City Research Online

City, University of London Institutional Repository

Machine learning techniques for the prediction of systolic and diastolic blood pressure utilising the photoplethysmogram

by

Chadi El Hajj

A thesis submitted in partial fulfilment for the
degree of Doctor of Philosophy

Research Centre for Biomedical Engineering

School of Mathematics, Computer Science & Engineering

Northampton Square, London, EC1V 0HB

Feb 2022

Declaration

I, Chadi El Hajj, declare that the work presented in this thesis titled, 'Machine learning techniques for the prediction of systolic and diastolic blood pressure utilising the photoplethysmogram', is entirely my own, expect those parts duly identified and referenced in my submission. I confirm that the work was carried out wholly during my candidature at City, University of London. I further confirm that this thesis has not been submitted to any other academic programme or award.

Signature: Chadi El Hajj

Date: 01/11/2021

Acknowledgements

First and foremost, I would like to express my deep appreciation and gratitude to my supervisor Professor Panicos A. Kyriacou for all his guidance, insightful feedbacks and support in every step of my PhD period. I am also thankful for his encouragement, being inspiration, and most importantly for giving me this opportunity. My heartfelt thanks goes to my second supervisor Professor Artur Garcez for providing me with the skills needed to pursue this PhD.

I would also like to thank all my colleagues at City, university of London, particularly, at the Research Centre for Biomedical Engineering, for their motivation, support, friendship and for sharing their valuable experience with me.

Last but not least, I am grateful to my parents and my friends for their moral support, encouragement and help throughout my journey.

Contents

List of Tables

List of Figures

Abbreviations

BP   Blood Pressure

ABP   Arterial Blood Pressure

CV   Cardiovascular

CVD   Cardiovascular Disease

PPG   Photoplethysmograph

ECG   Electrocardiogram

BCG   Ballistocardiograph

SBP   Systolic Blood Pressure

DBP   Diastolic Blood Pressure

PP   Pulse Pressure

PTT   Pulse Transit Time

PWV   Pulse Wave Velocity

PAT   Pulse Arrival Time

PEP   Pre-Ejection Period

WHO   World Health Organisation

LED   Light Emitting Diode

HR   Heart Rate

SW   Systolic width

DW   Diastolic width

ST   Systolic upstroke Time

DT   Diastolic Time

PI   Pulse Interval

PIR   PPG Intensity Ratio

TPR   Total Peripheral Resistance

MIMIC II   Multiparameter Intelligent Monitoring in Intensive Care II

AAMI   American National Standards of the Association for the Advancement of Medical Instrumentation

BHS   British Hypertension Society

MSE   Mean Squared Error

MAE   Mean Absolute Error

ME   Mean Error

SD   Standard Deviation

RMSE   Root Mean Square Error

ADAM   Adaptive Moment estimation

ANN   Artificial Neural Network

ReLu   Rectified Linear Unit

LR   Learning Rate

MLR   Multiple Linear Regression

SVM   Support Vector Machine

SVR   Support Vector Regression

RNN   Recurrent Neural Network

BiRNN   Bidirectional Recurrent Neural Network

LSTM   Long Short-Term Memory

BiLSTM Bidirectional Long Short-Term Memory

GRU Gated Recurrent Units

BiGRU Bidirectional Gated Recurrent Units

MLP Multilayer Perceptron

Glossary

Renal disease: relating to kidney disease or failure.

L1 regularisation: adds the sum of absolute values of weights as penalty to a loss function.

L2 regularisation: adds the sum of squared values of weights as penalty to a loss function.

Hidden layer units: neural network activation units in layers between the input and output layers.

SVR kernel function: function that maps input variables from low dimension onto a higher dimension space.

Python: Programming language.

Scikit-learn: Python library that implements classical machine learning models.

Temporal feature: Refers to time domain, amplitude, ratios and width-related features extracted from the PPG waveform and its derivatives.

List of Publications

Journals

- El-Hajj, C. and Kyriacou, P.A., 2020. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. *Biomedical Signal Processing and Control, 58,* p.101870.

- El-Hajj, C. and Kyriacou, P.A., 2021. Deep learning models for cuffless blood pressure monitoring from PPG signals using attention mechanism. *Biomedical Signal Processing and Control, 65,* p.102301.

- El-Hajj, C. and Kyriacou, P.A., 2021. Cuffless blood pressure estimation from PPG signals and its derivatives using deep learning models. *Biomedical Signal Processing and Control, 70,* p.102984.

Conferences

- El Hajj, C. and Kyriacou, P.A., 2020, July. 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)* (pp. 4269-4272). IEEE.

- El Hajj, C. and Kyriacou, P.A., 2021, Nov. 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.

Book chapter

- Mejia-Mejia E., Allen J., Budidha K., **El-Hajj C.**, Kyriacou P.A. and Charlton P.H., Photoplethysmography Signal Processing and Synthesis. In Photoplethysmography; Kyriacou, P.A., Allen, J., Eds.; Elsevier, 2021.

Abstract

Blood pressure (BP) is one of the four primary vital signs that provides important information regarding patients' cardiovascular system conditions. Continuous and regular blood pressure monitoring is essential for the early diagnosis, prevention and management of cardiovascular disease (CVD) and haemodynamic diseases (hypertension and hypotension). Current clinical blood pressure measurement techniques are either invasive or cuff-based, which can be impractical, intermittent, and uncomfortable for patients during frequent measurements. Considering these challenges, several studies have suggested new non-invasive and cuffless blood pressuring measurement techniques using physiological signals, such as, the Electrocardiogram (ECG) and the Photoplethysmogram (PPG). In particular, indirect cuffless BP measurement techniques using pulse transit time and pulse arrival time have been extensively investigated over the last few decades. However, these techniques require two measurement sensors, frequent calibration, and hence, they are also impractical and inconvenient for continuous BP measurements. More recently, with the advancement of computational techniques, including machine learning and artificial intelligence, a new simple and innovative approach using only PPG signals have been proposed in the literature for cuffless and continuous monitoring of blood pressure. However, the majority of these studies have been unable to achieve acceptable accuracies that comply or satisfy the international standards for cuffless BP monitoring. Thus, further investigations are required to realise this approach.

In this research, a total of 52 features have been extracted from the PPG and their individual impact on BP have been rigorously evaluated using several statistical and machine learning techniques. As a result, only the most important features for estimation of BP were selected, effectively reducing the input vector by more than half. Two datasets were created to accommodate the two input feature vectors. The

PPG and reference BP signals were derived from the publicly available MIMIC II database. In order to estimate BP, a total of nine machine learning and neural network models have been implemented and evaluated on the two datasets. Out of the nine models, four are widely used classical machine learning models, and five neural network models, three of which are conventional models and two advanced models have been proposed for BP estimation using only one PPG signal. The results of all these models have also been compared against well established studies in the literature.

The results obtained using the classical machine learning models, namely, multi-linear regression, random forest, adaboost and support vector machine, were poor and inferior to all the neural network models. A slight performance improvement was achieved using the non-recurrent multi-layer perceptron, however, the error was still much higher than the internationally acceptable range. On the other hand, a significant improvement was achieved for the first time by using the recurrent neural network models, namely, Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). The two proposed neural network models further enhanced the BP estimation accuracies and were able to reduce the mean absolute error (MAE) to a range below 5 mmHg. In particular, the best performing model was the one bidirectional GRU layer, followed by two unidirectional GRU layers, and an attention layer. The obtained MAE and standard deviation (SD) were 4.79±8.08 mmHg for systolic BP (SBP) and 2.77±4.72 mmHg for diastolic BP (DBP). Furthermore, the DBP estimation were well below the internationally acceptable limits (referring to the AAMI standards of mean error (ME), ME±SD less than 5±8), while the ME for the SBP estimation were acceptable but the SD exceeded the limits by only 1.34 mmHg.

This research has successfully demonstrated that advanced neural network models can be used for the non-invasive and cuffless prediction of BP utilising the PPG.

Chapter 1

Introduction

1.1 Clinical need for cuffless and continuous blood pressure measurement

Blood pressure (BP) is the pressure exerted by the circulation of blood against the wall of the artery. It is one of the four vital signs of the human body along with body temperature, pulse rate, and respiration, therefore it is an important indicator for assessing the health conditions of a person [6]. In particular, BP measurement is used for monitoring the cardiovascular functions of the body. Abnormal BP levels increases the risk of cardiovascular disease (CVD), damages the internal organs, and eventually leads to fatal consequences [7]. For example, chronic hypertension or high BP, can incite several diseases, such as, kidney disease, heart attacks, and strokes [8]. In fact, hypertension is one of the most critical risk factors for CVD [9], and according to the World Health Organisation (WHO), CVD is a leading cause of death world wide [10, 11]. Thus, cuffless and continuous BP monitoring is essential for early detection, management and treatment of CVD as well as hemodynamic diseases [12].

Current clinical BP measurement methods are either invasive or cuff-based. The former is known as catheterisation [13], it provides direct continuous BP measurement for every heartbeat. Although this method is considered the most accurate BP measurement, such a technique is only applicable to patients undergoing surgical interventions. However, the most commonly used BP measurement techniques in health-care are based on conventional cuff-type devices [14]. The cuff-based mea-

1

surements can provide Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP) values for a given point in time noninvasively. Yet, despite this advantage, cuff-based BP devices are mostly intermittent, tedious, and inconvenient for patients due to repeated cuff inflation and deflation during measurement.

Cuffless BP measurement has attracted a lot of attention over the past decades and has advanced greatly in every aspect. Several of such approaches are mainly based on the utilisation of physiological signals, such as, the Photoplethysmogram (PPG) and the Electrocardiogram (ECG). Pulse Transit Time (PTT) is one approach that has been studied extensively [15, 16, 17, 18, 19]. There is an inverse correlation that often exists between BP and PTT [20]. PTT is the time delay for the pressure wave to travel between two sites on the body. Pulse Arrival Time (PAT) is another cuffless approach and refers to the time interval for the blood pulse to travel between the aortic valve opening and the arrival of the pulse wave to a peripheral point on the body, such as, fingertip, toe, earlobe, etc [21]. The major drawback of the PTT and PAT approaches is the fact that two sensors are needed for simultaneous measurement at two different locations on the body. This is not easy to implement and requires more effort and attention during setup, since the obtained signals must be perfectly synchronised for accurate peak detection, which is vital for estimating these parameters. Another important drawback is that they require the calculation of the pre-ejection period (PEP) as well. Moreover, these methods depend on complicated arterial wave propagation models (i.e elasticity of artery determines the speed at which the pressure pulse travels) and require individual calibration due to their dependency to patients' physiological characteristics. All these reasons prevent PTT and PAT from being reliable practical indicators, and replacement of traditional BP measurement methods [22, 23].

In an effort to overcome the aforementioned limitations, estimating BP using a single physiological signal, namely, PPG, has gained an increasing amount of attention over the past few years, due to its simplicity, and ability to provide continuous and cuffless measurement. In this approach, raw PPG signals or features extracted from its waveform and derivatives form the input for machine learning and neural networks models. Various estimators have been employed including linear regression, support vector regression (SVR), random forest (RF), feedforward neural network,

convolutional neural network (CNN) as well as a combination of CNN and other recurrent neural network models [24, 25, 26, 27, 28]. However, many of these studies did not focus on optimising the feature set nor study the impact of each feature on the target outputs. Insignificant and redundant features may have negative impact on the model, and subsequently reduce its estimation accuracy. Furthermore, only few studies have considered modelling the variations in the PPG features with respect to time, however, the majority were based on the PTT or PAT approaches [29, 30, 31], hence, two signals were required. Additionally, the global standard for cuffless BP measurement, set by the American National Standards of the Association for the Advancement of Medical Instrumentation (AAMI), was not strictly followed in terms of the minimum number of subjects required for evaluation. The AAMI requires the mean error (ME) and standard deviation (SD) to be no greater than 5±8 mmHg evaluated on at least 85 subjects. However, the majority of published research studied have so far failed to satisfy both AAMI requirements i.e. evaluation of the proposed technique on 85 subjects and achieving an error below 5±8 mmHg. Hence, a simple reliable technique for cuffless and continuous BP monitoring using a single sensor has not yet been established in clinical practice. Therefore, the PPG approach requires further investigation in order to find the most relevant set of features for this task as well as develop an accurate BP estimation model that meets the AAMI requirements with an acceptable accuracy on a large dataset.

1.2   Aims and objectives of the study

The main aim of this thesis is to investigate the feasibility of establishing a reliable alternative for the traditional cuff-based BP measurement devices. In particular, this thesis aims to develop and evaluate a data driven machine learning model for non-invasive, cuffless and continuous BP monitoring using only PPG signals. Therefore, in order to achieve this aim, the objectives of this thesis are set as follows:

- Analyse the existing gap between the current proposed cuffless BP measurement techniques and the demand for cuffless continuous BP measurement.

- Identify the limitations and obstacles preventing the current available cuffless methods from being deployed for continuous BP measurement.

- Acquire PPG and reference BP signals, from a sufficient number of subjects,

from the most commonly used publicly available dataset (MIMIC II database).

- Implement and evaluate a noise-robust algorithm for the accurate extraction of feature values from different PPG morphologies including first and second derivatives.

- Analyse the impact and influence of each PPG feature on the target BP outputs using statistical methods (collinearity test and maximum information coefficient) and machine learning techniques (recursive feature elimination), for selecting only the most effective features for estimation.

- Create two datasets: one containing the full feature set before feature elimination and the second one containing the refined reduced feature set obtained from the feature elimination step.

- Implement, train, and optimise seven classical machine learning and non-recurrent as well as recurrent neural network models, on the two datasets. The implemented models are, namely, multi-linear regression, support vector regression, random forest, adaptive boost, non-recurrent feedforwards neural network, recurrent Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU).

- Evaluate and compare the performance of the seven models on the two datasets using two commonly used statistical metrics including the global standard metric for cuffless BP measurement. Further analyse the results obtained on the reduced feature set using three graphical evaluation tools, namely, histograms, Bland-Altman and regression plots.

- Investigate the impact of modelling the temporal variation in the PPG features, using the LSTM and GRU, on the estimation accuracy.

- Improve the performance of the conventional LSTM and GRU models by developing, training and optimising an architecture that: 1) exposes the networks to more context/information by taking full advantage of the available input sequence, and 2) reduces the search space for information by only considering the most important parts from the hidden states sequence that contribute the most on the target systolic and diastolic BP estimation.

4

- Evaluate and compare the performance of the two proposed models, on the two datasets, against the seven conventional models.

- Evaluate and compare the results of the best performing model against the results obtained using different BP measurement methods, presented in well-established published BP studies, evaluated on the same dataset.

1.3 Contribution to knowledge

The main contributions towards cuffless BP estimation using only PPG signals are:

- Extraction of a large number of features from the PPG waveform and its derivatives, and evaluate their relationship and influence on the BP estimation.

- Development and evaluation of several classical machine learning models for exploring the prediction of SBP and DBP across all BP categories.

- Development and evaluation of various recurrent and non-recurrent neural network models for overcoming the shortcomings of classical machine learning models.

- Further advancement of computational models in an effort to improve and align the predictions with internationally acceptable standards for non-invasive cuffless BP estimation.

- Provide a thorough comprehensive performance comparison between all the aforementioned models using various popular evaluation metrics.

1.4 Thesis outline

The thesis is organised as followed:

Chapter 1 introduces the thesis by describing the motivation behind the research project and outlines the aims, objectives and contributions.

Chapter 2 provides the context required for understanding the work carried out in this thesis. Particularly, it provides the fundamental background knowledge of the circulatory system, BP physiology, and the importance of frequent monitoring for

early detection and control of abnormal BP. It also reviews the current clinical as well as research based BP measurement methods. This chapter also introduces the BP measurement approaches based on both the wave propagation theory using two signals and the pulse wave analysis approach using only a single PPG sensor. It also presents the advantages and limitations for the current available techniques.

Chapter 3 provides a detailed literature review of existing cuffless BP measurement techniques using data driven machine learning and neural network models utilising physiological signals. It also discusses the advantages and challenges for every study.

Chapter 4 describes the dataset used, signal preprocessing and data preparation. It also presents the features extracted from the PPG and their relationship with BP. Additionally, it describes the feature reduction process and the impact of each feature on BP.

Chapter 5 presents the four most frequently encountered classical machine learning models, their optimisation, in addition to their advantages and disadvantages.

Chapter 6 presents a detailed overview of three conventional neural network models and the different processing units that comprise these models. It also describes in detail how these models work, their capabilities and their advantages and disadvantages. This chapter also focuses on the development on the two proposed models, its components and their significance. Additionally, it presents the loss function, and optimiser used in order to train all the neural network models.

Chapter 7 presents, evaluates and compares the results obtained using all nine models on both datasets. It also describes the different performance evaluation metrics, both statistical and graphical representations, used for assessing the performance for every machine learning model. Additionally, this chapter focuses on analysing the performance using five different metrics (namely mean absolute error, mean error, histograms, Bland-Altman and regression plots), in order to select the best estimation model for further comparison against well established studies.

Chapter 8 concludes the thesis by providing a summary of the main findings,

as well as discussion and comparison against other related works. It also presents the main strength and limitations of the study including suggestions for future work.

Chapter 2

Research Background

This main aim of this chapter is to introduce the physiological and technological concepts that form the foundation for understanding the work carried out in this thesis. In particular, the first section describes and provides useful information about the circulatory system, blood pressure physiology, clinical significance and importance of regular BP monitoring, main physiological factors affecting BP, different BP parameters as well as abnormal categories and their clinical relevance. The second section introduces the current invasive, occlusive cuff-based (oscillometric and auscultatory techniques) and semi-occlusive (volume clamp and tonometery) BP measurement devices and their limitations. The third section focuses on the main cuffless BP measurement techniques, that are mainly based on the theory of wave propagation velocity, derived from physiological signals.

2.1 Blood pressure and its significance

2.1.1 Cardiovascular system

The cardiovascular (CV) system is a network of organs and blood vessels that serve to provide rapid delivery of oxygen, vitamins, amino acids, glucose, hormones, and other vital nutrients and substances to all cells in the body. It is also responsible for the removal of carbon dioxide and other metabolic waste products [32]. The CV system consists of the heart and blood vessels which form the circulatory routes that transport blood to the whole body [33]. The heart acts as a periodic pump that provides the force for carrying the oxygen-rich blood to each organ. In sys-

temic circulation, the heart ejects blood from the left ventricle into the main artery (aorta) during ventricular contraction. From the aorta, the blood then flows through branches of hundreds of large arteries which lead to countless arterioles and smaller vessels that end in a network of capillaries [34].

2.1.2 Blood pressure physiology

BP is the force exerted by the blood flow against the walls of the blood vessels. It is a vital physiological parameter and one of the most frequently measured in clinical practice. With each beat of the heart, blood flows from the heart to the peripheral vasculature causing the pressure to vary between systolic (maximum pressure) and diastolic (minimum pressure) [35]. The pressure generated by the intermittent flow is closely related to the function of the heart as well as the thickness and elasticity of the blood vessels [36]. Moreover, owing to the sequential contraction of the pumping heart, the pressure and flow are pulsatile in many parts of the CV system. The unit of measurement for BP is millimetres of mercury (mmHg).

As the left ventricle contracts, the blood is ejected into the aorta causing its pressure to rise to approximately 100 mmHg. Two pressure waves are associated with every contraction, namely, forward wave and reflection wave. The former is the pressure wave travelling from the heart towards the peripheries, while the latter is the reflected wave travelling back from the peripheries towards the heart [37, 38]. The pressure wave transmitted by the aorta travels fast along the arterial tree. The speed at which the pressure wave propagates is associated with the properties of the arterial wall. It is well established that increased arterial stiffness implies an increase in wave propagation speed [39].

The BP waveform is constantly changing in terms of amplitude and frequency. These fluctuations may last for different time windows, varying from minutes, hours, or days, occurring over the course of the time [40]. The changes in BP occur in response to different daily events and is further influenced by activities such as exercise, stress, food intake, medication or sleep patterns. The responses, and subsequently variations in BP, are triggered by autonomic nervous system in order to maintain CV homeostasis.

2.1.3 Factors influencing BP

There are several factors affecting BP, particularly, those that are related to CV activity. The heart rate (HR) influences the volume of blood prior to heart's contraction (known as preload) and the volume of blood ejected i.e. cardiac output (CO), which positively impact BP. Cardiac output varies in proportion to the metabolic demand of the body. It is the product of HR and stroke volume (blood volume per ejection), hence, HR is proportional to CO [41]. In addition to CV activity, there are several physiological factors that cause changes to BP in a person over time, namely, ageing, physical activity, gender, and changes in posture.

Ageing effect on BP

BP will inevitably rise with age, even in healthy subjects [42]. This has been repeatedly proven and recognised as an aspect of human ageing [42, 43]. Moreover, human ageing is associated with increased arterial stiffness [44]. Is it known that increased stiffness of the arteries influences the transmission velocity of the pressure pulse, affecting CO and increasing BP [45].

The large elastic arteries help dampening the pulsation delivered by the heart. However, with age, the large arteries lose their distensibility, resulting in higher SBP and pulse pressure (PP) [45]. This is due to the fact that stiffening in the arteries increases the propagation speed of the pressure wave, and this leads to earlier return of the reflected pressure wave from the periphery back to the heart. The reflected wave occurs during systole in older individuals. Thus, the amplitude of the pressure will be higher, and consequently SBP will be higher. On the other hand, in young individuals, the reflected wave occurs during diastole [39, 46]. It is acknowledged that stiffening of the large arteries leads to unfavourable CV outcomes, irreversible deadly complications and cardiac failure [45, 47].

Gender differences in BP

It has been demonstrated that there is gender associated difference in blood pressure [48, 49]. A few studies have shown that men have a higher BP than women of similar

ages [50, 51, 52]. Hence, in general, men are at higher risk of CV and kidney disease in comparison to women [53]. It is well established that women have less stiff arteries compared with age-matched men [54, 55]. However, due to increased arterial stiffness occurring after menopause, BP rises in women to similar levels or even higher as in men have been reported. Furthermore, the Framingham Heart Study showed that hypertension is more common among elderly women than elderly men [56].

Effect of physical activities on BP

Physical exercise is considered an effective drug-free approach for controlling hypertension. Incorporating physical exercise in the daily routine is a helpful strategy for delaying or even protecting against developing hypertension. Studies have shown that regular aerobic activity decreases resting SBP and DBP [57, 58]. Additionally, it has been reported that an increased volume of physical exercise is associated with a decrease in CV and all-cause mortality [59, 60, 61]. However, failing to maintain frequent and regular physical training will cause BP to return back to baseline levels [62]. The physical activity for health guidelines published by the World Health Organisation (WHO) recommend 150 minutes of moderate physical exercise or 75 minutes of vigorous physical exercise a week or a combination of both moderate and high intensity exercise [63].

Effect of postural change on BP

A change in posture such as transitioning from sitting in a chair or supine position to standing up position also causes a change to blood volume and pressure distribution in the body [64]. Due to the gravitational force, the blood volume shifts to the lower extremities, leading to a decrease in venous return [64]. This results in a decrease in stroke volume, and subsequently a decrease in CO and mean arterial blood pressure (MAP). Furthermore, reduction in MAP also reduces the blood flow to the brain.

2.1.4   Importance of regular BP monitoring

Though BP fluctuates due to short term CV response to different behaviour occurring during the day, evidence suggests that increased BP variations may have important clinical significance as it may reflect inherent changes in the CV system or outcome of other pathological conditions. Indeed, it has been demonstrated that

11

there is a positive correlation between enhanced BP variations and increased risk of CV events, internal organ damages, and CV mortality [65, 66].

Moreover, irregular BP (low or high BP) also leads to several pathologies relating to renal diseases and CV diseases such as heart attack and stroke as well as other life threatening complications [67]. Thus, there is no doubt that regular BP monitoring is essential for accurate assessment including early detection, diagnosis and treatment of BP variations and abnormal BP. However, the most common BP measurement devices can only provide intermittent BP measurement with repeated arm cuff inflations. This presents an obstacle for regular BP monitoring and cannot capture the BP variations accurately during daily life. Additionally, intermittent BP measurement devices often fail to capture true BP values due to masked hypertension [68] and white coat hypertension syndrome [69], which may lead to misdiagnosis. These challenges can be avoided by developing continuous, cuffless, and accurate self-measuring BP method for the early diagnosis, control and management of abnormal BP behaviour.

2.1.5 BP parameters

The definition of BP parameters are as follows:

Systolic blood pressure (SBP): is the maximum pressure in the blood vessels, as a result of the rapid flow of blood, caused by the systolic contraction of the heart.

Diastolic blood pressure (DBP): is the low pressure, in the relaxed blood vessels, when the heart rests between beats.

Mean arterial pressure (MAP): is the average BP in a cardiac cycle. It is calculated as:

$$MAP = DBP + 0.33 \times (SBP - DBP) \tag{2.1}$$

Pulse Pressure (PP): is the difference in pressure between SBP and DBP. It is calculated as:

$$PP = SBP - DBP \qquad (2.2)$$

2.1.6 Abnormal BP categories

BP levels fall into three main categories, namely, hypotensive, normotensive and hypertensive. Figure 2.1 presents these categories, as well as the SBP and DBP ranges in each category. Normotensive refers to healthy normal BP levels, ranging between 90 to 139 mmHg for SBP and 60 to 89 mmHg for DBP [1]. BP values lower or higher than these ranges are considered abnormal and can be dangerous if left untreated.
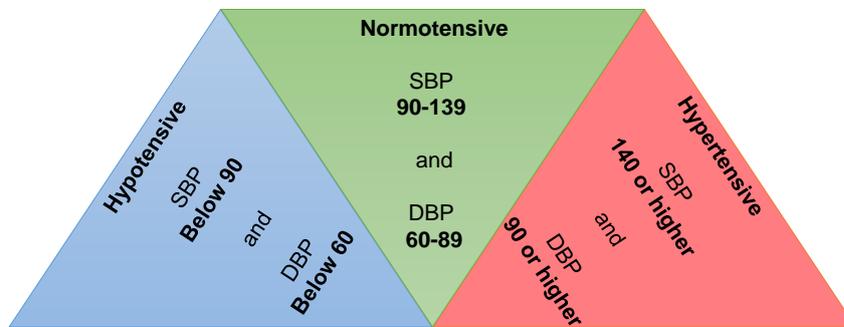
Figure 2.1: Blood pressure levels [1]

Hypotension

Hypotension refers to low BP levels. In this condition, the SBP and DBP ranges are below 90 mmHg and 60 mmHg, respectively. Hypotension leads to a decrease in blood flow resulting in a mismatch between the supply and demand of oxygen in the body [70]. As a result, hypotensive subjects may experience blurry vision, dizziness, fatigue, nausea, etc. There are a range of different conditions that cause hypotension such as dehydration, certain medicines e.g., BP control medicines, blood stream infection, shock, loss of blood due to injuries, and pooling of blood in the lower extremities upon transitioning from a sitting position to standing up.

Hypertension

Hypertension refers to high BP levels, where the SBP is above 140 mmHg and DBP is above 90 mmHg. In chronic hypertension, BP is elevated at rest. This medical condition can damage the internal body organs when left untreated and poses several health risks with potentially fatal consequences [71]. Early stages of hypertension often do not display any symptoms, hence patients are unaware of their health condition until later stages. For this reason, it is known as the "silent killer" and could lead to irreversible health complications [72].

High BP adds an excessive amount of force against the blood arteries. Over time, this strain may cause arteries to become less flexible with narrower interior space (lumen), which in turn increases the chances of having a clot. This leads to various pathological conditions relating to cardiovascular disease (CVD), such as stroke and heart attack [67]. Moreover, it is known that hypertension is a major risk factor for CVD [73]. Alarmingly, according to the WHO reports, more than one billion people are affected by hypertension [10]. This makes hypertension a global public health issue [74].

2.2 Current invasive and non-invasive BP measurement techniques

Several BP measurement devices are available in both clinical practice and research. These BP measuring methods and/or devices can be categorised into invasive and non-invasive. The non-invasive methods can further be divided into cuff-based and cuff-less. This section describes these methods along with their advantages and disadvantages. BP measurement categories and their limitations are presented in Figure 2.2.
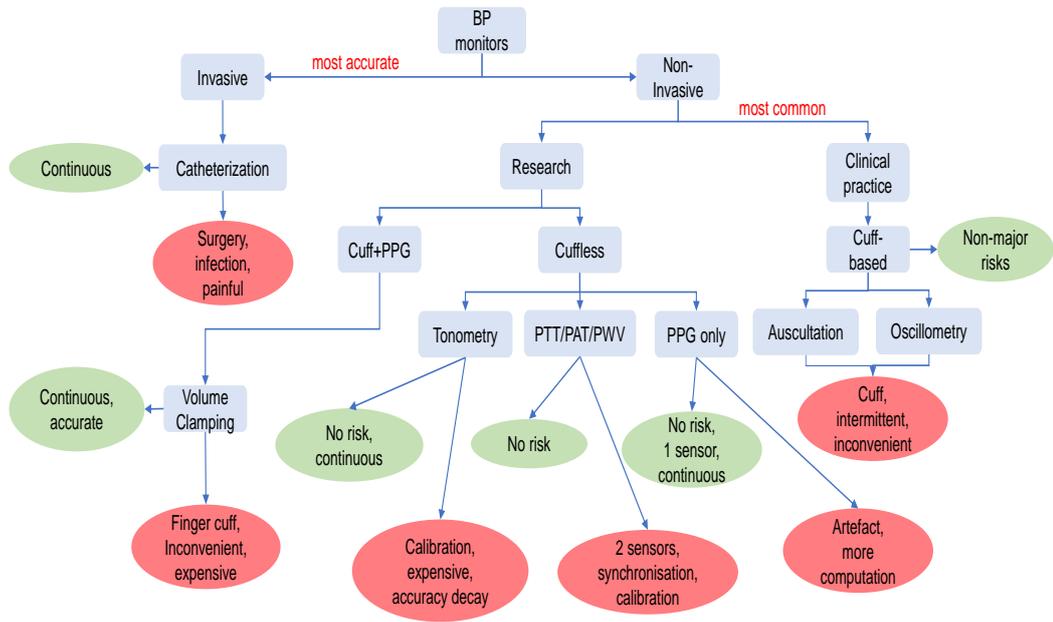
Figure 2.2: Blood pressure measurement methods. Green and red colour indicate their advantages and disadvantages.

2.2.1 Invasive measurements

The intra-arterial catheters are the primarily measurement methods used in hospital for providing continuous measurement of BP, when close monitoring is required in extremely sensitive and unstable cases [75]. The invasive intra-arterial method works by inserting a catheter incorporating a blood pressure sensor (electronic pressure transducer) directly into a pulsating blood vessel (e.g. brachial, axillary or radial artery) to measure the arterial pressure [13], as shown in Figure 2.3.

Advantages and limitations

BP measured invasively is continuous in nature, clinically accepted and the most accurate, hence, it is recognised as the gold standard for blood pressure measurement internationally [76]. However, there are several disadvantages associated with this method. First, it requires a trained professional to carefully place the catheter. Second, it is severely invasive, painful, carries multiple risks such as infection, and bleeding. Third, its application is strictly limited for monitoring critically ill patients in hospitals, primarily used during surgical procedures and in intensive care

units [77, 35].
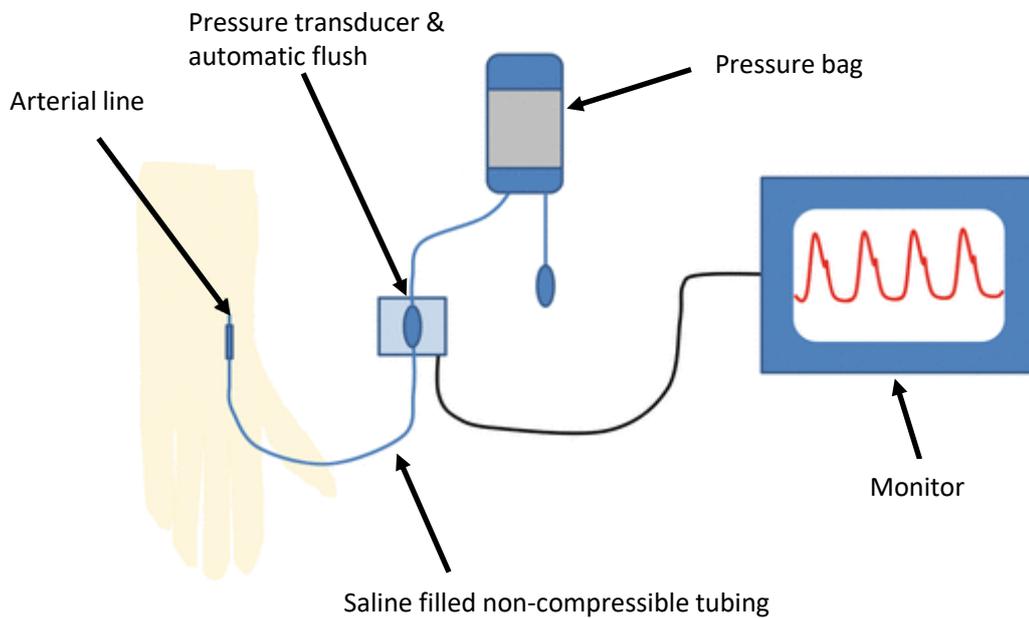
Figure 2.3: Invasive arterial blood pressure monitor.

2.2.2 Noninvasive- cuff-based and cuffless techniques

Cuff-based - auscultation

Auscultation is a non-invasive cuff-based method for measuring systolic and diastolic BP [78]. It is the standard and the most widely used BP measurement technique in clinical practice. Additionally, it is commonly used as reference method for calibrating as well as validating BP derived from physiological signals (PPG, ECG, BCG etc) using methods such as pulse transit time and pulse arrival time. The auscultation technique requires an inflatable cuff for occluding the blood flow in the artery, a stethoscope for listening to the Korotkoff sounds and a manometer for indicating the pressure values at each sound. The Korotkoff sounds are the changing arterial sounds heard using a stethoscope as the flow of blood resumes into the limb, when the cuff begins to deflate. Typically, a trained personnel is required to place the cuff around the arm and properly inflate the cuff to occlude the brachial artery. When the cuff pressure decreases during cuff deflation, the operator uses a stethoscope to identify the Korotkoff sounds generated by the return of the turbulent blood flow that corresponds to SBP and the laminar blood flow (which is silent) that corre-

16

sponds to DBP [78], as depicted in Figure 2.4.
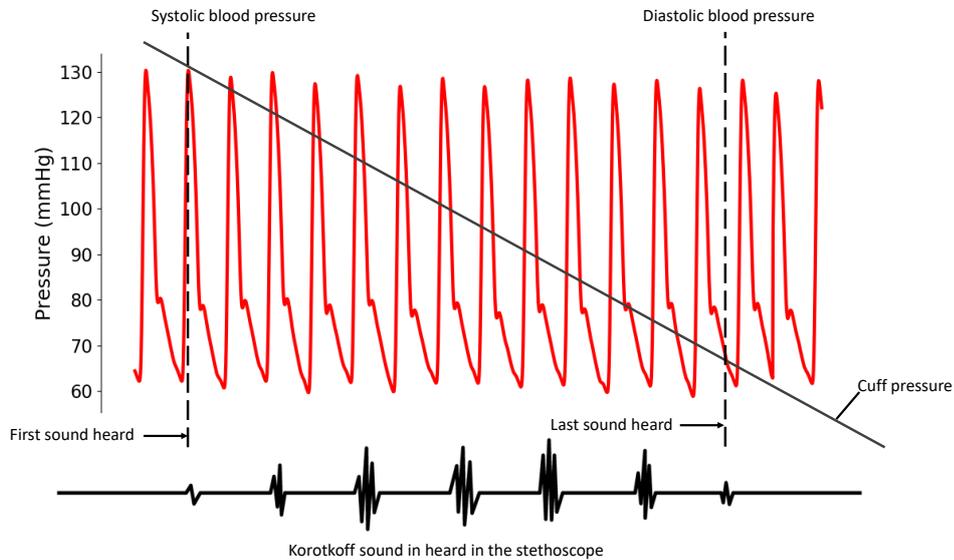
Figure 2.4: Auscultatory BP measurement method.

Advantages and limitations

The advantage of this technique is that it can provide SBP and DBP without any major risks as opposed to the invasive measurement. However, it can only provide BP measurement intermittently, with at least 2 minutes interval between measurements. Additionally, it is not suitable for regular self-measurement as it requires a trained professional to wrap and inflate the cuff properly around the arm and listen to the Korotkoff sounds. Moreover, patients have reported discomfort and inconvenience when measuring BP using these devices due to the compression of the artery by the cuff. Furthermore, it may result in inaccurate BP measurement associated with "masked hypertension" [68] or the "white coat hypertension" syndromes [69]. The former scenario occurs when a normal BP is observed for a patient when their actual BP is high, as opposed to the latter where high BP is observed for a nervous or anxious patient with actually a normal BP.

Cuff-based - oscillometric

Oscillometry is a popular automated non-invasive BP measurement technique used in clinical practice and a simpler monitoring technology in comparison to the auscultation method [79]. Similar to auscultation, this method also requires an inflatable

17

cuff for occluding blood flow. However, unlike auscultation, the cuff has a built-in pressure sensor for detecting pulsations, which are represented as oscillations. The oscillation amplitude varies with the applied cuff pressure and mean arterial pressure (MAP) forming a quasi-symmetric waveform from which BP values can be estimated using empirical fixed ratios principle [80]. As the cuff pressure decreases, the pulse amplitudes increase until it reaches a maximum, that corresponds to MAP, as illustrated in Figure 2.5. Afterwards, the pulse amplitudes decrease until they reach a minimal. Particularly, the oscillation amplitudes increase when the cuff pressure is in range of MAP and DBP, while the oscillation amplitudes decrease when the cuff pressure is between the MAP and SBP.
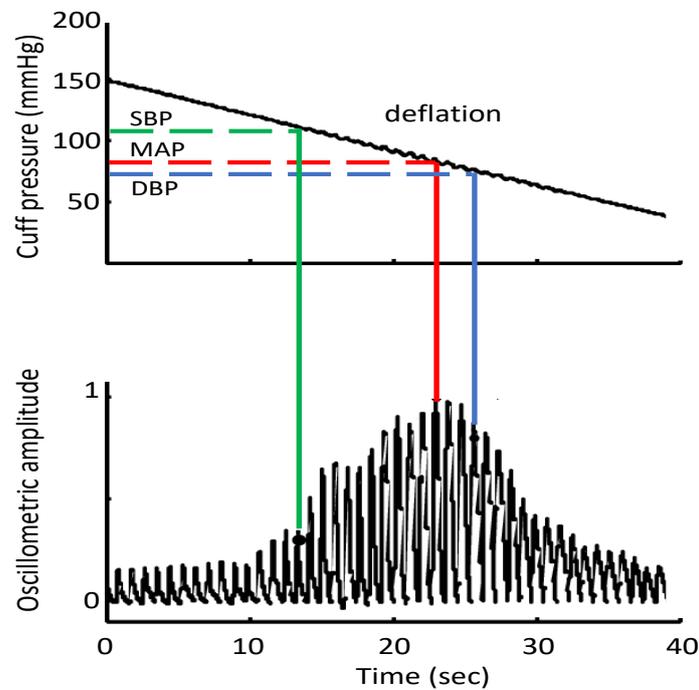
Figure 2.5: Determining BP using oscillometry [2].

Advantages and limitations

Similar to the auscultation method, this technique provides intermittent non-invasive BP measurement without any severe side effects. It also provides additional features in comparison to the auscultation method, such as, automatic BP measurement, and does not require an operator. Thus, it enables individuals to self monitor their BP and overcome the inaccuracies associated with white-coat hypertension or masked hypertension syndromes. However, it shares most of the disadvantages associated

18

with cuff based methods, namely, BP measurement are not continuous, uncomfortable, tedious, inconvenient and can be inaccurate if used improperly [81].

Finger cuff - volume clamp

Unlike the previous two full occlusion cuff-based methods (auscultation and oscillation), volume clamp is a non-invasive semi-occlusion finger-cuff based BP measurement device that is mainly available in research [82]. Volume clamp, also known as method of Penaz, utilises multiple sensors to provide continuous BP measurement, including BP waveform. Essentially, it uses a finger cuff with a built-in PPG sensor [83], as shown in Figure 2.6. The PPG sensor serves to detect volumetric change in blood flow, while the finger cuff, which is connected to a rapid servo controller system, is used to keep the blood volume constant with each cardiac cycle, through the rapid inflation and deflation of the cuff [84]. In particular, the external pressure applied by the inflatable cuff is equivalent to the arterial BP at the finger. This keeps the transmural pressure equal to zero, which is referred to as the set point. Thus, the dynamically controlled pressure applied by the rapid servo pump corresponds to the arterial pressure at the finger at the unloaded state. This in turn provides a continuous pressure pulse waveform with each heartbeat, from which continuous BP is then calculated using empirical models. An example of a commercially available device that is based on the volume clamp method is the CNAP monitor (CNSystems Medizintechnik GmbH, Austria) [85].

Advantages and limitations

The main advantage of this method is that it provides non-invasive continuous BP measurement through a small finger cuff, which permits long term recording. Additionally, this method is accurate for tracking changes in BP [86]. However, the device is large, very expensive, the finger cuff pressure can be inconvenient, and might lead to venous congestion and numbness around the finger during long term recording. Additionally, evidence suggests that this method can overestimate SBP [87]. Moreover, it can produce inaccurate BP measurement with improper set up.

Figure 2.6: Volume clamp for BP monitoring using a finger cuff [3].

Cuffless- tonometery

Unlike all previously mentioned methods, arterial tonometery is a non-invasive and cuffless technique providing continuous measurement of BP, as well as the BP waveform signal [88]. This method provides instantaneous BP by pressing a sensitive transducer directly on an artery close to a rigid surface, as illustrated in Figure 2.7. Favourable measurement sites that are close to an underlying bone, providing easy access and large diameter are the dorsalis pedis, carotid, brachial and radial arteries. During measurement, the arterial wall must be flattened, therefore, the pressure sensor must be placed accurately above the artery and lightly pressed so that the vessel is compressed against the underlying bone. This keeps the transmural pressure equal to zero. Under these conditions, the pulse pressure acquired by the sensor, in each cardiac cycle, translates into electrical signals depicting arterial pressure pulse. The maximum pulse pressure determines the MAP. Additionally, SBP and DBP can be determined by processing the acquired arterial pressure waveform using an underlying algorithm.

Advantages and limitations

Figure 2.7: Applanation tonometry for cuffless continuous BP measurement.

Arterial tonometery offers continuous measurement of BP non-invasively and without an inflatable cuff. However, manual and automatic measurements have been proven difficult to implement for two main reasons: 1) correct positioning of the sensor over the artery is highly critical for accurate measurement [89] and 2) the sensor is highly sensitive to arm movements and muscle contractions [90], thus it requires constant precise control of the sensor positioning. Furthermore, continuous pressure against the artery can be uncomfortable for patients and impracticable for long term recording. Additionally, the device is expensive and requires a trained operator to take the measurement, hence it is not affordable nor practical for continuous self BP measurement.

2.3 BP estimation using physiological signals

Over the past several decades, many research groups across the world have devoted a lot of time and considerable effort to provide non-invasive cuff-less and continuous BP monitoring. The motivation behind this work is to replace the current cuff-based BP devices. Cuff-based devices often need trained personnel, and they can cause irritation and inconvenience for patients due to cuff inflation and deflation. Cuff-based methods do not provide continuous BP measurements and are sometimes inaccurate. Consequently, current clinical cuff-based BP devices are not suitable for providing continuous BP monitoring which could play a significant role in the early detection of cardiovascular diseases amongst many other applications. One way to overcome these challenges is the photoplethysmography approach. The PPG approach allows

for the estimation of BP non-invasively by observing two waveforms obtained from signals such as two PPG signals from two anatomical locations or a combination between a PPG signal and the electrocardiogram (ECG). Most recently, there has been more traction towards monitoring BP using only one PPG waveform. Several methods for estimating BP from physiological signals have emerged in the literature such as PWV, PTT, PAT and PWA. This section introduces the PPG and describes these methods.

2.3.1 Introduction to photoplethysmograph (PPG)

Photoplethysmography is a simple, low cost and non-invasive optical method used for measuring volumetric changes in blood flow per pulse [91]. It is primarily utilised in healthcare for measuring heart rate and blood oxygen saturation using a pulse oximeter [92]. However, given the recent technological advancements in signal processing techniques, wearable technologies and the rise of machine learning, numerous research groups around the world have recognised the remarkable potential that these algorithms can bring to the healthcare industry for improving patient's wellbeing. As such, there is a lot of interest in utilising the PPG as a possible non-invasive solution to a wide range of challenging tasks such as determining haemoglobin levels [93, 94], blood glucose [95, 96] and oxygen saturation levels [97, 98], coronary artery disease [99, 100], stress levels [101, 102, 103], respiratory rate [104, 105], PPG signal quality assessment [106, 107] and most recently for continuous cuffless BP measurement [108]. Although there is a continually increasing interest in the PPG, the origins of the components of the PPG signal are still not clear [91]. Nonetheless, the PPG is considered a useful physiological signal since it can provide fundamental information about the cardiovascular system [91].

The principle of photoplethysmography

The PPG sensor consists of two components: a Light Emitting Diode (LED) to illuminate the skin surface and a photodetector for measuring the changes in light absorption or reflection, from the vasculature, over a period of time. The amount of reflected or absorbed light is governed by the blood circulation in the illuminated skin [109]. There are two different measurement configurations for recording the

PPG signal, namely, transmission and reflectance. In the transmission mode, the LED is place on one side of the tissue, while the photodetector is on the opposite side. Thus, the light intensity is measured by the photodetector after it had gone through the tissue. The measurement sites for this mode are mainly limited to the fingertip, toe, and ear lobe. However, this limitation can be avoided using the reflectance mode. In this mode, both the LED and photodiode are adjacent and placed side by side on the skin. The photodiode detects some portion of the reflected backscattered light emerging from the skin next to the LED. The measurement sites can be the wrist, chest, neck, forehead, legs or any other location with adequate capillary bed [109].

PPG waveform

The PPG waveform is divided into an AC and DC components. The high frequency part of the PPG signal, corresponding to the AC component, contain information regarding heart pulsation [91]. This component is superimposed onto a large non-pulsating, slow varying, lower frequency part known as DC component. The DC part is affected by various factors such as respiration, absorption from non-vascular tissue, and sympathetic nervous system activity [91]. The amplitude of the pulsatile AC component varies with each cardiac cycle. When the heart contracts during systole, blood volume and BP increase as a result of the blood pulse propagating from the heart towards the periphery, and subsequently light absorption increases. Conversely, when the blood volume and BP decline during heart diastole, light absorption also decreases. The PPG signal is synchronous to the blood pulse and has a wavelike form that resembles the BP waveform. Consequently, the repeated systole and diastole of the heart result in a periodic PPG signal exhibiting one systolic and one diastolic peak in every single cardiac cycle. The turning point or the valley between the systolic and diastolic peaks is called dicrotic notch. A typical PPG cycle is shown in Figure 2.8. The systolic peak is normally clear and easy to detect while the diastolic peak and dicrotic notch are normally visible mostly in healthy and young people and become undetectable in older and hypertensive patients. Although the relationship between the PPG and BP is not yet clear, the small variation in blood volume reflected in the PPG signal appears to be correlated with BP [91].
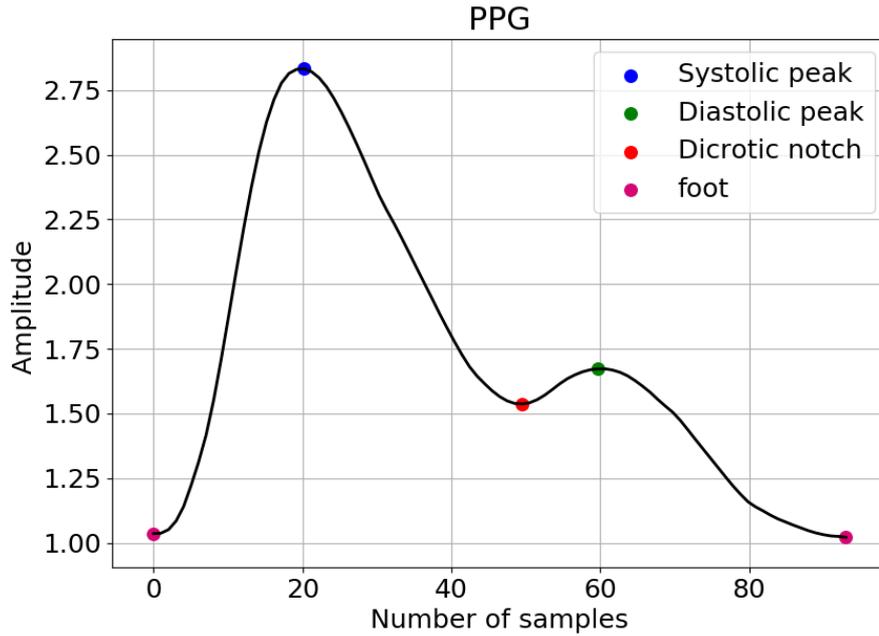
Figure 2.8: Typical PPG cycle with key points.

2.3.2 Pulse Wave Velocity (PWV)

Pulse wave velocity is an alternative cuff-less method for BP estimation [110]. It is the speed of the pressure wave propagation in the blood vessels which is based on the theory of wave propagation for fluids in elastic pipes. The motivation behind this approach is that BP can be determined from the velocity of the heartbeat pulse. The heart initiates the pressure pulse, in turn the blood is pushed or propagates through central arteries down to smaller distal arteries by expanding and contracting during systole and diastole respectively [111]. This phenomena results in the changes of the vessels wall elasticity and highly affects the velocity of the pressure pulse. Particularly, the elasticity of the arteries determines the speed at which the pulse wave travels [76]. This relationship can be illustrated using Moens-Kortweg equation [112] given as follows:

$$PWV = \sqrt{\frac{E \times H}{2 \times R \times P}} \qquad (2.3)$$

Where E, H, R, and P represent the incremental elastic modulus, ratio of arterial

wall thickness, radius of the vessel, and blood density, respectively.

PWV is measured using two PPG sensors located on the same arterial branch with a known distance apart. PWV can be obtained by dividing the artery length (D) between the two references by pulse transit time (PTT) as follows:

$$PWV = \frac{D}{PTT} \qquad (2.4)$$

For example, McCombie et al (2006) [112] took advantage of the relationship between BP and vessels elasticity to derive BP through PWV approach using two PPG signals. The artery length was measured as a distance, while the PTT was measured as the time difference for the pressure wave to travel from the previous PPG sensor to the leading PPG sensor. This method is difficult to perform non-invasively as several challenges occur during the calculation of PWV. It requires two measurements from two sensors. The arterial elasticity varies between individuals and is highly dependent on factors such as age, diet, height, etc. The length of the artery mandatory for the equation above also varies from one person to another. Therefore, it requires frequent calibration due to different physiological parameters between individuals as well as the expiration of the calibration in a short period of time [113]. This concern is the bottleneck preventing PWV from being used in health care. Calibration procedures are not permitted by health care standards as part of physiological measurement [114]. Hence, PWV is not practical nor a suitable replacement for cuff-based BP devices.

2.3.3 Pulse Transit Time (PTT)

Pulse transit time is yet another approach for cuff-less non-invasive BP measurement [21]. It is defined as the time that it takes for the pressure wave to travel between two arterial sites [20]. There is an inverse proportional relationship often found between PTT and BP. PTT can be measured using two PPG sensors located on two distant sites in the body. The parameter is estimated as the time delay between the proximal and distal PPG waveforms. It should be noted that many papers in

the literature refer to pulse arrival time as PTT [115, 15, 116], however, PTT is hardly investigated [76, 117]. The most common sites for PTT measurements are fingers, ears and toes [118]. Different measurement points have been investigated in the literature. For example, Chen et al [17] found that there is a strong relationship between PTT measured as foot-to-foot time delay and invasive Diastolic Blood Pressure (DBP), however, another study suggested that is not always the case [18]. PTT can also be measured as the time difference from mid-point of the falling edge of the proximal PPG to the peak of the peripheral PPG or as the time difference from dicrotic notch of the proximal PPG and the peak of the peripheral PPG [19], as shown in Figure 2.9.
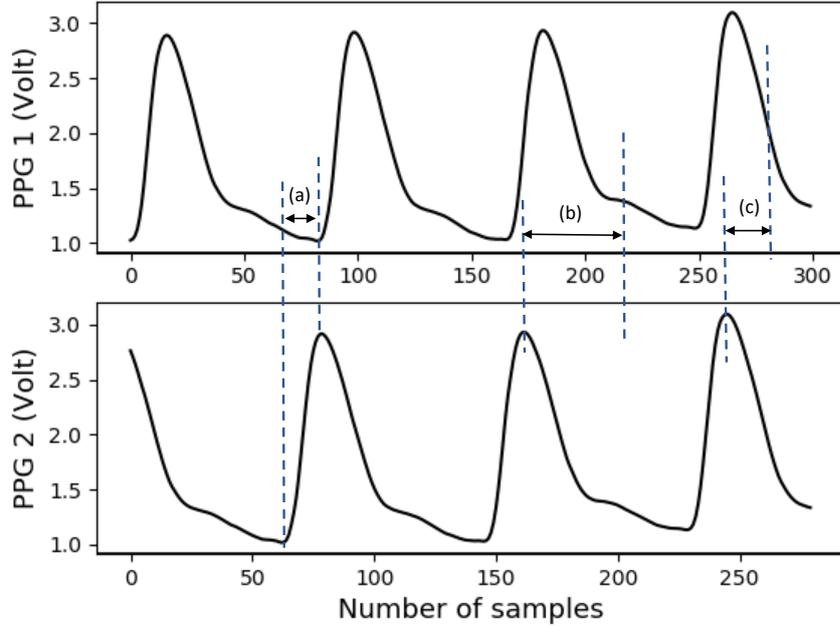
Figure 2.9: Different PTT measurement points between two PPGs. (a) foot-to-foot time delay, (b) peak-to-dicrotic notch time delay, and (c) peak to mid-point of the falling edge time delay.

BP measurements using the PTT based approach typically involves three steps: two PPG sensors for measuring proximal and distal PPG waveforms, calculation of the PTT parameter, and calibration. Hence, there are several obvious disadvantages when using the PTT approach for estimating BP. Firstly, two sensors are needed for the estimation. PPG sensors are very sensitive to patient/probe movement which results in motion artefacts in the waveforms [91]. Consequently, signal processing

needs to be done on both waveforms for smoothing and motion artefacts removal etc., whilst keeping the recordings in sync. Additionally, it is affected by physiological parameters of individuals and thus requires per person calibration [20].

2.3.4 Pulse Arrival Time (PAT)

Pulse arrival time (PAT) is defined as the time interval between the electrical activation of the heart and arrival of the pulse wave at a location on the body like the finger, toe, and forehead. In other words, PAT is the sum of PTT in addition to the ventricular electromechanical delay and isovolumic contraction period, commonly known as Pre-ejection Period (PEP) delay [20]. PEP can be influenced by stress, age, emotion, and movement. PAT is measured using two sensors, an ECG and a PPG sensor. It is estimated as the time difference between the R peak of the ECG and a point on the PPG rising edge [20]. Three characteristic points on the PPG waveform have been considered to calculate the time delay such as the foot of the PPG [119], mid-point on the rising edge [120] and peak of the PPG [121], as shown in Figure 2.10. Although it was found that PAT can reduce the diastolic pressure accuracy [122], it is still used in the literature for its simplicity. Some studies show that PAT is an inadequate surrogate for PTT for systolic and diastolic blood pressure [22], however, others suggest that PAT improves Systolic Blood Pressure (SBP) [32]. This method shares the same disadvantages mentioned in the PTT section. For instance, PAT is measured using two different sensors, namely, ECG and PPG, that often recorded with different sampling frequencies. Meaning that the signals may have different number of data samples per second, therefore they are not time aligned, and this requires an additional preprocessing step such as up-sampling or down-sampling to one of the signals before any time domain analysis can be made. Additional, accurate parameter estimation is highly dependent on precise synchronisation of the two signals for peak detection. This is highly demanding in practice. Moreover, recoding the ECG signal requires attaching several electrodes to the chest which is inconvenient. Furthermore, both sensors are prone to motion artefacts and require signal processing which is not straight forward especially if continuous monitoring of blood pressure, and intermittent measurement is desirable. PAT also requires calibration for different individuals and the calculation of PEP parameter.
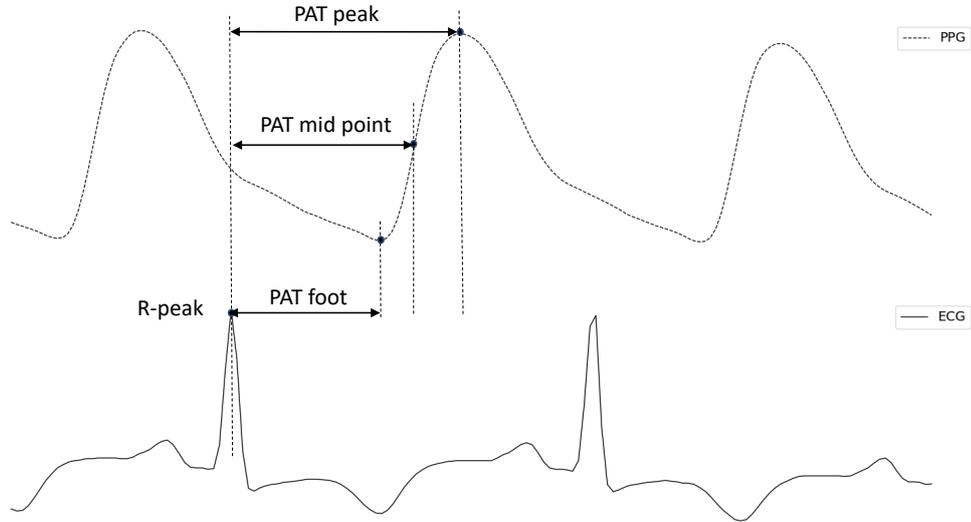
Figure 2.10: PAT measurement points, from R-peak of the ECG to foot, mid-point on the rising edge and peak of the PPG.

2.3.5 Pulse Wave Analysis (PWA)

Pulse Wave Analysis (PWA) refers to signal processing and extractions of certain characteristic features from the PPG waveform. This method requires only one measurement sensor, the PPG. Development in computing and data analysis tools have made it easier to pre- and post-process physiological signals such as the PPG and ECG. Signal processing like filtering and feature extraction have been employed in PPG pulse wave analysis. Although the PPG waveform looks similar to a BP waveform, these two waveforms are not the same [123, 124]. However, numerous PPG morphological features, including features extracted from first and second derivatives of the PPG, have been found useful in quantifying import characteristics about the cardiovascular system, such as features describing arterial tone changes, total peripheral resistance and arterial stiffness which are all directly related to BP [123, 124]. These features are typically used for creating models using machine learning and deep neural networks for estimating blood pressure. Several studies have investigated the feasibility of cuff-less and continuous BP predictions using only one PPG sensor [108]. This approach is very promising and appealing as the PPG technology used for such applications is relatively simple and inexpensive, plus the acquisition of

the PPG signals is straight forward, assuming that the sensor is placed on a vascular tissue. Additionally, it has been shown that the PPG and BP morphology undergoes similar changes in vascular disease and the peripheral volumetric changes are correlated with BP. The main disadvantages attached to this approach is that the PPG is prone to motion artefacts due to movements, and the relationship between the BP and the PPG waveform is not fully understood in comparison with PTT/PAT approaches [91]. Hence, more research is certainly needed to better understand this approach.

2.4 Summary

Abnormal BP levels increase the risk of future health problems and can incite several irreversible life-threatening diseases. BP is a direct indicator for hypertension and hypotension, as such, regular continuous BP monitoring is pivotal for the early detection, prevention and management of hemodynamic diseases as well as CVD. However, current most commonly used clinical BP measurement methods do not permit continuous self-monitoring of BP. Additionally, the current cuffless BP methods introduced in research are either expensive, tedious, impractical, or inaccurate. Recent advancements in signal processing techniques, including machine learning and artificial intelligence, have also opened up exciting new horizons for cuffless and continuous monitoring of BP using physiological signals, in particular using the PPG signal. The next chapter introduces the cuffless BP measurement methods using machine learning utilising the PPG signal as well as other signals such as ECG.

Chapter 3

A review of existing Cuffless BP Estimation Methods Using Machine Learning

Traditional BP monitoring devices have failed to provide ease of use for cuffless, continuous, self-monitoring, and management of BP. Most recently, the rise of machine learning and advances in wearable technologies have given way for new research towards cuffless BP estimation using a combination of signals, such as ECG, PPG and ballistocardiograms (BCG) or using only a single PPG sensor. Several studies have tried to improve on the traditional PTT/PAT approaches by estimating BP using data driven machine learning models, while other research studies have attempted to estimate BP by focusing on the analysis of the PPG waveform that generally capture the shape of the PPG signal. This chapter provides a comprehensive review of non-invasive cuff-less blood pressure estimation using PPG signals as well as other signals along with their advantages, challenges and limitations. Table 3.1 summarises all attempted methods in a chronological order from the earliest publication to the most recent.

Table 3.1: Summary of all machine learning and neural network based BP estimation models using physiological signals, presented in this chapter

Authors	Dataset	Model	Input	SBP Error (mmHg)	DBP Error (mmHg)
		ECG and PPG sesnors			
Hassan et al (2008) [125]	own data, 10 subjects	Linear regression	feature-based	NA	NA
Wong et al (2009) [116]	own data, 14 subjects	least-square	feature-based	ME±SD: 1.4±10.2	ME±SD: 2.1±7.3
Shen et al (2015) [126]	own data, 10 subjects	stepwise regression	feature-based	MAE±SD: 5.67±6.08	MAE±SD: 4±3.82
Kachuee et al (2015) [113]	MIMIC II, 851 subjects	SVR	feature-based	MAE±SD: 12.38±16.17	MAE±SD: 6.34±8.45
Kachuee et al (2017) [114]	MIMIC II, 942 subjects	AdaBoost	feature-based	MAE±SD: 11.17±10.09	MAE±SD: 5.35±6.14
Miao et al (2017) [127]	own data, 73 subjects	SVR	feature-based	ME±SD: 0.85±5.78	ME±SD: -1.24±4.63
Wang et al (2018) [128]	MIMIC, 72 subjects	MLP	feature-based	ME±SD: 4.02±2.79	ME±SD: 2.27±1.82
Shimazaki et al (2018) [129]	own data, 1363 subjects	Autoencoder-MLP	raw ECG, PPG	MAE±SD: NA±11.86	NA
Tanveer and Hasan (2018) [30]	MIMIC, 39 subjects	MLP-LSTM	raw ECG, PPG	MAE±SD: 1.1±1.56	MAE±SD: 0.58±0.85
Su et al (2018) [29]	own data, 84 subjects	LSTM	feature-based	RMSE: 3.9	RMSE: 2.66
Chen et al (2019) [130]	MIMIC, 772 records	Genetic-SVR	feature-based	MAE±SD: 3.27±5.52	MAE±SD: 1.16±1.97
Ripoll and Vellido (2019) [131]	MIMIC, 250 subjects	RBM	feature-based	MAE: 3.7	MAE: 1.01
Li et al (2020) [31]	MIMIC II, 678202 cycles	LSTM	feature-based	ME±SD: 4.63±14.50	ME±SD: 3.15±6.44
		ECG, PPG and BCG sesnors			
Eom et al (2020) [132]	own data, 15 subjects	CNN-GRU	raw ECG,PPG,BCG	MAE±SD: 4.06±4.04	MAE±SD: 3.33±3.42
Lee et al (2020) [133]	own data, 18 subjects	BILSTM-MLP	feature-based	ME±SD: -0.07±7.3	ME±SD: -0.17±6.4
		Single PPG sensor			
Teng and Zhang (2003) [24]	own data, 15 subjects	MLR	feature-based	MAE±SD: 0.21±7.32	MAE±SD: 0.02±4.39
Suzuki and Oguri (2009) [134]	own data, 368 subjects	AdaBoost	feature-based	ME±SD: 1.2±11.7	NA
Ruiz-Rodriguez et al (2013) [135]	own data, 572 subjects	DBN-RBM	raw PPG	ME±SD: -2.98±19.35	ME±SD: -3.65±8.69
Kurylyak et al (2013) [25]	MIMIC II, 15000 heartbeats	MLP	feature-based	ME±SD: 3.8±3.46	ME±SD: 2.21±2.09
Suzuki and Ryu (2014) [136]	own data, 80 subjects	MLR	feature-based	MAE: 4.04	NA
Choudhury et al (2014) [137]	Queensland, 32 subjects	MLR	feature-based	MAE±SD: 0.78±13.1	MAE±SD: 0.59±10.23
Datta et al (2016) [138]	own data, subjects unspecified	MLR	feature-based	ME: 1	ME: -2
Duan et al (2016) [139]	Queensland, 32 subjects	SVR	feature-based	ME±SD: 4.77±7.68	ME±SD: 3.67±5.69
Xing and Sun (2016) [28]	MIMIC+ own data, 92 subjects	MLP	feature-based	ME±SD: -1.67±2.46	ME±SD: -1.29±1.71
Gaurav et al (2016) [140]	MIMIC, 3000 records	MLP	feature-based	ME±SD: 0.16±6.85	ME±SD: 0.03±4.72
Gao et al (2016) [141]	own data, 65 subjects	SVR	feature-based	ME±SD: 4.9±4.9	ME±SD: 4.3±3.7
Liu et al (2017) [27]	MIMIC II, 910 cycles	SVR	feature-based	ME±SD: 8.54±10.9	ME±SD: 4.34±5.8
khalid et al (2018) [26]	Queensland, 32 subjects	Regression tree	feature-based	ME±SD: -0.1±6.5	ME±SD: 0.6±5.2
Dey et al [142]	own data, 205 subjects	Lasso regression	feature-based	MAE±SD: 6.9±9	MAE±SD: 5±6.1
Fujita et al (2019) [143]	own data, 265 subjects	partial least square	feature-based	ME±SD: -0.28±17.92	NA
Hasanzadeh et al (2019) [144]	MIMIC II, 942 subjects	AdaBoost	feature-based	MAE±SD: 8.22±10.38	MAE±SD: 4.17±4.22
Slapnicar et al (2019) [145]	MIMIC III, 510 subjects	Spectro-temp ResNet	raw PPG	MAE: 15.41	MAE: 12.38
Wang et al (2020) [4]	MIMIC, 20 subjects	CNN-GRU	raw PPG	MAE±SD: 3.81±4.28	MAE±SD: 1.99±2.57
Chowdhury et al (2020) [146]	126 subjects, 2.1 s frames	Gaussian process regression	feature-based	ME±SD: 3.02±9.29	ME±SD: 1.74±5.54
Sadrawi et al (2020) [147]	own data, 18 subjects	Genetic autoencoder CNN	raw PPG	MAE: 2.54	MAE: 1.48
Ibtehaz and Rahman (2020) [148]	MIMIC II, 942 subjects	U-net + MultiResnet	raw PPG	ME±SD: -1.58±10.68	ME±SD: 1.61±6.85
Schrumpf et al (2021) [149]	MIMIC III, 4000 records	ResNet	raw PPG	MAE±SD: 12.51±12.61	MAE±SD: 8.3±9.84
Aguirre et al (2021) [150]	MIMIC III, 1100 subjects	seq-to-seq GRU + attention	raw PPG	MAE±SD: 12.08±15.67	MAE±SD: 5.56±7.32
Harfiya et al (2021) [151]	MIMIC II, 942 subjects	autencoder LSTM	raw PPG	MAE±SD: 4.05±4.42	MAE±SD: 4.6±3.47
Rong and Li et (2021) [5]	MIMIC II, 1000 subjects	2-CNN + LSTM	raw PPG	MAE±SD: 5.59±7.25	MAE±SD: 3.36±4.48

3.1 Cuffless BP estimation using physiological signals and machine learning

The earliest attempt to measure BP using only PPG signals was done by Teng and Zhang (2003) [24]. In their study, the relationship between arterial BP and four PPG features were evaluated using a linear regression model. The authors collected their own PPG and BP signals from 15 healthy male subjects aged 24-30. The features selected from the PPG signals were: width at 1/ 2 and 2/ 3 amplitude,

systolic upstroke time and diastolic time, shown in Figure 3.1. Two challenges were reported in the process of extracting these features. In some PPG signals, they have experienced a shift in the foot position and in other signals the foot position was not clear due to poor signal recording. It is crucial for the position of the peak and foot to be clear and consistent for extracting the correct feature values. Consequently, Continuous Wavelet Tranform (CWT) was employed to overcome the two aforementioned challenges. Additionally, the correlation between the extracted features and BP values were evaluated, and only features with the highest correlation to BP were selected for regression analysis. It was found that the diastolic time has a higher correlation with Systolic blood pressure (SBP) and diastolic blood pressure (DBP) than other features. The mean error and standard deviation between the estimated BP values and the reference BP values were 0.21±7.32 mmHg and 0.02±4.39 mmHg for SBP and DBP respectively. According to the American National Standards of the Association for the Advancement of Medical Instrumentation (AAMI), the mean difference and standard deviation of non-invasive BP should not exceed 5±8 mmHg from a reference BP evaluated on no less than 85 patients. However, the relationship between BP and PPG is not always linear [25, 129], and this study was conducted on only 15 young healthy male volunteers. This suggests that there is a low variability in terms of BP range between volunteers which may explain the reason behind the low estimation error using a linear model. Moreover, estimating SBP and DBP using classical machine learning methods requires two different models, one for each objective. In this case, DBP and SBP were strongly correlated [30], thus learning both objectives using one model structure would further improve the estimation by learning shared data representations. This can be achieved using a neural network for estimating both SBP and DBP simultaneously using one model.

In 2008, Hassan et al [125] derived a regression model for estimating only SBP based on the PTT method without the need for calibration for every individual subject. In their study, PPG and ECG signals were collected from 10 healthy subjects with BP reference values measured by a sphygmomanometer with a cuff attached to the subject's right arm. The peripheral pressure pulses were measured at the fingertip using a PPG sensor. Both PPG and ECG were recorded simultaneously for 45 seconds sampled at 1 kHz using AD instruments followed by the calculation of the PTT values. Regression models were then established for each subject and then
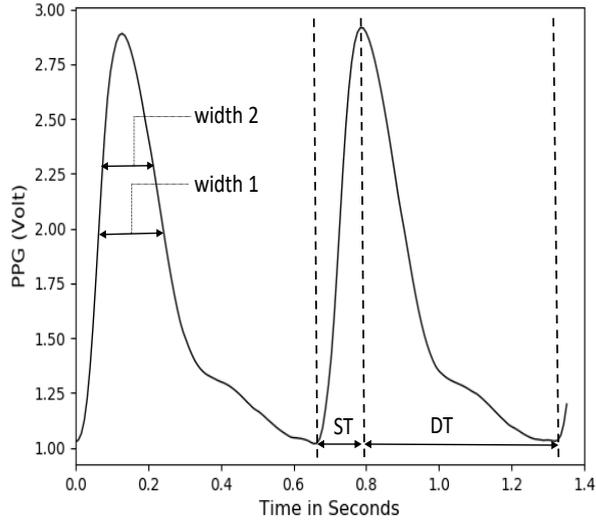
Figure 3.1: Four PPG features: Systolic upstroke Time (ST), Diastolic Time (DT), width 1 refers to width at 1/2 amplitude and width 2 refers to width at 2/3 amplitude.

combined together to get a new regression estimation model that fits all subjects. Subsequently, the average of the slope values from all individual regression models becomes the new reference slope for the new regression model. The results were below the 5±8 mmHg error rate set by the AAMI, however the study was performed on a very small dataset of only 10 all male subjects, while the AAMI requires at least 85 subjects. Having a diverse range of BP values that truly represent the population (males and females from different age range) would increase model generalisation for accurate BP monitoring. PPG features have also been proved to correlate favourably with BP [24, 123], thus including PPG features would improve the model's generalisation. The linear model performed relatively well given the linear relationship between PTT and BP when evaluated on this dataset. However, this model does not account for the temporal variation in the extracted features which should be modelled for continuous, accurate, and well generalised BP prediction. A simple non-recurrent model only considers the relevant features and with no feedback loop from previous cycles. This can be attained using a recurrent neural network represented by a recurrent link passing information learnt from previous time steps along with the present input features for estimating BP values.

Wong et al (2009) [116] investigated the correlation between BP and PTT under different circumstances, i.e., pre-exercise, post-exercise, etc. In this study, the model

was evaluated on 14 normotensive subjects with no history of cardiovascular disease using least-square regression. The authors designed an in-house circuitry to detect the derivative of the ECG (dECG) and PPG. The dECG and PPG were sampled by an analogue to digital converter (ADC) at 1 kHz and brachial BP was recorded intermittently on the subject's right arm with an automatic BP monitor. Beat-to-beat PTT parameters were calculated from the peak of the dECG to the peak of the PPG derivative. Two tests were carried out using the same model coefficient six months apart. The least square model from their first experiment was applied with the same coefficient half a year later to predict BP of a different pressure baseline. The results show that arterial BP increased and PTT decreased sharply after exercise and a high correlation between SBP and PTT was found. However, the regression coefficient obtained from the first study (6 months before) failed to predict BP well in all subjects when the blood pressure values changed in the second experiment. The mean error and standard deviation resulted from the aforementioned experiment were 1.4±10.2 mmHg for SBP and 2.1±7.3 mmHg for DBP. This drawback is due to the fact that PTT requires calibration when the blood pressure baseline changes between different subjects. Additionally, the least-square regression cannot estimate both SBP and DBP simultaneously and requires implementing two different models to learn each objective separately. As mentioned previously, these two objectives are correlated and thus should be estimated using one model to improve the estimation precision. Also, this technique requires two sensors for measuring the PTT parameters. It has been shown in the literature that the PTT parameters expire one day after the initial calibration which in turn would increase the estimation error [127]. For all these reasons, this technique is not reliable for long term continuous BP monitoring.

Suzuki and Oguri (2009) [134] presented a technique for measuring SBP using only a PPG sensor. In their study, SBP was estimated using error-correcting output coding method based on an aggregation of AdaBoost as a binary classifier machine. This method was evaluated on 368 volunteers. Individual information and characteristic features from their PPG waveform were used to calculate BP. The reference BP values were measured every minute with a cuff placed on the right brachial using a commercial BP device and the PPG signals were measured at the left finger by a commercial sensor. The PPG features selected in this study were percussion

wave, tidal wave, dicrotic notch and dicrotic wave. The mean error and standard deviation were 1.2±11.7 mmHg. Hence, the results were highly variable, did not satisfy the AAMI requirements and could only provide intermittent SBP measurement. Moreover, although AdaBoost is a non-linear function, it is not appropriate for time domain analysis and for handling the complexity of the task, therefore, its accuracy will decrease for multiday BP estimation. Additionally, the feature set is relatively small and should be enhanced further for effectively modelling the relationship between PPG features and BP.

Ruiz-Rodriguez et al (2013) [135] introduced a continuous cuff-less BP monitoring using a deep neural network, namely, Deep Belief Network- Restricted Boltzmann Machine (DBN-RBM). The authors collected their PPG signals through pulse oximetry with reference to invasive BP values. PPG and BP measurement devices were attached to a processing module connected to a General Electric (GE) Datex Ohmeda device. The signals were recorded for a period of 30 minutes. PPG and BP signals that exhibited anomalies such as overdamping or underdamping phenomena, motion artefacts (due to unexpected patient movement, cough, etc.) or extrasystole, were excluded during signal analysis and therefore ensuring that signal quality was optimal. Each 30-minute signals were then segmented into 10 seconds frames. The neural network model applied in this study, DBN-RBM, belongs to a family of networks that build probabilistic generative models. Values of SBP, MAP, and DBP were obtained through a mathematical algorithm that detects the maximum amplitude of the PPG oscillations. This promising method estimates BP continuously without a cuff and does not necessitate calibration. The advantages of this study is that it can model SBP, MAP and DBP using one structure, thus allowing the model to capture the strong correlation between the three objectives. However, the results of SBP, DBP and mean arterial pressure (MAP) predictions were highly variable which in turn caused the standard deviation to exceed 8 mmHg limit imposed by the standards of the AAMI. In particular, the standard deviations were 19.35, 10.35 and 8.69 mmHg for SBP, MAP, and DBP, respectively. It was stated that the high variability might be influenced by the respiratory variability in the PPG signals. Also, the PPG processing module significantly affected the results since it changes the shape of the obtained PPG pulse. The results of this study might be improved by providing a feedback link from previous cycles to the input layer to account for

the temporal dependencies in the PPG features.
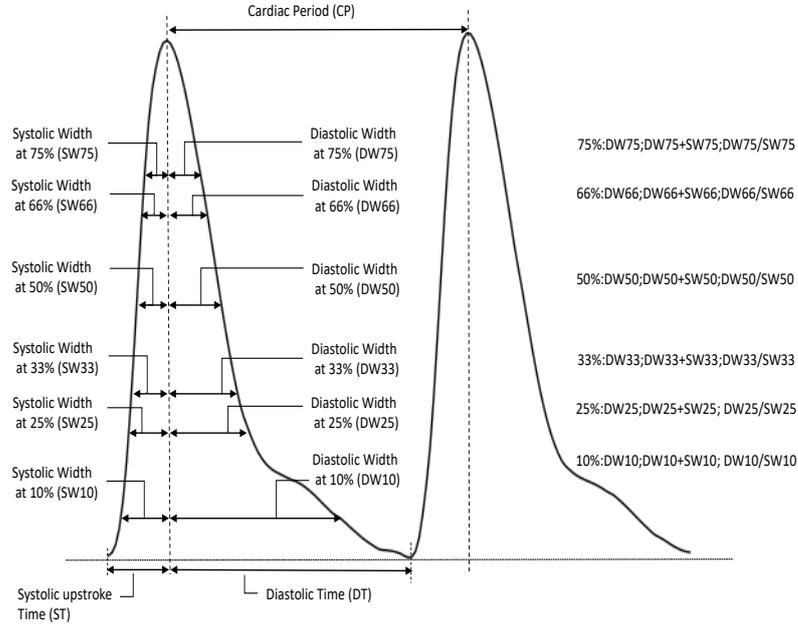
Figure 3.2: Potential 21 temporal PPG features for BP estimation.

In Kurylyak et al (2013) [25], another type of neural network was employed for estimating SBP and DBP using just PPG signals. More than 15000 heartbeats were analysed from PPG signals extracted from the MIMIC database [152]. This study improves on Teng and Zhang (2003) [24] by using 21 temporal features instead of four features extracted from a much bigger and more representative dataset (patients under treatments, drugs, elderly, etc). A feedforward neural network with 21 input vector was applied to estimate SBP and DBP, these features are shown in Figure 3.2. The SBP and DBP results from the neural network outperformed the linear regression method tested on the MIMIC dataset and satisfies the AAMI standards. The results of this study can be enhanced by adding information about the peripheral resistance, arterial stiffness, cardiac output and blood volume that strongly affect BP [123]. Additionally, similar to the previous method used in [135], the feedforward model is not suited for time domain tasks as it is not equipped with a feedback link and a memory or gating mechanisms to carry information from previous time steps for more accurate predictions. Recurrent neural networks (RNN) are built specifically to handle time domain data by providing a better control for

the flow of information much more efficiently for this task and hence is a more appropriate choice for reliable continuous BP monitoring.

Suzuki and Ryu (2014) [136] proposed a PPG feature selection method for estimating SBP. The data were acquired using a cuff-based BP device attached to the left upper arm and PPG recording acquired from the index finger of the right hand from 80 healthy subjects aged between 20 to 60 years old. Their method uses an orthogonal array and the signal-to-noise ratio (SNR) obtained from the Taguchi method for selecting PPG features that are robust against noise for multiple regression analysis. After calculating the SNR and an orthogonal array, seven features were selected -from the first and second order derivative of the normalised PPG signal - all with a positive SNR and influence for estimation of BP. It was found that the Taguchi method improved the effectiveness of the feature selection method for estimating SBP at the presence of large individual variability. However, the linear predictive model utilised here shares the same limitation mentioned before, which is not ideal for time series tasks for long term monitoring. Furthermore, this study only estimates SBP from data collected from healthy volunteers which is not optimal for detecting cardiovascular diseases. Therefore, this model is not clinically reliable for continuous BP measurements.

Choudhury et al (2014) [137] introduced a method that maps PPG features with intermediate patient specific latent parameters which are then used to derive the SBP and DBP values. For this study, PPG and BP signals were extracted from The University of Queensland Vital Signs Dataset [153]. The PPG signals were filtered using a band-pass filter with a cut-off frequency of 0.7 Hz and 3 Hz. Maximal Information Coefficient has been used as feature selection method to reduce the number of feature extracted from each PPG cycle. The final PPG features selected for the regression analysis were: systolic upstroke time, diastolic time, sum of systolic and diastolic width at 33% and at 75% amplitude. After feature extraction, outliers were removed from the dataset using a threshold-based approach. The Windkessel model was then applied to estimate total peripheral resistance and arterial compliance for individuals followed by a linear regression model for estimating BP values. This method provided non-invasive cuff-less BP values. The estimated error values were below 0.8 for both SBP and DBP, however, their model predictions were highly

variable. This caused the standard deviation to be high ±13.1 mmHg for SBP and ±10.23 mmHg for DBP. Moreover, this technique could not predict very high or very low BP values and was not validated on a multiday BP dataset for continuous monitoring.

Shen et al (2015) [126] proposed a stepwise regression BP model based on five features extracted from ECG and PPG signals for estimating SBP, DBP, and MAP. The PPG, ECG and continuous BP were collected from 10 healthy subjects. The ECG and PPG were collected using a multi-channel physiological instrument sampled at 1 kHz and the BP references were recorded using Finapres (Finapres Medical Systems B.V., Netherlands). All the recordings were done simultaneously for a period of 10 minutes. Given the high sampling frequency and the recording instrument utilised in the data collection, filtering was not necessary. The feature vector consisted of PTT values, systolic time, diastolic time, PPG area, and ECG time interval of a single cardiac cycle. The mean error and standard deviation were less than 6±6.5 mmHg for all the BP values, evaluated on a very small dataset of 10 healthy volunteers, and thus does not satisfy the AAMI standards. Also, stepwise regression is not the best option for sequential time domain data, especially when continuous BP monitoring is desired since it does not store knowledge learnt from previous cycles. Besides, it requires three different models for predicting SBP, DBP and MAP separately, meaning that it does not capture the strong correlation between them for increasing the BP prediction accuracy. The three objectives, SBP, DBP and MAP, can be estimated in parallel using more advanced models such as neural networks.

In 2015, Kachuee et al (2015) [113] proposed a calibration free BP estimation using the PTT approach. The data was acquired from the MIMIC online database. Signal pre-processing was applied to the PPG signals from which several whole base PPG features were extracted and combined with PTT parameters. Signal pre-processing included: simple averaging filter to smooth the signals, removing signal block with irregular and unacceptable BP and heartrate values, removing signal block exhibiting motion artefacts and calculating PPG signal autocorrelation. In total, ten features were used as input for a regularised linear regression model as well as two non-linear models, namely feedforward neural network and a support vector machine (SVM). Although these methods provide cuff-less, continuous and calibration free

BP measurements, it still bears many disadvantages mentioned earlier. It requires two sensors, and their neural network and SVM model do not explicitly model the temporal dependencies in the data resulting in long-term inaccuracy. Moreover, the results for SBP, DBP, and MAP did not meet the AAMI standards. Therefore, further improvement in terms of choices of models and features are needed for accurate BP estimation.

Datta et al (2016) [138], used a combination of machine learning and mathematical modelling for calculating SBP and DBP from PPG signals. The authors acquired their own data in an effort to include a wide range of BP values and proof the effectiveness of their noise cleaning techniques. PPG signals were measured from the right hand index finger using a fingertip pulse oximeter sampled at 60 Hz. BP signals on the other hand were recorded using a digital BP monitoring device directly after the PPG signal acquisition. Their proposed method introduces noise cleaning techniques to reduce the noise of the PPG signals. The following processing steps were applied to help reduce the noise in the PPG signals: mean subtraction normalising the PPG signals, band-pass filter to remove low and high frequency respiratory movement and jitters respectively, baseline correction to bring the end cycles to the same level, topline correction for removing random fluctuations in the signal amplitudes and finally cycle selection. After processing, the most relevant PPG features were selected, namely: age, weight, systolic upstroke time, systolic area, time between cycle onset to dicrotic notch, width at 50% and 75% amplitude. Subsequently, the latent parameters of the Windkessel were modified based on those PPG features. A linear regression model was applied for estimating the latent parameters. This study claims that the overall BP estimation error was within 10% of a commercially available digital BP monitoring device. Nonetheless, the relationship between some of the used features and the BP is not always linear, hence the results can be improved further with non-linear recurrent functions such as recurrent neural networks.

Sideris et al (2016) [154] introduced a cuff-less continuous BP measurement for remote health monitoring systems. The data was collected from ICU patients from the MIMIC database with PPG measured by pulse oximetry and referenced to invasive BP. A Long Short Term Memory (LSTM) network was applied on the PPG signals to estimate SBP, DBP, and mean arterial blood pressure. LSTM is the state of the

art recurrent neural network that takes into consideration pervious states or events in the prediction process, and therefore, leverage long-term pattern to deliver more accurate BP estimations. Unlike most studies that utilise domain knowledge for extracting features from each PPG cycle, in this study the input data for the LSTM network were overlapping PPG windows/frames. However, the evaluation metric used for the evaluation was root mean squared error and the model was only tested on 42 patients. As such, their model evaluation did not follow the conventional standard set by the AAMI or British Hypertension Society, hence the results are not comparable with other studies that mostly use mean absolute error. Moreover, the authors stated that the model optimisation i.e. number of cells, hidden layers, window size etc. was beyond the scope of their study.

Duan et al (2016) [139] proposed a feature exploration method for cuff-less BP estimation using just a PPG sensor. The University of Queensland Vital Signs Dataset was adopted for the evaluation. For improving the PPG signals quality over noise, a wavelet transform and average filtering were applied first to remove the noise. Subsequently, several analytical techniques were utilised, such as random error elimination, adaptive outlier removal and maximum information coefficient and Pearson's correlation for features selection. Three separate sets with eleven features, each was proposed to predict SBP, DBP and MAP out of fifty-seven possible feature candidates. A support vector machine was used as an estimator model. The results of this study satisfy the AAMI standards in terms of error rate, however, this dataset contains only 32 cases and SVM is not suitable for long term continuous BP measurements and hence the accuracy will decrease over time. As mentioned earlier, non-recurrent models cannot estimate BP efficiently since they do not provide feedback from previous events, as it is the case for regulating the arterial pressure in the human body that involves multiple feedback control loops. Therefore, the history of BP events affects future values.

Unlike the previous PPG time domain approaches, Xing and Sun (2016) [28] introduced a frequency domain methodology for extracting certain features from the PPG signals. Fast Fourier transformation (FFT) was applied on the PPG to extract fundamental features such as amplitude and shape information. FFT uses a small number of parameters to keep most of the information relating to the PPG wave-

form. This method applied a feedforward neural network to estimate the BP and was evaluated on 69 patients collected from the MIMIC database and 23 volunteers. Signals were first pre-processed as follows: PPG and BP signals were aligned to remove their phase lag. Only good quality signals were selected based on predefined criteria. PPG signals were then normalised and analysed in the frequency domain. This was followed by extraction of both amplitude and phase features from the waveform using FFT. The authors reported that this method performed well for BP estimation. However, they also suggested that FFT features are not sufficient markers for building a BP estimation model, and hence, a more efficient model is required to take into consideration the essential PPG waveform characteristics. Moreover, their feature extraction method has some limitations. When rapid changes occur in BP values, the features will be influenced by adjacent beats, which will lead to a decrease in the accuracy of BP estimation. Additionally, a feedforward neural network is not suitable for continuous long term BP monitoring as the estimation error will increase for longer estimation period.

Gaurav et al (2016) [140] used only PPG signals to estimate SBP and DBP. Their work combines PPG based features with Heart Rate Variability (HRV) related features in an effort to enhance the input feature vector for a more accurate BP estimation. The data was derived from the MIMIC online database from which 3000 PPG and BP signals were extracted. Signals were then pre-processed to remove inconsistent windows, and irregular BP and heartrate values. The BP and PPG were also aligned for feature extraction. PPG windows obtained from the previous step were normalised using min-max scaler. Afterwards, 8 PPG features were extracted from the magnitude and temporal information of each PPG window. Furthermore, 19 features were extracted from the filtered second derivate PPG signals. Additionally, 8 non-linear cardiac cycle time ratio based features were also extracted along with 11 HRV features from 10 consecutive peak intervals of the PPG. All these features combined together constitute the input vector for three feedforward neural networks for each systolic and diastolic BP. The mean error for the SBP and DBP reported were 0.16 ± 6.85 mmHg and 0.03 ± 4.72 mmHg respectively. Hence, as a result this method met the AAMI and presented significant improvement on previous methods published in the literature tested on large datasets. Nonetheless, this method is computationally expensive given the fact that 46 features were derived from PPG

and its second derivative along with HRV features which was then fed into 6 neural networks. Tuning model's parameter is time consuming, consequently, finding the best parameters and architectures for 6 models is very complex. Selecting the right model such as recurrent networks can further reduce the variability and enhance estimation precision.

Gao et al (2016) [141] developed a method for estimating SBP using only PPG signals. Their method uses a non-linear SVM with discrete wavelet transformation. It was found to be robust against small irregularities in the PPG waveform which enabled them to use PPG signals obtained from a pulse oximeter and phone. The PPG signals were collected using an Android application and Discrete Wavelet Transform was used for extracting temporal characteristics. The feature set includes: systolic upstroke time, diastolic time, age, gender along with thousands of features extracted from the obtained DWT coefficients. Afterwards, a forward feature selection technique was utilised to include only those features that have an effect on the BP estimation. Test results from PPGs obtained from both a pulse oximeter and a phone were within the limits imposed by the AAMI. The error estimation can be enhanced by estimating not only SBP but DBP as well, given the strong correlation between the two. This could be done by providing a feedback from the DBP model to the SBP model or simply using neural networks allowing for SBP and DBP estimating using one model. Furthermore, refining the feature set would also improve the prediction by taking into account information such as peripheral resistance and vessel elasticity from the first and second derivative of the PPG [123]. Also, this technique was evaluated on 65 subjects with no history of cardiovascular disease, hence, a larger and more diverse dataset will enhance model generalisation for early detection of cardiovascular diseases.

Liu et al (2017) [27] proposed a cuff-less BP measurement based on PPG and its second derivative. This work attempts to enhance the SBP and DBP prediction by combining the 21 features used in Kurylyak et al (2013) [25] along with 14 features from the second derivative of the PPG (SDPPG), shown in Figure 3.3. SDPPG contain information about the aortic compliance and stiffness which is highly related to BP. A support vector machine was applied as a BP estimator. This study reported a 40% accuracy improvement in BP estimation when taking the SDPPG features into
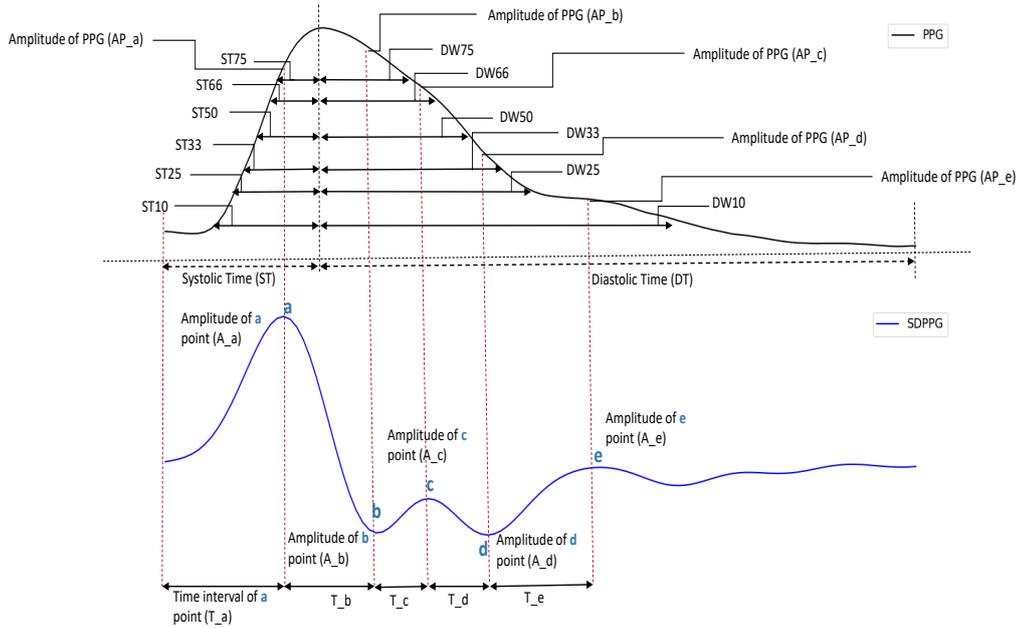
Figure 3.3: Temporal PPG features and potential Second Derivative PPG (SDPPG) features.

account -using 35 features- as opposed to 21 features and a neural network applied in Kurylyak et al. Experimental results (ME±SD) based on the 35 features using an SVM were 8.54±10.9 mmHg for SBP and 4.34±5.8 mmHg for DBP compared to 13.4±11.6 mmHg and 6.9±5.9 mmHg for SBP and DBP respectively, evaluated on a neural network. The major challenge in this technique is extracting the correct SDPPG features values which are entirely dependent on the visibility of the five peaks a,b,c,d and e as shown in Figure 3.3. These five peaks, which mark the 'W' shape in the SDPPG signal, were not visible for all patients, and hence, the models were evaluated on a very small dataset consisting of 910 PPG cycles from the MIMIC II dataset. The overall performance of the model on this dataset is poor in terms of accuracy and the study did not satisfy the AAMI requirements for a reliable cuff-less BP estimation.

In another publication, Kachuee et al (2017) [114] proposed methods to estimate BP using PAT for continuous, cuff-less and calibration-free estimation for SBP, DBP and MAP. PPG, ECG and BP signals were collected from the MIMIC database. The ECG and PPG signal processing consisted of motion artefacts removal and de-noising using discrete wavelet composition. In this study, the proposed method was

based mainly on PAT features along with other whole-based PPG features. Two types of features were extracted, namely, physiological parameters (e.g., heartrate, arterial stiffness index, augmentation index) and other features that describe the shape of the PPG waveform. For reducing the effect of collinearity between the features and reducing the dimensionality of 190 extracted features, PCA has been utilised to reduce the dimension down to 15 new features while preserving 98% variance in the data. Several traditional machine learning techniques, such as linear regression, SVM, Random Forest, and Adaboost, have been used to achieve better accuracy model. The results show that there exists a considerable non-linearity in this task which can be inferred from the superior performance displayed by the non-linear functions such as SVM and ensemble learning methods. Their techniques met the standards of the AAMI for DBP and MAP. However, it suffered from several disadvantages, such as the need for two sensors, the SBP prediction did not satisfy the AAMI standards and their DBP and MAP accuracy were just on the acceptable limit of the AAMI standards, which will eventually deteriorate for longer continuous BP monitoring using the applied models.

Miao et al (2017) [127] proposed a beat-to-beat BP estimation method using a combination of data mining techniques and a mechanism-driven model. For this study, the data was collected from 73 healthy subjects for a static BP estimation experiment, 35 healthy subjects for dynamic BP estimation where the subjects had to exercise for 5 minutes and 10 subjects for a follow up to test the robustness of their models. The PPG signals were collected from the left index finger and the BP reference was collected using a Finapres machine. Fourteen features were first extracted from (the first and second derivative) PPG and ECG signals followed by a genetic algorithm-based feature selection method for selecting the most influential features for BP estimation for each subject. As a result, features with the highest effect on SBP and DBP estimation were selected. A multivariate linear regression (MLR) and support vector regression were established to evaluate the effectiveness of the genetic algorithm. A reported 2 mmHg reduction in standard deviation for different calibration time intervals, compared to PTT-based model, was achieved by their approach. The results from their experiments show that the SVR outperformed the MLR model signifying the non-linear relationship between the features and BP. Furthermore, it was found that the error of the proposed models deterio-

rates significantly one day after the model construction (with initial calibration) but stabilises afterwards for longer calibration periods, as opposed to the PTT-based approach where the error continuous to increase 1 to 3 days after the initial calibration. Moreover, their proposed approach still requires two sensors, and their models are not suitable for handling time series data for continuous BP estimation.

In 2018, Wang et al [128] proposed a BP estimation based on features extracted from PPG signal using the multitaper method (MTM). PPG and ECG signals were acquired from the MIMIC database and a total of 58795 valid PPG intervals were extracted from 72 subjects. MTM was employed to extract the spectral components that were combined with two features from the PPG waveform to constitute the input feature vector for a feedforward neural network. The advantage of this approach is that it can provide cuff-less continuous BP measurements with acceptable results using one sensor, mean error of 4.02±2.79 mmHg for SBP and 2.27±1.82 mmHg for DBP. This level of accuracy is achieved while the feedforward neural network does not incorporate temporal dependencies in its estimation which prevents long term accurate predictions. The incorporation of such features, using recurrent neural networks may further improve the accuracy of predictions.

Shimazaki et al (2018) [129], introduced a BP estimation model from PPG features extracted using an autoencoder. For this study, the data was collected from 687 healthy male subjects and 676 healthy female subjects. The PPG signals were measured from the left index finger, while the BP reference was measured every minute using a sphygmomanometer with a cuff attached to the right upper arm. An autoencoder was applied as an alternative for conventional feature engineering/extraction. An autoencoder is a neural network algorithm that can reconstruct a better version of its input vector by extracting complex features and adding new ones. The resulting features obtained from the autoencoder were then passed into a feedforward neural network for estimating SBP. It was found that the features learnt by the autoencoder are effective for BP estimation and the non-linear learner outperformed the linear regression model. However, the standard deviation of the error for this method was 11.86 mmHg, exceeding the ±8 mmHg standard deviation set by the AAMI, and therefore the results were highly variable and were not deemed reliable for clinical BP measurements. This suggests that the new constructed fea-

tures obtained from the autoencoder are not optimal for BP estimation, since the autoencoder alters the original PPG waveform which might have caused the high variability in the BP estimation.

In a comparison study between different machine learning approaches, khalid et al (2018) [26] extracted three features from the PPG waveform. The University of Queensland vital sign dataset (contains 32 cases) has been acquired for evaluating their proposed technique. The PPG signals were filtered using Savitzky-Golay filter and the baseline wandering was also removed. This was followed by a 2-dimentional normalisation for both amplitude and width. The PPG signals were also segmented into 5 seconds frame and manual check was performed for removing bad quality segments or segments with no reference to BP values. Pulse area, pulse rising time and width at 25% were used as features for three traditional machine learning models. Regression tree, multiple linear regression and support vector machine were established for prediction of SBP and DBP. The models were analysed for three BP categories: normotensive, hypertensive and hypotensive. The decision tree outperformed both SVR and the linear regression models for both SBP and DBP. The mean error difference only for the regression tree for normotensive people were within the AAMI standard and the rest of the models had standard deviation above 8 mmHg. Additionally, since only intermittent non-invasive BP references were available for this study, the BP estimation was implemented on the basis of each PPG segments. Hence, this method does not allow continuous beat-to-beat BP estimation.

In 2018, Dey et al [142], developed an ensemble of BP estimation models based on demographical and physiological features. A unique set of PPG features were also incorporated in the models for estimating SBP and DBP using lasso regression model. The authors collected their own PPG signals from 205 volunteers of diverse demographical and physiological profiles. PPG signals were recorded for 15 minutes using a phone Heart Rate sensor sampled at 125 Hz. Each PPG waveform was segmented into 15 seconds window where each window is interpolated to a fixed length (using cubic spline interpolation) and normalised using min-max normalisation before the correlation coefficients were determined. BP values were collected before and after each PPG recording using a mercury cuff-based device. A total of 233 time and frequency domain features were extracted from a single heartbeat

(one PPG pulse). The first four derivatives of the PPG signals were considered for feature extraction. Features from each individual cycle were averaged over the 15 seconds window. Other features, demographic and physiological information (age, height, weight, gender) were also used as independent features. The Lasso regression model was applied to first estimate the DBP value which were then used as input along with other features for estimating SBP values. In an effort to take full advantage of the demographic and physiological feature, the dataset was partitioned separately based on age (young < 40 and old ≥ 40), gender (female and male), and BMI (underweight BMI $< 24\text{Kg/m}^2$ and overweight BMI $\geq 24\text{Kg/m}^2$). Afterwards, the regression models were applied on each individual partition. The results show that the addition of multiple independent partitions on the basis of demographic and physiological features can further improve the BP estimation values. The systolic and diastolic values of the combined model were 6.9±9 mmHg and 5.0±6.1 mmHg, respectively. This study demonstrates that it is possible to utilise PPG signals collected from a phone for the estimation of BP values and incorporating demographic and physiological information can further enhance estimation accuracy. However, in terms of model precision, the results did not satisfy the AAMI standards for SBP and further improvements are needed such as the use of more advanced models and particularly recurrent neural network which are capable of processing more data, specifically time domain data. Additionally, evident from the results, the feature extraction process is somewhat overcomplicating the task. Optimising the input feature set is crucial for improving model performance in terms of accuracy and time complexity.

Tanveer and Hasan (2018) [30] proposed a two hierarchy levels model to estimate BP using ECG and PPG signals. The lower level is an artificial neural network (ANN) used to extract morphological features. The ANN is connected with an upper level which consists of two stacked long short term memory (LSTM) layers to take into consideration the temporal variation for the features extracted by the ANN in the lower level. This method, similar to the one proposed by Shimazaki et al (2018) [129], uses a neural network to extract features instead of using traditional feature engineering techniques used by most researchers. This paper argues that it is hard to obtain correct features from the ECG and PPG signals since the waveform contour changes from one subject to another, and hence the position of these features

varies or maybe not be visible for all patients. Therefore, it is not certain that all these features can be extracted from all patients for a complete and reliable dataset. To overcome this challenge, the author applied an ANN on a small set of PPG and ECG signals that were collected from 39 subjects acquired from the MIMIC database. Both PPG and ECG signals were pre-processed for removing the baseline wandering and high frequency noise. This was achieved by a bandpass filter using Tunable-Q wavelet transform. Both signals were segmented into a fixed length of three consecutive peaks to avoid varying number of cycles per fixed number of seconds between subjects. This was followed by normalisation and resampling for both signals to a length of 256 samples per segment. The concatenated PPG and ECG segments constitute the input feature vector for the ANN model. The results suggest that, compared to traditional feature engineering-based model, this automatic feature extraction technique combined with an LSTM model provides a much better accuracy. The SBP and DBP mean absolute error values were 1.1 mmHg and 0.85 mmHg respectively, however, the AAMI requires at least 85 patients. Moreover, this method requires two signals sampled at two different sampling rates and involves fine tuning two models which can be difficult, time consuming and varies depending of the data size.

In Su et al (2018) [29], a four-layer LSTM network was employed to estimate SBP and DBP from ECG and PPG signals. This method built an LSTM model with (1) a bidirectional structure to access larger scale context information of input sequence and (2) residual connections to allow the gradient in the LSTM network to propagate efficiently. The ECG and PPG signals were recorded with a Biopac system while the BP signals were recorded with a Finapres at the same time. All these signals were sampled at 1000 Hz for a period of 10 minutes from 84 healthy subjects at rest position. Another dataset was also collected from 12 subjects for a multi-day continuous BP consisting of 8 minutes recording for each signal. Since the main focus of the paper was to demonstrate the importance of modelling the time variation of the input features, the authors simply selected seven features from the PPG and ECG signals, such as PTT values, Heart Rate, systolic upstroke time, etc. It was reported that the results of this method outperformed all previous models with significant improvement for multi-day BP datasets and root mean squared error of 3.9 mmHg and 2.66 mmHg for SBP and DBP respectively, on the static 84 subject BP dataset.

Therefore, suggesting that modelling the temporal dependencies leads to a much accurate prediction for long-term BP measurements compared to classical models. Optimising the feature set, which was not the focus of this paper, would increase the precision and performance of the model. However, it is difficult to compare the results with other methods since this study did not follow the conventional metric, mean absolute error and standard deviation set by the AAMI.

Fujita et al (2019) [143] proposed a cuff-less SBP estimation method using partial least-square (PLS) regression. Their multivariate estimation method used Level Crossing Features (LCF) extracted from the contour lines randomly drawn on the PPG's second derivative. The authors collected their own signals from 265 subjects with SBP 133.1±18.4 mmHg and aged 62.8±16.8 years participated in the study. The SBP reference values were acquired from the left upper arm using an automatic BP monitor at rest position and the PPG signals were recorded immediately after the BP recording for a period of 20 seconds. The PPG signals were pre-processed by a first order low-pass filter and a finite impulse response filter to remove high frequency noise. The LCF features were extracted from the second order derivative PPG of which two types of features were obtained, namely, the number of crossing and the length of the curve line. This paper attempted to simplify the SBP estimation by using a very small set of input data comprising six LCF features evaluated on PLS regression. Only 38% of the subjects had their SBP estimation below 5 mmHg. Consequently, this method received grade D using the British Hypertension Society BP metric, suggesting that this technique is not fit for clinical trial. Additionally, it also shows that PLS is not appropriate for long-term BP measurements and the LCF features are not ideal for BP estimation. Besides, their dataset comprised only healthy subjects in a resting position, hence the dataset does not contain diverse BP values that represent the population.

Chen et al (2019) [130] proposed SBP and DBP estimation models based on PTT approach in addition to PPG waveform characteristics. The impact value for each feature was investigated and a genetic algorithm was also used for fine tuning model parameters. SVM and multivariate linear regression models were established to predict BP values evaluated on the MIMIC dataset. A total of 772 sets of waveforms were extracted from the MIMIC database containing ECG, PPG and BP

signals. The PPG and ECG signals were first cleaned from motion artefacts, irregular segments and missing waveforms. Furthermore, both ECG and PPG signals were denoised using wavelet threshold denoising method and cubic spline interpolation respectively. Fourteen features were extracted such as PTT, heart rate, and other features describing the shape of the PPG waveform. All these features were normalised using min-max scaler and the importance of each feature was investigated using mean impact value (MIV) for removing redundant features and reducing the input dimension. The results from their proposed SMV method were 3.27±5.52 mmHg for SBP and 1.16±1.97 for DBP, hence it satisfies the AAMI requirement for non-invasive cuff-less BP estimation. The results can be further improved by taking into account demographical features such as age, gender, weight etc. Additionally, evaluation for long term BP prediction i.e. one week, one month or six months should be conducted to test the model's performance for long term measurement. The experimental results from [127] shows that the performance of PTT based models decrease for long term monitoring due to the expiration of the PTT parameter and the inability of the SVM to perform well for multiday estimation.

In a study published by Hasanzadeh et al (2019) [144], four different classical machine learning models were tested on the MIMIC II dataset for estimating SBP, DBP and MAP using only PPG signals. The PPG signals extracted from the MIMIC dataset were denoised and the baseline wandering was removed using a forward-backward Type I Chebyshev low-pass filter. Additionally, in order to improve signal quality for feature extraction, the PPG signals were upsampled to 500 Hz. In an effort to further improve the feature extraction process, this study takes advantage of the first and second derivatives for improving the detection of key points in the PPG waveform, such as dicrotic notch, diastolic peak, slope, and inflection point. Features used for estimating BP includes, heart rate, area related features, modified normalised pulse volume, amplitude features, heart rate variability properties, and time related features. The extracted features were evaluated using linear regression, decision tree, random forest and AdaBoost. The best results were obtained using the AdaBoost model. The reported MAE±SD were 8.22±10.38 mmHg for SBP, 4.17±4.22 mmHg for DBP and 4.58±5.53 mmHg for MAP. The results for DBP and MAP are reasonable, however, the error was considerably higher for SBP. Furthermore, the reported correlation coefficient between the estimated and reference

values were 0.78, 0.72 and 0.75 for SBP, DBP and MAP, respectively. This does not show a significant positive correlation in the BP estimation against the ground truth values. These results can be further improved using a suitable model for handling sequential data.

A more advanced model was employed in Slapnicar et al (2019) [145] for estimating SBP and DBP using the PPG approach. In this study, the PPG and BP signals were extracted from 510 subjects from the MIMIC III database. Several steps were performed in order to clean and preprocess the data, including, removing empty or less than 10 minutes samples, normalising the PPG to zero mean and units variance, high and low frequency noise removal using 4th order bandpass filter and finally Hampel filter for eliminating outliers. After data cleaning, 700 hours of signals were obtained and segmented into 5 s segments for further analysis. PPG segments along with their first and second derivatives were used as input for the spectro-temporal ResNet proposed by the authors. The proposed model is highly complex, in particular, it consists of three networks for processing each signal on its own (PPG and two derivatives). Each network contains a stack of five ResNet blocks and a separate spectro-temporal block (comprising spectrogram and GRU layers). The outputs of the ResNet block of all the networks are concatenated together and fed into a GRU layer followed by a batch normalisation layer, while the output of the three separate spectro-temporal blocks are concatenated and fed into a batch normalisation layer. Afterwards, the outputs were concatenated into a single layer, out of which the SBP and DBP values are calculated. The results were evaluated using only the MAE. The reported error for SBP was 15.41 mmHg and 12.38 mmHg for DBP while the standard deviation was not reported. Hence, considering the complexity of the model and results obtained, it can be said that the performance was very poor and the proposed model over complicated the task.

In Ripoll and Vellido (2019) [131], a Restricted Boltzmann Machine (RBM) was established as a proof of concept for estimating SBP and DBP values. The RBM-BP estimation model was based on the PTT approach. The data used from 250 patients were collected from the MIMIC database. All PPG and ECG signals were segmented into a 5 seconds window. Motion artefacts and noisy waveforms were also removed from the dataset. Three different RBMs were established for estimating

SBP and DBP using one input feature each, namely, PTT, 1/PTT and log (PTT). The performance of this method decreased as the measurement parts from the calibration point. The overall results from this experiment were acceptable and received grades A and B according to the British Hypertension Society (BHS) metric. However, there are several limitations to this method, such as the need for two sensors, the accuracy of the model decreases after 6 minutes of the initial calibration and therefore, as a result it necessitates calibration. Also, the RBM is not capable for estimating continuous BP since it will suffer from vanishing or exploding gradient in long term continuous BP prediction. Consequently, utilising a recurrent model should further increase the accuracy for longer tracking capabilities for BP values. Estimation accuracy can be further increased by including demographical features and whole-base PPG features.

In Wang et al (2020) [4], an end-to-end BP estimation model was established using a neural network consisting of several depth-separable convolutional layers and a GRU layer. The convolutional layers were used to automatically extract features from the PPG waveform, instead of the more conventional feature engineering approach. The output of the convolutional layers were then fed into a GRU layer for modelling time related information. This model was evaluated on PPG and BP signals extracted from the MIMIC database. In order to make the data ready for analysis, first, the authors ignored segments where either the PPG or BP signal was missing. Second, the data segments were divided into single cycles and each PPG cycle was normalised to fixed length of 100 data points. The normalised single cycle PPGs formed the input vector for the proposed model, as shown in Figure 3.4. The reference SBP and DBP values correspond to the highest and lowest (end diastole) values of the BP signal, respectively. For evaluation, the model was trained and tested on 20 subjects individually. Meaning that the model was evaluated on one subject at a time, for 20 individuals. In particular, for every experiment, the signals of one subject were divided into 75% for training, 20% for validation and 15% for testing. The performance was judged using mean absolute error and standard deviation. The reported average error of the 20 experiments were 3.81±4.28 mmHg for SBP and 1.99±2.57 mmHg for DBP. Although the results obtained are reasonable, there are several drawbacks to this study. First, the dataset belongs to 20 subjects only, hence, the authors did not adhere to the AAMI requirements. Second, in ev-

ery experiment, the test set belonged to the same subject it was trained on. Third, the segments were divided into single cycles, hence, the temporal variations between consecutive cycles were not captured using the GRU layer. The recommended length for input vector is 5 to 10 s in order to take full advantage of neural network models.
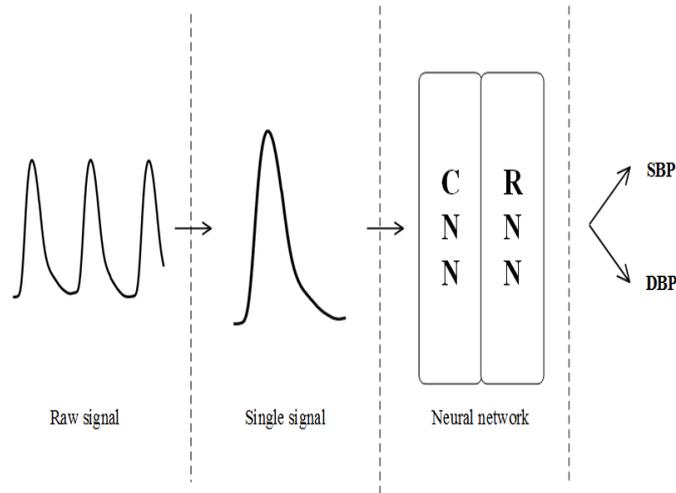
Figure 3.4: Proposed method by Wang et al (2020) [4], for estimating SBP and DBP using a combination of CNN and GRU from single PPG cycles

An end-to-end neural network model for continuous BP estimation was introduced by Eom et al (2020) [132]. The deep learning model consists of a combination of convolutional neural network and GRU. The authors used the raw PPG, ECG as well as ballistocardiograms (BCG) signals as input for the deep learning model. The signals were acquired from a total of fifteen subjects recruited for this study. All signals including the BP were recorded simultaneously for a period of 30 minutes while the subjects were sitting in an armchair. In terms of data preprocessing, a second order Butterworth bandpass filter was applied for noise and baseline wandering removal. Afterwards, all signals were resampled to 125 Hz and segmented into 5 s frames. The three raw signals formed the input for a model consisting of ten layers CNN, followed by a bidirectional GRU layer, attention layer and a fully connected feedforward layer. The performance of the model was assessed using the MAE±SD, R-squared and AAMI standards. The reported results were 4.06±4.04 mmHg for SBP and 3.33±3.42 mmHg for DBP. While the R-squared (measured of explained variance) were 0.52 and 0.49 for SBP and DBP, respectively. This means

that the features automatically extracted by the CNN layers could not effectively explain the changes in the target SBP and DBP outputs. Additionally, the results obtained were within the acceptable AAMI standards, however, they did not satisfy the required number of subjects set by the AAMI. Although estimating BP using raw signals is easier compared to the feature engineering approach, it generally adds more complexity to the model (more data points to process) and the results have so far not been superior compared to the domain knowledge feature-based approach.

Several classical machine learning models were evaluated, by Chowdhury et al (2020) [146], for estimating BP using only PPG signals along with demographical features. The PPG signals were collected from 219 subjects derived from an open source dataset [155]. A total of 657 PPG recording, sampled at 1000 Hz, were extracted and each PPG record was 2.1 s long. The dataset also contained valuable information, such as, age, gender, height and weight. After data preprocessing, poor quality PPG recordings were ignored, and only 222 signals extracted from 126 subjects were used. The data cleaning and preprocessing included, removing bad quality segments, z-score normalisation, low-pass Butterworth filter and baseline removal. Afterwards, a total of 107 features were extracted from the PPG and its derivatives, such as, width related features, demographic features, time and frequency domain features as well as statistical features. However, after feature elimination, roughly a dozen features remained and were used for further analysis. The authors used three different feature elimination techniques for reducing redundancy and avoid overfitting the model. These techniques are namely, correlation-based feature selection, relief feature selection and minimum redundancy maximum relevance algorithms. The resulting features were evaluated using Gaussian process regression and ensemble trees regression for estimating SBP and DBP values. The best performance was obtained using the Gaussian process regression, where the ME±SD for SBP was 3.02±9.29 mmHg and 1.74±5.54 for DBP. Although the reported results satisfy the AAMI standards, the dataset used for evaluation was considerably small. In particular, the test set contained roughly 20 data points for SBP and DBP. Additionally, the classical models employed in this study are not fit for long term measurement, hence, their performance will diminish over time. Furthermore, two different models were established for estimating SBP and DBP, which is not practical nor efficient since there is a strong correlation between SBP and DBP. This issue can be resolved

using recurrent neural networks.

In a published study by Sadrawi et al (2020) [147], the authors attempted to estimate the ABP waveform using deep convolutional autoencoder (DCAE) utilising only PPG signals. Two variants of the convolutional neural network were tested, namely, LeNet-5 and U-Net, abbreviated LDCAE and UDCAE, respectively. These models were also compared to a genetic based deep-convolutional autoencoder (GD-CAE). All the aforementioned models were evaluated on signals collected during surgical operation from 18 subjects. The PPG and ABP signals were recorded using MP60 IntelliVue Patient Monitor and sampled at 128 Hz. All PPG and ABP signals were divided into 5-s frames. The raw PPG segments formed the input feature vectors for the neural network models. In terms of performance evaluation, the reported MAE show that the GDCAE outperformed both the LDCAE and UDCAE. Additionally, the study reported high correlation between the estimated and target SBP and DBP values. While the model demonstrated reasonable MAE with 2.54 mmHg and 1.48 mmHg for SBP and DBP, respectively, it was only evaluated in a very small number of patients. Overall, the results of this study satisfied the AAMI standards in terms of acceptable error, however, it did not meet the required number of subjects.

Another end-to-end model was proposed by Ibtehaz and Rahman (2020) [148] for estimating ABP waveform using only PPG signals. The established model consists of two neural network models, one was used as an initial ABP waveform approximation network, while the second was used as a refinement network for improving the approximated ABP waveform. In particular, the first model was a U-net network composed of a symmetric Encoder-Decoder network constructed using a CNN layers. The output ABP waveform of this model was fed into a MultiResNet neural network model for further improving the quality of the initial estimation. The MultiResNet was essentially an enhanced version of the approximation network also constructed using CNN layers. The proposed network was evaluated on 942 subjects extracted from the MIMIC II database. In terms of data processing, BP signals with very high and very low values were ignored and no preprocessing was done on the PPG signals. The PPG and BP signals were then segmented into 10 s frames for analysis, hence, the length of input vectors were 10 s raw PPG segments. The performance

of the model was judged using the MAE and the AAMI standards. The reported MAE was 3.45 mmHg and 5.73 mmHg for DBP and SBP, respectively. In terms of AAMI standard, the ME±SD were -1.582±10.688 mmHg for SBP and 1.619±6.859 mmHg for DBP. Hence, only the DBP estimation error was within the acceptable AAMI standard range. Furthermore, in terms of complexity of the model, both the U-Net and MultiResNet are very complex and takes a lot of time to train, especially on the 942 subjects dataset. This requires a lot of processing power, which makes it almost impossible to deploy the proposed model on wearable devices. However, the overall performance on this large dataset is impressive and promising towards cuffless and continuous BP estimation.

In Li et al (2020) [31], a deep learning BP estimation model was proposed using features extracted from the ECG and PPG signals. The proposed model consists of a one bidirectional LSTM layer followed by a stack of four LSTM layers with residual connections. The PPG, ECG and reference BP signals were extracted from the MIMIC II dataset. The PPG and ECG signals were preprocessed using Fast Fourier Tranform for removing baseline wandering and low frequency noise. Additionally, noisy and corrupt segments were eliminated. Afterwards, the PPG and ECG amplitudes were normalised using min-max normalisation method. Also, very high and very low BP values were removed (e.g. $180 \leq SBP \leq 80$ and $130 \leq DBP \leq 60$). After preprocessing, 678202 single cycles remained and used for further analysis. The authors extracted seven features from the PPG and ECG signals, namely, PTT, heart rate, reflection index, systolic and diastolic volume, time difference between the start of the cycle and peak as well as time difference between the start and dicrotic notch. The performance of the model was evaluated using MAE±SD and the AAMI metric. The reported MAE±SD for SBP was 6.726±14.505 mmHg and 2.516±6.442 mmHg for DBP. Additionally, the ME±SD were 4.638±14.505 mmHg and 3.155±6.442 for SBP and DBP, respectively. Thus, the DBP estimation was acceptable by the AAMI standards, while the SD of error for the SBP estimation failed to meet the AAMI requirement. However, the ME reported in this study was relatively higher than the ME reported in the previously mentioned studies. Also, the reported SD for both ME and MAE were exactly the same, which suggests that one of the reported metrics was unreliable. The results can be further enhanced by including more features from the PPG as well as demographical features.

In a recently published study by Lee et al (2020) [133], features extracted from three physiological signals were evaluated on a deep learning model for estimating cuffless and continuous SBP and DBP. The PPG, ECG and BCG signals were acquired from 18 volunteers. All signals were recorded simultaneous for 30 minutes and sampled at 1000 Hz. In terms of signal processing, a 2nd order Butterworth filter was applied removing noise and baseline wandering. The signals were then segmented into frames of 10 cardiac cycles each. Seven features were extracted including PTT, ECG R-R peak interval, interval between ECG to BCG peak, interval between the BCG notch to first derivative PPG peak, and amplitudes for ECG, BCG and first derivative PPG. The proposed model consisted of two bidirectional LSTM layers and two fully connected feedforward layers. The first fully connected layer was used to include personal information, such as body mass, weight, gender, age and height. The performance of the model was evaluated using MAE as well as the ME±SD of the AAMI standard. The reported MAE for SBP and DBP were 5.82 mmHg and 5.24 mmHg, respectively. In terms of evaluation against the AAMI standard, the results achieved by this model were -0.07±7.3 mmHg and -0.17±6.4 for SBP and DBP, respectively. Hence, in terms of error, the results obtained using this model were within the acceptable AAMI range, however, it was evaluated on only 18 subjects. Another drawback is that three signals were required for BP estimation using the proposed model. This is not straightforward since all signals must be synchronised and three sensors are attached to the subject's body, making this approach uncomfortable and impracticable for long term BP measurement.

In a study published by Schrumpf et al (2021) [149], three different neural network models have been adopted from the literature and evaluated on PPG signals derived from pulse oximeter as well as camera. The pulse oximeter PPG signals and their corresponding reference BP were acquired from the MIMIC III database, while the camera based remote PPG (rPPG) were derived from videos of patients' face and upper body, collected by the authors using a USB camera. The 4000 records of PPG signals obtained from the MIMIC dataset were filtered using 4th order Butterworth bandpass filter. Also, PPG frames with signal-to-noise (SNR) ratio less than 7 dB were eliminated. Lastly, the PPG segments were normalised. As for the rPPG signals, the authors used plane-orthogonal-to-skin to derive the PPG signals from the

pixels of the camera recordings. After manually examining signals obtained from 50 subjects, only 14 were deemed valid signals for further analysis. Additionally, rPPG signals with less than 7 dB SNR were dropped and the signals were then segmented into 7 heartbeat frames. The authors first evaluated AlexNet (consists of CNN layers), ResNet and the model implemented in [145] on the MIMIC dataset. Afterwards, the authors took advantage of transfer learning to fine tune only the last layer of the pretrained models using the rPPG signals. Additionally, this study investigated the effect to calibration on the BP estimation. The performance of the models were evaluated using the MAE±SD. The best performance was achieved using a calibration-based ResNet, and the reported SBP error were 12.51±12.61 mmHg and 8.3±9.84 mmHg for DBP. The main findings of this study are: raw derivative signals and window length do not impact or enhance the BP estimation precision, while calibration can further improve the prediction. Overall, the performance of the models were poor and unreliable for BP estimation. This suggests that estimating BP using PPG derived from a camera is less feasible compared to PPG collected using pulse oximeter. Furthermore, none of the results obtained using all the evaluated models, with and without calibration, were able to meet the AAMI requirements. Lastly, it should be noted that calibration is not acceptable in routine clinical practice.

In another attempt at providing cuffless BP estimation using only raw PPG signals, Aguirre et al (2021) [150] established a sequence-to-sequence model with attention mechanism. The PPG and their corresponding reference BP signals along with the age and gender were derived from 1100 subjects extracted from the MIMIC III database. The length of the PPG and BP were at least 15 minutes long. Additionally, poor quality PPG segments were excluded from further analysis. Afterwards, PPG segments were filtered using a bandpass Butterworth filter. The results of pre-processing was a dataset of 6478 segments of 13 s each. However, the input feature vector was 5 s frames of PPG and its first derivative signal. The proposed model was an encoder-decoder RNN network with GRU units. The encoder consists of three bidirectional GRU layers. From this part, two outputs were obtained,the first was the output of the whole network which is fed into an attention layer, and the second output was the last hidden layer states, which formed the input for the decoder network. The second part of the network was a decoder consisting of three GRU

layers. Its output was sent to the attention layer and at the same time to a fully connected feedforward layer (L1). From the attention layer, a context vector was calculated and concatenated with the output of the decoder network. The concatenated output goes to another fully connected feedfoward layer (L2). Finally, L1 and L2 were concatenated together along with the age and gender information. The final output of the entire network was the ABP waveform. This study evaluated both calibration based and calibration free approaches for BP estimation. The output of the proposed model was assessed using MAE±SD for both SBP and DBP. The best performance was achieved using the calibration based approach. The reported results for SBP was 12.08±15.67 mmHg and 5.56±7.32 mmHg for DBP. There are several drawbacks associated with this study. Firstly, the performance of the model was very poor as depicted by the reported error, even when using the calibration based approach, which as mentioned previously prevents this approach from being deployed in clinical practice. Secondly, the model was evaluated on a very small subset of the MIMIC III (only 6478 segments). Thirdly, the results were not compared against the AAMI standard.

Another cuffless continuous BP estimation model, using only raw PPG signals, was proposed by Harfiya et al (2021) [151]. The PPG and BP signals were extracted from the MIMIC II database. The authors extracted 12000 records, belonging to 942 subjects, out of which only 5289 records were valid segments. In order to make the data ready for analysis, the PPG signals were filtered using a bandpass filter, PPG segments with undetected systolic peak were removed, and finally, the PPG segments were normalised using z-score normalisation. Additionally, BP segments with high and low BP values were removed (e.g. $180 \leq SBP \leq 80$ and $130 \leq DBP \leq 60$). In order to estimate the BP waveform, the authors introduced a LSTM-based autoencoder. However, the traditional autoencoder was replaced by LSTM layers in order to allow the network to learn time domain representations from the sequential data. The output of the first encoding layer was a reconstruction of the PPG signals. The second part of the network was a decoder comprised also of LSTM units, and the output of this layer was a sequential BP waveform. The feature vector for this model was a stack of PPGs along with first and second derivative signals. The reported MAE±SD for SBP was 4.05±4.42 mmHg and 4.6±3.47 mmHg for DBP. Additionally, the results for both SBP and DBP satisfied the AAMI stan-

dards. Thus, considering the size of the dataset and the model employed, it can be said that the results achieved in this study are acceptable and potentially promising towards cuffless and continuous long term BP monitoring.
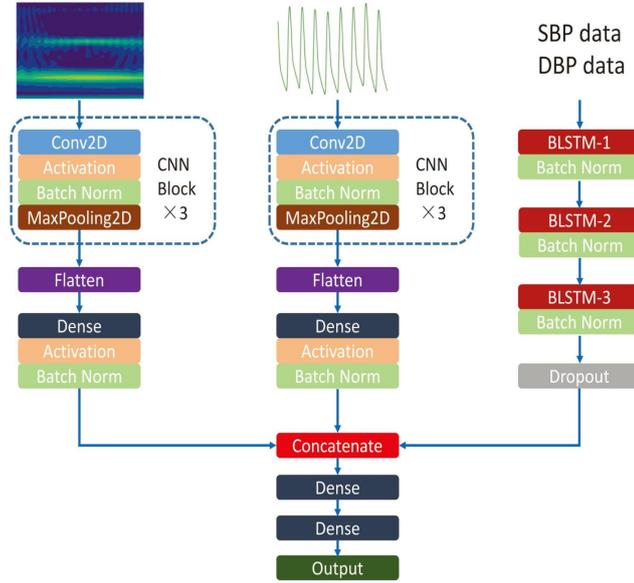
Figure 3.5: Model architecture proposed by Rong and Li et (2021) [5] for estimating BP.

In a more recently published study, Rong and Li et (2021) [5] introduced a multi-type feature fusion neural network for estimating BP using the PPG approach. Three types of features form three input vectors for training three neural network models, one model for each feature type. Morphological features and frequency spectrum features were used for training two CNN models separately, while temporal PPG features were used for training a one bidirectional LSTM layer. The CNN models, both consisted of three two-dimensional CNN blocks, and were trained to automatically extract morphological and frequency spectrum features from images, and the temporal input features were automatically extracted from the raw PPG signal by a three bidirectional LSTM layers. The output of each model was then concatenated together in order to estimate the SBP and DBP values. The model architecture is presented in Figure 3.5. The PPG and BP signals used in this study were derived from 1000 subjects extracted from the MIMIC II dataset. The PPG signals were filtered using a Butterworth bandpass filter and segmented into 680 data points each. A total of 11546 segments were selected for analysis. Afterwards, the PPG segments were converted into images, forming the morphological input features for the first CNN network. The input for the second CNN network was generated by a

continuous wavelet transform and saved as images. The third input feature vectors were raw PPG segments (actual data points). The proposed architecture achieved a MAE±SD of 5.59±7.25 mmHg for SBP and 3.36±4.48 mmHg for DBP. Additionally, the performance was assessed against the AAMI requirements and satisfied standards for both SBP and DBP in terms of error. However, only a total of 11546 5 s segments were used for evaluation. Moreover, the preprocessing steps were not straightforward as it required transforming the PPG into images, which also requires a lot of space. Furthermore, the model structure consisted of three separate neural network models. This means that the model complexity is too high, and thus requires a lot of processing power. The following Table 3.1 summaries all the aforementioned methods presented in this chapter.

3.2   Summary

Cuffless BP estimation using a single sensor/signal is appealing and has become more feasible with the recent advancement in computational models and wearable technologies. Although some studies have shown remarkable performances using the PPG approach, this was mostly achieved using a relatively small number of subjects, hence, the majority of these studies did not adhere to the strict requirement (as stipulated by AAMI) of 85 subjects for evaluation. Furthermore, uncertainty still remains around the accuracy of the PPG approach which is evident given the wide variation in the reported results. For this reason, more research is certainly needed in order to find the best combination of features and computational model(s) in order to bring the BP estimation capability to an internationally acceptable standard.

Chapter 4

Dataset, Signal Pre-processing, Feature Extraction and Selection

The aim of this chapter is to introduce the data source, from which the PPG and their corresponding reference BP signals were acquired. It also describes in detail the signal pre-processing steps, including filtering, baseline wandering removal, normalisation, and segmentation, in addition to presenting the extracted PPG features and the assessment of their relationship as well as individual influence on SBP and DBP estimation. Additionally, it also describes the methods used and processing steps taken in order to reduce the input feature vector dimension and select only the most effective features for BP estimation.

4.1 MIMIC II

In this research, the data used for evaluating the BP estimation algorithms were derived from the MIMIC II dataset [152]. This dataset is provided online by Physionet and contains several waveforms measured simultaneously from thousands of patients in intensive care units (ICU). It includes signals such as PPG recorded from the fingertip, invasive ABP, ECG, etc. Both PPG and ABP signals are sampled at 125 Hz. The MIMIC dataset provides a variety of subjects from different age groups and gender with potentially varying range of BP values, which are of great value for this study. In addition to this and the gold standard invasive BP reference, the

MIMIC is one of the most commonly used dataset for evaluating cuffless BP estimation algorithms, allowing comparisons to be made between results obtained in this research with other research studies.

4.1.1 Data collection

The original raw signals acquired from the MIMIC II are noisy and sometimes corrupt and require a lot of pre-processing prior to their utilisation in the analysis. Fortunately, Kachuee et al [113] have published clean and pre-processed data from the MIMIC II and hence, we adopted their version of the MIMIC II dataset that are currently presented in the University of California, Irvine (UCI) Machine Learning Repository.

4.1.2 Data structure

The records in this dataset are extracted from 942 patients. Kachuee et al [113] stored the data in a very efficient and convenient format for analysis. The data is stored in four parts, each part contains three thousand recordings. Each recording consists of three signals, PPG, ECG and ABP stored in a cell array of matrices where each cell corresponds to one record. Each row in each matrix represents one signal channel. For this project only the PPG and the reference ABP signals were used.

4.2 Signal pre-processing

4.2.1 Filtering

Since the dataset utilised in this research was compiled and published by Kachuee et al [113], the filtering and denoising step is the identical to the method presented in their paper. Several pre-processing techniques were tested in their study, and after analysis, the discrete wavelet decomposition (DWT) was selected given its performance advantages over the others in terms of complexity, phase response and handling of different levels of noise in the signals. The DWT was applied to the signals performing ten decomposition levels, with Daubechies 8 (db8) set as the mother wavelet. Afterwards, the decomposition coefficients corresponding to the

very low (0 to 0.25 Hz) and high (250 to 500 Hz) frequency components were set to zero in order to eliminate them. The wavelet denoising was then performed with soft Rigrsure thresholding on the rest of the decomposition coefficients. The pre-processed PPGs were then obtained by reconstructing the decomposition.

4.2.2 Baseline wandering removal

It is essential for the peak and foot of the PPG to be clear for proper detection of single cycles and accurate feature extraction. Generally, a PPG cycle contains at least one peak (systolic), that is normally clearly visible, and two feet. However, the foot of the PPG are not always clear, aligned and may vary greatly in amplitude. This makes it difficult to properly interpret the PPG and to clearly identify the feet at the onset and end of each cycle. During acquisition, the low frequency baseline wandering of the PPG could be introduced by respiration, sensor and body motion. In this study, the baseline wandering was removed using adaptive iteratively reweighted penalised least squares [156]. The effect of baseline wandering removal on the PPG signal is illustrated in Figure 4.1.

4.2.3 Normalisation

In order to make the feature extraction process more accurate and robust, the PPG's amplitude was normalised to a range of (0,1). The amplitude of the PPG varies greatly between subjects and this introduces two challenges. First, the arbitrary change in the PPG amplitude makes the cycle per cycle segmentation (for feature extraction) more difficult since it depends on a fixed PPG peak detection threshold. This in turn may cause cycles with lower PPG peaks than the peak threshold to go undetected and, hence losing valuable data and as a result reduce the BP range in the final dataset. Second, some of the extracted features may depend on the scale or amplitude of the PPG and its derivatives. Therefore, all PPG signals were normalised in order to make the feature extraction process more efficient and ensure that the extracted feature values are meaningful. The PPG signals were normalised using the min-max normalisation method. Figure 4.1 presents an example of a pre-processed PPG segment.
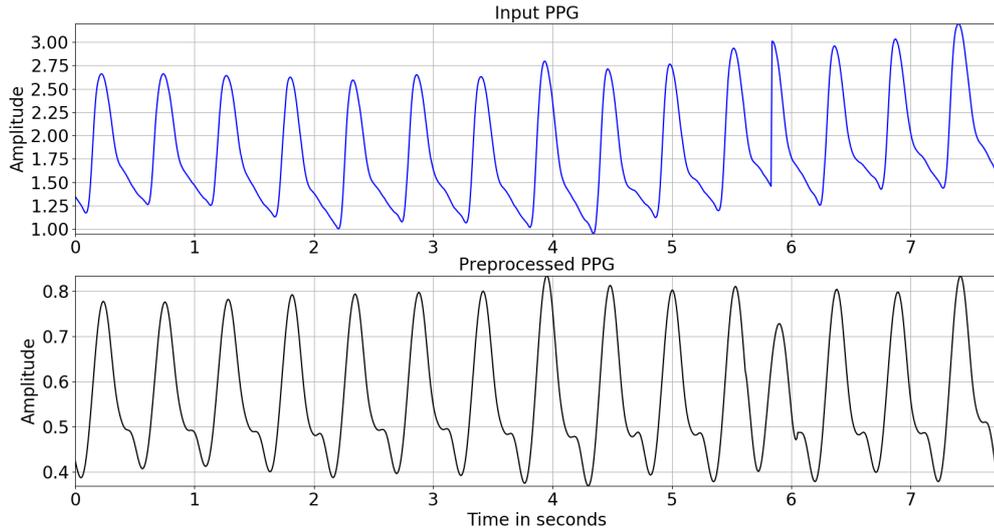
Figure 4.1: PPG signal before (top) and after (bottom) pre-processing including filtering, baseline wandering removal and normalisation

4.2.4 Segmentation

The signals were segmented into 10 second frames in order to take full advantage of the LSTM and GRU structures in processing sequential data for modelling the time domain variation in the PPG features. During data segmentation, low quality signals, including irregular and distorted segments, affected by motion artefacts were excluded from the final dataset, as well as PPG segments with missing BP reference segments. Irregular and corrupt segments were identified as having less than 8 or more than 16 cycles per 10 seconds during signal processing. Figure 4.2 shows examples of bad and inappropriate quality PPG segments that have been excluded from feature extraction. Low quality and inappropriate signals increase the chances of having outliers in the data, which in turn can negatively impact the data driven BP estimation models, and hence decreases its prediction precision and accuracy.

Additionally, sequences that correspond to very high or very low SBP and DBP were also excluded (e.g. $SBP \geq 180$, $DBP \geq 130$, $SBP \leq 80$, $DBP \leq 60$). These ranges were excluded from the analysis due to lack of sufficient number of segments containing very high and very low BP values. The ground truth SBP and DBP values were extracted from the ABP signal, which was not pre-processed nor filtered in order to keep the peaks and valleys as accurate as possible. The SBP and DBP correspond to the average of the detected peaks and valleys, respectively, in the 10 s sequence. The peak is the highest value in each single cardiac cycle whereas the
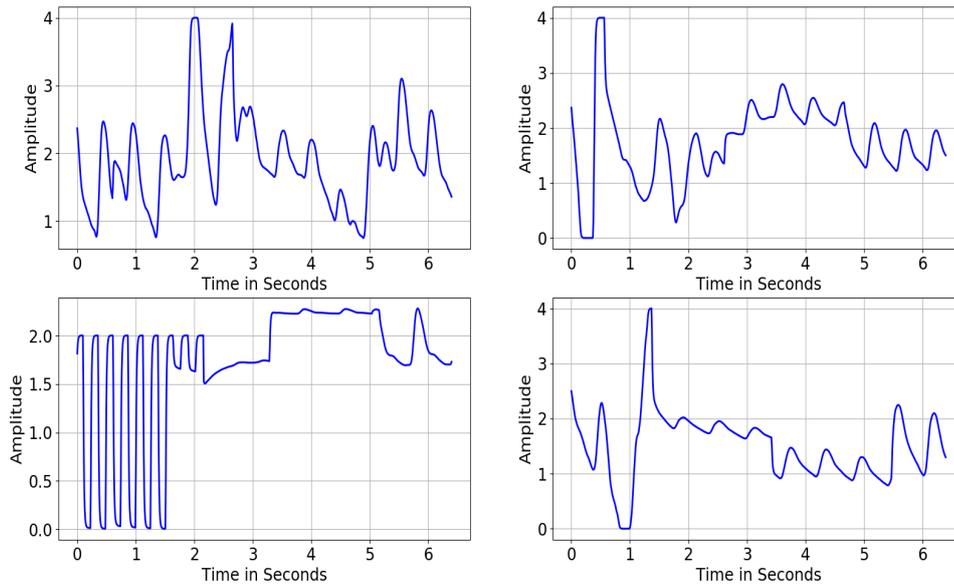
Figure 4.2: Examples of inappropriate PPG segments that were excluded from further analysis.

valley is the end-diastole value. Lastly, given the large size of the dataset presented by Kachuee et al [113] (more than 740 hours in total) and in anticipation of the computational complexity of the deep learning models, the size of the dataset was further reduced in an effort to speed up the training and alleviate the computational requirements. The final dataset comprises of roughly 80000 10-second good quality PPGs and their corresponding ABP segments. Table 4.1 and Figure 4.3 present the statistical information about the ranges of 10-s averaged SBP and DBP values in terms of min, max, mean and standard deviation in the dataset.

Table 4.1: Statistics for the SBP and DBP values in the final dataset

	Min (mmHg)	Max (mmHg)	Mean (mmHg)	SD (mmHg)
SBP	80.09	179.9	134.3	19.8
DBP	60.00	129.58	73.48	10.04

4.3   Feature extraction

In the PPG approach, BP is commonly estimated using a set of features computed from the PPG contour describing its morphology on a per-cycle basis (i.e., features
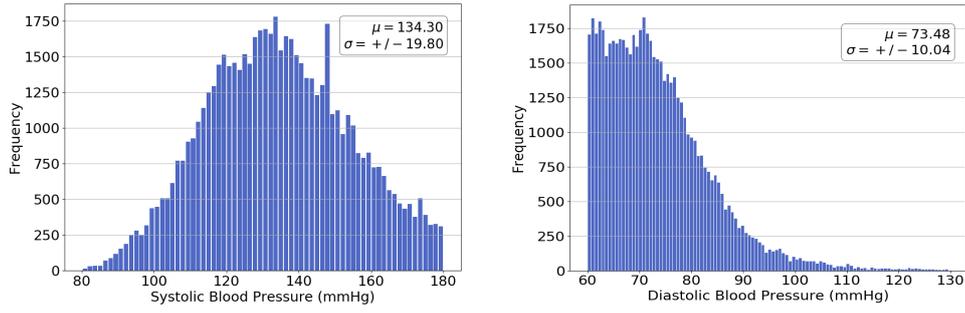
Figure 4.3: SBP and DBP ranges in the dataset.

extracted from every cardiac cycle), particularly those that can best interpret the changes in BP. Hence, the estimation accuracy of this approach relies heavily on the quality of the features, coupled with a high-performing prediction model.

4.3.1 PPG waveform characteristics

The PPG pulsatile component is characterised by its height and the duration of specific components of the cardiac cycle. However, as mentioned previously, there is no explicit knowledge about a known set of PPG features that correlate directly to BP, and many research groups around the world are still currently investigating this relationship. This, and the fact that the PPG waveform contour varies between individuals, make it difficult to accurately detect and extract some features from all PPG waveforms. For example, in a perfect PPG waveform, as shown in Figure 4.4 (a), the dicrotic notch is clearly visible. Hence, many features, such as, the time from the cycle onset to the dicrotic notch, time from the dicrotic notch until end diastole and other relevant features, can be easily extracted from this kind of waveform. However, generally, this kind of contour shape can only be seen in young and healthy subjects [123]. In another example, illustrated in Figures 4.4 (b) and (c), the dicrotic notch is not as clear or distinct but can be approximated accurately from the inflection point in the diastolic part of the PPG wave, through the first and second derivative of the PPG [123]. Another example, shown in Figure 4.4 (d), given the sharp decay in the diastolic portion of the waveform, the dicrotic notch is non-existent and completely invisible. This makes is very difficult to establish a reliable and error free algorithm for extracting all different PPG features systematically. These are some of the main limitations of this approach.
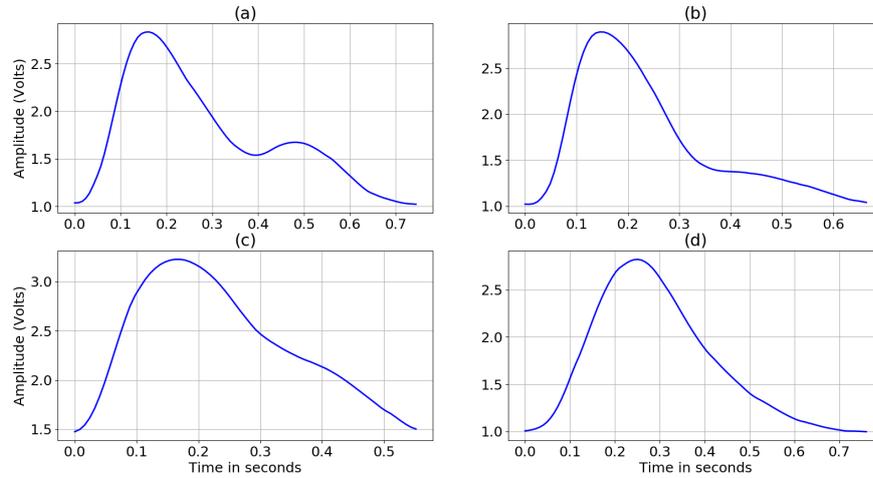
Figure 4.4: Different PPG waveform contours. (a) shows clear dicrotic notch. (b) and (c) the dicrotic notch is not distinct, thus relevant features cannot be easily extracted but can be approximated from the inflection point. On the other hand, in (d) the dicrotic notch is completely invisible thus relevant features cannot be extracted.

4.3.2 Extracted features

The signals provided by the MIMIC dataset are collected from ICU patients (i.e., patients under medication which could influence the BP and the PPG waveforms, as well as other factors such as diseases and age), thus the dicrotic notch and diastolic peak may not be dominant or visible in all signals. This is also reflected in the PPG' and PPG" (the first and second derivative of the PPG signal respectively) making it very difficult to compute related features such as delta T (time between systolic and diastolic peaks), dicrotic notch time, augmentation index, large artery stiffness index and many other important features. Therefore, all features related to dicrotic notch and diastolic peak were excluded from this study. It is essential, however, to have two clear feet and one systolic peak since all extracted features from the PPG, PPG' and PPG" depend on them. In total, 52-features were computed which are presented in Table 4.2 and shown in Figure 4.5.

4.3.3 Relationship with BP

Several of the features computed in this study have already been linked or proven to have influence on BP. It has been suggested that the systolic peak is rather more

68

Table 4.2: Features extracted from the PPG, PPG' and PPG"

Feature	Definition (and abbreviation/labels for future analysis)
PPG - 34 features	
Systolic peak	Peak amplitude of the waveform (ppg_peak)
Pulse width and ratios @:	
10 % amplitude	DW10; DW10+SW10 (sdw10); DW10/SW10 (d_s10)
25 % amplitude	DW25; DW25+SW25 (sdw25); DW25/SW25 (d_s25)
33 % amplitude	DW33; DW33+SW33 (sdw33); DW33/SW33 (d_s33)
50 % amplitude	DW50; DW50+SW50 (sdw50); DW50/SW50 (d_s50)
66 % amplitude	DW66; DW66+SW66 (sdw66); DW66/SW66 (d_s66)
75 % amplitude	DW75; DW75+SW75 (sdw75); DW75/SW75 (d_s75)
Systolic area (A1)	Area under curve between start and peak of the waveform (s_area)
Diastolic area (A2)	Area under curve from the peak to the end of the waveform (d_area)
Pulse area	A2/A1
Pulse interval (t_pi)	Distance between the start and end of the cycle (pi)
t_pi/ ppg peak	Ratio of the pulse interval to its systolic peak (pi_peak)
Heart rate	60/peaks interval (hr)
Systolic upstroke time	Time interval between the start and peak of the waveform (st)
Diastolic time	Time interval between the peak and end of the waveform (dt)
ST/t_pi	Ratio of the systolic upstroke time to its pulse interval (st_pi)
DT/t_pi	Ratio of the diastolic time to its pulse interval (dt_pi)
Main wave rising slope	Value of the PPG at the peak index of its first derivative (slope)
Slope/ppg peak	Relative height of the slope point (slope_peak)
t_slope	Time interval between the slope point and the peak (t_slope)
t_slope/t_pi	Ratio of time interval between the slope point and the peak to its pulse interval (t_slope_pi)
PPG intensity ratio	Ratio of the peak point intensity to the foot (valley) intensity (pir)
PPG first derivative (PPG') − 5 features	
Peak amplitude (a1)	Intensity of the first maximum peak of the PPG' (dev1_peak)
Peak time	Time interval between the beginning and a1 peak (t_a1)
Valley time	Time interval from the peak (a1) to first valley (t_b1)
t_a1/t_pi	Ratio of the PPG' peak time to its PPG pulse interval (t_a1_pi)
t_b1/t_pi	Ratio of the PPG' first valley time to its PPG pulse interval (t_b1_pi)
PPG second derivative (PPG") − 13 features	
First peak amplitude (A_a)	Intensity of the first maximum peak a-wave (dev2_peak)
First valley (A_b)	Intensity of the first valley of the PPG" after A_a (dev2_min)
A_b/A_a	Ratio of the first valley intensity to the first peak intensity of the PPG" (b_a)
AP_b/AP_a	Ratio of the PPG amplitude at the index of A_b to the amplitude of the PPG at the index of A_a (apb_apa)
A_e/A_a	Ratio of the second peak intensity e-wave A_e to the first peak of the PPG" (ae_aa)
AP_e/AP_a	Ratio of the PPG amplitude at the index of A_e to the amplitude of the PPG at the index of A_a (ape_apa)
t_a2	Time interval from the start of the PPG" to its first peak A_a
t_b2	Time interval between A_a to the first valley A_b
t_c2	Time interval between the first valley A_b to the second peak A_e
Total PPG" intensity	Total intensity of PPG" between peak A_a and valley A_b (dev2_height)
t_a2/t_pi	Ratio of t_a2 to the pulse interval of the PPG waveform (t_a2_pi)
t_b2/t_pi	Ratio of t_b2 to the pulse interval of the PPG waveform (t_b2_pi)
t_c2/t_pi	Ratio of t_c2 to the pulse interval of the PPG waveform (t_c2_pi)

suitable for continuous BP estimation than PAT [123]. In another investigation published by Awad et al [157], pulse width at 50% amplitude has been found to be correlated with total peripheral resistance or systemic vascular resistance. Pulse area is another indicator of total peripheral resistance as reported in a previously
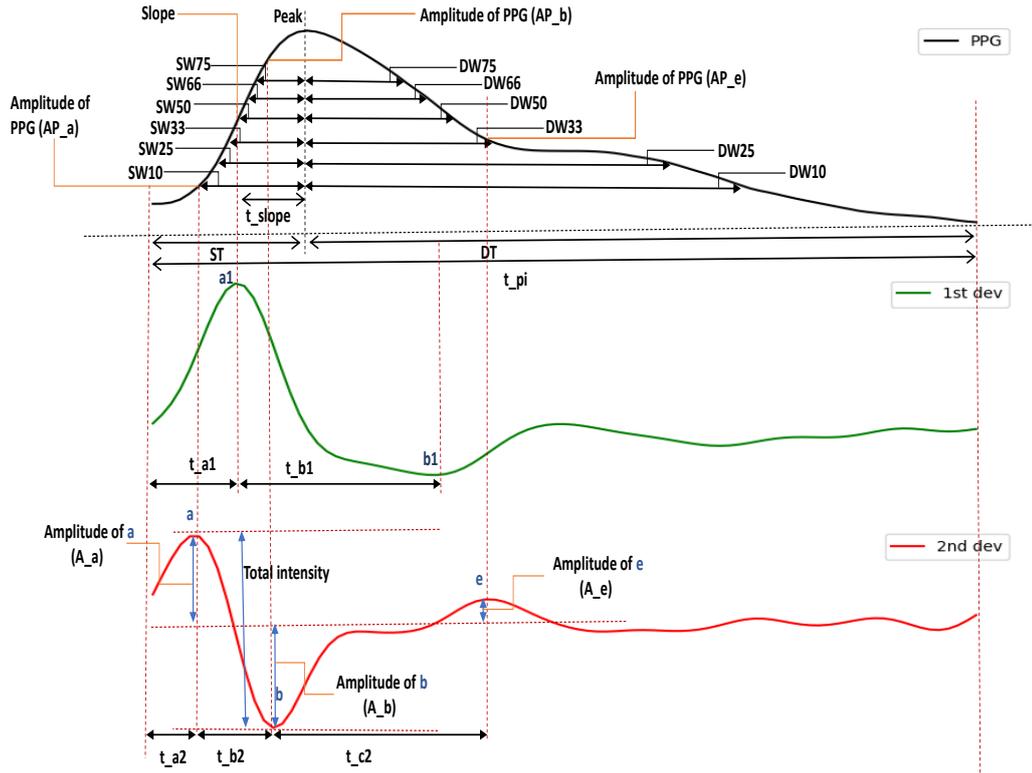
Figure 4.5: Illustration of temporal features from the PPG and its derivatives.

published study by Wang et al [158]. Kurylyak et al [25] expanded on the findings of Awad et al study by considering a total of 21 width-related features from the systolic and diastolic phases and their ratios at 10%, 25%, 33%, 50%, 66% and 75% amplitude of the PPG, as well as systolic time, diastolic time and cardiac period. Their study aimed to include as many PPG features as possible in order to capture the shape of the waveform accurately and improve the estimation performance. It has been suggested by Poon et al [159] that the ratio of pulse interval to its peak amplitude provides an understanding of the properties of the cardiovascular system of a person. Systolic upstroke time (ST), referring to the time between the start of the PPG cycle and the systolic peak point reflects changes in BP and can be used for cardiovascular disease classification [160]. PPG intensity ratio (PIR) has been used in several studies as a potential indicator for BP estimation [127, 161]. PIR is correlated with the arterial diameter change, which has an impact on the total peripheral resistance [161]. Moreover, height ratios of the PPG" extracted from the a-wave, b-wave and e-wave denoted as A_b/A_a and A_e/A_a are linked to arterial distensibility, and age. For example, higher A_b/A_a ratio reflects an increased

arterial stiffness which increases with age, while higher A_e/A_a ratio indicates decreased arterial stiffness hence decreases with age [123]. Both arterial distensibility and total peripheral resistance are direct influencers of BP.

4.4 Dimensionality reduction

Feature selection or dimensionality reduction is an essential step that aims to reduce the complexity of the model by eliminating redundant or irrelevant features which in turn helps to avoid the risk of overfitting the machine learning algorithms. The outcome of this process is a reduced set of features with the highest impact and influence on the BP estimation. In order to achieve this objective, several steps are involved which are described in the following subsections.

4.4.1 Feature normalisation

The features extracted from the PPG represent different BP indicators. Therefore, their scale may differ significantly in value within the same input feature vector dimension. For example, systolic upstroke time has incredibly small values compared to the heart rate values. Hence, these features are not comparable, which leads to imbalanced analysis, and introduce errors and abnormality during training. This is due to the fact that machine learning models consider higher feature values as more important and assign higher weights to those features during processing, and as a result this weakens and undermines the actual impact and influence of smaller value features on BP estimation. This bias is troublesome and highly affects model precision, performance, as well as complexity. Thus, the dataset must be normalised so that the model treats all features as equally important and valuable for BP approximation. Another advantage of normalisation is that it can suppress the adverse effect of outliers by scaling down all features to the same range of (0,1). This obviously has a trivial benefit in terms of time complexity as well, since it is easier to process (add and multiply) smaller numbers which speeds up the computation, and in turn it translates to faster predictions. In this work, the normalisation was done on the entire dataset together before the train validation and test slip to effectively and accurately bring down input feature values to the same range. The normal-

71

isation method used here is the min-max normalisation which follows this simple mapping equation:

$$X_{norm} = \frac{x - min(x)}{max(x) - min(x)} \tag{4.1}$$

4.4.2 Collinearity test and feature elimination

In the second step, two methods were applied to assess the correlation among the extracted features and their relationship to BP. The methods involved in this process are namely Pearson's correlation coefficient and maximum information coefficient (MIC) [162]. The former is used to evaluate the linear relationship between each pair of features, while the latter is used to evaluate both linear and nonlinear relationships between the extracted features and BP. In the first part of the analysis, Pearson's correlation was applied to identify each pair of features with high linear correlation in an attempt to reduce collinearity in the input vector. Table 4.3 presents the pair of features with high collinearity in the dataset.

Collinearity introduces redundancy in the data, and this increases the computational complexity (more space and time needed) and the risk of overfitting the model. In this study, two features are considered to be highly correlated if their Pearson's correlation coefficient is bigger than or equal to 0.9. Afterwards, the results of the MIC between all features and BP were analysed for both SBP and DBP, separately. Figures 4.6 and 4.7 depict the strength of the relationship between the extracted features and SBP and DBP, respectively. MIC values were between 0 and 1, and similar to Pearson's correlation coefficient, a higher value means higher correlation/dependency between the pair. Therefore, among each collinear pair, the feature with the smaller MIC value against BP (SBP or DBP) was dropped, while the other one was retained for further analysis. Table 4.4 shows the features with weaker correlation (among the collinear pair) to SBP and DBP, based on the results obtained from MIC. Those features were dropped separately, and as a result, two datasets were created. The first dataset contains the remainder of the 52 collinear-free input features for SBP, while the second one contains the remainder of the 52 collinear-free input features for DBP.

Table 4.3: Pairs with high collinearity in the input feature vector

d_s33	d_s25
d_s75	d_s66
dev2_height	dev1_peak
dev2_peak	dev1_peak
dev2_height	dev2_min
dev2_height	dev2_peak
sdw10	DW10
DW33	DW25
sdw25	DW25
sdw33	DW33
sdw50	DW50
DW75	DW66
dt	hr
dt	t_pi
hr	t_pi
sdw33	sdw25
DW33	sdw25
sdw50	sdw33
sdw66	sdw50
sdw75	sdw66
dt_pi	st_pi
t_a2_pi	t_a2
s_area	t_slope

Table 4.4: Highlighted features with weaker correlation to SBP and DBP based on the MIC analysis

DBP		SBP	
d_s33	d_s25	d_s33	d_s25
d_s75	d_s66	d_s75	d_s66
dev2_height	dev1_peak	dev2_height	dev1_peak
dev2_peak	dev1_peak	dev2_peak	dev1_peak
dev2_height	dev2_min	dev2_height	dev2_min
dev2_height	dev2_peak	dev2_height	dev2_peak
sdw10	DW10	sdw10	DW10
DW33	DW25	DW33	DW25
sdw25	DW25	sdw25	DW25
sdw33	DW33	sdw33	DW33
sdw50	DW50	sdw50	DW50
DW75	DW66	DW75	DW66
dt	hr	dt	hr
dt	t_pi	dt	t_pi
hr	t_pi	hr	t_pi
sdw33	sdw25	sdw33	sdw25
DW33	sdw25	DW33	sdw25
sdw50	sdw33	sdw50	sdw33
sdw66	sdw50	sdw66	sdw50
sdw75	sdw66	sdw75	sdw66
dt_pi	st_pi	dt_pi	st_pi
t_a2_pi	t_a2	t_a2_pi	t_a2
s_area	t_slope	s_area	t_slope
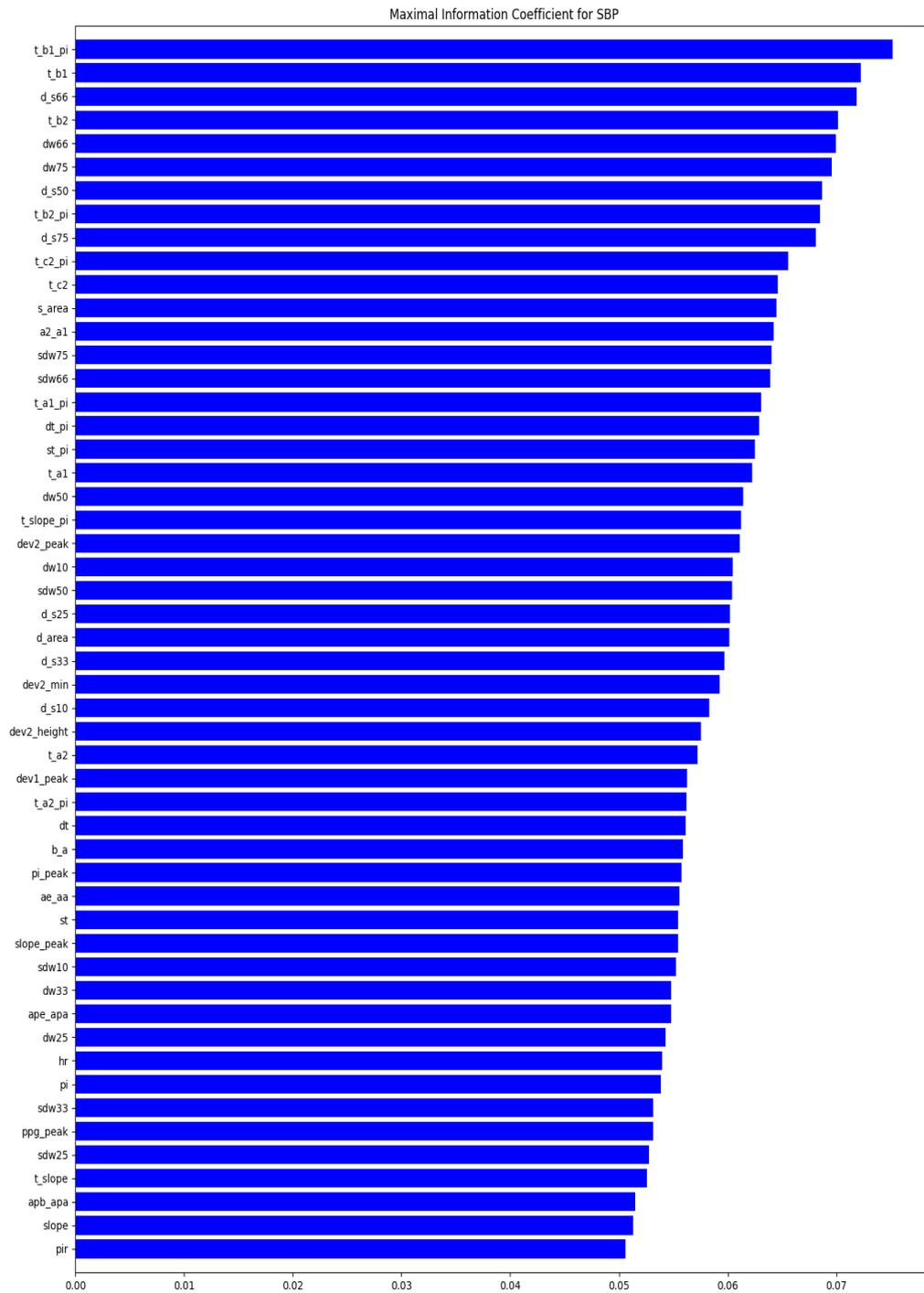
Figure 4.6: Maximal information coefficient between feature variables and SBP.
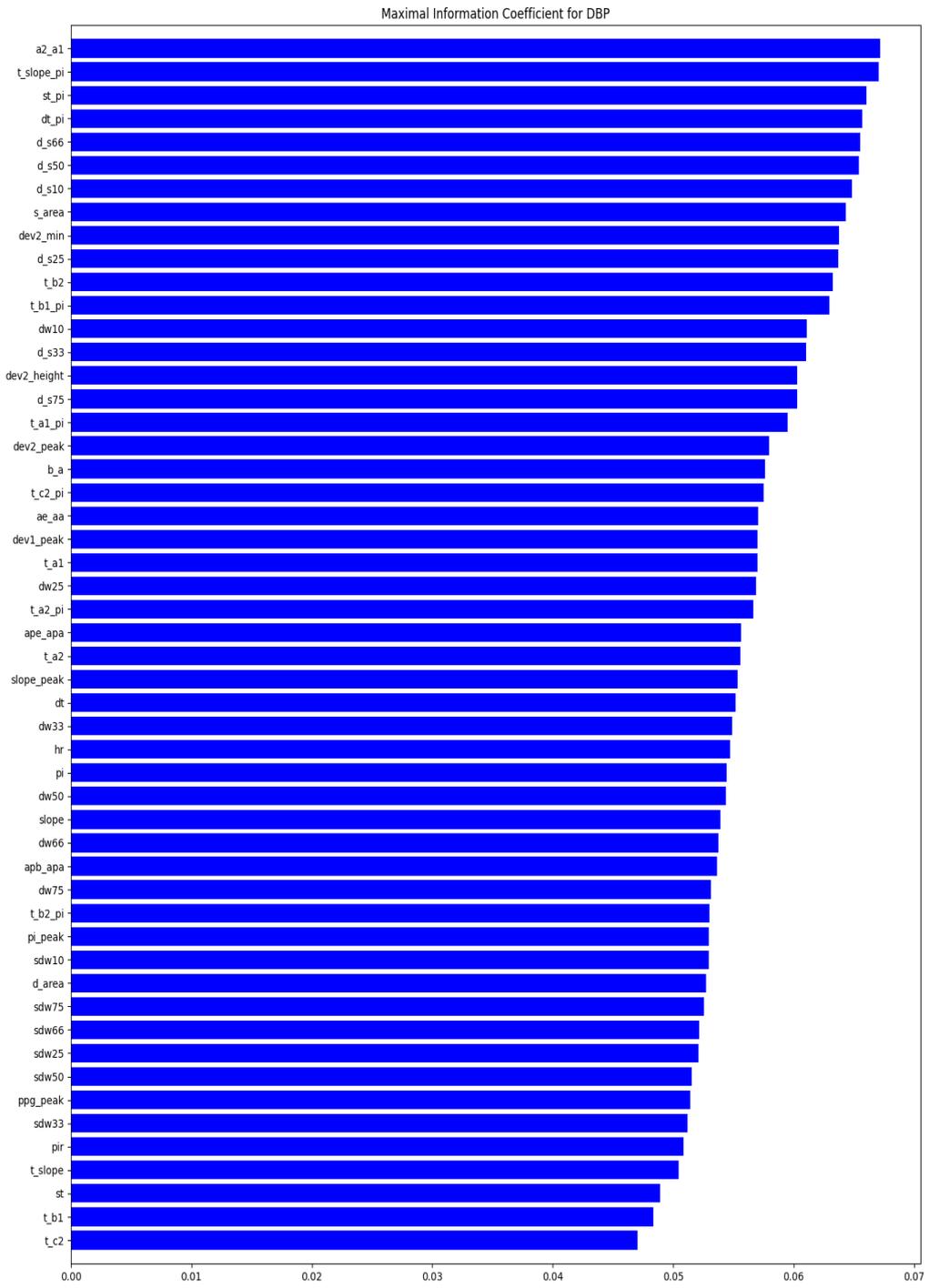
Figure 4.7: Maximal information coefficient between feature variables and DBP.

4.4.3 Feature selection

The input feature vectors for both SBP and DBP obtained from the previous step were further reduced using recursive feature elimination (RFE) [163]. RFE is a popular wrapper-type feature selection method. Wrapper methods incorporate a machine learning algorithm, used in its core, during the feature selection process. RFE starts with all feature-set and iteratively removes features until the optimum or specified number of features is reached. RFE works by fitting a learner model, ranks the features by importance based on the results of the regression model, eliminates the features with the least contribution on the target estimation and refits the model. This process is repeated until the optimum number of features is achieved. In this study, the learner model wrapped in the core of RFE is random forest. The RFE has two configurations, either the user specifies the number of desired features or the algorithm finds the optimal number of features. The latter configuration was chosen, and as a result, the RFE selected 22 features for DBP and 23 features for SBP as optimal number of predictors. After analysing the results, it was found that most of the selected features were common for SBP and DBP but with different importance. These features are listed in Table 4.5 and their relative strength/influence is presented in Figures 4.8 and 4.9. The reduced input feature vectors for both SBP and DBP were combined together to form the final input vector, comprising 24 features in each cycle, for the deep learning models. However, it should be noted that for the classical machine learning models employed in this study, the input feature vectors for the SBP and DBP were tested separately. In other words, classical machine learning models were trained for one objective at a time (either SBP or DBP estimation), while the deep learning models learn both objectives simultaneously. Moreover, all models utilised here (described in the next two chapters) were evaluated and compared on the full feature set and the reduced feature sets obtained from this step. Thus, several datasets were created to accommodate all these combinations, which are described below:

For classical machine learning models, the input features extracted from each cycle of the 10 second PPG sequence were averaged to form one cycle, and the following datasets were generated:

- Full feature set containing input vector of 52 features

- Reduced feature set containing input vector of 23 features for SBP, listed in Table 4.5

- Reduced feature set containing input vector of 22 features for DBP, listed in Table 4.5

For the deep learning models, the input vector is a sequence of 10 second PPG cycles, and the following number of features were extracted from each cycle:

- Full feature set containing 52 features extracted from each cycle

- Reduced feature set containing 24 features from each cycle, representing the combined features for both SBP and DBP, listed in Table 4.5

Table 4.5: List of the best features selected by the RFE for SBP and DBP separately and combined input features vector for deep learning models

# of features	SBP	DBP	Combined
1	ae/aa	ae/aa	ae/aa
2	t_b1	d_s25	d_s25
3	dt	dt	dt
4	DW25	t_slope_pi	t_slope_pi
5	t_b2	dev2_min	dev2_min
6	t_a1_pi	t_a1_pi	t_a1_pi
7	t_b1_pi	DW25	DW25
8	d_s66	t_c2	t_c2
9	DW10	DW10	DW10
10	t_c2	d_s50	d_s50
11	t_slope_pi	t_b2	t_b2
12	b/a	t_a1	t_a1
13	d_s50	t_b1_pi	t_b1_pi
14	d_s25	t_b1	t_b1
15	st	b/a	b/a
16	t_c2_pi	sdw75	sdw75
17	sdw75	DW50	DW50
18	d_s10	ppg_peak	pi_peak
19	t_a1	t_c2_pi	t_c2_pi
20	DW50	d_s66	d_s66
21	t_b2_pi	st	st
22	dev2_min	d_s10	d_s10
23	pi_peak		t_b2_pi
24			ppg_peak
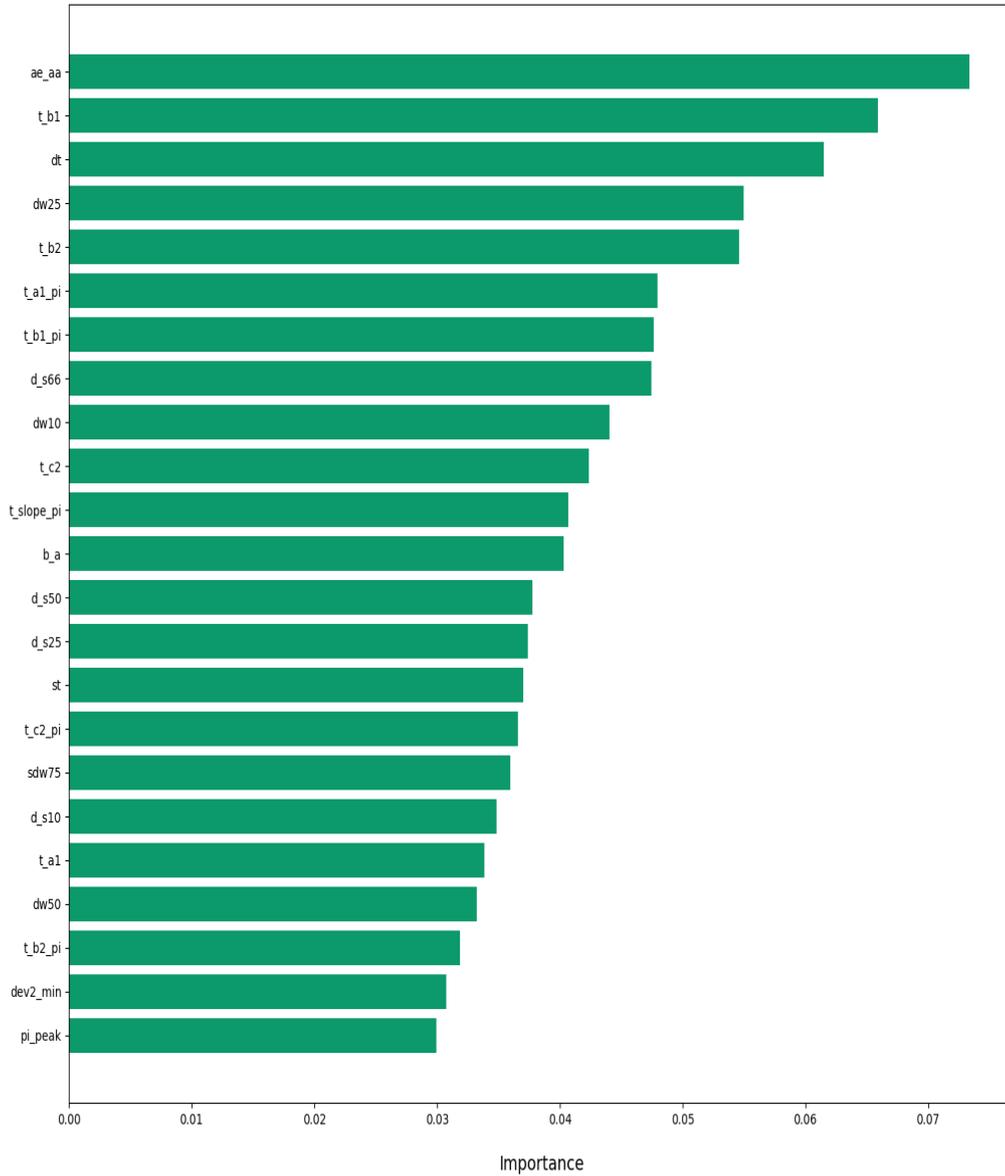
Figure 4.8: Ordered list of optimum features selected by the RFE method for SBP, arranged by their importance.
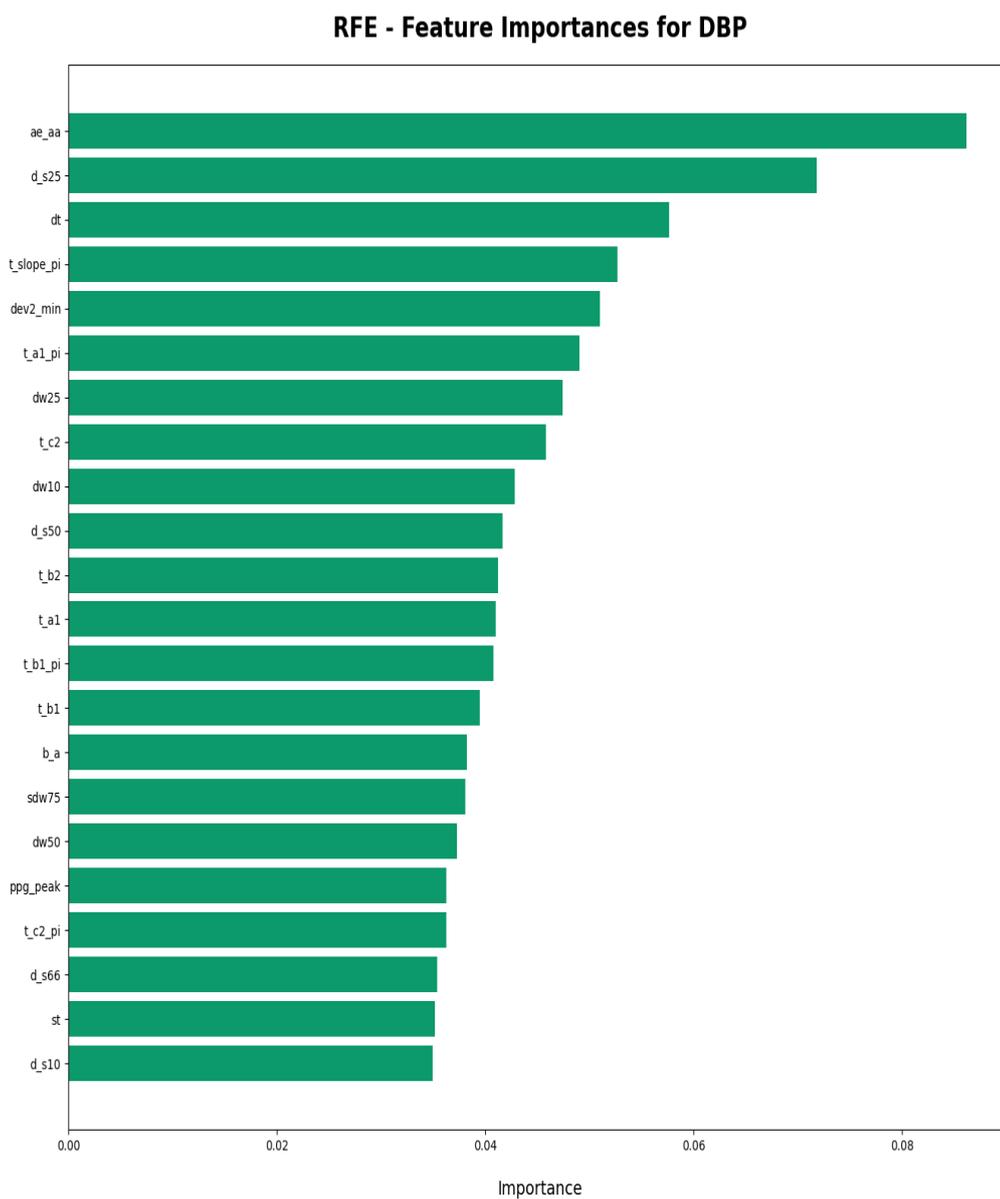
Figure 4.9: Ordered list of optimum features selected by the RFE method for DBP, arranged by their importance.

4.5 Summary

The quality of the PPG signals have a direct impact on the BP estimation, hence, signal pre-processing is an integral part for accurately extract the PPG features. Additionally, data processing including feature normalisation and dimensionality reduction, are also essential in order to reduce effect of outliers and eliminate redundant insignificant features in the dataset. This in turn helps improve the estimation accuracy of the model and reduce its complexity. The analysis provided in this chapter identifies and run the features that are most prominent for SBP and BDP which will be considered and utilised in the development of the BP estimation models.

Chapter 5

Classical Machine Learning Regression Algorithms

Machine learning algorithms refer to statistical estimation models that can learn to approximate a target output from past observations, through mathematical operations. This behaviour resembles the humans' ability to adapt and learn from past examples and experiences. Machine learning tasks are divided into supervised and unsupervised learning. In supervised learning, the training data is labelled. This means that the target output is known or available in the dataset, hence, the algorithm learns a mapping function from input vector (independent variables) to target output (dependent variable). On the other hand, in unsupervised learning tasks, the target output is not known in the dataset. In this type of tasks, the algorithm is forced to learn patterns from its environment and generate new representations. This project is concerned with supervised learning since the labels or reference values (SBP and DBP) are available in the training dataset.

Supervised learning deals with two types of problems: classification and regression problems. Classification task refers to a problem where the objective or target output of the predictive model is to estimate discrete values, labels or categories. While regression task refers to a problem where the objective of the predictive model is to approximate continuous values. However, in both tasks, classification and regression, the relationship or the mapping function between the input variables and the target output is learned by data driven models. In this project, SBP and DBP values in the dataset are continuous values, hence, this research problem is a supervised machine

learning regression task. The following sections describe multiple classical machine learning regression models that were employed in this project for estimating SBP and DBP using features extracted from the PPG waveform.

5.1 Multiple Linear Regression (MLR)

Linear regression is one of the most frequently encountered regression models for BP estimation and regression tasks in general [108]. It is the simplest and the most used approach for learning a mapping function between one independent input feature variable X and one dependent target output Y (SBP or DBP). The relationship between the input variable and the output variable is assumed to be linear and is modelled given the equation of the line as follows:

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon \tag{5.1}$$

where β_0 is the intercept, β_1 is the slope coefficient of the independent variable X and ε is the random error. A change in the input variable X_1 causes changes to the associated coefficient β_1. The random error represents the difference in value between the ground-truth output and the estimated output. Training the linear regression model starts with random initialisation of its parameters. These values are updated/optimised iteratively during the training phase according to the obtained value of the cost/loss function, which is the mean squared error (MSE). The linear regression model tries to minimise the MSE between the estimated value and reference value. In other words, the linear regression model tries to find the best coefficients for which the MSE between the predicted value and the reference value is minimum. The algorithm is optimised when the lowest MSE is reached, such point is called global minima.

The linear regression model in its simplest form takes only one independent variable to estimate one target output. However, in this project, the input feature vector consists of multiple features, as mentioned in the previous chapter. Therefore, its popular variant, the Multiple Linear Regression (MLR) was employed instead. MLR models the relationship between multiple independent variables $(X_1, X_2, ..., X_n)$ and one dependent variable Y, such that Y is a linear combination of the independent variables. MLR is calculated given the following equation:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \qquad (5.2)$$

The advantage of the MLR lies in its simplicity, which makes it an appealing approach, particularly when the relationship between the input and the target is linear. Additionally, MLR is a useful and efficient technique for modelling non-complex relationship on small datasets. This also makes it easier to examine its parameters in order to figure out which features have more influence on the target estimation. However, despite these advantages, its performance it not as effective on large datasets, especially when the relationship between the input and the target output gets more complex and non-linear, as is the case for the large MIMIC dataset and the inherently complex relationship between the extracted PPG feature vector and BP. Another disadvantage of the MLR is that it does not allow the estimation for both SBP and DBP simultaneously, which necessitate establishing two models for estimating each objective separately. However, this further degrades the estimation precision as both SBP and DBP strongly correlate with each other [108], hence estimating both objectives using one model structure would enhance the estimation accuracy. Furthermore, the PPG is a time series signal, thus the input data is sequential. Yet, the MLR does not account for this when processing the data, which will further hinder the estimation precision.

5.2 Support Vector Regression (SVR)

Support Vector Regression (SVR) [164] is a non-parametric machine learning algorithm that is widely used in the literature for estimating cuffless BP values using physiological signals [108]. This algorithm is an extension of the Support Vector Machine (SVM) [165], which is a popular algorithm used for classification tasks. SVR utilises two types of loss functions, namely, epsilon-intensive (also known as Vapnik's ε-intensive or L1 regularisation) and Huber's (combines L1 and L2 regularisations) loss functions. These well regularised loss functions make the SVM/SVR more resistant to overfitting. However, it all depends on careful fine-tuning of the regularisation parameters, hyper-parameters and the choice of kernels.

The motivation behind the SVR is to find an optimised threshold or a bound on the generalisation error based on a margin. This essentially means that rather than minimising an error rate, the SVR tries to approximate a threshold to fit the data points within a certain distance from a separation line (known as hyperplane), such that the error is minimum. Particularly, data points that fall within the ε-boundaries, do not contribute to the error rate, i.e. their error is ignored. Because of this important feature -structural risk minimisation, defined by the epsilon intensive loss function, the SVR does not face the local minima issue associated with traditional error functions.

Figure 5.1 illustrates how the SVR calculate its error based on a margin. The black line in the middle is the hyperplane, whereas the two parallel dotted lines represent the decision boundaries. These boundary lines create the margins at $\pm\varepsilon$ distance from the hyperplane. Data points, represented as green squares, can fall on either side of the boundaries. Furthermore, data points that fall on the boundaries or outside it are called support vectors. The main objective for the SVR is to find the best decision boundaries, such that the largest number of data points that are closest to the hyperplane fall within these $\pm\varepsilon$ margins. Thus, only data points that fall outside these margins are penalised and taken into consideration for computing the error rate. For this reason, these margins can be seen as margins of tolerance.
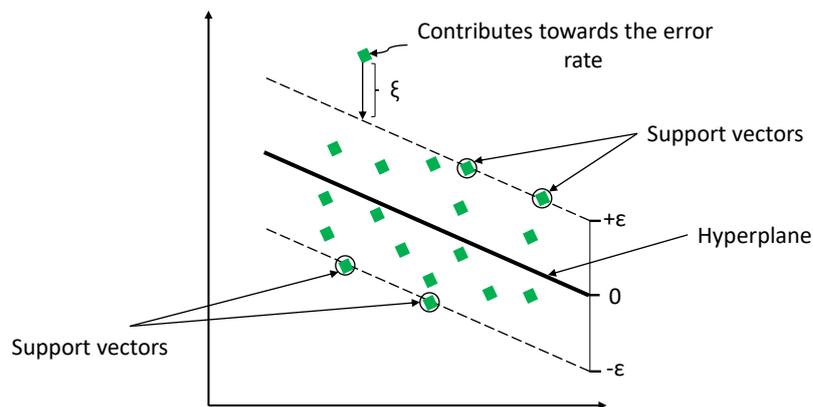
Figure 5.1: SVR: data points represented as green squares separated by the hyperplane. The dotted lines are the boundaries located at ε distance from the hyperplane.

The SVR is essentially a linear function, however, it can be applied for both linear and non-linear tasks. In this project, the relationship between the features and BP is complex and highly non-linear. Hence, in order for the SVR to handle the

non-linearity, it converts the data into linear by transforming the input feature from low dimension space onto a high-dimension space by a non-linear mapping achieved using a kernel function. Afterwards, linear regression can be applied in the new high-dimension space. For example, given a time series sample (x_i, y_i), where x_i are the independent PPG feature variables and y_i are the target SBP and DBP outputs, the SVR first maps the data onto a high dimension feature space $< \varphi_x, \varphi_y >$. However, instead of explicitly computing a mapping function, it utilises non-linear kernels, such as polynomial kernels or radial basis function (RBF) kernel, which allows the SVR to then take non-linear decisions. In this project, the kernel (K) employed is the RBF, defined in the equation below:

$$K \ (x, y) = \exp\left(-\frac{1}{2\sigma^2} \ ||x - y||^2\right) \qquad (5.3)$$

There are several advantages to the SVR. First, the regularised $\pm\varepsilon$-intensive loss function which penalise errors situated outside the ε boundaries, leads the spare representation of the decision rule given. Second, SVR is robust to outliers. Third, the complexity of the SVR is independent from input vector dimension. Nonetheless, there are also several disadvantages associated with the SVR, such as, poor performance on large datasets, does not handle time series data, and cannot estimate two target outputs simultaneously.

5.3 Random Forest

Random Forest (RF) [166] is a popular nonparametric tree-based supervised machine learning technique that is capable of performing both classification and regression tasks. Hence, RF can process both categorical and continuous data. Moreover, this technique belongs to the ensemble learning algorithms. These algorithms construct multiple machine learning models and combine their outputs in order to make a more robust decision/estimation than any underlying model could achieve on its own. In this case, RF establishes an ensemble of decision trees regression [167] from different subsets from the dataset and averages the predictions of the constituent trees to compute the final target output.

Decision tree regression is an extension of decision tree classifier, and therefore works in a similar fashion: establishing a tree starts from the top or root node and iteratively split the node into a left and right sub-nodes according to a decision function, until a stopping criteria is satisfied, creating terminal node where the decision (in this case SBP or DBP values) is typically presented. Particularly, the first step is to determine the best feature value of all its predictors that would lead to the best split i.e. the root node at which the dataset will be split. This is done by examining all the independent variables and create multiple splits for each predictor value or data point. The aim is to minimise the mean square error (i.e., cost function). Therefore, at each split predictor value, the MSE is calculated between the ground-truth BP values and the predicted value. Each node of the predictor value has a left and right sub-node. The predictor value that leads to a split that minimises the sum of squared error is selected as the root node. This process is repeated until a terminal node or a stopping criteria (e.g. max depth) is reached. In comparison to the SVR, decision tree regression is relatively simpler in terms of computation, however, it can easily overfit the data.

Random forest (RF) is the most popular tree-based algorithm. It consists of an ensemble of decision tree regression. The advantage of RF is that it introduces randomness every time a tree is created. This helps increase its accuracy, and subsequently its generalisation. RF works the following way:

- Build n number of regression trees from a random number of training examples and select a random subset of features for splitting.

- Each tree produces its own estimate.

- The predictions of all individual trees are averaged out to produce a final output for the RF.

This algorithm has several advantages. Such as, the averaging of prediction from all the trees reduces the risk of overfitting and increases its robustness. Additionally, RF provides the relative feature importance or impact on the target output. Furthermore, the RF does not make any assumption regarding the distribution or relationship of the data. Hence, it can handle linear and non-linear data. Moreover, RF can process discrete values for classification and continuous values in the case

of regression. However, despite its advantages, there are few drawbacks to the RF. For example, it is fairly complex and computationally demanding, and also is not suitable for time series estimation.

5.4   Adaptive Boost (AdaBoost)

AdaBoost [168] is another ensemble learning method that can be applied for classification and regression tasks. This method combines multiple under-performing learners to establish a stronger learner and boost its accuracy. It can be used to enhance the performance of decision trees or any machine learning model. Moreover, AdaBoost belongs to sequential ensemble learning group. This means that the learner models are created and added sequentially, and the shortcomings of previous weak models are learned by the following model. In this project, the selected based learner for the AdaBoost algorithm is the decision tree regression presented in section 7.3.

The algorithm works the following way:

- Initially, a random subset of training data is selected by the model.

- The (weak) learner model starts by estimating the output of the training samples. Each sample or observation is associated with a weight that indicates the importance of correctly estimate this sample. Initially, all samples are assigned equal weights.

- Incorrectly estimated observations are assigned higher weights in order to increase their probability of being correctly estimated in the next iteration.

- In each iteration, a weight is also assigned to the trained learner according to its prediction accuracy. However, unlike the previous step, here a higher weight is assigned to the more accurate estimators.

- Being an iterative process, these steps are repeated until all targets have been estimated or a specified number of learners have been reached.

The AdaBoost algorithm shares some the advantages and disadvantages presented in the previous sections. For instance, it can be used for both classification and

regression on linear and non-linear data, provided that the machine learning base learner can handle both cases. Another advantage is that it can be used with any machine learning algorithm such as SVR, RF, decision trees, etc. Additionally, it requires less parameters to tune in comparison to SVR. Moreover, AdaBoost can be used to boost the performance of underperforming weak models and enhance their accuracy. However, just like the previous models, AdaBoost is not suitable for time series modelling and requires establishing two separate models for estimating SBP and DBP. Furthermore, this algorithm is very sensitive to noisy data, and it is also highly affected by outliers.

5.5 Implementation and optimisation

Finding the best model parameters is a critical step for establishing an accurate machine learning estimator. As previously mentioned, the idea behind the optimisation is to simply find the optimised parameters for which the cost function is at its lowest value i.e., the global minima point. Lower error rate and fine-tuned parameters translate into higher estimation accuracy.

All the aforementioned models in this chapter were implemented in python, particularly using the scikit-learn library. In addition to the models, this library also implements a grid-search method over specified parameter values, which provides ease of use for users to find the optimum combination of parameters for the models on a given dataset. In this project, each statistical model was fine-tuned for each objective (i.e. one model for SBP and one for DBP) on the 52-feature (presented in chapter 4, section 4.3.2) and the reduced feature sets (presented in chapter 4, section 4.4.3). Table 6.1 provides the best parameters obtained from the gird-search.

Table 5.1: Parameters grid-search results for the statistical models

Models	parameters grid-search	Full feature set		Reduced feature set	
		SBP	DBP	SBP	DBP
MLR	-	-	-	-	-
SVR	C:[1,10,100,1000], Kernels:[linear, polynomial, radial basis function (rbf)]	C:1000, kernel: rbf	C: 1000, kernel: rbf	C:1000, kernel: rbf	C:1000, kernel: rbf
RF	Max depth:[10,20,30,40,50], # of estimators: [10,20,30,40,50,60,70,80,90,100,110,120,130,140,150]	max depth:10, # of estimators: 100	max depth:10, # of estimators: 100	max depth: 10, # of estimators: 100	max depth: 10, # of estimators: 100
AdaBoost	Learning rate (LR) :[0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8], # of estimators: [10,20,30,40,50,60,70,80,90,100]	LR: 0.1, # of estimators: 60	LR: 0.1, # of estimators: 10	LR: 0.1, # of estimators: 50	LR: 0.1, # of estimators: 20

5.6 Summary

There are several types of classical machine learning models that are intended to handle different tasks and data types. The models described in this chapter can be used for supervised classification and regression tasks. Also, these four models are the most popular estimators utilised in the literature for the BP regression task, where each has its own advantages and limitations. However, none of the afore-mentioned models are suitable for continuous BP estimation due to the challenges presented in the previous chapters, such as, the need for two models, cannot handle very large datasets and diminishing accuracy. The neural network models described in the next chapter can overcome these limitations and are more appropriate for this task.

Chapter 6

Neural Network Algorithms and Architectures

The aim of this chapter is to introduce the neural networks used in this research, their internal computational units, conventional models and proposed architectures. In particular, it starts with general overview of neural network, basic units of computation and their associated weights, as well as the different types activation units and optimisers. Additionally, this chapter describes in detail the multi-layer perceptron, Long-Short Term Memory, and Gated Recurrent Units. It also introduces the components of the proposed architecture, namely, Bidirectional connections and attention layer. Lastly, this chapter presents the object/cost function, optimiser, back-propagation method and implementation.

6.1   Introduction to neural network

Artificial neural networks are defined by Haykin (1999) [169] as "massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use". The characteristics of neural network resemble the characteristics of the human brain in two aspects:

- The network acquires knowledge through a learning process.

- The weighted links or connections between the neurons store the acquired knowledge.

An artificial neural network consists of processing units and weighted interneuron connections. The processing units are called nodes, neurons or simply units. Each layer contains a number of neurons that are connected to all neurons of the previous and the following layers through edges. There are various types of neural networks that were established to tackle different real world problems. For example, recurrent neural networks (such as LSTM and GRU) were built to handle time series sequential data, convolutional neural network are specialised in image processing, and encoder-decoder for machine translation, etc. The neural network models employed in this project along with the proposed architectures are explained in detail in the following sections.

The neurons are the basic units of computation in the neural network. Neurons receive inputs from previous neurons along with their associated strength or importance, relative to other inputs, in the form of weights, as shown in Figure 6.1. Each neuron computes an activation a, that is the weighted sum of incoming inputs multiplied by their associated weights. The output of the neuron is computed by a function $\mathbf{f}(a)$, formally called an activation function. The activation function for the hidden layer units is non-linear, while the output layer units can be non-linear in the case of classification tasks or linear for regression tasks. Introducing non-linearity to the output of the hidden layer neurons allows the network to learn the non-linear representation in the dataset.
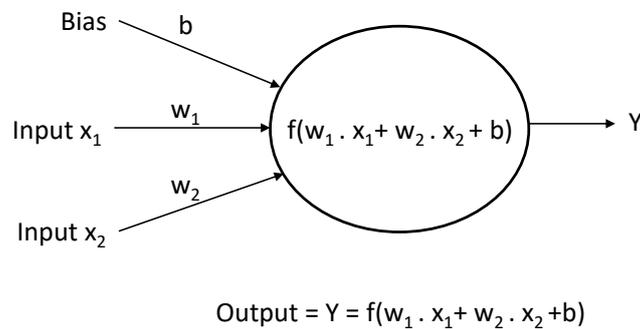
Figure 6.1: Example of a simple neural network processing unit.

There are several types of activation functions, some are more suitable than others

for certain tasks. However, they all receive a single input value and perform a set of mathematical operations to compute a single output value. The most frequently encountered activation functions in current practice are:

- **Sigmoid**: takes an integer value and normalises it to a range of [0,1], as shown in Figure 6.2. Sigmoid is non-linear, monotonic, easy to implement, outputs values within a fixed range and continuously differentiable. It is calculated using the following equation:

$$\sigma(x) = \frac{1}{(1 + e^{-x})} \tag{6.1}$$
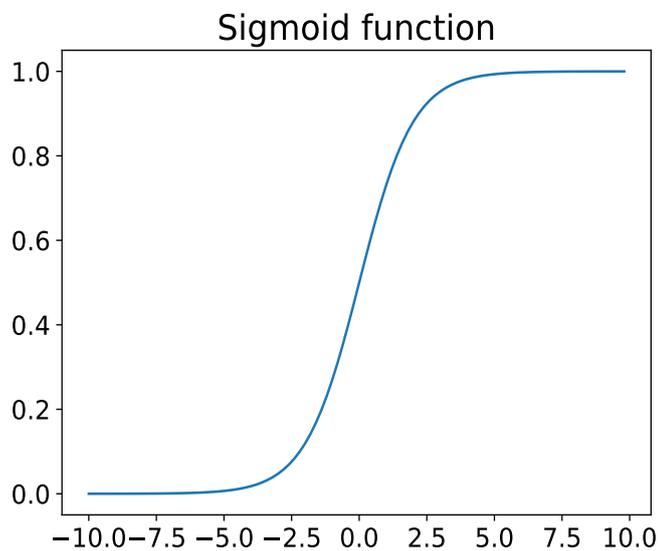
Figure 6.2: Plotting the graph for the sigmoid function.

- **Hyperbolic tangent (tanh)**: takes an integer value and normalises it to a range of [-1,1]. Figure 6.3 depicts the graph for the tanh function. tanh is also non-linear, however, its more popular in natural language processing and machine translation. tanh is computed given the following equation:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{6.2}$$

- **Rectified Linear Units (ReLu)**: takes a real valued input, replaces all negative values with zero, and outputs a positive number. The graph for the ReLu function is presented in Figure 6.4. ReLu is non-linear, simple and easy
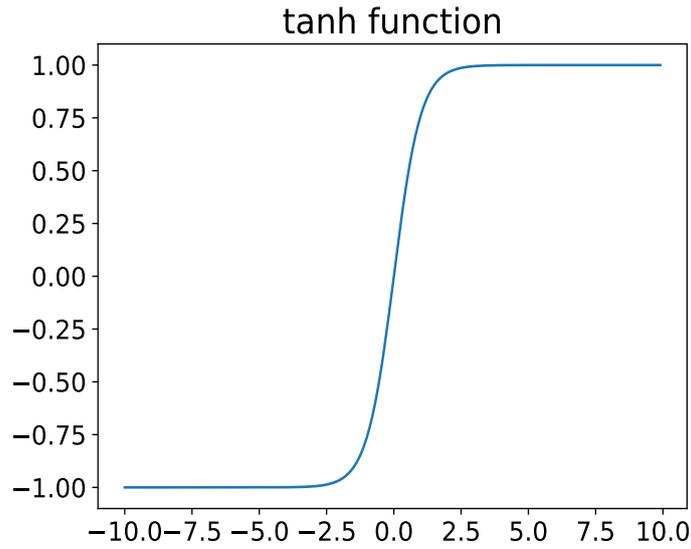
Figure 6.3: Plotting the graph for the Hyperbolic Tangent (tanh) function.

to implement. Despite its simplicity, currently it is the most popular activation function. ReLu is calculated as follows:

$$F(x) = max\ (0, x) \tag{6.3}$$
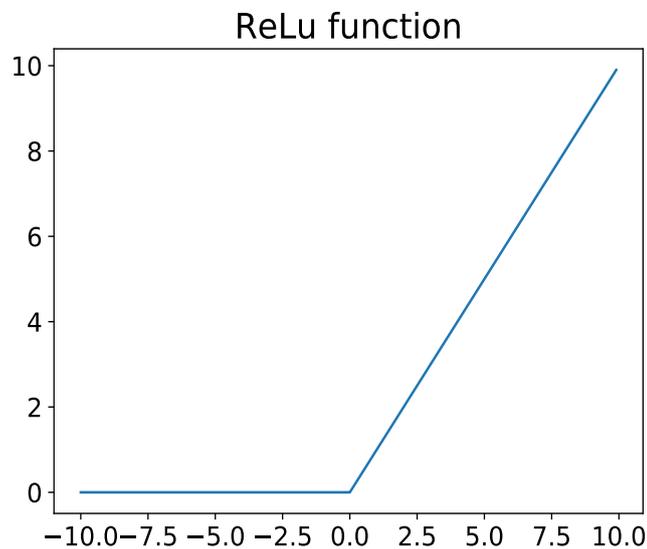
Figure 6.4: Plotting the graph for the Rectified Linear Unit (ReLu) function.

In terms of training a neural network, when the ground-truth values are given in the training data, it is called supervised learning. During training phase, the network learns from the available examples to adjust its parameters so that its response

(estimated target output) matches, as close as possible, the ground-truth output. There are several learning algorithms (optimisers) that can achieve this objective using different approaches. Some of the most popular optimisers are gradient descent [170], stochastic gradient descent, Adagrad [171], Adadelta [172], RMS, and ADAM [173], etc. For this project, ADAM was selected as the optimiser to train the developed neural network models since it converges much faster and requires less training epoch. This algorithm is explained in detail in section 6.9.2.

6.2   Multi-Layer Perceptron (MLP)

A feedforward neural network [174], more commonly known as Multi-Layer Perceptron (MLP), is a simple neural network made of processing units arranged in two or more layers, as shown in Figure 6.5. MLP is one of the primary nonlinear machine learning functions used for both regression and classification tasks in supervised learning. In this network, neurons in one layer are connected to every neuron in the next layer via weighted links, that stores the learnt information. This means that the outputs of all neurons in layer *n-1* form the input vector for neurons in layer *n*. The network is called feedforward since there are no feedback connections where the output is fed back to previous layers i.e., the connections form a directed acyclic graph. The information flows forward through the nonlinear activation function in the hidden layer, until it reaches the top output layer, where the SBP and DBP are estimated simultaneously using a linear activation function. The activation a is first computed as follows:

$$a = \Sigma_i \ \ x_i \ w_i + b \tag{6.4}$$

Where x_i is the input, w_i is the weight associated with each input link and b is the bias. The term a is the sum of all the incoming input connections multiplied by their associated weights w in addition to the bias. This is then passed through the nonlinear activation function F(a) in the hidden layer units to learn a new data representation and pass it on to the next layer. It was empirically found that ReLu leads to better performance compared to sigmoid and tanh. Hence, F(a) in this study is defined by the following equation:

$$F(a) = max \ (0, a) \tag{6.5}$$
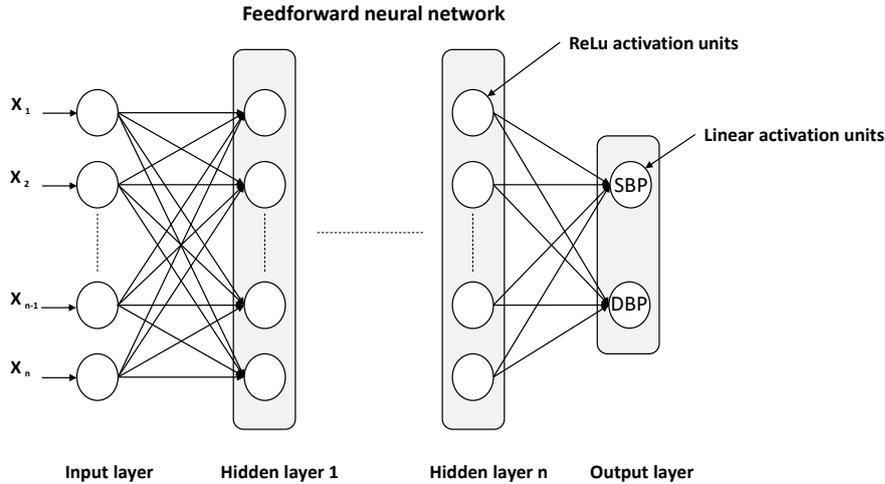
Figure 6.5: Multilayer perceptron with n hidden layers and two output nodes in the output layer. The links between the neurons represent the weights.

6.3 Long Short Term-Memory (LSTM)

Unlike the network described in the previous section, Long Short Term-Memory (LSTM) [175] is a variation of Recurrent Neural Networks (RNNs), a family of neural networks where the connections between the neurons form a directed cycle. This feature enables the network to learn time domain patterns in sequential data. LSTM is the state-of-the-art RNNs, specialised in processing temporal sequential data with the capability of learning long-term dependencies. This network was designed to address the vanishing gradient associated with traditional RNNs. This problem occurs when training a deep neural network or RNN with activation function such as sigmoid or hyperbolic tangent (tanh). Training a neural network involves computing the prediction error at the output layer and estimating a gradient. This gradient error is essentially used by the network to update or optimise the weights at each layer as it propagates backward from the final layer to the input layer. However, with deep neural networks i.e., many layers, the gradient decreases dramatically as it moves backwards through the layers and gets very small as it approaches the initial layers, this results in little or no update to their weights and biases. The LSTM resolves this issue by replacing the conventional tanh RNN hidden unit by a memory cell state C_t that preserves information from previous time steps using three multiplicative units. The memory cell acts as a conveyor belt that passes the information from one-time step to another, while the flow of information is controlled by mul-

tiple gating mechanisms. This allows the LSTM to handle long term dependencies in sequential data more effectively than conventional non-recurrent neural networks and RNNs. The internal LSTM cell is illustrated in Figure 7.6. The LSTM hidden state h_t is calculated by the following equations:

$$f_t = \sigma(W_f \; x_t + U_f \; h_{t-1} + b_f) \tag{6.6}$$

$$i_t = \sigma(W_i \; x_t + U_i \; h_{t-1} + b_i) \tag{6.7}$$

$$o_t = \sigma(W_o \; x_t + U_o \; h_{t-1} + b_o) \tag{6.8}$$

$$\hat{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{6.9}$$

$$c_t = f_t \odot c_{t-1} + \; i_t \odot \hat{c}_t \tag{6.10}$$

$$h_t = \; o_t \odot tanh(c_t) \tag{6.11}$$
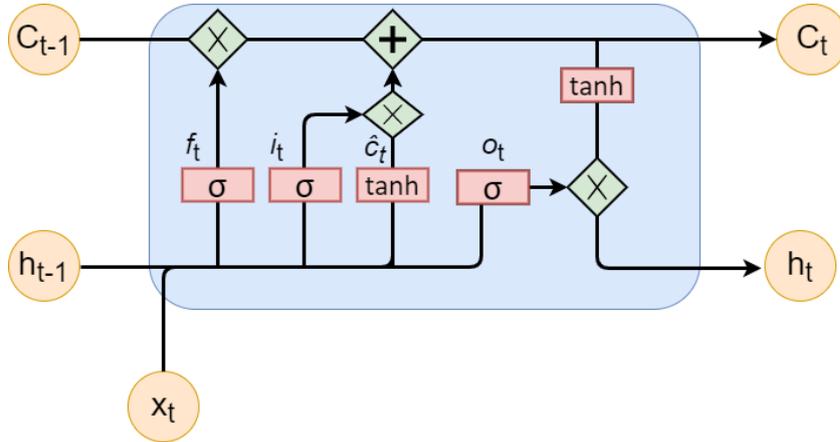
Figure 6.6: Internal LSTM cell structure.

where the f, i and o represent the forget, input and output gates, respectively. The W, U and b terms denote weight matrices for the input, previous hidden states and bias, respectively. σ denotes the logistic sigmoid function, tanh stands for the hyperbolic tangent function and $\odot$ represents the element wise multiplication.

The forget gate controls how much information should be thrown away or forgotten. The input and output gates decide how much information should be accumulated/stored and outputted, respectively. In particular, at time step t, the LSTM takes the current input X_t and the output from the previous hidden state h_{t-1} and pass it through the forget gate f_t, which is a sigmoid function (σ). The sigmoid function outputs values between 0 and 1 for each element of the previous cell state C_{t-1}. The information is kept as the output of f_t gets closer to 1 and removed as the output gets closer to 0. In the next step, the input gate i_t and the new candidate vector $\hat{c}_t$ (output of the tanh layer) are multiplied together to create an update to the cell state C_t. The input gate decides which values are kept from the $\hat{c}_t$. Afterwards, C_t is computed by multiplying C_{t-1} by the forget gate and add it to $i_t * \hat{c}_t$. Finally, the output gate o_t determines what information should be carried to the next hidden state. h_t is calculated by multiplying the output of o_t by $\tanh(C_t)$.

Considering the equations presented, it is noticeable that the LSTM is far more computationally demanding in comparison to MLP or standard RNNs. Nevertheless, LSTMs have shown a significant improvement compared to the previous models. As a result, LSTMs are currently the state of art technique for time series analysis such as BP.

6.4   Gated Recurrent Unit (GRU)

Another variant of the RNN that has been more recently proposed to solve the long-term dependency issue is the Gated Recurrent Unit (GRU) [176], illustrated in Figure 7.7. Similar to the LSTM, the GRU regulates the flow of information using multiples gates embedded in its cell and provides superior performance in comparison to conventional RNNs. However, the GRU has a few advantages over the LSTM. Firstly, the GRU has only two gates, namely, an update gate and a rest gate. Secondly, the GRU cell does not contain a memory cell and uses the hidden state to pass the information instead. Therefore, the GRU is computationally more efficient in terms of required space and time, and as it has less parameters, hence it trains faster than LSTMs. Nonetheless, their simplicity should not be underestimated as

it provides competitive performance in many tasks. The GRU hidden state h_t is generated by the following equations:

$$r_t = \sigma(W_r \ x_t + U_r \ h_{t-1} + b_r \ ) \tag{6.12}$$

$$z_t = \sigma(W_z \ x_t + U_z \ h_{t-1} + b_z \ ) \tag{6.13}$$

$$\hat{h}_t = tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h \tag{6.14}$$

$$h_t = (1 - z_t \ ) \odot h_{t-1} + z_t \odot \hat{h}_t \tag{6.15}$$

Where r and z denote the reset and update gates, respectively. $\hat{h}_t$ is the new candidate activation, σ and tanh stands for sigmoid and hyperbolic tangent functions, respectively. And W, U and b denote the weight matrices and bias.

The GRU works in a similar fashion as the LSTM, where at each time step the GRU cell has two inputs which include the current input X_t and the output of the previous hidden state h_{t-1}. The rest gate r_t decides how much information to forget from the previous hidden state output h_{t-1}. The update gate z_t determines how much the unit updates its activation. z_t primarily combines the role of the input and forget gates of the LSTM which decides what information to discard and what to add to the new hidden state h_t. h_t is a linear interpolation between the previous hidden state h_{t-1} and the new candidate vector $\hat{h}_t$.

6.5   Difference between LSTM and GRU

The GRU and LSTM belong in the same class of neural networks but utilise different gating mechanisms to control flow of information which aims at preventing the vanishing gradient problem. Some of key differences between the LSTM and GRU units are as follow:

- In order to regulate the flow of information, the LSTM utilises three gates, namely, input, output and forget gates. Whereas the GRU utilises two gates, namely, reset and update gates. The input and forget gates of the LSTM are
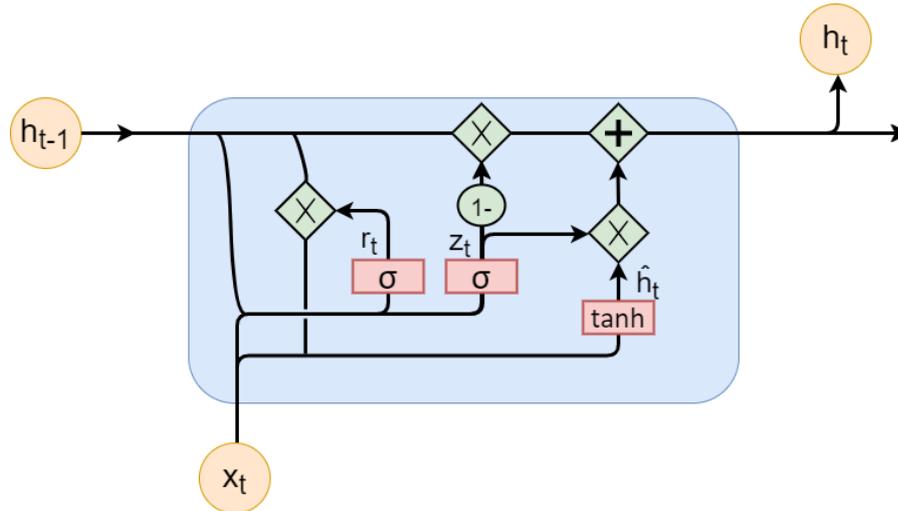
Figure 6.7: Internal GRU cell structure.

merged together to form the update gate in the GRU unit.

- Unlike the LSTM unit, the GRU does not have a memory cell, but instead the GRU exposes the content of the hidden state, in every time step, without any control.

- The GRU is considered relatively new, proposed in 2014, whereas the LSTM was first introduced in 1997. Nonetheless, the GRU provides competitive performance, and its computationally more efficient as it requires less computation (less gates and no memory cell) compared to the LSTM.

- The GRU trains faster and can outperform the LSTM using less training data.

6.6 Bidirectional connections

Traditional unidirectional (layer with forward connection only) LSTM and GRU networks, at their core, try to capture information from the input data that had already been processed through their hidden state, from past history h_{t-1} along with the present input observation x_t. Thus, the unidirectional connection can only preserve past information from the input.

In order to provide additional context to the network and increase its accuracy, one can incorporate past input (history) $x_1, x_2, ..., x_{t-1}$, present input x_t and near future information $x_{t+1}, ..., x_n$ from input sequence where all time steps are available. This
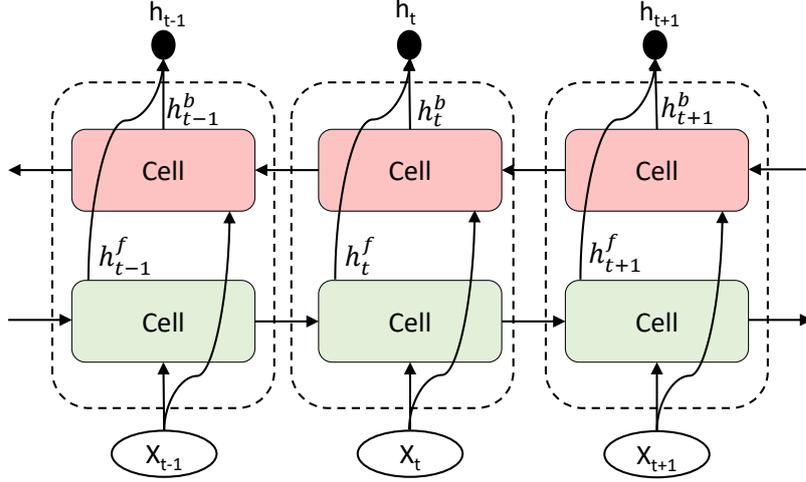
Figure 6.8: Example architecture of a bidirectional layer with three-time steps. The conventional forward layer (green) and backward layer (red) are concatenated in each time step (black dot) to form the final output of a bidirectional layer.

can be achieved through bidirectional RNN (Bi-RNN) connections [177], where the network creates two hidden layers each time a hidden layer is added to it, as opposed to just one unidirectional forward layer in their conventional structures. Each bidirectional layer has a forward and a backward layer. This layer processes the input sequence in a forward direction, using the first layer, and in backward direction using the second layer. The forward layer allows the hidden state to consider past information, while the backward layer allows the hidden state to consider near future information. Thus, having two layers expose the hidden state to more information and in turn this allows the network to improve its ability in handling long-term dependencies and increases its performance. A bidirectional layer containing a forward layer and backward layer is shown in Figure 6.8. On one hand, the forward layer processes the input in a normal time order, and the forward hidden vector h_t^f is calculated in a similar way as described earlier in the conventional unidirectional LSTM and GRU layer. On the other hand, the backward hidden vector h_t^b is obtained by processing the input in reverse order. The final output of the bidirectional hidden vector h_t is a concatenation of the forward hidden sequence h_t^f and the backward hidden sequence h_t^b. h_t is calculated as follows:

$$h_t = W^f h_t^f + W^b h_t^b + b_h \tag{6.16}$$

101

$$h_t^f = f\left(W_h^f x_t + \ U_h^f h_{t-1}^f + \ b_f\right) \tag{6.17}$$

$$h_t^b = f\left(W_h^b x_t + \ U_h^b h_{t+1}^b + \ b_b\right) \tag{6.18}$$

Where f is computed by the equations presented in section 6.3 for the Bi-LSTM layer and the equations presented in section 6.4 for the Bi-GRU layer.

6.7 Attention mechanism

Attention mechanism [178, 179] is a major breakthrough in deep learning, particularly, in natural language processing (NLP). Its application lead to improvements on the encoder-decoder based networks in neural machine translation [178, 180]. A common implementation of an encoder-decoder architecture uses two RNNs [176, 181]. It is known that the RNNs suffer from the vanishing gradient problem associated with the time series tasks. This limitation can be avoided using the LSTM, that is capable of capturing longer range dependencies better than the RNNs. However, it has been observed that in some cases, particularly in very long sequences, the LSTM does not always properly memorise very long interdependencies/correlations [178, 182]. This causes the performance of the encoder-decoder network to degrade over time resulting in what is known as the long-range dependency problem [182]. The attention mechanism alleviates this problem, by retaining all source input hidden states sequence, at every time step, and place more attention/importance to the most relevant parts of the sequence, in order to produce an output.

The application of attention mechanism has been extended to domains other than NLP, such as computer vision [183]. To the best of our knowledge, this is the first time attention mechanism has been applied for cuffless BP estimation using a single PPG sensor. The advantages of attention in this context is that the large number of feature vectors of the RNN or Bi-RNN (thereafter (Bi)-RNN for simplicity) may have different impact on the BP estimation and attention mechanism allows the network to effectively focus on the more important hidden states or feature vectors, automatically, in each time step. Those selected hidden states have been identified as more valuable during the mapping between input sequence and the target output.

Therefore, larger weights are assigned to the most important hidden states that hold the significant information. Furthermore, attention mechanism also provides a reduced sum (weighted combination) of the hidden states dimension. Attention layer reduces the (Bi)-RNN outputs with attention vector. By doing this, it effectively reduces the search space for information by only considering the most effective parts from the sequence.

In [178], attention mechanism works by computing a context vector V_t that preserves information from all (Bi)-RNN output hidden states and calculates their alignment scores with the current target output. V_t is calculated by the following equation:

$$v_t = \sum_{i=1}^{t-1} \alpha_{ti} h_i \tag{6.19}$$

$$\alpha_{t,i} = align(y_t, x_i) = \frac{\exp\left(f\left(h_i, \quad h_t\right)\right)}{\sum_{j=1}^{t-1} \exp\left(f\left(h_j, \quad h_t\right)\right)} \tag{6.20}$$

where $\alpha_{t,i}$ is the weight assigned to the pair of input x at position i and output y at position t. α_{ti} is computed by a softmax function. f is a single layer feedforward neural network, with nonlinear tanh activation function, as implemented in [178]. The output score of function f attempts to capture the relevance/correlation between the source input and target output. The obtained attention weights vector is then passed to the softmax function to normalise its values to a range between (0,1). The final output vector V_t is generated by computing the weighted sum of the attention weight vector $\alpha_{t,i}$ and their hidden state vector h_i.

6.8 Proposed architectures

This study provides cuffless and continuous estimation for SBP and DBP using PPG signals only. The proposed architecture of the deep learning models consist of one bidirectional RNN layer (Bi-RNN) at the bottom, followed by a stack of n unidirectional RNN (uni-RNN) layers and one attention layer at the top layer before the output layer. The proposed architecture is presented in Figure 6.9. The bidirectional layer helps the network to capture more information by processing the input sequence in a forward and backward order. Additionally, in order to solve the vanishing gradient issue that occurs during the training of a deep RNN, the conventional RNN units were replaced by LSTM and GRU units, each tested separately on the proposed architecture. This study attempted to further improve the performance of the model by training the network to focus on the hidden states with significant information using attention mechanism. The SBP and DBP were then estimated using a linear activation function in the output layer. Two deep learning recurrent models were proposed and evaluated in this research using the same neural network architecture, one model is equipped with LSTM cells in its hidden units while the other is equipped with GRU cells in its hidden units. The number of units and the number of stacked unidirectional layers were optimised during training.
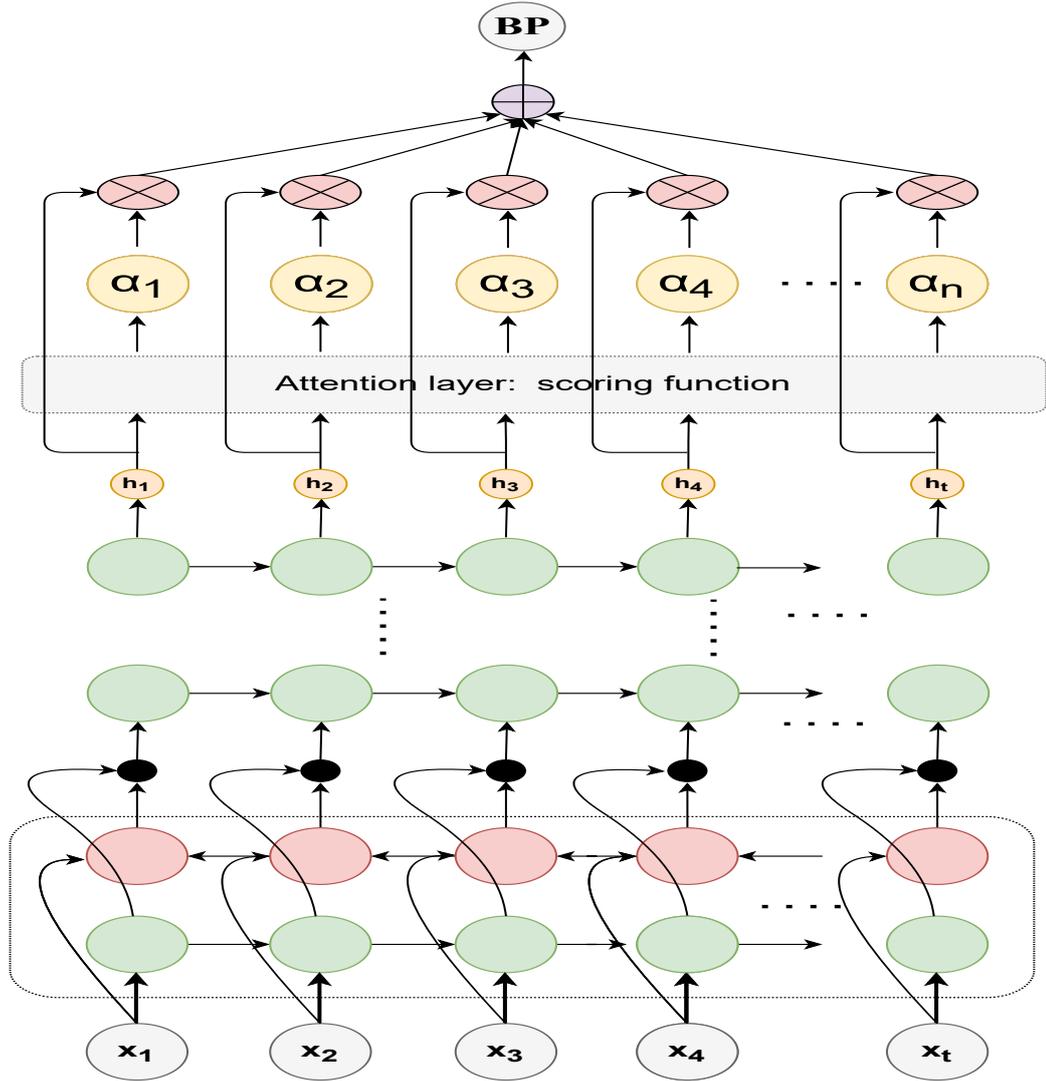
Figure 6.9: The architecture of the proposed models: the dotted block represents the Bi-RNN layer which consists of a forward (green) and backward (red) layer, and the output of the two layers are concatenated together represented by a black dot. This is then followed by one or more unidirectional (forward direction) RNN layers (green). The output of the last RNN hidden layer goes through an attention layer to calculate the context vector and afterwards the SBP and DBP values are calculated using linear activation. The internal RNN cells are replaced by the LSTM and GRU units presented in Figures 7.6,7.7, respectively

6.9   Neural network training

6.9.1   Cost or objective function

Training a neural network is crucial in order to find the optimum set of parameters that can best approximate the target output. This process primarily involves modifying or more accurately optimising the model's weights and biases. To achieve this goal, during training, the model aims to minimise an objective function, also known

as loss or cost function. The values of the cost function indicate how closely is the network able to estimate the target output on a given dataset. Thus, the objective of training a model is to find a particular set of parameters -weights and biases, corresponding to the lowest value (error) of the cost function. Such point is referred to as local or global minima.

In this project, the objective function to minimise during the training stage is the Mean Squared Error (MSE) function. MSE is the most commonly used objective function for regression tasks, and is calculated using the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6.21}$$

where y_i is the ground truth value for the ith observation in the dataset and $\hat{y}_i$ is the estimated value for the ith observation. MSE computes the average squared difference between the ground truth value and the predicted value. When the MSE score decreases, the performance of the model improves, and thus the estimation accuracy increases.

6.9.2 ADAM optimiser

The optimiser is the learning algorithm that is used during the training stage to update and optimise the weights and biases of a neural network model. In this project, the learning algorithm utilised is the Adaptive Moment Estimation (Adam) [173]. This optimiser is computationally and memory efficient, suitable for tasks that deals with large dataset, and simple to implement [173]. It is a stochastic first-order gradient-based optimiser that utilises historical and statistical information of the gradient of the cost function to update the weights and biases of the model. Adam can outperform and converge to a minimum (lowest error value) quicker than popular optimisers such as Stochastic Gradient Descent (SGD) or Root Mean Squared (RMS) optimisers. This is attributed to the fact that Adam computes an adaptive learning rate (i.e., size of the step taken towards finding a minimum), and bias-correction and momentum for each parameter as opposed to SGD and RMS, respectively. [173].

The algorithm utilises the computed gradient of the cost function (g) with respect to model parameters. g is calculated using the following equation:

$$g_t = \nabla_\theta J(\theta_t) \tag{6.22}$$

where ∇_θ is the partial derivative of the cost function $J(\theta)$ parametrised by the model parameters (θ) at time step t. Afterwards, the learner algorithm uses the computed gradient to calculate an exponentially decaying average of the past gradient and the past squared gradient given the following equations:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{6.23}$$

$$v_t = \beta_2 m_{t-1} + (1 - \beta_2) g_t^2 \tag{6.24}$$

Where m_t and v_t are the mean and variance of the gradient, respectively. And β is the decaying rate. The inventors of Adam observed during the initialisation of m_t and v_t, where the values are set to zero, that the optimiser was biased towards zero. Therefore, to overcome this issue, a bias corrected mean and variance were computed as follows:

$$m_t' = \frac{m_t}{1 - \beta_1^t} \tag{6.25}$$

$$v_t' = \frac{v_t}{1 - \beta_2^t} \tag{6.26}$$

The weights and biases of the neural network model are then updated using Adam's update rule:

$$\theta_{t+1} = \theta_t - \frac{n}{\in + \sqrt{v_t'}} m_t' \tag{6.27}$$

where θ and n represent the model parameters that need to be updated and the learning rate, respectively.

6.9.3 Back-propagation

Back-propagation is an elegant method for computing the gradient of the cost function for neural network models [184]. The learning algorithms, such as Adam, Gra-

dient Descent (GD), SGD, RMS optimisers etc., often take advantage of this efficient computational trick to calculate the gradient of the error function, required for the update rule, in order to modify the model's parameters- weights and biases, as described in section 6.9.2 for Adam optimiser's update rule. The significance of back-propagation is that it offers a simple straightforward implementation of the chain rule of derivatives. This essentially allows all partial derivatives, required for the gradient computation, to be computed in linear time in terms of the graph size. The time and memory required for computing the gradient using the traditional method increases exponentially as the size of the graph increases.

Back-propagation works the following way. The initialisation of the neural network starts with random weights and biases. Afterwards, a feedforward pass is performed, where every observation in the dataset is propagated forward through the network, from the input layer towards the output layer. At each layer, every neuron receives input values from the previous layer. For every neuron, an activation value is calculated from the incoming inputs and their associated weights. For the hidden layer neurons, this activation value is fed into a non-linear activation function, such as ReLu, sigmoid, or tanh, etc., to calculate an output value (h) for every neuron. The output is then propagated forwards to neurons of the next layer. This process is repeated, for every observation in the dataset, at every layer, until the final output layer is reached. The final target outputs are calculated in a similar fashion i.e., using the incoming inputs and weights, however, since this is a regression task, the activation function for the output neurons is linear. Afterwards, the estimated output values (SBP and DBP) are observed and compared against the reference (ground-truth) values. The cost function (MSE) calculates the error value between the predicted output (obtained through the forward propagation) and the reference output. The error is then propagated back through every layer of the network. In order to update the weights for every neuron across all layers, the back-propagation algorithm calculates the gradient of the MSE with respect to model parameters. The optimiser then updates the model parameters θ, given the update rule described in section 6.9.2, for every layer until the first layer is reached. This process is repeated until all observations have been processed.

6.9.4 Implementation and optimisation

Unlike the classical models implemented in the previous chapter, neural network models can estimate SBP and DBP in one model structure, hence both objectives were approximated simultaneously. All neural network models described in this chapter were implemented in python using tensorflow. In terms of optimisation, the main parameters that need to be fine-tuned are almost similar for all models, such as batch size, number of epochs, learning rate, number of hidden layers and number of units in each layer. The batch size and number of epochs were fixed to 64 and 300 respectively. As for the remaining parameters, Table 6.1 provides the best combination obtained from the grid search on each dataset (full and reduced feature sets).

Table 6.1: Best parameters for each neural network model on the 52 and 24-feature set

		best parameters	
Models	parameters grid-search	52-feature set	24-feature set
MLP	{# of layers; # of units in each layer; learning rate}	{3; 100,250,500; 0.01}	{3; 100,250,500; 0.01}
LSTM	{# of layers; # of units in each layer; learning rate}	{3; 512,512,512; 0.001}	{4; 512,512,512,512; 0.0001}
GRU	{# of layers; # of units in each layer; learning rate}	{4; 512,512,512,512; 0.001}	{3; 512,512,512; 0.001}
Bi-LSTM + LSTM+ attention	{# of Bi-LSTM units; # of LSTM units in each layer; learning rate}	{512; 512,512,512,512; 0.0001}	{512; 512,512,512,512; 0.0001}
Bi-GRU + GRU + attention	{# of Bi-GRU units; # of GRU units in each layer; learning rate}	{512; 512; 0.001}	{512; 512,512; 0.001}

6.10 Summary

Similar to classical machine learning models, there are different types of neural networks that were established to handle various data types (sequential, images, text, etc) and regression as well as classification problems. In this study, the PPG signal is a time series signal, hence, it requires a special type of neural networks, namely, recurrent neural networks. However, traditional RNNs suffer from the vanishing or exploding gradient. This can be avoided using the LSTM and GRU units instead of the tanh unit found in the traditional RNNs. Furthermore, this study implements bidirectional connections in order to allow the network to access more information. Additionally, the attention mechanism forces the network to focus its search on the most relevant units for the BP estimation. The next chapter assesses the performance of all the aforementioned models and compares them against the performance of the classical machine learning models.

Chapter 7

Performance Metrics and Evaluation of Machine Learning and Neural Network Algorithms on the MIMIC II

This chapter introduces the evaluation metrics used for assessing the estimation performance for all the classical machine learning and neural network models described in chapters 5 and 6. It starts by presenting the most common statistical metrics used in the literature, namely, mean absolute error and standards deviation, as well the international standard metric followed in this study. It also presents various graphical performance evaluation tools for further analysing the estimation precision. The results for all models, on both datasets, are then presented and compared against each other.

7.1 Data partitioning

Data partitioning is a common technique in supervised machine learning used for training, optimising and validating a predictive model. There are several ways to partition the dataset into smaller subsets, such as, splitting the data into train and test or train, validation and test or cross validation and test or leave-one-out and test, etc. Selecting the right partitioning option mainly depends on the size of the available dataset. For example, partitioning into train and test sets is the best prac-

tice when the dataset is extremely large. Conversely, leave-one-out or 10-fold cross validation are better options when the data is scarce or relatively small. However, the common denominator between all these partitioning techniques is the reservation of a small separate subset of the dataset out of the analysis, known as a test set, for validating the final optimised predictive model on unseen data. Thus, partitioning the dataset using any of the aforementioned techniques serves the same purpose, that is avoiding overfitting or underfitting the data during training, and subsequently examining the validity or generalisation of the model on unseen data. In this research project, the datasets were divided into 70% train, 15% validation and 15% test sets. The training set was used for optimising the model parameters. While the test set was reserved for the final evaluation of the optimised model and remained completely disjoint from the training data. The model selection criterion was based on the lowest error on the validation set.

7.2 Performance metrics

Several evaluation metrics were used to assess the estimation accuracy and the generalisation of the developed models. These metrics can be divided into descriptive statistics and graphical analytical representations. The former comprises the most common evaluation metrics in the literature, such as, mean absolute error, mean error and their standard deviation. While the latter consists of popular graphical evaluation methods, such as Bland-Altman plots, regression plots and histogram of error distribution. These methods are described in more detail in the following subsections. However, taking into consideration the trade-off between complexity and accuracy, as well as the fact that the results obtained from the 52 feature set and the reduced feature set are comparable with insignificant difference (as seen in the following sections), the graphical analytical tools will only be applied on the reduced feature set for further analysis.

7.2.1 Mean absolute error

Mean absolute error (MAE) is a widely used BP estimation metric in the literature, hence, it has been selected in order to allow fair comparisons to be made between the results obtained in this research against other works in this field. The MAE is the

average of all absolute differences between the estimated BP values and the ground truth (reference) BP values, in millimetre of mercury (mmHg). MAE is calculated using the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{7.1}$$

where n is the total number of samples, y_i and $\hat{y}_i$ are the reference and estimated BP values for the ith observation, respectively.

7.2.2 Standard deviation

The standard deviation (SD) is a statistical measure of the variation of a set of values relative to their mean value. In this research, SD is the measure of the deviation of the error (MAE and ME) relative to the mean error. SD is calculated as the square root of the variance using the following formula:

$$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{7.2}$$

where x_i and $\overline{x}$ are the error of the ith observation and mean error, respectively.

7.2.3 Cuffless BP global standard- AAMI

The American National Standards of the Association for the Advancement of Medical Instrumentation (AAMI) sets the global standard of acceptable non-invasive cuffless BP using the mean error difference (ME) and SD metric. According to the AAMI criterion, the ME and SD of the non-invasive cuffless BP measurement, evaluated on at least 85 subjects, should be less than or equal to 5 mmHg and 8 mmHg, respectively. The ME is calculated as follows:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{7.3}$$

7.2.4 Bland-Altman Plot

Bland Altman plot [185] is a useful analytical tool for evaluating the relationship between a pair of variables. Particularly, it displays the difference between a pair of variables (i.e., difference between estimated BP and reference BP) over the mean of all pair of variables, and constructs ± 1.96SD lines parallel to the mean, corresponding to the limit of agreements interval, within which 95% of the differences between the estimated and reference values fall. Thus, providing a simple graphical representation for visually assessing mean differences as well as evaluating the limits of agreement (confidence interval) between the estimates of the predictive model and the reference method. An example of a Bland Altman plot is shown in Figure 7.1. The x-axis represents the mean of the pair of variables, while the y-axis corresponds to the difference between these pairs. The two dashed lines show the constructed 95% confidence interval.
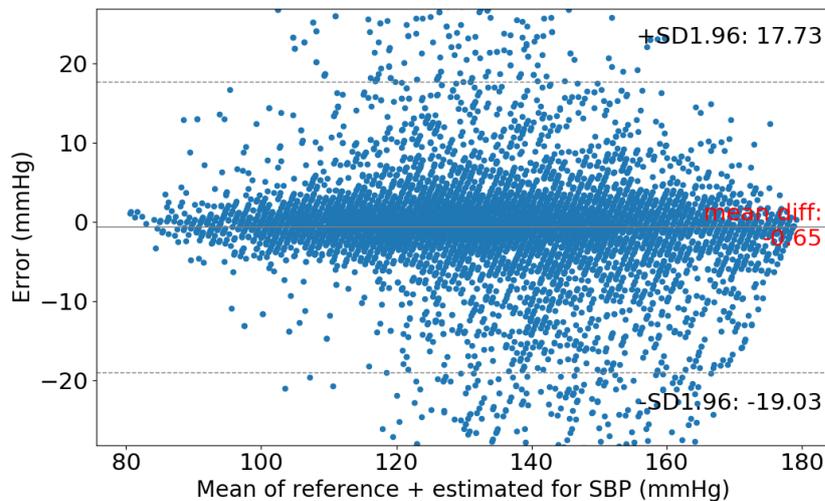
Figure 7.1: Graphical representation showing an example of Bland-Altman plot. The line in the centre represents the mean difference between the estimated and reference values, while the two parallel dashed lines represent the confidence interval or limits of agreement at ± 1.96SD from the mean difference

7.2.5 Regression Plot

Regression plot is another way of assessing the strength of the relationship between the estimated BP (independent variable) and reference BP (dependent variable).

These plots can also help determine how closely the model was able to estimate the target output, and subsequently determine consistency and goodness fit of the model. For example, Figure 7.2 shows a regression plot between the reference SBP and estimated SBP. The figure clearly shows a strong positive linear correlation between the two variables, which is quantified by the R-value (Pearson's Correlation coefficient). Hence, in this case, the plot shows that the model was well fitted on the data.
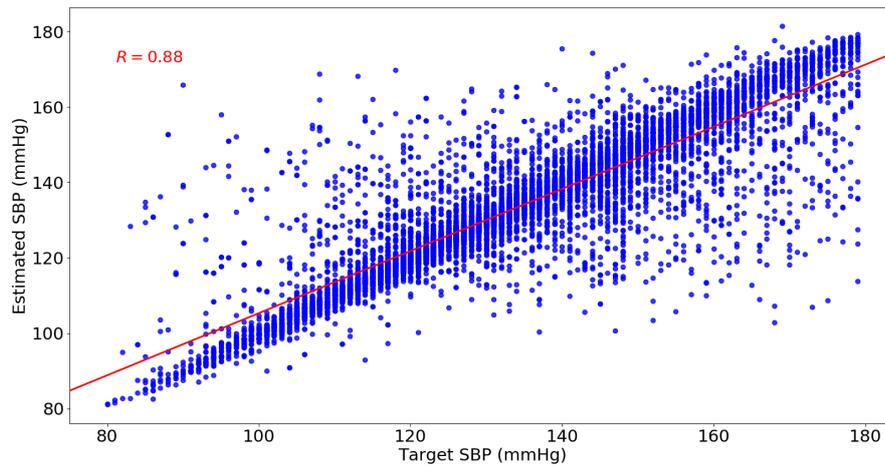
Figure 7.2: Graphical representation showing an example of a regression plot between the estimated and reference SBP with a linear regression (red) line. R represents the Pearson's correlation coefficient between the two variables.

7.3 Results from MLR

7.3.1 Evaluation using MAE, SD and AAMI

Table 7.1 presents the results obtained from the MLR model on the 52-feature set (listed in Table 4.2) and the reduced feature set (presented in Table 4.5) using the MAE±SD mmHg and AAMI (ME±SD mmHg). It can be seen from Table 7.1 that the overall performance of the model is very poor. Particularly, the MAE±SD for the SBP and DBP estimation on the 52-feature set are 14.86±10.88 mmHg and 7.14±6.3 mmHg, respectively. Whereas the MAE±SD obtained from the reduced feature set is 15.11±10.95 mmHg for SBP and 7.42±6.6 mmHg for DBP. Thus, it can be seen that there is an insignificant change in the overall performance of the model on the reduced feature set compared to the performance on the full 52 feature set. Furthermore, the results signify that the MLR is incapable of accurately capturing

or modelling the relationship between the extracted PPG features and the target BP variables, which suggests that there is an inherently complex nonlinear relationship between the input variables and the dependent variables in the dataset. This is further supported by the performance evaluation using the AAMI metric. Table 7.1 shows that the ME±SD for SBP and DBP failed to meet the AAMI requirements on both 52 feature set as well as the reduced feature set. For example, the SD for the SBP vastly exceeded the ±8 mmHg limit set by the AAMI. High SD indicates high variability in the estimation, and in turn low precision. On the other hand, DBP estimation was slightly better, as indicated by the lower MAE±SD and ME±SD in comparison to SBP. This is due to the fact that the range of DBP values is much smaller compared to the SBP values in the dataset. Nonetheless, the SD for both SBP and DBP were higher than the acceptable range set by the AAMI standard. The following subsections will further analyse the results obtained on the reduced feature set.

Table 7.1: Results of the multi-linear regression (MLR) model on the 52 and reduced-feature set using MAE±SD mmHg and AAMI (ME±SD mmHg)

Model: MLR	SBP				DBP			
Results	MAE	±SD	ME	±SD	MAE	±SD	ME	±SD
52-features	14.86	10.88	-0.24	18.42	7.14	6.3	0.19	9.52
reduced set	15.11	10.95	-0.02	18.66	7.42	6.6	-0.08	9.93

Figure 7.3 present the histograms of the distribution of the estimation error for the SBP and DBP values using the MLR model on the reduced feature set. It can be seen from the histograms that the error is centered around zero but distributed over a large range of roughly (-50,+50) and (-40,+20) for SBP and DBP, respectively. Moreover, the SBP error is normally distributed, while the DBP is left skewed. This means that higher DBP values were underestimated by the MLR model. Overall, given that the error range is too large, it can be said that the MLR was unable to predict the SBP and DBP values correctly.
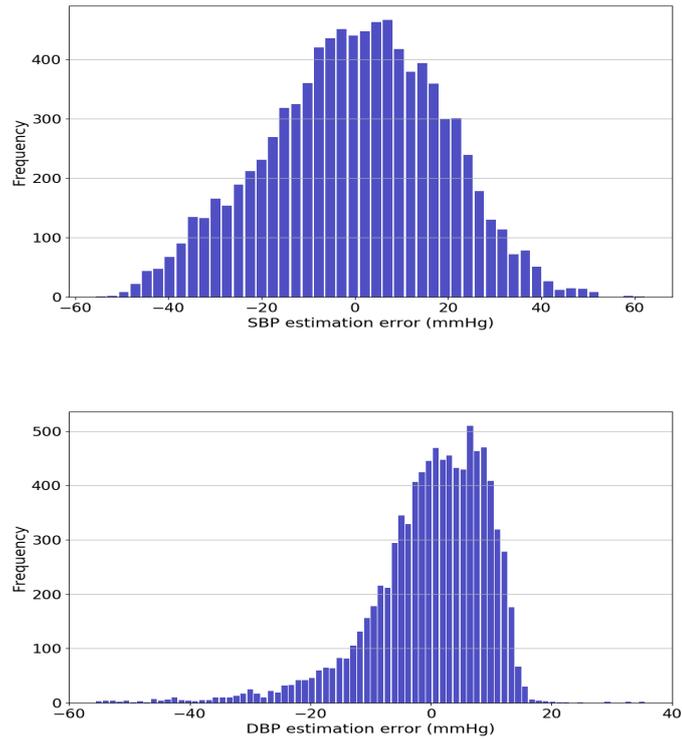
Figure 7.3: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the multilinear regression model.

7.3.2 Evaluation using Bland-Altman plots

Bland-Altman plots is another visual analytical tool used for assessing the performance of the MLR model. Ideally, in case of high estimation accuracy, the data points should be spread across the whole range of target values in the dataset on the x-axis with a narrow error range on the y-axis. However, the Bland-Altman plots in Figure 7.4 shows the opposite scenario for the MLR model. First, it can be seen that the range of target values on the x-axis is limited to roughly (110,170) mmHg for SBP and (65,90) mmHg for DBP. Thus, the MLR was unable to estimate low SBP and high DBP values. Furthermore, the data points are spread diagonally over a large y-axis range, which indicates a high variance in the estimation error. In particular, the upper and lower limits of agreements are (-36.6,36.55) for SBP and (-19.55,19.38) for DBP, respectively. The limits of agreements are beyond the acceptable AAMI limits, which in this case are, $\pm 8*1.96$ (± 15.68 mmHg). However, the mean difference for both SBP and DBP are within the AAMI limits i.e. below 5 mmHg.
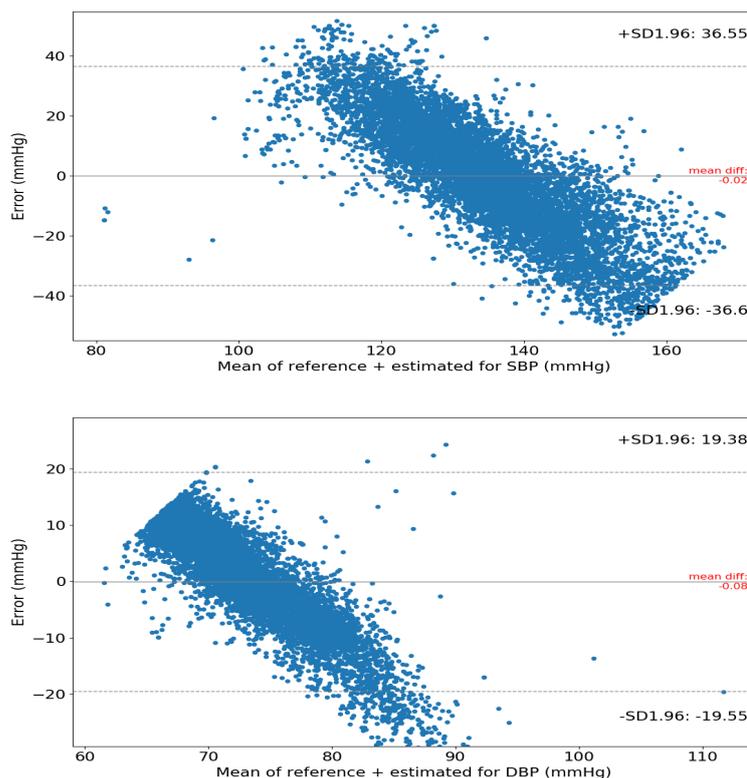
Figure 7.4: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the multilinear regression model.

7.3.3 Evaluation using Regression plots

Figure 7.5 presents the regression plots between the reference and estimated values for both SBP and DBP using the MLR model. These figures clearly show a very weak correlation between the estimated and reference values for SBP and DBP. This is evident by the almost horizontal linear regression line in both figures, in addition to the very low correlation coefficient R of 0.34 and 0.26 for SBP and DBP, respectively. Thus, given all the previously mentioned indicators and analysis, it can be said that the MLR model is not a suitable estimator for continuous non-invasive and cuffless BP.

7.4 Results from SVR

7.4.1 Evaluation using MAE, SD and AAMI

The results obtained using the SVR model on both 52-feature set and reduced feature set are presented in Table 7.2. At first glance, it can be seen from this table that there is a slight improvement over the results obtained using the MLR model
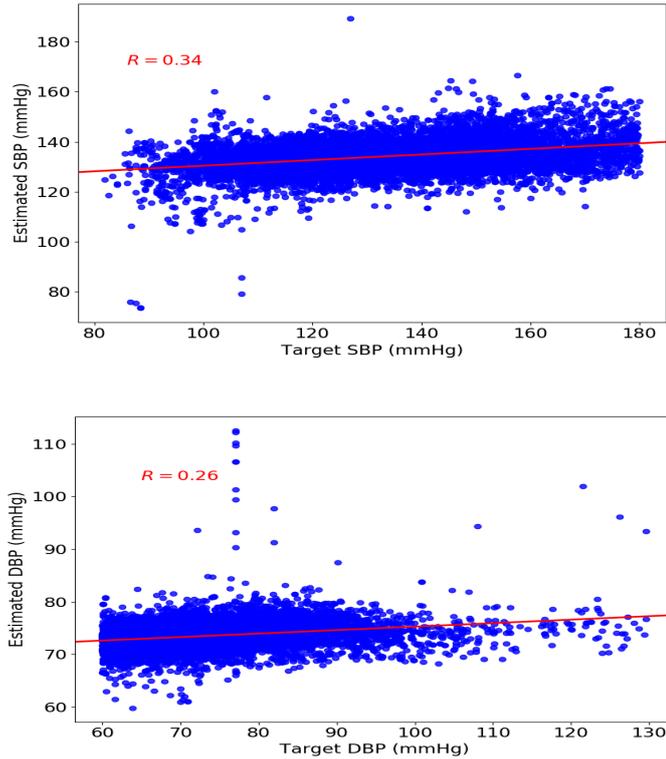
Figure 7.5: Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the multilinear regression model.

in Table 7.1. Table 7.2 shows a reduction of roughly 2 mmHg in MAE for SBP and a reduction of around 1.3 mmHg in MAE for DBP, on both dataset, compared to the results presented in Table 7.1. However, overall the MAE and SD for both SBP and DBP are still very high. In particular, the MAE±SD on the 52-feature set is 12.49±10.99 mmHg for SBP and 5.82±6.24 mmHg for DBP. Furthermore, the results on the reduced feature set are also within the same range with negligible difference, with 12.95±11 mmHg and 6.11±6.7 mmHg for SBP and DBP, respectively. Thus, the estimation accuracy for the SVR is low, especially for the SBP due to the larger range of values in the datasets. In terms of ME and SD, both SBP and DBP did not meet the AAMI requirements on all datasets. Table 7.2 shows that the SD of the ME for the SBP is twice as high as the AAMI limit, while the SD of ME for the DBP fell short by less than 1 mmHg of the acceptable limit.

The distribution of the estimation error for SBP and DBP on the reduced feature set are presented in Figure 7.6. Similar to the MLR model in the previous section, it can be seen that the error is distributed around zero with a relatively large range of (-60,60) and (-40,20) for SBP and DBP, respectively. Additionally, the error

Table 7.2: Results of the SVR model on the 52 and reduced-feature set using MAE±SD mmHg and AAMI (ME±SD mmHg)

Model: SVR	SBP				DBP			
Results	MAE	±SD	ME	±SD	MAE	±SD	ME	±SD
52-feature set	12.49	10.99	-0.53	16.63	5.82	6.24	-1.07	8.46
reduced set	12.95	11	-0.31	16.99	6.11	6.7	-1.35	8.97

for the SBP is normally distributed, while the estimation error for the DBP is left skewed (i.e., SVR underestimated high DBP values). Therefore, these histograms show inconsistency and low estimation accuracy.
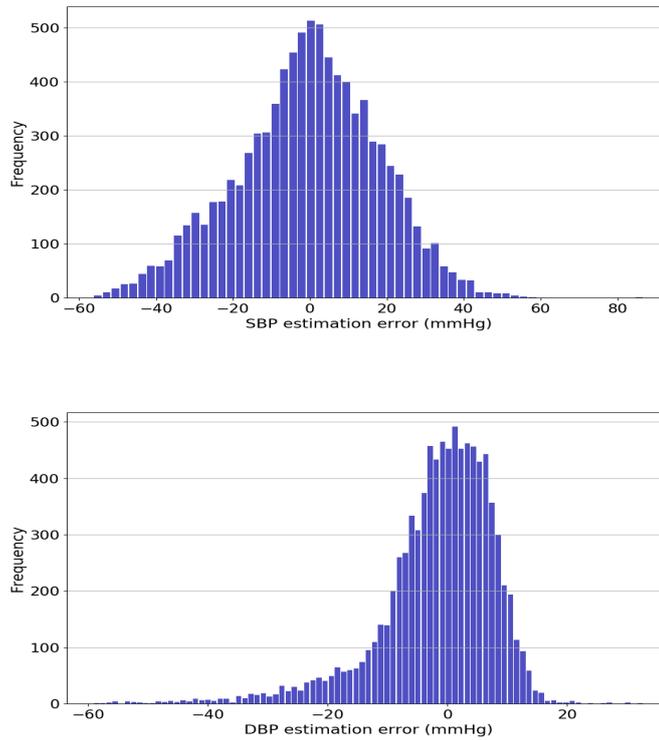
Figure 7.6: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the SVR model.

7.4.2 Evaluation using Bland Altman plots

The Bland-Altman plots between the estimated and reference values for both SBP and DBP are shown in Figure 7.7. These plots show that the data points are not spread across the whole range of SBP and DBP values in the dataset but instead are spread diagonally over a wide range of y-axis (representing the error range). This further proves that the SVR failed to accurately estimate the target values in the dataset. The mean difference for SBP and DBP are -0.5 mmHg and -1.58

mmHg, respectively. Whereas the limits of agreement are (-36.45,35.44) for SBP and (-20.6,17.44) for DBP. Similar to the MLR model, the limits of agreement exceeded the 15.68 mmHg (8x1.96) limits of the AAMI standards. This confirms the low estimation precision of the SVR on this dataset.
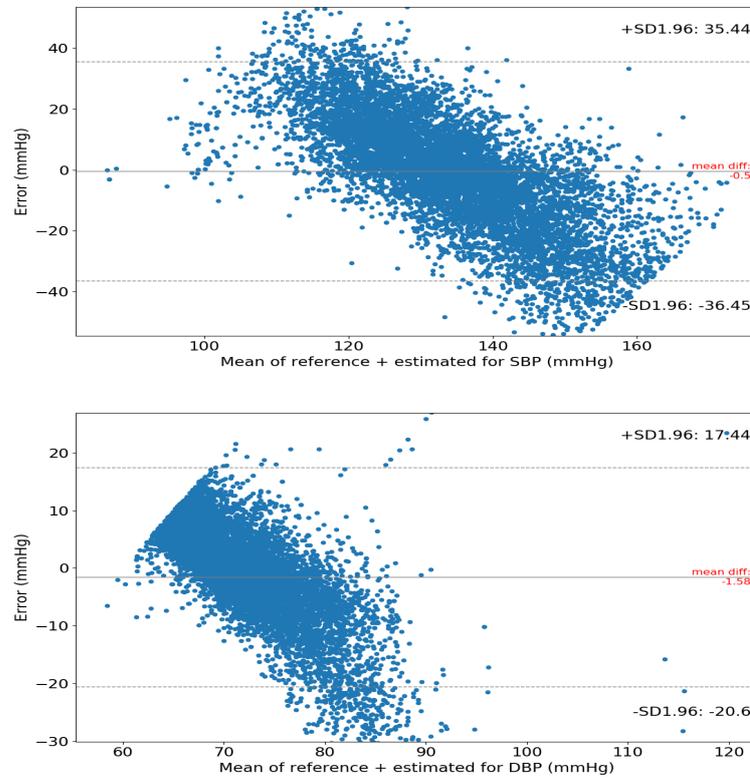

Figure 7.7: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the SVR model.

7.4.3 Evaluation using Regression plots

The relationship between the estimated and reference values for both SBP and DBP using the SVR model are illustrated in Figure 7.8. The regression plots show a very weak correlation between the estimated and reference values. This is depicted by the horizontal linear regression line and the low correlation coefficient. The Pearson's correlation coefficient R is 0.34 and 0.26 for SBP and DBP, respectively. Thus, the estimated SBP and DBP values are not in accordance with reference target values in the dataset. Therefore, the performance of the SVR model is poor, and subsequently it is not suitable for non-invasive, cuffless and continuous BP estimation.
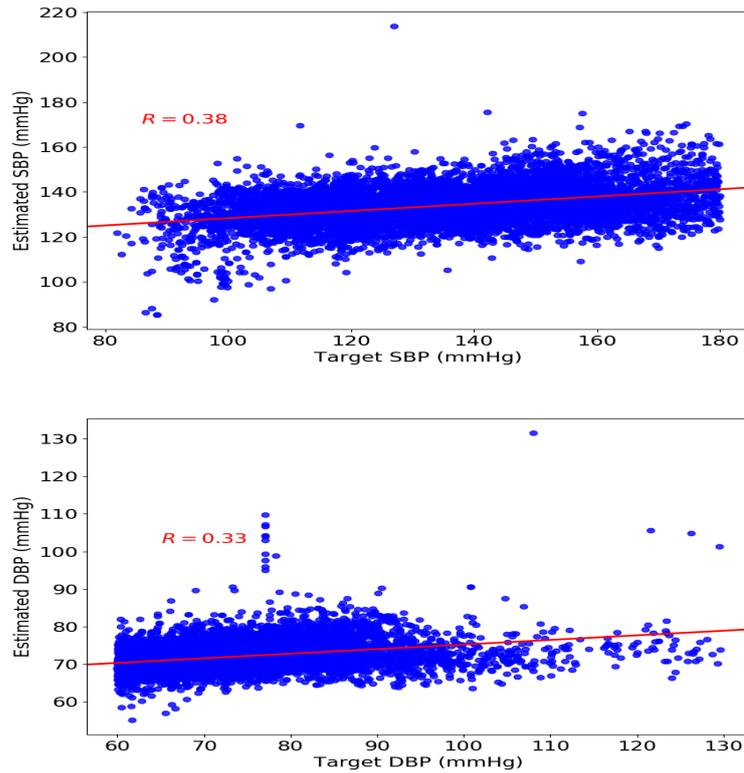
Figure 7.8: Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the SVR model.

7.5 Results from Random forest

7.5.1 Evaluation using MAE, SD and AAMI

The results obtained using the random forest (RF) model, on both datasets, are presented in Table 7.3. It can be seen that there is a noticeable improvement in comparison to the results obtained from the MLR model in Table 7.1 and the SVR model in Table 7.2. For example, the MAE for SBP is reduced by more than 3 mmHg and 1 mmHg compared to the MLR and SVR models, respectively. The MAE±SD for SBP estimated by the RF model is 11.36±9.5 mmHg on the 52 feature set and 11.47±9.4 mmHg on the reduced feature set. On the other hand, the MAE±SD for the DBP is 5.5±5.26 mmHg and 5.58±5.7 mmHg on the 52 feature set and reduced feature set, respectively. Again, similar to the previously mentioned models, there is a negligible difference between the results obtained on the reduced feature set and the 52 feature set. Nonetheless, overall the MAE and SD for both SBP and DBP are high, even with the performance improvement compared to the last two models. As for the performance evaluation using the AAMI standard, the SBP failed to meet the requirements on both dataset, while the ME±SD for the DBP where within the

acceptable limits (below 5±8 mmHg) on both full and reduced feature sets.

Table 7.3: Results of the random forest (RF) model on the 52 and reduced-feature set using MAE±SD mmHg and AAMI (ME±SD mmHg)

Model: RF	SBP				DBP			
Results	MAE	±SD	ME	±SD	MAE	±SD	ME	±SD
52-features	11.36	9.5	-0.2	14.8	5.5	5.26	0.12	7.61
reduced set	11.47	9.49	-0.09	14.89	5.58	5.7	-0.1	7.98

Figure 7.9 shows the estimation error distribution for SBP and DBP on the reduced feature set. The histogram plot for the SBP estimation error is normally distributed around zero over a range of (-40,40) mmHg. As for the DBP estimation error, the histogram shows that the error is almost normally distributed around zero with a range of (-20,20). Thus, compared to the previous model, the plots in Figure 7.9 have a smaller error range for both SBP and DBP. However, these large error ranges illustrate the low estimation precision of the RF model on this dataset.
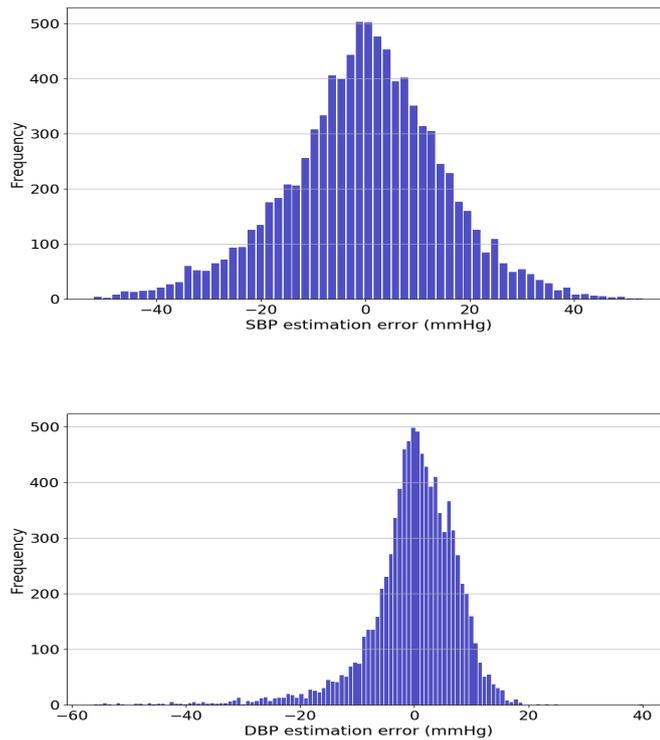

Figure 7.9: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the random forest model.

7.5.2 Evaluation using Bland Altman plots

The plots in Figure 7.10 present the Bland-Altman graphs between the reference and estimated SBP and DBP values using the RF model. These plots do not show a significant difference compared with the Bland-Altman plots presented in the previous sections. The mean difference between the estimated and reference SBP values is -0.09 mmHg, and lower and upper limits of agreements are (-29.27,29.09), where 95% of the error falls. Furthermore, the mean difference for the DBP is -0.1 mmHg and at 95% confidence interval, the limits of agreements were (-15.74,15.54). This further proves that the DBP estimation were within the acceptable limits of the AAMI standards. However, these figures show high variance for both SBP and DBP estimation, and subsequently, the overall performance of the RF model is weak.
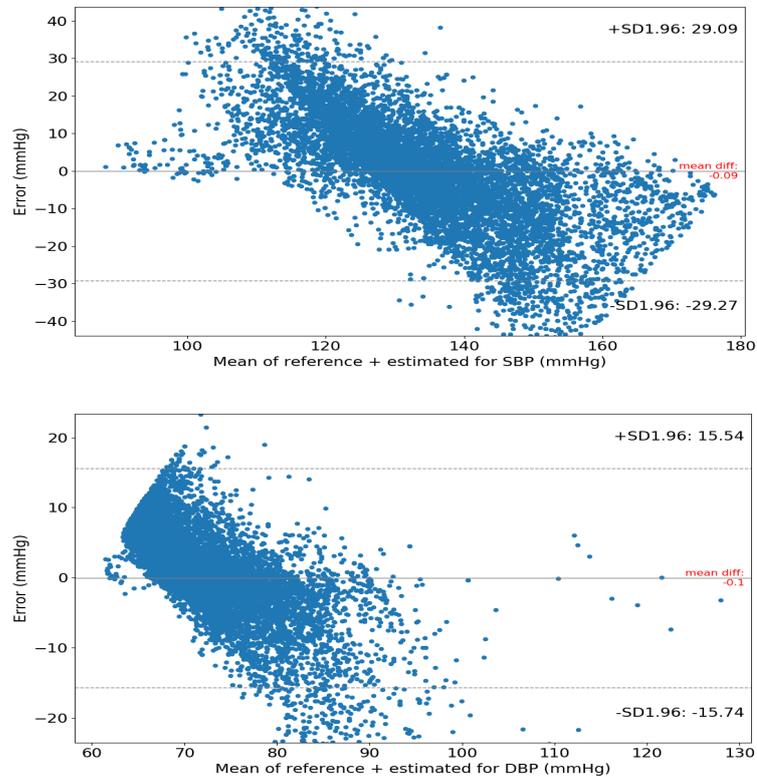

Figure 7.10: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the random forest model.

7.5.3 Evaluation using Regression plots

The regression plots for SBP and DBP using the RF model are presented in Figure 7.11. These plots show a better correlation between the reference and estimated values in comparison to the previous two models (MLR and SVR). In particular, these

plots illustrate a slightly stronger positive correlation, which is further supported by the higher Pearson's correlation coefficient R (almost double that of the previous models). The R value for the SBP is 0.68 and 0.65 for the DBP. Though the R value is showing more strength compared the MLR and SVR model, the R-squared (explained variance) for both targets are still very low. Therefore, given all the previously mentioned evaluation metrics, the RF model did not accurately capture the relationship between the independent variables and the target variables in the dataset. Thus, this model is not reliable for cuffless and continuous BP estimation.
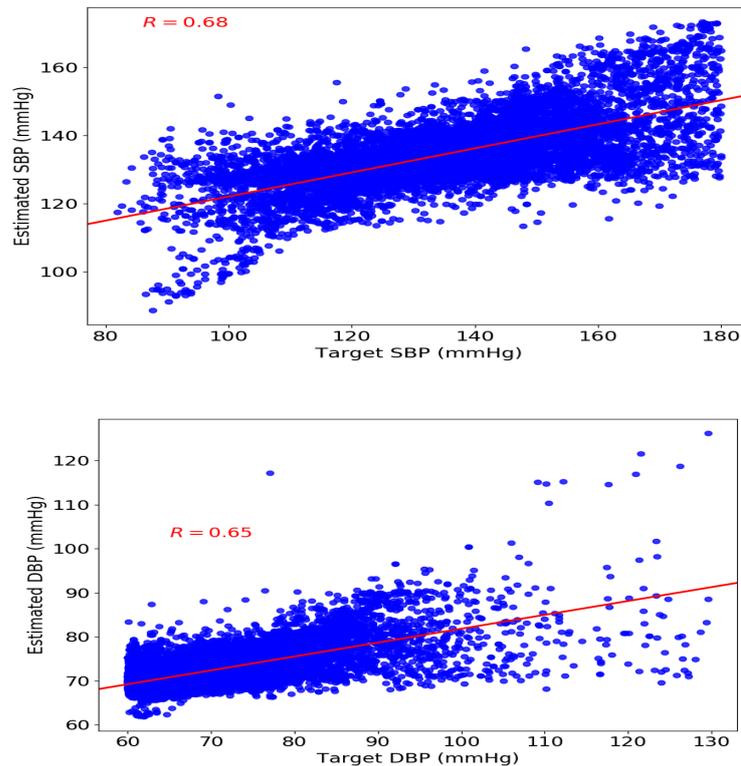
Figure 7.11: Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the random forest model.

7.6 Results from AdaBoost

7.6.1 Evaluation using MAE, SD and AAMI

The performance evaluation of the AdaBoost model, for SBP and DBP, using the MAE and SD, as well as the AAMI metric are presented in Table 7.4. It can be seen from this table that the results obtained using the AdaBoost model are the worst compared to the results obtained using the models mentioned in the previous sections, including the MLR model. On the 52 feature set, the MAE±SD for the

SBP and DBP are 15.15±10.68 mmHg and 7.4±6.15 mmHg, respectively. Additionally, the MAE±SD results obtained on the reduced set are 15.18±10.69 mmHg for SBP and 7.57±6.4 mmHg for DBP. Hence, the performance is poor and the model was unable to estimate the target outputs accurately. Furthermore, the ME and SD did not statisfy the AAMI requirements for both SBP and DBP on the 52 and reduced feature sets. Thus, given the weak performance and low estimation precision, AdaBoost is not a suitable estimator for cuffless and continuous BP monitoring.

Table 7.4: Results of the AdaBoost model on the 52 and reduced-feature set using MAE±SD mmHg and AAMI (ME±SD mmHg)

Model: AdaBoost	SBP				DBP			
Results	MAE	±SD	ME	±SD	MAE	±SD	ME	±SD
52-features	15.15	10.68	-0.17	18.54	7.4	6.15	0.46	9.61
reduced set	15.18	10.69	-0.06	18.56	7.57	6.4	0.74	9.88

The Histograms in Figure 7.12 present the estimation error distribution for SBP and DBP on the reduced feature set. Similar to previous models, the error is normally distributed around zero with a range of (-45,45) for SBP, while the distribution for the DBP is left skewed with a range of (-40,20). The wide error further confirms the low estimation precision obtained using the AdaBoost model.
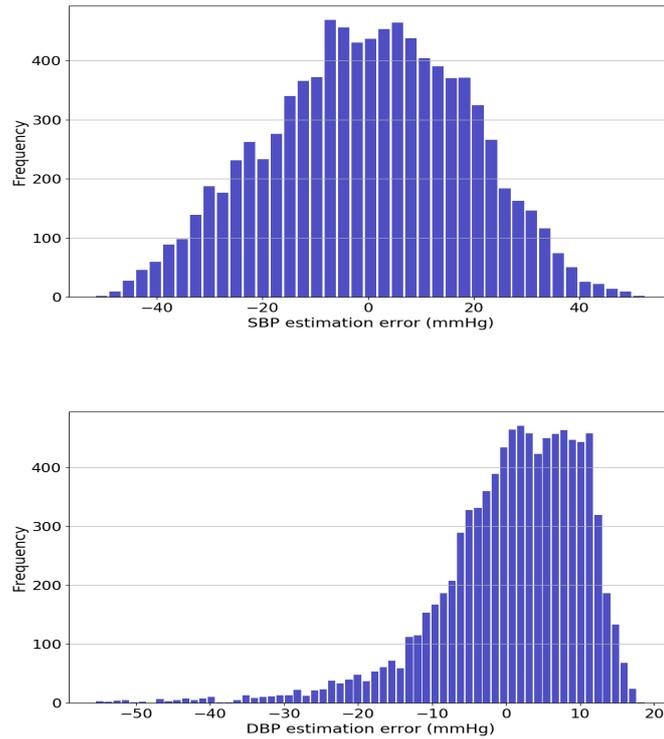
Figure 7.12: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the AdaBoost model.

7.6.2 Evaluation using Bland Altman plots

The performance evaluation using the Bland-Altman plots between the estimated and reference values on the reduced feature set are shown in Figure 7.13. The mean difference and limits of agreements for the SBP are -0.06 mmHg and (-36.44,36.32), respectively. While the mean difference and limits of agreements for the DBP are 0.74 mmHg and (-18.63,20.11), respectively. These plots illustrate a narrower range of values on the x-axis and a wider range of error compared to the previous models described in this chapter.
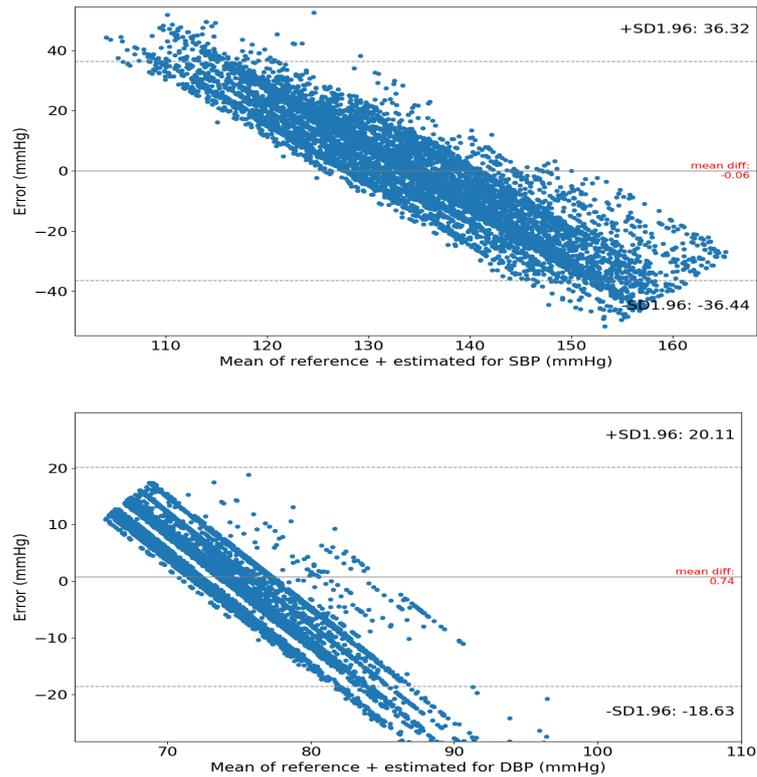
Figure 7.13: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the AdaBoost model.

7.6.3 Evaluation using Regression plots

The regression plots between the estimated and reference values on the reduced dataset are presented in Figure 7.14. These plots clearly show a very weak correlation between the values estimated by the AdaBoost model and the reference target values. Furthermore, it can be seen that the predictions are not in-line with the reference values. This claim is supported by the low Pearson's correlation coefficient R of 0.39 for SBP and 0.29 for the DBP. Thus, the AdaBoost is not a suitable estimator for the cuffless and continuous BP monitoring.
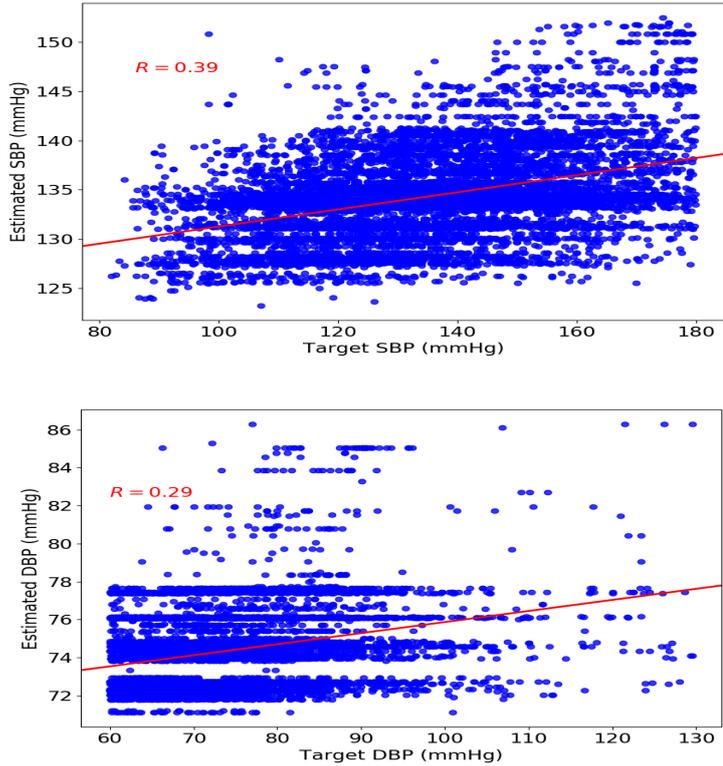
Figure 7.14: Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the AdaBoost model.

7.7 Results from multilayer perceptron (MLP)

7.7.1 Evaluation using MAE, SD and AAMI

Table 7.5 presents the results obtained using the MLP neural network on the 52 feature set and the reduced feature set. It is obvious from this table that the MLP provided a superior performance in comparison to the classical machine learning models. This is noticeable by the considerable reduction of error in Table 7.5 compared to the results produced by the MLR in Table 7.1, SVR in Table 7.2, RF in Table 7.3, and particularly the AdaBoost in Table 7.4. The MAE±SD obtained by the MLP for the SBP and DBP on the 52 feature set are 9.16±8.91 mmHg and 5.26±5.39 mmHg, respectively. Additionally, almost the exact same results were obtained on the reduced feature set, where the MAE±SD for the SBP is 9.21±8.84 mmHg and 5.16±5.41 mmHg for DBP. Though the results on both datasets were similar, the complexity is significantly reduced when the input feature dimension is cut by more than half. In terms of performance evaluation against the AAMI metric, the DBP estimation were within the acceptable limits, while the SBP did not

satisfy the requirements. Consequently, given the relatively low accuracy on both datasets, and the fact that the MLP cannot fully capture the relationship between the temporal input features and the target outputs, this model is deemed unreliable for continuous BP estimation.

Table 7.5: Results of the multi-layer perceptron (MLP) model on the 52 and 24-feature set using MAE±SD mmHg and AAMI (ME±SD mmHg)

Model: MLP	SBP				DBP			
Results	MAE	±SD	ME	±SD	MAE	±SD	ME	±SD
52-features	9.16	8.91	-0.11	12.78	5.26	5.39	-1.05	7.45
24-features	9.21	8.84	0.33	12.76	5.16	5.41	-0.19	7.47

The histogram plots in Figure 7.15 present the estimation error distribution for the SBP and DBP on the 24 feature set. These plots show a slight improvement compared to previous models, as illustrated by the narrower range of error and normal distribution, as opposed to the left skewed distribution for the DBP and wider error range in the previous cases. The estimation error is centered around zero mainly between the range of (-25,25) and (-20,20) for SBP and DBP, respectively.
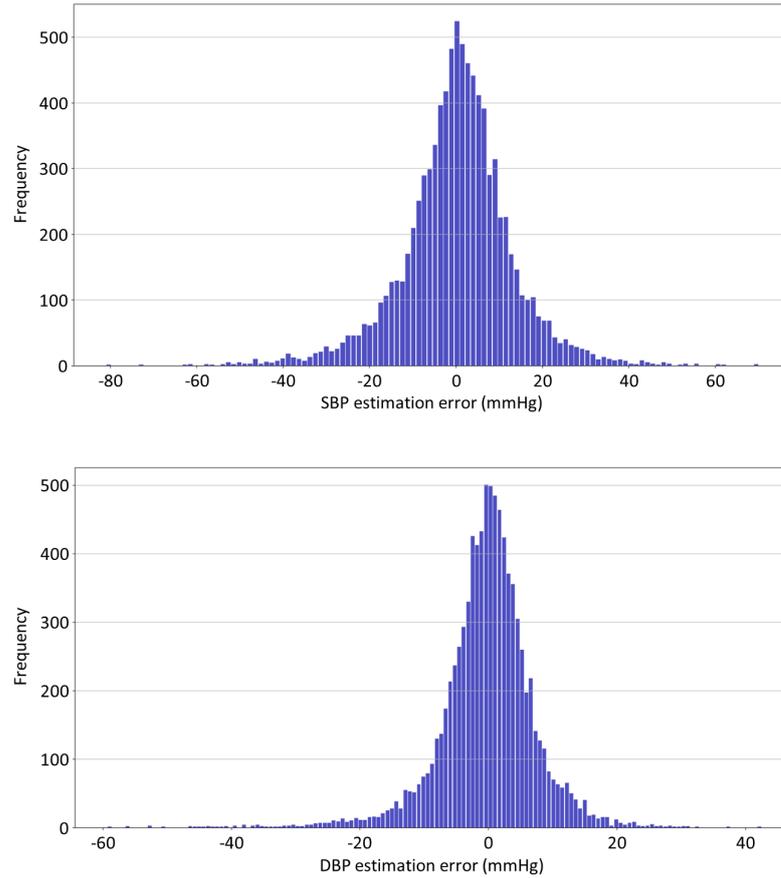
Figure 7.15: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the MLP network.

7.7.2 Evaluation using Bland Altman plots

The evaluation between the estimated and reference values using the Bland-Altman plots are shown in Figure 7.16. The plots show two main differences in comparison to the classical machine learning models: 1) the upper and lower limits of agreements for both the SBP and DBP are smaller, and 2) the data points are spread across the entire range of target values in the dataset i.e. (80,180) mmHg for SBP and (60,130) mmHg for DBP. The mean difference between the estimated and reference values are 0.33 for SBP and -0.19 for DBP. At 95% confidence, the limits of agreements for SBP is (-24.7,25.35) and (-14.8,14.46) for DBP. Thus, as illustrated by these plots, the ME for both SBP and DBP were within the required limits of the AAMI standards, however, only the SD of error for estimated DBP was within the acceptable limits.
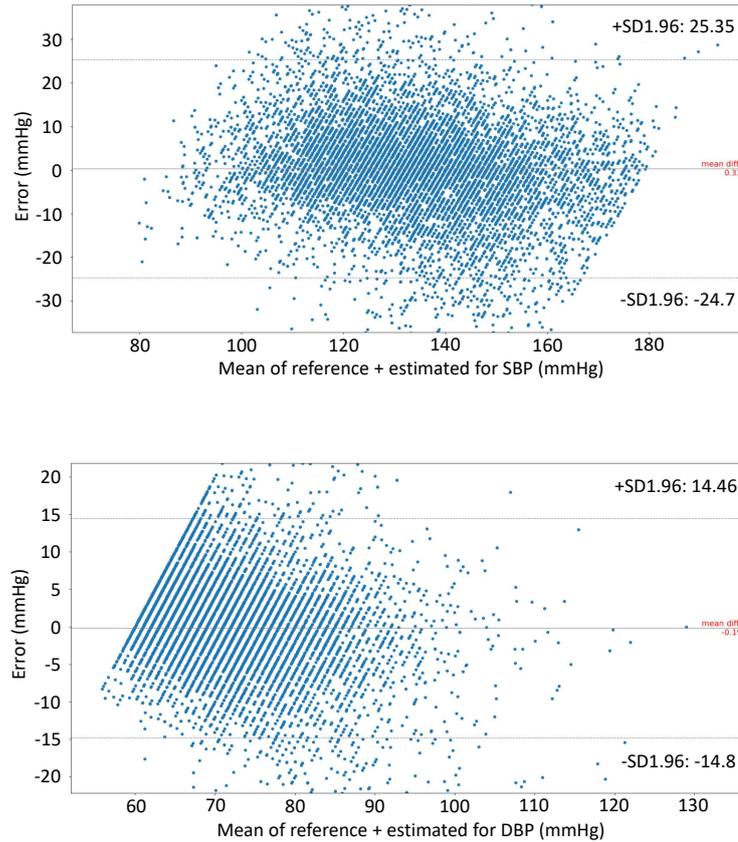
Figure 7.16: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the multi-layer perceptron (MLP) neural network.

7.7.3 Evaluation using Regression plots

The regression plots, in Figure 7.17, between the estimated and ground-truth values also present a large improvement compared to previous plots. It can be seen from these figures that the SBP and DBP estimation were in accordance with the reference values on the 24 feature set. Additionally, these plots show a stronger positive correlation between the predicted and ground truth values, as illustrated by the linear regression line and the correlation coefficient. The estimated SBP had a stronger correlation with the reference SBP values (R = 0.78), while the DBP had slightly less positive correlation with the reference DBP value (R = 0.7). Overall, the MLP model outperformed all the previous models by a large margin. Nonetheless, the MLP estimation accuracy is not reliable for accurate long term prediction.
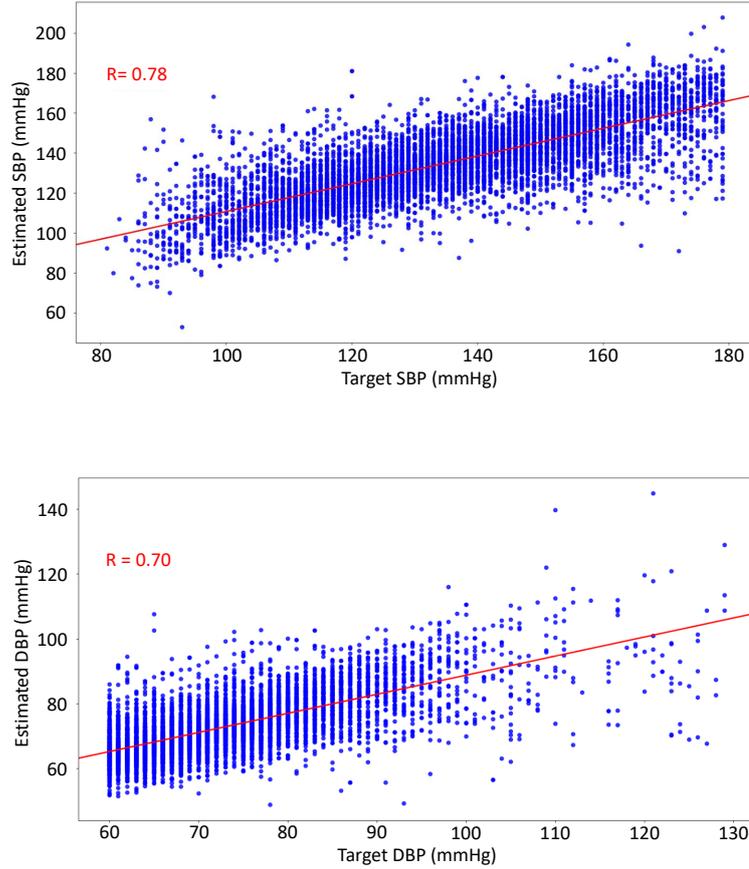
**Figure 7.17:** Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the MLP network.

7.8 Results from LSTM

7.8.1 Evaluation using MAE, SD and AAMI

The results obtained using the LTSM model on both datasets are presented in Table 7.6. The performance was assessed using the MAE±SD and ME±SD (AAMI metric) on both datasets. The results presented in Table 7.6 show a significant improvement compared to the results obtained using the classical machine learning models and MLP. For example, the MAE for SBP was reduced more than 62% in some case (such as AdaBoost and MLR), while the MAE for the DBP was reduced by up to 58%. The LSTM achieved a MAE±SD of 5.7±8.38 mmHg for the SBP and 3.17±4.77 mmHg for DBP on the 52 feature set. On the other hand, the MAE±SD on the 24 feature set were 5.97±8.81 mmHg for SBP and 3.27±4.99 mmHg for DBP. Considering the trade off between the complexity and accuracy, the time taken to

train the model on the reduced feature set was reduced while the estimation accuracy was almost unaffected. Furthermore, compared to all previously mentioned models, the evaluation metric using the AAMI depict a further reduction to the SD of error for both SBP and DBP on both datasets. The ME and SD for the DBP estimation are well below the limits set by the AAMI standards (± 8 mmHg), while the SD of the ME for the SBP fell short by only 2.62 mmHg on the reduced 24 feature set and only 2 mmHg on the full 52 feature set. Given the significant accuracy improvement obtained using the LSTM on both dataset, it can be said that modelling temporal dependencies between the input features is essential for more accurate long term estimation.

Table 7.6: Results of the LSTM model on the 52 and 24-feature set using MAE$\pm$SD mmHg and AAMI (ME$\pm$SD mmHg)

Model: LSTM	SBP				DBP			
Results	MAE	$\pm$SD	ME	$\pm$SD	MAE	$\pm$SD	ME	$\pm$SD
52-features	5.7	8.38	-1.3	10	3.17	4.77	-0.55	5.69
24-features	5.97	8.81	-0.56	10.62	3.27	4.99	-0.55	5.94

The estimation error distribution for the SBP and DBP using the LSTM model on the 24 feature set are shown in Figure 7.18. As expected, the histograms show a narrow error range (narrow base of the histogram) and a higher frequency around zero. This indicates that the error is centered around (the mean) zero and that the majority of the error lies within a close range of the mean. This can be easily distinguished when compared to the histograms in Figures 7.3, 7.6, 7.9, 7.12 and 7.15. The plots in Figure 7.18 show a normal distribution around zero with an estimation error range of mainly (-15,15) mmHg for SBP and roughly (-10,10) mmHg for DBP.
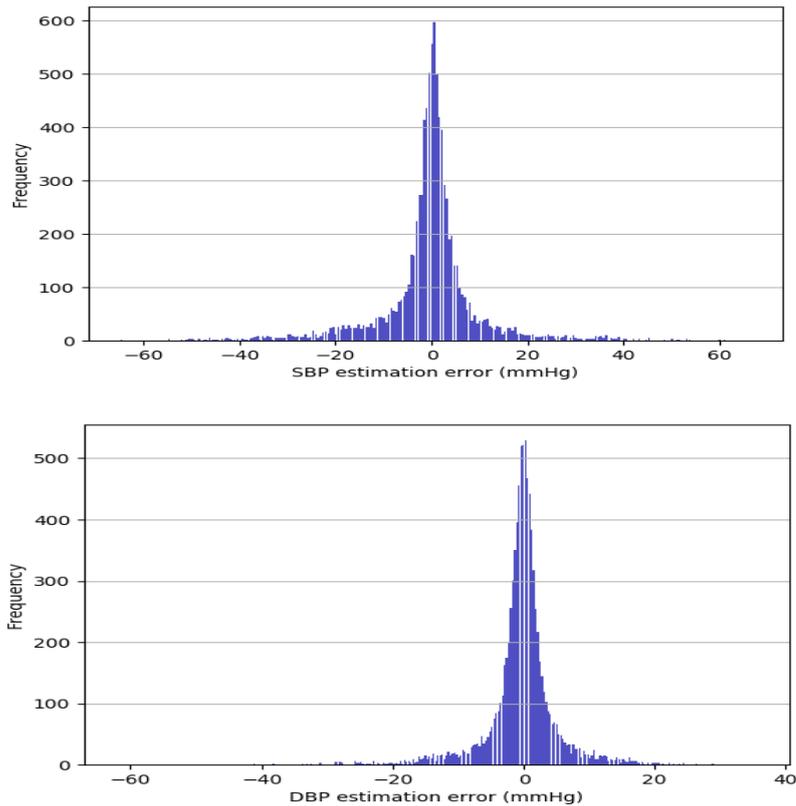
Figure 7.18: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the conventional LSTM.

7.8.2 Evaluation using Bland Altman plots

The Bland-Altman plots between the estimated and reference values for SBP and DBP on the reduced feature set are shown in Figure 7.19. These plots show a noticeable difference in comparison to Bland-Altman plots presented in the previous sections, such as Figures 7.4, 7.7, 7.10, and 7.13, etc. The main difference is that the data points in Figure 7.19 are spread horizontal along the entire range of BP values in the dataset and concentrated around the error zero on the y-axis with a narrow range, as opposed to the previous Bland-Altman figures were the data points are spread vertically (much larger error range on the y-axis over a smaller range of BP values on the x-axis). Furthermore, the upper and lower limits of agreements are now reduced over a smaller range. For example, the mean difference and [lower,upper] limits of agreements between the estimated and ground-truth values are -0.56 and (-21.3,20.26) for SBP, and -0.55 and (-12.2,11.09) for DBP. Although the limits of agreement where 95% of the error lies might be large, it is obvious from the plots that the data points are mainly concentrated within (-10,10) and (-7,7) for SBP and

134

DBP, respectively. Additionally, these plots clearly show that the ME for both SBP and DBP satisfy the AAMI standards, while the SD for the SBP was higher than the acceptable limits.
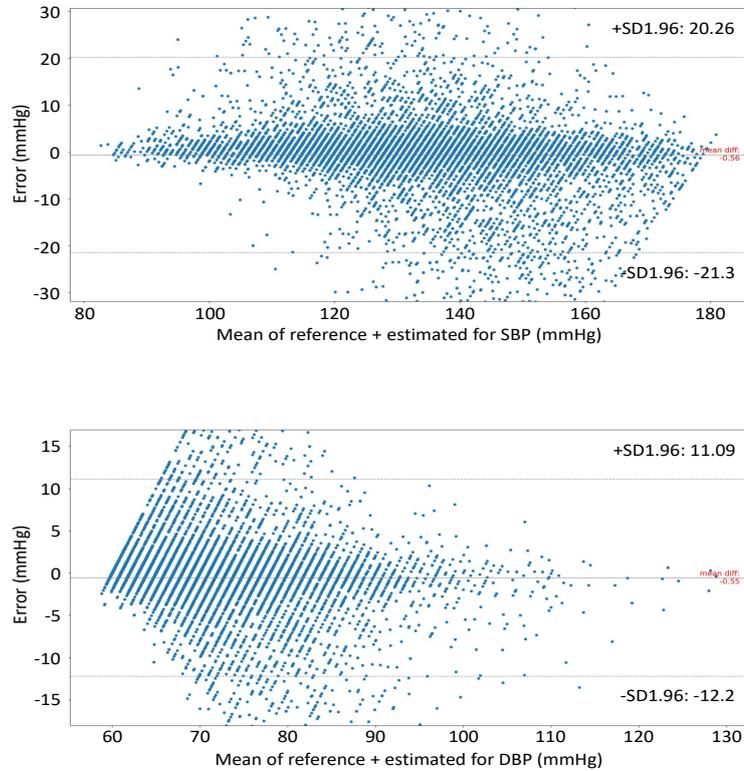
Figure 7.19: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the LSTM network.

7.8.3   Evaluation using Regression plots

The regression plots for the estimated and reference SBP and DBP values are presented in Figure 7.20. These plots indicate a strong positive correlation, where the estimated BP values are clearly in accordance with the ground-truth BP values. The strong correlation is also indicated by Pearson's correlation coefficient R of 0.84 for SBP and 0.81 for DBP. In comparison to the regression plots presented in the previous sections, the LSTM presented the strongest correlation between the predicted and reference BP values. Overall, the results obtained using the LSTM model are promising towards cuffless and continuous BP estimation.
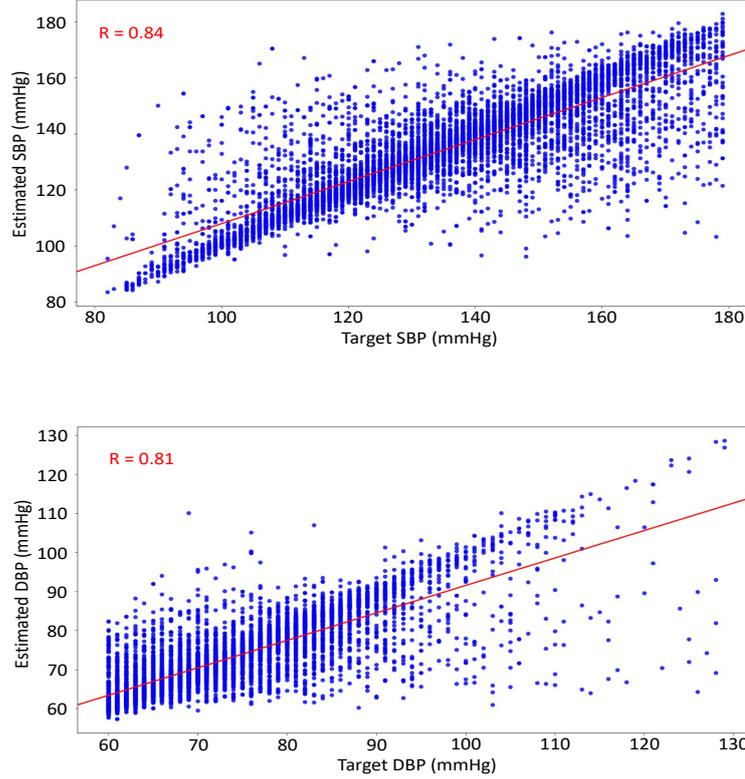
Figure 7.20: Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the LSTM network.

7.9 Results from GRU

7.9.1 Evaluation using MAE, SD and AAMI

Table 7.7 presents the MAE±SD and AAMI performance evaluation of the GRU model on both datasets. The results of the GRU model are comparable to the results obtained using the LSTM model in the previous section. Thus, similarly, the GRU outperformed all the classical machine learning models and the MLP model by a huge margin. The MAE±SD achieved by the GRU on the 52 feature set are 5.7±8.42 mmHg and 3.3±4.69 mmHg for SBP and DBP, respectively. These results are almost identical to the LSTM results presented in Table 7.6, on the 52 feature set. Furthermore, the results obtained on the 24 feature set using the LSTM (Table 7.6) and GRU (Table 7.7) are also comparable within a very close range. The MAE±SD for the SBP and DBP using the GRU model on the reduced feature set are 5.77±8.52 mmHg and 3.33±5.02 mmHg, respectively. Additionally, the performance evaluation against the AAMI standards show that the DBP estimation were acceptable on both 52 and 24 feature sets, while the SD of the ME for the SBP were

roughly 2 mmHg higher than the ±8 mmHg limits. This again proves the significance of selecting the right estimator for handling time series data, and subsequently increase estimation precision.

Table 7.7: Results of the GRU model on the 52 and 24-feature set using MAE±SD mmHg and AAMI (ME±SD mmHg)

Model: GRU	SBP				DBP			
Results	MAE	±SD	ME	±SD	MAE	±SD	ME	±SD
52-features	5.7	8.42	-0.99	10.15	3.3	4.69	0.09	5.73
24-features	5.77	8.52	-1.24	10.21	3.33	5.02	-0.44	5.7

The histograms in Figure 7.21 present the estimation error distribution for SBP and DBP using the GRU model on the reduced feature set. As expected, these histograms resemble the ones displayed in the previous section, in Figure 7.18, for the LSTM model. Thus, similar to the LSTM error histograms, the error distribution for both SBP and DBP are normally distributed around zero. Additionally, the errors are mainly concentrated between (-15,15) for SBP and (-10,10) for DBP with the highest clusters closer to zero.
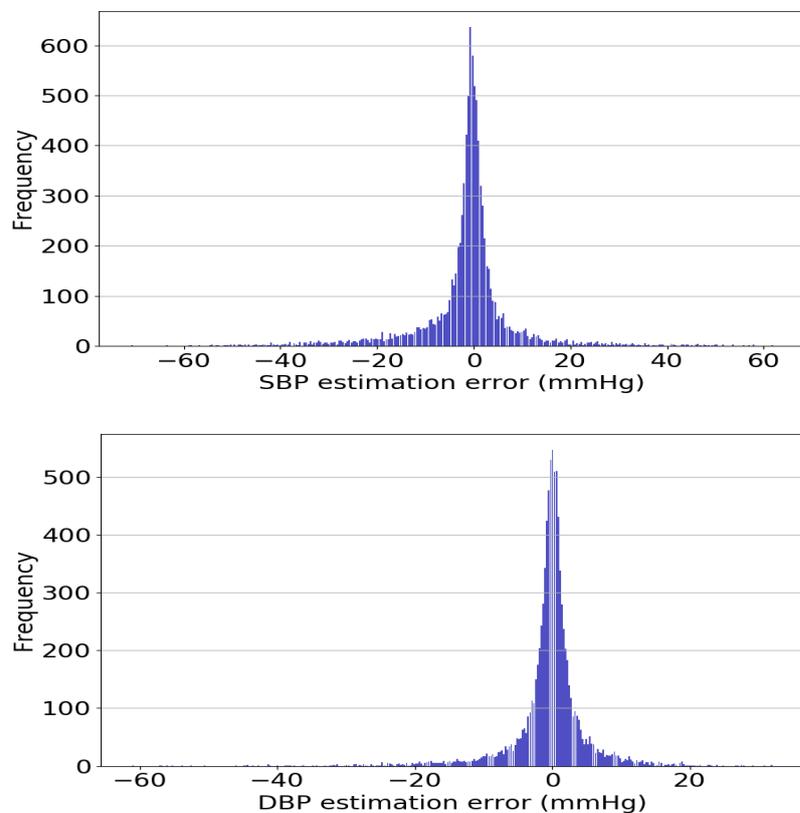

Figure 7.21: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the conventional GRU.

7.9.2 Evaluation using Bland Altman plots

Figure 7.22 presents the Bland-Altman plots between the estimated and reference values on the reduced feature set using the GRU model. Again similar to the LSTM model, these plots show a significant improvement compared to the Bland-Altman plots for models such as the MLR in Figures 7.4, SVR in Figure 7.7, random forest in Figure 7.10, and even the feedforward network in Figure 7.16. The mean difference and [lower, upper] limits of agreements are -1.24 and (-20.5,18.08) for SBP and -0.44 and (-11.6, 10.75) for DBP. The 95% confidence interval for the DBP indicates that the results fall within limits set by the AAMI standards. Moreover, although the results for SBP were slightly above the acceptable range, it can be seen that the error is more concentrated within a range of (-10,10).
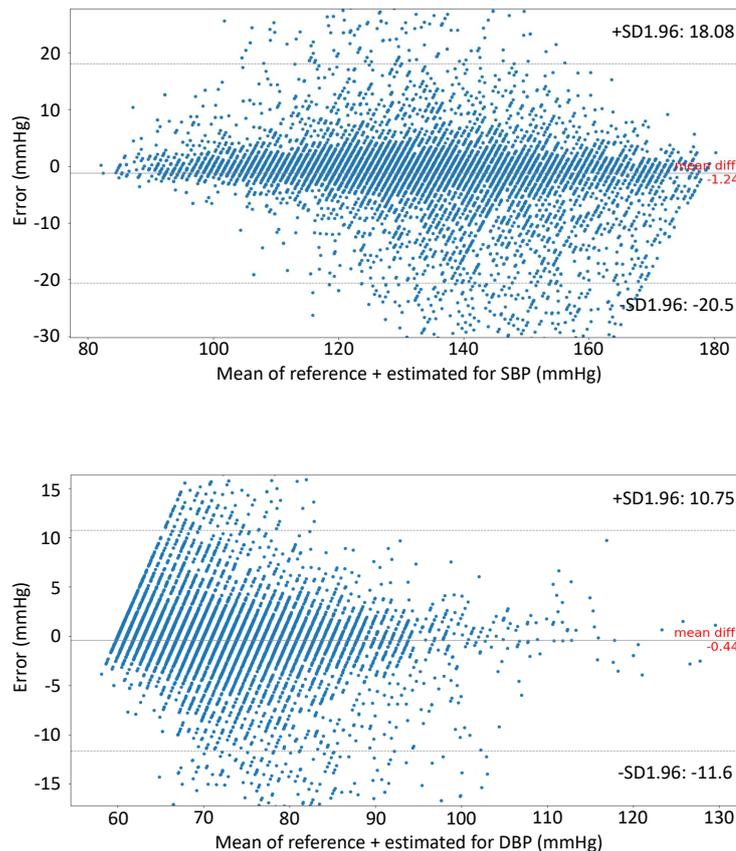

Figure 7.22: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the GRU network.

7.9.3 Evaluation using Regression plots

The regression plots for the SBP and DBP using the GRU model are shown in Figure 7.23. These plots suggest that the estimated SBP and DBP values were in-line with the reference values in the dataset with a very strong positive correlation. As indicated by Pearson's correlation coefficient (R), the strength of the correlation between the estimated and reference values are 0.85 and 0.81 for SBP and DBP, respectively. In comparison to all previously mentioned models, except the LSTM, it is clear that the GRU provided a superior performance with a much higher degree of estimation accuracy. Thus, the overall performance of the GRU is acceptable.
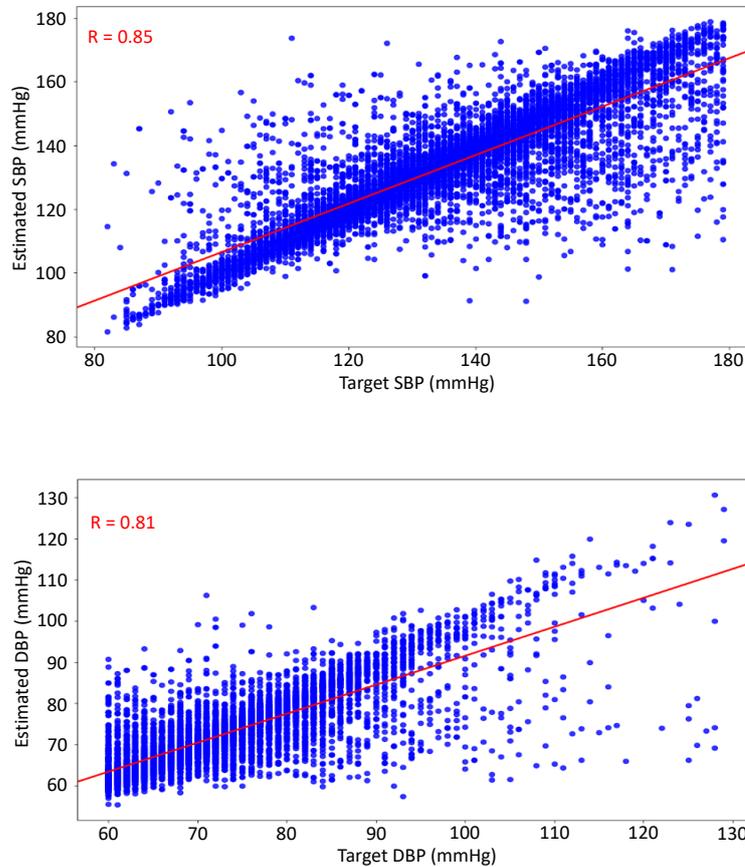

Figure 7.23: Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the GRU network.

7.10 Results from the proposed Bi-LSTM and attention

7.10.1 Evaluation using MAE, SD and AAMI

The performance of the first proposed model (i.e. Bi-LSTM layer followed by a stack of unidirectional LSTM layers and an attention layer) evaluated using the MAE±SD and AAMI standards are presented in Table 7.8. The results obtained using this proposed model vastly outperformed the classical machine learning models, where the MAE for SBP and DBP were reduced by up to 69% and 64%, respectively. Also, this model provided superior performance in comparison to the MLP neural network model, where the MAE for SBP and DBP were reduced by up to 50%. Additional, this model further reduced the error obtained using the conventional LSTM and GRU models. As shown in Table 7.8, this model achieved a MAE±SD of 4.51±7.81 mmHg for SBP and 2.6±4.41 mmHg for DBP, on the 52 feature set. Also, comparable results were achieved on the 24 feature set, where MAE±SD for SBP and DBP are 4.86±8.43 mmHg and 2.83±4.86 mmHg, respectively. Thus far, this model provided the best performance with the lowest estimation error in comparison to all other models presented earlier. In terms of performance evaluation against the global standards set by the AAMI, the ME for both SBP and DBP were well below the 5 mmHg limit, on both datasets. Additionally, the SD for the DBP on both datasets were within the required limits. However, the SD for the SBP were a little over 1 mmHg above the acceptable limit of 8 mmHg. Hence, the DBP estimation did satisfy the AAMI standards, while the SBP estimation fell short by a small margin on the SD. Nonetheless, the performance of this model is very promising towards non-invasive cuffless and continuous BP estimation.

Table 7.8: Results of the proposed first model (Bi-LSTM + LSTM + attention) on the 52 and 24-feature set using MAE±SD mmHg and AAMI (ME±SD mmHg)

Model: Bi-LSTM+ LSTM+Attention	SBP				DBP			
Results	MAE	±SD	ME	±SD	MAE	±SD	ME	±SD
52-features	4.51	7.81	-0.48	9.15	2.6	4.41	-0.49	5.1
24-features	4.86	8.43	-0.79	9.69	2.83	4.86	-0.35	5.61

The SBP and DBP estimation error distribution, on the 24 feature set, are presented in the histogram plots in Figure 7.24. These plots show a normal distribution, where the errors are mainly spread around zero. In particular, the errors are mainly

concentrated between [-10,10] for both SBP and DBP. The sharp pointy shape in both plots indicate the remarkably small range of estimation error achieved by this architecture.
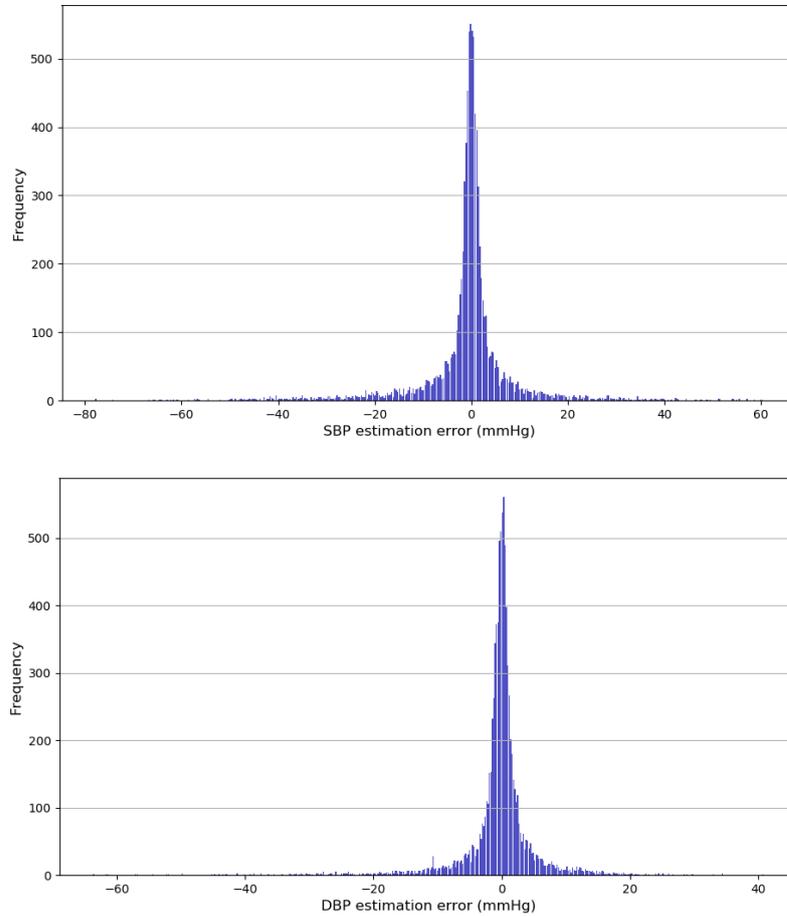
Figure 7.24: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the proposed architecture with LSTM units.

7.10.2 Evaluation using Bland Altman plots

The Bland-Altman plots between the estimated and ground-truth values, using the proposed model, on the reduced feature set, are shown in Figure 7.25. It can be seen from these plots that the data points cover the entire range of SBP and DBP values available in the dataset. Additionally, the plots show that the small error range is mainly between (-10,10) and (-5,5) for SBP and DBP, respectively. The mean difference between the estimated and reference values are -0.79 for SBP and -0.35 for DBP. Furthermore, the limits of agreements are (-19.79,18.22) and (-11.36,10.66) for SBP and DBP, respectively. This proves that the model was able to estimate SBP and DBP with a good estimation precision and low error rate.
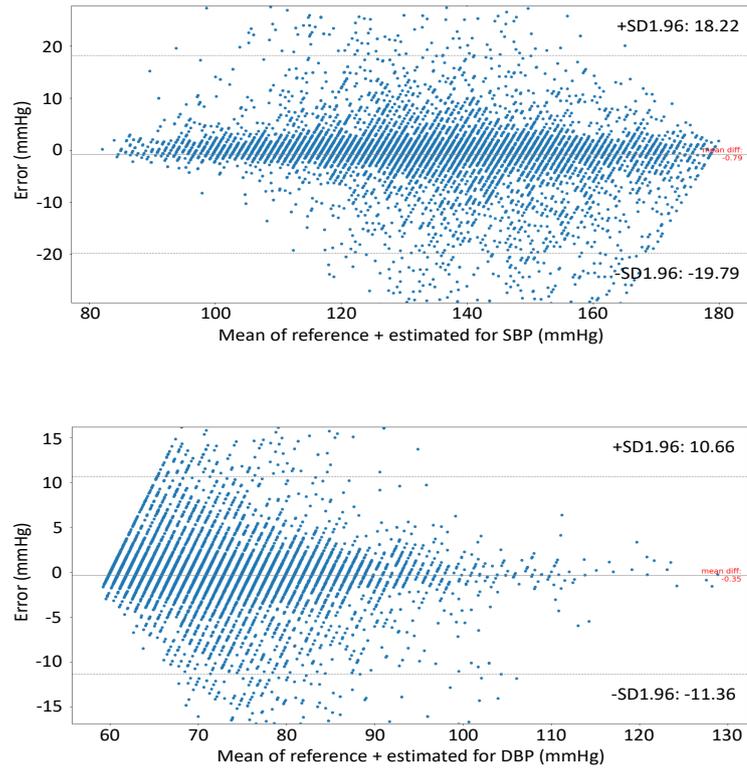
Figure 7.25: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the proposed architecture with LSTM units.

7.10.3 Evaluation using Regression plots

The relationship between the estimated and reference values were assessed using the regression plots presented in Figure 7.26. These plots depict a strong positive linear correlation between the two variables. The strength of the correlation is quantified using Pearson's correlation coefficient. The R value for the SBP is 0.87 and 0.84 for DBP. This indicates that the estimated values were in-line with the reference values in the dataset with a high correlation.
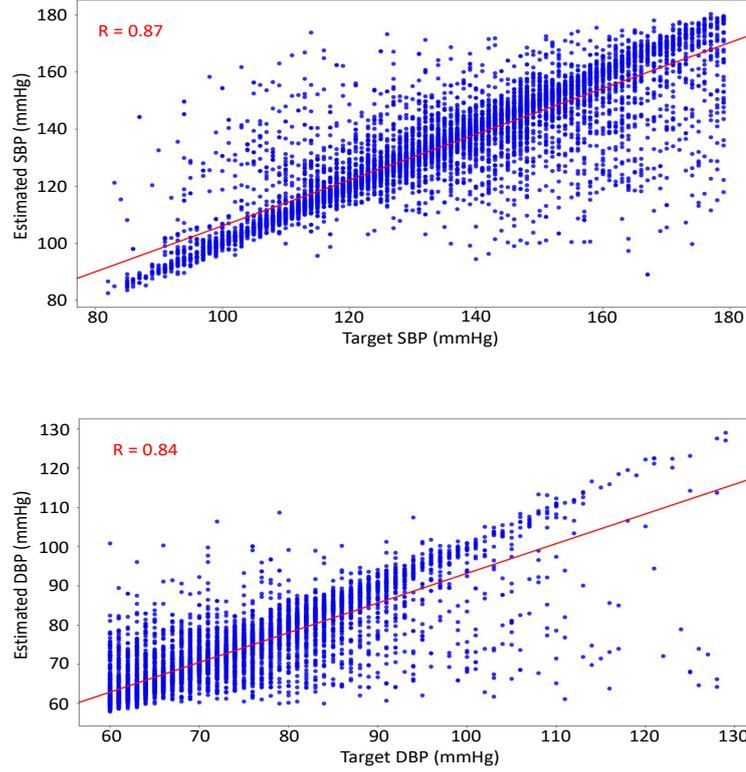
**Figure 7.26:** Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the proposed architecture with LSTM units.

7.11 Results from proposed Bi-GRU and attention

7.11.1 Evaluation using MAE, SD and AAMI

Table 7.9 presents the results obtained using the second proposed model (i.e., Bi-GRU layer followed by a stack of unidirectional GRU layers and an attention layer). This table shows that the results of this model are very similar to the ones obtained using the first proposed model (Bi-LSTM), with almost a negligible difference in the reported errors. However, the performance of this model significantly outperforms the classical machine learning models and the MLP model. It also provides better estimation accuracy and lower error in comparison to the conventional unidirectional LSTM and GRU models. In particular, the MAE±SD for the SBP and DBP on the 52 feature set are 4.69±7.76 mmHg and 2.68±4.39 mmHg, respectively. As for the performance on the 24 feature set, the MAE±SD are 4.79±8.08 mmHg for SBP and 2.77±4.72 mmHg for the DBP. The advantage of using the 24 feature set is that it reduces the complexity and subsequently time taken to train while at the same

time achieve comparable results to the ones obtained using the full 52 feature set. Furthermore, in terms of evaluation against the AAMI standards, as expected the ME and SD for the DBP satisfied the AAMI standards, while SD for the SBP did not fall below the acceptable limits. However, the overall performance indicates a very strong potential towards non-invasive cuffless BP estimation device.

Table 7.9: Results of the proposed second model (Bi-GRU + GRU + attention) on the 52 and 24-feature set using MAE±SD mmHg and AAMI (ME±SD mmHg)

Model: Bi-GRU+ GRU+Attention	SBP				DBP			
Results	MAE	±SD	ME	±SD	MAE	±SD	ME	±SD
52-features	4.69	7.76	-0.24	9.12	2.68	4.39	-0.37	5.17
24-features	4.79	8.08	-0.91	9.34	2.77	4.72	-0.44	5.45

The histograms in Figure 7.27 present the distribution of error for the SBP and DBP estimation on the 24 feature set. Similar to the previous model, the errors are distributed around the mean zero with a very small range of (-15,15) for SBP and (-10,10) for DBP. The highest frequencies were within a very close proximity to the mean, which shows that the differences between the estimate and reference values are very small. This indicates a high estimation precision and outstanding performance.
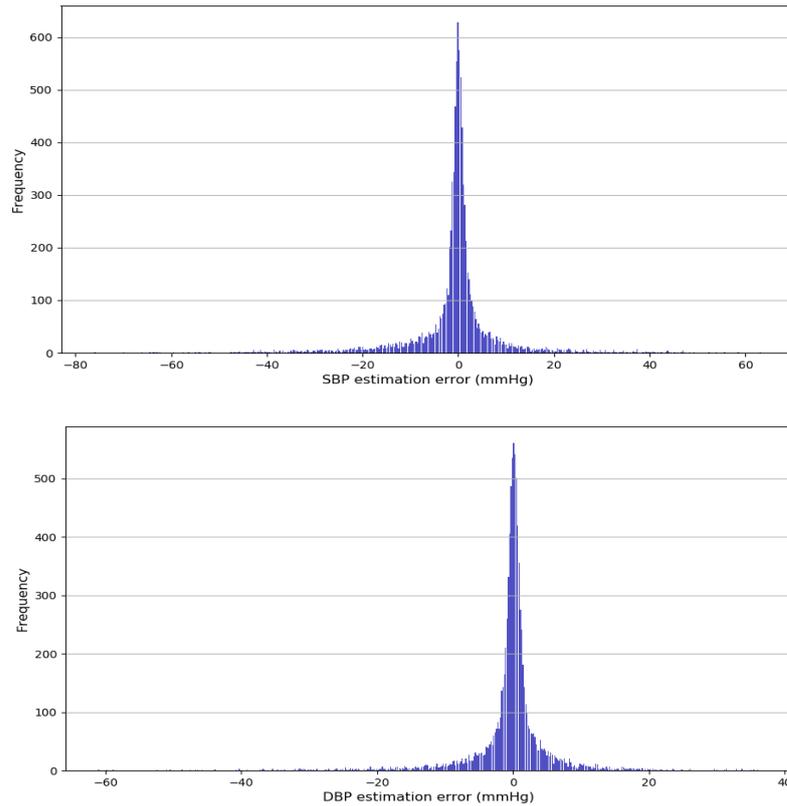
Figure 7.27: Histogram of error distribution for the estimated SBP (top) and DBP (bottom) using the proposed architecture with GRU units.

7.11.2   Evaluation using Bland Altman plots

The bland-Altman plots between the estimated and reference BP values on the reduced feature set are shown in Figure 7.28. The mean difference for SBP is 0.91 and -0.44 for DBP. Additionally, [lower,upper] limits of agreements were 95% of the error fall are (-19.22,17.41) and (-11.13,10.25) for SBP and DBP, respectively. As seen in these plots, the majority of the errors are spread horizontally within a small distance from the mean, over the entire range of BP values in the dataset. This means that the model was able to predict the low and high BP values in the dataset with good precision.

145
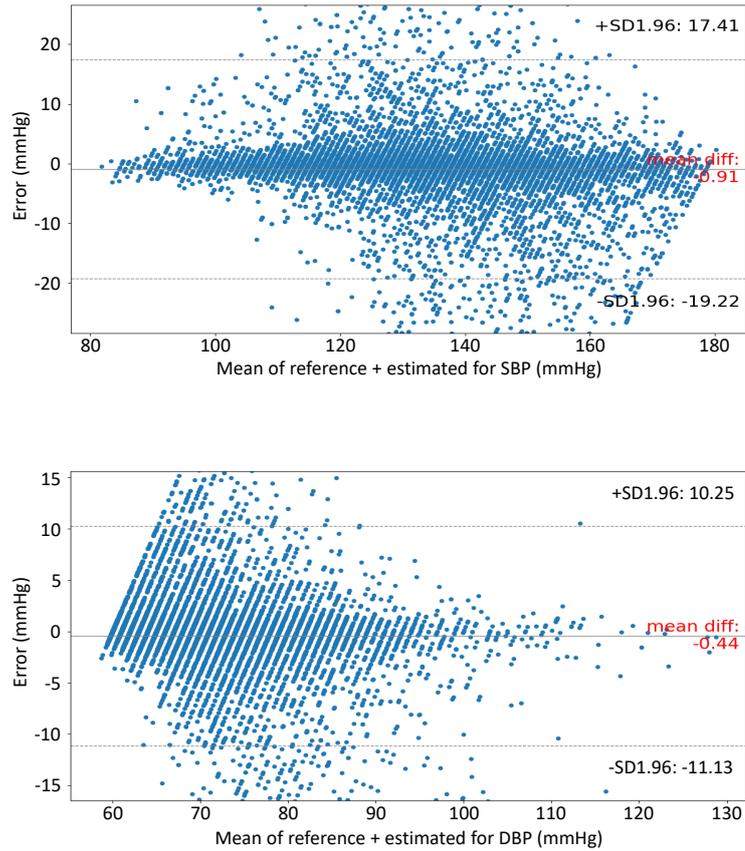
Figure 7.28: Bland Altman plots between the estimated and ground-truth values for SBP (top) and DBP (bottom) using the proposed architecture with GRU units.

7.11.3 Evaluation using Regression plots

The regression plots between the estimated and reference BP values are presented in Figure 7.29. It can be seen in these plots that the estimated values where in accordance with the BP values in the dataset. Furthermore, Pearson's correlation coefficient shows a very strong correlation of 0.88 for SBP and 0.84 for DBP. Visually, the plots indicate a positive correlation between the estimated and reference values. Overall, this model had the highest correlation coefficient R, lowest SD of ME as well as lowest MAE±SD amongst all models explored in this chapter, hence, this model provided the best estimation accuracy, and subsequently was selected as the best performing model in this research project.
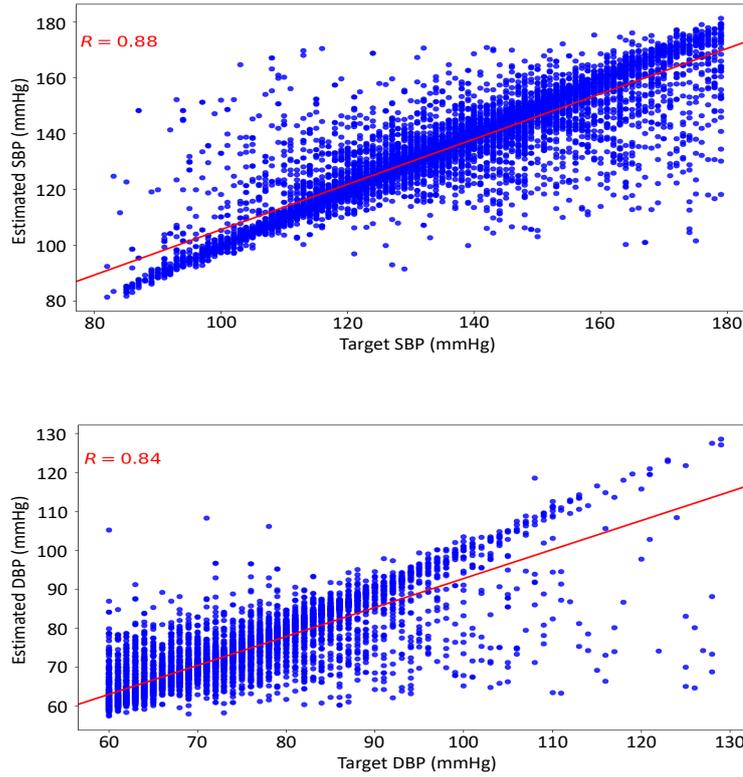
Figure 7.29: Regression plots with a linear regression line for the estimated and ground-truth SBP (top) and DBP (bottom) using the proposed architecture with GRU units.

7.12 Summary

In summary, the results achieved by the LSTM and GRU units using the proposed architecture outperformed all other models implemented in this research. Furthermore, the results obtained using the linear and nonlinear classical machine learning models as well as the feedforwad network (MLP) failed to achieve acceptable accuracies for the SBP and DBP estimation. This suggests that the inherently non-linear relationship between BP and the PPG features requires more advanced models. The results also confirm that the non-recurrent models are not suitable for handling the time series data. This is evident by the superior performance achieved by all the recurrent models. Moreover, the bidirectional connection layer and the attention mechanism further improved the estimation precision and brought the accuracies closer to the internationally acceptable standards. Table 7.10 presents a summary of the results obtained in this chapter on the 24 feature set using MAE±SD mmHg and AAMI standards.

Table 7.10: Summary of the results achieved on the 24-feature set by all the models implemented in this chapter using MAE±SD mmHg and AAMI (ME±SD mmHg)

Algorithms	SBP		DBP	
	MAE±STD (mmHg)	ME±STD (mmHg)	MAE±STD (mmHg)	ME±STD (mmHg)
Linear regression	15.11±10.95	-0.02±18.66	7.42±6.6	-0.08±9.93
Support Vector Machine	12.95±11	-0.31±16.99	6.11±6.7	-1.35±8.97
Random forest	11.47±9.49	-0.09±14.89	5.58±5.7	-0.1±7.98
AdaBoost	15.18±10.69	-0.06±18.56	7.57±6.4	0.74±9.88
MLP	9.21±8.84	0.33±12.76	5.16±5.41	-0.19±7.47
LSTM	5.97±8.81	-0.56±10.62	3.27±4.99	-0.55±5.94
Bi-LSTM+LSTM+Attention	4.86±8.43	-0.79±9.69	2.83±4.86	-0.35±5.61
GRU	5.77±8.52	-1.24±10.21	3.33±5.02	-0.44±5.7
Bi-GRU+GRU+Attention	4.79±8.08	-0.91±9.34	2.77±4.72	-0.44±5.45

Chapter 8

Discussion and Conclusion

8.1 Summary of the current progress towards cuffless BP estimation

Currently, cuff-less non-invasive BP measurements can be divided into techniques that use only a PPG sensor and techniques that use a hybrid approach namely a PPG sensor and the ECG or two PPG sensors. The hybrid approach is mainly based on PWV through PTT or PAT. PTT is the time that takes the blood pressure wave to travel between two points on the body and is inversely correlated with BP. PAT is defined as the time interval between the electrical activation of the heart and arrival of the pulse wave at a location on the body peripheral. PAT is PTT in addition to Pre-ejection Period. PAT can be measured using two sensors, an ECG sensor and a PPG sensor. It is based on the time difference between the R peak of the ECG and a point on the PPG rising edge. Although both PTT and PAT are well established techniques, these methods are not easy to implement and have several practical challenges. PTT and PAT parameters expire after a short period causing the estimation accuracy to deteriorate as it parts away from the initial calibration. Moreover, both methods require two measurement sensors that need to be synchronised for accurate peak detection, and placed on fixed positions on the body which is difficult and inconvenient for patients to maintain during measurements. Additionally, both sensors have different sampling rates in real time. Furthermore, PPG and ECG sensors are very sensitive to motion artefacts due to movements during the recording which in turn require rigorous signal processing before the signals can be used in a BP study. All these challenges make the PTT and PAT approaches less

appealing and non-practical for everyday use.

To overcome some of these challenges, researchers introduced a pulse wave analysis method. Pulse wave analysis approach can be used to estimate BP using only one PPG sensor without an ECG. Even though the origin of the PPG components are not fully understood, it is acceptable that PPG can provide information about the cardiovascular system. This approach is simple, inexpensive and more convenient for patients during measurement since it only uses one photosensor. In this approach several temporal features are extracted from the PPG and used as input data for machine learning and neural network models for BP estimation. The main obstacle for BP monitoring using PPG is accuracy. Different guidelines were created for researchers to follow in order to compare their methods to reference invasive BP values. These guidelines were set by the British Hypertension Society (BHS) and AAMI. According to AAMI, the mean error difference between estimated and reference should not exceed 5 mmHg, and the standard error deviation should not exceed 8 mmHg for 85 patients.

Over the last 15 years, linear and non-linear models have been employed for estimating BP. In some cases, where the dataset belongs to healthy individuals, some of the linear models were able to achieve reasonable and acceptable results, as reported in chapter 3. However, other studies show that when these models are evaluated on new subject data (not seen before) they fail to provide acceptable BP estimates that comply with the AAMI standards. As a result, many non-linear models have also been employed such as, support vector machine, random forest, AdaBoost, feedforward neural network, etc. In many cases, the non-linear models outperformed the linear models but again depending on the dataset and approach used i.e. PTT, PAT or PWA (PPG only). More advanced methods have also been proposed, such as, convolutional neural networks, ResNet, Unet etc. However, these models are not suitable for long term estimation, and as shown in chapter 3, their complexity did not improve the BP estimation precision and in all cases failed to provide acceptable results that satisfy the AAMI standards. It has been shown in the literature that of all the models that have been tested, one particular category stands out by providing the best reported estimation accuracies, and that is the recurrent neural network class, particularly the GRU and LSTM which have a huge advantage over

150

the previously mentioned models. These models are equipped with the ability to model the variation of the extracted features with respect to time. Studies have reported an improvement in the BP estimation using these models and the possibility to potentially employ recurrent models for long term continuous measurements.

8.2   Summary of thesis and findings

The main aim of this research project is to establish a reliable algorithm for non-invasive, cuffless and continuous estimation of SBP and DBP using only PPG signals. In order to achieve this aim, two models were proposed and evaluated on a large dataset. The proposed models are deep learning models consisting of one bidirectional RNN layer, followed by a series of stacked unidirectional RNN layers, and an attention layer. However, the traditional tanh activation function in the RNN hidden layers units were replaced by the LSTM and GRU cells, as described in section 6.1.7 of chapter 6. The models were evaluated on two datasets derived from 942 subjects extracted from the MIMIC II database. The first dataset consists of 52-feature set extracted from each cycle of the 10 s PPG segments and its first two derivatives (presented in Table 4.2, section 4.3.2), while the second dataset consists of a reduced feature set containing features with the highest impact on the target outputs (features are listed in Table 4.5, section 4.4.3). These features were selected based on rigorous feature selection and elimination techniques using a combination of statistical (Pearson's correlation and maximum correlation coefficient- MIC) and machine learning (recursive feature elimination) approaches, as explained in sections 4.4.2 and 4.4.3. Removing unreliable and insignificant (noisy) features not only reduces the complexity of the task but also increases the estimation accuracy of the model and enhances its generalisation. The result of the 3-steps feature elimination process is a refined 24-feature set containing only the most significant PPG features for BP estimation. The performances on the two datasets (52 vs 24 input features) were comparable and in most cases the difference in accuracy was insignificant, however, the models yielded slightly higher variance using the 24 features, as shown in the performance analysis in chapter 7. This might suggest that adding redundant features or bias to the model could reduce the variance. Nonetheless, reducing the input feature dimension by more than half, helped reduced the time

and space needed for training the models whilst at the same time maintaining a good performance.

The performance of the proposed models was compared to seven frequently used models in the literature. These models can be divided into classical machine learning models (multi-linear regression, support vector regression, random forest and AdaBoost) and neural network models (feedforward- MLP, LSTM and GRU). The neural network models can be further divided into non-recurrent (MLP) and recurrent models (LSTM and GRU). The results and performance analysis are presented in chapter 7. The results show that the MLR and AdaBoost produced the worst results in comparison to all other models. A slight performance improvement was achieved by the SVR as well as the random forest- which produced the best accuracy among the classical machine learning models. However, the poor performance and high error obtained using these models indicate that they cannot capture the complex relationship between the input features and target outputs, and subsequently are not suitable for this task. On the other hand, a noticeable improvement was achieved by the neural network models. For example, the feedforward non-recurrent model outperformed all classical machine learning models, while the recurrent neural network models outperformed all the classical and non-recurrent model. Moreover, the proposed models further enhanced the estimation accuracy and reduced bias and variance compared to the conventional LSTM and GRU. Thus, the two proposed models were able to estimate SBP and DBP with the highest precision in comparison to the most commonly encountered models in the literature. This proves that incorporating past and future information through the bidirectional connections, and allowing the network to learn more intensively by automatically selecting the more important hidden states through attention mechanism, can improve the estimation accuracy. In terms of performance comparison between the two proposed models, the architecture equipped with GRU cells was selected as the best performing model on the reduced feature set, as shown in chapter 7, section 11. The reasons for selecting this model as the best estimator are based on both accuracy and complexity. Firstly, the MAE and SD is lower compared to the results achieved by the BiLSTM model (section 10 of chapter 7). Secondly, the SD of ME is also lower (ME was ignored since both models achieved an acceptable ME by AAMI standards). Thirdly, the regression plots showed that the predicted values had the best correlation to

the reference values as indicated by the Pearson's correlation coefficient (chapter 7, section 7.11.3). Fourthly, the GRU cells require less computation, and thus it is computationally more efficient. Fifthly, DBP estimation met the requirements of the international standard, while the SBP fell short by only 1.34 mmHg for the SD limits.

8.3   Comparisons with other related works

Table 8.1 presents a comparison between the best proposed deep learning model (in chapter 7, section 7.11) and some of the related works in the literature. It should be noted that it is not possible nor fair to make a direct and accurate comparison between the work done in this research and other related work since the data, acquisition devices and evaluation metrics are not consistent in all studies along with many other important information such as calibration. A large number of studies collect their own private data, while others use publicly available datasets with insufficient or unspecified number of subjects and potentially different subjects even when the same database is used. Furthermore, the evaluation metric also varies between studies, such as MAE, ME, root mean square error (RMSE), R-squared, etc. All these reasons make it very difficult to compare our work with other studies. It was observed that, on one hand, reasonable estimation accuracy was achieved on small datasets or limited number of patients. On the other hand, the performance worsens as the number of patients or the size of the data increases. Therefore, the results obtained in this research will only be compared to well-established studies (highly cited) that adopted the MIMIC dataset, calibration free approach, and where the MAE or ME is used as an evaluation metric.

In Kachuee et al [113, 114], several classical machine learning models were employed for estimating BP using PTT or PAT related features along with PPG features. The models were evaluated on a large number of subjects collected from the MIMIC II dataset. There are several limitations associated with this approach such as it required two signals, often from two different measurement sensors, and both SVR and AdaBoost used were not suitable for modelling the variation in the input features with respect to time, which generally enhances the prediction accuracy. Furthermore, the reported MAE and SD for both SBP and DBP were high, as shown in

Table 8.1: Performance comparisons with related works evaluated on the MIMIC database

Authors	Dataset	Approach	Error (mmHg)
Kachuee et al [114]	851 subjects, (MIMIC II)	PTT feature-based, (SVR, calibration free)	MAE±SD for SBP: 12.38±16.17 DBP: 6.34±8.45
Kachuee et al [113]	942 subjects, (MIMIC II)	PAT feature-based, (Adaboost, calibration free)	MAE±SD for SBP: 11.17±10.09 DBP: 5.35±6.14
Tanveer et al [30]	39 subjects, (MIMIC I)	PTT raw signals, (ANN+LSTM, calibration free)	MAE for SBP: 0.93 DBP: 0.52
Kurylyak et al [25]	15000 heartbeats, (MIMIC II)	PPG features-based, (ANN, calibration free)	ME±SD for SBP:3.8±3.46 DBP: 2.21±2.09
Slapnicar et al [145]	510 subjects, (MIMIC III)	PPG+ derivatives of raw signals, (ResNet, calibration free)	MAE for SBP: 15.41 DBP: 12.38
This work	942 subjects, (MIMIC II)	PPG + first and second derivatives features, (BiGRU+GRU+attention, calibration free)	MAE±SD for SBP: 4.79±8.08 DBP: 2.77±4.72 ——— ME±SD for SBP: -0.91±9.34 DBP: -0.44±5.45

Table 8.1. In comparison, our results were superior using only one PPG sensor. Tanveer et al [30] estimated BP using the raw ECG and PPG signals collected from 39 subjects from the MIMIC I. The BP estimation model used a feedforward layer for extracting features from the input signals automatically, which were then fed into a LSTM model to account for the temporal variations in the data. This method achieved excellent results for both SBP and DBP. However, this approach still requires two signals, and the window length of the ECG and PPG was 40 s i.e., 5000 data points. Additionally, the accuracy also relied on the perfect synchronisation between the PPG and ECG signals. Furthermore, the model was evaluated on a small number of patients, which is insufficient for meeting the required standard set by AAMI for cuffless BP evaluation and the standard deviation was not reported. In our experiment, the model was evaluated on 942 subjects and the PPG features were extracted from a 10 s segment, hence the input feature vector was much smaller. In another study, Kurylyak et al [25], employed a feedforward model for estimating SBP and DBP using features extracted from the PPG. The results were reasonable in terms of ME and SD, however, the dataset was very small (15000 cycles), and the number of patients is unspecified. Slapnicar et al [145] proposed a ResNet based model for estimating SBP and DBP using the raw PPG including first and second derivatives signals as input. The model was evaluated on 510 subjects and the win-

dow length of each signal was 5 s. The network architecture is relatively complex, each of the three input signals are processed by five ResNet blocks and a spectro-temporal block with a GRU layer in parallel. This means that the complexity of the model is high. Additionally, the reported performance was poor compared to other reported studies in the literature. The MAE for SBP was 15.41 mmHg and 12.38 mmHg for DBP while the standard deviation was not reported. As shown in Table 8.1, in general, studies reported higher error and standard deviation on larger datasets. Overall, considering the large number of patients used in this research study and the fact that only PPG signals were utilised, the model performance was relatively good. This again confirms that it is possible to effectively estimate BP using only a single, simple and inexpensive PPG sensor with a reasonable accuracy, even when compared to the well-established PTT approach.

8.4 Strengths and limitations of this study

Strengths

The strengths of this study are as follows:

- The dataset used in this study is a large public dataset that contains bio-signals with reference to gold standard invasive arterial BP signals collected from a diverse population with exposure to a wide range of BP values.

- A large number of features extracted from the PPG and its derivatives were explored. Afterwards, redundant, unreliable and insignificant PPG features were effectively eliminated whilst keeping the estimators' accuracy intact.

- A total of nine models were implemented, evaluated and compared using the most common regression metrics, graphical analytical tools, as well as the global BP measurement standards set by the AAMI. Out of the nine models, seven were used as baseline models for comparisons against the two models proposed in this research.

- Classical machine learning models, non-recurrent as well as recurrent neural network models were all evaluated on two datasets: one containing 52 feature set and the other one containing the reduced feature set (24 most significant features).

155

- The proposed models provided superior performance compared to all baseline models that are frequently encountered in the literature.

- The proposed algorithms have the ability to learn temporal dependencies in successive PPG cycles during the mapping between the input and target. Additionally, the bidirectional connections expose the network to more context by processing the input vector in a forward and backward manner, effectively allowing the network to consider past as well as near future information, which in turn enhanced the BP estimation.

- The proposed models are capable of consistently providing good performance whilst avoiding the vanishing gradient problem associated with deep recurrent neural network models.

- The proposed algorithms were trained to attend to the hidden states with the most significant influence on BP, in each time step, using attention mechanism to enhance its estimation performance.

Limitations

The limitations of this project are associated with the PPG approach, which are primarily related to data and feature extraction. These main challenges are:

- The quality of the PPG signals in the dataset is relatively poor, and contains noisy and corrupt segments. Additionally, the PPG and BP signals are not perfectly synchronized which required more attention.

- The data recordings in the MIMIC II are collected from ICU patients, and therefore, the subjects are generally sick, potentially older than the average population age, and under heavy medications.

- This approach relies on an accurate extraction of PPG features that correlate to BP, however, not all features mentioned in the literature can be extracted due to varying PPG waveform contours between individuals owing to different age, disease, medication, etc. This makes it impossible to extract all relevant features from the PPG morphology since some of these features were invisible and hard to detect, such as, dicrotic notch and diastolic peak related features as well other features from the PPG derivatives.

For example, the dicrotic notch or information relating to arterial stiffness from the PPG's second derivative that were used in [47], which were not visible for all patients in this study.

- The signals extracted from the MIMIC database contain more normotensive and hypertensive than hypotensive. Therefore, the models are biased towards the two main BP categories.

- Demographical information about the patients such as age, gender, height and weight are not available for all patients, hence, these features were excluded. However, demographical information can further improve model performance and produce more reliable BP estimation as it exposes the model to more input features and important information/context that have direct influence on BP. Additionally, demographical information allows for more features to be extracted such as large artery stiffness which requires knowing patient's height.

- In terms of complexity, the proposed models take more time to train compared to classical machine learning, feedfoward neural network and conventional recurrent neural networks.

8.5 Conclusion

Overall, the PPG is a promising technology with a great potential for offering BP measurements in a non-invasive, continuous and cuff-less manner. Such a device will have significant and transformative impact in the monitoring of patients, especially those who are at risk of cardiovascular disease. It is encouraging to see so much global interest by researches and industry alike in this field. There are still challenges to be resolved, however if the momentum of this research topic continuous in the same trajectory as it is now it is very hopeful that a PPG based non-invasive, cuff-less and continuous BP monitoring device could be commercialised in the near future.

References

[1] Guidelines Committee et al. 2003 european society of hypertension–european society of cardiology guidelines for the management of arterial hypertension. *Journal of hypertension*, 21(6):1011–1053, 2003.

[2] D Zheng, JN Amoore, S Mieke, FE Smith, ST King, and A Murray. Automated blood pressure measurement: reasons for measurement variability uncovered. In *2009 36th Annual Computers in Cardiology Conference (CinC)*, pages 21–24. IEEE, 2009.

[3] Lysander WJ Bogert and Johannes J van Lieshout. Non-invasive pulsatile arterial pressure and stroke volume changes from the human finger. *Experimental physiology*, 90(4):437–446, 2005.

[4] Cuicui Wang, Fan Yang, Xueguang Yuan, Yangan Zhang, Kunliang Chang, and Zhengyang Li. An end-to-end neural network model for blood pressure estimation using ppg signal. In *Artificial Intelligence in China*, pages 262–272. Springer, 2020.

[5] Meng Rong and Kaiyang Li. A multi-type features fusion neural network for blood pressure prediction based on photoplethysmography. *Biomedical Signal Processing and Control*, 68:102772, 2021.

[6] Jeffrey D Stanaway, Ashkan Afshin, Emmanuela Gakidou, Stephen S Lim, Degu Abate, Kalkidan Hassen Abate, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):1923–1994, 2018.

[7] Gerard J Tortora and Bryan H Derrickson. *Principles of anatomy and physiology*. John Wiley & Sons, 2018.

[8] Carlene MM Lawes, Stephen Vander Hoorn, Anthony Rodgers, et al. Global burden of blood-pressure-related disease, 2001. *The Lancet*, 371(9623):1513–1518, 2008.

[9] Stephen S Lim, Theo Vos, Abraham D Flaxman, Goodarz Danaei, Kenji Shibuya, Heather Adair-Rohani, Mohammad A AlMazroa, Markus Amann, H Ross Anderson, Kathryn G Andrews, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The lancet*, 380(9859):2224–2260, 2012.

[10] World Health Organization. *World health statistics 2015*. World Health Organization, 2015.

[11] Teemu Koivistoinen, Leo-Pekka Lyytikäinen, Heikki Aatola, Tiina Luukkaala, Markus Juonala, Jorma Viikari, Terho Lehtimäki, Olli T Raitakari, Mika Kähönen, and Nina Hutri-Kähönen. Pulse wave velocity predicts the progression of blood pressure and development of hypertension in young adults. *Hypertension*, 71(3):451–456, 2018.

[12] Xiao-Rong Ding, Ni Zhao, Guang-Zhong Yang, Roderic I Pettigrew, Benny Lo, Fen Miao, Ye Li, Jing Liu, and Yuan-Ting Zhang. Continuous blood pressure measurement from invasive to unobtrusive: Celebration of 200th birth anniversary of carl ludwig. *IEEE journal of biomedical and health informatics*, 20(6):1455–1465, 2016.

[13] Leslie Alexander Geddes. *Handbook of blood pressure measurement*. Springer Science & Business Media, 2013.

[14] Daniel Levy, Martin G Larson, Ramachandran S Vasan, William B Kannel, and Kalon KL Ho. The progression from hypertension to congestive heart failure. *Jama*, 275(20):1557–1562, 1996.

[15] CCY Poon and YT Zhang. Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time. In *2005 IEEE engineering in medicine and biology 27th annual conference*, pages 5877–5880. IEEE, 2006.

[16] Niranjan Kumar, Amogh Agrawal, and Sujay Deb. Cuffless bp measurement using a correlation study of pulse transient time and heart rate. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1538–1541. IEEE, 2014.

[17] Yan Chen, Changyun Wen, Guocai Tao, Min Bi, and Guoqi Li. Continuous and noninvasive blood pressure measurement: a novel modeling methodology of the relationship between blood pressure and pulse wave velocity. *Annals of biomedical engineering*, 37(11):2222–2233, 2009.

[18] Meir Nitzan, B Khanokh, and Y Slovik. The difference in pulse transit time to the toe and finger measured by photoplethysmography. *Physiological measurement*, 23(1):85, 2001.

[19] Yan Chen, Changyun Wen, Guocai Tao, and Min Bi. 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. *Annals of biomedical engineering*, 40(4):871–882, 2012.

[20] Ramakrishna Mukkamala, Jin-Oh Hahn, Omer T Inan, Lalit K Mestha, Chang-Sei Kim, Hakan Töreyin, and Survi Kyal. Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice. *IEEE Transactions on Biomedical Engineering*, 62(8):1879–1901, 2015.

[21] LA Geddes, MH Voelz, CF Babbs, JD Bourland, and WA Tacker. Pulse transit time as an indicator of arterial blood pressure. *psychophysiology*, 18(1):71–74, 1981.

[22] Guanqun Zhang, Mingwu Gao, Da Xu, N Bari Olivier, and Ramakrishna Mukkamala. Pulse arrival time is not an adequate surrogate for pulse transit time as a marker of blood pressure. *Journal of applied physiology*, 111(6):1681–1686, 2011.

[23] RA Payne, CN Symeonides, DJ Webb, and SRJ Maxwell. Pulse transit time measured from the ecg: an unreliable marker of beat-to-beat blood pressure. *Journal of Applied Physiology*, 100(1):136–141, 2006.

[24] XF Teng and YT Zhang. Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach. In *Proceedings of the*

25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439), volume 4, pages 3153–3156. IEEE, 2003.

[25] Yuriy Kurylyak, Francesco Lamonaca, and Domenico Grimaldi. A neural network-based method for continuous blood pressure estimation from a ppg signal. In *2013 IEEE International instrumentation and measurement technology conference (I2MTC)*, pages 280–283. IEEE, 2013.

[26] Syed Ghufran Khalid, Jufen Zhang, Fei Chen, and Dingchang Zheng. Blood pressure estimation using photoplethysmography only: comparison between different machine learning approaches. *Journal of healthcare engineering*, 2018, 2018.

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

[28] Xiaoman Xing and Mingshan Sun. Optical blood pressure estimation with photoplethysmography and fft-based neural networks. *Biomedical optics express*, 7(8):3007–3020, 2016.

[29] Peng Su, Xiao-Rong Ding, Yuan-Ting Zhang, Jing Liu, Fen Miao, and Ni Zhao. Long-term blood pressure prediction with deep recurrent neural networks. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 323–328. IEEE, 2018.

[30] Md Sayed Tanveer and Md Kamrul Hasan. Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ann-lstm network. *Biomedical Signal Processing and Control*, 51:382–392, 2019.

[31] Yung-Hui Li, Latifa Nabila Harfiya, Kartika Purwandari, and Yue-Der Lin. Real-time cuffless continuous blood pressure estimation using deep learning model. *Sensors*, 20(19):5606, 2020.

[32] Abraham Noordergraaf. *Circulatory system dynamics*, volume 1. Elsevier, 2012.

[33] Elaine Nicpon Marieb and Katja Hoehn. *Human anatomy & physiology*. Pearson education, 2007.

[34] John E Hall and Michael E Hall. *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2020.

[35] John KJ Li. *Dynamics of the vascular system*, volume 1. World scientific, 2004.

[36] Lukáš Peter, Norbert Noury, and M Cerny. A review of methods for noninvasive and continuous blood pressure monitoring: Pulse transit time method is promising? *Irbm*, 35(5):271–282, 2014.

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

[38] Nicolaas Westerhof, Nikolaos Stergiopulos, Mark IM Noble, and Berend E Westerhof. *Snapshots of hemodynamics: an aid for clinical research and graduate education*. Springer, 2018.

[39] Jacques Blacher, Roland Asmar, Saliha Djane, Gérard M London, and Michel E Safar. Aortic pulse wave velocity as a marker of cardiovascular risk in hypertensive patients. *Hypertension*, 33(5):1111–1117, 1999.

[40] Gianfranco Parati, Juan E Ochoa, Carolina Lombardi, and Grzegorz Bilo. Assessment and management of blood-pressure variability. *Nature Reviews Cardiology*, 10(3):143, 2013.

[41] ARTHUR C Guyton. The relationship of cardiac output and arterial pressure control. *Circulation*, 64(6):1079–1088, 1981.

[42] Michael Gurven, Aaron D Blackwell, Daniel Eid Rodríguez, Jonathan Stieglitz, and Hillard Kaplan. Does blood pressure inevitably rise with age? longitudinal evidence among forager-horticulturalists. *Hypertension*, 60(1):25–33, 2012.

[43] Susan Cheng, Vanessa Xanthakis, Lisa M Sullivan, and Ramachandran S Vasan. Blood pressure tracking over the adult life course: patterns and correlates in the framingham heart study. *Hypertension*, 60(6):1393–1399, 2012.

[44] Carmel M McEniery, null Yasmin, Ian R Hall, Ahmad Qasem, Ian B Wilkinson, John R Cockcroft, and ACCT Investigators. Normal vascular aging:

differential effects on wave reflection and aortic pulse wave velocity: the anglo-cardiff collaborative trial (acct). *Journal of the American College of Cardiology*, 46(9):1753–1760, 2005.

[45] Raymond R Townsend, Ian B Wilkinson, Ernesto L Schiffrin, Alberto P Avolio, Julio A Chirinos, John R Cockcroft, Kevin S Heffernan, Edward G Lakatta, Carmel M McEniery, Gary F Mitchell, et al. Recommendations for improving and standardizing vascular research on arterial stiffness: a scientific statement from the american heart association. *Hypertension*, 66(3):698–722, 2015.

[46] Kozo Hirata, Masanobu Kawakami, and Michael F O'Rourke. Pulse wave analysis and pulse wave velocity a review of blood pressure interpretation 100 years after korotkov. *Circulation journal*, 70(10):1231–1239, 2006.

[47] Bryan Williams, Giuseppe Mancia, Wilko Spiering, Enrico Agabiti Rosei, Michel Azizi, Michel Burnier, Denis L Clement, Antonio Coca, Giovanni De Simone, Anna Dominiczak, et al. 2018 esc/esh guidelines for the management of arterial hypertension: The task force for the management of arterial hypertension of the european society of cardiology (esc) and the european society of hypertension (esh). *European heart journal*, 39(33):3021–3104, 2018.

[48] Jan A Staessen, G Ginocchio, Lutgarde Thijs, and Robert Fagard. Conventional and ambulatory blood pressure and menopause in a prospective population study. *Journal of human hypertension*, 11(8):507–514, 1997.

[49] D Martins, K Nelson, D Pan, N Tareen, and K Norris. The effect of gender on age-related blood pressure changes and the prevalence of isolated systolic hypertension among older adults: data from nhanes iii. *The journal of gender-specific medicine: JGSM: the official journal of the Partnership for Women's Health at Columbia*, 4(3):10–3, 2001.

[50] Niels Wiinberg, Asbjørn Høegholm, Hanne R Christensen, Lia E Bang, Kim L Mikkelsen, Poul Ebbe Nielsen, Tage L Svendsen, Jens P Kampmann, Niels H Madsen, and Michael W Bentzon. 24-h ambulatory blood pressure in 352 normal danish subjects, related to age and gender. *American journal of hypertension*, 8(10):978–986, 1995.

[51] Sleman Khoury, Steven A Yavows, Thomas K O'Brien, and James R Sowers.

Ambulatory blood pressure monitoring in a nonacademic setting: effects of age and sex. *American journal of hypertension*, 5(9):616–623, 1992.

[52] Vicki L Burt, Paul Whelton, Edward J Roccella, Clarice Brown, Jeffrey A Cutler, Millicent Higgins, Michael J Horan, and Darwin Labarthe. Prevalence of hypertension in the us adult population: results from the third national health and nutrition examination survey, 1988-1991. *Hypertension*, 25(3):305–313, 1995.

[53] Jane F Reckelhoff. Gender differences in the regulation of blood pressure. *Hypertension*, 37(5):1199–1208, 2001.

[54] Tamara K Waddell, Anthony M Dart, Christoph D Gatzka, James D Cameron, and Bronwyn A Kingwell. Women exhibit a greater age-related increase in proximal aortic stiffness than men. *Journal of hypertension*, 19(12):2205–2212, 2001.

[55] Harold Smulyan, Roland G Asmar, Annie Rudnicki, Gerard M London, and Michel E Safar. Comparative effects of aging in men and women on the properties of the arterial tree. *Journal of the American College of Cardiology*, 37(5):1374–1380, 2001.

[56] Stanley S Franklin, William Gustin IV, Nathan D Wong, Martin G Larson, Michael A Weber, William B Kannel, and Daniel Levy. Hemodynamic patterns of age-related changes in blood pressure: the framingham heart study. *Circulation*, 96(1):308–315, 1997.

[57] Elisabete Pinto. Blood pressure and ageing. *Postgraduate medical journal*, 83(976):109–114, 2007.

[58] Seamus P Whelton, Ashley Chin, Xue Xin, and Jiang He. Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. *Annals of internal medicine*, 136(7):493–503, 2002.

[59] I-Min Lee and PATRICK J Skerrett. Physical activity and all-cause mortality: what is the dose-response relation? *Medicine and science in sports and exercise*, 33(6; SUPP):S459–S471, 2001.

[60] Marc Nocon, Theresa Hiemann, Falk Müller-Riemenschneider, Frank Thalau, Stephanie Roll, and Stefan N Willich. Association of physical activity with

all-cause and cardiovascular mortality: a systematic review and meta-analysis. *European Journal of Preventive Cardiology*, 15(3):239–246, 2008.

[61] Amanda Rossi, Anastasia Dikareva, Simon L Bacon, and Stella S Daskalopoulou. The impact of physical activity on mortality in patients with high blood pressure: a systematic review. *Journal of hypertension*, 30(7):1277–1288, 2012.

[62] Mitsugi Motoyama, Yoshiyuki Sunami, Fujihisa Kinoshita, Akira Kiyonaga, Hiroaki Tanaka, Munehiro Shindo, Takashi Irie, Hidenori Urata, Jun Sasaki, and Kikuo Arakawa. Blood pressure lowering effect of low intensity aerobic training in elderly hypertensive patients. *Medicine and Science in sports and Exercise*, 30(6):818–823, 1998.

[63] World Health Organization. *Global recommendations on physical activity for health*. World Health Organization, 2010.

[64] Mette S Olufsen, Johnny T Ottesen, Hien T Tran, Laura M Ellwein, Lewis A Lipsitz, and Vera Novak. Blood pressure and blood flow variation during postural change from sitting to standing: model development and validation. *Journal of applied physiology*, 99(4):1523–1537, 2005.

[65] Gianfranco Parati, Guido Pomidossi, Fabio Albini, Daniele Malaspina, and Giuseppe Mancia. Relationship of 24-hour blood pressure mean and variability to severity of target-organ damage in hypertension. *Journal of hypertension*, 5(1):93–98, 1987.

[66] Gianfranco Parati, Juan E Ochoa, and Grzegorz Bilo. Blood pressure variability, cardiovascular risk, and risk for renal disease progression. *Current hypertension reports*, 14(5):421–431, 2012.

[67] Patricia M Kearney, Megan Whelton, Kristi Reynolds, Paul Muntner, Paul K Whelton, and Jiang He. Global burden of hypertension: analysis of worldwide data. *The lancet*, 365(9455):217–223, 2005.

[68] Guillaume Bobrie, Pierre Clerson, Joel Menard, Nicolas Postel-Vinay, Gilles Chatellier, and Pierre-Francois Plouin. Masked hypertension: a systematic review. *Journal of hypertension*, 26(9):1715–1725, 2008.

[69] Stephen K Glen, Henry L Elliott, Joan L Curzio, Kennedy R Lees, and John L Reid. White-coat hypertension as a cause of cardiovascular dysfunction. *The Lancet*, 348(9028):654–657, 1996.

[70] Peter B Raven and Mark W Chapleau. Blood pressure regulation xi: overview and future research directions. *European journal of applied physiology*, 114(3):579–586, 2014.

[71] Maria-Cláudia Irigoyen, Kátia De Angelis, Fernando Dos Santos, Daniela R Dartora, Bruno Rodrigues, and Fernanda Marciano Consolim-Colombo. Hypertension, blood pressure variability, and target organ lesion. *Current hypertension reports*, 18(4):31, 2016.

[72] World Health Organization et al. A global brief on hypertension: silent killer, global public health crisis: World health day 2013. Technical report, World Health Organization, 2013.

[73] Jiang He and Paul K Whelton. Elevated systolic blood pressure and risk of cardiovascular and renal disease: overview of evidence from observational epidemiologic studies and randomized controlled trials. *American heart journal*, 138(3):S211–S219, 1999.

[74] Manish Hosanee, Gabriel Chan, Kaylie Welykholowa, Rachel Cooper, Panayiotis A Kyriacou, Dingchang Zheng, John Allen, Derek Abbott, Carlo Menon, Nigel H Lovell, et al. Cuffless single-site photoplethysmography for blood pressure monitoring. *Journal of clinical medicine*, 9(3):723, 2020.

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

[76] Manuja Sharma, Karinne Barbosa, Victor Ho, Devon Griggs, Tadesse Ghirmai, Sandeep K Krishnan, Tzung K Hsiai, Jung-Chih Chiao, and Hung Cao. Cuff-less and continuous blood pressure monitoring: a methodological review. *Technologies*, 5(2):21, 2017.

[77] Nicola Pirrone and Kathryn R Mahaffey. *Dynamics of mercury pollution on regional and global scales: atmospheric processes and human exposures around the world*. Springer Science & Business Media, 2005.

[78] Dorothee Perloff, Carlene Grim, John Flack, Edward D Frohlich, Martha Hill, Mary McDonald, and Bruce Z Morgenstern. Human blood pressure determination by sphygmomanometry. *Circulation*, 88(5):2460–2470, 1993.

[79] Gary Drzewiecki, Rebecca Hood, and H Apple. Theory of the oscillometric maximum and the systolic and diastolic detection ratios. *Annals of biomedical engineering*, 22(1):88–96, 1994.

[80] Bruce S Alpert, David Quinn, and David Gallick. Oscillometric blood pressure: a review for clinicians. *Journal of the American Society of Hypertension*, 8(12):930–938, 2014.

[81] P Bonnafoux. Auscultatory and oscillometric methods of ambulatory blood pressure monitoring, advantages and limits: a technical point of view. *Blood pressure monitoring*, 1(3):181–185, 1996.

[82] J Penaz. Photoelectric measurement of blood pressure, volume and flow in the finger. *International Conference on Medical and Biological Engineering*, 104, 1973.

[83] Ben PM Imholz, Wouter Wieling, Gert A van Montfrans, and Karel H Wesseling. Fifteen years experience with finger arterial pressure monitoring: assessment of the technology. *Cardiovascular research*, 38(3):605–616, 1998.

[84] Gianfranco Parati, Roberto Casadei, Antonella Groppelli, Marco Di Rienzo, and Giuseppe Mancia. Comparison of finger and intra-arterial blood pressure monitoring at rest and during laboratory testing. *Hypertension*, 13(6_pt_1):647–655, 1989.

[85] Julia Y Wagner, Ileana Negulescu, Miriam Schöfthaler, Alexander Hapfelmeier, Agnes S Meidert, Wolfgang Huber, Roland M Schmid, and Bernd Saugel. Continuous noninvasive arterial pressure measurement using the volume clamp method: an evaluation of the cnap device in intensive care unit patients. *Journal of clinical monitoring and computing*, 29(6):807–813, 2015.

[86] BPM Imholz, W Wieling, GJ Langewouters, and GA Van Montfrans. Continuous finger arterial pressure: utility in the cardiovascular laboratory. *Clinical Autonomic Research*, 1(1):43–53, 1991.

[87] Heather L Ristuccia, Paul Grossman, Lana L Watkins, and Bernard Lown. Incremental bias in finapres estimation of baseline blood pressure levels over time. *Hypertension*, 29(4):1039–1043, 1997.

[88] Gary M Drzewiecki, Julius Melbin, and Abraham Noordergraaf. Arterial tonometry: review and analysis. *Journal of biomechanics*, 16(2):141–152, 1983.

[89] Koen Matthys and Pascal Verdonck. Development and modelling of arterial applanation tonometry: a review. *Technology and Health Care*, 10(1):65–76, 2002.

[90] Clifford L Cua, Kristi Thomas, David Zurakowski, and Peter C Laussen. A comparison of the vasotrac with invasive arterial blood pressure monitoring in children after pediatric cardiac surgery. *Anesthesia & Analgesia*, 100(5):1289–1294, 2005.

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

[92] Meir Nitzan, Yair Adar, Ellie Hoffman, Eran Shalom, Shlomo Engelberg, Iddo Z Ben-Dov, and Michael Bursztyn. Comparison of systolic blood pressure values obtained by photoplethysmography and by korotkoff sounds. *Sensors*, 13(11):14797–14812, 2013.

[93] A Reşit Kavsaoğlu, Kemal Polat, and M Hariharan. Non-invasive prediction of hemoglobin level using machine learning techniques with the ppg signal's characteristics features. *Applied Soft Computing*, 37:983–991, 2015.

[94] Hongyun Liu, Fulai Peng, Minlu Hu, Jinlong Shi, Guojing Wang, Haiming Ai, and Weidong Wang. Development and validation of a photoplethysmography system for noninvasive monitoring of hemoglobin concentration. *Journal of Electrical and Computer Engineering*, 2020, 2020.

[95] Enric Monte-Moreno. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. *Artificial intelligence in medicine*, 53(2):127–138, 2011.

[96] Swathi Ramasahayam, Lavanya Arora, and Shubhajit Roy Chowdhury. Fpga based smart system for non invasive blood glucose sensing using photoplethys-

mography and online correction of motion artifact. In *Sensors for Everyday Life*, pages 1–21. Springer, 2017.

[97] Swaathi Venkat, Mohamed Tanveejul PS Arsath PS, Annamol Alex, SP Preejith, DJ Christopher, Jayaraj Joseph, Mohanasankar Sivaprakasam, et al. Machine learning based spo 2 computation using reflectance pulse oximetry. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 482–485. IEEE, 2019.

[98] Sricharan Vijayarangan, Prithvi Suresh, SP Preejith, Jayaraj Joseph, and Mohansankar Sivaprakasam. Robust modelling of reflectance pulse oximetry for spo 2 estimation. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 374–377. IEEE, 2020.

[99] Rohan Banerjee, Ramu Vempada, KM Mandana, Anirban Dutta Choudhury, and Arpan Pal. Identifying coronary artery disease from photoplethysmogram. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1084–1088, 2016.

[100] Neeraj Paradkar and Shubhajit Roy Chowdhury. Coronary artery disease detection using photoplethysmography. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 100–103. IEEE, 2017.

[101] Aamir Arsalan, Muhammad Majid, Syed Muhammad Anwar, and Ulas Bagci. Classification of perceived human stress using physiological signals. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1247–1250. IEEE, 2019.

[102] Yekta Said Can, Niaz Chalabianloo, Deniz Ekiz, and Cem Ersoy. Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors*, 19(8):1849, 2019.

[103] Dongrae Cho, Jinsil Ham, Jooyoung Oh, Jeanho Park, Sayup Kim, Nak-Kyu Lee, and Boreom Lee. Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine. *Sensors*, 17(10):2435, 2017.

[104] Ainara Garde, Walter Karlen, Parastoo Dehkordi, John Mark Ansermino, and Guy A Dumont. Empirical mode decomposition for respiratory and heart rate estimation from the photoplethysmogram. In *Computing in cardiology 2013*, pages 799–802. IEEE, 2013.

[105] Vignesh Ravichandran, Balamurali Murugesan, Vaishali Balakarthikeyan, Keerthi Ram, SP Preejith, Jayaraj Joseph, and Mohanasankar Sivaprakasam. Respnet: A deep learning model for extraction of respiration from photoplethysmogram. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5556–5559. IEEE, 2019.

[106] Simhadri Vadrevu and M Sabarimalai Manikandan. Real-time ppg signal quality assessment system for improving battery life and false alarms. *IEEE transactions on circuits and systems II: express briefs*, 66(11):1910–1914, 2019.

[107] Tania Pereira, Cheng Ding, Kais Gadhoumi, Nate Tran, Rene A Colorado, Karl Meisel, and Xiao Hu. Deep learning approaches for plethysmography signal quality assessment in the presence of atrial fibrillation. *Physiological measurement*, 40(12):125002, 2019.

[108] C El-Hajj and Panayiotis A Kyriacou. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. *Biomedical Signal Processing and Control*, 58:101870, 2020.

[109] John G Webster. *Prevention of pressure sores: engineering and clinical aspects*. CRC Press, 1991.

[110] Brian Gribbin, Andrew Steptoe, and Peter Sleight. Pulse wave velocity as a measure of blood pressure change. *Psychophysiology*, 13(1):86–90, 1976.

[111] Jessica E Wagenseil and Robert P Mecham. Elastin in large artery stiffness and hypertension. *Journal of cardiovascular translational research*, 5(3):264–273, 2012.

[112] Devin B McCombie, Andrew T Reisner, and H Harry Asada. Adaptive blood pressure estimation from wearable ppg sensors using peripheral artery pulse wave velocity measurements and multi-channel blind identification of local

arterial dynamics. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3521–3524. IEEE, 2006.

[113] Mohamad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. In *2015 IEEE international symposium on circuits and systems (ISCAS)*, pages 1006–1009. IEEE, 2015.

[114] Mohammad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuffless blood pressure estimation algorithms for continuous health-care monitoring. *IEEE Transactions on Biomedical Engineering*, 64(4):859–869, 2016.

[115] Heiko Gesche, Detlef Grosskurth, Gert Küchler, and Andreas Patzak. Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method. *European journal of applied physiology*, 112(1):309–315, 2012.

[116] Mico Yee-Man Wong, Carmen Chung-Yan Poon, and Yuan-Ting Zhang. An evaluation of the cuffless blood pressure estimation based on pulse transit time technique: a half year study on normotensive subjects. *Cardiovascular Engineering*, 9(1):32–38, 2009.

[117] Mingwu Gao, N Bari Olivier, and Ramakrishna Mukkamala. Comparison of noninvasive pulse transit time estimates as markers of blood pressure using invasive pulse transit time measurements as a reference. *Physiological reports*, 4(10):e12768, 2016.

[118] J Allen and A Murray. Age-related changes in peripheral pulse timing characteristics at the ears, fingers and toes. *Journal of human hypertension*, 16(10):711–717, 2002.

[119] W Chen, T Kobayashi, S Ichikawa, Y Takeuchi, and T Togawa. Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration. *Medical and Biological Engineering and Computing*, 38(5):569–574, 2000.

[120] Mohamad Forouzanfar, Saif Ahmad, Izmail Batkin, Hilmi R Dajani, Voicu Z Groza, and Miodrag Bolic. Model-based mean arterial pressure estimation us-

ing simultaneous electrocardiogram and oscillometric blood pressure measurements. *IEEE Transactions on Instrumentation and Measurement*, 64(9):2443–2452, 2015.

[121] T Ma and Yuan-Ting Zhang. A correlation study on the variabilities in pulse transit time, blood pressure, and heart rate recorded simultaneously from healthy subjects. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 996–999. IEEE, 2006.

[122] Chang-Sei Kim, Andrew M Carek, Ramakrishna Mukkamala, Omer T Inan, and Jin-Oh Hahn. Ballistocardiogram as proximal timing reference for pulse transit time measurement: Potential for cuffless blood pressure monitoring. *IEEE Transactions on Biomedical Engineering*, 62(11):2657–2664, 2015.

[123] Mohamed Elgendi. On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews*, 8(1):14–25, 2012.

[124] Sandrine C Millasseau, Franck G Guigui, Ronan P Kelly, Krishna Prasad, John R Cockcroft, James M Ritter, and Philip J Chowienczyk. Noninvasive assessment of the digital volume pulse: comparison with the peripheral pressure pulse. *Hypertension*, 36(6):952–956, 2000.

[125] Muhamad Khairul Bin Ali Hassan, MY Mashor, NF Mohd Nasir, and S Mohamed. Measuring blood pressure using a photoplethysmography approach. In *4th Kuala Lumpur International Conference on Biomedical Engineering 2008*, pages 591–594. Springer, 2008.

[126] Zixiao Shen, Fen Miao, Qinghan Meng, and Ye Li. Cuffless and continuous blood pressure estimation based on multiple regression analysis. In *2015 5th International Conference on Information Science and Technology (ICIST)*, pages 117–120. IEEE, 2015.

[127] Fen Miao, Nan Fu, Yuan-Ting Zhang, Xiao-Rong Ding, Xi Hong, Qingyun He, and Ye Li. A novel continuous blood pressure estimation approach based on data mining techniques. *IEEE journal of biomedical and health informatics*, 21(6):1730–1740, 2017.

[128] Ludi Wang, Wei Zhou, Ying Xing, and Xiaoguang Zhou. A novel neural net-

work model for blood pressure estimation using photoplethesmography without electrocardiogram. *Journal of healthcare engineering*, 2018, 2018.

[129] Shota Shimazaki, Shoaib Bhuiyan, Haruki Kawanaka, and Koji Oguri. Features extraction for cuffless blood pressure estimation by autoencoder from photoplethysmography. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2857–2860. IEEE, 2018.

[130] Shuo Chen, Zhong Ji, Haiyan Wu, and Yingchao Xu. A non-invasive continuous blood pressure estimation approach based on machine learning. *Sensors*, 19(11):2585, 2019.

[131] Vicent Ribas Ripoll and Alfredo Vellido. Blood pressure assessment with differential pulse transit time and deep learning: a proof of concept. *Kidney Diseases*, 5(1):23–27, 2019.

[132] Heesang Eom, Dongseok Lee, Seungwoo Han, Yuli Sun Hariyani, Yonggyu Lim, Illsoo Sohn, Kwangsuk Park, and Cheolsoo Park. End-to-end deep learning architecture for continuous blood pressure estimation using attention mechanism. *Sensors*, 20(8):2338, 2020.

[133] Dongseok Lee, Hyunbin Kwon, Dongyeon Son, Heesang Eom, Cheolsoo Park, Yonggyu Lim, Chulhun Seo, and Kwangsuk Park. Beat-to-beat continuous blood pressure estimation using bidirectional long short-term memory network. *Sensors*, 21(1):96, 2021.

[134] Satomi Suzuki and Koji Oguri. Cuffless blood pressure estimation by error-correcting output coding method based on an aggregation of adaboost with a photoplethysmograph sensor. In *2009 Annual international conference of the IEEE engineering in medicine and biology society*, pages 6765–6768. IEEE, 2009.

[135] Juan C Ruiz-Rodríguez, Adolf Ruiz-Sanmartín, Vicent Ribas, Jesús Caballero, Alejandra García-Roche, Jordi Riera, Xavier Nuvials, Miriam de Nadal, Oriol de Sola-Morales, Joaquim Serra, et al. Innovative continuous non-invasive cuffless blood pressure monitoring based on photoplethysmography technology. *Intensive care medicine*, 39(9):1618–1625, 2013.

[136] Arata Suzuki and Kazuteru Ryu. Feature selection method for estimating systolic blood pressure using the taguchi method. *IEEE Transactions on Industrial Informatics*, 10(2):1077–1085, 2013.

[137] Anirban Dutta Choudhury, Rohan Banerjee, Aniruddha Sinha, and Shaswati Kundu. Estimating blood pressure using windkessel model on photoplethysmogram. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4567–4570. IEEE, 2014.

[138] Shreyasi Datta, Rohan Banerjee, Anirban Dutta Choudhury, Aniruddha Sinha, and Arpan Pal. Blood pressure estimation from photoplethysmogram using latent parameters. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2016.

[139] Kefeng Duan, Zhiliang Qian, Mohamed Atef, and Guoxing Wang. A feature exploration methodology for learning based cuffless blood pressure measurement using photoplethysmography. In *2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 6385–6388. IEEE, 2016.

[140] Aman Gaurav, Maram Maheedhar, Vijay N Tiwari, and Rangavittal Narayanan. Cuff-less ppg based continuous blood pressure monitoring—a smartphone based approach. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 607–610. IEEE, 2016.

[141] Shi Chao Gao, Peter Wittek, Li Zhao, and Wen Jun Jiang. Data-driven estimation of blood pressure using photoplethysmographic signals. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 766–769. IEEE, 2016.

[142] Jishnu Dey, Aman Gaurav, and Vijay N Tiwari. Instabp: Cuff-less blood pressure monitoring on smartphone using single ppg sensor. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5002–5005. IEEE, 2018.

[143] Daisuke Fujita, Arata Suzuki, and Kazuteru Ryu. Ppg-based systolic blood

pressure estimation method using pls and level-crossing feature. *Applied Sciences*, 9(2):304, 2019.

[144] Navid Hasanzadeh, Mohammad Mahdi Ahmadi, and Hoda Mohammadzade. Blood pressure estimation using photoplethysmogram signal and its morphological features. *IEEE Sensors Journal*, 20(8):4300–4310, 2019.

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

[146] Moajjem Hossain Chowdhury, Md Nazmul Islam Shuzan, Muhammad EH Chowdhury, Zaid B Mahbub, M Monir Uddin, Amith Khandakar, and Mamun Bin Ibne Reaz. Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques. *Sensors*, 20(11):3127, 2020.

[147] Muammar Sadrawi, Yin-Tsong Lin, Chien-Hung Lin, Bhekumuzi Mathunjwa, Shou-Zen Fan, Maysam F Abbod, and Jiann-Shing Shieh. Genetic deep convolutional autoencoder applied for generative continuous arterial blood pressure via photoplethysmography. *Sensors*, 20(14):3829, 2020.

[148] Nabil Ibtehaz and M Sohel Rahman. Ppg2abp: Translating photoplethysmogram (ppg) signals to arterial blood pressure (abp) waveforms using fully convolutional neural networks. *arXiv preprint arXiv:2005.01669*, 2020.

[149] Fabian Schrumpf, Patrick Frenzel, Christoph Aust, Georg Osterhoff, and Mirco Fuchs. 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*, pages 3820–3830, 2021.

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

[151] Latifa Nabila Harfiya, Ching-Chun Chang, and Yung-Hui Li. Continuous blood pressure estimation using exclusively photopletysmography by lstm-based signal-to-signal translation. *Sensors*, 21(9):2952, 2021.

[152] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

[153] David Liu, Matthias Görges, and Simon A Jenkins. University of queensland vital signs dataset: development of an accessible repository of anesthesia patient monitoring data for research. *Anesthesia & Analgesia*, 114(3):584–589, 2012.

[154] Costas Sideris, Haik Kalantarian, Ebrahim Nemati, and Majid Sarrafzadeh. Building continuous arterial blood pressure prediction models using recurrent networks. In *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 1–5. IEEE, 2016.

[155] Yongbo Liang, Guiyong Liu, Zhencheng Chen, and Mohamed Elgendi. Ppg-bp database, 2018.

[156] Zhi-Min Zhang, Shan Chen, and Yi-Zeng Liang. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135(5):1138–1146, 2010.

[157] Aymen A Awad, Ala S Haddadin, Hossam Tantawy, Tarek M Badr, Robert G Stout, David G Silverman, and Kirk H Shelley. The relationship between the photoplethysmographic waveform and systemic vascular resistance. *Journal of clinical monitoring and computing*, 21(6):365–372, 2007.

[158] L Wang, Emma Pickwell-MacPherson, YP Liang, and Yuan Ting Zhang. Non-invasive cardiac output estimation using a novel photoplethysmogram index. In *2009 annual international conference of the IEEE engineering in medicine and biology society*, pages 1746–1749. IEEE, 2009.

[159] Carmen CY Poon, XF Teng, YM Wong, C Zhang, and YT Zhang. Changes in the photoplethysmogram waveform after exercise. In *2004 2nd IEEE/EMBS International Summer School on Medical Devices and Biosensors*, pages 115–118. IEEE, 2004.

[160] Stephen R Alty, Natalia Angarita-Jaimes, Sandrine C Millasseau, and Philip J Chowienczyk. Predicting arterial stiffness from the digital volume pulse waveform. *IEEE Transactions on Biomedical Engineering*, 54(12):2268–2275, 2007.

[161] Xiao-Rong Ding and Yuan-Ting Zhang. Photoplethysmogram intensity ratio: A potential indicator for improving the accuracy of ptt-based cuffless blood pressure estimation. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 398–401. IEEE, 2015.

[162] David N Reshef, Yakir A Reshef, Hilary K Finucane, Sharon R Grossman, Gilean McVean, Peter J Turnbaugh, Eric S Lander, Michael Mitzenmacher, and Pardis C Sabeti. Detecting novel associations in large data sets. *science*, 334(6062):1518–1524, 2011.

[163] Pablo M Granitto, Cesare Furlanello, Franco Biasioli, and Flavia Gasperi. Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products. *Chemometrics and intelligent laboratory systems*, 83(2):83–90, 2006.

[164] Harris Drucker, Chris JC Burges, Linda Kaufman, Alex Smola, Vladimir Vapnik, et al. Support vector regression machines. *Advances in neural information processing systems*, 9:155–161, 1997.

[165] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[166] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[167] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.

[168] Y. Freund and R. Schapire. Experiments with a new boosting algorithm. In *ICML*, 1996.

[169] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, USA, 2nd edition, 1998.

[170] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[171] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[172] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[173] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[174] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[175] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[176] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[177] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.

[178] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[179] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[180] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

[181] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *arXiv preprint arXiv:1409.3215*, 2014.

[182] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

[183] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057. PMLR, 2015.

[184] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[185] Douglas G Altman and J Martin Bland. Measurement in medicine: the analysis of method comparison studies. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 32(3):307–317, 1983.