Politecnico di Torino

Politecnico di Torino

Master Degree in Biomedical Engineering

A.a. 2020/2021

Graduation session December 2021

A novel approach for blood pressure prediction using Machine Learning techniques

Supervisor:
Chiasserini Carla
Co-Supervisor:
Pagana Guido

Candidate:
Centonze Ilenia

Abstract

Sedentary lifestyle, unbalanced diet, stress and alcohol abuse are the risk factors of many diseases of worldwide population. Among them hypertension is known as "the silent killer" being the leading cause of premature death according to World Health Organization (WHO). Hypertension is the primary cause of the development of cardiovascular, kidney and brain diseases. After the development of a pathological cardiovascular condition in which blood pressure is constantly too high, the subject does not manifest immediately significant symptoms, and that is why it often leads to further complications.
Although the data show a particularly high incidence (1 in 5 women and 1 in 4 men) hypertension is a risk condition widely preventable and modifiable through interventions on the population and individuals at high risk. Thus, it is an objective of contemporary studies to use the most efficient and innovative technologies that allow society to prevent and monitor the risks and the progress of the problem. In recent years, with the development of miniaturized sensors and increasingly refined measurement techniques, this has become possible.

This thesis aims to provide a simple and effective method that will lay the foundation for future improvements in the prevention and monitoring of hypertension. In the current state of the art, it is not possible to have direct and non-invasive measurements of pressure in a reliable and continuous way. For this reason, the research is evolving by exploiting the non-invasive collection of other physiological signals of human body, such as photoplethysmogram (PPG) and electrocardiogram (ECG).

In this application, the development of an intelligent system starts from the study of the signals provided by MIMIC III database, which is a widely furnished online database about physiological data collected in ICU (Intensive Care Unit). It is proposed the processing of the input data by exploiting modern signal processing techniques, which have been implemented in Python programming language. A new approach for informative and consistent dataset construction was performed for the application of regression techniques, such as Linear Regression, Ridge Regression, Support Vector Regression and Random Forest Regression. The dataset contains data of Heart rate (HR) and Pulse Transit Time (PTT) related to specific

time windows has been used to train and then predict blood pressure of patients in a continuous way. Unfortunately, it is not possible to provide a model that could be suitable for every patient, and due to the dependence of blood pressure on other external factors, such as age and drug intake, the algorithm itself needs periodic re-calibrations to maintain its degree of accuracy.

The final results show prediction errors that are in accordance with the Association for the Advancement of Medical Instrumentation (AAMI) guidelines, obtaining a mean absolute error (MAE) of $1{,}96 \pm 1.44$ mmHg on the Diastolic Blood Pressure (DBP) values and 3.11 ± 1.89 mmHg on the Systolic Blood Pressure (SBP) values. These encouraging results promise future developments of the algorithm for wearable devices, and lay the groundwork for the improvement of an automatic, generic and high-performanced system.

Acknowledgements

Before proceeding with the discussion, I would like to dedicate a few lines to all those who have been close to me on this path of personal and professional growth. First of all, I would like to thank the LINKS Foundation for giving me the opportunity to work on this project and for allowing me to grow. I would also like to thank Professor Guido Pagana for having enthusiastically welcomed me into the team and believing in me. I would also like to thank Valeria Figini for her availability, her support and her kindness. It has been a fascinating journey, and thanks to you I have learned the beauty of always being fascinated and passionate about your work.

I would also like to thank my parents, who were a source of inspiration and courage. Despite the distance from home, your support has never been missing. Mum and Dad, you are and always will be the idea of the person I would like to be when I grow up.
To my sister, who supported me in moments of confusion and the desire to give up. You are for me the example of a determined person.

To my university friends, who during the pandemic period were with me behind a screen making university and life plans. I also wish you to be as happy as I am today, and in the future as we imagined during those endless Skype calls.

Finally, I would like to thank the Politecnico di Torino, for having encouraged my passions and for having, in these five years of university, provided opportunities for growth not only in teaching but also in social contexts of listening and sharing, art and science.

Contents

List of Figures

List of Tables

Acronyms

AAMI Association for the Advancement of Medical Instrumentation

ABP Arterial Blood Pressure

BP Blood Pressure

CVD Cardiovascular diseases

DBP Diastolic Blood Pressure

ECG Electrocardiogram

ESH European Society of Hypertension

ICU Intensive Care Unit

HR Heart Rate

KDE Kernel Density Estimation

MAE Mean Absolute Error

PTT Pulse Transit Time

PWT Pulse Wave Velocity

STD Standard Deviation

SVR Support Vector Regression

SBP Systolic Blood Pressure

PPG Photoplethysmogram

WHO World Health Organization

Chapter 1

The context

Coronary heart disease, cerebrovascular disease, rheumatic heart disease are part of a group of disorders called Cardiovascular diseases (CVDs), that, according to World Health Organization (WHO) data, are the main causes of death globally (are estimated 17.9 million lives each year) and in Europe causes more than 29 million Disability Adjusted Life Years (DALYs) [22][23]. Several epidemiologist's studies [24][25][26] have demonstrated that CVD shares a major risk factor: hypertension, that consists in continuously high values of one or both Systolic Blood pressure (SBP) and Diastolic blood pressure (DBP), which this often leads to debilitating cardiovascular and renal complications [27]. In light of these complications, prevention and control of hypertension become highly recommended.
At present non-pharmacologic strategies for the blood pressure decrease in hypertensive patients, such as a change in diet or lifestyle, can reduce blood pressure [26]. Sometimes, however, anti-hypertensive drug treatment is necessary, which decreases the complications of hypertension [26].

Current clinical practice for the diagnosis of hypertension (Figure 1.1) provides continuous blood pressure monitoring [28]. The methods used [29][30] are Ambulatory Monitoring (ABPM) and Home Monitoring (HBPM) : both are non-invasive methods that provide a sufficient clinical picture to identify high blood pressure values throughout the day [28]. Furthermore, sometimes hypertension does not manifest clearly to the patient: it is the case of "white coat hypertension" and "masked hypertension", that manifest high SBP or DBP in specific situation discussed in more detail in Chapter 2. It is also supported by major international guidelines [31][32][33] that Out-of-office measurements (i.e. measures taken in different context during the day) provide better prognostic information. Anyway, ABPM and HBPM also have a number of limitations, such as patient comfort, sleep disturbance, availability and cost [34].
Today, new technologies have been introduced and research is evolving toward measurement methods that affect the patient less and less: wearable sensors [35]. They

allow a more frequent monitoring of the vital signs with minimal stress on the patient. In addition, wearable sensors and devices was firstly thought for measuring blood pressure during exercise [35] and for this reason they often record other features, such as monitoring of environmental conditions, which can be useful for cross-analysis between the blood pressure detected and the context [34].
Recently, research is driven into use of Machine Learning (ML) technology [36]. In fact Blood Pressure is influenced by a variety of factors and basing on big data, through ML, it is possible to extract the optimal features required to monitor SBP and DBP and develop algorithms that produce values that meet system validation requirements defined by the Association for the Advancement of Medical Instrumentation (AAMI).

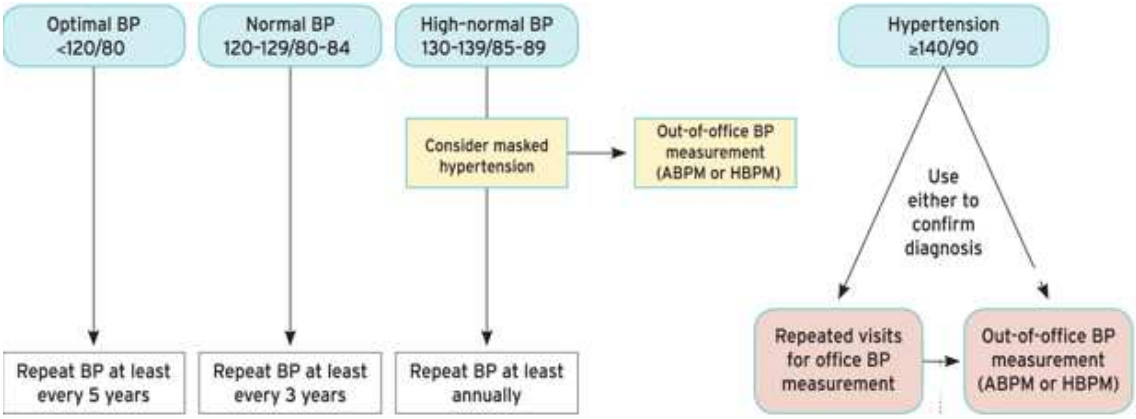
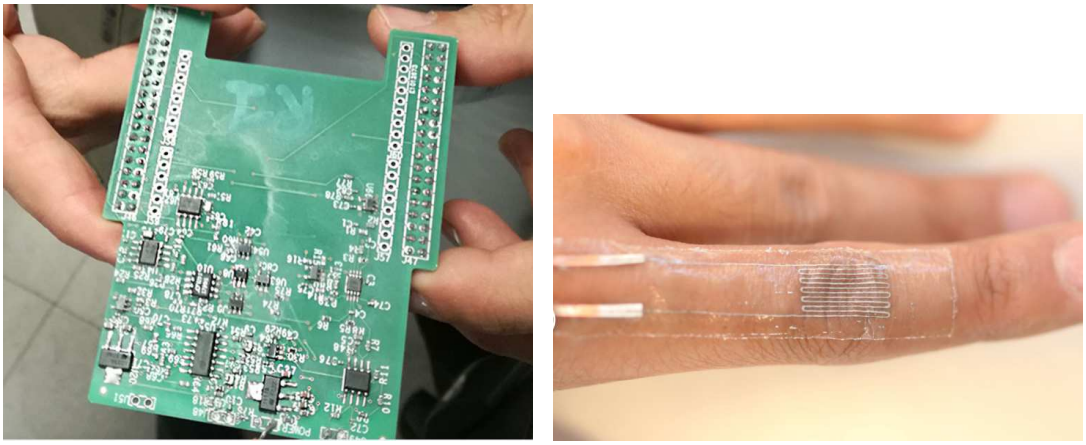


Figure 1.1: Clinical practice during Blood Pressure measurement for the management of arterial hypertension according to ESC/ESH Guidelines [1].

1.1 Thesis aims and objectives

This thesis work aims to develop a patient-specific Blood Pressure prediction algorithm from ECG and PPG for SINTEC devices with the purpose of monitoring hypertension.
The SINTEC project [2] is a European project that aims to obtain a soft, sticky, and stretchable sensor patches to be used for the recording of vital signal and their elaboration. The SINTEC device is designed for use in both clinical environment (e.g. medical technology) and performance assessment of athletes (e.g. in preventive care, sports and fitness).
This thesis work is one of the possible application of the smart patch, that will improve the everyday life of hypertensive patient, helping in monitoring Blood Pressure values. The expected software starts from the state of art to obtain a

reliable prototype for the real application. Signal processing and machine learning techniques have been used to achieve real-time measurement and allowing continuous monitoring of blood pressure. Since the patch is not an invasive or bulky device for the person wearing it, it is suitable to be worn at any time of the day and activity.
Focusing on the device, it can be seen in the Figure 1.2 that the smart patches are developed by a circuit board technology: small modules of common electronic components are mounted in a soft rubber material and connected with fluid alloy conductor tracks. So, the technological impact of this application is precisely in the use of the most novel artificial intelligence techniques combined with the design of the device itself.



(a) (b)

Figure 1.2: SINTEC device [2]: (a) the rigid sensor, (b) smart sensor patch.

Chapter 2

Introduction

The major anatomical components and physiological processes that regulate and influence blood pressure status are investigated. In addition, the fundamentals of machine learning and its biomedical areas of application are explored.

2.1 The cardiovascular system

The cardiovascular system is a closed system of vessels (the blood vessels) in which blood is moved through the force given from the heart: the blood pressure. It is formed by two main components:

- The systemic circulation (represented by the red pathway in Figure 2.1). It starts with oxygen entering in the system through lungs in the pulmonary capillaries. Then, it is delivered to body tissues via the systemic capillaries, covering a functional surface area in a range between $90m^2$ and $200m^2$.

- The small circulation, also called the pulmonary circulation (represented by the blue path in Figure 2.1). It forms a closed circuit between the heart and the lungs and begins in the right ventricle, from which the carbon dioxide rich blood is pumped. Firstly, it is injected into the pulmonary artery, which divides itself into two branches, each directed to a lung. Where the branches terminate to form capillaries that collect oxygen at the level of the alveoli. The oxygenated blood is then pumped into larger and larger vessels until it flows into the pulmonary veins and then returning to the heart and the systemic circulation.

Also blood vessels [37][38] have different names and characteristics and they are distinguished in arteries, capillary and veins (Figure 2.2). Main features are:

- Arteries are thick-sided tubes made of elastic tissue and muscle fibers (tunicas). Blood is injected into arteries with high pressure (normal ranges from

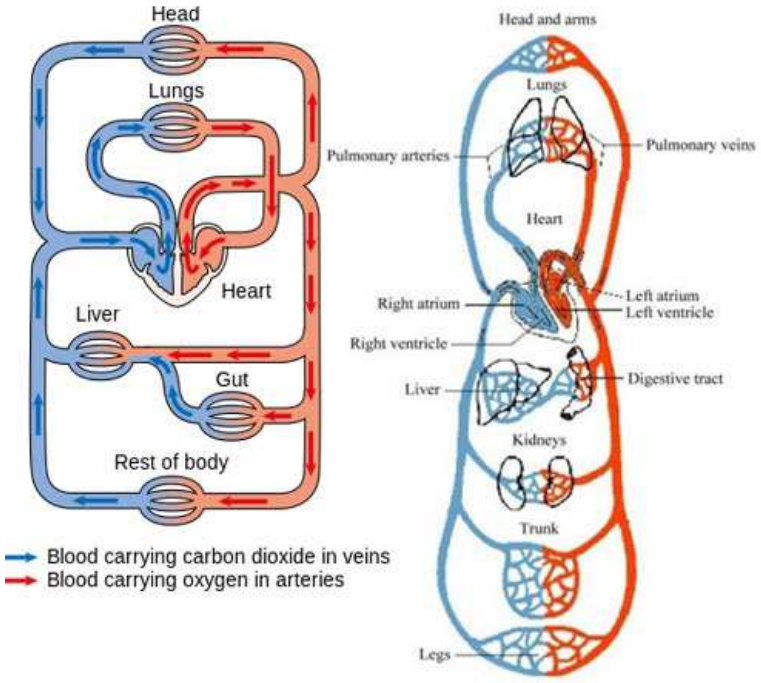


Figure 2.1: Schematic view of the cardiovascular system and blood flow in human body [3].

80 mmHg to 120 mmHg). From heart to periphery of the body or organs, arteries gradually branch into thinner blood vessels, eventually giving origin to arterioles through which blood is conveyed into capillaries.

- Capillaries form a dense network within all body tissues. The pressure in the capillaries ranges from 25 mmHg on the arterial side and 15 mmHg on the venous end [38]. They have thin walls, which is permeable by nutrients, gases and waste products. Capillaries channel blood to small blood vessels, venules, which converge to form veins.

- Veins are part of the pulmonary circulation and carry deoxygenated blood. Compared to arteries, veins have thinner and less well organised walls, larger lumens and bigger diameters [39]. The venous system collectively has approximately two thirds of all the blood. Because of the absence of a pumping force the phenomenon of blood backflow (or reverse flow) is prevented by venous valves in the extremities [38].

Hence, the cardiovascular system allows blood to circulate through the body for enabling transport of nutrients, hormones, oxygen, carbon dioxide, and blood cells to and from the cells. As consequence, the body is supplied with nourishment and can dispose of metabolic waste products, stabilize temperature and pH, and

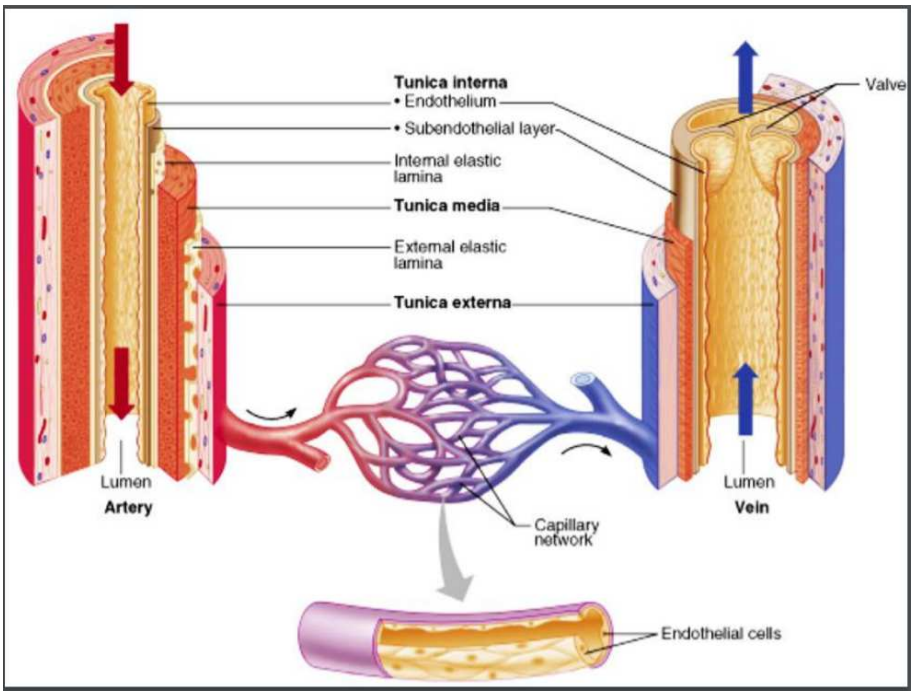


Figure 2.2: Arteries, capillaries and veins communication and components representation.

maintain homeostasis [40][41].

2.1.1 The heart

The heart is one of the most important organs of the human body and it is the engine of the cardiovascular system. It is an involuntary muscle with a unique property: it generates by itself the electric signal need for the contraction [42]. The heart is located in the centre of the chest between the right and left lung, just below the sternum.

As shown in Figure 2.3 the internal anatomy of the heart consists of four chambers: two atria, called upper chambers, and two ventricles called lower chambers. The atria are receiving chambers. Four pulmonary veins enter the left atrium coming from the lungs, two of them are visible in the anterior section of the heart in Figure 2.3. In the right atrium, instead, blood enters through three veins:

- The superior vena cava, that returns blood from body regions superior to the diaphragm.
- The inferior vena cava, that returns blood from body areas below the diaphragm.

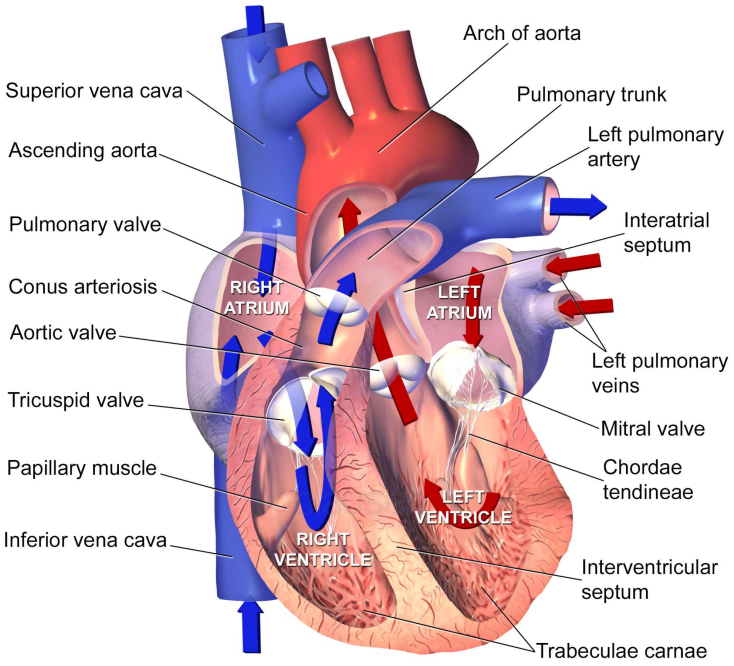


Figure 2.3: Explained anterior Heart Sectional Anatomy [4].

- The coronary sinus, that collects blood draining from the myocardium.

Atria are small and thin-walled chambers because they need a minimal contraction to push blood into the ventricles, due to gravity.
Ventricles are separated by the interventricular septum and make up most of the volume of the heart. They operate as discharging chambers: the right ventricle pumps blood into the pulmonary trunk, which sends the blood to the lungs, where gas exchange occurs, and the left ventricle ejects blood into the aorta, the largest artery in the body.

There is a unidirectional flow of blood through the heart; this is accomplished by four valves, which separate chambers [4]. One valve lies between each atrium and ventricle, and they are called atrioventricular valves (AVs) (tricuspid or bicuspid), the other valves rest at the exit of each ventricle, and they are called semilunar valves (pulmonary or semilunar).
These hollow compartments are delimited by heart walls (Figure 2.4):

- Endocardium is the epithelial tissue lining in contact with the blood and helps the contraction of the middle myocardium layer [43].

- Myocardium is the thickest layer and is composed by cardiomyocytes, the cells that act the contraction and the heart beating, and by pacemaker cells [44] .

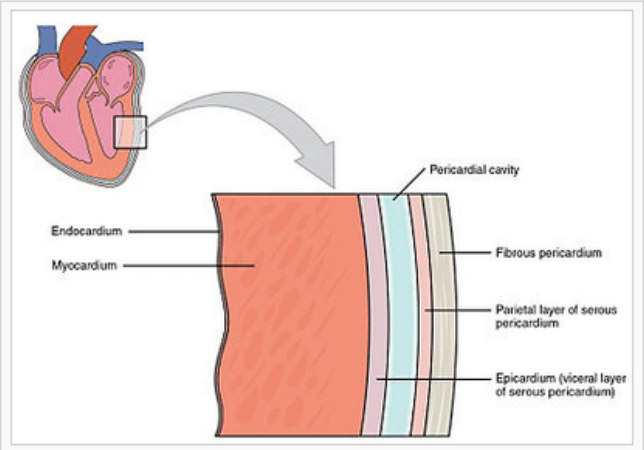


Figure 2.4: Cross-section of heart walls [5].

- Pericardium covers the heart and works as protection layer from infections coming from other organs. Moreover, it prevents excessive dilation of the heart in case of volume overload and the setting of the heart in the mediastinum [45]. It is composed by epicardium, pericardial cavity, a parietal layer and another fibrous one.

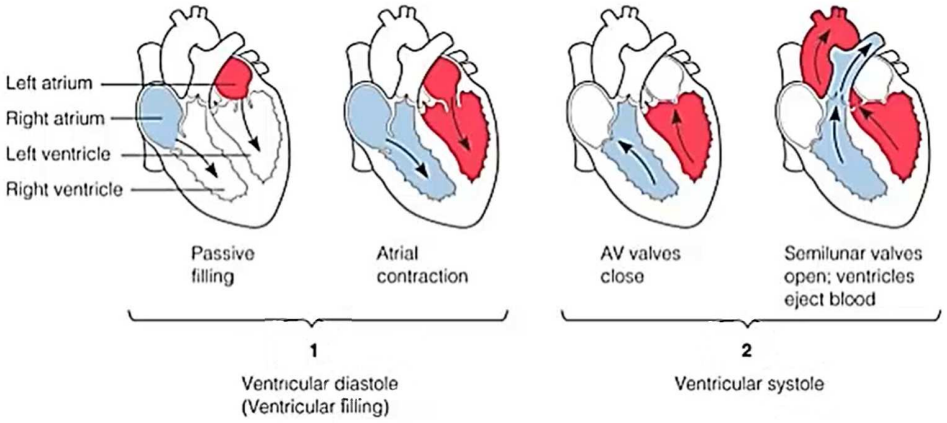


Figure 2.5: Main steps of an heartbeat.

Synergic work above all the components of the heart is the cardiac cycle or heartbeat and it consists in two main phases: the systole (the cardiac contraction) and the diastole (the cardiac relaxation). An heartbeat can be detailed in four main steps (Figure 2.6) [46]:

1. Isovolumic relaxation or passive filling: aortic valve is closed and the atria are filling;

2. Atrial systole and ventricular filling: the atria contract and the AV valves is opened to pump blood into the ventricles;

3. Isovolumic contraction: all the valves are closed until the pressure of the heart is higher than the pressure in the aorta;

4. Ejection: blood is pumped from ventricles to the aorta and pulmonary arteries through their own valves.

The time that elapses between two cardiac cycles is assessed by a frequency value: the heart rate (HR). It is used as a parameter of performance of human heart, widely used by clinicians in defining the clinical picture. In a healthy adult is about 70 beats per minute during resting activity. However, it can change in order to satisfy the body needs [46]. The mechanism of the adjustment of the Heart Rate is influenced by environment: the heart cells receive nerve signals from the vagus nerve and from nerves arising from the sympathetic trunk [42][47][48]:

- The vagus nerve of the parasympathetic nervous system acts to decrease the heart rate.

- Nerves from the sympathetic trunk, which form a network of nerves that lies over the heart called the cardiac plexus, act to increase the heart rate.

There are several types of reflex that induce chronotropic effects (i.e. heart rate modifying) through increases or decreases in sympathetic and parasympathetic activity. Reflex can be transmitted by:

- Baroreceptors, i.e. receptors sensitive to blood pressure, in the aortic arch, carotid sinus and cardiac chambers.

- Chemoreceptors, which are receptors sensitive to arterial oxygen and carbon dioxide concentration, also found in the aortic arch and carotid sinus.

The effects of changing heart rate, therefore the sympathetic-vagal stimulation, affects venoconstriction, vasoconstriction and cardiac contractility in order to maintain blood pressure and blood flow levels appropriate to the metabolic needs of each organ and tissue [47].
In resting situations the decreasing of heart rate implied reduced force of contraction, constricted coronary arteries, thus saving energy.

2.2 Bio-signals

2.2.1 Electrocardiography

Electrocardiography is a non-invasive method that allows to obtain a representation of the electrical activity of the heart and is one of the most widely used tools for the diagnosis of various disorders, even those not directly related to the activity of the pumping system of our body [49].
For the collection of this particular waveform there is an instrument called electrocardiograph which allows to obtain the electrocardiogram (ECG), which is the final representation of a series of complex physiological and technological processes.

As shown in Figure 2.6, the heartbeat begins at the right atrium called Sino

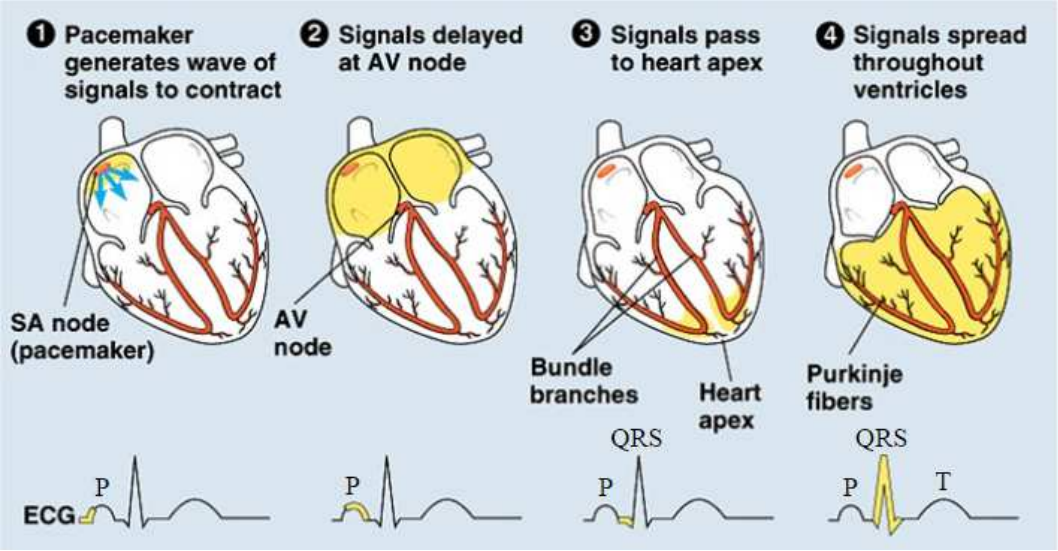


Figure 2.6: Scheme of heart electrical conduction pathway [6].

atria (SA) node and the pacemaker cells transmits these electrical signals across the heart cells. Then, this signal moves from atria to the atrioventricular (AV) node. The AV node connects to a group of fibers, called Purkinje fibers, through the Bundle branches (or Bundle of His), in ventricles that conducts the electrical signal and sends the impulse to all parts of the ventricles (third and fourth steps). So, during the cardiac cycle the electrical field in and around the heart change with time, due to these currents, that are synchronized by cardiac activation and recovery sequences [50].

This electrical field is perturbated from numerous other structures, including the lungs, blood and skeletal muscle before the currents reach the skin. The signal is detected attenuated by electrodes placed in specific locations [7]. The outputs are

amplified, filtered and displayed by a variety of devices to produce an electrocardiographic recording. These signals, if they are processed by a computerized system, are digitized, recorded, stored and elaborated by pattern recognition software. Diagnostic criteria are then applied, either manually by the clinician or with the aid of a computer, to produce an interpretation [50][51].
The ECG typically has an amplitude of the order of mV, in particular it is larger than 0.5 mV, the upper limit of the amplitude is about $2.5 - 3.0$ mV.
The following features can be identified in the cardiac cycle (Figure 2.7):

- P wave: represents atrial depolarization;
- PR interval: time elapsed between atrial depolarization and the onset of ventricular depolarization;
- QRS complex: ventricular depolarization;
- QT interval: time taken by the ventricles to depolarize and repolarize;
- ST interval: plateau phase of action potentiation at the ventricular level;
- T wave: ventricular repolarization. It has a positive deflection because the last cells to depolarize are the first ones to repolarize;
- U wave: last signal of ventricular repolarization, it is present only in particular clinical situation.

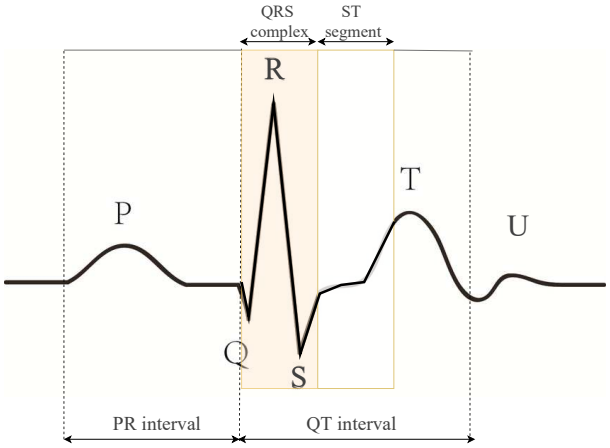


Figure 2.7: Schematic ECG representation of an heartbeat in normal sinus rhythm.

The one shown in Figure 2.7 is one of the possible representation of the ECG, and it depends on the configuration of the electrodes on the skin of the patient and the aim of the recording [52]. The most common configuration is based on the Einthoven's triangle, in which the position of the body of the patient outlines three

extremity points, where electrodes will be placed, which connection forms an imaginary triangle (Figure 2.8). The three electrodes are connected to a device that measures the voltage and the potential difference of each pair of electrodes: the resulting signal defines a derivation or lead [7]. Different proposes mean different electrodes configurations, in number and location: it can be reached the number of twelve leads with ten electrodes for the deepest analysis of the heart activity [52].

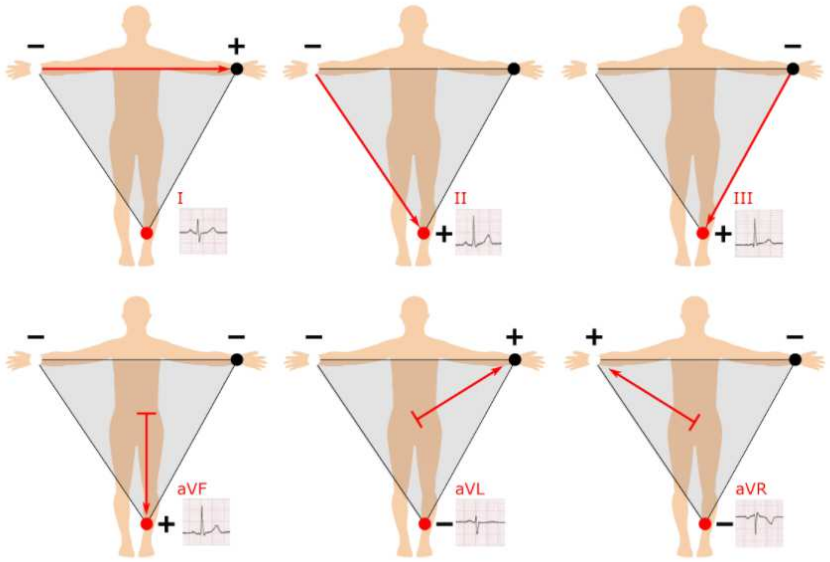


Figure 2.8: Eithoven's triangle [7].

From ECG waveform can be computed the HR of the i_{th} cardiac cycle using the time distance between two R peaks:

$$HR_i = \frac{1}{R_i - R_{i-1}} \tag{2.1}$$

It is expressed in beat per minute (bpm) and it is a frequency measure.
The goal of performing an ECG aim to obtain indicators for the evaluation of a suspected myocardial infarction (heart attack), diagnosis of cardiac stress, measuring the heart rhythm and any increase in the size of the heart chambers, diagnosing latent heart disease, monitoring of symptoms such as shortness of breath and arrhythmia [52][53].
The clinical usefulness of the ECG is developed in more than one hundred years of studies, and it aim to reach more accurate diagnostic criteria exploiting the full application of the capabilities of the computerized system [50]. The most recently reported application is in the field of wearable technologies, which have enabled continuous monitoring of the electrocardiogram signal, without the need for any laboratory settings.

2.2.2 Photoplethysmography

Each cardiac cycle the heart pumps oxygenated blood to the periphery for the nourishment distribution, until it reaches the skin, causing the relaxation of arteries and arterioles in the subcutaneous tissue [54]. This phenomena are exploited for the recording of the photoplethysmography (PPG), a non-invasive technique used to make measurements at the skin surface to detect blood volume variations [55]. The main factors that affect these changes are [56]:

- Breathing;
- Vasomotor activity;
- Sympathetic nervous system activity;
- Thermoregulation;
- Micro-vascular bed of tissue changing in absorption.

The device used for collecting PPG use a sensor to detect the change in volume caused by the pressure pulse by illuminating the skin with the light from a light-emitting diode (LED). The LED can be set to wavelengths ranging from 532 nm (green light) when it is necessary to stay on the surface, in applications such as quantifying oxygen in haemoglobin, to 1064 nm (infrared or IR) when it is necessary to go deep, in applications such as measuring vascular branch volume [57]. In fact, it is demonstrated that lights with longer wavelengths penetrates more deeply into the tissue. However, infrared light is more susceptible to motion artifacts. For PPG collection IR is employed. The light emitted by the LED is then rejected by the skin and a photo-detector (PD) measures its intensity (Figure 2.9). Even if the light output by the led is absorbed by all the tissues beam through, the absorption by blood is higher than the others and the changes in blood volume can be detected by a proportional dependency [58]. Because the skin is richly perfused, small changes in blood volume can be detected using this method, although it ca not be used to quantify the amount of blood.

So, it is simple to detect the pulsating component of the cardiac cycle.
The PPG waveform is unpacked in an alternating (AC) and a direct (DC) components [59]:

- The AC component represents blood volume cardiac variation at each heartbeat and is attributed to the pulsating behaviour of heart;
- The DC component is highly correlated to central and periphery venous pressure (respiration, sympathetic nervous system activity, and thermoregulation).

The AC component of the signal depicts changes in blood volume that can be studied by monitoring adjustments in systolic and diastolic phases (Figure 2.9).

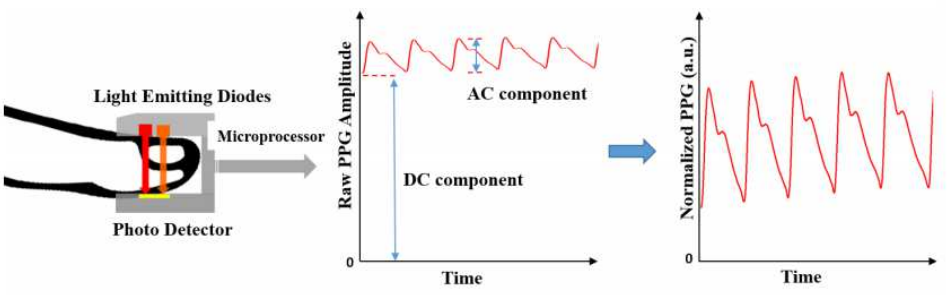


Figure 2.9: Scheme of a PPG instrument design, the signal collection and decomposition [8].

As shown in Figure 2.10, a typical PPG waveform of a healthy and compliant vessel network has a systolic phase (A area in Figure 2.10) with a systolic peak a dicrotic notch and a diastolic phase (B area in Figure 2.10) with a diastolic peak. Using these parameters may be derived other morphological features including area under the curve, ratio between amplitudes (a/b measure in Figure 2.10), pulse width, etc. that allow to obtain vital information including cardiac output, arterial ageing, endothelial function, and autonomic function [55].

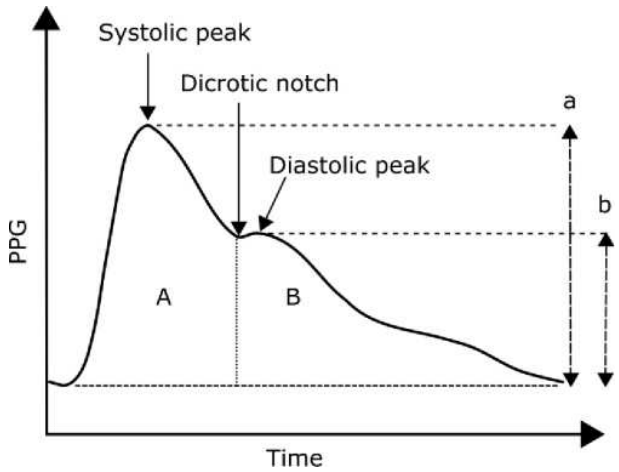


Figure 2.10: Scheme of a typical PPG waveform [9].

One of the most popular usage by PPG waveform is the heart rate monitoring, especially as alternative to ECG in less invasive application, such as personal monitoring devices in sports, since PPG sensor can be placed on diverse body regions such as fingertips, wrists, or thighs [59].
Due to recent COVID-19, research interest in PPG sensors increases [60][61]. In early stages of COVID-19 disease, in fact, the oxygen saturation level of a patient decreases to a specific deficient level (less than 60%) without any symptom, such

as shortness of breath (silent hypoxia), leading to severe lung damage [62]. So, since the PPG technique extracts oxygen saturation rate, a device continuously monitoring PPG possibly detects silent hypoxia [63].

In this work, PPG signal is used in combination with the ECG are both collected by a wearable sensor (one spot for each sensor) and used for the extraction of the Pulse Transit Time (PTT).
PTT represents the time that the blood pressure pulse takes to travel from a site to another into the arteries. It is determined as the interval between onset of cardiac ejection (approximated by the R peak in the ECG) and the arrival of the systolic peak in the PPG wave (Figure 2.11) and measured with the Equation 2.2 for each cardiac cycle:

$$PTT_i = SP_i - R_i \tag{2.2}$$

it is a delay measure: SP_i is the Systolic Point of the PPG in the *i-th* cardiac cycle and R_i is the time at which the arrival of the R peak is detected, both are expressed in milliseconds.
PTT is mainly influenced by: (i) the stiffness of the arterial wall, (ii) age, (iii) stress level, (iv) temperature, (v) posture.

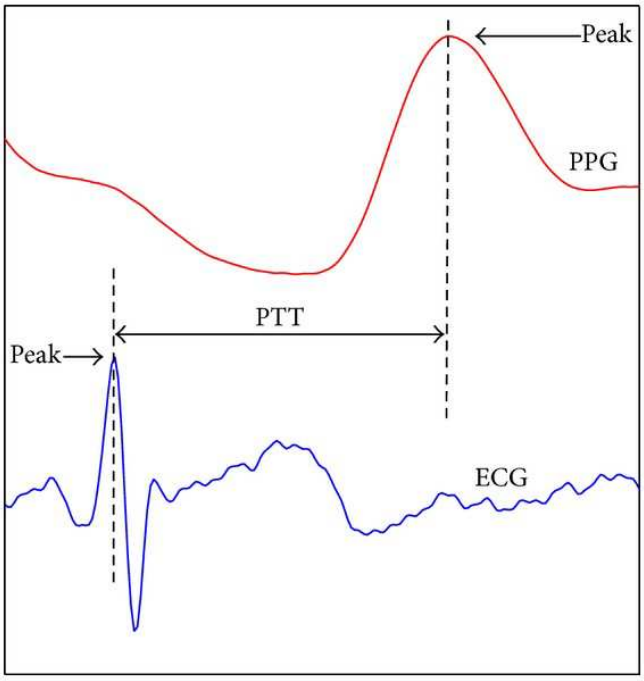


Figure 2.11: Graphical representation of PTT calculation [10].

The importance of PTT in this work is led back to the following assumption:

- It is measured with non-invasive method: suitable for this application, because the subjects have to be continuously monitored.
- Its relationship with the BP: it is demonstrated the possibility to measure SBP through the Pulse Wave Velocity (PWV), that derives from PTT according to the Equation 2.3:

$$PWV = \frac{K}{PTT} \tag{2.3}$$

where coefficient K represents the distance travelled by blood between two arterial locations [64]. In particular, the PWV is the velocity of propagation of the pressure pulse from the aortic valve to the entire arterial tree (typical values are greater than $4ms^{-1}$). PWV can be calculated with the Moens-Korteweg equation, a bio-mechanical formula that establishes the PWV dependencies:

$$PWV = \sqrt{\frac{hE}{\rho d}} \tag{2.4}$$

where h is the wall thickness, d the diameter, ρ the blood density and E the vase elastic modulus is calculated using the Geddes formulation:

$$E = E_0 e^{\gamma p} \tag{2.5}$$

where E_0 is the elastic modulus at zero pressure, p the pressure and γ a constant. Thus, through the equations 2.3 and 2.4:

$$\frac{K}{PTT} = \sqrt{\frac{hE_0 e^{\gamma p}}{\rho d}} \tag{2.6}$$

BP can be estimated.

Hence, from the Equation 2.6, it can be assumed that PTT decreases with blood pressure for two reasons [65]:

- Arterial compliance decreases with increasing pressure due to the curvilinear relationship between arterial pressure and volume;
- Volume increases with increasing pressure (the artery dilates).

So, it is possible to trace back a reliable relationship between BP and PTT.

2.2.3 Arterial Blood Pressure

Blood Pressure (BP) is one of the vital signs used by physicians for the evaluation of patient's health, together with respiratory rate, heart rate, oxygen saturation, and body temperature and it reflects the status of the circulatory system [66].

During a cardiac cycle, the blood pressure increase due to the contraction of the heart to push blood into the aorta, then it decreases when the heart is filling up [67]. An increase or decrease of blood pressure may be associated with the change in activity status, emotional state or the patient's health status, working in synergy with cardiac output, vascular system resistance and arterial stiffness [68].
Blood pressure is measured in millimetres of mercury (mmHg) and can be divided into Venous (VBP) and Arterial Blood Pressure (ABP):

- Venous pressure is affected by the cardiac cycle as well as the respiratory cycle: during inspiration intrathoracic pressure becomes negative, instead it is positive during exhalation. The typical values of central venous pressure vary, in an healthy person, from -4 to +8 mmHg [69].

- Arterial pressure, also commonly referred to just as blood pressure, fluctuates according to the cardiac cycle between two values known as "maximum", called Systolic Blood Pressure (SBP), and "minimum", called Diastolic Blood Pressure (DBP). Values between these two peaks correspond to the cardiac systolic and diastolic phases, respectively. SBP represents the pressure in the vessel when the heart is contracting and lies between 100-140 mmHg and DBP is the pressure in the vessel when the heart is relaxing and it is between 60-90 mmHg in a healthy adult heart; the two phases are separated by a Dicrotic Notch (Figure 2.12) [70].

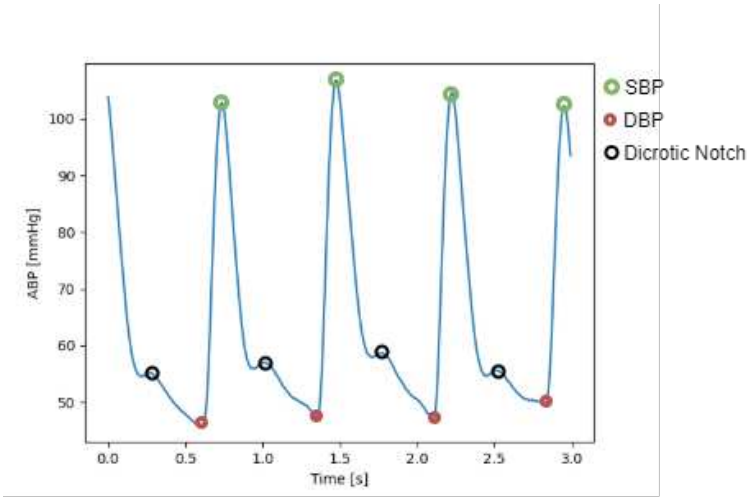


Figure 2.12: ABP waveform of an adult.

Since veins do not have sufficient pressure, blood pressure is measured in arteries, mostly using a sphygmomanometer [71].
The sphygmomanometer is a blood pressure cuff located around the arm following a specific procedure, as shown in Figure 2.13:

- the arm is relaxed and the blood flow laminar;
- the cuff is inflated to well above expected systolic pressure and cuts off circulation to a region (typically the brachial artery in the arm);
- then, the cuff pressure is gradually released. Once the pressure in the cuff becomes equal to the systolic pressure, the blood flow exceeds the cuff, creating a turbulent flow. This phenomenon is manifested by a pulse that can be audibly detected with a stethoscope: the Korotkoff sounds;
- the pressure continues to be released from the cuff until a pulse can no longer be audibly detected (diastolic pressure) [72].

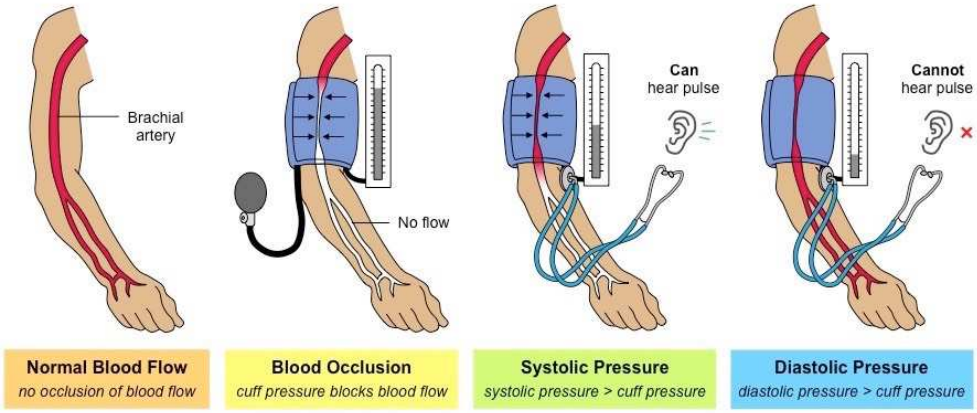


Figure 2.13: Steps of measuring ABP using a sphygmomanometer [11].

Blood pressure derived two main features: the pulse pressure (PP) and the mean arterial pressure (MAP) that allow the clinicians to have a view of hemodynamic of systemic arterial pressure of a patient. The first one is the difference between the amplitude of the systolic and diastolic pressure (Equation 2.7), and it represents the heartbeat and it is typically about 40 mmHg [66][67].

$$PP = SBP - DBP \tag{2.7}$$

The second one is the average of blood pressure over a cardiac cycle and depends on the cardiac output, systemic vascular resistance, and central venous pressure [72]. The MAP calculation, with some simplifications, can be faithfully represented by the following equation:

$$MAP = DBP - \frac{SBP - DBP}{3} \tag{2.8}$$

Typically clinicians evaluate as normal value MAP between 70 and 100 mmHg [73]. BP measurements undergo several fluctuations during the day and has physiologically lower values during sleep and in the early morning hours, but sometimes

these changes of BP could be linked to pathological condition [74]. The method previously described is non-invasive, and have to be done in presence of a clinician. Another method is called Ambulatory Blood Pressure Monitoring (ABPM) or Pressure Holter and provides various blood pressure measurements with different clinical implications such as daytime blood pressure during activities and nighttime blood pressure during sleep because it collects data in 24 or 48 hours. It is largely used for hypertension diagnosis [75].
Patient can also measure ABP in autonomy at home using a device that uses the same technology as ABPM monitors, but allows patients to monitor BP as often as they wish. It is called Home Blood Pressure Monitoring (HBPM) and it is a non-invasive method with a sphygmomanometer cuff and a electronic pressure sensor: the SBP and DBP values are visualized on a screen (Figure 2.14).
Another non-invasive device, arrived on the market just in 2019, is Omron HeartGuide and it is the first smartwatch that complies the guidelines for the ABP measurement of medical devices [76]. Its user interface is an App, that show on the screen of the smartwatch the results of the software processing: such as the measurements of BP and other vital signs like HR and the number of steps.
In addiction, for cases in which the patient need a continue and accurate monitoring of BP (e.g. in Intensive Care Unit) it can be used an invasive technique: Intra-arterial BP monitoring, that requires that BP is measured directly by inserting a catheter in an artery. This method is the best in terms of accuracy, but it has some disadvantages due to the applicability of the instrument due to its invasive nature [77]:

- It is not possible to take measurements for excessively long times.
- The patient is exposed during the collection of the signal to the risk of infection and other vascular damage.
- It can not be performed safely in obese patients or with other pathologies.

2.3 Cardiovascular diseases

CVDs have proven to be preponderant in current society, leading to the highest number of premature deaths per year [22].
Monitoring the parameters that allow the physicians to identify a critical situation since first stages is crucial. For the prevention of onset of CVD the approach is lifestyle changing: there is a high correlation between CVDs development and an unhealthy diet, physical inactivity, stress, tobacco use and harmful use of alcohol [22] whose effects may show up in individuals as: (i) raised blood pressure, (ii) raised blood glucose, (iii) raised blood lipids, (iv) overweight, (v) obesity. The main non-modifiable risk factors for cardiovascular disease are advanced age, family history of cardiovascular disease, gender and race. So, the primary prevention of

Figure 2.14: (a) Example of device using the oscillometric method for home measurements. (b) Omron HeartGuide [12].

CVDs includes actions related to improving the diet and increasing physical activity levels of the population, decreasing exposure to alcohol and smoking, and preventing the development of metabolic conditions such as obesity, type 2 diabetes, and hypertension (HT) [78]. Above these disorders, hypertension is the most important preventable risk factor according to WHO [79].

2.3.1 Hypertensive disease

The term hypertension refers to persistently high blood pressure in the arteries over time, in fact it is also called high blood pressure (HBP). HBP typically does not cause symptoms and for this reason the only way to diagnose the hypertension is to continuously monitor ABP[80]. Continuous monitoring is particularly important for identifying situations such as white-coat and masked hypertension.
White-coat hypertension [81] is manifested by a temporary rise in blood pressure above normal levels (from 4 to 75 mmHg in SBP values and from 1 to 36 mmHg in DBP values) that occurs only under certain circumstances, mainly linked to emotion. In most cases, this type of hypertension emerges when the patient undergoes a normal blood pressure control carried out by clinician (hence the name "white-coat"). The state of anxiety that this practice causes in some subjects can lead to tachycardia and an increase in pressure that is detected by the sphygmomanometer. In order to be certain of the nature of the clinical condition and thus distinguish it from hypertension, the subject undergoes blood pressure monitoring for a period of about 24 hours. It is therefore essential that during self-measurement in a familiar environment the subject has normal values. However, this condition should not be considered a "false alarm", as clinical evidence has shown that white coat hypertension increases the risk of developing stable hypertension [82].
Masked hypertension [83] is defined as an elevated BP outside the clinic (diurnal-ABPM or HBPM measurements over the day greater than 135/85 mmHg) but a

normal blood pressure in the clinic or office (greater than 140/90 mmHg). It is also called "inverse white-coat hypertension" [84]. It can occur in 10% of the population, and is important because it is not diagnosed by routine medical investigations. On the contrary, it carries an adverse prognosis, both in terms of increased cardiovascular events and target organ damage. The clinical practice leads to the suspect and diagnosis of masked hypertension in individuals who have a history of occasional high BP readings, but who are apparently normotensive when checked in the office.

Moreover, hypertension is very related to arteriosclerosis, that occurs when the blood vessels, that carry oxygen and nutrients from the heart to the rest of the body, become thick and stiff because of the fatty deposits [85] (Figure 2.15). The process become a vicious circle:

1. Plaques are formed at the level of the arterial walls with a consequent decrease in their elasticity and a decrease in their lumen diameter.
2. Blood pressure becomes higher.
3. The decrease in the lumen increases resistance to the passage of blood.
4. The atherosclerotic process is triggered.

The two phenomena amplify the effects of the other, progressively worsening the health conditions of the person suffering from hypertension, accelerating the process that leads to heart failure.

Hypertension has to be monitored, for avoiding the development of new, and worse, complications. Moreover, monitoring of the blood pressure levels is usually a routine for elderly citizens. However, for many adults, blood pressure measurement is not taken regularly. Yet, early diagnosis will help them get rid of their cardiovascular risk for heart disease later in life.
The American Heart Association recommends ABPM to measure Blood Pressure and, for the evaluation of the measured values a looking up table (Table 2.1) [86]). It is used, in clinical practice, for detect the presence and eventually the stage of hypertension. As shown the state of hypertension can be divided into 3 different scenarios (or stages) [80]:

- Stage 1: it covers the most cases (about 90%) and it is typically caused by genetic or lifestyle factors;
- Stage 2: it is related to a specific disorder, such as renovascular hypertension, associated with kidney dysfunction due to inappropriate secretion of the endocrine system [87]. In these cases, identification and removal of the causes (i.e., treatment of the underlying disease) may be accompanied by normalization of blood pressure values [88];

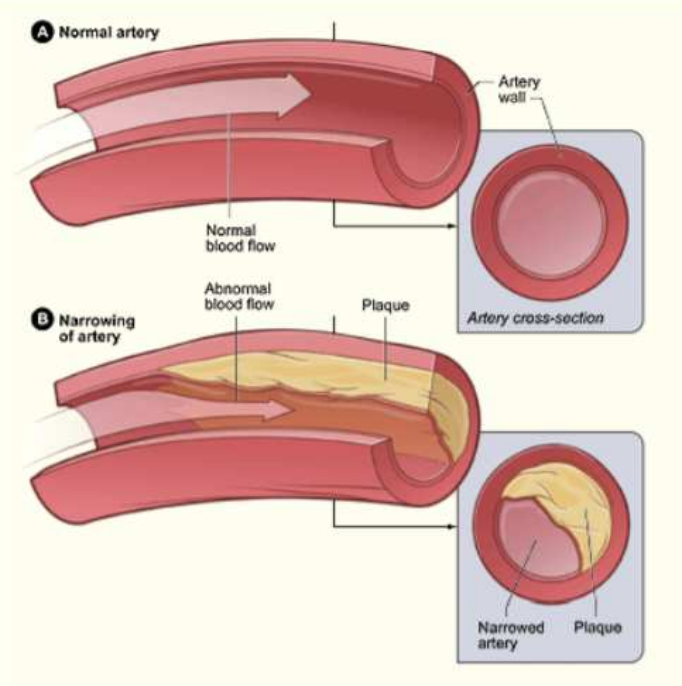


Figure 2.15: Schematic representation of normal artery (A) and the narrowing effect due to plaque (B) [13].

BP Category	**SBP** [mmHg]		**DBP** [mmHg]
Normal	< 120	and	< 80
Elevated	$120 - 129$	and	< 80
Hypertension			
Stage 1	$130 - 139$	or	$80 - 89$
Stage 2	≥ 140	or	≥ 90
Hypertensive crisis	> 180	and/or	> 120

Table 2.1: Classification of BP values into categories according to ANSI/AAMI.

- Hypertensive crisis: is an emergency situation that requires urgent medical care. Sometimes it can be associated with chest pain, vision problems, numbness or weakness, breathing difficulty, or other signs and symptoms of a stroke or heart attack.

The US Association for the Advancement of Medical Instrumentation (AAMI), the British Hypertension Society, the European Society of Hypertension (ESH), International Organization for Standardization (ISO) and the Working Group on Blood Pressure Monitoring agreed to develop a universal standard for BP measuring device validation [89]. The outcome is that the measuring error is considered

tolerable if it is below 10 mmHg and a device complies it if the estimated probability that the error (evaluated with reference to a gold standard) of measurement is equal or lower than 10 mmHg on at least 85% of samples.

2.4 Introduction to Machine learning

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that gives calculators the ability to learn to perform a specific task without being explicitly programmed to do so, that consist into the development of computational models of learning. [90].
The objective of Machine Learning algorithms is to associate every x_i item (input) to an y_i (output), such that $y_i = f(x)$, where f is the model. The model is built from the first observation and updated to obtain the best possible representation, based on whole available dataset. In fact, from a computational point of view, machine learning refers to the ability of a machine to improve its performance based on previous results.
ML is necessary in cases when: (i) human expertise does not exist; (ii) human are unable to explain their expertise; (iii) human expertise exists but is unreliable (e.g. results may be affected by subjectivity); (iv) human expertise exists, but it is unworkable (e.g. too many data or too expensive); (v) solution may change in time; (vi) solution may need to be adapted to particular cases. In these situations, machine learning is a smart solution because it allows the problem to be solved in a less expensive and automatic way.
Three main phases are needed for design a Machine Learning algorithm [91]:

- Problem Conceptualization. Starting from the problem it is necessary to extract information about the availability of the data, the uncertainties of the data themselves (measurement errors, personal judgements from borderline situations) and the manual approaches used.
- Construction Process. It includes:
  - Dataset construction, that starts from the available data.
  - Analysis of the data, for identifying the main characteristics (such as the type of the input data, the membership range, etc.). At the end of the data analysis, two sub-sets of data are identified: the training set, which will be used to train the algorithm, and the test set, which will be used in the tuning phase to make the model interact with an unknown data set.
  - Identification of a possible model for the problem: there are different models that can be used, such as the Gaussian model.
  - Selection of the method, among several that can be used for the predicted model, an example in the case of a Gaussian model a Support Vector Machine with Gaussian kernel can be used.

  - System tuning, the control parameters of the method are changed, in a controlled way, in an attempt to obtain a model that gives good results and is robust.

  At the end a model is given as output.

- Validation Process: that include verification (check that the model meets the specifications) and validation (checking that the product satisfies the user).

The steps are not rigid, in fact it is possible to go back to the conceptualisation of the problem if the next steps are not passed (e.g. if the tuning of the model parameters does not give acceptable results or if the verification phase is not passed).
At first glance, ML algorithms are distinguished in three families:

- Supervised learning: the target to which the training set data belongs is known. It can be used to do:
  - Regression: the goal is to model the relationship that exists between a continuous dependent variable (the target) and a set of independent variables;
  - Classification: the goal is to identify the class (the target) to which a new item belongs based on previous observations (Figure 2.16). For classification aim the training phase consist in the identification of the decision boundary, i.e. the separation space between the classes, which in the n-dimensional space of the model looks like an n-dimensional curve.
- Unsupervised learning: recognize patterns in data where I don't know the class they belong to. It is used to do clustering, that means to divide my training set into subgroups (clusters) so that the elements of each cluster are very similar to each other and elements of different clusters are very different from each other (Figure 2.16). It is also used for the dimensionality reduction, that is applied to a dataset for the extraction of the smaller set of variable to retain the informative value of the dataset and allow an easier representation and increase the computational speed.
- Reinforced learning: learning is based on feedback derived from the external environment (typically used in video games).

Outputs can be different, and diverse methods use diverse output type, as can be seen in the Figure 2.17. Discrete outputs include categorical, codified or specific numbers; continuous outputs are numbers in a non-specific interval of values.

The application of ML to real-world problems and the efficiency of the results produced has made these algorithms attractive for biomedical applications. Today, ML in medicine includes (i) applications to support clinical decision-making, (ii) diagnosis, (iii) monitoring, (iv) patient treatment. In particular can be outlined four macro areas of application:

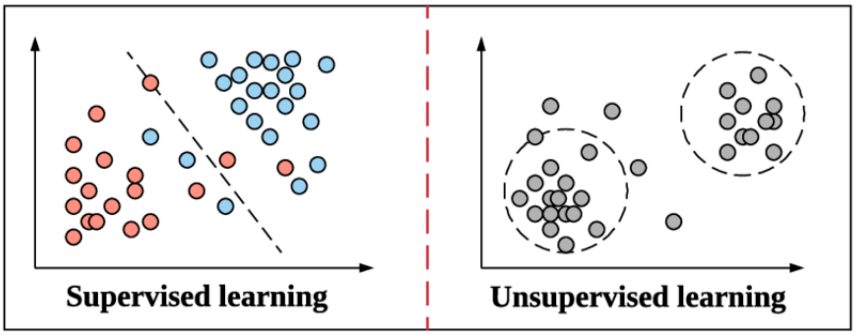


Figure 2.16: Example of supervised learning as Classification: data are mapped into a 2-dimensional space and they belong to 2 different classes, that are identified by blue and red dots. In this example, the classification algorithm identifies a decision boundary for that data that include some error of prediction (blue dots in red region and red ones in blue region) for data modelling. Example of Unsupervised Learning as Clustering: the data mapping in the space identify two regions in which can be identified low distance among data in it, and pretty separated [14].

	Discrete	*Continuous*
Supervised Learning	Classification	Regression
Unsupervised Learning	Clustering	Dimensionality reduction

Figure 2.17: Types of output in ML methods.

- Prediction of disease onset: e.g. screening tests or estimating the probability of success of a treatment before starting it (e.g. in the case of cancer) [92][93];
- Prediction of possible adverse events, complications [94];
- Assisting the clinician during the treatment of the patient, e.g. CAD (computerized systems that process images and then highlight any areas that are more likely to become cancerous areas) [95];
- Personalized medicine: understanding what are the mechanisms of a given pathology (not yet used in the clinic, only research) and then identify the best treatment for that pathology on the specific patient [96].

Current algorithms are able to perform many problems much faster than humans,

but they should not replace the human being but act as a support.

Chapter 3

Materials and methods

3.1 Overview of the state of Art of ABP estimation

Such as explained in the Chapter 2, BP can be measured by several devices, each one thought for a specific application and the need of a specific accuracy of measurement in order to satisfy the application needs. Summarizing, the BP can be obtained by invasive methods or non-invasive methods.
The invasive methods are the first methods used in medicine [71]. It was in the eighteenth century when Stephen Hales (1677-1751) inserted a tube, opened to the end, in the carotid artery of a horse and measuring the height reached by the horse's blood in the tube, and, just exploiting the Bernoulli's equation, the blood pressure was obtained [97]. After three thousand years, invasive methods have reached high sensibility, and they are the gold standard of BP measurements [89]. The pros of using invasive method are the higher accuracy and the possibility of beat-to-beat measuring of the pressure values. Consequently, it is deployed for investigating critical fluctuations of pressure or for monitoring pressure during the day. Rather, the cons are related to the measuring system, in fact devices could be uncomfortable for the patient, and also it could lead to infections.
On the other hand, non-invasive methods involve fewer complications for the patient and include clinical devices (ABPM and HBPM) and wearable devices. Comparatively, while ABPM provides BP information at many timepoints on a particular day during unrestricted routine daily activities, HBPM provides BP information obtained under fixed times and conditions over a long period. Thus, HBPM gives stable readings with high reproducibility and has been shown to be as reliable as ABPM [98]. Figure 3.1 shows a table that summarized the main advantages and limitation of using HBPM instead of ABPM. It can be noticed that the use of HBPM has led to an empowerment of the patient and has allowed the involvement of more modern technologies, such as telemedicine. On the other hand, self-measurement

has disadvantages if not performed correctly. Wearable devices entered this scenario

Advantages	Limitations
• Can take multiple readings over an extended period of time • Avoids white-coat reaction to BP measurement • Reproducible • Predicts CV morbidity and mortality better than office BP • Can diagnose white-coat and masked hypertension • Allows patients to better understand hypertension management • Telemonitoring allows remote monitoring by healthcare professionals • Detects increased BP variability	• Some devices have been found to be inaccurate • Cuff placement can affect accuracy • May induce anxiety and excessive monitoring • Risk of treatment change by patients based on casual home measurements without doctors' guidance • Lack of nocturnal recording • Not yet reimbursed by insurance companies in many countries

Figure 3.1: Advantages and Limitations of Home Blood Pressure Monitoring [15][16]. CV = cardiovascular.

in 2019 and right from the start they have proven to be devices that best suits the needs of the market. Advantages are in terms of ease of use, calculation capacity and minimal invasiveness [3]. The capabilities that make them attractive are the possibility to connect to the internet (that allows IoT) and the high calculation levels, that widen their use during years in application as remote home monitoring, especially for chronic diseases [99]. Despite this, they incorporate the use of a cuff in their working principles, and it means limitations:

- The size of the cuff [100]: it must not be too small, because it led to overestimation of BP. At the same time this limitation renders the reduction of size, cost and power consumption in these devices very difficult.
- Limited recording [101]: they are not preferred to be used frequently. It is because the cuff led to the occlusion of the artery and it increases the workload of the heart and causes circulatory interference at the measurement site.

The overcoming of these limitations led to the design of the cuffless methods, that base the BP estimation on other physiological parameters. Research in this direction implies the need to make the ABP measurement's system more portable and

to contribute to early disease prediction and intervention [102].
Finally, for the hypertensive propose, compared to the cuff-based approaches, a cuffless technique is more desirable for the evaluation of pressure values in the prognosis, diagnosis and treatment of the pathologies associated to it. This thesis work aims to study and to develop novel approaches for the cuffless estimation of BP for SINTEC device applied to hypertensive patients [2].

3.1.1 State of Art of cuffless methods for BP estimation

A fundamental feature of our cardiovascular system is its complex dynamic self-regulation involving multiple feedback control loops in response to BP variation (i.e., pulse pressure, heart rate, and arterial stiffness) [103]. It is possible to exploit this feature to investigate BP estimation indirectly, i.e. cuffless. Anyway indirect measurement of blood pressure is a very recent tool.
Mathematically, a turning point was defined by studies on the stiffness of arteries and other mechanical properties of vessels on the self-regulating mechanism of blood flow.
It has been demonstrated that the velocity of a longitudinal pressure wave is related to the elasticity of the arterial vessel and the vessel size by the Moens and Korteweg equation (Equation 2.4) and the elastic module of the vessel wall as a function of the blood pressure described by Geddes (Equation 2.5).
In literature early studies reported involve PTT and in particular its derivative PWT for BP prediction. In 2000, Chen et al. carried out a study on 20 patients during a cardiovascular surgery using an electrocardiograph and an oximetry sensor. They measured, in a synchronous way and beat by beat, three different signals (i.e. ECG, PPG, ABP), obtaining results that showed correlation coefficients value between SBP and PTT higher than 0.96 [104]. Early approaches to BP estimation using PTT employed pattern recognition techniques and machine learning techniques such as Support Vector Machine (SVM) or K-Nearest Neighbors (KNN). However, they required continuous calibration on the individual patient because of model instability, due to the wrong assumption that the arterial stiffness is constant in time [105].
In subsequent years, the goal of the studies was to find a faithful approximation for obtaining SBP using PTT, and to test its reliability for determining directly SBP and DBP using a nonlinear algorithm and a more stable calibration model. Three main models have been proposed: all of them start from the Equation 2.6 and, then, define an initial calibration and do some approximation. In fact, in these models, the h/d value, the distance K and the elasticity parameter E_0 are considered constant. So, it has been possible to estimate BP values through different formulations:

- The linear model: Mccarthy et al. (2011) used to calibrate the linear algorithm with Omron M6. Anyway, just doing every 6 minutes the calibration process

it is in accordance with the AAMI guidelines [106].

$$BP = a * PTT + b \tag{3.1}$$

- The logarithmic model: proposed by Hughes and used by Soerensen et al. (2017) adopting specific parameters evaluated with least squares method (a and b). However, this study showed a tendency towards PTT and weak and inconsistent blood pressure [107].

$$BP = a * \log(PTT) + b \tag{3.2}$$

- The inverse linear model: used by Shrimanti et al. (2016) [108] for the estimation of BP evaluating different position (recumbent, seated, standing, walking, cycling) [108]. The results, in this application, are shown to be reasonably accurate during rest but not during movement.

$$BP = \frac{a}{PTT} \tag{3.3}$$

More recent studies related to BP estimation are reported below.
In 2019, Lazzazzera introduces in the market a new smartwatch called CareUp ®. This smartwatch is able to provide blood pressure measures in real-time using two photoplethysmograms (PPG) [109].A PPG waveform is acquired by placing the index finger of the hand without the watch on the oximeter sensor and the second is taken from the rear front sensor of the watch in contact with the skin of the wrist (Figure 3.2). After signal acquisition and processing of PPG waveform they have

Figure 3.2: How to use CareUp®for the PPG collecting [17].

extracted HR and PTT values that are used as input in a linear model of the kind shown in Equation 3.4.

$$BP = a * PTT + b * HR + c \tag{3.4}$$

Where a,b and c are the patient-specific parameters. Prediction error results almost in agreement with the American Association for the advancement of the medical instrumentation standard [17]. Another recent work has been conducted by Also Zhang et al. (2019) [110]. The dataset includes PTT, HR, PPG, I, II, III, aVF, aVR, aVL, SpO2 for the BP prediction. They have used a non-linear model: Support Vector Regression (SVR) with radial basis function (RBF). Results were acceptable by the global standards set by the AAMI for cuffless blood pressure estimation.

3.2 Machine learning regression techniques

The Regression [111] is a statistical process that allows us to estimate relationships between variables. It determines the functional link between the independent variables (x vectors) and the dependent variables (y vector or target). As mentioned, in regression the target values (y) are numerical and continuous (in $\mathbb{R}$), and the goal of training is to learn a function f, such that $f(x) \rightarrow y$ and it is assumed that the independent variable is exact while the dependent variable is affected by error (e.g. measurement inaccuracy).
The easiest method for regression is the simple linear regression [112]. It is a model that establishes a linear predictor function between the dependent (Y) and the independent (X) variables. There are many ways to find the best regression line that fits the data distribution, and they are based on the minimization of the distance. The formula for simple linear regression is (Equation 3.5):

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{3.5}$$

where α and β are the parameter of the curve. The objective of the regression is to find the best $\hat{\alpha}$ and $\hat{\beta}$ for the fitting of the data that minimize the error term Q, i.e. the sum of the squared residuals ϵ_i [111] (Equation 3.6).

$$Q(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^{n} \hat{\epsilon}^2 \tag{3.6}$$

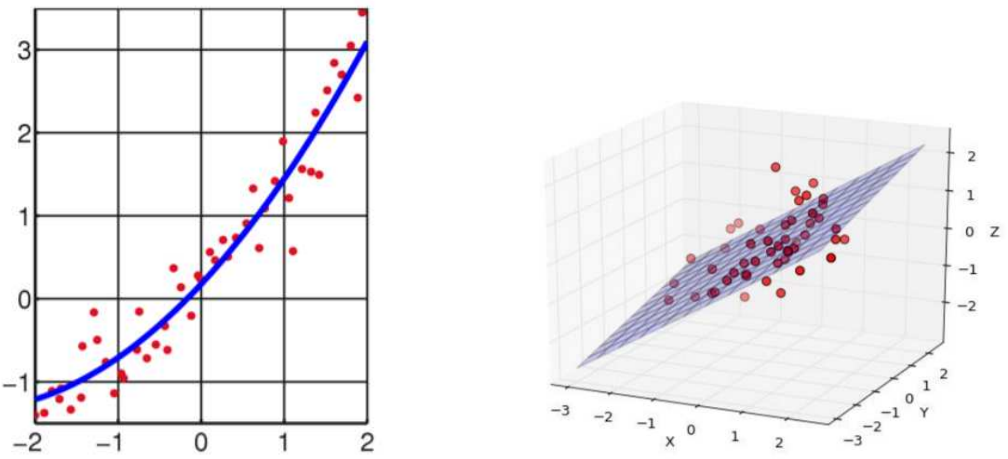
Metrics for the residual calculation (ϵ) are: (i) the least squares, (ii) the least absolute deviation, (ii) the least square cost function (ridge regression), (iii) the lasso (or L1-norm). Simple linear regression needs assumptions such as:

- Linearity: the relationship between the dependent and the independent variables is linear;

- Homoscedasticity: the data around the regression line have the same variance for all values of X;
- The errors of prediction are distributed normally: the distributions of deviations from the regression line are normally distributed [112].

For complex applications, where simple linear regression is not enough for modelling the data, there are other types data representations in the space. The goal is to have the best fitting to the data and the best regression results:

- **linear/non linear**: the f function can be a first grade curve (e.g. regression line) or higher than one curve (e.g. quadratic regression example in Figure 3.3(a));
- **simple/multiple**: the x_i vector can be described by one element or more, if the vector has higher dimensionality it is called multiple (e.g. regression hyper-plane for two-dimensions vector example in Figure 3.3(b)).



(a) (b)

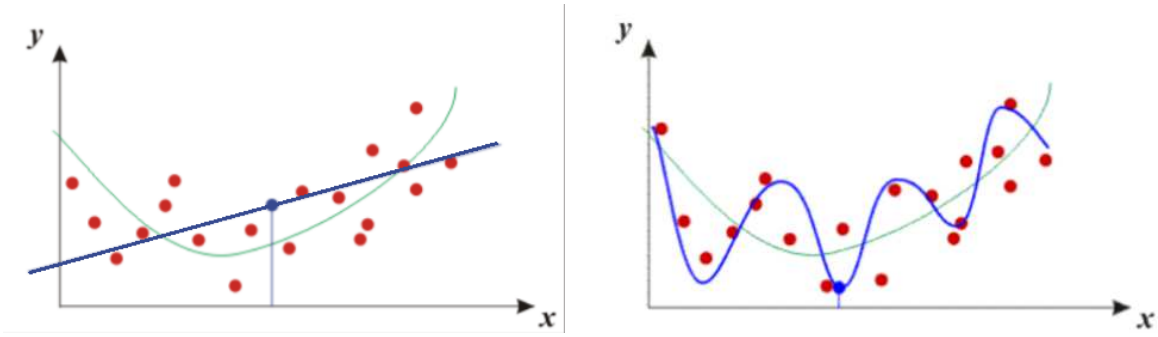
Figure 3.3: (a) Quadratic regression. (b) Two-dimensional regression [18].

Furthermore, as it can be seen from the Figure 3.4, the curve representing the regression model never assumes the true shape of the data. In fact, approximations are essential to avoid creating models that do not take into account limitation of the data used for construction of the algorithm. Two concepts are useful when designing a regression algorithm:

- Bias: error due to inaccurate assumption or simplification made by the model. It is a measure of how much does the average model over all training sets from the true model;

- Variance: metric for evaluating the dependence of the model from the input data. It is a measure of how much models estimated from different training sets differ from each other.

When a model has high bias and low variance means that there are few parameters for the modelling: this is underfitting (Figure 3.4(a)), instead a low bias and a high variance indicate overfitting (Figure 3.4(b)) and it happens when the model is too accurate just for the training data. There is not a model that works best for every problem but when a regression technique is applied it is important to make a variance and bias trade-off.



(a) Example of underfitting model

(b) Example of overfitting model.

Figure 3.4: Two different regression model applied to the same data.

The best combination of these two parameters allows generalization, i.e. the ability of the algorithm to give accurate output on new data.

3.3 MIMIC III database and recordings selection

The construction of a significant database of patients with hypertension should consist in physiological signals (such as ECG, PPG and ABP) and other biographical information such as age and sex.
Currently, it has not yet been possible to collect this type of information, since SINTEC has not yet built a prototype of the product. Therefore, the data used in this thesis were collected from the database MIMIC III, which provides publicly available data on physiological signals recorded at the same time, of over sixty thousand patients. Data came from patients who stayed in intensive care units (ICU) of the Beth Israel Deaconess Medical Center between 2001 and 2012. A subset of these recordings includes physiological waveforms obtained from bedside monitors, such as electrocardiograms, photoplethysmograms and blood pressure waveforms, which are of interest to this study [113].
Signals's features:

- Recordings are about 60s long;
- The sampling frequency is 125Hz.

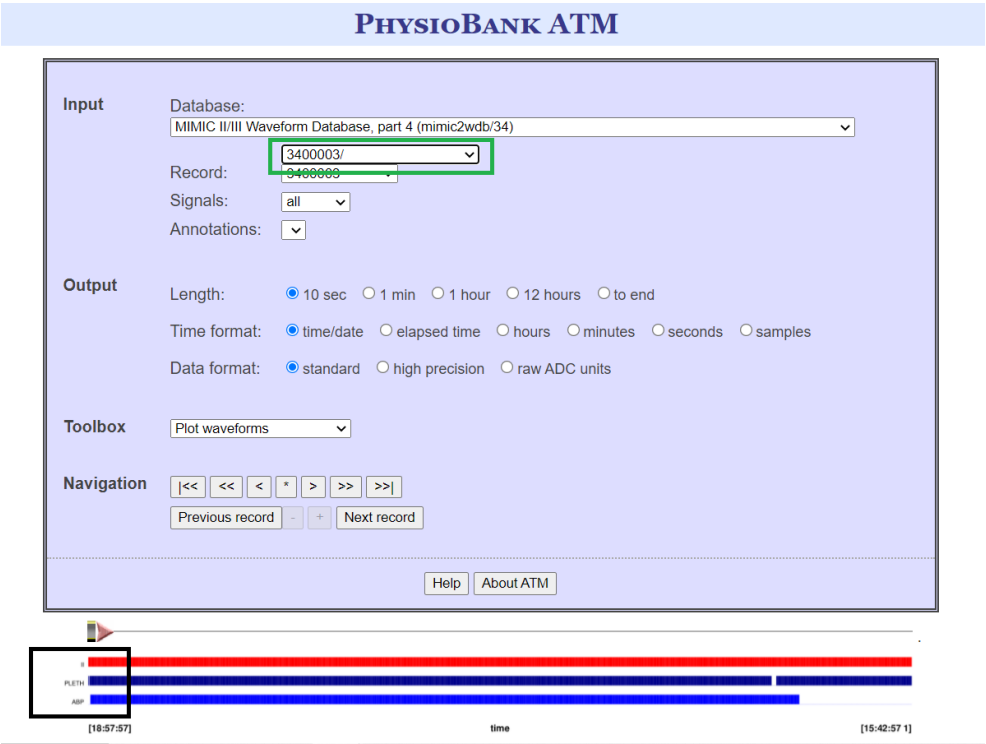


Figure 3.5: Example of MIMIC III database access page: PhysioBank ATM. In the image the green box indicate the position of the patient ID and the black box the signals present in the recording [19].

The steps of the recording selection from the online database have been the follow:

1. Selection of the patient, choosing by the ones presents in the drop-down window in the green box in the Figure 3.5;
2. Graphical observation of waveforms in the recording (Figure 3.6);
3. Check the presence of ABP, ECG and PPG signals, as shown in the black box in the Figure 3.5, it can be noticed that it has been used the second derivative of ECG ("II") and PPG is indicated as "PLETH";
4. Download the file in .csv format, an example is shown below.

```
'Time','II','PLETH','ABP'
'hh:mm:ss.mmm','mV','NU','mmHg'
'[12:33:05.136]',-0.062,0.573,63.281
```

```
'[12:33:05.144]',-0.094,0.592,61.719
'[12:33:05.152]',-0.094,0.608,60.938
...
```

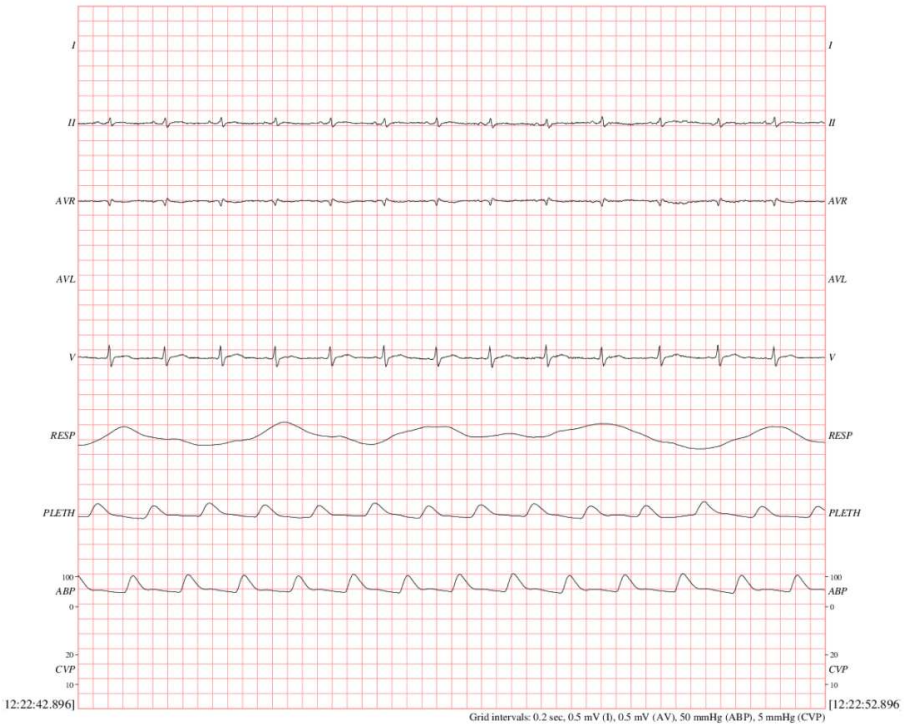


Figure 3.6: Signals of patient 3000714 visualized from MIMIC III.

However, the data collected for MIMIC III have not the same purpose of this work and some recordings have to be dropped after the download for different reasons:

- The absence of continuous recording: sometimes files were empty, and other times one or more signal had no values or not enough (more than 5 seconds of recording missing) (Figure 3.7).
- The waveform morphology is noised: the plotted waveforms underline some anomalies. It has been reported in Figure 3.9 an ideal PPG waveform compared with the cases found that were not suitable for our application: i.e. absence of the diastolic peak, presence of more than two peaks or systolic wave phase too wide. Sometimes, it has an abnormal morphology that can be traced back to sensor-off or sensor movements/misplacing.
- Presence of significant changes of ABP values, that was due to drug intake (the records are collected on ICU patients) (Figure 3.8).

At the end of this process 139 recordings have been selected.

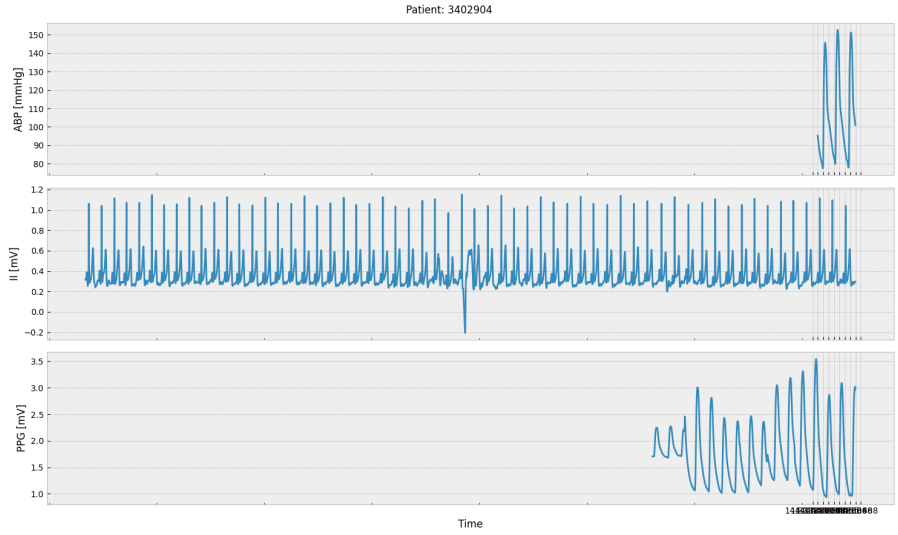


Figure 3.7: Example of missing values in the recordings.

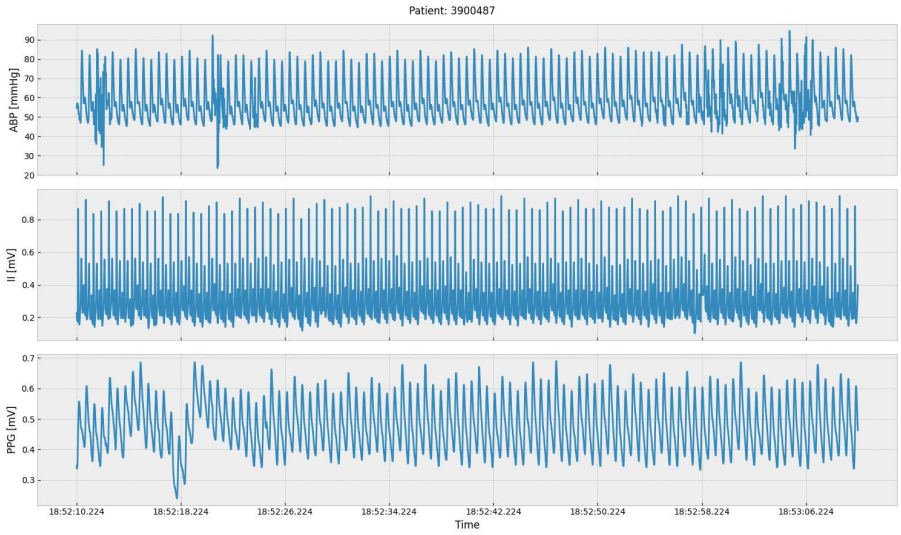


Figure 3.8: Example of abnormal fluctuations of BP values.

3.4 Workflow of implemented methods

In the Figure 3.10 are traced the main steps of the algorithm. Only ECG, PPG and ABP signals are saved from each subject file, as it may contain other signals besides those useful for our application, which has been discarded. The second phase focuses on the signal filtering. Then, the filtered signals are submitted to the peak identification phase, that is followed by the feature extraction from the obtained data. The dataset is constructed and it is ready for the training of different

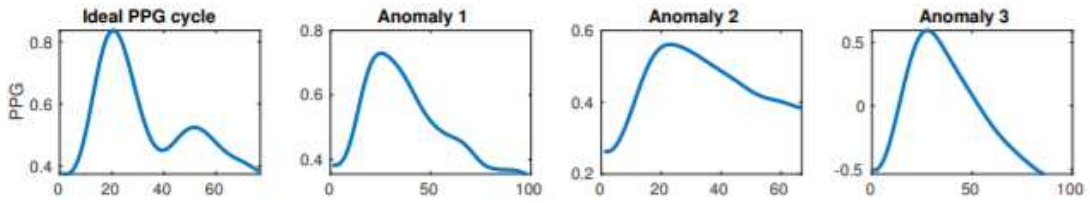


Figure 3.9: Example of PPG morphological anomalies [20].

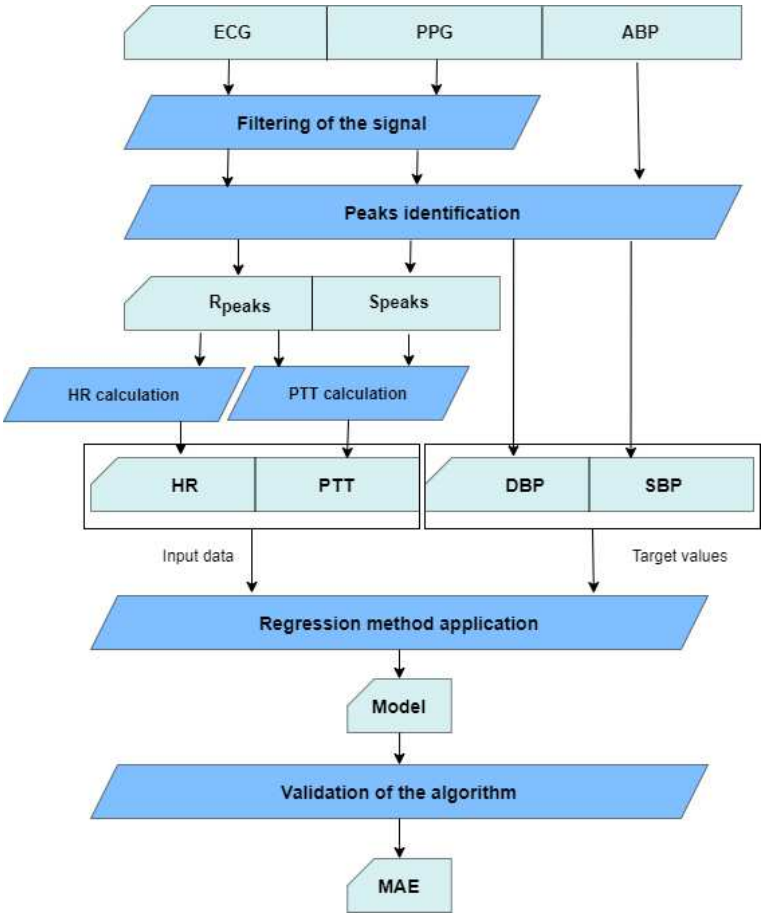


Figure 3.10: Scheme of the main steps of the work.

regression models. Finally, the system is validated to verify its capacity to generalize not reaching overfitting. Results are evaluated through the Mean Absolute Error (MAE).

3.5 Preprocessing

ECG and PPG signals are processed for obtaining signals with higher quality, removing background noise and artifacts. The ABP signal, instead, does not need this operation because the signal was collected for MIMIC invasiveness, which limits the presence of artifacts.
The ECG signal suffer of the following interferences:

- Baseline drifting or baseline wandering: it is due to improper electrodes (electrode-skin impedance) and patient's movement and breathing. Graphically it led the entire signal to shift from its normal base (x-axis). It occurs under 0.5 Hz, so it is easily removable with a high pass filter.

- Electromyography (EMG) Noise: it is an effect of muscle contraction under the electrodes. It affects the signal morphology when the subject is in moving. The spectral content of muscle activity overlaps the PQRST complex, so It can not be easily removed. Since the ECG is a repetitive signal, ensemble averaging can be used to reduce muscle noise. Anyway, it is not applied in our application (ICU patients).

- Powerline Interference: it is recognised by 50 or 60 Hz sinusoidal interference, possibly accompanied by several harmonics. To remove powerline interference from ECG signals it can be used a low pass filter with 50Hz cutoff frequency.

- Electrode Motion Artifacts: motion modifies the contact impedance of the electrodes that lead to anomalous manifestations on the signal. Since the interference band concentrates around 1 Hz - 10 Hz this interference can not be removed.

Therefore, ideal filtering would maintain frequencies between 0.5 Hz and 50 Hz.
The PPG signal has its main spectrum between 0.5 Hz and 4 Hz [63]. The recording system introduces different types of noise [114]:

- Baseline drift: small repetitive movements of the subject, like respiration, led to the signal drift. Typically, it occurs around 0.4 Hz and it can be removed through a high pass filter.

- Powerline Interference: it corrupts the signal at a frequency of 50 Hz or 60 Hz, so it can be easily eliminated by filtering it with a low pass filter.

- Movement artifact: due to the easiness of the movement of the PPG sensor because of its position, typically on the wrist or finger. The spectrum of this artifact is concentrated between 0.05 Hz and 8 Hz and for this reason it is difficult to remove without altering the information of the signal.

The best way for the signal filtering is having two separated filters, in order to maintain the target band and remove the noisy components of the specific signal. However, it has been considered that parameters are time-based, so it is crucial to maintain the time positioning of the peaks. As known from digital filter theory, FIR filters, i.e. with finite impulse response, need higher order filters [115]. On the other hand, infinite impulse response (IIR) filters introduce a phase delay, but, for this application, it is essential to preserve the morphology component in time of the signal and at the same time maintain coherence between the two signals. If the phase of the harmonics has changed, the spectral content remains the same, but the morphology of the signal in the time domain is altered [116].
As a consequence of these observations, the best strategy is to apply the same filter to the PPG and ECG signal, which is of IIR type but applied with the double-pass filtering technique, which allows us to obtain already the phase rotation compensated filtered signal.
With the conservative method it is necessary to design a filter that would allow us to not attenuate the frequencies between 0.5 Hz and 50Hz (minimum and maximum of the two signals), and do the opposite with the others. Filters have been implemented using the Python function `scipy.signal.butter()` and some of the implemented filter transfer functions are reported in the Bode diagrams in Figure 3.11. Among all, the frequency response given by the filter Figure 3.11(a) is the most suitable, as it allowed to sufficiently attenuate the external band and preserve the internal one in the previously established band (0.5 H - 50Hz, represented in the figures by the red lines). So, the chosen designed filter characteristics are:

- Order: 5;
- Band-pass with cut-off frequencies: 1Hz and 10Hz;
- IIR.

In this way, the need to use the same filter for both signals has been satisfied in order not to introduce different delays. At this point, however, it has been built an IIR filter, which could create problems in the evaluation of the morphology of the signals themselves. So the next step was to implement a double pass filter, using the Python function `scipy.signal.filtfilt()` when filtering. An example of the results of the pre-processing phase is reported in Figure 3.12. As it can be seen, both signals lose their off-set component. Furthermore, the PPG signal shows a greater stability in the filtered waveform.

3.5.1 Features Extraction

The next step is to identify diastolic and systolic peaks from the blood pressure signal, R peaks from the ECG signal, and systolic peaks from the PPG signal. In fact, in this application it is better to use as features for prediction HR and PTT

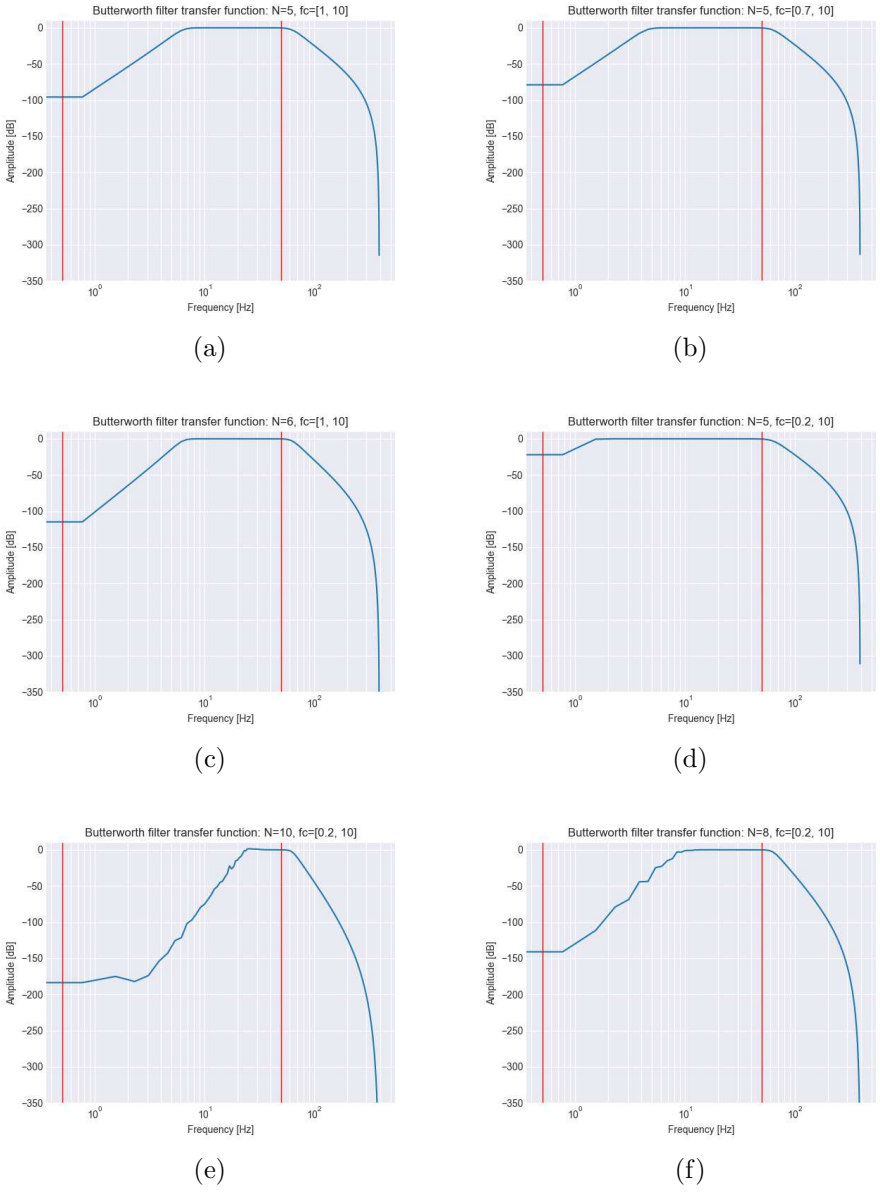


Figure 3.11: Bode diagrams of different designed filters. N: is the order of the filter, fc: is the cut-off frequencies given to the python function. Red vertical lines represent 0.5Hz and 50 Hz.

only. This is because it is not yet possible, since the prototype of the device is not available, determine the degree of accuracy will have the signal taken from the SINTEC sensors. The most sensitive points, however, are the peaks of the signals, which will certainly be recognizable. Future applications, once the prototype is

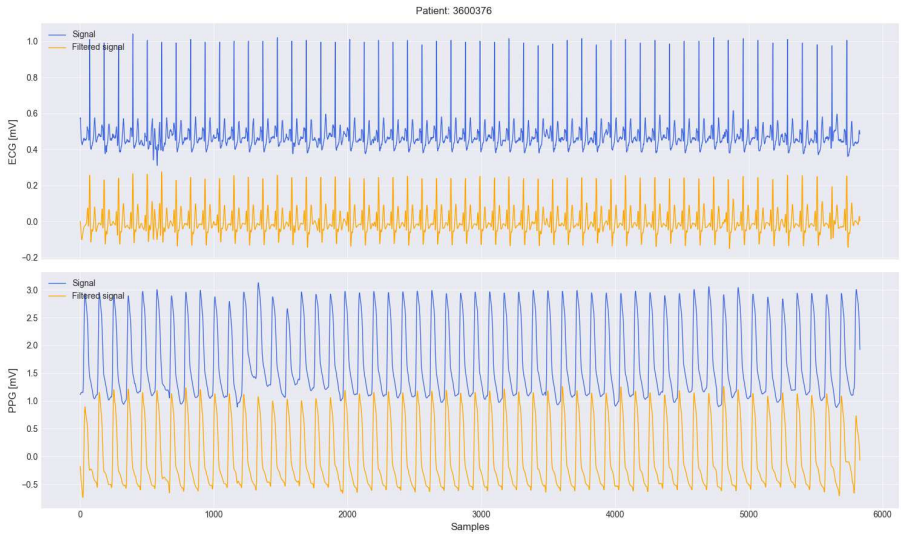


Figure 3.12: Example of a patient ECG and PPG pre-processed signals (blue) and post-processed signals (orange).

ready, will allow to add other parameters such as the MAP and the PP, previously described.
So, once the points of interest have been identified, it will be possible to calculate the features to be extracted from the signal (HR, PTT, SBP and DBP).
For both ABP and ECG, peak identification was not difficult. In fact, since both signals are characterized by a significant prominence and reduced width cycles, using the python function `scipy.signal.find_peaks()`, it has been possible to obtain, with a high level of accuracy, the peaks present in the recording. However, it has been necessary to take some precautions in order to obtain reliable results. In fact, this function takes a finite window of the signal and finds all local maxima by simple comparison of neighbourhood values.
For the systolic peak, for example, a peak width of minimum 10 samples has been defined, to avoid identifying as a *peak* noisy values. Then, a prominence equal to 0.5 and a minimum distance, in samples, between a peak and another equal to 60 has been set: these parameters allow the algorithm not to stop at a local maximum, which can be identified for example in the point of the dicrotic notch or in points of signal alterations due to residual noise.
At the end the results show a good algorithm performance (Figure 3.13).

PPG signal needs more processing since each signal of the MIMIC III database has a different amplitude (probably due to the use of different sensors). Moreover, the waveform has a larger amplitude between the maximum peak and the diastolic minimum, as was shown in the previous chapter. This implies that the function

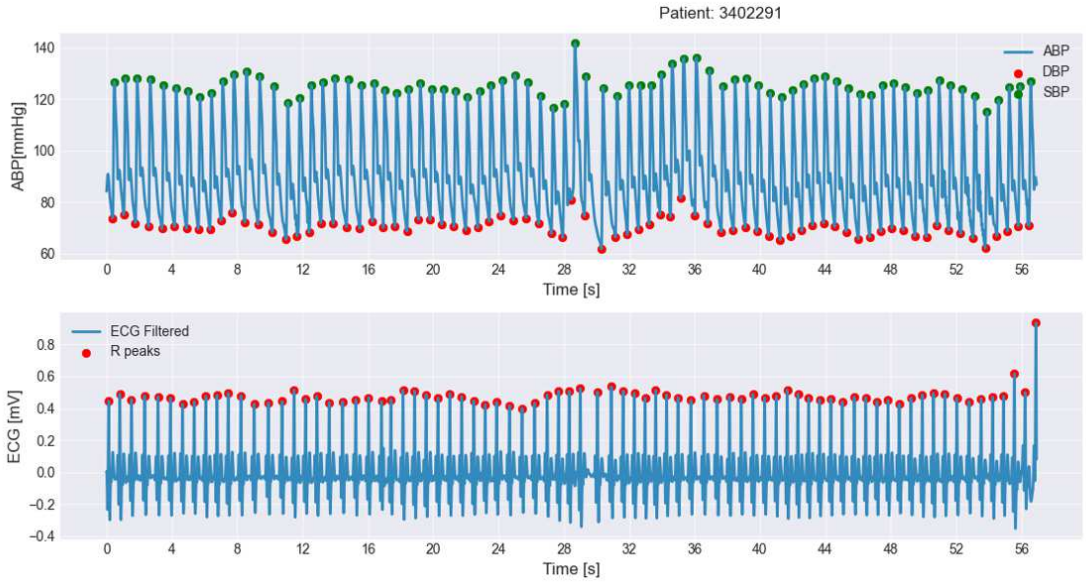


Figure 3.13: Example of peak recognition phase for ABP and ECG.

on its own, while using parameters appropriate to the signal, tends to detect too many maxima, but by increasing the specificity to the peaks, it would become too tailored to only a few recordings. A statistical analysis was therefore conducted on the distribution of the signal and peaks. Steps for the systolic peaks on PPG has been the following:

1. Identification of systolic peaks in PPG filtered (Figure 3.14(a));
2. Computation of the histogram of these peaks with a Gaussian kernel (Figure 3.14(b));
3. Study of the Gaussian distribution peaks with a Kernel Density estimation (KDE) and identification of the threshold (Figure 3.14(c)).

As it can be noticed from the third image, the threshold for the peaks identification is based on the first minimum value of the Gaussian distribution calculated.
However, it has been observed that there are exceptions, i.e. some KDEs present a peak earlier than the main one (Figure 3.15). For this reason, to make the algorithm more generic, a check was carried out on it. After the identification of the KDE minima of the PPG peaks, the algorithm looks if the first highest amplitude peak is less than only 90% of the highest peak in the signal. If the closeness condition is met, the second minimum is the one that is identified as the selection threshold. This behaviour has been found in cases where the signal has not a higher variability, but there are a few numbers of higher magnitude peaks, such as the one shown in the example in Figure 3.15.

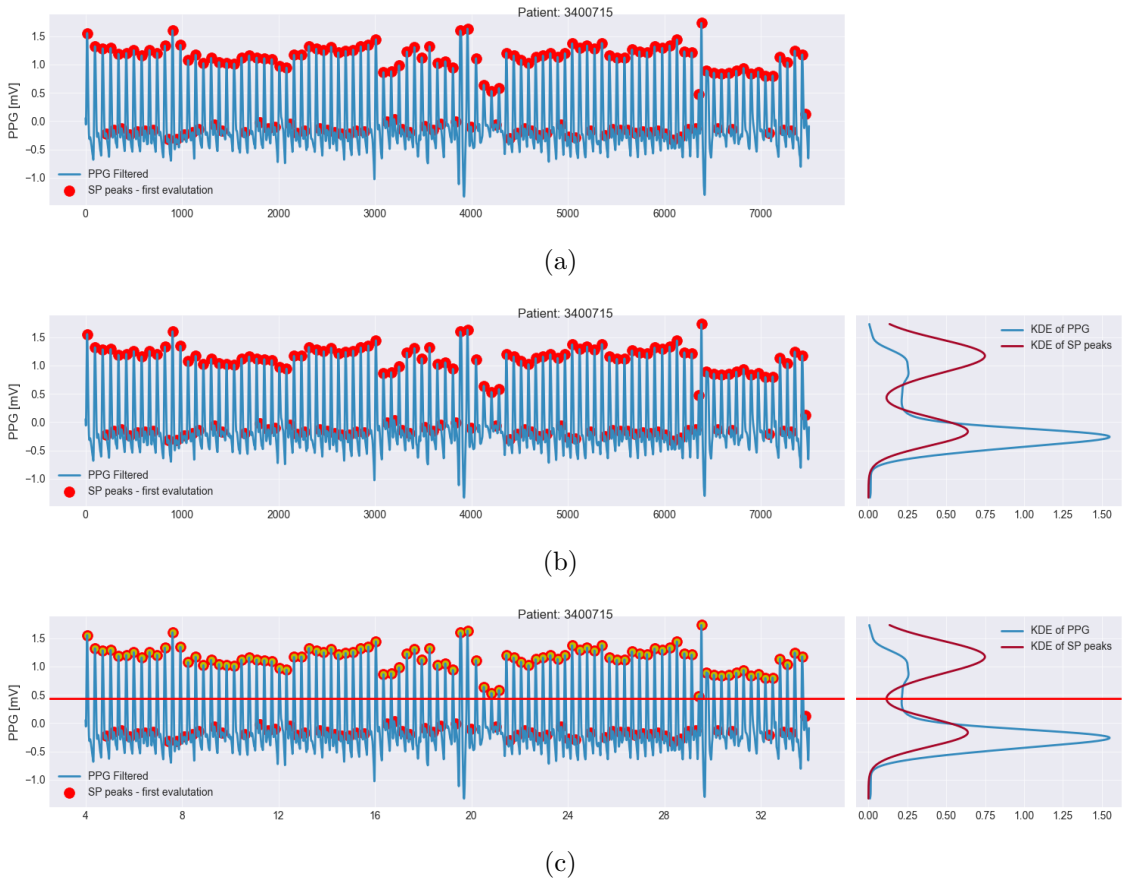


Figure 3.14: Identification of systolic peaks in PPG: (a) First step; (b) Second step; (c) Third step.

The set of ECG, PPG and ABP signals relating to a single patient, or just 'batch', it has been discarded if one or more of the following conditions were met:

- Absence of any peaks (leading to erroneous estimation of HR and PTT features) was noticed for more than 1 second.
- HR or PTT estimated values after this process were physically impossible (e.g. DBP reaching 0 mmHg).
- Possible lack of synchronization between the signals.

So, at the end of the filtering and peak recognition phase, 90 patients suitable for a reliable regression study were recognized.

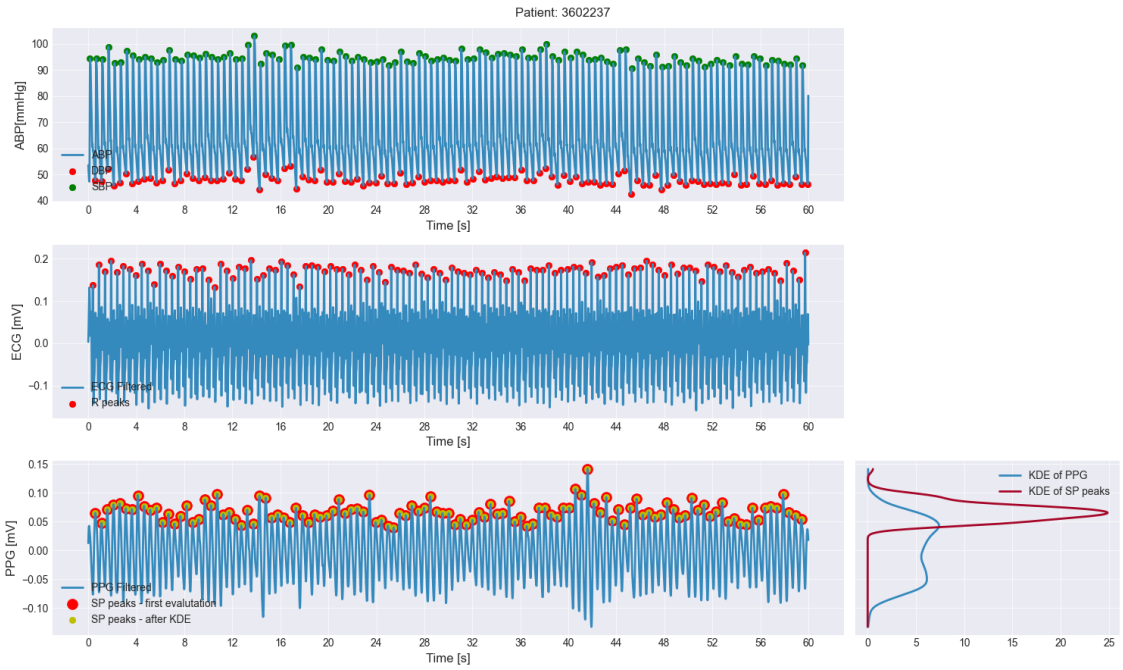


Figure 3.15: Example of second minimum used as threshold for the PPG peaks selection.

3.5.2 Dataset Construction

It is now possible to extract features from the collected data. The implemented calculations require that the HR vector is constructed by going to a difference between successive values of R peaks over time (Figure 3.16) and obtaining the R-R interval, then the HR is calculated according to the Equation 2.1.

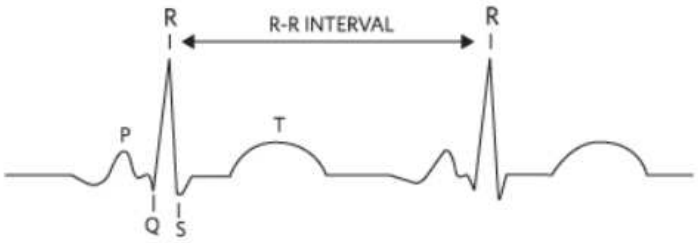


Figure 3.16: Scheme of R-R interval location [21]

For PTT evaluation, instead, it has been implemented the following steps:

1. Pattern verification: there is an S peak following the R peak. Sometimes this

is not verified due to a lack of recognition of one of the two peaks, so the sequence in time is S-S or R-R.

2. PTT calculation: the identified R-S pairs allow to obtain the PTT value according to the Equation 2.2.

Once the features values are obtained, they undergo another processing step. In fact, it should be kept in mind that this thesis work wants to build a machine learning approach suitable for future software development for the SINTEC device, whose data will be taken from the device itself. With this work therefore it has been attempted to build a dataset suitable for the prediction of blood pressure, although the signals were not of optimal quality and of limited duration. For these reasons, the obtained HR and PTT values are evaluated by standard deviation over the observation window. Taking into account that in one minute of recording (duration of the signals in the MIMIC database) the physiological parameters can be considered stationary, any sudden and isolated variations of the parameter can be traced back to an error in the identification of peaks [117]. Both vectors are evaluated by following these steps:

1. Observation of a time window equal to $\frac{1}{5}$ of the signal length.

2. Calculation of the standard deviation within the observation window.

3. If the standard deviation is greater than a threshold (equal to 3 for HR and 0.5 for PTT) then the outlier value is replaced by the polynomial approximation, evaluated in the window.

4. Observation window of the next window, overlapped by half to the previous one. So, continue with the step 2 until the end of the vector.

For each subject has been evaluated a total of 10 observation windows. As can be seen from Figure 3.17, abnormal heart rate values for the parameter trend are reported in ranges more likely to be realistic. Similarly, it can be observed an example for the PTT trend in Figure 3.18.

As it can be noticed from the previous graphs (Figure 3.17 and Figure 3.18) HR and PTT values are represented as continuous in the time. After the identification of a reliable vector in time, the dataset has been enlarged. In fact the following steps have been applied:

1. Reconstruction of the time axis from 0 to 60 seconds with a delta time of 0.1 seconds.

2. HR, PTT, DBP and SBP values identified at a specific t are placed at the time instant to which they belong. Rows corresponding to time instant for which no peak is present are filled with NaN.

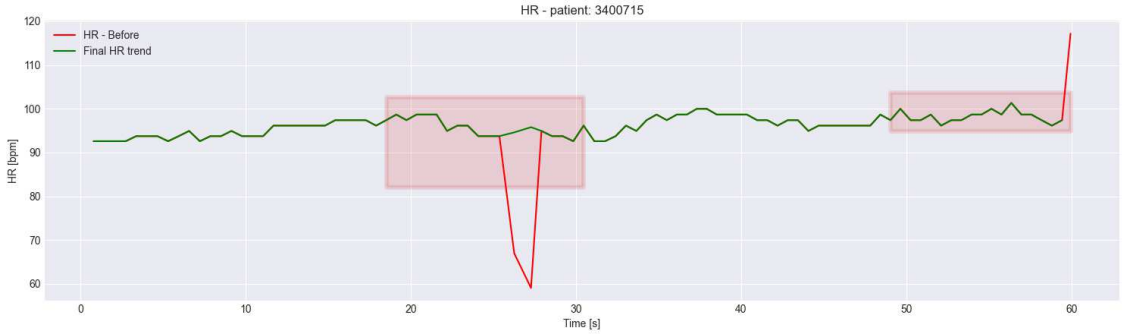


Figure 3.17: Example of HR interpolation. The pink rectangles are the observation windows where the algorithm identifies anomalies.

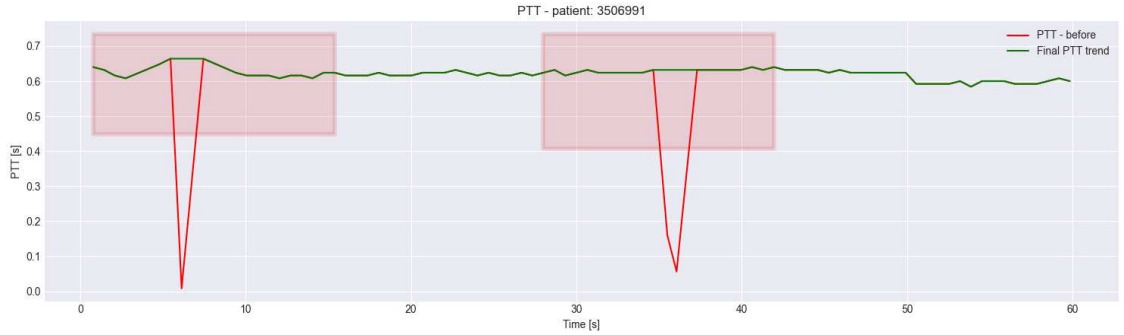


Figure 3.18: Example of PTT interpolation. The pink rectangles are the observation windows where the algorithm identifies anomalies.

3. Application of the interpolating function `DataFrame.interpolate()` to fill the rows with NaN with the corresponding calculated value of the first-order polynomial approximation of the series (for each parameter).

4. Intermediate values, which are reasonably possible but not calculated, are obtained between two successive measurements for all parameters. The dataset therefore consists of 600 values for each parameter.

This is a way of enhancing the size of the dataset by exploiting a feasible approach. In fact, the intermediate steps of evaluating the KDE, the standard deviation analysis and this last interpolation are not suitable for a real time application. However, these steps are only used during the calibration phase to allow the construction of a predictive model based on a larger number of possible intermediate plausible data. In this way, a more robust model is created.

3.6 Regression section

90 subjects (Table 3.1) and 90 dataset with dimensions 600x5 was then built for the regression of SBP and DBP parameters. For this application BP should be

N	Patient ID	N	Patient ID	N	Patient ID
1	3000714	31	3600293	61	3609839
2	3001203	32	3600490	62	3700665
3	3001689	33	3600620	63	3700837
4	3402291	34	3601272	64	3703763
5	3402408	35	3602237	65	3703856
6	3403213	36	3602521	66	3703872
7	3403274	37	3602666	67	3704307
8	3502786	38	3602766	68	3704658
9	3503404	39	3602772	69	3704803
10	3503406	40	3603256	70	3705715
11	3503726	41	3603658	71	3705993
12	3503945	42	3604217	72	3800183
13	3505101	43	3604352	73	3800350
14	3505162	44	3604404	74	3900487
15	3505174	45	3604430	75	3901160
16	3506991	46	3604660	76	3901339
17	3507993	47	3605724	77	3901654
18	3508009	48	3605744	78	3902124
19	3508299	49	3606358	79	3902445
20	3508696	50	3606882	80	3902729
21	3509498	51	3606909	81	3902894
22	3509505	52	3607077	82	3902994
23	3510820	53	3607464	83	3903282
24	3511504	54	3607634	84	3904246
25	3512125	55	3607711	85	3904308
26	3513230	56	3608436	86	3904396
27	3513631	57	3608706	87	3904550
28	3513879	58	3609155	88	3905695
29	3515650	59	3609182	89	3905772
30	3516310	60	3609463	90	3907039

Table 3.1: List of Patients ID used for the regression process.

measured continuously during diverse activities by means of innovative devices. In order to improve performances, it is proposed the use of HR and PTT features in a specific window length. Each SBP and DBP measurement is predicted based on a

PTT and HR window of about two cardiac cycles (with an HR of 90 *bpm* it means T= 1.5 *s*). So, after the interpolation, in a time window of 1.5 *s* are present 15 successive measurements (Figure 3.19).

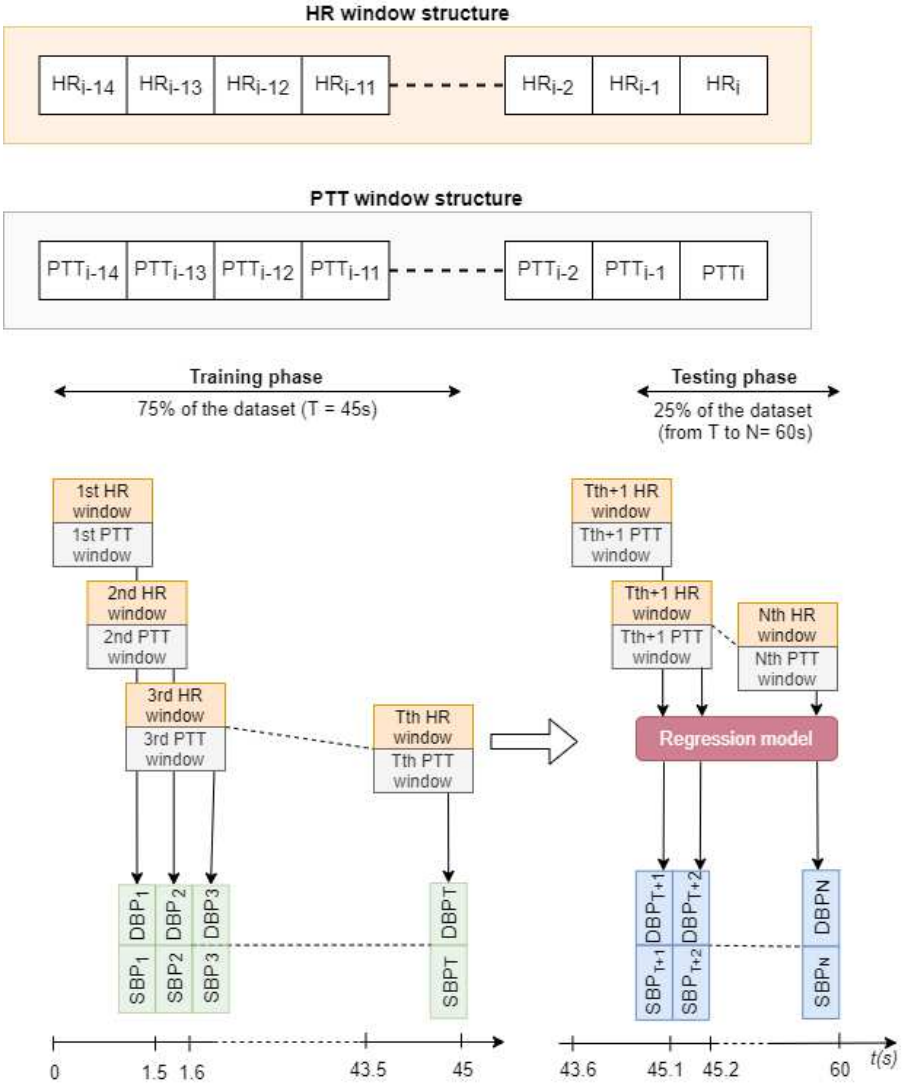


Figure 3.19: Scheme of dataset organization and the regression process.

3.6.1 Regression analysis

Dataset is divided into two subsets: a certain amount of data is used for training the algorithm, and the remainder for performance verification. Initially, this division was performed for each patient:

- Training Set: 75% of the dataset (approximately from 0.12s to 45s);

- Test Set: the remaining 25% of the dataset (approximately from 45s to 60s).

The Training Set has been considered from the second 1.5 s because the first value of SBP and DBP obtained needs the previous 15 samples of HR and PTT as previously explained. As it can be seen from the histogram in the Figure 3.20, the membership ranges and the distribution of HR and PTT values are very different. So, once the dataset is divided into two subsets, they has been normalized using the data of the training set, in order to have a more reliable regression analysis. It has been chosen a different method from the classic min-max scaling, in fact it has been applied the python function `sklearn.preprocessing.RobustScaler()` that removes the median and scales the data according to the Interquartile Range (IQR). The IQR is the range between the twenty-fifth quantile and the 75th quantile. This scaler allows a more robust normalization, especially in the presence of outliers (which have more weight with the classical evaluation).

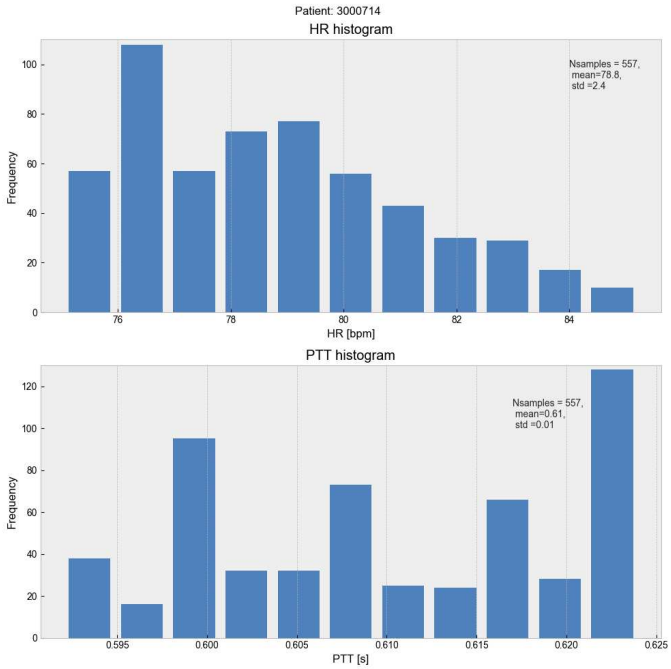


Figure 3.20: Histogram of Data distribution of a example patient.

So, it has been implemented different automatic supervisioned algorithms for the BP measures: SBP and DBP. Furthermore, linear model is evaluated through the Adjusted R2 score and correlation score.

The first regression technique used in this application is linear regression. The DBP and SBP values were then estimated using the following formulations 3.7 and

3.8:

$$SBP = [a_S] * [PTT] + [b_S] * [HR] + c_S \tag{3.7}$$

$$DBP = [a_D] * [PTT] + [b_D] * [HR] + c_D \tag{3.8}$$

Where $[a_S]$, $[b_S]$ are the systolic coefficient vectors for each $[PTT]$ and $[HR]$ values of the time window, and $[a_D]$, $[b_D]$ the diastolic ones. The coefficients of the equation are calculated, for each patient in the training phase of the algorithm, with the aim of minimizing the quadratic error of the identified line and data. The linear relation between a feature of the dataset and the target is evaluated through the Pearson's correlation coefficient. It is a measure of linear correlation between two sets of data, according to the formula 3.9:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^{n}(x_i - x)^2}\sqrt{\sum_{i=1}^{n}(y_i - y)^2}} \tag{3.9}$$

thus it is essentially a normalised measurement of the covariance, the result always has a value between -1 and 1.
The goodness of the fit shows that it is not a good linear correlation between the single feature and the output (Table 3.2), as far as the r_{xy} is about zero.

	r_{xy}	
	Mean	STD
DBP	0.25	0.037
SBP	0,24	0,015

Table 3.2: Mean and standard deviation of Pearson Coefficient r_{xy} between the features values and the output value.

Moreover, a study has also been performed regarding the linear correlation between all model parameters and the output. In this case, therefore, having a number of features greater than one, the Adjusted R-squared has been employed. That metric compares the descriptive power of regression models, when more than one predictor is involved, and it is used for multiple regression analysis. It increases when the new term enhances the model more than would be expected by chance. It decreases when a predictor improves the model fewer than expected. Results show that Linear model is not suitable for this application. In fact, also considering multiple features simultaneously the linear correlation is low: in this case the mean value obtained averaging the adjusted R-squared among patients, it is minus than 0 (Table 3.3).

The parameters used for linear regression, however, are affected by multicollinearity, or near-linear dependence. The multicollinearity is a statistical phenomenon

	$R_{squared_{xy}}$	
	Mean	STD
DBP	−3,51	9,84
SBP	−3,89	11,69

Table 3.3: Mean and standard deviation of adjusted R-squared calculated among patients.

that is established when the independent variables of a linear regression model are correlated with each other [118]. In fact, the calculation of Heart Rate is strongly correlated to that of Pulse Transit Time. Hence, the ill-conditioning of the problem is reflected in the estimation of the coefficients of the linear regression. Ridge Regression (RR) has been developed as a regression technique for overcoming the imprecision of the least square estimators, i.e. linear regression [119][120]. This model solves a regression model where the loss function is the linear least squares function and regularization is given by the L2-norm: the normalization term allows, through a control parameter, to have a mean squared error for the ridge estimator smaller than the corresponding least squares estimator quantity [120]. For implementing the Ridge Regression it is used the Python function `sklearn.linear_model.Ridge()`, different control parameter (alpha) has been tried for the best fitting to the data:

- $\alpha = 0.001$;
- $\alpha = 0.01$;
- $\alpha = 1$;
- $\alpha = 10$.

Another regression method that has been implemented for this application is the Support Vector Regressor (SVR). This technique allows to individualize the hyperplane, in the N-dimensional space of the input features. The constraint is to minimize the error between the predicted value of the function for a given input and the actual output [110]. Using a simple example, it can be seen in the Figure 3.21 a linear SVR representation. The objective distance ϵ identify the decision boundary and margins (ξ) are the over boundary values that are tolerated for the modelling. Margins allow the flattening of the line [121].

The optimization problem is made up by solving the formula 3.10:

$$MIN(\frac{1}{2}\|w^2 + C\sum_{i=1}^{N}\xi_i) \tag{3.10}$$

where C determines the trade-off between the margins and the error rate. SVR can also modelling data that have a non-linear distribution, using the kernel [121].

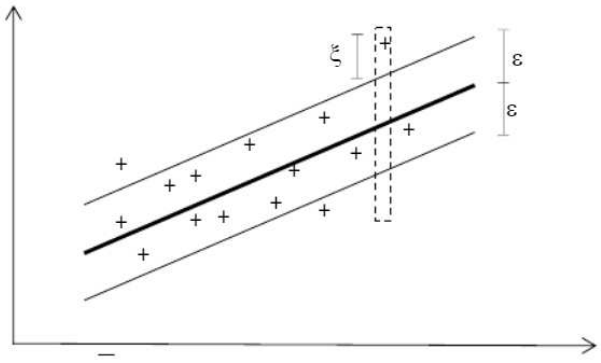


Figure 3.21: Scheme of linear SVR method.

SVR has different parameters to be customized for having the best fitting curve. It has been considered for this application:

- Epsilon: fixed to 0.2;
- Kernel function: Linear, Radical Basis Function (RBF), Polynomial;
- C value: 1, 10, 20, 30, 50.

Furthermore, it has been exploited the Epsilon-Support Vector Regression implemented in Python with the function `sklearn.svm.SVR()`.
Stating that also non-linear regressor can be suitable for this application, another method is applied.
Random Forest Regressor is an ensemble learning technique based on decision trees. To build a Random Forest, an algorithm is trained by learning from multiple decision trees, driven on slightly different subsets of data [122]. Decision trees prediction is combined by the bagging method (Figure 3.22). Bagging is made up by three main stages:

1. Choose the number of trees of the forest ($nTree$) and the dataset size given in input (uniform sampling with replacement);
2. Construction of a certain number of estimators with $nTrees$ for the ensemble forecasting;
3. Prediction using aggregation of results that allows to reduce the variability of the results by averaging.

. This technique has considerable advantages over its predecessors. In fact, since it is an ensemble method, it is much less likely to be overfitted to the training data and can work with large dataset. On the other hand, it is much more computationally

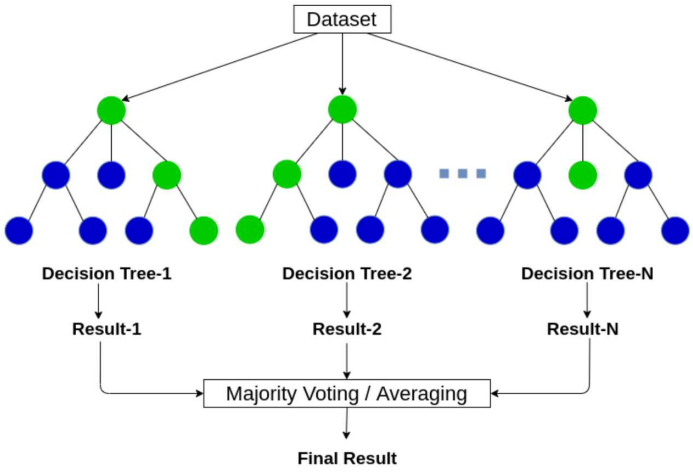


Figure 3.22: Scheme of Bagging process in Random Forest.

expensive. This last observation does not have decisional importance on the choice of the algorithm because the dataset needed for the calibration process allows to maintain short times.
For the algorithm implemented in this work, the Python function `sklearn.ensemble.RandomForestRegressor()` was used. MAE is used as metrics for the quality of a split, 100 estimators for each model and different numbers of decision trees in each estimator are tuned:

- nTrees = 10;
- nTrees = 20;
- nTrees = 50;
- nTrees = 100;
- nTrees = 200;

Chapter 4

Results

It is now necessary to identify which of the selected methods is most suitable for blood pressure prediction. The obtained results in the testing phase for each regression algorithm are observed. The guidelines provided by AAMI recommend an error assessment using the MAE as a metric. The minimum performance requirements are [89]:

- At least 85 records;
- Error of less than 10 mmHg on both SBP and DBP;
- Error tolerance on 85% of the dataset.

4.1 Regression performance

Each batch, i.e. is the recording of the specific patient, has been processed and the dataset built. The forecasting process gives for each line of the dataset a predicted value y_i. The error between the y_i and the expected value y has been calculated for each subject using MAE, so at the end of the regression process 90 MAE values are obtained for each algorithm. Mean and standard deviation calculated on these 90 values is used for methods comparing. Firstly, each algorithm was constructed using the specifications explained in the previous chapter. Then, the closest version of the model to the actual data representation was identified.
The first model, i.e. Linear Regression, does not need any control parameter, so the obtained results of testing phase are fixed. As shown in Table 4.1, the model satisfies the AAMI guidelines on error tolerance.
Using Ridge regression instead, there is the need to select the best tuned parameter for this application. It can be observed in Figure 4.1, that as the control parameter alpha increases, the mean error on each patient decrease. However, there is no significant difference between the results to prefer one to another using only this observation. On the other hand the algorithm, both on DBP and SBP, with higher

	MAE values [mmHg]
DBP	2.02 ± 1.44
SBP	3.22 ± 2.21

Table 4.1: Performance evaluated with Mean Absolute Error (MAE) for linear regression.

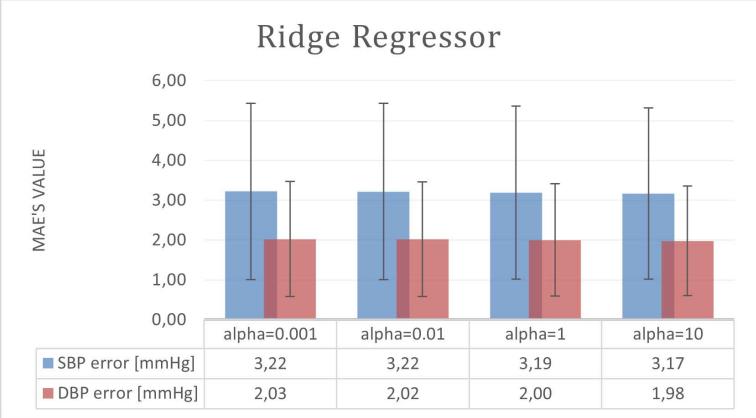


Figure 4.1: Graphical representation of MAE's value in Ridge Regression with different alphas.

alpha seems to reach a good recognition of the trend together with a greater stability to changes (Figure 4.2). Alpha equal to 10 is chosen for the next step.

The third algorithm trained is the SVR. It can be noticed (Figure 4.3) that the polynomial kernel is not suitable for the data modelling. Instead, RBF (Figure 4.3(b)) and Linear (Figure 4.3(a)) ones have very low MAE's values. Moreover, As it can be noticed from the Figure 4.4 all the configuration have almost the same performances. The lowest error is identified in the configuration setting of an SVR with a RBF kernel and a C equal to 1, that also allow to save computational time.

Finally, the Random Forest regressor is used tuning the number of trees parameter and the results are reported in the Figure 4.5. It can be seen that even in this case, the mean prediction error across all patients is approximately constant and after twenty trees the MAE's value is under 2 mmHg on DBP estimation. This means that the simplest model can be used for blood pressure estimation (nTree = 20).
Performances are then evaluated in order to identify the best model among the previously described.
First of all, it can be noticed in the Figure 4.6 that the algorithms satisfy the AAMI requirements. Indeed, the MAEs are low and similar among the four model constructed. For this reason it is necessary to evaluate deeper parameters for the selection of the algorithm. Figures below are structured as the following one (Figure

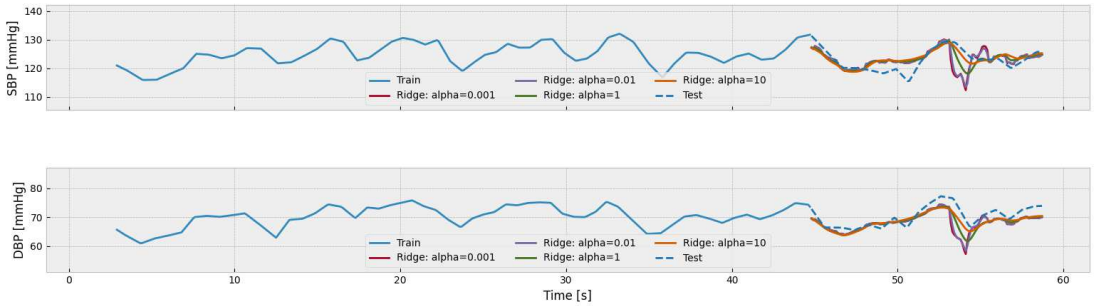
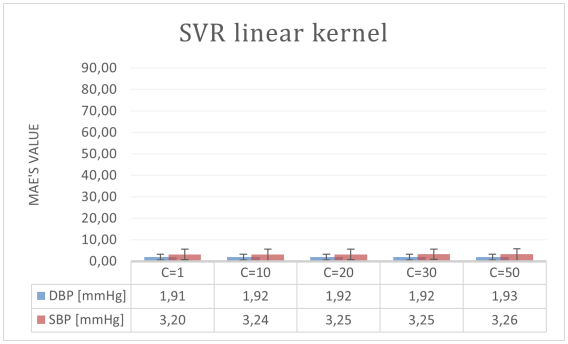
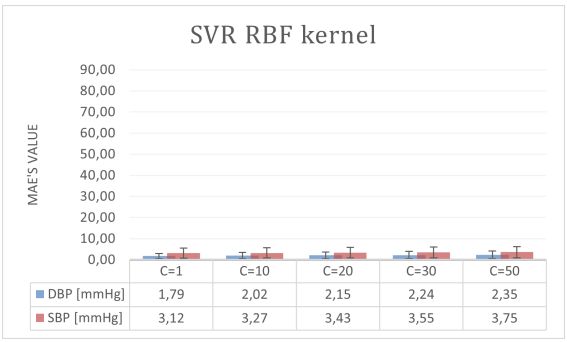


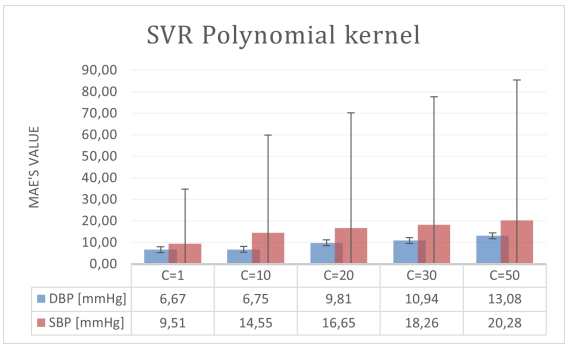
Figure 4.2: Example of trend recognition and stability for different alphas in Ridge regression.



(a)



(b)



(c)

Figure 4.3: Results of SVR testing phase.

4.7):

- The first graph shows the development of PTT and HR values interpolated over time.
- The second graph shows the development of SBP values interpolated over time,

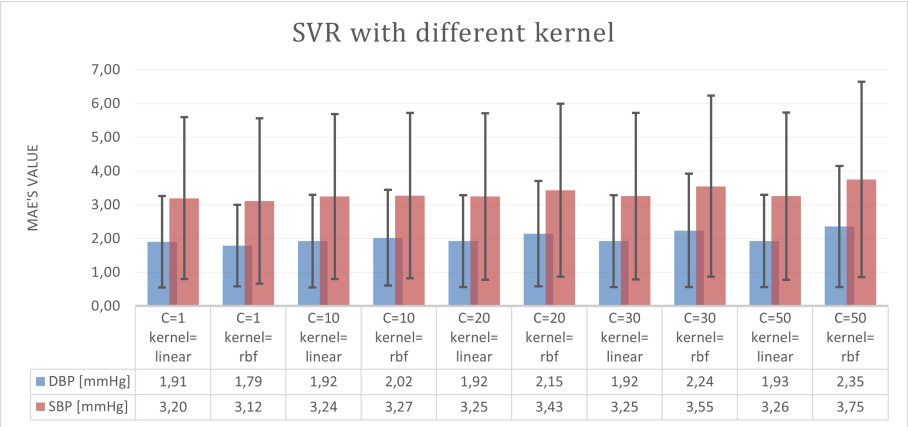


Figure 4.4: Comparison Linear and RBF SVR errors.

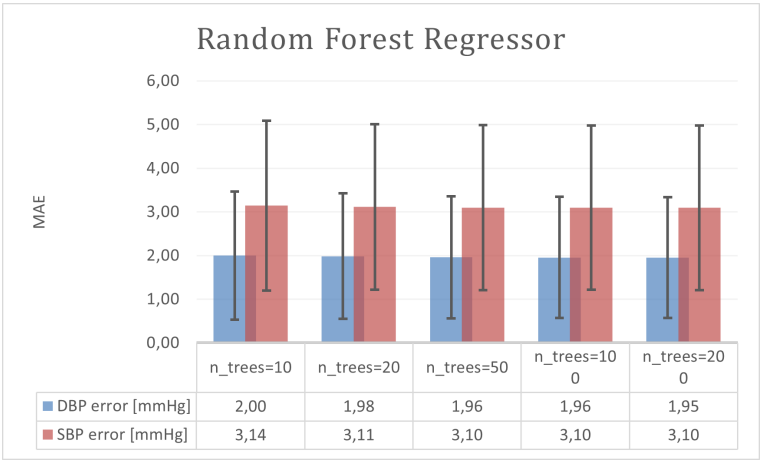


Figure 4.5: Results of Random Forest testing phase.

and it is divided into the Training phase (from 0s to 45s) and testing phase (from 45s to 60s).

- The third graph shows the development of DBP values interpolated over time, and it is also divided into the Training phase and testing phase.
- The fourth graph indicates the MAEs values for each algorithm.

Three main and recurrent observations has been done analysing the interpolation of the time series during Test phase on the 90 subjects:

- When the input data are stable, all the models work well (Figure 4.8): great trend recognition and low error;
- Random Forest Regressor has been more precise in recognize trends (Figure 4.9 in SBP prediction);

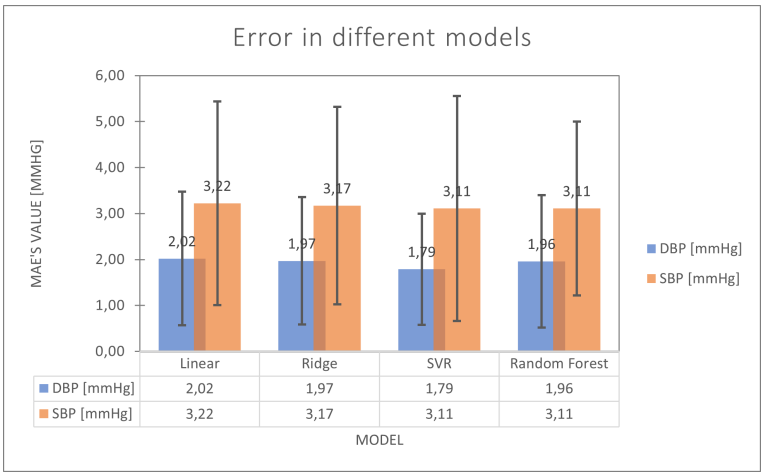


Figure 4.6: Graphical representation of MAEs obtained in the testing phase for comparison aim.

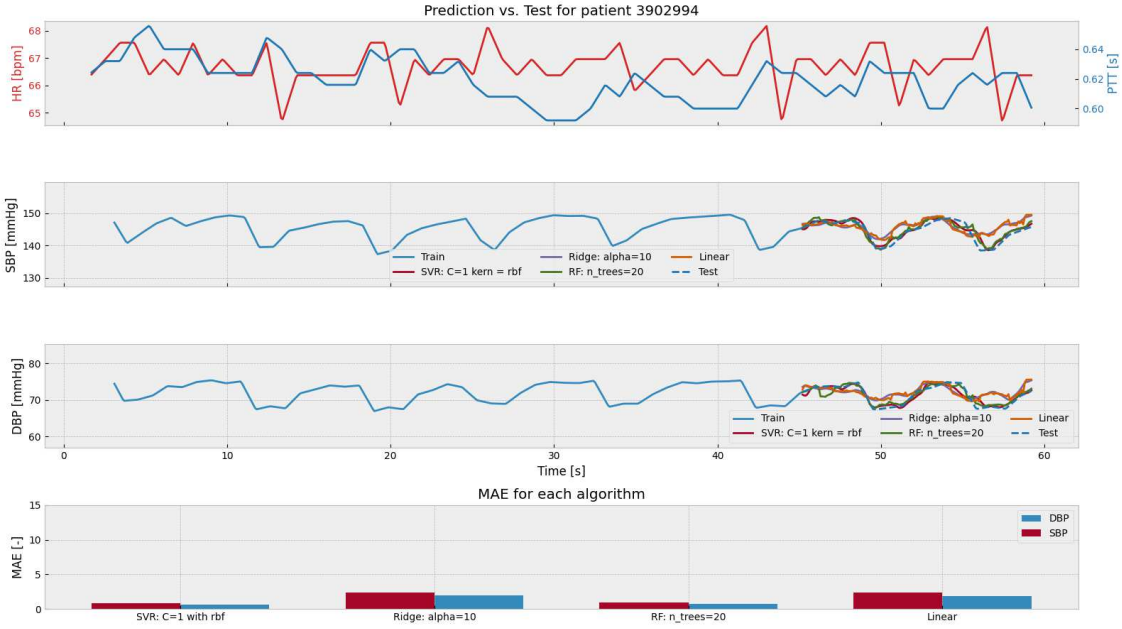


Figure 4.7: Example of training and testing phase done with the four selected algorithms.

- SVR has a higher tendency to have an averaging behaviour (Figure 4.9);
- Linear Regressor has been more sensible in PTT/HR rapid changing (Figure 4.10) and it led to abnormal forecasted values.

For these reasons and since he AAMI expectations has been satisfied (Table 4.2),

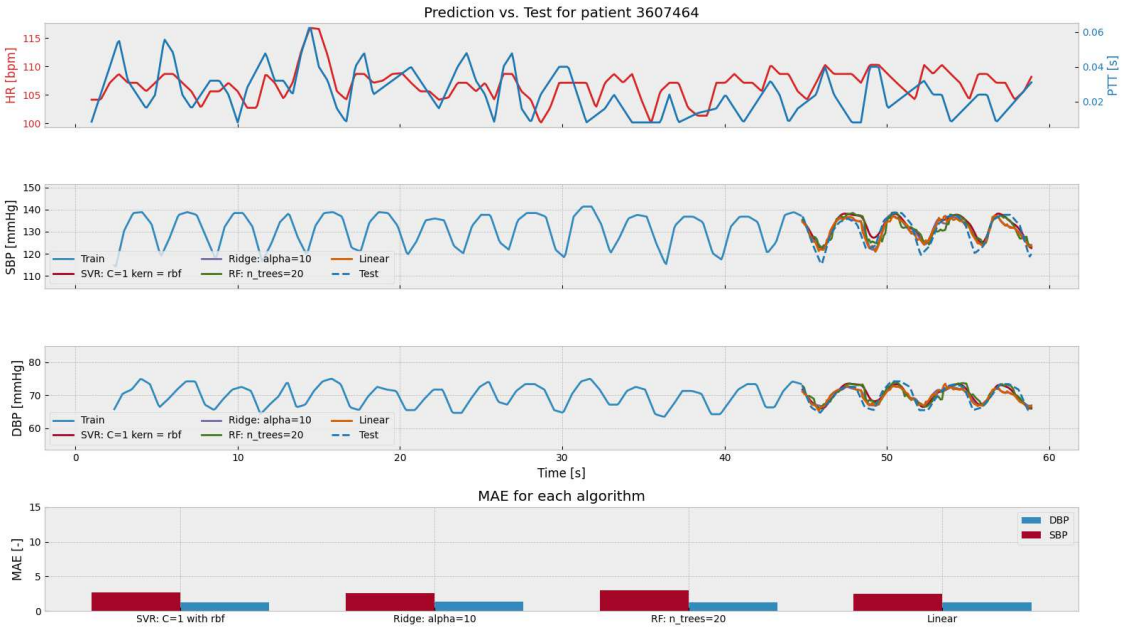


Figure 4.8: Example of prediction that underline the trend recognition.

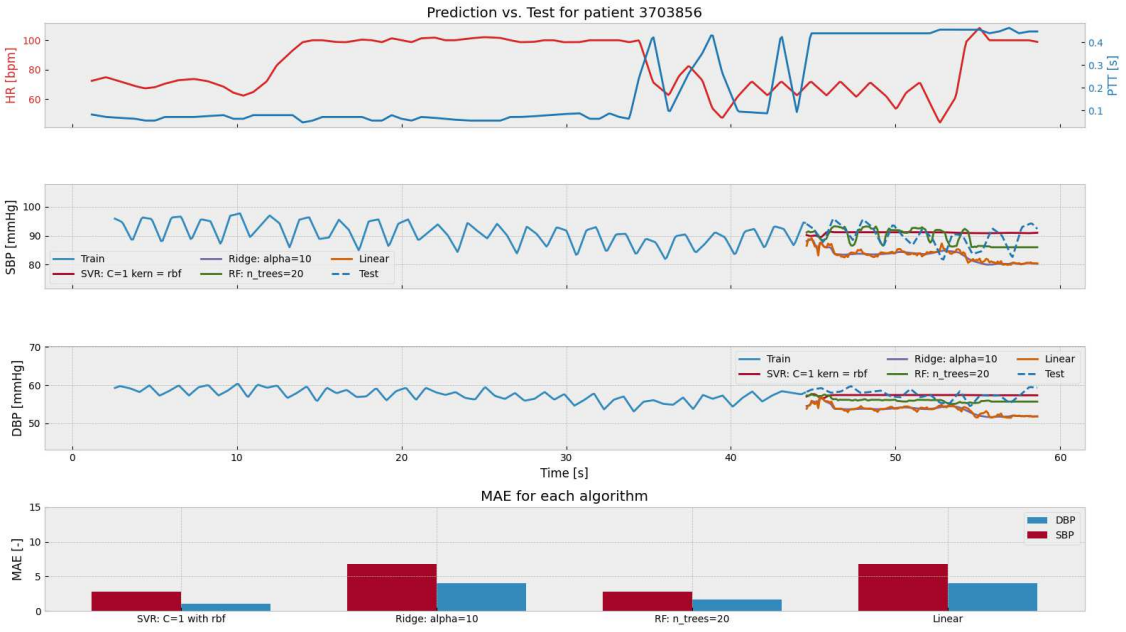


Figure 4.9: Example where is clear the SVR's averaging tendency.

Random Forest model has been chosen for the validation process.

MAE's values have been reported in two histogram plots (for both DBP and

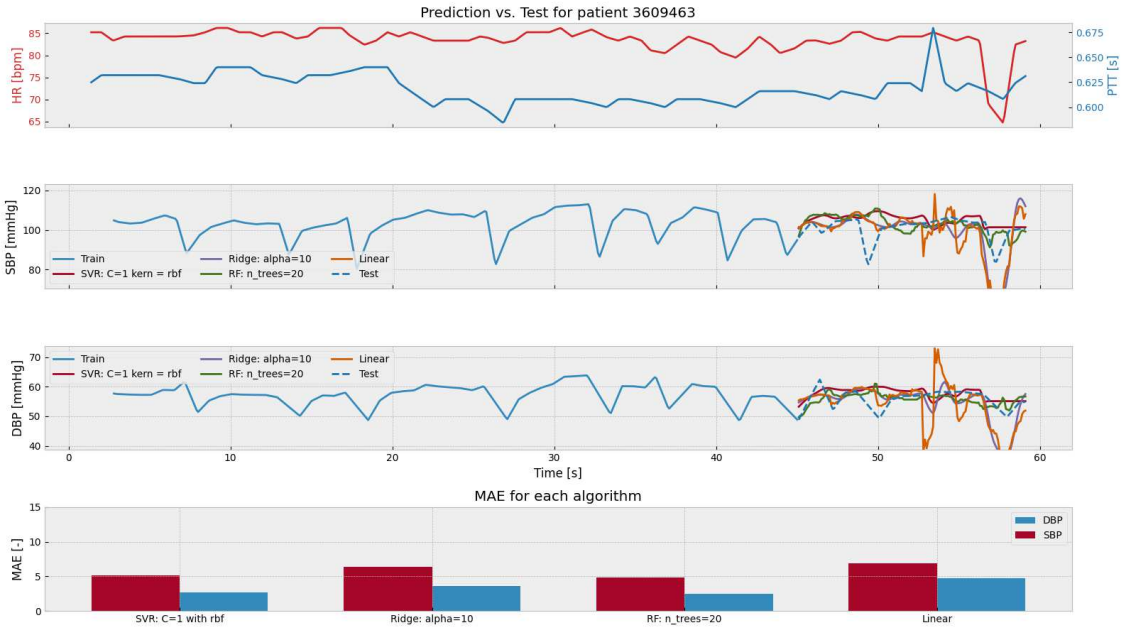


Figure 4.10: Example that underline the instability of linear regressor.

	DBP error [mmHg]	**SBP error** [mmHg]	**Number of subjects**
Random Forest	$1{,}96 \pm 1.44$	3.11 ± 1.89	90
AAMI guidelines	≤ 10.00	≤ 10.00	≥ 85

Table 4.2: Comparison of Random Forest (nTree = 20) performances and AAMI guidelines

SBP) for observing the error distribution among subjects. The graphs in Figure 4.11 and Figure 4.12 confirm that forecasting main error is located in ≤ 3 mmHg area for DBP and in ≤ 5 mmHg area for SBP, instead higher values belong very few subjects.

4.2 Validation process

The chosen algorithm was validated on the dataset. The methods employed are:

- AAMI validation;
- Wilcoxon Signed rank test;
- Standard error;
- K-fold cross validation for time series.

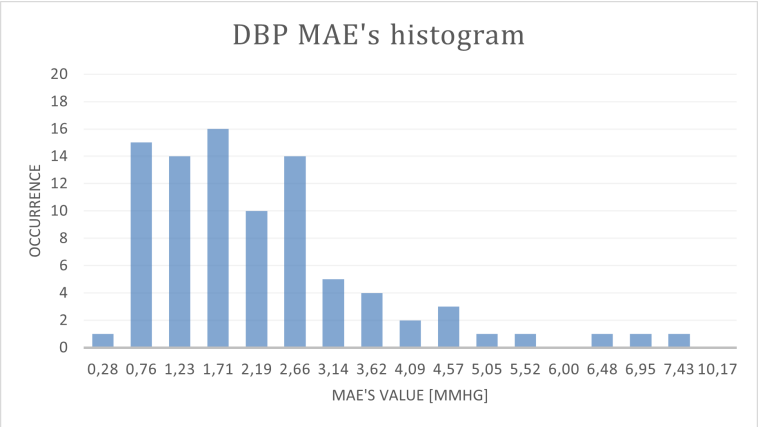


Figure 4.11: DBP errors histogram.

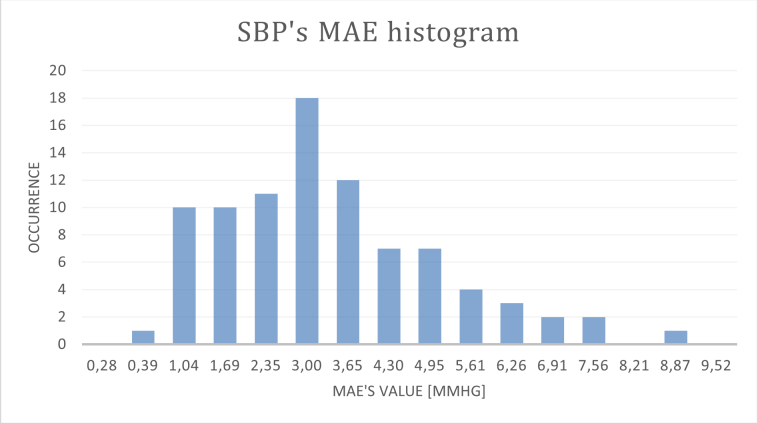


Figure 4.12: SBP errors histogram.

In order to satisfy the AAMI guidelines, the chosen algorithm have to meet also the goal of 85% of samples with error lower than 10 mmHg. The error of prediction as follows:

$$|\hat{y} - y| \tag{4.1}$$

where $\hat{y}$ is the predicted value and y the target one. For this aim, every batch is submitted to this analysis. The results (Table 4.3) show that some subjects did not meet the requirements. In particular, two subjects (3509590 and 3704307) have not be validated on both DBP and SBP. For investigate the reasons of this behaviour the starting signals used to compose the dataset are observed.

Subject 3509590. In Figure 4.13 it can be noticed that the extraction of peaks is consistent with the shape of the signal, as well as the choice of threshold for the detection of systolic peaks from the PPG (third graph) is correct. So, going further, in Figure 4.14 it is shown the interpolation of HR and PTT values: even in this

	Subjects	Validated subjects
DBP	90	87
SBP	90	85

Table 4.3: Starting subjects and Validated subjects (more than 85% of samples under the error threshold) for DBP and SBP.

case, no anomalies can be discerned, e.g. in the range of parameters or the trend.

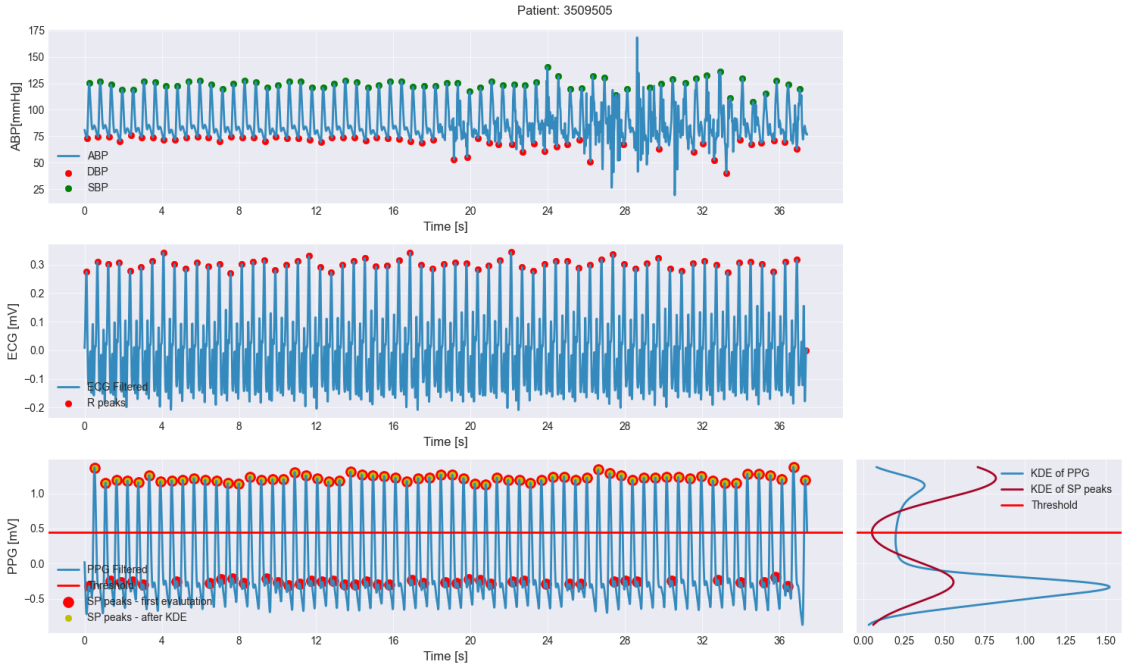


Figure 4.13: ABP, ECG and PPG signal of subject 3509590 with the found peaks.

The conclusion in this case is that the model is not able to fit the data given as input.

Subject 3704307 In Figure 4.15 the input and target data are shown. In this situation, however, it is clear that the peaks related to the PPG are not correctly identified, as there are windows (shorter than 5s) in which no peak is detected. However, this weakness is overcome by the statistical analysis of the stability of the parameter (Figure 4.16). The prediction error demonstrates an inability of the regression model to fit the data and provide results that meet the expectation. Anyway, AAMI guidelines are respected for both models: it is acceptable if a tolerable error ($\leq$ 10 mmHg) is at least on 85% of the dataset in 85 subjects.

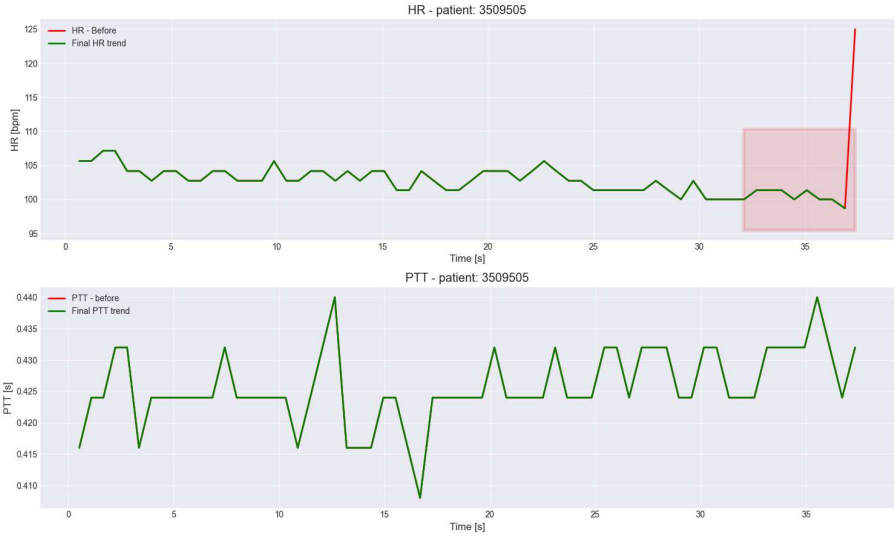


Figure 4.14: Interpolation of HR and PTT values of subject 3509590.

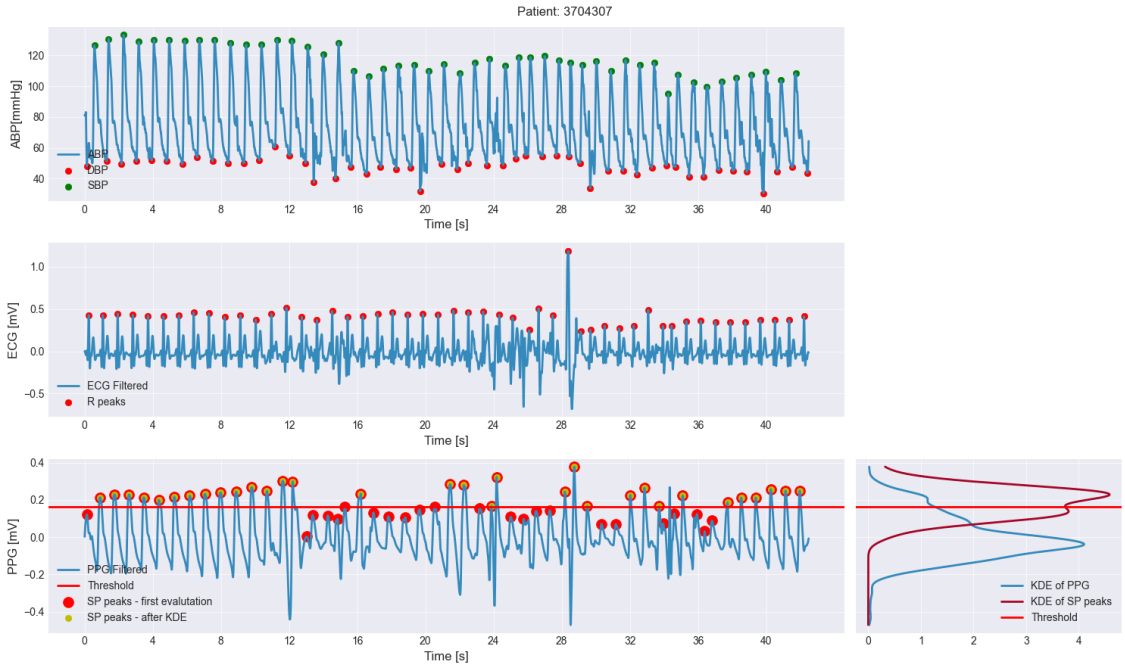


Figure 4.15: ABP, ECG and PPG signal of subject 3704307 with the found peaks.

The analysis proceeds with Wilcoxon Signed rank test [123], that is one of Non parametric Statistical Significance Tests and do not assume a specific distribution to the data. The objective was to test whether there was a significant difference between

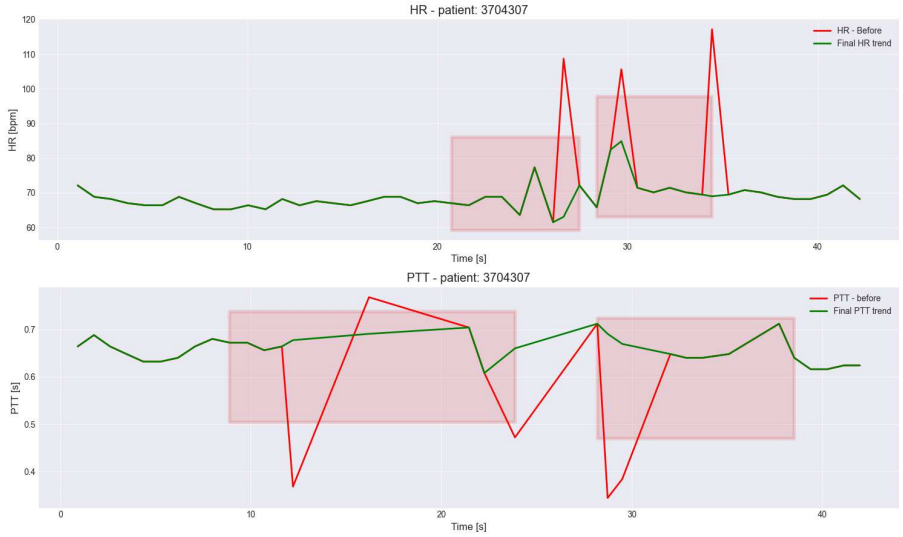


Figure 4.16: Interpolation of HR and PTT values of subject 3704307.

the DBP and SBP values obtained manually and automatically. The null hypothesis H_0 assumes that the two distributions, i.e. the target values and the forecasted ones, do not differ significantly from each other. In particular, it tests whether the distribution of the differences $SBP_{target}/DBP_{target}$ - $SBP_{forecasted}/DBP_{forecasted}$ is symmetric about zero.
From a bivariate distribution (i.e. paired samples), given n independent samples (x_i, y_i), it is computed the differences $d_i = x_i - y_i$. The only one assumption of the test is that the differences are symmetric. It is applied a two-sided test, so it has the null hypothesis that the median of the differences is zero against the alternative that it is different from zero.
The criterion of assessment is based on the computation of a tail probability (of the data distribution) which is called the P-value. Assuming the null hypothesis to be true, the P-value is the probability of obtaining results that are equal to or less likely than those observed during the test [124]. It has been then assessed by setting a test significance level $\alpha\ of 5\%$ prior to its calculation (it has been used the *exact* distribution since it is recommended for sample sizes of up to 25).
So, once the P-value has been calculated for the observed data, it is possible to act as follows:

- If P-value $> \alpha$ the empirical evidence is not contrary enough to the null hypothesis, which therefore cannot be rejected;

- If P-value $\leq \alpha$ the empirical evidence is strongly against the null hypothesis and must therefore be rejected. In this case, the observed data are said to be statistically significant.

In this application, the outcomes (Table 4.4) reveal that the null hypothesis can not be rejected, that means that it is not possible to assume that there is a statistical difference between the values obtained by the signal and the predicted one.

The third assessment is the Standard Error (SE) (Equation 4.2):

$$SE = \frac{SD}{\sqrt{N}} \tag{4.2}$$

where SD is the standard deviation and N the samples size. The standard error falls as the sample size increases, as the extent of chance variation is reduced [125]. This value gives an indication about the representativeness of the sample in respect of the overall population: e.g. high SE low representation of the sample in the distribution. In particular, for regression analysis, the standard error assesses the precision of the predictions, more than other methods, such as R-squared metric. It has been obtained a very small value of SE for both ABP peaks and it indicates

	P-value [%]	**SE**
DBP	7.8	0.016
SBP	5.4	0.021

Table 4.4: Outcomes of statistical analysis of data.

that the observations are closer to the fitted model.

Finally, Cross-Validation (CV), or out-of-sample testing, is any of various similar model validation techniques for assessing how the results of a statistical analysis will generalize to an independent dataset [126]. There are many ways to split data into training and test sets to avoid model overfitting, to standardize the number of groups in test sets, and, in general, to set up a CV [127]. This application relates to time series. For this reason, CV has more specific characteristics, as it is necessary to maintain the information on the progressiveness and evolution of the samples over the time. The method is called Split Time Cross Validation and starts with a small subset of data for training purpose, forecast for the later data points and then checking the accuracy for the forecasted data points. The same forecasted data points are then included as part of the next training set and subsequent data points are forecasted. So, four iterations steps have been constructed. The training set fold consists of 20% of the batch samples in the first iteration. At each iteration it increases by 20% until it reaches 80% of samples (maintaining the chronological order). While, the size of the test set remains fixed at 20% of the batch (about 12 seconds): the test set samples follow those of the training set in each iteration. The percentage is evaluated on the whole dataset (Figure 4.17).

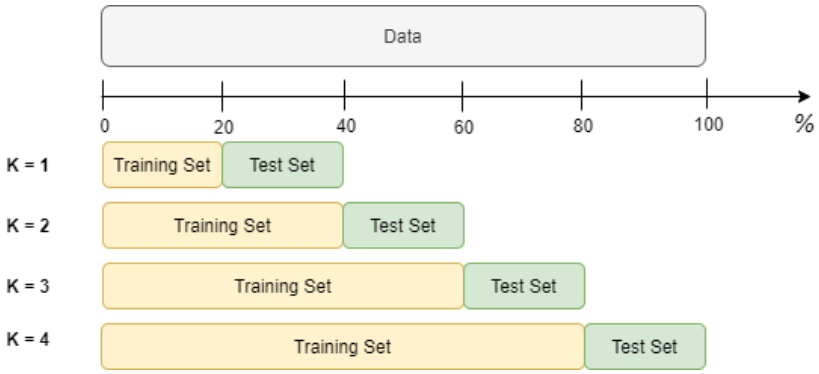


Figure 4.17: Scheme of Cross-validation for time series.

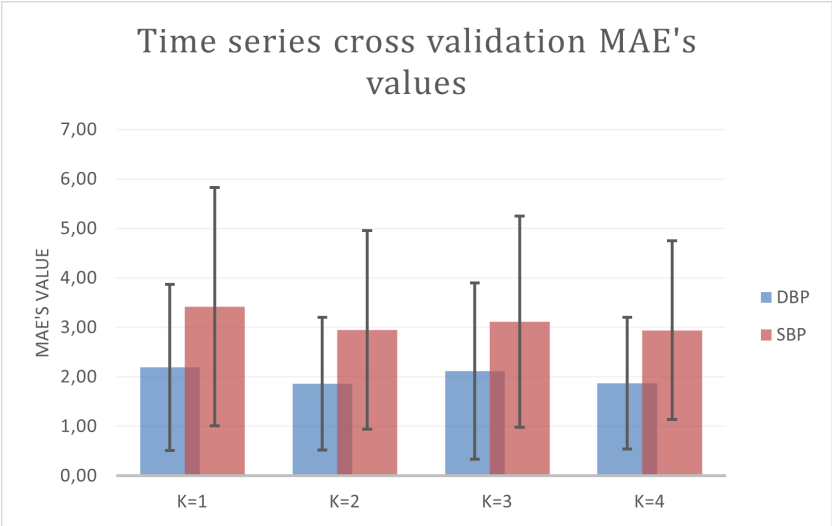


Figure 4.18: Results obtained from the Split time CV.

Below, in graph shown in the Figure 4.18, the results obtained are reported.

The achieved results show that the algorithm is stable with respect to the training set and its size. Therefore, it can be stated that it is highly robust with respect to generalization.

Chapter 5

Conclusions

The work carried out in this thesis has produced robust results on the prediction of maximum and minimum values of blood pressure (SBP and DBP). The search of interested points to be used includes signal analysis and peak detection. Then, the values obtained has been set as targets for regression algorithms, which have been trained to build a model using fixed time window of two parameters: HR and PTT.
The calculation of these two parameters has been preceded by signal cleaning techniques and statistical analysis of the reliability of peak extraction. Accordingly, from the ECG signal, the time between two successive heartbeats was obtained by identifying the position of the R-peaks and locating them in time, in order to calculate a measure of frequency (HR). From the PPG signal, the position of the systolic peak in time was obtained to calculate the PTT parameter, using the position of the R peak for the same beat. The design of a regression model that lead to the presented results confirmed that the dataset constructed with HR and PTT has the ability to capture the information contained in the spectrum of the ABP signal, as witnessed by previous works. In addition, the observation windows of two heart cycle used in this work improve the forecasting performances of the algorithm.
However, the work was carried out on the basis of a dataset from MIMIC III database and the signals of which were taken from invasive instruments on ICU patients. So, the signals used belong to persons who are older than the average target, are in critical health conditions, are administered with drugs of various types. Moreover, healthy people, also with hypertension, do not have radical changes in BP as the ones present in MIMIC III, but have a more stable trend in blood pressure.
Other innovative methods introduced in this work are:

- Automated calculation of peaks, with no need to set up a range in which to search for them. This makes the algorithm more robust and versatile.
- Use of the hypothesis of stationarity of the physiological parameters to correct

possible errors in the calculation of the features through the estimation of the Gaussian density kernel (KDE).

The resulting software is hence designed to be calibrated on the subject to obtain a regression model, using Random Forest Regression, for continuous prediction of blood pressure and help the patient to monitor and control hypertension.

5.1 Future of the work

Having a device like the one designed for the SINTEC project in the future will mean that hypertensive patients will also be able to benefit from a continuous monitoring service using healthcare facilities and staff resources more efficiently and with less recurrence [2]. The picture of the work in progress is:

- In the future, the algorithm must in future be implemented on the SINTEC device, which will work in real time with sensors created specifically for the application:
  - Critical aspects such as the different dynamics of the signals in the database will no longer be a problem for the actual application, because the same devices will be used for all patients.
  - The statistical analysis made on HR and PTT measurements will not be necessary as the signal sampling will provide higher quality data than the database used.
- The prediction is not intended to be beat-to-beat but will be the result of a larger number of observations, that lead to less noised outputs. Moreover, this approach allows us to maintain double-pass filtering, which cannot be used for real-time solutions.
- The use of observation windows of 1.5 seconds does not compromise the real-time development of the algorithm as this only affects the first measurement provided. Subsequent measurements will be available almost synchronously to the signal recording.
- Currently, the algorithm is already achieving very good performance. However, it would be very useful to have more information about the subject:
  - Each model is created on the individual patient. It would be optimal to be able to go beyond the initial patient-specific calibration phase. A successful strategy, which is unfortunately not yet available, could be to be able to add patient biographical information, such as age and gender, to the currently used dataset, together with other information about the patient's body, such as weight and height.

  - In this work the calibration phase is based on Blood Pressure values collected with through an invasive method. It is reasonable to assume, therefore, that by obtaining the pressure values with a sphygmomanometer (a non-invasive and less accurate device) for the calibration phase, the performance of the algorithm itself could be affected by a major error.

- Finally, the proposed application proposes the basis for a telemedicine application. In fact, the data collected and estimated can be used by the general practitioner or specialist to obtain long-range statistical analyses of the patient's health conditions. It is also possible to think that these will be able to predict any worsening of the hypertensive condition, and allow the doctor to intervene in a minimum of time.

The proposed application, with the methods developed, is therefore the result of research work in its final stages. The encouraging results are a good starting point for the development of an effective device for the continuous monitoring of the state of health of hypertensive patients, which will allow the customisation of the patient's diagnostic pathways and an improvement in their well-being.

References

[1] B. Williams, G. Mancia, and W. e. a. Spiering, "2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH)," *European Heart Journal*, vol. 39, pp. 3021–3104, 08 2018.

[2] "Sintec project." `https://www.sintec-project.eu/`.

[3] Wikilectures — Biomechanics of blood circulation. `https://www.wikilectures.eu/w/Biomechanics_of_blood_circulation`.

[4] A. J. Weinhaus and K. P. Roberts, "Anatomy of the human heart," in *Handbook of cardiac anatomy, physiology, and devices*, pp. 51–79, Springer, 2005.

[5] A. Zafar, M. Hamid, A. Khan, M. Usman, and U. Shahid, *Portable Cardiac Monitoring System*. PhD thesis, 05 2016.

[6] R. Santi, "Conduzione del potenziale d'azione e attività elettrica del cuore." `https://slideplayer.it/slide/190667/`.

[7] E. Electrodes, "12-lead ecg placement." `www.emtresource.com`.

[8] F. Badilini, T. Erdem, W. Zareba, and A. J. Moss, "Ecgscan: a method for conversion of paper electrocardiographic printouts to digital electrocardiographic files," *Journal of Electrocardiology*, vol. 38, no. 4, pp. 310–318, 2005.

[9] M. Paul, A. F. Mota, C. H. Antink, V. Blazek, and S. Leonhardt, "Modeling photoplethysmographic signals in camera-based perfusion measurements: optoelectronic skin phantom," *Biomed. Opt. Express*, vol. 10, pp. 4353–4368, Sep 2019.

[10] H. T. Ma, "A blood pressure monitoring method for stroke management," *BioMed research international*, vol. 2014, 2014.

[11] "Bioninja — blood pressure." `https://ib.bioninja.com.au/standard-level/topic-6-human-physiology/62-the-blood-system/blood-pressure.html`.

[12] Omron. `https://www.omron-healthcare.it/it/misuratori-di-pressione/heartguide.html`.

[13] SwipeRX, "What you need to know about hypertension." `https://mobile.swiperxapp.com/what-you-need-to-know-about-hypertension/`.

[14] B. Qian, J. Su, Z. Wen, R. Yang, A. Zomaya, and O. Rana, "Orchestrating the development lifecycle of machine learning-based iot applications: A taxonomy and survey," 10 2019.

[15] S. A. Yarows, S. Julius, and T. G. Pickering, "Home blood pressure monitoring," *Archives of Internal Medicine*, vol. 160, no. 9, pp. 1251–1257, 2000.

[16] G. Parati, G. S. Stergiou, R. Asmar, G. Bilo, P. De Leeuw, Y. Imai, K. Kario, E. Lurbe, A. Manolis, T. Mengden, *et al.*, "European society of hypertension practice guidelines for home blood pressure monitoring," *Journal of human hypertension*, vol. 24, no. 12, pp. 779–785, 2010.

[17] R. Lazazzera, Y. Belhaj, and G. Carrault, "A new wearable device for blood pressure estimation using photoplethysmogram," *Sensors*, vol. 19, p. 2557, 06 2019.

[18] "Regressione e pattern recognition." `http://bias.csr.unibo.it/maltoni/ml/DispensePDF/5-ML-Regressione.pdf`.

[19] "Physiobank atm." `https://archive.physionet.org/cgi-bin/atm/ATM`.

[20] G. Slapničar, N. Mlakar, and M. Luštrek, "Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network," *Sensors*, vol. 19, no. 15, p. 3420, 2019.

[21] `https://www.maximintegrated.com/cn/design/technical-documents/design-solutions/7/7157.html`.

[22] WHO, "Cardiovascular diseases." `https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1`.

[23] Eurostat. `https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20190928-1`.

[24] R. S. Vasan, M. G. Larson, E. P. Leip, J. C. Evans, C. J. O'Donnell, W. B. Kannel, and D. Levy, "Impact of high-normal blood pressure on the risk of cardiovascular disease," *New England journal of medicine*, vol. 345, no. 18, pp. 1291–1297, 2001.

[25] P. K. Whelton, "Epidemiology of hypertension.," *Lancet (London, England)*, vol. 344, no. 8915, pp. 101–106, 1994.

[26] J. Kumar, "Epidemiology of hypertension," *Clinical Queries: Nephrology*, vol. 2, no. 2, pp. 56–61, 2013.

[27] R. S. Vasan, M. G. Larson, E. P. Leip, J. C. Evans, C. J. O'Donnell, W. B. Kannel, and D. Levy, "Impact of high-normal blood pressure on the risk of cardiovascular disease," *New England journal of medicine*, vol. 345, no. 18, pp. 1291–1297, 2001.

[28] A. S. Meidert and B. Saugel, "Techniques for non-invasive monitoring of arterial blood pressure," *Frontiers in medicine*, vol. 4, p. 231, 2018.

[29] D. Shimbo, M. Abdalla, L. Falzon, R. R. Townsend, and P. Muntner, "Role of ambulatory and home blood pressure monitoring in clinical practice: a narrative review," *Annals of internal medicine*, vol. 163, no. 9, pp. 691–700, 2015.

[30] J. A. Staessen, J. Wang, G. Bianchi, and W. H. Birkenhäger, "Essential hypertension," *The Lancet*, vol. 361, no. 9369, pp. 1629–1641, 2003.

[31] G. Parati, G. Stergiou, and E. O'Brien, "Cardiovascular variability, european society of hypertension working group on blood pressure monitoring. european society of hypertension position paper on ambulatory blood pressure monitoring," *J Hypertens*, vol. 31, no. 9, pp. 1731–1768, 2013.

[32] C.-E. Chiang, T.-D. Wang, K.-C. Ueng, T.-H. Lin, H.-I. Yeh, C.-Y. Chen, Y.-J. Wu, W.-C. Tsai, T.-H. Chao, C.-H. Chen, *et al.*, "2015 guidelines of the taiwan society of cardiology and the taiwan hypertension society for the management of hypertension," *Journal of the Chinese Medical Association*, vol. 78, no. 1, pp. 1–47, 2015.

[33] NICE - National Istitute for Health and Care Excellence, "Hypertension in adults: diagnosis and management." `https://www.nice.org.uk/guidance/ng136/resources/hypertension-in-adults-diagnosis-and-management-pdf-66141722710213`.

[34] K. Kario, "Management of hypertension in the digital era: small wearable monitoring devices for remote blood pressure monitoring," *Hypertension*, vol. 76, no. 3, pp. 640–650, 2020.

[35] T. Arakawa, "Recent research and developing trends of wearable sensors for detecting blood pressure," *Sensors*, vol. 18, no. 9, p. 2772, 2018.

[36] S. Rastegar, H. GholamHosseini, and A. Lowe, "Non-invasive continuous blood pressure monitoring systems: Current and proposed technology issues and challenges," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 1, pp. 11–28, 2020.

[37] A. Noordergraaf, *Circulatory System Dynamics*. Biophysics and Bioengineering, Elsevier Science, 2012.

[38] P. I. Aaronson, J. P. Ward, and M. J. Connolly, *The cardiovascular system at a glance*. John Wiley & Sons, 2020.

[39] R. N. Mitchell and F. J. Schoen, "Blood vessels," *Robbins and Cotran: Pathologic Basis of Disease.(8th edition) Saunders Elsevier, Philadelphia, US*, pp. 516–17, 2010.

[40] Wikipedia — Circulatory system. `https://en.wikipedia.org/wiki/Circulatory-system`.

[41] Wikipedia — Blood . `https://en.wikipedia.org/wiki/Blood`.

[42] C. B. Higgins, S. F. Vatner, and E. Braunwald, "Parasympathetic control of the heart," *Pharmacological reviews*, vol. 25, no. 1, pp. 119–155, 1973.

[43] M. Meo, *Spatio-temporal characterization of the surface electrocardiogram for catheter ablation outcome prediction in persistent atrial fibrillation*. PhD thesis, 12 2013.

[44] Wikipedia — Heart Wall . `https://en.wikipedia.org/wiki/Heart#Heart_wall`.

[45] Wikipedia — Pericardium . `https://en.wikipedia.org/wiki/`

Pericardium-Functions.
[46] Wikipedia —Cardiac cycle . https://en.wikipedia.org/wiki/Cardiac_cycle.
[47] N. R. Colledge, B. R. Walker, and S. H. Ralston, *Davidson's principles and practice of medicine*. Elsevier Limited, 2010.
[48] J. G. Betts, *Anatomy & physiology*. 2013.
[49] M. AlGhatrif and J. Lindsay, "A brief review: history to understand fundamentals of electrocardiography," *Journal of Community Hospital Internal Medicine Perspectives*, vol. 2, no. 1, p. 14383, 2012. PMID: 23882360.
[50] D. M. Mirvis and A. L. Goldberger, "Electrocardiography," *Heart Disease. A Textbook of Cardiovascular Medicine, 6th ed. Philadelphia: WB Saunders*, pp. 82–126, 2001.
[51] H. Nagendra and S. Mukherjee, "Application of wavelet techniques in ecg signal processing: an overview,"
[52] Wikipedia — Electrocardiograph. https://en.wikipedia.org/wiki/Electro-cardiography.
[53] Humanitas — Elettrocardiogramma. https://www.humanitas.it/visite-ed-esami/elettrocardiogramma/.
[54] L. M. Miller and A. Gal, "Cardiovascular system and lymphatic vessels," *Pathologic basis of veterinary disease*, p. 561, 2017.
[55] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.
[56] Wikipedia — Photoplethysmogram. https://en.wikipedia.org/Photoplethysmogram, 2021.
[57] J. Lee, K. Matsumura, K.-i. Yamakoshi, P. Rolfe, S. Tanaka, and T. Yamakoshi, "Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 1724–1727, IEEE, 2013.
[58] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *International journal of biosensors & bioelectronics*, vol. 4, no. 4, p. 195, 2018.
[59] D. e. a. Castaneda, "A review on wearable photoplethysmography sensors and their potential future applications in health care.," *International journal of biosensors & bioelectronics*, vol. 4, p. 7, 2018.
[60] J. Peck, M. Wishon, H. Wittels, F. Hasty, S. Hendricks, S. Lee, and S. Wittels, "Covid-19 causes changes in photoplethysmography," 2021.
[61] J. Peck, M. Wishon, H. Wittels, F. Hasty, S. Hendricks, S. Lee, and S. Wittels, "Covid-19 causes changes in photoplethysmography," 2021.
[62] U. Guler, I. Costanzo, and D. Sen, "Emerging blood gas monitors: How they can help with covid-19," *IEEE Solid-State Circuits Magazine*, vol. 12, no. 4,

pp. 33–47, 2020.

[63] D. Seok, S. Lee, M. Kim, J. Cho, and C. Kim, "Motion artifact removal techniques for wearable eeg and ppg sensor systems," *Frontiers in Electronics*, vol. 2, p. 4, 2021.

[64] B. McCarthy, B. O'Flynn, and A. Mathewson, "An investigation of pulse transit time as a non-invasive blood pressure measurement method," in *Journal of Physics: Conference Series*, vol. 307, p. 012060, IOP Publishing, 2011.

[65] Wikipedia — Pulse Wave Velocity. `https://en.wikipedia.org/wiki/Pulse-wave-velocity`.

[66] Wikipedia — Blood pressure. `https://en.wikipedia.org/wiki/Blood_pressure`.

[67] "Physiology, pulse pressure.." `https://www.ncbi.nlm.nih.gov/books/NBK482408/`.

[68] R. Klabunde, *Cardiovascular physiology concepts*. Lippincott Williams & Wilkins, 2011.

[69] "Venous blood pressure." `https://med.libretexts.org/@go/page/7858`.

[70] NHS. `https://www.nhs.uk/common-health-questions/lifestyle/what-is-blood-pressure/`.

[71] J. Booth, "A short history of blood pressure measurement," 1977.

[72] K. RE, "Cardiovascular physiology concepts —- mean arterial pressure." `https://www.cvphysiology.com/Blood%20Pressure/BP006.html`.

[73] "Healthline — understanding mean arterial pressure.." `https://www.healthline.com/health/mean-arterial-pressure`.

[74] weNurse24.

[75] J. J. W. Group *et al.*, "Guidelines for the clinical use of 24 hour ambulatory blood pressure monitoring (abpm)(jcs 2010)," *Circulation Journal*, vol. 76, no. 2, pp. 508–519, 2012.

[76] Omron, "Heartguide." `https://www.omron-healthcare.it/it/misuratori-di-pressione/heartguide.html#prefn1=device_type&prefv1=Misuratore-%20di%20pressione%20arteriosa%20indossabile&start=1`.

[77] E. O'brien, R. Asmar, L. Beilin, Y. Imai, J.-M. Mallion, G. Mancia, T. Mengden, M. Myers, P. Padfield, P. Palatini, *et al.*, "European society of hypertension recommendations for conventional, ambulatory and home blood pressure measurement," *Journal of hypertension*, vol. 21, no. 5, pp. 821–848, 2003.

[78] E. Commission. `https://knowledge4policy.ec.europa.eu/health-promotion-knowledge-gateway/cardiovascular-diseases-prevention_en`.

[79] A. D. Lopez, C. D. Mathers, M. Ezzati, D. T. Jamison, and C. J. Murray, "Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data," *The lancet*, vol. 367, no. 9524, pp. 1747–1757, 2006.

[80] Wikipedia, "Hypertension." https://en.wikipedia.org/wiki/Hypertension.

[81] P. Verdecchia, J. A. Staessen, W. White, Y. Imai, and E. O'Brien, "Properly defining white coat hypertension," *European heart journal*, vol. 23, no. 2, pp. 106–9, 2002.

[82] S. S. Franklin, L. Thijs, T. W. Hansen, E. O'brien, and J. A. Staessen, "White-coat hypertension: new insights from recent studies," *Hypertension*, vol. 62, no. 6, pp. 982–987, 2013.

[83] T. G. Pickering, K. Eguchi, and K. Kario, "Masked hypertension: a review," *Hypertension Research*, vol. 30, no. 6, pp. 479–488, 2007.

[84] G. Bobrie, P. Clerson, J. Ménard, N. Postel-Vinay, G. Chatellier, and P.-F. Plouin, "Masked hypertension: a systematic review," *Journal of hypertension*, vol. 26, no. 9, pp. 1715–1725, 2008.

[85] J. P. O'HARE and W. G. WALKER, "ARTERIOSCLEROSIS AND HYPERTENSION," *Archives of Internal Medicine*, vol. 33, pp. 343–349, 03 1924.

[86] Heart.org. https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings.

[87] C. Clinic. https://my.clevelandclinic.org/health/diseases/16459-renal-hypertension.

[88] Humanitas — Ipertensione. https://www.humanitas.it/malattie/ipertensione/.

[89] G. S. Stergiou, B. Alpert, S. Mieke, R. Asmar, N. Atkins, S. Eckert, G. Frick, B. Friedman, T. Graßl, T. Ichikawa, *et al.*, "A universal standard for the validation of blood pressure measuring devices: Association for the advancement of medical instrumentation/european society of hypertension/international organization for standardization (aami/esh/iso) collaboration statement," *Hypertension*, vol. 71, no. 3, pp. 368–374, 2018.

[90] X.-D. Zhang, "Machine learning," in *A Matrix Algebra Approach to Artificial Intelligence*, pp. 223–440, Springer, 2020.

[91] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[92] K. Worden and G. Manson, "The application of machine learning to structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 515–537, 2007.

[93] K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research-commentary," *Biomedical engineering online*, vol. 13, no. 1, pp. 1–9, 2014.

[94] A. Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, and G. Hartvigsen, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes," *Artificial intelligence in medicine*, vol. 98, pp. 109–134, 2019.

[95] H.-P. Chan, R. K. Samala, and L. M. Hadjiiski, "Cad and ai for breast cancer—recent development and challenges," *The British journal of radiology*, vol. 93, no. 1108, p. 20190580, 2019.

[96] J. C. Weiss, S. Natarajan, P. L. Peissig, C. A. McCarty, and D. Page, "Machine learning for personalized medicine: predicting primary myocardial infarction from electronic health records," *Ai Magazine*, vol. 33, no. 4, pp. 33–33, 2012.

[97] L. O. *Journal of human hypertension*, pp. 865—-871, 1994.

[98] L. J. Appel and W. B. Stason, "Ambulatory blood pressure monitoring and blood pressure self-measurement in the diagnosis and management of hypertension," *Annals of internal medicine*, vol. 118, no. 11, pp. 867–882, 1993.

[99] "Business insider - latest trends in medical monitoring devices and wearable health technology." `https://www.businessinsider.com/wearable-technology-healthcare-medical-devices?r=US&IR=T`.

[100] A. V. Chobanian, G. L. Bakris, H. R. Black, W. C. Cushman, L. A. Green, J. L. Izzo Jr, D. W. Jones, B. J. Materson, S. Oparil, J. T. Wright Jr, *et al.*, "The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the jnc 7 report," *Jama*, vol. 289, no. 19, pp. 2560–2571, 2003.

[101] C. Poon and Y. Zhang, "Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time," in *2005 IEEE engineering in medicine and biology 27th annual conference*, pp. 5877–5880, IEEE, 2006.

[102] P. Su, X.-R. Ding, Y.-T. Zhang, J. Liu, F. Miao, and N. Zhao, "Long-term blood pressure prediction with deep recurrent neural networks," in *2018 IEEE EMBS International conference on biomedical & health informatics (BHI)*, pp. 323–328, IEEE, 2018.

[103] M. C. Khoo, *Physiological control systems: analysis, simulation, and estimation*. John Wiley & Sons, 2018.

[104] W. Chen, T. Kobayashi, S. Ichikawa, Y. Takeuchi, and T. Togawa, "Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration," *Medical and Biological Engineering and Computing*, vol. 38, no. 5, pp. 569–574, 2000.

[105] H. Gesche, D. Grosskurth, G. Küchler, and A. Patzak, "Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method," *European journal of applied physiology*, vol. 112, no. 1, pp. 309–315, 2012.

[106] B. McCarthy, C. Vaughan, B. O'flynn, A. Mathewson, and C. Ó. Mathúna, "An examination of calibration intervals required for accurately tracking blood pressure using pulse transit time algorithms," *Journal of human hypertension*, vol. 27, no. 12, pp. 744–750, 2013.

[107] K. Soerensen, A. K. Verma, A. Blaber, J. Zanetti, S. E. Schmidt, J. J. Struijk,

and K. Tavakolian, "Challenges in using seismocardiography for blood pressure monitoring," in *2017 Computing in Cardiology (CinC)*, pp. 1–4, IEEE, 2017.

[108] S. Ghosh, A. Banerjee, N. Ray, P. W. Wood, P. Boulanger, and R. Padwal, "Continuous blood pressure prediction from pulse transit time using ecg and ppg signals," in *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*, pp. 188–191, 2016.

[109] R. Lazazzera, Y. Belhaj, and G. Carrault, "A new wearable device for blood pressure estimation using photoplethysmogram," *Sensors*, vol. 19, no. 11, p. 2557, 2019.

[110] B. Zhang, H. Ren, G. Huang, Y. Cheng, and C. Hu, "Predicting blood pressure from physiological index data using the svr algorithm," *BMC bioinformatics*, vol. 20, no. 1, pp. 1–15, 2019.

[111] G. A. Seber and A. J. Lee, *Linear regression analysis*, vol. 329. John Wiley & Sons, 2012.

[112] D. M. Lane, D. Scott, M. Hebl, R. Guerra, D. Osherson, and H. Zimmer, *Introduction to statistics*. Citeseer, 2017.

[113] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-Wei, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[114] T. Bhowmik, J. Dey, and V. N. Tiwari, "A novel method for accurate estimation of hrv from smartwatch ppg signals," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 109–112, IEEE, 2017.

[115] A. Fettweis, "Wave digital filters: Theory and practice," *Proceedings of the IEEE*, vol. 74, no. 2, pp. 270–327, 1986.

[116] A. de Cheveigné and I. Nelken, "Filters: when, why, and how (not) to use them," *Neuron*, vol. 102, no. 2, pp. 280–293, 2019.

[117] A. E. Aubert, B. Seps, and F. Beckers, "Heart rate variability in athletes," *Sports medicine*, vol. 33, no. 12, pp. 889–919, 2003.

[118] J. I. Daoud, "Multicollinearity and regression analysis," in *Journal of Physics: Conference Series*, vol. 949, p. 012009, IOP Publishing, 2017.

[119] "Wikipedia - ridge regression." `https://en.wikipedia.org/wiki/Ridge_regression`.

[120] A. E. Hoerl and R. W. Kennard, "Ridge regression: applications to nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 69–82, 1970.

[121] Q. T. Nguyen, R. Fouchereau, E. Frenod, C. Gerard, and V. Sincholle, "Comparison of forecast models of production of dairy cows combining animal and diet parameters," *Computers and Electronics in Agriculture*, vol. 170, p. 105258, 2020.

[122] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[123] A. King and R. Eckersley, "Inferential statistics iii: nonparametric hypothesis testing," *Statistics for Biomedical Engineers and Scientists, AP King and RJ Eckersley, Eds*, pp. 119–145, 2019.

[124] D. W. Marquardt and R. D. Snee, "Ridge regression in practice," *The American Statistician*, vol. 29, no. 1, pp. 3–20, 1975.

[125] D. G. Altman and J. M. Bland, "Standard deviations and standard errors," *Bmj*, vol. 331, no. 7521, p. 903, 2005.

[126] "Wikipedia — Cross Validation (statistics)." `https://en.wikipedia.org/wiki/Cross-validation_(statistics)`.

[127] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.