POLITECNICO DI TORINO

DEPARTMENT OF MECHANICAL AND AEROSPACE ENGINEERING

Master Degree Course in Biomedical Engineering

Master Degree Thesis

A machine learning approach for non-invasive blood pressure estimation

Supervisors
Prof. Danilo Demarchi
Prof. Guido Pagana

Candidate
Andrea Tiloca

March 2020

Abstract

The increase in the production and employment of wearable devices in everyday life, as well as the improvement of their refinement and precision, encourage the development prospects of this market in the future. The wearable market amount is estimated to reach 57,653 million dollars in 2022. This is due to their continuous increase in easy to use, flexibility, convenience and the possibility of returning the collected data in real-time. Nevertheless, there are still limits, such as the life of their battery, which is holding back their position on the market.
On the other hand, the study of new methods for estimating vital parameters with non-invasive techniques is encouraged by the increased computational capacity of these devices and by their facility to be integrated with complex algorithms, such as machine learning algorithms. In fact, the objective of this thesis work consists in the use of a wearable device for blood pressure measurements.
One of the biggest causes of death around the growing world in recent decades are cardiovascular disease (CVD). The number of people suffering from it increases year by year and one of the risk factors that compete with the appearance of the pathology is chronic hypertension, characterized by high variations in pressure over prolonged periods of time.
Therefore, estimating blood pressure is pivotal in order to monitor the health status of the population. Currently the measurement are mainly performed through auscultatory devices such as the sphygmomanometer, which can be digital or classic, requiring in the second case the presence of a trained operator, and they both require static conditions for the subjects.
Encouraged by this, with the increase in the refinement of wearable devices and the implementation of machine learning algorithms, it has been explored other physiological parameters that allow to measure blood pressure via indirect measurement. An example is the Pulse Wave Velocity (PWV), in which the variation in the propagation velocity of the pressure wave generated by the passage of blood in the vessels is observed. PWV, together with the variation of vessels diameters, can be therefore used to estimate pressure in a specific observation window, that in the proposed study consists in 10 seconds.
In this thesis work, a model was implemented to estimate pressure through the collection of data (ECG, PPG and ABP signals) taken from online databases (MIMIC

data).
àFrom electrocardiogram (ECG) and photoplethysmography (PPG) signals the best morphological or temporal characteristics are identified in order to train a linear regression algorithm that allows to estimate blood pressure values.
The results are achieved by implementing the Random Forest as a linear regression model and by splitting the data-set into training set and test set. A quadratic error and a standard deviation of approximately 8 ± 10 mmHg for SBP, 6 ± 9 mmHg for DBP and 7 ± 9 mmHg for MAP are obtained. Finally, the system was validated through 10-fold cross-validation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Wearable technologies application

Nowadays wearable technologies represent one of the pillars of the electronic devices market. Their market is estimated to reach 57,653 millions of dollar in 2022 [1]. The main slice of this market is represented by devices related to fitness and well-being (Figure 1.1).

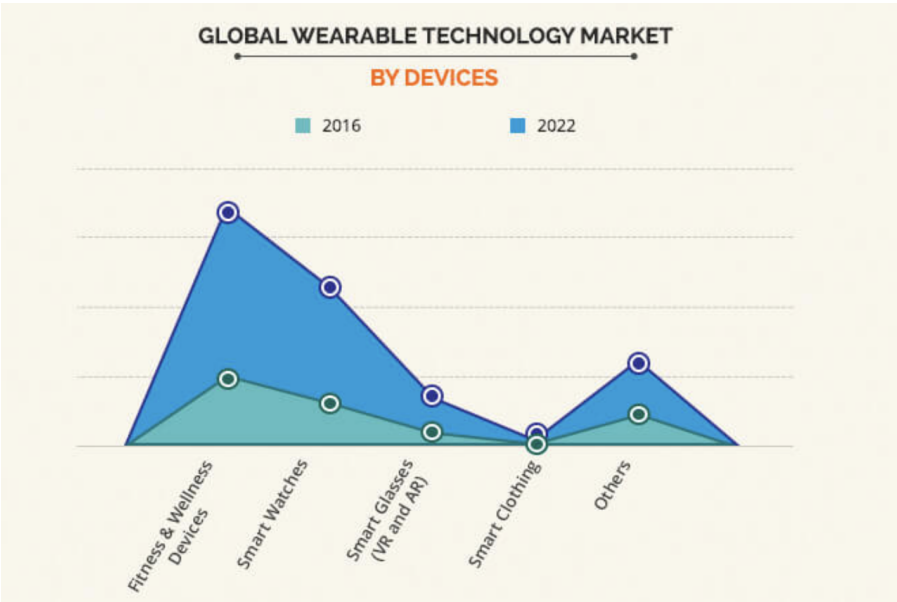

Figure 1.1: Principal market areas of wearable devices [1].

The increase in chronic diseases and cases of obesity simultaneously to the increase in popularity of these devices have led to a diffused use of wearable devices,

such as activity and movement trackers.
This family of devices can provide information about various vital parameters by returning daily reports on the health status of those who use them.
With wearable it is intended a technology device "that can be worn, such as clothing or glasses, that contain computer technology or can connect to the internet" [3]. These devices can reach a computing capacity similar to the ones of laptops or mobile phones. In addiction, wearable devices have the possibility of incorporating sensors capable of taking biofeedback and monitoring physiological functions. Moreover, they allow the creation of protocols for monitoring the condition of the subject remotely while performing normal activities at home or at work.
The diffusion of these technologies allows to have the availability of huge amounts of data, even outside the environments in which they could usually be obtained, such as the heart rate (HR) or the cadence of the user's pace or the levels of stress. This encourages to search new methods for estimating important vital parameters by minimizing the footprint with which they are obtained and by limiting the awareness of the user, while improving the quality of the measurements.
In addiction, new data communication systems through the Internet and the possibility of storing them to online archives, such as the Cloud, must be considered. This new family of devices has taken the name of Internet of Things (IoT), which has encouraged researchers to seek solutions in which different devices communicate directly with the internet by collecting data of different origin. These new wearable devices are forming a new market segment called: Wearable IoT, for their ability to integrate sensing, calculation and communication [6].

On the other hand, the increase in the use of medical wearable devices poses new challenges on the management of this enormous amount of sensitive data, with the consequent need to place stakes that allow the protection of consumers through their awareness of their correct use. An example may be the case of fitness tracking Strava © App, in which it was found by the Pentagon that through that application it was possible to identify the movements of US military forces in Syria [4].

1.2 Objective of the work

One of the most important parameters to be monitored and measured in a non-invasive way, but at the same time in a serious and accurate way, is blood pressure (BP). BP is mainly measured to control patients in serious health conditions such as hypertension or following a heart attack.
In addition to clinical practices, BP control is a useful parameter to monitor also during sports practices, to monitor progress and the possible presence of pathologies not yet identified. This is because during physical activity, although an increase of more than 200 mmHg in the maximum pressure is tolerable for those who do sports,

an increase of about 15 mmHg for the minimum pressure can be an indication of coronary heart disease [2].
Classically BP is measured through the sphygmomanometer, a medical device consisting of a sleeve in which air is blown through a pump and a stethoscope to auscultate the heart beat, to understand when the flow through the arm (where the measurement typically takes place) is interrupted after inflating the cuff. This measurement method requires training by the person who performs it and it entails the impossibility of carrying the measurements out while the subject is in motion. For this reason, one of the topics on which researchers are focusing is the estimation of pressure through different methods. This thesis work has the goal to estimate BP through the study of the correlation between Pulse Wave Velocity (PWV) and BP variations.
More precisely, two signals are needed to do this, such as ECG and PPG, without the need for other means or devices. Thanks to this system it is possible to carry out the measurement in a simpler and safer way for the subject and giving information on his health, the risk of heart attacks and the stress conditions following physical activity.
Given the potential market in which this technology can be used, it is useful to invest both in the realization of medical devices and in the implementation of devices usable by simple consumer, who can use them without the intervention of a professional medical figure.
This is also due to the strong restrictions on the legislation that regulates the accuracy and reliability of the measure that a device must have in order to be considered a medical device.
This study focuses on the realization of an algorithm for processing the signals and the data obtained from them through computational tools such as Matlab in order to estimate BP in a non-invasive way through the study of PWV. Online databases such as MIMIC II and MIMIC III, from the free access portal PhysioNet [62], are used for the realization of the model, since a large enough database from devices was not available. This work shows the feasibility of a new method of measuring BP, adopting strategies that can make it automatic and easy to perform.

The implemented algorithm is based on the following steps:

- Initial analysis and feasibility study: in the first phase of the research the problem was analyzed, the feasibility of the method that was chosen to implement was investigated and the state of the art was studied. In particular, the following researches were made:

    1. Physiology studies, in particular regarding the cardiovascular system and blood pressure.
    2. Studies on the role and useful information of the signals involved such as ECG, PPG and BP.

3. Research in literature on works concerning the estimation of BP through these signals.

- Choice of sources from which to obtain the signals: in this phase the MIMIC II and MIMIC III online databases were chosen for their immediate availability and the unnecessary implementation of denoising or signal alignment algorithms.

- Implementation of the model through Matlab tools: in which it is implemented a mathematical model for the first signals elaboration and analysis, followed by the feature's extraction (like the PTT, the PIR etc. used for their correlation with BP) and finally by the realization of a machine learning regression algorithm based on Random Forest model for BP estimation.

- Validation of the algorithm: the last step regards the validation of the created model in order to establish its level of reliability and the quality of the reached results. This stage is implemented with the k-fold cross-validation method and the hypothesis test.

Chapter 2

Physiological background

2.1 Heart physiology

The heart is a muscle that generates the force necessary to circulate blood in the blood vessels [10], situated in the thoracic cavity, just above the diaphragm, which separates the thoracic cavity from the abdominal ones. Anatomically and functionally, the heart is divided into right and left portions, that are separates by a longitudinal wall called septum, which prevents the blood in the two chambers from mixing. Within these two areas, a further wall separates the upper and lower portions, forming a total of four cavities with different sizes: two upper, the atria, which receive the venous blood returning from the circulation and that perform a contraction action to push the blood into the two lower cameras, and the ventricles, which receive the blood from the atria and generate the pressure necessary to push the blood out of the heart into the large arteries [10].
The contraction of the heart conveyed by the transmission of electrical impulses generated by pacemaker cells and transmitted through the muscles of the myocardium is called myogeny and is independent of the commands of the central nervous system. The conduction system causes a conduction wave that guides the impulse first through the atria, which contract simultaneously as a unit, and subsequently to the ventricles, ending with their depolarization and then contraction, putting blood in the blood vessels. The heart then works as two intermittent pumps that support the circulation of the blood. To do this, the two pumps must remain synchronized [7]. Cardiac contraction is called systole, while the release phase is called diastole.

The action potential is an event supported by ionic currents that depolarize and repolarize the cell [7]:

- the repolarization phase is generated by a progressive reduction of the passage inside the cell of Sodium and Calcium ions and by an output current of positive charges (K+ ions) that make the inside of the cell negative. This generates

the formation of the resting potential, in which the amount of K+ ions inside the cell is lower than that of Na+ ions outside;

- the depolarization phase is characterized by an inversion of polarization and therefore by a massive entry of positive charges inside the cell. Calcium ions go into the cell generating an action potential. The impulse from the sinoatrial node (SA), located on the outer wall of the right atrium, reaches the atrioventricular node (AV) located above the plane of the tricuspid valve, where conduction slows down. The contraction of the atria occurs before the ventricular one and it allows the atrium to fill the left ventricular chamber. From the AV node, the His beam originates running along the wall of the interventricular septum. Reaching the distal portions, it gives rise to two branches, the right branch and the left one. These branches spread into the peripheral system of conduction (Purkinje fibers) which allows the almost simultaneous transmission of the pulse (depolarization waves) to the endocardium of the two ventricles [7].

The heart cycle is the period defining the time interval between a systole and the next one. It consists of four phases [10]:

1. ventricular filling: the blood reaches the ventricles thanks the opening of the Atrioventricular valves between the atria and the ventricles. Since the heart is relaxed, the diastolic phase occurs. The increase of the ventricular volume at the end of the filling phase determines the closure of the valves, marking the beginning of the systolic phase;

2. isovolumetric contraction: in this phase all the valves are closed. The pressure inside the ventricle rises rapidly due to isometric contraction. When the value of the ventricular pressure exceeds that of the aorta, the semi-moon valves open;

3. ejection: with the opening of the semi-moon valves, the blood is expelled from the contracting ventricle. Ventricular pressure continues to rise (because it continues to contract) while volume decreases rapidly;

4. iso-volumetric release: once the aortic valve is closed, the pressure inside the ventricular chamber decreases.

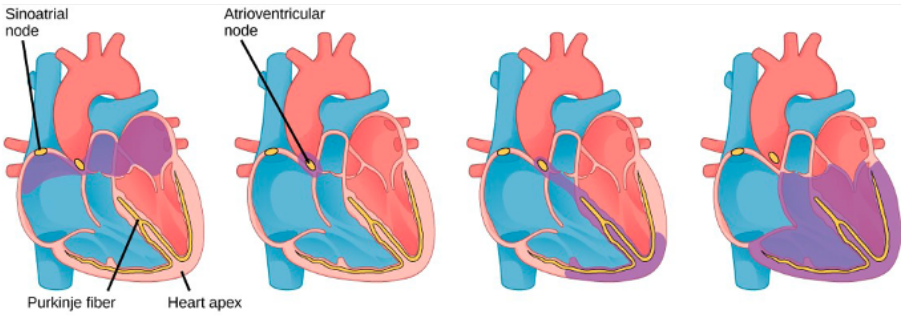

Figure 2.1: Representation of heart cycle [8].

The cardiovascular system, consisting of the heart and blood vessels, is in charge of releasing nutrients and oxygen to the body cells.

The blood vessels can be distinguished in:

- arteries: vessels that transport nutrients from the heart to the tissues. They have a thickness of the walls and a diameter of the relatively large vases [10]. One of the main characteristics of the larger arteries is to offer little resistance to the passage of blood so they have a task mainly of conducting the flow;

- veins: which have about the same diameter as the arteries but half the thickness of the walls. Their thickness is justified by the need for veins to have significantly lower blood pressure than that in the arteries. Their job is to bring venous blood to the heart.

Finally, there are two types of blood circulation [9]:

- the systemic ones connect the heart to all tissues of the body. The blood moves along the arteries, according to decreasing diameters. In the capillary network, it gives oxygen and nutrients to the tissues. Recharging itself with waste substances and carbon dioxide, the blood reaches the right ventricle through the veins;

- the pulmonary circulation connects the heart to the lungs. The blood rich in carbon dioxide moves from the right ventricle to the pulmonary artery. Latter branches out into smaller vessels to form a network around the pulmonary vesicles. Here, the blood recharges with oxygen, releasing carbon dioxide.

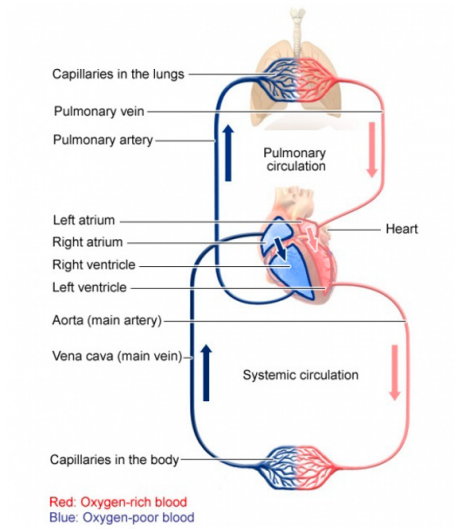

Figure 2.2: Representation of the two type of blood circulation.

2.2 Electrocardiogram

The electrocardiogram (ECG) is the recording and reproduction of the electrical and chemical activity of the cardiac muscle fibers during the cardiac cycle [10]. It is a registration of bio-potential produce by continuous depolarization and repolarization of cardiac muscle fibers during the cardiac cycle.
It can provide information useful to establish the condition of the patient, based on the analysis on the intervals it is composed of.
Specifically, it is formed by the following waves and segments [65]:

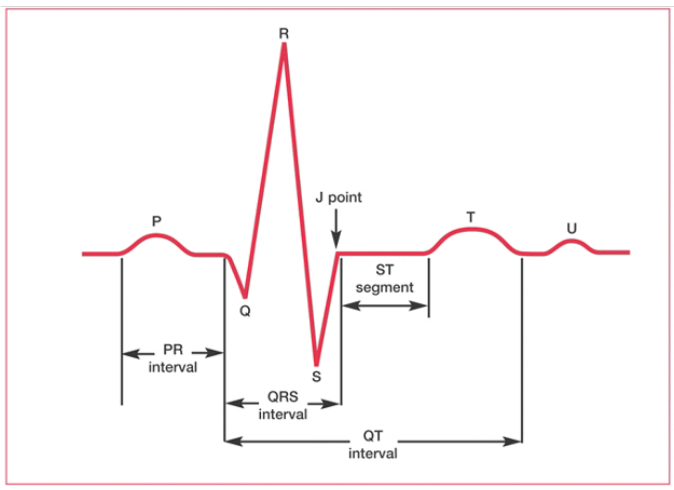

Figure 2.3: Electrical activity pattern of the heart [67].

- the P wave is a deflection wave that represents the depolarization phase of the atria [10], as previously seen, the depolarization of the atria starts from the atrial sinus node. However, there are not enough cells to produce a depolarization important enough to be detected by the sampling system;
- the QRS complex is generated after atrial depolarization and consists of 3 waves constituting the ventricular depolarization;
  - the Q wave is usually thin and small and is related both to the depolarization of the interventricular septum and to the respiration. They give information about the myocardial infarction;
  - the R wave is the largest because it reflects the depolarization activity of a large portion of the ventricle;
  - the S wave is generated by the final action of ventricle depolarization [10]. When the complex is ended the ventricular chambers are completely depolarized and the atriums are repolarized;
- finally, the T wave which represents the repolarization phase of the ventricle;
- the PR interval is the time measured between the atrial depolarization and the followed ventricular ones;
- the ST segment is the portion between the end of the ventricular depolarization, with the end of QRS complex, and their repolarization.

Electrodes placed on the skin are typically used to record the ECG signal, the difference in potential due to the travel of these signals along the body is measured. The differences with which the signal occurs are not only due to the intensity of the contraction of the heart muscle but also to the position of the electrodes on the subject's skin. In fact, the waveform of the ECG appears in different shapes based on the position as well as on the number of electrodes that perform the sampling. For this reason, standard configurations have been introduced with respect to which to take the signal.

Bipolar Limb

The most common configuration is that which respects the structure of the Einthoven triangle (which has the name of the German scientist who described it first), in which an imaginary triangle is formed that widens up to reach its apexes at the hands and at the left foot. The triangle may not be so extensive but also more concentrated in the chest area only. The electrodes that make up the corners are connected to a device that measures the voltage. The potential difference of each pair of electrodes is measured and the resulting signal defines a derivation [10]. Any potential difference is defined as Leads [12]:

- Lead I: obtained as the difference between the electrode potential towards the left arm (LA) and the right arm (RA):

$$I = LA - RA; \tag{2.1}$$

- Lead II: obtained as the difference between the electrode potential towards the left foot (LL) and the right arm (RA):

$$II = LL - RA; \tag{2.2}$$

- Lead III: relating to the voltage difference between the electrode of the left foot (LL) and the left arm (LA):

$$III = LL - LA; \tag{2.3}$$

In addition to this configuration there is the unipolar, in which the limb electrodes refer to a single reference electrode and the precordial leads, also a unipolar configuration in which six electrodes are placed on the lungs that measure the electrical activity on the transverse the front one. In particular, the latter configuration allows to study with particular accuracy the possible damages following a damage thanks to the proximity to the cardiac chambers.

2.3 Blood Pressure

The blood flowing inside their vessels generates a pressure on the walls proportional to the force that the heart uses to pump the blood into the vessels, which is called blood pressure. In general, the law that best describes the phenomenon is the flow law [10], from which the flow of the fluid is obtained as the ratio between the pressure gradient, with it is directly proportional, and the resistance to the flow, with which it is inversely proportional.

$$Flow = \frac{\Delta P}{R} \tag{2.4}$$

The pressure gradient is the force that pushes the blood along the vessels while resistance is the set of factors that obstruct the flow.
As shown in Figure 2.4 the beginning of the systemic circulation, the average pressure in the aorta throughout the cardiac cycle is about 80-120 mmHg, while in the large veins of the thoracic cavity, that bring the blood back into the heart chambers, is between 2-8 mmHg and about 0 mmHg in the vena cava. This pressure difference allows us to assume that the pressure gradient that allows blood to circulate is virtually equal to the average pressure.

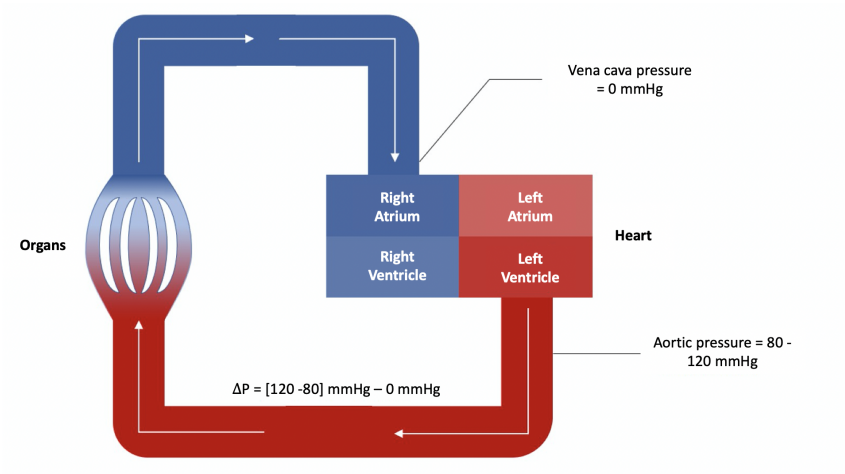

Figure 2.4: BP gradient which guides the blood flow.

The resistance understood as the degree with which the flow in the vessel is delayed is influenced by several factors such as:

- radius of the vessel, which affects resistance more than any other. A decrease in the radius of the vessel is called vasoconstriction while an increase in vasodilation;
- length of the pot, in which a greater length lends itself to an increase in the resistance to the flow. Nonetheless, the blood vessels do not vary in length except in periods of growth;
- blood viscosity, as a more viscous fluid increases the friction within the fluid itself. It is influenced by the presence of figured elements of the blood and plasma proteins. However, their variation is a slow process.

The mathematical aspects linked to this phenomenon are expressed by Poiseuille's law, in which resistance is obtained as:

$$R = \frac{8t\eta}{\pi r^4} \tag{2.5}$$

where η is the viscosity of the fluid flowing through the tube, with length L and radius r. As it can be see, the component that most influences the resistance is the radius of the tube, which assumes a prevalent importance.

Depending on the type of vessel in which it is measured, this is called venous pressure or arterial pressure. The tone of the venous pressure is lower than the arterial one [10], since in this tract the blood does not receive the cardiac thrust. For this reason, the measurement of venous pressure is carried out only in the case of particular pathologies [13].

Arterial Pressure

The role of the arteries is to lead blood from the heart to the tissues. This is made possible by the thickness of their walls, together with the presence of elastic fabric which gives it a certain rigidity and the ability to expand and contract following each ventricular contraction, which allows it to be defined as a pressure reservoir [10].

The arteries are structurally organized in 4 different layers: [40]:

1. Endothelium, which serve like a wall for blood flow inside the vessels. It has superficial importance for the mechanical properties of the arteries;

2. Elastin which sizable elastic properties that can produce tension on the walls. It is responsible of elastic response of the arterial vessels at low BP values [40];

3. Collagen, stiffer than the elastic layer and exerts more tension when the arterial vessel became more stretched. It gives an important contribution to the arterial wall property for high BP values [40];

4. Smooth-muscle to modulate the elasticity of the arteries producing tension [41].

Disease like aging contributes changes mechanical properties of arterial walls strongly and this can be effects to cardio-circular performance.

When the arterial walls expand during systole the elastic fibers expand like a spring, while when there is no more blood flow during the diastole the walls passively retract by elastic recall, allowing the blood to continue to circulate also in this phase.

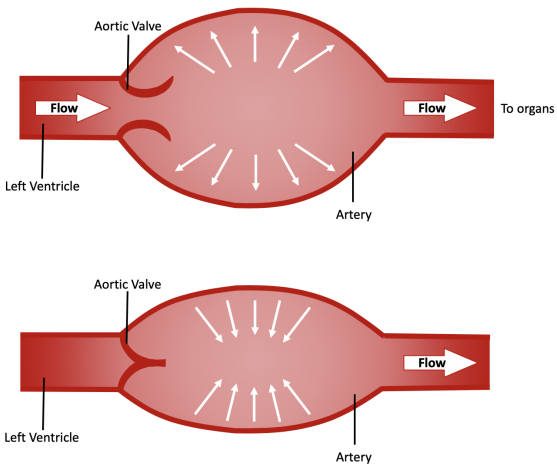

Figure 2.5: Aretery like pressure tank.

The blood pressure depends on the blood flow at a given time: an increase in blood volume causes an increase in blood pressure. Since the blood is introduced from the heart into the large arteries during systole, the arteries will contain more blood in the systolic period than in the diastolic one. Therefore, four arterial pressure values can be distinguished, all resulting from the same periodic phenomenon. Each one, however, depends on specific hemodynamic parameters [15]:

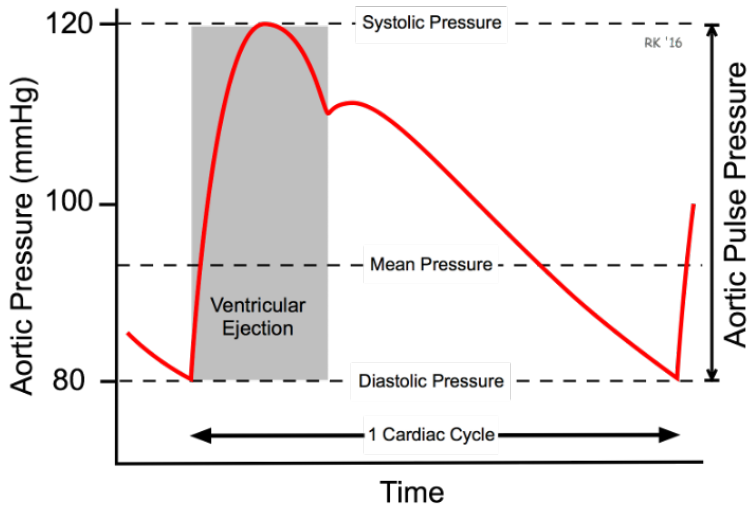

Figure 2.6: Representation in time of arterial blood pressure [16].

Mean atrial pressure (MAP), is the pressure that a continuous and non-pulsatile flow system would have, to keep the blood flow constant. It depends on cardiac output and vascular resistance, which in turn are determined by vessel size and by the number of small arteries and arterioles.

- *Diastolic blood pressure* (DBP), also called minimum pressure. It is generated when the heart is not contracted but released. Its value is always lower than the systolic one [13]. The hemodynamic parameters that influence this measure are the peripheral resistance to the sliding generated by the small arterioles, the rigidity of the arteries and the duration of the diastolic phase [57].

- *Systolic blood pressure* (SBP), also called maximum pressure, is generated by the flow of oxygen-rich blood into the vessels during the contraction phase of the heart. It depends on the left ventricular ejection rate, vascular resistance and peripheral reflection waves. Indeed, if the cardiac output increases, the systolic blood pressure augments contrary to the diastolic one, which does not undergo significant variations. An increase in peripheral resistance correspond to a systolic pressure increasing, but also to a more marked increase in the diastolic one. The decrease in arterial compliance increases systolic pressure due to two mechanisms: the decrease in the damping capacity of the systolic wave

as a result of the more rigid arterial walls, and the earlier arrival of peripheral reflection waves that are added to the incident wave generated by the left ventricle. Changes in the characteristics of the reflected waves (amplitude of the reflected wave and location of the reflection site) cause some modifications of the arterial pressure, especially in the central arteries [57].

- Pulse pressure represents the aortic pressure pulse during systole. In fact, the aorta undergoes a great variation in pressure during the ejection phase, due to the high blood flow that passes through its vessel walls [17]. Pulsatory pressure, also called differential pressure, is defined as the variation of blood pressure values around the mean pressure values. It can be measured through the difference between systolic and diastolic pressures [18]. The pulse pressure is mainly influenced by the viscoelastic properties of the wall of arteries with more or less large diameters. When the blood flow is constant, there is a reduction in the viscoelastic properties of the wall of the large arteries, causing an increase in systolic pressure and a decrease in the diastolic one, so the non-variation of the mean pressure and the increase of the pulse pressure [57].

Table 2.1: American Heart Association [25] rages defined of pressure.

	Systolic (mmHg)	**Diastolic (mmHg)**
normal	below 120	below 80
elevated	120 - 129	below 80
high BP (stage 1)	130 - 139	80 - 89
high BP (stage 2)	140 or higher	90 or higher
hypertension crisis	above 18	above 120

2.3.1 Hypertension

Hypertension, associated with pressure values above 120/80 mmHg, is a disease that affects at least 60 million people in the United States, about 1 in 6 adults is affected [10]. The main forms with which it occurs are two: primary and secondary hypertension. Of the first, also called essential hypertension, which affects about 90 % of hypertensives, the causes are not well defined, but it is certainly associated with risk factors such as obesity, stress, smoking and genetic predisposition. The second is associated with secondary pathologies, such as renal hypertension, associated with kidney dysfunction due to inappropriate secretion of the endocrine system.
Arteriosclerosis is related to hypertension, in which plaques are formed at the level of the arterial walls with a consequent decrease in their elasticity and a decrease in their lumen diameter. In particular, the decrease in the lumen increases resistance to the passage of blood by triggering the arteriosclerotic process. This triggers a vicious circle in which the two phenomena amplify the effects of the other, progressively worsening the health conditions of the person suffering from hypertension.

In addition to this, the pathology can lead to other cardiovascular disorders, such as the increased risk of myocardial infarction due to the cardiac activity required to balance its effects on the vessels. Heart failure may occur due to the increase in end-diastolic volume due to the increase in the volume of blood required. Or there may be other consequences for the kidneys, such as kidney failure, or retinal level, with loss of vision.

2.4 PPG sensor

This technology, based on optical technique, make possible to detect the blood volume changes in peripheral body areas. It is a non-invasive kind of technology composed by two mainly parts:

- a LED (Light-Emitting Diode) which convert electrical energy into light. It is characterized by a narrow single bandwidth (50 nm typically) [19]. It has a high operating longevity (>105 h) and good mechanical robustness and reliability. They are used for the ability to maintain a constant level of brightness intensity low enough to reduce at the same time the heating of the tissues and the danger of damaging the tissues with non-ionizing radiation;
- a corresponding photodiode that matches the light source spectral characteristic, to measure blood volume changes at the arteries, converting light energy into electrical current [19].

2.4.1 Physical principles

Light propagating through a medium interacts with it in different ways. It can be partly reflected, absorbed, scattered or transmitted. It can pass through a medium as a ballistic component or it can be split into different components (Figure 2.7) [14].

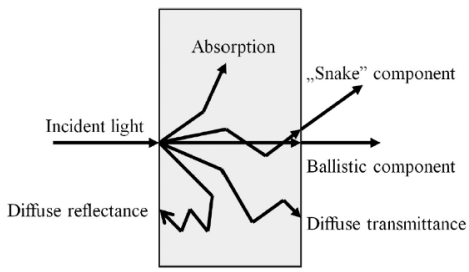

Figure 2.7: Physical response of objects affected by radiation.

At the base of this phenomena there are four parameters:

1. The refractive index, which define the speed of light and its variation give rise to scattering, refraction and reflection phenomena. It usually occurs when light is incident to the boundaries between two media characterized by two different refractive indices;

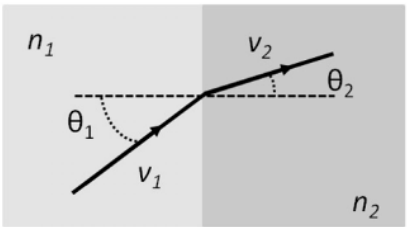

Figure 2.8: Example of light refraction corresponding to the boundary that separates two media described by different refractive indices (n1 < n2) (Haahr, 2006).

2. Absorption coefficient, in which the loss of photonic energy during interaction with the media is converted into either heat or into light of longer wavelengths. It is wavelength dependent and it is defined as a probability of photon absorption per unit path length, which the photon passed through medium;

3. Scattering coefficient, which involves the deflection of the beam due to the interaction with the medium. Among the factors that influence this parameter are the wavelength of the light source and the variation of the refractive indices of the different tissues crossed by the light beam.

2.4.2 Sensor configuration

To minimize these effects and optimize the light transmission system, it is necessary to identify the wavelength window that is affected to a lesser extent by the absorption and scattering effects by the tissues, mostly composed of water, which absorbs a lot of radiation in the infrared at higher wavelengths and in ultraviolet.

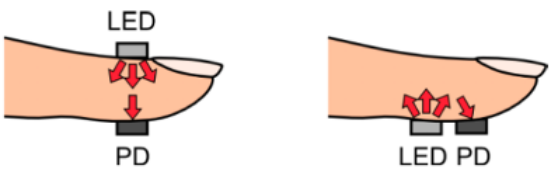

Figure 2.9: Possible configurations of the PPG sensor.

The PPG signal can be measured in two configurations: in transmission or reflection mode.

- In the first configuration, the light radiation that passes through the medium is measured by the photodiode, placed in a position opposite the LED. This, however, limits the possible sites of signal collection, the finger would be the best, and suffers from the interference due to the light present in the surrounding environment, as well as to clutter the daily or sports activities that you want to improve in the case of positioning on the finger;

- In the second configuration the photodiode receives the reflected or backscattered radiation from the medium. This gives the possibility of having multiple sensor application sites for measurement. On the other hand, it is mainly affected by the movement artifacts and by the contact pressure variations which translate on the signal quality.

2.4.3 Influencing factor

Light intensity reflected by the blood vessels, bones and skin is traduced in variation of PPG intensity signal. In general, the significant variations are duo to the change of amount of blood in the vessels and this permit to study the performance of vascular system. However, the amount of light that is usually studied does not coincide with the incident one, but it oscillates between 1 % and 5 % of the initial radiation and this increase the difficult to distinguish the light absorption by blood from that caused by other anatomical components [20]. The light transmitted or reflected is influenced by different factors like:

- skin properties: skin structure and composition, blood oxygen saturation, temperature and blood flow variation;

- number of red blood cells in circulation, orientation and agglomeration;

- speed of contraction and dilation of arterioles and their capillaries.

The structure of the skin produces a minimum dispersion of light like the 6 % of the total reflected.
In addition to melanin and water, there are also other substances which affect the absorption spectrum of this layer, including hemoglobin, fibrous proteins, collagen, fat, bilirubin and carotene.
The measurement of non-absorbed light, i.e. reflected or transmitted light, can be used to estimate the amount of blood volume in the tissues, since they are inversely related to each other [21].

Beer-Lambert's law (Equation 2.6) can be used to estimate the physical correlation with the biological behavior of light diffusion through the body [47]:

$$I_{out}(\lambda) = I_{in}(\lambda)e^{-\mu_a d} \tag{2.6}$$

It is influenced by different factors like: light intensity of the light, the wavelength of the incident light, the length of the path and the composition of the structure in which pass the radiation.

2.4.4 PPG signal morphology

The waveform of the PPG signal consists of two components: AC and DC component. The AC component is attributed to the pulsatile behavior of the cardiovascular system and synchronous cardiac variations in blood volume at each heartbeat [35]. This component is superimposed to the DC one, which represents the continuous component linked to the average volume of blood. The latter changes slowly over time and is influenced by breathing, vasomotor activity, sympathetic nervous system activity, vasoconstrictor waves as well as thermoregulation [19].

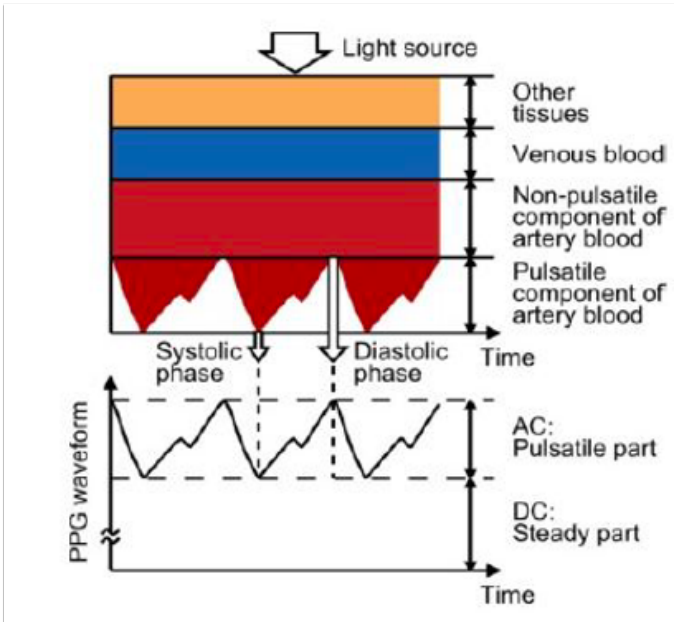

Figure 2.10: Components of PPG signal.

The first part of the waveform is called anacrotic phase or systolic tract and is obtained by the passage of blood from the aortic root to the peripheral site on which the sensor is placed. The second part, the catacrotic phase, includes the diastolic tract and is obtained by the transmission of pressure from the ventricle, along the aorta, to the lower part of the body [21].

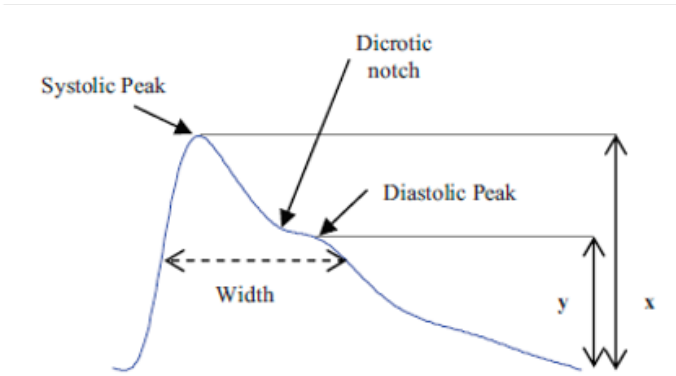

Figure 2.11: Typical waveform of PPG signal [26].

The features of the signal are:

- systolic peak: obtained when the pressure pulse reaches the peripheral site from the left ventricle;
- diastolic peak: generated when the pressure wave through the arteries and reaches the lower part of the body;
- systolic-diastolic peak-to-peak time (SDPPT): measured as the interval time between the systolic peak and the diastolic one [24];
- dicrotic notch: located at the diastolic tract, it is generated in the instant of closing of the aortic valve [35]. This episode causes a retrograde flow and an increase of the blood volume in the arteries for a short time [26]. Its amplitude varies with the elasticity of the arterial vessels and depends on the pressure wave that is reflected in the peripheral arteries [35];
- pulse width: defined like the width of the signal obtained considering the time distance of the two points of the PPG signal with a value equal to half of the systolic peak. This point relates more precisely the PPG signal with the resistance of the vascular system than the peak reached in the systolic phase;
- pulse area: defined as the total area underlying the PPG signal curve. The Inflection Point area (IPA) is an indicator of the resistance of peripheral vessels [43] and is obtained by the ratio of the two areas underlying the curve separated by the dichrotic notch;
- peak to peak interval: time interval between two systolic peaks of the PPG signal. This value corresponds to the duration of the cardiac cycle and is comparable to the RRI of the ECG signal;

- pulse interval: which measures the distance between the start and end point of the PPG signal. This is a time interval related to some parameters of the heart system, but the precise relationship is not yet known [26].

- photoplethysmography intensity ratio (PIR): which is the relationship between the peaks and the valleys of the signal inside the observation window [43].

2.5 The ABP signal

To complete, you also need to deepen your atrial blood pressure signal (ABP). The source of this signal is the same as for the PPG signal, i.e. the heart. In fact, the two signals have a similar morphology, with particular differences due to the technique and the collection area of the respective signals.
For the ABP signal, the signal acquisition site assumes considerable importance, as already seen above, becouse the characteristics of the vessels have a marked heterogeneity between the proximal areas of the heart, characterized by a marked elasticity due to the thickness of the walls, and the distal ones, characterized by narrower vessels that offer greater resistance to flow [28].

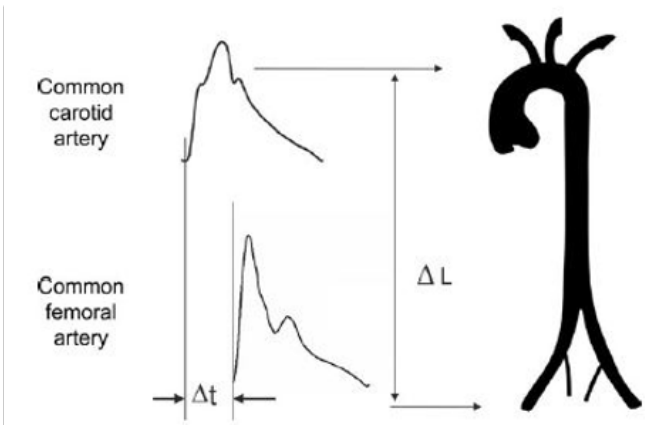

Figure 2.12: Comparison between pressure waveform from acquired from different sites [26].

The signal is represented by a pressure wave, which moves through the arteries and will have different diffusion speed and morphology based on the area crossed, as shown in Figure 2.13. A pressure wave that diffuses through a viscoelastic tube is progressively attenuated with an exponential decrease in speed, but if the tube has different branches in which to divide, reflection phenomena occur which translate into an amplification of the signal.

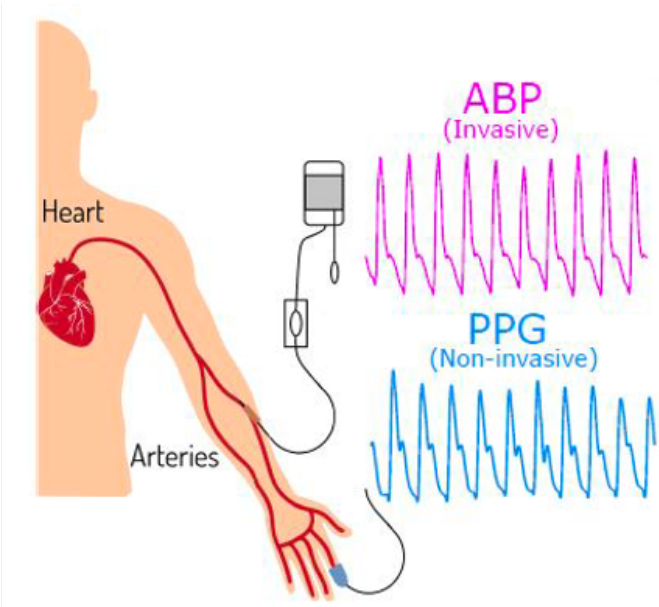

Figure 2.13: Comparison between the ABP and PPG signals [5].

The contribution of reflection changes the morphology of the signal collected at the arteries from the recorded near the heart. In the aorta for example, the reflection is negligible as it is less rigid.
Aortic pressure can also be measured at the radial artery using a transfer function, or in carotid artery.
The difference between ABP and PPG is related also to the different technique used to acquire the signal. The aortic pressure wave can be acquired using a non-invasive technique, which uses a pencil type probe with a built-in transducer [28]. The acquisition of the PPG signal is different, as already explained. The contribution of absorption and reflection of the light generated by the different tissues is added.

Chapter 3

State of the art

3.1 Conventional method for BP measuring

Blood pressure can be measured in invasive and non-invasive way. In the following sections both methods are described.

3.1.1 Invasive method

Early studies, since the second half of the XVIII century, aimed the possibility of blood pressure estimation. The discovery was reached by Stephen Hales (1677 - 1761) [50], who firstly measure the mean BP by inserting a tube, opened to the end, in the carotid artery of a horse and measuring the height reached by the horse's blood. The discover gives the possibility to measure BP and after it, multiple study and device were implemented to reach a major precision, safety and comfort.
Actually, invasive methods remain the gold standard for BP measures, for its accuracy and the possibility of beat-to-beat measuring of the pressure. Despite this, it is used only in the case of wide fluctuations in the subject's pressure values or if a continuous control of his pressure values is necessary or subjects which can't use non-invasive methods like obese people.
The measurement is performed through a needle normally inserted in an artery, usually the radial artery because of its superficiality and accessibility, connected through a catheter to a pressure transducer, who traduced a physical quantity in input, from saline in the latter part of the tube, in an electrical one with respect to the atmospheric pressure. This measuring system needs to be calibrated respect the atmospheric pressure.
The device measures the wave of blood hitting the catheter through the needle and the blood pressure waveform is displayed on a monitor, after being amplified and filtered.

Because of its invesiveness and uncomfortable for the patient, this type of measurement is disadvantaged in the long run, with the risk for the patient to contract an infection where the catheter is inserted or bleeding due to its sudden detachment. In fact, this procedure is used for hospitalized patients in intensive care unit or in the operation room, where a continuous control is needed. Figure 3.1 depicts a pressure transducer typically used for monitoring BP with invasive measurements.

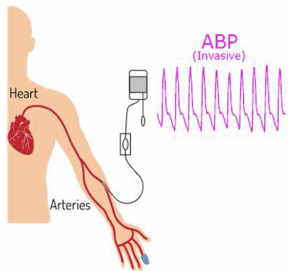

Figure 3.1: Invasive blood pressure measurement [5].

3.1.2 Non-invasive methods

Figure 3.2 shows different technique to measure the BP in non-invasive way which remain the best way to measure the BP in term of safety and usability outside clinics.

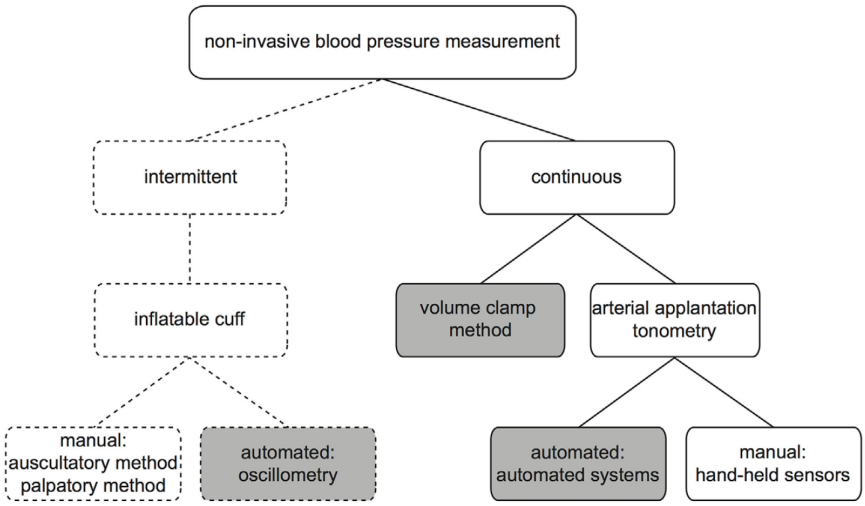

Figure 3.2: Tree schema of different ways to compute the blood pressure measurements [23].

Inflatable cuff

The first and the most popular measuring method is based on occlusive method. The occlusion is provided on the artery, typically the brachial artery, by an external pressure imposed from the outside. This vessel is used because it is at the same level of the heart. As a consequence, in this way, the hydrostatic pressure component is delated, not affecting the measurement.
The device that performed the measure is the sphygmomanometer, that was perfected for clinical environments by prof. Riva Rocci in 1896. Even today, this method is the one most used in BP measurements. It consists of an inflatable cuff connected to both the pressure measurement system and the pump with a valve that regulates the amount of air present in the cuff.
The cuff, posed on the upper harm, is inflated to reach a pressure superior of the systolic pressure and to interrupt the blood flow. Then the pressure inside the cuff is reduced by means of the valve until the flow is restored.
By analyzing the restored process, is possible to measure the SBP and DBP. The most used sphygmomanometer is mercury based. The mercurial sphygmomanometer consists on a cloth cuff connected through a tube to an external pump with a valve and a mercury column. Another pressure measuring system is the aneroid sphygmomanometer, in which the mercury column is substituted by a clock manometer with mobile needle and a graduate scale in mmHg. This system does not maintain the stability of the measurements over time and it needs to be calibrated.
New kind of sphygmomanometer, called hybrid sphygmomanometer, try to reproduce the best features of mercury ones and replace with electrics component, with a display instead a manometer of the mercury ones and it replaces [49].
Due to the time interval between measurements, this measurement method cannot measure the pressure minute by minute, for each heartbeat, therefore it is not a continuous, but an intermittent measurement method.
Figure 3.3 shows a mercury based sphygmomanometer.

Figure 3.3: Example of mercury based sphygmomanometer [51].

Auscultatory method

With a stethoscope, placed on the antecubital fossa of the elbow, the presence or absence of a beat is auscultated during air inflection and deflection of the cuff. When the pulsatile blood flow is restored reducing the air in the cuff the stethoscope makes the possibility to check the sounds generated by the combination of the turbulent blood flow and the oscillations of the arterial wall [29].
The sounds are called Korotkoff sounds, in honor of the Russian physicist who firstly study the relation between blood pressure and blood pressure cuff. He divided the detected sound in five phases [30]:

- Phase 1: the first phase corresponds to the appearance of a clear, palpable and repetitive pulse. The pressure measured is the systolic one;
- Phase 2/Phase 3: the second one have sounds becoming less intense and more durable, with the presence of slight noise until the third one, in which the beats are less clear and low;
- Phase 4: the sounds become softer and less distinct from each other;
- Phase 5: the final corresponds to the last audible sound. The pressure measured in this phase is the diastolic one.

This method can suffer from human errors because of the need of a trained operator to perform the measurement.

Oscillometric method

The oscillometric method is performed to reduce the error computed by the auscultatory method [31]. It is a derivative measurement technique consisting in a pneumatic circuit connected to a cuff, with similar functions to the cuff of the sphygmomanometer, but with internal pressure sensors connected to electrical circuit to detect the pulse pressures inside.
The measurement is performed automatically by the device positioned in correspondence of the writs or of the biceps. It stars with rapid inflating air until the lumen of the artery is completely compressed and the flow is interrupted. In this way the pressure wave is completely reflected and diffuses in the cuff.
The cuff is subsequently deflated in steps of 5 mmHg and the sequence considers that the pressure in the cuff decreases by one step when two pulses of equal amplitude are detected. The cuff is subsequently deflated in steps of 5 mmHg and the sequence considers that the pressure in the cuff decreases by one step when two pulses of equal amplitude are detected. At each step, the system acquires the pressure and amplitude of the pulses. Until the measure system reaches a maximum the amplitude of the beats increasing during the cuff deflates. Once the maximum is exceeded, the detection of the beats stops permanently at the moment in which

normal flow is restored. The pressure variations are produced by the passage of the pressure wave in the artery and they are recorded thanks to the counterposed action between the blood pressure and the inflated cuff pressure. These variations induce small changes in the volume in the cuff, and then in the pressure inside it. In the condition in which there is the maximum variation in the section of the artery during the cardiac cycle, pressure pulsations of maximum amplitude are found.
When the maximum amplitudes are detected the pressure inside the cuff correspond to the mean arterial pressure (MAP). Typically, the SBP and DBP are measured by algorithm way and often the corresponds to the 45 % - 55 % of the pulse amplitude maximum for the SBP and to the 74 % - 82 % for the DBP.
Figure 3.4 represent an example of oscillometric shygmomanometer.

Figure 3.4: Oscillometric sphygmomanometer [52].

3.1.3 Tonometric method

Tonometric measurement is a continuous cuff-less technique to detect blood pressure. It is based on strain gauge pressure sensor (piezoresistive crystals) placed over placed over the skin, in correspondence of blood vessels with to superficial arteries located near bone structures, generally the radial or carotid arteries because their easily access and large diameter.
It is normally applied on the right wrist and receive proportionally pulse and it receives the ones produced inside its. Unlike the measurement systems seen so far, this technique does not determine a complete occlusion of the flow of the vessels on which the sensor is applied. In this way it is possible to detect the pulse wave of the blood flow and to obtain the consequent pressure values during the acquisition. The system needs to be calibrated with SBP and DBP values measured with traditional methods, such as oscillometric or auscolturory methods.
The calibration must be perform for each patient and this does not make it ideal for hospital measurements. Furthermore, the quality of the registered signal is affected by the correct positioning of the sensor and by its high sensitivity to movement artifacts [32].

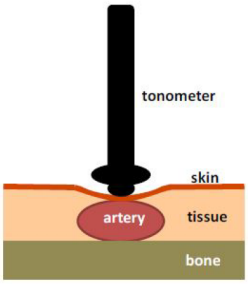

Figure 3.5: Tonometric method for BP estimation [55].

3.1.4 Volume clamp method

The technique is based on measuring blood pressure through a cap placed on a finger that gives its air flow so that the diameter of the vessels never changes during the cardiac cycle. In addition, there is a PPG system that allow to detect variations in the blood volume in the vessels during the different phases of cardiac cycle.
This system allows to automatically adjust the response to changes in diameter and to detect the pressure wave inside, therefore the values of SBP and DBP. The control works with negative feedback. When the systolic peak is detected which leads to an increase of blood volume detected by the PPG system, the servo pressure system responds quickly increasing the volume of air inside the cuff to maintain the blood volume constant [37]. In this way the beat-by-beat pressure can be measured. The continuous pressure variations inside the cuff to impede the blood vessels diameter changes corresponds to the MAP.
This system is a good solution for healthy subjects, but it is not optimal in case of subjects with bad peripheral perfusion.

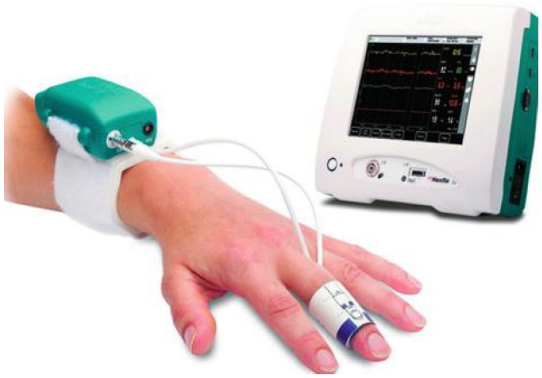

Figure 3.6: Volume clamp device for BP estimation [54].

3.2 Wearable devices for blood pressure estimation

As mentioned in the Introduction, the wearable device market is growing rapidly. There are many advantages that these devices offer in terms of ease of use, calculation capacity and minimal invasiveness. Furthermore, the possibility of being connected to the internet and the emergence of new communication systems, such as 5G, will strengthen this family of technologies in the Internet of Things (IoT). One of the challenges that is being faced now is to increase its charge capacity while keeping the calculation levels high. These technologies have the ability to collect data and to extract variables with refined accuracy.
Currently the devices of this family used both in the clinic and for remote home monitoring are based on

- tonometric measurement;
- analysis of the PWV.

The first one uses the technique showed below but in automatic way. It can be associated to a watch with a display that showed the measured values. Furthermore, it can send the collected data through a Bluetooth connection to the smartphone which a specific App.
The first of the devices falling into this class is the BP Pro developed by Healthstats (30 Tai Seng Street, #09-03 BreadTalk IHQ, Singapore 534013). It is constituted by a sensor module with a piston (the sensore plugger) that exercise a constant compression on the radial artery during the sampling period (10 s). It is based on the applanation tonometry to acquire the pulse waves [39], detected from the radial waveform, and it gives a brachial blood pressure measurement after every sampling.
The measurement is automatically computed from the matching of pressure waveform sampled with template waveform, obtained from several indices and coefficients. When the sampled waveform coincides with one of the templates it is collected and the MAP calculated. The SPB and DBP are extracted from derived coefficients by detecting the calibrated level through change in arterial waveform [48].

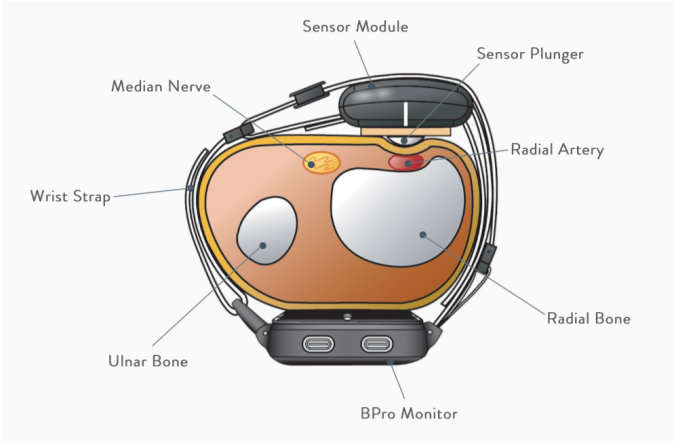

Figure 3.7: Volume clamp device for BP estimation [39].

This device has shown excellent results in terms of accuracy, with an error within 5 mmHg from auscultation method, defined by the guidelines for use in the clinical environment, by various validation tests. In addition, it allows rapid measurements repeated several times throughout the day, up to 96 times in 24 hours. Despite this, it has not yet obtained FDA approval and it is classified like cosmetic device.

Another example of device is the Omron pressure measurement device called HeartGuide. It is the only model who receive the FDA approval as a medical device until now. The device needs to work of an the specific Omron App to send the registration [53]. The App elaborates the acquisition and gives the measurements of BP and other vital signs like HR and the number of steps.

Figure 3.8: HeartGuide watch developed by Omron [53].

To collect pressure data, the subject must be stationary and bring the arm to which the watch is applied at chest height, keeping it still. The detection takes a few minutes and an accelerometer checks if you maintain a correct posture and

if you are moving. If excessive movement is detected, the detection is canceled. During the detection, initially the cuff in the strap inflates to obtain the width of the arm, then deflates and inflates again until it reaches a clear detection of the pulse wave in the vessels. Finally, it deflates again and the detection can be read both from the watch and from the device to which it is connected via Bluetooth.
In occasion of the annual congress of the European Society of Hypertension (ESH) in Athens 2014 , prof. Parati present a validation study for SOMNOtouch^{TM} NIBP accuracy for non-invasive BP estimation through the PTT [56]. The device has been designed to continuously monitor patient pressure in a clinical environment by minimizing invasiveness. It is made up of a central body for data collection and processing to which the electrode cables for the collection of the ECG signal and the PPG sensor are connected. During the validation tests, the device's accuracy was compared with the measurements performed with the auscultory method, demonstrating an accuracy within the criteria of the ESH-IP 2010 for both SBP and DBP.
This encourages the deepening and development of devices based on this technology for the definitive affirmation of this new family, capable of performing BP detections in multiple environments (clinical, sports, etc.) in a continuous and non-invasive way.

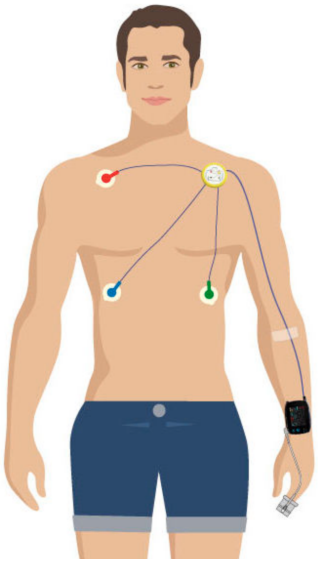

Figure 3.9: Architecture of SOMNOtouch^{TM} device [56].

3.2.1 Blood pressure estimation using other physiological parameters

In addition to classical methods, in literature, there has been a huge attention for the study of other physiological parameters. This is because they try to make the system portable, non-invasive and that allows the continuous detection of pressure values.
The goal is to make detection as invasive measurement method as possible continuously due to the importance of long-term measurement to observe changes in pressure values and to make a prognosis, diagnosis and treat the pathologies associated with hypertension in time [43].
In this section, attention is placed on the study of the Pulse Wave Velocity (PWV) with the Pulse Transit Time (PTT) and of the parameters obtainable from the PPG signal for the estimation of blood pressure. Other studies have already identified the correlation between PTT and BP since the 2000 [43]. PTT consists of the delay between the detection of the R wave of the ECG signal and the systolic peak detected in the PPG signal of the pulsed wave transmitted through the circulatory system to the peripheral sites. Chen et al. [44] for example have demonstrated the possibility to measure the SBP through the pulse arrival time with successive calibration. Another case is the study conducted by Poon et al. [45] for a PTT initial calibrated model for the SBP and DBP estimation.

Despite its merits, the use of this technique is still limited above all by the need for continuous system calibrations and therefore difficult to apply to the long-term monitoring of the subjects. Furthermore, the need for continuous calibration of the instrument does not allow this technology to replace the traditional measurement methods, as the standards do not foresee the need for calibration for this class [42].

The Pulse Transit Time

The PPT is the measure of the time interval necessary to the pressure pulse to spread into peripheral blood vessels. It begins when the pulse pressure is detected in aortic vessel, due to systolic phase of the heart cycle, and it finishes when the pulse is detected in the distal part of the vessels.

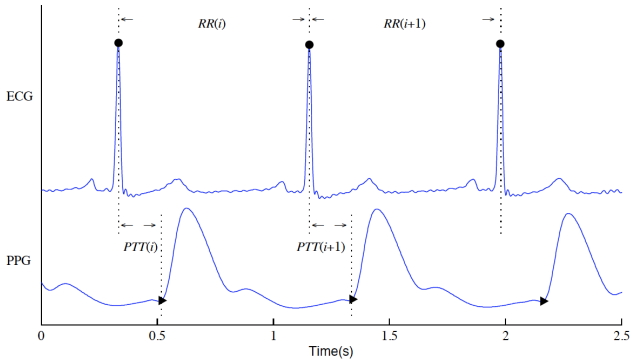

Figure 3.10: Example of measure of the pulse transit time between ECG R-peak and the initial stage of anacrotic phase of PPG signal [58].

Essentially the elasticity of the vessels has been traced back to the volume of blood circulating in the vessel [38]. When the pressure increases, the compliance of the vessel decreases and therefore the temporal window between the beat and the arrival of the pressure wave to the peripheral areas decreases, decreasing the PTT.

The most effective mathematical models behind these phenomena are studied by Brawell-Hill and Moens-Kortwege, which have been widely used and they have been widely used and extended. Moens-Kortwege describes the relation between the PWV and the elasticity of the arteries as reported in Equation 3.1:

$$PWC = \sqrt{\frac{tE}{\rho d}} \tag{3.1}$$

Where E is the elastic module of the vessel wall, t the vessel thickness, ρ the blood density and d the arterial diameter. Geddes described the elastic module of the vessel wall as a function of the blood pressure (Equation 3.2) [46].

$$E = E_0 e^{\alpha P} \tag{3.2}$$

Where E_0 is the elastic module at zero pressure, α is a constant variable with a range values between 0.016 – 0.018 mmHg^{-1} and P is the BP. Furthermore, the relation between PTT e PWV can also be expressed according to Equation 3.3:

$$PWV = \frac{A}{PTT} \tag{3.3}$$

Where the numerator parameter A is intended like the distance between the heart and the peripheral sample site [40]. Finally, through Equation 3.1 and Equation 3.3, BP can be estimated:

$$\frac{A}{PTT} = \sqrt{\frac{tE}{\rho d}} \tag{3.4}$$

To extract BP is necessary to define an initial calibration and the t/d value relation constant. Furthermore, also considering constant the distance A and the elasticity parameter E_0, the BP can be estimated through a logarithmic relationship:

$$BP = a * (PTT) + b \tag{3.5}$$

From this equation we can define three subject-constants a, b and c to obtain a regression relationship between BP and PTT, that also include the HR as second variable:

$$BP = a * (PTT) + b * HR + c \tag{3.6}$$

Because of to the approximation computed to the variables and the necessity of repetitive calibrations, the BP measured with the PTT technique has normally higher error levels than classical methods. Furthermore, the HR has positive and negative relation with the BP measures: positive for normal conditions but negative under baroreflex activity [73]. Finally, another parameter which influences the error is the distance between the heart and the site if detection of the signal: in particular, which decrease when it is large [33].
Given the need for an initial calibration, it is assumed that the estimate through the PTT can record, with this setting, only the pressure variation values (relative pressure). Therefore, it is necessary to initially estimate pressure with classic methods and to repeat the operation over time. This is the case of another work focused on the theme, proposed by I. Sharifi et al., in which the values of SBP and DBP are estimated as follows:

$$DBP = MAP_0 + \frac{2}{\gamma} \log \frac{PTT_0}{PTT} - \frac{1}{3}(SBP_0 - DBP_0) * \frac{(PTT_0)^2}{(PTT)^2} \tag{3.7}$$

$$SBP = DBP + (SBP_0 - DBP_0) * \frac{(PTT_0)^2}{(PTT)^2} \tag{3.8}$$

where PTT_0, SBP_0, DBP_0 and MAP_0 are the initial conditions. In addition to these calibration-based methods, new methods have been studied to overcome this limit by switching from automatic learning algorithms that can vary their configuration according to the data they receive at the input [59], [60]. Particular importance in the case of supervised learning is the necessity to provide multiple systems with input to these systems so as to make the algorithm capable of adapting to the various cases and responding with a possible contained error.

There are several possibilities to calculate the PTT:

- **I-PTT** in which it is measured the time interval between aortic blood flow (BF) and the temporal BP in distal positions (see Figure 3.11);

- **PPG-PTT** which measures the time interval between two characteristic points of the PPG signal. The time interval can be identified through a two-channel PPG system where the signal is picked up in two different points, one more proximal and one more distal [22]. It is calculated like the interval between two points of the pulse wave detected;

- **PAT** that is the time interval from the R-peak of ECG signal and a specific point of PPG signal (like systolic peak) detected distally to the heart.

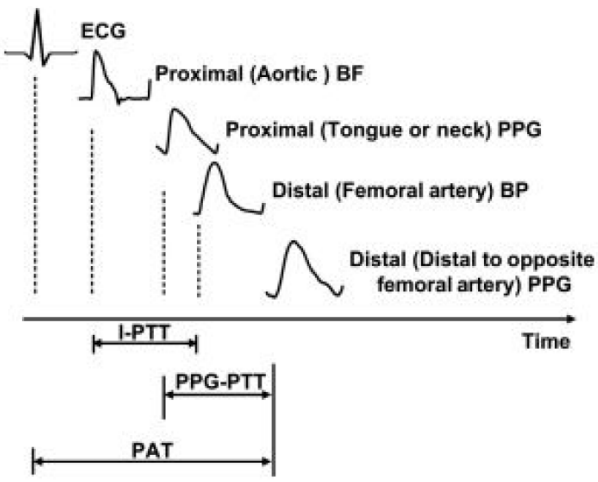

Figure 3.11: Different PTT measures in different points of the body [34].

The last method is the one often used for BP estimation. For the minimum width of the window in which the measurement is made, all the small contributions of delay can give rise to errors which, if accumulated, can give rise to errors of large proportions. This is the case of the Pre-Ejection-Period (PEP), which is considered as the delay between the depolarization of the left ventricle and the opening of the aortic valve [36]. The factors that influence this delay are different and they change from subject to subject. This delay can result in pressure estimation errors. To eliminate this error, additional complex instrumentation is required and for this reason the error is normally overlooked.
Normally three different points of PPG signal are used for the PTT estimation:

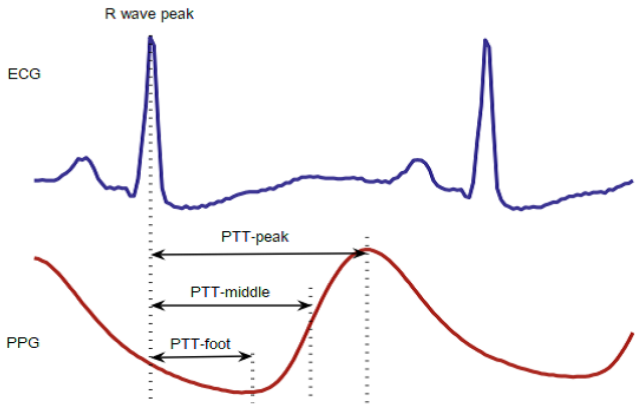

Figure 3.12: Different PTT computed between ECG and PPG signals [35].

- PTT-peak: the distance between the R-peak of ECG signal and systolic peak of the PPG signal representing the same cardiac cycle [35];
- PTT-middle: the distance between the R-peak of ECG signal and the maximum slope of PPG signal in the same cardiac cycle [35];
- PTT-foot: the distance between the R-peak of ECG signal and the minimum of PPG signal in the same cardiac cycle.

To influence the PTT there are several parameters like:

- Rigidity of arterial walls;
- Subject's age;
- Stress level;
- Posture;
- Temperature;

3.3 Employed tools

For this work, mainly two computational environments were used. The Physionet portal, available online, for the collection of data, which have been processed through the second environment: Matlab^{TM}, that allowed the realization of the algorithm and its validation.

3.3.1 Physionet

The Physionet online portal was born in 1999 with the aim of collecting biomedical signals useful for researchers and students to conduct their researches. It therefore allows free access to sensitive data, whose privacy is guaranteed, such as physiological and clinical data, as well as the related open-source software [61].
Through the database, it is possible to access more than 50 different databases, containing cardiac, pulmonary, neural and other parameters. In the case of MIMIC databases, it is possible to download the single 24-48 hour tracing of more than 100 patients with multiple signals collected simultaneously. Thanks to these peculiarities, the portal was essential, as it provided the source of the data necessary for the realization of the algorithm.

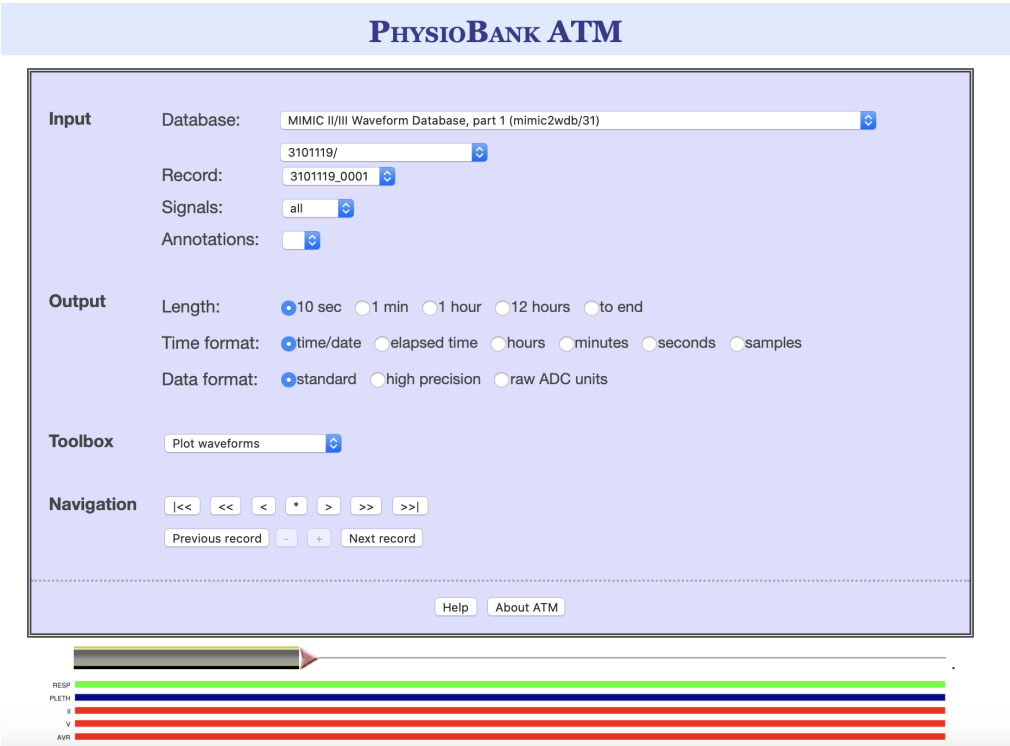

Figure 3.13: Physionet online ATM [61]

3.3.2 Matlab^{TM}

Matlab^{TM} is a numerical computing, programming and visualization environment that uses high level language. Through the software it is possible to perform calculations, measurements and visualization of the signals. In addition, it can be used for the implementation of elaborate algorithms for data processing and rapid visualization of the results achieved. In this case it is through this software that our model was created. In this case, the model was implemented through this software.

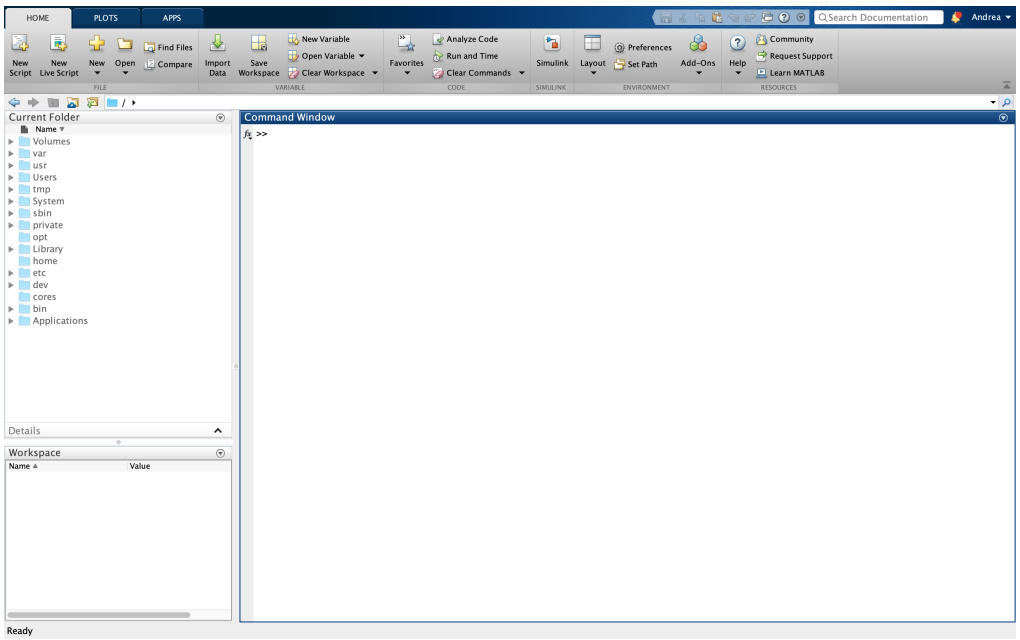

Figure 3.14: Different panels of Matlab^{TM} enviroment

Chapter 4

Materials and Methods

4.1 MIMIC databases

This work would require the collection of multiple physiological parameters registered simultaneously from healthy and hypertensive subjects, whose pathology had been claimed by medical evaluations. In this way the algorithm can be trained and tested with specific data relating to age, sex and habits of selected subjects. Unfortunately, it was not possible to implement this protocol and this required the search for an alternative source of information.
MIMIC is an open access online database containing multi-parameter biomedical data collected by the MIT Lab for Computational Physiology [62].
It collects data from intensive care from around 60,000 hospitalizations. It is a huge source of free access information for carrying out studies in a wide range of clinical fields, from cardiology to pneumatology. The available data concerned only signal recordings, therefore the work focused only on the information available, not considering information such as age and gender for the realization of the model.
In carrying out this study, the subjects whose parameters were necessary for the implementation of the system were searched within the database. In particular it was request that ECG, PPG and ABP signals were simultaneously recorded for each subject.
However, the collection of this information was not born with the same purpose of this thesis, therefore the quality of the data collected was not particularly taken care of, which in this case have a central role. This is because using mainly morphological information, especially on the PPG signal. In fact, a clear and prolong repetition of the characteristic waveform, which is difficult, especially for the sampling sites that are typically used in the collection of PPG signals such as the finger, which offers, albeit minimally, a footprint for the subject in performing even the simplest actions during the day.
The collected data for the creation of the database come from the Intensive Care

Unit (ICU) and this adds up to problems relating to the quality of the signal, because most of the subjects whose data has been collected is over 50 years of age and it is subjected to drugs that alter their pressure values. Moreover, age influences the elasticity of the vessels: in particular, vessels of older people are characterised by less elasticity and, consequently, higher stiffness. This is traduced in changes of the PPG signal waveform as a reduction in the time interval between the systolic peak and the diastolic peak, as it is shown in Figure 4.1 [64].

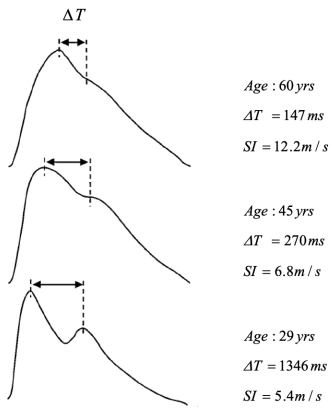

Figure 4.1: Example of the effect of advancing age on signal quality [64].

In addition, the signal waveform can present abnormal and noisy recordings, as missing peak, pulsus bisferiens, interference to the sensor (sensor-off) and others.

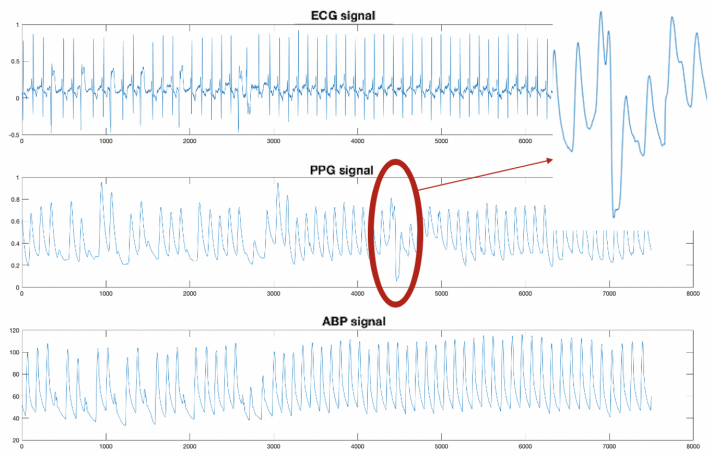

Figure 4.2: PPG signal of the subject p000333 with pulsus bisferiens.

In healthy subjects, significant changes in blood pressure occur with a time

interval of at least 10 minutes. For this reason, during the first conception of the work, it was evaluated to obtain 5-minute signal windows with respect to which to evaluate the pressure values. However this was not possible due to the quality of the signals obtained from the database, which suffered in some cases from a not negligible alteration of the measurements, which did not allow the use of such large windows (Figure 4.3). In addition, very long acquisitions would have been required in order to train the algorithm and the length of the acquisitions available in the database was not enough. For this reason it is used a modified window length to reach a compromise between samples quantities and quality of the signals contained.

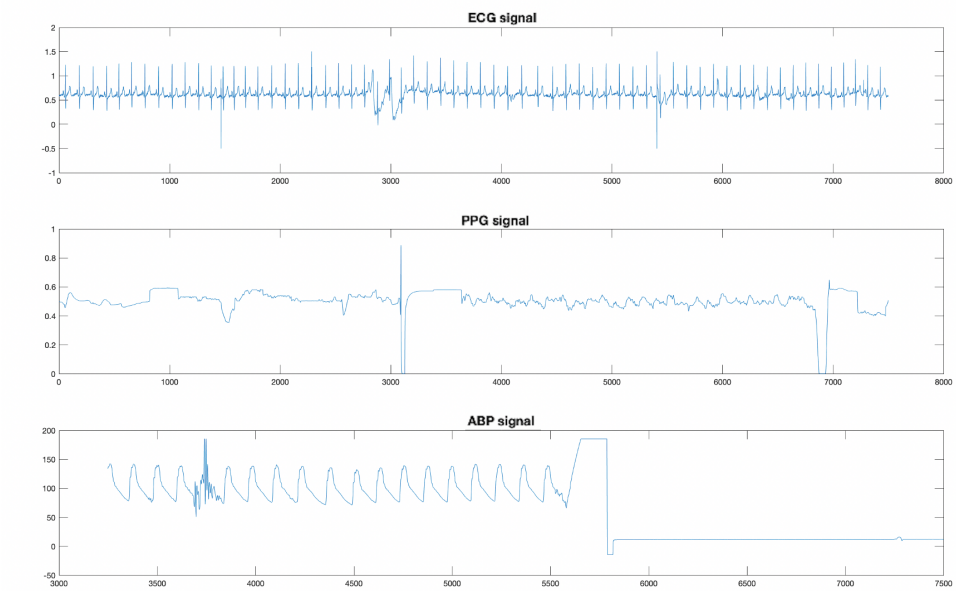

Figure 4.3: ECG, PPG and ABP signals of the subject p000160 with significant disturbances.

4.2 Algorithm description

Figure 4.4 represents the used flowchart. It shows the steps that have been taken in the implementation of the algorithm.

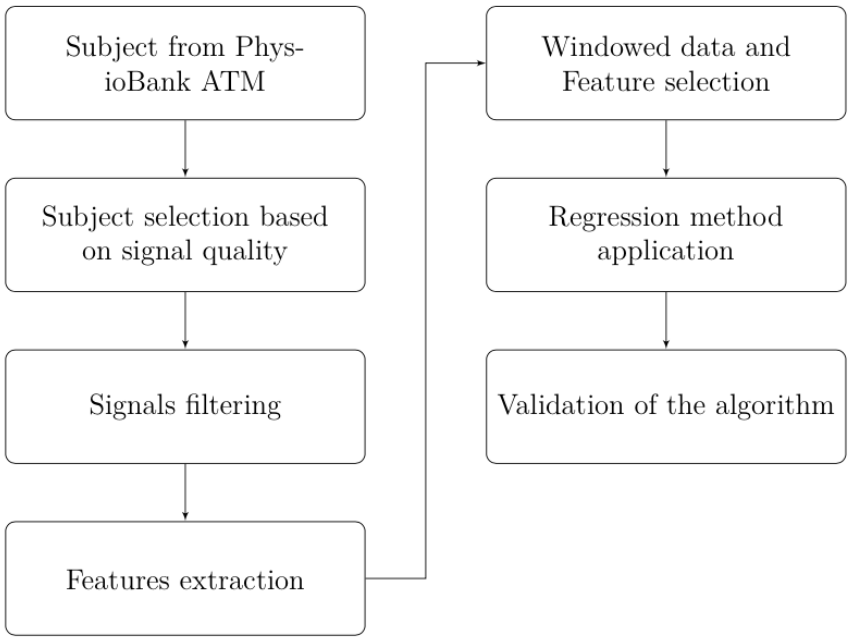

Figure 4.4: Flow chart of developed algorithm.

The first phase focuses on the selection of signals with minimum noise and a sufficiently large window. The second phase focuses on the signal pre-processing, which is followed by the extraction of the defined features. The third part instead concerns the organization of the dataset and the application of the regressive method. Finally,the system is validated to verify its capacity to generalize not reaching overfitting.

4.2.1 Subjects selections

The first step of the algorithm involves identifying the subjects in the database that are adequate in terms of the presence of the necessary signals with an adequate level of quality of the trace to correctly measure the sought parameters.
Given the structure of the online portal, to identify the subjects that contained the signals, the plots were examined and observed directly on the Physionet portal. The research took a long time due to the huge amount of subjects collected and to the absence of prior informations on the types of signals that correspond to each subject (derivation of the ECG, ABP, PPG, Respiration Rate (RESP) signal, etc.).

After having identified the subjects and having downloaded their tracks, the quality of the contained signals is investigated respecting two phases:

1. The ABP signal is studied to eliminate parts of the track where there are alterations of unknown origin that do not allow it to be used for our model.

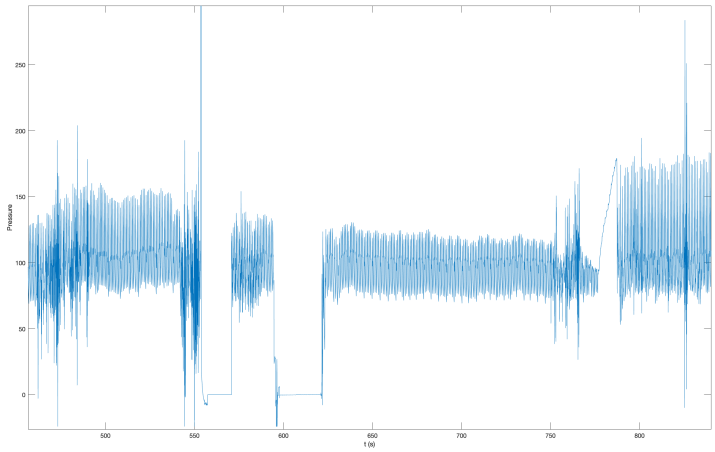

Figure 4.5: Noisy ABP signal (Subject ID: 3702767)

2. The ECG, PPG signals were broken into 1 minute fragments, assessing the quality of the traces. In many cases the quality of the signal was poor, especially for what concern the PPG signal, since high noise levels were present. This resultes in the exclusion of the segments.

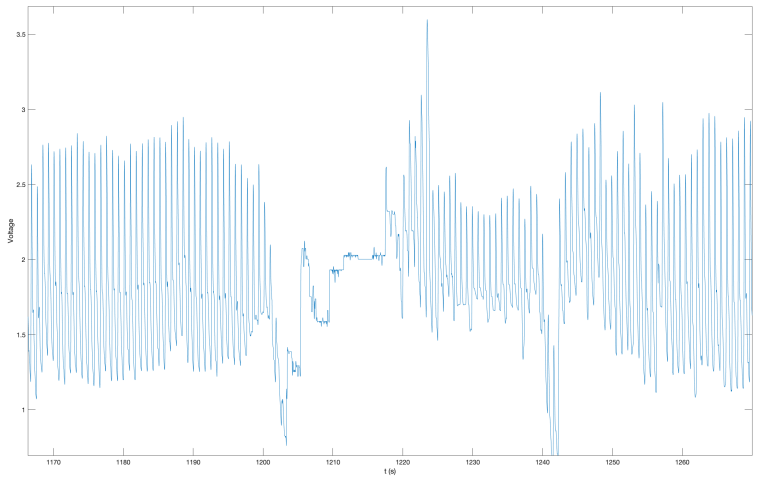

Figure 4.6: Noisy segment of PPG signal from Subject 3702767

If the number of segments is considered acceptable (at least more than 50 segments not excluded), then the subject goes to the second phase. In this case, the analysis of the signals of the subjects collected led to 24 of them. The Figure 4.7 shows the example of a subject with a clear presence of noisy corrupted signals.

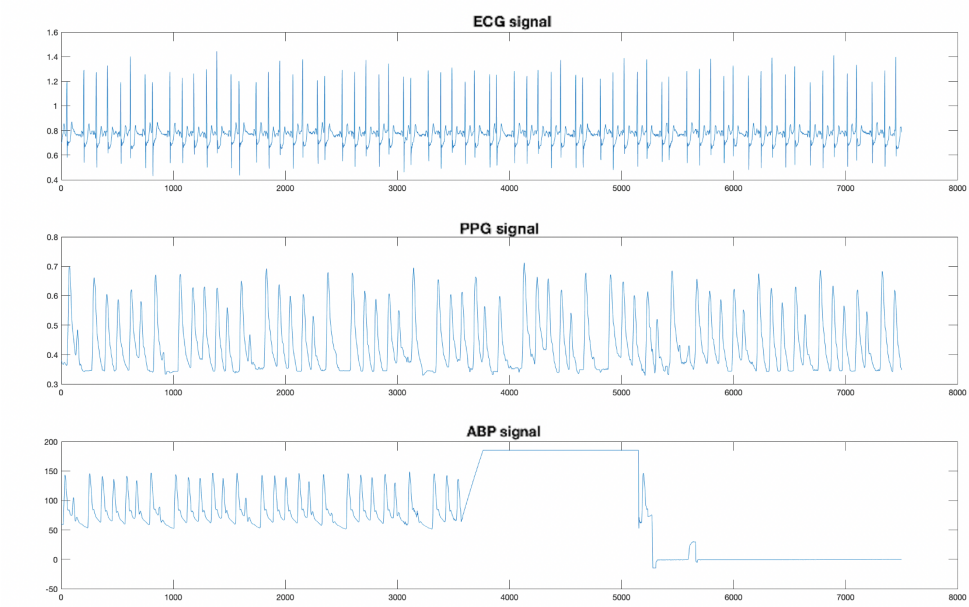

Figure 4.7: Example non acceptable signals quality with consequence elimination of the subject (ID: 3106311).

4.2.2 Signal processing

To remove the distortion effects of the signals caused by background noise and artifacts, a pre-processing phase was implemented which included the filtering and denoising of the ECG and PPG signals. In the case of the ABP signal this operation was not necessary because there was a clear and distinct waveform in the segments that passed the initial selection, also thanks to the invasiveness of the sampling which limits the presence of artifacts.

PPG signal filtering

The PPG signal suffers from noisy contributions and interference, which mainly translate into the morphology of the waveform. This is also due to the fact that the signal is characterized by low amplitudes.
The frequency content of the signal is concentrated in a band between 0.01 Hz and 3 Hz [68]:

- **low frequencies** (0.01 Hz - 0.04 Hz), which contains information about the sympathetic and parasympathetic activity of the nervous system;

- **medium frequencies** (0.04 Hz - 0.15 Hz);

- **high frequencies** (0.15 Hz - 0.5 Hz), affected by parasympathetic system activity and respiration;

- **heart rate** (0.5 Hz - 3 Hz).

The main noise contributions that sum up to the PPG signal originate from:

- **movement artifact**, due to the position of the sensor (typically on the wrist or finger) which is exposed to the movements of the subject, also involving exposure to light sources that contribute to signal alterations. The artifact is concentrated between 0.05 Hz and 8 Hz and for this reason it is difficult to remove without altering the high frequency information of the signal [68]. To remove it, various removal methods have been proposed, among which the one proposed by Ju-Won Lee et al. analyzing the signal of the acellerometer contained in the PPG sensor.

- **line interference**, which in many cases affects the signals taken by means of instruments because of the presence of cables and connections to the line. It corrupts the signal at a frequency of 50 Hz or 60 Hz, which can be easily eliminated by filtering it with a low pass filter.

- **baseline dift**, due to small repetitive movements of the subject like respiration [40]. Typically it occurs around low frequencies (0.4 Hz [69]) and it can be removed through a high pass filter.

In order to comply with what has been discussed, it was decided to digitally filter the signals using a finite response band-pass filter (FIR) with a bandwidth between 0.5 Hz - 5 Hz to contain the information about cardiac activity, removing low-frequency drifts, network interference and high frequency movement artifacts. The FIR filter was chosen because it does not alter the morphological components of the signal, leaving them unchanged. However, its limitation consists in the introduction of a delay on the signal proportional to the order of the filter, which in this case was chosen to be equal to 4 as suggested by the work of Y. Liang et al., who compared the performance of different filters. They proposed the parameters for the filters that reached the best values in term of signal quality index [70]. Given the high precision required for the timing of the signals, double pass filtering was used, in order to remove the delay by using double pass filtering [71].

Table 4.1: Filter characterization

Filter type	**Order**	**Band (Hz)**
FIR (Butterworth [70])	4	0.5 - 5

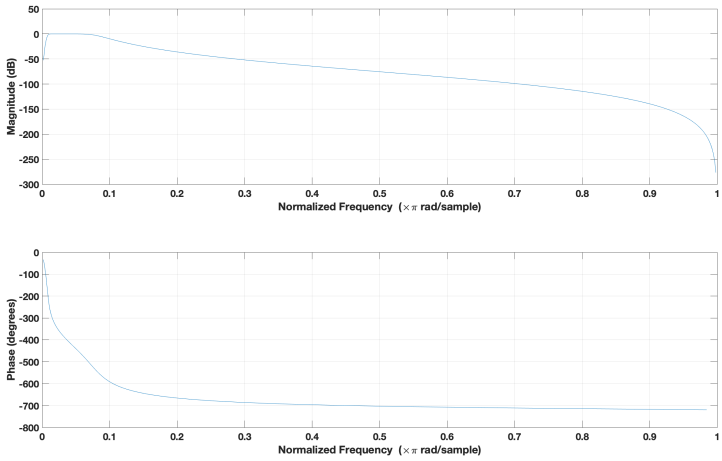

Figure 4.8: Band-pass filter transfer function implemented.

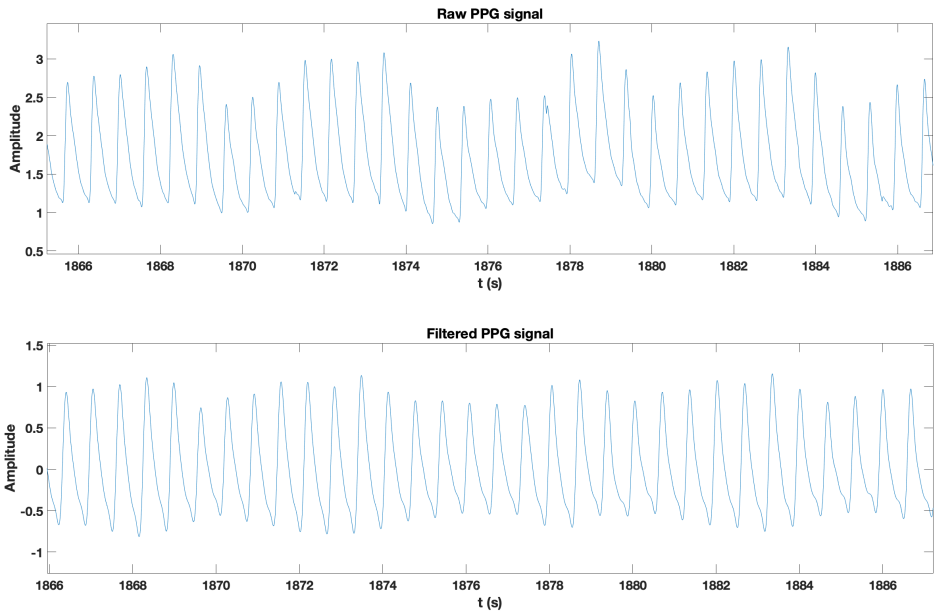

Figure 4.9: Filtered PPG signal.

ECG signal filtering

The ECG signal is the result of the sum of several waveforms originating at different moments from the heart muscle. Each waveform is characterized by different frequency bands:

- **P wave** (0.5 Hz - 8 Hz);
- **T wave** (1 Hz - 7 Hz);
- **R wave** (5 Hz - 15 Hz).

The amplitude of the ECG signal is between $10\mu V$ - 4 mV: it is very small and for this it suffers the interference on the quality of the signal originated by different sources of noise [72].

- **network interference**, the main source of disturbance for the ECG signal due to the presence of cables and network connections. Focusing around 50- or 60 Hz, it alters the morphology of the signal. To remove it, the signal is filtered with a low pass filter with a 50 Hz cutoff frequency.
- **baseline wander**, caused by subject movement artifacts such as breathing, which by concentrating around low frequencies (0.5 Hz) can be easily eliminated through a high pass filter [72].
- **muscle contractions**, being a sample made superficially on the skin, the signal may be affected by the contribution of the electromyographic signal (EMG) originated from muscle contraction. The EMG signal has a band around 0.05 Hz - 300 Hz [66] overlapping that of the ECG signal, around the band containing the different waveforms that characterize it. For this reason, the contribution of the EMG cannot be totally removed and it can give problems especially in the detection of the ECG during physical activity [76].
- **movement of the electrodes**, which modifies the contact impedance of the electrodes with anomalous manifestations on the signal. Since the interference band concentrates around 1 Hz - 10 Hz this interference cannot be removed [76].

To minimize the noisy components, the signal was filtered through a FIR bandpass filter, of order 4 and bandwidth between 5 Hz - 30 Hz. In this way, any low frequency drifts, network interference and the informations about the others ECG waves are limited by concentrating the information relating to the peak R of the signal.

Table 4.2: Filter characterization

Filter type	**Order**	**Band (Hz)**
FIR (Butterworth)	4	5 - 30

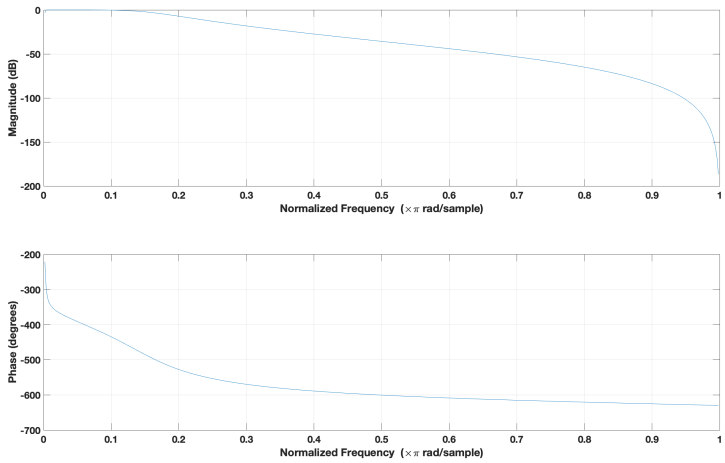

Figure 4.10: Band-pass filter transfer function implemented.

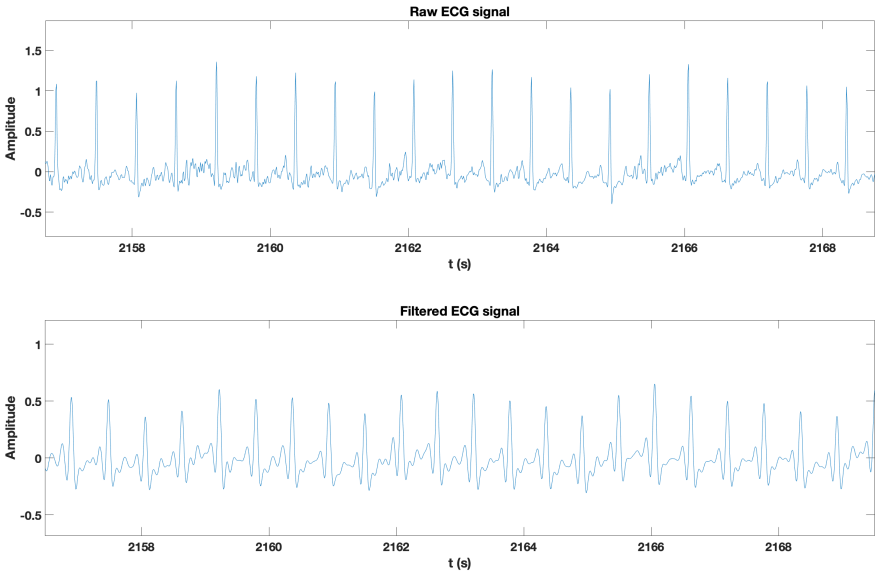

Figure 4.11: ECG signal filtered with incremented quality.

4.2.3 Features extraction

According to what has been found in the literature, multiple information extracted from the PPG and ECG signals have been explored to measure the pressure through this signal as accurately as possible. The main limit found for all the examined cases is the quality of the PPG signal, that can present itself with an altered morphology also because of the areas from which it is normally registered. Even in the case of the database used, it was possible to observe this problem, as explained in the previous section, and this led to the necessity of choosing which signal characteristics could be studied by reaching adequate levels of precision.
Initially, one of the parameters that was extracted was the distance between the systolic peak reached by the PPG signal and the diastolic one, as this distance is directly correlated with the quality of the cardiac contraction and varies with a certain proportion to the variation of the BP. However, due to the variety of morphology with which the signal occurred, it was not always possible to distinguish the diastolic peak for all subjects and this distance was not considered, wanting to maintain a minimally significant number of subjects (and therefore samples) to implement the algorithm.
For all the signals processed, a 10-second window was used in which to extract the signal features.

PPG features extraction

To extract all the parameters from the PPG signal, the systolic peaks of the signal were searched by imposing a minimum threshold equal to 30% of the maximum of the signal identified inside the window and a minimum interval between one peak and and it was considered at least 0.33 seconds between one peak and the next one, according to the cardiac property of one heart beat for second about. Once all the signal peaks were identified, the minimum points between two systolic peaks were searched related to the end of diastolic phase, by imposing the presence of a valley between two consecutive peaks.

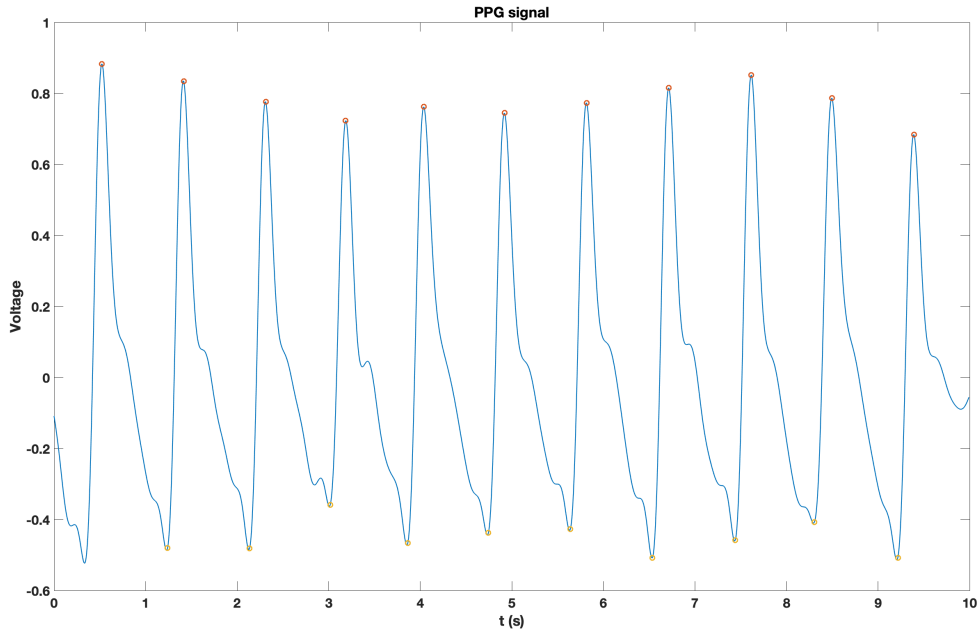

Figure 4.12: Identification of the systolic peak and diastolic valley of the PPG signal (Subject ID: 3702767).

The characteristics of the PPG signal that have been extracted were mediated with the time window considered and they are the following:

1. the systolic upstroke time (SUT), meaning the time interval between the systolic peak and the previous diastolic valley at the end of previous diastolic phase;

2. the diastolic time (DT), as the time interval between the systolic peak and the successive diastolic valley;

3. the photoplethysmography intensity ratio (PIR), which is calculated as the difference between the maximum and minimum peak of the PPG signal for each cardiac cycle [43]. This parameter can theoretically varies according to the variations of vessels diameters, as expressed in Equation 4.1.

$$PIR = e^{\beta \Delta d} \tag{4.1}$$

   where β is a function of the path covered by the light beam. PIR can capture the low frequency information of the ABP signal spectrum, originating from arterial compliance and cardiac output [43]. As seen in the previous paragraphs, the arteries vary their diameters during the different phases of the cardiac cycle (Figure 4.13), increasing simultaneously with the systolic phase of arterial BP and decreasing it in the diastolic phase ($D_d < D_s$).

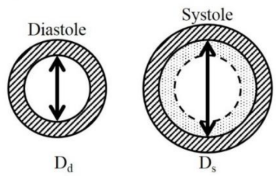

Figure 4.13: Artery diameter variation during cardiac cycle [63].

This involves variations in the amount of light absorbed by the blood tissues, giving the possibility of obtaining this parameter as a function of the difference in intensity of the PPG signal at the relative maximum and minimum points [63]. According to Figure 4.14, PIR is measured as follow:

$$PIR = \frac{I_H}{I_L} \tag{4.2}$$

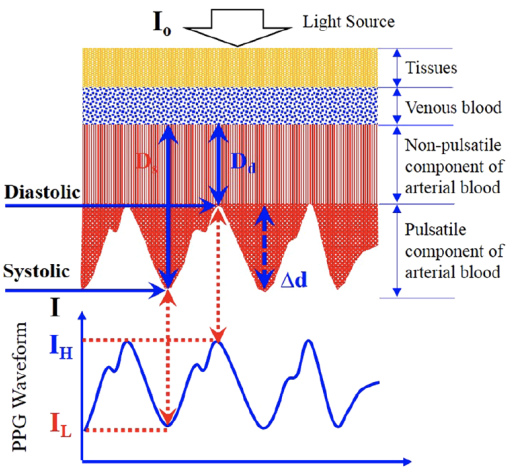

Figure 4.14: Diagram for PIR evaluation and relation with physiological background [74].

4. rate peaks alternating component of the signal, since eliminating the continuous component, that brings information about changes in blood volume pulses, which is synchronized to the heartbeat [73]. Once the continuous component of the signal is set to zero, the average intensity of the systolic peaks of the AC signal is calculated.

5. the mean intensity of the systolic peak found in the window;

6. the mean intensity of the diastolic valley of the signal in the window.

ECG features extraction

The only sought information for the ECG signal is the R peak of the QRS complex, relating to the depolarization of the ventricles. Given the relative good quality of the ECG traces present in the database for the identification of the QRS complex, the method developed by Pan and Tompkins in 1985 was used. This method is based on the analysis of the slopes of the QRS complex in terms of duration and amplitude. It consist in the following steps [75]:

- **Band-pass filtering**, the signal is filtered through a band-pass filter, which increases its quality in terms of signal to noise ratio (SNR) (done in the previous signal processing phase).

- **Derivative**, in which the filtered signal is differentiated to identify the slopes of the QRS complex. This step enhances the high frequency components of the signal, as in the case of the waveform of the R peak which is concentrated around the higher ones with respect of the other ECG waves.

- **Squaring**, in which each component of the differentiated signal is squared to emphasize the high frequency components of the signal, bringing them to positive values. The operation is performed according to Equation 4.3.

$$Y(nT) = [X(nT)]^2 \tag{4.3}$$

  where T is the sampling period of the signal.

- **Moving Average**, in which a moving average window flows along the signal. This operation smoothes the output signal from the previous phase, reducing the QRS complex to a single peak, easy to identify. The most important choice concerns the size of the window, as a larger window certainly includes the QRS complex, but it could also involve the T wave, while a too small window can lead to the appearance of narrower peaks, with the risk of having more than one them in the same QRS complex. In this work, a window of 150 ms was assumed, in agreement with the average amplitude of the QRS complex considering a minimum margin.

The result of this method is shown in Figure 4.15.

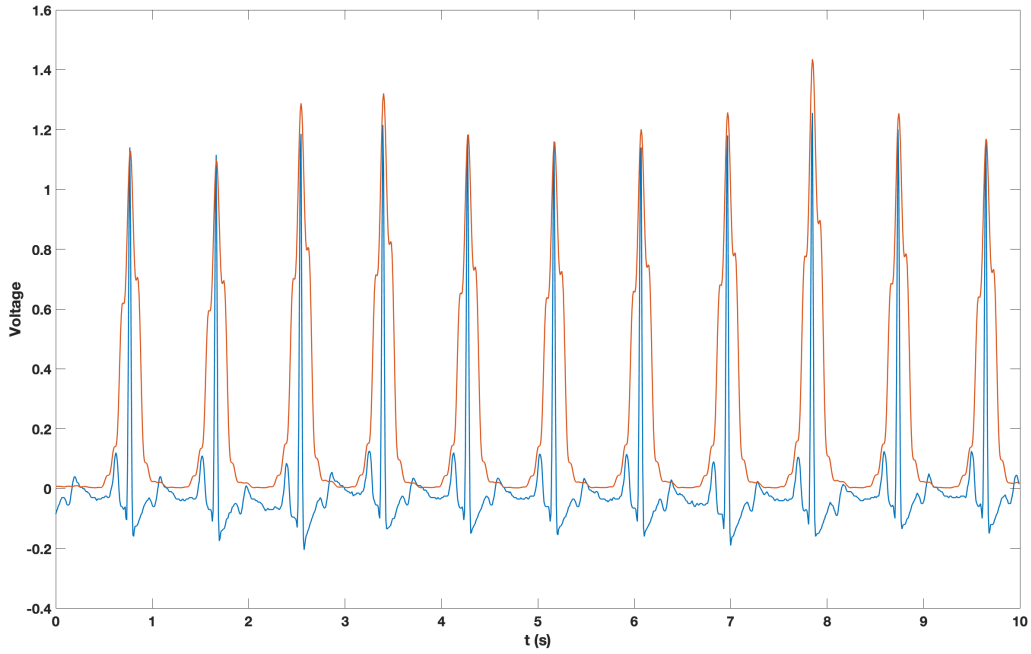

Figure 4.15: ECG signal (blu) (Subject ID: 3702767) with superimposed output of the moving average window (red).

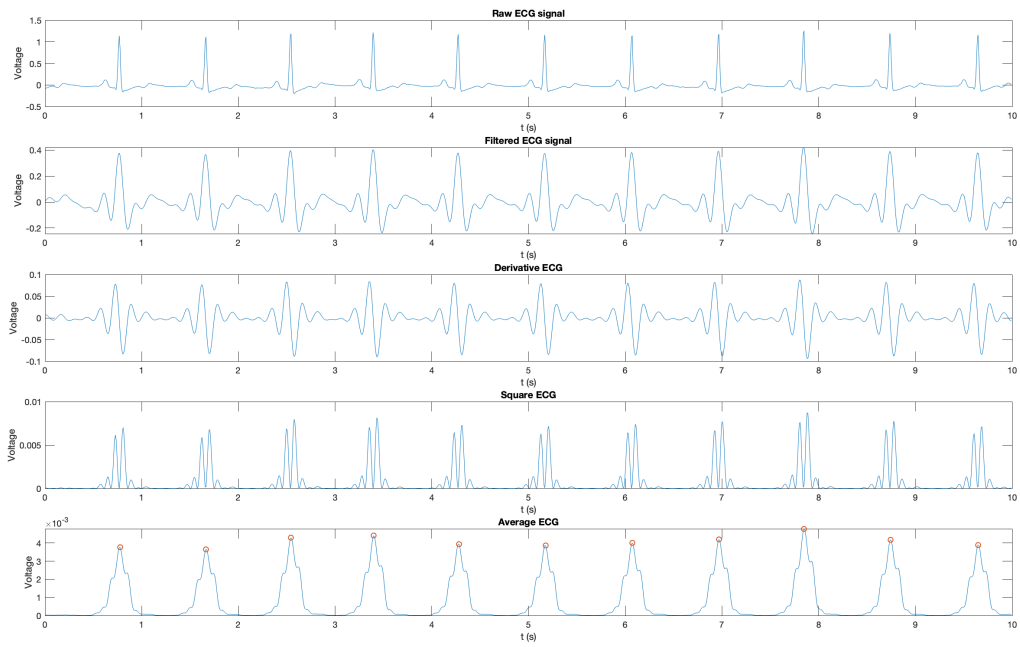

Figure 4.16: Representation of the results of every algorithm's phases.

Once the peak of the R wave was obtained, it was possible to pass to the estimate of the heart rhythm inside the window and to the estimate of the PTT.

Pulse Transit Time estimation

The importance of the correct estimate of this parameter has already been explained. The PTT can capture the high frequency information of the ABP signal spectrum, but only after the necessary starting approximations.
Among the seen measurements, the PTT middle was chosen, meaning the distance between the peak of the R wave of the ECG signal and the corresponding maximum slope point of the PPG signal, as shown in Figure 4.19.
Before implementing the steps leading to the PTT estimate, it is imposed that a maximum slope point must be contained between two consecutive R peaks, the first considered and its next.

$$t_{R-peak}(i) < t_{PPG-peak}(i) < t_{R-peak}(i+1) \tag{4.4}$$

Once the validity criterion has been passed, the following steps are applied:

1. Individuation of the R-wave peaks of the ECG signal through the method discussed in the previous section. The result of this first step is reported in Figure 4.16;

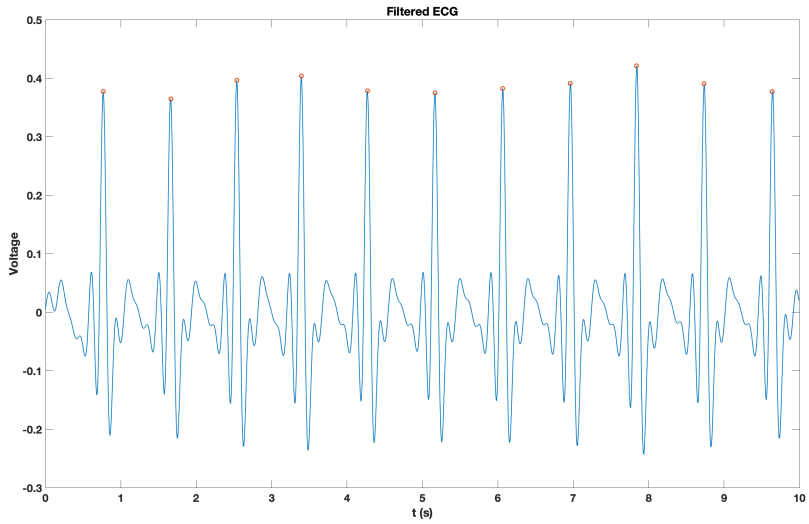

Figure 4.17: R peak from filtered ECG signal.

2. PPG signal derivation. This step is necessary because the inversion points relative to the maximum slope of the signal can be obtained from the maximum of the derivative waveform of the PPG signal. For graphic convenience, points below zero are set to zero (Figure 4.17);

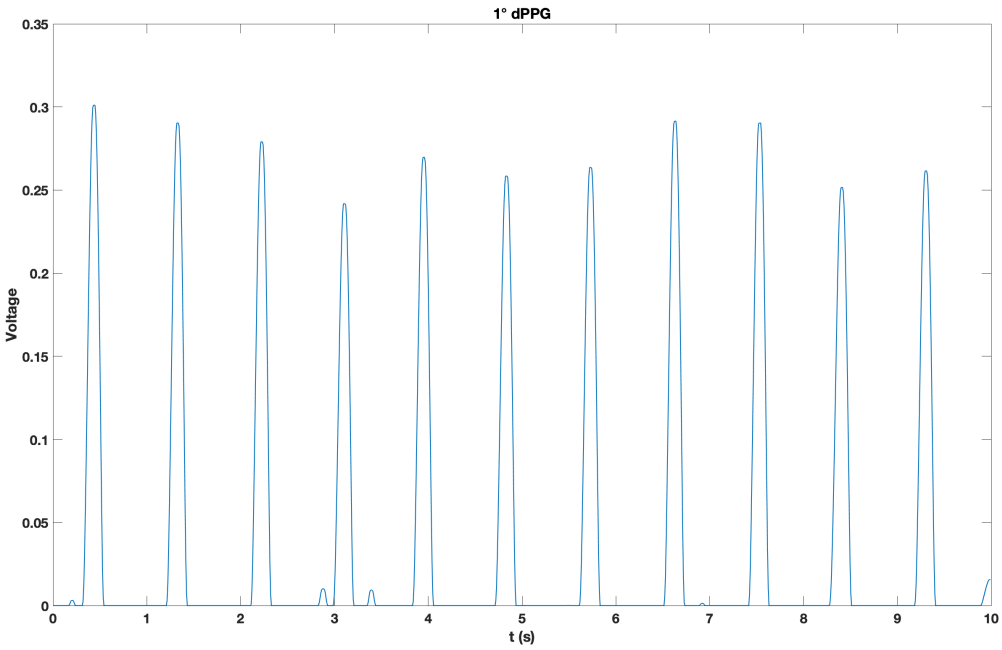

Figure 4.18: 1th dPPG signal with points below zero are set to zero.

3. Identification of the maximum points of the derivative signal by imposing a threshold equal to 30% of the maximum of the embankment, with an interval between two successive peaks around 0.33 seconds;

4. It is verified that the first R-wave peak is before the maximum slope point of the PPG signal, if not, the calculation starts from the second point;

5. Finally, calculation of the distance between the peak R and the maximum slope identified in the maximum of the derivative PPG signal (Figure 4.19) for the points that comply with the initial condition within the window and calculate its average value, if not, the calculation starts from the second peak.

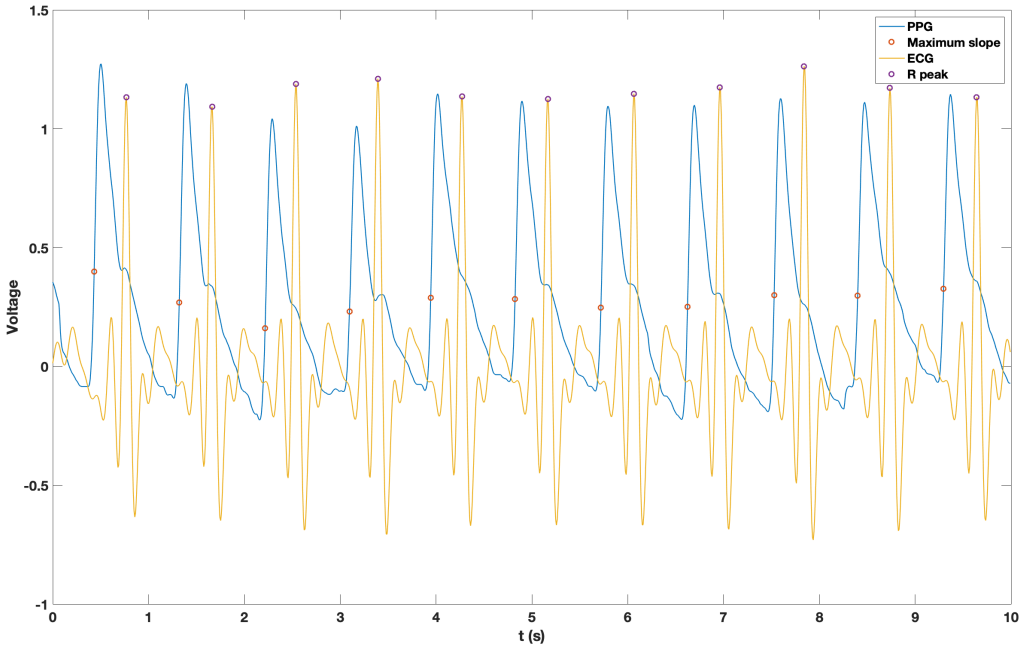

Figure 4.19: ECG and PPG signals with the respective points with respect to which the PTT measurement is performed.

Pressure values extraction

The pressure values relating to all the intervals of the various signals are obtained in the form of an average value with respect to the window. Pressure measurements concern both the SBP and the DBP, but also the mean arctic pressure (MAP), obtained through the Thirty-Three percent formula [77] reported in Equation 4.5.

$$MAP = \frac{2 * DBP + SBP}{3} \tag{4.5}$$

Which is simplified as:

$$MAP = DBP + 0.33 * PP \tag{4.6}$$

considering PP as:

$$PP = SBP - DBP \tag{4.7}$$

To detect the maximum and minimum points of the ABP signal, it is proceeded as in the PPG systolic point individuation, with the exception of the fact that given the invasiveness of the sampling, the only artifacts that were presented were those caused by the movement of the catheter. Since these kind of artifacts at the beginning, it was possible, at this point, to work directly on the signal, without other intermediate steps.

1. Identification of the maximum points of the ABP signal by imposing a minimum threshold equal to half of the maximum of the signal inside the window and an interval between one peak and the next one of approximately 0.33 seconds. These points are related to SBP.

2. Identification of the minimum points between two systolic peaks, related to the DBP values.

3. MAP estimation through Thirty-Three percent formula.

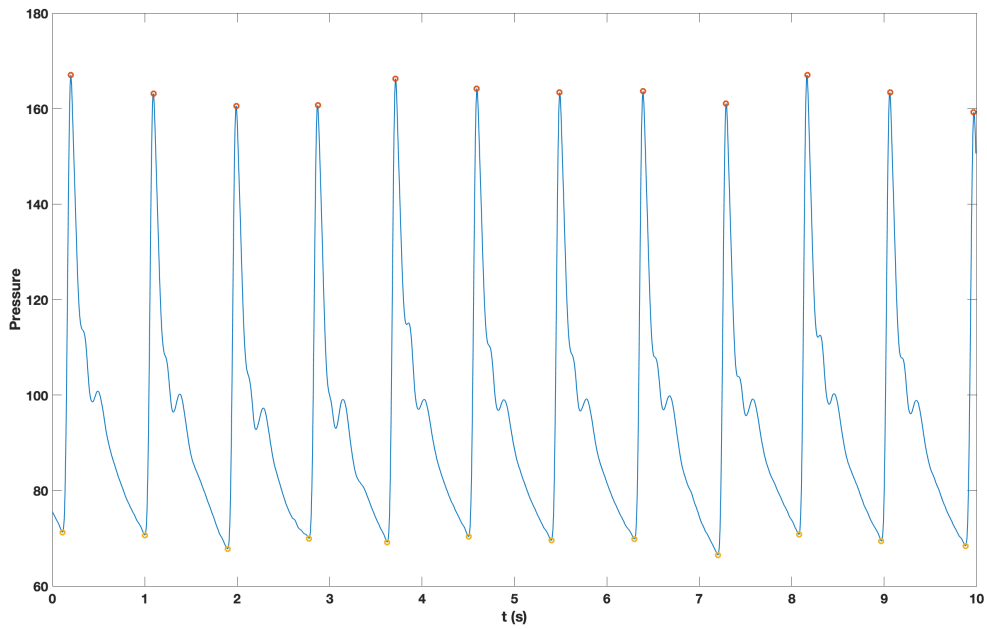

Figure 4.20: ABP signals with the respective SBP (red) and DBP (yellow) points.

4.3 Data organization

Data organization represented one of the central phases of the work.
Each subject whose signals have been used to extract the features is identified by an integer number to replace the ID (list of subject in Appendix A) and the matrices containing the samples of each feature are concatenated to obtain the complete dataset.

Table 4.3: Regression Model characterization.

	Mean	σ
PIR	3.23	0.55
PTT	0.54	0.46
HR	13.21	2.82
HI	2.61	9.01
HL	2.78	9.09
meu	0.98	3.34
SUT	0.22	0.06
DT	0.36	0.20

By observing the central values reported in Table 4.3 assumed by the different features and their deviation from these values, to the features have been imposed reasonable margins to be considered physiological valid, given the impossibility of excluding errors made during data collection.

- PTT:

$$0.2s < PTT < 0.8s \tag{4.8}$$

  the minimum limit is imposed because the pressure wave will not reach the finger in a shorter time than the electrochemical conduction that leads to the contraction of the heart, while the maximum limit is imposed if there are anomalies on the path that lead to overestimate the time lapse;

- SUT e DT:

$$0.1s < SUT < 0.4s \tag{4.9}$$

$$0.1s < DT < 0.8s \tag{4.10}$$

  Obtained by observing the mean values of all samples and their respective standard deviations;

- SBP and DBP values:

$$SBP > 80mmHg \tag{4.11}$$

$$30mmHg < DBP < 130mmHg \tag{4.12}$$

  Due to uncontrolled point changes in the signal that led to incorrect estimation within the observation interval according to literature.

After this phase, Subject 18 was eliminated, because it did not presented a sufficient number of samples. As a consequence, 23 subjects were then analyzed, for a total of more than 10600 samples.

4.4 Feature selection

The next step concerned the correlation between the information taken by the features. This step is necessary because a strong correlation between two features may be due to the fact that one of the two features generates the other or the two of them are connected [78]. In this case, if the increase or decrease of one feature affects the increase or decrease of the other one respectively, the two features are said to be linearly related. It is possible that two features are correlated in a negative way, that is, the growth of one corresponds to a decrease of the other and vice versa.
Figure 4.20 shows an example of correlation plots.

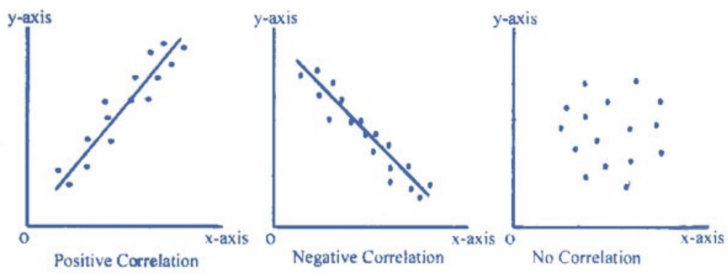

Figure 4.21: Example of attended correlation plot [79].

The correlation score is between -1 and 1 and the values that lead to affirm a linear correlation between different features are around 0.5 - 0.7. Values around the unit mean a strong linear correlation, while values around -1 a strong negative correlation.
If all the features within the dataset were strongly correlated with each other, a problem called **multicollinearity** would be encountered [78]. In this case, a redundancy of information having the same origin or that carries the same information has an impact on the quality of the estimate, also giving misleading results.
If the features were not correlated, it would mean that their values are not influenced by common factors and that therefore each of them brings additional information that can be useful for classification/regression purposes.
Although in the case of tree decision (presented in the next section) this problem does not alter the measurements, since once it decides to separate its limb, it will choose only one of the most correlated features. It was decided to investigate the correlation graph of the features through the correlogram and a limit of 0.8 was imposed, to limit the feature strongly correlated.

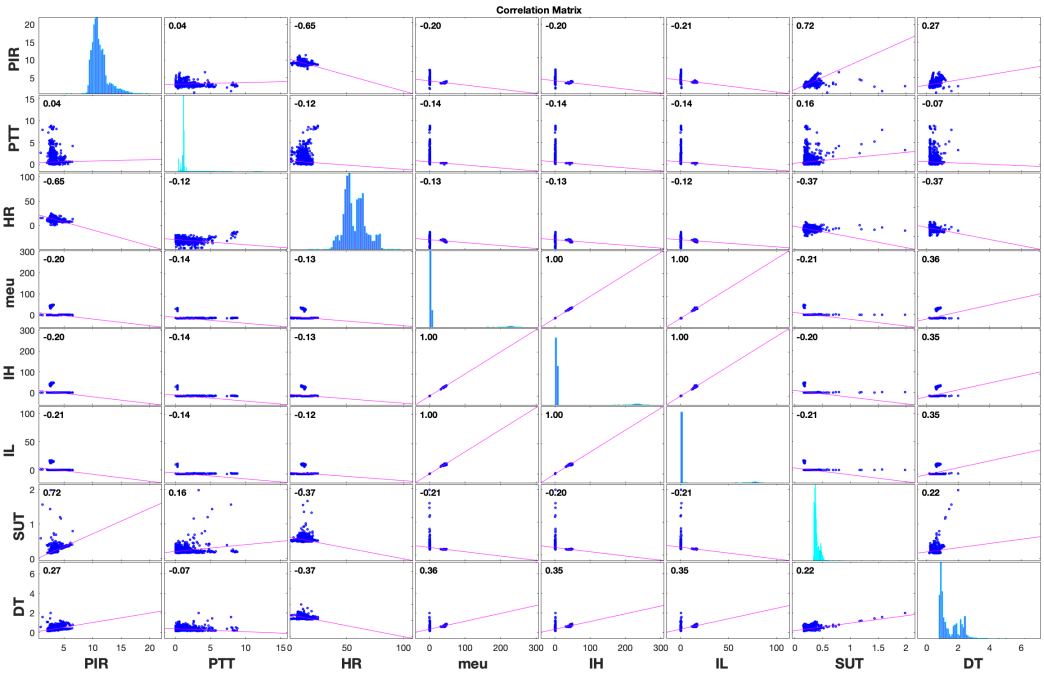

Figure 4.22: Correlation graphic of the feature extracted.

As shown in the graph, the features that had the highest correlation values between them are:

- the meu with the IH and IL respectively: $r = 1.00$ $r = 1.00$;
- the IH and IL: $r = 1.00$.

This means that these three features bring the same information and that they influence each other. For this reason just the meu was preserved.
As a result of this phase, we will therefore maintain a total of 6 features, optimizing the information brought by the available dataset:

1. PIR;
2. PTT;
3. HR;
4. meu;
5. SUT;
6. DT.

4.5 Regression analysis

Regression analysis consists in creating a predictive model in which the relationship between dependent (target) and independent (predicted) variables is investigated [81].

Figure 4.23: Example of regression analysis to extract the fitting curve (red) of target data (black) [81].

Assuming that the data y depend on a number of variables i: $f = f(x1, ..., x_i)$, it is possible to derive the relationship between the two of them through the regressive model that allows to minimize the residual error or the quadratic error between prediction and target [71]. A linear regression model can be [71]:

$$\hat{y} = \beta_0 + \beta_1^T * x + \epsilon \tag{4.13}$$

where β_0 and β_1^T are the scalar and the vector to choose, while ϵ is the error between the target and the model prediction.
Minimizing the mean square error means minimizing the following expression:

$$L = \sum_i (y - \beta_0 - \beta_i^1 x_i)^2 \tag{4.14}$$

In this configuration, by minimizing the mean square error between the estimated values and the reference target, the fitting quality operated can be estimated and the best coefficients of the linear system can be chosen [71].

4.5.1 Random Forest

For this work the Regression model adopted is the **Random Forest**.

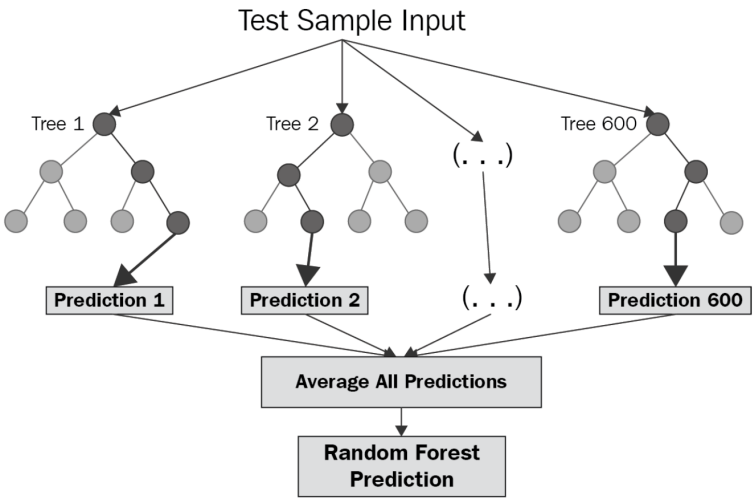

Figure 4.24: Structure of Random Forest [80].

The Random Forest is the result of a set of decision methods whose prediction result is the result of the combination of the different predictions of the individual decision-makers within the system [40]. In making the decision, each tree works independently in parallel with the other trees without the slightest exchange of information between them.
In the realization of its structure it uses random subset of the dataset to train and identify the one that minimizes the quadratic error of each tree [81]. This random choice allows to minimize the risk of model overfitting. In addition to this, to prevent the system in its construction from focusing on a single feature of the dataset, a limited number of functions are used for each node. By imagining the space on which the dates are distributed, the use of multiple decision-makers allows to identify the plans with respect to which they are linearly correlated with the output. The advantage of using this regression method lies in its high accuracy and ease of implementation. However, it is extremely sensitive to the dataset, therefore, in the case of noisy datasets, it can be affected in terms of performance, also leading to the overfitting of the model.

Model implementation and testing

The available dataset consists of a total of approximately 10,600 samples for each of the 6 features selected from the previous phase. To implement the model, the dataset is separated into two parts:

1. the first to train the system, the 70% of the subjects, therefore in the first approximation 16 subjects;
2. the second to test the accuracy achieved by the algorithm, equal to the remaining 30%, i.e. 7 subjects.

The subjects dedicated to train and to test the system were randomly chosen from those present in the dataset. For each of the measurements that have been taken, a single regressive Decision Tree corresponds, all having the same settings showing in Table 4.4.

Table 4.4: Regression Model characterization.

Model	**Method**	**# Tree**	**Leaf Size**
Random Forest	Regression	200	5 (Default)

Table 4.5 shows the identifiers of the subjects used to train the system and the ones of those used to test it.

Table 4.5: Subject selected to implement training and test phases.

Training	**Test**
Subject 8	Subject 9
Subject 17	Subject 22
Subject 24	Subject 21
Subject 4	Subject 7
Subject 20	Subject 23
Subject 11	Subject 10
Subject 12	Subject 5
Subject 1	
Subject 3	
Subject 2	
Subject 13	
Subject 14	
Subject 19	
Subject 15	
Subject 6	
Subject 16	

After having chosen which subjects to use in the two phases of the regression, the data of the two groups were randomly mixed to minimize the risk of excellent premises due to slightly variable trends in pressure values.

Chapter 5

Results

This chapter reports the results achieved by the algorithm implemented for the estimation of the three BP values. As explained in the previous chapter, the available dataset has been filtered and highly correlated features have been eliminated to avoid a redundancy of information. To implement the regression, the data set is divided into two parts, respectively intended to train and to test the network, with respect to the subjects treated. After training the system,the samples of subjects that were not used in the first phase are used to test it.
The following parameters have been used to evaluate the fit quality achieved by the algorithm:

- **Route-Mean-Square error** (RMSE), which represents the quadratic mean difference between the target and the estimated value [82]. It is useful to measure the quality of the prediction. In the case of n predicted values, with y corresponding to the vector of the target values and $\hat{y}$ to the vector of the values predicted by the algorithm, the RMSE is obtained according to Equation 5.1 [82].

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5.1}$$

- **Standard Deviation** (σ), is a statistical dispersion index that measures the variability of a population. In general, it is used as a means of measuring the dispersion of data with respect to a defined position index, which is typically represented by the average value [82].

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}} \tag{5.2}$$

- **Determination coefficient** (R^2), The coefficient is defined as the proportion of the estimate variance of the regression. This makes it useful for evaluating

the quality of the estimate of the variables (output) dependent on the independent variables [83]. It can assume values in the range between 0 and 1, which indicate respectively a very bad and an excellent fitting of the data by the implemented regression system.

The graphs of the SBP, DBP and MAP pressure values obtained for each signal window are shown below (Figure 5.1). A part of these signals was used for training (blue) and a part for testing (red). In order to obtain excellent results, the data used in the training phase must be as representative as possible of the total set of data in order to obtain a reliable forecast.

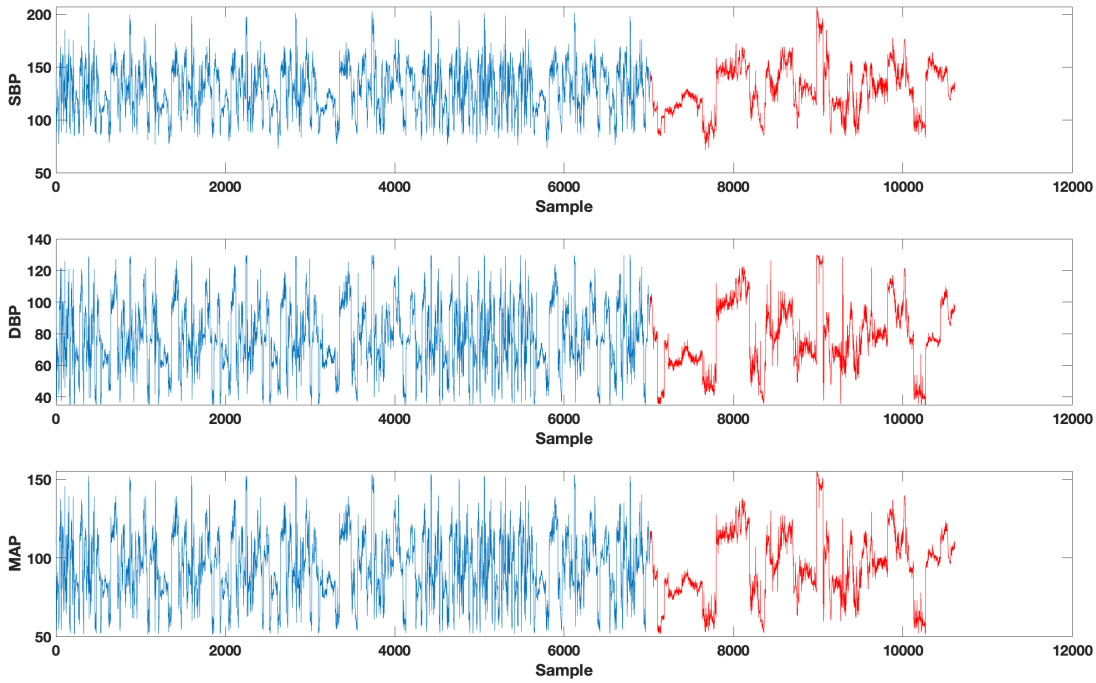

Figure 5.1: Graphs of the outputs for used in the training phase (blue) and of the outputs used in the test phase (red) to verify the quality of the fitting.

Figures 5.2, 5.3, 5.4 show the values predicted by the regression models (red), corresponding to the different pressure measured from the ABP signals window (blu).

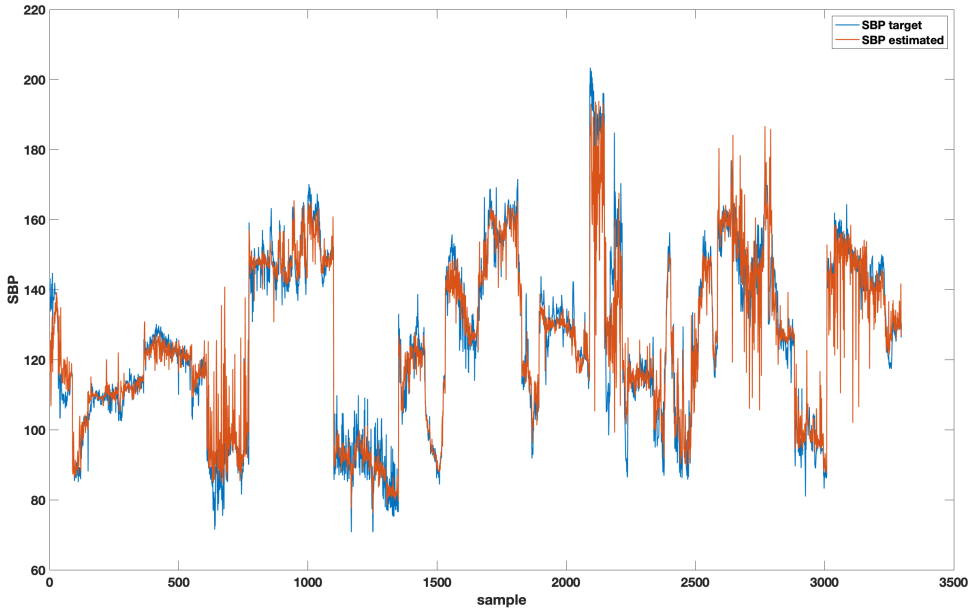

Figure 5.2: Estimated SBP superimposed on target values obtained from the ABP signal.

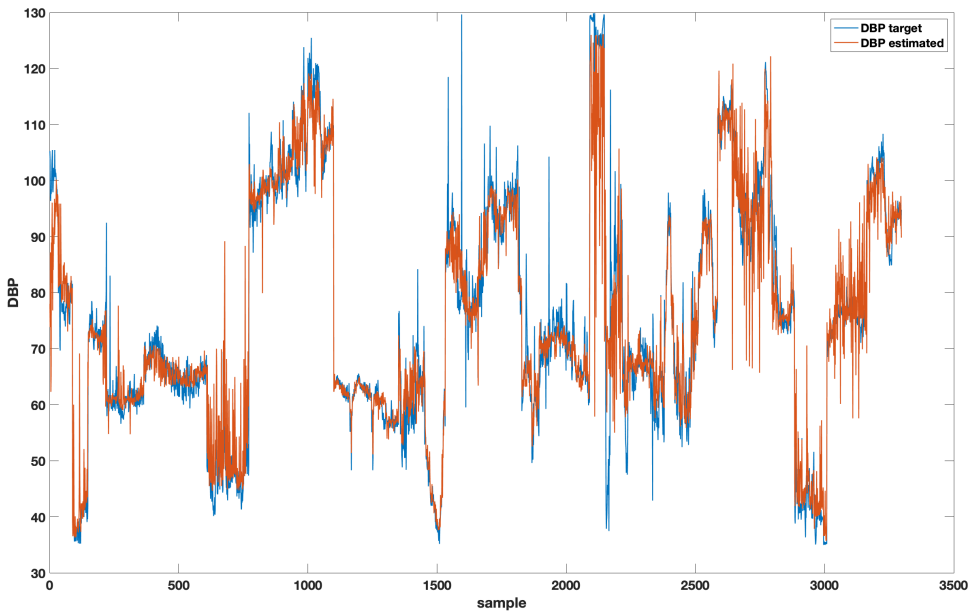

Figure 5.3: Estimated DBP superimposed on target values obtained from the ABP signal.

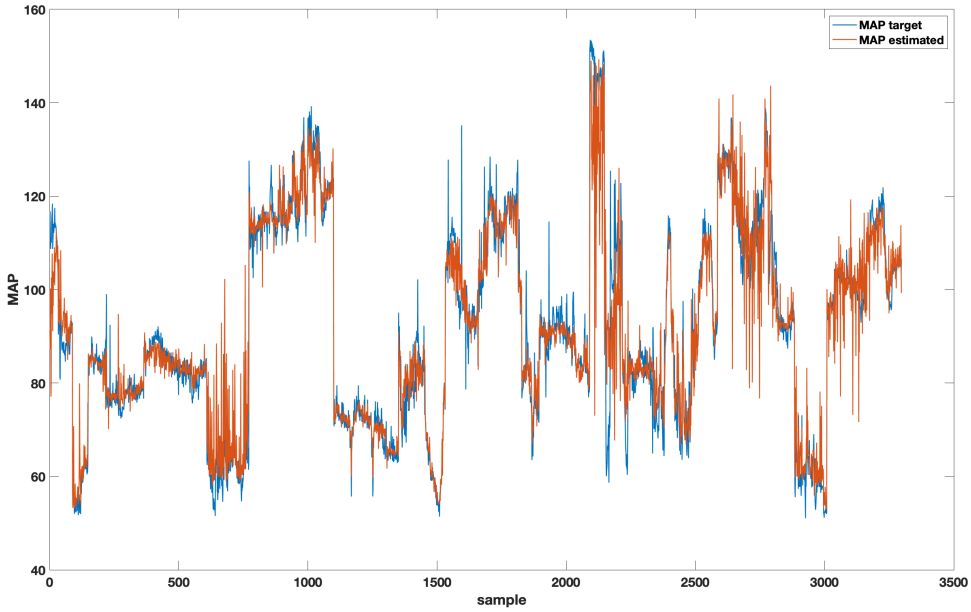

Figure 5.4: Estimated MAP superimposed on target values obtained from the SBP and DBP values with Thirty-third percent formula.

Table 5.1: Performance reached in terms of accuracy and variability of the regression system

Pressure measured	**RMSE (mmHg)**	σ **(mmHg)**
SBP	8.36	± 10.5
DBP	6.4	± 9.45
MAP	6.72	± 9.45

The Table 5.1 shows the result of the predictions in term of RMSE and σ. The best result in term of readout error and variability is reached by the DBP.

The determination coefficient was calculated to evaluate the statistical correction of the model respect the data available. The values obtained for the three estimators are shown in Table 5.2. For all three cases, significant values were reached, demonstrating a good adaptation of the data to the regression systems.

Table 5.2: Determination coefficient reached by the three regression systems.

	SBP	**DBP**	**MAP**
R^2	0.89	0.90	0.90

As reported, the best results in terms of reduced error have been achieved for the estimation of diastolic pressure. This is evidenced by the Figures 5.5, 5.6, 5.7 which show the error distribution for the individual estimators of SBP, DBP and MAP respectively. As it can be seen from the histograms, most of the errors made by the estimators are concentrated in the range around ±10 mmHg, with few values far outside this range, especially for SBP.

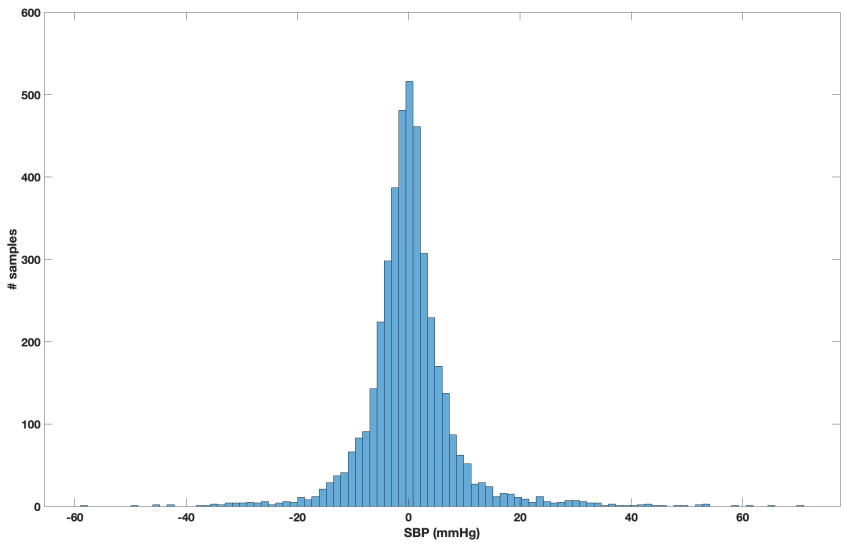

Figure 5.5: Error histogram of SBP.

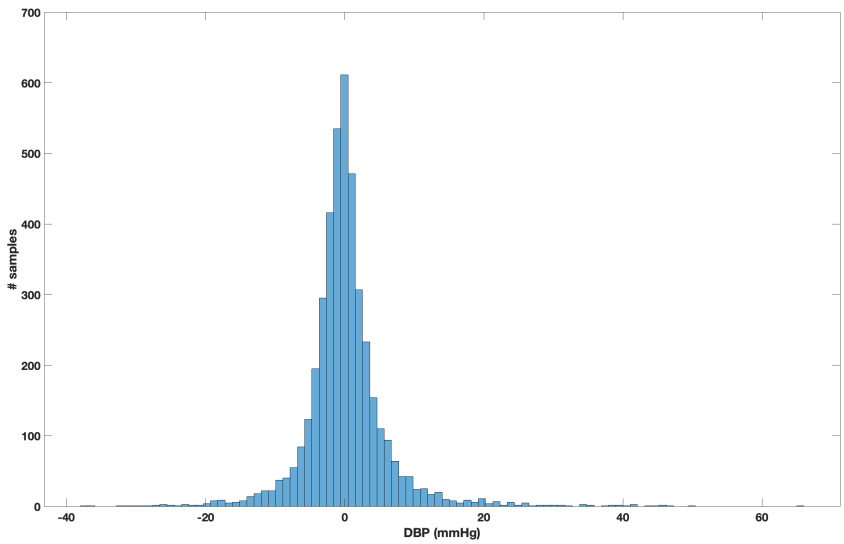

Figure 5.6: Error histogram of DBP.

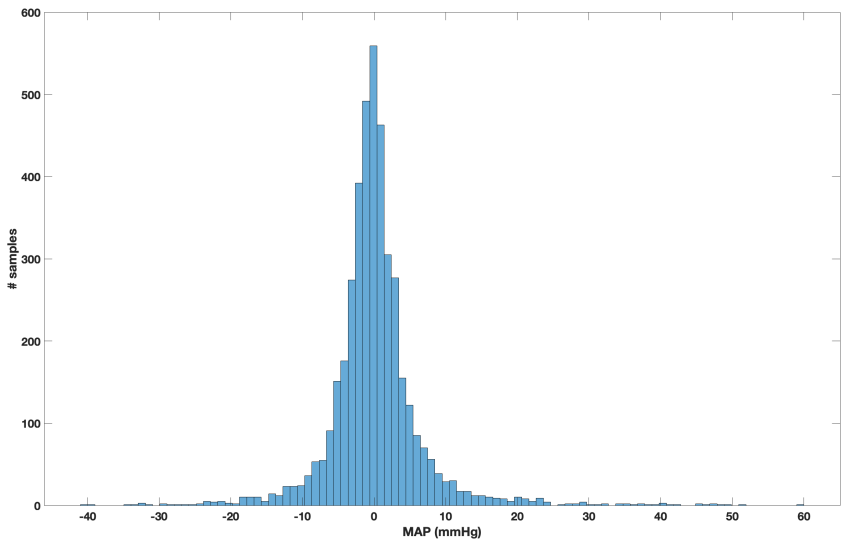

Figure 5.7: Error histogram of MAP.

Observing the scatter plots in Figure 5.8, it is possible to assess that the estimates made for all the implemented models manage to concentrate along the central straight line, indicating an almost linear correlation between the target and the estimated values.
In special, central pressure values for each predictor have good precision related to target pressure values expect for the extremes, which present high variation. This is due to the insufficient number of samples to prepare the system toward this possibility, as can be seen in Figure 5.9, which represent all BP values used for training and testing the systems. The graph shows high number of samples for central values but not enough for the high ones. This is traduced in an accentuated error compared to the average related the central values.

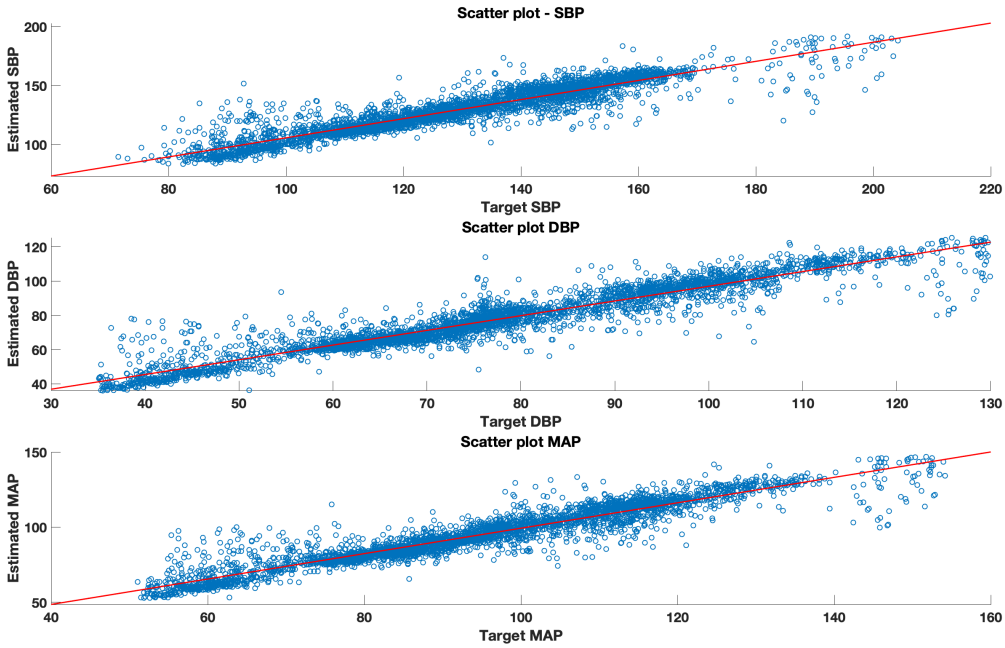

Figure 5.8: Three estimator's Scatter plot.

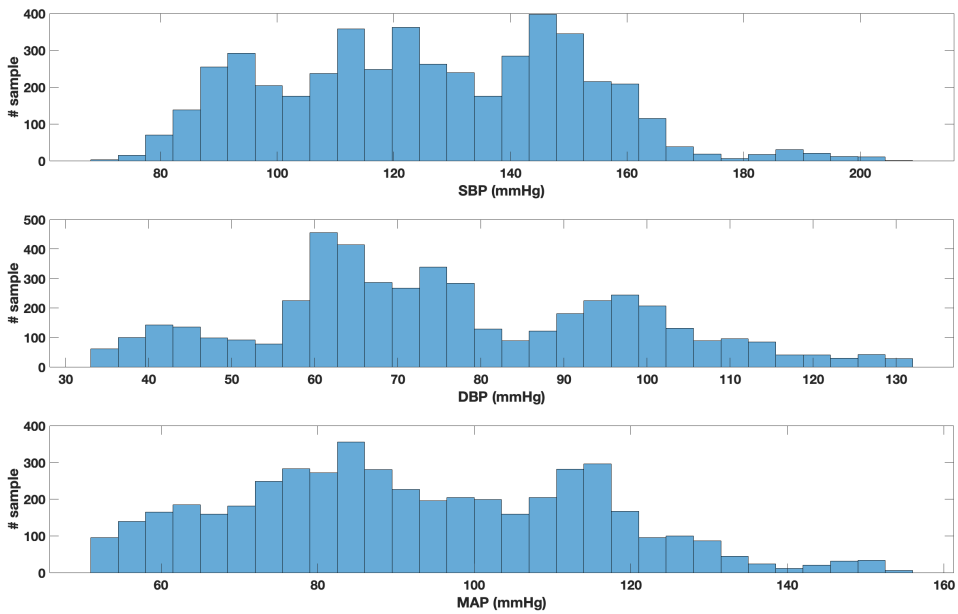

Figure 5.9: Distribution histogram respectively of all data-set SBP, DBP and MAP values.

By examining the trend of the estimated blood pressure samples with respect to the targets, it is possible to observe that there are cases in which there is a very similar trend (Figure 5.10) and cases in which the regressive system have more difficulties to estimate the correct trend (Figure 5.11).

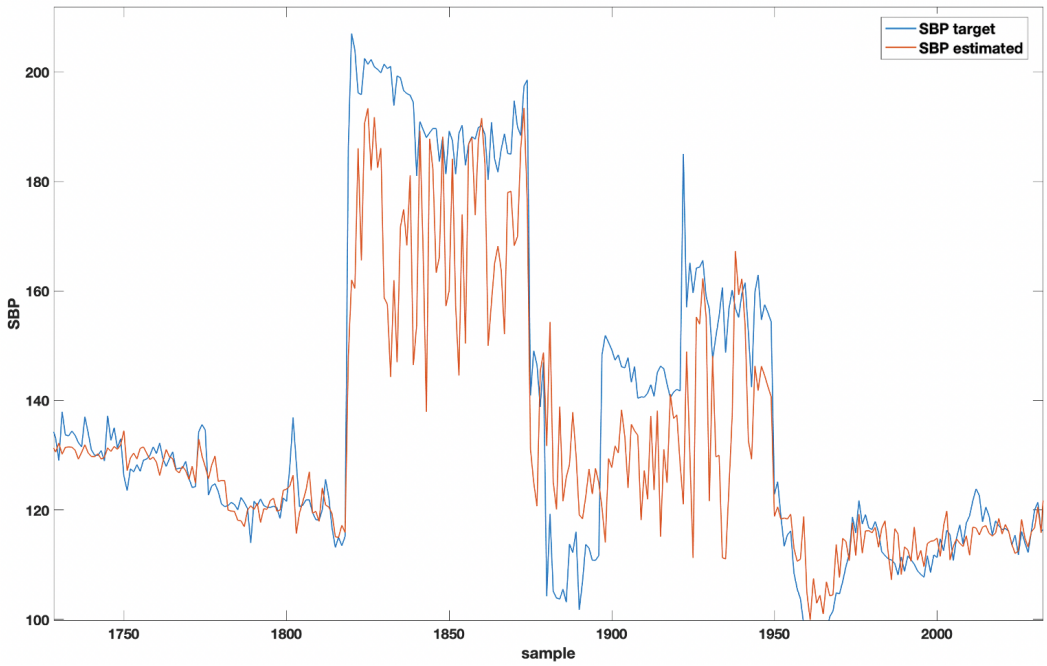

Figure 5.10: Estimate range of SBP values of test-set in which the system makes large errors (> 10 mmHg).

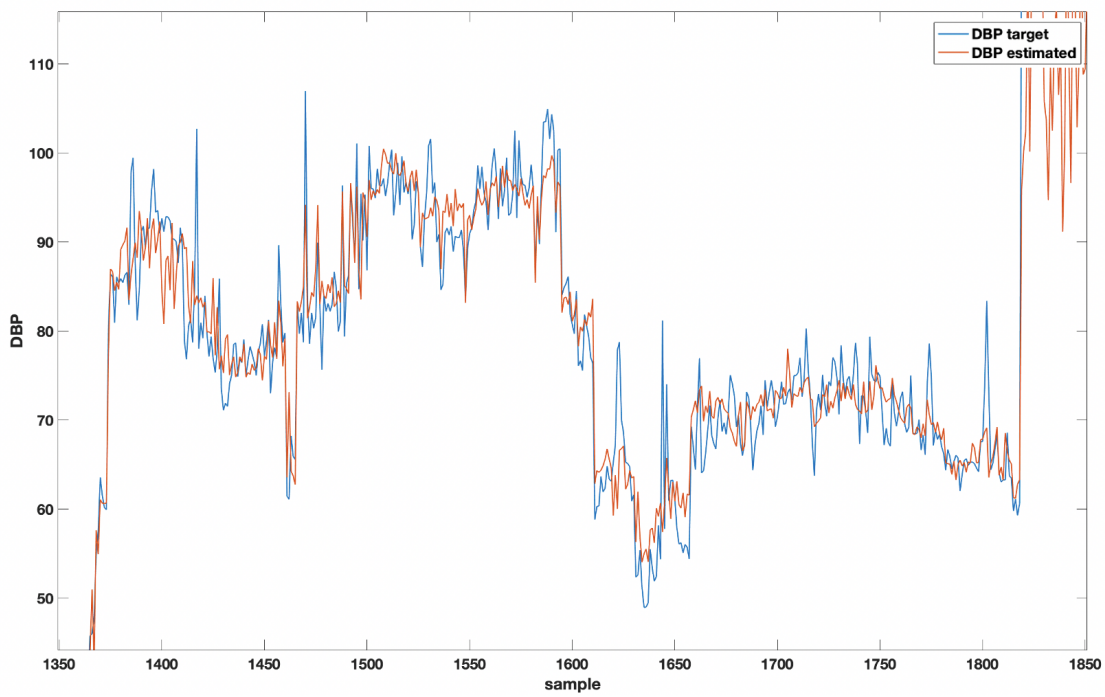

Figure 5.11: Part of the SBP signal of test-set in which the predictor is able to estimate with contained error with respect to the target values.

Figures 5.2, 5.3, 5.4 represent the behavior of the system with the data mixed for the individual subjects randomly selected for the test session.
Now it is investigated how the system behaves with respect of individual subjects. As a consequence, the performance evaluation scores were obtained individually rather than analyzing the entire test set. The results obtained for the subjects forming the test set are reported in Table 5.3.

Table 5.3: Performance reached in terms of route mean square error and standard deviation of the regression system for individual subjects of test-set.

Subject	**SBP** RMSE (mmHg) $\pm\sigma$(mmHg)	**DBP** RMSE (mmHg) $\pm\sigma$(mmHg)	**MAP** RMSE (mmHg)$\pm\sigma$(mmHg)
Subject 9	5.98 ± 5.27	6.75 ± 8.15	6.03 ± 7.11
Subject 22	2.36 ± 0.68	1.57 ± 0.44	1.50 ± 0.45
Subject 21	7.98 ± 6.78	5.19 ± 4.21	5.74 ± 5.02
Subject 7	5.76 ± 13.26	4.48 ± 11.12	4.50 ± 11.61
Subject 23	12.75 ± 10.15	9.42 ± 8.22	10.22 ± 8.83
Subject 10	6.41 ± 10.63	5.24 ± 8.51	5.40 ± 9.16
Subject 5	9.30 ± 10.15	6.94 ± 10.20	7.44 ± 9.84

For all subjects, a limited error is observed except for Subject 23, which presents a wide range of pressure values with high number of samples that reaches high pressure level (in some sections the SBP exceeds 200 mmHg and the DBP exceeds 110 mmHg) as shown in the histogram (Figure 5.12), compared to the subjects with less RMSE and standard deviation (Figure 5.13,5.15). The error could have been limited by training the system with data that involve a larger group of subjects and that covering a wide range of cases. This was not possible in this study because of the nature of the MIMIC database, which was born with another objective.

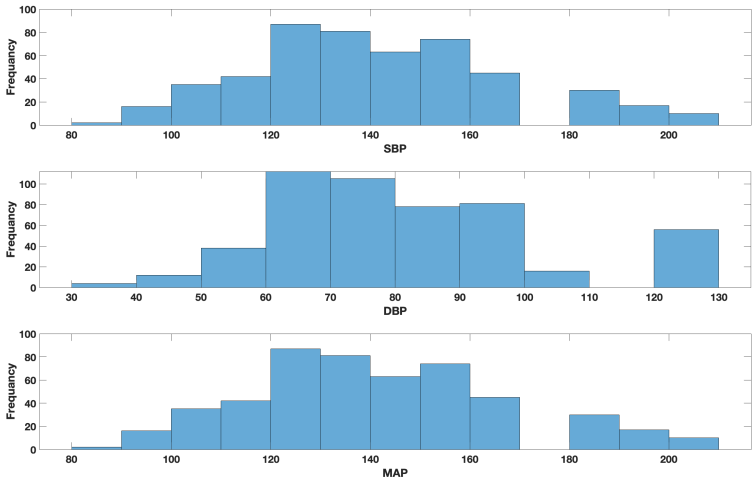

Figure 5.12: Distribution of BP values for Subject 19.

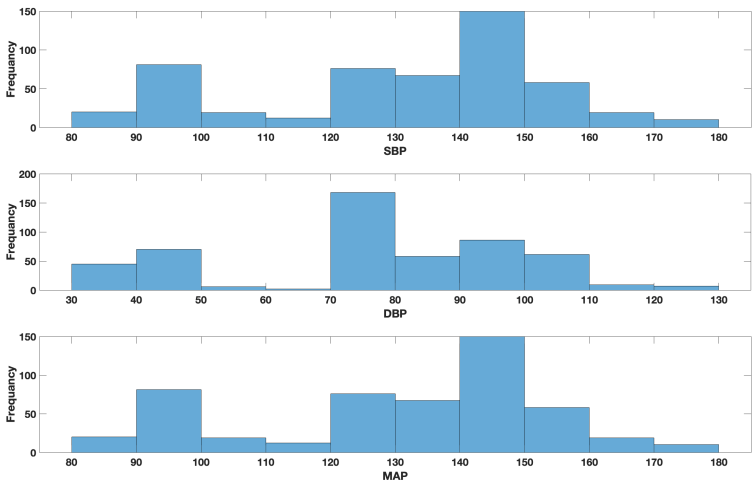

Figure 5.13: Pressure histogram for Subject 21.

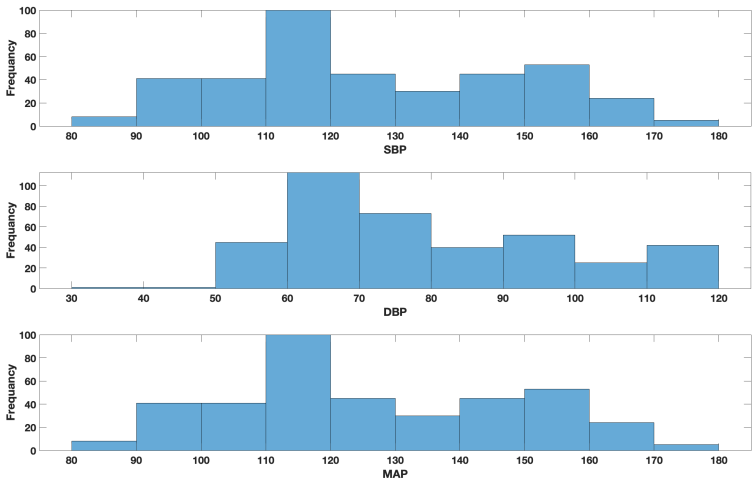

Figure 5.14: Pressure histogram for Subject 10.

Finally, the weight of the single feature was investigated to discover which one have major decision importance. In other words, this part served to understand which features had the greatest decision-making influence, affecting the output more. This phase is also necessary to refine the estimate of those parameters that can most influence the quality of the model. The obtained results are represented in Figure 5.15.

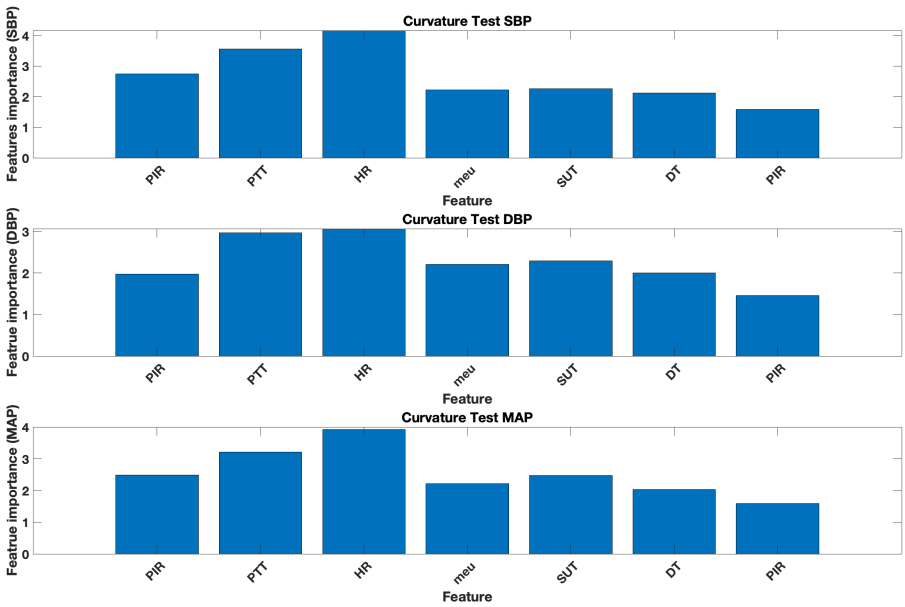

Figure 5.15: Histogram of importance of single features.

As it can be seen from the Figure 5.15, the most important features are those

reported in the data organization phase, i.e. the HR, the PTT and the PIR and each of these has a specific role in capturing information from the spectrum of the pressure signal, as reported in the previous sections.

5.1 Validation

Given the size of the dataset and observing the prediction performance on the test set, it was decided to validate the system to identify problems such as overfitting and inability to generalize the system in the presence of data different from the ones forming the training set. The **K-fold cross validation** method was used to implement this phase.
The K-fold consists in dividing the dataset into K subset and repeating the training and classification operation K times [84]. For each iteration, a subset of the available K is used to test the system, while the remaining K-1 are used to train it [84].

A K equal to 10 is used. Therefore the data-set (randomly mixed) is divided into 10 subsets and in each turn a different subset is used in the test phase, while the remaining 9 are used to train the different regression systems.
To evaluate the results, the RMSE is calculated at each iteration and if the error values between all the sessions are similar, there is no overfitting by the system. The obtained results are reported in Table 5.4.

Table 5.4: Performance reached in terms of RMSE and standard deviation of the regression system for individual subjects of test-set.

SBP RMSE (mmHg)	**DBP** RMSE (mmHg)	**MAP** RMSE (mmHg)
8.06	6.22	6.41
8.34	6.23	6.56
7.91	6.26	6.49
7.78	6.36	6.43
7.85	6.26	6.43
7.77	5.08	6.06
7.27	5.92	5.99
8.26	6.57	6.77
8.18	6.63	6.82
8.65	6.88	7.13

From Table 5.4 it is possible to assess that the error settles about the same value with minimal variations in all the three cases. This indicates that the system achieves a good level of consistency and repeatability of the performances achieved.

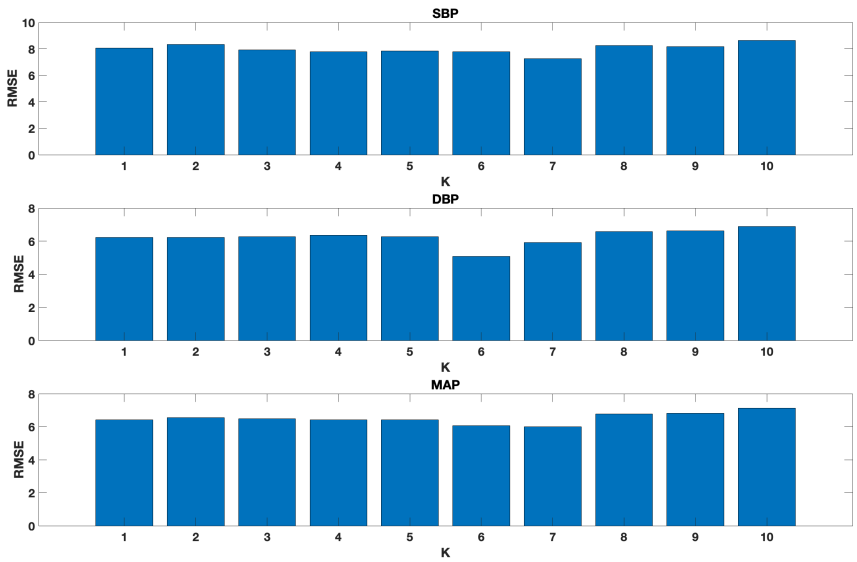

Figure 5.16: Error distribution for each K.

The k-fold cross-validation is used also to find the best configuration of the Forest in term of number of tree. The number of tree varies between 50 to 300 and for each configuration the k-fold is executed. For each session with different tree configuration the mean RMSE is taken and finally, the performances are studied to choice the best possible configuration. Table 5.5 shows the mean RMSE for each cross-validation session. As it is reported, DBP and MAP reach the best one with a number of tree equal to 200 and the RMSE of SBP varies negligibly between the different configurations. For this reason the best configuration has a number of 200 tree for each BP values calculation.

Table 5.5: K-fold cross-validation to investigate the best configuration.

Pressure	$\overline{\text{RMSE}}$ # Trees = 50	$\overline{\text{RMSE}}$ # Trees = 100	$\overline{\text{RMSE}}$ # Trees = 150	$\overline{\text{RMSE}}$ # Trees = 200	$\overline{\text{RMSE}}$ # Trees = 250	$\overline{\text{RMSE}}$ # Trees = 300
SBP	8.07	8.01	8.00	7.99	7.99	7.97
DBP	6.35	6.29	6.28	6.28	6.28	6.28
MAP	6.60	6.54	6.55	6.50	6.52	6.51

The last step investigated the statistical correlation between the measurements obtained by invasive detection method for ABP and the measurements obtained by regression, investigating the PWV and the morphological characteristics of the PPG signal.
To this purpose, the **hypothesis test**, or **T-test**, was carried out, in order to investigate if the population of the data estimated through the regression algorithm is statistically relevant with the one obtained with invasive detection method. If they

are relevant, it means that they bring two different information and therefore these two methods lead to unrelated results. Otherwise, they carry the same information and the achieved results are similar. The results of the T-test are reported in Table 5.6.

Table 5.6: T-test on estimated data (PWV) vs. target data (ABP).

Pressure	**H**	p
SBP	0	0.82
DBP	0	0.61
MAP	0	0.68

H is the null hypothesis and p is the p-value. The null hypothesis is not rejected for all regression systems ($H = 0$), with a level of significance (p-value) far above the threshold $\alpha < 5\%$ .This means that the two methods carry statistically similar information and that therefore the non-invasive method implemented through the study of the PWV and the use of prediction algorithms can prove to be a valid alternative to the classical methods.

Chapter 6

Conclusions

6.1 Conclusion of present work

The implemented method in this thesis focused on estimating systolic, diastolic and average blood pressure through the study and analysis of Pulse Wave Velocity as an alternative method to the "classic" ones, that, for many years, have been known as more reliable. The possibilities of estimating the different blood pressure values were mainly studied through the Pulse Transit Time, which required the implementation of a method for identifying the QRS complex of the ECG signal, in particular the R-peak, and the Photoplethysmogram signal. From the latter morphological characteristics are extrapolated, such as the Photoplethysmogram Intensity Ratio. This one, together with the PTT, have shown in literature the ability to capture the information contained in the spectrum of the ABP signal [40] and this is also confirmed by the results achieved by the algorithm implemented in this thesis.
One of the main advantages of the proposed study, compered to those seen in literature, is the possibility to perform measurements without the necessity of using a calibration system for each different subject.
Some conclusion can be summarized as follow:

- The algorithm lends itself to an easy and quick estimate of the different blood pressure values compared to a large dataset. Ease of implementation was one of the objectives pursued to limit computational cost, in an application perspective on wearable devices.

- As far as the explored signals are concerned, the algorithm has excellent analysis capabilities of the treated signals. Despite this, the algorithm is not appropriate to noisy signals and to signals affected by marked interference, due to the choice to investigate the method in a context of normal operation to study its effectiveness on the signals involved in a context.

- The observation window of the signals is too short and is not appropriate for the type of estimate that was wanted, as the blood pressure values vary with a time interval of about a few minutes. This is due to the reduced amount of data in the MIMIC database which do not show a long enough observation interval without interference, since the purpose of this database is not the same to the one of this work. In addition to the limits of the model itself, the limits due to the treated database are added, since MIMIC is a collection of data from the Intensive Care Unit (ICU) department, in which patients are subjected to drugs that alter blood pressure values.

- In terms of accuracy and variability of the results of the algorithm, encouraging performances have been achieved, although, for clinical applications, they do not meet the performances imposed by the American National Standards by the Association for the Advancement of Medical Devices (AAMI), the organization responsible for the safety and accuracy of medical devices. The standard imposes an error of the device below 5 mmHg , with a standard deviation of 8 mmHg compared to the already approved devices available on the market [85]. However, this study is not born with the purpose of developing a medical device, but rather a tool for the consumer, for whom regulations are less restrictive and they require less amount of time for their approval.

- One of the limitations of the proposed thesis work is the need to implement 3 different regression systems, one for each individual pressure studied, especially treating the minimum and maximum pressure as if they were unrelated, while this is not true in reality.

- Another innovative approach of this thesis work concerns the in-depth analysis of the regression system applied to subjects of whom, once trained, it does not know any information regarding blood pressure values, age, sex, physical characteristics and without an initial calibration.

It is then possible to conclude that the system in its current state returns encouraging results if it is intended for the consumers, thinking to develop a wearable device capable of obtaining ECG and PPG signals with good accuracy and returning blood pressure values during daily activities.

6.2 Future work

Based on the results achieved in this thesis work, several changes and improvements are encouraged:

- Tailored with data collected from ECG, PPG and BP by subjects in different situations (healthy, sick, young, elderly) with the direct purpose of studying BP through PWV.

- Exploring other features of PPG signal, such as the T1 interval (the distance between the systolic peak and the diastolic peak), which was not possible to study due to the quality of the signals extracted from the subjects present in the database.
- Exploration of other regression methods, possibly also nonlinear, in order to compare the results. In addition, algorithm that returns multiple outputs simultaneously for different kind of pressure should be considered.
- Testing the method on devices from the market to improve the develop model in order to overcome the problems that concern real-time acquisitions, such as the synchronization of the devices and the quality of the obtained signals.

Appendix A

List of subject from MIMIC II/III

Below, in Table A.1 is the complete list of subjects present in the MIMIC II / III database used in this thesis work to implement the algorithm. The subject ID is represented by the numbers on the left while what is after the "_" indicates the specific acquisition of the subject. In some cases, multiple acquisitions were processed for the same subject.

Table A.1: List of subject and related acquisition.

Subject ID
3100033_0008m
3106149_0023m
3106150_0029m
3106152_0004m
3106152_0007m
3106152_0010m
3106186_0033m
3106263_0012m
3106263_0016m
3106380_0007m
3400086m
3400290m
3500192m
3500228m
3501358_0005m
3501441_0010m
3501441_0011m

Continued on next page

Table A.1 – continued from previous page

Subject ID
3501441m
3503726_0008m
3702767_0014m
3703802_0009m
3500015_0002m
3500015m
3500192_0002m
3503654_0013m
3702767m
3703856_0001m
3703856m
3703872_0004m

Bibliography

[1] "Wearable Technology Market by Device (Smart watch, fitness wellness devices, Smart glasses, Smart clothing), by Product type (Wrist wear, eyewear, hearables, neckwear, bodywear) and Geography," *Global Opportunity Analysis and Industry Forecast*, 2014-2022.

[2] Wilmore JH and Costill DL. (2005) Physiology of Sport and Exercise: 3rd Edition. - Champaign, IL: Human Kinetics

[3] Cambridge Dictionary.
https://dictionary.cambridge.org/dictionary/english/wearable

[4] Ecco come una app per il fitness può rivelare basi americane e percorsi dei soldati USA:
https://www.lastampa.it/esteri/2018/01/30/news/ecco-come-un-app-per-il-fitness-puo-rivelare-basi-militari-e-percorsi-dei-soldati-usa-1.33973562

[5] Gloria Martínez, Newton Howard, Derek Abbott, Kenneth Lim, Rabab Ward and Mohamed Elgendi, "Can Photoplethysmography Replace Arterial Blood Pressure in the Assessment of Blood Pressure?," School of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada;

[6] Shivayogi Hiremath, Geng Yang, Kunal Mankodiya, *Wearable Internet of Things: Concept, Architectural Components and Promises for Person-Centered Healthcare* - January 2014

[7] D. Girola Cardiologia fitness - Alea, 1999

[8] Mammalian heart and blood vessels.
https://courses.lumenlearning.com/os-biology/chapter/mammalian-heart-and-blood-vessels/

[9] How does the blood circulatory system work?
https://www.ncbi.nlm.nih.gov/books/NBK279250/?report=classic

[10] C. L. Stanfield - Fisiologia

[11] MIMIC database: https://mimic.physionet.org

[12] Shimmer user guide
http://www.shimmersensing.com/support/wireless-sensor-networks-documentation/

[13] What is blood pressure and how is it measured?
https://www.ncbi.nlm.nih.gov/books/NBK279251/

[14] Agnieszka Gorczewska, "Influence of sensor design and optical properties of tissue on the photoplethysmographic signal," Gdansk University of Technology, Department of Biomedical Engineering, 11/12 Gabriela Narutowicza Street, 80-233 Gdansk-Wrzeszcz, Poland

[15] Zanichelli, Scienze integrate. Franco Lucisano Editore, 2010

[16] Arterial blood pressure.
https://www.cvphysiology.com/Blood%20Pressure/BP002

[17] P. Salvi, Onde di polso. Springer, 2012.

[18] Arterial and aortic pulse pressure.
https://www.cvphysiology.com/Blood0Pressure/BP003.htm

[19] John Allen, "Photoplethysmography and its application in clinical physiological measurement", Regional Medical Physics Department, Freeman Hospital, Newcastle upon Tyne NE7 7DN, UK

[20] M. J. Hayes, Peter R. Smith, “Artefact reduction in photoplethysmography,” Appl. Opt. 37, 7437-7446 (1998).

[21] R. Gonzalez, A. Manzo, J. Delgado, J. Padilla, B. Trénor, and J. Saiz, “A computer based photoplethysmographic vascular analyzer through derivatives,” in 2008 *Computers in Cardiology*. IEEE, 2008, pp. 177–180.

[22] V. Jadav and H. Sitapra, “A novel approach for designing of pulse transit time measurement device for physiological applications,” *J. Inf*, vol. 2, pp. 779–782,

2013.

[23] Non-invasive methods BP measure.
https://www.researchgate.net/figure/Non-invasive-blood-pressure-monitoring-techniquesfig1322314441

[24] D. McDuff, S. Gontarek, and R. W. Picard, "Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 12, pp. 2948–2954, 2014.

[25] Understanding Blood Pressure Readings.
https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings

[26] G. Joseph, A. Joseph, G. Titus, R. M. Thomas, and D. Jose, "Photoplethysmogram (ppg) signal analysis and wavelet de-noising," in *2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/ iCMMD)*. IEEE, 2014, pp. 1–5.

[27] J. L. Moraes, M. X. Rocha, G. G. Vasconcelos, J. E. Vasconcelos Filho, V. H. C. De Albuquerque, and A. R. Alexandria, "Advances in photopletysmography signal analysis for biomedical applications," *Sensors*, vol. 18, no. 6, p. 1894, 2018.

[28] S. Laurent, J. Cockcroft, L. Van Bortel, P. Boutouyrie, C. Giannattasio, D. Hayoz, B. Pannier, C. Vlachopoulos, I. Wilkinson, and H. Struijker-Boudier, "Expert consensus document on arterial stiffness: methodological issues and clinical applications," *European heart journal*, vol. 27, no. 21, pp. 2588–2605, 2006.

[29] G. Fortino and V. Giampà, "Ppg-based methods for non invasive and continuous blood pressure measurement: an overview and development issues in body sensor networks," in *2010 IEEE International Workshop on Medical Measurements and Applications*. IEEE, 2010, pp. 10–13.

[30] D. Sahu and M. Bhaskaran, "Palpatory method of measuring diastolic blood pressure," *Journal of anaesthesiology*, clinical pharmacology, vol. 26, no. 4, p. 528, 2010.

[31] L. Pilotto, "Percorso diagnostico dei pazienti ipertesi. misurazione della pressione arteriosa: metodi e strumenti," *Ital Heart J*, vol. 1, no. 5, pp. 32–35, 2000.

[32] T. Sato, M. Nishinaga, A. Kawamoto, T. Ozawa, and H. Takatsuji, "Accuracy of a continuous blood pressure monitor based on arterial tonometry." *Hypertension*, vol. 21, no. 6pt1, pp. 866–874, 1993.

[33] M. Sharma, K. Barbosa, V. Ho, D. Griggs, T. Ghirmai, S. K. Krishnan, T. K. Hsiai, J.-C. Chiao, and H. Cao, "Cuff-less and continuous blood pressure monitoring: a methodological review," *Technologies*, vol. 5, no. 2, p. 21, 2017.

[34] M. Gao, N. B. Olivier, and R. Mukkamala, "Comparison of noninvasive pulse transit time estimates as markers of blood pressure using invasive pulse transit time measurements as a reference," *Physiological reports*, vol. 4, no. 10, p. e12768, 2016.

[35] J. L. Moraes, M. X. Rocha, G. G. Vasconcelos, J. E. Vasconcelos Filho, V. H. C. De Albuquerque, and A. R. Alexandria, "Advances in photopletysmography signal analysis for biomedical applications," *Sensors*, vol. 18, no. 6, p. 1894, 2018.

[36] Paola A. Lanfranchi and Virend K. Somers, "Cardiovascular Physiology: Autonomic Control in Health and in Sleep Disorders," *Cardiovascular Physiology: Autonomic Control in Health and in Sleep Disorders*, Chapter 20.

[37] Shouping Wang, Jia Liu, Pandeng Zhang, Xiaodan Yang, Lu Huang, "Digital Control Method of Volume Clamp Method," *International Conference on Mechatronics*, Electronic, Industrial and Control Engineering (MEIC 2015)

[38] X.F. Teng and Y.T. Zhang. "Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach". In: *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE Cat. No.03CH37439) 4 (2003), pp. 3153– 3156. issn: 1094-687X. doi: 10.1109/IEMBS.2003.1280811.

[39] BP Pro. http://www.bpro.ie

[40] Mohammad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade and Mahdi Shabany, "Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring," *IEEE transactions on biomedical engineering*, VOL. 64, NO. 4, April 2017.

[41] R. Mukkamala et al., "Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8,

pp. 1879–1901, Aug. 2015.

[42] M. Moghadam and M. H. Moradi, “Model based blood pressure estimation during exercise test using modified fuzzy function,” in *Proc. 21st Iranian Conf. Biomed. Eng.*, 2014, pp. 69–73.

[43] Iman Sharifia, Sobhan Goudarzib, Mohammad Bagher Khodabakhshic, “A novel dynamical approach in continuous cuffless blood pressure estimation based on ECG and PPG signals,” *Artificial Intelligence In Medicine*, 97 (2019) 143–151.

[44] Chen W, Kobayashi T, Ichikawa S, Takeuchi Y, Togawa T., "Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration". Med Biol Eng Comput 2000;38(5): 569–74.

[45] Poon C, Zhang Y., Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time. Engineering in Medicine and Biology Society, 2005. IEEEEMBS 2005. *27th Annual International Conference of the 2006*:5877–80.

[46] Geddes L, Voelz M, James S, Reiner D., Pulse arrival time as a method of obtaining systolic and diastolic blood pressure indirectly. Med Biol Eng Computì 1981;19(5):671–2.

[47] George Wypych, Photophysics, in *Handbook of Material Weathering* (Sixth Edition), 2018.

[48] D Nair, S-Y Tan, H-W Gan, S-F Lim, J Tan, M Zhu, H Gao, N-H Chua, W-L Peh and K-H Mak, "The use of ambulatory tonometric radial arterial wave capture to measure ambulatory blood pressure: the validation of a novel wrist-bound device in adults," *Journal of Human Hypertension* (2008) 22, 220–222; doi:10.1038/sj.jhh.1002306; published online 8 November 2007.

[49] R. Wang, W. Jia, Z.-H. Mao, R. J. Sclabassi, and M. Sun, “Cuff-free blood pressure estimation using pulse transit time and heart rate,” in *2014 12th international conference on signal processing (ICSP)*. IEEE, 2014, pp. 115–118.

[50] History of the Sphygmomanometer https://www.adctoday.com/learning-center/about-sphygmomanometers/history-sphygmomanometer.

[51] Mercury based sphygmomanometer. https://www.my-personaltrainer.it/benessere/sfigmomanometro.html

[52] Oscillometric BP devices.
https://pdfserv.maximintegrated.com/en/an/AN4689.pdf

[53] HeartGuide.
https://omronhealthcare.com/products/heartguide-wearable-blood-pressure-monitor-bp8000m/

[54] S. i Carós and J. Maria, “Continuous non-invasive blood pressure estimation,” Ph.D. dissertation, ETH Zurich, 2011.

[55] C. M. Ehrenfeld J., Noninvasive Arterial Pressure Monitoring. Springer, 2014.

[56] Grzegorz Bilo, Cristina Zorzi, Juan E. Ochoa Munera, Camilla Torlasco, Valentina Giuli and Gianfranco Parati, "Validation of the Somnotouch-NIBP noninvasive continuous blood pressure monitor according to the European Society of Hypertension International Protocol revision 2010," *Journal Blood Pressure Monitoring*, 2015, 20:291-294

[57] R. Asmar, “Pression artérielle. régulation et épidémiologie. mesures et valeurs normales,” *Néphrologie thérapeutique*, vol. 3, no. 4, pp. 163–184, 2007.

[58] Guang Zhang, Chengyu Liu, Lizhen Ji, Jing Yang and Changchun Liu, "Effect of a Percutaneous Coronary Intervention Procedure on Heart Rate Variability and Pulse Transit Time Variability: A Comparison Study Based on Fuzzy Measure Entropy," *Entropy* 2016, 18, 246; doi:10.3390/e18070246

[59] Ümit Şentürk, İbrahim Yücedağ, Kemal Polat, "Repetitive neural network (RNN) based blood pressure estimation using PPG and ECG signals,"978-1-5386-4184-2/18/$31.00 ©2018 IEEE

[60] Cong Yan, Zhenqi Li, Wei Zhao, Jing Hu, Dongya Jia, Hongmei Wang and Tianyuan You, "Novel Deep Convolutional Neural Network for Cuff-less Blood Pressure Measurement Using ECG and PPG Signals," 978-1-5386-1311-5/19/$31.00 ©2019 IEEE

[61] Physionet portal.
https: //physionet.org/about/physionet

[62] MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available

at: http://www.nature.com/articles/sdata201635

[63] Xiao-Rong Ding,Yuan-Ting Zhan, "Photoplethysmogram Intensity Ratio: A Potential Indicator for Improving the Accuracy of PTT-Based Cuffless Blood Pressure Estimation," 978-1-4244-9270-1/15/$31.00 ©2015 IEEE

[64] Mohamed Elgendi. "On the Analysis of Fingertip Photoplethysmogram Signals". In: *Current Cardiology Reviews* 8.1 (2012), pp. 14–25. issn: 1573403X. doi: 10.2174/157340312801215782.

[65] M. Y. Rudenko, O. K. Voronova, and V. A. Zernov, "Ecg periodic table: a new ecg classification based on heart cycle phase analysis," *Cardiometry*, no. 2, p. 19, 2013.

[66] Laxmi Shaw, Sangeeta Bhaga, " Online EMG Signal Analysis for diagnosis of Neuromuscular diseases by using PCA and PNN," *International Journal of Engineering Science and Technology* (IJEST), Vol. 4 No.10 October 2012

[67] S. Vandeput, "Heart rate variability: linear and nonlinear analysis with applications in human physiology," 2010.

[68] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Current cardiology reviews*, vol. 8, no. 1, pp. 14–25, 2012.

[69] Tanmoy Bhowmik, Jishnu Dey and Vijay Narayan Tiwari, "A Novel Method for Accurate Estimation of HRV from Smartwatch PPG Signals," 978-1-5090-2809-2/17/$31.00 ©2017 IEEE

[70] Yongbo Liang, Mohamed Elgendi, Zhencheng Chen Rabab Ward, "Analysis: An optimal filter for short photoplethysmogram signals," *Scientific Data*, 5:180076, DOI: 10.1038/sdata.2018.76

[71] L. Mesin, Introduction to biomedical signal processing

[72] Aifeng Ren, Zhenxing Du, Juan Li, Fangming Hu, Xiaodong Yang, and Haider Abbas, "Adaptive Interference Cancellation of ECG Signals," *Sensors* (Basel). 2017 May; 17(5): 942.

[73] R. Lazazzera, Y. Belhaj and and G. Carrault, "A New Wearable Device for Blood Pressure Estimation Using Photoplethysmogram," *Sensors* (Basel). 2019 Jun; 19(11): 2557.

[74] Xiaorong Ding, Bryan P. Yan, Yuan-Ting Zhang, Jing Liu, Ni Zhao Hon Ki Tsang, "Pulse Transit Time Based Continuous Cuffless Blood Pressure Estimation: A New Extension and A Comprehensive Evaluation," *Scientific Reports* 7: 11554 DOI:10.1038/s41598-017-11507-3

[75] D. Martinez, Arthur Edwards-Block, Andrés Felipe Murillo, "Wireless Sensor Networks for Ambient Assisted Living," *Sensors* 2013, 13, 16384-16405; doi:10.3390/s131216384

[76] J. Kameenoff, "Signal processing techniques for removing noise from ecg signals," *Biomedical Engineering and Research*, vol. 1, no. 1, p. 1, 2017.

[77] V. Prabhu, P. G. Kuppusamy, A. Karthikeyan M. Sucharitha, "A novel approach for non-invasive measurement of mean arterial pressure using pulse transit time," *Multimed Tools Appl* 79, 3775–3789 (2020). https://doi.org/10.1007/s11042-018-6971-8

[78] Why Feature Correlation Matters .... A Lot! https://towardsdatascience.com/why-feature-correlation-matters-a-lot-847e8ba439c4

[79] Positive and Negative Correlation. https://fisherdong.wordpress.com/2011/03/25/positive-and-negative-correlation/

[80] Random Forest. Regression:https://towardsdatascience.com/random-forest-and-its-implementation-71824ced454f

[81] Fundamentals of Regression Analysis. https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/

[82] Mean Square Error. https://en.wikipedia.org/wiki/Meansquarederror#Predictor

[83] Inge S. Helland, "On the Interpretation and Use of R^2 in Regression Analysis," *Biometrics*, Vol. 43, No. 1 (Mar., 1987), pp. 61-69

[84] K-Fold Cross Validation: https://medium.com/datadriveninvestor/k-fold-cross-validation-6b8518070833

[85] Association for the Advancement of Medical Instrumentation. https://www.aami.org/Standards/