Non-invasive Continuous Blood Pressure Measurement using Machine Learning

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Informatik und Angewandte Kognitionswissenschaften
der Universität Duisburg-Essen

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

genehmigte Dissertation

von

Carolin Wuerich

aus

Stuttgart

Datum der Einreichung: 14.06.2023
Referent: Prof. Dr. rer. nat. Gregor Schiele
Koreferent: Prof. Dr. rer. nat. Anton Grabmaier
Tag der mündlichen Prüfung: 20.11.2023

DOI: 10.17185/duepublico/81920
URN: urn:nbn:de:hbz:465-20240617-125011-7

Acknowledgements

I would like to express my gratitude and appreciation to all those who have supported me throughout the journey of completing this dissertation.

First and foremost, I am thankful to Prof. Dr. Gregor Schiele, my supervising professor, for his invaluable guidance and scholarly expertise. His insightful feedback has been instrumental in shaping the quality of this research. My gratitude also goes to my supervisor Dr.-Ing. Christian Wiede for his mentorship and expertise. His valuable insights and feedback have helped me navigate the field and contributed to the development of this work.

I would also like to extend my appreciation to my second supervisor, Prof. Dr. Grabmaier, and the members of my dissertation committee, Prof. Dr. Jens Krüger and Prof. Dr. Josef Pauli, for their time, expertise, and constructive criticism. Their valuable input and rigorous examination have contributed to the refinement of this dissertation.

My acknowledgement goes to the staff and colleagues at Fraunhofer IMS, who provided the research environment and access to essential resources. I am grateful for the opportunities I had to attend conferences and seminars that enriched my knowledge and allowed me to connect with other researchers in my field.

Moreover, I would like to express my sincere appreciation to Felix, Belmin, Gongbo, Lukas, Lars and Johannes, my colleagues and friends who embarked on the path of completing their dissertations alongside me. Your shared experiences, encouragement and camaraderie have made this process more manageable. The gratitude also goes to the rest of my outstanding team, including Wolfgang, Maarten, Dennis, Niels, Hüseyin, Diana, Burkhard and Cornelia. I am especially thankful to Luise and Belmin for revising my thesis and offering valuable feedback.

Next, I would also like to thank all the students - Robin, Xiaowen, Philip, Eva-Maria, Judith, Simon, Kira, Roswitha, Fang, Omar, Navraj und Kusay - whom I had the privilege of supervising for their theses during the course of my own dissertation.

Furthermore, I am deeply grateful to my family and friends for their love, understanding and support throughout this demanding journey. Your belief in me and your continuous encouragement have been a driving force behind my perseverance.

Lastly, I extend my appreciation to all the participants who generously contributed their time for this study. Their cooperation and willingness to share their data have been crucial in providing a robust foundation for this research.

Although it is impossible to mention everyone who has contributed to my professional and personal growth, please know that your support has not gone unnoticed and unappreciated. To all those mentioned above and countless others who have played a role, however small, in this dissertation, completing this dissertation would not have been possible without your support and encouragement. Thank you.

Abstract

Blood pressure is an important clinical parameter that is used to assess cardiovascular health. Deviations from the normal range can lead to severe organ damage and increased mortality, while sudden changes can indicate cardiovascular events. Therefore, a regular ambulatory blood pressure measurement as well as a reliable monitoring in clinical settings are essential for an accurate diagnosis and treatment of such cardiovascular diseases and events. However, the commonly used cuff-based measurement devices exhibit several limitations regarding accuracy, temporal resolution and measurement comfort.

Therefore, this thesis investigates new methods for camera-based beat-to-beat blood pressure monitoring. Such methods rely on the extraction of remote Photoplethysmogram (rPPG) signals from skin pixels of the video. To overcome illumination and movement artifacts as well as an insufficient Signal-to-Noise Ratio (SNR) for darker skin tones, various colour model transformations and rPPG extraction algorithms are evaluated with respect to the application for remote blood pressure measurement. The experiments show that Plane Orthogonal to Skin (POS) performs best under difficult measurement conditions and that considering skin pixels of the palm as opposed to the face significantly improves SNR for darker skin.

Moreover, a method based on hand-crafted rPPG features and a Random Forest Regression (RFR) model is proposed. To obtain a light-weight model and increase prediction accuracy, a Sequential Forward Selection (SFS) is performed. The prediction accuracy could be improved to an Mean Absolute Error (MAE) $\pm$Standard Deviation (SD) of 11.91 $\pm$9.66 mmHg for Systolic Blood Pressure (SBP) and 7.92 $\pm$6.02 mmHg for Diastolic Blood Pressure (DBP) on a wide range of blood pressure values outperforming comparable studies. An analysis of the feature selection results is provided to enhance model interpretability for

medical applications and aid future developments.

Next, based on the ResNet-50 architecture, a more complex Convolutional Neural Network (CNN) is developed to automatically extract features from the raw rPPG signals. Due to the small size of the rPPG data set, the model is developed and pre-trained on the MIMIC III waveform data base. This Photoplethysmogram (PPG)-based beat-to-beat prediction method reaches an MAE $\pm$SD of 8.73 $\pm$7.36 mmHg for SBP and 8.07 $\pm$6.86 mmHg for DBP which is comparable to related studies that rely on longer PPG sections or an additional ECG signal.
Further analyses underline the potential of model personalisation and the importance of a balanced fine-tuning data set, since the results of personalisation strongly depend on the selected tuning data and is prone to overfitting when using sequential tuning samples. Therefore, different strategies are derived for balancing the tuning data set in real-world applications.
Finally, transfer-learning from the PPG domain to the rPPG domain is assessed and shows encouraging results on the rPPG data set with prediction errors close to the feature-based method.

Zusammenfassung

Der Blutdruck ist ein wichtiger klinischer Vitalparameter, der häufig zur Beurteilung der kardiovaskulären Gesundheit herangezogen wird. Starke Abweichungen vom Normalbereich führen zu schweren Organschäden und erhöhter Sterblichkeit, und plötzliche Schwankungen können auf kardiovaskuläre Ereignisse hinweisen. Daher sind eine regelmäßige ambulante Blutdruckmessung sowie eine zuverlässige Überwachung im klinischen Umfeld für eine genaue Diagnose und Behandlung solcher Herz-Kreislauf-Erkrankungen unerlässlich. Die gemeinhin verwendeten, auf Blutdruckmanschetten basierenden Messsysteme weisen jedoch einige Einschränkungen hinsichtlich Genauigkeit, zeitlicher Auflösung und Messkomfort auf.

Daher werden in dieser Arbeit neue Methoden zur kamerabasierten Beat-to-Beat-Blutdruckmessung untersucht. Diese Methoden beruhen auf der Ableitung von kontaktlosen Photoplethysmogramm-Signalen (rPPG) aus Hautpixeln des Videos. Um Beleuchtungs- und Bewegungsartefakte zu reduzieren sowie ein unzureichendes Signal-zu-Rausch-Verhältnis für dunklere Hauttöne zu beheben, werden verschiedene Farbmodelltransformationen und rPPG-Extraktionsalgorithmen hinsichtlich der Anwendung zur kontaktlosen Blutdruckmessung evaluiert. Die Experimente zeigen, dass POS unter schwierigen Messbedingungen am besten abschneidet und dass die Berücksichtigung von Hautpixeln der Handfläche im Gegensatz zu denen des Gesichts das SNR für dunklere Haut deutlich verbessert.

Außerdem wird eine Methode entwickelt, die auf manuell definierten rPPG-Merkmalen und einem Random-Forest-Modell basiert. Um ein leichtgewichtiges Modell zu erhalten und die Vorhersagegenauigkeit zu erhöhen, wird ein sequentielle Merkmalsselektion durchgeführt. Die Vorhersagegenauigkeit konnte auf einen MAE ±SD von 11,91 ±9,66 mmHg für systolischen Blutdruck und 7,92 ±6,02 mmHg für diastolischen Blutdruck verbessert werden und übertrifft damit die Ergebnisse vergleichbarer Studien. Zusätzlich wird eine Analyse der Ergebnisse

der Merkmalsauswahl durchgeführt, um die Interpretierbarkeit des Modells für medizinische Anwendungen zu verbessern und zukünftige Entwicklungen zu unterstützen.

Anschließend wird auf Grundlage der ResNet-50-Architektur ein komplexeres Faltungsnetzwerk entwickelt, um automatisch Merkmale aus den rohen rPPG-Signalen zu extrahieren. Aufgrund der geringen Größe des rPPG-Datensatzes wird das Modell auf der MIMIC-III-Signaldatenbank entwickelt und vortrainiert. Diese PPG-basierte Beat-to-Beat-Prädiktion erreicht einen MAE ±SD von 8,73 ±7,36 mmHg für systolischen Blutdruck und 8,07 ±6,86 mmHg für diastolischen Blutdruck, was vergleichbar ist mit verwandten Studien, die auf längere PPG-Abschnitte oder eine zusätzliches EKG-Signal angewiesen sind.
Weitere Analysen belegen das Potenzial der Modellpersonalisierung und die Relevanz eines ausgeglichenen Finetuning-Datensatzes: So sind die Ergebnisse der Personalisierung stark von den ausgewählten Tuning-Daten abhängig und bei der Verwendung sequenzieller Tuning-Samples zeigt sich eine hohe Anfälligkeit für eine Überanpassung des Modells. Daher werden verschiedene Strategien für einen ausgeglichenen Finetuning-Datensatz in realen Anwendungen abgeleitet.
Eine abschließende Bewertung des Transfer-Lernens von der PPG-Domäne auf die rPPG-Domäne zeigt vielversprechende Ergebnisse auf dem rPPG-Datensatz mit Vorhersagefehlern, die nahe an denen der merkmalsbasierten Methode liegen.

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

AAMI	US Association for the Advancement of Medical Instrumentation
ABP	Arterial Blood Pressure
AC	Alternating Current
ADC	Analog Digital Converter
ANN	Artificial Neural Network
BMI	Body Mass Index
BP	Blood Pressure
BSH	British Society of Hypertension
BSS	Blind Source Separation
CCD	Charge-Coupled Device
CIE	Commission Internationale de l'Éclairage (engl. International Commission on Illumination)
CMOS	Complementary Metal-Oxide-Semiconductor
CMYK	Cyan Magenta Yellow Key
CNN	Convolutional Neural Network
CSRT	Channel and Spatial Reliability Tracker
DANN	Domain-Adversarial Neural Network
DBP	Diastolic Blood Pressure
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
ESH	European Society of Hypertension
dPTT	differential Pulse Transit Time
FFT	Fast Fourier Transformation
FIR	Finite Impulse Response
HOG	Histogram of Oriented Gradients
ICA	Independent Component Analysis
ICU	Intensive Care Unit

IoU	Intersection over Union
KCF	Kernelized Correlation Filter
LED	Light Emitting Diode
LOSO CV	Leave-One-Subject-Out Cross Validation
LSI	Laser Speckle Imaging
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAP	Mean Arterial Pressure
ME	Mean Error
MIL	Multiple Instance Learning
MLP	Multi-Layer Perceptron
MLR	Multiple Linear Regression
MOSSE	Minimum Output Sum of Squared Error
NIR	Near Infrared
NMS	Non-Maximum Suppression
PAT	Pulse Arrival Time
PC	Principal Component
PCA	Principal Component Analysis
PEP	Pre-Ejection Periode
POS	Plane Orthogonal to Skin
PP	Pulse Pressure
PTT	Pulse Transit Time
PPG	Photoplethysmogram
PWV	Pulse Wave Velocity
RFR	Random Forest Regression
RGB	Red Green Blue
RMSE	Root Mean Squared Error
ROI	Region of Interest
rPPG	remote Photoplethysmogram
SBP	Systolic Blood Pressure
SD	Standard Deviation
SFS	Sequential Forward Selection
SM	Smooth-Muscle
SNR	Signal-to-Noise Ratio
SPG	Speckleplethysmogram
SQI	Signal Quality Index
SSD	Single-Shot Detection
SSR	Spatial Subspace Rotation
YOLO	You Only Look Once

Symbols

a	learning rate
A	area

b_d	diffuse reflection
b_n	quantisation noise of the image sensor
b_s	specular reflection
BB	sum of pixels belonging to the predicted bounding box
c	(RGB vector of skin pixel)
c_{dob}	de-oxygenated blood concentration
c_m	melanin concentration
c_{ob}	oxygenated blood concentration
C	compliance of vessel
D	training data sample
e_d	number of detection failures
e_t	number of tracking failures
E	combined elastic modulus of arterial walls
E_e	elastic modulus of elastin
E_{hf}	elastic modulus of arterial walls at high frequencies
E_{lf}	elastic modulus of arterial walls at low frequencies
E_N	elasticity of vessel walls in radial and circumferential directions
E_{pc}	elastic modulus of parallel collagen
E_{rZ}	shear modulus of vessel walls
E_{sc}	elastic modulus of serial collagen
E_{sm}	elastic modulus of smooth muscle
f	continuous variable of frequency
f_{acti}	activation function
f_d	dominant frequency
f_{high}	upper cut-off frequency
f_{low}	lower cut-off frequency
f_s	sampling rate
F_f	resistance force
F_N	restoring force
F_S	shear force
g	number of input channels
GT	sum of pixels belonging to the ground truth bounding box

h	wall thickness
h_{BB}	height of bounding box
h_{ROI}	height of region of interest
H	residual function
i_a	aggregated inputs to a neuron
i_l	input to the l-th residual unit
i_n	input to a neuron
I	arterial inertance
J	Jaccard coefficient
K	number of pixels in region of interest
$\hat{K}$	kernel size
k	absorbed energy during vessel dilation
l	length of signal
lr	learning rate
L	loss function
m	identity mapping
M	momentum
n	batch size
N	number of harmonics
$\hat{o}_i$	expected output of a neuron in the final layer
o_l	output of the l-th residual unit
o_m	output of a neuron
p_{PPG}	photoplethysmography signal
p_{rPPG}	remote photoplethysmography signal
P	pressure
P_a	arterial pressure
P_{map}	mean arterial pressure
Q	perceived luminance intensity
$\bar{Q}$	temporal average of luminance intensity
$\hat{Q}$	spatial average of pixel intensity
Q_0	intensity of light reflected at tissue surface
r	radial dilation of the vessel
r_0	inner radius of vessel in static condition
R	correlation coefficient

s	stride
$\tilde{S}(f)$	frequency spectrum
SNR	signal-to-noise ratio
SQI	signal quality index
t	continuous variable of time
$\bar{t}$	average processing time
t_{PTT}	pulse transit time
T_c	time of cardiac cycle
U	total number of frames
v_{PWV}	pulse wave velocity
v_r	radial velocity
V_0	non-pulsatile blood volume
w_{BB}	width of bounding box
w_{ROI}	width of region of interest
W	set of all weights in a neural network
W_l	set of all weights in a residual unit
W_n	weight of a connection in a neural network
x_{BB}	x-axis position of upper left corner of bounding box
x_{ROI}	x-axis position of upper left corner of region of interest
X	longitudinal coordinate along arterial segment
y_{BB}	y-axis position of upper left corner of bounding box
y_{ROI}	y-axis position of upper left corner of region of interest
Z_c	tube characteristic impedance
Z_l	load impedance
β	magnitude of amplitude modulation
γ	length of arterial segment
Γ	reflection coefficient
η_i	blood viscosity
η_{sm}	viscosity of smooth muscle tissue
κ	modulus of absorbed energy during vessel dilation
μ	density of arterial wall
$\hat{\mu}$	mean
θ	angular coordinate
ρ	blood density

σ	standard deviation
ω_0	fundamental vibration frequency
ω_{HR}	harmonics of heart rate
ω_n	nth harmonic

Chapter 1

Introduction

Blood Pressure (BP) is a vital biosignal of the human body and a biomarker for many diseases. High BP significantly increases the risk for strokes, heart attacks, ischaemic heart disease, dementia and chronic kidney disease [1]. With over 25% of adults suffering from hypertension (i.e., Systolic Blood Pressure (SBP) $\geqslant$ 140 mmHg or Diastolic Blood Pressure (DBP) $\geqslant$ 90 mmHg), it is a major contributor to cardiovascular diseases which are the leading cause of premature death worldwide [1]. Hypertension is also called "the silent killer" since it comes without symptoms and slowly damages organs and the cardiovascular system. According to the World Health Organisation, 46% of adults with elevated BP are unaware of their condition and only one in five affected adults have their hypertension under control [2].

Moreover, in the clinical setting, changes in BP can indicate a hypovolemic shock, hypoxemia, infections or other conditions that require immediate treatment. Therefore BP is closely monitored especially during surgeries and in the Intensive Care Unit (ICU). During these high risk situations, an invasive measurement is employed which is considered to be the gold-standard. For the invasive measurement, a catheter with a pressure sensor is introduced into the patient's artery such that BP can be monitored continuously and at various locations along the arterial tree. Despite the high accuracy and continuity of these measurement systems, they are limited to critical clinical applications since they entail a high risk for thrombosis and infections.

Hence, for common BP monitoring and ambulatory BP measurement, auscultatory and oscillometric cuff-based devices are employed instead, which come

without complications but exhibit inferior accuracy. Moreover, cuff-based systems are discontinuous with measurement durations up to a minute and cause discomfort in many patients.

Other less established methods are vascular unloading, tactile sensors and volume clamps, each with their own set of disadvantages.

1.1 Motivation

In order to overcome drawbacks of existing methods, recent research investigated surrogate-based BP measurement approaches that take other biosignals such as Photoplethysmogram (PPG) and Electrocardiogram (ECG) into account. PPG is a non-invasive optical technique that measures changes in blood volume in peripheral tissues. For BP estimation, PPG has proven to be the most practical surrogate signal due to the close physiological relationship with BP and the requirement of only a single sensor [3]. However, conventional contact-based PPG sensors apply constant pressure to ensure skin contact, which in the long term changes the blood flow in the affected tissue.

In contrast, remote Photoplethysmogram (rPPG), a contactless approach based on the same optical measurement principle, could enable a comfortable and safe beat-to-beat BP monitoring in clinical settings. Further, contactless measurement is more hygienic as no physical contact is required between patient, device and clinician, reducing the risk of cross-contamination and infection transmission. It reduces time and effort needed for BP measurement especially for patients in isolation rooms since clinicians can avoid changing into and out of protective clothing. This allows medical staff to allocate their time more efficiently to other critical tasks.

Moreover, ambulatory BP monitoring would be facilitated. Camera-based methods can be integrated into smartphone applications enabling ubiquitous BP screening without the need for assistance or a healthcare professional. Such contactless methods might encourage more regular measurements and enable better management of chronic conditions like hypertension.

Contact with healthcare professionals or the physical sensation of a tight cuff can induce anxiety in some individuals leading to elevated BP readings, a phenomenon known as white coat hypertension. By minimising the factors that

contribute to stress, contactless devices can help patients feel more at ease during the measurement process, resulting in more accurate readings and reducing the risk of misdiagnosis or overtreatment.

Hence, the development of contactless novel BP measurement techniques holds great potential for overcoming the limitations of existing methods and revolutionizing the field of clinical and ambulatory BP monitoring.

1.2 Scope of Work and Contributions

The work at hand aims at advancing such non-invasive camera-based BP measurement from rPPG signals. This includes the generation of a suitable data set for developing and testing the proposed methods, as well as analyses and optimisations of the contactless PPG measurement as the basis for subsequent signal processing. The work aims at enhancing the rPPG signal for BP prediction while considering difficult measurement conditions.

The main interest lies in the development of suitable machine learning models for estimating BP from these rPPG signals. A leight-weight feature-based model as well as a deep CNN are developed and tested on the acquired data set. Both models are further used to analyse relevant PPG waveform characteristics for BP prediction in order to increase model interpretability. Finally, fine tuning for model personalisation is investigated to compensate the influence of individual cardio-vascluar physiology.

All analyses are carried out under consideration of an application in real clinical practice and target a beat-to-beat BP measurement from a single rPPG signal. Although the focus is a camera-based method, the development of prediction models on conventional PPG signals is used as a stepping stone for the contactless approaches. Chapter 3.6 elaborates the specific research questions.

The contributions of this thesis are as follows:

- Conducting a comprehensive comparative analysis of multiple rPPG methods and colour models to refine rPPG extraction for accurate BP prediction across various measurement conditions.

- Exploring and comparing skin regions to determine the most suitable ROIs

for rPPG extraction, enhancing BP prediction accuracy across several measurement conditions.

- Developing hand-crafted features and employing feature selection techniques to create a lightweight regression model for BP prediction.
- Analysing the results of feature selection to gain deeper understanding of underlying relevant PPG characteristics.
- Creating, training, and testing a CNN for beat-to-beat BP prediction from raw PPG signals, achieving state-of- the-art results with reduced input information.
- Conducting analysis and evaluation of fine tuning to personalise the model and account for individual physiological differences, enhancing its predictive capabilities.
- Deriving several strategies to enhance the personalisation process, mitigating overfitting on the tuning dataset.
- Increasing transparency of the model's decision-making process by analysing layer activation patterns and correlations between PPG waveforms.
- Investigating and evaluating the potential of transfer learning from the PPG to rPPG domain, enabling the application of knowledge learned from one domain to enhance performance in the other domain.

1.3 Structure

Chapter 2 introduces relevant physiological and technical topics related to BP estimation from rPPG signals using machine learning methods. This is followed by Chapter 3 which outlines previous studies and related work in this field. Chapter 4 details the first approach of this thesis. It entails information about the recorded rPPG data set, the image processing pipeline with the comparison of detection and tracking algorithms as well as the development and results of the feature-based BP measurement. In Chapter 5, the analyses and selection of rPPG extraction methods and skin regions with respect to BP estimation performance are described. Next, Chapter 6 comprises the CNN-based approach which is used

for investigating model personalisation and transfer learning between PPG and rPPG models. Finally, Chapter 7 provides a summary and discussion of the thesis and its results as well as an outline of future work.

Chapter 2

Theoretical Background

In this chapter, relevant biomedical and methodical background information is provided. Generally, BP is an important physiological parameter which belongs to and strongly depends on the cardiovascular system. Hence, in Section 2.1 the cardiovascular system will be outlined with respect to its interrelations with and influences on BP. In Section 2.2, the measurement principle of photoplethysmography and related optical effects are described. This is followed by methodological background on deployed signal processing (Section 2.3), machine learning (Section 2.4) and image processing algorithms (Section 2.5).

2.1 Cardiovascular System

The cardiovascular system is a self-contained transport system which is driven by the heart. A distinction is made between the two primary circulatory loops, i.e., the pulmonary circulation loop and the systemic circulation loop [4]. The larger systemic circulation loop connects the left ventricle with the right atrium to supply the entire body (except heart and lungs) with oxygenated blood and at the same time to remove waste products from body tissues. The de-oxygenated blood is then returned to the heart where it is pumped to the pulmonary circulation. Here, de-oxygenated blood is transported through the lungs where it picks up oxygen and returns to the left side of the heart, i.e., the start of the systemic circulation. The transportation of blood is essential for the supply, detoxification and hormonal communication of all organs and cells in the body.

In the following Subsection 2.1.1, fundamentals of BP are introduced, followed

by Subsection 2.1.2 and 2.1.3 in which the impact of the cardiac system on the BP is described along the vascular tree and at the level of blood vessels, respectively.

2.1.1 Blood Pressure

The muscle contraction of the four-chambered heart results in simultaneous pumping of both sides of the heart. Each heart beat consists of the contraction phase, the systole, and the relaxation phase, the diastole. During systole, the heart pumps out the blood from the ventricles into the circulatory loops. This is followed by the diastole in which the heart fills with blood again. Due to inertia of the blood already contained in the aorta, the ejection of the cardiac blood volume produces a pressure wave which is propagated through the whole arterial system [4]. Thereby, BP is defined as the pressure that is imposed by the blood

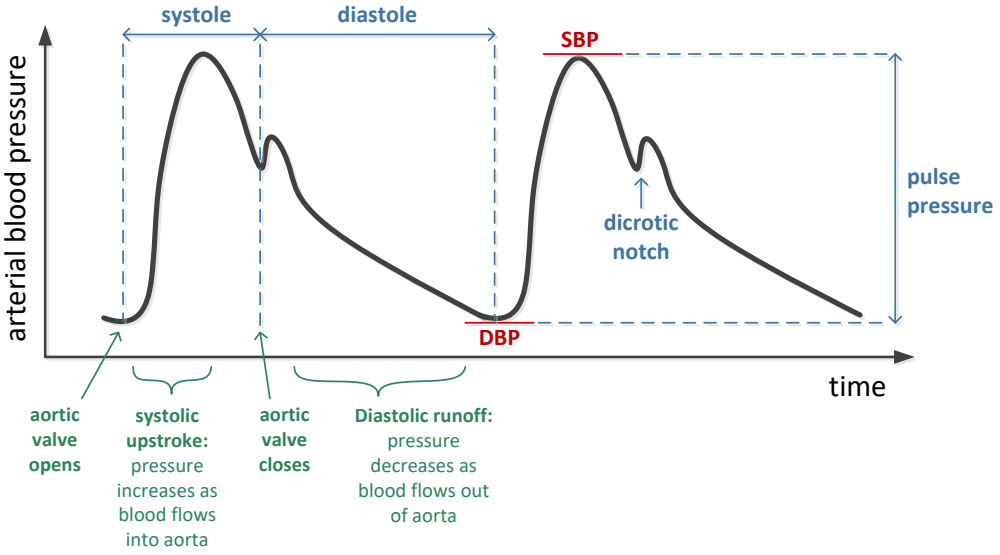


Figure 2.1: Schematic image of a BP waveform.

on the walls of the vessels [5]. Due to the pumping of the heart, the arterial pressure P_a changes periodically and thus, is typically indicated by two values as illustrated in Figure 2.1. These BP values are commonly quantified in the unit *mmHg* (millimetres of mercury). It originates from the use of mercury columns in early sphygmomanometers, where the height of the mercury column indicated the pressure in the cuff and hence, the arteries.

The SBP represents the peak value of the pressure pulse during systole while

the DBP value refers to the minimum pressure during diastole. The difference between these two values is referred to as Pulse Pressure (PP) [6]. For healthy adults, normotensive arterial pressure values lie below 120/80 mmHg, while values up to 140/90 mmHg are considered pre-hypertensive and above latter limit hypertensive [7]. In contrast, hypotension is defined as pressure values below 90/60 mmHg. Deviations from normotensive pressure values can impact the health negatively by damaging organs due to increased stress on their blood vessels, or, in case of hypotension, caused by insufficient blood perfusion. In clinical settings, the Mean Arterial Pressure (MAP) is used as well to assess BP levels and organ perfusion [5]. From the arterial pressure curve P_a, the MAP P_{map} within a cardiac cycle T_c is derived as:

$$P_{map} = \frac{1}{T_c} \int_0^{T_c} P_a(t)dt \tag{2.1}$$

Individual variations in BP result from the combination of heart rate, stroke volume, blood viscosity and the peripheral resistance of the vascular system [5]. For every person, the BP exhibits natural fluctuations throughout the day and reacts to postural changes, stress, activity, drugs and disease [8]. Moreover, SBP and DBP change along the arterial tree as detailed in Subsection 2.1.2. Therefore, medically relevant BP values are measured at the arteries.

2.1.2 Vascular Tree

Blood circulates through the body or the circulatory loops via the vascular tree, which consists of three general vessel types, the arteries, veins and capillaries. The arteries and arterioles carry blood away from the heart. In the systemic circulatory loop, this blood is highly oxygenated, whereas in the pulmonary loop, it is de-oxygenated. Arterioles are smaller vessels branching off from the end of arteries and connecting them to capillaries. Since in the circulation loop, arteries are located right after the heart followed by arterioles, they have to withstand high levels of BP. Therefore, these arterial vessels including the left ventricle are defined as the high-pressure system [4]. Due to the force they are subjected to, they are more elastic and have thicker walls than other vessel types. By dialation during systole, the aorta and arteries are able to absorb around half of the cardiac stroke volume which is released again during diastole due to the restoring force of the elastic vessel walls [4]. Hence, the discontinuous blood flow produced by

the heart is transformed into a continuous one by the so-called Windkessel effect.

The corresponding Windkessel model describes the cardiovascular system as a closed hydraulic system. An analogy is drawn to a closed circuit with a water pump connected to a chamber. The chamber is filled with water and air such that pumping of the water compresses the air, which in turn continuously pushes the water into the connected tube [9].

Capillaries are the smallest and most highly branched blood vessels. They are running through almost every tissue in order to exchange nutrients, gases, hormones and waste products. At the end of capillaries, venules begin and connect them to veins. Veins return blood to the heart acting as a counterpart to arteries. Since arteries, arterioles and capillaries absorb most of the pressure pulse, veins and venules do not have to withstand the same force. Thus, they belong to the low-pressure system together with the capillaries [4].

Pulse Pressure Amplification along the Arterial Vascular Tree

The arterial vascular tree is a complex network with branchings and taperings where reflection occurs at arterial terminations and sites of impedance mismatch [10]. At the end of the arteries, the pressure wave is reflected and reflection at both ends will occur until extinction of the wave. Therefore, the pressure wave at a given location appears as the sum of all forward and backward travelling waves. The backward wave is related to the forward wave by the reflection coefficient [10]

$$\Gamma(f) = \frac{Z_l(f) - Z_c}{Z_l(f) + Z_c} \tag{2.2}$$

where load impedance Z_l depends on frequency f, and the tube characteristic impedance Z_c is described as $Z_c = (I/C)^{1/2}$ with C denoting compliance of the vessels and I being a constant representing arterial inertance. Superposition of the backward wave and the forward wave increases peripheral BP, since at the reflection sites, there is no time delay between both waves. On the other hand, central BP is less affected, since the backward wave is shifted by $2 \cdot t_{PTT}$, where t_{PTT} is the transit time of the pulse between heart and observed reflection site. Hence, the BP waveform is undergoing amplification with increasing distance to the heart as shown in Figure 2.2. This process is called pulse pressure amplification, where the pressure levels change along the vascular tree. During

amplification, first, the SBP increases and the DBP decreases gradually towards the smaller arteries, then both values drop and end up around 15mmHg at the capillaries [5]. The MAP falls slowly but the pulse pressure increases until the saphenous artery as illustrated in Figure 2.2.

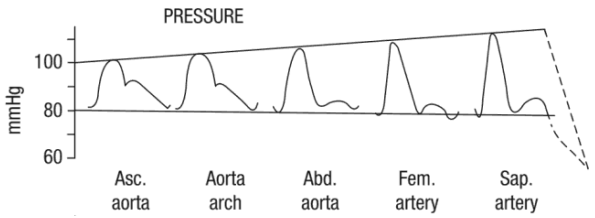


Figure 2.2: The arterial pressure undergoes amplification along the arterial tree caused by superposition of reflected pressure waves. SBP increases while MAP slowly decreases. [11]

Pressure Waveform Distortion along the Arterial Vascular Tree

Further, amplification leads to morphological changes of the pressure pulse along the arteries. Thereby, the shape of the waveform strongly depends on the distance to the reflection sites and the arterial wall elasticity. Figure 2.3 depicts invasively measured BP waves at various distances from the aortic arch visualising the spatial dependency of the waveform morphology. At the end of the systole, the pressure in the ventricle drops below the aortic pressure and briefly causes a blood reflux which is immediately stopped by the cardiac valves [11]. As shown in Figure 2.3, close to the heart, this leads to a small notch called incisura, which disappears with increased distance from the heart. However, it is replaced by the dicrotic notch which emerges due to superposition of reflected pulse waves.

Distortion in shape can further be attributed to BP-dependent compliance C and arterial wall elasticity [10]. Compliance $C = dA/dP_a$ is defined as the change in cross-sectional area A divided by change in BP P_a, and generally decreases as P_a increases following [10]

$$C(P_a) = \frac{A_s}{\pi P_1 \left[1 + \left(\frac{P_a - P_0}{P_1}\right)^2\right]} \tag{2.3}$$

for central arteries. Thereby, P_0, P_1 and A_s are subject-specific constants. Hence, pressure wave morphology depends on BP along with subject-specific parameters.

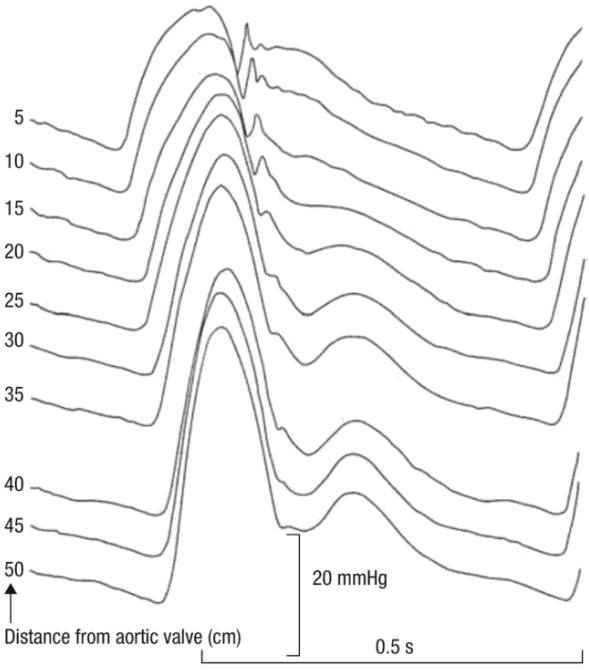


Figure 2.3: The arterial waveform changes along the arterial tree due to superposition of reflected pressure waves. Whereas in the ascending aorta, the incisura is distinct, at larger distances, it is attenuated by arterial wall elasticity. Instead, the dicrotic notch appears. Its position depends on the distance from the heart since it is caused by reflection off the aortic valve. [11]

2.1.3 Blood Vessels

Arterial walls consist of four layers which are responsible for hemodynamic regulation of the BP: Endothelium, elastin, collagen and Smooth-Muscle (SM). The innermost layer, endothelium, serves as a boundary for the blood flow, but its effects on arterial wall mechanical properties are neglegible [12]. Elastin is a highly elastic tissue which is responsible for tension and arterial elasticity at low BP levels as soon as the arterial wall expands. Collagen, on the other hand, is about 400 times [10] stiffer than the elastin layer and produces tension only when the arterial wall is stretched, i.e. at higher pressure values. SM tissue is mechanically relevant mainly in the peripheral arteries where it can modulate the arterial elasticity dynamically [12]. The SM of arterioles and smaller arteries can contract (vasoconstriction or stenosis) to regulate blood flow to different parts of the body to adapt to varying circumstances. This regulation also affects BP since

contracted arterioles decelerate blood flow out of the arteries and thus, increase pressure on arterial walls.

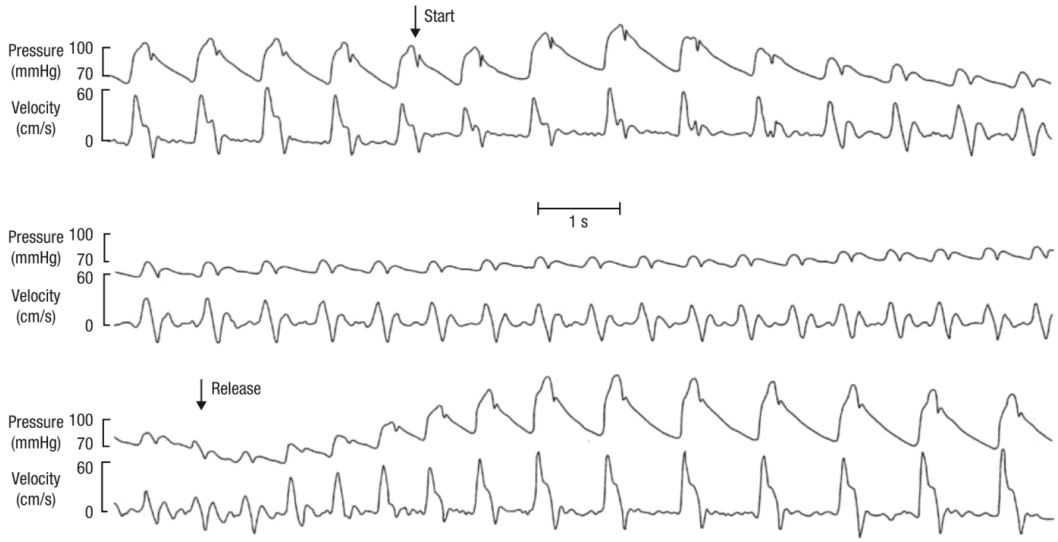


Figure 2.4: BP and blood flow in a human artery during a Valsalva manoeuvre. The BP waveform varies at different pressure values. [13]

During a Valsalva manoeuvre, vasoconstriction and vasodialation lead to a brief increase in BP followed by a significant drop and an overshoot above baseline after releasing. Mills et al. [13] recorded ABP during a Valsalva manoeuvre capturing pressure waveforms at various BP levels as depicted in Figure 2.4. Different morphologies can be observed in the waveform for different BP levels which is caused by elastic properties of arterial vessel walls.

Arterial Wall Elasticity Model

Due to their different tissue layers, blood vessels have highly non-linear elastic properties [14]. The force on the arterial wall is mainly carried by collagen, elastin and SM. In this context, their combined Young's elastic modulus E represents the overall stiffness of the vessels. To better understand the arterial wall elasticity, Mukkamala et al. [10] model the elastic properties under consideration of arterial wall mechanics and wave propagation in the arteries.

Central arteries are sparse in SM and without the influence of SM serial collagen is not recruited. Hence, central arterial elasticity is mainly influenced by BP, where for low BP, only elastin determines the elastic modulus E of the arterial

walls, and for high BP, parallel collagen contributes in addition to elastin.

In contrast, peripheral arterial elasticity is influenced by BP as well as by SM contraction. During SM contraction, serial collagen is recruited. Again, parallel collagen plays a roll in high-extension states only, hence, for high pressure, E is determined by elastin E_e and serial collagen E_{sc}, and for low pressure, by elastin, serial collagen and parallel collagen E_{pc} [10].

Further, their model considers viscous properties based on several *in vitro* and *in vivo* studies [10]. A variable viscosity η_{sm} is introduced in the SM model in parallel with the spring, such that for SM contraction, the arterial wall is viscoelastic, i.e., its elastic modulus is frequency-dependent. η_{sm} increases with SM contraction. Increased strain is immediately carried by elastin and all recruited collagen, but then, SM gradually expands until the stress is equally distributed between SM and serial collagen. Thus, the elastic modulus E of any artery can be described as a function of frequency, where elasticity at low frequencies E_{lf}

$$E_{lf} = E_e + E_{pc} + E_{sc} \cdot \frac{E_{sm}}{E_{sc} + E_{sm}} \tag{2.4}$$

is smaller than elasticity at high frequencies E_{hf}

$$E_{hf} = E_e + E_{pc} + E_{sc} \tag{2.5}$$

Due to the low E_{sc} in central arteries, frequency changes have a relatively low impact on its elastic modulus. However, in the peripheral arteries, E_{sc} is large and depends strongly on frequency [10].

Blood Pressure Propagation Model

Considering the mechanical properties of the arterial walls, the vascular system and the propagation of the BP wave can be modelled by the propagation of a pressure wave in a cylindrical elastic tube network [6]. The energy coming from the heart is transformed into elastic energy resulting in the periodic dilation of the vessel walls which travels along the arteries. Looking at the longitudinal coordinate X along the artery, the radial dilation r of the arterial vessel can be described as a function of coordinate X and time t, $r(X,t)$, where r_0 denotes the inner radius of the vessel in the static condition. h is the thickness of the vessel wall. A small element of the vessel tube of length ΔX and circumferential angle $d\theta$ is considered. The following forces and moments are acting on the element [15]:

1. The restoring force F_N is exerted from the deformation of the vessel wall and acts in the radial direction.

$$F_N = -E_N \frac{\Delta r}{r_0} \Delta O r d\theta \tag{2.6}$$

where elasticity modulus E_N denotes the elasticity of vessel walls in radial and circumferential directions.

2. The shear force F_S is acting parallel to the X-axis restoring the wall elements orientation.

$$F_S = E_{rX} \frac{dr}{dX} h r d\theta \tag{2.7}$$

where E_{rX} denotes the shear modulus of the vessel wall.

3. The resistance force F_f caused by the blood viscosity ηi and the adherent vessel wall and acts in the radial direction. F_f is considered to be proportional to radial velocity $v_r = dr/dt$.

$$F_f = -\eta_i \Delta X r d\theta \frac{dr}{dt} \tag{2.8}$$

4. The momentum M of the element in the radial direction is

$$M = \mu \Delta X r d\theta \frac{dr}{dt} \tag{2.9}$$

with μ describing the density of arterial wall and adherent blood.

Based on Newton's second law, from Equations 2.6 - 2.9 follows [15]

$$M = F_N + F_R + F_S + F_f \tag{2.10}$$

such that [15]

$$d\theta \mu \Delta X r \frac{d}{dt}(r \frac{dr}{dt}) = d\theta \left(-\frac{E_N}{r_0} \Delta X r \Delta r + (E_{rX} h r \frac{dr}{dX}) - \eta_i \Delta X r \frac{dr}{dt} \right) \tag{2.11}$$

With cancelling $d\theta$ in Equation 2.11 and further manipulations, for $\Delta X \rightarrow dX$ in a uniform elastic vessel follows [6]

$$\mu \frac{d^2 P_a}{dt^2} + \eta_i \frac{dP_a}{dt} + P_a \frac{E_N}{r_0} = E_{rX} h \frac{d^2 P_a}{dt^2} \tag{2.12}$$

Equation 2.12 approximates the propagation of the BP wave in the arterial system considering uniform elastic vessels. When further introducing radial oscillation

of the elastic wall with N harmonics ω_n of fundamental vibratory frequency ω_0, Equation 2.12 can be solved by separating spatial and temporal dependences [6]

$$P_a(X,t) = \sum_{n=0}^{N} P_n(X)\cos(\omega_n t) \tag{2.13}$$

The dominant oscillation frequency ω_0 in Equation 2.13 of the pressure pulse follows the heart rate, whereas the pressure waveform consists of other-order harmonics which result from pressure wave reflections and modulations due to respiration. To account for pressure pulse modulation and amplification, adoption of amplitude modulation of propagating radio signals has been suggested [6]. Hence, while satisfying Equations 2.12 and 2.13, BP at different sites in the arterial system X_i and X_{i+1} is described by [6]

$$P_a(X_{i+1},t) = P_a(X_i,t)\left[\beta\cos(\omega_{HR}t) + (1-\beta)\right] \tag{2.14}$$

where β lies between 0 and 0.5 and indicates the magnitude of amplitude modulation of the pulse pressure.

2.1.4 Blood Pressure Surrogates

From the described physiological properties and relationships of BP, surrogate parameters can be derived to measure BP indirectly. Some common surrogates are derived from the Pulse Wave Velocity (PWV), such as the Pulse Arrival Time (PAT) and Pulse Transit Time (PTT). The relation between PWV v_{PWV} and BP is described by the Moens-Korteweg relation [16]

$$v_{PWV} = \sqrt{\frac{E \cdot h}{2r \cdot \rho}} \tag{2.15}$$

where E denotes the combined Young's elastic modulus of the vessel wall, h denotes the wall's thickness, r refers to the vessel radius and ρ is the blood density. Thereby, elasticity modulus E is a function of BP (see Section 2.1.3). In an *in vitro* study, Hughens et al. [17] investigated the relationship between E and the transmural arterial pressure P_a and published the widely used empirical formula

$$E = E_0 e^{\alpha P_a} \tag{2.16}$$

where E_0 and $\alpha > 0$ are constants that depend on the measurement location and individual subject. A different modelling approach by other research groups [18, 19] can be summarised as [14]

$$E = (1 - \sigma^2) \cdot \frac{r_0^2}{h_0} \cdot \frac{\partial P_a}{\partial r_0} \tag{2.17}$$

where σ is a subject-specific constant, h_0 denotes the thickness of the vessel wall and r_0 denotes the mean radius of the blood vessel. Equation 2.16 and 2.17 both show that elasticity modulus E increases with BP P_a.

PTT t_{PTT} represents the propagation time of the pressure pulse travelling with PWV v_{PWV} along an arterial segment of length γ [5].

$$t_{PTT} = \frac{\gamma}{v_{PWV}} \tag{2.18}$$

Equation 2.15 and 2.18 show that t_{PTT} decreases with increasing elastic modulus E. In turn, elastic modulus E increases as BP increases (see Equation 2.16 and 2.17). From Equations 2.15, 2.16 and 2.18 the relationship between BP and PTT can be derived as [16]

$$P_a = \alpha_1 \ln t_{PTT} + \alpha_2 \tag{2.19}$$

where α_1 and α_2 are subject specific constants depending on the individual's physiology. In contrast to PTT, PAT comprises the Pre-Ejection Periode (PEP) since it is measured between the electric excitation of the heart and the arrival of the pressure pulse at an arterial site.

Some studies have been developing and investigating surrogate parameters that are not derived from the PWV but rely on the PPG waveform analysis [20], heart sound characteristics, impedancecardiography or ballistocardiography [5]. Relevant approaches related to the PPG waveform analysis are presented and discussed in Chapter 3.

The previous sections have established the relationship between BP level and BP waveform, where the waveform also strongly depends on individual physiological factors such as the compliance of the vessels, arterial elasticity and the distance of reflection sites to the heart (see Section 2.1.3). Further individual influences on the PPG waveform as a BP surrogate emerge from the optical measurement principle of the PPG signal which are detailed in the following Section 2.2.

2.2 Photoplethysmography

Photoplethysmography is a method to optically measure blood volume changes in subcutaneous tissue by taking advantage of reflective and absorbent properties of skin tissues and blood components. Thereby, the local blood volume strongly correlates with the BP pulse due to their direct physiological relationship. In this section, the measuring principle is explained (2.2.1) and the optical properties of human skin are detailed with respect to PPG measurement (2.2.2).

2.2.1 Measuring Principle

A PPG is typically obtained using a pulse oximeter which consists of a Light Emitting Diode (LED) (red, infrared and/or green light) and corresponding photo diodes. There is the transmissive method, often applied at the finger or earlobe, where the LEDs and photo diodes are placed on opposite sides of the tissue to measure the intensity of transmitted light. In contrast, for the reflective method, the photo diode is placed next to the LED to measure reflected photons. Thereby,

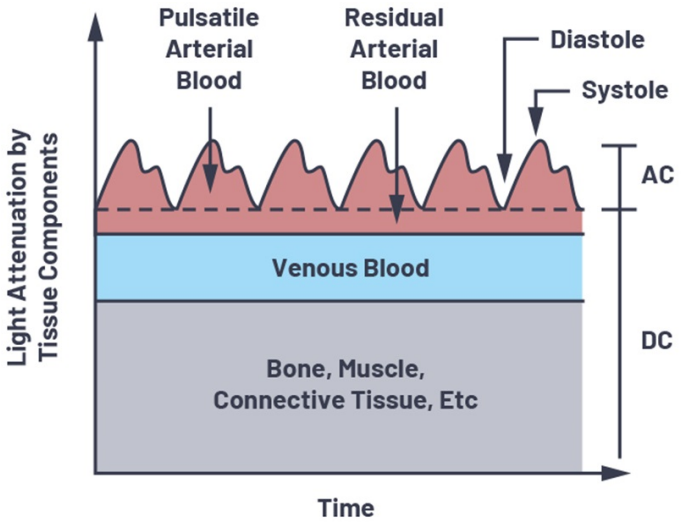


Figure 2.5: Absorption of light in tissue components. [21]

the intensity of light returned to the skin surface strongly depends on the amount of blood, since haemoglobin is one of the main absorbers and fluctuating with every pressure pulse. This changing part of blood volume is mainly caused by the pulsation of BP in the arteries and dilation of the vessels as described in Section 2.1.2. Figure 2.5 illustrates the absorption of light by tissue components over

time. The relationship between BP $P_a(t)$ and light intensity sensed by the photo diode Q is established as [6]

$$Q(t) = -\kappa C P_a(t) + Q_0 \tag{2.20}$$

where κ is the modulus of absorbed energy during vessel dilation, C is the compliance of the vessel and Q_0 a fitting constant for reflections at tissue layers.

To select the optimal wavelength for a rPPG measurement, the optical properties of the different skin layers are considered in the following subsection.

2.2.2 Optical Properties of Human Skin

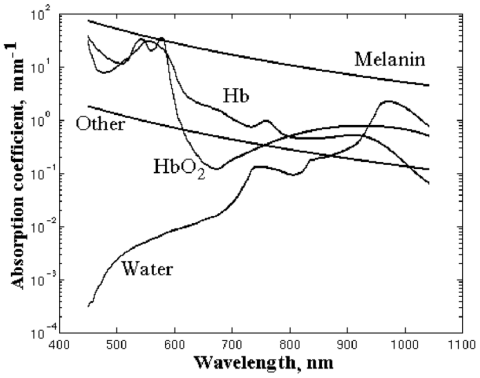
On the stratum corneum, the surface of the skin, approximately 4 to 7% of the light is lost due to specular reflection caused by the refractive index change at the air skin interface [22]. The remaining light is further affected by absorption, scattering and diffuse reflection, whereby the magnitude of these events depend on the tissue characteristics and light source (e.g. wavelength, polarisation). The following paragraphs describe these optical events in more detail and outline further influences on optical parameters of skin.

Light Absorption

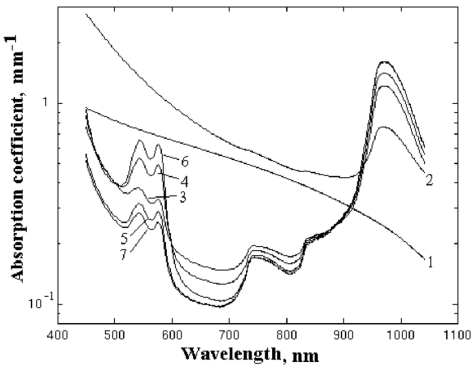
The main absorbers in superficial skin tissue are melanin, haemoglobin and subcutaneous lipids [22]. The highest melanin absorption occurs at wavelengths below 510nm. While the absorption is still significant for red light, it exponentially decreases towards the Near Infrared (NIR) spectrum [23]. In contrast, maximal lipid absorption takes place at longer wavelengths around 760 and 930nm [22]. Another dominant absorber is haemoglobin with three peaks in the visible spectrum, the most dominant one in the blue region and the other two between 500 and 600nm wavelength [25]. Further, water in skin tissue absorbs light increasingly at wavelengths above 800 nm [23]. Figure 2.6a depicts the absorption spectra of melanin and blood components, while Figure 2.6b shows that absorption in deeper skin layers is highly related to the absorption coefficient of haemoglobin.

Light Scattering

Light scattering describes changes in direction, polarisation or phase of photons [25] and is caused by differences in refraction indices of the tissues and their



(a) Absorption coefficients of melanin, oxyhaemoglobin, deoxyhaemoglobin and water. [24]



(b) Simulated absorption coefficients of skin layers: (1) stratum corneum, (2) living epidermis, (3) papillary dermis, (4) upper blood net dermis, (5) reticular dermis, (6) deep blood net dermis, (7) subcutaneous lipid layer. [24]

Figure 2.6: Absorption coefficients of 2.6a melanin and blood components, and 2.6b of skin layers as a function of wavelength.

surroundings. Cell membranes, melanin, haemoglobin, platelets, blood vessels and fat cells are forward scattering elements. This means, that photons which return to the skin surface have undergone a large number of scattering events. Due to the wavelength dependence, on average, red and NIR light will have travelled deeper than blue and green light [25]. Therefore, deeper lying blood vessels only absorb light with longer wavelengths. Cellular organelles and collagen fibre, on the other hand, scatter in all directions. In general, scattering in human skin layers is greater for red light than for NIR wavelengths [22]. Scattered photons may eventually be absorbed by one of the skin compounds, or after multiple scattering events, may escape from the skin surface as diffuse reflection [23].

Influences on Optical Parameters

Nussbaum et al. [23] examined factors that affect the light penetration of skin. They showed that in contrast to 840 nm light, the transmittance of 660nm light significantly decreases with increasing skin darkness, i.e. density of melanin particles. However, this impact diminishes with target depth and increasing wavelength [22]. Whereas the transmission rate of red light is not significantly affected

by age and gender, for NIR (840 nm) the average transmission rate is only 66.4% for females compared to that of males [23]. Further, Nussbaum et al. [23] showed that the transmission rate is almost halved when having a 2-mm distance between skin and light source compared to contact-based irradiation using pressure. As a possible reason for the last effect, they suggest an reduction in blood volume, but also the reflection at the air interface.

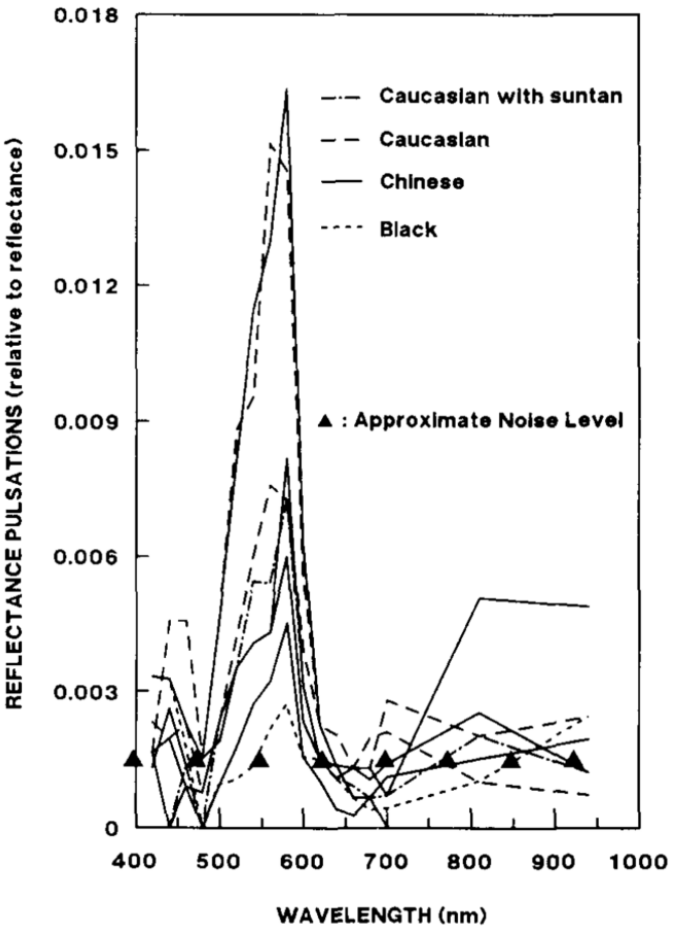


Figure 2.7: Reflectance pulsation spectrum relative to mean reflectance of the skin, measured at the lower leg. [26]

Cui et al. [26] measured *in vivo* reflectance of blood and skin tissue at various wavelengths. They observed a maximum in the reflectance pulsation spectrum around 575nm with a peak 10 times larger than the modulation for red and NIR light. Skin colour does not significantly affect the shape of the reflectance pulsation spectrum, but reduces the total reflectance pulsation magnitude. Another study [24] confirms those results with a simulation of the reflectance spectra for

0% and 40% blood volume in the upper blood net dermis.

2.3 Signal Processing

For this work, an additive noise model is assumed where the user signal is superimposed by noise. Most biomedical signals have small amplitudes and thus, are sensitive to inferences of other signals and artifacts. In biomedical signals, noise can originate from thermal noise due to vibration of electrons within the electrical conductor [27]. Thermal noise is proportional to temperature and can be modelled as white noise. In contrast to thermal noise, flicker noise inversely correlates with frequency and is caused by non-constant flow of electrons at interfaces of two materials. Another common source of noise is the power-line interference which introduces a 50-60 Hz frequency due to electromagnetic fields of nearby electrical equipment. CMOS image sensors suffer from thermal noise in the output amplifier and flicker noise as well [28]. Further, noise is introduced as photo detector shot noise which is caused by the random generation and flow of electrons and relates to the discrete nature of electronic charge. Lastly, Analog Digital Converter (ADC) can create quantisation noise at each pixels sensory input, where it is assumed to be stochastic and uncorrelated with the user signal.

This section outlines signal processing fundamentals including filters, spectral analysis as well as the multi-variate analysis method of Independent Component Analysis (ICA).

2.3.1 Filters

A Finite Impulse Response (FIR) filter is a discrete time filter, where the output y is computed as the sum of weighted past input values x [29]

$$y[n] = \sum_{k=0}^{K} b_k x[n-k] \tag{2.21}$$

where b_k is the weight coefficient assigned to the k-th previous input. Hence, the FIR transfer function $H[z]$ can be described as

$$H[z] = \sum_{k=0}^{K} b_k^{-k} \tag{2.22}$$

An advantage of FIR filters is their linear phase response, since they have a zero-phase distortion. Therefore, they are well suited for removing artifacts and noise from biomedical signals.

A common frequency domain filter is the Butterworth filter [30] due to its close-to-linear phase response and sharp frequency roll-off for high filter orders. It's amplitude response is defined as

$$|H(j\omega)| = \frac{1}{\sqrt{1 + (\omega/\omega_c)^{2n}}} \tag{2.23}$$

with ω_c being the cut-off frequency and n the filter order.

2.3.2 Fourier Analysis

Fourier analysis is the study of the frequency spectrum of a signal based on the Fourier transform [31]. Mathematically, the Fourier series $F(f)$ as a function of frequency f is defined as

$$F(f) = \int_0^T x(t)\mathrm{e}^{-2\pi i f t} dt \tag{2.24}$$

where $x(t)$ is the input signal as a function of time t, and i denotes the imaginary unit. Respectively, the Discrete Fourier Transform (DFT) for discrete periodic signals is

$$F[k] = \sum_{n=1}^{N} x[n]\mathrm{e}^{-2\pi i k n/N} \tag{2.25}$$

where k is the counter of frequencies in the signal. Thereby, the highest frequency f_{max} in signal $x(t)$ must be lower than half of the sampling frequency f_s. Hence, the frequency resolution f_r is

$$\frac{f_s}{N} \leqslant f_r < \frac{f_s}{2} \tag{2.26}$$

The Fast Fourier Transformation (FFT) [32] is an algorithm for computation of the DFT which reduces complexity by recursively breaking DFT down into smaller parts such that complexity is reduced from $O(N^2)$ to $O(N \cdot \log N)$.

2.3.3 Independent Component Analysis

ICA [33] is a signal processing method that aims at separating a multivariate signal into its additive components, hence following the general objective of blind

source separation. It is based on the assumption that these components are statistically independent and non-Gaussian.

The j input signals $x_j(t)$ are assumed to be a linear combination of the n original independent signals s_n, where $j = n$. Dropping the time index, $x_j = a_{j1}s_1 + a_{j2}s_2 + ... + a_{jn}s_n$ are weighted sums of n independent components. For an input matrix $\mathbf{X} = (x_1, x_2, ...x_j)^T$ and independent component matrix $\mathbf{S} = (s_1, s_2, ...s_n)^T$, the generative model is given as

$$\mathbf{X} = \mathbf{AS} \tag{2.27}$$

with $\mathbf{A}$ being the mixing matrix that defines how the independent components $\mathbf{S}$ are combined to form the input signals in $\mathbf{X}$. Thereby, the model assumes linear combinations of the sources of $\mathbf{X}$. To reconstruct the original signals in $\mathbf{S}$, a unmixing matrix $\mathbf{W}$ is computed which is the inverse of mixing matrix $\mathbf{A}$.

$$\mathbf{S} = \mathbf{WX} = \mathbf{A}^{-1}\mathbf{X} \tag{2.28}$$

2.4 Machine Learning

Machine learning is a branch of artificial intelligence that enables algorithms to improve automatically by the use of data. Machine learning algorithms optimise a model based on training data by using statistical methods or exploiting recurrent patterns. A trained model then can make decisions or predictions for unknown input data without being explicitly programmed. Since training data are the basis from which the algorithms deduce their models, it is essential to have a relevant, well balanced and pre-processed data set prepared. In this section, Artificial Neural Network (ANN)s and Random Forest Regression (RFR) are described followed by feature selection and dimensionality reduction methods.

2.4.1 Artificial Neural Networks

ANNs are a subset of machine learning which emulate biological neural networks. It consists of multiple processing elements, called neurons or perceptrons, that receive an input and produce an output depending on a pre-defined activation function. Typically, these neurons are arranged in layers such that outputs of one layer serve as inputs to the subsequent layer. Each neuron typically receives mul-

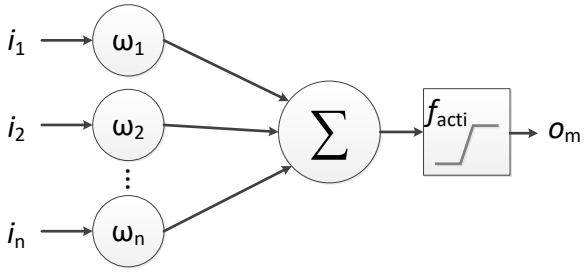


Figure 2.8: A single neuron of an ANN. All weighted inputs to a neuron are aggregated and passed onto its activation function.

tiple inputs from previous layers. All n input values i_n of a neuron are multiplied by its associated connection weights W_n and then aggregated (e.g. summation, averaging, maximum) to a single input value i_a for the neuron (see Figure 2.8). From this input value, the neuron computes its output o_m by using an activation function f_{acti}, such as a sigmoid, hyperbolic-tangent or other non-linear activation function. This process takes place in each neuron of a layer and is propagated through every layer until the output layer is reached.

A traditional feed forward neural network consists of an input layer where the signal is received, an arbitrary number of hidden layers that learn the mapping of input to output and can approximate any continuous function, and finally the output layer which returns the prediction result. A neuron is usually connected to all or multiple neurons of the previous and of the subsequent layer. As for all network types, the size and dimension of the input dictates the number of neurons in the input layer. Architectures made of these three types of layers and a forward data flow are commonly referred to as Multi-Layer Perceptron (MLP).

The neurons are trained with a back-propagation learning algorithm. During the forward phase of the training process, the weights are fixed and input data is propagated through the network. From the predicted output and the ground truth output, an error signal is computed with respect to the defined loss function $L(D, W)$ with D being a training sample and W being the set of all weights. Subsequently, this error signal is propagated through the ANN in backward direction while the weights are adjusted to minimise the error. By calculating the change

in error with respect to the weight, the algorithm can determine how much each weight W_n needs to be adjusted:

$$W_n = W_n - a\frac{\partial L(D, W)}{\partial W_n} \tag{2.29}$$

where a denotes the learning rate. For every batch, the error is propagated through the ANN updating the weights until the model stops converging.

While the described fully-connected layers are the most common type, there are various other layer types as well. For image processing, convolutional layers are deployed. Here, an n-dimensional set of weights, called kernel, is applied to the input in a sliding-window fashion. At each position, an element-wise multiplication with the input is performed and the output is then obtained by summation of these products. The kernels describe the probabilities of a given pattern of pixels representing a feature. Convolutional layers are usually paired with pooling layers to reduce the size of the passed data.

When designing a neural network, several aspects need to be considered. The network type depends on the input data (e.g. separate features, time series, signal, image etc.) and the task that should be accomplished (e.g. time series prediction, classification, regression, sequence to sequence prediction, image to text etc.). These networks can become arbitrarily complex and for each task there may be many different approaches that could lead to the desired result. In general, when deciding for an architecture, performance and computational complexity play a role. For feed-forward ANNs or MLPs, it has been mathematically proven that with a single hidden-layer, any continuous function can be learned. The optimal number of hidden layers and neurons has to be determined empirically. Whereas too few neurons or layers limit learning capabilities, too many neurons or layers produce overfitting. This consideration is also known as the bias-variance trade-off and depends on the complexity of the relationship to be learned and the quality and size of training data. The data set plays an essential role in the successful training of an ANN and has to be representative of all target output values. More hyper-parameters to be tuned in search of the best MLP model are listed and explained in Table 2.1.

Table 2.1: Hyper-parameters of MLPs.

Hyper-parameter	**Description**
Number of hidden layers	More hidden layers can model more complex functions, but tend to overfit.
Neurons per layer	More neurons per layer can model more complex functions, but tend to overfit.
Activation functions	Activation function of the neurons, where $o_m = i_a \cdot f_{acti}$
Regularisation	Reduces overfitting, but limits learning capabilities if too strong.
Optimiser	Learning algorithm. Each has its strengths and weaknesses.
Loss function	For calculating the loss in the training process after the final layer.
Learning rate	Step size for updating the weights in the learning process.
Batch size	Number of data samples passed through the ANN at once.
Number of epochs	Number of times the ANN sees the whole data set (if steps per epoch = number of batches).

2.4.2 Random Forest Regression

RFR is a supervised machine learning algorithm and belongs to the ensemble learning methods which consist of multiple weaker models [34]. In ensemble learning, the results of multiple weaker classifiers or regressors is averaged in order to form a single strong model. The assumption for this is that the errors of each model are independent and cancel each other out.

A RFR deploys decision trees as base models which resemble a tree-like structure. Each of these decision tree starts with a single node which splits into two paths (branches) based on a specific feature's value. This structure is repeated recursively until a terminal node (leaf) is reached. The leaf nodes hold the decision trees' output values which finally are aggregated among all trees forming the overall output. All calculations are performed in parallel and the decision trees are build independently from each other.

For training, the data set is split into randomly sampled subsets for each of the decision trees to build. At each new node for the decision tree, the feature out of a random feature subset is chosen as criterion that minimises the prediction error most. This recursive partitioning process is repeated until a stopping criterion is met. Due to the ensemble learning technique, RFR has a low risk of overfitting. Further, it can handle missing values and non-linear parameters efficiently while beeing robust to outliers and noise.

2.4.3 Feature Selection

In most data sets, not all features contribute to prediction accuracy of the model. Therefore, feature selection algorithms are deployed which find those features that are most relevant for predicting the target variable. Thereby, they remove irrelevant and redundant features which may decrease the model's accuracy. In general, fewer features are desirable due to a reduced model complexity and hence, less training data is required. There are three types of feature selection algorithms: filter, wrapper and embedded methods.

Filter methods use statistical measures to assign scores and rank the features. These methods usually are univariate and consider each feature independently or with respect to the target variable. Therefore, they do not remove multicollinearity or redundant features [35].

Wrapper methods select features by evaluating combinations of features on a prediction model. Different methodical, stochastic or heuristic search processes can be used to add or remove features iteratively, and each combination is evaluated based on model accuracy. This makes wrapper methods computationally very expensive, but enables them to find the best feature subset. Sequential Forward Selection (SFS) is a wrapper method where features are sequentially added, starting with a single feature and terminating when no further improvement is obtained by adding more features. In every iteration, model accuracy is evaluated with each of the remaining features as addition into the selected feature subset.

Finally, embedded methods select features while the model is being created [35]. A popular example is to introduce regularisation to penalise complex models by discarding irrelevant features. Another example for embedded methods is feature importance in tree-based algorithms.

Feature selection as well as dimensionality reduction methods reduce the number of features that will be used as input to the model. However, feature selection algorithms pick the most relevant features without changing them, whereas in dimensionality reduction, the information of all features are condensed by combining them into new features.

2.4.4 Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction method for a data set with interrelated variables, where the coordinate space is transformed such that the new hyperspace maximises the variance of the data [36]. Thereby, the first Principal Component (PC) PC_1 is the line that best approximates all data in the original coordinate space. Respectively, the second PC PC_2 is the line that best approximates the data's variation on an axis orthogonal to PC_1, and so on. Hence, the PCs represent a new set of uncorrelated features (or dimensions) each consisting of a linear combination of the original features (or dimensions).

To obtain these PC, a covariance matrix $\mathbf{B} = f \times f$ (where f is the number of original features) is created from the standardised data set and describes the covariance of each possible combination of two features $\mathbf{B}_{x,y}$. From this, the eigenvectors and eigenvalues of $\mathbf{B}$ are computed. The eigenvectors then indicate the directions of variance in the data set and the eigenvalues their magnitude. Consequently, the eigenvector belonging to the highest eigenvalue is PC_1. To reduce dimensionality, eigenvectors with low eigenvalues are eliminated.

2.5 Image Processing

BP estimation methods in this work are based on image data. Therefore, image processing forms an essential part in the methodology and is described in this section. Specifically, theory of object detection and tracking algorithms are outlined in subsections 2.5.1 and 2.5.2, respectively, followed by colour models in subsection 2.5.3.

2.5.1 Object Detection

Object detection refers to the localisation (typically indicated by a bounding box around the object) and classification of objects in images. In general, object detection methods employ a type of feature descriptor to compress the image information and characterise objects. Based on these feature descriptors and their typical values for each object, a classification algorithm is trained and can be deployed for object detection in new images. Feature descriptors are usually applied to small image patches of various sizes to enable detection of objects at

different scales. In this subsection, five object detection algorithms are briefly described which are considered or deployed for detecting a hand or face for rPPG extraction from video data.

Viola-Jones Object Detector

The Viola-Jones face detection method [37] was one of the first approaches on digital pattern recognition and face detection in the machine learning domain and takes an ensemble approach. The method employs a cascade of weak classifiers, where each classifier is based on Haar features. Haar features are pre-defined convolution kernels represented by subdivided rectangles. The difference between the sums of pixel intensities within these subregions form the convolution's output.

Histogram of Oriented Gradients

In contrast, Histogram of Oriented Gradients (HOG) [38] is a feature descriptor which considers occurrences of different gradient orientations. For small image sections, histograms are computed that detail the distribution of gradient magnitudes for the angles from 0° to 180° (with a step size of 20°) within this region.

YOLO

You Only Look Once (YOLO) [39] and Single-Shot Detection (SSD) [40] are CNN-based one-stage object detectors which compute bounding boxes and class labels at once. For YOLO, the input image is divided into uniform subregions, for each of which an object class together with its probability and bounding box position is computed. Since multiple subregions might produce differing bounding boxes for the same object, Non-Maximum Suppression is performed selecting only the one with the highest probability. Subsequently, those bounding boxes are eliminated that have the highest overlap with the currently selected one until the final predictions are obtained.

Single-Shot Detector

SSD [40] is a CNN-based multi-box object detector, too, but consists of more convolutional layers. These are used as multi-scale feature maps to obtain a varying

number of subregions and thus bounding boxes at multiple scales. Unlike methods that require an additional object proposal stage, SSD eliminates this proposal generation and subsequent pixel or feature resampling stage by encapsulating all computations in a single network.

SSD defines the output space of bounding boxes as a set of standard boxes across different aspect ratios and scales per feature map location. During prediction, the network generates scores for the presence of each object category in each standard box and produces adjustments to the box to better represent the object's shape. In addition, the network combines predictions from multiple feature maps with different resolutions to handle objects of different sizes in a natural way. To remove redundant detections, Non-Maximum Suppression (NMS) is applied. NMS retains the bounding box with the highest confidence score and removes all other bounding boxes that exceed a certain overlap threshold with this box.

Hand Landmarks

Whereas the previously described object detection algorithms are trained to detect entire objects, landmark detection [41] is designed to find the positions of several unique feature points of a non-rigid object. Therefore, it is well suited for facial feature (mimic) tracking or for classification of hand gestures and poses. It is based on CNNs as well.

2.5.2 Object Tracking

Object tracking is the localisation of a specific image region over a sequence of video frames. This image region might be an object of which the position in the first image is given. In this subsection, five object tracking algorithms are briefly described which are considered for the BP estimation from video data.

MIL

The Multiple Instance Learning (MIL) tracker [42] uses the "tracking by detection" approach. Such methods consist of a discriminative classifier which is trained online on the specific object by extracting positive and negative examples from the current frame. Thereby, the MIL tracker overcomes the problem of drifting due to slight inaccuracies in previous frames by employing a MIL classifier.

A MIL classifier uses labelled bags of instances (where the positive labelled bag contains at least one positive instance) instead of individually labelled instances, and thus, is more robust against incorrect class assignments.

Median Flow

In contrast, the median flow tracker [43] is based on optical flow, i.e. a vector field of apparent motion of objects relative to the scene. The Lucas-Kanade method [44] estimates optical flow under the assumption that the intensity of pixels remains constant for small movements and short time periods. Based on this, they approximate trajectories of 3×3 pixel patches using the least square method. Considering an object, the median flow tracker estimates its movement by the median x-shift and median y-shift of its feature points. For the calculation of this median trajectory, those feature points are used that exhibit the lowest forward-backward errors. The forward-backward error is computed by comparing the feature points' trajectories in forward direction and the trajectories in backward direction of the image sequence.

MOSSE

Another group of object tracking algorithms is based on adaptive correlation filters. Here, the correlation is used to compute the similarity between two images, specifically, between the object to be tracked in the previous frame and possible new object locations in the current frame. Hence, these methods aim at finding the pixel in the new frame that maximises correlation ρ between the object's filter matrix $\mathbf{H}$ and the intensity matrix $\mathbf{J}$ of the new image section. For efficiency, correlation is performed in the frequency domain, where according to the convolution theorem, a convolution can be performed by element-wise multiplication. For transformation of the image matrices an FFT is used [45]:

$$\rho(\mathbf{J}, \mathbf{H}) = FFT^{-1}(FFT(\mathbf{J}) \cdot FFT(\mathbf{H})) \tag{2.30}$$

The Minimum Output Sum of Squared Error (MOSSE) algorithm [45] employs the correlation filter $\mathbf{H}$ on a gray scale version of the image. In the first frame, this adaptive filter matrix $\mathbf{H}$ is initialised by the image section to be tracked by

minimising the error between the current prediction and ground truth [45]:

$$\min_{\mathbf{H}^*} \sum_i |FFT(\mathbf{J}_i) \cdot FFT(\mathbf{H}^*) - FFT(\mathbf{G}_i)|^2 \tag{2.31}$$

where $\mathbf{G}_i$ is the ground truth tracking output of the i-th training sample. In subsequent frames, correlation filter $\mathbf{H}$ is adapted to the new appearance of the depicted object. This way, the method is robust against changes in illumination, size, orientation and non-rigid transformations.

KCF

The Kernelized Correlation Filter (KCF) method [46] is based on correlation filters as well, but it uses kernel trick and circulant matrices. In circulant matrices, rows of pixels in the image section of the object are shifted and thus, can be used as negative training samples. Further, KCF trackers additionally apply HOG features to improve prediction.

CSRT

Another correlation filter-based tracker is the Channel and Spatial Reliability Tracker (CSRT) [47] which adjusts the filter map to focus on the part of the object that is suitable for tracking. Thereby, the reliability score considers channel-wise quality of the learned filters and is employed to weight the corresponding features during tracking.

2.5.3 Colour Models

Colour videos usually are recorded and saved in the Red Green Blue (RGB) colour model which is an additive model, i.e. red, green and blue light intensities are added together in different ratios to obtain the whole visual colour spectrum. Corresponding RGB cameras use Complementary Metal-Oxide-Semiconductor (CMOS) or Charge-Coupled Device (CCD) sensors and operate with some variation of the RGB model where its grid consists of arrangements of red, green and blue light detectors. Based on this, digital RGB images are saved in three channels, red, green and blue with each colour's intensity represented by a 8-bit value from 0 to 255, 0 being the lowest intensity. This colour representation is related to the human eye. The human visual system, or more specifically the

retina, consists of three types of cone cells which respond to light of different wavelengths.

However, there are several other colour models that introduce dimensions for brightness or lightness separating them from colour information [48]. This subsection describes some alternative colour models which are relevant to this work.

CMYK

In contrast to the RGB model, the Cyan Magenta Yellow Key (CMYK) colour representation is a subtractive model commonly used for printing. The idea of this colour model is based on the absorption of certain wavelengths to obtain the target colour in the reflected light. The four parameters Cyan C, Magenta M, Yellow Y and Black K can range from 0 to 100 where white is obtained by $C = M = Y = K = 0$.

CIE Lab and CIE Luv

The colour model CIE Lab is represented in a three-dimensional coordinate system, where L represents lightness, with $L = 0$ being black and $L = 100$ referring to white [48]. The values $a \in [-128, 127]$ and $b \in [-128, 127]$ indicate colour, where $a > 0$ is red and $a < 0$ is green. Accordingly, $b > 0$ refers to yellow and $b < 0$ to blue. The CIE Luv colour model is similar to the CIE Lab model, however, it exhibits a reduced green area and an increased blue area. Further, CIE Luv deploys Judd-type white point adaptation, whereas CIE Lab uses a Kies transform. The Lab and Luv colour model can be transformed from the Cartesian coordinates into polar coordinates as the cylindrical models CIE LCh(ab) and LCh(uv). These CIE LCh colour models are different from HSV or HSL, since LCh models are perceptually uniform and use four instead of three colours.

HSL and HSV

HSL and HSV colour models use cylindrical models represented by polar coordinates, where for both models, the angular dimension H indicates the hue and $S \in [0, 1]$ the saturation of the colour [48]. $L \in [0, 1]$ in HSL corresponds to lightness, while V in HSV stands for "value" and can be interpreted as brightness. The difference between these two colour models is that in HSL, a colour

with maximum lightness is pure white regardless of the other two values, whereas in HSV, a colour with maximum value has maximum brightness in the sense of shining a strong light on a coloured object. Hence, pure saturated colours in HSL have lightness $L = 0.5$, whereas in HSV they have value $V = 1$.

YCrCb, YUV and YIQ

YCrCb is a three-dimensional Cartesian colour model often deployed in video and digital photography. $Y \in [0, 1]$ indicates the luma, i.e. brightness in videos, and the chromaticity values $Cr \in [-1, 1]$ and $Cb \in [-1, 1]$ correspond to red/cyan and blue/yellow components [48]. YUV deploys a similar colour model representation, however, U and V indicate blue-luminance and red-luminance respectively. Likewise, the YIQ colour model can be derived from YUV by rotating 33° along the Y-axis. Hence, $I \in [-1, 1]$ represents the orange/blue range and $Q \in [-1, 1]$ indicates the purple-green range.

2.5.4 Camera-based PPG extraction methods

There are several methods for extraction of PPG signals from videos, i.e. rPPG. Two of the currently most advanced techniques are the Plane Orthogonal to Skin (POS) algorithm [49] and the Chrom algorithm [50].

Chrom Algorithm

The Chrom algorithm [50] was specifically designed for rPPG measurement. It models the pulse as a linear combination of the three RGB channels and operates under the assumption of a standardised skin colour profile to achieve white balancing of the video frames.

The algorithm constrains all colour variations to the pulsatile direction by expressing its chrominance signals as a projection of $\tilde{\mathbf{C}}_n(t)$ onto the plane orthogonal to the variation direction of specular reflection. To handle variations caused by different lighting conditions, Chrom adopts a standardized skin-tone vector $\mathbf{u}_{skin} = \mathbf{M}^{-1} = [0.77, 0.51, 0.38]^{\mathrm{T}}$ for white balancing, ensuring independence from the colour of illumination and lighting conditions. Here $\mathbf{M}$ represents the diagonal mapping matrix for $\tilde{\mathbf{C}}_n(t)$. Furthermore, the algorithm incorporates a 2×3 projection matrix $\mathbf{P}_{Chrom}$ to define a plane orthogonal to the variation

direction of specular reflection within the temporally normalized RGB space:

$$\mathbf{P}_{Chrom} = \begin{pmatrix} 3 & -2 & 0 \\ 1.5 & 1 & -1.5 \end{pmatrix} \tag{2.32}$$

The two resulting signals, denoted as $S_1(t)$ and $S_2(t)$, are generated through this projection matrix, revealing motion-induced and pulse-induced variations in anti-phase:

$$\mathrm{S}(t) = \mathbf{P}_{Chrom} \cdot \mathbf{M} \cdot \tilde{\mathbf{C}}_n(t) \tag{2.33}$$

Finally, an estimate of the pulse signal $p(t)$ is computed using the formula $p(t) = S_1(t) - \alpha \cdot S_2(t)$, where α is a parameter derived from the standard deviations of $S_1(t)$ and $S_2(t)$.

Notably, the algorithm strategically eliminates specular reflection by relying on illumination intensity for alpha tuning, thus, effectively minimising the impact of lighting conditions and enabling the precise estimation of pulse-induced variations in colour information.

Plane Orthogonal to Skin Algorithm

In contrast to Chrom, the POS algorithm [49] operates by creating a virtual plane in the temporally normalized RGB space that is orthogonal to the skin tone. This strategic placement facilitates the isolation and removal of reflections from the skin. POS defines its projection matrix $\mathbf{P}_{POS}$ in such way that it identifies two axes that directly compute in-phase signals:

$$\mathbf{P}_{POS} = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{pmatrix} \tag{2.34}$$

Based on this, it employs a two-step process to enhance signal quality. Initially, it eliminates distortions caused by light intensity, and subsequently, it utilizes specular reflections for alpha-tuning to improve the algorithm's performance. Hence, while POS is intentionally designed to be less affected by movement, it exhibits a higher sensitivity to variations in lighting conditions compared to the Chrom method.

To further refine the analysis, POS initially divides the input signal into overlapping subsequences to be processed, resulting in the synthesis of a final signal. This approach significantly contributes to improving the Signal-to-Noise Ratio (SNR), enhancing the overall robustness of the algorithm.

Chapter 3

Related Work

In recent years, increased attention is paid to new approaches on cuff-less BP measurement based on surrogate parameters. The most investigated methods consider the PWV by measuring PAT or PTT. However, with the rise of Machine Learning, also more complex models based on signal morphology were proposed. This chapter outlines the state-of-the-art related work on methods for optical BP estimation. Although there are many approaches on BP estimation from PPG and ECG signals, the work at hand focuses on methods that are only PPG based. This allows completely contactless BP estimation methods when measured by camera, whereas ECG requires multiple electrodes with skin contact.
The chapter is divided into contactless PTT-based approaches in Section 3.1, contactless feature-based approaches in Section 3.2 and contactless deep-learning-based methods in Section 3.3, followed by methods for remote photoplethysmography in Section 3.4 and BP estimation from PPG signals in Section 3.5.

3.1 PTT-based Blood Pressure Estimation from rPPG

PTT that is based on two different skin locations, i.e. the difference in arrival time of the pressure pulse between two distal locations, is also referred to as differential Pulse Transit Time (dPTT). Most methods employing dPTT are following camera-based approaches, since only a single sensor is required for both PPG signals. The rPPG (in literature also image PPG (iPPG) or camera PPG (cPPG)) is based on the same measuring principle as a pulse oxymeter by cap-

turing the reflected light from the skin tissue (see Section 2.2).

For dPTT calculation, a large distance between the two skin areas is beneficial. Previous studies often extract the signals from the face and hand [16, 51, 52, 53], since these locations are not covered normally. Though, Murakami et al. [54] take advantage of a larger distance between wrist and ankle. They obtain mean peak-to-peak dPTT values of 128.1 $\pm$ 6.2 ms for healthy subjects compared to 97.2 $\pm$ 10.6 ms for hypertensive subjects and attain a correlation coefficient $R_{SBP} = -0.88$ on SBP estimation. In contrast, Nakano et al. [16] obtain dPTTs between face and palm of 48 to 17 ms and a correlation coefficient $R_{SBP} = -0.79$ for SBP estimation with logarithmic regression.

Whereas Murakami et al. [54] and Nakano et al. [16] only apply a band-pass filter and then base the dPTT on the time difference between two PPG minima [16] or peaks [54], Fan et al. [51] reconstruct the true peak from the filtered signal. Due to the application of a strong filter, the dicrotic notch of the rPPG wave is often buried into the main peak and causes a so called peak shift. Hence, Fan et al. [51] construct an adaptive Gaussian model to obtain the true peak location. They show that linear regression performs better with the adaptive Gaussian model.

Khong et al. [55] perform a Haar wavelet decomposition to eliminate the high frequency noise and scale the measured chest-to-forehead peak-to-peak PTTs by the person's height. With a polynomial model, they obtain Pearson's correlation coefficients of $R_{SBP} = 0.89$ and $R_{DBP} = 0.82$.

Since there is no public data base available with video data and BP reference, all studies are conducted on their own data sets. These data sets are usually very small, ranging from five [16] to ten [54] subjects. Huang et al. [52] address this limit by enabling incorporation of the MIMIC data set [56] into the training process. The MIMIC data set is a large public database comprising biosignals from ICU patients such as ECG, PPG and a reference ABP signal. Huang et al. propose a transfer learning approach where the ABP signal of the MIMIC data is used as a proximal PPG signal and the MIMIC PPG as the distal signal, respectively. Transfer learning could reduce RMSE from 16.5 mmHg to 14.0 mmHg for SBP test data.

Jeong and Finkelstein [53], however, observed different slopes of the dPTT-vs.-SBP regression curve for their subjects, which they attribute to individual

characteristics of cardiovascular systems and pulse wave propagation. They conclude that individualised calibration procedures are required for such methods shifting the issue from the lack of data for universal models to individualised models.

Besides the problem of individual calibration, dPTT-based methods suffer from orthostatic dependencies. The relative positional relationship between the two skin areas needs to be maintained in order not to influence pulse propagation properties. Further, with a regression based on a single input value, only one linearly independent target value, i.e. SBP or DBP, can be estimated. Zhang et al. [57] showed that PTT predicts DBP and MAP significantly better that SBP.

3.2 Feature-based Blood Pressure Estimation from rPPG

Feature-based approaches consider various characteristics of the PPG morphology and are therefore able to create more complex models than the PTT-based methods. Table 3.1 gives a comparative overview of BP prediction methods from a single rPPG including feature-based approaches. Again, a major limitation is an impeded comparability of the studies due to the lack of public data sets. The proprietary data sets used instead exhibit large differences in size, number of subjects and BP range as well as recording frame rate and resolution.

Jain and Subramanyam [58] published one of the first feature-based estimation methods using only a camera for rPPG extraction. Features are extracted from the best part of the signal and include mostly time domain features and the dominant frequency. With a polynomial kernel regression on data of 45 normotensive subjects, they obtain a MAE of 3.9 $\pm$ 5.4 mmHg for SBP and 3.7 $\pm$ 5.1 mmHg for DBP.

Fang et al. [66] integrate a palm-to-cheek dPTT into a feature-based method. A total of 24 features are extracted from both signals and fed to a three-layered feed-forward ANN. For SBP, Fang et al. obtain a RMSE of 11.2 mmHg. By using all hand-crafted features as input to the ANN, also irrelevant information might be fed to the model, thus hindering the training process and limiting prediction performance.

Lou et al. [59] published a similar method in which 126 rPPG features are

Table 3.1: Related work on video-based BP prediction from a single rPPG area.

Reference	Input Type	Regression Model	Data Set	Result
Jain and Subramanyam, 2016 [58]	time domain features, dominant frequency	polynomial kernel regression	proprietary: 45 normotensive subjects	MAE: 3.9 ± 5.4 mmHg (SBP), 3.7 ± 5.1 mmHg (DBP)
Luo et al., 2019 [59]	126 time and frequency domain features; dPTTs between different facial regions; 29 external features	PCA, MLP	proprietary: 1328 normotensive subjects	ME: 0.4 ± 7.3 mmHg (SBP), -0.2 ± 6 mmHg (DBP)
Sugita et al., 2019 [20]	degree of distortion of rPPG			-
Tran et al., 2020 [60]	face and palm rPPG signals, 9 external features	MLP	proprietary: 164 videos from 82 subjects	ME: 3.1 ±7.3 mmHg (SBP), 2.6 ±7.5 mmHg (DBP)
Schrumpf et al., 2021 [61]	raw rPPG segments	adapted AlexNet [62] and ResNet [63]	MIMIC-III: 1.5 million PPG data samples from >4000 subjects, proprietary: videos of 25 subjects	MAE: 13.0 ±10.6 mmHg (SBP), 9.8 ±9.9 mmHg (DBP)
Wu et al., 2022 [64]	three facial rPPG signals, 7 external and signal features	MLP and CNN	proprietary: videos of 1143 subjects	MAE: 11.54 ±10.57 mmHg (SBP), 8.09 ±6.65 mmHg (DBP)
Zhuang et al., 2022 [65]	raw rPPG segments	adapted ResNet and Long Short-Term Memory (LSTM)	proprietary: videos of 124 subjects	MAE: 12.4 mmHg (SBP), 10.0 mmHg (DBP)

extracted such as pulse amplitude, shape and frequency related features. This further comprises dPTTs between different facial regions. Additionally, they include 29 external features of subjects' physical characteristics and ambient information such as room temperature and light intensity. Then, Lou et al. [59] employ PCA for dimensionality reduction producing 30 eigenvectors as input for an MLP. On a large dataset of 1328 normotensive subjects, they obtain a ME of 0.4 ± 7.3 mmHg ($R_{SBP} = 0.67$) for SBP and -0.2 ± 6 mmHg ($R_{DBP} = 0.47$) for DBP. However, the meta-data on physical characteristics and ambient information can be difficult to access limiting the practicality of the method. Further, a PCA reduces dimensionality by compressing all information into a smaller feature space. Hence, irrelevant information is still not eliminated from the dataset.

Another camera-based approach was proposed by Sugita et al. [20]. They calculate the degree of distortion of the rPPG signal as a time difference between

the minima of the raw and the band-pass filtered rPPG signal. This feature shows the highest correlation coefficient with BP when measured at the palm compared to forehead and cheek. However, the correlation exhibited large inter-subject differences.

In general, the studies on remote feature-based BP prediction define their own sets of hand-crafted features which are then used as model input without evaluation of feature relevance or redundancy. Especially since hand-crafted features are employed instead of an automatically learned feature representation, selection of these features with respect to model performance is necessary and might even provide new insights for an advanced waveform analysis.

3.3 Deep-Learning-based Blood Pressure Estimation from rPPG

Deep-Learning-based methods usually consist of deeper ANNs which take a segment of the rPPG signal as input and automatically extract the relevant features for BP estimation.

Tran et al. [60] use an MLP with 9 layers. Their method requires two rPPG signals, from face and palm, and 9 external subject-specific features. For BP estimation, they obtain a ME ± SD (RMSE/MAE) of 3.1 ±7.3 mmHg (7.9/6.6) for SBP and 2.6 ±7.5 mmHg (7.9/6.4) for DBP on a data set of 164 videos from 82 subjects. The authors do not describe the input data in more detail and do not specify if the selection of the test data guarantees a strict separation between training and test subjects. Further, the requirement of signals from face and hand reduces usability of the method and suffers from orthostatic dependencies.

Schrumpf et al. [61] adapted the AlexNet [62] and ResNet [63] architectures to the BP regression task to study the effect of signal length on the prediction accuracy. These analyses were performed on MIMIC-III PPG data from more than 4000 subjects. The input segment length did not influence the performance and they further showed that the use of the first and second derivative of the PPG signal does not significantly improve their predictions. Finally, the ANNs trained on PPG data are fine-tuned on rPPG signals of 25 subjects. ResNet obtained the overall best results with a MAE ± SD of 13.0 ±10.6 mmHg for SBP and 9.8 ±9.9 mmHg for DBP. With personalisation of the model by including test subjects'

data in the training set, the ResNet model improved by 0.5 and 1.29 mmHg for SBP and DBP, respectively.
In an extension of their study, Schrumpf et al. [67] investigated different data split configurations for personalisation and transfer learning from PPG to rPPG signals. With a random sample-based data split instead of a subject-based split, they obtain a significantly improved model performance, and suggest that some previous studies report overestimated prediction performance due to data leakage between training and test set. Further, they show that after training on PPG data, fine-tuning the model with rPPG data improves prediction accuracy for the video-based signals.

Wu et al. [64] employ the POS method [49] (see Section 2.5.4) for rPPG extraction from three ROIs on the forehead, cheek and nose. The three signals are band-pass filtered and serve as input to an ANN consisting of six residual blocks with 2D convolution layers. A second ANN with fully connected residual blocks is trained to predict BP values from 7 physiological indicators such as heart rate, PTT and BMI. Both models are combined to a single network with two auxiliary loss functions. To optimise the training process, Wu et al. [64] filter the rPPG signals with different cut-off frequencies for the bandpass filter (0.5-3 Hz, 0.5-5 Hz, 0.5-7 Hz and 0.5-10 Hz) obtaining signals with various amounts of useful information and noise. They show an improvement in prediction accuracy for their proposed model as well as for the models proposed by Schrumpf et al. [61] using this training data augmentation method. The authors tested their model on a data set containing video records from 1143 subjects with a mean SBP ($\pm$ SD) of 120.8 $\pm$17.7 mmHg and a mean DBP of 73.4 $\pm$11.3 mmHg and obtained a MAE $\pm$ SD of 11.54 $\pm$10.57 mmHg for SBP and 8.09 $\pm$6.65 mmHg for DBP.

Zhuang et al. [65] transfer the ROI data to the YUV colour model to mitigate signal contamination caused by external illumination changes. Then, the ROI is divided into several subregions in order to form a new spatio-temporal representation of the image sequence. To do so, the V-channel is spatially averaged for each subset of ROIs. Zhuang et al. designed a deep ANN with the ResNet architecture as backbone. Its outputs are fed to a subsequent LSTM obtaining features that are used for BP classification. The classification result is integrated with the final predictor of the BP values. The proposed method was trained and

tested on a data set of 124 subjects containing 1-minute videos at a frame rate of 30 fps. In order to deal with the unbalanced distribution of BP values in their data set, an oversampling strategy was adopted. With a signal length of $t = 15$s, they obtained a MAE of 12.4 mmHg for SBP and 10.0 mmHg for DBP during a 5-fold cross validation.

Deep-learning-based methods show promising results for BP estimation from video data as feature representations are learned automatically by the ANNs. However, the presented results should be considered carefully since many previous studies suffer from data leakage due to sample-based data splits and thus overestimate prediction performance. A common issue in many studies are the small sizes of the rPPG data sets as well as low rPPG signal quality. The transfer learning approach adopted by Schrumpf et al. [67] could mitigate the problem of small training data while providing high quality signals during the source domain training process.

3.4 Remote Photoplethysmography

Especially morphology-based methods rely on a clear PPG with well preserved signal morphology. However, rPPG captured by camera is very sensitive to illumination changes, movement and sensor noise. Therefore, various approaches have been proposed in order to extract and reconstruct the rPPG signal.
Wang et al. [49] describe the reflection of a skin pixel p in a RGB image sequence as

$$c_p(t) = Q(t) \cdot (b_s(t) + b_d(t) + b_n(t)) \tag{3.1}$$

where $c_p(t)$ is the column vector containing RGB values of the pth skin pixel. $Q(t)$ is the luminance intensity level which depends on the distances between light source, skin tissue and camera sensor, and it is modulated by specular reflection $b_s(t)$ and diffuse reflection $b_d(t)$. Both components are time dependent due to movement and pulsatile blood, respectively. $b_n(t)$ denotes quantisation noise of the image sensor. Whereas specular reflection usually is the largest component by far [49], only diffuse reflection is associated with the pulsatile blood volume changes in rPPG. The amount of haemoglobin and melanin in blood tissue lead to a specific chromaticity for $b_d(t)$, that is to be determined for rPPG extraction.

Most studies rely on spatial averaging of the green channel's pixel values [54, 53, 52, 20] which coincides with the absorption spectrum of haemoglobin and was proven to contain the strongest reflection pulsation (compare to Section 2.2.2) but does not specifically consider specular reflection and diffuse reflection. By spatially averaging a sufficiently high number of pixels, the quantisation error of the camera sensor $b_n(t)$ can be reduced.
Fan et al. [51] average the intensities in the red Q_r and the green channel Q_g over each frame, and normalise them by dividing them by their temporal average $\bar{Q}_r$ and $\bar{Q}_g$. Their rPPG is defined as

$$rPPG = \frac{Q_g \bar{Q}_g^{-1}}{Q_r \bar{Q}_r^{-1}} - 1 \tag{3.2}$$

The rPPG extraction task can be represented as a linear mixture model of a stationary signal, a movement induced signal and the pulse signal [49] which suggests the application of a Blind Source Separation (BSS) technique.
Oiwa et al. [68] spatially average each channels' pixels and then apply ICA to R, G and B channel signals for BSS. The independent component with the highest power between 0.7 and 2.0 Hz in the frequency spectrum was chosen as rPPG. These methods, however, cannot deal with periodic motions, e.g. during exercising, as they rely on determining the rPPG signal as the most periodic ICA component. Jain and Subramanyam [58] use the red channel for spatio-temporal BSS. Red channel information of all frames in the video is transformed and stacked, such that a matrix $\mathbf{A} \in \mathbb{R}^{U \times K}$ is obtained, where U is the total number of frames and K is the number of pixels in the Region of Interest (ROI). A PCA is applied to matrix $\mathbf{A}$. They obtain the rPPG signal as the reconstruction error between $\mathbf{A}$ and the reconstructed matrix $\mathbf{A}'$.
Lou et al. [59] transformed the RGB channels, which are encoded in an 8-bit colour stack, into 8 separate bit layers. For pulse signal extraction, they isolate layers where the bits fluctuate along with the reference BP.

In contrast to the BSS-based techniques, model-based methods [50, 49, 16] apply knowledge about the colour vectors and signal components in order to control the demixing. By temporal normalisation of the RGB signals, the average skin reflection colour, i.e. specular reflection, is eliminated while preserving the

signal's AC components. Nakano et al. [16] transform RGB components into skin chromophore concentrations: melanin concentration c_m, oxygenated blood concentration c_{ob} and de-oxygenated blood concentration c_{dob}. For this purpose, RGB values are transformed into CIE XYZ tristimulus values under standard illuminant D_{65} by using a multiple regression analysis to determine the RGB-to-CIE XYZ transformation matrix. Next, they transform these CIE XYZ values into the skin chromophore concentrations by computing a second transformation matrix considering a Monte Carlo simulation of light-skin tissue interaction. From this, the total blood volume is obtained as the sum of c_{ob} and c_{dob}.

The chrominance-based method Chrom [50] introduced in Section 2.5.4 is restricting all colour variations to the pulsatile direction by defining its chrominance signals as a projection of $\tilde{\mathbf{C}}_n(t)$ onto the plane orthogonal to the variation direction of specular reflection. To be independent of the colour of illumination, a standardised skin-tone vector $\mathbf{u}_{skin} = \mathbf{M}^{-1} = [0.77, 0.51, 0.38]^{\mathrm{T}}$ is employed for white balancing the images, with $\mathbf{M}$ denoting the diagonal mapping matrix for $\tilde{\mathbf{C}}_n(t)$. Further, Chrom uses a 2×3 projection matrix $\mathbf{P}_{Chrom}$ (see Eq. 2.32) to define a plane in the temporally normalised RGB space. As a result, two projected signals $\mathrm{S}(t)$ are created, where motion-induced and pulse-induced variations appear in anti-phase (see Eq. 2.33) These properties can be used to compute an estimate of the pulse signal $\hat{p}(t)$ as

$$\hat{p}(t) = S_1(t) - \alpha \cdot S_2(t) \tag{3.3}$$

with $\alpha = \frac{\sigma(S_1)}{\sigma(S_2)}$. $\hat{p}(t)$ is a good estimate for the pulse signal as long as standard deviation $\sigma(S_1) \neq \sigma(S_2)$.

Similarly to Chrom, the POS algorithm [49] (detailed in Section 2.5.4) creates a plane orthogonal to the skin tone in a temporally normalised RGB space to eliminate specular reflection $b_s(t)$. One difference compared to Chrom is the projection matrix $\mathbf{P}_{POS}$ (see Eq. 2.34), as POS finds two projection axes that directly compute in-phase signals. Whereas Chrom removes specular reflection $v_s(t)$ and uses intensity for alpha tuning, POS first eliminates intensity distortions and uses specular reflections for alpha tuning. That makes Chrom more vulnerable to changes in direction of specular reflection, i.e. movement, and POS more sensitive to inhomogeneous illumination spectra. Furthermore, the input

signal of POS is divided into sub-sequences such that the final signal is obtained by overlapping of these shorter segments and exhibits an improved SNR.

Wang et al. [49] compared various rPPG methods such as the green RGB channel, PCA, ICA, Chrom and POS in terms of SNR and obtain the overall best performance for the POS method across different skin tones, changing luminance and movement. Also Fang et al. [66] compared the green channel to POS, Chrom and motion resistant rPPG [69], where POS yielded the best SNR. Moreover, several comparative analyses of rPPG methods have been conducted for heart rate estimation [70, 71, 72]. Tsouri and Li [70] and Ernst et al. [71] compared channels of seven and eight different colour models, respectively, and both found the best heart rate estimation accuracy for the hue channel of the HSV model. Boccignone et al. [72] consider the same rPPG extraction methods as Wang et al. [49] across 15 datasets and showed that performances of POS, Chrom, PCA and Spatial Subspace Rotation (SSR) [73] are statistically not different.

For BP estimation from rPPG data, preservation of a more detailed signal morphology is necessary than for heart rate measurement. Therefore, it should be considered that the proposed rPPG extraction methods and colour models might perform differently for this use case, and further experiments with respect to BP estimation should be conducted.

3.5 Blood Pressure Estimation from PPG

Whereas BP estimation from video data still remains a challenging task, there have been some advances in BP prediction from conventional PPG signals. However, many studies suffer from methodological drawbacks such as data leakage [74] or unspecified sample-based data splits [75, 76, 77, 78, 79, 80]. Some remaining publications present approaches that incorporate a calibration procedure to overcome the generalisation issues of universal regression models.

Haddad et al. [81] perform an individual calibration of their Multiple Linear Regression (MLR) model by computing the offset between the predicted BP of the calibration signal with the true BP and explicitly adding this offset value to all subsequent predictions of the same subject.

Schlesinger et al. [82] perform personalisation by employing a Siamese network architecture, where one network predicts feature values from the current PPG

signal and the other one computes calibration values from the patient's example PPG. With an MAE ±SD of 6.0 ±6.7 mmHg for SBP and 3.4 ±4.0 mmHg for DBP they show improved results compared to the same base network without calibration. However, their approach requires a 30-second signal in form of a spectrogram and therefore, does not allow a beat-to-beat measurement.

Zhang et al. [83] developed a Domain-Adversarial Neural Network (DANN) which uses an adversarial training approach to extract domain-invariant features from bioimpedance sensors. The model is personalised for each patient using 5 minutes of beat-to-beat data, and obtains an RMSE of 6.8 mmHg for SBP and 4.5 mmHg for DBP. Even though the final results are not comparable due to the narrow BP range in their proprietary data set, Zhang et al. showed that an DANN approach can overcome overfitting for models with very few training subjects.

Qin et al. [84] and Ibtehaz et al. [85] both proposed an ANN for signal translation from PPG to ABP waveforms. Qin et al. [84] employ domain-adversarial training for an auto-encoder network which is subsequently fine-tuned for individual patients with one quarter of the patients test signal. Their results show that this training strategy only marginally improves generalisation capabilities of the model whereas fine-tuning for individual calibration can further reduce MAE by -2.5 mmHg to 5.4 mmHg for SBP and by -1.0 mmHg to 3.1 mmHg for DBP. Ibtehaz et al. [85] presented a two-stage model consisting of two U-nets, where the first one approximates the BP signal and the second one refines its waveform. Using 8-second signals, they obtain an MAE ±SD of 5.7 ±9.2 mmHg for SBP and 3.4 ±6.1 mmHg for DBP. Their data split remains unclear since in an earlier arXiv preprint version [86] of their paper with exactly the same results, a random sample-based split is described whereas in the newer publication, they claim to split the sequentially ordered data set in such way that training and test patients remain separate.

These PPG-based models show significant improvements in prediction performance after individual model calibration. The most promising method is fine tuning with data of the new subject. However, while all studies compare their overall results to the AAMI and BSH standards, more extensive analyses of prediction errors for different BP levels are missing to evaluate the personalisation results.

3.6 Research Gaps

This review of related work shows that there has been substantial research on PAT and PTT-based BP estimation, however, these studies revealed several drawbacks of this approach, particularly the generalisation problem of its models. In contrast, feature-based and deep-learning-based methods allow more complex models since they are considering the entire rPPG/PPG waveform. Whereas studies on rPPG data are difficult to evaluate and compare due to large differences in data sets regarding size, BP range and video specifications, comparable feature-based and deep-learning-based approaches for conventional PPG signals show promising results on public data sets. However, in both domains, most published studies suffer from data leakage and unsuitable sample-based data splits leading to overestimated model performances [67]. Hence, subject-based data splits are required to conduct fair and realistic model evaluations that reflect prediction accuracy for real-world applications. Considering this data handling strategy, in the work at hand, the following research gaps will be addressed.

Can feature selection improve BP prediction and provide a deeper understanding of relevant rPPG characteristics? What are relevant rPPG features and waveform characteristics for BP estimation?
Previous remote feature-based methods use the entire hand-crafted feature set as input to the model, thus not eliminating any irrelevant information and hampering the training process. In addition to improving prediction performance, a feature selection could also provide insight into relevant rPPG waveform characteristics and interrelations with SBP and DBP.

Which method and location for rPPG extraction can enhance rPPG signal quality for BP estimation?
For contactless BP estimation, the rPPG measurement still poses a large problem, since measuring from a distance degrades the PPG signal significantly. Due to illumination and movement artifacts as well as sensor noise, additional signal processing steps are required and regression methods need to be adapted to the amount of information which is preserved from the original PPG. Whereas several rPPG extraction methods have been proposed, extensive comparisons of

these algorithms have been performed only for heart rate estimation. Further, these studies consider either only colour model transformations or only specific rPPG extraction methods. As BP estimation relies on more details of the signal waveform, performances of both the rPPG models and colour model transformations need to be evaluated for BP specifically. An additional aspect in rPPG measurement that has not been investigated with respect to BP prediction performance is the optimal skin location from which the signal is to be extracted.

Is fine tuning for model personalisation also applicable in the rPPG domain and can it improve prediction results of beat-to-beat data?
Another unresolved issue in BP estimation from rPPG and PPG data are significant inter-individual differences in signal morphology [12, 20, 77, 87]. Therefore, universal models still do not reach medical standards and some PPG-based methods try to mitigate these morphological differences by integrating an individual personalisation process. Especially fine-tuning for personalisation shows improved prediction performance compared to general models, but the number of studies on this topic is still very small and it has not yet been applied to rPPG models or beat-to-beat data. Hence, it needs to be evaluated if beat-to-beat information and rPPG signal quality suffice to derive the relevant individual characteristics for personalisation from a small data subset. An analysis of inter- and intra-individual waveform correlations could be helpful to understand the results of model personalisation.

Can large public PPG data bases enable more complex rPPG-based models by providing a suitable source domain for transfer learning? Does pre-training on PPG data improve model convergence for rPPG-based BP estimation?
Due to the lack of a publicly available rPPG data set, a common limitation of self-recorded rPPG data sets is the small sample size. However, the MIMIC III waveform data base [56] provides a large number of records including PPG and ABP signals. Despite significant morphological differences between rPPG and PPG data [88], both signals capture the same blood volume changes. Thus, their basic behaviour with respect to BP is comparable, and the larger PPG data base could be employed as source domain for transfer learning from PPG to rPPG.

This might overcome limitations of small sample sizes and enable more complex models.

Chapter 4

Feature-based Blood Pressure Measurement

The first approach of this thesis on contact-less BP prediction is based on hand-crafted features that are extracted from the rPPG signal. Hand-crafted features are beneficial to reduce input size to machine learning models, especially if a large input signal contains few useful information, or if the relevant characteristics of the input are known. Since the waveforms of PPG and BP are closely related as established in Section 2.2.1, and BP signal morphology changes with the pressure level (see Section 2.1.3), rPPG waveform morphology is expected to contain useful information for deduction of the current BP. Under this condition, feature selection of the hand-crafted features can then provide further insight into the BP dependent characteristics of rPPG waveforms.
In Section 4.1, the proposed feature-based method is outlined, followed by detailed descriptions of each processing step and its analyses in Sections 4.1.1 to 4.1.5. Subsequently, in Section 4.2, the self-recorded rPPG data set is described, and in Section 4.3, results of the different BP prediction models are analysed and discussed.

4.1 Feature-based Method

The pipeline of the proposed method is visualised in Figure 4.1. It starts with an image processing part to obtain the rPPG signal from the video. To do so, the signal is extracted from skin pixels whose position is determined by face (or

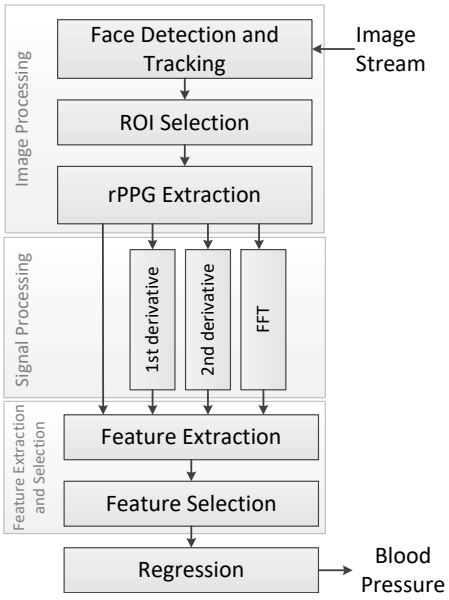


Figure 4.1: Complete pipeline for feature-based contact-less BP estimation. [89]

hand) detection and tracking. Next, the pixels' colour information is spatially averaged and further processed in order to obtain an rPPG signal over time. After rPPG extraction, BP is to be predicted from the signal waveform. For this, 120 morphological features are computed from the rPPG signal in the time and frequency domain as input candidates for the regression model. Since not all of these features might contain relevant information for BP estimation, feature selection is performed. Finally, a regression model is trained to predict BP based on the selected features. These processing steps are described in more detail in the following Sections.

4.1.1 Face/Hand Detection and Tracking

rPPG information can be extracted from image pixels which are associated with the skin of the subject. To determine which pixels are, face and hand detection algorithms are employed. Hand and face are the two skin regions which are most commonly exposed and therefore suitable for this task. Once the object is detected, tracking of this region instead of re-detection might improve processing time and performance.

As the rPPG signal is obtained by spatial averaging of the determined skin region, it is important for the detection algorithm to be accurate and for the object

tracking or re-detection not to exhibit large displacements. Strong shifts in the determined skin region could create artifacts due to different illumination of this area or overlap with non-skin pixels such as background, eyes or hair. Therefore, we compare three pre-trained face detection and five hand detection algorithms in terms of processing time, detection accuracy and detection failures, as well as five tracking algorithms in terms of processing time, stability and tracking failures on our specific task. For detection, we consider the Viola-Jones face detection [37], HOG [38], YOLO [39], Hand Landmarks [41] and SSD [40], and for tracking, we compare Median Flow [43], MOSSE [45], KCF [46], MIL [42] and CSRT [47] which are described in Sections 2.5.1 and 2.5.2.

Experimental Setup for Detection and Tracking Assessment

For evaluation of these detection and tracking algorithms, we recorded a total of 80 videos of five subjects (three female, two male). Eight videos per subject were captured by an Allied Vision Manta G-201-30fps RGB camera and further 8 videos were recorded in the NIR spectrum with an Allied Vision Alvium 1800 U-501m NIR. For both cameras, the resolution was set to 352 x 198 pixels. Whereas the RGB videos were taken in a naturally illuminated room, for the NIR videos, we mounted two 850 nm lamps in a dark room with 1.5 m distance to the subject. The subjects were seated in front of the camera in the same way the BP measurement would be conducted. They were asked to sit still and raise the right hand next to their heads with the palm facing the camera as shown in Figure 4.4.

For evaluation of the detection algorithms, we used 30 frames per recording which results in 1200 frames per image type. For each frame, a ground truth bounding box GT for face and hand was defined manually, to which the algorithm's bounding boxes BB are compared using the Jaccard coefficient J which measures the Intersection over Union (IoU) [90].

$$J(GT, BB) = \frac{|GT \cap BB|}{|GT \cup BB|} \tag{4.1}$$

The tracking algorithms were evaluated on 150 frames per recording resulting in 6000 frames per image type. For the first frame, the ground truth bounding boxes for hand and face are given as the regions to be tracked. The tracking stability is calculated as the Jaccard coefficient between the new and the previous bounding box. Since subjects are sitting still, the tracked bounding box should

rarely move as well. Further, we introduced one video of 150 frames at a frame rate of 5 fps, where the subject was moving randomly in all directions while face and palm were facing the camera at all times. For this video, tracking stability is evaluated against a manually defined ground truth bounding box in each frame.

In the following, face and hand detection algorithms are compared in terms of processing time $\bar{t}$, detection accuracy J_d and detection failures e_d, as well as the tracking algorithms in terms of processing time $\bar{t}$, tracking stability J_t and tracking failures e_t on the specific task. For this, RGB and monochrome (NIR) images are evaluated separately.

Assessment of Object Detection Algorithms

Table 4.1 for face detection and Table 4.2 for hand detection detail the average and Standard Deviation (SD) for each of the evaluation criteria. For both RGB and NIR images, the SSD algorithm [40] obtains the highest face detection accuracy with $\overline{J}_d = 85.1\%$ and $\overline{J}_d = 86.8\%$. However, it has to be noted, that for NIR images, SSD exhibits a high failure rate of $e_d = 75.6\%$. There are no face detection failures for the RGB images in our setup, but all three algorithms show difficulties performing face detection on the monochromatic NIR images since these models are trained on the three RGB colour channels.

For RGB images, HOG [38] exhibits the lowest processing time with $\bar{t} = 16.7$ ms followed by the Viola-Jones method [37] which almost doubles the time. In contrast, for NIR images, HOG takes the longest processing time while Viola-Jones performs best with $\bar{t} = 22.5$ ms.

Since hand detection always fails on the NIR images with the Viola-Jones and SSD algorithms, additional two models are evaluated for hand detection. Whereas YOLO [39] is not able to detect the hands in the NIR images either, the hand landmarks [41] succeed in each frame with a detection accuracy of $\overline{J}_d = 71.8\%$. For RGB images, hand landmarks perform best as well and obtain an accuracy of $\overline{J}_d = 84.2\%$ without any failures. However, this algorithm takes over $\bar{t} = 7$s for processing a single frame and thus, is not suitable for any real-time prediction task.

Small variations in detection accuracy can be denoted to different sizes of bounding boxes. Since the predicted bounding boxes are compared to the ground

Table 4.1: Evaluation results of the face detection algorithms

Detection Algorithm	**Image type**	**Processing Time [ms]**		**Detection Accuracy [%]**		**Detection Failures [%]**
		$\overline{t}$	SD_t	$\overline{J}_d$	SD_J	e_d
Viola-Jones [37]	RGB	32.2	5.6	75.31	4.64	**0**
	NIR	**22.5**	5.2	57.89	5.74	66.3
HOG [38]	RGB	**16.7**	0.5	70.50	5.88	**0**
	NIR	82.2	2.9	67.37	5.90	**11.1**
SSD [40]	RGB	69.5	3.2	**85.09**	3.42	**0**
	NIR	73.5	8.5	**86.79**	3.85	75.6

truth box using the Jaccard coefficient, a prediction with the exact same centre as the ground truth might not reach the highest value of $J_d = 100\%$ if it is sized differently. Bounding boxes by the Viola-Jones face detection model tend to be bigger, whereas those by the HOG face detection model are comparatively small. For the hand detection models of these two algorithms, the effect is reversed. The Viola-Jones bounding box only spans the fingers and half of the palm, and the HOG bounding box contains a lot of free space above and below the hands in many cases. Both are not ideal for the next step of defining the ROI since their properties make the relative positioning of the ROI sensitive to the hand's pose and rotation.

Assessment of Object Tracking Algorithms

Table 4.3 details the average and SD for each of the tracking evaluation criteria. All tested algorithms exhibit an acceptable tracking stability of $\overline{J}_t > 98\%$ and tracking failures only occur occasionally for Median Flow [43] and KCF [46] on NIR images. The high values for tracking stability indicate that there are no abrupt displacements or changes in position of the bounding box. Although, this evaluation method (except for the video with movement) does not consider if the tracking algorithm follows the target object correctly, we assume that the object to be tracked is not changing position since the subjects are sitting relatively still. Hence, a stable tracking reduces the risk of a corrupted rPPG due to tracking inaccuracies. Despite the static measurement scenario, one video with substantial

Table 4.2: Evaluation results of the hand detection algorithms

Detection Algorithm	**Image type**	**Processing Time [ms]**		**Detection Accuracy [%]**		**Detection Failures [%]**
		$\overline{t}$	SD_t	$\overline{J}_d$	SD_J	e_d
Viola-Jones [37]	RGB	**5.0**	0.7	53.94	9.23	16.7
	NIR	**13.5**	1.2	-	-	100
HOG [38]	RGB	14.6	0.6	65.95	10.17	13.7
	NIR	72.5	2.1	66.15	10.01	7.7
SSD [40]	RGB	52.8	3.1	69.85	9.09	**0**
	NIR	56.5	6.6	-	-	100
YOLO [39]	RGB	282.9	15.1	59.55	13.53	6.3
	NIR	332.5	36.6	-	-	100
Hand Landmarks [41]	RGB	7666.6	290.2	**84.24**	5.12	**0**
	NIR	7028.2	227.3	**71.79**	24.04	**0**

movement and per-frame ground truth annotations was included to account for movement robustness. Moreover, to rule out a drift in the tracking bounding box, a visual inspection of the last frame of the image series is conducted. None of the compared algorithms exhibited a noticeable drifting. The largest differences between the algorithms can be found in terms of processing time which for RGB images ranges from $\overline{t}$ = 1.6ms with MOSSE [45] up to $\overline{t}$ = 128.0ms with MIL [42].

Selection of Object Detection and Tracking Algorithms

Based on the presented evaluation results, we can select the hand and face detection and tracking algorithm best suited for our setup. For face detection on RGB images, SSD has proven to be the most precise and without failures. It takes longer to process the image, but detection is needed only in the first frame and after a tracking failure, whereas for subsequent frames, object tracking takes over. Considering the high failure rate of the SSD model on NIR images, HOG will be employed for face detection in NIR videos.

For hand detection, hand landmarks perform best with respect to accuracy and detection failures. However, due to its extremely long processing time of $\overline{t}$ = 7.35s, we select SSD for RGB images and HOG for NIR images instead.

Finally, MOSSE will be employed for object tracking since it is fast with $\overline{t}$ =

Table 4.3: Evaluation results of the tracking algorithms

Tracking Algorithm	**Image type**	**Processing Time [ms]**		**Tracking Stability [%]**		**Tracking Failures [%]**
		$\overline{t}$	SD_t	$\overline{J}_t$	SD_J	e_t
Median Flow [43]	RGB	1.7	0.1	99.78	0.04	**0**
	NIR	**4.7**	0.4	99.30	0.25	8.2
MOSSE [45]	RGB	**1.6**	0.2	**99.98**	0.01	**0**
	NIR	9.2	1.3	**99.99**	0.01	**0**
KCF [46]	RGB	15.2	6.8	98.12	1.38	**0**
	NIR	169.9	143.3	99.39	0.55	1.6
MIL [42]	RGB	128.0	4.5	99.73	0.27	**0**
	NIR	155.2	29.9	99.88	0.10	**0**
CSRT [47]	RGB	58.6	4.0	98.59	0.70	**0**
	NIR	133.1	26.3	98.37	0.63	**0**

5.4ms and exhibits a high tracking stability without any failures on both RGB and NIR videos. Based on the tracking evaluation results, we can assume that artifacts and noise in the rPPG signal are not caused by tracking inaccuracies but originate from subsequent processing steps or external influences. Such influencing factors and the effect of difficult measurement conditions on the quality of the rPPG signal will be investigated in Chapter 5.

4.1.2 Definition of the Region of Interest

Based on the bounding boxes obtained by SSD and MOSSE, we define the ROI from which the rPPG will be extracted. The ROI is shifted and sized relatively to the bounding box ensuring the right size and position of the ROI independently of the patient's size or distance to the camera. On the one hand, the ROI has to contain only skin pixels and should be universally applicable to each subject. On the other hand, the ROI should be as large as possible to reduce the effect of the camera's sensor noise.

The ROI is defined by the position of the upper left corner (x_{ROI}, y_{ROI}) and its width w_{ROI} and height h_{ROI}. Equivalently, the bounding box is given by the position of the upper left corner (x_{BB}, y_{BB}) and its width w_{BB} and height h_{BB}.

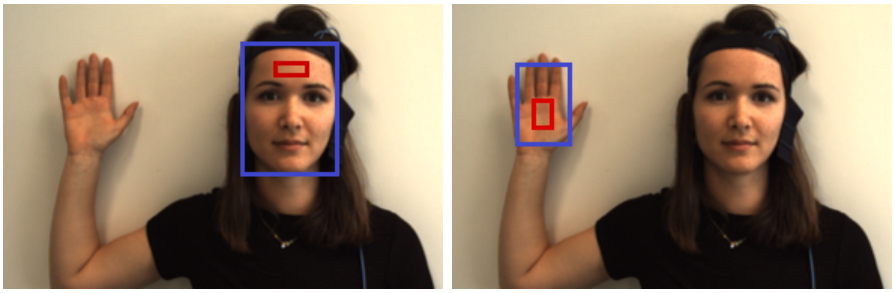


Figure 4.2: The ROI in red is computed relatively to the position and size of the bounding box in blue.

Empirical analysis led to the following definition of the facial ROI:

$$x_{ROI,face} = x_{BB,face} + (0.35 \cdot w_{BB,face}) \tag{4.2a}$$

$$y_{ROI,face} = y_{BB,face} + (0.15 \cdot h_{BB,face}) \tag{4.2b}$$

$$h_{ROI,face} = 0.3 \cdot h_{BB,face} \tag{4.2c}$$

$$w_{ROI,face} = 0.1 \cdot w_{BB,face} \tag{4.2d}$$

And the ROI on the palm is defined as

$$x_{ROI,palm} = x_{BB,hand} + (0.35 \cdot w_{BB,hand}) \tag{4.3a}$$

$$y_{ROI,palm} = y_{BB,hand} + (0.5 \cdot h_{BB,hand}) \tag{4.3b}$$

$$h_{ROI,palm} = 0.3 \cdot h_{BB,hand} \tag{4.3c}$$

$$w_{ROI,palm} = 0.3 \cdot w_{BB,hand} \tag{4.3d}$$

An example of the resulting ROIs for forehead and palm is given in Figure 4.2 where the bounding boxes are depicted in blue and the corresponding ROIs in red.

4.1.3 Signal Extraction and Pre-processing

Once the position of the ROI has been computed, the rPPG signal is extracted from the green channel of the ROI pixels, since for haemoglobin, the absorption of the light peaks in the green spectrum (see Section 2.2.2). For this reason, the signal is obtained as the average pixel intensity $\hat{Q}_G(t)$ of the green channel G of all K pixels i within the ROI over time t:

$$\hat{Q}_G(t) = \frac{1}{K} \sum_{i \in ROI} I_{G,i}(t) \tag{4.4}$$

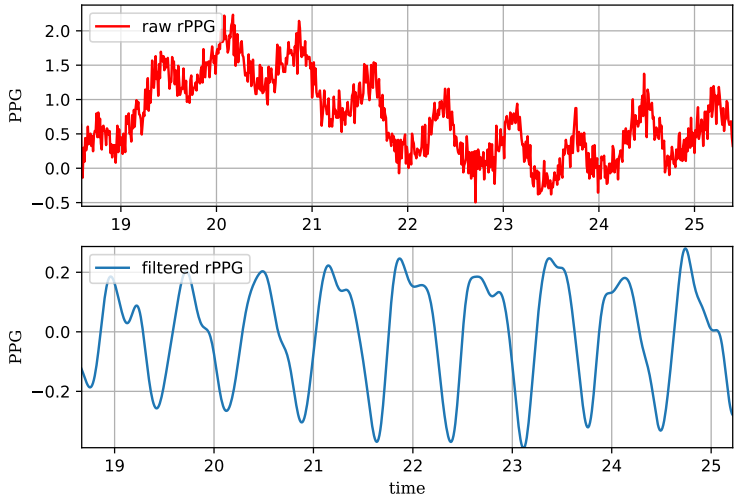


Figure 4.3: The upper plot shows the raw rPPG signal while the lower one depicts the filtered and inverted signal which serves as the basis for subsequent processing steps. [89]

The obtained raw rPPG signal contains a lot of noise and the blood volume pulses are modulated by the respiratory signal as shown in Figure 4.3. To reduce noise and level the signal, we apply an FIR band-pass filter with filter order $k = 64$. The lower cut-off frequency is set to $f_{low} = 0.3$ Hz such that the respiration signal up to 20 breaths per minute is eliminated, and the upper cut-off frequency of $f_{high} = 6.0$ Hz removes high-frequency noise while preserving the rPPG waveform.

4.1.4 Feature Extraction and Selection

To reduce the ANN input size and gain insight into BP related characteristics of the rPPG waveform, a total of 120 hand-crafted features are defined to describe the rPPG morphology. For time domain features, feature extraction is performed for each rPPG cycle, whereas for frequency domain features, 5-second segments are considered. Cycle segmentation is based on the steepest points of the systolic up-slopes, which are determined by the maxima positions of the first derivative rPPG'. From the dominant frequency f_d in the Fourier-transformed signal, a sliding window of size $1.5/f_d$ ensures that only the largest maxima in the neighbourhoods are kept. Thus, detection of diastolic peaks or potential double peaks

are suppressed. Finally, the minimum before the selected systolic rise is defined as the start of a cycle and the following maximum as the systolic peak, respectively. The dicrotic notch is identified as the last minimum previous to the beginning of the next cycle. Alternatively, if no minimum is found for inflection type dicrotic notches, it is determined as last maximum in the second derivative of the rPPG signal. For further processing, only those cycles are chosen that exhibit a dicrotic notch and whose length is between $0.67/f_d$ and $1.67/f_d$, in order to eliminate corrupted pulses. The values are chosen empirically.

After cycle segmentation, features are extracted. The defined features can be categorised into pulse widths, amplitudes, areas under the curve, FFT magnitudes, maximum power frequencies and body features. Pulse widths comprise the systolic and diastolic width, the width of the whole cycle and ratios between those. Moreover, systolic and diastolic pulse widths at various heights between the minimum and the systolic peak are considered, as well as their ratios. Amplitudes and their ratios are computed from the rPPG signal as well as from its first and second derivative. The areas under the curve are defined as the integral of the signal for the systolic and diastolic area, the whole cycle and ratios between them.

Frequency domain features include amplitudes in the signal's frequency spectrum up to 4.6 Hz and the frequencies with the highest spectral densities in descending order. Finally, body features are retrieved as well, such as the subject's sex, height, weight and age. The complete list of features and their descriptions can be found in Table A.1.

Since not all of the defined rPPG features contribute relevant information for BP estimation, a sequential forward feature selection is performed. In the process, SBP and DBP are considered separately for training single-output regression models, but also to gain insight into overlaps or differences in relevant rPPG waveform characteristics.

4.1.5 Regression Models

Firstly, joined regression models with two outputs are compared to separate single-output models for SBP and DBP. Additionally, MLP and RFR are considered as regression methods for this feature-based approach. In a grid search for the SBP single-output MLP, we obtain a model of 25 inputs and four hidden

layers with two layers of 500 neurons followed by two layers of 100 neurons, and for DBP, the same architecture with 22 inputs performed best. The RFR models have a maximal depth of 200 nodes and are limited to 25 features and 100 estimators.

All models are trained and tested in a nested Leave-One-Subject-Out Cross Validation (LOSO CV). To do so, one subject's data is reserved as test data. The remaining data is used for feature selection, where for each number of selected features, grid search is performed with a LOSO CV. The best hyper-parameter combination from the inner LOSO CV is subsequently evaluated on the test data of the outer loop.

4.2 Remote PPG Data Set

For developing methods on remote optical BP estimation, a large data set is required comprising videos of the patients' skin and reference BP. However, no relevant public data sets are known at the time of project initiation. Each research group conducts studies on their own recorded data set, but publication of this data is usually not possible due to privacy issues. Therefore, a new data set was acquired for the work at hand containing videos and reference measurements. This section outlines the experimental setup and data acquisition and subsequently details the statistics of the obtained data set.

4.2.1 Experimental Setup and Data Acquisition

For training and testing of the proposed methods, a data set was created containing video data and simultaneously recorded PPG reference signals as well as reference BP values. Overall, a total of 118 videos of 30 subjects were acquired. The videos were captured using an Allied Vision Manta G-040C RGB camera, with an exposure time of 4 ms and a resolution of 600×388 pixels. Each video has a duration of 20 s, sampled at a rate of $f_s = 240$ fps. The measurement setting was artificially illuminated at approximately 1100 lux. During the recording of the videos, the subjects were seated in front of a plain wall at 1.5 m distance to the camera, while holding their right hands next to their faces. Figure 4.4 depicts a single frame from a video from the data set.

Figure 4.4: Example frame of the recorded videos.

Before and after every recording, reference BP was measured using a sphygmomanometer boso medicus X by BOSCH+SOHN GmbH. This sphygmomanometer has been clinically validated according to protocol by the European Society of Hypertension (ESH) and exhibits a SD of ± 3 mmHg [91]. Further, simultaneously to the videos, reference PPG signals were recorded from the finger and from the forehead of the subjects. For this, a NeXus-10 MKII by Mind Media BV was employed which samples the signals at a rate of $f_s = 128$ Hz.

All subjects (17 male, 13 female) were healthy and without diagnosed (pre-)hypertension. The represented age groups range from 20 to 61 years with an average $\pm$ SD of 29.0 ± 5.4 years. Additionally, the subjects' weight and height were queried. Age, weight and height values were quantified in steps of 5, 5 and 10, respectively, in order to reduce the risk of ANN models recognising individual subjects, i.e. overfitting.

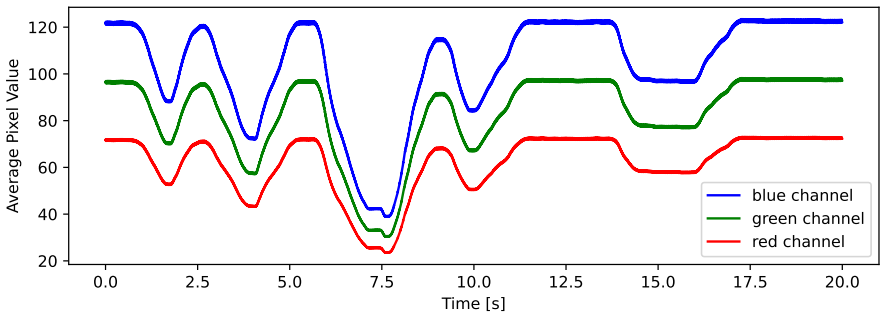


Figure 4.5: Spatially averaged pixel values of an example video with illumination changes.

Each subject has undergone the measurement four times, where the first iteration consisted of a regular measurement, and in the second one, the aperture of the camera was manipulated in order to simulate illumination changes. The

illumination changes can be characterised by an average overall intensity change in pixel values of −37% with intensity reductions of up to 57%. Figure 4.5 visualises the spatially averaged pixel values of an exemplary video with illumination changes. To achieve a wider range of BP values, in the third measurement, the subjects were instructed to focus on deep and slow breathing. These breathing exercises have been shown to lower BP immediately by an average of 6.4 mmHg for normotensives and 9.6 mmHg for hypertensives [92]. Moreover, prior to the final measurements, the subjects had to do physical exercise (1 min plank or squats) to increase BP.

4.2.2 Data Set Statistics

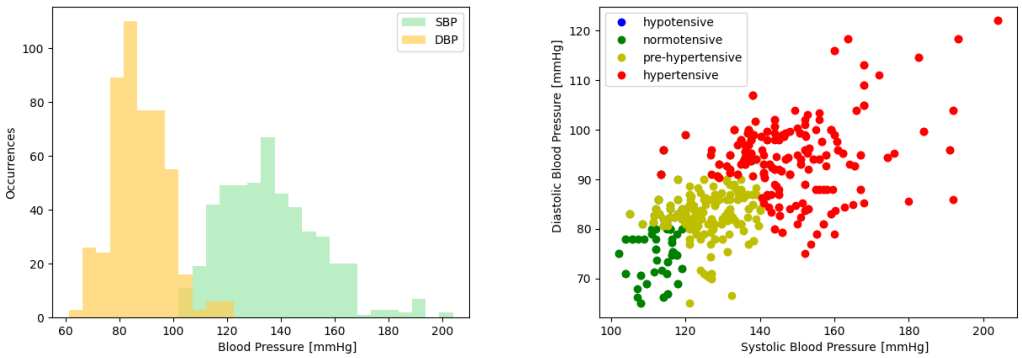


Figure 4.6: Distribution of measured BP values in the recorded data set. Hypotensive BP (SBP $<$ 90 mmHg and DBP $<$ 60 mmHg) is depicted in blue, normotensive BP (90 mmHg $\leqslant$ SBP $<$ 120 mmHg and 60 mmHg $\leqslant$ DBP $<$ 80 mmHg) in green, pre-hypertensive BP (120 mmHg $\leqslant$ SBP $<$ 140 mmHg and 80 mmHg $\leqslant$ DBP $<$ 90 mmHg) in yellow and hypertensive BP (140 mmHg $\leqslant$ SBP and 90 mmHg $\leqslant$ DBP) in red.

As a result, the recorded dataset includes DBP values from 65 to 122 mmHg with an average ± SD of 87.3 ± 9.0 mmHg and SBP values from 102 to 204 mmHg with an average ± SD of 136.0 ± 14.7 mmHg, respectively. Figure 4.6 shows the distribution of SBP and DBP in the data set. The data set does not comprise any hypotensive BP sample but exhibits 9.5% normotensive records, 39.7% pre-hypertensive records and 50.7% hypertensive records. The Pearson correlation

between systolic and diastolic pressure is $R = 0.77$. Both, SBP and DBP appear to be strongly correlated with sex ($R_{SBP} = 0.54$ and $R_{DBP} = 0.47$), height ($R_{SBP} = 0.49$ and $R_{DBP} = 0.46$) and weight ($R_{SBP} = 0.48$ and $R_{DBP} = 0.46$), however, to a lesser extend with Body Mass Index (BMI) ($R_{SBP} = 0.30$ and $R_{DBP} = 0.30$).

4.3 Results and Discussion

Based on this data set, in the following, results of the feature selection are given and analysed in Section 4.3.1 and performances of the regression models for BP estimation are assessed and discussed in Section 4.3.2.

4.3.1 rPPG Features Selection Results

Feature selection results can provide insight into the relevant characteristics of rPPG signals based on ranking and order of the selected hand-crafted features. The SFS for rPPG features is performed in a LOSO CV , and the number of selected features for SBP and DBP lie between 21 and 25. The complete set of selected features is listed and described in Table 4.4 for DBP and Table 4.5 for SBP.

For DBP estimation, the first features being selected are amplitudes in the Fourier-transformed signal and maximum power frequencies. The frequencies included lie between 0.8 and 1.6 Hz. This frequency range might indicate the overall pulse shape and the heart rate which has been shown to influence peripheral BP and central diastolic pressure [93]. In the time domain, the systolic time and ratios between time and area under the curve are selected. These area-time ratio features represent the average amplitude during the considered time period. For the closely related BP waveform, this is the definition of MAP which typically is closer to DBP than to SBP [94]. Moreover, systolic and diastolic pulse width at 75% of the total amplitude is considered as well as their sum and ratio. The diastolic width contains information about the temporal position of the dicrotic notch where a delay is related to low DBP [95]. Further, a shortening of the systolic time indicates vasodilation and thus decreased DBP [94]. Next, the first two amplitudes of the first and second derivative, i.e. velocity and acceleration of the blood volume changes, are chosen. A steep slope of the pressure pulse can

Table 4.4: Full list and description of the selected rPPG features for DBP prediction.

Feature Name	**Description**
Ts	duration of the systole (systolic time)
SysWidth_75	systolic pulse widths at 75% of pulse amplitude
DiaWidth_25, DiaWidth_75, DiaWidth_90	diastolic pulse widths at 25%, 75% and 90% of pulse amplitude
PulseWidth_75	pulse width at 75% of pulse amplitude
Div_SysDiaWidth_75	ratio of systolic and diastolic pulse widths at 75% of pulse amplitude
Div_AsTs, Div_AT	ratio of systolic area and systolic time, ratio of cycle area and cycle time
Amp_FFT0, Amp_FFT4	amplitudes of the Fourier-transformed signal at indices 0 and 4, i.e. 0.8 and 1.6 Hz
MaxPow_FFT1, MaxPow_FFT3, MaxPow_FFT6, MaxPow_FFT8	frequencies of the 1st, 3rd, 6th and 8th highest amplitudes in the Fourier-transformed signal
d1_HPIP_width	time between first peak to following inflection point in the first derivative of the rPPG signal
d1_Amp1, d1_Amp2	amplitudes of the first and second peak in the first derivative of the rPPG signal
d2_Amp1, d2_Amp2	amplitudes of the first and second peak in the second derivative of the rPPG signal
Amp_DN	amplitude of the dicrotic notch
weight	weight of the subject

reflect vessel stiffness [95]. Finally, the amplitude of the dicrotic notch is added as well.

In contrast to DBP, SBP estimation relies on higher frequencies from 2.0 to 4.0 Hz which contain information about systolic peak and dicrotic notch morphology. In general, frequency features show a comparatively strong correlation with both

Table 4.5: Full list and description of the selected rPPG features for SBP prediction.

Feature Name	Description
T, Ts	duration of a whole cycle (cycle time), duration of the systole (systolic time)
A, Ad	area under the curve for whole cycle (cycle area), area under the curve for diastole (diastolic area)
Div_AsAd, Div_AAd, Div_AAs	ratio of systolic area and diastolic area, ratio of cycle area and diastolic area, ratio of cycle area and systolic area
SysWidth_10, SysWidth_25, SysWidth_33, SysWidth_50	systolic pulse widths at 10%, 25%, 33% and 50% of pulse amplitude
Div_SysDiaWidth_10, Div_SysDiaWidth_50, Div_SysDiaWidth_67	ratios of systolic and diastolic pulse widths at 10%, 50% and 67% of pulse amplitude
Amp_FFT6, Amp_FFT7, Amp_FFT10, Amp_FFT15, Amp_FFT16	amplitudes of the Fourier-transformed signal at indices 6, 7, 10 and 15, i.e. 2.0, 2.2, 2.8, 3.8 and 4.0 Hz
MaxPow_FFT0, MaxPow_FFT1	frequencies of the highest and second highest amplitudes in the Fourier-transformed signal
d1_HPIP_height	height difference between first peak and following inflection point in the first derivative of the rPPG signal
d2_Div_Amp1Amp2	ratio of the first and second amplitude in the second derivative of the rPPG signal
d2_Sum_Amp1Amp2	sum of the first and second amplitude in the second derivative of the rPPG signal
age, height	age and height of the subject

DBP and SBP which is in accordance with the work by Luo et al. [59]. For SBP, further, multiple systolic pulse widths and ratios between systolic and diastolic

width are chosen from the lower half of the pulse at 10, 25, 33 and 50% of the amplitude. Finally, the cycle time and area as well as ratios between the systolic, diastolic and whole cycle area are selected. Area and width features comprise indicators for stroke volume and vascular capacitance, which are the major influences on SBP [95].

In the SFS results for a joint model, time domain features are mostly similar to the selected features for DBP. The results comprise diastolic area under the curve and time, the ratio between cycle area and diastolic area, and various diastolic widths. Moreover, amplitudes of the first and second derivative as well as the dicrotic notch height are chosen. In return, selected frequency domain features resemble the selection results for SBP prediction as they include higher frequencies between 2.0 and 4.0 Hz besides the most dominant frequency, i.e. heart rate.

4.3.2 BP Prediction Accuracy

Table 4.6 shows the LOSO CV results on the dataset described in Section 4.2 for all regression models as well as further error metrics. SD_{MAE} and SD_{RMSE} refer to the SD of MAE and RMSE among all iterations in the LOSO CV, i.e. among all subjects. For SBP, the separate RFR model performs best with an MAE ±SD of 11.91 ±9.66 mmHg, and for DBP, the joint RFR model outperforms other configurations with an MAE ±SD of 7.92 ±6.02 mmHg.

Comparing the results from MLP and RFR, all RFR models perform slightly better than their MLP counterparts. This can be attributed to the embedded feature selection which occurs during the learning process of each RFR model. In contrast to SFS for MLP which was performed for all LOSO CV models, RFR integrates a new feature selection into each training process. Furthermore, the ensemble learning strategy followed by RFR enables the regression trees to reduce variance and compensate each others prediction errors. In general, RFR is very robust against outliers.

For MLP, the joint model performs better for both SBP and DBP than the single-output architectures. The joint model seems to benefit from the additional backpropagated output information and the correlation between the two output variables. In contrast, the SBP prediction error of the joint RFR model increases compared to the separate model. With both regression methods, SBP and DBP

Table 4.6: Prediction results from the MLP and RFR for rPPG features extracted from the forehead.

DBP					
	Separate Models		**Joint Model**		
	MLP	**RFR**	**MLP**	**RFR**	Mean Reg.
ME ±SD	0.15 ±11.78	0.12 ±10.02	-0.37 ±10.26	-0.18 ±9.94	0.46 ±12.31
MAE ±SD	8.19 ±8.46	7.81 ±6.27	8.12 ±6.29	7.92 ±6.02	8.58 ±8.83
SD_{MAE} among subjects	±3.87	±4.40	±4.47	±4.45	±5.17
RMSE	11.77	10.01	10.27	9.95	12.31
SD_{RMSE} among subjects	±4.12	±4.76	±4.69	±4.73	±5.59
Pearson coeff.	0.37	0.36	0.32	0.37	-0.26
BSH[a] Grade	D	C	D	C	D
SBP					
	Separate Models		**Joint Model**		
	MLP	**RFR**	**MLP**	**RFR**	Mean Reg.
ME ±SD	-1.67 ±17.78	0.44 ±15.34	-2.3 ±17.04	0.64 ±15.40	-1.50±19.41
MAE ±SD	13.61 ±11.54	11.91 ±9.66	13.48 ± 10.67	11.96 ±9.72	14.29 ±13.19
SD_{MAE} among subjects	±7.18	±5.27	±6.22	±5.27	±15.89
RMSE	17.84	15.33	17.20	15.41	19.43
SD_{RMSE} among subjects	±7.84	±5.86	±6.83	±5.85	±16.51
Pearson coeff.	0.41	0.54	0.41	0.54	0.59
BSH[a] Grade	D	D	D	D	D

[a]British Society of Hypertension

need to share each decision node or neuron in a joint model. However, in RFR, decision trees follow one path for both variables using the splitting criteria that reduces the average reduction across both outputs. Hence, only in the leave nodes, the two output variables are differentiated, whereas in MLP, the output variables can be decoupled at an earlier layer by adjusting the weights accordingly. In case of the RFR BP prediction model at hand, DBP benefits from the correlation with SBP, but hinders optimisation for SBP.

Overall, the BP estimation still exhibits large deviations from the reference values with only a small advantage over the mean regressor. None of the results is within the standard set by the US Association for the Advancement of Medical Instrumentation (AAMI) of a maximum mean error < 5 mmHg and standard deviation < 8 mmHg [96] and most models obtain grade D (see Table 4.6) ac-

cording to Criteria of British Society of Hypertension (BSH) [97]. These are the two most widely used protocols which are also referenced and employed by the ESH. However, state-of-the-art methods from related works obtain similar errors, though, comparability is limited since each study uses their own dataset for training and validation of the proposed methods. Older literature [58, 59] even uses data only from normotensive subjects (DBP from 60 to 90 mmHg and SBP from 100 to 140 mmHg [59] or from 95 to 130 mmHg [58]). If we reduce the BP range to 60 to 90 mmHg for DBP and 100 to 140 mmHg for SBP as well, the errors decrease as expected since training data and BP predictions are limited to a smaller range accordingly. Moreover, the sparsely represented range in our data set of DBP $>$ 105 mmHg and SBP $>$ 170 mmHg is excluded. We obtain an MAE $\pm$SD of 6.27 $\pm$4.58 mmHg for SBP and 3.91 $\pm$2.79 mmHg for DBP, and an ME $\pm$SD of -0.11 $\pm$7.76 mmHg and 0.45 $\pm$4.78 mmHg, respectively. These error values are comparable to the results obtained by Jain and Subramanyam [58] and Luo et al. [59], even though different BP distributions of the proprietary data sets need to be considered.

The proposed approach in this chapter is based on hand-crafted features and RFR, and shows improved performance compared to previous studies. While models based on hand-crafted features may benefit from a reduced input size and previous domain knowledge, important features and signal characteristics might be missed. In contrast, an automated feature extraction by employing convolutional layers on the raw signal input has the potential of finding new relevant features. Still, with this method, the feature extraction and selection process revealed relevant characteristics of the rPPG signal. Especially the frequency domain showed to contain important information for BP prediction and should be considered for the CNN-based approach as well.

Furthermore, rPPG signals of videos with changing illumination were strongly corrupted such that after automatic selection of appropriate pulses, the number of rPPG cycles was significantly reduced. Hence, rPPG extraction needs to be improved. For this purpose, transformations of the RGB colour space will be analysed in the following chapter. Moreover, the forehead might not be an appropriate measurement site for rPPG-based BP since blood circulation in the face is subject to sympathetic and parasympathetic activation [20]. The palm

will be considered as an alternative which exhibits a high vessel density [26] and is commonly exposed.

Chapter 5

Optimizing Remote Photoplethysmography

The first processing step for camera-based BP estimation is the extraction of the rPPG signal from the video. This signal is the basis for all further processing and thus, should be as clear and noise-free as possible. The signal is extracted from skin pixels whose position is determined by hand or face detection as described in Chapter 4. Next, the pixels' colour information is spatially averaged and further processed in order to obtain an rPPG signal over time. While in the previous Chapter 4.1.1, selection of a suitable tracking algorithm ensures a stable ROI, in this chapter, various colour representations of the images are analysed for different measurement scenarios in order to increase robustness against movement and illumination artifacts (Section 5.2), and signal quality is compared for ROIs at the hand and face (Section 5.3).

5.1 Interfering Factors in rPPG Measurement

As established in Section 2.2, light with a wavelength around 575 nm is suited best for rPPG measurement due to the strong reflectance pulsation magnitude [26]. For RGB cameras, this usually corresponds to the spectrum of the green channel as shown in Figure 5.1 which depicts the quantum efficiency of the Allied Vision Manta G-040C RGB camera used in this work. In this section, the quality of the rPPG signal extracted from the green channel is evaluated as a baseline, where also non-ideal measurement conditions are considered.

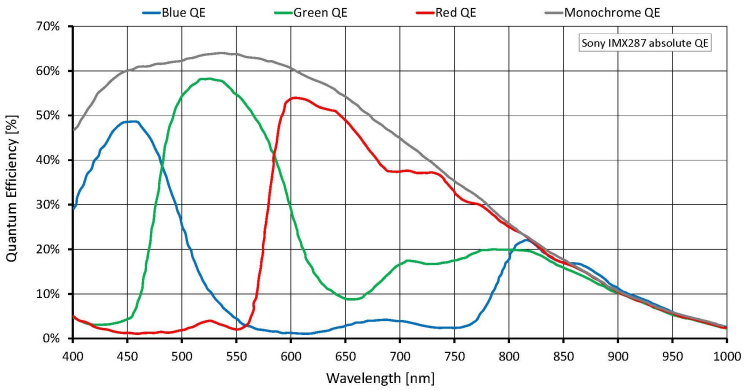


Figure 5.1: Quantum efficiency of the Allied Vision Manta G-040C RGB camera. The optimal wavelength for rPPG measurement (575 nm) falls into the green RGB channel. [98]

Depending on the target application of the rPPG-based BP estimation, different interfering factors might influence signal quality. In this work, four major factors are considered: Illumination changes, dark skin, overlap and movement.
When employed in a naturally illuminated environment, illumination changes cannot be avoided and lead to alterations in the intensity values of all three RGB channels. When only looking at the green channel signal, such illumination changes cannot be differentiated from absorption changes caused by pulsatile blood.
Next, a darker skin poses the challenge of a lower mean reflectance and a lower reflectance pulsation accordingly (see Chapter 2.2.2), such that the rPPG signal quality is more sensitive to noise. The rPPG extraction method should produce acceptable signal quality across all skin colours to enable an application for a diverse population.
Further, although the ROI is adjusted relatively to the position and size of the bounding box corresponding to face or hand, it might overlap with non-skin pixels such as hair, eyebrows, glasses or textile. By spatially averaging all pixel values in the ROI, this can strongly influence the resulting signal. Overlap with overexposed regions and skin reflections will have a similar effect since skin pixels are replaced by pixels that do not exhibit rPPG information.
Finally, movement of subjects might be the most complex case: Even if the tracking of the ROI is exact, the movement can introduce other interfering factors such

as illumination changes due to a new relative orientation of the face to the light source. Hence, there might be alterations in specular reflection, as well as appearances of new or disappearances of current overexposed regions and skin reflections within the ROI. Moreover, faces are not perfectly spherical and non-rigid, such that movement can lead to changes in the skin (and maybe non-skin) region that corresponds to the ROI pixels.

In the following, the signal quality of the green-channel rPPG is established as a baseline taking into account the previously described interfering measurement situations.

5.1.1 Experimental Setup

Based on correlation and SNR, the quality of the green-channel rPPG signal is analysed for the whole dataset as well as separately for videos with different interfering factors. The considered conditions are changing illumination, darker skin

Table 5.1: Investigated rPPG measurement scenarios and their definitions. These conditions are visually assessed in the recorded videos.

Measurement category	**Definition**
illumination changes	purposefully introduced illumination changes during the whole video. Illumination changes were achieved by manipulating the aperture of the camera during recording resulting in an average overall intensity change in pixel values of -37% with intensity reductions of up to 57% (see Chapter 4.2).
dark skin	skin tone $\geqslant 4$ on the Fitzpatrick scale [99] as opposed to light skin tones $\leqslant 3$.
overlap	significant overlap ($\geqslant 10\%$) of the ROI with hair, glasses, headband or eyebrows during at least half of the video, or same conditions for overexposed regions and skin reflections within the ROI
movement	at least five clearly visible head movements in any direction

colour, videos with noticeable head movement and videos where the ROI creates overlap with eyebrows, hair, the PPG head band or overexposed regions due to skin reflections. The measurement scenarios with the different interfering factors are defined in Table 5.1. The videos are visually assessed after face detection and positioning of the ROI to assign these measurement categories.

Correlation Analysis

For evaluation of robustness against the interfering factors, the recorded videos are compared to the corresponding PPG references from the forehead (see Chapter 4.2). For the rPPG signals, face detection and tracking is performed with HOG and MOSSE, respectively, according to the results in Section 4.1.1. Afterwards, the ROI is defined as described in Section 4.1.2. Then, both rPPG and PPG signals are re-sampled to $f_s = 100$Hz and filtered with a 32-order low-pass at $f_{high} = 6.0$Hz and a 512-order high-pass at $f_{low} = 0.3$Hz. Finally, a cross correlation $(p_{rPPG} * p_{PPG})$ between rPPG and PPG is calculated as

$$(p_{rPPG} * p_{PPG})[n] = \sum_{m=0}^{N} p^*_{rPPG}[m] \cdot p_{PPG}[m+n] \tag{5.1}$$

with p^*_{rPPG} being the complex conjugate of p_{rPPG}. Since rPPG and PPG were measured by two separate systems which were triggered manually, the temporal synchronicity of the two signals is not ensured. Hence, to account for a time delay between the PPG and rPPG signal, correlation R between the two signals is defined as

$$R = \max((p_{rPPG} * p_{PPG})[n]) \tag{5.2}$$

for $-0.5\text{s} * f_s < n < 0.5\text{s} * f_s$.

Signal-to-Noise Ratio

Moreover, the SNR of the extracted rPPG signal is assessed. Since the noise level for an ROI at a skin area cannot be measured without the rPPG signal and since the original signal is not available, the spectral method by De Haan and Van Leest [100] is employed. To estimate the SNR, the ratio of the energy around the fundamental heart rate frequency as well as its second harmonic (in-band frequencies) and the remaining energy (out-of-band frequencies) within $f_{low} = 0.8$

Hz and $f_{high} = 5.0$ Hz is computed, as defined by:

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_{f_{low}=0.8}^{f_{high}=5.0} (U_t(f)\tilde{S}(f))^2}{\sum_{f_{low}=0.8}^{f_{high}=5.0} (1 - U_t(f)\tilde{S}(f))^2} \right) \tag{5.3}$$

where $\tilde{S}(f)$ denotes the spectrum of the raw rPPG signal and $U_t(f)$ is a binary window to pass in-band frequencies while blocking out-of-band frequencies.

Blood Pressure Prediction Accuracy

BP prediction performance is evaluated for all measurement scenarios, using the model obtained in Chapter 4.3. To do so, a LOSO CV is performed on the data set presented in Chapter 4.2.

Rejection Rate

Requirements for passing the quality assessment for rPPG pulses are a wavelength similar to the rest of the signal section, a dominant systolic peak and a dicrotic notch in form of another peak or inflection point (see Chapter 4.1.4). If any of these criteria is not met, the rPPG cycle is discarded and not used for BP estimation. Hence, as another quality measure, the percentage of signal sections that passed these quality criteria is evaluated for the different types of measurement scenarios.

5.1.2 Baseline Performance during Interfering Conditions

Figure 5.2 depicts the cross-correlation results and SNR for the baseline green-channel rPPG signal. The correlation between reference PPG and rPPG is low in general with $R_{RGB,all} = 0.52$. While the videos with dark skin, overlap or movement obtain slightly lower Pearson correlation coefficients, rPPG signals for videos with illumination changes exhibit a strongly reduced correlation with the reference signals of $R_{RGB,i} = 0.16$. This can be attributed to the dependence of each RGB colour channel with overall brightness leading to strong artifacts in the rPPG signal.

This limitation of the RGB representation in dealing with changing illumination is also reflected by the low SNR. For RGB signals, the noise level is above the signal's energy level for all interfering measurement conditions where SNR

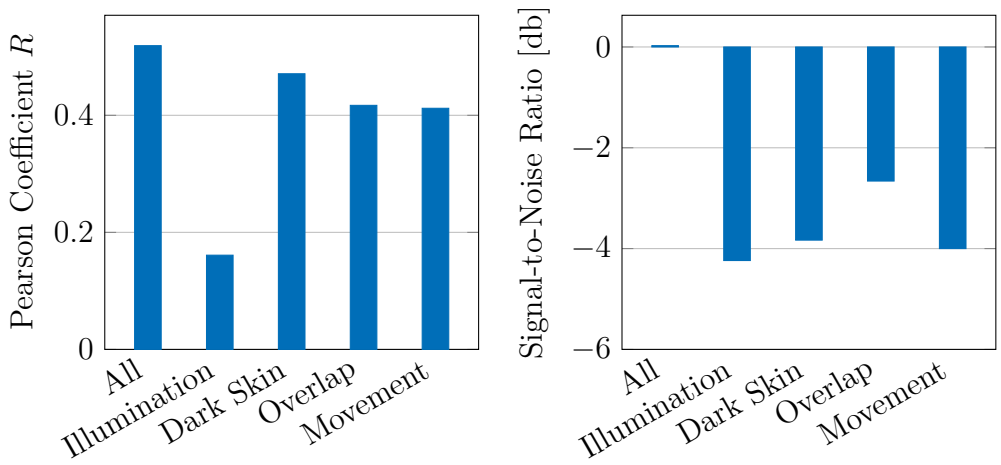


Figure 5.2: Cross-correlation results between reference PPG and green-channel rPPG signals (left) and SNR of the green-channel rPPG signal (right). Both quality assessments are performed for the entire data set as well as for the different interfering measurement scenarios separately.

is reduced most significantly for videos with illumination changes, but also for dark skin and movement. The high photon absorption by melanin leads to a low reflection rate for darker skin causing a low SNR. Melanin is mainly present in superficial skin tissue, thus reducing light intensity for photons entering the skin and again, reducing photons that were reflected by other skin components, such that noise levels in specular reflection exceed the rPPG signal. Moreover, movement can introduce artifacts of different sources, most importantly from variations in specular reflection due to changes in the relative angle to the light source. All measurement scenarios exhibit a strongly reduced SNR compared to the reference PPG with $SNR_{ref} = 4.17$db.

After signal processing, individual pulses are segmented and assessed in terms of wavelength and the existence of a dominant systolic peak and a dicrotic notch. That way, corrupted pulses are sorted out and are not used for BP estimation. Figure 5.3 shows the percentage of passed data samples for the green-channel baseline in each of the interfering measurement scenarios. Despite the low SNR, rPPG segments from videos with dark skin have the lowest rejection rate of 12.5%. In contrast, movement and illumination changes produce the most unsuitable signal sections with 31.8% and 31.7% rejected, respectively.

Finally, BP prediction accuracy is evaluated for each interfering factor. Figure

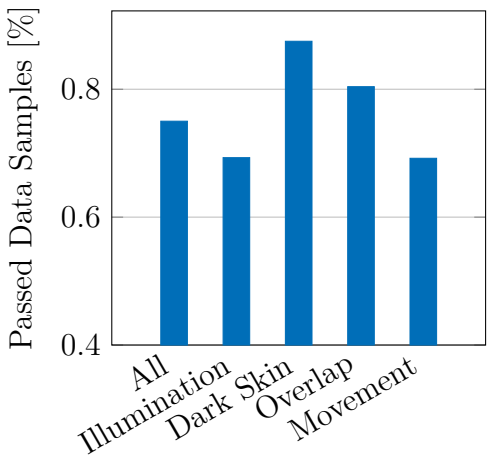


Figure 5.3: Percentage of signal sections that passed the pre-processing quality criteria. Evaluated for green-channel rPPG signals during different recording scenarios.

5.4 details the RMSEs and MAEs for SBP and DBP. Note that due to the study design, the different measurement scenarios are not equally distributed across the BP values Illumination changes were introduced in the second of four recordings per person such that mainly normal BP levels are included whereas the other categories are distributed arbitrarily among all videos. Therefore, these values serve as baseline for the rPPG extraction method under various measurement conditions on this data set, but should not be compared among measurement conditions.

5.2 Colour Model Transformation

As shown in the previous section, especially illumination changes but also movement, overlap and dark skin degrade the signal quality of rPPG signals extracted from the green RGB channel. This can be attributed to the RGB colour model representing illumination changes by a simultaneous increase (or decrease) in pixel values of all three channels.

Several other colour models, on the other hand, separate brightness, luminance or lightness parameter from chromaticity values by introducing different dimensions. A transformation to such colour model might eliminate illumination artifacts in the rPPG signal. Therefore, transformations of the ROI pixels to five

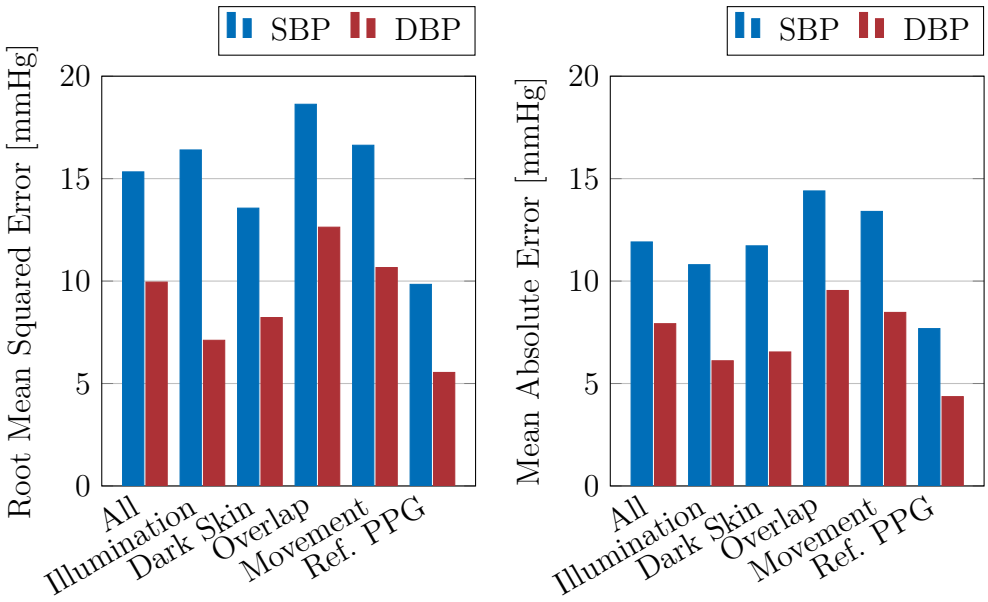


Figure 5.4: RMSE in BP estimation from green-channel rPPG signals (left) and MAE in BP estimation from green-channel rPPG signals (right). Both error metrics are calculated for the entire data set as well as for the different interfering measurement scenarios separately. For comparison, results are also given for a model trained and tested on the reference PPG signals.

different colour models is performed. Besides RGB, we consider HSV, CMYK, Lab, YCrCb and YUV (see Section 2.5.3) as well as Chrom [50] and POS [49]. The rPPG signals are obtained by spatially averaging the new chromaticity channels that are associated with the green colour dimension.

Subsequently, these signals are band-pass filtered to eliminate high-frequency noise as well as respiration and movement artifacts. Finally, they are compared to the corresponding PPG reference signals by cross correlation. Moreover, SNRs are assessed and BP prediction performance for each of the colour representations is evaluated on the RFR models of the feature-based approach of Section 4.3. All results are compared to the green-channel rPPG baselines which were established for the whole data set as well as for the different interfering factors in Section 5.1.

5.2.1 Experimental Setup

To evaluate robustness of the new colour channels and suitability for rPPG extraction, the recorded videos and PPG references from the forehead (see Chapter

4.2) are compared. The same processing steps are applied as for the RGB-based rPPG signal extraction used in Section 5.1. After face detection and tracking with HOG and MOSSE, respectively (see Section 4.1.1), the ROI is defined as described in Section 4.1.2 and its pixel values are transformed to the colour models HSV, Lab, YCrCb, YUV and CMYK. For each of the colour models, we compute rPPG signals from the chromaticity channels that are associated with the green colour component, i.e. HSV-H, Lab-a, YCrCb-Cr, YUV-V, and CMYK-M. After re-sampling rPPG and PPG signals to $f_s = 100$Hz and eliminating high frequency noise and modulation by the respiratory signal (see Section 5.1.1), cross correlation ($p_{rPPG} * p_{PPG}$) is calculated according to Equations 5.1 and 5.2.

Moreover, the SNR of the extracted rPPG signal is assessed for each of the colour models using the spectral method by De Haan and Van Leest [100]. The method is described in Section 5.1.1 and the SNR is computed according to Equation 5.3.

Analogous to the green-channel baseline in Section 5.1, the quality of the rPPG signals of different colour models is analysed for the whole dataset as well as separately for videos with the previously defined measurement scenarios (see Table 5.1). The same process is performed for the rPPG extraction methods Chrom [50] and POS [49]. Finally, BP prediction performance is evaluated for all colour models across the measurement scenarios, using the model obtained in Section 4.3.

The following section provides results and discussions which are divided into the rPPG signal assessment regarding correlation with the reference signals and SNR in Section 5.2.2, evaluation of the colour models considering BP prediction accuracy in Section 5.2.3 followed by analyses of the number of usable signal sections in Section 5.2.4.

5.2.2 rPPG Quality Assessment

Correlation Analysis

The correlation results between contactless measured rPPG and reference PPG signals under the defined measurement conditions are shown in Figure 5.5. Transformations to Lab, YCrCb and YUV do not improve correlation with the reference signal in any video category, in contrast to HSV, CMYK, Chrom and POS. Most

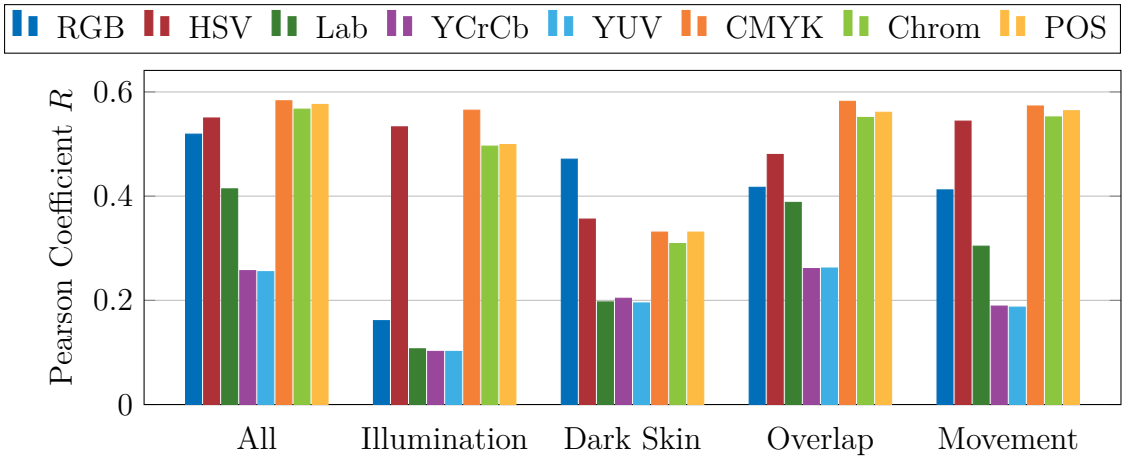


Figure 5.5: Cross-correlation results between reference PPG and rPPG signals from different colour models or rPPG models.

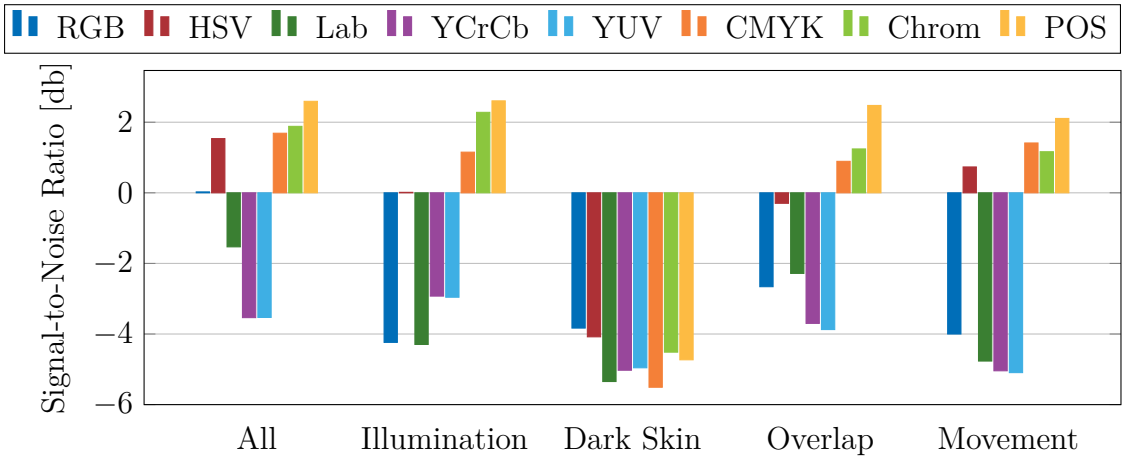


Figure 5.6: SNR of rPPG signals from different colour models or rPPG models.

significant is the improvement with latter transformation models for changing illumination. Whereas the RGB signals exhibit a strongly reduced correlation with the reference signals from $R_{RGB,all} = 0.52$ to $R_{RGB,i} = 0.16$, correlation decreases for Chrom and POS only by 0.07 and 0.08, respectively. Least signal degradation is observed for a colour transformation to the CMYK colour model with $R_{CMYK,all} = 0.58$ and $R_{CMYK,i} = 0.57$.

The same trend is observable for ROIs overlapping with eyebrows, hair, head band or reflections as well as for movement, since these two scenarios cause sudden changes in ROI pixel values and manifest as similar artifacts in the overall rPPG signal. Overlap with hair, objects and reflections reduces the skin area

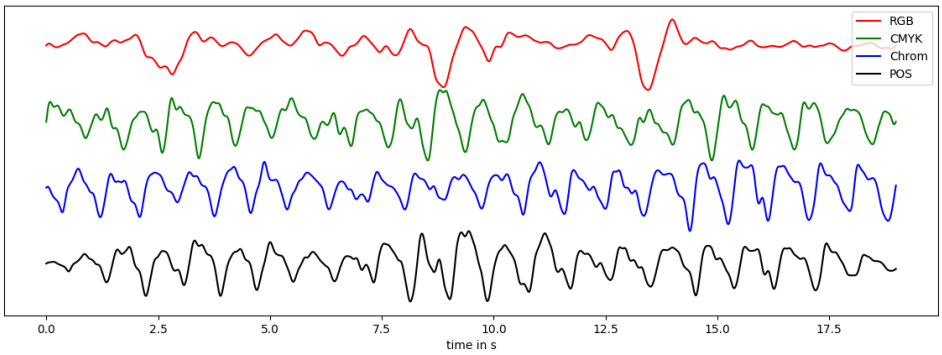


Figure 5.7: rPPG signals from different colour models, recorded under illumination changes.

within the ROI and replaces a subset of the ROI pixels by constant values. Due to minor head movements, the number of non-skin pixels is subject to small variations that introduce artifacts in the spatially averaged ROI values. Especially skin reflection and overexposed regions but also most hair colours do not cause significant changes in chromaticity, thus mimicking illumination changes in the rPPG signal.
For overlap with coloured objects, the rPPG extraction methods Chrom and POS have an advantage since they estimate the light intensity and colour of the specular reflection in order to separate this signal from the rPPG signal. Hence, overlap with coloured objects will affect the estimated spectrum of specular reflection such that its influence on chromaticity and on the rPPG is mostly eliminated.
Movement, on the other hand, causes changes in the direction of specular reflection and some minor morphological changes of the skin surface within the ROI. Variations in direction of specular reflection lead to perceived illumination changes, whose influence can be mitigated by colour model transformations.

However, across all colour models, rPPG signal quality is strongly reduced for darker skin tones since absorption by melanin is increased and the proportion of by haemoglobin reflected photons becomes even smaller. For these measurements, RGB signals exhibit the strongest correlation with the reference PPG of $R_{RGB,s} = 0.47$ and colour model transformations do not improve signal quality.

Signal-to-Noise Ratio

Figure 5.6 depicts the SNRs for rPPG signals of all colour models and for different measurement conditions. It is evident that the POS method obtains the highest SNR across all measurement scenarios with the exception of darker skin tones, providing an SNR closest to the reference PPG ($SNR_{ref} = 4.17$db). CMYK and Chrom, too, exhibit an SNR above 0.9 db for illumination changes, overlap and movement. Also HSV shows a significant improvement over the RGB baseline for all categories except dark skin. Figure 5.7 depicts an example signal with illumination changes for different rPPG extraction methods. It illustrates the limitations of the RGB representation in dealing with changing illumination due to the dependence of each colour channel with overall brightness. Interestingly, in terms of SNR, CMYK outperforms other colour models which separate brightness, luminance or lightness from chrominance information, although it is a subtractive colour model. Lab, YCrCb and YUV exhibit negative SNRs even for videos without any noteworthy artifacts. These three colour models do not show any improvement over the RGB colour representation regarding correlation with the reference signal and SNR, and thus, will not be considered in further analysis.

As for RGB-based rPPG signals, the generally low photon reflection rate for darker skin causes low SNRs across all colour models. In this case a transformation of the colour space does not improve signal quality. However, Shirbani et al. [101] showed in their analysis of ambient light and skin tone for the estimation of heart rate, that at brighter light intensity, the heart rate prediction error was independent of the skin tone. Since the interaction of light intensity and skin tone can not be investigated with the data set at hand, its assessment with respect to BP estimation should be part of a future study.

It has to be noted that the employed SNR method only considers the heart rate and second harmonic to belong to the rPPG signal. However, the rPPG waveform also contains higher frequency components, e.g. belonging to the dicrotic notch. Colour models that better preserve these higher frequency features might obtain a reduced SNR.

In general, when comparing the contact-less measured rPPG to the reference PPG signals, a different waveform morphology can be observed although both measurement locations are right next to each other at the forehead. The pres-

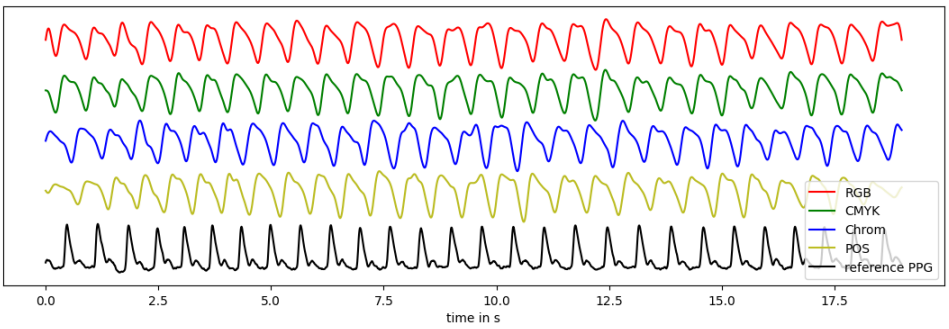


Figure 5.8: rPPG signals from different color models in comparison to the reference PPG (black).

sure (and thus absorption) decrease during diastole is slower for the rPPG signal resulting in a less pronounced dicrotic notch as shown in Figure 5.8. This could be attributed to the rPPG originating from more peripheral vessels due to lower light intensity at the measurement site and thus, an overall more shallow light penetration of the skin. As covered in Section 2.1, arteries and arterioles absorb around half of the cardiac stroke volume during systole such that peripheral arterioles and capillaries exhibit less pronounced BP pulses. This might be reflected by the rPPG signals. The differences in pulse morphology between PPG and rPPG cause a generally low correlation coefficient. It is to be evaluated whether rPPG signals still contain sufficient information about the underlying BP. For further assessment of colour model transformed rPPG signals, prediction performance is analysed for the colour models that provided the strongest correlation with the reference signals and the highest SNR, i.e. RGB, CMYK, Chrom and POS.

5.2.3 BP Prediction Accuracy

For each colour model, general BP prediction performance is evaluated in a LOSO CV on the data set presented in Chapter 4.2. Table 5.2 for DBP and Table 5.3 for SBP detail the corresponding evaluation results. For DBP prediction, Chrom signals perform best and obtain an MAE ±SD of 7.54 ±6.12 mmHg closely followed by POS and RGB with 7.73 ±6.43 mmHg and 7.92 ±6.02 mmHg respectively. For SBP, RGB signals obtain the lowest MAE ±SD with 11.91 ±9.66 mmHg, but in terms of RMSE, the Chrom method shows a slight advan-

Table 5.2: DBP prediction results for separate and joint RFR and different colour spaces. (Forehead)

Joint RFR (DBP)					
	RGB	**CMYK**	**Chrom**	**POS**	Mean Reg.
ME ±SD [mmHg]	-0.18 ±9.94	0.39 ±10.64	-0.07 ±9.71	-0.12 ±10.06	0.46 ±12.31
MAE ±SD [mmHg]	7.92 ±6.02	8.02 ±7.00	7.54 ±6.12	7.73 ±6.43	8.58 ±8.83
SD_{MAE} among subjects [mmHg]	±4.45	±5.08	±5.14	±5.58	±5.17
RMSE [mmHg]	9.95	10.65	9.71	10.06	12.31
SD_{RMSE} among subjects [mmHg]	±4.73	±5.42	±5.83	±6.06	±5.59
Pearson coeff.	0.37	0.25	0.22	0.28	-0.26
BSH[a] Grade	C	D	C	C	D
separate RFR (DBP)					
	RGB	**CMYK**	**Chrom**	**POS**	Mean Reg.
ME ±SD [mmHg]	0.12 ±10.02	-0.61 ±10.29	0.15 ±10.07	0.17 ±10.26	0.46 ±12.31
MAE ±SD [mmHg]	7.81 ±6.27	7.86 ±6.67	7.84 ±6.31	7.94 ±6.48	8.58 ±8.83
SD_{MAE} among subjects [mmHg]	±4.40	±4.59	±5.17	±4.81	±5.17
RMSE [mmHg]	10.01	10.30	10.06	10.25	12.31
SD_{RMSE} among subjects [mmHg]	±4.76	±4.95	±5.50	±5.21	±5.59
Pearson coeff.	0.36	0.30	0.14	0.25	-0.26
BSH[a] Grade	C	C	D	D	D

[a]British Society of Hypertension

tage over RGB signals with 15.28 mmHg on the joint RFR model. In comparison, when trained and tested with almost ideal signals, i.e. with the reference PPG signals, the RFR model obtains an MAE ±SD of 4.36 ±3.42 mmHg for DBP and 7.68 ±6.16 mmHg for SBP.

Based on the correlation results, the RGB representation was not expected to exhibit lower prediction errors than any of the other rPPG extraction methods. This outcome can be explained by the signal quality assessment that is performed prior to feature extraction. Features are only extracted from rPPG cycles that exhibit a similar wavelength as the rest of the signal section, a dominant systolic peak and a dicrotic notch in form of another peak or inflection point (see Section 4.1.4). Hence, the influence of signal quality on the prediction performance is reduced by these pre-processing and signal selection steps. Therefore, the percentage of signal sections that passed these quality criteria is evaluated for each of the rPPG extraction methods and for the different types of challenging measurement scenarios in Section 5.2.4.

Table 5.3: SBP prediction results for separate and joint RFR and different colour spaces. (Forehead)

Joint RFR (SBP)					
	RGB	**CMYK**	**Chrom**	**POS**	Mean Reg.
ME ±SD [mmHg]	0.64 ±15.40	0.60 ±15.62	0.59 ±15.27	0.80 ±15.37	-1.50±19.41
MAE ±SD [mmHg]	11.96 ±9.72	12.27 ±9.68	12.06 ±9.39	12.36 ±9.17	14.29 ±13.19
SD_{MAE} among subjects [mmHg]	±5.27	±6.17	±5.88	±6.44	±15.89
RMSE [mmHg]	15.41	15.63	15.28	15.39	19.43
SD_{RMSE} among subjects [mmHg]	±5.85	±6.64	±6.28	±6.71	±16.51
Pearson coeff.	0.54	0.47	0.50	0.51	
BSH[a] Grade	D	D	D	D	D
separate RFR (SBP)					
	RGB	**CMYK**	**Chrom**	**POS**	Mean Reg.
ME ±SD [mmHg]	0.45 ±15.34	0.81 ±15.77	0.35 ±15.57	0.62 ±15.32	-1.50±19.41
MAE ±SD [mmHg]	11.91 ±9.66	12.35 ±9.85	12.38 ±9.46	12.17 ±9.32	14.29 ±13.19
SD_{MAE} among subjects [mmHg]	±5.27	±6.06	±6.10	±6.68	±15.89
RMSE [mmHg]	15.33	15.73	15.58	15.33	19.43
SD_{RMSE} among subjects [mmHg]	±5.88	±6.52	±6.52	±6.88	±16.51
Pearson coeff.	0.54	0.46	0.47	0.52	0.59
BSH[a] Grade	D	D	D	D	D

[a]British Society of Hypertension

In general, the differences in BP prediction accuracy between the colour models are very small when looking at performances on the whole data set. Thus, for a more differentiated evaluation of the colour models, their performances are considered separately for each of the different measurement conditions. For SBP, the separate RFR model is employed, and for DBP, the joint RFR model is used, based on the results in Section 4.3.

Table 5.4 shows the RMSE values of BP prediction on the colour model transformed signals for illumination changes, darker skin, overlap with hair or textiles and movement as well as the results on the whole data set. Table 5.5 contains results of the same experiments in terms of MAE and SD. Whereas the error values on the whole data set are very similar for all colour models and do not allow a clear decision, they exhibit significant differences within each of the measurement scenarios.

The large differences in error values across the measurement scenarios can be attributed to the distribution in BP values. Illumination changes were introduced

Table 5.4: RMSE for BP prediction for different colour models and recording scenarios. SBP is computed by an single-output RFR model, and DBP is obtained by a joint RFR model.

SBP - RMSE in [mmHg]					
	RGB	**CMYK**	**Chrom**	**POS**	Mean Reg.
All	**15.33**	*15.73*	15.58	**15.33**	19.43
Illumination Changes	*16.40*	**9.53**	10.39	12.36	9.51
Dark Skin	13.56	13.31	**13.29**	*13.68*	13.38
Overlap	**18.63**	*20.59*	19.00	18.77	22.71
Movement	16.63	*17.73*	16.89	**15.60**	20.60
DBP - RMSE in [mmHg]					
All	9.95	*10.65*	**9.71**	10.06	12.31
Illumination Changes	*7.11*	6.87	**4.60**	5.62	4.11
Dark Skin	*8.22*	6.73	**6.16**	6.61	5.21
Overlap	**12.63**	*13.64*	13.02	12.92	13.88
Movement	10.66	*11.29*	10.53	**9.73**	8.26

in the second of four recordings per person such that mainly normal BP levels are included whereas the other categories are distributed arbitrarily among all recordings. Therefore, error values are compared only within each recording category.

For illumination changes, CMYK obtains the lowest RMSE for SBP with 9.53 mmHg closely followed by the Chrom model, while for DBP, Chrom performs best with an RMSE of 4.60 mmHg. If considering the MAE instead, Chrom appears to be most robust against illumination changes for both SBP and DBP. RGB performs worst with an RMSE of 16.40 mmHg for SBP and 7.11 mmHg for DBP. In contrast to RGB, who's channels have no separation between chrominance and brightness information, Chrom was specifically developed for rPPG extraction and only considers intensity variations that are orthogonal to the direction of specular reflection in the projection space [50].

For dark skin colour, there is not much difference between the results of the colour models. Chrom exhibits a small advantage in both SBP and DBP measurement, and RGB obtains a significantly higher RMSE and MAE for DBP than the other options.

For overlap of the ROI with non-skin pixels, RGB and POS perform slightly

Table 5.5: MAE for BP prediction for different colour models and recording scenarios. SBP is computed by an single-output RFR model, and DBP is obtained by a joint RFR model.

SBP - MAE ±SD in [mmHg]					
	RGB	**CMYK**	**Chrom**	**POS**	Mean Reg.
All	**11.91** ±9.66	12.35 ±9.85	*12.38* ±9.46	12.17 ±9.32	14.29 ±13.19
Illumination Changes	*10.80* ±12.34	7.47 ±5.92	**7.21** ±7.49	9.67 ±7.69	8.22 ±4.79
Dark Skin	**11.72** ±6.82	11.92 ±5.93	11.80 ±6.12	*12.22* ±6.17	9.32 ±9.61
Overlap	14.40 ±11.81	*16.23* ±12.68	14.68 ±12.05	**14.14** ±12.34	18.10 ±13.71
Movement	13.40 ±9.84	*14.17* ±10.68	13.74 ±9.82	**12.60** ±9.20	15.89 ±13.12
DBP - MAE ±SD in [mmHg]					
All	7.92 ±6.02	*8.02* ±7.00	**7.54** ±6.12	7.73 ±6.43	8.58 ±8.83
Illumination Changes	*6.11* ±3.64	5.67 ±3.57	**4.08** ±2.11	4.76 ±2.99	3.68 ±1.83
Dark Skin	*6.54* ±4.99	5.74 ±3.52	**4.94** ±3.69	5.36 ±3.87	4.41 ±2.76
Overlap	9.54 ±8.28	*10.24* ±9.01	**9.31** ±9.10	9.45 ±8.80	10.26 ±9.35
Movement	8.47 ±6.46	*9.18* ±6.57	8.36 ±6.40	**7.64** ±6.03	6.43 ±5.19

better than Chrom, while CMYK exhibits the largest RMSEs with 20.59 mmHg for SBP and 13.64 mmHg for DBP.

Finally, POS shows to be the most robust colour representation against movement of the subject, while CMYK obtains the highest errors for both SBP and DBP on these videos.

Overall, videos with ROI overlap obtain the highest errors compared to other measurement scenarios. In contrast to illumination changes, overlap with hair or textiles can introduce changes in chrominance depending on the colour of these objects. Hence, this might be the most challenging measurement scenario for the colour models, where they might not be able to compensate these artifacts. However, a segmentation of skin pixels can eliminate this issue and should be considered in future implementations.

Across all measurement conditions, Chrom appears to be the most robust rPPG model closely followed by POS which exhibits the highest invariance against movement of the subject. As discussed before, this performance assessment is based on BP prediction errors for signal sections that passed the rPPG quality criteria before feature extraction. For a more holistic assessment, the ratio of signal sections that are considered unsuitable and thus are not used for BP

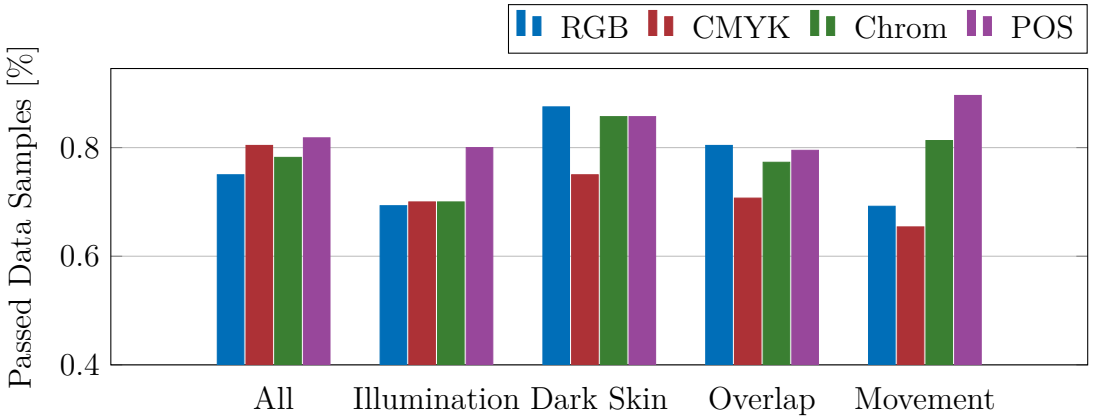


Figure 5.9: Percentage of signal sections that passed the pre-processing quality criteria. Evaluated for rPPG signals from different colour spaces and for the different recording scenarios.

estimation needs to be quantified.

5.2.4 rPPG Rejection Rate

Requirements for passing the quality assessment for rPPG pulses are a wavelength similar to the rest of the signal section, a dominant systolic peak and a dicrotic notch in form of another peak or inflection point (see Chapter 4.1.4). If any of these criteria is not met, the rPPG cycle is discarded. Hence, in order to determine which colour model is best suited for BP prediction from video data, the percentage of signal sections that passed these quality criteria is evaluated for each of the rPPG extraction methods and for the different types of measurement scenarios.

The results are presented in Figure 5.9 and show the same trend as the signal correlation analysis in Section 5.2. On the whole data set, RGB signals are discarded most often with 25 %, whereas POS signals are the most robust with 18 % rejected. Across all challenging measurement scenarios, CMYK keeps the lowest number of signal sections. In contrast, POS signals are rejected the least for videos with illumination changes and movement, while also performing well for dark skin and ROI overlap. In latter two cases, RGB achieves the lowest rejection rate with a small improvement over POS and Chrom.

5.2.5 Conclusion on the Selection of the Colour Model

Overall, the colour model transformations can achieve an improvement over the green-channel baseline in terms of signal quality and prediction performance for the majority of investigated interfering measurement scenarios. Considering the BP prediction errors from Section 5.2.3 and the rPPG rejection rates, POS is the clear choice for videos containing movement since POS achieves the lowest prediction errors while keeping the most signal sections. For the other measurement scenarios, the decision is a trade-off between prediction accuracy on the current model and rejection rate. Whereas Chrom leads in terms of accuracy, POS appears to provide a more stable signal quality. Taking the high SNRs of the POS signals into account, POS is the rPPG extraction method that will be employed for following experiments.

5.3 Region of Interest

So far, the ROI for rPPG extraction was defined at the forehead. A previous study [102] compared various facial subregions and showed that the forehead and upper cheeks provide the best rPPG signals. However, concerns were raised in a related study [20] that the forehead might not be an appropriate measurement site for rPPG-based BP due to the impact of sympathetic and parasympathetic activation on the blood circulation in the face. Therefore, the palm will be considered as an alternative measurement site which exhibits a high vessel density [26] and is commonly exposed. Furthermore, the effect of the size of the ROI on the signal quality is investigated.

5.3.1 Experimental Setup

The two ROIs (palm and forehead) are compared in terms of rPPG SNR and BP prediction accuracy on the data set described in Chapter 4.2. A reliable detection and stable tracking of the hand was ensured previously in Section 4.1.1 to provide the same preconditions as the ROI at the forehead. The definition of the size and position of the ROI is described in Section 4.1.2. For comparison of the prediction performance of the two skin regions, the POS method is utilised for rPPG extraction from the video data, whereas SNR is also evaluated for

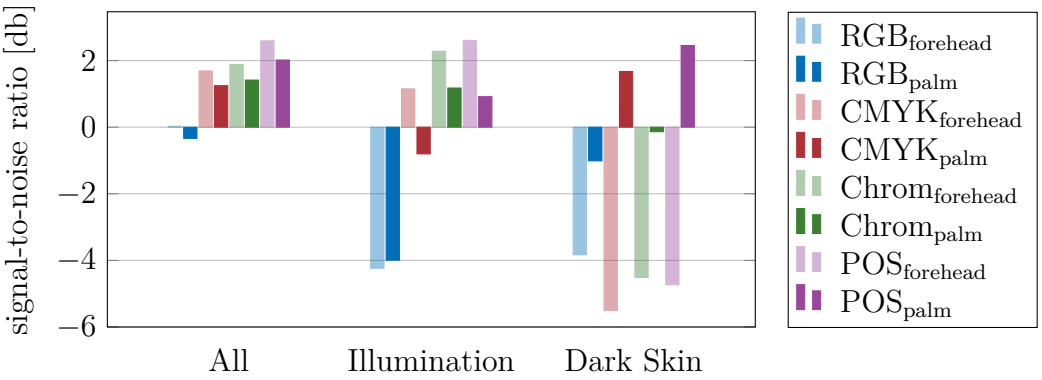


Figure 5.10: SNR of rPPG signals from different colour models and ROIs.

RGB, CMYK and Chrom to confirm previous forehead-based results and avoid selection bias. The SNR is computed according to Equation 5.1, while prediction accuracy is evaluated by a LOSO CV on a single-output RFR for SBP and a joint RFR for DBP, based on the outcomes in Section 4.1.5. To evaluate the effect of the ROI's size on the signal quality, ROIs spanning different numbers of pixels $200 < K_{pixels} < 600$ are analysed regarding rPPG SNR.

In the following, this section is divided into the assessment of ROI positioning in Section 5.3.2 and the analysis of ROI size in Section 5.3.3.

5.3.2 Comparison of ROI Positions

Figure 5.10 depicts the SNRs for the colour models RGB, CMYK, Chrom and POS with the transparent colours representing the values for rPPG signals extracted from the forehead and the bold colours for signals from the palm. Only the measurement scenarios with illumination changes and dark skin are analysed since for these two categories, the contained artifacts are comparable between the two ROIs. On the whole data set as well as on the videos with illumination changes, palm rPPG signals obtain a lower SNR for all colour models. Only the RGB signal with illumination changes shows a small advantage over the forehead-based rPPG, though the SNR is still at SNR = −4.0 db. For dark skin, all colour models show a significant improvement with the palm-based rPPG signal. Whereas RGB and Chrom are still in the negative range, CMYK and POS reach an SNR of 1.7 db and 2.4 db, respectively. This effect can be attributed to the lower melanin concentration at the palm compared to the face [103], and

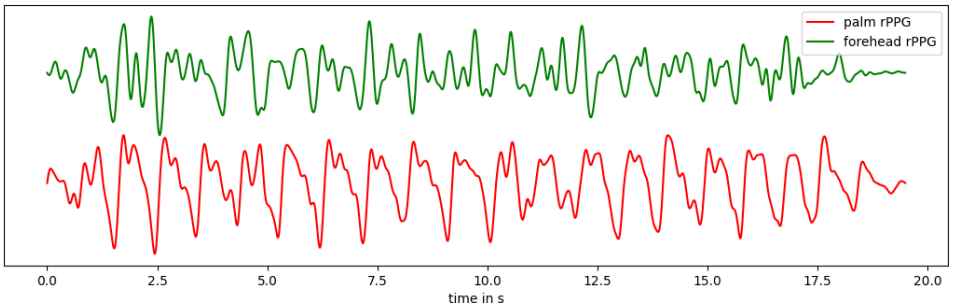


Figure 5.11: rPPG signals from a subject with darker skin. In such cases, signals from the hand exhibit a higher quality whereas signals from the forehead suffer from a low SNR due to the higher concentration of light absorbing melanin.

Table 5.6: SBP and DBP prediction results from the RFR for rPPG signals extracted from different ROIs.

	SBP		**DBP**	
ROI position	**forehead**	**palm**	**forehead**	**palm**
ME ±SD [mmHg]	0.62 ±15.32	0.67 ±16.84	-0.12 ±10.06	-0.51 ±9.55
MAE ±SD [mmHg]	12.17 ±9.32	13.28 ±10.37	7.73 ±6.43	7.14 ±6.35
SD_{MAE} among subjects [mmHg]	±6.68	±7.23	±5.58	±4.61
RMSE	15.33	16.85	10.06	9.56
SD_{RMSE} among subjects [mmHg]	±6.88	±7.74	±6.06	±5.03
Pearson coeff.	0.52	0.32	0.28	0.36
BSH[a] Grade	D	D	C	C

[a]British Society of Hypertension

the results suggest to use the palm as ROI for rPPG extraction for persons with darker skin. A set of example rPPG signals from palm and forehead of a subject with darker skin are shown in Figure 5.11. Both signals were obtained with the POS method. It is evident that the forehead-based signal is distorted by noise such that the true rPPG pulses are not recognisable, whereas in the palm-based signal, the rPPG waveform is clearly visible.

Prediction performance results for the palm and forehead ROIs on a single-output RFR for SBP and a joint RFR for DBP are presented in Table 5.6. It shows that SBP estimation exhibits lower errors with rPPG signals extracted from the forehead whereas DBP estimation performs better on signals extracted

Table 5.7: RMSE for BP prediction for rPPG signals extracted from different ROIs.

	SBP [RMSE in mmHg]				**DBP** [RMSE in mmHg]			
	RFR		**MLP**		**RFR**		**MLP**	
	Joint	**Sep.**	**Joint**	**Sep.**	**Joint**	**Sep.**	**Joint**	**Sep.**
Forehead	**15.40**	**15.33**	**16.27**	18.32	10.06	10.25	10.51	10.10
Palm	16.11	16.85	16.41	**17.23**	**9.56**	**9.96**	**9.95**	**9.99**

Table 5.8: MAE for BP prediction for rPPG signals extracted from different ROIs.

	SBP - MAE ±SD in [mmHg]				**DBP** - MAE ±SD in [mmHg]			
	RFR		**MLP**		**RFR**		**MLP**	
	Joint	**Sep.**	**Joint**	**Sep.**	**Joint**	**Sep.**	**Joint**	**Sep.**
Forehead	**12.35** ±9.19	**12.17** ±9.32	12.65 ±10.23	14.44 ±11.28	7.84 ±6.45	7.94 ±6.48	7.98 ±6.84	7.74 ±10.11
Palm	12.63 ±10.00	13.28 ±10.37	**12.60** ±10.52	**13.11** ±11.17	**7.14** ±6.35	**7.44** ±6.34	**7.70** ±6.30	**7.58** ±6.12

from the palm. To rule out the influence of the specific models' properties, both ROIs are further evaluated on the models tested in Section 4.1.5, i.e. a separate (single-output) and joint (two-output) MLP as well as a separate and joint RFR. The results of these tests are shown in Table 5.7 for RMSE values and in Table 5.8 for the corresponding MAE ±SD. For all models, DBP prediction performs better on rPPG signals at the palm. In contrast, the results for SBP are less consistent. Both RFR models obtain better results with forehead-based rPPG signals, whereas the separate MLP obtains lower errors on palm-based rPPG signals. For the joint MLP, the outcome depends on the error metric.

Pasyk et al. [104] analysed cutaneous tissue of 20 different body areas, and found the highest capillary density of the papillary dermis at the palms with 5.3 %, while the facial areas exhibit a value of only 3.6 %. Although facial areas show a higher capillary density in the reticular dermis, the palm still exhibits the highest density overall [104]. Therefore, this area might be best suited for rPPG extraction as reflected by the prediction accuracies for DBP. However, the results for SBP do not coincide. Especially SBP is affected by pressure amplification along the arterial tree and might be more sensitive to subject-individual distances of the measurement location to the heart. Thereby, the distance and variation

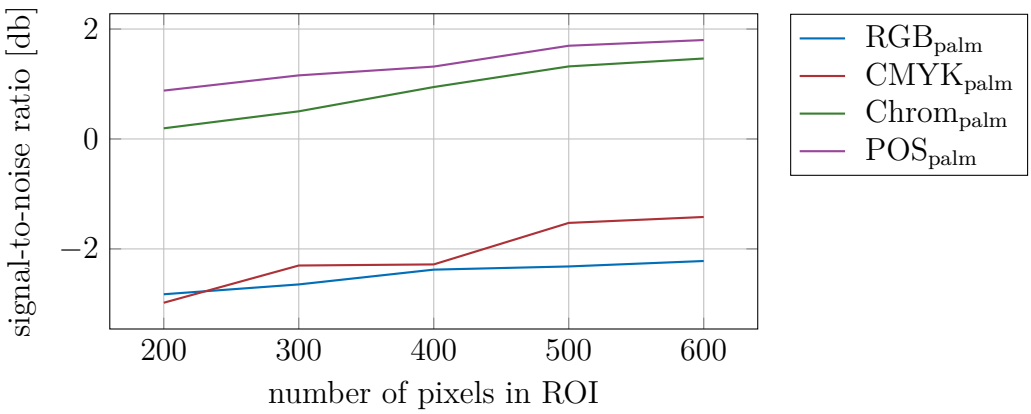


Figure 5.12: SNR of rPPG signals extracted from differently sized ROIs.

in distance to the palm is greater than the distance to the forehead and could explain the lower error on forehead-based rPPG signals.

Differences in signals from palm and forehead caused by sympathetic and parasympathetic activation [20] could not be observed nor were they deliberately provoked during data acquisition. Sympathetic activation leads to a systemic BP increase which is mainly driven by vasoconstriction in cutaneous vessels in the extremities whereas facial vasoconstriction remains comparatively low [105]. It has been shown that the fingers' and other extremities' sympathetic responsiveness is more pronounced than the head's sympathetic responsiveness [106, 107]. In contrast, facial regions are subjected to parasympathetic vasodilation [20]. Their influences on video-based BP measurement should be investigated in future studies.

5.3.3 Analysis of ROI Size

The influence of ROI size on the signal quality is presented in Figure 5.12. All investigated colour models show an improvement with a higher number of pixels, e.g. POS obtains an SNR of $SNR_{600} = 1.80$ db for 600 pixels compared to $SNR_{200} = 0.88$ db for 200 pixels. This shows that sensor noise is better averaged out within a larger ROI. On the other hand, a too large ROI would contain rPPG signals with big phase differences such that spatial averaging of these pixel values would affect the waveform and signal level. Chen et al. [108] measured PWV invasively and obtained values between 8 $\frac{m}{s}$ and 30 $\frac{m}{s}$. Depending on the exact distance to the camera, the ROIs spanning 600 pixels have an estimated height

of 5 cm, such that a slow pulse wave with $v_{PWV} = 8\frac{\mathrm{m}}{\mathrm{s}}$ would take 6.25 ms to pass the ROI. Hence, even for a high heart rate of 160 bpm and a low PWV, the phase difference is only 6° and thus not affecting the signal waveform significantly.

5.3.4 Conclusion on the Selection of the Region of Interest

Although forehead-based rPPG signals exhibit an overall lower SNR, prediction errors show a clear preference of palm-based signals for DBP. In contrast, for SBP prediction, mostly forehead-based rPPG signals perform better. Considering the quality improvement with palm-based signals for darker skin tones, the palm should be chosen as ROI to provide a BP estimation method applicable for a diverse population. Further, a large ROI has proven to increase SNR and will be applied in the following approach.

Chapter 6

Convolutional Neural Networks for Blood Pressure Estimation

To rule out possible error sources in BP estimation from rPPG data, the model proposed in this chapter incorporates the following changes compared to the feature-based model in Chapter 4:

1. A CNN is employed as regression method for automatic feature extraction from rPPG signals. This allows finding more relevant features than the hand-crafted ones.

2. To further rule out a negative impact of low rPPG signal quality, the proposed method is first validated on conventional PPG signals. For this, the MIMIC III data set [56] is used which offers a much larger data base with more reliable, invasive BP references. The resulting PPG-based model is finally used for transfer learning to the rPPG signals.

3. On the above mentioned large PPG data set, the generalisation capabilities of a universal model are evaluated, and the effect of individual calibration of such model is explored.

6.1 Method

By applying transfer learning from the PPG to the rPPG domain, we can take advantage of large amounts of data while benefiting from a higher signal quality in PPG signals as well as in the BP references. Since the MIMIC III data set

[56] used for this purpose consists of raw biosignals from ICU patients, it also comprises pathological BP and PPG signals. Moreover, many signals suffer from defective sensor attachment or movement artefacts. Therefore, an extensive signal filtering and preprocessing is performed as detailed in Section 6.1.1 which is adapted to the shortcomings of this data set. Further, different time and frequency domain representations of the segmented PPG pulses are considered as input to the ANN.

The regression ANN is described in Section 6.1.2 and consists of several convolutional layers to learn a feature representation of the input automatically. The network is first trained as a universal model on the data of the training subjects. Then, the resulting model is calibrated individually on held off fine-tuning data from the test subjects. This calibration process is analysed for PPG signals and finally also adopted to test the performance of the proposed method on rPPG data by transfer-learning. To do so, the last layers need to be fine-tuned due to the differences in signal morphology between PPG and rPPG (see Section 5.2.2).

6.1.1 Signal Preprocessing

As a real world data set, the MIMIC III waveform data base [56] contains a large part of unusable signals due to technical issues, such as sensor attachment problems and movement artifacts, as well as pathological alterations of the signals caused by the medical condition or drugs. Since these influences each manifest differently, a more extensive preprocessing is performed compared to the rPPG data set. Figure 6.1 shows the signal processing pipeline for the MIMIC III waveform data.

First, the PPG signal is smoothed by applying a Savitzky-Golay filter [109] with a window length of 31 and a polynomial order of three. In contrast to other filters, the Savitzky-Golay filter does not cut off high frequency components of the signal, but includes them into the calculation. Therefore, extrema are well preserved and do not suffer from flattening and shifting as caused by other methods. Subsequently, plateaus that are longer than 19 samples are deleted from the PPG and ABP signals, which can be caused by PPG sensor detachment or ABP sensor issues. To do so, a plateau is defined as no change up to the first decimal digit for ABP signals and up to the fourth decimal digit for PPG signals. With the sampling rate of $f_s = 125$ Hz, 19 samples denote 152 ms and have shown

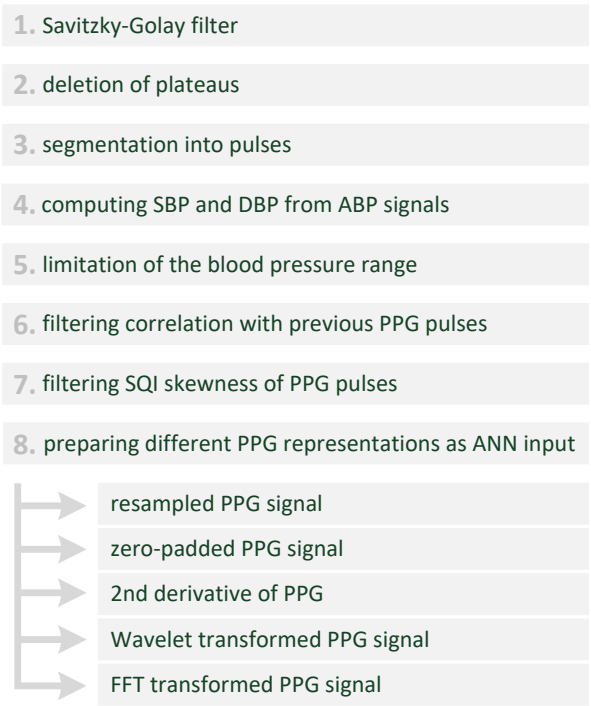


Figure 6.1: Complete pipeline for MIMIC PPG and ABP signal processing.

good accuracy in an empirical evaluation. The remaining signal is segmented into separate PPG pulses and corresponding ABP signal sections based on the PPG minima, i.e. diastoles.

In the next step, these ABP sections are examined for the correct number of peaks and for physiologically possible BP values. Pulses are discarded that exhibit more than one global minimum or maximum or BP values of SBP $<$ 60 mmHg, SBP $>$ 200 mmHg, DBP $<$ 50 mmHg or DBP $>$ 120 mmHg. This allowed BP range is larger than in comparable studies [83, 82, 84]. Further, each PPG pulse is correlated with its three predecessors and must meet an average correlation coefficient of $R_{PPG} \geqslant 0.3$.

A Signal Quality Index (SQI) based on skewness is the final criterion to evaluate signal quality, as increased skewness was found to be associated with corrupted PPG signals [110]. In a comparative study, Elgendi et al. [111] showed that skewness is the best metric to assess the quality of PPG signals compared to perfusion, relative power, kurtosis and entropy. Therefore, the SQI is defined

as

$$SQI = \frac{1}{l}\sum_{i=1}^{l}[u_i - \hat{\mu}_u/\sigma]^3 \tag{6.1}$$

where l denotes the number of samples in the PPG signal, and $\hat{\mu}_u$ and σ are mean and standard deviation of all samples u, respectively. Based on empirical analysis and a previous study [111], the SQI threshold is set to $SQI_{PPG} = 0.0$ for the MIMIC data. All remaining pulses are used for the data set. For rPPG signals, the threshold is relaxed to $SQI_{rPPG} = 0.2$ to avoid elimination of too many samples in the already small data set.

To find the best input representation for the ANN, the PPG pulses undergo further processing. Since all pulses need to have the same length for the input layer of the ANN, the two employed options are either to re-sample all pulses to a length of $l = 100$ samples or to maintain the same sampling rate for all signals and zero-pad the pulses such that all input data have the same length ($l = 100$ samples).

Another variation of the PPG pulses is their second derivative PPG" which represents the acceleration of changes in blood volume, thus possibly amplifying alterations in the PPG waveform. Finally, the PPG pulses are also transformed to the frequency domain using a single-level Discrete Wavelet Transform (DWT) with Haar wavelets and an FFT. In the feature selection in Chapter 4.3 as well as in previous studies [14, 59], the frequency domain has been shown to contain valuable information for BP prediction. Whereas a Fourier-transformed signal only provides the overall frequency composition, a Wavelet-transformation extracts both spectral and temporal information.

6.1.2 Neural Network

In two comparative studies on ANN architectures for PPG-based BP estimation [67, 64], ResNet architectures obtained superior results than AlexNet [62], LSTM networks and an MLP. A large model with three parallel ResNet architectures proposed by Slapnicar et al. [87] did not outperform the original ResNet-50 structure, and an additional use of derivatives did not improve performance [61]. Hence, for the proposed method, a ResNet-50 architecture is adapted to one-dimensional input signals, and the number of layers is reduced to avoid overfitting [112].

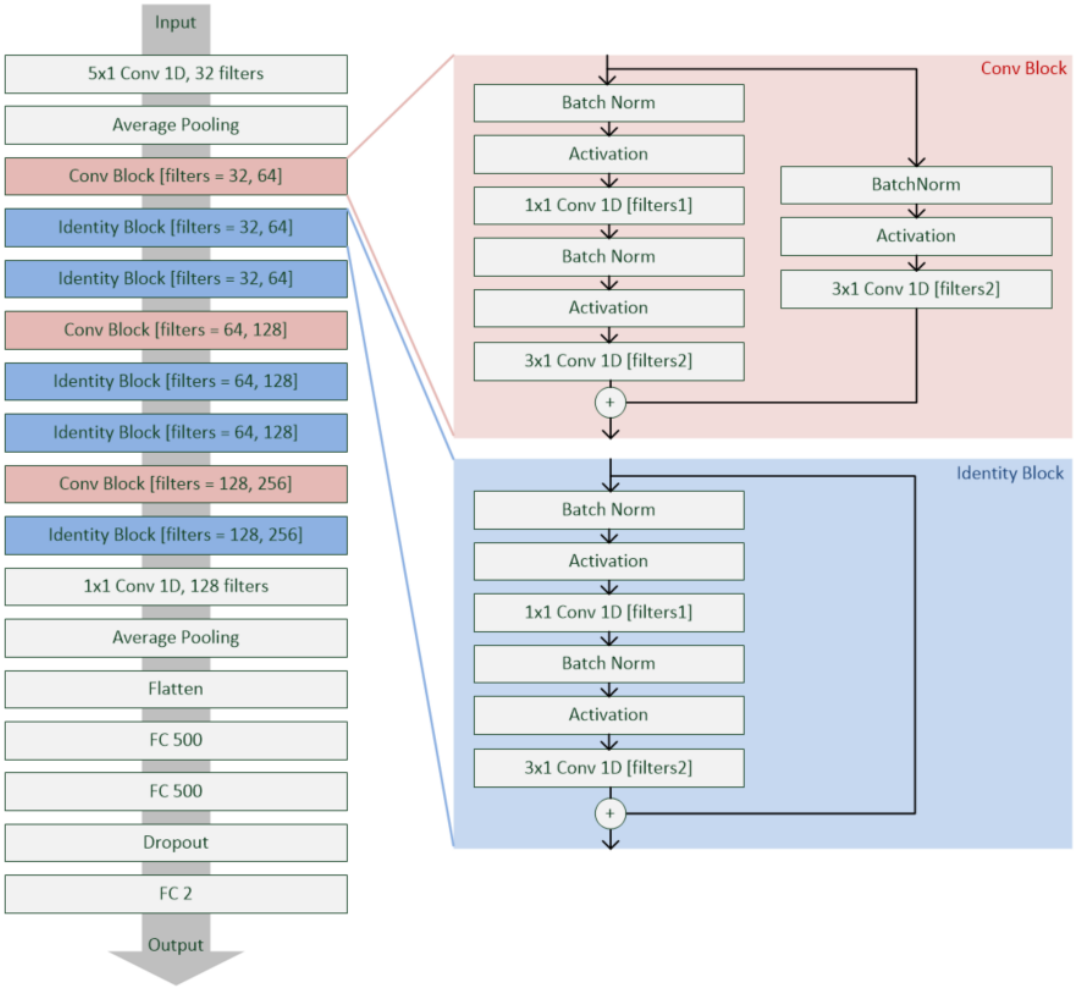


Figure 6.2: Deployed ANN architecture [112]. The proposed network consists of convolutional layers for an automatic feature extraction from the PPG signals and is structured in eight residual blocks.

For an automatic feature extraction, the ANN comprises one-dimensional convolutional layers. The layers are organised in residual blocks, as they have shown favourable convergence behaviours for deep neural networks [63]. The central idea of residual networks is to learn the additive residual function H with respect to the identity mapping $m(i_l) = i_l$, such that

$$o_l = m(i_l) + H(i_l, W_l) \tag{6.2}$$

where i_l and o_l are input and output of the l-th residual unit, and W_l its set of weights. This is realised by skip connections which create a direct propagation path from one residual unit to any other unit. He et al. [113] found that us-

ing identity mappings as skip connections and after-addition activation further improve training and generalisation of the model. Hence, their proposed residual unit structure is adopted in the ANN as shown in the network architecture illustrated in Figure 6.2. $\tanh(x)$ is selected as activation function for convolutional layers and activation layers in the residual blocks, whereas for the last fully connected layer, a linear activation function is used.

The first convolutional layer in each residual block has a kernel size of $\hat{K} = 1 \times 1 \times g_1$, and the second convolutional layer as well as those in the skip connection exhibit $\hat{K} = 3 \times 1 \times g_2$, with g_1 and g_2 being the numbers of input channels, respectively. Whereas $3 \times 1 \times g_2$-kernels learn spatial relationships, $1 \times 1 \times g_1$-kernels are used to reduce dimensions by combining cross-channel information into fewer channels. Only in the convolutional blocks, $1 \times 1 \times g_1$-kernels do not reduce the number of channels but serve as a cross-channel projection layers refining the already extracted features. In the proposed architecture, $3 \times 1 \times g_2$-kernels are used as a bottleneck layer to calculate features from the condensed low-dimension input and increase the number of channels again. Further, for all layers, stride is set to $s = 1$ to avoid loss of spatial resolution, and zero-padding is performed evenly on both sides such that each output has the same size as the corresponding input data, except the last pooling layer which exhibits $s = 2$.

The following loss function $L(\hat{o}, o)$ and Adam optimiser are employed with a start learning rate of $\mathrm{lr} = 0.0001$ and decay $\epsilon = 0.001$:

$$L(\hat{o}, o) = \frac{1}{n} \sum_{i=1}^{n} \log(\cosh(o_i - \hat{o}_i)) \tag{6.3}$$

where $\hat{o}$ is the expected output and o the predicted output of the ANN, and n denotes the batch size. The model is trained with batches of size $n = 128$ until the loss stops decreasing. Hyperparameter tuning is performed in a 5-fold cross-validation for the batch size, learning rate, dropout rate and pooling size as well as for the selection of the activation function and loss function.

The effect of fine tuning is analysed (1) for personalisation of the model for MIMIC test subjects and (2) for transfer learning from PPG to rPPG data. The best universal model is used for personalisation on test subjects by fine-tuning the last layers on a small part of the test subject's data. For fine tuning the network, the first layers and residual blocks are frozen such that only the weights of the last three fully connected layers are further adjusted during training. Hence, 251,700

parameters remain trainable. The initial learning rate is set to $lr = 0.00001$ as the layers have already been trained, and the batch size is reduced to $n = 60$ to account for the few tune data.

Finally, for transfer learning from the PPG to the rPPG domain, the same fine-tuning process is performed with the rPPG data set and four trainable layers, i.e., 3,452,002 trainable parameters. The weights of all fully connected layers are adjusted while the weights and kernels in the convolutional layers remain the same. The initial learning rate is defined as $lr = 0.00001$, and the batch size is set to $n = 60$. For further personalisation of this rPPG model, three layers are fine-tuned with a batch size of $n = 3$ due to the small number of rPPG tune data per patient.

6.2 Experimental Setup

The previously described method and ANN model are first trained and tested on the MIMIC III waveform data base described in Section 6.2.1. Subsequently, transfer learning is performed with the rPPG data set detailed in Section 6.2.2. To do so, the optimisation process described in Section 6.2.3 is followed.

6.2.1 PPG Data Set

After preprocessing 483 records of the MIMIC III waveform data base, 854,073 samples from 180 patients are obtained with a mean BP $\pm$ SD of 164.6 $\pm$ 23.25 mmHg for SBP and 78.8 $\pm$ 17.6 mmHg for DBP. The number of samples per patient was limited to 20,000 to reduce the influence of single patients.

The data set is split into 116 training and 64 test subjects as illustrated in Figure 6.3 obtaining 383,185 and 415,888 samples, respectively. Both subsets are further divided into 20% validation subjects and 80% actual training and test subjects. This split with two validation sets is chosen to account for the training and validation of the universal model as well as a subsequent optimisation of the fine-tuning process. The data set is split at patient-level, such that all data of each patient only appear in the training set, validation set A, validation set B or test set. In the sequentially ordered records, the first 300 PPG pulses of each test (and validation B) subject are held off for fine tuning of the model, and the subsequent 300 pulses are eliminated from the data set to avoid too much

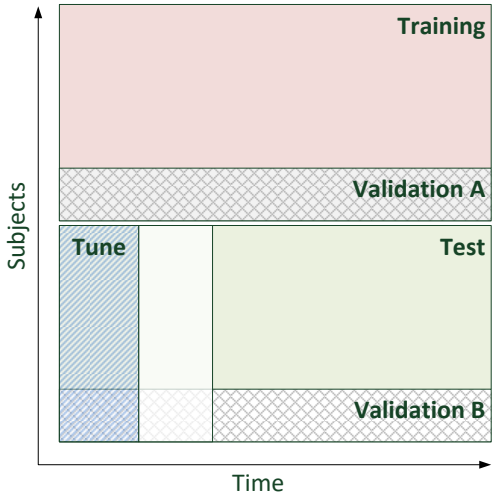
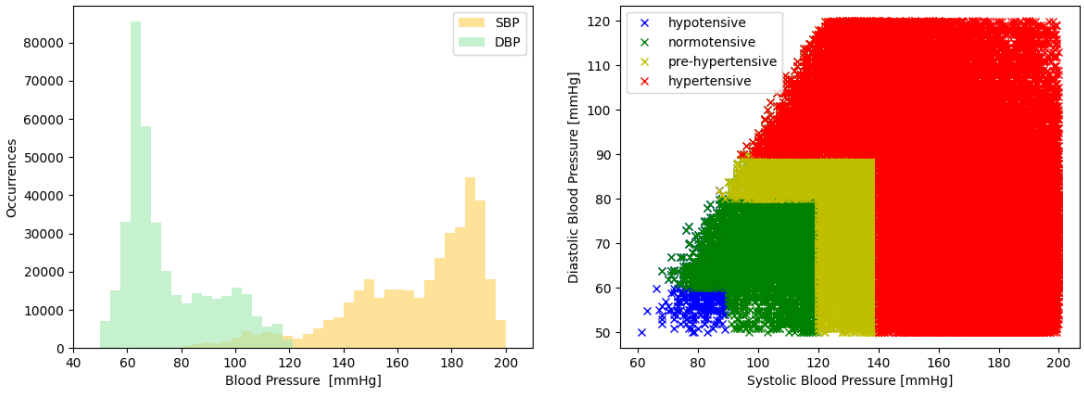


Figure 6.3: Split of PPG data set. Training and validation set A are used for development of the universal model, tune and validation set B are used for fine tuning and development of the calibrated models, and the test set is employed for evaluation of both. A subject-based data split is conducted, and the tune data is selected as the first 300 samples of each test subject.

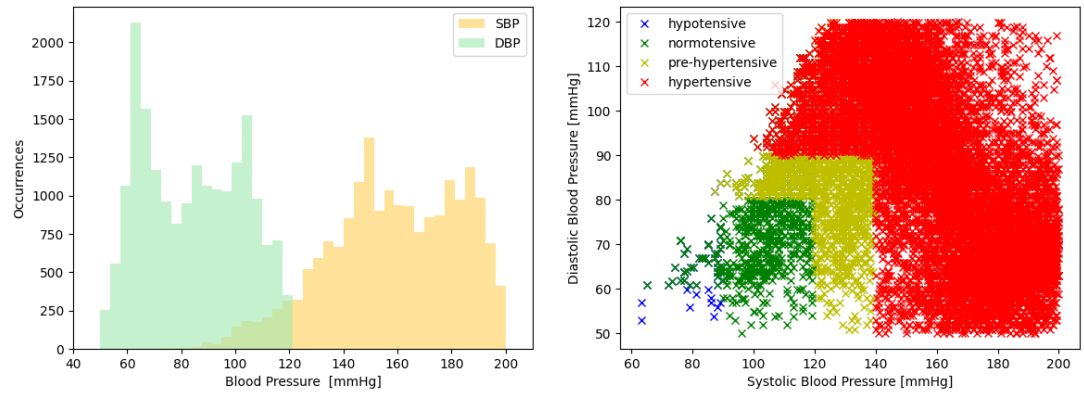
similarity between tune and test data and to simulate a real-world application. The rest of the test subjects' PPG pulses form the test set.

6.2.2 rPPG Data Set

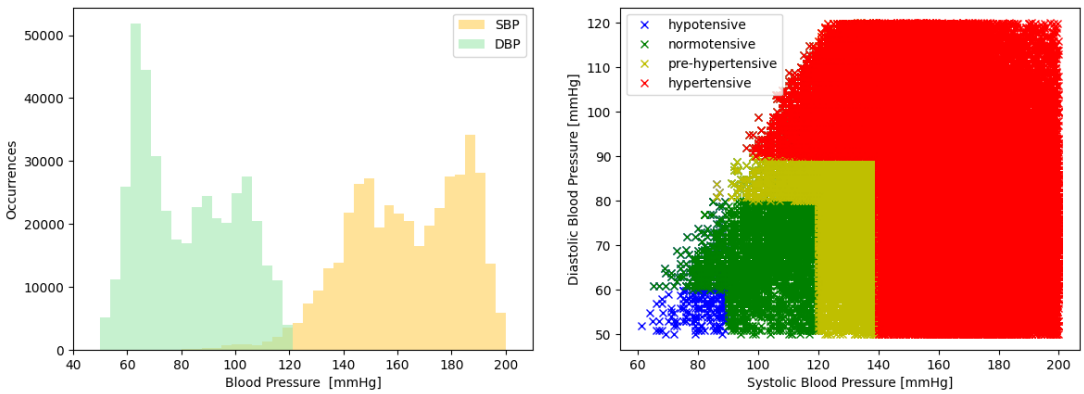
For evaluation of transfer learning capabilities from PPG to rPPG signals, the previously acquired data set described in Chapter 4.2 is used. The same signal processing pipeline (Section 6.1.1) is followed to ensure maximal input similarity. After pre-processing, a total of 709 samples from 28 subjects remain. Due to the very small number of samples per subject, the first 40% of each subject's samples, but at least 8 samples, are employed for tuning and the rest is assigned to the test data set.



(a) training data



(b) tune data



(c) test data

Figure 6.4: Histogramm and distribution of BP values in the PPG data set. Hypotensive BP (SBP $<$ 90 mmHg and DBP $<$ 60 mmHg) is depicted in blue, normotensive BP (90 mmHg $\leqslant$ SBP $<$ 120 mmHg and 60 mmHg $\leqslant$ DBP $<$ 80 mmHg) in green, pre-hypertensive BP (120 mmHg $\leqslant$ SBP $<$ 140 mmHg and 80 mmHg $\leqslant$ DBP $<$ 90 mmHg) in yellow and hypertensive BP (140 mmHg $\leqslant$ SBP and 90 mmHg $\leqslant$ DBP) in red.

6.2.3 Model Optimisation

A 5-fold cross-validation is performed on the training data set for determining the optimal configuration of the ANN in a grid search. Further, performances for different input representations of the PPG signals are compared. For this, the PPG and its second derivative are considered, where for each, a zero-padded and a resampled version is tested, as well as frequency domain representations obtained by an FFT and a single-level DWT.

For interpreting the results, layer activation of the ANN is analysed for various input samples. Moreover, a correlation analysis of PPG waveforms is performed. To do so, PPG pulses are grouped according to their reference BP in intervals of 5 mmHg, and the Pearson correlation is computed between the PPG waveforms for all combinations of BP levels. This analysis is conducted for inter-individual combinations as well as intra-individual ones.

To investigate the impact of individual calibration on the prediction accuracy, the best universal model according validation set A is fine-tuned individually on each subject's tune data and tested on the latter part of the signal, i.e., the test data. Different numbers of layers are fine-tuned.

Next, transfer learning from PPG signals to the rPPG domain is performed. To do so, a LOSO CV is conducted where the universal PPG model is fine-tuned on the rPPG data of 27 subjects and subsequently tested on the test data of the remaining subject. Again, different numbers of layers are tested for fine-tuning. In the next iteration, these new universal rPPG models are further personalised for each corresponding individual by fine-tuning on the subject's tune data.

6.3 Results and Discussion

In the following sections, the results of the previously described experiments are presented and analysed. First, Section 6.3.1 details the results of the universal model trained on the MIMIC III data set, where obtained results are compared to other state-of-the-art BP prediction models and an example record is assessed. Further, the CNN layer activation is analysed to increase understanding of the model's decision making and of relevant PPG features.

Subsequently, in Section 6.3.2, personalisation of the universal PPG model is investigated and analysed based on a detailed error distribution and predictions

of example records. The results are evaluated in comparison to the current state of research and strategies for enhancing the personalisation process are derived.

Finally, the results for transfer learning from the PPG to the rPPG domain are presented in Section 6.3.3, including an universal model as well as model personalisation.

6.3.1 Universal PPG Model

After grid search for hyperparameter optimisation, the best ANN on the validation set A is further tested for the different input representations. Table 6.1 details BP prediction errors on the test data set. Best prediction accuracies are achieved with zero-padded time domain signals, where the second derivative of the PPG (PPG") has a slight advantage over the PPG signal itself. The PPG" data representation obtains an MAE ±SD of 8.73 ±7.36 mmHg for SBP and 8.07 ±6.86 mmHg for DBP. Although both signals contain the same information, the second derivative amplifies differences in signal morphology.

Resampled signals and frequency domain representations on the other hand, exhibit prediction errors similar to the mean regressor with an MAE ±SD of 18.24 ±11.73 mmHg for SBP and 16.10 ±10.65 mmHg for DBP. This shows that the ANN relies on information about cycle lengths and heart rate which are lost during resampling the signal to a standardised length. Moreover, predictions based on frequency domain signals might suffer from the low frequency resolution due to the short signal sections. Xing et al. [14] showed that the frequency domain of PPG signals is well suited for BP estimation from longer input signals. However, for the purpose of beat-to-beat prediction, the resolution appears to be insufficient. The second derivative of the PPG further exhibits the lowest SD in prediction errors among subjects.

The obtained results are not within the AAMI standard [96] and obtain BSH [97] grade C for DBP and D for SBP. Still, the prediction errors of this work are lower than the results of [87, 114] as well as all four models tested by Schrumpf et al. [67] with their best results being MAE_{SBP} = 16.4 mmHg and MAE_{DBP} = 8.5 mmHg. Qin et al. [84] report significantly lower errors for DBP prediction with MAE_{SBP} = 7.9 mmHg and MAE_{DBP} = 4.1 mmHg, however, it has to be noted that they use 10-second signal sections as model input. Similarly, Schlesinger et al. [82] consider 30-second intervals and obtain MAE_{SBP} = 7.3 mmHg and

Table 6.1: Prediction results (in mmHg) from the universal CNN for different PPG input representations.

DBP							
	zero-padded				**resampled**		
	PPG	**PPG'**	**FFT**	**DWT**	**PPG**	**PPG'**	Mean Reg.
ME ±SD	-1.22 ±11.28	-1.95 ±10.41	-2.15 ±18.81	3.36 ±17.92	-3.06 ±18.33	-0.36 ±18.02	-6.98 ±18.00
MAE ±SD	8.73 ±7.25	8.07 ±6.86	15.50 ±10.88	14.54 ±10.49	15.09 ±10.84	14.65 ±10.50	16.10 ±10.65
SD_{MAE} among subjects	±3.10	±2.82	±4.41	±4.34	±4.31	±4.10	±5.85
RMSE	11.35	10.59	18.93	17.93	18.58	18.02	19.31
SD_{RMSE} among subjects	±3.41	±3.18	±4.37	±4.49	±4.47	±4.30	±5.90
BSH[a] Grade	D	C	D	D	D	D	D
SBP							
	zero-padded				**resampled**		
	PPG	**PPG'**	**FFT**	**DWT**	**PPG**	**PPG'**	Mean Reg.
ME ±SD	-0.54 ±11.92	-0.77 ±11.39	-0.21 ±20.79	-2.10 ±20.96	-0.23 ±21.92	-1.33 ±20.80	3.01 ±21.47
MAE ±SD	9.16 ±7.65	8.73 ±7.36	16.76 ±12.31	16.89 ±12.60	17.51 ±13.19	16.73 ±12.44	18.24 ±11.73
SD_{MAE} among subjects	±2.87	±2.89	±3.82	±4.33	±4.34	±3.89	±7.24
RMSE	11.94	11.42	20.79	21.07	21.92	20.85	21.68
SD_{RMSE} among subjects	±3.43	±3.33	±4.16	±4.67	±4.66	±4.26	±7.97
BSH[a] Grade	D	D	D	D	D	D	D

[a]British Society of Hypertension

MAE_{DBP} = 3.9 mmHg. Longer signal sections reduce the impact of noisy pulses in reference ABP and input PPG but do not allow a beat-to-beat measurement. Further, many related studies use proprietary data sets [83] or data with a smaller BP range [82, 83, 84]. SBP prediction with the proposed approach shows competitive results, even compared to methods that consider ECG in addition to PPG signals [12, 115, 116, 117].

Figure 6.5 depicts the true and predicted BP values of test patient 3403232 in the correct sequential order. The patient's ABP exhibits strong fluctuations over time with a drop of 80 mmHg within a few heart beats. Even though the predicted and true BP levels are not perfectly aligned with discrepancies of up to 20 mmHg in some sections, the plot shows that the predicted values follow the changes in the true ABP curve closely.

However, the plot of the true ABP values exhibits substantial noise as well. Smaller low-frequency variations, which can be observed in Figure 6.5 between sample 200 and 400, can be attributed to normal breathing. Respiration changes intratoracic pressure, thus affecting cardiac output and systemic pressure and leading to an inspiratory decrease in ABP of up to 10 mmHg [118]. Further influ-

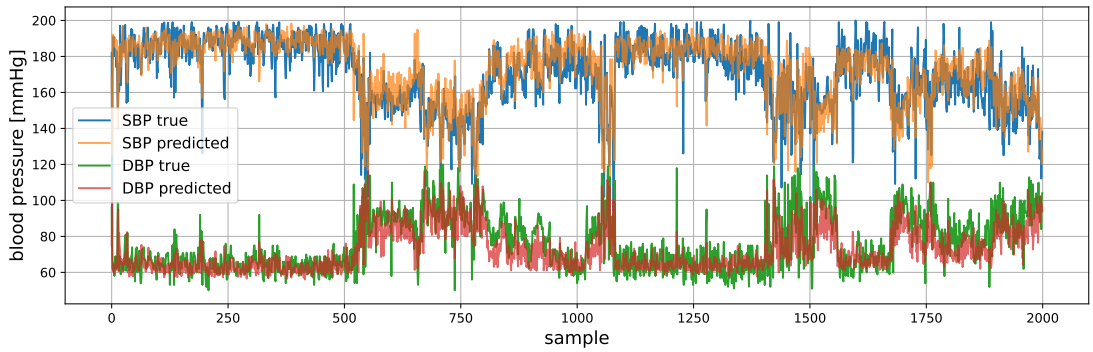


Figure 6.5: True and predicted BP values of a test patient in sequential order. [112]

ences on invasive ABP measurement are movement artefacts (patient or tubing system), disturbances of the sensor system by other electrical devices [118], inappropriate damping [119] and the employed filtering methods of the sensor system [120]. Other published methods on BP estimation from PPG signals mostly consider longer signal sections ranging from 8-second [85] to 30-second intervals [82]. Longer sections eliminate high-frequency noise and variance in the ABP reference signal and cancel out beat-to-beat variations in the predicted BP. Hence, these methods obtain lower prediction errors as opposed to the beat-to-beat approach presented in this work.

It should be noted that the model used in this work does not consider the temporal relation of adjacent PPG pulses but predicts each new pulse independently from previous input information. The time series nature of the BP signal suggests taking advantage of the predecessor information by employing recurrent neural networks. This could further improve BP prediction and smooth its noisy curve.

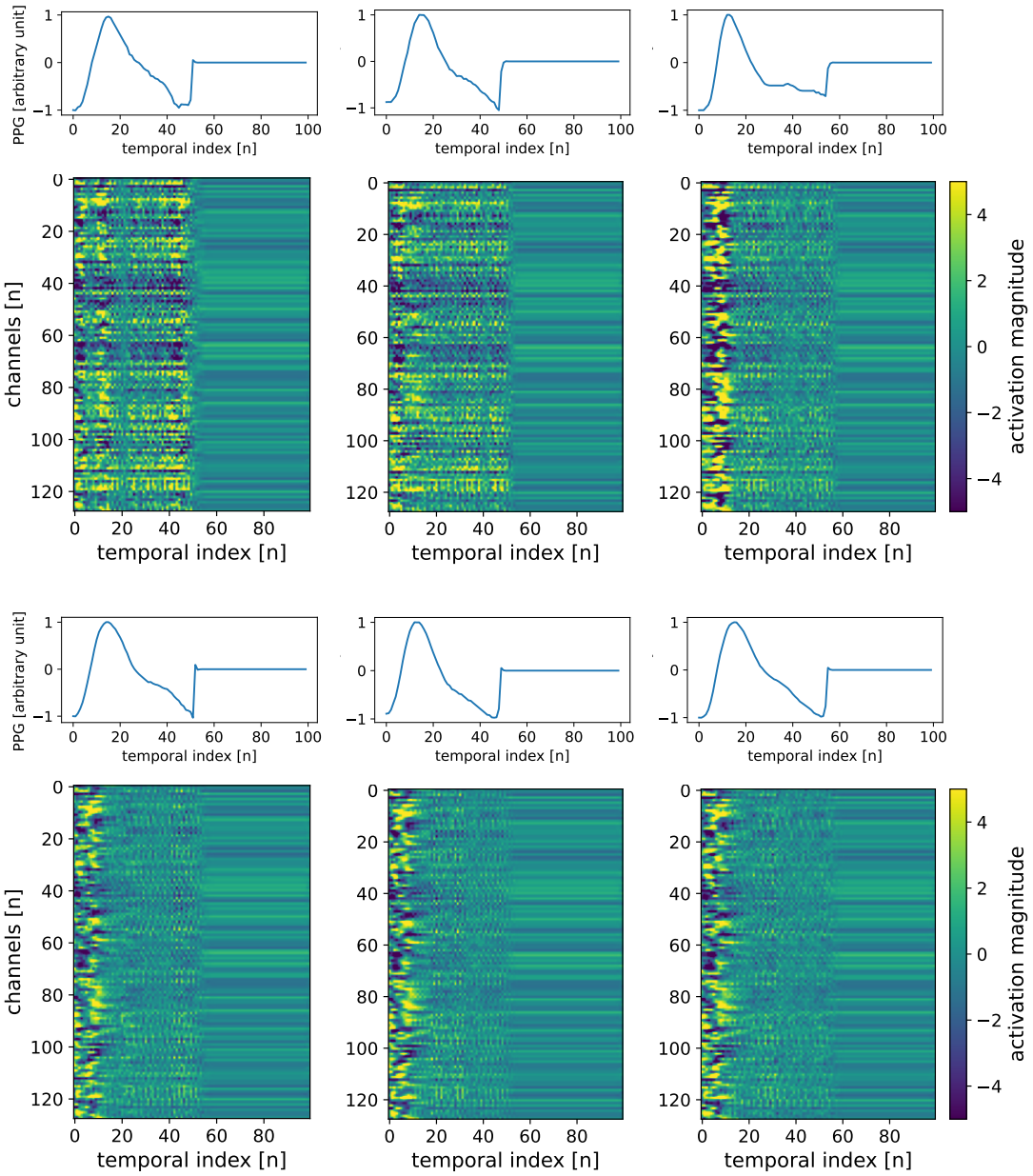


Figure 6.6: Layer activations (heat maps) of the 50th layer for inputs (line plots) with increasing SBP from left (SBP = 115 mmHg) to right (SBP = 180 mmHg).

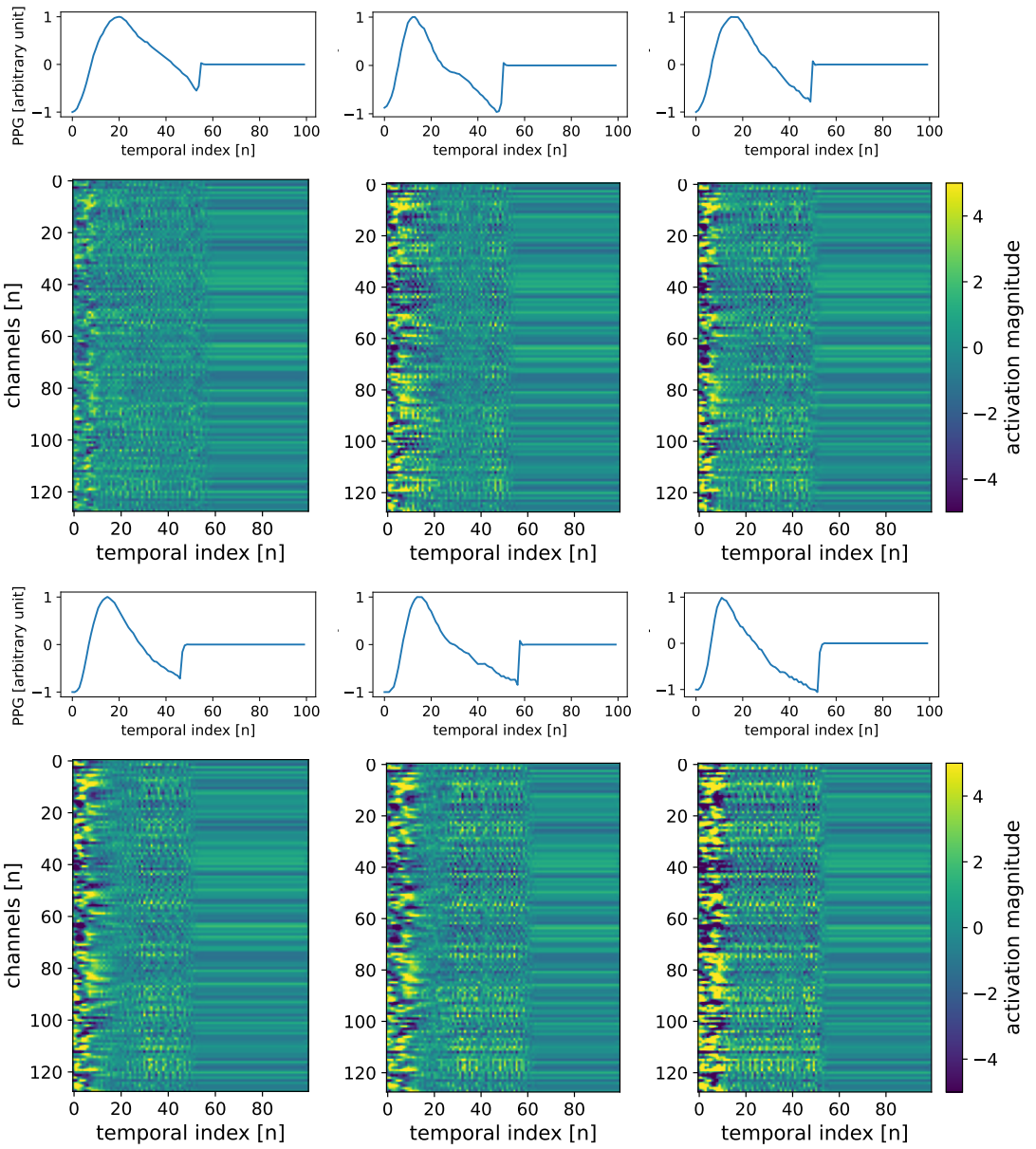
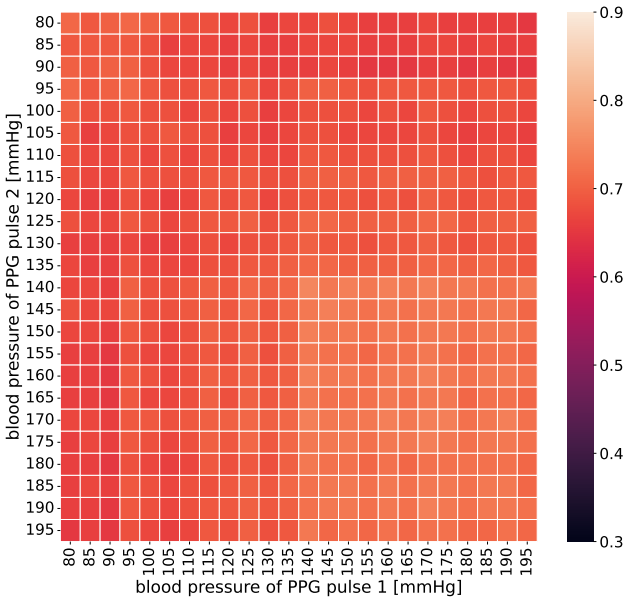


Figure 6.7: Layer activations (heat maps) of the 50th layer for inputs (line plots) with increasing DBP from left (DBP = 65 mmHg) to right (DBP = 103 mmHg).

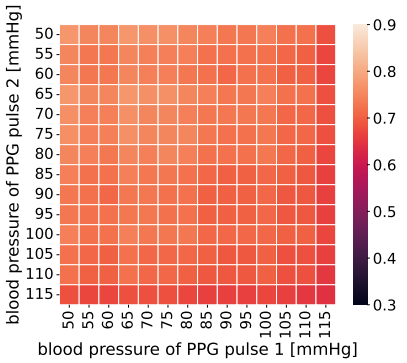
To better understand the models decision making, Figure 6.6 and 6.7 display the activations of layer 50 for test samples at various SBP and DBP levels. Figure 6.6 shows the layer activation for inputs of increasing SBP from left to right where DBP is kept almost constant. The output of each filter in the visualised layer is stacked on top of each other, with yellow representing strong positive activations and blue representing strong negative activations. In the process, the spatial information in preserved such that a direct connection can be made between the signal and the filter activation at the same index. In the first convolutional layers, a CNN learns to detect coarse-grain features like edges and gradients, and in the deeper layers, it learns more complicated features by combining features of earlier layers. The displayed activations of layer 50 show the outputs of the sixth ResNet block. It is evident that for lower SBP, a strong activation occurs along the whole PPG pulse, whereas for higher SBP, activation becomes sparser after the systolic peak. Figure 6.7 shows the layer activation for inputs with increasing DBP, respectively, where SBP is kept nearly constant. Differences in layer activation are mostly limited to the systolic upstroke, where activation becomes stronger with larger DBP values.

It seems counter-intuitive that DBP prediction relies on the systole, whereas for SBP prediction, the differences in filter activation are located in the diastole part of the PPG. However, this finding is in accordance with the feature selection results in Section 4.3, where the DBP features considering the systole can be related to the MAP, vasodilation and vessel stiffness. All three properties are strong indicators for the DBP level [94, 95]. On the other hand, area under the curve and similar features describing the whole waveform comprise indicators for stroke volume and vascular capacitance, which are the main influences on SBP [95].

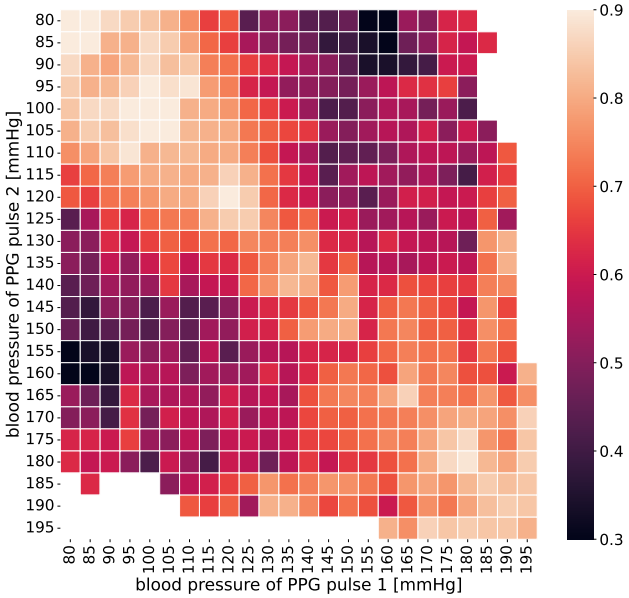
Figure 6.8 displays the average Pearson correlation coefficients between two PPG pulses corresponding to different BP levels. The correlation was calculated between pulses of the same subjects as well as pulses of different subjects. It can be observed that when considering the whole waveform, inter-individual PPG correlation is moderate across all combinations of BP levels. This shows that the same BP does not manifest as the same PPG waveform in different subjects, and a similar PPG morphology in two subjects might represent a very different BP level. Although an ANN will focus on more generally applicable characteristics, the



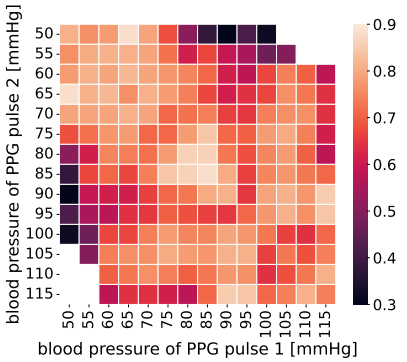
(a) inter-individual correlation sorted by SBP



(b) inter-individual correlation sorted by DBP



(c) intra-individual correlation sorted by SBP



(d) intra-individual correlation sorted by DBP

Figure 6.8: Correlation between PPG pulses of different persons sorted by SBP (a) and DBP (b), and correlation between PPG pulses of the same person sorted by SBP (c) and DBP (d).

intra-individual correlations in Figures 6.8c and 6.8d suggest that a personalised model might benefit from overall more distinctive waveforms for different BP levels and thus a clearer differentiation for the model.

Table 6.2: Prediction results (in mmHg) from the individually calibrated CNN for different PPG input representations.

DBP					
	PPG (zero-padded)		**PPG" (zero-padded)**		
	universal	**calibrated**	universal	**calibrated**	Mean Reg.
ME ±SD	-1.22 ±11.28	1.41 ±11.14	-1.95 ±10.41	1.12 ±10.87	-6.98 ±18.00
MAE ±SD	8.73 ±7.25	8.72 ±7.08	8.07 ±6.86	8.42 ±6.88	16.10 ±10.65
SD_{MAE} among subjects	±3.10	±2.74	±2.82	±2.66	±5.85
RMSE	11.35	11.23	10.59	10.92	19.31
SD_{RMSE} among subjects	±3.41	±3.03	±3.18	±2.96	±5.90
BSH[a] Grade	D	D	C	D	D
SBP					
	PPG (zero-padded)		**PPG" (zero-padded)**		
	universal	**calibrated**	universal	**calibrated**	Mean Reg.
ME ±SD	-0.54 ±11.92	-0.87 ±11.26	-0.77 ±11.39	-0.82 ±10.71	3.01 ±21.47
MAE ±SD	9.16 ±7.65	8.76 ±7.13	8.73 ±7.36	8.26 ±6.86	18.24 ±11.73
SD_{MAE} among subjects	±2.87	±3.19	±2.89	±2.94	±7.24
RMSE	11.94	11.29	11.42	10.74	21.68
SD_{RMSE} among subjects	±3.43	±3.77	±3.33	±3.39	±7.97
BSH[a] Grade	D	D	D	D	D

[a]British Society of Hypertension

6.3.2 Personalised PPG Models

All models from Section 6.3.1 are calibrated individually by fine tuning of the universal model on the test subject's corresponding tune data. Table 6.2 contains the results for zero-padded time domain signals.SBP prediction from PPG" is improved by individual calibration by a difference in MAE of -0.45 mmHg to 8.26 ±6.86 mmHg. For DBP however, the prediction error increases slightly to an MAE ±SD of 8.42 ±6.88 mmHg, and the bias shifts from an underestimation of $ME_{DBP,uni} = -1.95$ mmHg to an overestimation of $ME_{DBP,calib} = 1.12$ mmHg while the SD between subjects decreases to $SD_{MAE} = 2.66$ mmHg. The same tendency can be observed in the prediction accuracy of the original PPG signal. This can be explained by differences in the rather unbalanced data distributions of the training, tune and test sets that cause biases in the corresponding directions.

Figure 6.9 depicts the absolute prediction errors for different BP levels before

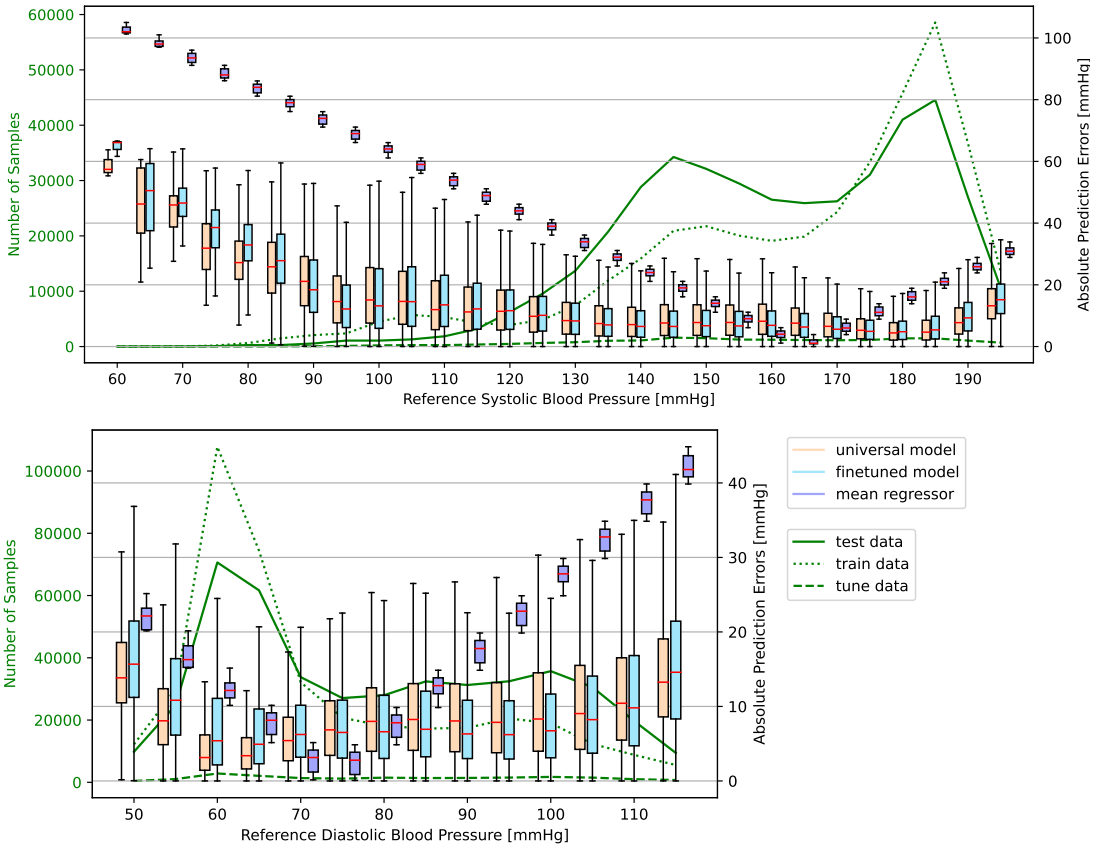


Figure 6.9: MAE for different BP levels before and after individual calibrations of the PPG-based model. [112]

and after individual calibration and in comparison to the mean regressor. The plot breaks down the overall error values into 5-mmHg steps of the corresponding BP reference values for a more detailed understanding of the error distribution. In general, the errors increase for the outer bounds of BP values in the data set since the number of samples strongly decreases in these ranges as indicated by the green lines in Figure 6.9, and a dependency of the prediction accuracy and data distributions can be observed.

To optimise the model personalisation process, different numbers of layers were frozen during fine-tuning. Overall best results are obtained with 3 trainable layers which includes the last fully connected layers resulting in 251,700 trainable parameters of a total of 3,986,242 parameters. The filters in the convolutional layers are not modified after optimisation on the training data set. If more than four layers are set trainable, convolutional layers are fine-tuned as well. Whereas

SBP slightly improves for a fine-tuning of 10 to 15 layers, DBP obtains the best results for the universal model without any fine-tuning at all.

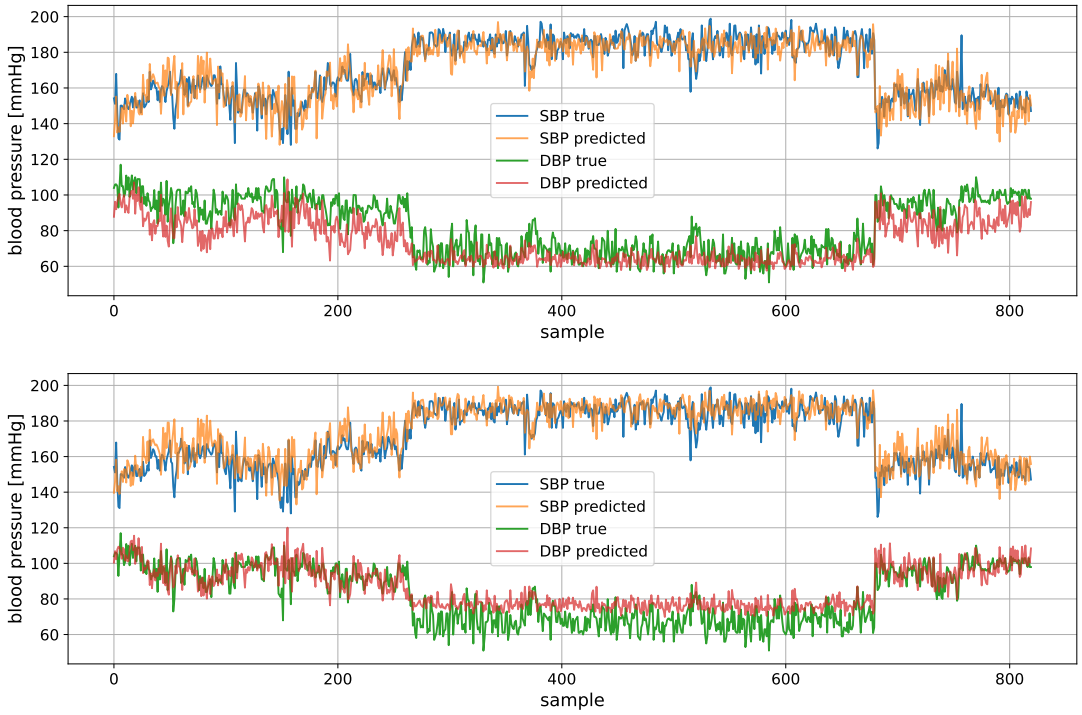


Figure 6.10: True and predicted BP values of a test patient before (upper plot) and after (lower plot) personalisation. [112]

To better understand the effect of model personalisation, BP predictions before and after fine-tuning of the model are compared. Figure 6.10 shows the true ABP and corresponding predicted values of test patient 3017781, where the predictions in the upper plot are obtained by the universal model and the lower plot shows the outputs after personalisation. The effect of personalisation can be best observed in DBP. For the first 280 samples, the offset between reference ABP and predicted DBP is eliminated after fine-tuning. However, the same baseline shift causes a larger discrepancy between the two curves for the next 400 samples where the signal exhibits a significantly lower ABP. By fine-tuning of the fully connected layers, the model adapts its PPG-BP mapping to the presented tune data. Despite the 300-sample distance between tune and test data (see Section 6.2.1), the tune samples most likely are in the same ABP range as the beginning of the test signal. Since the tune data are selected as consecutive pulses of a patient's signal and due to the small sample size, the tune data sets tend to be

very unbalanced. Hence, fine-tuning might introduce bias in un(der)-represented BP ranges.

Related studies that report successful personalisation of their general models employ proprietary data sets of healthy subjects who are less likely to show large changes in BP [83, 84] or eliminate signal intervals that comprise BP values which deviate more than 40 mmHg from the tune signal [82]. Therefore, these studies obtain higher prediction accuracies with their personalised models, since their test and tune data exhibit a lower BP variability. Zhang et al. [83] proposed a domain adversarial training specifically to overcome overfitting on new patients with few training data. Nevertheless, they report that for subjects with low variability, the model does not follow changes in BP. Thus, performance of their model strongly depends on the tune data set as well, and the reported prediction errors are lower due to low signal variability.

Overall, other published methods performing personalisation of their models report larger improvements over the universal model than the ones obtained in this work. Most of them employ the same fine-tuning strategy [83, 67, 84]. While the previously discussed BP variability greatly influences the personalisation outcome, also the data distribution, filtering, pre-processing and input length are factors that can have a large impact on a model's performance and vary a lot among those studies. Especially longer signal sections could be beneficial as they reduce the variance in reference BP and PPG waveform. A low variance and reduced noise are essential for the small tune data sets to provide conclusive and relevant data to the model. More comparable study set-ups are required to draw meaningful conclusions about the importance of each of the factors for personalisation in BP estimation.

Depending on the use case, different strategies could improve the personalisation process by balancing the tune data set: (1) If a beat-to-beat reference measurement is possible during acquisition of the tune data, a Valsalva manoeuvre could be performed to cover a larger range of ABP values. (2) During long-term monitoring, regular reference measurements could be included in the tune data set to continuously update the model until a sufficiently balanced data set is obtained. This would assimilate a random sampling of tune data from test subjects for which Schrumpf et al. [67] showed an improved performance compared to the sequentially split data set. (3) ABP values that are under-represented in the

subject's tune data set could be supplemented by the training data set of the universal model. Even though this might slow the personalisation process down, overfitting will be prevented. (4) Since arterial stiffness is a major contributor to changes in ABP waveform [121] and is strongly related to age, personal information could improve the universal model and possibly the personalised models as well.

Schrumpf et al. [67] showed that the personalisation success further depends on the chosen ANN architecture and is not correlated with the model's performance before fine-tuning. Also Qin et al. [84] obtained better personalised predictions when using a DANN as source model instead of a conventional ANN despite only marginal differences in prediction accuracy of these universal models. Hence, in the future, other architectures should be investigated with respect to personalisation.

6.3.3 Performance on rPPG Data

Estimating BP from rPPG data using the universal model trained on PPG signals results in a large overestimation of SBP with an ME ±SD of 40.91 ±21.55 mmHg and an underestimation of the DBP of -26.05 ±14.52 mmHg. It is evident that the PPG-based model is not suitable for the rPPG characteristics.

Hence, using the best universal PPG-based model, a LOSO CV is performed to evaluate suitability of transfer learning from the PPG to the rPPG domain. To do so, different numbers of trainable layers are compared. SBP benefits from retraining the whole network, thus adapting the weights of all filters to the target domain obtaining an MAE of 12.90 mmHg. In contrary, with an MAE of 7.55 mmHg, DBP obtains best results after adjusting only the 4 fully connected layers at the end of the ANN. This behaviour might be attributed to the significantly smaller DBP range and number of samples in the rPPG data set compared to the PPG data set, in addition to the high variability in rPPG waveform due to noise and artifacts. SBP and DBP both show only a small improvement over the mean regressor.

Next, the new rPPG models are personalised for each corresponding individual by fine-tuning on the subject's tune data. Since there are only an average of 39 data samples per subject, fine-tuning is performed just for the last three fully connected layers in order to reduce the risk of overfitting. Instead, the models

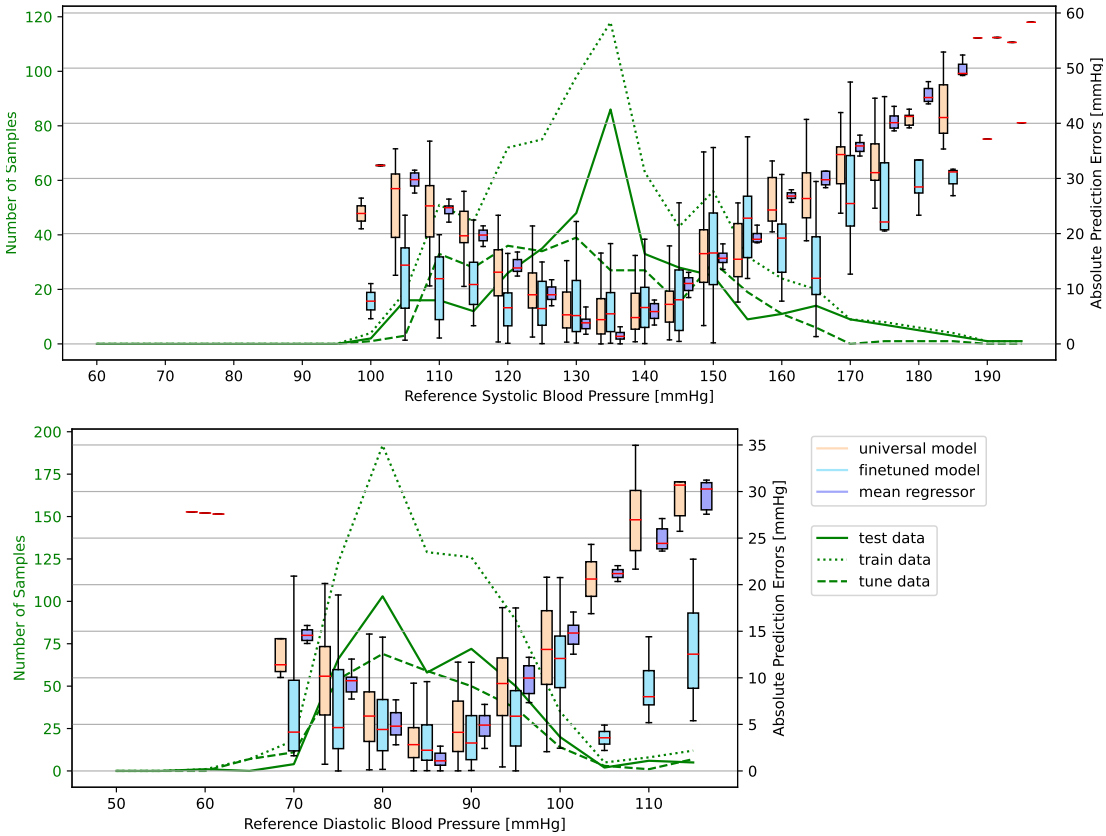


Figure 6.11: MAE for different BP levels before and after individual calibrations of the rPPG-based models.

of the previous LOSO CV with different numbers of rPPG-fine-tuned layers are used as initialisation. Models for which the last 40 layers were adjusted to the rPPG domain, exhibit the best SBP prediction error after personalisation with an MAE ±SD of 10.59 ±8.82 mmHg. With fine-tuning of 40 layers, the high-level features in the early convolutional layers remain the same while the fine-grain and more domain specific features in later convolutional layers are adapted. Again, DBP prefers the original filters and a transfer-learning only of the fully connected layers and obtains an MAE ±SD of 4.85 ±4.32 mmHg.

Figure 6.11 depicts the error distribution for the rPPG-based models (40 trainable layers during transfer learning) before and after personalisation as well as a mean regressor. It shows that overall the universal models obtain the same DBP prediction errors as the mean regressor and only slightly better results in SBP prediction. After personalisation, fine-tuning evens out the error distribution in SBP reducing prediction errors significantly in the lower range between 100 and

125 mmHg and above 160 mmHg. For DBP, fine-tuning improves prediction across all BP levels.

Although the overall error values look promising for BP estimation from rPPG, analysis of individual prediction and ground truth signals show that the universal rPPG model constantly predicts values close to DBP = 85 mmHg and SBP = 140 mmHg. This indicates that the model could not learn any underlying pattern to reliably estimate BP, but predicting an almost constant value reduced the loss most effectively. Subsequent personalisation of the model reduces the offset between predicted and true BP like in the personalisation process of the PPG-based model in Section 6.3.2, but the predictions do not follow changes in the reference BP.

Due to the lack of a standardised data set, prediction errors are very difficult to compare between studies on rPPG-based BP estimation. Still, related work on this topic shows similar difficulties regarding generalisation capabilities of the model. Many studies do not give details about the error distribution [58, 60, 64, 61] and obtain only minor improvements over their reported mean regression accuracy [61, 64]. Schrumpf et al. [61] show large intra-individual differences in their results even after personalisation of the model. Further, the Bland-Altman plot by Zhuang et al. [65] indicates a prediction of the mean BP value, and the plot by Luo et al. [59] shows almost no correlation between predicted and true values for DBP and only a weak correlation for SBP.

Many environmental influences such as lighting conditions, movement and skin properties cause distortions and artifacts in the rPPG waveforms that cannot be fully eliminated by existing rPPG extraction methods (see Chapter 5). Hence, rPPG pulses exhibit a larger variation in waveforms than PPG signals. Finding relevant patterns in the training data that contain information about BP is therefore more difficult for ANNs and too many faulty or noisy samples can prevent a successful training process.

Moreover, the small data set and low number of samples per subject in combination with the high variance in rPPG waveform appear to be insufficient to learn the complex task at hand. Especially in the hypo- and hypertensive BP ranges, the samples are sparse. Another major drawback of the used rPPG data set is the low frequency of reference measurements and the comparatively high bias of the reference device.

The proposed method should be tested with a larger rPPG data base. Even though rPPG signals constitute a more difficult domain for BP prediction as opposed to conventional PPG signals, a pre-trained model can overcome initialisation problems and convergence starts closer to the global minimum than when training a model from scratch. However, a careful data acquisition and pre-processing has to be performed such that for the training data set, variance in rPPG waveform caused by noise is reduced. That way, it is more likely for the model to find relevant hidden patterns in the data, and if successful, more challenging samples can be presented.

Chapter 7

Summary and Discussion

This chapter provides a comprehensive summary and discussion of the results and findings presented in this thesis. Additionally, it offers insights into potential directions for future research and applications in the field of BP estimation from video data.

7.1 Summary and Contributions

After introduction of the theoretical background in Chapter 2 and an extensive review of related work in Chapter 3, Chapter 4 describes the first approach of this thesis. It comprises the development of an image processing pipeline where several face and hand detection and tracking algorithms were evaluated for the specific measurement set-up to ensure a stable ROI placement and rule out any rPPG signal distortions due to tracking failures. SSD [40] has proven to be the most robust and precise face and hand detection algorithm while exhibiting an acceptable processing time. For tracking, MOSSE [45] was selected since it obtains the highest tracking stability as well as the lowest computation time among the compared methods.
Building on this, a BP prediction model was proposed which is based on 120 hand-crafted rPPG features and a small RFR regression model. The method was validated in a LOSO CV on a self-recorded data set containing videos and BP reference values of 30 subjects. By employing a SFS for selecting only the most relevant features for BP estimation, the prediction accuracy could be improved beyond the results of comparable studies with an MAE $\pm$SD of 6.27 $\pm$4.58 mmHg

for SBP and 3.91 ±2.79 mmHg for DBP on the reduced data set.
Moreover, an analysis and interpretation of the feature selection results for SBP and DBP was provided to aid future developments and enhance explainability of the model to potentially increase acceptance of such machine learning methods in clinical practice.

Whereas this first method considers the green channel of the RGB images, in Chapter 5 various other colour models and rPPG extraction methods were evaluated for BP estimation to overcome illumination and movement artifacts while increasing signal quality for persons with darker skin. Among these, Chrom [50] and POS [49] obtained similar results regarding correlation, SNR, BP prediction accuracy and signal rejection rate across most video recording conditions, but POS was shown to exhibit higher SNR and better performance for videos with movement.
Furthermore, dark skin was shown to benefit from an ROI selection at the palm reaching a significantly improved SNR due to a lower melanin concentration and higher vessel density. Also DBP prediction in general reached a higher accuracy with rPPG signals from the palm as opposed to the face. Depending on the suitability for the specific use case, the palm should be considered as an advantageous alternative to a measurement site in the face.

Under consideration of these findings, Chapter 6 introduced a CNN-based approach for BP estimation. To overcome the small size of the rPPG data set, this CNN model was developed and trained on the public MIMIC III waveform data base [56] and then used for transfer learning to the rPPG domain. In contrast to many related studies, a subject-based data split was followed.
After an extensive signal processing, different PPG representations from the time and frequency domain were tested as input to the CNN. Pulse length appeared to be an important feature since zero-padded PPG pulses obtained significantly better performance than resampled ones, and the resolution of frequency domain representations was shown to be insufficient for beat-to-beat data. The proposed method reaches an MAE ±SD of 8.73 ±7.36 mmHg for SBP and 8.07 ±6.86 mmHg for DBP and showed a good fit across a large range of BP levels. The fact that the obtained results are comparable to other PPG-based studies that rely on longer signal sections or an additional ECG signal shows that the proposed beat-to-beat prediction method is well suited to this task.

A layer activation analysis for inputs corresponding to different BP values was conducted to increase transparency of the model's decision making. The results were in accordance with the findings of the SFS analysis in Chapter 4. Moreover, an inter- and intra-individual correlation analysis of PPG waveforms of different BP levels showed that among different persons, the PPG waveforms are not distinctive for specific BP levels suggesting personalisation of the model.
Hence, fine tuning was employed for model personalisation. Despite an overall improved prediction accuracy, closer evaluation of the test results revealed tendencies for overfitting of the models due to the naturally unbalanced tune data set. We have shown that the outcome of model personalisation through fine-tuning is highly dependent on the selected tuning data and that there is a high risk of overfitting when a limited number of sequential tuning data are used. Model personalisation can be further improved by ensuring a balanced tune set across all BP levels, for which several strategies were proposed.
Finally transfer learning from the PPG to the rPPG domain was performed and evaluated by retraining several layers of the PPG model with rPPG data. With an MAE of 12.90 mmHg for SBP and 7.55 mmHg for DBP, the accuracy was comparable to the results of the feature-based method in Chapter 4. While the small rPPG data set was not sufficient to fully assess the capabilities of the proposed methods, both methods improved upon the mean regressor and delivered encouraging results for video-based BP estimation.

7.2 Discussion and Future Work

The following section provides insights and analysis regarding the findings and limitations of this work, highlighting valuable areas for further research.

The remotely measured rPPG signals still show significant degradations compared to the conventional PPG signals. Whereas dominant frequencies such as the heart rate and respiration are well preserved in various colour models, BP estimation relies on more fine-grain details in the rPPG waveform that are very susceptible to noise.

Differences in skin colour across individuals can affect the quality and consis-

tency of rPPG measurements. While this work has demonstrated that the positioning of the ROI can positively influence the rPPG quality, a study by Shirbani et al. [101] has shown that skin tone dependencies can be further reduced by applying a higher light intensity to the measurement site.

Movement and illumination artifacts, on the other hand, cannot be fully eliminated by existing rPPG extraction methods. The sensitivity of rPPG to these conditions arises from the fact that the technique relies on subtle changes in skin reflectance caused by blood volume variations. The modulation of ambient lighting and direction of specular reflection can interfere with these reflectance changes and introduce artifacts into the rPPG signal.

In conclusion, the current rPPG signal quality raises doubts regarding the feasibility of utilising rPPG signals for accurate BP estimation. Hence, to enable a reliable contactless BP estimation, rPPG signal quality needs to be enhanced further.

This could be achieved by more advanced signal processing methods, such as deep-learning-based PPG reconstructions from rPPG signals proposed by Song et al. [122] or Bousefsaf et al. [123]. While they show promising results, it needs to be evaluated if BP-relevant features are preserved. Moreover, these methods come with an increased computational cost, potentially imposing limitations on the real-time applicability.

Another approach could be the usage of Laser Speckle Imaging (LSI) for Speckleplethysmogram (SPG) measurement instead of rPPG. In contrast to PPG and rPPG signals that describe blood volume changes, SPG measures the flow rate of red blood cells and exhibit a much larger SNR [124, 125]. Herranz et al. [125] have shown a very high correlation between rPPG and remote SPG signals, but due to the difference in measured parameter the applicability for BP estimation needs to be evaluated first.

Comparing the CNN to the feature-based BP prediction approach proposed in this work, their results show that both methods improve upon the mean regressor baseline accuracy providing encouraging results in regard to the feasibility of BP estimation from rPPG signals. However, the small rPPG data set is not sufficient to determine which method is most accurate or feasible. In DBP prediction, the universal models of both methods exhibit comparable error values and obtain

BSH grade C. In SBP prediction, the feature-based method has a slightly lower MAE and a smaller SD among subjects, thus showing a minor advantage for this method. Despite the lack of a definite conclusion in the comparison of the two methods due to the limited data set, results suggest that further investigation of these methods might be beneficial.

The feature-based method may be computationally more efficient and could provide a preferable solution for resource limited systems while still obtaining comparable performance. On the other hand, the CNN is a more powerful tool for extracting features from the signals and therefore has more potential for improving on a larger data set. More data and further research is needed to fully evaluate the capabilities of both approaches.

Using a PPG data set, the proposed CNN for beat-to-beat BP estimation obtained good results across a large range of BP values and demonstrated remarkable accuracy in following strong and abrupt BP changes when observed over time, indicating its potential for tracking dynamic variations in BP.

Still, Chapter 6 has shown that even in BP estimation from conventional PPG signals, prediction accuracy needs to improve to comply with clinical standards. Related studies consider longer signal sections benefiting from reduced variance and noise in the reference values in addition to the increased amount of input information. To maintain the beat-to-beat resolution of the proposed model while achieving similar effects, a recurrent network architecture, e.g. LSTM, could be employed, allowing for the incorporation of temporal dependencies from both PPG and BP signals. Thus, the features extracted by the proposed CNN architecture could be further utilised in subsequent LSTM layers to enhance overall performance of the prediction model.

Moreover, transfer learning is a promising approach to leverage knowledge from the PPG domain to improve performance for rPPG signals since both signals are obtained by the same measurement principle capturing blood volume changes. It can enable overcoming differences in signal morphology of the two domains while taking advantage of the larger PPG data base.

Nevertheless, in the present work, the investigation and evaluation of transfer learning from PPG to rPPG was limited due to the small size of the rPPG

dataset. PPG and rPPG signals have shown significant differences in their pulse morphology and therefore, the features and patterns learned from PPG may not directly translate to rPPG. Moreover, the sensitivity of rPPG signals to noise and artifacts leads to a greater waveform variation compared to PPG signals. This increased variation makes it more challenging for ANNs to identify relevant patterns in the training data that contain information about BP. As a result, a larger data set might be required to adapt and fine-tune existing knowledge to the new domain.

The process of model personalisation has shown to be prone to overfitting on sequentially acquired tuning-data, since typically no major BP changes are captured within this short time period. Consequently, the personalised CNN models in this work do not outperform the general CNN model they are based on. To address this limitation, we proposed several strategies that can be implemented in real-world scenarios to obtain a more balanced tuning data set. Evaluating the effectiveness of these strategies in improving the performance of personalised models should be a focus of future investigations.

Alternatively, it may be worthwhile to explore the development of general models tailored to specific population groups. For instance, investigating whether there are differences in the relationship between BP and PPG waveforms among various groups such as males/females, different age groups, or individuals with specific pathological conditions could provide insights into the potential benefits of using separate regression models for each group. Existing research indicates that men and women have different aortic compliances [126, 127] generally leading to an underestimation of SBP in women and an overestimation of DBP if measured with cuff-based devices [128]. Additionally, Picone et al. [129] describe four BP amplification phenotypes characterised by different levels of amplification along the arterial tree. Since cuff-based devices usually measure brachial or radial BP, this causes large deviations in the estimation of the more relevant central BP. PPG- and rPPG-based BP prediction is probably impacted by the same phenomenon due to the peripheral measurement and could benefit from the identification of different population groups and development of separate regression models.

Although many studies have been working on rPPG-based BP estimation, there is still a lack of public data sets and standardised protocols for validation and comparison across different methods and devices. Consequently, for the work at hand, a new data set had to be generated to address this issue. However, limited resources have resulted in certain limitations associated with this data set, including its small size, sparse representation in the hypo- and hypertensive BP ranges, and a limited number of reference values. These factors need to be acknowledged as constrains that influenced the findings and conclusions of the study.

Moreover, discrepancies in the represented BP range, demographic, ethnicities and underlying health conditions in the data sets of different studies, as well as technical properties such as the recording frame rate, image resolution and controlled measurement conditions make it impossible to conduct reliable and fair comparisons between the proposed methods of different research groups. Furthermore, the majority of published studies on rPPG-based BP estimation does not follow a subject-based data split, leading to overly positive results that are not applicable for real-world scenarios. Hence, creating public data sets and establishing validation procedures and benchmarks will be crucial for evaluating the reliability and accuracy of rPPG-based BP estimation methods.

Due to the absence of a standardized dataset, comparing prediction errors in studies focusing on rPPG-based BP estimation is challenging. However, similar difficulties related to the generalization capabilities of the models as encountered in this work have been observed in related research. Many rPPG-based studies lack details regarding the error distribution, but those who do provide information report only marginal improvements over the mean regression accuracy. While this work has shown good results on conventional PPG signals, enhancing rPPG signal quality will be essential for achieving a comparable prediction performance on video data.

In summary, the findings in this work have shown new state-of-the-art results for beat-to-beat BP estimation from a single PPG signal. Integrating this novel approach into devices such as pulse oximeters, smartwatches, or medical sensor patches could facilitate clinical and ambulatory BP monitoring. This advance-

ment has the potential to enhance early detection of hypertension and related diseases while enabling beat-to-beat BP monitoring in standard hospital wards.

The integration of PPG-based BP estimation capabilities into widely available devices holds significant implications for healthcare. Pulse oximeters, commonly used for measuring oxygen saturation levels, could now provide additional valuable information about a patient's BP. Smartwatches, worn on a daily basis, could enable continuous and unobtrusive BP monitoring, enhancing the understanding of BP patterns and aiding in personalised disease management. Medical sensor patches, designed for long-term monitoring, could offer a non-invasive solution for patients requiring extended BP monitoring, improving patient comfort and adherence to the monitoring process.

Although the feasibility of BP-prediction from video data remains questionable and may be limited by the low rPPG signal quality, this thesis has shown some encouraging new results in this field. Due to its contactless nature, rPPG-based BP monitoring could provide several advantages over PPG-based methods. By eliminating the need for additional sensors or devices to be attached to the patient's body, the potential for contamination and associated infection would be significantly reduced. This aspect is crucial in healthcare environments where patients may already be vulnerable to infections.

Moreover, the absence of additional sensors on the patient's body alleviates any discomfort and eliminates concerns regarding sensor displacement or detachment. This would make the method also applicable to special patient populations. Neonates, individuals with sensitive skin, or burn victims may have difficulties wearing sensors directly on their skin, such that a rPPG-based measurement could be a suitable alternative for these specific groups.

Additionally, rPPG-based BP estimation enables accessibility and cost-effectiveness. As nearly everyone has access to a smartphone, leveraging rPPG through smartphone applications would make BP monitoring accessible and affordable to a larger population.

These benefits highlight the potential of rPPG-based BP estimation as a valuable tool in healthcare, enabling contactless, continuous and convenient monitoring of BP for various applications, ranging from clinical settings to personal health management. Hence, further research, along with the acquisition of large data

sets, will be essential in fully realizing the capabilities of the proposed approach and its potential for enhancing healthcare practices.

7.3 List of Publications

7.3.1 Journal Articles

Wuerich, C.; Wiede, C.; Grabmaier, A. "A comparative analysis of fall risk factors in the elderly and their automatic assessment." In: *International Journal of Biomedical Engineering and Technology* 41.3 (2023): 203-220.

7.3.2 Conference Proceedings

Wuerich, C.; Wichum, F.; Wiede, C.; Grabmaier, A. "Contactless optical respiration rate measurement for a fast triage of SARS-CoV-2 patients in hospitals." In: *Proceedings of the International Conference on Image Processing and Vision Engineering* (2021)

Wuerich, C.; Rademacher, R.; Wiede, C.; Grabmaier, A. "PTT-based contact-less blood pressure measurement using an RGB-camera." In: *Current Directions in Biomedical Engineering* 7.2 (2021): 375-378.

Wuerich, C.; Humm, E. M.; Wiede, C.; Schiele, G. "A Feature-based Approach on Contact-less Blood Pressure Estimation from Video Data." In: *30th European Signal Processing Conference (EUSIPCO)* (pp. 1343-1347). IEEE, 2022.

Wuerich, C.; Wichum, F.; El-Kadri, O.; Ghantawi, K.; Grewal, N.; Wiede, C.; Seidl, K. "Blood Pressure Estimation based on Electrocardiograms." In: *Current Directions in Biomedical Engineering* 8.2 (2022): 53-56.

Wuerich, C.; Wiede, C.; Schiele, G. "Cuffless Beat-to-Beat Blood Pressure Estimation from Photoplethysmogram Signals." In: *36th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2023.

Appendix A

Implementation Details

Table A.1: Full list and description of all rPPG features.

Feature Name	**Description**
T, Ts, Td	duration of a whole cycle (cycle time), duration of the systole (systolic time), duration of the diastole (diastolic time)
A, As, Ad	area under the curve for whole cycle (cycle area), area under the curve for systole (systolic area), area under the curve for diastole (diastolic area)
Div_TsTd, Div_TTd, Div_TTs	ratio of systolic time and diastolic time, ratio of cycle time and diastolic time, ratio of cycle time and systolic time
Div_AT, Div_AsTs, Div_AdTd	ratio of cycle area and cycle time, ratio of systolic area and systolic time, ratio of diastolic area and diastolic time
Div_AsAd, Div_AAd, Div_AAs	ratio of systolic area and diastolic area, ratio of cycle area and diastolic area, ratio of cycle area and systolic area

Table A.1: Full list and description of all rPPG features.

Feature Name	**Description**
SysWidth_10, SysWidth_25, SysWidth_33, SysWidth_50, SysWidth_67, SysWidth_75, SysWidth_90	systolic pulse widths at 10%, 25%, 33%, 50%, 67%, 75% and 90% of pulse amplitude
DiaWidth_10, DiaWidth_25, DiaWidth_33, DiaWidth_50, DiaWidth_67, DiaWidth_75, DiaWidth_90	diastolic pulse widths at 10%, 25%, 33%, 50%, 67%, 75% and 90% of pulse amplitude
Div_SysDiaWidth_10, Div_SysDiaWidth_25, Div_SysDiaWidth_33, Div_SysDiaWidth_50, Div_SysDiaWidth_67, Div_SysDiaWidth_75, Div_SysDiaWidth_90	ratios of systolic and diastolic pulse widths at 10%, 25%, 33%, 50%, 67%, 75% and 90% of pulse amplitude

Table A.1: Full list and description of all rPPG features.

Feature Name	**Description**
Amp_FFT0, Amp_FFT1, Amp_FFT2, Amp_FFT3, Amp_FFT4, Amp_FFT5, Amp_FFT6, Amp_FFT7, Amp_FFT8, Amp_FFT9, Amp_FFT10, Amp_FFT11, Amp_FFT12, Amp_FFT13, Amp_FFT14, Amp_FFT15, Amp_FFT16, Amp_FFT17, Amp_FFT18, Amp_FFT19	amplitudes of the Fourier-transformed signal at indices 0 to 19, i.e. 0.8 to 4.6 Hz
MaxPow_FFT0, MaxPow_FFT1, MaxPow_FFT2, MaxPow_FFT3, MaxPow_FFT4, MaxPow_FFT5, MaxPow_FFT6, MaxPow_FFT7, MaxPow_FFT8, MaxPow_FFT9	frequencies of the highest amplitudes in the Fourier-transformed signal
d1_Amp1, d1_Amp2, d1_Amp3	amplitudes of the first, second and third peak in the first derivative of the rPPG signal
d2_Amp1, d2_Amp2, d2_Amp3	amplitudes of the first, second and third peak in the second derivative of the rPPG signal

Table A.1: Full list and description of all rPPG features.

Feature Name	**Description**
d1_HPIP_height, d1_HPIP_width	height difference between first peak and following inflection point in the first derivative of the rPPG signal, time between first peak and following inflection point in the first derivative of the rPPG signal
d2_HPIP_height, d2_HPIP_width	height difference between first peak and following inflection point in the second derivative of the rPPG signal, time between first peak and following inflection point in the second derivative of the rPPG signal
d1_time_Amp1Amp2, d2_time_Amp1Amp2	time between first and second peak in the first derivative of the rPPG signal, time between first and second peak in the second derivative of the rPPG signal
d1_Div_Amp1Amp2, d2_Div_Amp1Amp2, d2_Div_Amp1Amp3	ratio of the first and second amplitude in the first derivative of the rPPG signal, ratio of the first and second amplitude in the second derivative of the rPPG signal, ratio of the first and third amplitude in the second derivative of the rPPG signal
d1_Sum_Amp1Amp2, d2_Sum_Amp1Amp2, d2_Sum_Amp1Amp2Amp3	sum of the first and second amplitude in the first derivative of the rPPG signal, sum of the first and second amplitude in the second derivative of the rPPG signal, sum of the first, second and third amplitude in the second derivative of the rPPG signal
d1_Sum_Amp1Amp2, d2_Sum_Amp1Amp2	sum of the first and second amplitude in the first derivative of the rPPG signal, sum of the first and second amplitude in the second derivative of the rPPG signal
Amp_DN	amplitude of the dicrotic notch
sex, height, weight, age	biological sex, height, weight and age of the subject

Bibliography

[1] Zhou, Bin ; Perel, Pablo ; Mensah, George A. ; Ezzati, Majid: Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. In: *Nature Reviews Cardiology* 18 (2021), Nr. 11, S. 785–802

[2] Vision, Allied (Hrsg.): *Hypertension*. `https://www.who.int/news-room/fact-sheets/detail/hypertension`. Version: 2021, Abruf: 13.02.2023 (fact-sheets)

[3] Chao, Paul C-P ; Wu, Chih-Cheng ; Nguyen, Duc H. ; Nguyen, Ba-Sy ; Huang, Pin-Chia ; Le, Van-Hung: The machine learnings leading the cuffless PPG blood pressure sensors into the next stage. In: *IEEE Sensors Journal* 21 (2021), Nr. 11, S. 12498–12510

[4] Grissmer, S: Blutkreislauf. In: *Physiologie* 2 (2010), S. 111–162

[5] VDE DGBMT: Surrogate Based Continuous Noninvasive Blood Pressure Measurement. (2020)

[6] Huang, Sheng-Chieh ; Hung, Pei-Hsuan ; Hong, Chung-Hung ; Wang, Hui-Min: A New Image Blood Pressure Sensor Based on PPG, RRT, BPTT, and Harmonic Balancing. In: *IEEE Sensors Journal* 14 (2014), Nr. 10, S. 3685–3692. `http://dx.doi.org/10.1109/JSEN.2014.2329676`. – DOI 10.1109/JSEN.2014.2329676. – ISSN 1530–437X

[7] Elliott, William J. ; Black, Henry R.: Prehypertension. In: *Nature Clinical Practice Cardiovascular Medicine* 4 (2007), Nr. 10, S. 538–548

[8] Guyton, Arthur C. ; Coleman, Thomas G. ; Cowley, Allen W. ; Scheel, Konrad W. ; Manning, R. D. ; Norman, Roger A.: Arte-

rial pressure regulation:: Overriding dominance of the kidneys in long-term regulation and in hypertension. In: *The American Journal of Medicine* 52 (1972), Nr. 5, S. 584–594

[9] Catanho, Marianne ; Sinha, Mridu ; Vijayan, Varsha: Model of aortic blood flow using the Windkessel effect. In: *University of California of San Diago, San Diago* (2012)

[10] Mukkamala, Ramakrishna ; Hahn, Jin-Oh ; Inan, Omer T. ; Mestha, Lalit K. ; Kim, Chang-Sei ; Töreyin, Hakan ; Kyal, Survi: Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice. In: *IEEE transactions on bio-medical engineering* 62 (2015), Nr. 8, S. 1879–1901. http://dx.doi.org/10.1109/TBME.2015.2441951. – DOI 10.1109/TBME.2015.2441951

[11] Vlachopoulos, Charalambos ; O'Rourke, Michael ; Nichols, Wilmer W.: *McDonald's blood flow in arteries: theoretical, experimental and clinical principles*. CRC press, 2011

[12] Kachuee, Mohammad ; Kiani, Mohammad M. ; Mohammadzade, Hoda ; Shabany, Mahdi: Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring. In: *IEEE transactions on bio-medical engineering* 64 (2017), Nr. 4, S. 859–869. http://dx.doi.org/10.1109/TBME.2016.2580904. – DOI 10.1109/TBME.2016.2580904

[13] Mills, CJ ; Gabe, IT ; Gault, JH ; Mason, DT ; Ross Jr, J ; Braunwald, E ; Shillingford, JP: Pressure-flow relationships and vascular impedance in man. In: *Cardiovascular Research* 4 (1970), Nr. 4, S. 405–417

[14] Xing, Xiaoman ; Sun, Mingshan: Optical blood pressure estimation with photoplethysmography and FFT-based neural networks. In: *Biomedical optics express* 7 (2016), Nr. 8, S. 3007–3020. http://dx.doi.org/10.1364/BOE.7.003007. – DOI 10.1364/BOE.7.003007. – ISSN 2156–7085

[15] Wang, Yuh-Ying L. ; Chang, CC ; Chen, JC ; Hsiu, H ; Wang, WK: Pressure wave propagation in arteries. A model with radial dilatation for

simulating the behavior of a real artery. In: *IEEE Engineering in Medicine and Biology Magazine* 16 (1997), Nr. 1, S. 51–54

[16] Nakano, Kazuya ; Ohnishi, Takashi ; Nishidate, Izumi ; Haneishi, Hideaki: Noncontact sphygmomanometer based on pulse-wave transit time between the face and hand. In: Coté, Gerard L. (Hrsg.): *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, SPIE, 27.01.2018 - 01.02.2018. – ISBN 9781510614871, 34

[17] Hughes, DJ ; Babbs, Charles F. ; Geddes, LA ; Bourland, JD: Measurements of Young's modulus of elasticity of the canine aorta with ultrasound. In: *Ultrasonic imaging* 1 (1979), Nr. 4, S. 356–367

[18] Bergel, DH: The static elastic properties of the arterial wall. In: *The Journal of physiology* 156 (1961), Nr. 3, S. 445

[19] Peterson, Lysle H. ; Jensen, Roderick E. ; Parnell, John: Mechanical properties of arteries in vivo. In: *Circulation Research* 8 (1960), Nr. 3, S. 622–639

[20] Sugita, Norihiro ; Yoshizawa, Makoto ; Abe, Makoto ; Tanaka, Akira ; Homma, Noriyasu ; Yambe, Tomoyuki: Contactless Technique for Measuring Blood-Pressure Variability from One Region in Video Plethysmography. In: *Journal of Medical and Biological Engineering* 39 (2019), Nr. 1, S. 76–85. http://dx.doi.org/10.1007/s40846-018-0388-8. – DOI 10.1007/s40846–018–0388–8. – ISSN 1609–0985

[21] Chatterjee, Ayan ; Roy, Uttam K.: Algorithm To Calculate Heart Rate & Comparison Of Butterworth IIR and Savitzky-Golay FIR Filter. In: *J Comput Sci Syst Biol* 11 (2018), S. 171–177

[22] Souza-Barros, Leanna ; Dhaidan, Ghaith ; Maunula, Mikko ; Solomon, Vaeda ; Gabison, Sharon ; Lilge, Lothar ; Nussbaum, Ethne L.: Skin color and tissue thickness effects on transmittance, reflectance, and skin temperature when using 635 and 808 nm lasers in low intensity therapeutics. In: *Lasers in surgery and medicine* 50 (2018), Nr. 4, S. 291–301. http://dx.doi.org/10.1002/lsm.22760. – DOI 10.1002/lsm.22760

[23] Nussbaum, Ethne L. ; van Zuylen, Jeff: Transmission of phototherapy through human skin: dosimetry adjustment for effects of skin color, body composition, wavelength, and light coupling to skin. In: Hamblin, Michael R. (Hrsg.) ; Waynant, Ronald W. (Hrsg.) ; Anders, Juanita (Hrsg.): *Mechanisms for Low-Light Therapy*, SPIE, 2006 (SPIE Proceedings), S. 61400H

[24] Meglinski, Igor V. ; Matcher, Stephen J.: Quantitative assessment of skin layers absorption and skin reflectance spectra simulation in the visible and near-infrared spectral regions. In: *Physiological measurement* 23 (2002), Nr. 4, S. 741–753. http://dx.doi.org/10.1088/0967-3334/23/4/312. – DOI 10.1088/0967–3334/23/4/312. – ISSN 0967–3334

[25] Lister, Tom ; Wright, Philip A. ; Chappell, Paul H.: Optical properties of human skin. In: *Journal of biomedical optics* 17 (2012), Nr. 9, S. 90901–1. http://dx.doi.org/10.1117/1.JBO.17.9.090901. – DOI 10.1117/1.JBO.17.9.090901

[26] Cui, Weijia ; Ostrander, Lee E. ; Lee, Bok Y.: In vivo reflectance of blood and tissue as a function of light wavelength. In: *IEEE transactions on biomedical engineering* 37 (1990), Nr. 6, S. 632–639

[27] AlHinai, Noura: Chapter 1 - Introduction to biomedical signal processing and artificial intelligence. Version: 2020. http://dx.doi.org/https://doi.org/10.1016/B978-0-12-818946-7.00001-9. In: Zgallai, Walid (Hrsg.): *Biomedical Signal Processing and Artificial Intelligence in Healthcare*. Academic Press, 2020 (Developments in Biomedical Engineering and Bioelectronics). – DOI https://doi.org/10.1016/B978–0–12–818946–7.00001–9. – ISSN 25897527, 1-28

[28] Tian, Hui: *Noise analysis in CMOS image sensors*. Stanford University, 2000

[29] Zgallai, Walid A.: *Biomedical Signal Processing and Artificial Intelligence in Healthcare*. Academic Press, 2020

[30] Butterworth, Stephen u. a.: On the theory of filter amplifiers. In: *Wireless Engineer* 7 (1930), Nr. 6, S. 536–541

[31] Cochran, William T. ; Cooley, James W. ; Favin, David L. ; Helms, Howard D. ; Kaenel, Reginald A. ; Lang, William W. ; Maling, George C. ; Nelson, David E. ; Rader, Charles M. ; Welch, Peter D.: What is the fast Fourier transform? In: *Proceedings of the IEEE* 55 (1967), Nr. 10, S. 1664–1674

[32] Cooley, James W. ; Tukey, John W.: An algorithm for the machine calculation of complex Fourier series. In: *Mathematics of computation* 19 (1965), Nr. 90, S. 297–301

[33] Jutten, Christian ; Herault, Jeanny: Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. In: *Signal processing* 24 (1991), Nr. 1, S. 1–10

[34] Breiman, Leo: Random forests. In: *Machine learning* 45 (2001), S. 5–32

[35] Kumar, Vipin ; Minz, Sonajharia: Feature selection: a literature review. In: *SmartCR* 4 (2014), Nr. 3, S. 211–229

[36] Abdi, Herve ; Williams, Lynne J.: Principal component analysis. In: *Wiley interdisciplinary reviews: computational statistics* 2 (2010), Nr. 4, S. 433–459

[37] Viola, Paul ; Jones, Michael J.: Robust real-time face detection. In: *International journal of computer vision* 57 (2004), Nr. 2, S. 137–154

[38] Dalal, Navneet ; Triggs, Bill: Histograms of oriented gradients for human detection. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* Bd. 1 Ieee, 2005, S. 886–893

[39] Redmon, Joseph ; Divvala, Santosh ; Girshick, Ross ; Farhadi, Ali: You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, S. 779–788

[40] Liu, Wei ; Anguelov, Dragomir ; Erhan, Dumitru ; Szegedy, Christian ; Reed, Scott ; Fu, Cheng-Yang ; Berg, Alexander C.: Ssd: Single shot multibox detector. In: *European conference on computer vision* Springer, 2016, S. 21–37

[41] Simon, Tomas ; Joo, Hanbyul ; Matthews, Iain ; Sheikh, Yaser: Hand keypoint detection in single images using multiview bootstrapping. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, S. 1145–1153

[42] Babenko, Boris ; Yang, Ming-Hsuan ; Belongie, Serge: Robust object tracking with online multiple instance learning. In: *IEEE transactions on pattern analysis and machine intelligence* 33 (2010), Nr. 8, S. 1619–1632

[43] Kalal, Zdenek ; Mikolajczyk, Krystian ; Matas, Jiri: Forward-backward error: Automatic detection of tracking failures. In: *2010 20th international conference on pattern recognition* IEEE, 2010, S. 2756–2759

[44] Lucas, Bruce D. ; Kanade, Takeo u. a.: An iterative image registration technique with an application to stereo vision Vancouver, 1981

[45] Bolme, David S. ; Beveridge, J R. ; Draper, Bruce A. ; Lui, Yui M.: Visual object tracking using adaptive correlation filters. In: *2010 IEEE computer society conference on computer vision and pattern recognition* IEEE, 2010, S. 2544–2550

[46] Henriques, Joao ; Caseiro, Rui ; Martins, Pedro ; Batista, Jorge: High-speed tracking with kernelized correlation filters. In: *IEEE transactions on pattern analysis and machine intelligence* 37 (2014), Nr. 3, S. 583–596

[47] Lukezic, Alan ; Vojir, Tomas ; Cehovin Zajc, Luka ; Matas, Jiri ; Kristan, Matej: Discriminative correlation filter with channel and spatial reliability. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, S. 6309–6318

[48] Ibraheem, Noor A. ; Hasan, Mokhtar M. ; Khan, Rafiqul Z. ; Mishra, Pramod K.: Understanding color models: a review. In: *ARPN Journal of science and technology* 2 (2012), Nr. 3, S. 265–275

[49] Wang, Wenjin ; Den Brinker, Albertus C. ; Stuijk, Sander ; De Haan, Gerard: Algorithmic principles of remote PPG. In: *IEEE Transactions on Biomedical Engineering* 64 (2016), Nr. 7, S. 1479–1491

[50] De Haan, Gerard ; Jeanne, Vincent: Robust pulse rate from chrominance-based rPPG. In: *IEEE Transactions on Biomedical Engineering* 60 (2013), Nr. 10, S. 2878–2886

[51] Fan, Xijian ; Ye, Qiaolin ; Yang, Xubing ; Choudhury, Sruti D.: Robust blood pressure estimation using an RGB camera. In: *Journal of Ambient Intelligence and Humanized Computing* 11 (2020), Nr. 11, S. 4329–4336. `http://dx.doi.org/10.1007/s12652-018-1026-6`. – DOI 10.1007/s12652–018–1026–6. – ISSN 1868–5137

[52] Huang, Po-Wei ; Lin, Chun-Hao ; Chung, Meng-Liang ; Lin, Tzu-Min ; Wu, Bing-Fei: Image based contactless blood pressure assessment using Pulse Transit Time. In: *2017 International Automatic Control Conference (CACS)*, IEEE, 112017. – ISBN 978–1–5386–3900–9, S. 1–6

[53] Jeong, In C. ; Finkelstein, Joseph: Introducing Contactless Blood Pressure Assessment Using a High Speed Video Camera. In: *Journal of medical systems* 40 (2016), Nr. 4, S. 77. `http://dx.doi.org/10.1007/s10916-016-0439-z`. – DOI 10.1007/s10916–016–0439–z

[54] Murakami, Kenta ; Yoshioka, Mototaka ; Ozawa, Jun: Non-contact pulse transit time measurement using imaging camera, and its relation to blood pressure. In: *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, IEEE, 52015. – ISBN 978–4–9011–2214–6, S. 414–417

[55] Khong, Wei L. ; Rao, Nittala Surya Venkata K. ; Mariappan, Muralindran: Blood pressure measurements using non-contact video imaging techniques. In: *2017 IEEE 2nd International Conference on Automatic Control and Intelligent Systems (I2CACIS)* IEEE, 2017, S. 35–40

[56] Moody, B ; Moody, G ; Villarroel, M ; Clifford, G ; Silva III, I: *MIMIC-III Waveform Database (version 1.0)*. 2020

[57] Zhang, Guanqun ; Gao, Mingwu ; Da Xu ; Olivier, N. B. ; Mukkamala, Ramakrishna: Pulse arrival time is not an adequate surrogate for pulse transit time as a marker of blood pressure. In: *Journal of*

applied physiology (Bethesda, Md. : 1985) 111 (2011), Nr. 6, S. 1681–1686. http://dx.doi.org/10.1152/japplphysiol.00980.2011. – DOI 10.1152/japplphysiol.00980.2011

[58] Jain, Monika ; Deb, Sujay ; Subramanyam, A V.: Face video based touchless blood pressure and heart rate estimation. (2016), S. 1–5

[59] Luo, Hong ; Yang, Deye ; Barszczyk, Andrew ; Vempala, Naresh ; Wei, Jing ; Wu, Si J. ; Zheng, Paul P. ; Fu, Genyue ; Lee, Kang ; Feng, Zhong-Ping: Smartphone-Based Blood Pressure Measurement Using Transdermal Optical Imaging Technology. In: *Circulation. Cardiovascular imaging* 12 (2019), Nr. 8, S. e008857. http://dx.doi.org/10.1161/CIRCIMAGING.119.008857. – DOI 10.1161/CIRCIMAGING.119.008857

[60] Tran, Quoc-Viet ; Su, Shun-Feng ; Tran, Quang-Minh ; Truong, Vi: Intelligent Non-Invasive Vital Signs Estimation From Image Analysis. In: *2020 International Conference on System Science and Engineering (ICSSE)* IEEE, 2020, S. 1–6

[61] Schrumpf, Fabian ; Frenzel, Patrick ; Aust, Christoph ; Osterhoff, Georg ; Fuchs, Mirco: Assessment of deep learning based blood pressure prediction from PPG and rPPG signals. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, S. 3820–3830

[62] Krizhevsky, Alex ; Sutskever, Ilya ; Hinton, Geoffrey E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems* 25 (2012)

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

[64] Wu, Bing-Fei ; Wu, Bing-Jhang ; Tsai, Bing-Ruei ; Hsu, Chi-Po: A Facial-Image-Based Blood Pressure Measurement System Without Calibration. In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), S. 1–13

[65] Zhuang, Jialiang ; Li, Bin ; Zhang, Yun ; Zheng, Xiujuan: InsightNet: non-contact blood pressure measuring network based on face video. In: *arXiv preprint arXiv:2203.03634* (2022)

[66] Fang, Yu-Fan ; Huang, Po-Wei ; Chung, Meng-Liang ; Wu, Bing-Fei: A feature selection method for vision-based blood pressure measurement. In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* IEEE, 2018, S. 2158–2163

[67] Schrumpf, Fabian ; Frenzel, Patrick ; Aust, Christoph ; Osterhoff, Georg ; Fuchs, Mirco: Assessment of non-invasive blood pressure prediction from ppg and rppg signals using deep learning. In: *Sensors* 21 (2021), Nr. 18, S. 6022

[68] Oiwa, Kosuke ; Bando, Shizuka ; Nozawa, Akio: Contactless blood pressure sensing using facial visible and thermal images. In: *Artificial Life and Robotics* 23 (2018), Nr. 3, S. 387–394. `http://dx.doi.org/10.1007/s10015-018-0450-1`. – DOI 10.1007/s10015–018–0450–1. – ISSN 1433–5298

[69] Feng, Litong ; Po, Lai-Man ; Xu, Xuyuan ; Li, Yuming ; Ma, Ruiyi: Motion-resistant remote imaging photoplethysmography based on the optical properties of skin. In: *IEEE Transactions on Circuits and Systems for Video Technology* 25 (2014), Nr. 5, S. 879–891

[70] Tsouri, Gill R. ; Li, Zheng: On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras. In: *Journal of biomedical optics* 20 (2015), Nr. 4, S. 048002

[71] Ernst, Hannes ; Scherpf, Matthieu ; Malberg, Hagen ; Schmidt, Martin: Color spaces and regions of interest in camera based heart rate estimation. In: *2020 11th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO)* IEEE, 2020, S. 1–2

[72] Boccignone, Giuseppe ; Conte, Donatello ; Cuculo, Vittorio ; d'Amelio, Alessandro ; Grossi, Giuliano ; Lanzarotti, Raffaella: An open framework for remote-PPG methods and their assessment. In: *IEEE Access* 8 (2020), S. 216083–216103

[73] Wang, Wenjin ; Stuijk, Sander ; De Haan, Gerard: A novel algorithm for remote photoplethysmography: Spatial subspace rotation. In: *IEEE transactions on biomedical engineering* 63 (2015), Nr. 9, S. 1974–1984

[74] Liu, Mengyang ; Po, Lai-Man ; Fu, Hong: Cuffless blood pressure estimation based on photoplethysmography signal and its second derivative. In: *International Journal of Computer Theory and Engineering* 9 (2017), Nr. 3, S. 202

[75] Kurylyak, Yuriy ; Lamonaca, Francesco ; Grimaldi, Domenico: 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, S. 280–283

[76] Hassani, Atefe ; Foruzan, Amir H.: Improved PPG-based estimation of the blood pressure using latent space features. In: *Signal, Image and Video Processing* 13 (2019), Nr. 6, S. 1141–1147

[77] Mousavi, Seyedeh S. ; Firouzmand, Mohammad ; Charmi, Mostafa ; Hemmati, Mohammad ; Moghadam, Maryam ; Ghorbani, Yadollah: Blood pressure estimation from appropriate and inappropriate PPG signals using A whole-based method. In: *Biomedical Signal Processing and Control* 47 (2019), S. 196–206

[78] Hsu, Yan-Cheng ; Li, Yung-Hui ; Chang, Ching-Chun ; Harfiya, Latifa N.: Generalized deep neural network model for cuffless blood pressure estimation with photoplethysmogram signal only. In: *Sensors* 20 (2020), Nr. 19, S. 5668

[79] El Hajj, Chadi ; Kyriacou, Panayiotis A.: Cuffless and continuous blood pressure estimation from PPG signals using recurrent neural networks. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* IEEE, 2020, S. 4269–4272

[80] El Hajj, Chadi ; Kyriacou, Panayiotis A.: Recurrent Neural Network Models for Blood Pressure Monitoring Using PPG Morphological Features. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* IEEE, 2021, S. 1865–1868

[81] Haddad, Serj ; Boukhayma, Assim ; Caizzone, Antonino: Continuous ppg-based blood pressure monitoring using multi-linear regression. In: *IEEE Journal of Biomedical and Health Informatics* 26 (2021), Nr. 5, S. 2096–2105

[82] Schlesinger, Oded ; Vigderhouse, Nitai ; Eytan, Danny ; Moshe, Yair: Blood pressure estimation from PPG signals using convolutional neural networks and Siamese network. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* IEEE, 2020, S. 1135–1139

[83] Zhang, Lida ; Hurley, Nathan C. ; Ibrahim, Bassem ; Spatz, Erica ; Krumholz, Harlan M. ; Jafari, Roozbeh ; Bobak, Mortazavi J.: Developing personalized models of blood pressure estimation from wearable sensors data using minimally-trained domain adversarial neural networks. In: *Machine Learning for Healthcare Conference* PMLR, 2020, S. 97–120

[84] Qin, Keke ; Huang, Wu ; Zhang, Tao: Deep generative model with domain adversarial training for predicting arterial blood pressure waveform from photoplethysmogram signal. In: *Biomedical Signal Processing and Control* 70 (2021), S. 102972

[85] Ibtehaz, Nabil ; Mahmud, Sakib ; Chowdhury, Muhammad E. ; Khandakar, Amith ; Salman Khan, Muhammad ; Ayari, Mohamed A. ; Tahir, Anas M. ; Rahman, M S.: PPG2ABP: Translating Photoplethysmogram (PPG) Signals to Arterial Blood Pressure (ABP) Waveforms. In: *Bioengineering* 9 (2022), Nr. 11, S. 692

[86] Ibtehaz, Nabil ; Mahmud, Sakib ; Chowdhury, Muhammad E. ; Khandakar, Amith ; Ayari, Mohamed A. ; Tahir, Anas ; Rahman, M S.: Ppg2abp: Translating photoplethysmogram (ppg) signals to arterial blood pressure (abp) waveforms using fully convolutional neural networks. In: *arXiv preprint arXiv:2005.01669* (2020)

[87] Slapnicar, Gasper ; Mlakar, Nejc ; Lustrek, Mitja: Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. In: *Sensors* 19 (2019), Nr. 15, S. 3420

[88] Wu, Bing-Fei ; Chiu, Li-Wen ; Wu, Yi-Chiao ; Lai, Chun-Chih ; Chu, Pao-Hsien: Contactless Blood Pressure Measurement via Remote Photoplethysmography with Synthetic Data Generation Using Generative Adversarial Network. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, S. 2130–2138

[89] Wuerich, Carolin ; Humm, Eva-Maria ; Wiede, Christian ; Schiele, Gregor: A Feature-based Approach on Contact-less Blood Pressure Estimation from Video Data. In: *2022 30th European Signal Processing Conference (EUSIPCO)* IEEE, 2022, S. 1343–1347

[90] Jaccard, Paul: The distribution of the flora in the alpine zone. 1. In: *New phytologist* 11 (1912), Nr. 2, S. 37–50

[91] Co. KG, Bosch + Sohn G. (Hrsg.): *BOSO Medicus X*. `https://www.boso.de/fileadmin/common/pdf/gebrauchsanweisungen/boso_medicus_X.pdf`. Version: 2019, Abruf: 22.05.2023 (Gebrauchsanweisungen)

[92] Mori, Hisao ; Yamamoto, Hareaki ; Kuwashima, Masaomi ; Saito, Saburo ; Ukai, Hiroshi ; Hirao, Kouichi ; Yamauchi, Mikio ; Umemura, Satoshi: How does deep breathing affect office blood pressure and pulse rate? In: *Hypertension research* 28 (2005), Nr. 6, S. 499–504

[93] Wilkinson, Ian B. ; MacCallum, Helen ; Flint, Laura ; Cockcroft, John R. ; Newby, David E. ; Webb, David J.: The influence of heart rate on augmentation index and central arterial pressure in humans. In: *The Journal of physiology* 525 (2000), Nr. Pt 1, S. 263

[94] Bos, Willem J. ; Verrij, Elisabeth ; Vincent, Hieronymus H. ; Westerhof, Berend E. ; Parati, Gianfranco ; Van Montfrans, Gert A.: How to assess mean blood pressure properly at the brachial artery level. In: *Journal of hypertension* 25 (2007), Nr. 4, S. 751–755

[95] Nirmalan, Mahesh ; Dark, Paul M.: Broader applications of arterial pressure wave form analysis. In: *Continuing Education in Anaesthesia, Critical Care & Pain* 14 (2014), Nr. 6, S. 285–290

[96] Association for the Advancement of Medical Instrumentation and others: American national standard. Electronic or automated sphygmomanometers. In: *ANSI/AAMI SP10-1992/A1* (1996)

[97] OBrien, Eoin ; Petrie, James ; Littler, WA ; Swiet, Michael de ; Padfield, Paul L. ; Altman, Douglas ; Bland, Martin ; Coats, Andrew ; Atkins, Neil u. a.: The British Hypertension Society protocol for the evaluation of blood pressure measuring devices. In: *J hypertens* 11 (1993), Nr. Suppl 2, S. S43–S62

[98] Allied Vision (Hrsg.): *Manta G-040*. `https://www.alliedvision.com/en/camera-selector/detail/Manta/G-040#`. Version: 2022, Abruf: 10.06.2022 (Datasheet)

[99] Fitzpatrick, Thomas B.: The validity and practicality of sun-reactive skin types I through VI. In: *Archives of dermatology* 124 (1988), Nr. 6, S. 869–871

[100] De Haan, Gerard ; Van Leest, Arno: Improved motion robustness of remote-PPG by using the blood volume pulse signature. In: *Physiological measurement* 35 (2014), Nr. 9, S. 1913

[101] Shirbani, Fatemeh ; Hui, Nicholas ; Tan, Isabella ; Butlin, Mark ; Avolio, Alberto P.: Effect of ambient lighting and skin tone on estimation of heart rate and pulse transit time from video plethysmography. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* IEEE, 2020, S. 2642–2645

[102] Kim, Dae-Yeol ; Lee, Kwangkee ; Sohn, Chae-Bong: Assessment of ROI Selection for Facial Video-Based rPPG. In: *Sensors* 21 (2021), Nr. 23, S. 7923

[103] Firooz, Alireza ; Sadr, Bardia ; Babakoohi, Shahab ; Sarraf-Yazdy, Maryam ; Fanian, Ferial ; Kazerouni-Timsar, Ali ; Nassiri-Kashani, Mansour ; Naghizadeh, Mohammad M. ; Dowlati, Yahya: Variation of biophysical parameters of the skin with age, gender, and body region. In: *The Scientific World Journal* 2012 (2012)

[104] Pasyk, Krystyna A. ; Thomas, Steven V. ; Hassett, Cheryl A. ; Cherry, George W. ; Faller, Richard: Regional differences in capillary density of the normal human dermis. In: *Plastic and reconstructive surgery* 83 (1989), Nr. 6, S. 939–45

[105] Pape, HC ; Kurtz, A ; Silbernagl, S: *Physiologie. 8., unveränderte Auflage*. 2018

[106] Awad, Aymen A. ; Ghobashy, M Ashraf M. ; Ouda, Wagih ; Stout, Robert G. ; Silverman, David G. ; Shelley, Kirk H.: Different responses of ear and finger pulse oximeter wave form to cold pressor test. In: *Anesthesia & Analgesia* 92 (2001), Nr. 6, S. 1483–1486

[107] Budidha, Karthik ; Kyriacou, Panayiotis A.: Photoplethysmography for quantitative assessment of sympathetic nerve activity (SNA) during cold stress. In: *Frontiers in Physiology* 9 (2019), S. 1863

[108] Chen, Yan ; Wen, Changyun ; Tao, Guocai ; Bi, Min: Continuous and noninvasive measurement of systolic and diastolic blood pressure by one mathematical model with the same model parameters and two separate pulse wave velocities. In: *Annals of biomedical engineering* 40 (2012), Nr. 4, S. 871–882

[109] Savitzky, Abraham ; Golay, Marcel J.: Smoothing and differentiation of data by simplified least squares procedures. In: *Analytical chemistry* 36 (1964), Nr. 8, S. 1627–1639

[110] Krishnan, Rajet ; Natarajan, Balasubramaniam ; Warren, Steve: Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data. In: *IEEE transactions on biomedical engineering* 57 (2010), Nr. 8, S. 1867–1876

[111] Elgendi, Mohamed: Optimal signal quality index for photoplethysmogram signals. In: *Bioengineering* 3 (2016), Nr. 4, S. 21

[112] Wuerich, Carolin ; Wiede, Christian ; Schiele, Gregor: Cuffless Beat-to-Beat Blood Pressure Estimation from Photoplethysmogram Signals. (2023), S. 305–310

[113] He, Kaiming ; Zhang, Xiangyu ; Ren, Shaoqing ; Sun, Jian: Identity mappings in deep residual networks. In: *European conference on computer vision* Springer, 2016, S. 630–645

[114] Aguirre, Nicolas ; Grall-Maës, Edith ; Cymberknop, Leandro J. ; Armentano, Ricardo L.: Blood pressure morphology assessment from photoplethysmogram and demographic information using deep learning with attention mechanism. In: *Sensors* 21 (2021), Nr. 6, S. 2167

[115] Slapničar, Gašper ; Mlakar, Nejc ; Luštrek, Mitja: Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. In: *Sensors* 19 (2019), Nr. 15, S. 3420

[116] Baek, Sanghyun ; Jang, Jiyong ; Yoon, Sungroh: End-to-End Blood Pressure Prediction via Fully Convolutional Networks. In: *IEEE Access* 7 (2019), S. 185458–185468. http://dx.doi.org/10.1109/ACCESS.2019.2960844. – DOI 10.1109/ACCESS.2019.2960844

[117] Thambiraj, Geerthy ; Gandhi, Uma ; Mangalanathan, Umapathy ; Jose, V Jeya M. ; Anand, M: Investigation on the effect of Womersley number, ECG and PPG features for cuff less blood pressure estimation using machine learning. In: *Biomedical Signal Processing and Control* 60 (2020), S. 101942

[118] McGhee, Beate H. ; Bridges, Elizabeth J.: Monitoring arterial blood pressure: what you may not know. In: *Critical care nurse* 22 (2002), Nr. 2, S. 60–79

[119] Rook, WH ; Turner, JD ; Clutton-Brock, TH: Analysis of damping characteristics of arterial catheter blood pressure monitoring in a large intensive care unit. In: *Southern African Journal of Critical Care* 33 (2017), Nr. 1, S. 8–10

[120] Pasma, Wietze ; Peelen, Linda M. ; Van Buuren, Stef ; Van Klei, Wilton A. ; De Graaff, Jurgen C.: Artifact processing methods influence on intraoperative hypotension quantification and outcome effect estimates. In: *Anesthesiology* 132 (2020), Nr. 4, S. 723–737

[121] Wu, Hsien-Tsai ; Lee, Chun-Ho ; Liu, An-Bang ; Chung, Wei-Sheng ; Tang, Chieh-Ju ; Sun, Cheuk-Kwan ; Yip, Hon-Kan: Arterial stiffness using radial arterial waveforms measured at the wrist as an indicator of diabetic control in the elderly. In: *IEEE Transactions on Biomedical engineering* 58 (2010), Nr. 2, S. 243–252

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

[123] Bousefsaf, Frédéric ; Djeldjli, Djamaleddine ; Ouzar, Yassine ; Maaoui, Choubeila ; Pruski, Alain: iPPG 2 cPPG: reconstructing contact from imaging photoplethysmographic signals using U-Net architectures. In: *Computers in Biology and Medicine* 138 (2021), S. 104860

[124] Dunn, Cody E. ; Lertsakdadet, Ben ; Crouzet, Christian ; Bahani, Adrian ; Choi, Bernard: Comparison of speckleplethysmographic (SPG) and photoplethysmographic (PPG) imaging by Monte Carlo simulations and in vivo measurements. In: *Biomedical optics express* 9 (2018), Nr. 9, S. 4306–4316

[125] Herranz Olazábal, Jorge ; Wieringa, Fokko ; Hermeling, Evelien ; Van Hoof, Chris: Camera-Derived Photoplethysmography (rPPG) and Speckle Plethysmography (rSPG): Comparing Reflective and Transmissive Mode at Various Integration Times Using LEDs and Lasers. In: *Sensors* 22 (2022), Nr. 16, S. 6059

[126] Gatzka, Christoph D. ; Kingwell, Bronwyn A. ; Cameron, James D. ; Berry, Karen L. ; Liang, Yu-Lu ; Dewar, Elizabeth M. ; Reid, Chris M. ; Jennings, Garry L. ; Dart, Anthony M. u. a.: Gender differences in the timing of arterial wave reflection beyond differences in body height. In: *Journal of hypertension* 19 (2001), Nr. 12, S. 2197–2203

[127] Hayward, Christopher S. ; Kelly, Raymond P.: Gender-related differences in the central arterial pressure waveform. In: *Journal of the American College of Cardiology* 30 (1997), Nr. 7, S. 1863–1871

[128] Picone, Dean S. ; Stoneman, Elif ; Cremer, Antoine ; Schultz, Martin G. ; Otahal, Petr ; Hughes, Alun D. ; Black, J A. ; Bos, Willem J. ; Chen, Chen-Huan ; Cheng, Hao-Min u. a.: Sex differences in blood pressure and potential implications for cardiovascular risk management. In: *Hypertension* 80 (2023), Nr. 2, S. 316–324

[129] Picone, Dean S. ; Schultz, Martin G. ; Peng, Xiaoqing ; Black, J A. ; Dwyer, Nathan ; Roberts-Thomson, Philip ; Chen, Chen-Huan ; Cheng, Hao-Min ; Pucci, Giacomo ; Wang, Ji-Guang u. a.: Discovery of new blood pressure phenotypes and relation to accuracy of cuff devices used in daily clinical practice. In: *Hypertension* 71 (2018), Nr. 6, S. 1239–1247