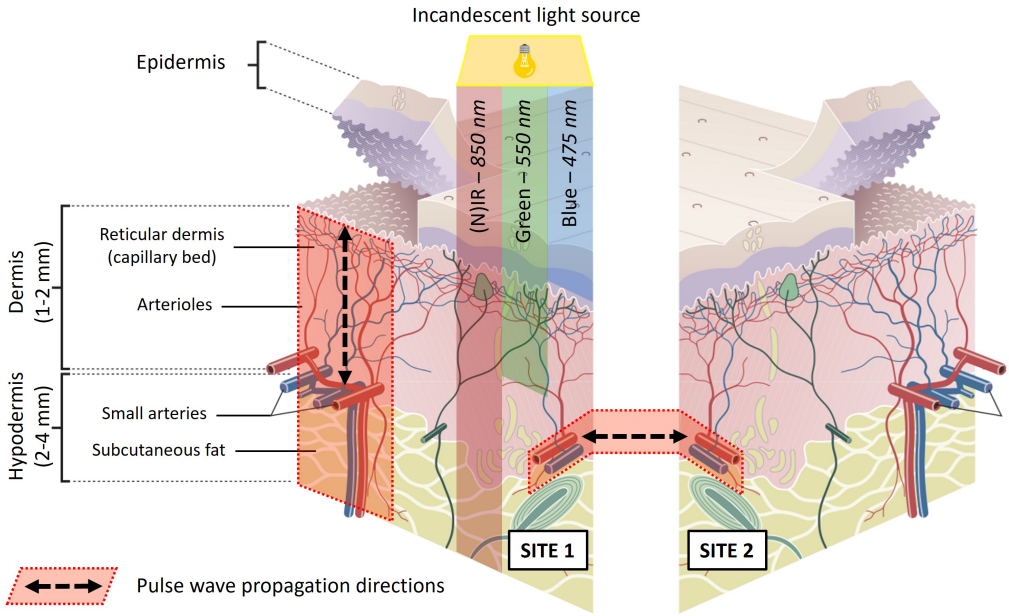

Figure 3.1: Different skin depth penetration of light with different wavelengths.

- **(N)IR and Thermal Cameras.** The second example of a relatively common hardware are those capable of capturing information beyond 700 nm, in the infrared (IR) part of the EM spectrum. This information corresponds to happenings in the deeper layers of the skin tissue and can probe beyond the capillary bed, reaching smaller arterioles [56]. Light in this part of the spectrum does not interfere with human vision, and, perhaps more importantly, such cameras can be used in low light conditions [84]. However, absorption of such near infrared (NIR) light by hemoglobin is also weaker compared to the green spectral band, so the obtained signal is expected to have a lower SNR [63]. Additionally, thermal cameras sensitive to a broader part of the IR spectrum are more expensive by orders of magnitude compared to RGB cameras.

- **Multi- and Hyper-Spectral Cameras.** Given the known advantages of certain spectral bands, resulting in better SNR of the obtained PPG signal or more motion and illumination robustness [85], one can deduce that measuring specific narrow-band parts of the spectrum would be optimal. This is made possible by multi- and hyper-spectral cameras, which are special cameras that collect information as a set of quasi images, each representing a narrow-wavelength spectral band. These quasi images are combined to form a three-dimensional (x, y, λ) hyperspectral data cube for processing and analysis, where x and y represent two spatial dimensions of the image, and λ represents the spectral dimension (comprising a range of wavelengths). However, the availability and practicality of such cameras are very limited, as this type of hardware is extremely expensive, specialized for specific tasks (e.g., physics spectroscopy) and generally difficult to use. An alternative is thus to instead use several cameras sensitive to specific parts of the spectrum and then fuse their information [86].

3.1.2.2 Measurable Parameters

Using the previously described optical PPG approach, most commonly obtained with an RGB camera, several physiological parameters were measured successfully.

- **Heart rate.** This is the simplest parameter to obtain from PPG, as it corresponds to the dominant frequency in the pulse signal. Its fundamentals lie in remotely obtaining high-quality PPG with high SNR from the camera RGB traces.

  An early approach was proposed by Verkruysse et al. [81] who used ambient lighting and a consumer RGB camera to obtain HR from the human face using mainly the green channel trace. Common updates to this method include blind source separation (BSS) methods like principal component analysis (PCA) and independent component analysis (ICA) which separate the temporal RGB traces into uncorrelated or independent signal sources to separate the pulse and the noise [87], [88]. This lets them obtain cleaner and more robust pulse waveform that allows for stable HR measurement.

  Another important algorithmic update that improved HR estimation under challenging conditions is the chrominance (CHROM) method [89], which linearly combines the three RGB color channels to build two orthogonal chrominance signals and also standardizes the skin-tone influence, making the method robust to various skin tones and illumination colors. Their method showed improvement over BSS methods – for a population of 117 stationary subjects the root mean squared error (RMSE) was lowered by a factor of 2 compared to BSS methods.

  Furthermore, Wang et al. [45] proposed a mathematical color model explaining different algorithmic choices made in previous approaches like CHROM, alongside a novel algorithm called plane orthogonal to skin (POS). It resembles CHROM, but alters the order in which the main expected color distortions are reduced, as their projection planes have different chromaticity distributions. For a detailed overview of differences we refer the reader to the original work [45]. Importantly, comparison with existing methods showed that POS outperforms other approaches in terms of SNR in a variety of situations, including different subject skin tones, luminance conditions, and subject state (stationary, during exercise, recovery after exercise).

  The above methods focused on fundamental understanding of remote PPG reconstruction and were developed early. Despite their consistent performance, as the field matured and more datasets became available, focus shifted towards deep learning approaches, with neural network (NN) architectures like 3D convolutional neural networks (CNNs) and long-short-term-memory (LSTM) networks with attention [90], [91]. Such end-to-end networks further decreased the errors in HR estimation, reaching mean absolute errors (MAEs) of under 2 bpm [92].

  The above methods mostly focused on either waveform quality in terms of SNR (early methods) or instantaneous HR estimation, meaning a single value at a single temporal window. However, HR changes in time and these changes carry important additional information [93] in the form of heart rate variability (HRV).

- **Heart rate variability.** HRV captures the changes of HR over time and is related to ANS, which dynamically adjusts HR to maintain homeostasis, as described previously. To compute it, one generally requires detecting the interbeat intervals (IBIs), which requires precise systolic peak localization [94], [95]. Recent work [96] identified weakness of systolic peak detection for IBI computation in low SNR cases and

proposed a frequency-based approach that uses an adaptive band-pass filter and discrete energy separation algorithm to determine instantaneous frequencies, which are then converted to IBIs based on changes. They validated the approach on a custom dataset of 16 subjects with different skin tones in terms of standard HRV metrics like root mean square of successive differences (RMSSD) and standard deviation of N-N intervals (SDNN), achieving 0.98 and 0.99 Pearson's correlation coefficient with contact-based ground truth.

- **Respiratory rate.** RR information was extracted from PPG alongside HR in one of the earliest works [81], showing multi-modal cardio-respiratory potential of the PPG. Early work proposed combining color channels of RGB camera to suppress noise, similarly to CHROM and POS, but with the aim of extracting respiratory signal rather than cardiac [97]. They reported MAEs of 1–2 breaths-per-minute, which surpassed the SOTA. More recently it was shown that analyzing the amplitude variations in reconstructed PPG can lead to RR [98], [99], achieving even lower reported errors of 1 breath-per-minute.

  Respiration can also be clearly observed using thermal cameras, as the exhaled air in the perinasal area has a higher temperature compared to the surrounding air, meaning a clear periodic signal can be reconstructed and RR estimated with RMSE of around 2 breaths-per-minute [100].

- **Blood oxygen saturation.** The ratio between oxygenated and deoxygenated hemoglobin was broadly investigated. One of the pioneering works in remote sensing by Wieringa et al. [101] proposed obtaining two optical PPGs from different wavelengths with a custom camera, which were correlated to respiration. The latter influences the amount of oxygen in the blood. While they proposed the fundamental idea, specific results on SpO2 were not reported. A decade later, Tarassenko et al. [102] showed in a clinical study involving over 500 patients that computing ratio of ratios feature from the AC (the pulsatile signal amplitude reflecting cardiac activity) and DC (non-pulsatile baseline variations or wandering of the signal) component of red and blue camera traces allows for SpO2 estimation. They reported the coefficient of determination of 0.64 between estimated and ground-truth SpO2 values on a large sample size.

- **Pulse transit time.** The clinical significance of PTT as a descriptor of hemodynamic state and its relation to ABP was described in detail in the previous chapter. While PTT was traditionally measured using two contact sensors at distant body locations [103], recent advances proposed contact-free camera-based solutions that record different exposed skin locations, such as palm and forehead or two regions of forehead, and reconstruct two PPG waveforms. The delay between reference points or phase shift is then computed and averaged in some window to obtain the PTT [85], [104].

- **Blood pressure.** Remote BP estimation is a highly valuable and desired measurement, however, there is some disagreement in the community regarding its measurement. Many works report successful estimation using multi-site PTT approach, where palm and forehead are used to reconstruct two distinct delayed waveforms and measure the PTT between them [105], [106]. While researchers reported the possibility of using morphological features of only a single waveform as an alternative, this was always done with contact sensors only [107], or with contact sensors as a basis for transfer learning (learned on contact waveforms, fine-tuned on contact-free waveforms) [108]. A very recent review claims that there exists little evidence

of possible BP measurement via remote PPG [63] – outside of the aforementioned multi-site PTT (as a proxy correlated with BP) and contact-sensor-based transfer learning approaches – as existing research is scarce and too limited, both in number and type of subjects. A reported outlier is a large recent study on contact-free BP estimation using a smartphone, which included 1328 normotensive subjects and reported mean errors (ME) of 0.39 ± 7.30 mmHg for SBP and -0.20 ± 6.00 mmHg for DBP, respectively [109]. While the MEs themselves are extremely low and fall well within major standards for clinical BP estimation devices, it is important to consider that they were obtained by averaging the MEs of individual instances, which meant that large positive and negative errors cancel each other out. It is thus important to consider standard deviation when MEs are reported. Generally such evaluation using MEs is not uncommon in related work, often even without standard deviations reported. It is however being increasingly supplemented with absolute errors [110], which give a more realistic overview of performance if standard deviations are not reported.

- **Perfusion index.** Perfusion index (PI) is derived from PPG signal and represents the ratio of pulsatile on non-pulsatile light absorbance or reflectance of the PPG signal [111]. It was proposed that when computed traditionally from contact PPG sensors, it can be used to monitor the interaction between peripheral and central haemodynamic characteristics, such as vascular tone and stroke volume. Several studies have proposed it as a valuable metric to assess regional or neuraxial block success, as well as haemodynamic monitoring in anaesthesia, perioperative and intensive care [111]. While its value with traditional sensors is well-established, it was also recently proposed to be estimated remotely from several smaller regions of tissue. Kossack et al. [112] showed that using a remote PPG (rPPG) reconstruction algorithm to obtain the rPPG waveforms and then compute PI at several scales (depending on region sizes) can be valuable in assessing sufficient tissue perfusion during intraoperative organ transplantation. Additionally, its value can be extended to liveliness assessment for the detection of presentation attacks to authentication systems (artificially generated videos of people).

The parameters listed above represent the set for which research is reasonably mature and the obtained results achieve performance that makes them feasible and informative for non-clinical monitoring [63]. Many are still measured only in controlled environments, although there is increased focus towards real-world applications [84]. We intentionally omitted parameters that are in early development phases and not yet thoroughly researched in terms of contact-free sensing, such as blood glucose [113].

3.2 Remote Monitoring of Complex States

We have seen that contact-free measurement of physiological parameters is a relatively mature topic in terms of related work quantity. Many physiological parameters are also known to be correlated with more complex (psychological or mental) states [114]–[119], thus computation of such parameters is often necessary and valuable beyond the parameters themselves. Naturally, researchers thus also attempted to detect more complex states in a contact-free manner. In this section, we will overview such contact-free approaches uniformly, not focusing on specific sensors or states as previously. This is because many works use several input modalities (e.g., facial expressions, physiological parameters, audio, etc. all obtainable from a single camera recording the face) and estimate several things at once (e.g., HRV and stress). We will focus on mental states that are well-established in

literature to be detectable via physiological parameters described in the previous section. Review of work dealing with complex states serves to underscore the additional value of contact-free monitoring of fundamental physiological parameters.

3.2.1 State-of-the-art Multi-modal Methods

Modern approaches commonly use a hybrid approach using several input signals and derived physiological parameters to estimate a variety of complex states. We will broadly divide them into those related to psychological (e.g., stress, cognitive load, etc.) and those related to physiological (e.g., diseases or other specific physiological conditions) phenomena.

3.2.1.1 Psychological Phenomena

Historically, the psychology community has estimated mental state objectively by monitoring the involuntary dilations and constrictions of the pupil using medical grade equipment known as pupillometers or by following eye gaze and focus using eye trackers [120]–[122]. Since then, this idea was explored further in the context of eye trackers and cameras [123], [124]. Such methods require specialized devices for eye gaze and pupil dilation measurement, which strictly restrict movement and position of the user and are almost akin to traditional medical devices, which require you to be positioned near a measuring device (e.g., ECG electrodes). Thus, recent research shifted towards using more widely available cameras, which do not focus merely on the eyes, but on the wider regions of the face and the additional information it provides.

McDuff et al. [125] investigated cognitive stress measurement by using a five band digital camera to first reconstruct PPG and estimate HRV and then use it to classify cognitive stress of 10 subjects at a distance of 3 m. Cognitive stress was described as a form of mental exertion or information processing, and is more commonly known as cognitive load. They reported 85% classification accuracy in a binary experiment (stress vs. rest). Importantly, they reported HRs to be similar between the two states, while HRV and respiration differed.

Kaur et al. [126] similarly used remote PPG to obtain HRV during rest and mental stress conditions. They used PCA and ICA to separate the cardiac signal and then computed HRV-based features such as RR-intervals. They further applied a set of classification algorithms, namely logistic regression and linear discriminant analysis, to distinguish between those two states.

Wu et al. [127] used remote PPG and facial expressions to estimate fatigue of athletes by computing a derived Rating of Perceived Exertion (RPE), which is a standard measure describing how hard your body works during physical activity, typically on a scale from 1 to 10. They trained a deep neural network from online videos showing fatigued facial expressions and reconstructed the PPG using a frequency-based method to discard movement and lighting variation artefacts. They also collected a dataset of 14 subjects and reported the Pearson's correlation coefficient of 0.82 between estimated and ground-truth RPE, using HR obtained from rPPG and features obtained from facial expressions.

Tan et al. [128] proposed a system which could describe the mental state of a driver, as traffic accidents and mental stress are strongly correlated (drivers under pressure more easily cause accidents). Their proposed driver monitoring system could not only estimate HR of the driver, but also indicated whether he is under pressure or not, as it delivered two outputs: HR and mental stress level (stress index), which is estimated from HRV features. PPG would be reconstructed remotely using an 18-bit camera to grab frontal facial frames. Importantly, this was a method proposal and did not report results of any validation yet.

Abdelrahman et al. [129] used a commercial thermal camera to monitor a person's forehead and nose temperature changes to estimate their cognitive load. To assess the effect of different levels of cognitive load on facial temperature they conducted a user study with 12 participants, showing that different levels of the Stroop test (a widely-used test where there is an intentional mismatch or conflict between information, for instance between the name of a color and the color it is printed in – the word "RED" printed using blue ink) and the complexity of reading texts affect facial temperature patterns, thereby giving a measure highly correlated with cognitive load. They focused on real-time performance, as their system detected temperature change with an average latency of 0.7 s, which is faster compared to other thermal imaging work.

Cho et al. [130] proposed a deep learning model which automatically recognises people's psychological stress level (mental overload) from their breathing patterns. Using a low cost thermal camera, they tracked a person's breathing patterns as temperature changes around his/her nostril. They transformed 1D respiratory signals into 2D respiration variability spectrograms on which a CNN was trained. They validated the approach on people exposed to two types of cognitive tasks – Stroop test and mental computation test – and achieved 85% accuracy based on self-reports of people on a two-level stress scale.

Radar was also reported as a useful sensor for mental state classification, again heavily based on respiration, HRV and its derived features. Han et al. [73] conducted a feasibility study of mental state estimation from HRV features obtained with a radar in a rest and sleep scenario. They proposed several methods for radar signal processing and then validated it on data from 10 subjects by training several classifiers using time and frequency domain features. In a robust leave-one-subject-out (LOSO) experiment they achieved 53% accuracy with the best-performing nearest neighbour model estimating four mental states (normal, fatigue, stress, sleep).

3.2.1.2 Physiological Phenomena

Our interest is mostly in contact-free monitoring of health-related or health-threatening physiological conditions, including a number of complex cardiovascular and respiratory diseases.

Common cardiovascular diseases are cardiac arrhythmias. These are irregularities in the heart rhythm, which can be split into several groups, each characterized by subtle differences in either the number of beats (too many vs. too few) or in which part of the heart it is beating incorrectly (atrial vs. ventricular). Researchers typically do not separate them in great detail, although atrial fibrillation (AF) was studied in great detail. The latter is a condition characterized by rapid and irregular beating of the atrial heart chambers.

Iyer et al. [131] proposed an RF-based approach using a mm-wave radar. They extracted heartbeat phases from the radar and fed it to a three-layer ANN to predict the presence of arrhythmia. The model was trained using the MIT-BIH database alongside several other contact signals, such as ECG. It achieved internal train accuracy of 93.9%. It was then evaluated on a test set of 15 subjects and achieved a test accuracy of 75% for binary arrhythmia classification.

Yan et al. [132] alternatively proposed an optical approach using a smartphone camera to detect and analyze rPPG signals from the face in order to detect AF. They collected such data alongside ground-truth ECG and auxiliary contact PPG data from 217 hospital patients, mostly elderly (mean age 70.3 years). Each patient recorded three successive 20-second recordings and ground-truth pulse irregularity was marked by experts. They confirmed the feasibility of discriminating AF from sinus rhythm remotely, with high sensitivity (95%) and specificity (96%), while the model was trained on contact ECG and PPG.

They expanded on their previous work in 2020 [133] by proposing an end-to-end deep learning approach which detects AF from facial video with 95% sensitivity and 98% specificity. Several other authors also recently reported similar success in remote AF detection using deep learning approaches [134], [135].

While most work focused on cardiac arrhythmias, the latter can also be respiratory, describing irregular respiratory patterns. An early RF-based approach for respiratory sinus arrhythmia detection via Doppler radar was proposed by Massagram et al. [136]. They extracted respiration signals of 12 human subjects directly from a radar system. The subjects were recorded in two positions, seated and supine. They showed high correlations with ground-truth respiratory piezoresistor chest belts in both positions, using Bland-Altman plots, but did not report correlations specifically.

An extreme form of respiratory irregularity is sleep apnea. Du et al. [137] proposed ApneaRadar, an early contactless sleep monitoring system to detect sleep apnea using a commercial off-the-shelf radar. They proposed signal processing techniques that allow for RR detection via FFT and validated the approach on a single apnea-diagnosed subject measured for 7 nights. They achieved 90% accuracy in apnea detection compared to ground-truth polysomnograph.

Other authors also proposed alternative RF-based systems for detection of apnea. Koda et al. [138] used support vector machines (SVMs), which were fed vectors representing spectrograms of the preprocessed radar signals. They validated the model on two apnea-diagnosed subjects achieving accuracy of 79.5%, and precision and recall of 71.2%, again compared to a ground-truth polysomnograph.

A large-scale study was conducted by Zhou et al. [139]. They again used a radar system for measurement of apnea-hypopnea index (AHI), which is the combined average number of apneas and hypopneas that occur per hour of sleep. Data of 176 participants were labelled by physicians for ground-truth AHI. They did not disclose the algorithmic details of AHI computation from the radar, but reported high correlation coefficient of 0.98 between radar and ground-truth AHI.

Some authors also investigated parameters relating to tachycardia. Xia et al. [140] proposed a radar system for precise beat-to-beat interval and cardiac timing monitoring. Specifically, they developed a decoding peak detection method to address precise heartbeat peak extraction problem by decoding the most likely sequence of states from the single-band frequency envelogram of the radar signal. They compared their method with the ECG as the gold standard. Experimental results of 6 subjects showed a high F1 score of 0.93 when classifying cardiac peaks, and also low mean relative errors between 0.37% and 1.15% for beat-to-beat interval estimation. They argue usefulness of such an approach for both tachycardia and arrhythmia detection, although their test set included very few subjects.

3.3 Synthesis and Hypotheses

Contact-free sensing is evidently a mature research field with plentiful work dealing both with fundamental cardio-respiratory parameter monitoring, as well as more complex state analysis, such as mental states and varying diseases. Contact-free monitoring of some parameters, such as HR, is becoming increasingly robust and is reaching near-clinical levels of performance with very low errors, even in challenging conditions [20]. However, for assessment of more complex parameters and states, the technology is not yet widely used in real-world scenarios, as limitations and challenges remain [63].

In light of the above, we identified two important areas that have a profound impact on the health and well-being of individuals, but so far lack robust solutions or often rely

on contact sensors for feasible monitoring, making them especially important for subjects unable to use such traditional approaches.

1. **RF-based sleep monitoring.** RF-based methods are arguably superior to optical methods for sleep monitoring due to several reasons. First, regular RGB cameras rely on visible light, which is obviously not available in sleep settings, so one must instead use infrared or thermal cameras, which are rarer and more expensive. Even so, direct body exposure in a bed is relatively poor due to blankets, and the body position can change frequently as the person turns during their sleep. Subsequently, a recent review [141] gave an overview of the more feasible RF-based sleep monitoring methods, focusing especially on sleep apnea, which is an important, potentially life-threatening condition. Some of the main deficits and required improvements that they identified were the lack of data, relatively simple and common methods used for classification, and narrow focus on signal processing of the respiratory signal in isolation. Such systems, while valuable, are only research prototypes and not commercially available, although large companies like Amazon were recently reported to focus precisely on development of radar-based apnea detection for their Alexa device [142].

   First of the deficits above was addressed by a dataset released in 2020 by Schellenberger et al. [143]. Their main purpose was to record and make available a robust dataset with many participants and varied accompanying signals, including a myriad of traditional contact sensors, as well as radar data. The subjects underwent several cardio-respiratory scenarios on a tilting table, which induced complex hemodynamic changes. We will describe the dataset in more detail in the following chapter. Importantly, this dataset offered a means to address the other issues mentioned previously – it allows for experimentation with more complex ML methods, including end-to-end neural networks, and it enables multi-modal learning and comparison between different modalities, especially contact and contact-free sensors. Such end-to-end methods have the potential for robust real-world implementation, as seen in many other domains [144].

2. **Blood pressure estimation.** BP estimation is especially interesting due to its previously described importance, however, despite notable efforts [40], cuffless BP estimation remains elusive due to both physiological factors – higher complexity of hemodynamics depending on many factors (vascular stiffness, blood composition, fat deposits, etc.) – as well as technical limitations – BP estimation from (r)PPG commonly relies on precise morphology, requiring diastolic peaks and notches to be precisely reconstructed [107].

   Several conceptual approaches have been proposed and mentioned in previous sections, some relying on precise waveform morphology possible to obtain from a single sensor, and others instead using the known PTT or PAT approach with two sensors (most often ECG and PPG). Each approach also has its own set of challenges and drawbacks.

   PPG morphology approaches [107] are used both with contact and contact-free sensors and rely on precise morphological analysis of the (r)PPG waveform on per-cycle basis. This means that there is reliance on consistently high-quality waveforms and not only a single prominent reference point (e.g., systolic peak), as many features require consistent detection of the diastolic peak and even dicrotic notch, which are often very difficult or impossible to obtain outside of highly-controlled lab environment with high-quality contact sensors [145]. This makes many widely-investigated

features such as augmentation index, stiffness index, systolic and diastolic times, amplitude ratios etc. [146] less useful when the waveform is not ideal. Furthermore, there is some debate in the community on which morphological features are universally performing well, as the underlying connection with BP is not as clear as with PTT [63]. To circumvent the required explicit definition and computation of features, many people rely on black-box models (neural networks) to internally compute features in end-to-end approaches, however, there is an even larger lack of well-understood connection between such neural-network-derived features and BP. Due to aforementioned factors, the performance and feasibility diminishes quickly when moving from high-quality contact sensors to (consumer) RGB cameras, as the waveform details become less apparent.

Traditional contact PTT approaches [147] using wearable sensors require two precisely synchronized sensors with good skin contact to obtain high-quality waveforms for reference point detection and subsequent PTT computation. This requires two devices at two skin locations and cannot be implemented on a single compact device. Furthermore, wearable sensors are battery dependant and cannot be used by people with specific skin conditions. Alternatively, contact-free PTT approaches [99] using cameras can be implemented with a single sensor (camera), but require a good consistent exposure of two pre-determined monitored regions of interest (ROIs, commonly forehead and palm), which imposes requirements and restrictions to a subject's positioning, making it impractical. Furthermore, when these ROIs are not precisely fixed in a camera frame, tracking and segmentation is required which introduces additional algorithmic and computational requirements.

In the domain of contact PTT measurement, some approaches were suggested to use signals from different wavelengths and tissue depths. Such MW PTT approaches [56] require specialized hardware, including image sensor sensitive to the required wavelengths and a light source capable of producing said wavelengths (typically narrowband light emitting diodes – LEDs). While this is an improvement over the traditional approach of using two contact sensors, as it omits the requirement for precise synchronization and additional power, it is not readily available in existing devices and still suffers from the skin contact requirement. Importantly, wearable sensors also inherently compress the skin slightly in an effort to maintain good skin-sensor contact. Such compression can substantially distort the PPG waveform [148], making it less reliable [104] especially in the upper skin layers corresponding to shorter wavelengths.

It would be sensible to investigate the feasibility of such a MW PTT approach in the contact-free domain, as it would resolve some of the issues mentioned above (e.g., skin compression), while offering a means for non-invasive BP estimation without wearables, based on the well-understood PWV and PTT principles.

3.3.1 Hypotheses

The two areas and challenges highlighted in the previous section are vital for cardiorespiratory monitoring and should be investigated further with the aim of developing feasible contact-free methods that would allow for continuous monitoring for subjects unable to use existing approaches. Accordingly, we set the following goals of our research and this dissertation:

1. Based on our detailed literature review alongside analysis and identified challenges of existing methods, we have found that RF-based sleep monitoring and apnea detection

relies almost exclusively on the reconstructed respiratory signal (chest displacement) or rather explicit features derived from such a signal. An end-to-end deep-learning-based approach using both raw data as well as derived signals could help train a more robust and accurate model, without the need for elaborate signal processing and explicit feature computation. We aim to design and robustly validate such a model.

2. In terms of BP estimation, many existing challenges could be resolved by merging contact-free and multi-wavelength approaches to derive a hybrid remote PTT measurement system that would allow for more robust and physiology-based feature computation, which could in turn be used for accurate BP estimation.

In line with this reasoning, we came to the following three hypotheses:

1. **Hypothesis 1:** Contact-free sensing can offer comparable performance to traditional contact approaches in monitoring of complex hemodynamic conditions and vitals (apnea detection via respiration and blood pressure estimation).

2. **Hypothesis 2:** Fusion of different sensor and input modalities (contact + contact-free, temporal + frequency) can achieve superior performance and better robustness compared to using individual modalities.

3. **Hypothesis 3:** PTT can be measured between different skin layers leveraging different penetration of light using a customized off-the-shelf RGB camera and used as an informative feature for single-site contact-free BP estimation, achieving established medical standards for BP estimation in terms of error.

The first hypothesis is general and is based on literature. The second and third follow directly from our goals, describing our expectations for the proposed approaches and the corresponding results obtained via robust validation.

All the hypotheses will be discussed alongside the results obtained in our experiments in Chapter 6. The obtained results will be the grounds on which the hypotheses will be accepted or rejected, and will be reported in detail together with their interpretation.

Chapter 4

RF-based Classification of Hemodynamic Scenarios

This chapter deals with RF-based radar monitoring of physiological states, where we specifically focused on sleep apnea (and other hemodynamic scenarios) detection using a novel multi-modal deep-learning-based approach [149]. Our goal here was to successfully and robustly detect complex scenarios present in the analyzed dataset, especially sleep apnea, which is an important condition with severe real-world implications [150].

4.1 Data and Materials

Importance of high-quality data for development and validation of ML models is known, and it further solidified with the rise of large deep learning models powered by plentiful data [151]. Generally there are many available datasets in popular ML domains, such as image analysis and natural language processing (NLP) [152]. However, quantity and availability diminish in more specialized domains, where the required hardware for the data collection becomes rarer and often more expensive. This becomes quite apparent in the RF sensing domain, as, despite the myriad of related work discussed previously, most data is proprietary, recorded with custom hardware, and not made publicly available. Furthermore, as such recording setups are usually specialized and unique within the research group conducting the experiments (prototypes, not designed for deployment at scale), data can reflect some sensor or environment specifics, and the subject count can also be relatively small, which is again undesired when evaluating feasibility or robustness of a method.

While data would ideally be consistent, plentiful, varied and of course readily available, the reality is that when novel methods are considered and proposed in the inception stage of research, this in turn sets the requirements for the data needed for evaluation, which then drives the researchers to collect such data. As the purpose of most research is to evaluate a method, the data are often not the ultimate goal, but the means to an end, which is again why many datasets used in related work are unique and not shared, as they are not the focus of the research.

Finally, whenever human subjects are involved, special attention and care must be given to anonymization and privacy. Even when those are considered, many people are still reluctant to have their health-reflecting data available online, which further limits availability.

This same reasoning applies to the work described in this dissertation as well. In order to develop and validate multi-modal end-to-end models for radar-based apnea detection and sleep monitoring, data from a relatively specific and complex setup is required. It should importantly include different continuous modalities, meaning both radar and traditional

contact sensors, as well as reflect different ground-truth hemodynamic and apnea-related scenarios of many subjects.

We already briefly touched on the radar dataset recorded and made available by Schellenberger et al. [143] in 2020. This dataset was pioneering in the sense of its availability, scope, quality and consistency, and served as the enabler of other RF-based physiological monitoring research afterwards [72], including the work described in this dissertation. We describe it in detail in the following sections.

4.1.1 Hardware and Radar Physics

For their [143] recording setup, a 24 GHz continuous wave radar system based on Six-Port technology was extended into a portable radar system and a bistatic antenna design was used to improve signal quality. The inclination angle of the antenna beams was $\pm 10°$ for transmitter (Tx) and receiver (Rx) antenna, respectively, with a focal point at 40 cm. The system was placed in front of the subject chest. For further technical implementation details of the recording setup we refer the reader to the original papers [143], [153].

Movement in front of the antennas causes a measurable phase change $\Delta\phi$ between the Tx and Rx signal, which can be converted into a displacement change Δx with the known wavelength λ of the Tx signal using the following equation:

$$\Delta x = \frac{\Delta\phi}{2\pi} \cdot \frac{\lambda}{2} \tag{4.1}$$

The radar however does not produce the displacement changes of interest (distance) directly, but rather two raw signal components I (In-phase) and Q (Quadrature) are initially digitized simultaneously using a 24 bit analog-to-digital converter with a sampling frequency of 2000 Hz. The I and Q signal are used to calculate $\Delta\phi$ by arctangent demodulation, after an initial compensation for nonidealities, called ellipse reconstruction, is made. Precise details are again described in the original work by Schellenberger et al. [143], [154].

In addition to the radar system, a reference system was used in parallel to record precise contact signals with electrodes placed on the upper body of each subject and cuffs placed on upper hands and fingers. The Task Force Monitor (TFM) 3040i from CNSystems Medizintechnik GmbH was used to record the following contact signals [143]:

- **Impedance of thorax (Z0):** Electrical resistive impedance changes in relation to pulmonary air volume. These were used to measure respiration.
- **Electrocardiogram (ECG):** A three-channel ECG was initially used to record electrical activity of the heart. Lead 3 and the augmented limb leads were calculated from two raw channels, but not exported to be used in the final dataset [143], so only two channels were ultimately available as seen in Figure 4.2.
- **Impedance cardiogram (ICG):** ICG provides insight into the impedance change of the thorax by applying alternating small current between two electrodes on the body [143]. The measured voltage is proportional to impedance based on Ohm's law.
- **Blood pressure (BP):** The TFM enables continuous non-invasive BP measurement called Continuous Noninvasive Arterial Pressure (CNAP), which is measured by combining the measurement of an oscillometric BP cuff and a cuff at the fingers measuring vascular unloading [155].

A block diagram showing the architecture of the measuring system is shown in Figure 4.1 and is adapted with permission from [143].

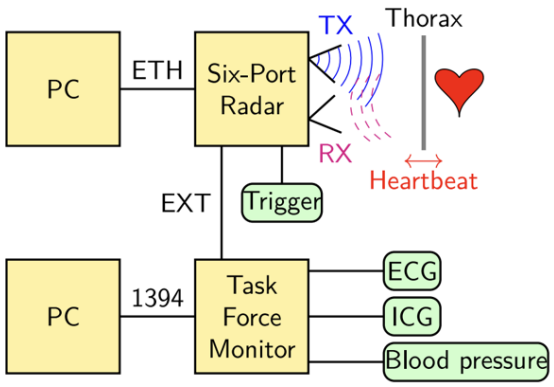

Figure 4.1: Block diagram showing the architecture of the data collection system. ETH is an ethernet connection and 1394 stands for IEEE 1394 (more commonly known as Firewire), which is mostly used to connect to peripherals. Credit to Schellenberger et al., used with their permission [143].

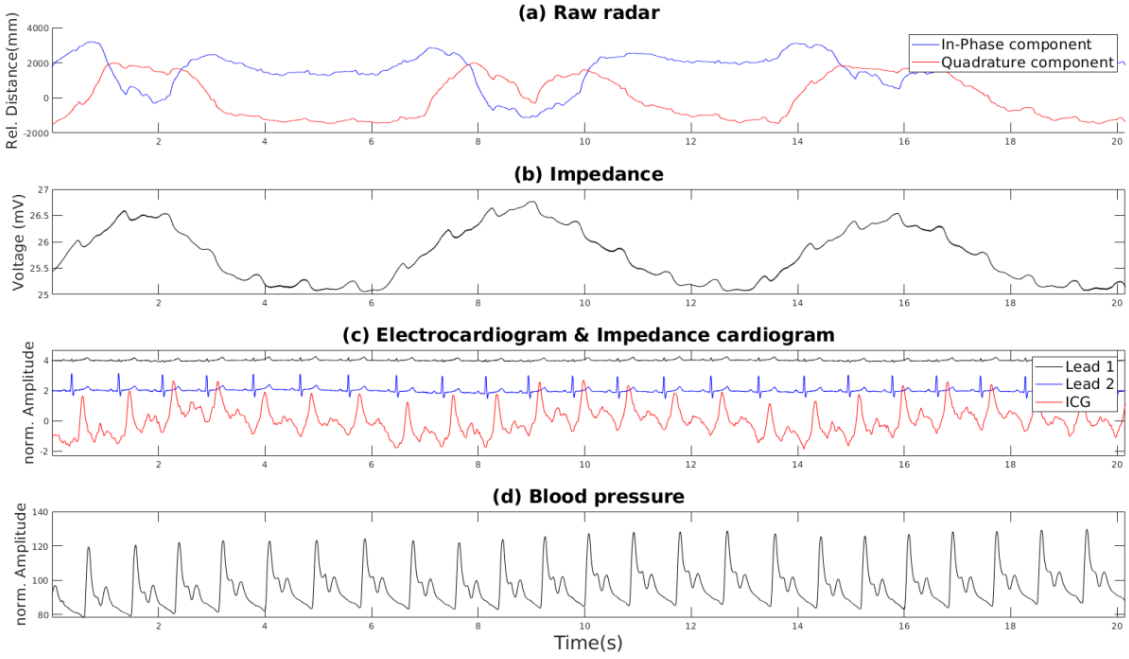

Figure 4.2: An example segment showing the initial raw signals in the dataset captured during a resting scenario. (**a**) shows the raw radar I and Q components, (**b**) shows the impedance Z0, (**c**) shows the cardiograms, and (**d**) shows the continuous BP.

An example 20-s period of these raw signals is shown in Figure 4.2 and the recording setup is shown in Figure 4.3, both adapted from the original paper with authors' permission.

Even though only 5 distinct signals were recorded in raw format, the authors of the dataset provided MATLAB scripts that allow for computation of several additional derived signals. For instance, the distance signal showing the chest displacement is computed from the raw I and Q components of the radar. This distance can in turn be used to estimate respiratory waveforms. Similarly, the raw radar components can be used to compute a rough approximation signal of cardiac activity captured by the radar. The derived signals that we computed with the accompanying MATLAB P-code (purposely obfuscated code

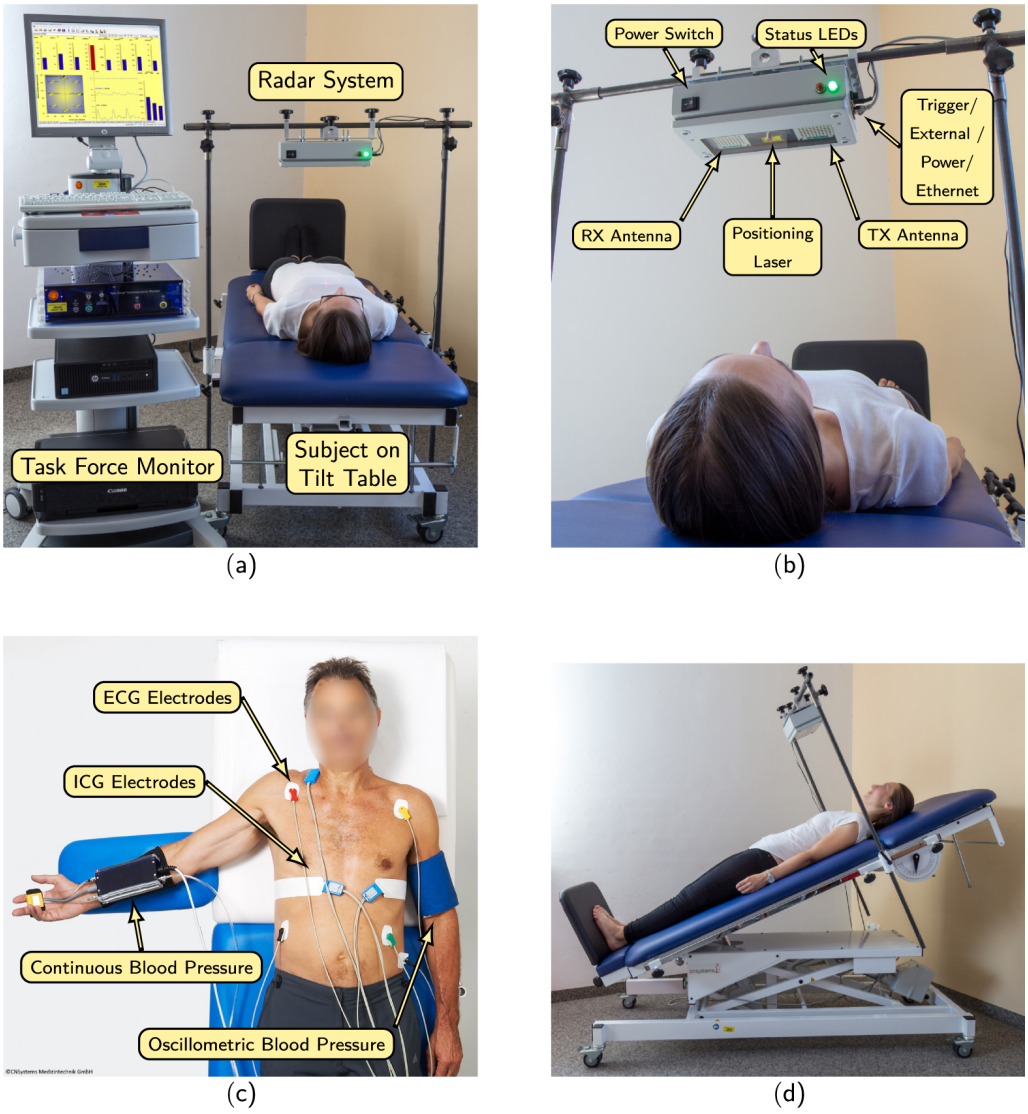

Figure 4.3: The recording setup which was used to collect the data. (**a**) Full system including reference TFM, radar setup and tilting table. (**b**) Details of the radar module. (**c**) Reference signals collected with contact sensors. (**d**) Subject in one of the tilting table scenarios. Credit to Schellenberger et al., used with their permission [143].

format for proprietary functions) provided by the authors [143] are the following:

- **Distance:** Change in distance corresponding to displacement of the chest wall, computed from the raw I and Q radar components.
- **Radar respiration:** Respiration based on the changes in distance computed previously.
- **Heart sound:** Rough approximation of the cardiac signal based on the raw radar components. Heart sounds are expected to be detected at measured sites in the form of vibrations along thorax and large vessel walls.
- **Pulse:** More refined cardiac signal based on the heart sound signal computed previously.

- **Contact respiration:** Respiratory signal based on the impedance recorded with the contact sensors.

This computation of additional derived signals gives additional data sources or modalities to be potentially used when training an ML model. Example segments of the derived signals are shown in Figure 4.4.

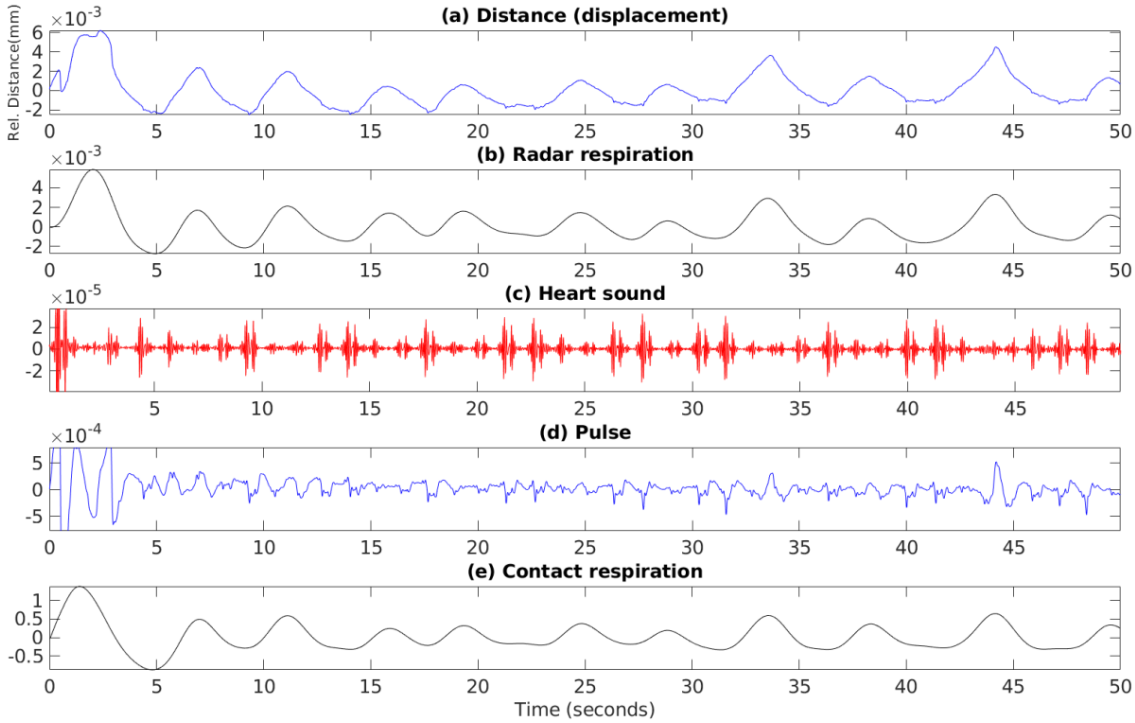

Figure 4.4: An example segment showing the derived signals in the dataset computed from the raw signals via scripts provided by the original paper authors [143]. (**a**) shows the raw distance or displacement of chest as computed from the I and Q radar components, (**b**) is the smoothed version of (**a**), (**c**) shows the heart sound estimation from the radar data, (**d**) shows an approximation of heart pulse computed from (**c**), and (**e**) shows the contact reference respiration. Note that some transitions or computations of these derived signals are part of proprietary code and the source code is not disclosed by the dataset paper authors [143].

Despite the fact that details on the computation of these derived signals are not known, as the source code was not available, we can still deduce a feasible interpretation on some signals from a visual inspection. For instance, looking at the signals shown in Figure 4.4, we can reasonably assume that the radar respiration and contact respiration are simply a filtered version of the distance and impedance signals, and are thus highly correlated and offer similar information.

4.1.2 Recording Setup and Data Collection Protocol

The data were collected in a controlled setting by physicians at the Department of Palliative Medicine at the University Hospital Erlangen [143]. They recorded 30 healthy subjects, 14 male and 16 female, with an average age of 30.7 ± 9.9 years and an average body mass index (BMI) of 23.2 ± 3.3 kg/m^2. These were generally healthy people without any known current or previous impactful medical conditions. In total, they recorded roughly

24 h worth of data, but not all scenarios were recorded for all people, as a few were either unavailable or incapable of performing certain scenarios.

The volunteers were strapped to a tilt table and were first subjected to an initial 10 min relaxation phase (lying down, calm breathing, minimal movement), which served to calibrate subjects to a default rest state. After the initial calibration phase, the subjects underwent the following 5 scenarios:

1. **Resting:** After the relaxation phase, the subject continued to lie relaxed with calm breathing, without any major movement.

2. **Valsalva maneuver:** In this scenario, the Valsalva maneuver is performed 3 times with pauses in between. The maneuver is defined as a forced expiratory effort against a closed airway [156], and was instructed to be performed for 20 s. After completion, the subject exhales and continues breathing normally. There is a 5 min recovery period in between each of the three Valsalva performances.

   This maneuver causes a complex cardiovascular response, reflecting in the BP, HR and RR, allowing for detection of potentially abnormal responses in subjects with possible medical conditions [156].

3. **Apnea:** In this scenario, the subjects held their breath twice, for as long as possible. In the first simulation, the subject inhaled completely before holding the breath (large pulmonary volume and extended lung), while in the second simulation, the subject exhaled completely (small pulmonary volume and contracted lung) before holding breath. This ensured apnea simulation in both scenarios, either with expanded or contracted lung.

   This simulation obviously causes important changes in RR, but also influences other hemodynamic parameters.

4. **Tilt up:** Here the tilt table was raised from a horizontal to a vertical position, which leads to changes in BP and HR. The measurement starts in a horizontal position, then the table is slowly raised to 70° and the measurement is continued for 10 min.

   Full-body tilt at slow speed also elicits a cardiovascular response, however the severity of it depends on the subject and their condition.

5. **Tilt down:** This is the opposite of the previous scenario. The upright table is lowered down to the starting horizontal position (simulating a slow backwards falling motion) and the measurement is again continued for 10 min.

Some summary metadata about the dataset is given in Table 4.1. It is important to note that this distribution is according to the original data, where a whole scenario session was marked with the same label, even if only parts of it were the actual simulations. For instance, people only performed the Valsalva maneuver three times, each repetition lasting for about 20 s, with 5 min recovery periods in between. However, the whole recording in such a session was marked as "Valsalva", even though the recovery period is likely something between "Valsalva" and "Resting" in terms of physiological state. We addressed this problem and discuss it more in the following chapter on data processing, also revisiting the distribution of data after this was taken into account.

In summary, this dataset has its pros and cons. On the positive side, it has a relatively high number of subjects recorded in a robust controlled sleep-like environment (imitation of lying in a bed) with many accompanying signals that offer a multi-modal cardiorespiratory overview. The subjects also underwent several complex hemodynamic scenarios that

Table 4.1: Metadata describing the quantity and distribution of data per-scenario. Subject count denotes how many subjects (out of 30) participated in a scenario.

Scenario	Subject Count	Duration	Fraction of All Data
Resting	30	5.3 h	22%
Valsalva	27	7.8 h	33%
Apnea	24	1.3 h	5%
TiltUp	27	4.8 h	20%
TiltDown	27	4.8 h	20%

elicited changes in HR, RR and BP, including sleep apnea simulation, which is an especially dangerous condition of interest. Finally the recorded data is also of exceptionally high quality in terms of captured waveform detail due to high-grade clinical devices and sensors used.

On the downside, such controlled environment does not simulate the real world very precisely, where there is usually more movement noise. Furthermore, all the subjects were healthy, so potential disease-induced anomalies in cardiovascular responses were not expected or accounted for. However, for early studies and evaluation of novel methods, it is arguably better to start with data from a controlled environment where things can be more clearly defined and observed.

4.2 Methodology

This section serves to describe the designed experiments in detail, together with reasoning for choices made and pitfalls to be wary of. While this section focuses on the methods, algorithms and experimental design, all of the results originating from these experiments are subsequently reported, interpreted and discussed in the following section.

4.2.1 Pipeline Overview

In order to use the radar dataset described in the previous chapter, we developed a robust pipeline that enabled reproducible evaluation of our proposed models. The input to the pipeline were the raw signals available in the dataset, alongside the corresponding labels of a specific scenario. Derived signals were first computed from the raw data using the authors' scripts, as described in the last chapter. All the signals were then preprocessed with various signal processing techniques, with the aim of standardizing the data in terms of sampling frequencies and removing unwanted and uninformative noise. The data were then segmented alongside the labels into discrete windows using several different windowing options. Distribution of the data in terms of labels was then investigated and modified with the aim of obtaining as uniform distribution as possible for training. Subsequently, a variety of neural network architectures were investigated in terms of different types, input modalities and topologies. Finally, the models were evaluated in a robust manner in order to avoid overfitting as much as possible and different classification performance metrics were monitored. This high-level pipeline is illustrated in Figure 4.5, and we will describe each part in detail in the following sections.

4.2.2 Data Preprocessing

We wanted our data to have the same unified sampling frequency, and to not contain any undesired and uninformative noise, while maintaining all the relevant physiological

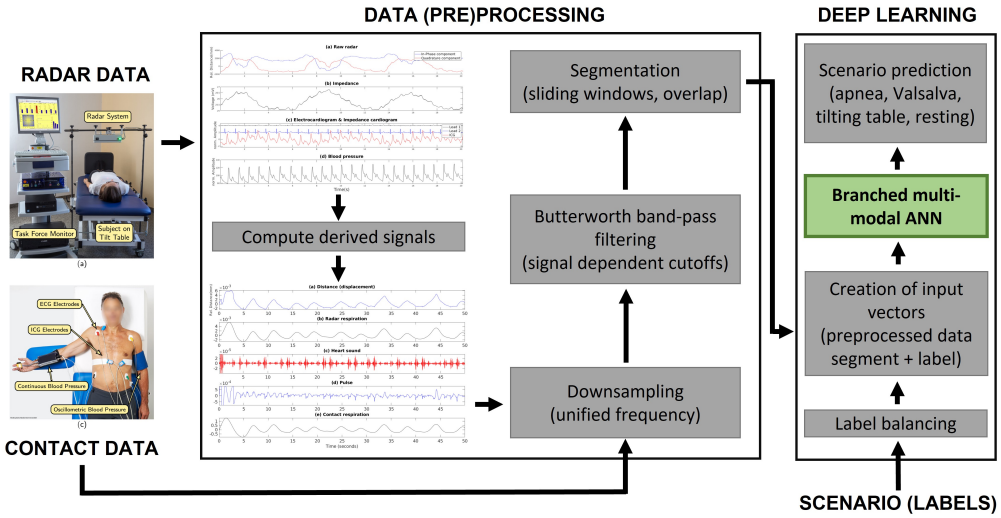

Figure 4.5: High-level pipeline of our proposed system for classification of hemodynamic scenarios using radar data. The green box denotes our main contribution in the form of a custom branched multi-modal ANN.

information in each waveform.

4.2.2.1 Downsampling and Filtering

The original data are sampled with relatively high and varied sampling frequencies from 100 Hz (impedance) to 2000 Hz (radar). While this is good in the sense of obtaining precise changes in the signals, it is not actually needed when considering the physiological phenomena of interest. Cardiorespiratory activity revolves around breathing and heart beats, both of which occur periodically with a much lower frequency. As the data are plentiful and this physiological information is not contained in high frequencies, we initially downsampled the signals to a unified lower sampling frequency of 100 Hz. While this value allowed us to keep the important frequency components related to RR (< 1 Hz) and HR (1 – 3 Hz), it still caused us to lose some very fine-grained information from the ECG [157]. This was a tradeoff we made in order to lower the size of inputs while keeping the intuitively important frequency components. However, the discarded high-frequency information in the ECG might also be valuable and should be considered in future work discussed in Section 7.2. We downsampled using the standard built-in MATLAB *resample()* function, which applies a final impulse response (FIR) Antialiasing Lowpass Filter to the input and compensates for the phase delay introduced by the filter [158].

Additionally, the data were filtered using a 4th order Butterworth band-pass filter with signal-specific cutoff frequencies. The lower cutoff was usually set at 0.1 Hz to remove the baseline drift while the upper cutoff ranged from 1 Hz for respiration-related signals to 25 Hz for the ECG, as the latter may contain information in the higher frequencies comprising the QRS complex [159].

4.2.2.2 Segmentation and Labelling

Recording session data were continuous and lasted from a few minutes to tens of minutes, depending on the session and scenario. We used a rather wide-spread and standard approach of first windowing the signal into shorter overlapping windows, which represent

instances in our machine learning (ML) pipeline. The window length (duration) and overlap amount are commonly discussed parameters and should be set correctly to capture the relevant information in each window. We investigated several different window lengths between 5 s and 20 s – the range was defined in accordance with the fact that we wanted the windows to capture at least one (but preferably a few) respiratory cycle and the corresponding frequency information. The upper limit was set to 20 s and not longer, as some simulation scenarios only lasted for that long. Additionally, we used 50% overlap between windows, which increases the number of instances but still keeps them varied. Finally, we also normalized all the input data to the same [0, 1] range, since this is common practice that helps with the training process, as the model weights are on the same scale every time, rather than being orders of magnitude apart.

After the signal data were windowed, a corresponding class label had to be assigned to each instance. This was not completely trivial due to the fact that a whole recording session was always marked with a single label (e.g., "Apnea" or "Valsalva"), but not all the signal data actually corresponded to this class. Instead, in some cases only a minority of the signal data in the whole recording corresponded to the actual marked scenario, like in the Valsalva scenario, where the maneuver was done on the scale of tens of seconds, while the recovery period was on the scale of minutes. A continuous electrical signal is provided with the dataset, in which button presses can be identified by sharp drops in the signal amplitude. A button was typically pressed at the start and end of the Valsalva maneuver and at the start and end of the tilting motion. For apnea, however, the button was pressed and *held* throughout the simulation. Illustration of presses and their meanings for each scenario is shown in Figure 4.6. On top of these differences between scenarios in button press durations, an additional concern was the fact that in a specific scenario, the number of button presses is not always the same and subsequently the meaning of button presses changes as well. This is a consequence of the authors' choice to change the number and meaning of button presses in later recording sessions compared to the early ones.

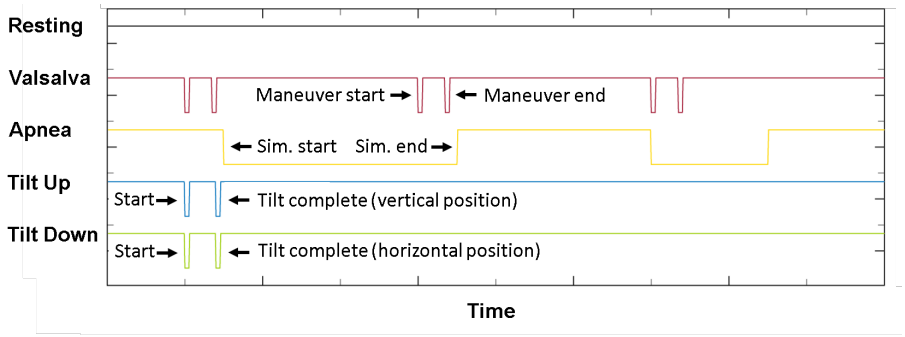

Figure 4.6: An example showing the button presses in the electrical signal and their meaning for each scenario (in majority of cases). Adapted from the original paper with authors' explicit permission [143].

Using the semantics of button presses described above, we assigned a class label to each segment or window. The periods in which scenarios were actually performed were clearly labelled as the actual scenarios while the periods in which the physiology was ambiguous were initially labelled with as "Other".

Going into more detail, we first created a discrete mapping between each scenario and an integer label (e.g., "Resting": 1, "Valsalva": 2, etc.). We then created a new ground truth array for each label signal originating from the button presses (unlike before, where the whole recording had a single uniform label). In each of these label arrays, we only

set the class value in periods where the scenario was actually being executed or simulated (e.g., where the Valsalva maneuver was performed, where the breath was held in Apnea simulation, etc.). All other periods (e.g., recovery periods, tilting table movement, etc.) are not clear in terms of physiology, as they might be influenced by the scenario but are also close to resting in nature. Due to this ambiguity we marked such periods with class "Other" to separate them from the clear classes. Due to this, there are some windows which overlap a period of "Other" and the actual scenario – in such cases we took the majority class for that instance.

After all the preprocessing, segmentation and labelling steps were complete, the new distribution of instances in terms of class is more heavily skewed than the one initially reported in Table 4.1, as seen in Table 4.2. Most importantly, the number of instances for "Valsalva" is actually much lower (closer to "Apnea") due to the long recovery periods being put into the "Other" class.

Table 4.2: Metadata describing the quantity and distribution of windowed instances per-scenario. It is quite different both in duration and distribution from the original labels due to the introduction of the label "Other" and overlapping windows.

Scenario	**Nr. Instances (20 s window)**	**Duration**	**Fraction of All Data**
Other (0)	3021	16.8 h	36%
Resting (1)	1865	10.3 h	22%
Valsalva (2)	173	1 h	2%
Apnea (3)	184	1 h	2%
TiltUp (4)	1575	8.7 h	19%
TiltDown (5)	1630	9.0 h	19%

4.2.2.3 Class Balancing

The emergent heavy imbalance in the data after segmentation is troublesome, since ML algorithms have difficulties learning from just a few instances of some compared to other classes. Naturally we wanted to keep as much data as possible, especially since we focused on DL approaches where more data are especially valuable, so we did not wish to undersample the data to fix the imbalance. Instead we decided to oversample the minority classes using the Synthetic Minority Oversampling TEchnique (SMOTE) [160]. It belongs to the group of magnitude-based pattern mixing with interpolation [161], meaning the generated synthetic instances will contain changes in magnitude, while preserving the temporal properties. Temporal warping, slicing or permutation of periodic time series data can be dangerous, since synthetic instances might contain anomalies that are impossible in real world, such as infeasible frequencies or sudden cut-offs in the waveform.

SMOTE synthesizes new unique instances of the undersampled class by selecting an instance and then finding k of its nearest neighbours. A randomly selected neighbour is then chosen and a synthetic instance is created between the two examples. This creates plausible new instances which differ from the original ones, especially if k is sufficiently large to avoid always picking nearly identical instances. The distance metric to select the nearest neighbours was the Euclidean distance, which is suitable for finding similar instances that are aligned (cycles at same locations), while punishing instances that have a phase shift (misaligned cycles).

This method was not originally developed to be used with time series, as generation of new instances was supposed to take place in the feature space, as is common with traditional ML algorithms. However, since then, many extensions and modifications have

been proposed that enable it to be used with a variety of data. Even though it is reported in literature to be extensively used with time series data in its default form [161], some considerations should be kept in mind when doing so. Creating new synthetic instances between two examples can yield nonsensical data, if for example the two instances are two inverted time series. This is usually avoided by the part of SMOTE that selects the nearest neighbour instances, which are the most similar. However, if the minority data is generally very different from one another, even the nearest neighbour might be problematic. In our case, this is not a problem, since the data is consistent and stable in time. For instance, during apnea simulation, the respiratory signal becomes an almost straight line, and traversing this with a sliding window yields similar nearest neighbour instances. The same holds for any other scenario, where the signal is consistently periodic, so again the nearest neighbours are quite similar. The downside of this is little variation between the instances used for synthetic data generation, meaning new ones will be similar to the existing ones [162].

Despite this, we wanted to avoid overwhelming similarity as much as possible, attempting to ensure that the created synthetic instances are not always almost copies of the originals. We thus increased the default value of parameter k to 6, meaning that six windows are considered for finding the nearest neighbour of the selected instance, and then a random one is chosen. Temporally speaking this covers quite a large range (depending on the window length), especially in scenarios which are short-lasting (e.g., Valsalva and Apnea simulation). However, we still wanted to limit the number of considered neighbours, as taking too many might cause problems mentioned earlier, where we cover too much of the waveform, making the worst nearest neighbours substantially different from the original. When such neighbours would inevitably be randomly chosen sometimes, the resulting instances would be nonsensical.

Typically oversampling is done on the training data only. This is to ensure robustness and generalization capabilities of the model, as one must be careful to not have similar synthetic instances in both train and test data, which would cause overfitting and inflated results. However, in experimental setups where training is expensive (e.g., $k-$fold cross validation (CV) with large k or leave one subject out), this requires oversampling to be done many times for different training sets. An alternative is to simply do oversampling once on the whole dataset, but this has to be taken into account when data is separated for training and testing.

In our case we used SMOTE to equalize the distribution of all classes in the whole dataset, with the main aim of oversampling the "Valsalva" and "Apnea" class. We wanted enough instances of these classes in the training data each time to allow the model to learn some characteristics of these classes, and not just ignore them due to their low representation and influence on the accuracy. Due to reasons mentioned prior, we were very careful with both oversampling and subsequent data splitting for training and evaluation. First, we performed per-subject SMOTE – this means that for each instance corresponding to subject S, we only considered instances of this same subject S when searching for nearest neighbours. We did this incrementally, instance by instance, uniformly growing the number of underrepresented class instances across all subjects. Vitally, our data was temporally stacked, meaning data of each subject was always together, starting from windows of their first recording and finishing with their last. When generating synthetic instances, these were inserted in temporal positions immediately after their originating instance. This way the underrepresented class instances remain close to their originating instance and always within each subject group of data. This prevents similar synthetic instances to be mixed between subjects, which is of vital importance for data splits, which we will describe in the following sections.

4.2.3 Model Description

Given the relatively plentiful available data and recent successes of deep learning (DL), we decided to primarily investigate neural networks as our classification models. The main advantage of such an approach compared to traditional ML is that with an end-to-end design, the signals can be input directly into the model, without the need for explicit feature engineering and extraction, as this is done implicitly by the network itself. The downside of this is that given the lack of manual expert feature design, the internal features of an ANN are not easily interpretable, as the model is black-box by nature.

The architectures of ANNs vary significantly, even for similar problems of the same type, such as time series analysis [163]. Many architectures were described for time series classification, such as fully-connected ANNs [164], CNNs [165], recurrent ANNs [166], encoder-decoder ANNs [167], and transformers [168].

Since related work [163] reports similar performance across different architectures for time series classification, we decided to focus on simpler ones that make sense with our input data, while also achieving consistently good performance in literature, such as fully-connected ANNs and 1D CNNs. Additionally, while architectures modelling temporal information (recurrent ANNs, transformers) can (with some modifications) be used for classification [166]–[168], they are vastly more popular in time series prediction (or extrapolation), rather than classification.

As described before, our data comes from several sources or sensors, which can broadly be categorized into those originating from the radar and those originating from the traditional contact TFM monitor. Thus we decided to use an independent-branch approach, where a branch is built for each of the input signals, the branches are then concatenated, and additional learning layers are put on top of it. Fully-connected and 1D convolutional networks were investigated in three major variants based on the input modality:

1. **Contact network:** A variant where a branch is built only for each of the contact sensors.

2. **Contact-free network:** A variant where a branch is built only for each of the radar sensors.

3. **Fusion network:** A variant where branches are built for all of the available input sensors, regardless of sensor modality.

The inputs into the ANN can be in either temporal (actual segments of the signals, most intuitive) or some other domain representation, such as frequency domain (via fast Fourier transform – FFT) or spectrograms. Initially our instances were temporal, however, we also wanted to investigate the mentioned alternative representations, such as the frequency domain. Thus, we also computed the squared absolute value of the FFT of each instance, which gave us the information about the dominant frequencies present in each instance. Such frequency spectrum representation does not include the phase information, which is still present in the temporal representation.

When we initially used temporal inputs, the size of the input was $(win\_len \times F_S))$, where win_len was the window length in seconds and F_S was the sampling frequency. In the frequency domain, we set the number of FFT coefficients such that we could represent frequencies up to Nyquist frequency $(F_S/2)$, which is the highest frequency that can be accurately represented in each window. The input size in such cases was thus $((win\_len \times F_S)/2)$, as only the first half of the coefficients are informative, while the second half is mirrored, duplicating the information.

This already shows that there was notable variation in the possible inputs (input modalities in terms of sensors, input types in terms of representation domain) and network ar-

chitectures. A general schematic of the proposed and investigated branched architecture alongside some relevant hyperparameters is shown in Figure 4.7. All investigated models follow this paradigm, but differ slightly in the architecture, hyperparameters and inputs, so we show a single generalized scheme for brevity.

Since we have many branches, the number of weights of the network is inherently large, however, this can cause overfitting to the dataset, limit the generalization capability, and increase the computational cost of training. We thus decided to limit the depth of the network and use suitable dropout mechanisms in an attempt to alleviate such problems.

This brings us to the investigation of the hyperparameters of the proposed models. It is known that the hyperparameter space of neural networks is large and it often comes down to researcher's experience and expert knowledge to narrow down the initial space and then converge to a good set. We employed a systematic approach – a list of hyperparameters and the values to investigate were initially defined as given in Table 4.3 and then a random search – which included running a 5-fold CV experiment on a randomly selected set of hyperparameters to evaluate them – was conducted each time. The search was done separately for contact, contact-free and fusion networks, for temporal and frequency inputs. Importantly, hyperparameter tuning was always done so that it did not include actual test data in a specific experiment. Naturally, the computational cost of such an experiment is substantial, so we limited the initial hyperparameter space and then further narrowed it down based on performance. The search and training were conducted on a workstation with 32 GB of random access memory (RAM) and nVidia Quadro P6000 graphics processing unit (GPU) with 24 GB of graphics double data rate 5 (GDDR5) video RAM (vRAM).

Table 4.3: The set of hyperparameters we investigated for our ANN models. Those with only a single value were fixed due to universal and common use in literature dealing with such problems.

Hyperparameter	**Investigated Values**
N_HIDDEN_LAYERS	[1, 2, 3]
N_UNITS	[32, 64, 128]
N_FILTERS	[16, 32, 64]
KERNEL_SIZE	[8, 16, 32]
ACTIVATION	[*relu*, *tanh*]
DROPOUT	[0.2, 0.3, 0.4]
LEARNING_RATE	[0.005, 0.01, 0.05]
OPTIMIZER	[*Adam*, *SGD*, *AdaDelta*, *RMSprop*]
REGULARIZER	[*l1*, *l2*]
REGULARIZATION	[0.001, 0.005, 0.01]
BATCH_SIZE	[32, 64, 128]
N_EPOCHS	100
LOSS	*categorical_crossentropy*
LAST_ACTIVATION	*softmax*
INITIALIZER	*GlorotNormal*

4.2.3.1 Architecture Comparison

As mentioned, out of many ANN architectures used for time series classification in other work, one stands slightly apart for being exceptionally common and consistently performing well – 1D CNN [169].

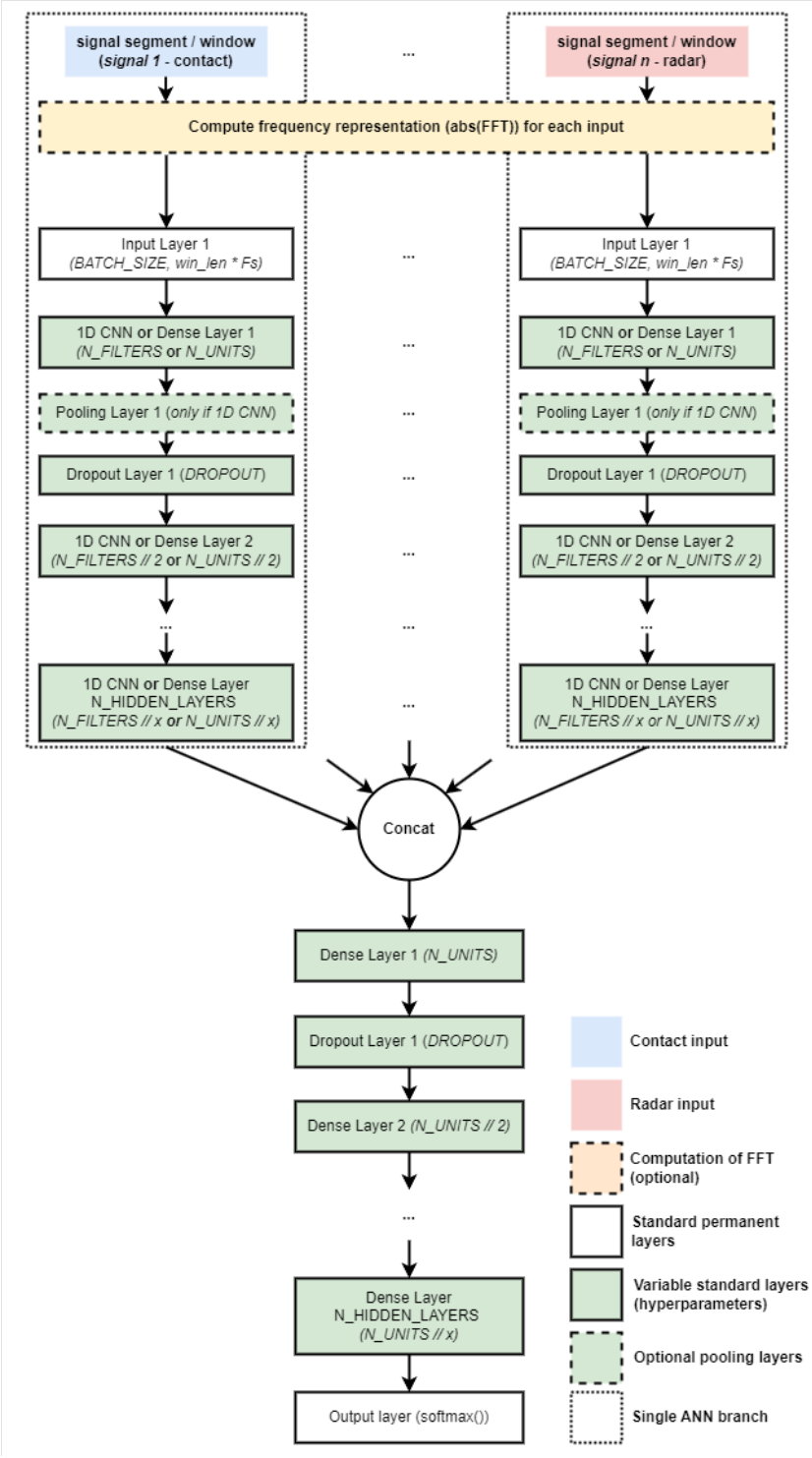

Figure 4.7: Conceptual schematic of the branched ANN architectural paradigm investigated in our work. There are different possible input modalities (contact, radar or fusion) and data types (temporal, frequency/FFT or fusion) and different possible hyperparameters.

The idea of 1D CNN is to have a convolutional filter or kernel W of size or width k slide along a single dimension – time – and calculate the output y based on the input x using the convolution operation. The kernel size represents the number of samples in a time series that are being used for convolution to produce the output. This is illustrated in Figure 4.8.

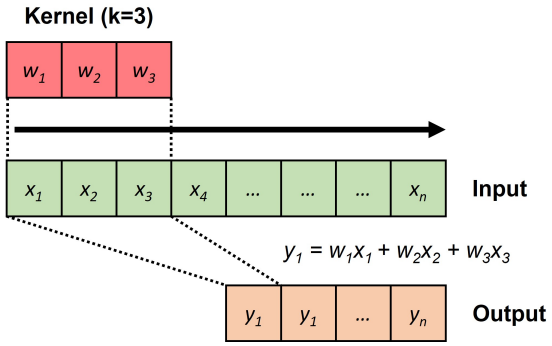

Figure 4.8: Graphical representation of 1D convolution.

This type of network allows for computation of temporally localized features of the time series, which are then commonly fed to a fully connected layer, which serves for learning the relationship between these features and class labels.

An alternative to this is to use only fully connected or dense layers. This approach is more common when features are already pre-computed and used as input to the network, which in our case corresponds to the frequency representation of the input signals. Dense layers can also be used directly on the input signals, although this is a less common approach.

4.2.4 Experimental Design and Evaluation

In standard ML notation, the input instances X to our model were the segments or windows of the input signals, some originating from the radar sensor (denoted only as "radar" from here on) and others originating from the contact TFM device (denoted only as "contact" from here on). We thus had two modalities, radar and contact, the former consisting of six total signals (I, Q, distance, radar respiration, heart sound and pulse) and the latter also consisting of six total signals (BP, two ECG leads, ICG, impedance, contact respiration). As we had two possible input data types, this multiplies the number of inputs by two, having a temporal and frequency representation of each instance for each signal.

For our target variable Y, we initially had six classes as defined in the preprocessing section, but we decided to remove the "Other" class from the evaluation, since it is semantically not clear what happens in those periods, as influences of several other classes can be present. Thus, we conducted our experiments with five classes defined in the original dataset, which were one-hot encoded.

In order to evaluate the classification performance of our models, a robust evaluation experiment was designed. First, classification metrics and a loss function were chosen, which in our case were the categorical accuracy, multi-class or macro F1 score, and categorical cross-entropy. Categorical accuracy (metric) simply checks the fraction of all one-hot encoded predictions that match the ground truth. Multi-class F1 score (metric) is calculated as the F1 score per-class in a one-vs-rest manner (simulating a binary scenario) and then averaged, as described in Equation 4.3.

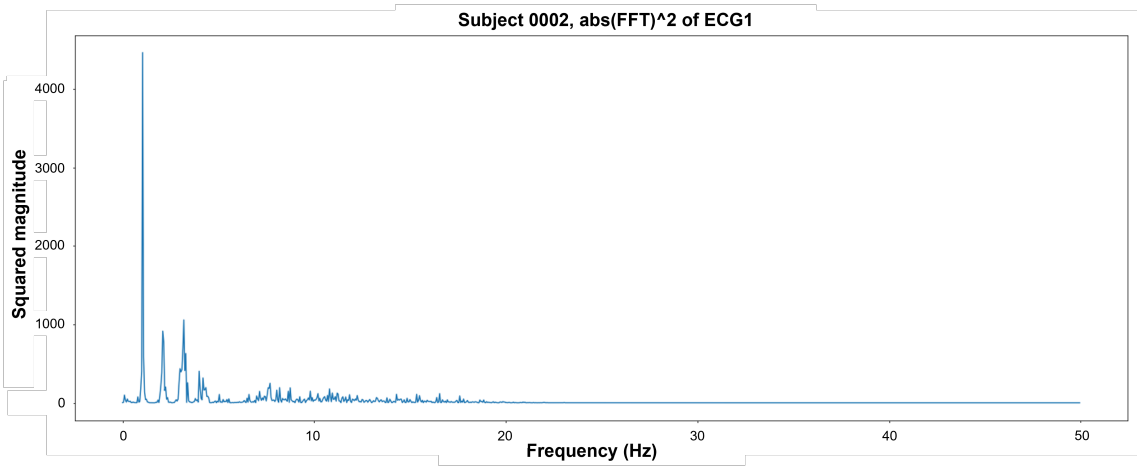

Figure 4.9: An example showing frequency domain input, as computed by the squared absolute value of the FFT.

$$\text{Precision}(C_i) = \frac{TP(C_i)}{TP(C_i) + FP(C_i)}$$

$$\text{Recall}(C_i) = \frac{TP(C_i)}{TP(C_i) + FN(C_i)} \tag{4.2}$$

$$\text{F1}(C_i) = \frac{2 \cdot Precision(C_i) \cdot Recall(C_i)}{Precision(C_i) + Recall(C_i)}$$

$$\text{F1}_{\text{macro}} = \frac{1}{n} \sum_{i=1}^{n} F1_i, \tag{4.3}$$

where $\text{Precision}(C_i)$, $\text{Recall}(C_i)$, $\text{F1}(C_i)$, $TP(C_i)$, $FP(C_i)$, $FN(C_i)$ are the precision, recall, F1 score, true positives, false positives and false negatives for i-th class in a one-vs-rest manner. F1_{macro} is the average F1 score across all n classes.

Categorical cross-entropy (loss) is based on checking how the probability distributions of predictions and ground truths match, and is defined in Equation 4.4.

$$CE(p, q) = -\sum_{i=1}^{n} p(i) \cdot \log(q(i)), \tag{4.4}$$

where CE is the cross-entropy, p and q are the true and predicted probability distributions and i goes over all class values in C. Categorical variant uses a softmax() activation in the final output layer, which ensures that the sum of probabilities in the one-hot encoded output vector is equal to 1.

This set of metrics and loss are commonly used in multi-class classification problems. Accuracy gives a very intuitive but potentially naive understanding of classification performance. To get a better understanding of actual per-class classifier performance, F1 score is commonly used to complement it, as it contains the merged precision and recall information. The best overview can be obtained directly from confusion matrices, which

show precise classification performance for each class and serve as the groundwork from which other metrics are derived.

We mentioned before that SMOTE was used to oversample the minority classes with the aim of equal class distribution. As discussed in detail previously, it is vital that the data of all subjects was stacked, ensuring the temporal order was kept and (synthetic) data of each subject also remained together. It is of utmost importance to also avoid shuffling the instances, as this would cause the internal results to be overly optimistic due to overfitting, since similar (synthetic) instances (of a single subject) would be present in both training and validation data.

A 5-fold CV experiment was done, in which 80% of the data were always taken for training and 20% was withheld for testing, with relatively balanced class distribution in both training and testing data. This was repeated five times, ensuring that each part of the data is independently tested on once, and the results are then averaged across all folds. We chose 5-fold instead of standard 10-fold as a reasonable balance between computational demands and robustness. As we focused on DL approaches, we additionally split the training data into internal training and validation, again in 80–20% ratio. The models were trained on the internal training data, and the purpose of validation data was to guide the model convergence and control overfitting to the training data, with the aim of achieving generalization capability of the model, which was then evaluated on the completely left out testing data.

This sort of $k-$fold CV is relatively robust with the exception of the border instances where the split is made. The split makes it so that some neighbouring instances of a single subject are put into train and test set, likely causing an overoptimistic performance for those few instances. In 5-fold CV this happens five times. In the first and the last fold, data of a single subject is split, while in the intermediate folds data of two subjects are split. However, each pair of subsequent intermediate folds shares one such subject. In total, this brings the number of subjects whose data gets split to 1–2 on average in each evaluation fold. However, all 30 subjects are eventually evaluated on and in aggregate only distinct 4 subjects data gets split. Furthermore, as data gets split again internally into training and validation, this further dilutes the number of problematic instances in the actual model training. When several tens of subjects and several thousands of instances are being considered, this influence is minimal and does not notably affect the results.

A way to circumvent this is to instead use the LOSO evaluation setup, where all but one subjects are used for training and the left out subject is used for testing, without any neighbouring instances appearing in both sets. The downside of the LOSO setup is that a model has to be trained and evaluated many more times compared to a $k-$fold CV in cases where the number of subjects is notably greater than k. This is especially impactful in DL, where the amount of data is typically large and the model training can take quite some time and require a lot of computational power. On top of this, considering investigation of different architectures and other hyperparameters, such evaluation becomes practically very challenging and limited by the available computing power and time.

We ran the described evaluation pipeline for different models, namely fully-connected ANNs vs. CNNs, radar vs. contact inputs, and temporal vs. frequency inputs. Detailed results, findings and comparisons are reported in the following chapter.

4.3 Results

We proposed and discussed a number of evaluation experiments and metrics RF-based physiological monitoring, highlighting the trade-off between robustness and computational requirements. This section details the results obtained in those experiments and offers

comparisons between different experiments and metrics of interest together with corresponding statistical significance. The reported results also answer our research questions and serve as the grounds for acceptance of rejection of hypotheses outlined in Chapter 3. While this section will present all the results and interesting observations, the final detailed interpretation and discussion will be given in Chapter 6.

4.3.1 Architecture and Hyperparameters

One of the early questions in the RF-based classification of hemodynamic scenarios were the specifics of the neural network architecture to be used. As discussed previously, related work reports a myriad of different architectures for time series classification, with 1D CNNs being at the forefront. As discussed in the previous Chapter, we decided to compare 1D CNNs with Dense ANNs in every experiment, as we had both temporal (raw) data and frequency representations available.

The downside of CNNs is that there is both theoretical and empirical evidence showing that CNNs generally require notably more time to train compared to Dense networks, even when the latter have substantially more parameters. In terms of Big O notation, the time complexity of all convolutional layers is reported [170] as given in Equation 4.5

$$O\left(\sum_{l=1}^{d} n_{l-1} \cdot s_l^2 \cdot n_l \cdot m_l^2\right), \tag{4.5}$$

where l is the index of the convolutional layer, d is the number of convolutional layers (depth), s_l is the filter size, n_l is the number of filters in the l-th layer, and m_l is the size of the output feature map. Finally, n_{l-1} is the number of channels, which in our case is 1, as we are dealing with a single channel (time series) or signal per branch.

In contrast to relatively high computational demands of CNNs, He et al. [170] report Dense networks to take only about 10% as much time as CNN networks with comparable depth. The difference comes mainly from the substantially larger number of multiplications required in the convolution. This can be a concern when investigating many architectures and hyperparameters and evaluating them using expensive experiments like k-fold CV and LOSO.

Despite the difference in time needed for evaluation, the reported state-of-the-art performance of CNNs cannot be ignored. We thus conducted our evaluations using both types of networks to estimate the baseline performance and compare the two across different input signal modalities and data types.

The architecture is defined in more detail by hyperparameters, such as the depth and width of the network and the connections between neurons, which directly define the network topology. However, in addition to hyperparameters defining the topology, there are other hyperparameters importantly influencing the network training and performance, such as the activation function, learning rate and regularization [171].

We described the search for their optimal values in the previous section. In summary, it included running a random search over a predefined search space, and conducting a 5-fold CV with a given set of hyperparameters each time, to evaluate it. Based on the number and density of given hyperparameter values, the search space can quickly become very large and checking all combinations is not feasible, so a random search was chosen over grid search. The best combinations found for each network type (Dense and 1D CNN) are given in Tables 4.4 and 4.5 for fully-connected and CNN networks, respectively.

Table 4.4: The best performing Dense ANNs with corresponding optimized hyperparameters.

	Best-Performing Dense ANN [4.7]	**Optimized Hyperparameters**
Contact	Signal inputs: 6 Temporal branches: 0 FFT branches: 6 Concatenation: 1 Output branches: 1 Window length: 20	N_HIDDEN_LAYERS: 1 N_UNITS: 32 ACTIVATION: *tanh* DROPOUT: 0.3 LEARNING_RATE: 0.005 OPTIMIZER: *Adam* REGULARIZER: *l2* REGULARIZATION: 0.001 BATCH_SIZE: 64
Radar	Signal inputs: 6 Temporal branches: 0 FFT branches: 6 Concatenation: 1 Output branches: 1 Window length: 20	N_HIDDEN_LAYERS: 1 N_UNITS: 64 ACTIVATION: *tanh* DROPOUT: 0.3 LEARNING_RATE: 0.05 OPTIMIZER: *Adam* REGULARIZER: *l2* REGULARIZATION: 0.001 BATCH_SIZE: 128
Fusion	Signal inputs: 12 Temporal branches: 12 FFT branches: 12 Concatenation: 1 Output branches: 1 Window length: 20	N_HIDDEN_LAYERS: 1 N_UNITS: 32 ACTIVATION: *tanh* DROPOUT: 0.4 LEARNING_RATE: 0.001 OPTIMIZER: *Adam* REGULARIZER: *l2* REGULARIZATION: 0.001 BATCH_SIZE: 32

4.3.1.1 Dense (fully-connected) Network

From Table 4.4 we can see that the best-performing hyperparameter configuration for Dense ANN networks uses frequency domain inputs computed via FFT instead of raw signals directly. This can be attributed to the fact that $abs(FFT)$ is a form of feature space representation of the inputs describing the frequencies present in a window, so it is more suitable for a Dense network input. Similarly, it was universally observed across different signal input modalities (contact, radar and fusion) that having a single hidden layer in each branch performed the best, although the width of the network differed – with radar inputs, a wider network starting with 64 neurons outperformed a narrower one starting with 32 neurons, which performed the best with contact and fusion inputs.

In terms of activation function, *tanh* outperformed *relu*, although the latter is simpler and faster to compute. Interestingly, the learning rate varied substantially between input modalities, although generally a relatively low value performed well. As for regularization, $l2$ outperformed $l1$, which is not uncommon in literature [172].

Penultimately, each input modality performed best with a different batch size, although the observed influence of batch size was overall minor in terms of accuracy, at least when not changed substantially or dynamically, as is common in recent works [173]. Finally, *Adam*

was dominantly the best performing optimizer across different sets of hyperparameters.

Table 4.5: The best performing 1D CNNs with corresponding optimized hyperparameters.

	Best-Performing 1D CNN [4.7]	**Optimized Hyperparameters**
Contact	Signal inputs: 6 Temporal branches: 6 FFT branches: 0 Concatenation: 1 Output branches: 1 Window length: 20	N_HIDDEN_LAYERS: 2 N_FILTERS: 64 KERNEL_SIZE: 32 ACTIVATION: *relu* DROPOUT: 0.3 LEARNING_RATE: 0.001 OPTIMIZER: *Adam* REGULARIZER: *l2* REGULARIZATION: 0.001 BATCH_SIZE: 64
Radar	Signal inputs: 6 Temporal branches: 6 FFT branches: 0 Concatenation: 1 Output branches: 1 Window length: 20	N_HIDDEN_LAYERS: 2 N_FILTERS: 64 KERNEL_SIZE: 32 ACTIVATION: *relu* DROPOUT: 0.3 LEARNING_RATE: 0.005 OPTIMIZER: *Adam* REGULARIZER: *l2* REGULARIZATION: 0.001 BATCH_SIZE: 32
Fusion	Signal inputs: 12 Temporal branches: 12 FFT branches: 0 Concatenation: 1 Output branches: 1 Window length: 20	N_HIDDEN_LAYERS: 2 N_FILTERS: 32 KERNEL_SIZE: 16 ACTIVATION: *tanh* DROPOUT: 0.4 LEARNING_RATE: 0.001 OPTIMIZER: *Adam* REGULARIZER: *l2* REGULARIZATION: 0.001 BATCH_SIZE: 32

4.3.1.2 1D Convolutional Neural Network

We can observe from Table 4.5 that in the case of CNNs, the temporal inputs – meaning signals directly in the time domain – consistently appear in the best-performing configurations. This is again not uncommon, as CNNs are uniformly reported to work best with direct inputs (be it 1D signals or 2D images), as the kernels compute features from the raw data internally.

In the CNN case we observed differences compared to best fully-connected configurations when looking at depths of individual branches, as now deeper branches with two hidden layers performed better. In terms of filter properties, higher number of filters and larger kernel size generally showed better performance.

Interestingly, *relu* activation function outperformed *tanh* in this case, however lower learning rates still performed best, indicating smaller changes to the model to be preferred.

Same as before, $l2$ regularization again proved superior to $l1$, while smaller batch sizes gave the best results. As before, *Adam* optimizer consistently performed the best.

4.3.1.3 Relation to Data Inputs

The topology-defining hyperparameters (e.g., the number and type of input branches) are closely related to our input data options. Despite the fact that some best-case configurations include only a single input data type (e.g., only temporal or FFT branches), we cannot be sure that this is not a random occurrence, even though random search is known to have a 95% probability of finding a configuration within 5% of optima with only 50 samplings [174], making it the preferred choice for hyperparameter space search due to lower computational demands than grid search. In order to get a complete exhaustive overview, we fixed the obtained best hyperparameters relating solely to the ANN and systematically investigated hyperparameters relating to data inputs, specifically window length, input modality and input data type. The complete overview is reported in Table 4.6 and accompanied with analysis of statistical significance between different values.

4.3.2 Classification Performance

Once the hyperparameters strictly relating to the model were chosen and fixed, we again evaluated the networks using the previously described 5-fold CV, which was carefully designed to minimize overfitting, by paying special attention to instance order and subsequent splits. Again, the two main monitored metrics were accuracy and F1 score. The former is highly intuitive but less robust with imbalanced data, and the latter offers a better view of performance when class distribution is taken into account. Averages across folds for different experiments with different values of input-data-related hyperparameters are reported in Table 4.6. We bolded and highlighted the results of the best performing configurations, while exhaustively reporting results for all possible configurations. Note that Table 4.6 is more extensive compared to Tables 4.4 and 4.5, as those two only showed input combinations of best-performing cases for brevity, while all possible combinations are reported here.

To initially verify the statistical significance of differences in accuracies and F1 scores between all possible configurations of independent variables (window length, network, modality, input data type), we tested a general linear model estimated with ordinary least squares, where we modelled accuracy and F1 score as a linear function of aforementioned independent variables. Such a test is equal to analysis of variance (ANOVA) and considers all independent variables at once within a single test. It provides the F statistic, which is a ratio of two variances – variance between sample means / variance within the samples – and corresponding p values relating to the level of significance [175]. The chosen method is parametric, meaning that it has higher statistical power, but requires some assumptions to be met. These are linear relationship between dependent and independent variables, independence of residuals, normal distribution of residuals, equal variance of residuals across different values of independent variables, and no perfect correlation between independent variables. These were checked and confirmed in the first stage. We always used all *per-fold* accuracies and F1 scores that were used to compute averages reported in Table 4.6, for all configurations of independent variables, as we wanted to maximize robustness and power of the tests. We always tested using the standard 5% level of significance.

Differences in accuracy and F1 score between different configurations were overall statistically significant with $F = 106.4$ ($p < 0.001$) and $F = 87.24$ ($p < 0.001$), respectively. With different configurations of independent variables, we can explain 93% variability in accuracy and 92% variability in F1 score.

Table 4.6: Accuracy and F1 score (Acc. / F1) for the investigated Dense and 1D CNN networks at different window lengths, input modalities, and input data types, always using the best-performing set of hyperparameters. Best results for each network architecture are bolded, and the overall best results are highlighted in green.

Window	Network	Modality	Temp. Data	Freq. Data	Temp. + Freq.
5 s	*Dense ANN*	Contact	0.69 / 0.68	0.68 / 0.66	0.69 / 0.66
		Radar	0.64 / 0.61	0.72 / 0.71	0.72 / 0.72
		Fusion	0.68 / 0.68	0.66 / 0.65	0.70 / 0.70
	1D CNN	Contact	0.65 / 0.65	0.67 / 0.65	0.70 / 0.69
		Radar	0.62 / 0.60	0.61 / 0.60	0.62 / 0.62
		Fusion	0.63 / 0.63	0.63 / 0.62	0.64 / 0.63
10 s	*Dense ANN*	Contact	0.78 / 0.77	0.80 / 0.80	0.80 / 0.79
		Radar	0.75 / 0.75	0.76 / 0.76	0.75 / 0.74
		Fusion	0.79 / 0.79	0.80 / 0.78	0.79 / 0.78
	1D CNN	Contact	0.76 / 0.76	0.73 / 0.71	0.75 / 0.74
		Radar	0.74 / 0.74	0.72 / 0.71	0.74 / 0.73
		Fusion	0.79 / 0.78	0.75 / 0.75	0.78 / 0.78
20 s	*Dense ANN*	Contact	0.84 / 0.84	**0.88 / 0.87**	0.87 / 0.86
		Radar	0.81 / 0.81	**0.83 / 0.83**	0.82 / 0.81
		Fusion	0.86 / 0.84	0.87 / 0.87	**0.88 / 0.87**
	1D CNN	Contact	**0.85 / 0.85**	0.82 / 0.80	0.84 / 0.83
		Radar	**0.82 / 0.82**	0.80 / 0.78	0.81 / 0.80
		Fusion	**0.86 / 0.84**	0.82 / 0.82	0.85 / 0.85

4.3.2.1 Window Length

When talking about signal preprocessing in Section 4.2.2, we already mentioned investigation of different window lengths between 5 s and 20 s when segmenting the signals. Specifically we checked the performance of 5 s, 10 s and 20 s windows, due to our assumption of them being long enough to contain at least 1 respiratory cycle, while being short enough to not overextend the short scenarios like Valsalva and apnea.

We assumed longer window lengths to be better suited for our problem. This assumption stems from the fact that a longer window captures several cycles, which means that frequency representation makes more sense compared to a short window length capturing only a single cycle. This is because periodicity (preferably with several periods) is a mandatory requirement when computing FFT to get meaningful representation. Moreover, we later found that while in most cases 5 s was enough to capture a single period or respiratory cycle, for some individuals even this was too short in the rest scenario, as their RR decreased to the point where a single cycle was longer than 5 s (e.g., a sort of meditative rest state). An additional benefit comes in the form of fewer instances, which lessened the high computational requirements.

Looking at Table 4.6, we can consistently observe the best performance of the longest 20 s window across both network architectures for different input signal modalities and data types, in line with expectations and reasoning given above.

Looking only at window length, we found it to statistically significantly influence the results with $F = 342.4$ ($p < 0.001$). More specifically, looking only at 20 s (long) window length (compared against other cases) within the general linear model, we observed that using a shorter window length on average decreases the accuracy of the model by 8% ($t = -11.1$, $p < 0.001$) or 18% ($t = -26.1$, $p < 0.001$) for 10 s (medium) and 5 s (short) window,

respectively. Similarly for F1 score, shorter window length on average decreases the F1 score of the model by 7% ($t = -9.7$, $p < 0.001$) or 18% ($t = -23.6$, $p < 0.001$). This confirms that using the longest window length achieves statistically significant highest accuracy and F1 score.

4.3.2.2 Network Type

Considering performance of different network types, the average difference between Dense ANN and 1D CNN was 3.2% and 3.1% for accuracy and F1 score, respectively. These changes were overall statistically significant with $F = 33.1$ ($p < 0.001$) and $F = 25.8$ ($p < 0.001$). Again looking more closely at specific network types in the general linear model, we observed an average increase of 3.2% in accuracy and F1 score when using the Dense network, with $t = 5.8$ ($p < 0.001$), concluding that using a Dense network statistically significantly improves performance.

4.3.2.3 Input Modality

Further looking at different modalities (contact, radar and fusion), we can see that the contact and fusion modalities achieved the best results, on average surpassing the radar modality by 5% and 4% for accuracy and F1 score, respectively. Overall we found differences in accuracy and F1 score due to changes of input modality to be statistically significant with $F = 11.7$ ($p < 0.001$) and $F = 8.8$ ($p < 0.001$). When again looking at specific values of input modality in the general linear model, we confirmed that using radar modality decreased the average accuracy and F1 by 4% compared to contact or fusion, with $t = -4.3$ ($p < 0.001$). However, there were no statistically significant differences between using contact or fusion modality.

4.3.2.4 Data Input Type

Finally, we can also observe that Dense network generally performs better with frequency inputs, although the performance was only marginally better compared to temporal inputs. The average improvement of Dense network when using frequency over temporal data input type was 1.8% for both accuracy and F1 score. Average improvement compared to fusion of input data types is negligible at almost 0%. On the other hand, the 1D CNN architecture was shown to perform better with temporal input data type, meaning direct signal segments, although the accuracies were lower compared to the fully connected network. The differences in accuracy and F1 score between different input types were shown to not be statistically significant, with $F = 1.7$ ($p < 0.2$) for accuracy, and $F = 1.5$ ($p < 0.24$) for F1 score. This means that even though we observed some networks performed better with one or other input type, the differences were overall not statistically significant.

4.3.2.5 Visualizations

In addition to numerical and statistical analysis of the results, we also checked confusion matrices of the overall best-performing cases in order to get a better understanding and overview of classification performance. These are shown in Figure 4.10 for each case highlighted in green in Table 4.6.

We can observe a similar performance between the first (contact input modality, frequency data type) and third (contact + radar input modalities, frequency + temporal inputs) row in Figure 4.10, where the model classifies Valsalva, Apnea and TiltUp scenarios very well, but has more trouble separating the Resting and TiltDown scenarios. Alternatively the second row (radar input modality, frequency data type) performs well

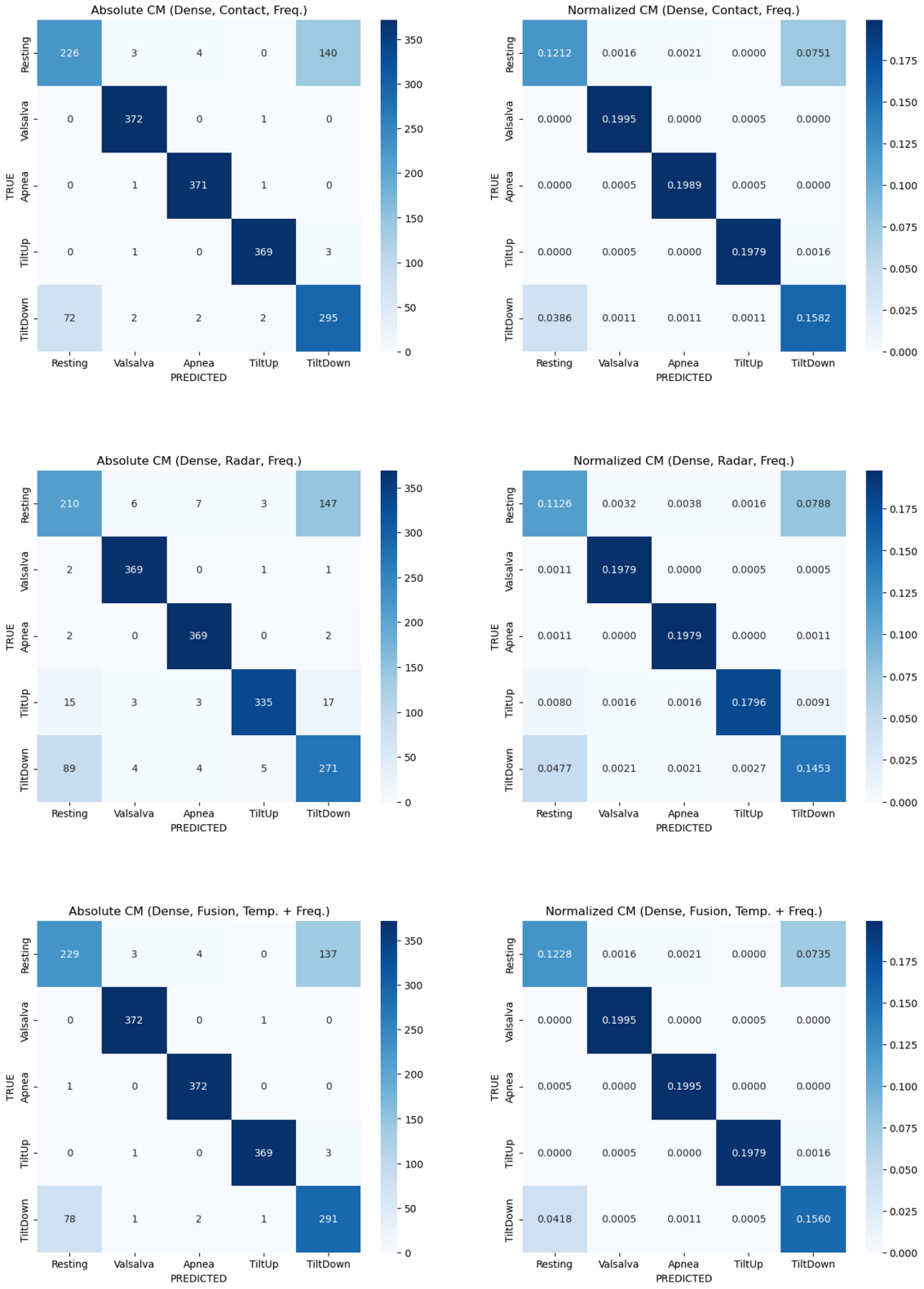

Figure 4.10: Absolute and normalized confusion matrices for the best-performing cases (bolded and highlighted in green) from Table 4.6. Network, input signal modality and input data type are given in each subplot title. Darker shades indicate a higher number of instances.

in classifying Valsalva and Apnea, but has more trouble separating the TiltUp scenario compared to the other two. Again, the separation of Resting and TiltDown is the poorest, being the worst out of all three cases.

Finally, for the overall best performing case, we also show how accuracy and loss were changing during the training process of the network for train and validation data. This can be seen in Figure 4.11, where we can observe consistent performance across different folds, without major variation in accuracy or loss, indicating good robustness regardless of the specific data split.

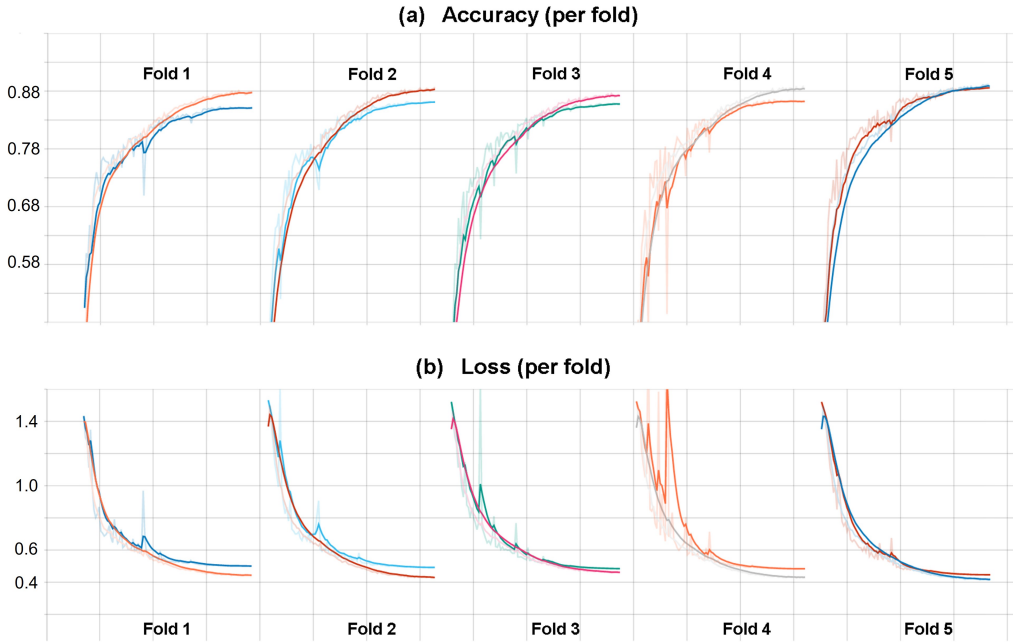

Figure 4.11: Changes in accuracy and loss for the best performing model between different folds. Bold lines are slightly smoothed and represent train and validation accuracy and loss. Corresponding vertical subplots (same fold) share same colors, otherwise colors are arbitrary.

We can conclude that a Dense network using contact modality and frequency data representation performed the best. The results of using fusion of input modalities and data representations equalled the best performance, but did not surpass it. This shows that little-to-no additional information can be obtained from the contact-free radar modality compared to using gold-standard contact sensors. However, the same network achieves only 4-5% worse performance when using only the contact-free radar input modality and frequency data representation. Importantly, such a network achieves almost identical performance in classification of Apnea and Valsalva scenarios, which are the highest interest for continuous non-invasive sleep monitoring [149]. We will discuss and interpret all the reported results in more detail in the following chapter.

4.4 Information and Privacy Preservation

We have seen that RF-based monitoring relies on the Doppler effect, providing information about changes in distance and velocity. This modality is fundamentally different from optical sensing, as the latter is arguably substantially richer with information and far more interpretable for humans. The signals obtained from a radar, especially in its raw form, are

in themselves not readable to humans, which makes this modality privacy preserving in the visual sense. From video or images, one can discern additional information, such as colors, facial expressions, age, gender, context information, etc., which can enrich physiological and health monitoring in general, at the cost of privacy preservation.

There are, however, ways to leverage optical modality in ways that attempt to preserve privacy as much as possible. Facial anonymization in the form of blurring or other similar mechanisms are an intuitive solution, but only (partially) attempt to solve identification concerns. Another approach is to simply use small regions or patches of exposed skin (not necessarily face) for obtaining the information about color change. This is how wearable PPG sensors work in a sense, as the photodiode monitors a small area such as a fingertip or wrist. In most extreme cases, such spatial privacy-preserving constraints were investigated using only a single pixel for monitoring [176], but such approaches are highly sensitive and SNR can be a concern.

In this chapter, we focused on remotely obtaining the anonymous movement information with a radar, however, the optical color and light absorption information should not be discarded. Based on the understanding of physiological mechanisms governing hemodynamics described in Chapter 2, optical information via light absorption changes would be useful for BP estimation, which will be the focus of the next Chapter.

Chapter 5

Optical Measurement of Multi-wavelength PTT and BP Estimation

This chapter deals with MW camera-based PTT measurement and subsequent BP estimation based on the underlying mechanisms described in Chapter 2, where fundamental feasibility of such a novel approach was investigated. Specifically, our goal was to estimate SBP and DBP using this approach, with errors as close to clinical standards as possible [177].

5.1 Data and Materials

For single-site contact-free multi-wavelength (MW) PTT measurement and BP estimation, hardware capable of capturing independent data from multiple wavelengths is needed, alongside varied ground-truth BP measurements [178]. Such data and the corresponding recording setups, alongside the data collection protocols employed for its collection, will be described in more detail in the following sections.

While a suitable dataset for RF-based monitoring existed and enabled our research in that direction, the same was not true for an optical video dataset that would initially allow for multi-wavelength depth-dependent rPPG reconstruction. This is not unexpected, as such an idea is novel and was previously only recently considered with contact wearable sensors [179]. Subsequently, we designed a data collection setup and protocol ourselves, which allowed for the collection of required data that met the needs for MW rPPG reconstruction, allowing for subsequent contactless MW PTT measurement and BP estimation [180].

We captured the data in a MW depth-dependant manner from different layers of skin of a single measurement site in an attempt to measure the subtle PTT between skin layers. As we mentioned in previous chapters, this approach is to the best of our knowledge unique, and offers several key advantages, which we will discuss in more detail in the next sections. One such advantage, relating to privacy discussion from the previous Chapter, is also privacy preservation, as this approach can be implemented in a way where camera records only a patch of (non-facial) skin, not revealing any ambient or easily discernible personal information.

5.1.1 Hardware and Camera Physics

Obtaining different wavelength information is an inherent property of RGB cameras, as each color pixel is sensitive to specific wavelengths. This is vital for obtaining delayed

PPG waveforms from different skin depths and computing the PTT as hypothesized in Chapter 3. However, the image sensor of an RGB camera is not perfect but is instead governed by the quantum efficiency of each wavelength pixels, which corresponds to the relative amount of specific wavelength photons successfully registered by the image sensor pixels and translated to a digital signal.

An example is shown in Fig. 5.1, which belongs to our specific image sensor, but also serves well for illustration purposes. The relative intensity of sensor response (photons registered and translated to digital signal) on the y axis is in the [0, 1] range, where 0 means no photons are registered and 1 means all of them are. The x axis simply shows different wavelengths of light in the visible and NIR spectrum.

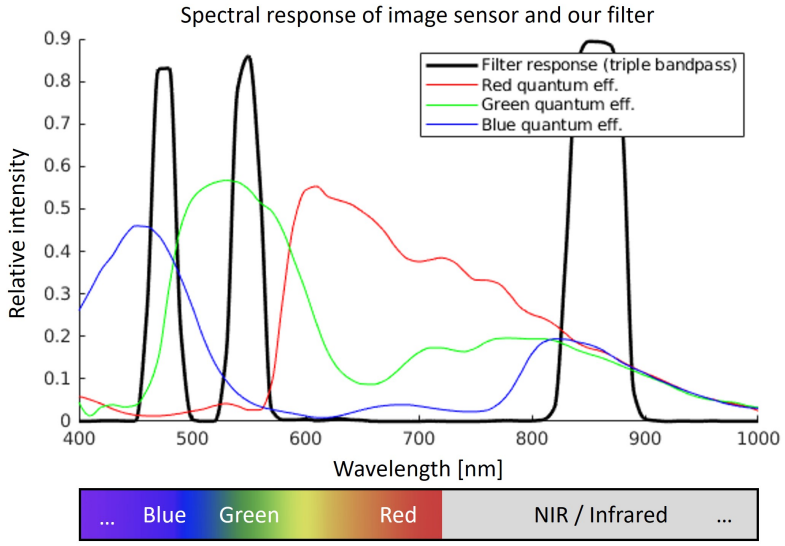

Figure 5.1: Quantum efficiency of the image sensor of our iDS 3040SE RGB camera. Superimposed black line shows the triple bandpass filter response, which helps with band overlapping at a hardware level. The red, green and blue lines correspond to pixel responses of the same color on the image sensor.

Looking at the quantum efficiency shown in Figure 5.1, we can already discern two important facts and corresponding challenges that must be taken into account when collecting our data with an RGB camera:

1. **Channel response overlap**: Looking at each specific channel response (e.g., green), we see that its pixels also respond to wavelengths in the neighbouring bands (e.g., in blue and red wavelengths), meaning that the image and subsequent rPPG trace obtained from such a channel are an impure mixture of several channels. We can also observe that all three color-specific pixels respond in the NIR band around 850 nm. Subsequently we can hypothesise that the PPG waveforms from different wavelengths are hard to distinguish, since rPPG traces obtained from pixels of each wavelength are in fact a mixture, which must first be separated.

2. **Lower relative response in the NIR**: The relative intensity of the pixel response falls substantially towards the longer wavelengths, reaching only around 20% quantum efficiency. This means that in order to obtain similar amplitudes and prominent systolic peaks in the NIR rPPG trace compared to other wavelengths, we must compensate by having higher energy in this part of the spectrum coming from the light

source. Additionally, the band of the triple bandpass filter we used is also wider in this range to additionally compensate for the lower quantum efficiency.

These challenges are fundamental to consumer RGB cameras and a vast majority of image sensors, and must therefore be understood and considered when dealing with MW approaches. We will revisit them and propose suitable solutions in the following chapter.

5.1.1.1 Camera Optics

We have seen in Chapter 2 that the skin layers are thin at 1–2 mm and blood traversal velocity in capillaries was reported to be around 12 mm/s [62], so the PTT measured between different skin layers is expected to be short at around 100 ms, with some variations due to changes in SVR and BP. Knowing this, we required an RGB camera capable of high-frequency recording, substantially above the standard 30 fps. Having these ranges in mind and also looking at related work dealing with contact MW PTT measurement [56], [179], we estimated that our camera should be capable of recording at around 250 fps. This allows measuring PTT with precision of up to 4 ms, which is more than an order of magnitude shorter than expected PTT, so it should be able to capture PTT variations related to BP. Additionally, any on-board image processing is undesired, since any changes to the pixel values might influence and distort the obtained rPPG waveforms.

Following these requirements we identified the iDS 3040SE-Q RGB camera with the Sony IMX273 1/3" complementary metal-oxide-semiconductor (CMOS) image sensor and iDS-5M23-C1618 16 mm lens as a suitable RGB camera for our data collection setup. This camera allows for variable frame rates of up to hundreds of fps while also offering programmatic access to the raw images as registered by the image sensor. For our experiments we fixed the frame rate to 250 fps due to aforementioned reasons and values reported in literature [56], [179].

As seen in Figure 2.6 and reported by related work [104], capillary loops dominate the papillary dermis layer reachable by shorter wavelengths, while smaller arterioles and arteries are found in the deeper layers of the skin, so we wanted to use light that also reaches those depths. NIR in the wavelengths around 850 nm is suitable for penetrating to deeper dermis layers, however, the information is not registered by the image sensor of a traditional RGB camera out of the box, since such cameras come equipped with a default factory IR filter with a cutoff at 650 nm. This filter is in place to modify the image produced by the camera to align with the way the human eye works and which light it can perceive. Since we are not interested in the natural look of images but obtaining the relevant information from our image, we modified our camera by first physically removing the on-board factory IR filter and then replacing it with a triple bandpass MidOpt TB475/550/850 filter, which allows only light in narrow bands of 475±10 nm (blue), 550±10 nm (green) and 850±22 nm (NIR) to pass. As mentioned before, the band is wider in the NIR due to lower sensitivity of the image sensor in this range, which can be seen from the quantum efficiency in Figure 5.1.

Furthermore, human skin often exhibits sweating, especially when exposed to heat or after physical activity, which can result in undesired specular reflections on its surface. These might distort our signal reconstructions, so we additionally used the MidOpt PR1000 VIS/SWIR Wire Grid Linear Polarizer together with the triple bandpass filter. This polarizer is effective in the range of 400–2000 nm, which covers our spectral range of interest.

The combination of these modifications gives us access to NIR information of arterioles in the deeper reticular dermis, while also partially addressing the first challenge mentioned above – channel response overlap. As the selected triple bandpass filter has narrow bands, it isolates the response of the pixels at image acquisition stage, producing a “cleaner” image

in terms of specific wavelength fidelity, albeit this does not solve the problem entirely. Additional separation is required and proposed in the form of channel separation algorithms described in the next chapter.

5.1.1.2 Light Source

In order to capture information from the NIR part of the spectrum, modifying just the camera is not enough. To obtain this information and to address the challenge mentioned several times – poor quantum efficiency in the NIR part of the spectrum – we also had to use a light source that emits the full spectrum of interest, as well as (ideally) increases in energy towards the NIR. Daylight has a suitable spectrum, however, we wanted good control and consistency in our light source, so we used an artificial source instead.

The spectrum requirement meant that conventional LED sources are not suitable, since they have distinct peaks in the blue and green part of the spectrum while emitting virtually no energy in the NIR. An obvious initial candidate was a traditional incandescent filament bulb, which has a suitable spectrum for our use case.

Upon making some initial recordings with a standard 230V 50W bulb and our high-frequency camera, we observed flickering in the recordings, which is a consequence of the 50 Hz alternating current. To circumvent this, we instead used a direct current power source – a 12V 650W power supply unit – alongside Osram Decostar 51ALU filament bulbs. These bulbs are coated with thin aluminum coating and unlike their dichroic counterparts, the heat is fully emitted in front, meaning we get a good amount of (N)IR emission, compensating for the lower sensor sensitivity in this part of the spectrum. We decided to use two such 50W bulbs to ensure enough light to get reasonable visibility, as the exposure time is very short at such high fps. The light was directed towards the ROI of the skin and the bulbs were placed perpendicular to one another in order to negate any shadows created by the uneven surface of the recording ROI.

In order to avoid any interference of other different light sources, we decided to record the subjects in a completely dark room (no windows, no other artificial sources), where only our selected light source provided visibility as it was directed towards the palm.

5.1.1.3 Blood Pressure Monitor

Ground-truth BP measurements were obtained with a clinical-grade Omron M10-IT cuff-based digital BP monitoring device in order to obtain precise and trustworthy measurements. The cuff was placed on the upper arm in accordance with the official guidelines. The cuff was always equipped on the opposite arm of the one being recorded with the camera, as compression of the tissue (as a consequence of the cuff inflation) influences the blood flow.

5.1.2 Recording Setup and Data Collection Protocol

Once our recording setup was finalized at the physical level, we also had to ensure that we removed any on-board default image processing in the software. We did this by developing a piece of recording software based on the official iDS software development kit (SDK) together with their Peak library. In short, as mentioned earlier, we set the fps to 250, which meant that we had to make a trade-off by setting the exposure time to only 4 ms. This in turn also meant that the light sources had to be strong enough for the image not to appear completely dark. We additionally turned all the white-balancing corrections off and set the color gains to the default value of 1.0.

The described and fine-tuned recording setup was used to record the central part of the palm skin of participating subjects. The palm was chosen for measurement since it is one

of the places where skin exhibits the most pulsatility (similar to fingertip and forehead, which are the most commonly used sites in literature) and is relatively comfortable to record and quite maneuverable. The spectral range between blue (450 nm) and NIR (850 nm) was selected as described previously. Note that such wavelengths penetrate between 0.5 and 4 mm deep in human skin tissue, which accounts for variations in skin thickness due to calluses or other anomalies [181]. Recordings lasted for 30-seconds, which is enough time for the digital sphygmomanometer to produce a single measurement via full cuff compression and decompression cycle. The recording setup with an example anonymous subject is shown in Figure 5.2.

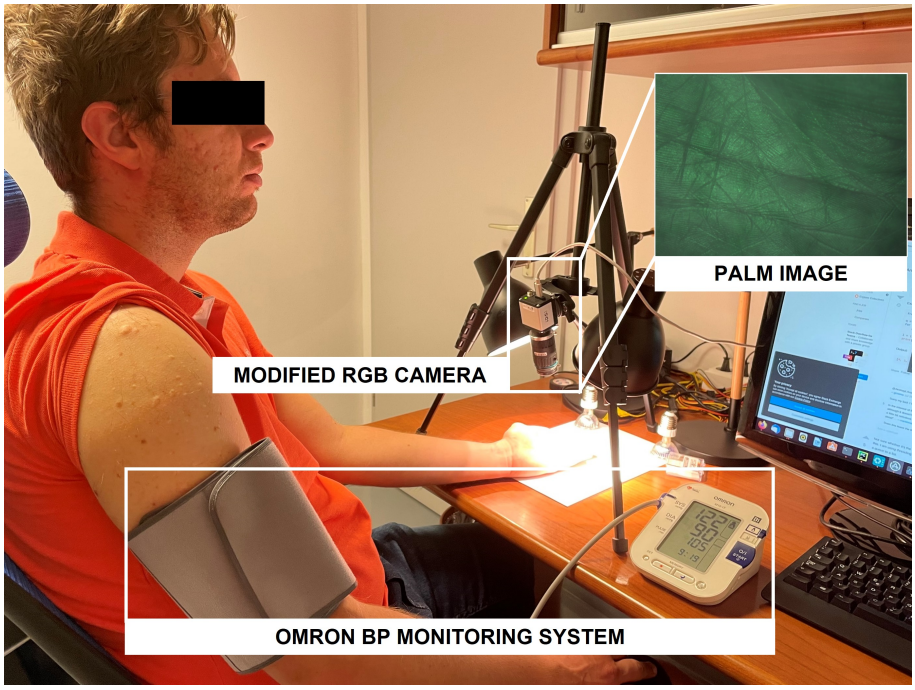

Figure 5.2: Anonymous subject being recorded with our custom camera and BP recording setup. Picture used with subject's explicit permission.

We collected data of 13 volunteers at the Department of Intelligent Systems, Jožef Stefan Institute (JSI). 10 subjects were male and 3 female with the mean age of 30 $\pm$ 3.2 years. Most were healthy young adults, with two exceptions being older and long-time smokers. All have given explicit consent to participation and their data was anonymized. We prepared the following two recording scenarios with the aim of inducing substantial cardiorespiratory and hemodynamic changes, which clearly reflect in the measured BP values:

1. **Resting.** The first scenario was resting in a seated position, where the subjects were taking deep breaths and relaxing with their eyes closed. After one minute, a 30-second recording was made and their ground-truth BP was measured with the clinical-grade Omron cuff-based sphygmomanometer. The aim of this scenario was to obtain a baseline measurement in a relaxed state without physical or mental exertion.

2. **Exercise.** The second scenario included intense physical activity, consisting of 1 minute of jumping jacks followed immediately by jump squats until failure. At failure, the subject was immediately measured, as they exhibit substantially elevated BP and HR due to intense activity and severe exhaustion.

We repeated this experimental protocol twice for each subject, preferably on different days (for some subjects this was impossible due to availability), to obtain more varied information and further validate robustness. This gave us at least two cases of each distinct hemodynamic state for each subject. Some subjects participated more (depending on availability), so we obtained more recordings for those.

Each 30-second recording was assigned the ground-truth BP and HR values of the Omron device. This makes sense since the measurement itself takes around 20 seconds, and since BP and HR do not change that rapidly outside of extreme circumstances like arterial bleeding or powerful medicine. A few recordings were also discarded from further evaluation, as in a couple cases the Omron device returned an error (usually when physical exertion was very high), meaning we did not have the ground truth BP for such a recording segment.

Similar pros and cons as were reported for the radar dataset in the last chapter can also be extended to this one. Namely, the collected data is relatively robust and unique, as it was recorded with custom modified hardware and a specific recording setup that enables further MW analysis of interest. Given the unique setup, a novel research direction, and obtained explicit permissions for data usage from all users, the collected dataset in itself is a contribution that can be useful to other researchers interested in MW monitoring. On the downside, this highly controlled and specific recording setup does not reflect everyday real-world scenarios. However, again, precise control over as many environmental variables as possible is a good thing when investigating fundamental feasibility of a novel approach.

5.2 Methodology

In previous chapters, we described the blood perfusion between different skin layers and the mechanisms relating PWV and PTT to BP. We also highlighted how a standard RGB camera with some modifications can be used to capture information from different wavelengths, allowing for reconstruction of depth-dependant rPPG waveforms. While the idea is well-grounded in physiology, the design of standard RGB image sensors poses many challenges, which we also listed before. In short, some of them are technical and resolvable with modifications of hardware (e.g., removal of IR filter, usage of DC power source, usage of filters on the lens, etc.), while others are fundamental and resolutions are not trivial. An important example of the latter is the imperfect spectral response of the image sensor in each wavelength band, causing the obtained signals to be a mixture containing additional undesired information. Since the phase delays (PTTs) of interest in the waveforms are extremely short and subtle, a resolution for this challenge is mandatory.

In this section, we again initially describe the data preprocessing methods used on our raw signals. We then propose models and algorithms for channel separation that allow us to obtain demixed waveforms from each specific spectral band and subsequently measure the PTT of interest. In the first phase, we evaluated the methods in terms of correlation with BP. In the second phase, we additionally proposed and validated a regression model for explicit SBP and DBP estimation. Similar to our previous experiments, we designed a robust experimental framework that allows for evaluation and validation of our novel approach. The high-level overview of this pipeline is shown in Figure 5.3.

5.2.1 Data Preprocessing

We based our experiments on the custom dataset described in the previous section. At this point we obtained raw 3D matrices corresponding to each frame, which we then manually Debayered as given in (5.1):

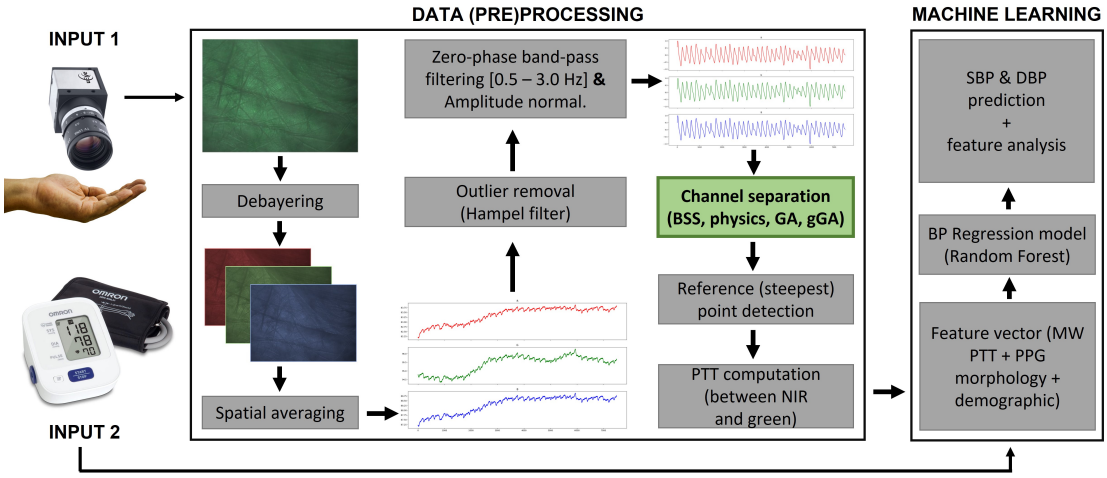

Figure 5.3: High-level pipeline of our proposed system for contact-free single-site measurement of MW PTT and subsequent BP estimation using a consumer RGB camera. The green box denotes our main unique contribution in the form of novel channel separation algorithms.

$$R, G, B \;=\; \begin{cases} R \\ round(\frac{G_1}{2} + \frac{G_2}{2}) \\ B \end{cases} \tag{5.1}$$

where R are the (nearinfra)red pixels, G_1 and G_2 are the green pixels and B are the blue pixels in a standard half green, one quarter red and one quarter blue Bayer mosaic pattern shown in Figure 5.4. Doing this lowered the resolution of the final frames (due to halving of the green pixels) to 540 x 720 pixels.

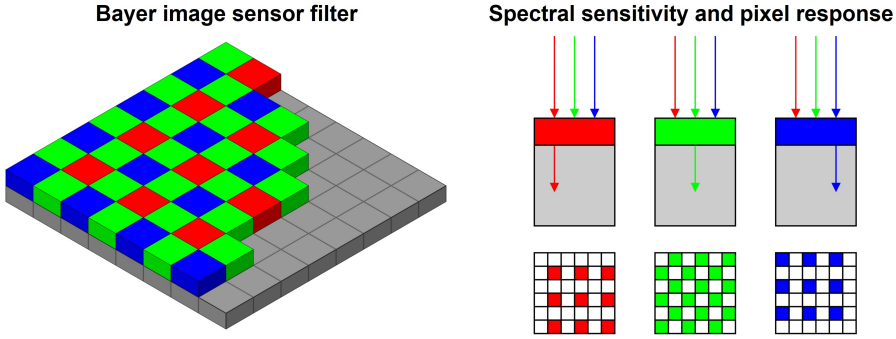

Figure 5.4: The Bayer filter design [182] commonly used by the vast majority of consumer RGB cameras. Distribution of pixels is standard, the order in which they appear is specific to camera or manufacturer.

The sequence of frames was first globally spatially averaged (taking all pixels, since only skin is present in each frame) in order to obtain three temporal RGB traces from the image pixels. These traces were then zero-phase filtered with a 2nd order Butterworth band-pass filter with cutoff frequencies of [0.5, 6.0] Hz. This removed low-frequency baseline wandering and eliminated high-frequency noise. Finally we normalized our signals amplitude to a constant range of $[-1, 1]$, as PPG amplitude in absolute sense has an arbitrary unit (no

unit) – only relative changes in a certain time period are informative [183]. At this point we obtained relatively clean traces with obvious cardiac pulsatility, as seen in Fig. 5.5.

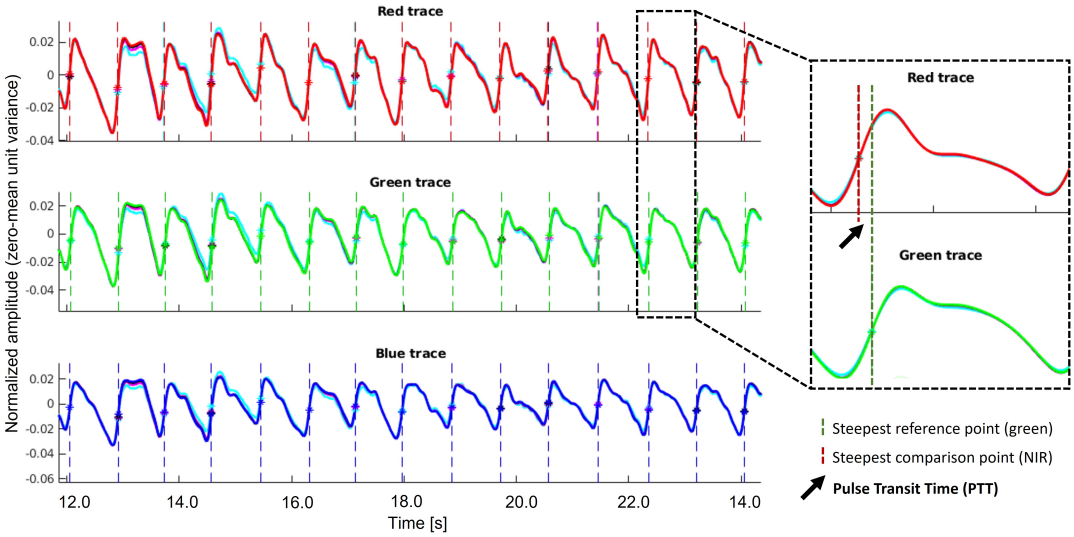

Figure 5.5: An example of all three rPPG traces after the preprocessing pipeline. Vertical dashed lines are the steepest systolic rise reference points. Zoomed frame highlights the subtle PTT between two cycles.

Methods used in this processing pipeline were universally applied to all signals, as they were carefully designed to not influence the waveforms in any way that would distort the relationship under investigation, especially the PTTs in the form of phase delays. More specifically, both the amplitude normalization and the zero-phase filtering were chosen specifically to be generally applicable and to not influence the temporal position of the systolic peak or the steepest systolic rise point, regardless of the subject.

5.2.2 Pulse Transit Time Measurement

In order to measure the PTT, we opted to use a reference-point method, meaning we used a single robustly detectable stable point per cardiac cycle, between which we measure PTT in different waveforms. Specifically we chose the systolic rise, as it is more apparent even in noisy signals, while systolic peaks are sometimes difficult to precisely determine even for state-of-the-art peak detection algorithms [184]. Stability is important in this case because even a slight temporal missdetection of a few samples can represent a large part of the very short PTT. We detected these reference points by using the derivative method, taking the more clearly-expressed peaks in the first derivative. As precision and stability of these reference points is vital, we cross-checked our detections with state-of-the-art systolic rise/peak detection algorithms proposed by Elgendi et al. and Han et al. [185], [186], which were specifically developed to tackle challenging conditions and cardiac anomalies respectfully. Since we ensured our signals were as clean as possible, by both stabilizing the measurement site, as well as using aforementioned preprocessing pipeline, all algorithms detected the same locations of reference points. Detected example reference points can again be seen in Fig. 5.5, represented by the star markers and their temporal locations marked by the vertical dashed lines.

At this point of the pipeline, an initial observation of the PTTs can be made by looking at the steepest points locations in each channel, and then computing the difference between these. To do this early estimation, we first took the green channel and its corresponding

steepest point detections as the reference, because green has been both historically and empirically (from our observations) identified – and also widely accepted in the field – as the trace with the most expressed and stable pulsatility. We then checked a very narrow area before and after these reference points by defining a short threshold of 20 samples or 80 ms in each direction, and searching for the steepest point detections in the other two traces (red and blue). If such points were found within the defined threshold, we computed the temporal difference (PTT), otherwise we ignored the cycle.

Looking at the waveforms obtained after the preprocessing pipeline, we came to two important initial observations. First, the waveforms were relatively clean and stable in NIR and green, while the blue often expressed very shallow and noisy pulsatility with low SNR. Second, the PTTs were extremely short, lasting just a sample or two, which is less than expected and reported in literature [179]. Given the former, while also knowing that blue wavelengths have the shallowest penetration mostly reaching the upper less blood-perfused epidermis, and concerns expressed in related work about lower wavelengths [104], we decided to simply remove the blue channel from further analysis, since it was too unstable. Following this, the PTT for one recording was computed as the average PTT between NIR and green reference points in the whole 30-second recording with a single reference ground-truth BP measurement.

Regarding the unobservable PTTs, this was in fact not entirely unexpected, due to the problematic inter-channel influence highlighted several times, which was only partially resolved at the hardware level by using the narrow-band triple band-pass filter.

5.2.3 Channel Separation

We devised several algorithmic approaches to resolve the persistent problem of cross-channel influence and allow for precise measurement of physiologically informative PTT [187]. In general we can assume, as seen in Figure 5.6, that each color trace obtained from an RGB image sensor is a combination of the actual response in the relevant wavelength as well as undesired response in the other wavelengths, especially when the IR-block filter is removed from the camera, as this unlocks the response around 850 nm. Our aim is thus to nullify this spectral band overlap and obtain clean traces from each spectral band of interest.

5.2.3.1 Blind Source Separation Methods (BSS)

A commonly used method for blind source separation of linear mixtures are PCA and ICA [45]. These are data-driven black-box methods and do not use any underlying understanding or models for the separation of channels. PCA requires the variation in the amplitude of the components to be sufficiently different to determine the eigenvector directions of demixing matrix, while ICA assumes that the sources are statistically independent and non-Gaussian [45]. As discussed previously, the resulting waveforms after preprocessing remain mixtures of overlapping channels, as those are not very different from one another in terms of frequency. Subsequently, PCA and ICA can then be used to separate the overlapping source channels present in the mixture [188].

However, we hypothesize that the above assumptions of PCA and ICA are problematic with respect to obtaining meaningful phase-delayed channel-separated traces, as such traces cannot be said to be sufficiently different nor independent. Based on skin physiology and hemodynamics shown in Figure 2.6 and described in related work [189], we know that the blood perfusion is interconnected and continuous throughout the cardiovascular system, including the skin layers, thus influencing one another.

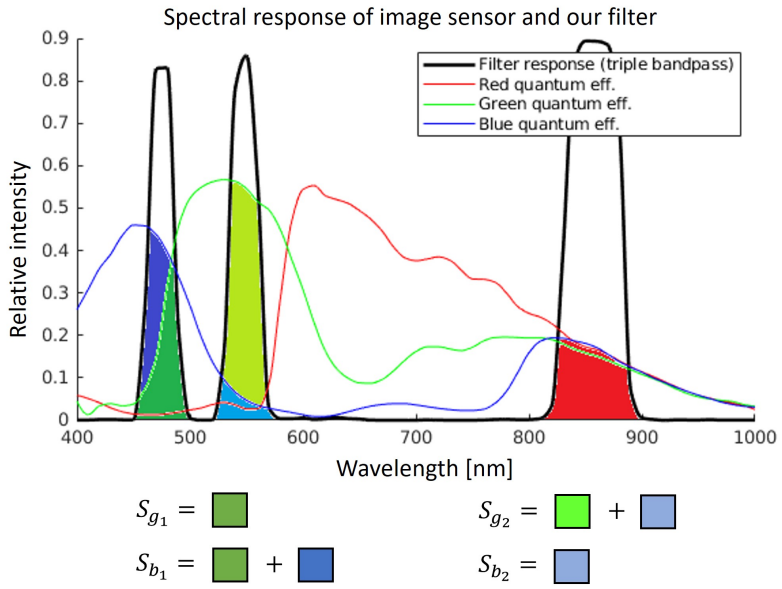

Figure 5.6: Quantum efficiency of the image sensor of our iDS 3040SE RGB camera. Superimposed black line shows the triple band-pass filter response, and the colored areas under the curves represent the ratios of channel mixture.

We investigated both PCA and ICA to confirm or reject our above hypothesis and as benchmark methods that can intuitively be considered as potential candidate methods to resolve the channel overlap in contact-free MW monitoring. Given the model we propose in Equation 5.11, we considered the preprocessed R, G and B traces obtained from the camera as input and then always assumed we also want three output source components corresponding to demixed pulsatile color signals.

5.2.3.2 Physics-based Channel Separation

Another intuitive approach to channel separation is to use the information explicitly available in the quantum efficiency of the camera, which shows the camera-specific representation of the inter-channel influence and spectral band overlap. In our initial physics-based approach, we started from Figure 5.6 and first assumed the response of all pixels in the IR part of the spectrum around 850 nm to be equal. Moreover, we also truncated the very minor response of the IR channel in other wavelengths to zero. This allows us to write the trace equations as follows:

$$\begin{aligned} R &= r \\ G &= g + x \cdot b + r \\ B &= b + y \cdot g + r \end{aligned} \tag{5.2}$$

where R, G and B are the red, green and blue values produced by the camera, and r, g and b are the corresponding channel-separated red, green and blue values. Parameters x and y describe the ratio of the corresponding color present in a mixture that must therefore be determined. Using areas under the quantum efficiency response curves marked in Figure 5.6, we can compute the two ratios as:

$$\begin{aligned} x &= \frac{S_{g_1}}{S_{b_1}} \quad in \quad [450, 500]\ nm \\ y &= \frac{S_{b_2}}{S_{g_2}} \quad in \quad [520, 570]\ nm \end{aligned} \tag{5.3}$$

where S_{g_1} and S_{b_1} are the areas under the curve in the blue band of the filter, while S_{b_2} and S_{g_2} are the areas under the curve in the green band of the filter. This essentially tells us the amount of one color present in the other one, where the unwanted "incorrect" color is the numerator and the dominant "true" color is the denominator, as shown in Figure 5.6. For our specific camera and image sensor we obtained $x = 0.61$ and $y = 0.13$.

Once x and y are computed via definite integration of the curves in Figure 5.6 bound by the filter bands, we can follow 5.2 to express channel-separated r, g and b traces as follows:

$$\begin{aligned} r &= R \\ g &= G - R - x \cdot b \\ b &= B - R - y \cdot g \end{aligned} \tag{5.4}$$

With some expression manipulation we can express the channel-separated traces as follows. First we express b by inserting g:

$$\begin{aligned} b &= B - R - y \cdot (G - R - x \cdot b) \\ b &= B - R - y \cdot G + y \cdot R + x \cdot y \cdot b \\ b &= \frac{B - R - y \cdot (G - R)}{1 - x \cdot y} \\ b &= \frac{1}{1 - x \cdot y} \cdot (B - R - y \cdot (G - R)) \end{aligned} \tag{5.5}$$

Inserting this back into g gives us:

$$\begin{aligned} g &= G - R - x \cdot b \\ g &= G - R - x \cdot \frac{B - R - y \cdot (G - R)}{1 - x \cdot y} \\ g &= \frac{G - R - x \cdot (B - R)}{1 - x \cdot y} \\ g &= \frac{1}{1 - x \cdot y} \cdot (G - R - x \cdot (B - R)) \end{aligned} \tag{5.6}$$

where all the variables are known and can be obtained from the system. The leading factor $\frac{1}{1-x \cdot y}$ is a camera-specific constant that only influences the amplitude, which was normalized, so we can omit it, giving us the final forms as:

$$\begin{aligned} r &= R \\ g &= G - R - x \cdot (B - R) \\ b &= B - R - y \cdot (G - R) \end{aligned} \tag{5.7}$$

Refined method. In the description above, we made simplifications by truncating the low response of red pixels in other parts of the spectrum to zero and assumed equal response of all pixels in the NIR part of the spectrum around 850 nm. In the next step we refined the physics-based method by removing these simplifications. This refined method supersedes the previous simplified version entirely, as it is a better representation of ground truth.

However, both are included in subsequent results to show evolution of the work, which started with the simplified method [178] and continued with the refined version. To refine the method, we introduced additional parameters describing the channel overlap in the initial equations, which are modified as follows:

$$\begin{aligned} R &= r + i \cdot g + j \cdot b \\ G &= w \cdot r + g + x \cdot b \\ B &= q \cdot r + y \cdot g + b \end{aligned} \tag{5.8}$$

where R, G and B are again the values obtained from the sensor, r, g and b are the corresponding channel-separated red, green and blue values, and i, j, w, x, q and y are the coefficients describing the ratio of the corresponding color present in the mixture. This system can be represented in matrix form as:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & i & j \\ w & 1 & x \\ q & y & 1 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix} \tag{5.9}$$

representing the system $d = W \cdot x$, where d is the data vector of our measured values, W is the weights matrix of our coefficients and x is the vector of our channel-separated values of interest. We solved this system using `sympy` library and obtained the following results:

$$\begin{aligned} r &= \frac{-B \cdot i \cdot x + B \cdot j + G \cdot i - G \cdot j \cdot y + R \cdot x \cdot y - R}{-i \cdot q \cdot x + i \cdot w + j \cdot q - j \cdot w \cdot y + x \cdot y - 1} = \\ &= \frac{R(1 - xy) - G(i - jy) - B(j - ix)}{1 + iqx + jwy - iw - jq - xy} \\ g &= \frac{-B \cdot j \cdot w + B \cdot x + G \cdot j \cdot q - G - R \cdot q \cdot x + R \cdot w}{-i \cdot q \cdot x + i \cdot w + j \cdot q - j \cdot w \cdot y + x \cdot y - 1} = \\ &= \frac{G(1 - jq) - R(w - qx) - B(x - jw)}{1 + iqx + jwy - iw - jq - xy} \\ b &= \frac{B \cdot i \cdot w - B - G \cdot i \cdot q + G \cdot y + R \cdot q - R \cdot w \cdot y}{-i \cdot q \cdot x + i \cdot w + j \cdot q - j \cdot w \cdot y + x \cdot y - 1} = \\ &= \frac{B(1 - iw) - G(y - iq) - R(q - wy)}{1 + iqx + jwy - iw - jq - xy} \end{aligned} \tag{5.10}$$

where all the values on the right-hand side of equations are either known or can be computed with the same process of definite integration of the curves in Figure 5.6 as before. Doing so, we computed $i = 0.07$, $j = 0.06$, $w = 0.90$, $x = 0.61$, $q = 0.98$ and $y = 0.12$.

We show and analyze the waveform effects of this physics-based channel separation in detail in the following Chapter. We focused especially on the analysis of PTT before and after this procedure and also on its correlation with BP, which is of vital importance for our end goal of explicit BP estimation from MW single-site PTT.

Such a method is fully dependant on precise quantum efficiency and spectrum information, which are unique for each image sensor and light source. Our aim was to generalize this method and make it fully data-driven in the sense that we propose an algorithm that can be used without the image sensor quantum efficiency and light source specifics known in advance.

5.2.3.3 Generalized Channel Separation

In the previous approach, we have read the information about the spectral overlap from the quantum efficiency in Figure 5.6. However, more generally we can assume that each color channel is a combination of all other channels. Thus the RGB traces obtained from a camera can generally be written as given in Equation 5.11

$$\begin{aligned} R &= a_1 \cdot r + a_2 \cdot g + a_3 \cdot b \\ G &= b_1 \cdot r + b_2 \cdot g + b_3 \cdot b \\ B &= c_1 \cdot r + c_2 \cdot g + c_3 \cdot b \end{aligned} \tag{5.11}$$

where R, G, B are the original channel-overlapped traces obtained from the camera, r, g, b are the actual pure channel-separated responses, and a_n, b_n and c_n are the coefficients representing the ratios of each response present in the overlap.

Understanding the underlying physiology of the skin tissue and pulse wave propagation properties described previously, we can make some assumptions that simplify the general system in Equation 5.11. As mentioned before, the blue trace can be discarded, since it exhibits relatively noisy signal with low SNR, as well as makes little physiological sense due to its low penetration depth and lack of perfusion in the epidermis [190]. Furthermore, we always consider the coefficient corresponding to the color we are trying to separate (e.g., a_1 for red and b_2 for green in Eq. (5.11)) to be 1. Finally, the remaining coefficients in Equation 5.11 can be constrained to the range [-1, 0], as we are always subtracting undesired response from the mixture, never adding it. Considering these constraints, our general model simplifies to the one in Equation 5.12

$$\begin{aligned} R &= r + a_2 \cdot g + a_3 \cdot b \\ G &= b_1 \cdot r + g + b_3 \cdot b \end{aligned} \tag{5.12}$$

where the remaining variables represent the same quantities as reported in Equation 5.11. These simplifications notably decrease the search space, but the system remains overdetermined and challenging to solve [191]. We decided to investigate genetic algorithms (GA) to find approximate solutions with respect to two different fitness functions and also compare against BSS methods that are common in literature [192].

Genetic Algorithm using BP (GA-BP). Our goal was to measure the PTT in order to ultimately train a regressor for BP estimation. It only makes sense to use such a regression model as a fitness function if the importance of the PTT to the model is high, meaning it can serve as a meaningful physiological feature for BP estimation.

We thus initially checked the average feature importances of the trained Random Forest regression models using the PTTs obtained after our initial physics-based channel separation. We ran a LOSO experiment with personalization, meaning that one instance of rest and one instance of activity of the left out subject were added to the training data. This simulates model calibration to a specific subject, which is widely accepted as a vital step to improve BP prediction performance [108]. We checked the average importance of the PTT for SBP and DBP estimation as per mean decrease in impurity (MDI), and found that the PTT is extremely dominant compared to other features, as shown in Figure 5.7.

This shows that the model relies heavily on the PTT and makes sense to use as a fitness function to evaluate the quality of the PTT, or rather the channel separation algorithm that facilitates its computation. In addition to the PTT, some commonly-used morphological and demographic features were added to each instance for the training of the Random Forest regressor. These are listed in Table 5.1 and explained in Figure 5.8.

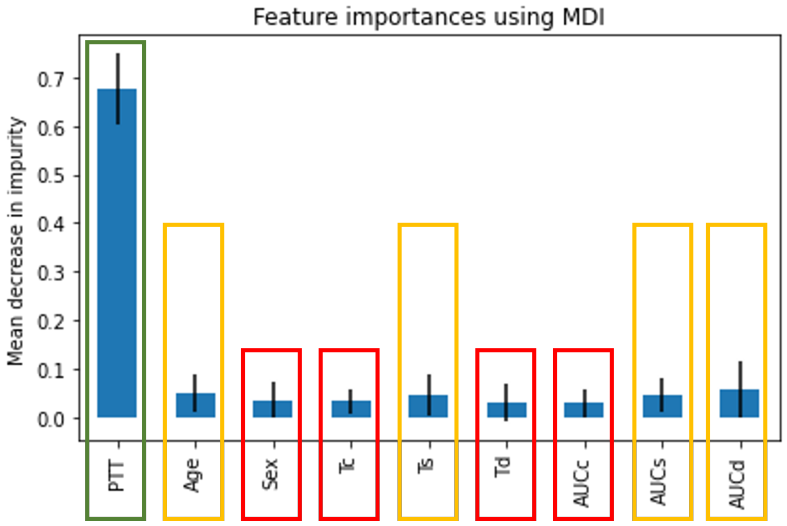

Figure 5.7: Comparison of average feature importances obtained from the Random Forest regressors for BP prediction. Green box denotes the most important feature, while orange and red boxes denote features with low to very-low importance.

Table 5.1: The set of morphological and demographic features that were used together with PTT in training a BP regression model.

Group	**Feat.**	**Description**
Morphological	T_c	Average cycle length
	T_s	Average systolic rise time
	T_d	Average diastolic fall time
	AUC_c	Average AUC of the whole cycle
	AUC_s	Average AUC of the systolic rise
	AUC_d	Average AUC of the diastolic fall
Demographic	Age	Subject age
	Sex	Subject sex

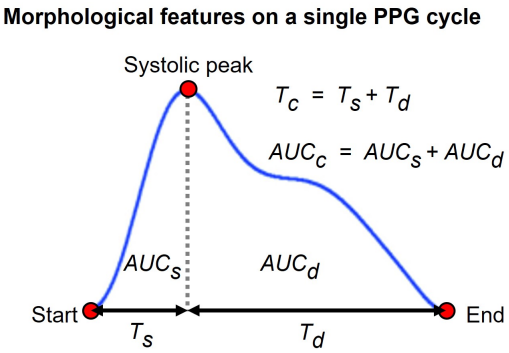

Figure 5.8: Visual representation of the morphological PPG features per cycle.

After PTT importance was confirmed, we framed the problem described in Equation 5.12 as an optimization problem, where one determines the optimal value of coef-

ficients with respect to a fitness function. When choosing the fitness function, we initially decided to use the average MAE of the accompanying trained Random Forest regressors predicting SBP and DBP. The regressor was evaluated each time in a 5-fold CV experiment without shuffling (instances of the same subject stayed together to avoid overfitting within the CV), using the PTTs computed with a given candidate vector.

Once the MAE of a regression model estimating BP was chosen as the fitness function, we attempted to find solutions to our optimization problem by using a genetic algorithm that trained a regression model for the computation of the fitness function of each candidate vector. For our case, we defined an initial population of $n = 100$ vectors containing random values (within previously described constraints) representing the channel separation coefficients we were optimizing. We optimized them through $g = 200$ generations. Tournament selection was used for creating the next generation offspring due to its simplicity and effectiveness in maintaining genetic diversity. We set the tournament size to $k = 5$, selecting k subjects at random for each tournament. Fitness of each subject in the tournament was then computed and the winner with the best fitness function was chosen for crossover until we had a sufficient number of selected parents. We used arithmetic crossover, which linearly combines two parent vectors to produce two new offspring using a weighted average of corresponding elements from the two parent vectors. We set the weight of each parent to $\alpha = 0.5$, making them both contribute equally to the computation of the offspring, which was done as given in Equation 5.13

$$o_i = \alpha \cdot a_i + (1 - \alpha) \cdot b_i, \tag{5.13}$$

where o_i is the i-th element of the offspring, α is the parameter influencing the contribution of each parent, and a_i, b_i are the i-th elements of each parent.

We also used standard random mutation, where we randomly added or subtracted a small random value in the range [0.01, 0.1] to a small subset of subjects. Such a GA approach is expected to converge towards channel-separation coefficients that minimize the BP estimation error (which is our fitness function and our ultimate goal), but does not ensure convergence to a global optimum. It also requires training a regression model with current PTTs each time a candidate vector is evaluated during the execution of the GA. This is why we decided to use a Random Forest regression model, as it can be trained quite quickly, while historically showing good and robust performance on a wide range of problems [193]. The pseudocode is given in Algorithm 5.1.

Generalized Genetic Algorithm using Phase Delay (GA-PD). The previously described algorithm has a major drawback in that it requires ground-truth BP measurements for the regression models training, which are in general not readily available. The evaluations are also computationally expensive, especially if the chosen regression model is complex (e.g., Support Vector Machine with polynomial kernel). We wanted to further generalize to achieve a solely input-data-driven approach, meaning a change in fitness function was required.

As the initial problem was rooted in the fact that PTTs are near-impossible to measure due to the mixture of traces, and we wanted to separate them as much as possible, an intuitive approach is to thus consider maximizing the phase delay between the channel-separated traces, assuming that this preserves the per-cycle PTT information in the sequence. This can be alternatively defined as minimization of the cross-correlation between the traces (up to a certain maximal lag threshold) [193].

The cross-correlation between signals x and y at different lags (phase delays) was computed using the MATLAB implementation in the function $xcorr(x, y)$. This returns correlations between the two signals at different possible delays or lags, up to a pre-defined

Algorithm 5.1: GA-BP using Random Forest regressor as fitness

INPUT: RANDOM POPULATION $n \times [1, a_2, a_3, b_1, 1, b_3]$
minimize MAE of regressors:
  for each vector:
    $RGB_i = channel\_sep(RGB, [1, a_2, a_3, b_1, 1, b_3]_i)$
    $PTT_i = compute\_PTT(G_i, R_i)$
    $morph_i = compute\_morph\_feats(G_i)$
    $feats = merge\_feats(PTT_i, morph_i, age, sex)$
    $MAEs = train\_test\_model(feats, BPs)$
    $fitness = avg(MAE_{SBP}, MAE_{DBP})$
  $selected\_vectors = tournament\_selection(fitness)$
  $offspring = arithmetic\_cross(selected\_vectors)$
  $mutate(offspring_{subset})$
repeat g times
OUTPUT: BEST COEFFICIENTS $[1, a_2, a_3, b_1, 1, b_3]_{best}$

threshold. It makes sense to set the threshold at average cycle length, as cross-correlation for periodic repetitive signals decreases up to a half cycle length and then begins to increase again, since the signals become re-aligned. Moreover, the delay we are interested in is physiological and should be notably less than one cycle. The phase delay at which the correlation is the lowest is then taken. Algorithm 5.1 thus gets modified as given in 5.2.

Algorithm 5.2: GA-PD using phase delay/xcorr as fitness

INPUT: RANDOM POPULATION $n \times [1, a_2, a_3, b_1, 1, b_3]$
minimize xcorr:
  for each vector:
    $RGB_i = channel\_sep(RGB, [1, a_2, a_3, b_1, 1, b_3]_i)$
    $fitness = min(xcorr(G_i, R_i, lags = avg\_cycle\_len))$
  $selected\_vectors = tournament\_selection(fitness)$
  $offspring = arithmetic\_cross(selected\_vectors)$
  $mutate(offspring_{subset})$
repeat g times
OUTPUT: BEST COEFFICIENTS $[1, a_2, a_3, b_1, 1, b_3]_{best}$

The proposed GA-PD Algorithm 5.2 substantially and importantly lessens the input requirements as it does not require the ground-truth BP to train a regression model for the fitness function computation each time. It also omits the potentially high time complexity of the specific regression model training. For a simple model, like Random Forest, the time complexity is $O(n \cdot log(n))$ [194], but this can increase dramatically for more complex models like SVM, reaching $O(n^2)$ or even $O(n^3)$ depending on the kernel used [195].

5.2.4 Experimental Design and Evaluation

Effects, performance and comparison between different proposed channel separation methods is reported in detail in the following chapter. Generally we expected the algorithms to converge to similar coefficient values for channel separation as those that were obtained explicitly from the quantum efficiency of our camera using the physics-based approach. Naturally the effects of the channel separation must also make sense in the scale of the ob-

tained PTTs (milliseconds) and importantly in the decreased MAEs of the BP estimation model.

To estimate the latter, we again devised a robust LOSO evaluation experiment with and without personalization. This experiment was chosen as it avoids overfitting or rather its personalization can be fully and clearly controlled. In the case without personalization, the model was trained on all subjects except the one it was tested on. As we mentioned previously, it was shown in related work that it is infeasible to train a generalized BP prediction model for general population, meaning that calibration/personalization should be considered to substantially improve the performance [108]. We thus investigated personalization by adding two instances of the left-out subject (one with elevated BP, one with resting BP) to the training data to personalize the model in each iteration of the LOSO experiment.

We compared MAEs computed in these two final experiments when using the channel separation coefficients from the physics-based approach, GA-BP and GA-PD, against standard BSS methods of PCA and ICA, and also against the baseline of no channel separation. Our results are reported in the next chapter, followed by discussion and interpretation of the obtained results.

5.3 Results

The optical monitoring experiments were split into two parts – PPG reconstruction and subsequent PTT measurement, and BP estimation with a regression model. Since we conducted experiments on a dataset collected by ourselves and no other work dealt with this specific methodology, let alone published a dataset on it, we had no ground truth for the PTT. This means that the first part of results presents the exploratory analysis of PTT. The observations are compared with hypotheses based on our understanding of underlying physiology and literature. For the BP estimation part we compared with the ground-truth clinical digital sphygmomanometer and placed the obtained results in the context of standards for clinical BP estimation devices.

5.3.1 Channel Separation Coefficients

For our specific experimental setup, the physics-based channel separation approach can be considered the ground truth in terms of channel separation coefficients, since the latter are computed directly using camera-specific quantum efficiency of the image sensor. Specifically, the most precise computation was done with the refined physics-based method described in Section 5.2.3.2. It thus makes sense to compare the channel separation coefficients obtained with the GA-based data-driven methods to those of the refined physics-based approach, to see if they converge to the same values. These are reported in Table 5.2 for all methods except PCA and ICA, which do not return coefficients but the channel-separated traces directly. We can observe close matches and very small differences between all considered methods, indicating convergence to the same values.

The channel separation coefficients of the original physics-based approach used a couple of simplifications, such as the response of the NIR channel in other wavelengths being truncated to zero, and the response of all channels in the NIR band (800–900 nm) being equalized. However, the refined variant, alongside the GA-BP and GA-PD algorithms, did not use these assumptions and thus all returned non-zero values for a_2 and a_3 as seen in Table 5.2. The coefficients obtained with the GA-BP and GA-PD algorithms were similar to those obtained with the original simplified physics-based approach, with a slight discrepancy – b_1 and b_3 differed by 0.06 and 0.045 on average, respectively, which is a

Table 5.2: Comparison of the channel separation coefficients obtained using the physics approach, the GA and the GA-PD. a_1 and b_2 were constant due to the constraints we used.

Coefficient	Physics (orig.)	Physics (refined)	GA-BP	GA-PD	Max diff. Δ
$a_1(constant)$	1.0	1.0	1.0	1.0	-
a_2	0.0	0.07	0.07	0.04	0.07
a_3	0.0	0.06	0.03	0.06	0.06
b_1	-1.0	-0.90	-0.95	-0.93	-0.10
$b_2(constant)$	1.0	1.0	1.0	1.0	-
b_3	-0.61	-0.61	-0.64	-0.67	-0.06

5.3% overall difference. However, comparing with the refined physics-based approach, we see these discrepancies decrease further, as b_1 and b_3 differ by only 0.04 and 0.045 on average, which is now a 4.3% difference. Furthermore, comparing values of a_2 and a_3 to the refined physics-based approach now shows minuscule differences of 0.015 on average, which is a substantial improvement compared to previous differences. This indicates that the GA approaches converged closer to the real ground truth represented by the refined physics-based approach.

5.3.2 PTT Analysis

We analyzed the PTTs between different color channel traces, as computed from the corresponding steepest systolic reference points defined in the previous section. Initially, before channel separation, the PTTs were miniscule and sometimes negative, indicating overlapping spectral response. As a consequence, the computed PTTs initially exhibited little-to-no variation, making them non-informative and physiologically meaningless. After the channel separation step, the PTTs became much more distinct and varied, while also appearing in the chronological order expected based on physiological background and related work [56], [179]. It is important to keep in mind that a single PTT value was computed for each signal segment as the average of per-cycle PTTs, as we had only a single ground-truth BP value corresponding to each recording. The average changes in PTTs before channel separation and after each proposed method are reported in Table 5.3.

Table 5.3: Overview of average PTT changes when using different channel separation methods.

Ch. Sep. Method	Avg. PTT	Δ from baseline	% in expected order
Baseline (no separation)	2.7 ms	-	61.4%
Physics-based (orig.)	46.9 ms	+42.2 ms	98.7%
Physics-based (refined)	47.4 ms	+44.7 ms	98.9%
GA-BP	44.8 ms	+40.1 ms	97.3%
GA-PD	49.2 ms	+44.5 ms	99.1%

We can observe that all the channel separation methods achieve similar average PTTs and increases compared to the baseline. Our initial hypothesis based on literature [73] and physiological background described in Section 2.2.3 was that the cardiac wave from the deeper vessels in the NIR range should be observed first, while those in the shallower layers like papillary dermis should be delayed. We defined our observed time as $PTT = timestamp\_steepest_{\text{green}} - timestamp\_steepest_{\text{NIR}}$, so we expected them to be positive, if in agreement with the order hypothesis.

The averages reported in Table 5.3 indicate that the channel separation algorithms

separate the color traces in a way that substantially increases PTT, however, it is of utmost importance when doing so to not lose the physiological information. Given our understanding of relationship between PWV, PTT and BP, we also expected the PTTs measured after activity scenario to be shorter compared to the rest scenario, as BP and SVR are higher, increasing the PWV and decreasing the PTT.

We checked the channel separation effects in more detail by creating boxplots of all per-cycle PTTs for each subject in both scenarios. These are shown for each method in Figure 5.9, confirming the average lower PTTs after activity while also showing a vast majority of individual measurements to be positive and in the expected order. In contrast, using no channel separation (baseline) yields low PTTs centered around 0, while also exhibiting much smaller differences between the rest and activity scenario.

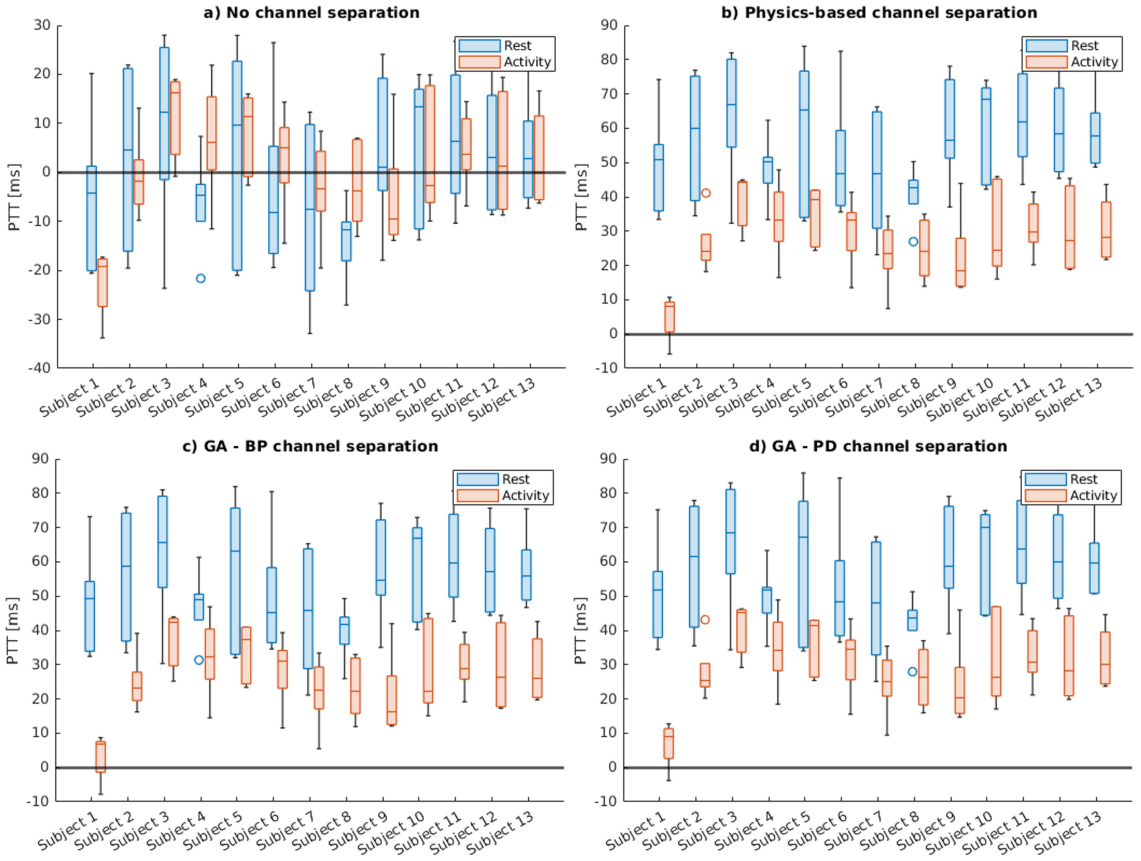

Figure 5.9: Boxplots of per-subject PTTs after different channel separation methods. The baseline without channel separation in subplot a) exhibits low and often negative PTTs. We only show a single plot for physics-based channel separation (refined version) approach due to high similarity between all non-baseline plots resulting from different channel separation methods.

Positive effects can also be observed on the waveforms themselves, as shown in Figure 5.10. We show an example only for the original physics-based channel separation, as the effects are quite similar with all the methods.

These initial results indicate channel separation effects in line with physiology and literature-based expectations, however, this in itself does not ensure the methods being useful, since the relation between PTTs and physiological changes may still lose important nuances.

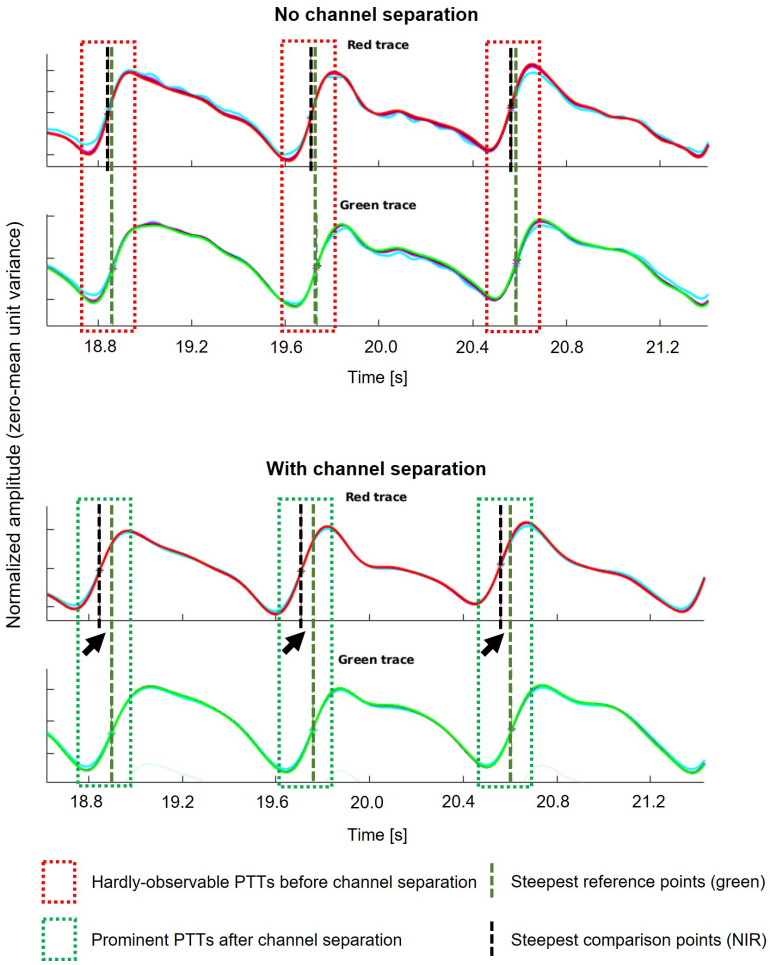

Figure 5.10: Visual presentation of channel separation effects on example rPPG waveforms from NIR and green part of the spectrum.

5.3.3 PTT and BP Correlation

To ensure that the effects of channel separation preserve the physiological information relating to PWV, SVR and ultimately BP, we investigated the correlations between the measured PTTs and corresponding ground-truth BP.

Initially we checked the correlations across our whole dataset, taking average PTTs from recordings of all subjects and both scenarios and their corresponding measured BP values. This is shown in Figure 5.11.

We can initially observe good separation between the blue and red group, corresponding to the rest and activity scenario. The blue group on average exhibits longer PTTs and lower BP values, while the red group exhibits the opposite characteristics – shorter PTTs and higher BP values. This holds for both SBP and DBP, although the variations in DBP are smaller compared to SBP. This is a known fact in medical literature, as DBP is generally more stable and changes less [196].

Additionally we can observe good correlation between PTT and BP, with Pearson's correlation coefficients (R) of -0.51 ($p < 0.001$) and -0.63 ($p < 0.001$) for DBP and SBP, respectively. P-values were computed using a t-test in which we tested the null hypothesis that there is no relationship between the observed phenomena.

While the above results give a reasonable overview of our whole dataset, we also inves-

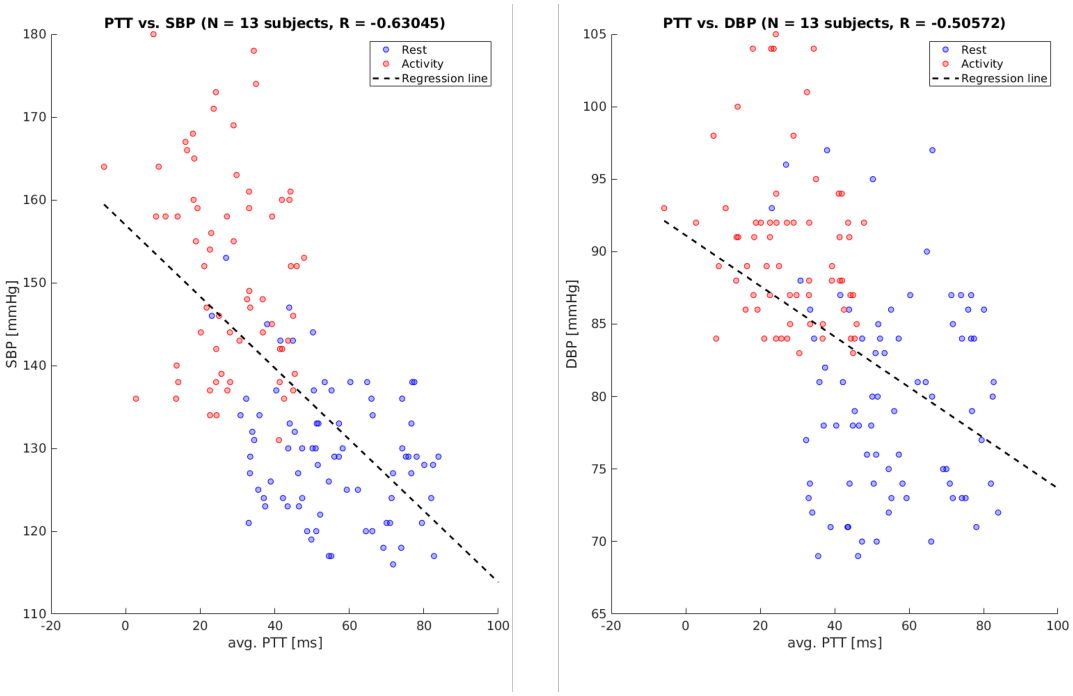

Figure 5.11: Correlations between average PTTs of all recordings and the ground-truth SBP and DBP values.

tigated the per-subject correlations. The latter have substantially lower statistical power compared to correlations on a larger sample size, but still give an intuitive overview of whether the correlation between MW PTT and BP holds generally, or if it is subject-dependent. These are shown in Figure 5.12, in which we can observe and confirm that the correlation also holds for each subject individually.

We have already observed previously in Figure 5.7 that a Random Forest regression model trained with such post-channel-separation MW PTT (alongside other demographic and morphological features) for SBP and DBP prediction assigns the highest feature importance to PTT. Detailed correlation analysis confirms the relationship between the two variables, further confirming the validity of the approach.

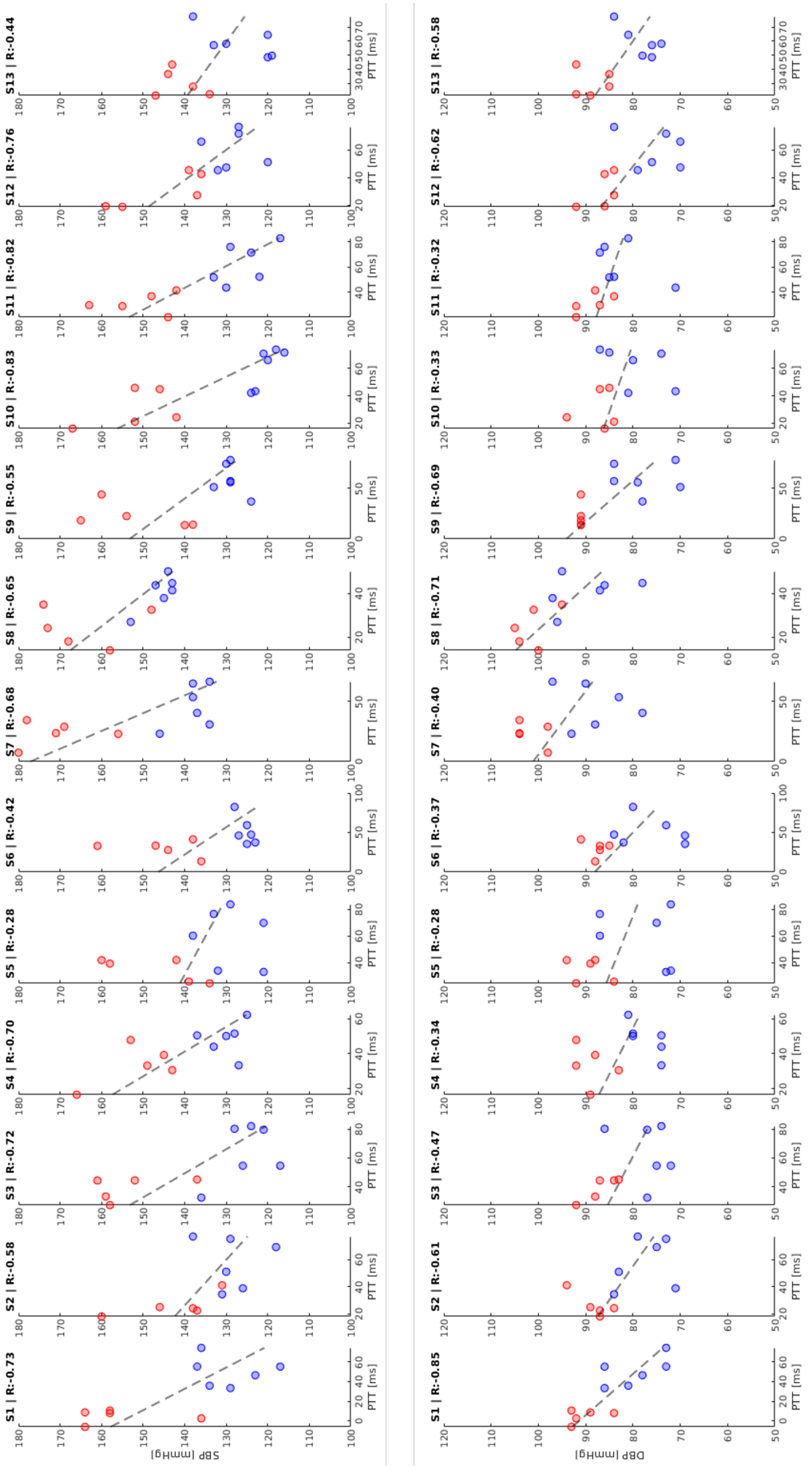

Figure 5.12: Correlations between average PTTs of each subject (S) and their ground-truth BP values.

5.3.4 BP Estimation

As described previously, we conducted a LOSO experiment to evaluate our regression model when using PTTs obtained from different channel separation algorithms.

It is a point of contention in literature whether a general BP estimation model is possible to train [108]. General relationship between PWV, PTT and BP holds true universally, however, as we saw in Chapter 2, many factors can contribute to changes in PWV and PTT. Based on our experience and related work [108], we decided to compare the performance of a general model to that of a personalized model, which uses some instances of the left-out subject to calibrate. This type of calibration by adding some instances of the test subject makes the evaluation optimistic compared to LOSO, but it is deliberate and based on physiological grounds.

The average MAEs for both types of experiments are reported in Table 5.4.

Table 5.4: Comparison of the final MAEs in mmHg for SBP and DBP estimation when using different channel separation algorithms. We compare against the baseline of using no channel separation. We report results for experiments with and without personalization in the final LOSO experiment.

	General regressor [mmHg]			*Personalized regressor [mmHg]*		
Algorithm	$\mathbf{MAE_{SBP}}$	$\mathbf{MAE_{DBP}}$	$\mathbf{MAE_{AVG}}$	$\mathbf{MAE_{SBP}}$	$\mathbf{MAE_{DBP}}$	$\mathbf{MAE_{AVG}}$
Baseline	*11.31±1.50*	*9.02±1.60*	*10.17±1.55*	*8.64±1.62*	*6.12±1.48*	*7.38±1.55*
PCA	10.22±1.31	8.91±1.19	9.57±1.25	8.01±1.25	5.99±1.35	7.00±1.30
ICA	9.81±1.20	6.97±1.10	8.39±1.15	6.83±1.10	5.75±1.30	6.29±1.20
Physics (or.)	7.72±1.00	5.46±0.98	6.59±0.99	4.78±0.96	3.89±0.97	4.34±0.97
Physics (ref.)	6.94±1.02	5.03±0.96	6.06±1.01	4.00±0.94	2.88±0.99	3.39±1.00
GA-BP	**6.89±0.95**	**4.91±0.98**	**5.90±0.97**	**3.48±1.02**	**2.61±0.90**	**3.05±0.96**
GA-PD	7.02±1.03	4.97±0.97	6.00±1.00	4.01±0.98	3.03±0.96	3.52±0.97

The errors indicate that a personalized regressor outperforms the general regressor on average by 2.53 mmHg. The differences are relatively consistent between different channel separation methods. To more robustly analyze its performance and show differences between a general and personalized model, we created a Bland-Altman plot, which is a common visual representation of means and variance of errors, alongside limits of agreement between actual and predicted values [197]. It is shown for the best-performing GA-BP algorithm in Figure 5.13.

When comparing the Bland-Altman plots of the best-performing algorithm for general (first column) and personalized (second column) models, we can observe similar distributions of errors with similar standard deviations for different means. Vast majority of errors fall within two standard deviations from the mean. The means between the general and personal model differ by 3.41 mmHg for SBP and 2.3 mmHg for DBP, whereas the standard deviations for the general model are ±0.95 mmHg and ±0.98 mmHg for SBP and DBP. For a personalized model, the standard deviations are similar at ±1.02 mmHg and ±0.90 mmHg. Considering this distribution of errors, and also looking at the Bland-Altman plots, we can see that the overlap between the general and personalized model is very small for SBP and slightly larger for DBP.

Another observation is that differences in errors are quite small between the both physics-based, GA-BP and GA-PD methods, while all of these differ more substantially from the baseline, PCA and ICA methods. The former group generally outperforms the latter, although when considering standard deviations, ICA becomes closer. Additionally, when no channel separation is used (baseline algorithm), the standard deviations are substantially larger compared to all other cases, indicating that the variance of the errors is

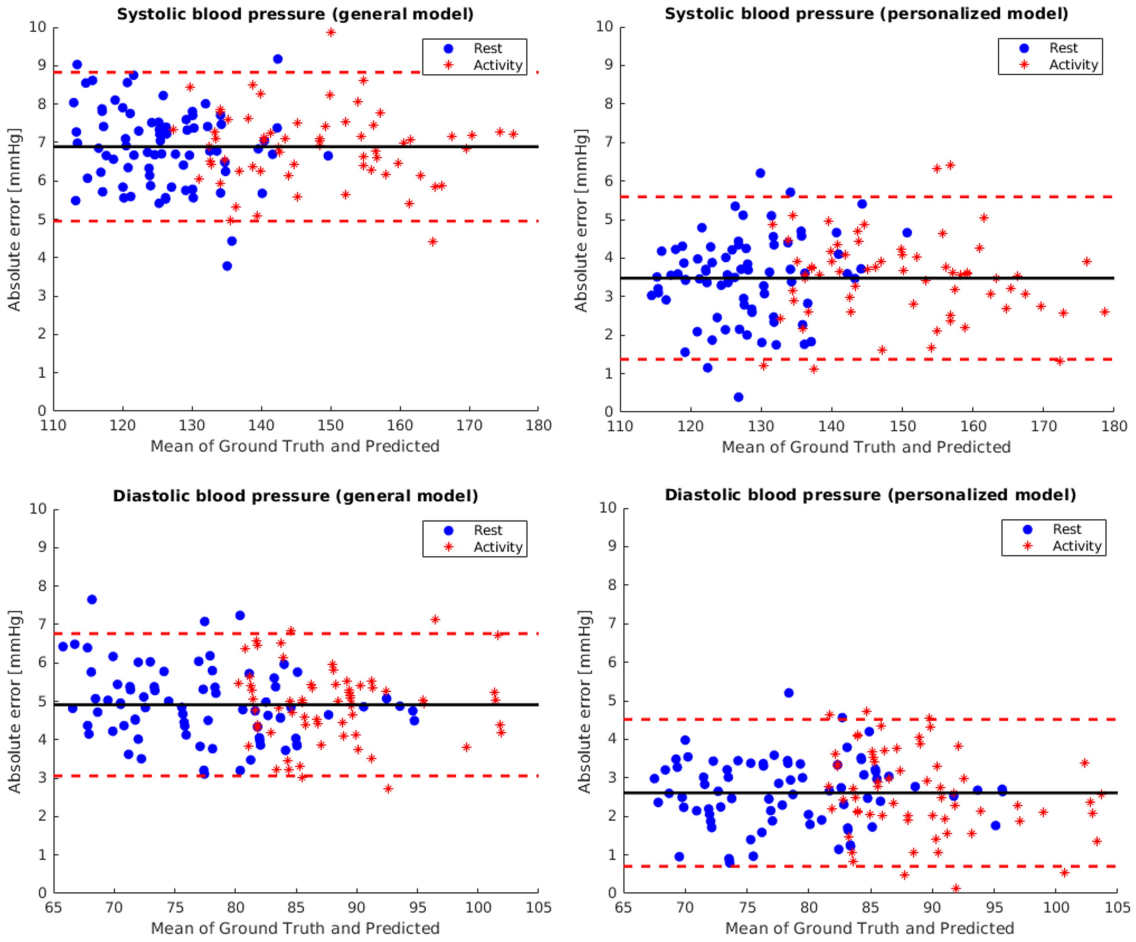

Figure 5.13: Bland-Altman plots after GA-BP channel separation for systolic (first row) and diastolic (second row) BP. Columns represent training a general or personalized regression model. Dashed lines are limits of agreement, defined as 1.96 times the standard deviation.

the largest in this case, and the mean values are least reliable.

Furthermore, we can observe that the distribution of errors is relatively equal across different values of SBP and DBP, observing no trends that would indicate a substantial increase of errors with lower or higher BP values.

In the next chapter, we will revisit the results reported in the current chapter, discuss and interpret their implications, practical feasibility, and place them in the context of standards where applicable. We will also accept or reject the hypotheses put forth in Chapter 3 in accordance with the obtained results.

Chapter 6

Discussion

The results reported in the previous chapter support several findings for both RF-based and optical monitoring of physiological parameters. We will discuss them individually per-topic, focusing especially on the implications for practical feasibility and potential in resolving challenges highlighted in the early chapters of this dissertation.

6.1 RF-based Classification of Hemodynamic Scenarios

We previously mentioned the main advantages of RF-based monitoring over optical, namely privacy preservation and potential for sleep monitoring due to no direct exposure and ambient lighting requirements. Apnea detection with a more comfortable method is especially valuable due to it being a dangerous condition that is not often monitored with conventional methods, as those are very obtrusive and difficult for the user (e.g., sleep laboratory or mask).

6.1.1 Interpretation of Results

Our results indicate that detection of complex hemodynamic and cardio-respiratory states, such as Apnea and Valsalva, which influence several physiological parameters at once, is feasible with a neural network achieving high accuracy of 88% using gold-standard contact sensor inputs. More importantly, this performance degrades only by 4-5% when using contact-free radar inputs instead of traditional contact sensors, showing that such detection remains accurate even when only radar is used. The performance degradation mostly occurs with similar scenarios, which are most often misclassified (tilting table and resting), while Valsalva and Apnea consistently maintain high accuracy and robust classification performance across different input modalities. We showed that results when using contact input modality are statistically significantly superior to those of radar input modality, which is expected, as both richer and more precise information is available in such sensors, such as ECG. Additionally, we saw that when fusion of both input modalities is used, there is no statistically significant difference between the performance. Again, this is not unexpected, since radar modality does not contribute additional information, meaning the performance stays at the level of contact input modality.

Additionally we showed that using longer signal segments as inputs statistically significantly improves performance by 15–20%. As explained earlier, this is likely due to a longer window better capturing the periodic nature of cardio-respiratory signals, thus offering a better representation of the physiological phenomena. This could also be a reason for slightly superior performance of frequency input data type, as the FFT makes more sense on more periodic signal segments. Moreover, we observed a slightly superior performance

of the fully-connected Dense network with frequency data representation, while 1D CNN worked best with temporal inputs. This could be interpreted as the fully-connected network being more suited for (frequency) feature-space representation of inputs, while CNN is better suited for raw temporal representation, allowing it to derive features on its own. However, the differences between different input data types proved to not be statistically significant, so this interpretation should be reconsidered.

Finally, the fully-connected Dense network overall slightly outperformed the CNN. More importantly, the training of such a network is less computationally expensive, meaning that even at similar performance, it is worth using.

When interpreting results of ANNs, special consideration must be given to generalization capabilities and overfitting, since such models have large expressive power. We opted for a tradeoff between computational complexity and robustness in carefully designing a 5-fold CV experiment that largely avoided overfitting. We also observed consistent performance across all folds, showing that classification performance is consistent regardless of the specific data split. While we withheld the test set every time, there is still the danger of generally overfitting to the dataset. The performance should thus be additionally validated either on additional datasets or better still in a real-world scenario. Only then can practical feasibility and generalization be completely robustly validated.

6.1.2 Significance and Value

Such a system could be implemented in a variety of scenarios, being beneficial to both average subjects as well as specific groups. We already mentioned its advantages of continuous non-invasive sleep monitoring, which greatly lessens the burdens of traditional highly-obtrusive setups. However, there are groups of people for which traditional monitoring is not only difficult, but completely infeasible, for instance people with PIMD [198], [199] or people with severe skin conditions (allergies, burns, etc.). A radar setup could be mounted statically in a configuration around the bed or other monitoring site, so that it would monitor cardio-respiratory movement in each direction, as a person can obviously turn even during sleep. Due to the aforementioned advantages of radar-based sleep monitoring, many manufacturers recently developed devices that facilitate it, such as Halo Rise from Amazon [142], [200] and One+ from Sleepiz [201].

The use cases are not limited only to medical scenarios, as for instance respiration is important in many fields. A non-medical example could be monitoring of respiratory patterns in musicians studying instruments that require specific skills and pulmonary manipulation, such as flute or trumpet. People studying such instruments are trained from a young age in regards to inhalation and exhalation patterns, which could be more easily traced and validated with a non-invasive radar system. Such a system could again be used both by teachers in schools, as well as practitioners in their home environment, where direct feedback from a teacher is not available.

Another popular use case could be meditation and mindfulness techniques often practiced by people to reduce stress and increase general well-being. These again heavily lean on correct respiratory patterns, which can be difficult to track and evaluate by practitioners, since it distracts them and thus defeats the whole purpose of such exercises.

6.1.3 Limitations

While the use cases are plentiful and performance on our dataset was good, there are some limitations that must be considered before real-world implementation. First, as briefly highlighted earlier, the whole evaluation was done on a single dataset. While we attempted to ensure robustness, validation should be conducted across different datasets.

Moreover, this dataset was closely resembling sleep monitoring, during which there are a lot less movement artefacts present. Movement artefacts are also not very problematic for potential applications in meditation and musician training, however, in many other everyday scenarios (e.g., exercise), a lot more movement noise is expected to be present.

Additionally, we relied on some derived signals as our inputs, which were computed in a black-box manner by the dataset authors. While we did interpret them and postulated them to be relatively simple to derive from the raw signals, this still requires both additional verification as well as computational efforts. Further generalization should be investigated in the sense of only using raw signal inputs, without their derivatives.

Finally, the proposed ANN approaches are black-box in nature, offering no interpretability or explainability. This could be partially alleviated using methods for model explainability, although these are largely designed in the context of 2D inputs, such as images. Still, a performance comparison with traditional ML approaches using hand-crafted features or applying explainability methods to existing black-box models could be valuable.

6.2 Optical Measurement of Multi-wavelength PTT and BP Estimation

When analyzing related work in earlier chapters, we highlighted the advantages of each existing approach, while also mentioning their disadvantages. Our proposed novel single-site MW PTT measurement and BP estimation approach merged the ideas of individual existing approaches with the aim of alleviating their disadvantages, while keeping their specific advantages. We showed that low errors can be obtained with such an approach, offering an attractive alternative to perform one of the most important and commonly performed physiological measurements.

6.2.1 Interpretation of Results

The reported results showed that when MW PTT is estimated remotely with an RGB camera, channel separation is a vital step. When using the quantum efficiency information of a specific image sensor, we can successfully isolate color channels of each spectral band of interest, increasing the measured PTT on average by 42 ms. This in turn also makes the steepest systolic reference points appear in the expected order. Given visual inspection of the rPPG waveforms, we saw that original waveforms are nearly identical in terms of positions of different reference points (steepest systolic point or systolic peak), which makes sense as they are a mixture of different spectral responses, which already have very subtle differences in isolation.

While channel separation can be done in various ways, the most robust starting point is to use the mentioned quantum efficiency information, as this gives precise information about the amounts of spectral overlap. In this sense it can be considered the ground truth for the specific image sensor. Comparing our GA algorithms for channel separation with the physics-based approaches representing the ground truth, we observed very close results both in terms of obtained channel separation coefficients, as well as BP prediction errors. We can conclude that both versions of GA converge towards physics-based ground truth, but do not reach precisely the same values. The channel separation coefficients of both GA approaches differ from the refined physics-based approach for about the same amount (difference of 2-3% between GAs), however, the GA-BP approach deviates in a way that allows for lower BP estimation errors, while GA-PD deviates in a way that allows for more apparent PTTs.

Looking at correlations between PTTs obtained with these channel separation approaches and the ground-truth BP values, we observed a strong correlation between the two. Importantly, we observed good separation between two groups – those with longer PTTs and lower BPs in the rest scenario and those with shorter PTTs and higher BPs after physical activity. This is in line with the physiological understanding reported in Chapter 2. Furthermore, the strong correlations between PTT and both SBP and DBP are statistically significant, confirming that PTT obtained after channel separation maintains physiological information and relationship with BP. Finally, we also checked the significance of PTTs via feature importances of a Random Forest regressor, which additionally confirmed its importance for BP estimation. Moreover, its importance was far larger compared to other simple demographic and morphological PPG features, which are often used in literature for BP estimation from PPG. Such morphological features are generally highly dependant on precise PPG morphology and subject of discussion in the field in terms of their general relationship with BP.

Moreover, superior and robust performance of the MW PTTs obtained with our proposed channel separation methods was confirmed in both general and personalized regression experiments. The former demonstrated higher errors on average, which is expected, since the model was completely subject-independent. When calibrating the model using only a few instances of the test subject, the errors decreased by 2-3 mmHg on average. We observed that the distribution of errors is stable across different BP values and standard deviations are relatively low at 1 mmHg on average.

There exist a number of standards for clinical BP monitoring devices. The two most widely used ones are the Association for the Advancement of Medical Instrumentation (AAMI) standard, which is commonly used in the U.S. for the performance and accuracy of blood pressure monitors, and the British Hypertension Society (BHS) standard, which is more common in Europe. Their requirements are as follows:

- **AAMI standard**: The AAMI standard (specifically the AAMI SP10) requires that blood pressure monitors should have a mean error within ± 5 mmHg and a standard deviation of < 8 mmHg when compared to a mercury sphygmomanometer.
- **BHS standard**: The BHS standards provide a grading system (A, B, C, worse) based on the accuracy of blood pressure monitors. Grades are based on the mean absolute difference between the predicted value and a mercury sphygmomanometer. All grades and their requirements are listed in Table 6.1.

Table 6.1: Grades of BP monitoring devices based on the BHS standard.

Grade	MAE ≤ 5 mmHg	MAE ≤ 10 mmHg	MAE ≤ 15 mmHg
A	60%	85%	95%
B	50%	75%	90%
C	40%	65%	85%
worse	$< 40\%$	$< 65\%$	$< 85\%$

Looking at the standards and the reported Bland-Altman plots in the previous chapter, we can conclude that the calibrated personalized model meets the AAMI SP10 requirements for both SBP and DBP, while the general model does not, especially for SBP. Furthermore, in terms of the BHS standard, the personalized model results would be placed in the A grade, while the general model would be borderline A grade for DBP and C grade for SBP. It is important to note that the SBP estimation is more difficult and important, since variation is greater, and SBP is the main early indicator of hypertension.

Another important consideration is the fact that our ground-truth BP was measured using a digital sphygmomanometer meeting these same standards (highest grade), which means that some error is also inherently present in our ground truth. It would be better to instead use a mercury-based sphygmomanometer, however, that would require a professional with experience (a medical worker) in order to obtain reliable measurements. Going even a step further, the most accurate ground truth would be obtained with an invasive catheter. However, it is easy to see why these two methods were not feasible in our experiments, as a trained professional was not available for extensive data collection and an invasive method is only plausible in a hospital setting. Should we obtain more precise ground truth, it is not unreasonable to expect our results to improve further, since PTT is expected to be even better correlated with actual BP than the one obtained with a digital sphygmomanometer.

6.2.2 Significance and Value

Due to the prevalence of hypertension and other cardiovascular conditions indicated by elevated BP, the significance of novel measuring methods is paramount. As before, the contact-free nature of measurement in itself offers important advantages, especially for sensitive subject groups that we discussed. Furthermore, it does not influence the physiology like a cuff does and potentially decreases the white coat syndrome effects due to its nature.

Compared to existing contact-free methods which rely on rPPG morphology, the PTT-based methods are much more robust, since the reconstructed rPPG waveform often loses detailed characteristics needed for feature computation (e.g., diastolic notch or peak). Compared to contact-free PTT methods, ours requires only a single measurement site, such as palm, making it more suited for privacy preservation (no face required) and more robust (no make up, glasses, hair, etc., which are common on the face).

Practical implementation would be valuable both in home environments and in clinician offices. An early prototype could consist of an enclosure (e.g., box) in order to isolate the dedicated light source and avoid interference and also to pre-define the distance between the camera and the palm. A patient could for example put their hand in such an enclosure before seeing his doctor, avoiding the necessary intervention of a nurse to place the cuff, while also obtaining a more relevant measurement (again avoiding the white coat syndrome and stress). The measurement could be forwarded to the doctor in advance, again saving time reporting the measured values.

In a home setting, this could potentially be extended for use with more wide-spread consumer devices, which are already omnipresent, such as webcams, phone cameras and bathroom mirrors (which can have a dedicated light source). If such measurements would be made daily, preferably several times, such measurements could be used for telemedicine and data aggregation across longer time periods, again without the need for wearing cuffs or other wearables. For such extensions, our proposed GA-PD channel separation approach would be especially valuable, as it would allow for implementation across different cameras without the need to calibrate it to a specific image sensor based on its quantum efficiency and without the need for ground-truth BP measurements to train a regression model in advance, making wide-spread adoption more feasible.

Such wide-spread home monitoring would be especially valuable for elderly people, who inevitably face hypertension and are often treated with prescription blood pressure medications [202].

6.2.3 Limitations

It should be kept in mind that all results were obtained in a highly-controlled prototype environment. As before, performance should be further validated on a larger dataset, ideally with subjects having various cardiovascular conditions. Moreover, the currently trained regression model is quite small and simple. A more powerful, potentially end-to-end model could be better suited for practical deployment.

As always, real-world applications also pose many challenges that were avoided via a controlled environment, namely environmental influence and more movement. The former could be circumvented with an enclosure as suggested, but the movement could be more problematic, introducing major artefacts in very subtle waveforms. A practical solution could be a mechanism for stabilization of the hand (e.g., memory foam within the enclosure) or further preprocessing dealing with major noise.

6.3 Hypotheses

In Chapter 3, we proposed three hypotheses based on our understanding of physiology and analysis of literature, which we now revisit:

1. **Hypothesis 1:** Contact-free sensing can offer comparable performance to traditional contact approaches in monitoring of complex hemodynamic conditions and vitals (apnea detection via respiration and blood pressure estimation).

2. **Hypothesis 2:** Fusion of different sensor and input modalities (contact + contact-free, temporal + frequency) can achieve superior performance and better robustness compared to using individual modalities.

3. **Hypothesis 3:** PTT can be measured between different skin layers leveraging different penetration of light using a customized off-the-shelf RGB camera and used as an informative feature for single-site contact-free BP estimation, achieving established medical standards for BP estimation in terms of error.

Evaluating them in the context of obtained results, we can come to the following conclusions:

1. **Hypothesis 1: Accepted.** We showed in Section 4.3 that using contact-free modalities we can achieve a performance within 5% of traditional contact sensors. Specifically, with radar signals we achieve accuracies 4-5% behind those obtained using contact signals for classification of five hemodynamic scenarios, including sleep apnea. We additionally demonstrated in Section 5.3 that using an RGB camera we can obtain (personalized) BP estimation models equalling the contact-PPG-based BP estimation performance reported in literature. Overall contact-free approaches were shown to achieve results rivalling traditional contact approaches.

2. **Hypothesis 2: Rejected.** We found equal performance between contact and fusion of input signal modalities when classifying five hemodynamic scenarios, as reported in Section 4.3. The minuscule differences of $\approx 1\%$ were not statistically significant. Importantly, we initially only hypothesized about the fusion of radar and traditional contact sensors, as a way to increase performance, and did not yet consider potential fusion of different contact-free sensors (e.g., radar and camera), as our data was not suitable. It is plausible that such fusion could increase performance and robustness.

3. **Hypothesis 3: Accepted.** We showed in Section 5.3 that with proper algorithmic channel separation we can obtain phase-delayed traces from different spectral bands, which in turn allow for pronounced and meaningful MW PTT measurement in accordance with literature. We showed that such a MW PTT exhibits the highest feature importance per mean decrease in impurity for both SBP and DBP. We used it to train regression models predicting SBP and DBP, which met the standards for clinical devices in terms of errors, when using personalization.

Chapter 7

Conclusions

In the early chapters of this dissertation we identified the need for contact-free physiological monitoring by highlighting many specific groups of people who would benefit greatly from such monitoring. Furthermore, advantages of unobtrusive physiological monitoring extend to the general population, especially in light of the world's aging population and inevitable rise of diseases that come with aging [203]. Here, we briefly summarize our findings and conclusions and propose directions for future work.

7.1 Summary

In our research we specifically focused on two parts of the EM spectrum that allow for unobtrusive privacy-preserving contact-free monitoring of physiological state, while being harmless to humans – radio frequency and visible + (near) infrared light. The former (RF-based) has its advantages in not requiring a light source and having better penetration, making it feasible for use even with barriers, but requires conspicuous movement to obtain meaningful information. The latter (optical) offers optical modality with richer information that allows for monitoring of parameters that are more difficult to observe, but requires a light source and direct exposure.

7.1.1 RF-based Classification of Hemodynamic Scenarios

We demonstrated that using a radar one can accurately detect major changes in hemodynamic condition of individuals, which also manifests in changes of the respiratory state. Complex states like apnea and Valsalva maneuver, where the circulatory system undergoes a cascade of deviations from normal functioning, are of particular interest and importance. Extreme states like apnea can lead to severe consequences and should be detected promptly and robustly. We found that training a branched ANN with radar data allows for highly-accurate detection of different hemodynamic states, including apnea, making it a good candidate for wide-spread adoption in sleep monitoring. Additionally, as briefly mentioned earlier, the use cases of unobtrusive hemodynamic monitoring (especially respiratory changes) extend beyond health applications, including monitoring and training for the general well-being (meditation) and specific activities where respiration is crucial (musicians, divers, etc.).

7.1.2 Optical Measurement of Multi-wavelength PTT and BP Estimation

In light of its overall importance as an indicator of cardiovascular state and many related diseases, BP is paramount. A lot of research effort in literature was directed towards

contact-free unobtrusive monitoring, as the awkward and user-unfriendly cuff-based methods remain prevalent. While direct BP estimation from rPPG is desired, there is little agreement in literature about the feasibility of such methods due to them requiring precise waveform morphology. Instead, a PTT-based approach is considered more feasible and robust, but suffers from the requirement of two synchronized sensors. MW approaches were proposed to circumvent this, but implementations on custom hardware are in early prototypes and importantly still have the traditional downsides of wearables (e.g., battery), while also compressing skin at the measuring site. We found that using a high-fps RGB camera modified for (N)IR recording it is possible to obtain traces from different skin layers at different depths. However, due to the design of modern image sensors, the obtained information is mixed between all three spectral bands, meaning the obtained rPPG signals initially exhibit no expected delay between corresponding reference points. We showed that using the precise information about the quantum efficiency of the image sensor allows for algorithmic separation of such signals based on camera physics, allowing for measurable PTTs. However, such an approach is camera-specific, so we instead proposed novel data-driven camera-independent approaches based on GA, which achieve nearly identical performance to the physics-based approach.

We further demonstrated that such MW PTT obtained from a single measuring site is the most important feature (compared to morphological and demographic features) when training a regression model to predict SBP and DBP. We finally found that personalized models achieve substantially better performance compared to general models, allowing for MW contact-free BP estimation borderline achieving AAMI and BHS standards.

While our initial prototype setup was highly-controlled, it is not infeasible to extended it to real-world applications in the form of enclosures for palm insertion (e.g., at a doctors office or in the waiting room, at home, etc.) and potentially also to non-dedicated sensors like smartphone cameras or bathroom mirrors, with suitable adjustments for loosened environments. While a dedicated device is still envisioned as an early prototype, its comfort, ease of use and speed arguably surpass that of the currently used cuff-based methods.

7.2 Future Work

Continuation of the work described in this dissertation can be based on the limitations of each approach described in Chapter 6. In short, the work we did investigated fundamental feasibility of proposed contact-free approaches for physiological monitoring, meaning the data originated from highly controlled settings in order to maximize its quality and observe the phenomena of interest. Extensions towards more varied and more challenging data from real-world scenarios are mandatory to confirm practicality and real-world potential of the proposed methods. Specifically, we expect the data to be much noisier in everyday life, meaning that additional robust preprocessing methods would likely be required.

On the other hand, such preprocessing could be done implicitly when using end-to-end deep learning approaches. In our RF-based monitoring scenario, we still included manual preprocessing in an attempt to increase SNR of the data being fed to the NN. However, such steps could potentially be excluded, as NNs were shown to be successful in dealing with noise present in the data and even increase robustness when noise is present [204]. In line with this reasoning, an end-to-end approach could also be applied to MW PTT and BP estimation. There are at least two ways to go about this. One could train a two-phase model attempting to extract MW PTT in the first phase and then estimate BP in the second phase, or a completely end-to-end NN could be trained, taking a sequence of images as input and attempt to predict BP directly. Naturally, in case of black-box approaches, especially when models have a high capacity for overfitting, an interpretability

and explainability analysis should accompany such attempts.

Another way to increase robustness is to obtain information from both radar and camera simultaneously or design a mechanism that selects the more adequate sensor based on contextual environmental conditions (e.g., radar at night, both during the day). One can envision the potential fusion of both sensors in a prototype device containing a radar and camera. It is plausible to assume this would enhance the performance and importantly increase robustness and real-world application potential.

Recently some ideas were also proposed in terms of cross-modal signal reconstruction, for instance reconstructing ECG (electrical modality) from PPG (optical modality) [205]. While such research is still in its early phases and its feasibility subject of discussion, a system could be proposed that would include a radar (RF modality) and a camera (optical modality) and then models could be investigated that would fuse available signals to reconstruct for example electrical modality as well (like ECG). In such a case, one could get more detailed QRS complexes that could allow for the detection of additional cardiac conditions and diseases, which manifest more subtly and cannot be observed in the PPG.

Additionally, some research recently proposed contact-free monitoring of novel parameters such as blood glucose [206], which is another exceptionally demanding invasive measurement that must be done frequently by diabetes patients. While some promising results were reported initially, it was later argued that feasibility of such monitoring is questionable [207]. Despite this, research in this direction continues.

Ultimately the goal of contact-free physiological monitoring is to ease the burden of demanding measurements, which must often be done regularly or even continuously, while nearing the accuracy and precision of current gold-standard wearable devices. Ideally, a system would be developed that would allow for completely non-intrusive and privacy-preserving monitoring of the widest array of parameters possible.

Privacy preservation is especially challenging, as it does not only relate to visual cues, which are more easily masked or circumvented, but physiological signature of individuals as well. We previously listed a number of applications where such a system could be useful, yet there are also many applications where it could be abused. Some examples include job (or other) interviews, games and sports (e.g., poker and chess), business (or other) negotiations, etc. These are all cases where information about a person's physiology, which could be obtained discreetly without them knowing, could provide important information not only about their physiological, but also psychological state, thus offering an advantage or leverage to one party over the other. To further extrapolate potential misuse of such technology, specific combinations of physiological parameters and signals allow for augmentation of already existing identification systems [208]–[210], which could lead to unwanted surveillance and tracking.

With suitable resolutions for the highlighted challenges (e.g., federated learning for model training, input signal transformations, etc.), such an unobtrusive system could immensely improve the speed, cost and comfort of physiological monitoring, both at home and in hospitals, and could include methods proposed in this dissertation to reach this goal.

References

[1] M. Haghi, K. Thurow, and R. Stoll, "Wearable devices in medical internet of things: Scientific research and commercially available devices," *Healthcare informatics research*, vol. 23, no. 1, pp. 4–15, 2017.

[2] T. Loncar-Turukalo, E. Zdravevski, J. Machado da Silva, I. Chouvarda, and V. Trajkovik, "Literature on wearable technology for connected health: Scoping review of research trends, advances, and barriers," *J Med Internet Res*, vol. 21, no. 9, e14017, 2019, ISSN: 1438-8871. DOI: 10.2196/14017. [Online]. Available: https://www.jmir.org/2019/9/e14017/.

[3] E. Sezgin and S. Ö. Yıldırım, "A literature review on attitudes of health professionals towards health information systems: From e-health to m-health," *Procedia Technology*, vol. 16, pp. 1317–1326, 2014.

[4] E. R. Dorsey and E. J. Topol, "Telemedicine 2020 and the next decade," *The Lancet*, vol. 395, no. 10227, p. 859, 2020.

[5] M. Rehman, R. A. Shah, M. B. Khan, *et al.*, "Contactless small-scale movement monitoring system using software defined radio for early diagnosis of covid-19," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17 180–17 188, 2021.

[6] M. B. Khan, Z. Zhang, L. Li, *et al.*, "A systematic review of non-contact sensing for developing a platform to contain covid-19," *Micromachines*, vol. 11, no. 10, p. 912, 2020.

[7] B. Calton, N. Abedini, and M. Fratkin, "Telemedicine in the time of coronavirus," *Journal of pain and symptom management*, vol. 60, no. 1, e12–e14, 2020.

[8] M. Weiser, "Some computer science issues in ubiquitous computing," *Communications of the ACM*, vol. 36, no. 7, pp. 75–84, 1993.

[9] K. Lyytinen and Y. Yoo, "Issues and challenges in ubiquitous computing," 2002.

[10] R. J. Martis, U. R. Acharya, and H. Adeli, "Current methods in electrocardiogram characterization," *Computers in biology and medicine*, vol. 48, pp. 133–149, 2014.

[11] S. M. Caples, A. S. Gami, and V. K. Somers, "Obstructive sleep apnea," *Annals of internal medicine*, vol. 142, no. 3, pp. 187–197, 2005.

[12] M. R. Pioli, A. M. Ritter, A. P. de Faria, and R. Modolo, "White coat syndrome and its variations: Differences and clinical impact," *Integrated blood pressure control*, pp. 73–79, 2018.

[13] Y. Sun, W. Wang, X. Long, *et al.*, "Respiration monitoring for premature neonates in nicu," *Applied Sciences*, vol. 9, no. 23, p. 5246, 2019.

[14] M. Kosiedowski, A. Radziuk, P. Szymaniak, *et al.*, "On applying ambient intelligence to assist people with profound intellectual and multiple disabilities," in *Intelligent Systems and Applications*, Y. Bi, R. Bhatia, and S. Kapoor, Eds., Cham: Springer International Publishing, 2020, pp. 895–914, ISBN: 978-3-030-29513-4.

[15] T. Hammann, J. Valič, G. Slapničar, and M. Luštrek, "The missing piece. physiological data as a factor for identifying emotions of people with profound intellectual and multiple disabilities," *International Journal of Developmental Disabilities*, pp. 1–17, 2022.

[16] K. Renaud and J. van Biljon, "Predicting technology acceptance and adoption by the elderly: A qualitative study," in *Proceedings of the 2008 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries: Riding the Wave of Technology*, ser. Saicsit '08, Wilderness, South Africa: Association for Computing Machinery, 2008, pp. 210–219. DOI: 10.1145/1456659.1456684. [Online]. Available: https://doi.org/10.1145/1456659.1456684.

[17] E. G. Nabel, "Cardiovascular disease," *New England Journal of Medicine*, vol. 349, no. 1, pp. 60–72, 2003.

[18] P. M. Kearney, M. Whelton, K. Reynolds, P. Muntner, P. K. Whelton, and J. He, "Global burden of hypertension: Analysis of worldwide data," *The Lancet*, vol. 365, no. 9455, pp. 217–223, 2005.

[19] T. Ferkol and D. Schraufnagel, "The global burden of respiratory disease," *Annals of the American Thoracic Society*, vol. 11, no. 3, pp. 404–406, 2014.

[20] A. Al-Naji, K. Gibson, S.-H. Lee, and J. Chahl, "Monitoring of cardiorespiratory signal: Principles of remote measurements and review of methods," *IEEE Access*, vol. 5, pp. 15 776–15 790, 2017.

[21] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 859–869, 2016.

[22] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders, "Towards mental stress detection using wearable physiological sensors," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Ieee, 2011, pp. 1798–1801.

[23] J. J. Batzel, F. Kappel, D. Schneditz, and H. T. Tran, *Cardiovascular and respiratory systems: modeling, analysis, and control*. Siam, 2007.

[24] G. J. Tortora and B. Derrickson, "The cardiovascular system: Blood vessels and hemodynamics," *Principles of anatomy and physiology*, pp. 610–635, 2012.

[25] W. Commons. "Blood flow through the heart." (2013), [Online]. Available: https://commons.wikimedia.org/wiki/File:2101%5C%5FBlood%5C%5FFlow%5C%5FThrough%5C%5Fthe%5C%5FHeart.jpg.

[26] M. Thiriet, *Biology and mechanics of blood flows: Part II: Mechanics and medical aspects*. Springer Science & Business Media, 2007.

[27] N. J. Langford, "Carbon dioxide poisoning," *Toxicological reviews*, vol. 24, pp. 229–235, 2005.

[28] J. Piiper and P. Scheid, "Respiration: Alveolar gas exchange," *Annual review of physiology*, vol. 33, no. 1, pp. 131–154, 1971.

[29] P. Safar, L. A. Escarraga, and J. O. Elam, "A comparison of the mouth-to-mouth and mouth-to-airway methods of artificial respiration with the chest-pressure arm-lift methods," *New England Journal of Medicine*, vol. 258, no. 14, pp. 671–677, 1958.

[30] W. Commons. "Respiratory system complete." (2007), [Online]. Available: https://commons.wikimedia.org/wiki/File:Circulatory%5C%5FSystem%5C%5Fen.svg.

[31] V. M. Remouchamps, N. Letts, D. Yan, *et al.*, "Three-dimensional evaluation of intra-and interfraction immobilization of lung and chest wall using active breathing control: A reproducibility study with breast cancer patients," *International Journal of Radiation Oncology* Biology* Physics*, vol. 57, no. 4, pp. 968–978, 2003.

[32] I. Milesi, M. Norgia, P. P. Pompilio, C. Svelto, and R. L. Dellaca, "Measurement of local chest wall displacement by a custom self-mixing laser interferometer," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 8, pp. 2894–2901, 2011.

[33] T. Leu, V. Schutzhold, J. Fandrey, and K. B. Ferenz, "When the brain yearns for oxygen," *Neurosignals*, vol. 27, no. 1, pp. 50–61, 2019.

[34] W. Commons. "Circulatory system." (2009), [Online]. Available: https://commons.wikimedia.org/wiki/File:Circulatory%5C%5FSystem%5C%5Fen.svg.

[35] J. Clarke, J. Shelton, G. Venning, J. Hamer, and S. Taylor, "The rhythm of the normal human heart," *The Lancet*, vol. 308, no. 7984, pp. 508–512, 1976.

[36] A. Kashani and S. S. Barold, "Significance of qrs complex duration in patients with heart failure," *Journal of the American College of Cardiology*, vol. 46, no. 12, pp. 2183–2192, 2005.

[37] W. Commons. "Phases of the cardiac cycle." (2013), [Online]. Available: https://commons.wikimedia.org/wiki/File:2027_Phases_of_the_Cardiac_Cycle.jpg.

[38] T. W. Secomb, "Hemodynamics," *Comprehensive physiology*, vol. 6, no. 2, p. 975, 2016.

[39] A. Sherwood, C. A. Dolan, and K. C. Light, "Hemodynamics of blood pressure responses during active and passive coping," *Psychophysiology*, vol. 27, no. 6, pp. 656–668, 1990.

[40] J. M. Solà i Carós, "Continuous non-invasive blood pressure estimation," Ph.D. dissertation, ETH Zurich, 2011.

[41] S. Fu, P. Ping, F. Wang, and L. Luo, "Synthesis, secretion, function, metabolism and application of natriuretic peptides in heart failure," *Journal of biological engineering*, vol. 12, pp. 1–21, 2018.

[42] J. Handler, "The importance of accurate blood pressure measurement," *The Permanente Journal*, vol. 13, no. 3, p. 51, 2009.

[43] E. Messas, M. Pernot, and M. Couade, "Arterial wall elasticity: State of the art and future prospects," *Diagnostic and interventional imaging*, vol. 94, no. 5, pp. 561–569, 2013.

[44] F. Callaghan, L. Geddes, C. F. Babbs, and J. Bourland, "Relationship between pulse-wave velocity and arterial elasticity," *Medical and Biological Engineering and Computing*, vol. 24, pp. 248–254, 1986.

[45] W. Wang, A. C. Den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2016.

[46] W. Commons. "Layers of the skin." (2012), [Online]. Available: https://en.wikipedia.org/wiki/File:Skin%5C%5Flayers.svg.

[47] J. McGrath, R. Eady, and F. Pope, "Anatomy and organization of human skin," *Rook's textbook of dermatology*, vol. 1, pp. 3–2, 2004.

[48] M. Brenner and V. J. Hearing, "The protective role of melanin against uv damage in human skin," *Photochemistry and photobiology*, vol. 84, no. 3, pp. 539–549, 2008.

[49] S. Sachdeva, "Fitzpatrick skin typing: Applications in dermatology," *Indian journal of dermatology, venereology and leprology*, vol. 75, p. 93, 2009.

[50] M. L. Hunter, *Race, gender, and the politics of skin tone*. Routledge, 2013.

[51] T. Lister, P. A. Wright, and P. H. Chappell, "Optical properties of human skin," *Journal of biomedical optics*, vol. 17, no. 9, pp. 090 901–090 901, 2012.

[52] A. Zirra, "The functional significance of the skin's stratum lucidum.," *Morphologie et embryologie*, vol. 22, no. 1, pp. 9–12, 1976.

[53] M. Wang-Evers, M. J. Casper, J. Glahn, *et al.*, "Assessing the impact of aging and blood pressure on dermal microvasculature by reactive hyperemia optical coherence tomography angiography," *Scientific reports*, vol. 11, no. 1, p. 13 411, 2021.

[54] W. Commons. "Arteries of a human left hand." (2015), [Online]. Available: `https://commons.wikimedia.org/wiki/File:Human%5C%5Fhand%5C%5Farteries.svg`.

[55] J. Habib, L. Baetz, and B. Satiani, "Assessment of collateral circulation to the hand prior to radial artery harvest," *Vascular Medicine*, vol. 17, no. 5, pp. 352–361, 2012.

[56] J. Liu, Y.-T. Zhang, X.-R. Ding, W.-X. Dai, and N. Zhao, "A preliminary study on multi-wavelength ppg based pulse transit time detection for cuffless blood pressure measurement," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Ieee, 2016, pp. 615–618.

[57] G. Davis and M. Bureau, "Pulmonary and chest wall mechanics in the control of respiration in the newborn," *Clinics in Perinatology*, vol. 14, no. 3, pp. 551–579, 1987.

[58] A. De Groote, M. Wantier, G. Chéron, M. Estenne, and M. Paiva, "Chest wall motion during tidal breathing," *Journal of Applied Physiology*, vol. 83, no. 5, pp. 1531–1537, 1997.

[59] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, R1–r39, 2007. DOI: `10.1088/0967-3334/28/3/r01`. [Online]. Available: `https://doi.org/10.1088%2F0967-3334%2F28%2F3%2Fr01`.

[60] J. Gordon, "Certain molar movements of the human body produced by the circulation of the blood," *Journal of anatomy and physiology*, vol. 11, no. Pt 3, p. 533, 1877.

[61] I. Sadek, J. Biswas, and B. Abdulrazak, "Ballistocardiogram signal processing: A review," *Health information science and systems*, vol. 7, pp. 1–23, 2019.

[62] P. Bedggood and A. Metha, "Direct measurement of pulse wave propagation in capillaries of the human retina," *Optics Letters*, vol. 46, no. 18, pp. 4450–4453, 2021.

[63] D. McDuff, "Camera measurement of physiological vital signs," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–40, 2023.

[64] M. Van Gastel, S. Stuijk, and G. de Haan, "Motion robust remote-ppg in infrared," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1425–1433, 2015.

[65] S. M. M. Islam, "Radar-based remote physiological sensing: Progress, challenges, and opportunities," *Frontiers in Physiology*, p. 2135, 2022.

[66] Y. S. Lee, P. N. Pathirana, C. L. Steinfort, and T. Caelli, "Monitoring and analysis of respiratory patterns using microwave doppler radar," *IEEE journal of translational engineering in health and medicine*, vol. 2, pp. 1–12, 2014.

[67] M. Zakrzewski, H. Raittinen, and J. Vanhala, "Comparison of center estimation algorithms for heart and respiration monitoring with microwave doppler radar," *IEEE Sensors Journal*, vol. 12, no. 3, pp. 627–634, 2011.

[68] Z. Zhang, Y. Liu, T. Stephens, and B. J. Eggleton, "Photonic radar for contactless vital sign detection," *Nature Photonics*, vol. 17, no. 9, pp. 791–797, 2023.

[69] Z. Ling, W. Zhou, Y. Ren, J. Wang, and L. Guo, "Non-contact heart rate monitoring based on millimeter wave radar," *IEEE Access*, vol. 10, pp. 74 033–74 044, 2022.

[70] D. Jung, S. Cheon, D. Kim, J. Yoon, and B. Kim, "Short time remote heart rate measurement based on mmwave fmcw radar frame structure," *IEEE Antennas and Wireless Propagation Letters*, 2023.

[71] M. Nosrati and N. Tavassolian, "High-accuracy heart rate variability monitoring using doppler radar based on gaussian pulse train modeling and ftpr algorithm," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 1, pp. 556–567, 2017.

[72] K. Shi, T. Steigleder, S. Schellenberger, *et al.*, "Contactless analysis of heart rate variability during cold pressor test using radar interferometry and bidirectional lstm networks," *Scientific reports*, vol. 11, no. 1, p. 3025, 2021.

[73] Y. Han, T. Lauteslager, T. S. Lande, and T. G. Constandinou, "Uwb radar for non-contact heart rate variability monitoring and mental state classification," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Ieee, 2019, pp. 6578–6582.

[74] Y. Xiao, C. Li, and J. Lin, "A portable noncontact heartbeat and respiration monitoring system using 5-ghz radar," *IEEE Sensors Journal*, vol. 7, no. 7, pp. 1042–1043, 2007.

[75] Y. Yang, J. Cao, X. Liu, and X. Liu, "Multi-breath: Separate respiration monitoring for multiple persons with uwb radar," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, Ieee, vol. 1, 2019, pp. 840–849.

[76] D. Buxi, J.-M. Redouté, and M. R. Yuce, "Blood pressure estimation using pulse transit time from bioimpedance and continuous wave radar," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 917–927, 2016.

[77] H. Zhao, X. Gu, H. Hong, Y. Li, X. Zhu, and C. Li, "Non-contact beat-to-beat blood pressure measurement using continuous wave doppler radar," in *2018 IEEE/MTT-S International Microwave Symposium-IMS*, Ieee, 2018, pp. 1413–1415.

[78] J. Zheng and S. Hu, "The preliminary investigation of imaging photoplethysmographic system," in *Journal of Physics: Conference Series*, IoP Publishing, vol. 85, 2007, p. 012 031.

[79] J.-P. Lomaliza, H. Park, and M. Billinghurst, "Combining photoplethysmography and ballistocardiography to address voluntary head movements in heart rate monitoring," *IEEE Access*, vol. 8, pp. 226 224–226 239, 2020.

[80] E. B. Blackford, J. R. Estepp, and D. J. McDuff, "Remote spectral measurements of the blood volume pulse with applications for imaging photoplethysmography," in *Optical diagnostics and sensing XVIII: toward point-of-care diagnostics*, Spie, vol. 10501, 2018, pp. 192–199.

[81] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light.," *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[82] J. Zheng, S. Hu, V. Chouliaras, and R. Summers, "Feasibility of imaging photoplethysmography," in *2008 International Conference on BioMedical Engineering and Informatics*, Ieee, vol. 2, 2008, pp. 72–75.

[83] S. Hu, V. A. Peris, A. Echiadis, J. Zheng, and P. Shi, "Development of effective photoplethysmographic measurement techniques: From contact to non-contact and from point to imaging," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Ieee, 2009, pp. 6550–6553.

[84] W. Wang and A. C. den Brinker, "Modified rgb cameras for infrared remote-ppg," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 10, pp. 2893–2904, 2020.

[85] J. Spigulis, "Multispectral, fluorescent and photoplethysmographic imaging for remote skin assessment," *Sensors*, vol. 17, no. 5, p. 1165, 2017.

[86] G. Scebba, G. Da Poian, and W. Karlen, "Multispectral video fusion for non-contact monitoring of respiratory rate and apnea," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 350–359, 2020.

[87] M. Lewandowska, J. Rumiński, T. Kocejko, and J. 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)*, Ieee, 2011, pp. 405–410.

[88] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE transactions on biomedical engineering*, vol. 58, no. 1, pp. 7–11, 2010.

[89] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[90] C. Zhao, M. Hu, F. Ju, Z. Chen, Y. Li, and Y. Feng, "Convolutional neural network with spatio-temporal-channel attention for remote heart rate estimation," *The Visual Computer*, pp. 1–19, 2022.

[91] H. Gao, X. Wu, J. Geng, and Y. Lv, "Remote heart rate estimation by signal quality attention network," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 2122–2129.

[92] R. Song, S. Zhang, C. Li, Y. Zhang, J. Cheng, and X. Chen, "Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7411–7421, 2020.

[93] C. M. van Ravenswaaij-Arts, L. A. Kollee, J. C. Hopman, G. B. Stoelinga, and H. P. van Geijn, "Heart rate variability," *Annals of internal medicine*, vol. 118, no. 6, pp. 436–447, 1993.

[94] A. P. Owens, "The role of heart rate variability in the future of remote digital biomarkers," *Frontiers in neuroscience*, vol. 14, p. 582 145, 2020.

[95] W.-H. Lin, D. Wu, C. Li, H. Zhang, and Y.-T. Zhang, "Comparison of heart rate variability from ppg with that from ecg," in *The International Conference on Health Informatics*, Springer, 2014, pp. 213–215.

[96] A. Pai, A. Veeraraghavan, and A. Sabharwal, "Hrvcam: Robust camera-based measurement of heart rate variability," *Journal of Biomedical Optics*, vol. 26, no. 2, pp. 022 707–022 707, 2021.

[97] M. Van Gastel, S. Stuijk, and G. de Haan, "Robust respiration detection from remote photoplethysmography," *Biomedical optics express*, vol. 7, no. 12, pp. 4941–4957, 2016.

[98] K. Selvakumar, E. V. Kumar, M. Sailesh, *et al.*, "Realtime ppg based respiration rate estimation for remote health monitoring applications," *Biomedical Signal Processing and Control*, vol. 77, p. 103 746, 2022.

[99] D. Shao, Y. Yang, C. Liu, F. Tsow, H. Yu, and N. Tao, "Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2760–2767, 2014.

[100] J. Ruminski, A. Bujnowski, K. Czuszynski, and T. Kocejko, "Estimation of respiration rate using an accelerometer and thermal camera in eglasses," in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Ieee, 2016, pp. 1431–1434.

[101] F. P. Wieringa, F. Mastik, and A. F. van der Steen, "Contactless multiple wavelength photoplethysmographic imaging: A first step toward "spo 2 camera" technology," *Annals of biomedical engineering*, vol. 33, no. 8, pp. 1034–1041, 2005.

[102] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiological measurement*, vol. 35, no. 5, p. 807, 2014.

[103] R. P. Smith, J. Argod, J.-L. Pépin, and P. A. Lévy, "Pulse transit time: An appraisal of potential clinical applications," *Thorax*, vol. 54, no. 5, pp. 452–457, 1999.

[104] A. Moço, S. Stuijk, M. van Gastel, and G. de Haan, "Impairing factors in remote-ppg pulse transit time measurements on the face," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1358–1366.

[105] I. C. Jeong and J. Finkelstein, "Introducing contactless blood pressure assessment using a high speed video camera," *Journal of medical systems*, vol. 40, pp. 1–10, 2016.

[106] P.-W. Huang, C.-H. Lin, M.-L. Chung, T.-M. Lin, and B.-F. Wu, "Image based contactless blood pressure assessment using pulse transit time," in *2017 International Automatic Control Conference (CACS)*, Ieee, 2017, pp. 1–6.

[107] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A neural network-based method for continuous blood pressure estimation from a ppg signal," in *2013 IEEE International instrumentation and measurement technology conference (I2MTC)*, Ieee, 2013, pp. 280–283.

[108] F. Schrumpf, P. Frenzel, C. Aust, G. Osterhoff, and M. Fuchs, "Assessment of non-invasive blood pressure prediction from ppg and rppg signals using deep learning," *Sensors*, vol. 21, no. 18, p. 6022, 2021.

[109] H. Luo, D. Yang, A. Barszczyk, *et al.*, "Smartphone-based blood pressure measurement using transdermal optical imaging technology," *Circulation: Cardiovascular Imaging*, vol. 12, no. 8, e008857, 2019.

[110] S. Kılıçkaya, A. Güner, and B. Dal, "Comparison of different machine learning techniques for the cuffless estimation of blood pressure using ppg signals," in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Ieee, 2020, pp. 1–6.

[111] M. Coutrot, E. Dudoignon, J. Joachim, E. Gayat, F. Vallee, and F. Depret, "Perfusion index: Physical principles, physiological meanings and clinical implications in anaesthesia and critical care," *Anaesthesia Critical Care & Pain Medicine*, vol. 40, no. 6, p. 100 964, 2021.

[112] B. Kossack, E. Wisotzky, P. Eisert, S. P. Schraven, B. Globke, and A. Hilsmann, "Perfusion assessment via local remote photoplethysmography (rppg)," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2192–2201.

[113] C. Zhang, E. Jovanov, H. Liao, *et al.*, "Video based cocktail causal container for blood pressure classification and blood glucose prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 2, pp. 1118–1128, 2022.

[114] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and biological engineering and computing*, vol. 42, no. 3, pp. 419–427, 2004.

[115] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.

[116] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: A review," in *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, Ieee, 2011, pp. 410–415.

[117] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device: In laboratory and real life," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 1185–1193.

[118] M. Zanetti, L. Faes, M. De Cecco, *et al.*, "Assessment of mental stress through the analysis of physiological signals acquired from wearable devices," in *Ambient Assisted Living: Italian Forum 2018 9*, Springer, 2019, pp. 243–256.

[119] I. Mohino-Herranz, R. Gil-Pita, J. Ferreira, M. Rosa-Zurera, and F. Seoane, "Assessment of mental, emotional and physical stress through analysis of physiological signals using smartphones," *Sensors*, vol. 15, no. 10, pp. 25 607–25 627, 2015.

[120] S. Rafiqi, C. Wangwiwattana, E. Fernandez, S. Nair, and E. Larson, "Work-in-progress, pupilware-m: Cognitive load estimation using unmodified smartphone cameras," in *2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems*, Ieee, 2015, pp. 645–650.

[121] J. F. Cavanagh, T. V. Wiecki, A. Kochar, and M. J. Frank, "Eye tracking and pupillometry are indicators of dissociable latent decision processes.," *Journal of Experimental Psychology: General*, vol. 143, no. 4, p. 1476, 2014.

[122] P. Ren, A. Barreto, J. Huang, Y. Gao, F. R. Ortega, and M. Adjouadi, "Off-line and on-line stress detection through processing of the pupil diameter signal," *Annals of biomedical engineering*, vol. 42, no. 1, pp. 162–176, 2014.

[123] Y.-M. Jang, R. Mallipeddi, S. Lee, H.-W. Kwak, and M. Lee, "Human intention recognition based on eyeball movement pattern and pupil size variation," *Neurocomputing*, vol. 128, pp. 421–432, 2014.

[124] G. Chernyshov, K. Ragozin, B. Tag, and K. Kunze, "Eog glasses: An eyewear platform for cognitive and social interaction assessments in the wild," in *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2019, pp. 1–5.

[125] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Ieee, 2014, pp. 2957–2960.

[126] B. Kaur, S. Moses, M. Luthra, and V. N. Ikonomidou, "Remote stress detection using a visible spectrum camera," in *Independent Component Analyses, Compressive Sampling, Large Data Analyses (LDA), Neural Networks, Biosystems, and Nanoengineering XIII*, International Society for Optics and Photonics, vol. 9496, 2015, p. 949 602.

[127] B.-F. Wu, C.-H. Lin, P.-W. Huang, T.-M. Lin, and M.-L. Chung, "A contactless sport training monitor based on facial expression and remote-ppg," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Ieee, 2017, pp. 846–851.

[128] L. Tan, M. Hakamata, C. Cao, K. Kagawa, N. Tsumura, and S. Kawahito, "A real-time driver monitoring system using a high sensitivity camera," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXVI*, International Society for Optics and Photonics, vol. 10883, 2019, 108831g.

[129] Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere, "Cognitive heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–20, 2017.

[130] Y. Cho, N. Bianchi-Berthouze, and S. J. Julier, "Deepbreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, Ieee, 2017, pp. 456–463.

[131] S. Iyer, L. Zhao, M. P. Mohan, *et al.*, "Mm-wave radar-based vital signs monitoring and arrhythmia detection using machine learning," *Sensors*, vol. 22, no. 9, p. 3106, 2022.

[132] B. P. Yan, W. H. Lai, C. K. Chan, *et al.*, "Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals," *Journal of the American Heart Association*, vol. 7, no. 8, e008585, 2018.

[133] B. P. Yan, W. H. Lai, C. K. Chan, *et al.*, "High-throughput, contact-free detection of atrial fibrillation from video with deep learning," *JAMA cardiology*, vol. 5, no. 1, pp. 105–107, 2020.

[134] B. Sadeh, I. Merdler, S. Sadon, *et al.*, "A novel contact-free atrial fibrillation monitor: A pilot study," *European Heart Journal-Digital Health*, vol. 3, no. 1, pp. 105–113, 2022.

[135] Y. Sun, Y.-Y. Yang, B.-J. Wu, *et al.*, "Contactless facial video recording with deep learning models for the detection of atrial fibrillation," *Scientific reports*, vol. 12, no. 1, p. 281, 2022.

[136] W. Massagram, V. M. Lubecke, A. HØst-Madsen, and O. Boric-Lubecke, "Assessment of heart rate variability and respiratory sinus arrhythmia via doppler radar," *IEEE Transactions on microwave theory and techniques*, vol. 57, no. 10, pp. 2542–2549, 2009.

[137] N. Du, K. Liu, L. Ge, and J. Zhang, "Apnearadar: A 24ghz radar-based contactless sleep apnea detection system," in *2017 2nd International Conference on Frontiers of Sensors Technologies (ICFST)*, Ieee, 2017, pp. 372–376.

[138] T. Koda, T. Sakamoto, S. Okumura, H. Taki, S. Hamada, and K. Chin, "Radar-based automatic detection of sleep apnea using support vector machine," in *2020 International Symposium on Antennas and Propagation (ISAP)*, Ieee, 2021, pp. 841–842.

[139] Y. Zhou, D. Shu, H. Xu, *et al.*, "Validation of novel automatic ultra-wideband radar for sleep apnea detection," *Journal of thoracic disease*, vol. 12, no. 4, p. 1286, 2020.

[140] W. Xia, Y. Li, and S. Dong, "Radar-based high-accuracy cardiac activity sensing," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[141] V. P. Tran, A. A. Al-Jumaily, and S. M. S. Islam, "Doppler radar-based non-contact health monitoring for obstructive sleep apnea diagnosis: A comprehensive review," *big data and cognitive computing*, vol. 3, no. 1, p. 3, 2019.

[142] J. Porter. "Amazon reportedly developing radar-equipped sleep apnea tracker." (2021), [Online]. Available: `https://www.theverge.com/2021/1/6/22216648/amazon-sleep-tracking-alexa-brahms-apnea-radar-device`.

[143] S. Schellenberger, K. Shi, T. Steigleder, *et al.*, "A dataset of clinically recorded radar vital signs with synchronised reference sensor signals," *Scientific data*, vol. 7, no. 1, p. 291, 2020.

[144] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, 2018.

[145] P. Li and T.-M. Laleg-Kirati, "Central blood pressure estimation from distal ppg measurement using semiclassical signal analysis features," *IEEE Access*, vol. 9, pp. 44 963–44 973, 2021.

[146] D. N. Dutt and S. Shruthi, "Digital processing of ecg and ppg signals for study of arterial parameters for cardiovascular risk assessment," in *2015 International conference on communications and signal processing (ICCSP)*, Ieee, 2015, pp. 1506–1510.

[147] X. Ding and Y.-T. Zhang, "Pulse transit time technique for cuffless unobtrusive blood pressure measurement: From theory to algorithm," *Biomedical engineering letters*, vol. 9, pp. 37–52, 2019.

[148] A. V. Moço, S. Stuijk, and G. de Haan, "Skin inhomogeneity as a source of error in remote ppg-imaging," *Biomedical optics express*, vol. 7, no. 11, pp. 4718–4733, 2016.

[149] G. Slapničar, W. Wang, and M. Luštrek, "Classification of hemodynamics scenarios from a public radar dataset using a deep learning approach," *Sensors*, vol. 21, no. 5, p. 1836, 2021. DOI: `10.3390/s21051836`.

[150] J. A. Dempsey, S. C. Veasey, B. J. Morgan, and C. P. O'Donnell, "Pathophysiology of sleep apnea," *Physiological reviews*, vol. 90, no. 1, pp. 47–112, 2010.

[151] X.-W. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE access*, vol. 2, pp. 514–525, 2014.

[152] L. Beyer, O. J. Hénaff, A. Kolesnikov, X. Zhai, and A. v. d. Oord, "Are we done with imagenet?" *arXiv preprint arXiv:2006.07159*, 2020.

[153] F. Michler, K. Shi, S. Schellenberger, *et al.*, "A clinically evaluated interferometric continuous-wave radar system for the contactless measurement of human vital parameters," *Sensors*, vol. 19, no. 11, p. 2492, 2019.

[154] C. Will, K. Shi, S. Schellenberger, *et al.*, "Radar-based heart sound detection," *Scientific reports*, vol. 8, no. 1, p. 11 551, 2018.

[155] J. Fortin, W. Marte, R. Grüllenberger, *et al.*, "Continuous non-invasive blood pressure monitoring using concentrically interlocking control loops," *Computers in biology and medicine*, vol. 36, no. 9, pp. 941–957, 2006.

[156] L. Pstras, K. Thomaseth, J. Waniewski, I. Balzani, and F. Bellavere, "The valsalva manoeuvre: Physiology and clinical examples," *Acta physiologica*, vol. 217, no. 2, pp. 103–119, 2016.

[157] O. Kwon, J. Jeong, H. B. Kim, *et al.*, "Electrocardiogram sampling frequency range acceptable for heart rate variability analysis," *Healthcare informatics research*, vol. 24, no. 3, p. 198, 2018.

[158] "Resample uniform of non-uniform data to new fixed rate - matlab." (2023), [Online]. Available: `https://www.mathworks.com/help/signal/ref/resample.html` (visited on 11/21/2023).

[159] N. V. Thakor, J. G. Webster, and W. J. Tompkins, "Estimation of qrs complex power spectra for design of a qrs filter," *IEEE Transactions on biomedical engineering*, no. 11, pp. 702–706, 1984.

[160] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[161] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *Plos one*, vol. 16, no. 7, e0254841, 2021.

[162] T. T. Um, F. M. Pfister, D. Pichler, *et al.*, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," in *Proceedings of the 19th ACM international conference on multimodal interaction*, 2017, pp. 216–220.

[163] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20 200 209, 2021.

[164] A. Borovykh, C. W. Oosterlee, and S. M. Bohté, "Generalization in fully-connected neural networks for time series forecasting," *Journal of Computational Science*, vol. 36, p. 101 020, 2019.

[165] L. Sadouk, "CNN approaches for time series classification," *Time series analysis-data, methods, and applications*, vol. 5, 2019.

[166] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "Lstm fully convolutional networks for time series classification," *IEEE access*, vol. 6, pp. 1662–1669, 2017.

[167] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Lstm-based encoder-decoder for multi-sensor anomaly detection," *arXiv preprint arXiv:1607.00148*, 2016.

[168] Q. Wen, T. Zhou, C. Zhang, *et al.*, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125*, 2022.

[169] N. Hatami, Y. Gavet, and J. Debayle, "Classification of time-series images using deep convolutional neural networks," in *Tenth international conference on machine vision (ICMV 2017)*, Spie, vol. 10696, 2018, pp. 242–249.

[170] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5353–5360.

[171] P. Probst, A.-L. Boulesteix, and B. Bischl, "Tunability: Importance of hyperparameters of machine learning algorithms," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 1934–1965, 2019.

[172] L. Moreno, D. Blanco, M. L. Muñoz, and S. Garrido, "L1–l2-norm comparison in global localization of mobile robots," *Robotics and Autonomous Systems*, vol. 59, no. 9, pp. 597–610, 2011.

[173] A. Devarakonda, M. Naumov, and M. Garland, "Adabatch: Adaptive batch sizes for training deep neural networks," *arXiv preprint arXiv:1712.02029*, 2017.

[174] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization.," *Journal of machine learning research*, vol. 13, no. 2, 2012.

[175] P. D. Berger, R. E. Maurer, and G. B. Celli, "Experimental design," *CA (USA): Wadsworth Group Belmont*, 2002.

[176] W. Wang, A. C. Den Brinker, and G. De Haan, "Single-element remote-ppg," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 7, pp. 2032–2043, 2018.

[177] A. De Greeff, J. Arora, S. Hervey, B. Liu, and A. H. Shennan, "Accuracy assessment of the tensoval duo control according to the british and european hypertension societies' standards," *Blood Pressure Monitoring*, vol. 13, no. 2, pp. 111–116, 2008.

[178] G. Slapničar, W. Wang, and M. Luštrek, "Feasibility of remote pulse transit time estimation using narrow-band multi-wavelength camera photoplethysmography," in *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2022, pp. 1–5. DOI: 10.1109/bhi56158.2022.9926828.

[179] J. Liu, B. P. Yan, Y.-T. Zhang, X.-R. Ding, P. Su, and N. Zhao, "Multi-wavelength photoplethysmography enabling continuous blood pressure measurement with compact wearable electronics," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 6, pp. 1514–1525, 2018.

[180] G. Slapničar, W. Wang, and M. Luštrek, "Feasibility of remote pulse transit time estimation using narrow-band multi-wavelength camera photoplethysmography," in *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '22 Adjunct, Association for Computing Machinery, 2023, pp. 115–117. DOI: 10.1145/3544793.3560339.

[181] C. Ash, M. Dubec, K. Donne, and T. Bashford, "Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods," *Lasers in medical science*, vol. 32, pp. 1909–1918, 2017.

[182] W. Commons. "Bayer pattern on sensor profile." (2006), [Online]. Available: https://commons.wikimedia.org/wiki/File:Bayer%5C%5Fpattern%5C%5Fon%5C%5Fsensor%5C%5Fprofile.svg.

[183] J. Park, H. S. Seok, S.-S. Kim, and H. Shin, "Photoplethysmogram analysis and applications: An integrative review," *Frontiers in Physiology*, vol. 12, p. 808 451, 2022.

[184] X. He, R. A. Goubran, and X. P. Liu, "Secondary peak detection of ppg signal for continuous cuffless arterial blood pressure measurement," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 6, pp. 1431–1439, 2014.

[185] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, "Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions," *PloS one*, vol. 8, no. 10, e76585, 2013.

[186] D. Han, S. K. Bashar, J. Lázaro, *et al.*, "A real-time ppg peak detection method for accurate determination of heart rate during sinus rhythm and cardiac arrhythmia," *Biosensors*, vol. 12, no. 2, p. 82, 2022.

[187] G. Slapničar, W. Wang, and M. Luštrek, "Feasibility of remote blood pressure estimation via narrow-band multi-wavelength pulse transit time," *ACM Transactions on Sensor Networks*, 2023. DOI: 10.1145/3597302.

[188] C. Serviere and P. Fabry, "Principal component analysis and blind source separation of modulated sources for electro-mechanical systems diagnostic," *Mechanical systems and signal processing*, vol. 19, no. 6, pp. 1293–1311, 2005.

[189] R. M. Berne, "Cardiovascular physiology," *Annual Review of Physiology*, vol. 43, no. 1, pp. 357–358, 1981.

[190] V. V. Tuchin, A. Bashkatov, E. Genina, *et al.*, "Finger tissue model and blood perfused skin tissue phantom," in *Dynamics and Fluctuations in Biomedical Photonics VIII*, Spie, vol. 7898, 2011, pp. 169–179.

[191] A. L. Buczak and J. J. Barrett, "Genetic algorithm method for determining temperature profiles," in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98TH8360)*, Ieee, 1998, pp. 142–147.

[192] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. J. Wang, "Video-based human heart rate measurement using joint blind source separation," *Biomedical Signal Processing and Control*, vol. 31, pp. 309–320, 2017.

[193] G. Slapničar, W. Wang, and M. Luštrek, "Generalized channel separation algorithms for accurate camera-based multi-wavelength PTT and BP estimation," *Biomed. Opt. Express*, vol. 15, no. 5, pp. 3128–3146, 2024. DOI: 10.1364/boe.518562.

[194] G. Louppe, "Understanding random forests: From theory to practice," *arXiv preprint arXiv:1407.7502*, 2014.

[195] L. Bottou, C.-J. Lin, *et al.*, "Support vector machine solvers," *Large scale kernel machines*, vol. 3, no. 1, pp. 301–320, 2007.

[196] A.-E. Stenehjem and I. Os, "Reproducibility of blood pressure variability, white-coat effect and dipping pattern in untreated, uncomplicated and newly diagnosed essential hypertension," *Blood pressure*, vol. 13, no. 4, pp. 214–224, 2004.

[197] D. Giavarina, "Understanding bland altman analysis," *Biochemia medica*, vol. 25, no. 2, pp. 141–151, 2015.

[198] G. Slapničar, E. Dovgan, P. čuk, and M. Luštrek, "Contact-free monitoring of physiological parameters in people with profound intellectual and multiple disabilities," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2019.

[199] E. Dovgan, J. Valič, G. Slapničar, and M. Luštrek, "Recognition of behaviour patterns for people with profound intellectual and multiple disabilities," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '21 Adjunct, Association for Computing Machinery, 2021, pp. 523–527. DOI: 10.1145/3460418.3479370.

[200] V. Song. "The amazon halo rise would be a neat sleep tracker, if it weren't for my cat." (2023), [Online]. Available: https://www.theverge.com/23571221/amazon-halo-rise-review-sleep-tracking-smart-lamp.

[201] Sleepiz. "Explore our solutions - sleepiz." (2024), [Online]. Available: `https://sleepiz.com/solutions/`.

[202] A. H. Association. "Types of blood pressure medications." (2024), [Online]. Available: `https://www.heart.org/en/health-topics/high-blood-pressure/changes-you-can-make-to-manage-high-blood-pressure/types-of-blood-pressure-medications`.

[203] B. J. North and D. A. Sinclair, "The intersection between aging and cardiovascular disease," *Circulation research*, vol. 110, no. 8, pp. 1097–1108, 2012.

[204] X. Liu, T. Xiao, S. Si, Q. Cao, S. Kumar, and C.-J. Hsieh, "How does noise help robustness? explanation and exploration under the neural sde framework," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 282–290.

[205] Q. Zhu, X. Tian, C.-W. Wong, and M. Wu, "Learning your heart actions from pulse: Ecg waveform reconstruction from ppg," *IEEE Internet of Things Journal*, vol. 8, no. 23, pp. 16 734–16 748, 2021.

[206] J. P. Devadhasan, H. Oh, C. S. Choi, and S. Kim, "Whole blood glucose analysis based on smartphone camera module," *Journal of biomedical optics*, vol. 20, no. 11, pp. 117 001–117 001, 2015.

[207] Y. Wang, W. Wang, M. van Gastel, and G. de Haan, "Modeling on the feasibility of camera-based blood glucose measurement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.

[208] D. Biswas, L. Everson, M. Liu, *et al.*, "Cornet: Deep learning framework for ppg-based heart rate estimation and biometric identification in ambulant environment," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 2, pp. 282–291, 2019.

[209] V. Jindal, J. Birjandtalab, M. B. Pouyan, and M. Nourani, "An adaptive deep learning approach for ppg-based identification," in *2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, Ieee, 2016, pp. 6401–6404.

[210] Y. Ye, G. Xiong, Z. Wan, T. Pan, and Z. Huang, "Ppg-based biometric identification: Discovering and identifying a new user," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Ieee, 2021, pp. 1145–1148.

Bibliography

Publications Related to the Thesis

Journal Articles

G. Slapničar, W. Wang, and M. Luštrek, "Classification of hemodynamics scenarios from a public radar dataset using a deep learning approach," *Sensors*, vol. 21, no. 5, p. 1836, 2021. DOI: 10.3390/s21051836.

G. Slapničar, W. Wang, and M. Luštrek, "Feasibility of remote blood pressure estimation via narrow-band multi-wavelength pulse transit time," *ACM Transactions on Sensor Networks*, 2023. DOI: 10.1145/3597302.

G. Slapničar, W. Wang, and M. Luštrek, "Generalized channel separation algorithms for accurate camera-based multi-wavelength PTT and BP estimation," *Biomed. Opt. Express*, vol. 15, no. 5, pp. 3128–3146, 2024. DOI: 10.1364/boe.518562.

Conference Papers

G. Slapničar, E. Dovgan, P. čuk, and M. Luštrek, "Contact-free monitoring of physiological parameters in people with profound intellectual and multiple disabilities," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2019.

E. Dovgan, J. Valič, G. Slapničar, and M. Luštrek, "Recognition of behaviour patterns for people with profound intellectual and multiple disabilities," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '21 Adjunct, Association for Computing Machinery, 2021, pp. 523–527. DOI: 10.1145/3460418.3479370.

G. Slapničar, W. Wang, and M. Luštrek, "Feasibility of remote pulse transit time estimation using narrow-band multi-wavelength camera photoplethysmography," in *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '22 Adjunct, Association for Computing Machinery, 2023, pp. 115–117. DOI: 10.1145/3544793.3560339.

G. Slapničar, W. Wang, and M. Luštrek, "Feasibility of remote pulse transit time estimation using narrow-band multi-wavelength camera photoplethysmography," in *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2022, pp. 1–5. DOI: 10.1109/bhi56158.2022.9926828.

Biography

Gašper Slapničar finished his bachelor's and master's degree at the Faculty of Computer and Information Science, University of Ljubljana. His master's thesis was on blood pressure estimation from photoplethysmography using neural networks, which resulted in a publication in the Sensors journal, receiving paper of the month and highly-cited paper designations. He continued his studies as a young researcher at the Jožef Stefan Institute (JSI) and Jožef Stefan International Postgraduate School (IPS) under the supervision of Prof. Dr. Mitja Luštrek and As. Prof. Dr. Wenjin Wang. In his PhD research, he focused on contact-free sensing of physiological parameters. He finished his IPS studies with a perfect GPA of 10.0.

During his studies he published several papers in peer-reviewed SCI journals and participated as a full-paper, short-paper and poster presenter in several leading conferences and workshops in his field, such as International Conference on Computer Vision (ICCV), IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) and UbiComp.

He also participated and continues to participate in cutting-edge international research projects as a member of the Department of Intelligent Systems (E9, JSI), dealing mainly with physiological signal analysis and deep learning. Additionally he participated as an outside advisor in the work of master's student Jie Su at The Eindhoven University of Technology, which resulted in a peer-reviewed SCI journal publication. Since 2022, he has also been participating as a program committee member in the Computer Vision for Physiological Measurement workshop, which is a leading workshop in the field of contact-free physiological sensing and is yearly organised as a part of ICCV or Computer Vision and Pattern Recognition (CVPR).