Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Structural Engineering

Kari Anne Dalheim and Anine Ahlsand

Informed machine learning models to predict blood pressure curves of hypertensive individuals after cumulative physical activity

Master's thesis in Engineering and ICT
Supervisor: Leif Rune Hellevik
Co-supervisor: Jacob Sturdy

June 2021

NTNU
Norwegian University of
Science and Technology

Kari Anne Dalheim and Anine Ahlsand

Informed machine learning models to predict blood pressure curves of hypertensive individuals after cumulative physical activity

Master's thesis in Engineering and ICT
Supervisor: Leif Rune Hellevik
Co-supervisor: Jacob Sturdy
June 2021

Norwegian University of Science and Technology
Faculty of Engineering
Department of Structural Engineering

NTNU
Norwegian University of
Science and Technology

Department of Structural Engineering
Faculty of Engineering
NTNU – Norwegian University of Science and Technology

ACCESSIBILITY: Open

MASTER THESIS 2021

SUBJECT AREA: Biomechanics	DATE: June 10th, 2021	NO. OF PAGES: 71

TITLE:

Informed machine learning models to predict blood pressure curves of hypertensive individuals after cumulative physical activity

BY:

Anine Ahlsand and Kari Anne Dalheim

SUMMARY:

This thesis implements and analyzes models that predict blood pressure curves of hypertensive individuals after a period of a given exercise. The objective of the study was to assess the application of informed machine learning (IML) in this context, and multiple models have been implemented to compare the applicability of different IML approaches.

The medical condition of hypertension is defined by persistently high blood pressure and increases the risk of lifestyle diseases. Norwegian University of Science and Technology, Trondheim, is currently working on a project called My Medical Digital Twin (MyMDT). The project's ambition is to develop a personal medical digital twin that monitors a hypertensive individual's blood pressure and gives individual-specific treatment advice. Our project aims to assist MyMDT by evaluating different machine learning models to predict the effect of physical activity on the blood pressure.

Previous studies have proven that incorporating physics-based knowledge into a machine learning model can encourage the model to make physically consistent and more accurate predictions. Additionally, literature shows that relevant knowledge may be valuable when dealing with insufficient amounts of data. Based on previous studies of IML, this project implements two approaches for IML prediction of blood pressure curves. The first method incorporates a physics-based loss function, and the second method predicts the residual of a physics model's estimate.

This project presents six different models predicting the blood pressure curve of individuals after an exercise period. Data was obtained from a 12-week study at St. Olavs University Hospital, Trondheim. A baseline model with linear regression has been implemented, in addition to two neural networks differentiated by a physics-based loss function. Further, two residual models are implemented and evaluated to examine the applicability of this IML approach. Lastly, a neural network based on large amounts of synthetic data is implemented to assess the effect of a greater amount of data than provided for this project.

Comparison of different models demonstrates the benefits of IML in the task of predicting blood pressure curves. The results show that the implemented IML approaches perform better than the standard machine learning models, and the analyses indicate that less data is required for IML to achieve equal performance quality.

SUPERVISORS: Professor Leif Rune Hellevik and Doctor Jacob Sturdy

CARRIED OUT AT: Norwegian University of Science and Technology

Engineering and ICT
TKT4950 Master's thesis

Informed machine learning models to predict blood pressure curves of hypertensive individuals after cumulative physical activity

Anine Ahlsand and Kari Anne Dalheim

Supervisors:
Professor Leif Rune Hellevik, Doctor Jacob Sturdy

Trondheim, June 10th, 2021

NTNU
Norwegian University of
Science and Technology

Faculty of Engineering
DEPARTMENT OF STRUCTURAL ENGINEERING

Preface

This Master's thesis was conducted during the spring of 2021 at Norwegian University of Science and Technology (NTNU), Trondheim, Norway. The thesis was motivated by a collaboration with My Medical Digital Twin (MyMDT), an ongoing multidisciplinary project at NTNU.

We would like to thank our supervisors Professor Leif Rune Hellevik and Doctor Jacob Sturdy for their engagement and support. We would also express our gratitude towards Nikolai Bjørdalsbakke from the MyMDT team for his valuable guidance and discussions throughout the semester.

Finally, we would like to thank NTNU and our fellow students for six great and memorable years.

The implemented code is accessable from: `https://doi.org/10.5281/zenodo.4917117`.

Kari Anne Dalheim and Anine Ahlsand,

Trondheim, June 10th, 2021

Contents

List of Figures

List of Tables

Acronyms

Adam adaptive moment estimation. 14, 35, 38, 39

AI artificial intelligence. 2, 10

API application programming interface. 30

BMI body mass index. 17, 18, 21

CO cardiac output. 9

CP cardiac power. 9

CV cross-validation. 30–32

DBP diastolic blood pressure. 4, 17, 19, 21, 23–26, 28, 32, 41, 42, 44, 48, 51

IML informed machine learning. 2, 14, 15, 29, 30, 47, 50, 52–54

MAP mean arterial pressure. 4, 9, 41, 42, 44

MSE mean squared error. xi, 13, 14, 35–39, 41, 43, 49, 50, 52, 53

MyMDT My Medical Digital Twin. 1, 9, 28, 33, 37, 38, 42, 47

NTNU Norwegian University of Science and Technology. 1

ODEs ordinary differential equations. 28

PAI personal activity intelligence. 19, 26, 27, 34

PC principal component. 11, 34, 48

PCA principal component analysis. 11, 34, 48

PP pulse pressure. 4, 32, 41, 42, 44, 49, 52

ReLU Rectified Linear Unit. 13, 35, 38, 39

RHR relative heart rate. 9

SBP systolic blood pressure. 4, 17, 19, 21, 23–26, 28, 32, 41, 42, 44, 48, 51

SGD stochastic gradient descent. 14

TNR true negative rate. 33

TPR true positive rate. 33

TRRA trust region reflective algorithm. 28

WHO World Health Organization. 1, 19

Abstract

This thesis implements and analyzes models that predict blood pressure curves of hypertensive individuals after a period of a given exercise. The objective of the study was to assess the application of informed machine learning (IML) in this context, and multiple models have been implemented to compare the applicability of different IML approaches.

The medical condition of hypertension is defined by persistently high blood pressure and increases the risk of lifestyle diseases. Norwegian University of Science and Technology, Trondheim, is currently working on a project called My Medical Digital Twin (MyMDT). The project's ambition is to develop a personal medical digital twin that monitors a hypertensive individual's blood pressure and gives individual-specific treatment advice. Our project aims to assist MyMDT by evaluating different machine learning models to predict the effect of physical activity on the blood pressure.

Previous studies have proven that incorporating physics-based knowledge into a machine learning model can encourage the model to make physically consistent and more accurate predictions. Additionally, literature shows that relevant knowledge may be valuable when dealing with insufficient amounts of data. Based on previous studies of IML, this project implements two approaches for IML prediction of blood pressure curves. The first method incorporates a physics-based loss function, and the second method predicts the residual of a physics model's estimate.

This project presents six different models predicting the blood pressure curve of individuals after an exercise period. Data was obtained from a 12-week study at St. Olavs University Hospital, Trondheim. A baseline model with linear regression has been implemented, in addition to two neural networks differentiated by a physics-based loss function. Further, two residual models are implemented and evaluated to examine the applicability of this IML approach. Lastly, a neural network based on large amounts of synthetic data is implemented to assess the effect of a greater amount of data than provided for this project.

Comparison of different models demonstrates the benefits of IML in the task of predicting blood pressure curves. The results show that the implemented IML approaches perform better than the standard machine learning models, and the analyses indicate that less data is required for IML to achieve equal performance quality.

Sammendrag

Denne oppgaven impementerer og analyserer modeller som predikerer blodtrykkskurver for hypertensive personer etter en periode med en gitt trening. Målet med studien var å vurdere anvendelsen av informert maskinlæring (IML) i denne konteksten, og flere modeller har blitt implementert for å sammenligne ulike IML-metoder.

Hypertensjon er definert som vedvarende høyt blodtrykk, og øker risikoen for livsstilssykdommer. Norges teknisk-naturvitenskapelige universitet, Trondheim, jobber for tiden med et projekt som heter My Medical Digital Twin (MyMDT). Prosjektets ambisjon er å utvikle en personlig medisinsk digital tvilling som overvåker en hypertensiv persons blodtrykk og gir individspesifikke behandlingsråd. Vårt prosjekt har som mål å bistå MyMDT ved å evaluere ulike maskinlæringsmodeller for å predikere effekten av fysisk aktivitet på blodtrykket.

Tidligere studier har vist at å inkludere fysikkbasert kunnskap i en maskinlæringsmodell kan hjelpe modellen til å gjøre fysisk riktige og mer nøyaktige prediksjoner. I tillegg viser litteraturen at relevant kunnskap kan være verdifullt når en jobber med utilstrekkelige mengder data. Basert på tidligere studier av IML implementerer dette prosjektet to forskjellige IML-metoder for å predikere blodtrykkskurver. Den ene metoden inkluderer en fysikkbasert tapsfunskjon, og den andre metoden predikerer feilen av etimatet til en fysisk modell.

Dette prosjektet presenterer seks ulike modeller som predikerer blodtrykkskurver til personer etter en treningsperiode. Dataen benyttet i prosjektet er hentet fra en 12-ukers studie ved St. Olavs Universitessykehus, Trondheim. En lineær regresjonsmodell er implementert som et sammenligningsgrunnlag, i tillegg til to nevrale nettverk med ulik tapsfunskjon, der den ene innholder et fysikkbasert ledd. Videre er to residualmodeller implementert og evaluert for å vurdere anvendbarheten til denne IML-metoden. Til slutt er et nevralt nettverk basert på store mendger syntetisk data implementert for å vurdere effekten av en større datamendge enn det som er tilgjengelig i dette prosjektet.

Sammenligning av ulike modeller viser nytten av IML i prediksjoner av blodtrykkskurver. Resultatene viser at de implementerte IML metodene presterer bedre enn de standard maskinlærlingsmodellene, og analysene indikerer at mindre data kreves for IML for å oppnå like gode prediskjoner.

Chapter 1

Introduction

This thesis presents a study on different methods to predict the blood pressure curve of hypertensive individuals after an exercise period. Chapters 1, 2 and 3 are revised and extended from prior work conducted in the project thesis written during the fall of 2020 (Dalheim and Ahlsand, 2020).

The medical condition of hypertension is defined as having a persistently blood pressure above $140/90\mathrm{mmHg}$. The high pressures put an extra load on the vessels, heart, and organs such as the brain and kidneys, and cause a higher risk for lifestyle diseases like heart failure, stroke, and kidney disease (Chobanian et al., 2003a). The World Health Organization (WHO) estimates that more than one billion people have hypertension, and the majority are unaware of the condition as the symptoms may not be present (WHO). Hypertension is the cause of around 9 million deaths every year (Haldar, 2013), and WHO considers it a major cause for premature deaths worldwide.

National Center for Health Statistics in the United States documented that more than half of the individuals treated with hypertension medication between 2013 and 2016 were still hypertensive in 2017 (CDC). A possible reason for this is incorrect medication as common guidelines often are used in treatment. Objective guidelines tend to draw sharp thresholds for treatment, ignoring the multiple interconnected risk factors which provide the need for an individualized treatment (Eddy et al., 2011).

My Medical Digital Twin (MyMDT) is an ongoing interdisciplinary project at Norwegian University of Science and Technology (NTNU), developing a personal medical digital twin. By continuously monitoring important health variables, the project aims to predict how an individual hypertensive individual will respond to a specific form of treatment (Hellevik et al.). This enables the possibility to tailor the treatment of individual profiles. A part of the project intends to fill a gap in previous research by recommending individual amounts of physical exercise, and predict the specific benefits due to it. This project aims to assist MyMDT by implementing models that predict the blood pressure after an exercise period with a given exercise amount.

A strong emerging trend in medicine is the support of artificial intelligence (AI) to help doctors make more personalized decisions. It has been shown that machine learning, which is a branch of AI, has multiple methods and implementations relevant for this use (Melville and Byrd, 2010; Dudchenko et al., 2020). Willard et al. (2020) presents a taxonomy of different methodologies to merge physics-based principles with machine learning. This approach is called informed machine learning (IML), and studies have proved that IML models have outperformed standard machine learning and made more physically consistent predictions. Research has shown that physical activity effectively lowers blood pressure levels (Pedersen and Saltin, 2015; Hackam et al., 2013), but no research has to our knowledge assessed the use of IML to examine this effect.

The objective of this study is to implement and analyze different models for predicting the blood pressure curve of hypertensive individuals after a period of a given exercise. A comparison of different model architectures is carried out to assess the benefits of IML in this context.

This report first presents an overview of the relevant theory. The theory includes physiology and physical principles for cardiovascular modeling, followed by an introduction to machine learning with focus on neural networks and IML. In Chapter 3, the available data for the project is presented, and analyses and processing of the data are described in detail. Further, Chapter 4 explains the architecture and implementation of the different models, and the results are presented and evaluated in Chapter 5. Chapter 6 presents comparison and discussion of the results, and finally the conclusion and proposal for future work are presented in Chapter 7.

Chapter 2

Theory

This chapter introduces the fundamental terminology and theory needed to comprehend the work done in this project. First, the background anatomy of the cardiovascular system is presented, followed by a brief introduction to mechanistic models of the system. Finally, different machine learning architectures are introduced.

2.1 The cardiovascular system

The cardiovascular system is the primary transport system of metabolic substances in the body, and consists of three main components; the heart, the blood vessels and the blood. The heart pumps blood contained in the system through the blood vessels in a closed loop. The blood delivers oxygen and other nutrients to keep the body function properly and gathers carbon dioxide and waste products that need to be removed (Alberta Health). The blood vessels consist of arteries and veins, where the arteries transport oxygen-rich blood from the heart to the body, and the veins transport oxygen-poor blood back to the heart. The heart is divided into two separate pumping systems; the right and left side. The right atrium receives oxygen-poor blood from the veins and pumps it to the right ventricle. The right ventricle ejects and pumps the blood further into the lungs. The lungs replenish the blood with oxygen, and the left atrium receives the re-oxygenated blood from the lungs and pumps it to the left ventricle of the heart. The left ventricle ejects, and the oxygenated blood is pumped to the main artery, the aorta, and further through the smaller arteries which leads it to the rest of the body.

2.1.1 Blood pressure and hypertension

Blood pressure is a result of the blood being pumped from the heart and pushed against the walls of the blood vessels. The highest pressure is achieved during

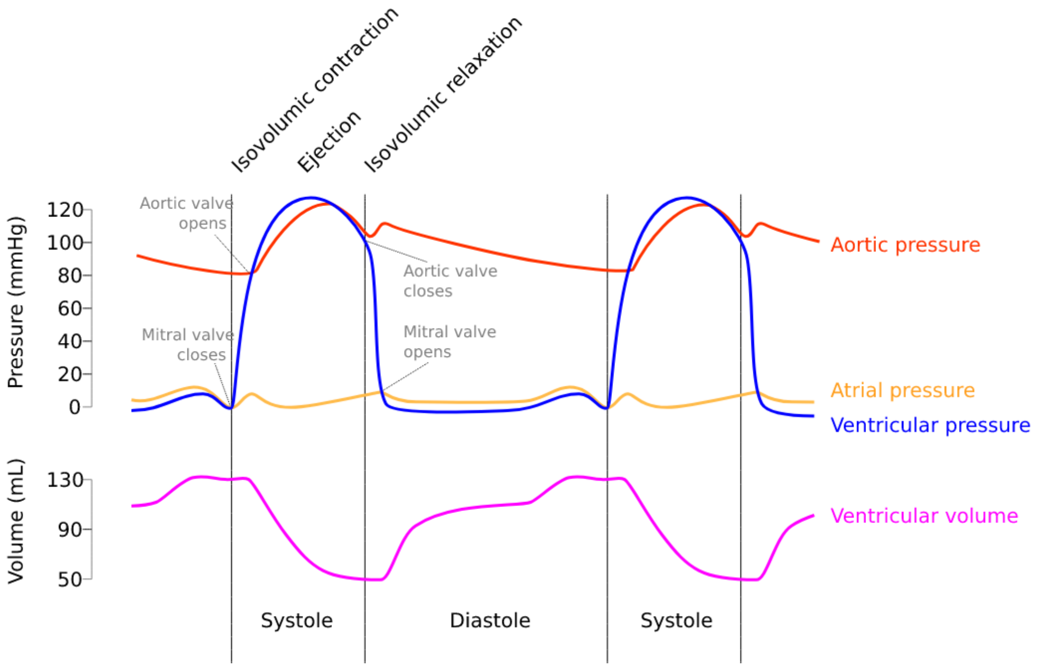


Figure 2.1.1: Illustration of the cardiac cycle with the curves of the aortic pressure, atrial pressure, ventricular pressure, and ventricular volume. Drawn by DestinyQx, redrawn and revised by xavax, DanielChangMD, and adh30, and by the authors of this thesis. Permission from CC BY-SA 4.0.

the active contraction of the left ventricle muscle, and is called the systolic blood pressure (SBP). In contrast, the vessels experience the lowest pressure, called the diastolic blood pressure (DBP), during left ventricle muscle relaxation. Hypertension is a state where a person has persistently high blood pressure. The blood pressure is measured in millimeters of mercury (mmHg), and a SBP persistently above 140 mmHg or a DBP persistently above 90 mmHg is diagnosed as hypertension (Ventura et al., 2005). The difference between SBP and DBP is called the pulse pressure (PP) and represents the force the heart generates when it contracts. The average pressure in a person's arteries during one cardiac cycle is called the mean arterial pressure (MAP). Compared to SBP and DBP, which defines the state of hypertension, MAP is a better indicator of perfusion to vital organs (Lee et al., 1999; Sesso et al., 2000). Figure 2.1.1 illustrates the blood pressure at three different locations in the heart; the aorta, the atriums, and the ventricles, in addition to the ventricular volume, during the cardiac cycle. The ventricular volume is the amount of blood in the ventricles, and is a time-varying variable changing in correspondence with the blood pressure. Each cardiac cycle has a diastole phase where the heart is filled with blood returning from the veins, and a systole phase where the heart contracts and pumps blood out to the arteries. The amount of blood ejected from the left ventricle during the systole phase is referred to as the volumetric flow and is measured in cubic metres per second (m^3/s).

Vascular impedance

Vascular impedance is a metric that represents the relation between corresponding harmonics of blood pressure and volumetric flow, and can be thought of as a frequency dependent resistance (Hellevik, 2015). The concept of impedance is of great importance in biofluid dynamics, and further follows an introduction of two definitions relevant for this project.

Wave transmission and reflection properties in a blood vessel can be derived from linearized mass and momentum equations. These equations are similar to those for so-called transmission-line theory, and characteristic impedance, Z_c, is the most important parameter for any transmission line (Hellevik, 2015; Wilson, 2012). Characteristic impedance is a function of both geometry and materials, and is a dynamic value independent of line length. It is a local characteristic property of the vessel wall and does not consider wave reflections. Characteristic impedance is defined as the ratio between the forward propagating pressure and the forward propagating flow, and is represented as

$$Z_c = \frac{\rho c}{A}, \tag{2.1.1}$$

where ρ denotes the fluid density, c the wave speed, and A the vessel cross-sectional area.

A second impedance definition is the input impedance, Z_i. The input impedance is similar to the characteristic impedance and is defined as the ratio of the pulsatile components of pressure, P, and flow, Q. In contrast, the input impedance is a global quantity and includes reflected wave components. It characterizes the properties in the vessel downstream to the point of measurement, and the cumulative effect of all distal contributions is incorporated in the property (Hellevik, 2015). The input impedance is defined as

$$Z_i(\omega_n) = \frac{P(\omega_n)}{Q(\omega_n)}, \tag{2.1.2}$$

where $P(\omega_n)$ and $Q(\omega_n)$ are the Fourier transforms of blood pressure and volumetric flow in the time domain. ω_n denotes the n different frequencies from the Fourier transforms.

2.2 Mechanistic modelling

Mechanistic models are based on fundamental laws of natural science and understanding of the behaviour of a system's components. The models aim to mimic real-life events through assumptions on prominent underlying mechanisms, typically involving the constructing of simplified mathematical formulations of causal mechanisms (Baker et al., 2018). Lumped models are mathematical, mechanistic models representing the load to heart, where the physics of the entire arterial system is represented by a few, lumped parameters (Hellevik, 2015).

2.2.1 The Windkessel model

The Windkessel model is a lumped parameter model based on conservation of mass for the cardiovascular system. The model was developed after an observation by Stephen Hales in 1733, where he found that blood flow in the peripheral arteries is approximately steady, despite the pulsating heart activity (Hellevik, 2015). In the 18^{th} century, Hales illustrated the principles of the interaction between the heart and arteries by utilizing a fire hose, an air chamber, and a water pump. The air chamber damps the pump's pulsatile work, and the elasticity of the large arteries is illustrated by the air chamber. In 1899 Otto Frank presented the two-element Windkessel model by formulating Hales' analogy mathematically. The model contains two elements, the peripheral resistance R_{sys}, representing resistance in the small arteries, and the total arterial compliance C_a, representing the larger vessels' elasticity. The mathematical representation of the Windkessel model is obtained by the requirement of mass conservation of a vessel, and can be expressed with the ordinary differential equation

$$
\begin{aligned}
Q_{in} &= Q_a + Q_{out} && (2.2.1)\\
&= \frac{\partial V}{\partial p}\frac{\partial p}{\partial t} + \frac{p}{R_{sys}} && (2.2.2)\\
&= C_a\frac{\partial p}{\partial t} + \frac{p}{R_{sys}}, && (2.2.3)
\end{aligned}
$$

where Q_{in}, Q_a, and Q_{out} represent the blood inflow to aorta, the stored volume in aorta, and outflow towards the periphery, respectively. Further, V denotes the volume of the vessel, t time, and p the pressure towards the vessel wall.

The two-element Windkessel model assumes that the pressure and flow are periodic functions, and thus can be expressed as Fourier series. With this assumption, the impedance of a two-element Windkessel model, Z_{WK}, can be derived from Equation 2.2.3, and formulated as

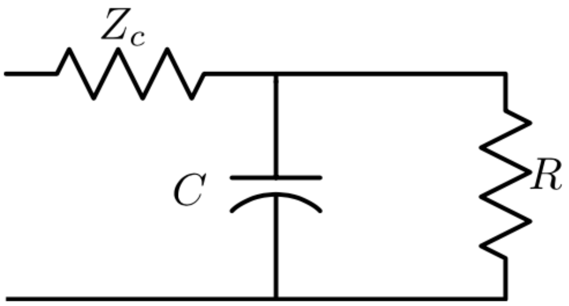


Figure 2.2.1: The three-element Windkessel model, where Z_C denotes the characteristic impedance, C the total arterial compliance, and R the total peripheral resistance. Figure from Hellevik (2015).

$$Z^n_{WK} = \frac{R_{sys}}{1 + j\omega_n R_{sys} C_a}, \tag{2.2.4}$$

where ω_n denotes the n different frequencies from the Fourier transform, and j denotes an imaginary number in the complex plane.

The two-element Windkessel model can be extended to a three-element model by introducing a third parameter, which leads to a more accurate description of the pressure-flow relation. The third parameter introduced is characteristic impedance, simulating the resistance to blood flow due to the characteristic resistance of the aorta. Characteristic impedance relates the geometry and the elastic properties of the vessel (Hellevik, 2015). Figure 2.2.1 illustrates the three-element Windkessel model, represented as an electric circuit relating the model's three elements.

2.2.2 The time-varying elastance model

The time-varying elastance model is a lumped parameter model that represents the pumping function of the left ventricle. Elastance, $E(t)$, explicitly reflects the contractile state of the heart (Suga and Sagawa, 1974; Suga et al., 1973), and represents the relation between the left ventricle volume, $V_{lv}(t)$, and the left ventricle pressure, $P_{lv}(t)$. The changes in elastance are due to the contractions of the left ventricle, and cycles from its diastolic to its systolic value during a cardiac cycle (Maksuti et al., 2016). The definition of the model is

$$E(t) = \frac{P_{lv}(t)}{V_{lv}(t) - V_0}, \tag{2.2.5}$$

where V_0 denotes the unloaded volume of the left ventricle.

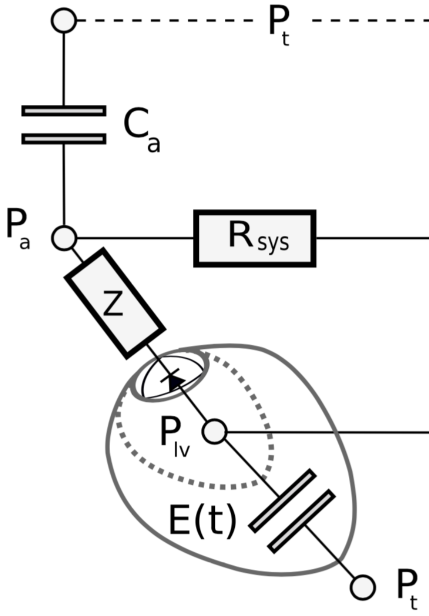


Figure 2.2.2: A schematic illustration of the lumped parameter model used in MyMDT's mechanistic model. C_a, P_a, C_v, and P_v represents the compliance and pressure in the arteries and veins, respectively. P_t denotes the thoracic pressure, and P_{lv} the left ventricle pressure. R_{sys} is the total peripheral resistance, R_m the resistance in the mitral valve, Z is aortic impedance, and $E(t)$ is the time-varying elastance.

2.2.3 MyMDT's mechanistic model

MyMDT has developed a mechanistic model that describes the dynamics of the blood flow in the cardiovascular system, and evaluates how changes in the cardiovascular system components contribute to the progression of hypertension. The model's intended application is to use blood pressure and volumetric flow curves of an individual, and estimate the curves after 12 weeks of a given exercise. The parameters expected to mediate the progression of hypertension are the left part of the heart, the peripheral resistance, the venous pressure, P_v, and the compliance in the arteries. These parameters and their relations are presented in the lumped model in Figure 2.2.2, which illustrates the hemodynamic part of the mechanistic model. In addition to the hemodynamic model, the mechanistic model consists of an acute exercise shift model, and a long-term remodeling of the cardiovascular system, described later in this section.

The curve of the blood pressure and flow for an individual at rest are inputs of an optimization algorithm that calibrates the mechanistic model's parameters. The hemodynamic model combines the time-varying elastance model representing the heart, and the three-element Windkessel model representing the arterial system, to process these parameters. A combination of the two models is previously described and validated by Stergiopulos et al. (1996), which is used as base for MyMDT's hemodynamic model. The main modification done by MyMDT is determining the venous pressure based on the balance of mechanical properties and total blood volume, instead of assuming a constant venous pressure. By recognizing relations in the data and simulating a curve to fit each individual's wave forms at rest, the hemodynamic model estimates internal parameters of the cardiovascular system, such as total arterial compliance and total peripheral resistance. These parameters are assembled in a vector called θ^0_{rest}, where 0 indicates that these are parameters estimated for the initial week. The parameters are further processed by the Acute Exercise Shift Model.

The exercise model estimates the hemodynamics during exercise, represented by $\theta^0_{exercise}$, given the values of θ^0_{rest} and an exercise pattern. The model uses population data from Chantler et al. (2008), that report a cross-sectional average of the relation between each parameter in θ_{rest} and $\theta_{exercise}$ at different intensities, to obtain the estimates. The $\theta^0_{exercise}$ parameters are further used to estimate how the physical activity will affect θ^0_{rest} over a period of 12 weeks. The model assumes that the resting parameters are modulated according to the stimulation of 12 weeks with different intensities of physical activity. The stimulation X is defined as a weighted sum of the time spent in different levels of physical activity (e.g. low, moderate, high). Cardiac power (CP) is defined as *MAP* $\times$ *cardiac output (CO)*, and is the basis of the weighting. The CP changes according to the relative heart rate (RHR), which is defined as the ratio between resting heart rate and maximum heart rate, and is used to determine heart rate zones during exercise. The stimulation X in MyMDT's mechanistic model is throughout this thesis referred to as exercise value, and is defined as

$$X = \int_0^T w(t) \times \frac{CP(RHR_i)}{CP_{rest}}\, dt, \tag{2.2.6}$$

where $w(t)$ is a weighting function, i indexes the sessions of physical activity, and thus RHR_i denotes the relative heart rate during the session. Further, T is the total period of 12 weeks. The weighting function is implemented to weight recent activity more than activities further back in time, according to the time of estimate.

Given the stimulus over 12 weeks, the exercise model estimates the values of θ^{12}_{rest}. To interpret the output of the model as blood pressure and flow curves, the hemodynamic model that initially was fitted to the curves at rest uses θ^{12}_{rest} to estimate the new curves. Consequently, the results of the mechanistic model are estimates of the blood pressure and flow curves at rest after 12 weeks of a given exercise pattern.

2.3 Machine learning

Artificial intelligence is a branch within computer science that can build smart machines capable of performing tasks that normally would require human intelligence (Helm et al., 2020). Machine learning is an application of AI and offers systems the ability to learn and improve from experience autonomously. Learning begins with observing data such as examples, direct experience, or instructions, that further can be used to find patterns and relations. When an algorithm is implemented and provided data, the algorithm aims to learn without any human interaction. After the training process, the model can make decisions and predictions on unseen data based on the patterns and experiences developed during training.

Machine learning is often categorized as supervised or unsupervised learning, based on how the model learns. Supervised learning is mainly used in this project and covers the learning of relations between a set of input variables and corresponding labels. The algorithm aims to learn a mapping function from input to output, where the goal is to approximate it to such an extent that the algorithm manages to predict the correct output on unseen data. In the context of this project, supervised machine learning will be performed to find potential patterns between personal parameters of an individual, a given exercise, and the individual's blood pressure curve after the exercise period. In unsupervised learning, labels are not provided, and the algorithm's task is to divide the data into clusters (Ghahramani, 2004). Unsupervised learning is utilized in one of the methods in this project to reduce the dimension of the blood pressure curves.

2.3.1 Linear regression

Regression is a study of dependence, and regression analysis is one of the most widely used techniques for analyzing multifactor data, and an important part of many research objects (Montgomery et al., 2021b; Weisberg, 2005). The most commonly used methods in regression analysis is linear regression, where the goal is to detect linear relationships between one or several predictors X, and a variable of interest Y, called a response variable. Such a linear relationships can be formulated as

$$y = \beta_0 + \beta_1 x, \tag{2.3.1}$$

where x denotes one different predictors, and the intercept β_0 and β_1 are unknown parameters. The exact nature of the considered linear relationship is not known, and thus the β parameters fulfills the equation (Jurgen, 2003). A method called least squares can be used to estimate the β parameters by observing n pairs of data, $((y_1, x_1), ..., (y_n, x_n))$. Least-squares method estimate

β so that the sum of the squares of the differences between the observations y_i and the straight line is minimum (Montgomery et al., 2021a). The least-squares criterion is

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{2.3.2}$$

In terms of predicting blood pressure curves, the observed pairs of data are personal features and their corresponding blood pressure curve. A linear regression model aims to approximate a linear relationship between these features.

2.3.2 Principal component analysis

Principal component analysis (PCA) is a standard statistical method often used for reducing the dimension of large data sets where variables are highly correlated, to a smaller set of variables. The new variables are linear combinations of the original variables, which are uncorrelated and explain most of the variation in the data. These variables are known as the principal components (PC). PCA can be used in machine learning to focus on a smaller number of independent variables, rather than a large number of original variables with complex correlations (Taylor et al., 2006).

The PCA method is based on the fact that any set of *m* variables (X_1, X_2, ... , X_m), can be transformed to a set of *m* orthogonal variables (S. Chaterjee, 1999). These variables are the principal components and are denoted C_1, C_2, ... , C_m. Each variable C_j is a linear function of the standardised variables $\widetilde{X}_1$, $\widetilde{X}_2$, ... , $\widetilde{X}_m$. Such as,

$$C_j = v_{1j}\widetilde{X}_1 + v_{2j}\widetilde{X}_2 + ... + v_{mj}\widetilde{X}_m, \tag{2.3.3}$$

where $j = 1, 2, ..., m$, and v_{ij} represents the elements of the eigenvector for every C_j. v_{ij} will throughout this thesis be referred to as the scores of the PCs. Let λ_j be the eigenvector, then it can be shown that the variance of each principal component C_j is

$$Var(C_j) = \lambda_j. \tag{2.3.4}$$

Further, the PCs are arranged so that $\lambda_1 > \lambda_2 > ... > \lambda_m$. Accordingly, the component containing most of the variance in the set of variables is the first component. The second component contains the second most variance, and so on. Thus, depending on the data set and the desired amount of variance conserved, one can reduce the data dimension by using the scores of a few of the first PCs for each individual, rather than the complete set of variables.

2.3.3 Neural networks

Neural networks are series of algorithms that endeavors to identify underlying patterns in a set of data. This is achieved through a process that mimics the way the human brain operates (Krogh, 2008). Neural network are neurologically-inspired systems, hence the name neural. Neurons are the basic building blocks of human intelligence, and the information messengers between different regions in the brain. In contrast, neurons in artificial neural networks are mathematical functions that collect and classify information according to a specific architecture (Koch and Segev, 2000).

Neural networks are built of three or more layers of interconnected neurons (Goodfellow et al., 2016). Each neuron is a perceptron, which is a type of linear classifier. The first layer in a neural network is an input layer, and the last layer is the output layer that holds the classifications or output signals. Between the input and output layers, the network may include several hidden layers. To find the ideal number of hidden layers and neurons is a challenging task that requires knowledge and experimentation (Abadi et al., 2015). Fully-connected layers are a common type of hidden layers, where each neuron receives connections from every neuron in the previous layer. The architecture of the network depends on the problem and dataset, and tuning of both the structure and hyperparameters are important to obtain a network with good performance.

The most popular form of neural networks are feed-forward, which are described in this section. One neuron can be connected to several neurons in both the layer before and after, where it respectively receives and transmits data (Beck, 2018). Each connection is assigned a weight by the neuron, which is multiplied with the data value received through that connection. In fully-connected layers, the total input of each neuron, z, is calculated as a weighted sum of the neurons in the preceding layer, and is further applied to the activation function, f, to get the neuron output, y. This process is the forward-pass in a neural network and can be described mathematically with the equations below (LeCun et al., 2015). Figure 2.3.1 illustrates the steps of the forward-pass, where one neuron receives weighted inputs from all neurons in the previous layer.

In a network with two hidden layers (H1 and H2), let i denote an input neuron, and x_i its value. Further, let j and k denote neurons in the first and second hidden layer, respectively, and l an output neuron. The weight between neuron i and j is denoted w_{ij}. The forward-pass is mathematically expressed as

$$\begin{aligned} z_j &= \sum w_{ij} x_i \\ y_j &= f(z_j), \end{aligned} \tag{2.3.5}$$

$$\begin{aligned} z_k &= \sum w_{jk} y_j \\ y_k &= f(z_k), \end{aligned} \tag{2.3.6}$$

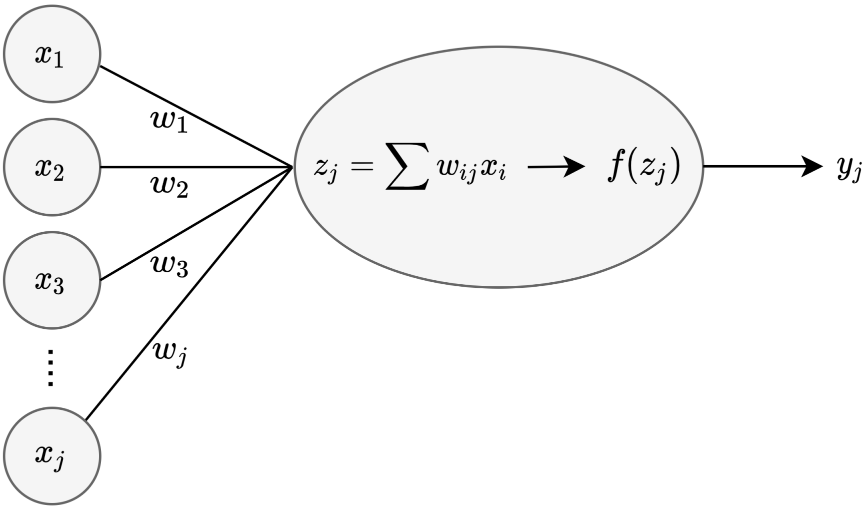


Figure 2.3.1: The steps of the forward-pass in a neural network. One neuron, receives the inputs from all neurons in the previous layer, $x_1 - x_j$, weighted by each connection's weight $w_1 - w_j$. Further, the neuron calculates the weighted sum, z_j, and applies the activation function f, which produces the neuron's output y_j.

$$\begin{aligned} z_l &= \sum w_{kl} y_k \\ y_k &= f(z_l). \end{aligned} \tag{2.3.7}$$

The choice of activation function in a neural network has a significant effect on the training dynamics and performance of the network (Ramachandran et al., 2018). A commonly used activation function is the Rectified Linear Unit (ReLU), which is defined as $f(x) = max(0, z)$. This allows the positive gradients that are inputs of the ReLUs to flow, and thus complex networks with this activation function is more easily optimized than those with functions with bounded output space. Multiple bounded activation functions only produces outputs between 0 and 1 (Sibi et al., 2013), and are more relevant for binary classification tasks. The nonlinear ReLU is common for hidden layers as it lets the network learn more complex function compared to a linear function. The linear activation function is a more suitable activation function for output layers as it outputs the weighed sum from the previous layer directly (Abadi et al., 2015). From this, ReLU and linear activation functions are deemed suitable for the hidden and output layers, respectively, in the implemented networks.

During training, the model receives training data in batches of samples (or instances), and predicts the label of each instance. The predicted label is compared to the true label for each instance, and the error is calculated by the loss function. A commonly used loss function is mean squared error (MSE), which is a quantitative score that describes the degree of similarity, or level of error, between two signals (Wang and Bovik, 2009). Let x_i and y_i be the values of the

ith samples in the signals x and y, respectively. Thus, the MSE between the two signals is defined as

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2. \tag{2.3.8}$$

Training of a neural network is an optimization process driven by an optimization algorithm, called an optimizer, that aims to minimize the loss. By calculating the error gradient and navigate it towards zero, the algorithm seeks to change the weights in the network so that the error is reduced for each evaluation.

Adaptive moment estimation (Adam) is an algorithm for efficient stochastic optimization first introduced by Kingma and Ba (2014), and is an extension of the stochastic gradient decent (SGD) algorithm. Adam is computationally efficient and requires little memory. It is well suited for problems with noisy or sparse gradients as the hyper-parameters often are intuitive and requires little tuning. Consequently, and after comparison with other algorithms, Adam is considered an appropriate algorithm for the models in this project.

With the process called the backward-pass, or back-propagation, the optimizer adjusts the weights and biases in the network after each batch. All batches are run through the network a defined number of times called epochs. Number of epochs should be defined according to the convergence of the calculated loss.

2.3.4 Informed machine learning

Informed machine learning (IML) is referred to as an umbrella term for machine learning methods that integrate additional knowledge into the model (von Rueden et al., 2020). IML extends standard machine learning as it learns from a hybrid information source consisting of both data and prior knowledge. Prior knowledge may include algebraic equations, logic rules, or simulation results, and is explicitly integrated into the machine learning pipeline separately from the rest of the data. How the knowledge is integrated into the model varies. One approach is to use models based on physical laws or relations to generate additional descriptive input data. A way of integrating this data is the concept of residual models, that aims to learn the biases of a physics-based model. Another method of IML is to integrate physical laws into the loss function of a neural network to encourage the model to make more physically consistent predictions.

There are several advantages with IML. Many machine learning algorithms are often referred to as black boxes as their predictions come without explanations and are based on underlying patterns that humans struggle to understand. Pure data-driven algorithms, for example a neural network, find complex correlations in the data, learn patterns, and provide an output for each given input.

The explanation for the output and how it was measured is not given. As additional knowledge is integrated into the learning process of IML, such a model can provide more insight between input and output. The additional knowledge may also increasing the chance of physically consistent predictions. Further, von Rueden et al. (2020) implies that IML may be favourable when dealing with insufficient training data.

Chapter 3

Data analysis and processing

This chapter presents the data utilized in this project. A comprehensive process to select feature space and standardize blood pressure curves have been carried out, and a detailed description of assessments and steps of the process follows.

3.1 Available data

For this project, we are provided data of hypertensive and prehypertensive individuals from a study performed by St. Olav's University Hospital, Trondheim, Norway. The data was collected from 26 individuals, 13 females and 13 males, ranging from 46 to 61 years old. The data contains basic characteristics such as age, gender, and height, together with a questionnaire regarding the individuals' lifestyle, a cardiopulmonary exercise test (CPET), as well as various other measurements from several clinical visits. The study aimed to investigate how the individuals, screened to be relatively inactive, responded to a higher level of physical activity over a 12-week period. The individuals' progress was recorded by performing measurements and tests at three clinical visits during the study period. The initial tests were performed before the exercise period began, the second tests were carried out halfway through the period, and the last tests after 12 weeks. At the beginning of the study, a more thorough screening was performed to map the initial physiology state of each individual. The screening included measurements of SBP, DBP, body mass index (BMI), and VO_2 max, among others. Furthermore, continuous measurements over a few seconds, including blood pressure measured at the finger and volumetric flow in the aorta, were performed at all three clinical visits.

The available data from the study at St. Olav's University Hospital contains some data that are irrelevant to this study, lack continuity, and was provided in multiple separate files. A comprehensive data preparation process was thus required to obtain a useful data set. Not all individuals completed the whole study, and some measurements are therefore missing. This reduced the amount of data available for use. To obtain more data for analyses and train the machine learning models, an additional dataset was created by dividing each individual into two samples. For each individual, the initial and half way measurements were used as pre-measurements for each sample. Further, the half way and final measurements were used as the samples' post-exercise measurements. It is important to note that with the dataset being divided as described, all instances were not independent. This limits the amount of additional information in contrast to 26 additional independent instances. Further, BMI and VO_2 max were not measured halfway through the exercise period, and the pre-measurements are thus the same for both instances of an individual. This leads to a limitation as the initial state of the second instance is not updated according to the its current activity level. A completely independent dataset is desired, however, the dependencies are thought to have less negative effect than the benefit a larger dataset provides.

As mentioned, some of the individuals had missing measurements that were deemed important in this study. It was therefore necessary to further remove some instances from both the full and divided dataset, resulting in a total of 20 and 37 instances, respectively.

3.2 Dimensionality reduction

The number of input features to a machine learning model is referred to as the dimensionality of the data. It is often beneficial to reduce the input dimension in order to improve the computational efficiency and the accuracy of the model (Cunningham, 2008). That is reducing the size of the feature space without impairing the data quality (Padmaja and Vishnuvardhan, 2016). A large feature space may cause a degradation in performance, as features may be redundant or provide too much information to learn for insufficient amount of training data. This problem is often referred to as the curse of dimensionality. The curse of dimensionality may lead to overfitting, which is a problem that can occur when irrelevant components are included (Hawkins, 2004). These problems are especially critical in cases with a small set of training samples, as the models depend highly on these samples. This project has limited data available, and it is therefore important to reduce the input dimension by selecting only the most significant features.

3.2.1 Feature selection

Features of interest have been selected according to the consensus in clinical practice and epidemiological studies about hypertension. Chobanian et al. (2003b) identify major risk factors that may affect the progression of hypertension. Among them are smoking, BMI, age, gender, and family history of premature cardiovascular disease. These features are therefore included in the features of interest. WHO (2015) presents various steps to minimize the odds of developing hypertension, thereby to stop tobacco use, have a healthy diet, limit alcohol intake, and manage stress. The questionnaire answered by the individuals included the weekly intake of glasses of wine, beer, and spirits, and if the individual is a smoker. These are all considered possible features of interest based on WHO's recommendations to reduce hypertension. WHO also states that physical activity is an important factor in reducing the risk of hypertension. VO_2 max is a measure of the maximum rate of oxygen consumption and is a good indication of a person's physical shape. The individuals in the study tested their VO_2 max prior to the exercise period, and this feature is considered interesting as it represents the individual's initial state of physical shape. In addition, Yang et al. (2019) suggest that VO_2 is a useful feature in prognostic indicators for predicting the effect of aerobic exercise for a hypertensive individual.

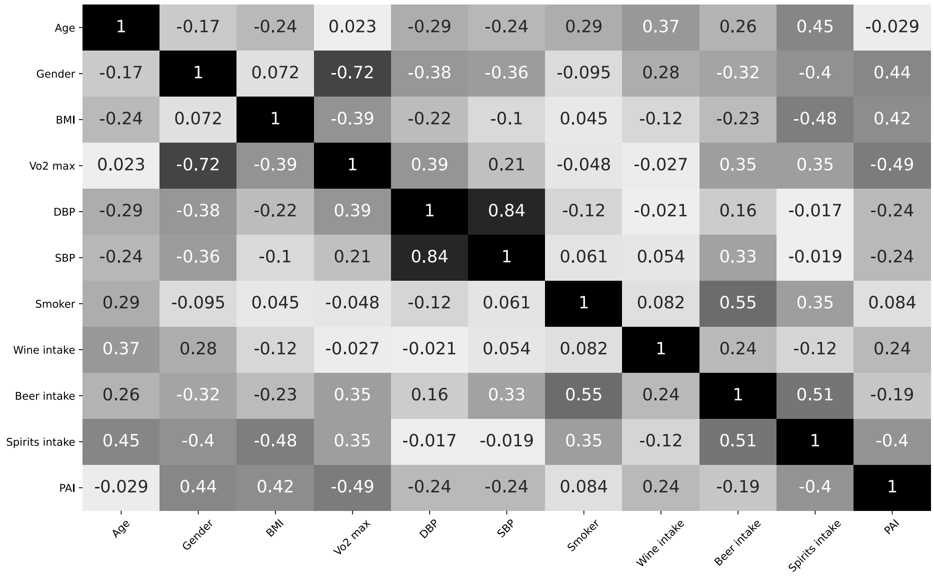
Ambulatory measurements of DBP and SBP are more comprehensive measurements of the blood pressure, calculating an average value of DBP and SBP measured continuously over 24 hours. These measurements are believed to avoid the raised values sometimes present in other measurements due to nervousness of being in a clinical setting. The ambulatory measurements are considered more accurate measurements for characterizing a person's typical blood pressure. Furthermore, the ambulatory blood pressure has shown to be a more sensitive risk predictor of coronary morbid events and stroke than clinical measurements (Banegas et al., 2018). The ambulatory pre-measurements of DBP and SBP are for these reasons considered features of interest. Other measurements deemed interesting are both pre and post-exercise measurements of the continuous waveform for blood pressure and flow.

Another important feature of this project is personal activity intelligence (PAI). PAI is a metric of physical activity, considering an individual's sex, age, and resting and maximal heart rate. The metric quantifies one's weekly physical activity from the heart rate pattern during physical activity (Nauman et al., 2019). Obtaining a PAI score higher than 100 every week over time is shown to reduce the risk of premature cardiovascular disease death (Nes et al., 2017) and is associated with a prolonged lifetime. The participants in the study by St. Olav's University Hospital were instructed to either exercise 150 minutes every week or gain over 100 in PAI score every week during the study period. In the available data, the PAI score for each individual is measured every day over the 12 weeks of the study. As PAI indicates the level of physical activity, this is a significant factor in our study, and the average score over the exercise period is calculated and included in the features of interest. A full list of the selected features of interest in shown in Table 3.2.1.

In this project, the relevant study group are people that have been diagnosed as hypertensive or prehypertensive. The individuals' family history is considered more relevant in cases where the probability to develop hypertension is

Feature of interest
Age
Gender
BMI
VO_2 max
DBP
SBP
Smoker
Wine intake
Beer intake
Spirits intake
PAI
Family history of hypertension

Table 3.2.1: A list of the analyzed and considered features of interest.



	Age	Gender	BMI	Vo2 max	DBP	SBP	Smoker	Wine intake	Beer intake	Spirits intake	PAI
Age	1	-0.17	-0.24	0.023	-0.29	-0.24	0.29	0.37	0.26	0.45	-0.029
Gender	-0.17	1	0.072	-0.72	-0.38	-0.36	-0.095	0.28	-0.32	-0.4	0.44
BMI	-0.24	0.072	1	-0.39	-0.22	-0.1	0.045	-0.12	-0.23	-0.48	0.42
Vo2 max	0.023	-0.72	-0.39	1	0.39	0.21	-0.048	-0.027	0.35	0.35	-0.49
DBP	-0.29	-0.38	-0.22	0.39	1	0.84	-0.12	-0.021	0.16	-0.017	-0.24
SBP	-0.24	-0.36	-0.1	0.21	0.84	1	0.061	0.054	0.33	-0.019	-0.24
Smoker	0.29	-0.095	0.045	-0.048	-0.12	0.061	1	0.082	0.55	0.35	0.084
Wine intake	0.37	0.28	-0.12	-0.027	-0.021	0.054	0.082	1	0.24	-0.12	0.24
Beer intake	0.26	-0.32	-0.23	0.35	0.16	0.33	0.55	0.24	1	0.51	-0.19
Spirits intake	0.45	-0.4	-0.48	0.35	-0.017	-0.019	0.35	-0.12	0.51	1	-0.4
PAI	-0.029	0.44	0.42	-0.49	-0.24	-0.24	0.084	0.24	-0.19	-0.4	1

Figure 3.2.1: Correlation between the features of interest. Darker color indicates a higher absolute value of the correlation coefficient, and thus a stronger correlation.

of interest, and not the progression of already hypertensive individuals. With family history excluded from the features of interest, eleven features are left for further analysis.

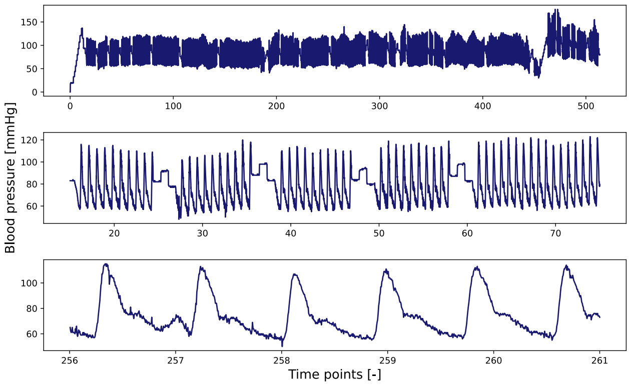
An input vector of eleven features may be challenging to fit for 20 and 37 instances, and may thus cause the curse of dimensionality as explained in Section 3.2. A correlation analysis has therefore been carried out to detect connections that may lead to redundancy. Figure 3.2.1 illustrates the correlation matrix describing the feature connections. The correlation coefficients vary from -1 to 1 which denotes maximum negative and positive correlation, respectively. As the figure shows, DBP and SBP are strongly correlated with a correlation coefficient of 0.84. This correlation is highly expected, but since they both describe the initial state of an individual's blood pressure, neither can be excluded from the feature space. $VO_2\max$ and gender also show a significant correlation of -0.72. Song et al. (2020) state that there exist gender differences in both risk factors and treatment of hypertension, and in addition to the already discussed importance of $VO_2\max$, both features are included in the feature space. Moreover, a small correlation was discovered between the intake of beer, and both smoking and intake of spirits. Additionally, Lahti-Koski et al. (2002) detected a further correlation between BMI and both alcohol consumption and smoking history. With the aim at reducing the dimension of the feature space, intake of all types of alcohol and smoking were excluded due to the discovered correlations with BMI.

The performed feature selection resulted in an input vector that was used throughout the modeling. The input vector is defined as

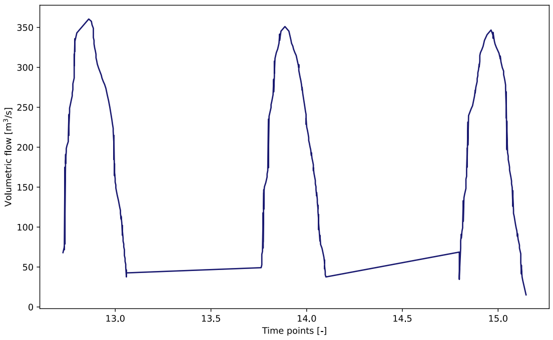
$$input = [age, gender, BMI, VO_2max, DBP, SBP, PAI]. \tag{3.2.1}$$

3.3 Processing of time series data

Blood pressure curves store more detailed information about a person's blood pressure than single values, and this was the basis for the choice of the curve as prediction target. The data of blood pressure and flow were represented as time series as they were measured over a period of time. Figure 1a and 1b present original signals of the blood pressure and flow of one individual. The duration of the signals ranged from a few seconds to several minutes, and varied duration required processing to obtain comparable data. The blood pressure signal consisted of the blood pressure cycle repeated throughout the measurement time. As the individuals remained in the same environment and sitting condition throughout the measurement, the consecutive cycles were expected to be similar. Still, signal noise and other disturbances such as measuring instrument calibration were common and made the cycles incoherent, and the quality varied throughout the signal. These quality differences can be observed in Figure



(1a) Original blood pressure signal



(1b) Original flow signal

1a. Only one blood pressure curve per clinical visit was desired as it was deemed sufficient to obtain an overall understanding of an individual's blood pressure. More than one cycle was considered excess information, and it was thus chosen to extract one representative cycle from the full signal. This approach required thorough analysis and processing of the signal, which is throughout the thesis referred to as a standardization process. A more detailed description of the process follows.

Like the blood pressure, the flow measurements also contained multiple consecutive cycles. Only one cycle was desired for flow as well, thus a standardization process was required for both measurements.

3.3.1 Standardization of blood pressure curves

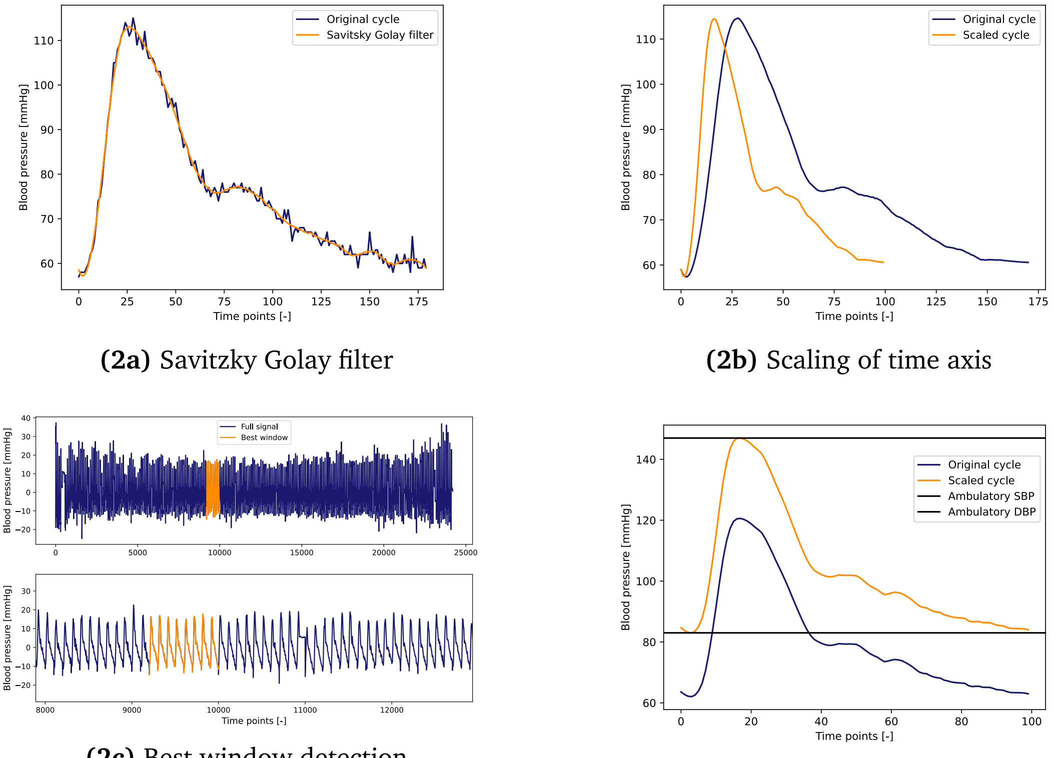
The full signals of the individuals' blood pressure were of varied duration, ranging from a few seconds to several minutes. They contained consecutive blood pressure cycles with varied quality, and the goal of the standardization process was to obtain one representative curve of good quality for each blood pressure measurement. The implemented code that performs the standardization process can be found in Appendix A.

The first step in the process was to split the full blood pressure signal into separate cycles. The transition from one cycle to the next was detected by locating local minima in the full signal, with a defined minimum distance between the points. The minimum distance was determined by exploration of different distances and subsequent observation of the detected minima. The signal was split at the minima, resulting in a set of cycles starting and ending at their lowest point, that is, the beginning of the systole phase.

By inspection, small, frequent waves were observed throughout most of the cycles. This is illustrated by the original cycle in Figure 2a, and were assumed to stem from noise. To reduce the overall noise, a Savitzky Golay filter was applied to smooth the data. A Savitzky Golay filter is a smoothing method based on local least-squares polynomial approximation, which reduces noise while maintaining the shape and height of waveform peaks (Schafer, 2011). A parameter called window length defines how many data points the polynomials are fitted to and affect the coarseness of the smoothing. The filter used on the blood pressure signal was applied with a relatively small window length to remove the small variations due to noise while maintaining a fine smoothing. The application of the Savitzky Golay filter is illustrated in Figure 2a.

For further analysis and comparison of the different cycles, cycles of the same length were considered beneficial. To achieve equal length cycles, all were scaled according to their time duration to 100 points, and the time aspect was thus further ignored. This simplification excludes the duration of the cardiac cycle, which may be a noteworthy limitation for further consideration. The scaling is illustrated in Figure 2b. Inspection proved that not all cycles were well-formed blood pressure curves but consisted also of multiple periods of measurement instrument calibration. To filter out the obvious calibration periods and other periods with significant noise present, constraints were defined and applied. The constraints delimited the minimum and maximum values, that is the DBP and SBP value, of each cycle. Cycles having unreasonable low values for DBP and SBP, respectively below 40 mmHg and 90 mmHg, were filtered out of the cycle collection. Furthermore, as the SBP occurs during the systole phase of the cardiac cycle, it is reasonable to expect that the peak time in a blood pressure curve is, within a buffer range, known. Thus, all cycles not having their peak between time point 15-30 were filtered out.

The remaining cycles were still of varying quality, thus more filtering was required to obtain a single representative cycle. The pressure level of the full signal showed tendencies to vary in waves, which was desirable to ignore when comparing the cycles. To solve this, the Savitsky Golay filter was again applied, this time to detect the trend of the signal. The window length was increased significantly which eliminated large parts of the pressure signal, and the remaining signal was a trend wave. The trend wave was then subtracted from the full signal, resulting in a signal with less global variations.



(2a) Savitzky Golay filter

(2b) Scaling of time axis

(2c) Best window detection

(2d) Scaling of pressure axis

Figure 3.3.2: Steps of the standardization process of the blood pressure signal, including Savitzky Golay filter, scaling of the time axis, best window detection, and scaling of the pressure axis.

The quality of the individual cycles varied in periods. Some periods consisted of multiple cycles of good quality consecutively, whilst other periods contained dominant noise or measurement instrument calibration. This lead to a window-based search to find periods of multiple good quality cycles consecutively. It was observed that between most of the calibration periods or dominant noise, eight or more distinct cycles occurred in a row, thus the window size was set to eight cycles. The search involved moving the window step wise through the signal. The total variance for each window was calculated point by point across the eight cycles, and the window with the least total standard deviation was considered to be of best quality. An illustration of a detected window is shown in Figure 2c. The cycle closest to the mean of the window was chosen as the representative cycle.

The pressure values in the chosen cycle showed in multiple cases to differ from the ambulatory measurements of DBP and SBP from the clinical visit. As introduced in Section 3.2.1, the ambulatory blood pressure is considered an accurate measurement for characterizing an individual's typical blood pressure.

To obtain a pressure curve with more correct values, the curve was scaled according to the ambulatory measurement of DBP and SBP, performed as

$$\hat{p_t} = SBP \times \frac{p_t - p\_min}{p\_max - p\_min} + DBP \times (1 - \frac{p_t - p\_min}{p\_max - p\_min}), \qquad (3.3.1)$$

where $\hat{p_t}$ denotes the scaled version of point p_t, and p_{min} and p_{max} are the extreme pressure values from the original curve. SBP and DBP denotes the ambulatory measurements. An illustration of the scaling is visualized in Figure 2d.

3.3.2 Standardization of flow curves

The participants' volumetric flow signal was based on continuous Doppler ultrasound measurements of the aorta, and consisted of flow values from three consecutive cardiac cycles. Figure 1b shows an example of a flow signal, and straight lines between the three curves can be observed. These lines are results of values below a detection threshold of the signal processing, and are hence minor and set to zero. The periods of missing values were used to split the signal into three separate curves. Further, the curves were scaled similar to the blood pressure curves, to 100 points according to their time duration. The mean maximum value of the three curves, and the mean of the peaks' time point were calculated. Further, all three curves were scaled to match their common means, and the final representative curve was the average of the scaled curves.

3.3.3 Feature analysis of blood pressure curves

Blood pressure curves store more detailed information about a person's blood pressure than single values. By analyzing the shape of the curve and how the shape changes with exercise over time, other details about the effect of exercise may be discovered. A detailed analysis of different features of the curve was considered interesting as consistent changes after the exercise period were deemed relevant. Which features to be analyzed was decided based on perceived differences in the collection of all individuals' blood pressure curves prior to and after the exercise period. The divided dataset was used in the analysis to obtain more samples to hopefully detect clearer patterns. The idea was to define distinct features and analyze their changes with exercise. If clearly affected features were discovered, the intent was to analyze their utility in prediction. The analyzed features are illustrated in Figure 3.3.3.

The first features analyzed were the already chosen input features DBP and SBP, as research has proven that physical activity effectively lowers the pressure levels (Pedersen and Saltin, 2015; Hackam et al., 2013). The analysis was motivated by the relevance of the available data to see if this small selection

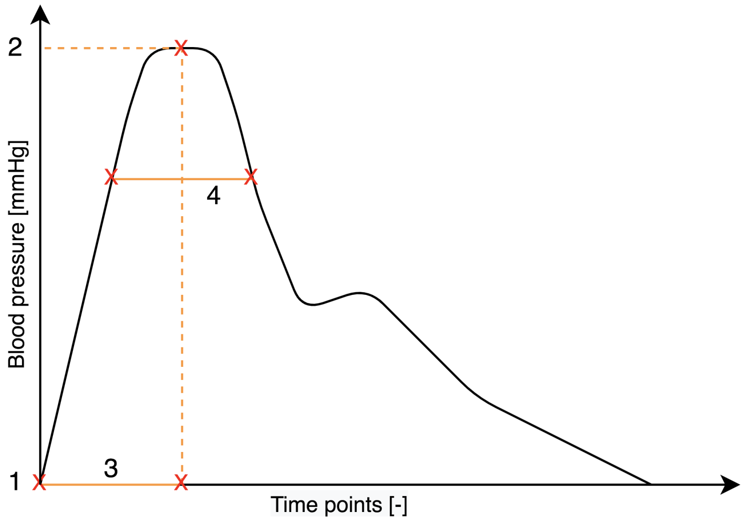


Figure 3.3.3: Visualization of the analyzed features of the blood pressure curves. Number 1 and 2 denote the DBP and SBP, respectively, 3 represents the time point of the systolic peak, and 4 denotes the peak width of the curve.

was representative and confirms literature. The analysis of changes in DBP and SBP did not have any obvious findings. 19 of the 37 individuals analyzed confirms the reviewed research and had a lower SBP after the exercise period. However, 18 individuals had an increased SBP value, and the analyses show that a low amount of exercise was not the grounds for this increase. Similarly, the DBP value decreased in 13 cases and increased in 14, and the rest remained unchanged. The overall averages for the DBP and SBP changes were $0.03 \pm 6.87\text{mmHg}$ and $-0.41 \pm 3.55\text{mmHg}$, respectively. The unanticipated variations detected in these analyses may be due to the division of the dataset, as half of the study period is little time for the blood pressure to change significantly. In addition, the amount of analyzed samples may be too small to detect global patterns.

The next feature analyzed was the time point of the systolic peak, where the average change over the exercise period was $-1.03 \pm 3.58\text{mmHg}$. The analysis showed that for 17 out of the 37 individuals, the time point of the systolic peak occurred earlier after the exercise period, whereas for 15 individuals it occurred later, and 5 showed no change. The changes were analyzed according to the individuals' PAI score, however no clear correspondence was detected.

Lastly, it was observed that the width of the systolic peak, illustrated in Figure 3.3.3, varied between individuals, and it was thus desired to assess whether it was affected by exercise. The number of time steps between a chosen point prior to the systolic peak, and the point with the same pressure value after the systolic peak, was utilized as the width metric. Different locations of the width measurement was considered and tested. Due to varying quality of the

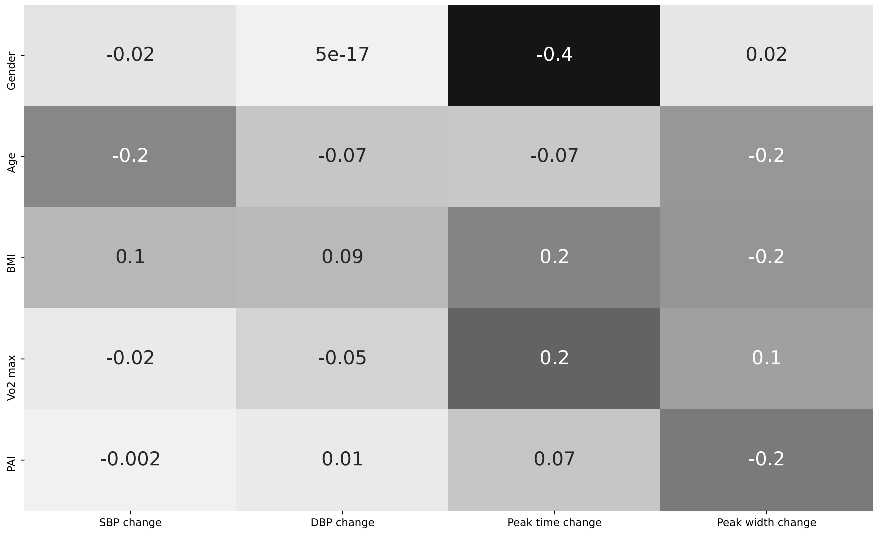


Figure 3.3.4: Correlation between the analyzed features and contextual data. Darker color indicates a higher absolute value of the correlation coefficient, and thus a stronger correlation.

blood pressure curves, the width leading to the most comparable measures was assumed calculated between the two points with pressure values closest to a value P, defined as

$$P = DBP + 0.8 \times (SBP - DBP). \tag{3.3.2}$$

The analysis showed that the peak width decreased for 20 individuals but increased for 15 individuals, with an average change of -0.89 ± 6.5 time points. Like the time point of systolic peak, the changes in peak width were also analyzed according the individuals' PAI score. No relationship between change in this peak width and PAI score was evident.

Figure 3.3.4 shows a correlation matrix between the analyzed features and contextual data for the individuals. As the figure indicates, no significant correlation is detected between any of the features. The strongest correlation coefficient is -0.4 between gender and change in time point of the systolic peak. However, -0.4 is considered a too low correlation coefficient for a final conclusions to be drawn from this analysis.

It is important to note that the presented analyses are based on a small number of samples, and thus involve a lot of uncertainty. Some global patterns are anticipated to exist for these, or other features, but may require more data for analyses.

3.4 Synthetic data

Limited amount of data may be the biggest obstacle in this project. By both reviewing literature and analyzing the available data, it was expected that a neural network would struggle to predict an individual's blood pressure curve with a satisfactory accuracy. To mitigate this problem, MyMDT's mechanistic model was used to generate large amounts of synthetic data. The model estimates the blood pressure curve of the individuals after 12 weeks of a given exercise value, as described in Section 2.2. To generate data, the unitless exercise value, defined in 2.2.6, was modified 150 times within a range between 0.25 and 7.5, and the estimated blood pressure curve changed accordingly. The mechanistic model is organized as a system of ordinary differential equations (ODEs), which is solved to steady state by a 4^{th} and 5^{th} order adaptive Runge-Kutta (Ixaru and Vanden Berghe, 2004). The initial curves of blood pressure and flow are specified as initial values for the model. The model dynamically generates arterial inflow by the heart model described in Section 2.2.2. The system is calibrated to reproduce real data by a non-linear least squares optimization method. The model is optimized with respect to synchronized flow and pressure traces over one heart cycle, and the trust region reflective algorithm (TRRA) is used to estimate parameters. The result is a synthetic data set of 150 instances per individual, only separated by the exercise value and blood pressure curve after the exercise period. The obtained synthetic dataset consists of 3450 instances in total.

A similar process was performed to generate an additional synthetic dataset. This dataset was also generated by MyMDT's mechanistic model, however modulated to only account for the part including the Windkessel model. In difference from the mechanistic model, the three-element Windkessel model has no dynamically generated arterial inflow, and was thus supplied with a fixed flow to generate the pressure response. The Windkessel model is, like the mechanistic model, formulated as a system of ODEs. The three arterial parameters from the full mechanistic model are also used to specify the parameters of the this model.

The synthetic blood pressure curves are similar to the true blood pressure curves, however the shapes are coarser and missing details compared to the true data. The pressure values increase and decrease according to the exercise value, and the curves have extreme values DBP and SBP within a wide range. As the synthetic data differs significantly from the true data, the datasets are kept separate and used for different models.

Chapter 4

Methods

In this chapter the implemented models are described in detail. First, a justification of the focus area of IML is presented, followed by a short presentation of the platform and libraries utilized for implementation. Further, the evaluation process and used metrics are presented, and descriptions of the model architectures follow.

4.1 Method selection

Previous studies show that machine learning models are computationally powerful and have the potential to make accurate predictions on different applications (Wu et al., 2014; Peng et al., 2019; Zhang and Lou, 2020; Li et al., 2017). However, these models depend on a sufficient amount of data for training, which is a limiting factor in many cases, as in this project. Literature shows that predictions from pure machine learning models may violate physical laws, and previous studies have proposed IML model frameworks that combine machine learning models with physics-based knowledge to make physically consistent predictions (Kissas et al., 2020; Beucler et al., 2020; Willard et al., 2020; Karpatne et al., 2017; Dourado and Viana, 2019). Willard et al. (2020) present an IML methodology that incorporates physical knowledge in the loss function of a neural network. This methodology is used by Kissas et al. (2020) who proposes a machine learning framework to predict arterial blood pressure with a constrained neural network. The constraints are incorporated in the loss function and include conservation of mass and momentum, available clinical measurement, and boundary conditions at interface points between arteries. Beucler et al. (2020) also introduce a neural network with enforced constraints in the loss function. In this study, conservation laws are incorporated, and a given α, varying between 0 and 1, assigns the different laws a weight.

A different IML methodology presented by Willard et al. (2020) is residual modeling. The key concept of residual modeling is for a machine learning model to learn the biases of a physics-based model and thereby learn to predict its error. Further, the predictions of the machine learning model can be added to the estimates of the physical model with the aim at obtaining more correct estimates. A limitation of residual models is their inability to enforce physical constraints as they model the error instead of the physical quantities (Willard et al., 2020).

In the context of this project, the key takeaway from IML studies is that incorporation of additional relevant knowledge can support models with limited data by directing them towards more physically consistent predictions. With the computational power potential machine learning models have proven to have, and the possible advantages of IML architectures, this study focuses on comparing IML models with standard machine learning. The goal of the comparisons is to examine possible contributions IML may have in the task of predicting blood pressure curves.

4.2 Platform and libraries

All code in this project is written with Pyhton programming language, and the models are built with the TensorFlows machine learning system. TensorFlow is an open source platform with tools, libraries and resources to help developers built machine learning applications (Abadi et al., 2015). Keras is a software library that runs on top of the TensorFlow platform and is an application programming interface (API) written in Python. The Keras library implements three APIs for creating TensorFlow models in Python. The Functional API is a flexible way to build models in Keras, and supports arbitrary model architectures. The API provides a way to create basic fully-connected layers called dense, which are used as the hidden layers for all models in this project.

4.3 Model evaluation

4.3.1 Evaluation process

To evaluate the models systematically, a common evaluation procedure was established. Cross-validation (CV) is a validation method that evaluates the performance of a model on unseen data. The method is based on splitting the data into a training set, used to train the algorithm, and a validation set, used to evaluate the performance of the algorithm (Browne, 2000). One data split yields a validation estimate, and an average over multiple splits yields a CV estimate.

To perform a CV of the models, the data was split into a training set and a validation set, and the validation set was further split into a test set and prediction set. The test and prediction set each consisted of one instance for the real data, and one individual of 150 instances for the synthetic data. After the splitting, the training set was randomly shuffled and used to train the models. The training process was carried out 20 times, where the test set was used to validate the performance of the resulting 20 models. Further, the best performing model on the test set was used to predict on the prediction set, and the prediction errors were the achieved results of this particular split. The evaluation algorithm utilized for each split is illustrated in Algorithm 1.

Different splits were evaluated to obtain a CV of the model. Table 4.3.1 presents the splitting of the dataset, and the results presented in Chapter 5 are the average results of the performed splits.

Train	Test	Predict
A	B	C
A	C	B
B	A	C
B	C	A
C	A	B
C	B	A

Table 4.3.1: The performed method to split the dataset during evaluation.

Algorithm 1: Evaluation of one split

```
A|B|C ← dataset;
train_set ← A;
test_set ← B;
prediction_set ← C;
best_model ← $\emptyset$;
best_error ← $\infty$;
while $i < 20$ do
    model ← new model;
    model ← model.$train()$;
    error ← model.$predict$(test_set);
    if error < best error then
        best model ← model;
        best error ← error;
    else
        continue
    i ← i + 1;
result ← best_model.$predict$(prediction_set);
return result;
```

The evaluation process differs from standard CV partly because the validation set is not randomly selected. The synthetic data used in this project consists of 150 similar instances for each individual, only distinguished by the exercise value. If the validation set was randomly selected, training instances would be derived from the same individuals as the validation instances. Consequently, the prediction samples would not be completely unseen data for the model, which could lead to a biased evaluation. The validation set was therefore selected manually for the models with synthetic data, and to obtain similar evaluation processes across the different model, this was also done for the models with real data.

4.3.2 Evaluation metrics

Pressure curve metrics

To evaluate and compare the different models, relevant evaluation metrics have been established. As the main objective is to predict the blood pressure curve, both the shape of the curve and its pressure values are deemed important. The mean error of the 100 points in the predicted curve captures the overall quality of the prediction. This metric is defined as the pointwise error. Further, the predicted DBP and SBP values, that is the minimum and maximum values of the curve, are important accuracies to remark. These values determine whether an individual is defined as hypertensive, thus the model's ability to predict them is crucial. Studies have found that PP is associated with, and may be a more accurate predictor of cardiovascular disease, than SBP or DBP alone (Lee et al., 1999; Sesso et al., 2000). Additionally, average SBP, DBP and MAP is shown to be associated with an increased risk of cardiovascular disease in younger men. These findings underline the importance of measuring the model's predictive power of PP and MAP in addition to DBP and SBP. Thus, all models are evaluated based on these pressure curve metrics.

Classification metrics

This section presents evaluation metrics relevant in cases of binary classification problems. The task of predicting blood pressure can be considered binary if the focus is the whether hypertension is present or not. The threshold between a positive and negative prediction are thus considered as

$$Y = \begin{cases} 1 & \text{if } DBP > 90 \text{ or } SBP > 140 \\ 0 & \text{otherwise.} \end{cases} \tag{4.3.1}$$

In binary classification problems, multiple evaluation metrics are frequently used in research to provide comprehensive assessments (He and Garcia, 2009).

A selection of these measures are accuracy, precision, recall, and specificity. Accuracy is the most frequent used metric and is a simple way of describing a classifier's performance as it assesses the overall effectiveness of the algorithm. Accuracy is defined as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{4.3.2}$$

where TP denotes true positives, FP denotes false positives, and TN and FN denote true negatives and false negatives, respectively.

A disadvantage of using only accuracy as evaluation metric is that is can be deceiving in certain situations and are highly sensitive to imbalanced data. For that reason, additional metrics help provide a more comprehensive evaluation of the classifier.

Precision is a measure of exactness and denotes the proportion of predicted positives that are correctly classified (Powers, 2020; Sokolova et al., 2006). Recall (true positive rate (TPR)) and specificity (true negative rate (TNR)) are measures of completeness and denotes the proportion of the true positives and true negatives, respectively, that are correctly classified. The definitions are

$$Precision = \frac{TP}{TP + FP}, \tag{4.3.3}$$

$$Recall(TPR) = \frac{TP}{TP + FN}, \tag{4.3.4}$$

$$Specificity(TNR) = \frac{TN}{TN + FP}. \tag{4.3.5}$$

In terms of classification metrics, these are not considered as relevant for the models with real data. The real data consists of only a few data samples, and consequently, multiple and varied data distributions of the prediction samples are required to obtain completely reliable evaluations. To obtain sufficient data distributions is complicated and time consuming, and we have thus chosen to focus on pressure curve metrics for these models. Only the synthetic data model is evaluated with classification metrics.

4.4 Models

A total of six models have been implemented. As a baseline, a regression model has been implemented to assess the necessity of a more complex machine learning model. The two first neural network models are similar but implemented with different loss functions. The objective with these models was to examine the effect of a physics-based loss function. The third neural network is a residual model that aims to predict the error of the estimates from MyMDT's mechanistic model. By adding the predicted error to the original estimates, the model was anticipated to correct, or improve the results of the mechanistic model. As limited data was available, two different residual models were implemented, with real and synthetic data, to examine the possible improvement of more data. The last model is a neural network implemented to predict synthetic data of blood pressure curves. With a large amount of synthetic data, the model aims to assess the significance of, and performance differences with different amounts of training data. The models trained on real data use the input vector defined in 3.2.1, whereas the models with synthetic data utilize the same input but replace the PAI score with exercise value, defined in Equation 2.2.6.

All model hyperparameters were optimized independently by exploring different numbers of layers and neurons, activation functions, optimizers, and learning rates. Batch sizes and number of epochs have been determined based on observations of the convergence of the loss, and all models were trained with 100 epochs. The achieved results are presented in Chapter 5.

4.4.1 Linear regression model

Linear regression assumes a linear relationship between features and corresponding labels, and estimates coefficients to give the best prediction of the label as a linear function of inputs. The implemented regression model is based on the presented input features, the regression coefficients β, and the calculated PCs of the blood pressure curves. An extended input space was implemented to include blood pressure curves, represented as PCs, to help the model discover more correlations in the data. First, PCA was applied to all blood pressure curves, and three components were required to capture the desired variance in the data of 95%. Further, these three PCs were multiplied with the input features in the training data, and the result of the calculation was the extended input space X. A linear regression model was fitted from X to the post-exercise pressure curves by the least-squares method, and the regression coefficients β were thereby obtained. The idea of this approach was to predict the amount of each PC in the post-exercise pressure curve.

To predict an unseen individual's post-exercise blood pressure curve, the input space of the new instance was extended as previously described, by multiplication with the selected PCs. Matrix multiplication with the extended input space and β was then performed, and the obtained vector compose the estimated blood pressure curve.

The prediction by the linear regression model can be expressed as

$$\begin{pmatrix} y_1 \\ \vdots \\ y_t \end{pmatrix} = \begin{pmatrix} x_1 c_{11} & \dots & x_p c_{31} \\ \dots & & \dots \\ x_1 c_{1t} & \dots & x_p c_{3t} \end{pmatrix} \begin{pmatrix} \beta_{11} \\ \vdots \\ \beta_{3p} \end{pmatrix} \tag{4.4.1}$$

for $0 < t < 100$, where $y_1 - y_t$ denotes the different points on the pressure curve.

The linear regression model was evaluated as described in Section 4.3, however without a test set. A linear regression model will achieve the same regression coefficients with the same training data, and multiple training processes is therefore unnecessary. The training set consisted of 36 instances, while the prediction set consisted of one instance. The linear regression model was used as a baseline for the neural network models to assess the benefit of a more complex model.

4.4.2 MSE loss model

A neural network that aims to predict the blood pressure curve of an individual after an exercise period has been implemented. The divided dataset described in Section 3.1 was used for training. The input of the MSE loss model is the defined input vector, and the model architecture consists of an input and output layer, and two hidden layers. The target value is a vector of 100 different numbers that form the predicted blood pressure curve. The output vector is denoted as

$$target = [P_1, P_2, ..., P_{100}]. \tag{4.4.2}$$

The hidden layers in the network consists of 25 and 50 neurons, respectively, and both implemented with ReLU is used as their activation function. The optimizer used in the network is Adam, with a learning rate of 0.001, and MSE is used as loss function. The batch size was set to one, as a small amount of training instances made a larger batch size less relevant.

Available training instances for the MSE loss model were in total 32, as 5 of the 37 from the divided dataset missed flow measurements necessary for the two loss models. The evaluation process was carried out as described in Section 4.3, where the test and prediction sets consisted of one instance each.

4.4.3 Custom loss model

The custom loss model is a further development of the MSE loss model. The architectures were initially close to identical, however the loss function in the custom loss model was physically constrained to direct the network towards more physically consistent predictions. By reusing the core architecture and training data as the MSE loss model, a comparison of the two can provide an understanding of the effect of a physics-guided loss function. During the model optimization, it was discovered that an increase in the number of layers resulted in better predictions, thus the model architecture was adjusted to consist of three hidden layers, with 25, 50, and 75 neurons, respectively. Apart from the loss function and number of layers, the rest of the model parameters and architecture are the same as the MSE loss model.

As introduced in Section 2.1.1, vascular impedance is a concept of great importance in biofluid dynamics. With input impedance being a property of the whole arterial system, and defining the relationship of periodic pressure and flow, it is a physical property that may be suitable to inform the neural network about the typical relationship between pressure and flow. The loss function in the custom loss model therefore includes a that quantifies how far the impedance of the predicted pressure is from the Windkessel impedance. The Windkessel model calculates impedance based on simplifications of the cardiovascular system. The predicted pressure values are desired to agree with this the Windkessel model as it is based on physical laws. By constraining the network to follow the laws incorporated in the Windkessel model, the predictions are directed towards more physical consistency. The impedance error is defined as the difference between the Windkessel impedance and the input impedance calculated as the ratio of the Fourier transform of the predicted pressure and true post-exercise flow. The sum of the magnitude of this difference make up the impedance loss term in the loss function. The definition is thus

$$Z_{loss} = \sum_{w} || \, Z_i - Z_{WK} \, ||, \tag{4.4.3a}$$

where

$$Z_i(\omega_n) = \frac{P(\omega_n)}{Q(\omega_n)}, \tag{4.4.3b}$$

with $P(\omega_n)$ and $Q(\omega_n)$ being the Fourier transforms of the predicted pressure curve and true post-exercise flow curve, respectively. ω_n denotes the n different frequencies from the transform. Further,

$$Z_{WK}^{n} = \frac{R_{sys}}{1 + j\omega_n R_{sys} C_a}, \tag{4.4.3c}$$

where

$$R_{sys} = \frac{MAP}{CO}, \tag{4.4.3d}$$

$$CO = \frac{SV}{T}, \tag{4.4.3e}$$

$$SV = \int Q_(t)dt, \tag{4.4.3f}$$

$$C_a = \frac{PP}{SV}, \tag{4.4.3g}$$

with MAP and PP being the mean arterial pressure and pulse pressure from the predicted pressure curve. T and $Q(t)$ denotes the cycle period and flow curve from the true post-exercise measurements.

The impedance difference $\hat{Z} - Z_{WK}$ is represented as a tensor consisting of 51 complex numbers $a + bi$, each representing the impedance error of a frequency. To capture all parts of the impedance error, the sum of the magnitude is calculated, and Equation 4.4.3a is the defined impedance loss. The magnitude of each of the 51 elements is calculated

$$|| \; a + bi \; ||= \sqrt{a^2 + b^2}. \tag{4.4.4}$$

In addition to the impedance term of the loss function, a term that comprises the MSE of the prediction is included. The significance of the different terms is determined through the assigned weights α and β that are optimized to obtain best possible predictions. Consequently, the loss function of the the custom loss model is

$$Loss = \alpha \times Loss_{MSE} + \beta \times Loss_{impedance}. \tag{4.4.5}$$

As the objective of the two models with different loss functions is to obtain an understanding of a physics-guided loss function, α was fixed at 1, and only β was varied during optimization. A β value of 30 was proven to achieve the best scores, thus the final loss function of the custom loss model is

$$\boxed{Loss = Loss_{MSE} + 30 \times Loss_{impedance}} \tag{4.4.6}$$

The evaluation process of the custom loss model was performed as described in Section 4.3. Like the MSE loss model, the test and prediction sets consisted of one instance each.

4.4.4 Residual models

As introduced by the taxonomy of Willard et al. (2020), residual models aim to learn the biases of physics-based models to be able to correct them. MyMDT's mechanistic model is constructed based on cardiovascular principles, and estimates the resting blood pressure curve after an exercise period with the stimulus defined in Equation 2.2.6. The estimates from the mechanistic model are not satisfactory, and the specific factors leading to the inaccurate and sometimes inadequate estimates are not known. By learning the limitations of the mechanistic model, a residual model's predictions added to the original estimates may improve the accuracy.

Two different residual models have been implemented, and aim to learn to error between two different blood pressure curves. The defined output of the two models is

$$target = [error_1, error_2, ..., error_{100}]. \tag{4.4.7}$$

Both models were constructed of three hidden layers, consisting of 50, 100, and 150 neurons, respectively, in addition to an input and output layer. The activation function passed to all hidden layers is the nonlinear function ReLU, and a linear activation function is used for the output layer. The network is further trained with the Adam optimizer with a learning rate of 0.001, and MSE as its loss function.

Residual real model

The first residual model is trained on real data, thereby referred to as the residual real model, and aims to learn the biases of MyMDT's mechanistic model. The model's input is the previously defined feature vector, and the target value is a calculated error curve between the mechanistic model's estimate and the true post-exercise blood pressure curve. The target value of the residual real model is defined as

$$error_i = P_i^{MM} - P_i, \tag{4.4.8}$$

where P_i^{MM} denotes the mechanistic model's estimated pressure value at point i, and P_i denotes the true pressure value at point i.

The residual real model was trained with a batch size of 5. Further, the model was evaluated as the process described in Section 4.3. The used dataset consisted of a total of 20 instances, where one instance was selected as the test set, one as the prediction set, and the remaining 18 constituted the training set.

Residual synthetic model

The second residual model was motivated by the limited available true data, and implemented to explore the possible improvement of a larger amount of training data. The model utilizes only synthetic data, and are thereby referred to as the residual synthetic data. The model aims to predict the error between the Windkessel model's estimate and the estimate from MyMDT's mechanistic model, where the latter is considered the true curve. The target value of the residual synthetic model is denoted

$$error_i = P_i^{WK} - P_i^{MM}, \tag{4.4.9}$$

where P_i^{WK} denotes the Windkessel model's estimated pressure value at point i.

The batch size was set to 32, a larger size than the residual real model due to a larger amount of training data. The full synthetic dataset of 3450 instances was used to train and validate the residual synthetic model. 3150 of these instances was used for training, and the remaining two individuals, consisting of 150 instances each, were used as test and prediction sets, as described in Section 4.3.

4.4.5 Synthetic data model

The limited data in this project may be an obstacle for the models described to perform satisfactorily. This motivated for the development of a model trained on large amounts of data, with the objective of evaluating amounts of data versus the accuracy of the model. The model is throughout the thesis referred to the synthetic data model. Different shares of the synthetic dataset described in Section 3.4 were randomly extracted and used to train the model. It was desired to see if the model could detect the effect of exercise on the synthetic blood pressure curve in combination with the individuals' personal features. As the objective of the model was to discover the required amount of data for predictions with a satisfactory accuracy, different amounts of data were in turns used to train the model.

The model architecture consists of an input and output layer, and two hidden layers. The two hidden layers in the network have 25 and 75 neurons, respectively, and the output layer has 100 neurons corresponding to the 100 points in the prediction. The input layer takes the defined input vector, which contains personal features for all training instances. The target value is a blood pressure curve represented as 100 discrete pressure values, defined in 4.4.2. ReLU is the implemented activation function of the hidden layers, and a linear activation function is used for the output layer. As optimizer, Adam is implemented with a learning rate of 0.01, and MSE defines the loss function. The batch size for training the synthetic data model was set to 32.

The synthetic dataset consists of 3450 instances in total. To evaluate the model, two individuals were used as test and prediction set as described in Section 4.3. Different shares of the training data, that is 1%, 20%, 40%, 60%, 80%, and 100%, were randomly selected and in turns used to train the model. The results of all shares are presented in Chapter 5.

Chapter 5

Results

This chapter presents the main results achieved from the models described in Chapter 4, and evaluated according to the metrics introduced in the same chapter. Firstly, the results from baseline model based on linear regression is presented, followed by the results from the MSE loss model and custom loss model. The results from the two residual models are presented accordingly, and lastly, the results from the synthetic data model trained with different amounts of data follows. Plots of predicted curves are summarized in Appendix A.

5.1 Linear regression model

The results for the implemented linear regression model are presented in Table 5.4.1, and plotted in Figure 5.4.1. The model achieved a pointwise error of $9.79 \pm 6.03\mathrm{mmHg}$, a DBP error of $5.41 \pm 5.97\mathrm{mmHg}$, and a SBP error of $13.09 \pm 7.47\mathrm{mmHg}$. The PP error was $10.17 \pm 7.79\mathrm{mmHg}$, and the MAP error $8.84 \pm 6.77\mathrm{mmHg}$.

5.2 MSE loss model

The model that predicts the blood pressure curve based on real data, with MSE as its loss function, achieved a pointwise error of $15.66 \pm 8.45\mathrm{mmHg}$. The DBP and SBP errors were $10.52 \pm 3.77\mathrm{mmHg}$ and $15.91 \pm 10.09\mathrm{mmHg}$, respectively, whereas the achieved PP error was $6.91 \pm 4.76\mathrm{mmHg}$. Lastly, the model's MAP error was $15.39 \pm 8.84\mathrm{mmHg}$. All the achieved scores are presented in Table 5.4.1 and a visual summary allowing comparison with other models is presented in Figure 5.4.1.

5.3 Custom loss model

The results from the model with a physics-guided loss function is presented and plotted in Table 5.4.1 and Figure 5.4.1. The model's achieved pointwise error was $9.14 \pm 4.35\text{mmHg}$. The DBP and SBP errors were $8.66 \pm 5.43\text{mmHg}$ and $12.95 \pm 8.15\text{mmHg}$, respectively, and the achieved PP and MAP errors were $8.27 \pm 6.78\text{mmHg}$ and $7.90 \pm 5.81\text{mmHg}$.

5.4 Residual models

5.4.1 Residual real model

The residual real model predicts the difference between MyMDT's mechanistic model's estimates and the true blood pressure curves. The achieved results are presented in Table 5.4.1 and illustrated in Figure 5.4.1. The model achieved a pointwise error of $12.65 \pm 8.27\text{mmHg}$, and DBP and SBP errors of $7.07 \pm 5.15\text{mmHg}$ and $15.17 \pm 7.25\text{mmHg}$, respectively. The PP error was $10.40 \pm 2.82\text{mmHg}$, and the MAP error was $10.68 \pm 9.75\text{mmHg}$.

5.4.2 Residual synthetic model

The residual synthetic model predicts the difference between MyMDT's mechanistic model and the Windkessel model, and achieved a pointwise error of $16.83 \pm 11.01\text{mmHg}$. The model's DBP error was $16.53 \pm 11.02\text{mmHg}$, and the SBP error $16.51 \pm 11.03\text{mmHg}$. A low error was achieved for the PP, that is $0.10 \pm 0.04\text{mmHg}$, while the MAP error was $16.54 \pm 11.01\text{mmHg}$. All pressure curve metric scores achieved by the model are presented in Table 5.4.1.

5.5 Synthetic data model

5.5.1 Pressure curve evaluation

As described in Section 4.4.5, different shares of the synthetic dataset were used to train the synthetic data model. Figure 5.5.1 presents a summary of the results in terms of feature errors, and are summed up in Table 5.5.1.

Firstly, only 1% of the total dataset was used, being 32 instances. The predictions with this model achieved a pointwise error of $18.91 \pm 5.44\text{mmHg}$, a DBP error of $16.81 \pm 5.87\text{mmHg}$, and a SBP error of $22.01 \pm 8.09\text{mmHg}$. Moreover, the PP and MAP errors were $14.75 \pm 5.08\text{mmHg}$ and $18.39 \pm 5.41\text{mmHg}$, respectively.

The model was further trained on 20% of the dataset, and resulted in decreased errors, that is $12.73 \pm 15.14\text{mmHg}$, $8.84 \pm 9.39\text{mmHg}$, $17.49 \pm 18.66\text{mmHg}$,

	Feature		
Model	**Pointwise error**	**DBP error**	**SBP error**
Linear regression	9.79 ± 6.03	5.41 ± 5.97	13.09 ± 7.47
MSE loss	15.66 ± 8.45	10.52 ± 3.77	15.91 ± 10.09
Custom loss	9.14 ± 4.35	8.66 ± 5.43	12.95 ± 8.15
Residual real data	12.65 ± 8.27	7.07 ± 5.15	15.17 ± 7.25
Residual synthetic data	16.83 ± 11.01	16.53 ± 11.02	16.51 ± 11.03
	PP error	**MAP error**	
Linear regression	10.17 ± 7.79	8.84 ± 6.77	
MSE loss	6.91 ± 4.76	15.39 ± 8.84	
Custom loss	8.27 ± 6.78	7.90 ± 5.81	
Residual real data	10.40 ± 2.82	10.68 ± 9.75	
Residual synthetic data	0.10 ± 0.04	16.54 ± 11.01	

Table 5.4.1: Achieved results from the baseline model with linear regression, the MSE loss model, the custom loss model, and the two residual models. The unit for all values is mmHg.

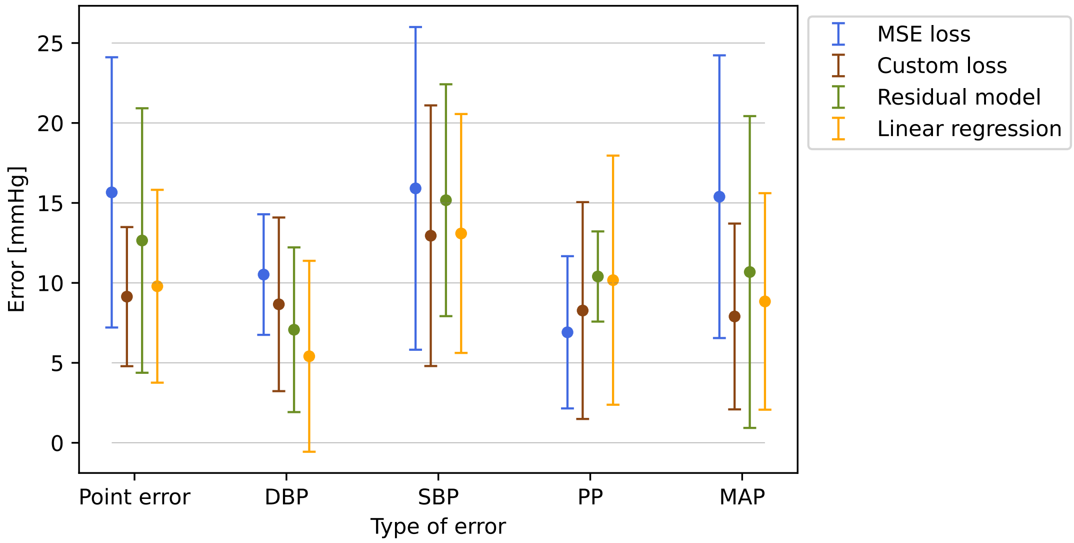


Figure 5.4.1: Errors and standard deviation of the error values of the MSE loss model and custom loss model, the residual real model, and the linear regression model.

$9.86 \pm 9.12\text{mmHg}$, and $12.04 \pm 15.66\text{mmHg}$ for the pointwise, DBP, SBP, PP, and MAP error, respectively.

A further increase in the dataset, to 40% of the full size, resulted in a pointwise error of $10.36 \pm 7.84\text{mmHg}$. The DBP and SBP errors were $9.35 \pm 7.13\text{mmHg}$ and $11.63 \pm 10.35\text{mmHg}$, respectively. Lastly, $11.36 \pm 5.71\text{mmHg}$ and $9.86 \pm 8.16\text{mmHg}$ were the achieved errors for PP and MAP.

With 60% of the full set, the pointwise error was $10.68 \pm 8.19\text{mmHg}$, the DBP error was $6.96 \pm 7.61\text{mmHg}$, and the SBP error was $11.26 \pm 10.12\text{mmHg}$. Further, the achieved PP and MAP errors were $9.51 \pm 6.76\text{mmHg}$, and $10.33 \pm 8.57\text{mmHg}$, respectively.

Next, 80% of the dataset was used to train the model. The achieved errors were $9.80 \pm 6.92\text{mmHg}$, $7.06 \pm 4.68\text{mmHg}$, $12.13 \pm 11.08\text{mmHg}$, $10.34 \pm 10.90\text{mmHg}$, and $9.18 \pm 7.29\text{mmHg}$, for the pointwise, DBP, SBP, PP, and MAP errors, respectively.

Lastly, the full amount of the synthetic dataset was used to train the model, that is 3150 instances. The pointwise error decreased to $7.90 \pm 4.93\text{mmHg}$, whereas the DBP and SBP errors decreased to $6.15 \pm 2.18\text{mmHg}$ and $9.63 \pm 5.66\text{mmHg}$, respectively. The achieved PP error was $9.86 \pm 4.79\text{mmHg}$, and the MAP error was $7.19 \pm 5.32\text{mmHg}$.

5.5.2 Classification evaluation

The classification performance of the models with synthetic data varied, and the progress of the different metrics according to the size of training set is illustrated in Figure 5.5.2.

The accuracies varied from 0.71 to 0.93, where the smallest dataset of 1% achieved the lowest score. A share of 20% achieved 0.87, whereas 40%, and 60% achieved accuracy scores of 0.91, and 0.92, respectively. The two largest datasets, 80% and 100% achieved the highest accuracy of 0.93.

The lowest precision score was, like the accuracy, achieved by the smallest dataset. A score of 0.61 was achieved for this model, whereas 40%, and 100% achieved 0.86, and 0.91, respectively. Further, 20% achieved 0.93, 80% achieved 0.94, and the best precision score was achieved by the model with a share of 60%, with a score of 0.97.

Recall denotes true positive rate, and the lowest achieved score was 0.55, being the model with only 1% of the total dataset. A score of 0.71 was achieved by the model with 20%, and scores of 0.80, 0.82, and 0.90, was achieved by the models with 60%, 80%, and 100%, respectively. A share of 40% outperformed the other models in terms of recall, achieving a score of 0.91.

The last classification metric evaluated is true negative rate. The smallest dataset achieved again the lowest score, that is 0.80. Further, 40% of the dataset achieved 0.91, and 100% achieved 0.95. Lastly, 20%, 80%, and 60%, achieved

Share of dataset	**Pointwise error**	**DBP error**	**SBP error**
1%	18.91 ± 5.44	16.81 ± 5.87	22.01 ± 8.09
20%	12.73 ± 15.14	8.84 ± 9.39	17.49 ± 18.66
40%	10.36 ± 7.84	9.35 ± 7.13	11.63 ± 10.35
60%	10.68 ± 8.19	6.96 ± 7.61	11.26 ± 10.12
80%	9.80 ± 6.92	7.06 ± 4.68	12.13 ± 11.08
100%	7.90 ± 4.93	6.15 ± 2.18	9.63 ± 5.66

Share of dataset	**Pointwise error**	**DBP error**	**SBP error**
1%	14.75 ± 5.08	18.39 ± 5.41	
20%	9.86 ± 9.12	12.04 ± 15.66	
40%	11.36 ± 5.71	9.86 ± 8.16	
60%	9.51 ± 6.76	10.33 ± 8.57	
80%	10.34 ± 10.90	9.18 ± 7.29	
100%	9.86 ± 4.79	7.19 ± 5.32	

Table 5.5.1: Results from training the synthetic data model on different amounts of data. The unit for all values is mmHg.

specificity scores of 0.97, 0.98, and 0.99, respectively.

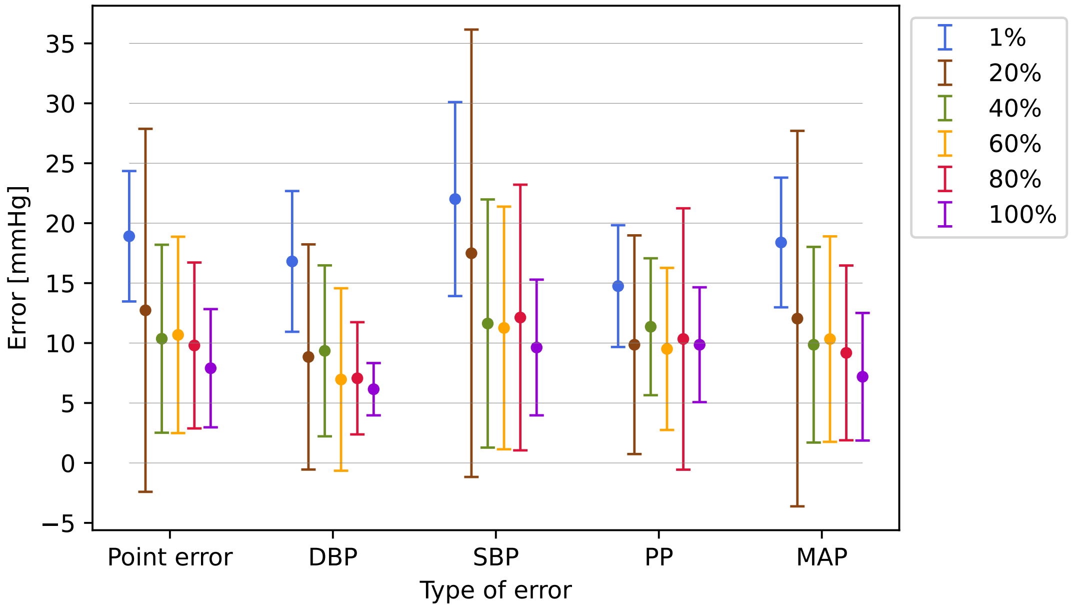


Figure 5.5.1: Error and standard deviation of the error values of the synthetic data model with different amounts of data.

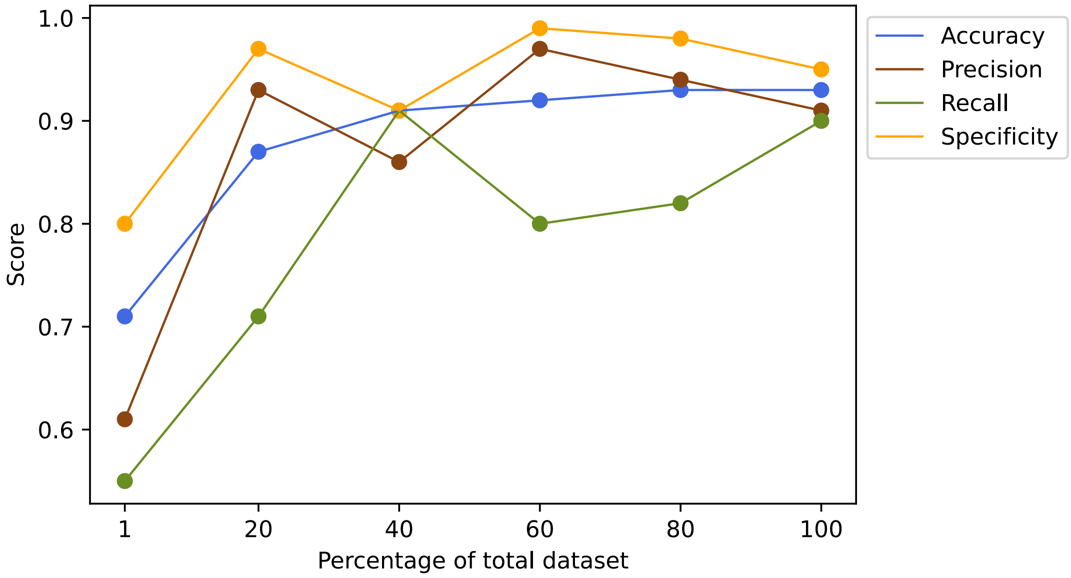


Figure 5.5.2: The progress of the different classification scores for the synthetic data model with an increased size of the training set.

Chapter 6

Discussion

In this chapter, uncertainties and sources of errors followed by the steps taken in this project are discussed. Further follows a discussion of the results and discovered effects of the applied IML methods, including assumptions about the required amount of data.

6.1 Data processing and the use of synthetic data

As described in Section 3.3, comprehensive data processing was performed on all blood pressure and flow measurements. The simplification of excluding the duration of the cardiac cycle and limit the curve to 100 points may have lead to lost information about a possible significant factor regarding the change in blood pressure. The limitation was still considered manageable as the benefit of equal length cycles and less data points were regarded as more valuable. It exists no defined solution on how to best obtain one representative curve from a full measurement signal. The standardization process applied in this project is one possible approach, and we acknowledge the possibility that there may exist more optimal solutions.

The synthetic data obtained from MyMDT's mechanistic model and the Windkessel model is based on the individuals' pressure and flow curves, an exercise value, and the models' estimated internal parameters. The models does not account for variations in age, gender, VO_2 max, or other features included in the input vector of the prediction models. This means that the features used to generate the synthetic data are not the same as the features used to predict it. Despite divergent input features, a machine learning model should be able to learn the determining factors with sufficient training data. However, some patterns must exist in the data for the model to learn. As the mechanistic and Windkessel model are based on the initial blood pressure and flow curves of an individual, a feature space including these parameters is assumed to provide clearer patterns.

Hawkins (2004) presents the problem of overfitting as the use of more terms or more complicated approaches than necessary, or including irrelevant components. This suggests that the models working with synthetic data may be prone to overfitting as they include features that the process creating the target data is ignorant about. However, it is not certain what are the determining features that realize the change in blood pressure, and we acknowledge that a more optimal feature space may exist for these models. Nevertheless, with the objective of comparing them to the models with real data, it was desired to utilize the same feature space for all models.

Additionally, synthetic data and its underlying patterns are not completely comparable to real data, hence uncertainties and sources of error will always be present when working with synthetic data.

6.2 Input and output of the models

When predicting the blood pressure curve of an individual several weeks ahead, it is reasonable to believe that the initial blood pressure curve is a relevant feature. Additionally, as the mechanistic and Windkessel model are based on the initial blood pressure and flow curves of an individual, a feature space including these parameters is assumed to provide clearer patterns for the models using synthetic data. Multiple ways of including the initial curve in the feature space have been discussed and tested. As the chosen curve consists of 100 points, the feature space would increase significantly if the entire curve was added as input. A different method considered was to create a PCA model based on all measured blood pressure curves, and calculate the scores of the PCs for each individual's initial curve. With 95% as the desired variance, three PCs was required, and the scores of the PCs for each individual were added to the feature space for testing. Evaluation showed that with scores of the PCs as a part of the input, the models' performance reduced, and consequently the initial blood pressure curve was excluded from the feature space. Some information about the initial blood pressure was however deemed crucial, and the ambulatory measurements of DBP and SBP were therefore included as input. Although an analysis was carried out during feature selection, we recognize, as mentioned, the possibility of a more optimal feature space, and see this as a possible optimization area.

The output of the models has also been reflected on. With the aim of predicting a blood pressure curve of 100 points, two different approaches were considered. The first approach was to predict a vector of 100 values. To predict 100 values based on seven input features seemed like a difficult task and a second approach was therefore tested. Each instance was divided into 100 separate instances, with one point on the pressure curve as each label. In addition to the input vector of seven features, a time feature representing the location on the curve was added. Prediction of 100 consecutive pressure values were used to evaluate the performance of the model. The evaluation of all models showed that prediction of one point on the pressure curve at a time achieved poorer results than the prediction of 100 points simultaneously. This may be because the division of the pressure curve deletes the connections between the points, which may impede the network from understanding the pressure values in a more complex context.

Other possible prediction outputs have been identified, such as a blood pressure curve of lower resolution than 100 points. This would simplify the target of the network but requires that a low resolution curve is within the desired outcome.

6.3 Informed machine learning in the prediction of blood pressure curves

6.3.1 Linear regression as a baseline

Figure 5.4.1 illustrates the results from the baseline model implemented with linear regression, in addition to the other models trained with real data. It can be observed that the linear regression model's performance had no major deviations compared to the other models. This may indicate that there exist linear relationships between the selected input features and the blood pressure curves. However, as the neural network models show lower errors for some pressure curve metrics, non-linear relations are also assumed present in the data. As discussed above, a further development of the feature space may increase the performance of the machine learning models. However, a more complex feature space may increase non-linear relations in the data, which a linear regression model would be unable to detect. This suggests that further optimization is feasible for the neural networks, while the linear regression model may have a limited performance. It is thus reasonable to believe that neural networks are preferable in the prediction of blood pressure curves.

6.3.2 The effect of a physics-guided loss function

The MSE loss model and custom loss model were trained on a small amount of data, and the results plotted in Figure 5.4.1 substantiates the hypothesis of limited data leading to relatively high variance. The results show that the model with impedance incorporated in the custom loss function achieved an overall better performance than the model with MSE loss. All pressure curve metrics, except the PP, achieved a lower error for the custom loss model. Additionally, the custom loss model achieved a lower standard deviation of the error values for three out of five metrics. The differences in errors were however not great, and the small amount of data leads to uncertainties, and the results are thereby not completely reliable. This hinders a direct conclusion, although the consistency of the comparisons may indicate that a physics-guided loss function can improve a neural network in the prediction of blood pressure curves. We acknowledge the narrow search of different physical contributions to the loss function, and we believe that the possibilities are unlimited. Impedance is only one of many possible ways of constraining a neural network, and both additional parameters and replacements may be worth exploring.

The impedance loss is calculated based on post-exercise pressure and flow measurements. With the comprehensive data processing done for both measurements, although they are measured simultaneously, the final curves are likely to be derived from different times during the measurement. This is a source of error when calculating the impedance as impedance denotes the relation between corresponding harmonics of pressure and flow. It would therefore be more optimal to utilize synchronized measurements.

6.3.3 The effect of residual modeling

The residual real model was implemented to assess the applicability of a residual model in the context of predicting blood pressure curves. The model aims to detect patterns between the input features and the difference between the mechanistic model's estimates and the true blood pressure curve. Figure 5.4.1 illustrates a comparison between the residual real model and the other models trained with real data. The residual model achieved lower errors than the standard machine learning model with MSE loss for four out of five pressure curve metrics. This may indicate that a pattern is plausible and again, that IML show tendencies of better training than standard machine learning models.

Compared to the IML model with custom loss, the residual real model showed poorer results to a small degree. The minor differences and the small amount of data complicates a conclusion about the best suitable IML approach in the task of predicting blood pressure curves. Additionally, only one physics-based model has been used as the residual target, and the possibilities of better suited models can thus not be ignored. Based on the achieved results, we recognize a potential in the concept of residual modeling, and thus believe that further exploration of the method may be valuable.

6.4 Amount of data vs. model performance

The results form the synthetic data model, in terms of pressure curve metrics, are illustrated in Figure 5.5.1, and verifies the hypothesis that an increase in training set size increases the performance accordingly. From the figure it can be observed that the model trained with only 1% of the full dataset achieved significantly higher errors. As 1% of the data, that is 32 instances, corresponds to the available amount of real data, a conclusion that the implemented models will increase performance by more training data can be drawn. Further, the errors decrease according to more training data, and with the full amount of the synthetic dataset, convergence of the errors is still not observed. This indicates that a further increase in the amount of training data will lead to better performance. Cornelissen and Smart (2013) show that an exercise period longer that four weeks reduces the DBP and SBP values of hypertensive individuals by about 8.3 mmHg and 5.2 mmHg, respectively. This implies that DBP and SBP errors of 5 mmHg, which is achieved for the model with 100% of the data, is too large when predicting blood pressure values. These findings suggests that more training data than 3150 instances, being the size of the synthetic dataset, is desired for the implemented models. However, further optimization may lead to better models, and an estimate of the required amount for satisfactory results is thus difficult to make.

The analysis of the performance of the synthetic data model at classification also holds some uncertainties. An average of random selected individuals of 150 instances was used to evaluate the model, and the distribution of positive and negative samples is thus random and may be skewed. The unknown distribution can affect the reliability of the classification metrics as a variety of data distributions is desired when performing such an analysis. With the uncertainties being acknowledged, we still recognize a value in the performed analysis.

From Figure 5.5.2 it is observed that the classification scores vary, but roughly increase according to an increased share of training data. The model with 100% of the dataset showed great stability in classification scores compared to the smaller shares of data. Some unexpected variations are discovered, and may be due to coincidences in terms of chosen prediction sets. A more detailed analysis of the splits and data distributions would be necessary to fully understand the reasons these variations.

A share of 60% achieved high scores for precision and specificity. High precision means that many of the samples predicted positive were indeed positive, and high specificity implies a high proportion of detected true negatives. Transferred to the real context, high precision results in few individuals falsely predicted to be hypertensive, and high specificity detects a large portion of the individuals being normotensive after the exercise period. These are important metrics as an effective training pattern is not desirable to go undetected.

6.4.1 Informed machine learning and required amount of data

The residual synthetic model was implemented to further assess the concept of residual modeling. The objective of the model was to test different sizes of training data and compare the performance with the synthetic data model. Such a comparison could asses if there were differences in the amounts of data required for a model based on IML and standard machine learning.

Initially, the residual synthetic model was trained with 100% of the synthetic data, and the results show that the residual real model outperformed the residual synthetic model. The PP was the only pressure curve metric with a lower achieved error for the model with a lot more training data. This was an unexpected result as the hypothesis that more data would increase the performance of the models with real data partly had been confirmed by the synthetic data model.

With a close look at the synthetic data from both the mechanistic and Windkessel model, it was observed that the estimated curves were not far from identical in shape for each instance. This results in an error curve close to a horizontal line, and not a complex target curve for the model that aims to predict it. This reflects the low PP error for the residual synthetic model, as a horizontal error curve added to the original estimate will preserve the shape of the curve. Apart from the PP error, the residual synthetic model performed significantly worse than the residual real model. The model was unstable and predicted pressure values far from the true values. The poor results may partly be explained by a trivial correlation between exercise value and the error between the mechanistic and Windkessel models' estimates. As exercise value is the only feature that differentiates instances derived from the same individual, a pattern between this value and error curve is essential to exist. Such a pattern has not been possible to detect by observing the data. If this is absent, the model that predicts the error between the mechanistic and Windkessel model is expected to encounter difficulties.

Due to the poor results from the residual synthetic model, the model could not be used as intended. A comparison with the synthetic data model was no longer relevant, and a conclusion of different amounts of data could thus not be drawn directly. However, the comparisons of the MSE loss model, with both the physics-based custom loss model and residual real model, favours the IML models. Additionally, with the benefit of a larger dataset, verified by the synthetic data model, it is reasonable to believe that IML approaches can support data science models and require less data to achieve equal performance.

Chapter 7

Conclusion and future work

7.1 Conclusion

The objective of this study was to implement and analyze different models for the task of predicting blood pressure curves of hypertensive individuals after a period of a given exercise. Five different neural networks were implemented, and no major deviations in performance was observed compared to the baseline model with linear regression. This indicates that there exist linear relationships between the chosen feature space of personal parameters and the post-exercise blood pressure curves. However, non-linear relations are also assumed present in the data. A further development of the feature space, which have been identified as an optimization area, may increase the non-linear relations and thus favor more complex machine learning methods than linear regression.

Comparison of different models demonstrated the benefits of IML in the task of predicting blood pressure curves. Two neural networks with different loss functions were implemented, and a loss function incorporating physics-knowledge proved to direct the network towards more accurate predictions than an ordinary MSE loss function. The physics-based loss function consisted of two weighted loss terms, MSE and impedance error, and although the impedance term did not affect the performance remarkably, the results indicate an improvement.

The use of IML was also assessed by implementing two residual models. These models aimed at predicting the error of physics-based models in order to correct their estimates. In comparison to the MSE loss model, the comparable residual model showed somewhat better results. However, a definite conclusion based on the small amount of real data in this project was not realistic. The second residual model was implemented with synthetic data to assess the possible improvement in performance with larger amounts of training data. The synthetic data was not of desired quality as patterns between the feature space and target curve were minimal. This lead to poor predictions of the residual model with synthetic data, and consequently no conclusion could be drawn.

To determine the amount of data required to obtain satisfactory results is not a straightforward task. A model was trained with different shares of a synthetic dataset, and the results confirm the hypothesis of insufficient amounts of real data in this project. The results show that an increase in the number of training samples increases the model's performance accordingly. IML approaches have in this project shown to perform better than standard machine learning models, and the analyses indicate that less data is required for IML to achieve equal performance quality. In conclusion, this project's experiences with IML indicates that it is a promising discipline in the context of blood pressure prediction. We therefore suggest further research of IML within this field.

7.2 Future work

We suggest multiple areas of focus for future work. Primarily, research of different physical contributions to incorporate in IML models may provide further knowledge regarding IML is this context. Impedance is only one of many possible ways of constraining a neural network, and both additional parameters and replacements may be worth exploring.

Further, an optimal feature space is crucial for a model to capture relevant patterns in the data. A specific proposal is to include the initial blood pressure curve in the input space as it is a highly descriptive feature of the initial state of the individual, and thus deemed important. Models with various architectures and target values require different feature spaces to obtain optimality, and consequently we suggest to strive for more tailored feature spaces.

An interpretable description of a discovered effective exercise is essential to give treatment advice to hypertensive individuals. The exercise value utilized in this project is an aggregated metric of the frequency, intensity, and duration of physical activity. To separate the quantities comprising this value could be valuable as a more comprehensible description of the exercise would be obtained.

Lastly, the synthetic data may be more realistic if new artificial individuals were constructed, instead of multiple training sessions for existing individuals. This would lead to totally independent instances, which is desired when training machine learning models.

Bibliography

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

AHS Alberta Health. How the heart works. https://myhealth.alberta.ca/Health/Pages/conditions.aspx?hwid=tx4097abc. Accessed: 2021-05-14.

Ruth E. Baker, Jose-Maria Peña, Jayaratnam Jayamohan, and Antoine Jérusalem. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol. Lett.*, 2018. doi: 1420170660.

José R Banegas, Luis M Ruilope, Alejandro de la Sierra, Ernest Vinyoles, Manuel Gorostidi, Juan J de la Cruz, Gema Ruiz-Hurtado, Julián Segura, Fernando Rodríguez-Artalejo, and Bryan Williams. Relationship between clinic and ambulatory blood-pressure measurements and mortality. *The New England Journal of Medicine*, 378:1509–1520, 2018. doi: 10.1056/NEJMoa1712231[doi].

M. W. Beck. Neuralnettools: Visualization and analysis tools for neural networks. *Journal of statistical software*, 85(11):1–20, 2018. doi: https://doi.org/10.18637/jss.v085.i11. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6262849/.

Tom Beucler, Michael Pritchard, Stephan Rasp, Jordan Ott, Pierre Baldi, and Pierre Gentine. Enforcing analytic constraints in neural-networks emulating physical systems, 2020.

Michael W Browne. Cross-validation methods. *Journal of mathematical psychology*, 44(1):108–132, 2000.

CDC. Estimated hypertension prevalence, treatment, and control among u.s. adults. `https://millionhearts.hhs.gov/data-reports/hypertension-prevalence.html`. Accessed: 2021-05-20.

Paul D. Chantler, Vojtech Melenovsky, Steven P. Schulman, Gary Gerstenblith, Lewis C. Becker, Luigi Ferrucci, Jerome L. Fleg, Edward G. Lakatta, and Samer S. Najjar. The sex-specific impact of systolic hypertension and systolic blood pressure on arterial-ventricular coupling at rest and during exercise. *American Journal of Physiology-Heart and Circulatory Physiology*, 295 (1):H145–H153, 2008.

Aram V Chobanian, George L Bakris, Henry R Black, William C Cushman, Lee A Green, Joseph L Jr Izzo, Daniel W Jones, Barry J Materson, Suzanne Oparil, Jackson T Jr Wright, and Edward J Roccella. The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the jnc 7 report. *JAMA*, 289(19):2560–2572, May 2003a.

Aram V Chobanian, George L Bakris, Henry R Black, William C Cushman, Lee A Green, Joseph L Jr Izzo, Daniel W Jones, Barry J Materson, Suzanne Oparil, Jackson T Jr Wright, and Edward J Roccella. The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the jnc 7 report. *JAMA*, 289(19):2560–2572, May 2003b.

Veronique Cornelissen and Neil Smart. Exercise training for blood pressure: A systematic review and meta-analysis. *Journal of the American Heart Association*, 2:e004473, 12 2013. doi: 10.1161/JAHA.112.004473.

Pádraig Cunningham. Dimension reduction, 2008.

Kari Anne Dalheim and Anine Ahlsand. Physics-guided neural network approaches to predict the blood pressure of hypertensive patients after 12 weeks of exercise. 2020.

Arinan Dourado and Felipe A. C. Viana. Physics-informed neural networks for corrosion-fatigue prognosis. *Annual Conference of the PHM Society*, 11(1), Sep 2019. doi: https://doi.org/10.36001/phmconf.2019.v11i1.814.

Aleksei Dudchenko, Matthias Ganzinger, and Georgy Kopanitsa. Machine learning algorithms in cardiology domain: A systematic review. *The Open Bioinformatics Journal*, 13:25–40, 2020. ISSN 1875-0362. doi: 10.2174/1875036202013010025.

David M. Eddy, Joshua Adler, Bradley Patterson, Don Lucas, Kurt A. Smith, and Macdonald Morris. Individualized guidelines: The potential for increasing quality and reducing costs. *Annals of Internal Medicine*, 2011. doi: https://doi.org/10.7326/0003-4819-154-9-201105030-00008.

Zoubin Ghahramani. *Unsupervised Learning*, pages 72–112. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. ISBN 978-3-540-28650-9. doi: 10.1007/978-3-540-28650-9_5. URL https://doi.org/10.1007/978-3-540-28650-9_5.

I. Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*. MIT press Cambridge, 2016.

Daniel G Hackam, Robert R Quinn, Pietro Ravani, Doreen M Rabi, Kaberi Dasgupta, Stella S Daskalopoulou, Nadia A Khan, Robert J Herman, Simon L Bacon, Lyne Cloutier, Martin Dawes, Simon W Rabkin, Richard E Gilbert, Marcel Ruzicka, Donald W McKay, Tavis S Campbell, Steven Grover, George Honos, Ernesto L Schiffrin, Peter Bolli, Thomas W Wilson, Ross D Feldman, Patrice Lindsay, Michael D Hill, Mark Gelfer, Kevin D Burns, Michel Vallée, G V Ramesh Prasad, Marcel Lebel, Donna McLean, J Malcolm O Arnold, Gordon W Moe, Jonathan G Howlett, Jean-Martin Boulanger, Pierre Larochelle, Lawrence A Leiter, Charlotte Jones, Richard I Ogilvie, Vincent Woo, Janusz Kaczorowski, Luc Trudeau, Robert J Petrella, Alain Milot, James A Stone, Denis Drouin, Kim L Lavoie, Maxime Lamarre-Cliche, Marshall Godwin, Guy Tremblay, Pavel Hamet, George Fodor, S George Carruthers, George B Pylypchuk, Ellen Burgess, Richard Lewanczuk, George K Dresser, S Brian Penner, Robert A Hegele, Philip A McFarlane, Mukul Sharma, Debra J Reid, Sheldon W Tobe, Luc Poirier, and Raj S Padwal. The 2013 canadian hypertension education program recommendations for blood pressure measurement, diagnosis, assessment of risk, prevention, and treatment of hypertension. *Can J Cardiol*, 29(5):528–542, May 2013.

Rana Haldar. Global brief on hypertension: Silent killer, global public health crisis. *Indian Journal of Physical Medicine and Rehabilitation*, 24:2–2, 03 2013. doi: 10.5005/ijopmr-24-1-2.

Douglas M. Hawkins. The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, 44(1):1–12, 2004. doi: 10.1021/ci0342472. URL https://doi.org/10.1021/ci0342472. PMID: 14741005.

Haibo He and E.a. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009. doi: 10.1109/tkde.2008.239.

L.R. Hellevik. Cardiovascular biomechanics. *NTNU*, pages 109–120, 11 2015.

L.R. Hellevik, K. Hveem, F. Lindseth, M. Steinert, I. Steinsland, and U. Wisløff. Kan en medisinsk digital tvilling gi deg bedre helse? https://www.ntnu.no/cerg/mymdt. Accessed: 2021-04-02.

J. Matthew Helm, Andrew M. Swiergosz, Heather S. Haeberle, Jaret M. Karnuta, Jonathan L. Schaffer, Viktor E. Krebs, Andrew I. Spitzer, and Prem N. Ramkumar. Machine learning and artificial intelligence: Definitions, applications, and future directions. *Current Reviews in Musculoskeletal Medicine*, 13 (1), Jan 2020. doi: https://doi.org/10.1007/s12178-020-09600-8.

Liviu Gr. Ixaru and Guido Vanden Berghe. *Runge-Kutta Solvers for Ordinary Differential Equations*, pages 223–304. Springer Netherlands, Dordrecht, 2004. doi: 10.1007/978-1-4020-2100-8_6. URL https://doi.org/10.1007/978-1-4020-2100-8_6.

Grob Jurgen. *Linear regression*. Springer, 2003.

Anuj Karpatne, William Watkins, Jordan S. Read, and Vipin Kumar. Physics-guided neural networks (PGNN): an application in lake temperature modeling. *CoRR*, abs/1710.11431, 2017. URL http://arxiv.org/abs/1710.11431.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

Georgios Kissas, Yibo Yang, Eileen Hwuang, Walter R. Witschey, John A. Detre, and Paris Perdikaris. Machine learning in cardiovascular flows modeling:predicting arterial blood pressure from non-invasive 4d flow mri data using physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 358:112623, 2020. ISSN 0045-7825. doi: https://doi.org/10.1016/j.cma.2019.112623. URL http://www.sciencedirect.com/science/article/pii/S0045782519305055.

C. Koch and I. Segev. The role of single neurons in information processing. *Nature Neuroscience*, 3:1171–1177, 2000. doi: https://doi.org/10.1038/81444. URL https://www.nature.com/articles/nn1100_1171.

A. Krogh. What are artificial neural networks? *Nature Biotechnology*, 26: 195–197, 2008. doi: https://doi.org/10.1038/nbt1386. URL https://www.nature.com/articles/nbt1386.

Marjaana Lahti-Koski, Pirjo Pietinen, Markku Heliövaara, and Erkki Vartiainen. Associations of body mass index and obesity with physical activity, food choices, alcohol intake, and smoking in the 1982–1997 FINRISK Studies. *The American Journal of Clinical Nutrition*, 75(5):809–817, 05 2002.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521 (7553):436–444, 2015.

Mei-Ling T. Lee, Bernard A. Rosner, and Scott T. Weiss. Relationship of blood pressure to cardiovascular death: The effects of pulse pressure in the elderly. *Annals of Epidemiology*, 9(2):101–107, 1999. ISSN 1047-2797. doi: https://doi.org/10.1016/S1047-2797(98)00034-9. URL `https://www.sciencedirect.com/science/article/pii/S1047279798000349`.

Xiaohan Li, Shu Wu, and Liang Wang. Blood pressure prediction via recurrent models with contextual layer. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, page 685–693, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee. ISBN 9781450349130. doi: 10.1145/3038912.3052604. URL `https://doi.org/10.1145/3038912.3052604`.

Elira Maksuti, Nicolaas Westerhof, Berend Westerhof, Michael Broomé, and Nikolaos Stergiopulos. Contribution of the arterial system and the heart to blood pressure during normal aging - a simulation study. *PloS one*, 11: e0157493, 06 2016. doi: 10.1371/journal.pone.0157493.

Sarah Melville and James Brian Byrd. Personalized medicine and the treatment of hypertension. *Current hypertension reports*, 21, 2010. doi: 10.1007/s11906-019-0921-3.

D.C. Montgomery, E.A. Peck, and G.G. Vining. *Introduction to Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, 2021a.

Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to linear regression analysis*. Wiley, 2021b.

Javaid Nauman, Bjarne M. Nes, Nina Zisko, Anders Revdal, Jonathan Myers, Leonard A. Kaminsky, and Ulrik Wisløff. Personal activity intelligence (pai): A new standard in activity tracking for obtaining a healthy cardiorespiratory fitness level and low cardiovascular risk. *Progress in Cardiovascular Diseases*, 62(2):179 – 185, 2019.

Bjarne M. Nes, Christian R. Gutvik, Carl J. Lavie, Javaid Nauman, and Ulrik Wisløff. Personalized activity intelligence (pai) for prevention of cardiovascular disease and promotion of physical activity. *The American Journal of Medicine*, 130(3):328 – 336, 2017.

D. L. Padmaja and B. Vishnuvardhan. Comparative study of feature subset selection methods for dimensionality reduction on scientific data. In *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, pages 31–34, 2016. doi: 10.1109/IACC.2016.16.

B K Pedersen and B Saltin. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scand J Med Sci Sports*, 25 Suppl 3:1–72, Dec 2015.

Han Peng, Weikang Gong, Christian F. Beckmann, Andrea Vedaldi, and Stephen M. Smith. Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, 68, Dec 2019. doi: 10.1101/2019.12.17.879346.

David M. W. Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, 2020.

Prajit Ramachandran, Barret Zoph, and Quoc Le. Searching for activation functions. 2018. URL `https://arxiv.org/pdf/1710.05941.pdf`.

B. Price S. Chaterjee, A.S. Hadi. *Regression analysis by example*. 1999.

R. W. Schafer. What is a savitzky-golay filter? [lecture notes]. *IEEE Signal Processing Magazine*, 28(4):111–117, 2011. doi: 10.1109/MSP.2011.941097.

Howard D. Sesso, Meir J. Stampfer, Bernard Rosner, Charles H. Hennekens, J. Michael Gaziano, JoAnn E. Manson, and Robert J. Glynn. Systolic and diastolic blood pressure, pulse pressure, and mean arterial pressure as predictors of cardiovascular disease risk in men. *Hypertension*, 36(5):801–807, 2000. doi: 10.1161/01.HYP.36.5.801. URL `https://www.ahajournals.org/doi/abs/10.1161/01.HYP.36.5.801`.

P Sibi, S Allwyn Jones, and P Siddarth. Analysis of different activation functions using back propagation neural networks. *Journal of theoretical and applied information technology*, 47(3):1264–1268, 2013.

Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: A family of discriminant measures for performance evaluation. In Abdul Sattar and Byeong-ho Kang, editors, *AI 2006: Advances in Artificial Intelligence*, pages 1015–1021, Berlin, Heidelberg, 2006. Springer Berlin Heidelber. ISBN 978-3-540-49788-2.

Juan-Juan Song, Zheng Ma, Juan Wang, Lin-Xi Chen, and Jiu-Chang Zhong. Gender differences in hypertension. *Journal of Cardiovascular Translational Research*, 13(1):47–54, 2020.

N Stergiopulos, J J Meister, and N Westerhof. Determinants of stroke volume and systolic and diastolic aortic pressure. *Am J Physiol*, 270(6 Pt 2):H2050–9, Jun 1996.

H Suga and K Sagawa. Instantaneous pressure-volume relationships and their ratio in the excised, supported canine left ventricle. *Circ Res*, 35(1):117–126, Jul 1974.

H Suga, K Sagawa, and A A Shoukas. Load independence of the instantaneous pressure-volume ratio of the canine left ventricle and effects of epinephrine and heart rate on the ratio. *Circ Res*, 32(3):314–322, Mar 1973.

James W. Taylor, Lilian M. de Menezes, and Patrick E. McSharry. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1):1–16, 2006. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2005.06.006. URL `https://www.sciencedirect.com/science/article/pii/S0169207005000907`.

Hector O. Ventura, Sandra J. Taler, and John E. Strobeck. Hypertension as a hemodynamic disease: The role of impedance cardiography in diagnostic, prognostic, and therapeutic decision making. *American Journal of Hypertension*, 18(S2):26S–43S, 02 2005. ISSN 0895-7061. doi: 10.1016/j.amjhyper.2004.11.002. URL `https://doi.org/10.1016/j.amjhyper.2004.11.002`.

Laura von Rueden et al. Informed machine learning – a taxonomy and survey of integrating knowledge into learning systems. 20:2, 2020.

Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009. doi: 10.1109/MSP.2008.930649.

Sanford Weisberg. *Applied Linear Regression*. Wiley, 2005.

WHO. Hypertension. `https://www.who.int/news-room/fact-sheets/detail/hypertension`. Accessed: 2020-09-18.

WHO. Q&as on hypertension, 2015. URL `https://www.who.int/news-room/q-a-detail/noncommunicable-diseases-hypertension`.

Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating physics-based modeling with machine learning: A survey, 2020.

Peter Wilson. Chapter 1 - grounding and wiring. In Peter Wilson, editor, *The Circuit Designer's Companion (Third Edition)*, pages 1–43. Newnes, Oxford, third edition edition, 2012. ISBN 978-0-08-097138-4. doi: https://doi.org/10.1016/B978-0-08-097138-4.00001-X. URL `https://www.sciencedirect.com/science/article/pii/B978008097138400001X`.

T. H. Wu, G. K. Pang, and E. W. Kwong. Predicting systolic blood pressure using machine learning. In *7th International Conference on Information and*

Automation for Sustainability, pages 1–6, 2014. doi: 10.1109/ICIAFS.2014.7069529.

G. Yang, X. Leng, F. Huang, M. V. Kasukurthi, Y. Huang, D. Li, J. Lin, S. Tan, G. Lu, R. Benton, G. M. Borchert, B. Ma, and J. Huang. Use cpet data to predict the intervention effect of aerobic exercise on young hypertensive patients. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1699–1702, 2019. doi: 10.1109/BIBM47256.2019.8983315.

D Zhang and S. Lou. The application research of neural network and bp algorithm in stock price pattern classification and prediction. *Future Generation Computer Systems*, 115:872–879, Oct 2020.

Appendix A

Standardization process

```
# Returns cycles_times, cycles_pressures
def split_signal(raw_time, pressure):
    pressure = pressure.T
    raw_time = raw_time.T
    raw_time = raw_time.to_numpy()
    pressure = pressure.to_numpy()
    pressure = pressure.reshape(-1)
    raw_time = raw_time.reshape(-1)
    peaks, _ = find_peaks(-1*pressure, distance=150)

    cycles_times = []
    cycles_pressures = []
    for i in range(1,len(peaks)):
        times = []
        pressures = []
        for j in range(peaks[i-1],peaks[i]):
            times.append(raw_time[j])
            pressures.append(pressure[j])
        cycles_times.append(times)
        cycles_pressures.append(pressures)

    return cycles_times, cycles_pressures

# Returns pressures_sav
def savitzky_golay(cycles, coarseness):
    pressures_sav = []
    for e in cycles:
        sav = savgol_filter(e, coarseness, 3)
        pressures_sav.append(sav)

    return pressures_sav

# Returns pressures_stand
def scale_x(cycles, cycles_times):
    pressures_stand = []
    for i in range(len(cycles)):
```

```python
            stand = []
            start = cycles_times[i][0]
            end = cycles_times[i][-1]
            for n in range(0,100,1):
                pressure = np.interp(start+n/100*(end-start),
                            cycles_times[i],cycles[i])
                stand.append(pressure)

            pressures_stand.append(stand)

        return pressures_stand

    # Returns pressures_lim
    def limit_cycles(cycles, dbp_min, sbp_min, sbp_min_i,
    sbp_max_i):
        pressures_lim = []
        for i in range(len(cycles)):
            if np.min(cycles[i]) > dbp_min and
            np.max(cycles[i]) > sbp_min:
                index, value = max(enumerate(cycles[i]),
                                key=operator.itemgetter(1))
                if index > sbp_min_i and index < sbp_max_i:
                    for w in np.arange(0,len(cycles[i])-29,10):
                        dev = np.std(cycles[i][w:w+30])
                        if dev < 0.6:
                            break

                    pressures_lim.append(cycles[i])

        return pressures_lim

    # Returns trend, stable
    def find_trend(cycles):
        signal = list(flatten(cycles))
        trend = savgol_filter(signal, 1001, 3)
        stable = signal - trend

        return trend, stable

    # Returns new_cycles
    def getCycles_without_trend(cycles):
        trend, stable = find_trend(cycles)
        new_cycles = []
        for n in np.arange(0,len(stable),100):
            cycle = []
            for i in range(100):
                cycle.append(stable[n+i])
            new_cycles.append(cycle)

        return new_cycles

    # Returns arrays
    def best_window(cycles_detrended):
        min_dev = math.inf
        min_w = 0
        for w in range(len(cycles_detrended)-8):
            a = list(flatten(cycles_detrended[w:w+8]))
            deviation = 0
            for p in np.arange(0,100,10):
                points = []
                for c in range(8):
                    points.append(a[c*100+p])
                dev = np.std(points)
                deviation += dev
            if deviation < min_dev:
                min_dev = deviation
                min_w = w

        arrays = [list(x) for x in cycles_lim[min_w:min_w+8]]

        return arrays, min_w

    # Returns cycles
    def split_window(arrays):
        cycles = []
        for n in range(len(arrays)):
            cyc = []
            for i in range(100):
                cyc.append(arrays[n][i])
            cycles.append(cyc)

        return cycles

    # Returns rep_cycle
    def find_representative_cycle(arrays):
        cycles_window = split_window(arrays)
        mean_curve = [np.mean(k) for k in zip(*arrays)]
        min_avvik = np.inf
        min_index = 0
        for c in range(len(cycles_window)):
            dev = 0
            for p in range(0,100,10):
                sa = abs(cycles_window[c][p] - mean_curve[p])
                dev += sa
            if dev < min_avvik:
                min_avvik = dev
                min_index = c

        rep_cycle = cycles_window[min_index]

        return rep_cycle


    # Returns scaled_cycle
    def scale_y(cycle, dbp, sbp):
        scaled_cycle = []
        peak = np.max(cycle)
        bottom = np.min(cycle)
        if sbp != np.nan and dbp != np.nan:
            for j in range(0, len(cycle), 1):
                y = cycle[j]
                norm_dist_from_bottom = (y-bottom)/(peak-bottom)
                norm_dist_from_top = 1-norm_dist_from_bottom
                new_y = sbp * norm_dist_from_bottom
                      + dbp * norm_dist_from_top
                scaled_cycle.append(new_y)

        return scaled_cycle
```

Listing A.1: Python code for the standardization process of the blood pressure curves

Appendix B

Custom loss function

```
def my_loss(y_pred, q, t):
    time = t.numpy()[0][0]
    y_pred = tf.dtypes.cast(y_pred, tf.float64)
    N = 100
    w = np.arange(N//2+1)*2*np.pi/time
    PP = tf.subtract(tf.reduce_max(y_pred,1),
            tf.reduce_min(y_pred,1))

    # Calculating integral of Q to find SO
    factor = q[0][0] + q[0][99]
    for i in range(1,100-1):
        factor += 2*q[0][i]
    SV = (time/100)*factor

    MAP = tf.reduce_mean(y_pred, axis=1)
    CO = SV / time
    R = MAP / CO
    C = PP / SV

    # Real and imaginary part of Z_model
    imag = -(tf.math.multiply
            (tf.math.multiply(tf.math.square(R),w),C))/
            (tf.math.multiply(tf.math.multiply
            (tf.math.square(R),tf.math.square(w)),
            tf.math.square(C))+1)

    real = R/(tf.math.multiply
            (tf.math.multiply(tf.math.square(R),
            tf.math.square(w)),tf.math.square(C))+1)

    z_model = tf.complex(real,imag)
    z_pred = tf.signal.rfft(tf.squeeze(y_pred,axis=2)) /
              tf.signal.rfft(q)
    z_diff = z_model-z_pred

    # Calculating magnitude of Z loss
```

```python
    a = tf.math.real(z_diff)
    b = tf.math.imag(z_diff)
    z_mag = tf.math.sqrt(tf.math.square(a)
                        + tf.math.square(b))
    z_sum = tf.math.reduce_sum(z_mag)

    total_loss = z_sum

    return total_loss
```

Listing B.1: Python code the for custom loss function

Appendix C

Prediction plots

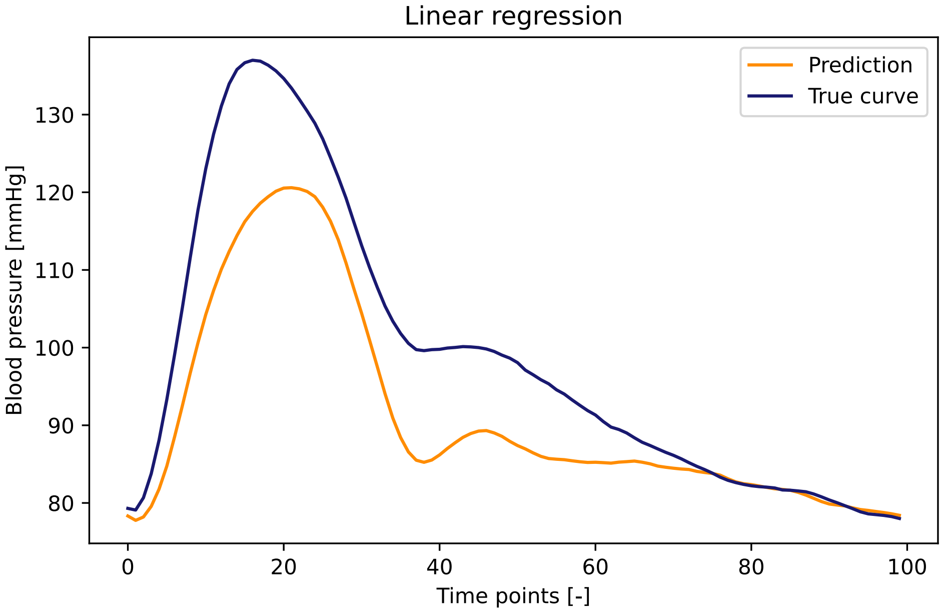


Figure C.0.1: Prediction by the baseline model with linear regression.

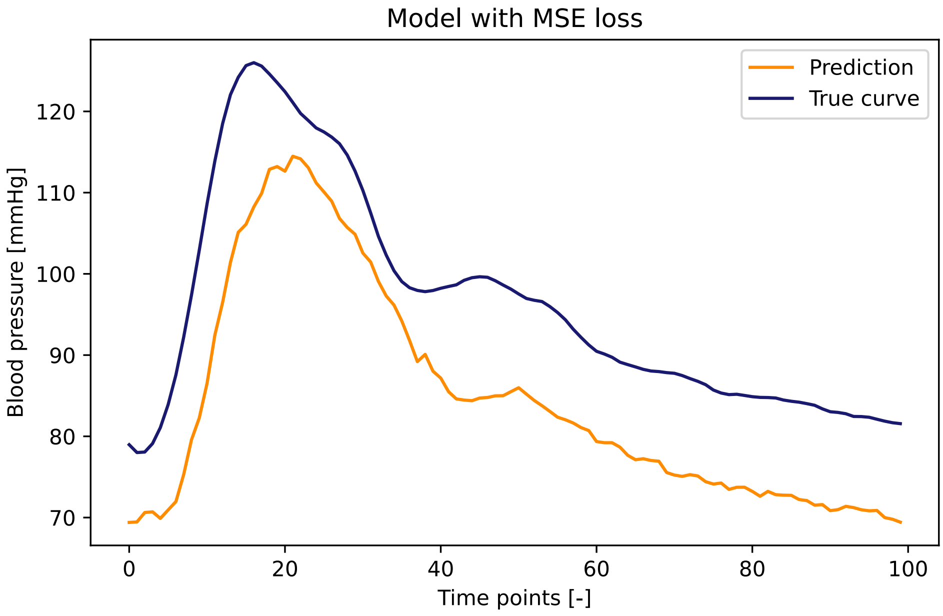


Figure C.0.2: Prediction by the MSE loss model.

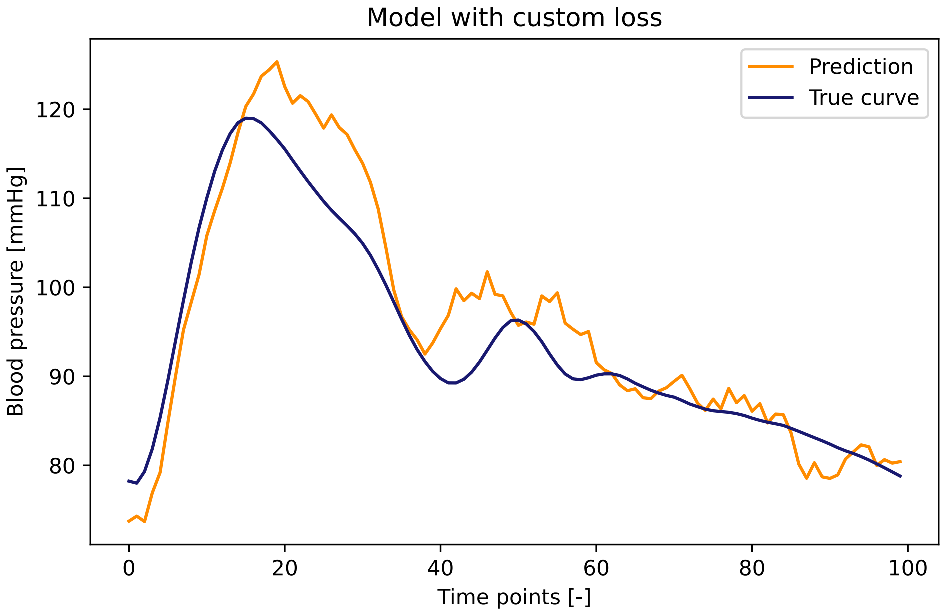


Figure C.0.3: Prediction by the custom loss model.

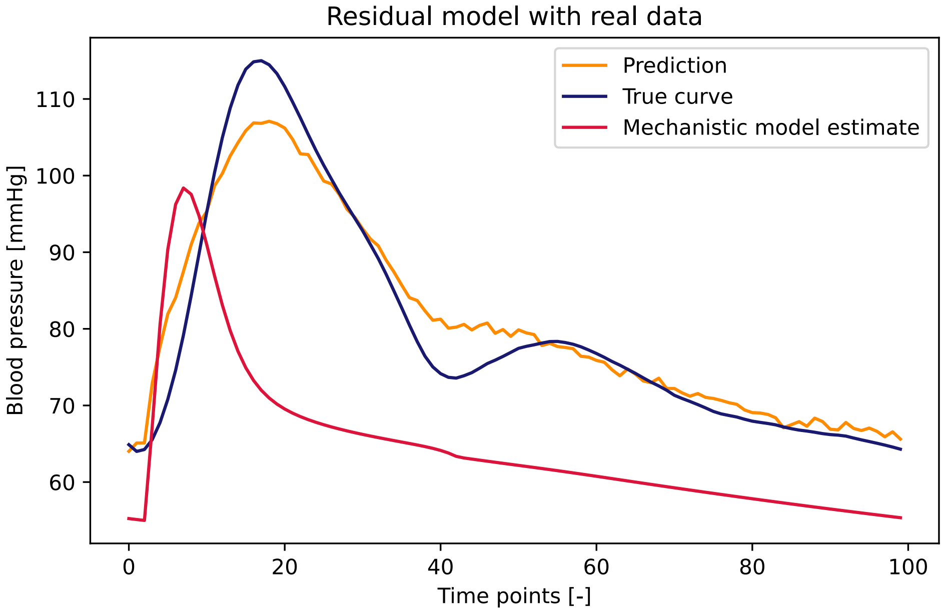


Figure C.0.4: Prediction by the residual real model.

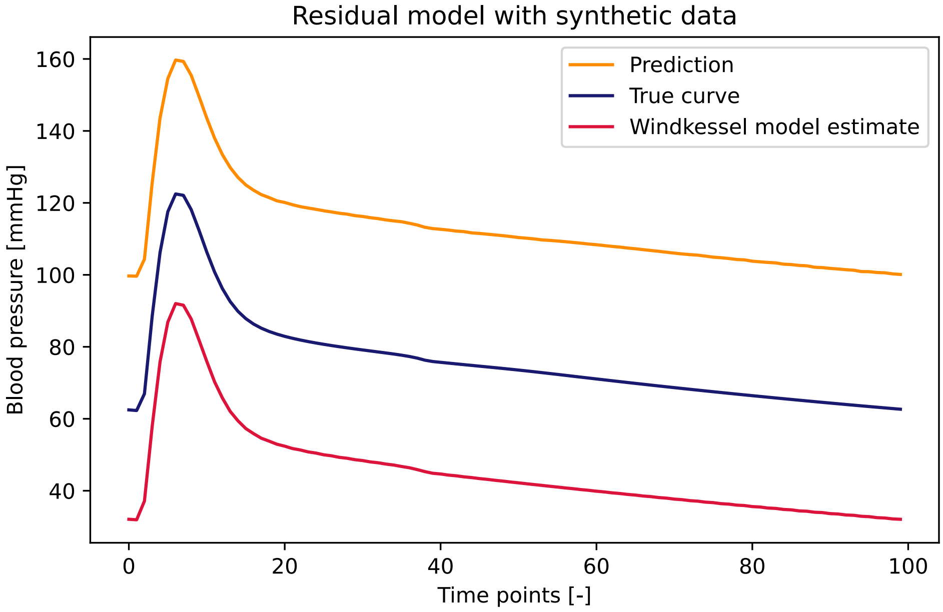


Figure C.0.5: Prediction by the residual synthetic model.

NTNU
Norwegian University of
Science and Technology