POLITECNICO DI TORINO

Master Thesis in ICT for Smart Societies

Blood Pressure retrieval in cuff-less devices through regression

Supervisor:
Prof. M.Visintin
Co-Supervisor:
Prof. G.Pagana

Candidate:
Silvia Bennici

Academic Year 2020-2021

Contents

Acronyms

WHO world health organization

CVD cardiovascular disease

BP blood pressure

ABP arterial blood pressure

SBP systolic blood pressure

DBP diastolic blood pressure

MAP mean arterial pressure

mmHg millimeters of mercury

PP pulse pressure

AAMI association for the advancement of medical instrumentation

ESH European society of hypertension

ISO international organization for standardization

ECG electrocardiogram

HR heart rate

PPG photoplethysmogram

PTT pulse transit time

PVW pulse wave velocity

ms milliseconds

ICU intensive care unit

MAE mean absolute error

MSE mean squared error

RMSE root mean squared error

CC cardiac cycle

Chapter 1

Introduction

Hypertension, or elevated blood pressure, is a serious condition since it significantly increases the risk of irreversible severe events and diseases like stroke, blindness, and dementia. Even though hypertension is underestimated by the majority of the population, health professionals say that it is one of the major causes of premature deaths in the world. It is usually so underestimated because it has no symptoms; for this reason it is called the *silent killer*. It is estimated that, in 2015, 1.13 billion people worldwide suffered from hypertension (it was 594 million in 1975), but less than 1 over 5 people had the problem under control[1]. To try to reduce the problem, many health professionals and organizations, among which there is also the World Health Organization (WHO), have launched important awareness campaigns.

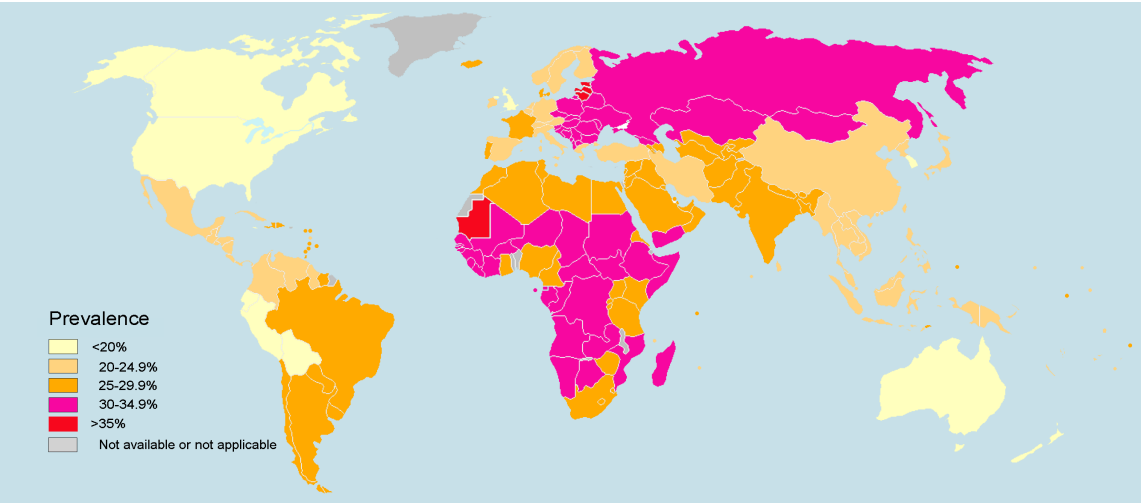


Figure 1.1: Prevalence of hypertension in the world [2]

According to the World Health Organization, hypertension has a higher incidence in males than females (1 over 4 and 1 over 5 respectively over a billion)

and is more frequent in low-income countries than in high-income ones (see Figure 1.1). The same organization reports a prevalence of 27% in Africa, 26% in Eastern Mediterranean, 25% in South East Asia, 23% in Europe, 19% in Western Pacific and 18% in the Americas. The target is to reduce the prevalence of hypertension by 25% (concerning the prevalence shown in 2010) by 2025.

In Europe, the 3.5% of the population with an age comprised between 25 and 34 years reported hypertensive diseases; it was the 53.3% for people aged $\leq$ 75 years. In 2016 there were 1.68 million deaths due to diseases linked to the circulatory system while the hospital discharge rates for patients with diseases of the circulatory system in Italy was of 1789 over 100,000 inhabitants [3].

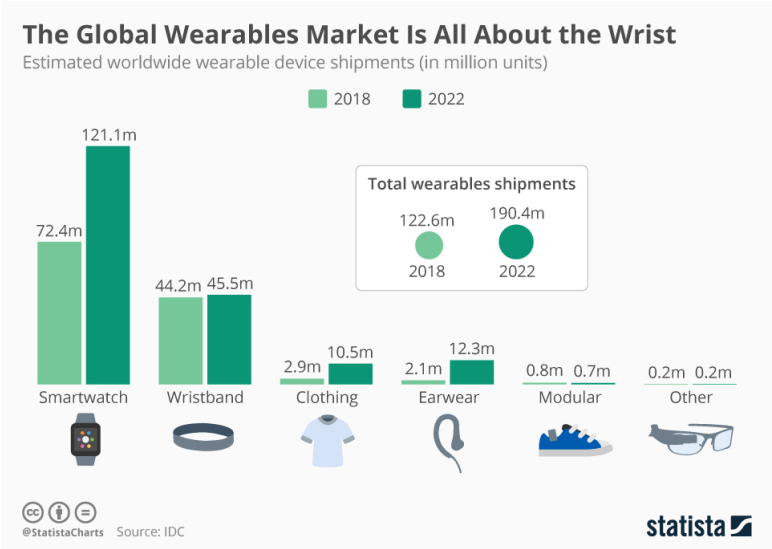


Figure 1.2: Wearable technologies share of the market

The only way to diagnose hypertension is to continuously monitor blood pressure. In the era of advanced technology, modern digital devices can be very helpful in measuring the patient's blood pressure throughout the day (hot just in the office) in a non-invasive and accessible way. A special mention is deserved by wearable technologies(including smartwatches and fitness trackers), which nowadays are flooding the market; these devices allow frequent monitoring of the vital signs with minimal stress on the patient. Moreover, since these kinds of devices collect a high number of different data, they also allow to correlate variations in blood pressure with daily stress and environmental changes [4].

It is expected that these devices will dramatically change our way of monitoring our heart's activity and prevent cardiovascular diseases. The interest on this share of the market is led both by researchers and by the interest of the users: the global smartwatch market was valued at 20.64 billion dollars in 2019 and is projected to reach 96.31 billion by 2027 [5](the statistics for years 2018 – 2022 are shown in

Figure 1.2).

Despite the great advancements of the latest years, the research in this field still has to go on, the major goal being to improve the accuracy of the vital signs estimation.

The present work is inserted in such framework, by trying to improve the estimation of the blood pressure from the electocardiogram (ECG) and the photoplethysmogram (PPG), which can be easily recorded non-invasively with cuff-less devices thus are suitable for continuous, affordable, non-invasive monitoring. Data are collected from an on-line public database (MIMIC3) and aggregated by patient. The algorithm is studied both by using the complete recordings of a single patient at once and simulating a real-time acquisition of the samples. Error sources and critical parameters are analyzed.

Chapter 2

Background

2.1 Physiological background

2.1.1 The cardiovascular system

The cardiovascular system is an organ system composed of the heart, blood, arteries, and veins and allows the circulation of blood through the vessels, allowing the transportation of blood cells, nutrients, hormones, and oxygen to each cell of the human body [6]. Of course, this circulation is led by the heart, which pumps the blood into the vessels.

The circulation can be split into two main parts (Figure2.1):

- *Pulmonary circulation* where blood rich in carbon dioxide flows from the right part of the heart through the lungs to be oxygenated and goes back to the left part of the heart.
- *Systemic circulation* where blood flows from the left part of the heart through the rest of the body to oxygenate it and retrieve carbon dioxide and goes back to the right part of the heart.

2.1.2 The heart

The heart is a muscle and it is considered one of the most important organs. Its size is about the one of a fist, its function is to pump blood into the vessels and it is located in the chest between the lungs, under the sternum [7]. The heart anatomy is shown in Figure 2.2; the muscle has four chambers: the two upper chambers are called left and right atria (LA,RA) and the two lower chambers are called left and right ventricles(LV,RV); left and right atria and ventricles are separated by a wall called septum. Notice that the atria are smaller than the ventricles since they receive blood (which in the final stage of the circulation has a lower pressure), while the ventricles are larger and they are stronger pumps; in particular, the left

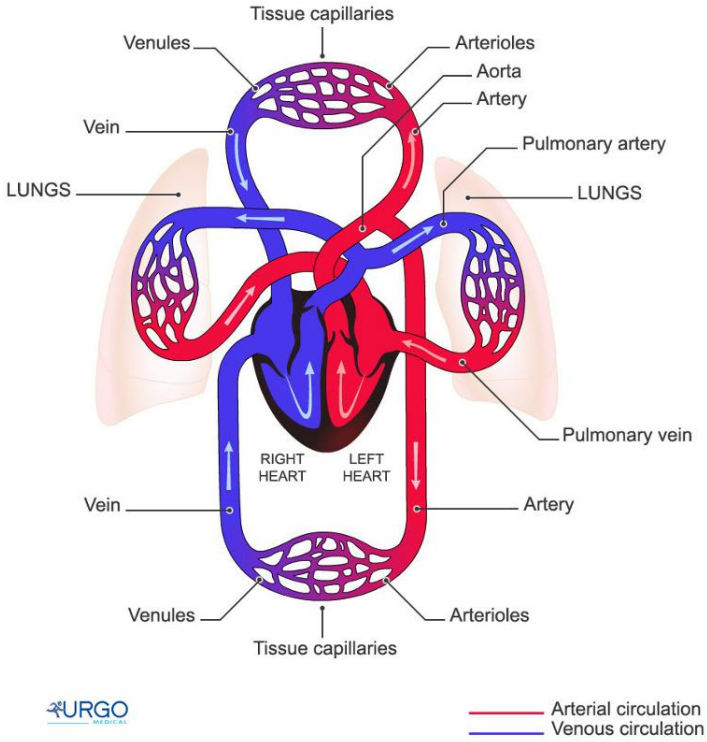


Figure 2.1: Pulmonary and systemic circulation

ventricle is the strongest one, since it has to pump blood to the whole body and for this reason its walls are thicker [8]. In normal condition, the four chambers work coordinately to keep oxygen-rich blood flowing in the body [10]. The flow of blood is regulated by four valves [11]:

- The *tricuspid* is placed between the right atrium and ventricle.
- The *pulmonary valve* is placed between the right ventricle and the pulmonary arteries, to start the pulmonary circulation.
- The *mitral valve* is placed between the left atrium and ventricle.
- The *aortic valve* is placed between the left ventricle and the aorta, to start the systemic circulation.

The blood flow inside the heart can be summarized in four steps:

1. The right atrium receives blood which is poor of oxygen and rich in carbon dioxide (which comes from the systemic circulation) and pumps it to the right ventricle through the tricuspid.
2. The right ventricle contracts to eject blood to the lungs (starting the pulmonary circulation) through the pulmonary valve.

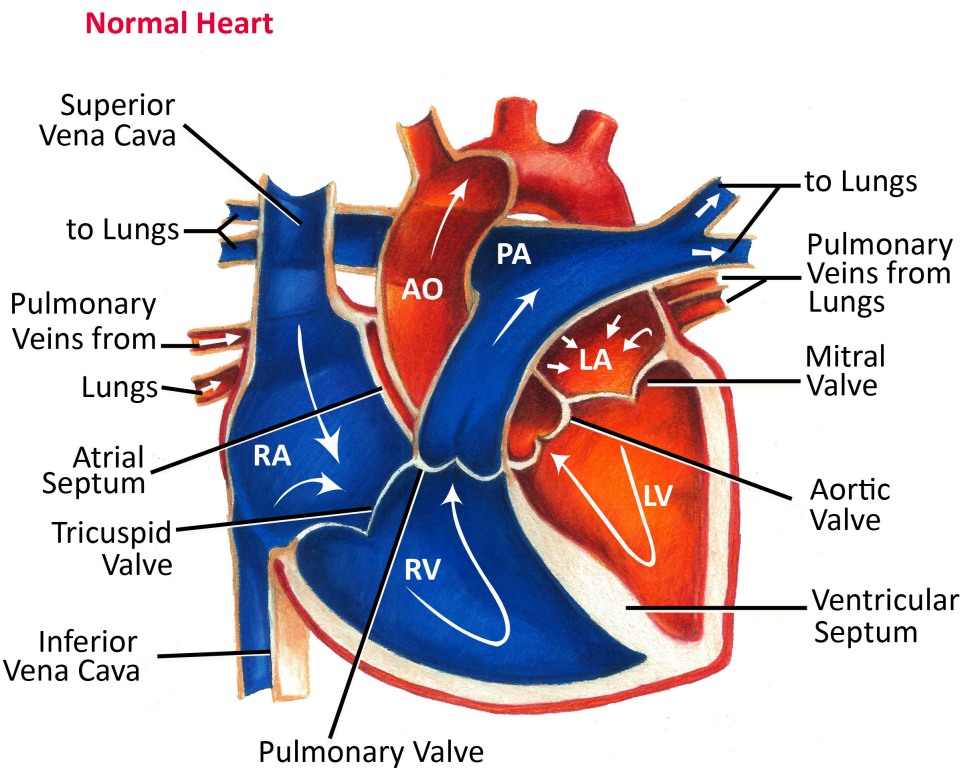
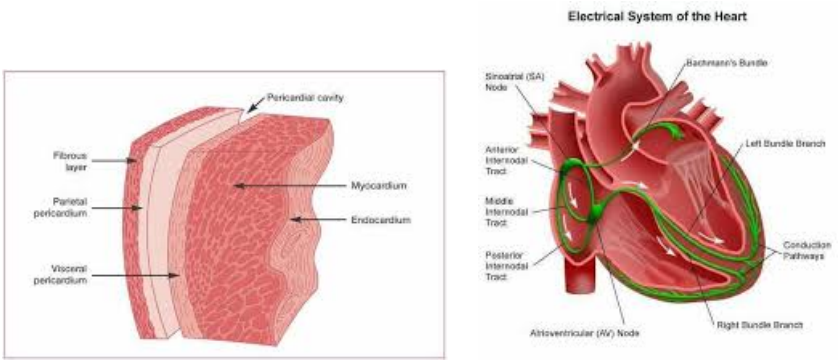


Figure 2.2: Heart anatomy [12]

3. The left atrium receives blood rich in oxygen and poor in carbon dioxide coming from the lungs (after the pulmonary circulation) and pumps it to the left ventricle through the mitral valve.

4. The left ventricle contracts to eject blood to the aorta (starting the systemic circulation) through the aortic valve.

The heart walls are made up of three layers (Figure 2.3a) which, from the inside to the outside, are: the endocardium (an epithelial layer), the myocardium (the actual muscle layer) and the epicardium; the heart is then surrounded by the pericardium, a membrane with a sac-shape. In the myocardium there are two kinds of cells: muscle cells (which can contract easily) and pacemaker cells (for a small 1%) [13]. In fact, even if it is classified as an involuntary muscle, the heart has the ability to generate by itself the electric signal need for the contraction. The signal for the contraction is generated by the Sinoatrial Node (SA, Figure 2.3b), located in the myocardial layer at the junction of the superior caval vein and the right atrium. This way the heart's electrical system can control the timing of the pump (the Heart Beat); the heart receives also nerve signals from the vagus nerve; these signals can influence (but not control) the heart rate. The conduction system makes first the



(a) Heart walls (b) Conduction system

Figure 2.3: Heart walls and conduction system [14]

atria contract simultaneously, and then the ventricles (again simultaneously) [9] Cardiac contraction is called systole, cardiac relaxation is called diastole [8].

A heartbeat consists of four phases [9]:

- Ventricular filling: blood enters the ventricles passing through the atrioventricular valves, making the ventricular volume rise. This happens during the diastolic phase; after the filling, the systolic phase begins.
- Iso-volumetric contraction: all the valves are closed and pressure inside the heart rises while the contraction begins. When the pressure in the heart is higher than the one in the aorta, the aortic and the pulmonary valves open.
- Ejection: blood is ejected from the aortic and the pulmonary valves. Ventricular pressure continues to rise, while volume decreases.
- Iso-volumetric release: after the closure of the aortic valve, the pressure inside the chambers decreases.

A cardiac cycle (the heart's performance between an heart beat and the following one) lasts about $0.6 - 1$ second.

2.1.3 Blood Pressure - BP

Together with respiratory rate, heart rate, oxygen saturation, and body temperature, Blood Pressure (BP) is a vital sign [16] (those signs that healthcare professionals measure in order to evaluate the patient's health) and it reflects the status of the cardiovascular system. It is due to the heart pumping blood into the arteries and has a cyclical path (as shown in Figure 2.4): it rises when the heart contracts to eject blood into the aorta and it falls when the heart relaxes to refill. Blood pressure is influenced by cardiac output, systemic vascular resistance, and arterial

stiffness (in particular it rises when at least one of these parameters rises) and varies depending on emotional state, physical activity and health state [15] [16].

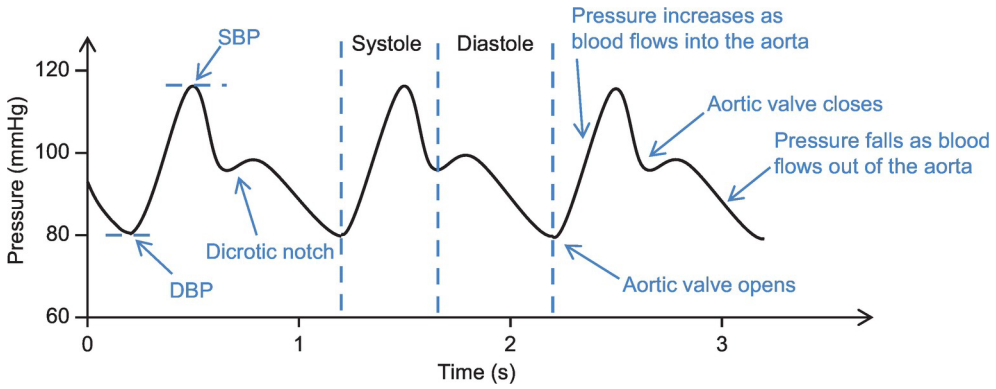


Figure 2.4: Arterial Blood Pressure [17]

Arterial blood pressure is to be distinguished from Venous Blood Pressure. The latter, in fact, is lower than the first the reason being that in these vessels the blood does not receive the heart's thrust and it is measured only in particular cases. That is why the veins' structure is different from the arteries' one: blood circulation is helped by the presence of one-way valves, that prevent the blood from flowing in the inverse direction. The structure of both arteries and veins is shown in Figure 2.5.

Blood pressure (BP) is associated with Arterial Blood Pressure (ABP) unless otherwise specified. Arteries are specifically made to sustain the pulse pressure: their walls are thicker and made of elastic tissue that allows them to expand and contract following each ventricular contraction. These properties of the arteries can deteriorate due to aging or pathological reasons thus worsening the circulatory performance [9].

Blood pressure is traditionally monitored non-invasively by auscultation (such method is still considered the gold standard of accuracy for non-invasive methods [16], the method is shown in Figure 2.6) and measured in mmHg (millimeters of mercury). Figure 2.4 shows a possible output of continuous monitoring of the BP: it has a cyclical path with a first part (Systole) corresponding to the pressure inside the artery when the heart contracts and a second part (Diastole) corresponding to the pressure inside the artery when the heart relaxes [18], separated by the Dicrotic notch. In particular, a cycle is characterized by two points:

- Diastolic Blood Pressure (DBP): also called *minimum pressure* is the pressure inside the arteries during the relaxation phase of the heart. In Figure 2.4 it is the minimum point of the diastolic phase in the ABP and in a healthy adult is about 80 mmHg.

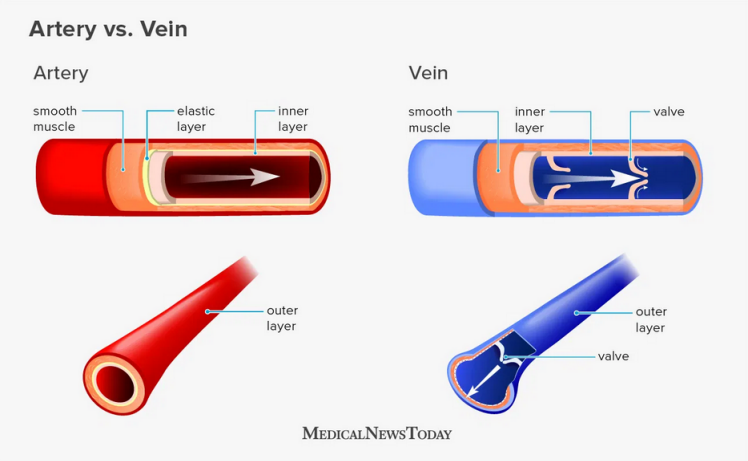


Figure 2.5: Artery and vein structure

- Systolic Blood Pressure (SBP): also called *maximum pressure*, is generated by the flow in the arteries of blood rich in oxygen during the heart's contraction (or systole). As shown in Figure 2.4 SBP is the maximum point of the systolic phase in the ABP and in a healthy adult is about 120 mmHg. If the cardiac output increases, the SBP increases as well (on the contrary, the DBP doesn't show significant variations).

Moreover, there are other parameters which could be considered, such as :

- Pulse Pressure (PP): is the difference between the systolic and diastolic pressure. Normal values are in the range of $40 - 60$ mmHg and tend to increase as the patient ages [19].

- Mean Arterial Pressure (MAP): is calculated as $(SBP + 2DBP)/3$ is the average blood pressure in a subject during a cardiac cycle [20].

Pressure parameters are usually indicated as SBP/DBP (120/80) mmHg. Notice that BP is the result both of the cardiac output and of the peripheral vascular resistance (which is influenced by the vessels' condition) so, for a given cardiac output, if the peripheral resistance increases this will lead to an increased BP (both systolic and diastolic) and conversely a lower peripheral resistance would lead to lower BP [21][22].

Blood Pressure is subject to physiological variations, up to 20 mmHg within a day [23]. For example, it follows a daily pattern: it is lower while the patient is asleep and starts rising a couple of hours before he/she awakens to reach the peak in the afternoon and start decreasing again. It could change also due to physical activity and daily stress (which can make BP rise) or outside temperature (cold temperatures narrow the blood vessels and hot temperatures enlarge them, causing BP to rise and fall respectively) [23].

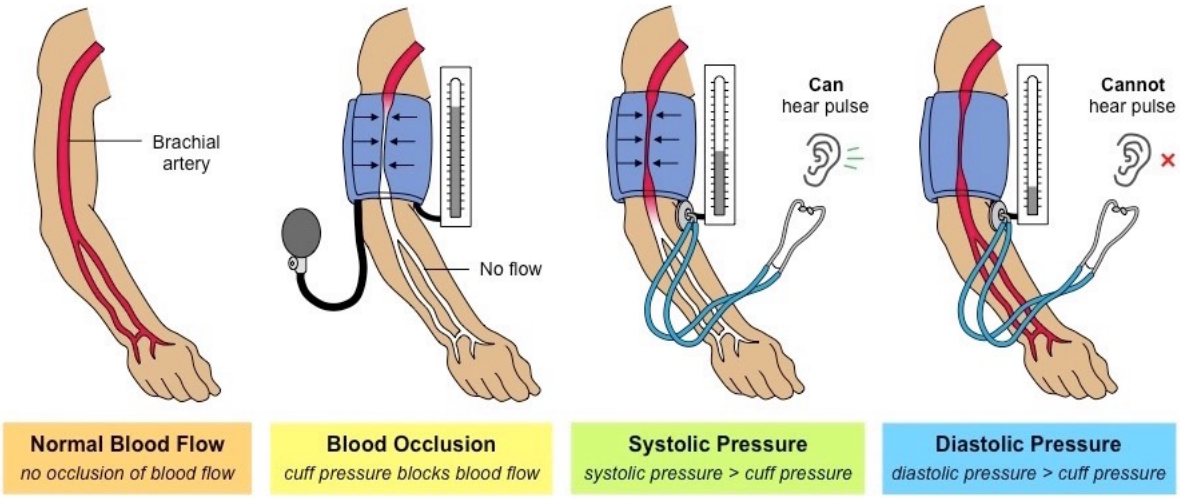


Figure 2.6: BP monitoring: mercury sphygmomanometer [24]

In order to keep the blood pressure under control, clinicians suggest to: do physical exercise, eat low-fat and low-sodium food, minimize alcohol, smoke, and daily stress [23].

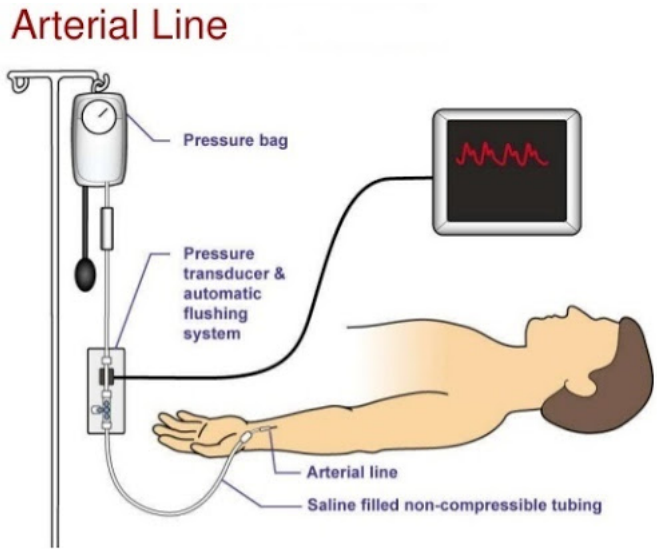


Figure 2.7: BP monitoring: arterial catheter [25]

As already said, the mercury sphygmomanometer (Figure 2.6) is the gold standard for blood pressure monitoring. Nowadays, though, mercury devices have been substituted also in hospitals by other instruments. The preferred location of the measurement is the brachial artery since BP varies by changing the location of the measurement [26]. The main methods known to measure one's BP are:

- The auscultatory method: is a non-invasive method performed when using a sphygmomanometer since the clinician has to listen for the Korotkoff sounds (Figure 2.6) [26].

- The oscillometric method: is a non-invasive method used in automatic devices, which have a sphygmomanometer cuff and observe cuff pressure oscillations through an electronic pressure sensor (Figure 2.8) [27].

- Intra-arterial BP monitoring: is an invasive technique used in the Intensive Care Unit. In this case, BP is measured directly by inserting a catheter in an artery (Figure 2.7). This method allows continuous and very accurate monitoring of the BP but the disadvantage is the risk to contract infections, the formation of hematoma, blood loss [28].

Sources of error in the measurement of BP in the upper arm are:

- Patient's posture: BP is usually measured while the patient is seated or supine. Diastolic BP is shown to change according to the posture and age (in particular in young subjects it is 10mmHg higher while supine; the discrepancy narrows for older patients) [26].

- Arm position: for example if the arm is moved from a horizontal to a vertical position, the BP increases by about $5 - 6$ mmHg.

- Patient's activity: if the patient has smoked, done physical exercise of ingested pressor substances before the measurement, or is anxious during the measurement or talks, the BP could be underestimated or overestimated respectively.

In particular, the condition in which the patient's BP is elevated during the measurement performed by a clinician due to anxiety or nervousness is called *white-coat* effect and in such event, the measures are totally unreliable. It is shown that Ambulatory Blood Pressure Monitoring (ABPM), that is measuring BP at regular intervals during the day by means of a portable device, is able to eliminate the white-coat effect and also detect masked hypertension (a condition in which the patient shows a normal BP when examined by a specialist, but abnormal BP values at home) [29].

In order to state if a device is suited for BP monitoring, it is necessary to validate it. The US Association for the Advancement of Medical Instrumentation (AAMI), the British Hypertension Society, the European Society of Hypertension (ESH) Working Group on Blood Pressure Monitoring, and the International Organization for Standardization (ISO) have collaborated to find a common way of validating BP measuring devices. The outcome was that the measuring error is considered tolerable if it is ≤ 10 mmHg and a device is accepted if the estimated probability

(a) Arm device (b) Wrist device

Figure 2.8: BP monitoring: home devices [34]

that the error of measurement is equal or lower than 10 mmHg is of at least 85% [30].

Studies say that automated and digital instruments have more or less the same accuracy; in particular wrist monitors (Figure 2.8) are particularly unreliable with respect to the gold standard mercury column manometer, while the arm monitors have higher accuracy than wrist ones [31] [32] [33]. This higher error is due to both a wrong positioning of the arm (in wrist devices the wrist has to be placed and heart level) and to the accuracy of the device itself, which is sensitive to cardiac arrhythmia and changes to the arteries in the wrist (so, for example, it is not recommended for the elders) [34]. Nevertheless, wrist BP monitors are useful in case the patient cannot wear the arm cuff due to particular conditions (like obesity or breast cancer surgery).

Figure 2.9: Omron Heartguide [23]: validated BP measuring smartwatch

In 2014 clinicians found an error in automated monitors pressure readings of

about 20 mmHg [35], while more recent articles talked about an error of 5 mmHg compared to the mercury sphygmomanometer [36], and of 0 ± 7 mmHg and -1 ± 6 mmHg [37]. A study published in 2020 in the British Journal of general practice [38], after validating many of the devices present in the market, concluded that clinically validated home devices for BP monitoring with an age lower than 4 years have an accuracy comparable to that of the professional ones (even if cuff failure is more frequent). For the sake of completeness, the measuring error declared by some BP monitor manufacturer was investigated. For the devices taken under exam (all validated devices) manufacturers declare a measuring error of 0 ± 3 mmHg independently of the price and type (upper arm or wrist) [23] [39] [40]. For some devices the measuring error was not found [34] [40]; in particular, it would have been interesting to know the accuracy of Omron Heartguide (Figure 2.9), which is the only validated smartwatch that is also able to measure BP by using a wrist cuff [23].

2.1.4 Hypertension

Hypertension is a condition in which the subject's blood pressure is constantly high. In this condition the heart and blood vessels are overworked; this usually does not lead to immediate symptoms, but over the long period could damage the cardiovascular system and lead to serious conditions [41].

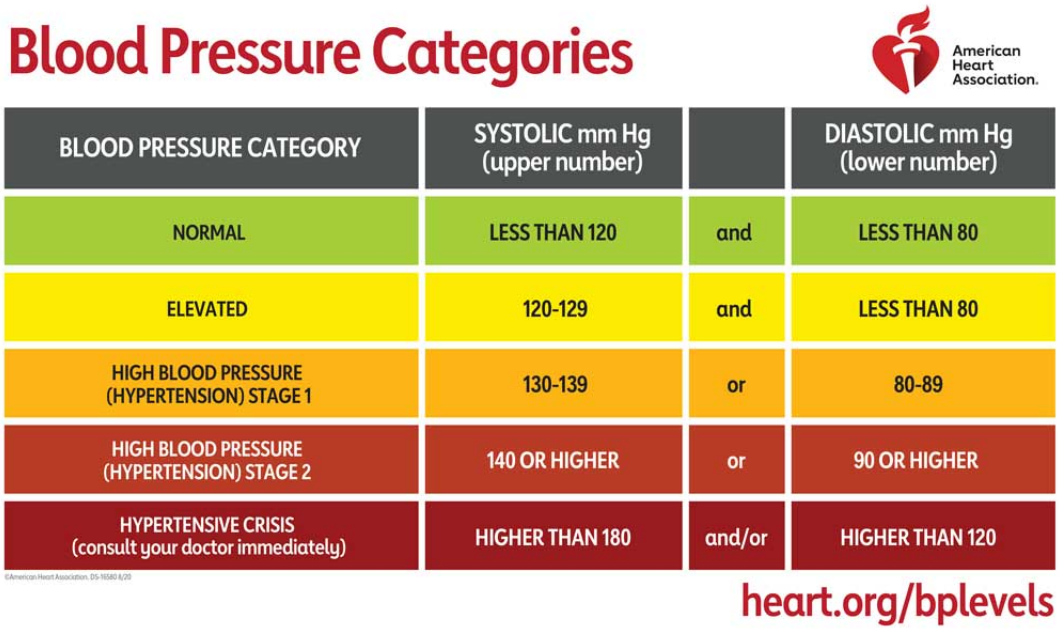


Figure 2.10: Categories of BP

As stated, hypertension is not associated with short-term symptoms: patients

report headaches, vertigo, tinnitus, or fainting episodes; these symptoms, however, could be associated also with anxiety and stress. In the long term instead, if not treated, hypertension could cause coronary artery disease, stroke, heart failure, atrial fibrillation, peripheral arterial disease, vision loss, chronic kidney disease, and dementia.

In order to diagnose hypertension, continuous monitoring of blood pressure is necessary , so for example, the American Heart Association recommends ABPM to confirm the diagnosis. Based on its values, blood pressure is classified in various categories as shown in Figure 2.10. Hypertension has two stages.

Stage 1 is the most common (it covers about $90-95\%$ of the cases) and is due to genetic and lifestyle factors (not to a specific disease) such as smoke, alcohol abuse, abuse of salt and caffeine in the diet, and overweight but also aging is a significant factor.

Stage 2 covers the $5-10\%$ of the cases and it can be related to a specific disorder such as chronic kidney disease (which is the most common cause), narrowing of kidney arteries, an endocrine disorder.

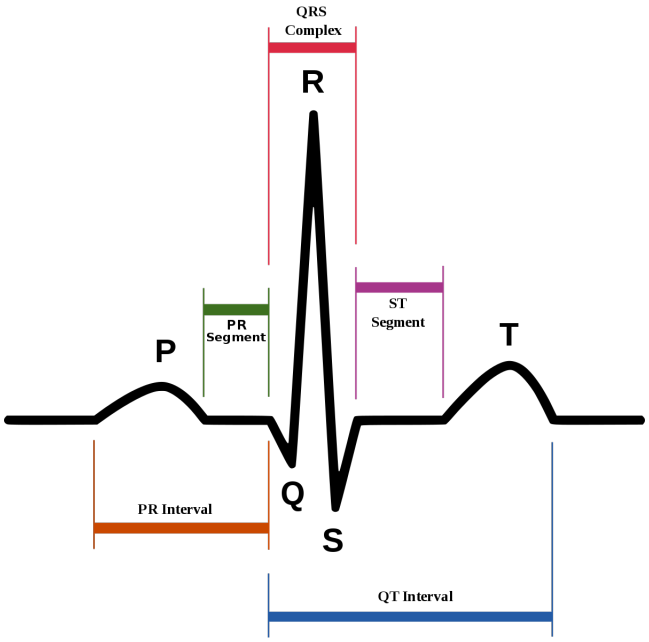
In case the blood pressure is severely high (SBP $\geq$ 180 mmHg and/or DBP $\geq$ 120 mmHg) clinicians talk about *hypertensive crisis*; in this case, some organs could be damaged (such as the brain, kidney, heart, and lungs) and it is of vital importance to act immediately to lower it. The symptoms, in this case, could be confusion, drowsiness, chest pain, and breathlessness.

In most people with Stage 1 hypertension, the cardiac output remains normal, while the total peripheral resistance is higher. Moreover, high blood pressure is also associated with decreased venous compliance and vasoconstriction of arterioles (arteries of smaller diameter) and high pulse pressure in older people.

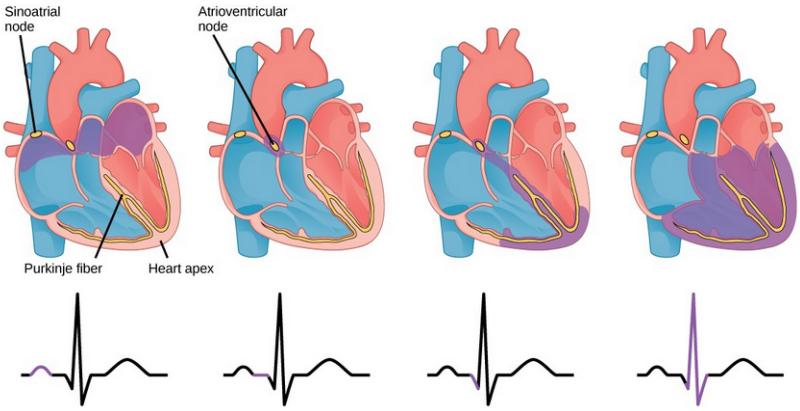
2.1.5 Electrocardiogram - ECG

The electrical impulses that allow the heart beating, generate a current that can be measured by putting some electrodes on the skin. The graph is drawn on graph paper with time on the abscissa (a second every 25mm) and amplitude on the ordinate (a milliVolt every 10 mm) [42]. The graph that reveals the electric activity of the heart (that is the summation of the electric activity of each cell in the myocardium) is the electrocardiogram (ECG). The ECG has a cyclical shape and it can be divided into four sections (as shown in Figure 2.11a):

As we have learned, the electric impulse starts from the sinoatrial node, it then spreads to the atria and makes them contract (this corresponds to the P wave). The signal is delayed by about 0.1 seconds by the atrioventricular node and this corresponds to the PR segment (this delay allows the atria to relax before the ventricular contraction); it is then brought to the heart apex (Q point) before the contraction (QRS complex) [8]. The ST segment represents the restoration of the basic electric conditions, while the T wave is due to the re-polarization of the



(a) ECG wave [42]



(b) ECG phases [8]

Figure 2.11: ECG

ventricles.

The ECG can be measured non-invasively at the skin level by placing some electrodes on the chest and limbs. The ECG can appear in different shapes depending both on the heart's activity and the position of the electrodes and their number; this is the reason why there are standard configurations with respect to which it is possible to take the signal [9]. The most known configuration is the *Einthoven's triangle* where a equilateral triangle is formed by the two wrists and the left ankle or by the two shoulders and the pubis, as shown in Figure 2.12. The electrodes are

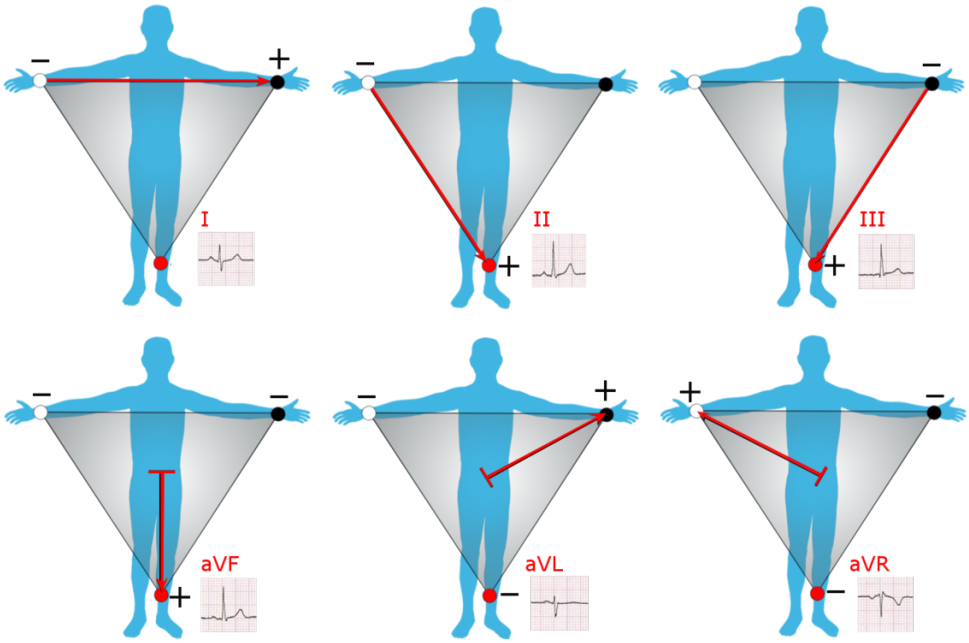


Figure 2.12: ECG electrodes placing [43]

placed at the vertices of the triangle and are connected to a device that measures the voltage. Literature reports examples of incorrect treatments due to inaccurate ECG monitoring caused by human errors (only 50% of the nurses and 20% of cardiologist under study were able to correctly place the electrodes) [44]; another study reports that the specificity of each patient is a huge source of error in reading the ECG [45]. Many portable ECG devices are available on the market (one of those is shown in Figure 2.13); some of them can also connect to the patient's smartphone to display the graph; manufacturers claim that the accuracy of the devices is the same as the one in the hospital.

Figure 2.13: Portable ECG device [34]

The Heart Rate is the number of heartbeats per second (bpm) and it is referred

to the ventricular beat. Normal values are between $60-100$ bpm: if they are lower, it is a case of bradycardia, if they are higher it is a case of tachycardia [46]. To measure the heart rate it is sufficient to see the time distance between two R peaks in seconds:

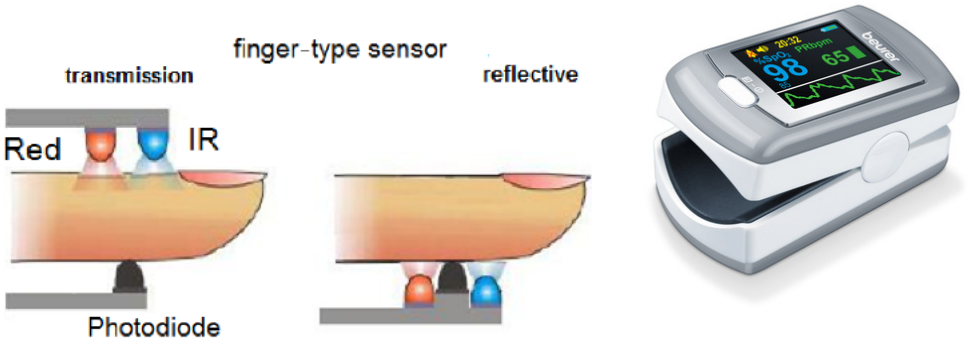
$$HR = \frac{60}{R_{t+1} - R_t} \tag{2.1}$$

where the denominator is in seconds.

2.1.6 Photoplethysmogram - PPG

A photoplethysmogram reveals the changes in blood volume in the microvascular bed of tissue. In human beings, the microvascular blood flow reflects, among others, the heart and pulmonary activity [47] [48]. It is a simple, low-cost, and non-invasive optical method that deploys the properties of the light reflected or transmitted by the surface of the skin.

A typical PPG sensor contains a led, which emits low-intensity infrared light, and a photodetector, that measures the light reflected by the skin; in fact, the light output by the led is absorbed by the tissues but, since the absorption by blood is stronger than the others, the changes in blood volume can be detected (the light reflected is proportional to blood volume variations). Notice that the infra-red diode is used to measure the blood flow while to measure the absorption of oxygen in hemoglobin it is better to use a green led, that penetrates more deeply the tissues [49]. A scheme of the PPG sensor and a device are shown in Figure 2.14.



(a) PPG sensor scheme [51] (b) PPG device [34]

Figure 2.14: PPG devices

Since the PPG is very simple to obtain, it is widely used in wearable devices (like smart-bands) and smartphones for the continuous and remote monitor the heart's activity [49] [50]. Because of its features it can be used for primary health care and remote monitoring [53].

The amplitude of the signal (shown in Figure 2.15) is directly proportional to the pulse pressure (the higher the peak, the stronger the pulse). PPG has two main phases: the *anacrotic* phase (the rising edge) is due to the heart's systole and it reflects mainly the state of the heart; the *catacrotic* phase (the falling edge) is caused by the diastole and it mainly reflects the elasticity of the circulatory system. In healthy subjects, the catacrotic phase also shows a dicrotic notch[48].

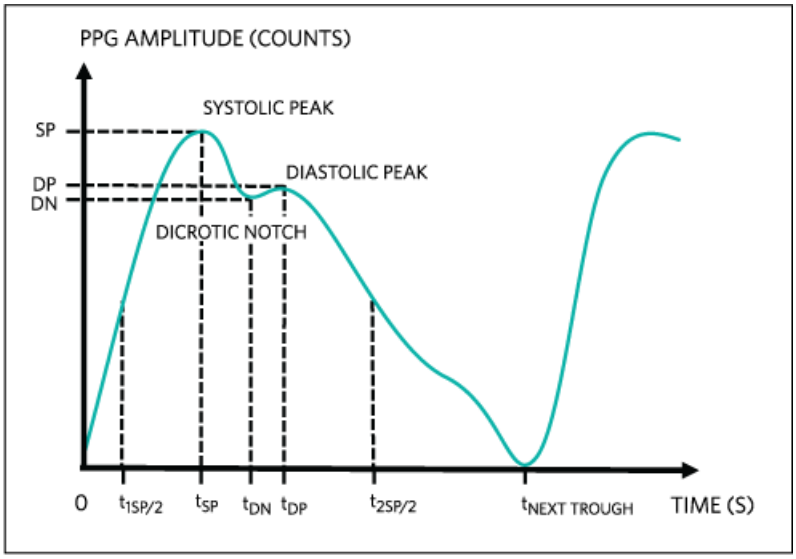


Figure 2.15: PPG waveform [52]

PPG is a very simple a low-cost method to monitor the circulatory system's condition, but the relation between the two (PPG and blood flow) is only qualitative and not quantitative [48].

Some of the factors that can influence the PPG are the skin color (due to the light properties), nail polish, external illumination, anemia, and the subject's movements; moreover, it is affected by heartbeat, hemodynamics, and properties of the vessels.

The quality of the signal is also influenced by the location of the sensor [49]; the most popular locations are:

- Wristband: it looks to be the most popular especially in wearable devices since it is very comfortable to wear for users. However, it has its limitations and several options have been proposed in order to improve its accuracy.

- Forehead: these kinds of sensors take advantage of the high reflectance of the skin in the forehead, caused by the high density of blood vessels. Another advantage of this position is that motion artifacts are mitigated, especially during physical activity.

- Earlobe: lobes contain a high quantity of blood vessels and are less affected by motion artifacts with respect to wrists and fingers.

.

Elgendi (2012) [53] makes an overview of the main artifacts which affect the signal; these are:

- Powerline interference: it is due to the instrumentation which probably picks ambient noise and other artifacts. It has a high frequency (higher than 50 Hz).

- Motion artifact: may be caused by bad contact between the skin and the sensor and it has a low frequency. It is associated with the patients' movements.

- Low amplitude signal: this is usually associated with the gain controller of the sensor, but may also be caused by low blood pressure or a constriction in the vessels supplying the skin.

- Premature ventricular contraction (PVC): this kind of arrhythmia interrupts the normal heartbeat rhythm.

The main characteristics of the PPG wave are [53]:

- Systolic amplitude: is the height of the peak; it is directly related to blood volume.

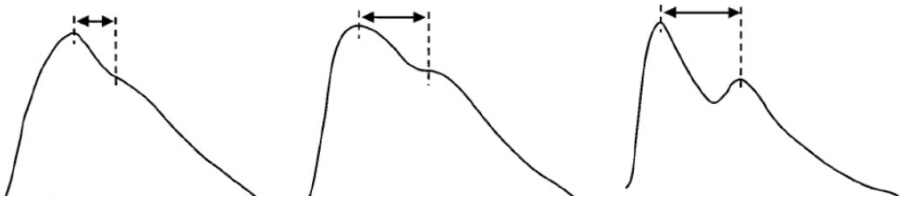
- Pulse width: it seems to be related to the systemic vascular resistance.

- Pulse area: is the area under the PPG curve. The ratio between the area before the dicrotic notch and the area after the dicrotic notch is related to the total peripheral resistance.

- Peak to peak interval: is the interval between two systolic peaks. It can be used to detect the heart rate, as the R-R interval of the ECG.

- Pulse interval: is the interval between the beginning and the end of the waveform; it could be used as an alternative method to measure Heart Rate Variability (HRV).

- Large Artery Stiffness Index: the time interval ΔT between the systolic and the diastolic peak is shown to be directly proportional to the subject height h. The ratio $SI = h/\Delta T$ is related to the large artery stiffness. Parameter ΔT is inversely proportional to the patient's age, and coherently SI is directly proportional his age, as shown in Figure 2.16.



(a) 60 years (b) 45 years (c) 29 years

Figure 2.16: PPG waveform for patients of different age [53]

Figure 2.17: Apple watch [57]

From the first and second derivative of the PPG, it is possible to evaluate other indexes that are shown to be related to other characteristics of the patients' health condition [53] [54] and that are very useful to clinicians in the diagnosis of cardiovascular diseases [49].

Bolanos et al. [55] claim that the PPG can be used effectively to replace the ECG for continuous monitoring of HRV, considering the advantages in flexibility, portability, and convenience. This statement looks to be in contrast with the results of other authors; Pietila et al. [56] tested the accuracy of two PPG sensors (PulseOn, PO, and Empatica E4, E4) in the evaluation of the heartbeat with respect to that on an ECG during different activities (sitting, standing, housework and cycling). The absolute error was lower than 10 bpm (w.r.t. the ECG) in the 90% of the cases for both sensors and on average, during normal activities, it rose of 2.5 bpm in PO and 3.7 bpm in E4 w.r.t. sitting. The RMSE in the measure of HRV was 3.5 ± 3.9 ms for PO and 10.2 ± 6.7 ms for E4 during sitting, 18.0 ± 10.9 ms for PO and 48.7 ± 21.8 ms for E4 during cycling. So the accuracy is quite good at least in the measure of HR when the quantity of movement is limited and varies

significantly by changing activity. This limitation due to motion artifact is stated also by Castaneda [49], even if the great potential of PPG is not questioned. Bent et al.[58] made a systematic study on the accuracy of PPG in measuring the heart rate (the reference was again the ECG) and found no statistical inaccuracy due to skin color, but significant differences in when varying the device and activity type; in particular, the sensor seems to be 30% more accurate when the subject is at rest. In particular, their tests showed that, at rest, the most popular consumer-grade devices had an error ($MAE \pm std$) of 7.2 ± 5.4 bpm, while the research-grade device had an accuracy of 13.9 ± 7.8 bpm, while during physical activity the accuracy was of 10.2 ± 7.5 bpm and 15.9 ± 8.1 bpm (this difference could be due to some biases intentionally inserted in consumer devices which are not present in research devices). The study also reports that in the presence of a rhythmic movement (like walking or running and interestingly also typing) the errors rise significantly (the only device which keeps the MAE low in every situation seems to be the Apple Watch, Figure 2.17).

.

2.1.7 Pulse Transit Time - PTT

The Pulse Transit Time is the time that the blood pressure pulse takes to travel from a site to another into the arteries. Typically, the two sites are detected by an ECG R-wave and any critical point of the PPG, as shown in Figure 2.18; this critical point is usually the systolic peak. PTT can be easily measured knowing the time delay between an R-peak and the corresponding Systolic peak in the PGG:

$$PTT = \frac{60}{SP_t - R_t} \tag{2.2}$$

where the denominator is in seconds, SP_t is the time in milliseconds when the systolic point (SP) of the PPG in the cardiac cycle t is recorded, and R_t is the time in milliseconds when the peak of the R-wave (R) of the PPG in the cardiac cycle t is detected.

Thanks to the PTT it is possible to investigate the changes in blood pressure and, since it can be measured non-invasively, it is suited for continuous monitoring of the BP [59].

The relation between PTT and BP is that, being blood a fluid, it propagates in the arteries with a certain velocity, called Pulse Wave Velocity (PWV), which is dependent also on the elasticity of the vessels (E), arterial thickness (t), arterial diameter (d) and blood density (ρ); notice that these characteristics are not only patient-dependent but also vary during his life-time. Given the assumption that the mentioned parameters don't vary significantly over a short period, the PWV is

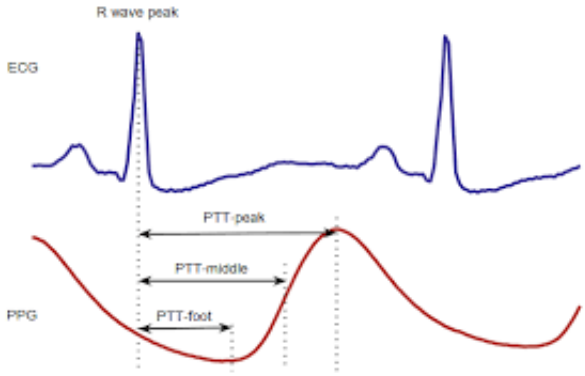


Figure 2.18: PTT [60]

calculated as follows (Moens–Korteweg equation):

$$PVW = \sqrt{\frac{gtE}{\rho d}} \tag{2.3}$$

where $g = 9.81 m/s^2$ and can be omitted (since it is constant) and, according to Hughes equation

$$E = E_0 e^{\gamma P} \tag{2.4}$$

being E_0 is the elastic modulus at zero pressure, γ is a coefficient and P is the blood pressure in mmHg. PVW and PTT are inversely proportional with coefficient K representing the distance traveled by blood between the two sites [59]:

$$PWV = \frac{K}{PTT} \tag{2.5}$$

The time occurrence of the peaks of the ABP, ECG, and PPG is shown in Figure 2.19: in a cardiac cycle, the first peak to be recorded is the R-wave, after that comes the SBP and last the systolic peak of the PPG.

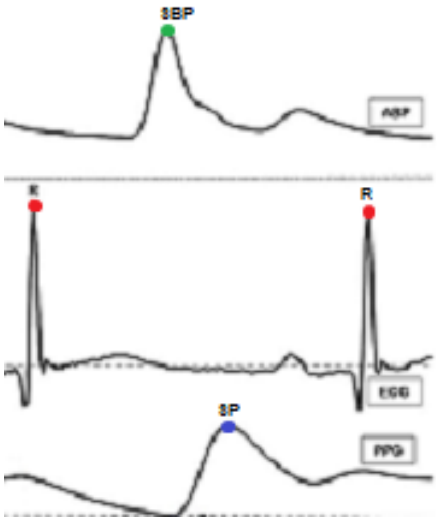


Figure 2.19: Time occurrence of R-wave peak, SBP and Systolic point

2.1.8 MIMIC database

MIMIC is a database included in PhysioNet, which is managed by the members of MIT's Lab for Computational Physiology [61]. PhysioNet is a website founded in 1999 to provide a public service to research institutions and at the same time stimulate new investigation in the study of biomedical signals.

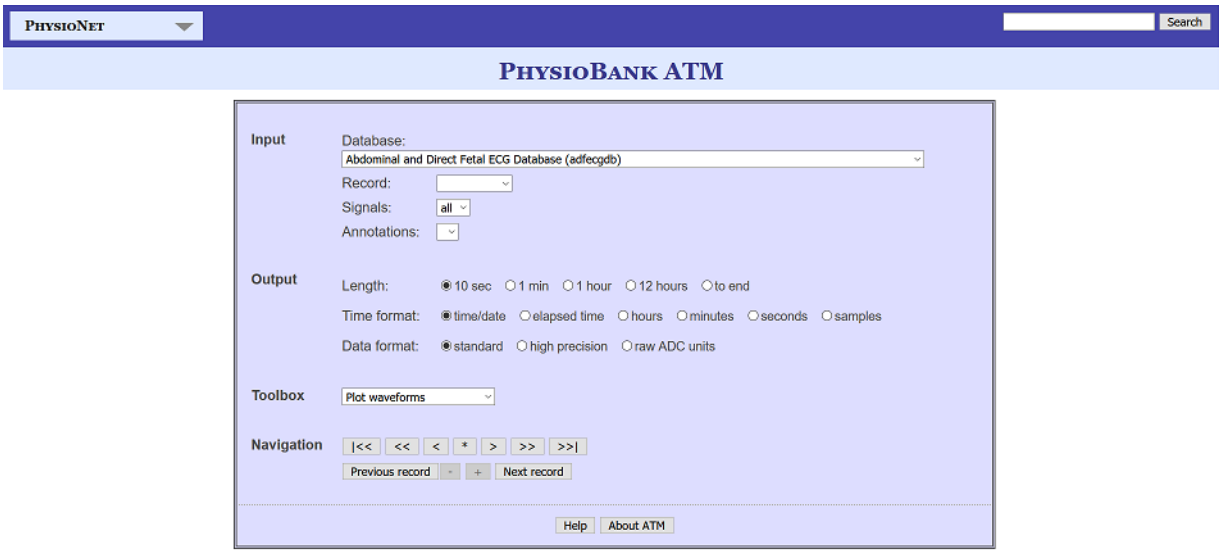


Figure 2.20: Physionet database interface

One of its components is PhysioBank, an archive containing digital recordings

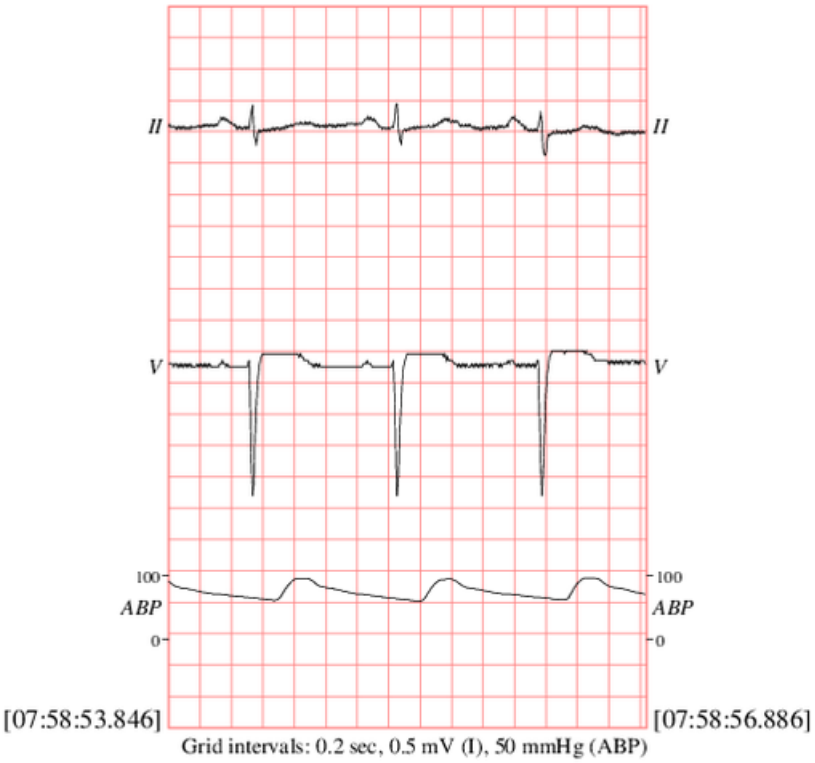
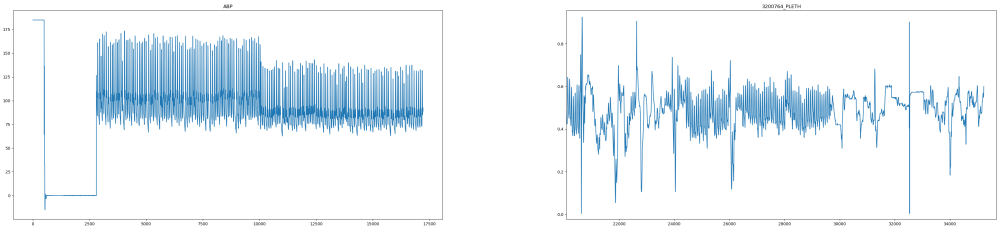


Figure 2.21: Waveform plot in Physionet

of physiological signals; the user interface of Physiobank is shown in Figure 2.20: from there it is possible to choose the specific dataset, the records, which signals to download, the length of the record and time and data format. Moreover, the interface allows the user to see the wave-forms before downloading them, by choosing `Plot waveform` in the toolbox. An example of plotted wave-forms can be seen in Figure 2.21: all the signals contained in the record are plotted on separate lines with the name specified. In the current case, the record contains two signals of ECG (II and V) and the ABP; at the bottom, the starting and ending time of the record (sharp to the millisecond) are indicated. In particular, MIMIC is a data-set that contains data of intensive care unit from approximately 60,000 hospitalizations.

MIMIC has the great merit to provide freely a huge number of real data; despite this, finding a suitable signal in the dataset may be not an easy task. Many series are only some seconds long or are not representative (see Figure 2.22): there are many missing data, flat portions of the signal, flat peaks, and not all the recordings have the same signals (for example some have ECG, PPG, ABP, and other signals and some have only ECG and ABP or a different combination).

It should be also noted that since the MIMIC database contains clinical data



(a) Missing data (b) Noisy data

Figure 2.22: Example of criticalities in MIMIC signals

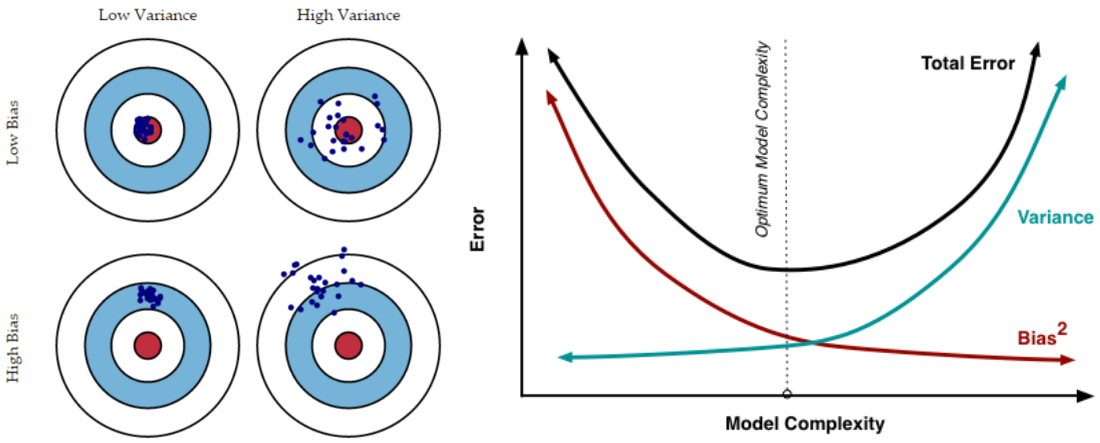
obtained from Intensive Care Units (ICU), all the vital signals contained in it are influenced by drugs that can potentially cause abnormal variations; besides, the average age of subjects is very high (65.8 years). As for the gender, 44.1% of patients are female and 55.9% are males and there is no additional information on the patient from whom the signal is recorded. These conditions may cause the data downloaded from this site to be not representative of the total population or the target population of a study.

2.2 Mathematical Background

Machine learning is a set of statistical techniques that use historical data to learn and make guesses on future or unknown events [62]. The learning can be both *supervised* and *unsupervised*. In the first case, input data $\mathbf{x}$ are associated to a specific output y, so there is an initial set $\mathcal{D} = \{(\mathbf{x_i}, y_i)\}_{i=1}^{N}$, called *training set*. The goal is to learn a function that best maps the input $\mathbf{x_i}$ to the correct output y_i. In the second case, there is no known output y, and the algorithm has to find the hidden structure within the set of data.

Supervised learning is usually seen in techniques like classification (when the output belongs to a discrete set) or regression (when the output belongs to a continuous set). The correctness of the output is determined during the training and validation phases, when the prediction result is compared to the given truth. Unsupervised learning is usually seen in techniques like clustering or dimensionality reduction. Since no output is provided, there is no specific way to determine the correctness of the algorithm.

When conducting this kind of learning, it is necessary to consider that data are typically composed of the actual information they bring, plus some noise (this term refers to irrelevant information or randomness in the dataset). It is vital that the noisy component in the input dataset is as low as possible, otherwise the underlying structure contained in the data would be covered and not recognized by the algorithm.



(a) (b)

Figure 2.23: Bias-variance trade-off (from [63])

Provided that the data have a sufficiently low noise, depending on how well the input data can describe the problem, there may be the occurrence of *underfitting* or *overfitting*. Underfitting occurs when the amount of data used to train the model is not sufficient to catch the underlying pattern of the data, or when the model

used is not suited to describe such pattern. The resulting prediction will have a low variance and high bias. Overfitting, on the other hand, occurs when the model is too complex and trained on too much data. The algorithm learns the noisy component (not only the underlying pattern) and it is not able to detect the actual structure of the data. The resulting prediction has low bias and high variance; the algorithm will perform incredibly well during training and poorly on new data (because it is not able to generalize).

We mentioned the concepts of *bias* and *variance*. A bias is a systematic error, it is not caused by randomness, thus its effect is not reduced when observations are averaged. Variance, instead, is a measure of variability, and it is an index of how much the model is sensitive to the system's noise. A model with high bias generalizes too much the input data, while a model with high variance does not make enough generalization (Figure 2.23a shows the effects of high and low bias or variance). According to Frotmann-Roe (2012) [63], as more parameters are added to the model, its complexity rises and the variance grows, while the bias falls (as in Figure 2.23b). To make a good prediction, it is crucial to balance bias and variance.

Figure 2.24: K-fold cross-validation, K = 5

Overfitting is tricky and much harder to detect than underfitting; one of the most used techniques to avoid it, is *cross-validation*. Cross-validation is a technique used to improve the ability of a model to generalize its prediction; it involves a partitioning of the initial dataset in training and validation set. A particular case is *K-fold cross-validation*: this method involves dividing the initial set in `K` sub-sets, of which `K-1` are used for the training, and 1 for the validation to test the model

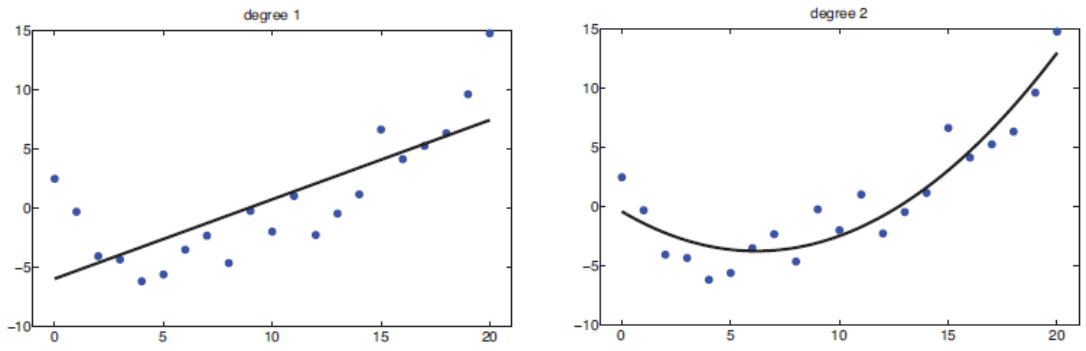
performance. The process is repeated `K` times, and each fold is used exactly once as validation fold. Figure 2.24 shows the `K-fold cross-validation` functioning when $K = 5$. As it is shown, each iteration yields a solution (S_i); the final result of the method (S) is calculated by averaging all the solutions obtained. In this way, the final solution will be less influenced by the noise present in the data.

This work will deal with supervised learning and, specifically, with regression.

2.2.1 Regression

Regression is a type of supervised learning technique where the output $y \in \mathbb{R}$. It is used to investigate and model the relationship between a dependent variable (which we want to predict) and one or more independent variables. The following assumptions are made:

- Samples are representative of the population.
- The parameters used in the prediction are linearly independent.
- The dependent variable is subject to white noise, while independent variables are not.



(a) (b)

Figure 2.25: From [62]: regression on same dataset using a polinomial of degree 1 (a) and degree 2 (b)

As shown in Figure 2.25, regression can be both linear and non-linear. In this work, we will focus on *linear regression*, which is one of the most popular models of regression.

Linear regression assumes that the dependent variable, or *regressand*, can be found as a linear combination of independent variables, or *regressors*. It solves a mathematical problem in the form:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \nu \tag{2.6}$$

where $\mathbf{y} \in \mathbb{R}^N$ is the regressand, N is the number of input data, $\mathbf{X} \in \mathbb{R}^{N \times F}$ is a rectangular matrix of regressors containing all the parameters taken in consideration in the model, F is the number of features considered, $\mathbf{w} \in \mathbb{R}^F$ is the optimum weight vector (the solution of the problem) and ν is the error vector (which, by hypothesis, has a Normal distribution and is negligible). The goal is to find the values of $\mathbf{w}$ that minimize the Mean Square Error $\frac{||\mathbf{Xw}-\mathbf{y}||^2}{N}$. All the variables are standardized (their mean is subtracted and they are divided by their standard deviation) before the analysis.

Ridge Regression is a linear regression technique used to analyze data that suffer from multicollinearity (or collinearity). Multicollinearity is the existence of a near-linear relationship among the independent variables; this can cause inaccuracies in the parameters' estimation and thus degrade the performance of the algorithm. Montgomery et al. [64] study the possible causes of this issue, which can be summarized in:

- *Bad data collection*: data were gathered from just a small sub-space of the independent variable. In this case, the dataset collected is not representative of the population; a more accurate collection will solve the problem.

- *Physical constraints*: collinearity exist in the model or population, independently from the sampling technique.

- *Bad parameters choice*: there is an excessive number of variables, and not all the parameters are independent of one another.

- *Outliers*: extreme values should be removed before regression is applied.

In absence of collinearity, the problem in Equation 2.6 can be solved directly by calculating the pseudoinverse of $\mathbf{X}$:

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{2.7}$$

and this is the case of the Linear Least Square (LLS) method. The most significant drawbacks are that this operation has a high computational cost and it is not robust in the case of multicollinearity. Ridge Regression can overcome these problems by calculating the weight vector $\mathbf{w}$ as follows:

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \tag{2.8}$$

where $\mathbf{I}$ is the identity matrix and λ is a constant (if $\lambda = 0$ the problem becomes equal to LLS). The optimum λ is the one for which the validation mean squared error is minimum, and its value depends on the specific function.

2.2.2 Error measures

We said that the correctness of a method can be stated by analyzing the error. The error is the measure of the deviation between the predicted variable and the observed one, but there is more than one way to measure it. In this thesis, we will consider the *Mean Absolute Error* (MAE) and the *Root Mean Squared Error* (RMSE).

The Mean Absolute Error (MAE) is a measure of the difference between two variables quantifying the same event; it is calculated as shown in Equation 2.9, where y is the observed variable, and $\hat{y}$ is the predicted one (notice that is always ≥ 0, and low values are desirable). A characteristic of MAE is that it has the same unit of measure as the data that are measured.

$$MAE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N} \tag{2.9}$$

The Mean Squared Error (MSE) indicates the average of the squares of the errors, as shown in Equation 2.10, where y is the observed variable, and $\hat{y}$ is the predicted one. MSE is thus always ≥ 0, and it should be as close to zero as possible; it has the unit of measure of the squared quantity that is measured. The squared root of MSE is used as a measure of error as well and it is called Root Mean Squared Error (calculated as shown in Equation 2.11); it has the same unit of measure of the data being measured.

$$MSE = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N} \tag{2.10}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{2.11}$$

Since in MSE and RMSE the error is squared, these measures strongly penalize large errors, so they are to be preferred in a situation where such deviations are particularly undesirable. On the other hand, they are more sensitive to outliers than MAE.

2.3 Related works

The possibility to exploit PTT to estimate a patient's blood pressure continuously has been already shown ever since 2000 by Chen et al. [65] who demonstrated the possibility to make such measure with a certain accuracy thanks to a calibration of the model which has to be repeated after some time due to the wrong assumption that arterial stiffness is constant in time. Mccarthy et al. (2011) [59] used this approach to calibrate the linear algorithm with Omron M6, but the calibration had to be done every 5 minutes. Problems related to the assumption of constant arterial stiffness have been found also in 2020 by El-Hajj et al. [66] who proposed to solve it using two PPG sensors instead of one PPG and one ECG, shifting the issue to the algorithm accuracy.

In the years, the linearity of the relationship between BP and PTT has been proven and several models have been proposed. Among these we cite the proportional model (Equation 2.12), the logarithmic model (Equation 2.13), the inverse model (Equation 2.14) and the inverse square model (Equation 2.15).

$$BP = a * PTT + b \tag{2.12}$$

$$BP = a * \log(PTT) + b \tag{2.13}$$

$$BP = \frac{a}{PTT} \tag{2.14}$$

$$BP = \frac{a}{PTT^2} + b \tag{2.15}$$

The accuracy of such models has been studied by Sharma et.al (2017) [67]; the results showed that $R^2 \in [0.02,0.97]$ (R^2 measures how well the observed variable is represented in the model, it belongs to the interval [0,1], and it should be as close to 1 as possible) and, to apply the models, it is necessary to train them over a large range of BP values, the prediction is more successful for MAP and DBP (0 ± 2.12 and 0 ± 2.13 respectively) while for SBP the error increases (1.3 ± 7.02) though staying always inside the AAMI standards.

Kachuee et al.(2017)[68] analyzed different regression methods and found that non-linear methods like the kernel method or ensemble regression (especially AdaBoost) yield better performance than a linear approximation, suggesting the complexity of the problem to be far higher than the one presented. Calibration-based methods outperform calibration-free methods and methods considering many parameters extracted from the PPG signal have the same performance as the ones considering only the pulse arrival time (PAT), which is defined as the time delay between the peak of the R-wave in the ECG and a systolic point of the PPG recorded with a finger device.

Among the noteworthy works done in the latest years, Shrimanti et al. (2016) [69] uses the inverse linear model (Equation 2.14) evaluating PTT as the time delay from the R-wave peak of the ECG and the systolic point of the PPG, using

signals recorded from 14 healthy people in 5 different positions (recumbent, seated, standing, walking, cycling). The prediction error was highest when walking (MAE 4.4 ± 20.9 mmHg) or cycling (MAE 10.2 ± 16.0 mmHg) since these activities introduce a baseline noise to the signal and highest when the patient was sitting (MAE 0.07 ± 5.8 mmHg).

Liang et al.(2018) [70] uses PAT and PPG features to classify BP in three categories, namely normotension, pre-hypertension, and hypertension using data downloaded from the MIMIC database. PPG features were shown to classify BP categories more accurately than PAT and the combination of both PAT and PPG features was able to improve the performance. Classification of normal blood pressure (normotension) against pre-hypertension reached a correctness of 84.34%, normotension against hypertension reached 94.84%, while normotension plus pre-hypertension against hypertension reached 88.49%.

Khalid et al. (2018) [71] use a single PPG signal to evaluate the patient's blood pressure. The three most significant features of the PPG are found to be pulse area, pulse transit time, and width. Patients are classified into three classes (normotensive, hypotensive, and hypertensive) using three classification methods: regression tree, multiple linear regression (MLR), and support vector machine (SVM). The best results are obtained with the regression tree; the error is -0.1 ± 6.5 mmHg for SBP and -0.6 ± 5.2 mmHg for DBP.

Lazzazzera (2019) [72] presents a new smartwatch (CareUp) able to measure blood pressure in real-time using two photoplethysmograms (PPG). The two signals are acquired, filtered, and cross-correlated to obtain PTT and HR; these data are used as input in a linear model of the kind shown in Equation 2.16.

$$BP = a * PTT + b * HR + c \tag{2.16}$$

The accuracy demonstrated almost matches the AAMI requirements; in particular, the error on DBP estimation is usually less than SBP. The procedure lasts about 45 seconds (of which 30 s are used to acquire a signal of sufficient length and 15 s to output the estimation); the algorithm has to be calibrated on each user (though no information about the frequency of the calibration is given).

Ding et al. (2019) [73] discuss the feasibility of exploiting PTT in continuous BP monitoring and the main challenges of the method. In particular, they state that non-invasive methods are not accurate enough to replace invasive measures. Among the main challenges of PTT-based methods, they point out that DBP estimation has a lower error than SBP (this might happen because DBP varies more slowly in peripheral arteries than SBP). Moreover, many studies try to estimate BP only using PTT even if this sole characteristic is not able to fully represent the problem and lastly he remarks on the already mentioned problem of variable arterial stiffness, not considered in the model and not solved even by adaptive algorithms like adaptive Kalman filter or recursive least square.

Chen et al. (2019) [74] use support vector regression (SVR) on data downloaded from the MIMIC dataset; the choice of the model is done to overcome errors caused by the use of a traditional linear model on a non-linear problem; moreover, a parameter optimization on the patient is performed. The algorithm is fed with 14 features related to blood pressure, which include PTT, HR, and characteristics of the pulse waveform. The authors obtain an error of 3.27 ± 5.52 mmHg for SBP and 1.16 ± 1.97 mmHg for DBP. The errors also meet the requirements of AAMI and the British Hypertension Society.

Also Zhang et al.(2019) [75] use an SVR model with radial basis function (RBF) with parameters optimization to predict blood pressure from PPG features, HR, and PTT. The optimization is done with a 10-fold cross-validation to improve the accuracy of the model.

Chapter 3

Objective and methodology

The objective of this thesis is to find an algorithm able to predict the patient's blood pressure given his/her ECG and PPG. The prediction error should be lower than 10 mmHg for 85% of the times, as indicated by the authorities for validated medical devices. The algorithm is intended to be used for the device developed in the SINTEC project [76], a European project that aims at developing soft, sticky, and stretchable sensor patches to be used for the recording of vital signal and their elaboration.

Since the SINTEC prototype hasn't been developed yet, the vital signs were downloaded by the MIMIC III database. Among the recordings, the ones downloaded needed to have the following characteristics:

- Contain at least ECG, PPG, and ABP; notice that the database already provides the signals aligned to the millisecond.
- Have few artifacts and sufficiently low noise.
- Have a length larger than 30 minutes.

All the recordings satisfying such criteria were downloaded as `.csv` files and aggregated for each patient as to have a single file of a larger length. Among these aggregated files, only the ones with a recording duration of one hour and a half were retained. Figure 3.1 shows a sample of dataset downloaded from MIMIC: the time format was set to `Elapsed time`; the first row stores the names of all the signals included in that record, while in the second row the respective units of measure are present. Missing values are denoted with dashes.

The algorithm used was a linear regression of the type shown in Equation 2.12, but considering also the heart rate of the subject, that is:

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

```
'Elapsed time','RESP','PLETH','ABP','II','III','V'
'seconds','pm','NU','mmHg','mV','mV','mV'
0.000,0.350,-,74.341,0.470,0.415,0.885
0.008,0.335,-,78.541,0.440,0.400,0.850
0.016,0.319,-,84.793,0.400,0.380,0.805
0.024,0.303,-,92.560,0.370,0.370,0.760
0.032,0.287,-,100.912,0.330,0.350,0.715
0.040,0.270,-,109.069,0.300,0.335,0.675
0.048,0.255,-,116.347,0.275,0.320,0.650
0.056,0.240,-,122.696,0.245,0.305,0.630
0.064,0.226,-,127.727,0.235,0.300,0.625
```

Figure 3.1: Structure of downloaded csv file

Considering that the prediction has to be performed separately for systolic and diastolic pressure, Equation 3.1 becomes:

$$\begin{cases} SBP = a_s * PTT + b_s * HR + c_s \\ DBP = a_d * PTT + b_d * HR + c_d \end{cases} \tag{3.2}$$

As mentioned in Section 2.1, the heart rate is found as the time delay between two consecutive peaks of the R-wave in the ECG, and PTT is found as the time delay between a peak of the R-wave in the ECG and a characteristic point of the PPG, here the systolic point is considered (notice that this choice reduces the computational load).

As mentioned in the related works, a major issue of this approach is that it requires a specific calibration on each user. Calibration allows to adequately correlate pulse transit time and blood pressure (both systolic and diastolic), thus finding the regression coefficients $a_s, b_s, c_s, a_d, b_d, c_d$ of Equation 3.1. In an ideal case, calibration is performed only once, when the patient starts using the device.

The work is performed using three approaches:

- *Whole file*, Figure 3.2: the signal (built as explained above) is pre-processed and processed as a whole; training and testing of the regression are done *beat-to-beat* (each cardiac cycle is evaluated separately from the others).
- *Real-time simulation*, Figure 3.3: the signal is subdivided into smaller segments, which are analyzed individually; training and testing of the regression are again done beat-to-beat.
- *Comparison with cuffed measurements*, Figure 3.4: the signal is subdivided into smaller segments, which are analyzed individually; the training phase of the regression is done beat-to-beat (because of a lack of data), but the test phase is performed *non-beat-to-beat* (cardiac cycles are aggregated as described in Section 3.3).

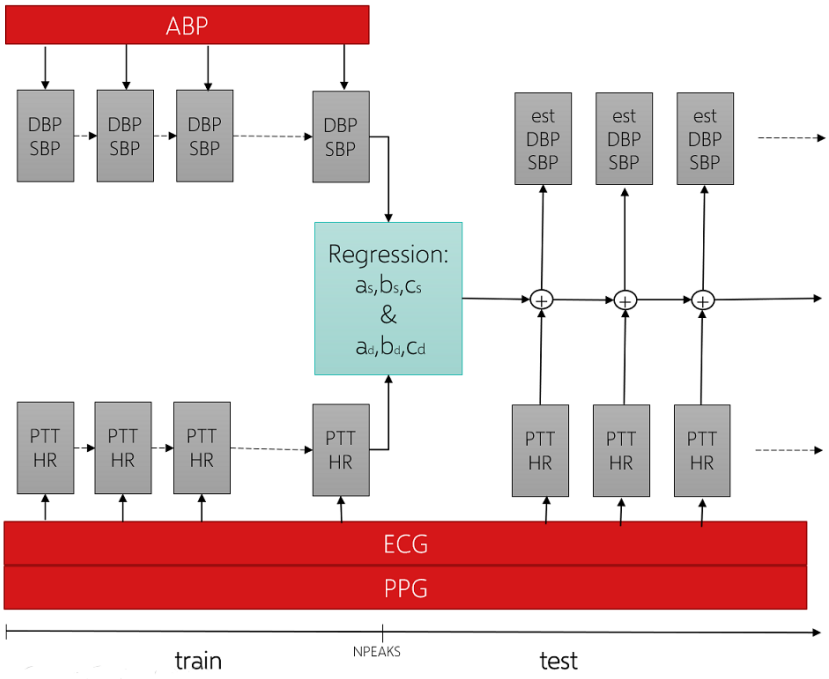


Figure 3.2: Block diagram of *whole file* approach

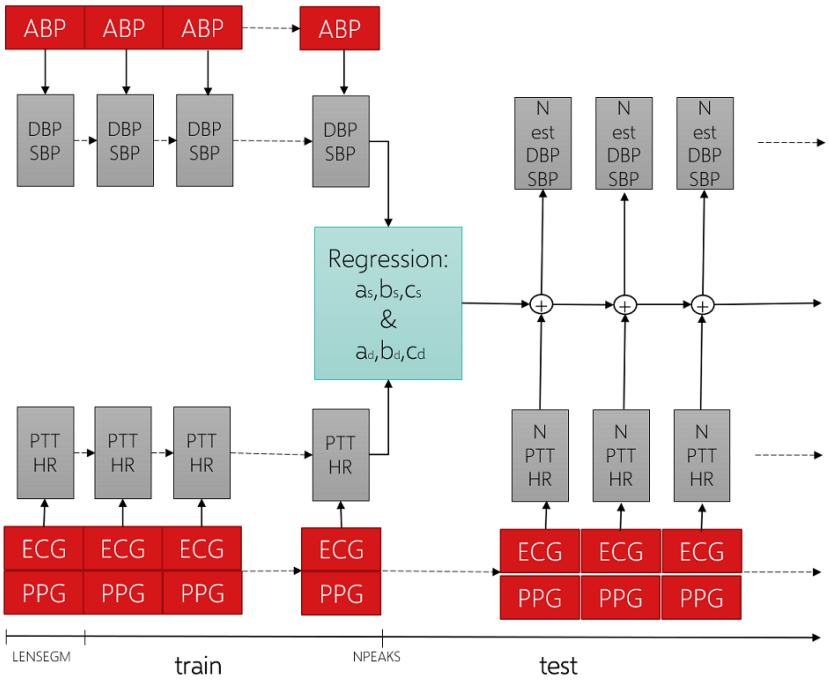


Figure 3.3: Block diagram of *real-time simulation* approach

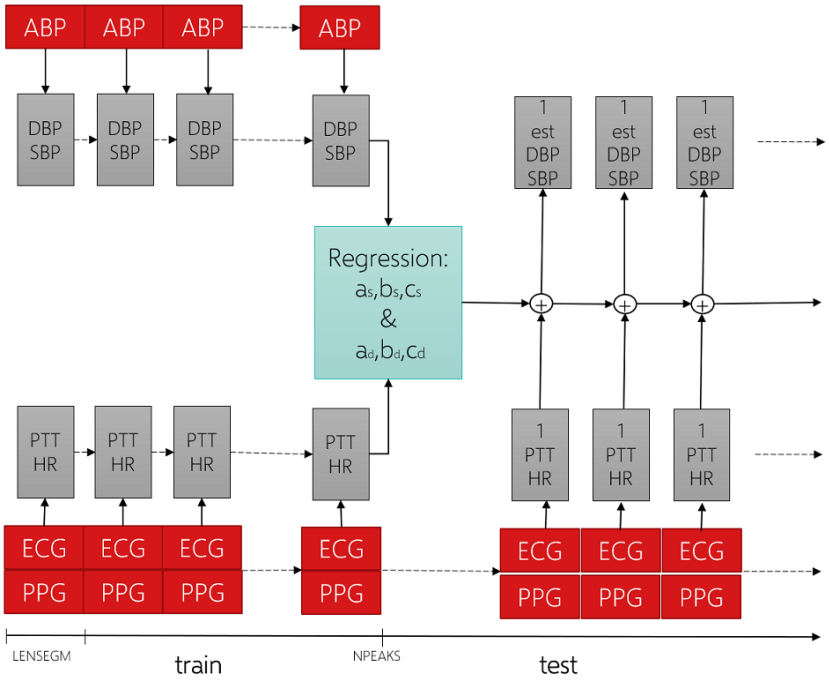


Figure 3.4: Block diagram of *comparison with cuffed measurements* approach

The algorithms are fully written in Python and the main libraries used are Numpy, Matplotlib, Scipy, and Scikitlearn.

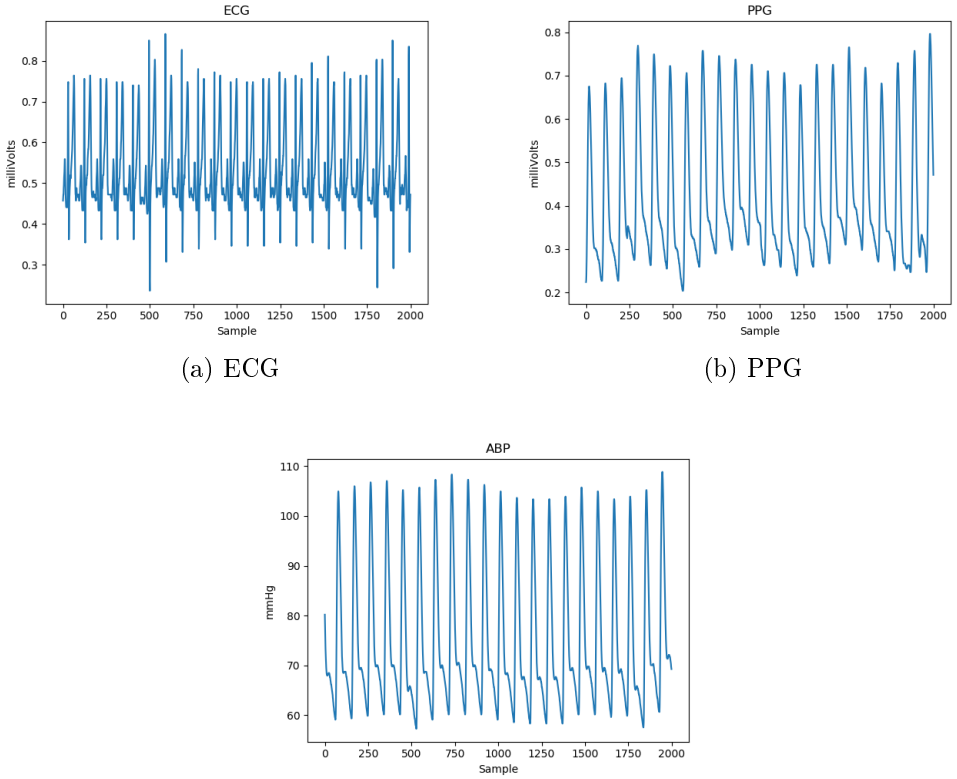
3.1 Whole file

The raw signals are shown in Figure 3.5; before the elaboration, a pre-processing of the signals is needed. ECG and PPG are thus interpolated and filtered to remove some noise (the ABP is not filtered, since it would impact the values of local maxima and minima).

3.1.1 Interpolation and filtering

The interpolation step is performed using `scipy.interpolate.interp1d` using a cubic interpolation; from one original sample, 10 samples are produced. Due to the previous step, the sampling frequency changes from the one used in MIMIC, i.e. 125 Hz, to 1250 Hz. The filtering step was performed using the Butterworth filter included in Scipy (`scipy.signal.butter`) with order 3 and band-pass frequencies equal to $[0.5 - 45]$ Hz for ECG and $[0.5 - 7]$ Hz for PPG; the Bode plots of the digital filter frequency response for ECG and PPG signals, are shown in Figure 3.6). The digital filter was applied using command `scipy.signal.filtfilt` because it

was seen that it applied the filter without shifting the signal on the x axis. The frequencies were chosen after a visual analysis of the Power Spectral Densities, found with `scipy.signal.welch` and shown in Figure 3.7; Figure 3.8 shows the difference between the original signal and the filtered one.

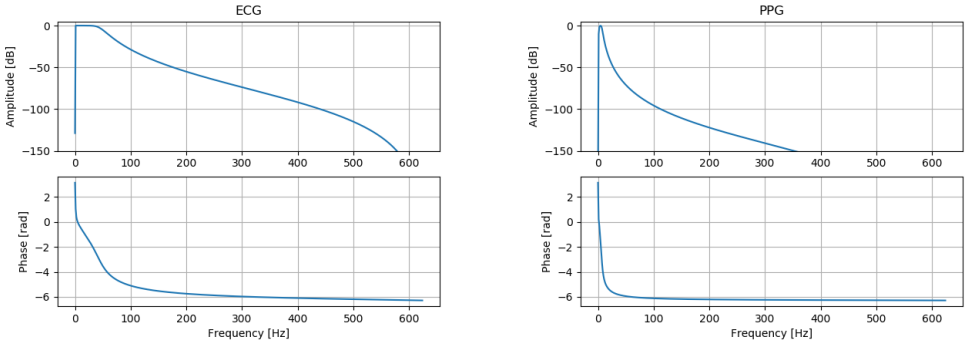


(a) ECG

(b) PPG

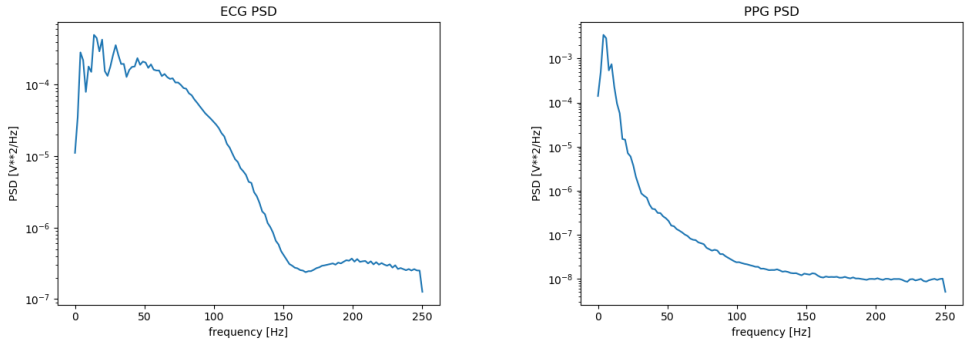
(c) ABP

Figure 3.5: Segments of raw signals



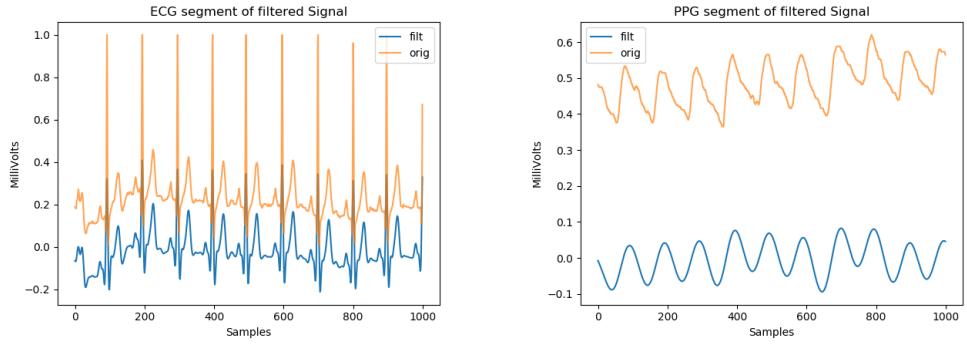
(a) ECG (b) PPG

Figure 3.6: Bode plot of digital filter frequency response



(a) (b)

Figure 3.7: Power spectral densities of the ECG and PPG



(a) (b)

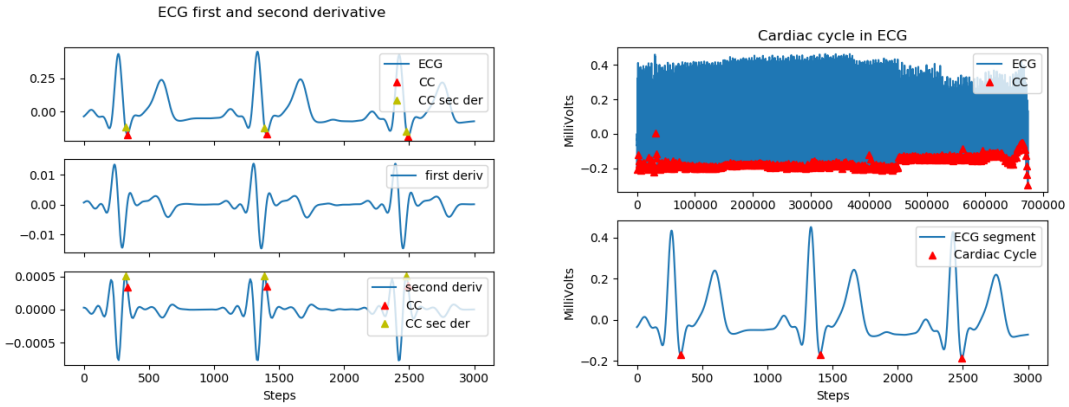
Figure 3.8: Segment of original and filtered signal for ECG and PPG

3.1.2 Finding local maxima

To find the local maxima of the signals, it is necessary to know the length of a cardiac cycle (CC). Such length is specific of the patient, and in this work is found as the average distance between the local minima of the ECG minus the standard deviation times a coefficient, as shown in Equation 3.3.

$$CC = mean_{CC_{ECG}} - std_{CC_{ECG}} * K \tag{3.3}$$

The ECG is chosen because its maxima and minima are easily recognizable; moreover, it is to be noticed that this step is performed to optimize the search of local maxima and minima in the following operations (the default CC used for this step considers that the minimum length of a cardiac cycle is of 330 ms). Mathematically, minima are found where the first derivative is equal to zero and the second derivative is larger than zero, but this operation requires a high computational cost. An alternative significant point can be found where the second derivative of a function has its local maxima; nevertheless, as shown in Figure 3.9a, local minima found as the maxima of the inverse funcion $-f(x)$ (named as `CC` in the picture) achieve a high accuracy, so points found as the maxima of the second derivative (named as `CC sec der` in the picture) are discarded. Peaks of the signal are found with function `scipy.signal.find_peaks`; the result of the above process is depicted in Figure 3.9b.

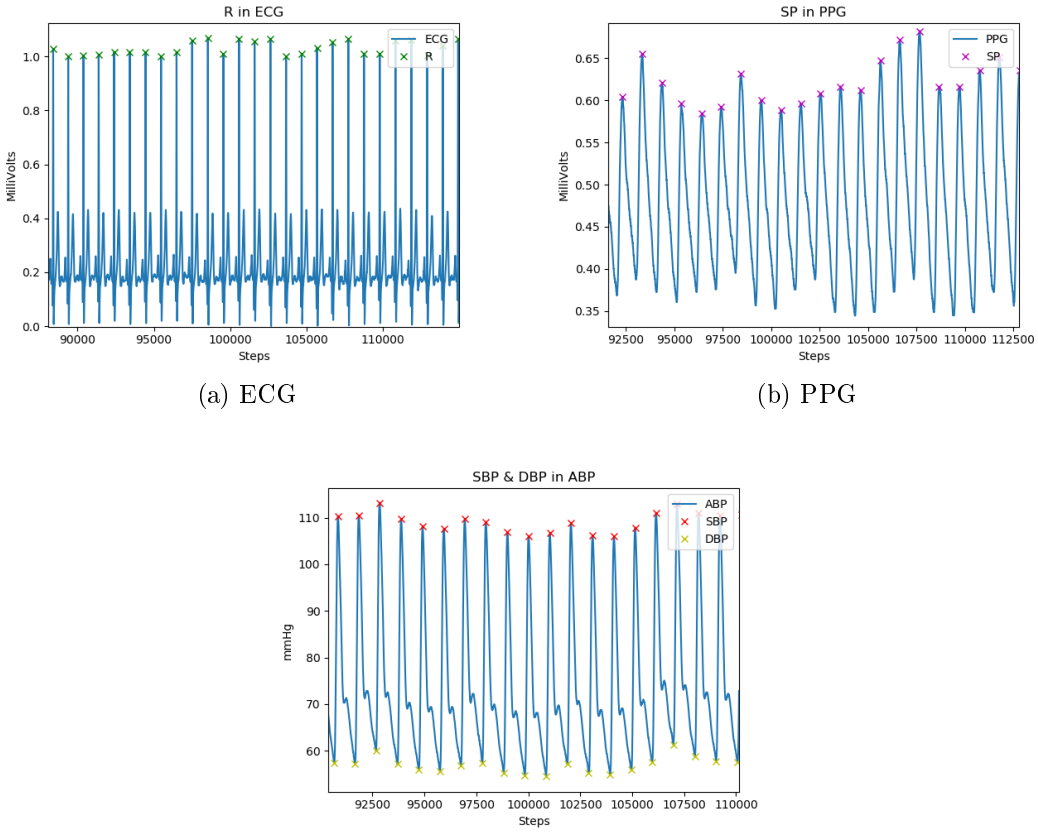


(a) (b)

Figure 3.9: Finding Cardiac Cycles in ECG

Knowing the value of CC, the R-peak of ECG, the Systolic Point of PPG, and SBP and DBP of ABP are found. As shown in Figure 3.10, such points are taken quite accurately except for corrupted portions of the signal. To achieve such accuracy it is necessary to set the optional parameters of function `find_peaks`, namely `distance` between two consecutive peaks (this parameter is the length

of a Cardiac Cycle, found in the previous step), `height` of the peaks $height = (h_{min}, h_{max})$, and for the PPG and ABP also the `width` of the wave-form $width = (w_{min}, w_{max})$. Such parameters are fundamental to detect artifacts in the signal especially in the PPG, which is heavily affected by them (see Figure 3.11a); notice that in MIMIC not all the recordings have the same scale, so it is necessary to modify the parameters for each data-set. As for the ABP, the parameters are set to cover the ranges $190 - 80$ for SBP and $120 - 45$ for DBP (so the whole range of pressure).

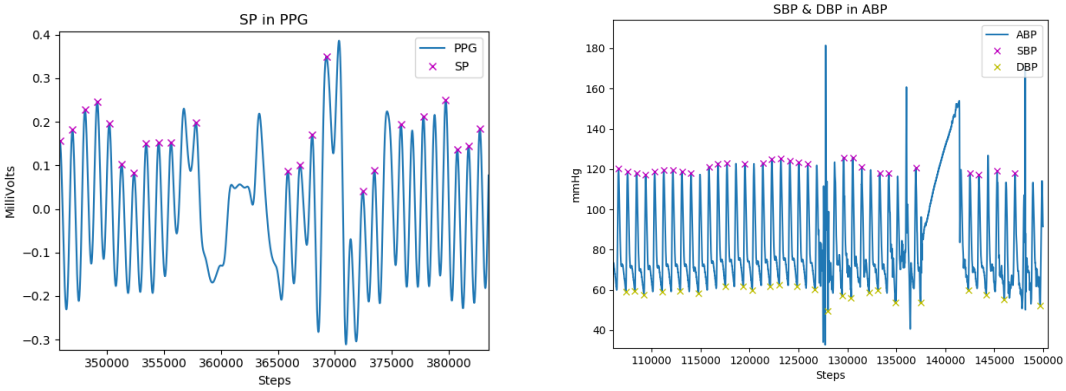


(a) ECG

(b) PPG

(c) ABP

Figure 3.10: Peaks in the series

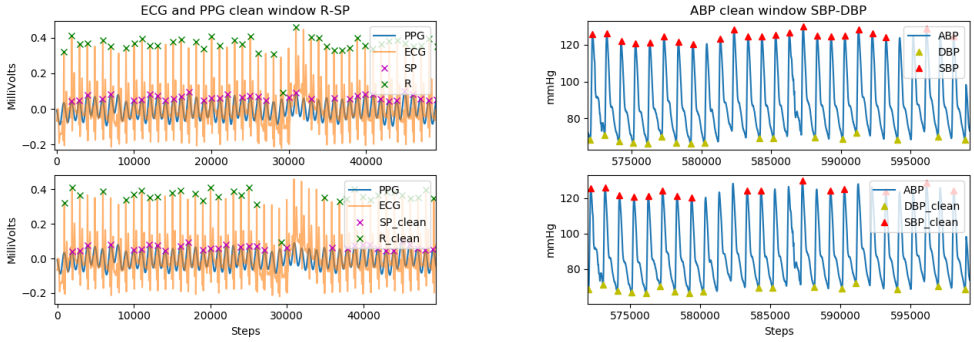


(a) (b)

Figure 3.11: Setting parameters in `scipy.signal.find_peaks`

3.1.3 Finding couples (R,SP) and (SBP,DBP)

At this point, a further step is made to retain only the couples (R,SP) where both points are present in the same cardiac cycle. This is done to avoid misalignment errors in the case where a point (R or SP) was correctly taken and the other is missing (if such an event occurs, the whole cycle is skipped). This process is shown in Figure 3.12a; a systolic point is considered to belong to the same cardiac cycle as the R-peak under exam if it occurs after the R-peak with a delay ranging from $dTep_{min} = 0.1$ s to $dTep_{max} = 1$ s where the mentioned constants are to be chosen depending on the distance between the ECG electrodes and PPG sensor and also depending on the patient's age and health status. In the present case, it was seen that [0.1,1] s was a good range to pair the correct points but, in other cases, the range could be narrowed if target patients are younger and in better health conditions. Figure 3.12c shows a pseudo-code implementing such a task. Notice that in Figure 3.12c R and SP are the sets containing the timestamp of the occurrence of the peaks (the timestamp is included in the file downloaded from Physionet). The same is done with SBP and DBP points in the ABP signal: (SBP, DBP) couples are retained only if both points were correctly taken in the same cardiac cycle (see Figure 3.12b).



(a) (R,SP) (b) (SBP,DBP)

- if SP[0] < R[0]:
  - SP.pop(0) """ start from the first R point """
- For i in R:
  - For j in SP:
    - if SP[j] > R[i] +dTepmin and SP[j] < R[i] + dTepmax:
      - Keep couple (R[i],SP[j])

(c) Pseudo-code

Figure 3.12: Cleaning couples (R,SP) and (SBP,DBP)

3.1.4 Evaluating PTT

The following step is to evaluate the PTT as the time delay between $R[i]$ and $SP[i]$ in milliseconds and the HR as the time delay between $R[i]$ and $R[i+1]$. Describing it at a high level, the cycle that takes care of this task takes as reference an SBP point and looks at all the R and SP points in an interval around the SBP. If it finds the couple (R,SP) belonging to the same cardiac cycle, it computes the PTT, HR and saves PTT, HR, SBP, and DBP values. Figure 3.13 shows a pseudo-code describing such process: the outer cycle scans all the SBP points and the inner cycle the R and SP points. If a match is found, the outer cycle moves to SBP_{i+1} while the inner cycle starts the search or the matching couple not from $j = 0$ but from $j = j + 1$. A match is found if the points occur in the correct order and in a "reasonable" interval of time, that is:

1. SP occurs after SBP but no more than 1.5 seconds after it

2. R occurs after SBP but no more than 0.8 s after it

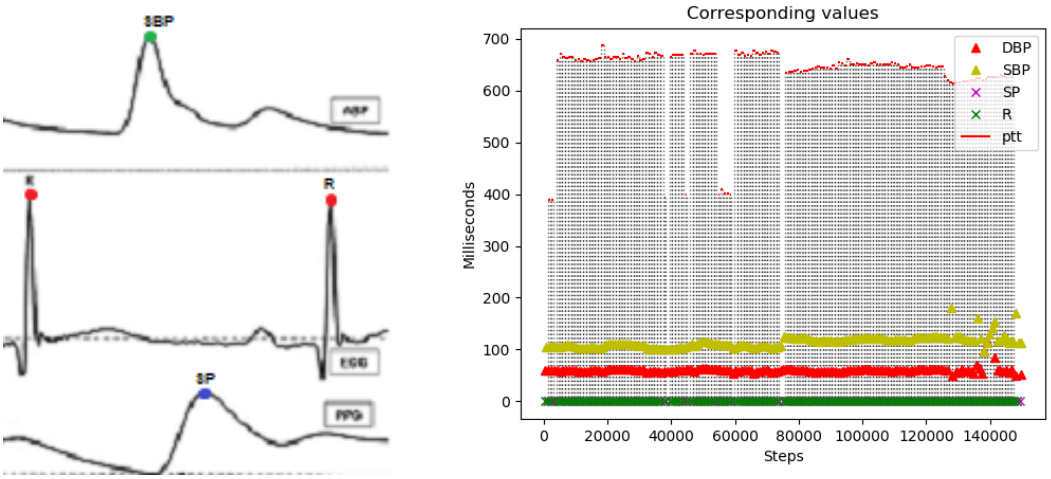
3. SP occurs at least 0.7 s after R

Notice that the above intervals of time are heavily influenced by the patient's health condition: considering that MIMIC3 deals with patients in Intensive Care Unit, the current values may be too high for healthy people. Moreover, the correct order of occurrence of the three peaks should be the one shown in Figure 3.14a, but such order was found to be inconsistent in the current data and no explanation was found in PhysioNet on the points where the signals are recorded.

```
start = 0
for i in range(SBP):
    found = 0
    j=start
    While found ==0 and j<len(R)-1 and ctr<=Nattempts
        ptt = SP_time - R_time
        If SP_time > SBP_time and SP_time < SBP_time + dTps and R_time >
        SBP_time and R_time < SBP_time + dTes and SP_time < R_time + dTpe :
            found = 1
            start = j+1
            hr = 60/(R_time[j+1]- R_time[j])
            save PTT,HR,SBP,DBP
```

Figure 3.13: Pseudo-code of cycle finding PTT

If the matching couple is found, PTT and HR are evaluated as shown in Equations 2.2 and 2.1 and saved along with SBP and DBP. If a matching couple is not found for a number of times equal to `Nattempts`, the couple (SBP, DBP) is discarded since it has no matching (R, SP). The value of Nattempts shall be chosen carefully, since a too low value could cause a misalignment of the two cycles even if there are few missing couples (SBP, DBP) and (R, SP), while a too high value would enlarge the computational time in case there are long portions of missing couples. A good compromise was found setting `Nattempts` $= 10$. Figure 3.14b shows the output of the portion of code described above: the patient has a PTT of 680 milliseconds on average and the evaluation was done correctly for each cardiac cycle, with only a few outliers.

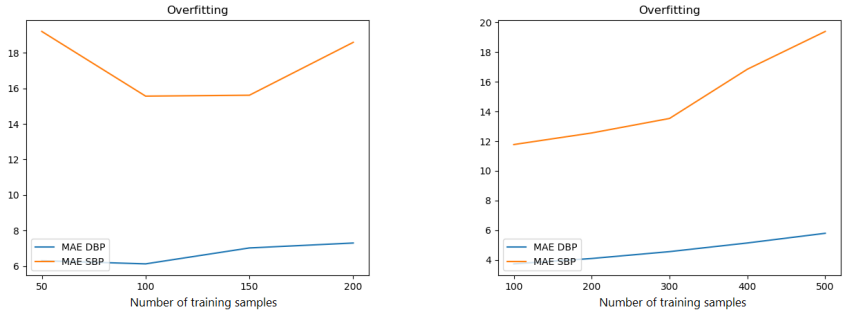


(a) (b)

Figure 3.14: (*a*) Time occurrence of R-wave peak, Systolic Blood Pressure, and Systolic point; (*b*) Finding PTT

3.1.5 Performing regression

The output of this step is the input of linear regression, performed with ridge-regression. The number of training samples is chosen after a study on the error variation. According to this study, almost all the series require a number of training samples of about $150 - 200$ samples and, for a larger number, the error starts to increase, as Figure 3.15 shows. This range seems to be the same for series of different lengths and quality; this may suggest that the problem is heavily subject to overfitting and may be an advantage in the case of calibration since the calibration phase would require less time. Nevertheless, the data available in the current work don't allow any deeper study on this matter. This happens because the patients from whom these signals are recorded are in a health status that is not representative of the population. This affects both the quality of the signal (in a not hospitalized subject, high and sudden variations of blood pressure are not expected) and the validity of the mathematical expression used (Equation 3.1), which is based on the strong assumption of constant arterial stiffness. In particular, the mentioned simplification is also the cause of the recurrent calibration needed in real applications but, in the present case, it is not possible to tell how much time should elapse between one calibration and the following due to the fragmentation of the data present in MIMIC. In a real case, the calibration should be performed by a professional (IT-specialist, a physician, or a nurse) but the time elapsing between a calibration and the other depends also on the performance of the device used. Unfortunately, the prototype of the SINTEC device is still not available.



(a) ECG (b) PPG

Figure 3.15: Study on error variation changing the number of training samples

Training is performed with K-fold cross-validation, being $K = 5$. The goal is to have at least a prediction error (MAE) lower than 10 mmHg at least the 85% of the times, as prescribed by the international organizations that validate medical devices.

3.2 Real-time simulation

The above process is performed also simulating a real-time acquisition of the data, by reading the `csv` file line by line. Figure 3.16 shows a pseudocode describing the process. As shown, samples are acquired until a segment of the signal of predefined length is built; such length is defined by `LENSEGM`. It is crucial to choose a suitable value because a too short segment would enlarge the error due to the start and end peaks of the segment and compromise the validity of the data, while a too long segment would impact the speed at which the prediction is done.

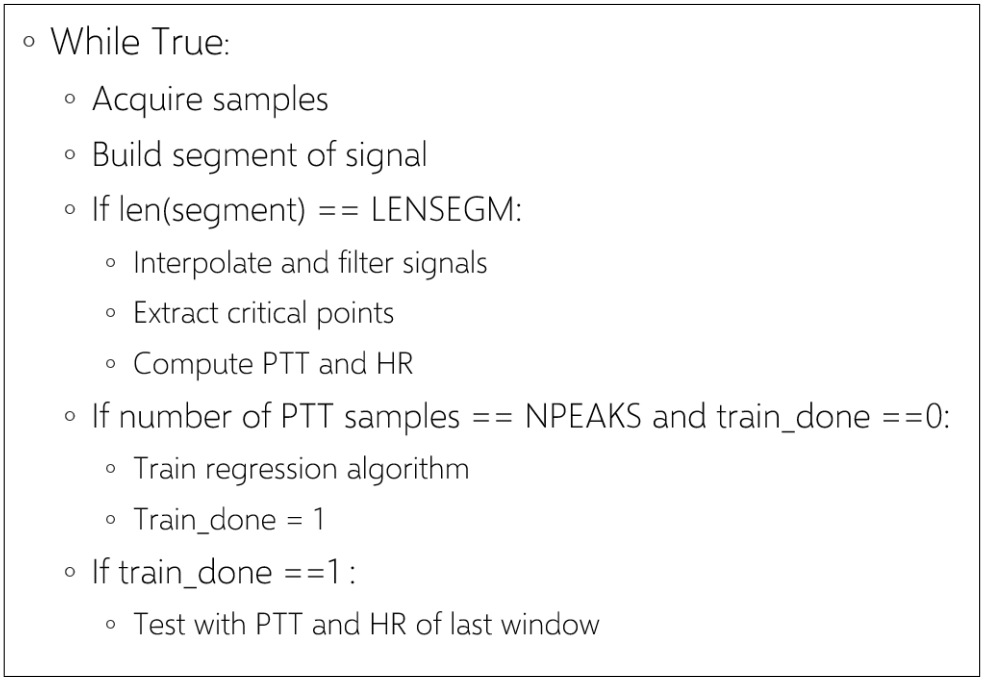


Figure 3.16: Pseudo-code describing the simulation of *real-time* acquisition

After the segment is built, differently from when the whole signal is processed, the quality of the segment is checked using two quality indicators: Noise to Signal Ratio (NSR) and Autocorrelation. These two indexes are calculated over the segment of PPG because, as explained in the previous chapter, it is much more subject to artifacts and noise. The threshold values for the quality indexes depend on the length of the segment; for a segment of 1250 samples, $NSR \leq 10$ and $AUTOCORR \leq 150$ (considering how Python computes them) are good values.

If the segment passes the quality check, it is processed as seen before (otherwise it is skipped): the segment is interpolated and filtered (with the same filter as before). After that, the critical points are extracted and coupled ((R, SP) and (SBP, DBP)) as before, and the PTT and HR are computed (again as already described). Notice that this process is repeated not only once (as done in the first

part) but for every single segment.

A further expedient is used to lower the error: when using the whole signal, the length of a cardiac cycle used to find the correct distance between two consecutive peaks is computed only once; instead, in the current approach it is adjusted for each segment taken before the training. The adjustment is done by computing a weighted moving average between the current value found(w_{curr}) and the one used in the previous step (w_{prev}), giving a higher weight to the previous ($w_{curr} = 1$, while w_{prev} is equal to the number of segments already examined).

When a `NPEAK` number of PTT and HR samples is acquired, the training is performed over those samples with K-fold cross-validation ($K = 5$) and using ridge-regression. The value of NPEAK is chosen considering the result of the study on overfitting performed when considering the signal as a whole. Since the algorithm shall work automatically, a unique value for all signals shall be found, and so the chosen value was `NPEAK` = 200 samples (as explained before, the majority of the signals were trained using about 200 samples).

After the training is performed, and the coefficients for the regression equation are found, all the samples of PTT and HR acquired in the subsequent segments are used for the test, again performed with ridge-regression.

3.3 Comparison with cuffed measurements

Up to this point, all the analysis has been performed considering each cardiac cycle singularly.

A further experiment was done by slightly modifying the *real-time simulation* algorithm to test the data not beat-to-beat, but averaging the critical values extracted from the segment under analysis (it is not done also for the training just because of a lack of data, though it is reasonable to have a training more accurate than the testing). For example, if, in the previous section, a segment produced 5 samples of PTT, 5 of HR, and 5 of BP; the 5 samples of PTT are averaged (the same goes for the 5 samples of HR and the 5 of BP), and the test is done only feeding the formula with the resulting sample of PTT and HR and comparing the estimated BP with the BP sample.

This approach emulates the functioning of cuff-based blood pressure monitors, that average the measure over about a minute and can yield very accurate results.

3.4 Summary of input parameters

The following parameters are to be inserted:

- `ECG`, `PPG`, `ABP`: input signals.
- `fs`: sampling frequency.
- `NPEAKS`: number of samples of PTT and HR to use for training.
- `LENSEGM`: length of the segment to acquire.
- `NSR`: threshold of noise to signal ratio; this value depends on `LENSEGM`.
- `AUTOCORR`: threshold of PPG autocorrelation coefficient; this value depends on `LENSEGM`.

To customize the algorithm to a category of target patients, it is possible to modify also other parameters:

- `K`: coefficient of K-fold cross-validation; this value depends on the quantity of samples used for the training.
- `Nattempts`: number of times a matching (SBP,DBP) (R,SP) is attempted; this value depends on the quality of the signal.
- $dTep_{min}$: minimum delay between a peak of the R-wave and a peak od the PPG belonging to the same cardiac cycle; this value depends on the characteristics of the patient.
- $dTep_{max}$: maximum delay between a peak of the R-wave and a peak od the PPG belonging to the same cardiac cycle; this value depends on the characteristics of the patient.
- $(height_{R_{min}}, height_{R_{max}})$: range of values that can be acquired by the R-wave peak; this values depend on the characteristics of the patient.
- $(height_{SP_{min}}, height_{SP_{max}})$: range of values that can be acquired by the systolic point; this values depend on the characteristics of the patient.
- $(width_{SP_{min}}, width_{SP_{max}})$: width of a PPG wave not affected by artifacts; this values depend on the characteristics of the patient.
- $(width_{BP_{min}}, width_{BP_{max}})$: width of an ABP wave not affected by artifacts; this values depend on the characteristics of the patient.

Chapter 4

Results

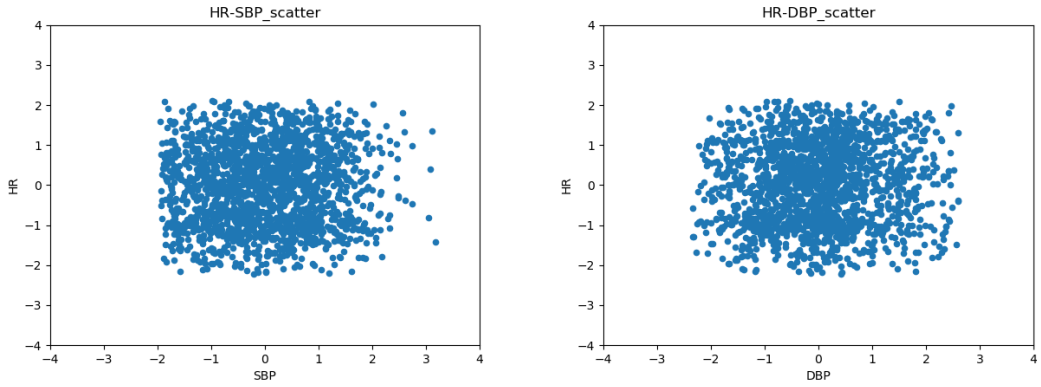
In this chapter, the results of the study described in Chapter 3 will be presented. Patient 3002229 is taken as a representative example.

As expected the use of the MIMIC dataset created some problems because:

- Portions of data recorded in different moments were pasted together, so the file showed some discontinuities.
- BP values in many cases are "peculiar" (i.e. particularly high or low, some times the value increases at high speed, some other times there is a very significant variability throughout the file) but there is no information on what caused the anomaly (some drugs, stoke, incorrect positioning of the sensor, motion artifact). This is a serious problem that prevents a reliable conclusion from being reached.
- The values of ECG and PPG had different scales, so it was not possible to write an algorithm that worked automatically for all the patients (we suppose that using the device developed by SINTEC the scale will be the same for every patient and this problem will be automatically overcome).

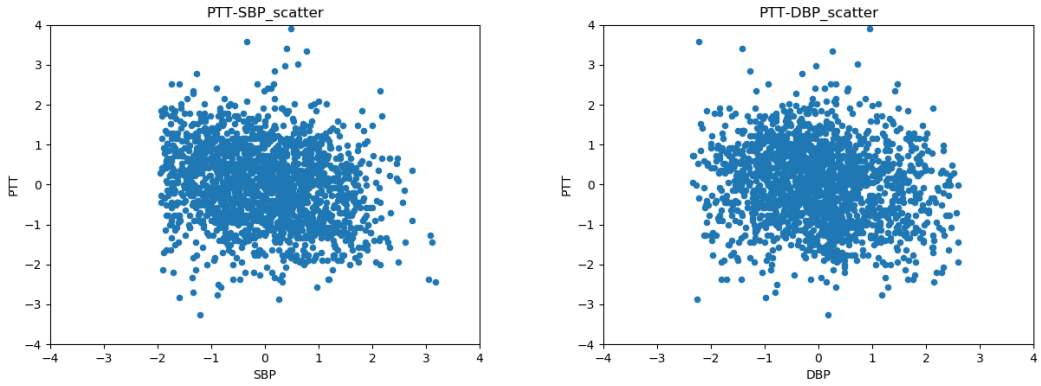
Since the algorithm is based on linear regression, the linearity of the problem has been declared. The scatter plots in Figure 4.1 4.2 show that there is no linear correlation, and a high variability is present (the standard deviation goes up to 2 without outliers). This means that the problem, by itself, already contains a high variability; for this reason, only points with a `z-score` ≤ 5 (`z-score` $= \frac{x-\mu}{\sigma}$, where x is the parameter under analysis, μ and σ are the mean and standard deviation of the considered parameter) were considered outliers and therefore not considered in the prediction.

The correlation heatmap in Figure 4.3 shows the correlation coefficient of all the variables considered. As expected, it is a symmetric matrix and all BP measures are highly positively correlated between themselves. Moreover, the heart rate is positively correlated to BP but with a low coefficient (this result was expected since



(a) (b)

Figure 4.1: Scatter plots of data: heart rate vs blood pressure



(a) (b)

Figure 4.2: Scatter plots of data: pulse transit time vs blood pressure

many other variables influence the heart rate, among which is to be mentioned also the mental state of the patient). Pulse transit time is, as expected, negatively correlated both with blood pressure variables and heart rate. In this case, a higher coefficient would have been expected, but the result could be explained considering the difficulties related to the use of the MIMIC dataset.

Figure 4.4 shows a detail of the signal recordings of patient 3002229. The waveforms show some irregularity (especially the ECG).

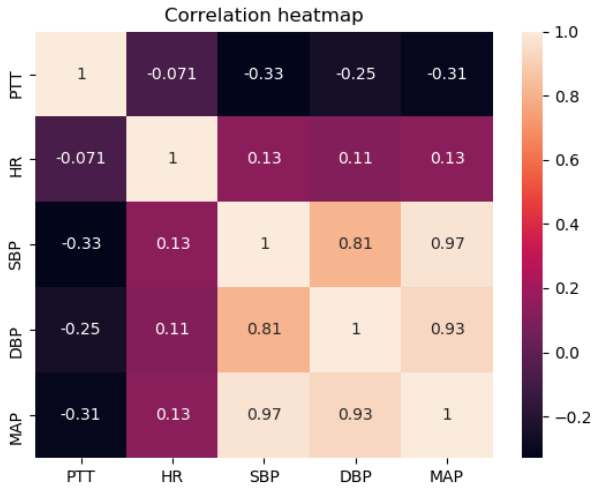


Figure 4.3: Correlation heatmap of the data

4.1 Whole file

Figure 4.5 shows the performance of the algorithm in predicting DBP values of a patient using the *whole file* approach. In particular, Figure 4.5a shows the performance of the training algorithm (fed with 200 samples, of which $\frac{1}{5}$th is used as a validation set, due to the K-fold cross-validation). Figure 4.5b is a histogram showing the prediction errors done in the training phase. The majority of the errors falls in the range [−5,0] mmHg; overall, the threshold of 10 mmHg is not exceeded. Figure 4.5c shows a zoom of the prediction performed in the test phase; it is noticeable that the predicted values follow the real pattern, even if with a certain discrepancy. Figure 4.5d reports the prediction errors of the test phase; the curve has a Gaussian shape, it is centered in 0, and the threshold of 10 mmHg is exceeded a negligible number of times (47).

Figure 4.6 shows the performance of the algorithm in predicting SBP values of a patient. Therefore, the algorithm is able to follow the real trend, both in the training and testing phase, even if with a discrepancy. This discrepancy is wider in the test phase than in training, and it is larger than the one shown in Figure 4.5. In this study, the error on SBP prediction is almost always larger than the error on DBP prediction; this could be explained considering that SBP is more unstable than DBP. Table 4.1 reports the numerical errors found for patient 3002229; the error is computed using both Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) and it is reported as $mean \pm std$.

Table 4.2 reports a summary of the results obtained by the regression performed on all the data extracted with the *whole file* approach. The regression settings were

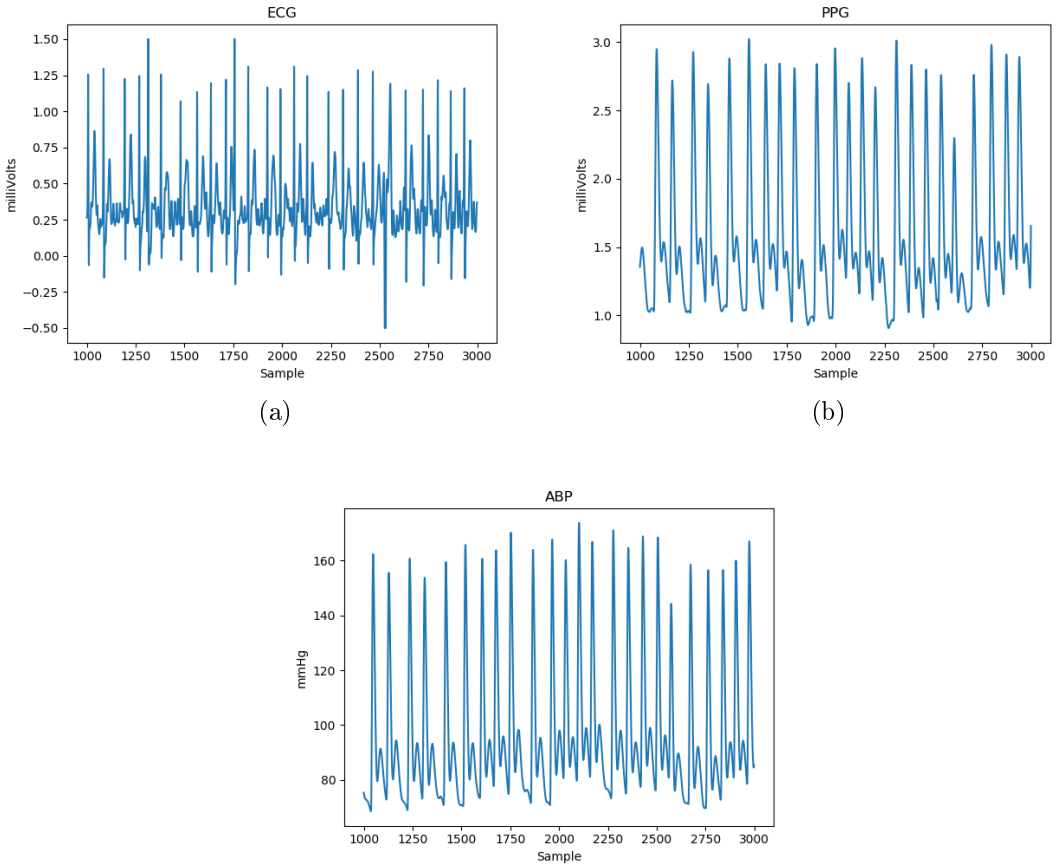
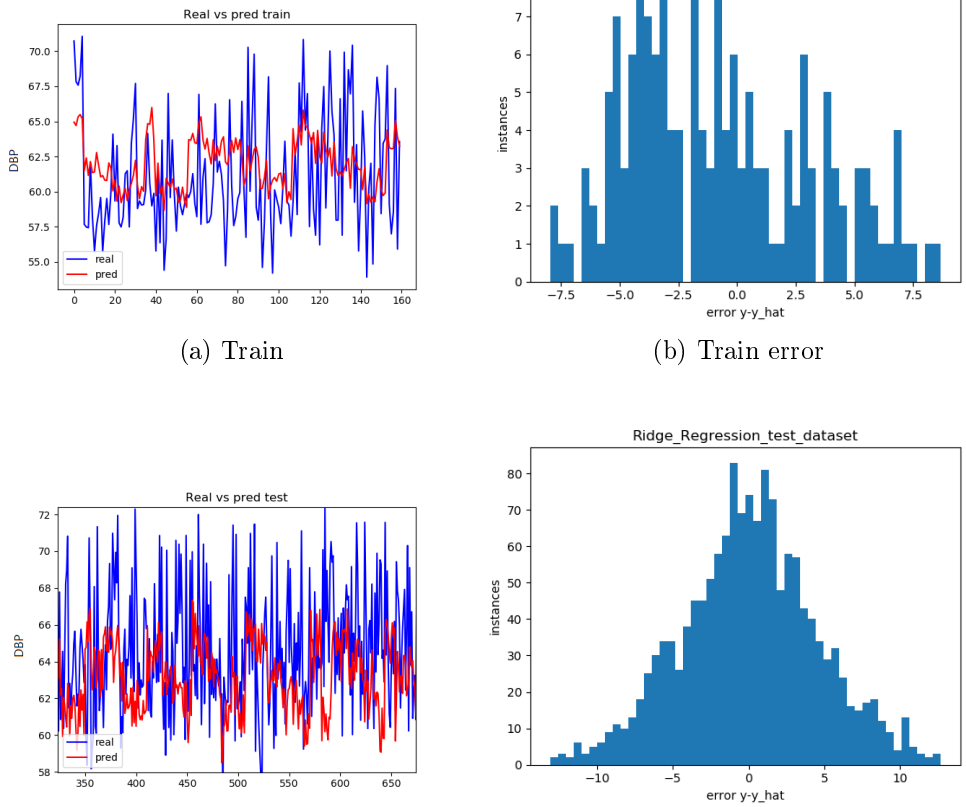


Figure 4.4: Zoom of ECG, PPG, and ABP recordings of patient 3002229, and overfitting

SBP				
patient	MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]
3002229	4.3	5.4	4.9	6
DBP				
patient	MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]
3002229	3.3	3.9	3.5	4.5

Table 4.1: Results of *whole file* for patients 3002229



(a) Train

(b) Train error

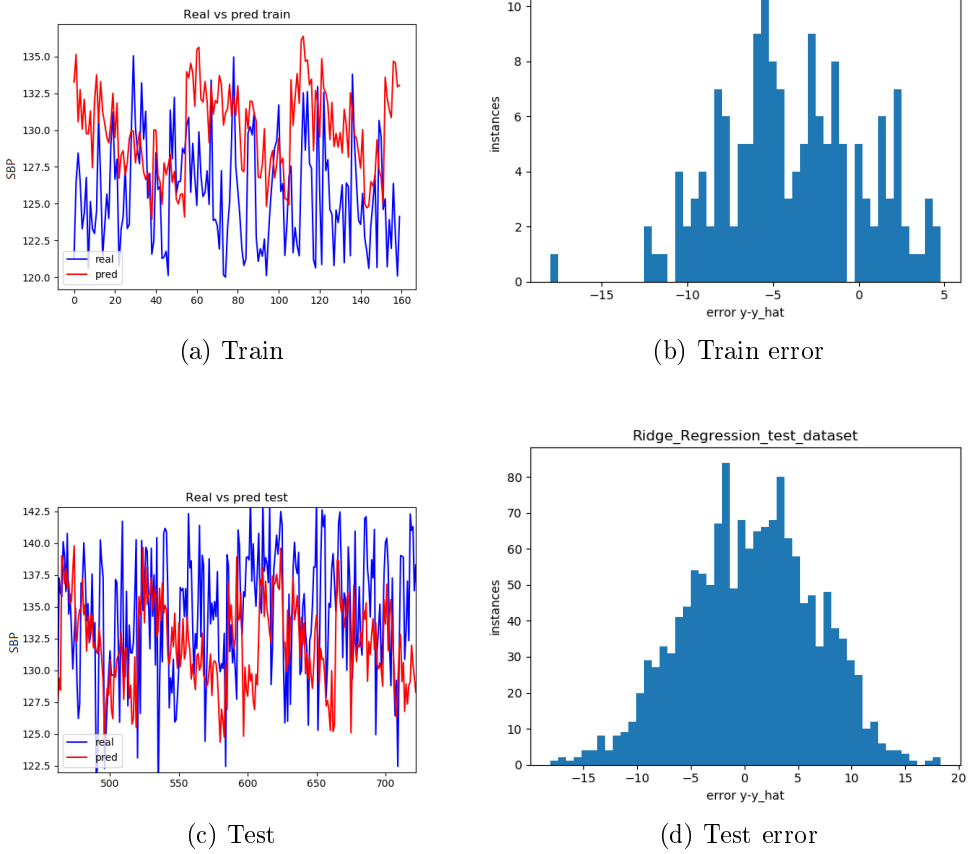
(c) Test

(d) Test error

Figure 4.5: *Whole file*: train and test performance on DBP values

already described in the previous chapter. These results shall be compared with the requirements for the validation of medical devices and with the error of cuff-based blood pressure monitors.

The first thing to notice is that cuff-based blood pressure monitors have a MAE equal to 0 ± 3 mmHg, which is much lower than the one that is presented in Table 4.2. As for the validation of a possible device, the measuring error (only MAE shall be considered) is above the threshold only in the field `MAE test`, which means that it could be improved to fit the requirements. Also the standard deviation is pretty high, especially in the SBP prediction.



(a) Train (b) Train error

(c) Test (d) Test error

Figure 4.6: Whole file: train and test performance on SBP values

SBP			
MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]
7.5 ± 5.3	8.7 ± 5.8	11 ± 6.4	13.1 ± 7
DBP			
MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]
3 ± 1.9	3.6 ± 2.3	4.6 ± 3.4	5.4 ± 3.6

Table 4.2: Results' summary of *whole file* performance for all patients

4.2 Real-time simulation

Figure 4.7 shows the performance of the *real-time simulation* approach in predicting a patient's blood pressure beat-to-beat, described in Section 3.2. As explained in the previous chapter, the training is performed always using the same amount of training data (200 samples) since, in a real case, it would be impossible to find *a priori* a personalized number of samples to use. The amount of data to use was determined as explained in Chapter 3.

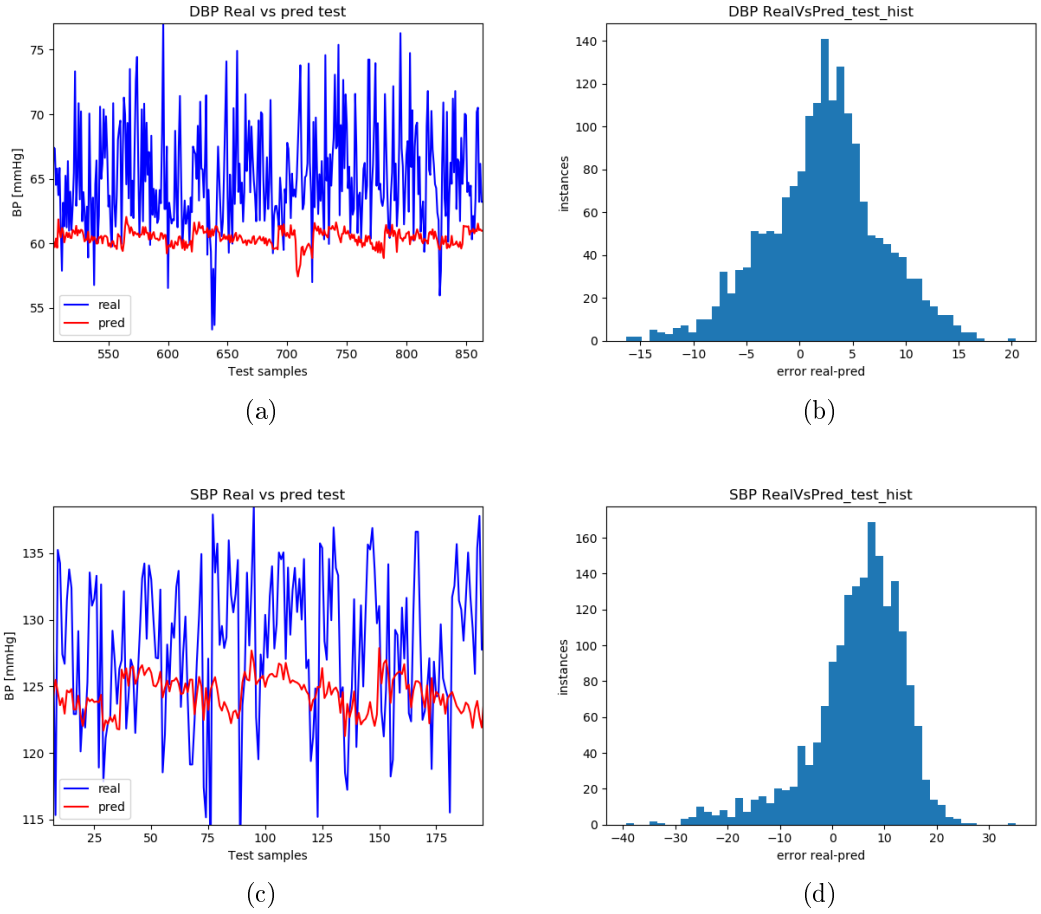


Figure 4.7: *Real-time simulation*: test performance on DBP and SBP values

The mentioned plots are obtained considering `NPEAK` = 200, `LENSEGM` = 10, `NSR` < 13, and autocorrelation coefficient `CORR` > 100; the signal quality was anyway high, since it had `NSR` ≈ 3 and `CORR` ≈ 1000; this means that no segment of the signal was discarded. Notice that the `NSR` is calculated as the sample mean divided by the standard deviation, while the autocorrelation coefficient is found with `np.correlate`; the corresponding threshold values were chosen because, after

an inspection, they were found to be a good quality compromise for segments of length `LENSEGM`.

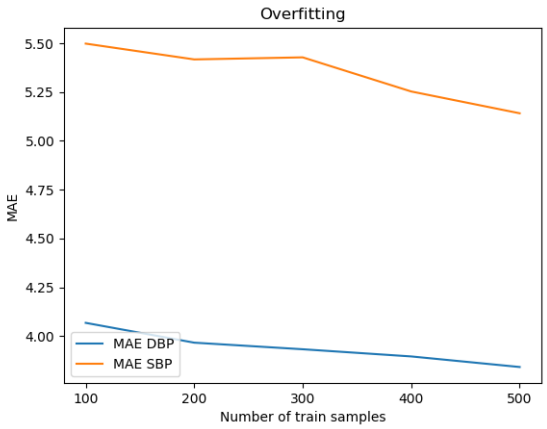


Figure 4.8: Prediction error varying `NPEAK`

Table 4.3 reports the numerical errors to compare with Table 4.1; from the comparison, it appears that the performance in the training phase is still good, but in the testing phase it is worse than with the *whole file* procedure (MAE almost doubles); the reason is unlikely to be an overfitting occurrence since, varying the number of `NPEAK` samples, the error is fairly constant (as shown in Figure 4.8), it decreases very little by increasing the number of train samples.

SBP					
patient	MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
3002229	3.3	4.0	8.0	9.6	40%
DBP					
patient	MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
3002229	2.2	3.0	5.0	6.2	7.8%

Table 4.3: Results of *real-time simulation* for patient 3002229

Table 4.4 shows how the error varies when changing the parameter `LENSEGM`; remember that `LENSEGM` is just a coefficient: the actual number of samples contained in the segment is equal to `LENSEGM` $\times fs$, where fs is the sampling frequency. The error in the training phase looks to be directly proportional to the parameter, while in the testing phase it looks to be indirectly proportional. Nevertheless, the variance is low, which suggests that the reason could be just the different number of samples

used in the respective phase; the same goes for the $err > 10$ mmHg field. The same behavior is shown by the data of the other patients. Values lower than `LENSEGM` = 10 are not reported since the number of peaks present in the segment was not enough for the algorithm to work, thus $10 \times fs$ seems to be the minimum length; in the final analysis, this minimum length was chosen to advantage the patient who, in a real case, has to remain stationary during the acquisition.

SBP					
LENSEGM	MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
10	3.3	4.0	8.0	9.6	40%
20	3.7	4.7	8.7	10.4	40%
30	3.7	4.6	8.9	10.5	42%
40	3.3	4.1	8.3	10.1	36%
50	3.7	4.7	8.7	10.3	41%
60	3.7	4.6	8.8	10.5	41%
DBP					
LENSEGM	MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
10	2.2	3.0	5.0	6.2	7.8%
20	2.7	3.4	4.6	5.7	7.3%
30	2.8	3.5	3.7	4.6	2.2%
40	2.6	3.3	5.1	6.2	8.8%
50	3.0	3.6	3.9	4.8	2.8%
60	3.1	3.7	3.4	4.5	3.3%

Table 4.4: Results of *real-time simulation* for patient 3002229 changing LENSEGM

Table 4.5 shows a summary of the results obtained with this approach on all the data tested. With respect to Table 4.2, an additional column reporting the percentage of cases where the error was larger than 10 mmHg is present. The calculation of such parameter was possible since each test was performed independently for each sample, as explained in Chapter 3. Notice that, even if the parameter is larger in SBP than in DBP (coherently with the other results), it is larger than the 15% allowed by the validation standards in both cases, and has a not negligible standard deviation. The overall performance follows the behavior already observed in patient 3002229. In particular, Table 4.6 reports the difference $MAE_{wf} - MAE_{rts}$ for both training and testing phase, where MAE_{wf} is the MAE obtained with the *whole file* approach, and MAE_{rts} is the MAE obtained with the *real-time simulation* approach; the results in the training phase are better, while in the testing phase they are worse.

SBP				
MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
5.8 ± 3.4	7.4 ± 4.4	13.4 ± 8.4	16 ± 9.5	$47.4 \pm 29.1\%$
DBP				
MAE train [mmHg]	RMSE train [mmHg]	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
2.4 ± 1.4	3.0 ± 1.7	6.7 ± 4.4	8.4 ± 5.7	$21.4 \pm 22.7\%$

Table 4.5: Results' summary of *real-time simulation* for all patients

MAE train SBP [$mmHg$]	MAE train DBP [mmHg]	MAE test SBP [mmHg]	MAE test DBP [mmHg]
1.7 ± 6.1	0.6 ± 2.1	-2.4 ± 7.4	-2.1 ± 5

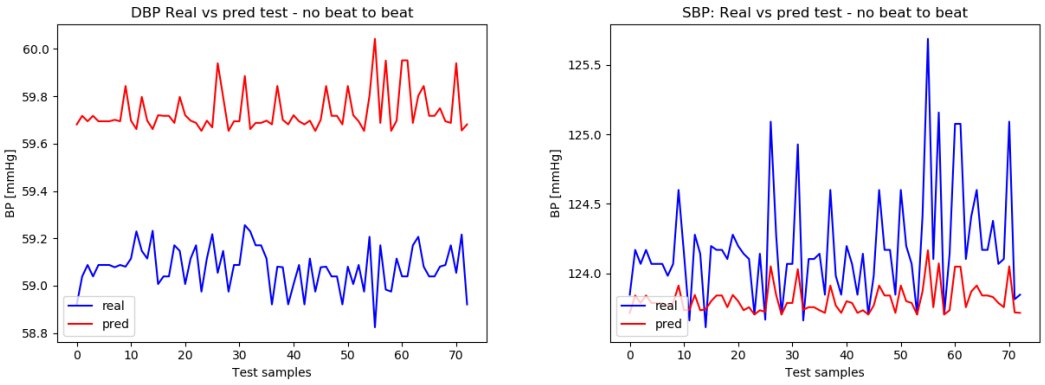
Table 4.6: $MAE_{wf} - MAE_{rts}$

4.3 Comparison with cuffed measurements

This section reports the results obtained with the approach described in Section 3.3: *Comparison with cuffed measurements*. The values of beat-to-beat blood pressure, pulse transit time, and heart rate are averaged over the segment and only one test is performed for each segment (the difference with respect to the *real-time simulation* approach is shown in Figures 3.3 and 3.4).

Figure 4.9 shows the prediction performance on the data of patient 3002229 with `LENSEGM` = 40 (as for the other parameters `NPEAK` = 200, `NSR` = 10, `CORR` = 10000), while Table 4.7 reports the error parameters changing `LENSEGM`. It is immediately noticeable that the error is incredibly low in all the parameters, it drops significantly when `LENSEGM` ≥ 40, and that it is never higher than 10 mmHg. As for the training, the parameters present in Table 4.3 are still valid, for the reasons explained in Chapter 3.

Table 4.8 reports a summary of the results obtained with the *comparison with cuffed measurements* approach for all the patients. Overall, the error is incredibly lower when using this approach with respect to the other two examined, and it is a bit lower when using a larger segment.



(a) (b)

Figure 4.9: *Comparison with cuffed measurements* simulation: test performance on DBP and SBP values

SBP			
LENSEGM	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
10	0.84	0.9	0%
20	0.83	0.88	0%
30	1.32	1.36	0%
40	0.39	0.48	0%
50	0.46	0.53	0%
60	0.23	0.32	0%
DBP			
LENSEGM	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
10	1.29	1.36	0%
20	1.67	1.68	0%
30	1.29	1.33	0%
40	0.66	0.67	0%
50	1	1.02	0%
60	0.55	0.57	0%

Table 4.7: Results of *comparison with cuffed measurements* for patient 3002229 changing LENSEGM

SBP			
LENSEGM	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
10	2.9 ± 2.9	3.2 ± 3.3	0%
40	2.4 ± 2.9	2.7 ± 3.4	0%
DBP			
LENSEGM	MAE test [mmHg]	RMSE test [mmHg]	err > 10 mmHg [-]
10	1.3 ± 1.1	1.4 ± 1.2	0%
40	0.9 ± 0.9	1 ± 1.1	0%

Table 4.8: Results' summary of *Comparison with cuffed measurements* for all patients

4.4 Remarks

Ideally, considering that the algorithm should be as automated as possible, the only parameters to choose should be `NPEAK`, `LENSEGM`, `NSR`, and `CORR`. Nevertheless, considering that data from MIMIC are used, it was necessary to adjust also the possible maximum and minimum value of the R-peak and Systolic Point, due to a very different scale in which the data are reported (in some data-sets the range of R belongs to the interval $[0.02, 1.5]$ and in some other case to the interval [1,3], thus it is not possible to find a single range). Moreover, probably due to their health condition, some patients have very high variability in the length of the cardiac cycle and others do not; for this reason, it was necessary to change also parameter `K` in Equation 3.3. Using a single device to record the data and limiting the target customer to people in good or medium health conditions, this problem should be automatically overcome.

After the analysis, most of the errors seem to be due to the conditions of the patients from whom the data are recorded: drugs, age, diseases influence the relation between blood pressure and pulse transit time shown in Equation 3.1, which is already based on the strong assumption of constant arterial stiffness. Moreover, some of these patients are subject to major pressure changes that make the training not valid. This statement is supported by the fact that Pulse Transit Time measured by the algorithm has extremely high variability.

	BP right [mmHg]	PTT right [ms]	BP wrong [mmHg]	PTT wrong [ms]
DBP	72.2 ± 3.2	640.8 ± 18.6	78.8 ± 8.4	648.8 ± 13
SBP	155.3 ± 2.8	613.38 ± 5.75	167 ± 11.3	649 ± 12.7

Table 4.9: Real values of BP and PTT of patient 3002094 divided according to the fact that PTT led to a right or wrong BP prediction

Table 4.9 shows the real values of Blood Pressure and Pulse Transit Time of patient 3002094 divided into *right* or *wrong* according to the fact that the BP prediction (using the *real-time simulation* approach and the corresponding PTT) had an error (`MAE`) lower than 10 mmHg (*right*) or larger (*wrong*). Considering that the plots and various checks show that BP and PTT were correctly extracted from the signals, it is noticeable that when the prediction is successful, the PTT is in a correct range (lower for SBP and higher for DBP) and that the ranges do not overlap for Systolic and Diastolic blood pressure. On the other hand, when the prediction is not successful, the blood pressure variation does not correspond to a variation in pulse transit time. In addition, notice that the standard deviation of PTT is very large (in some cases it is > 25 ms) which brings further uncertainty in the decision making.

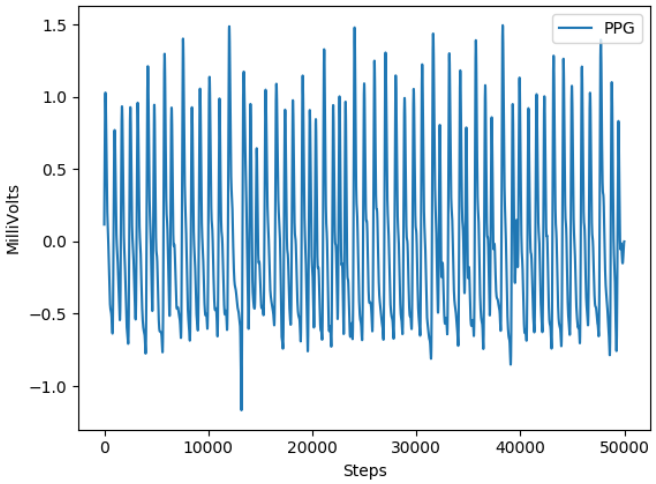


Figure 4.10: Segment of PPG signal

The cause of this high variability in the PPG is likely to be found in the patients' health condition; Figure 4.10 shows a sample of the PPG signal of patient 3002094, while other samples of signals of patient 3002229 are shown in Figure 4.4. Remembering that the PTT is evaluated as the time difference, in milliseconds, between a peak of the ECG and the peak of the PPG corresponding to the same cardiac cycle, it is evident that, to further reduce the error, the peaks should be more regularly distanced than they are (the irregularity seems not to be due to an artifact).

By comparing the results obtained with the three approaches, the best result is reached using the *comparison with cuffed measurements* method, while the worst one with the *real-time simulation* method. Considering the discussion above, this suggests that the problem is too complex to be correctly described by a regression when examining it in detail like in the *beat-to-beat* approach; while in the *non-beat-to-beat* approach, most of the second-order effects internal to the problem and summarized by parameter `K` in Equation 2.5, are damped.

Chapter 5

Conclusion

In this work, we presented an algorithm able to predict a patient's blood pressure given his ECG and PPG. By extracting some critical information (PTT and HR) from such signals, the algorithm, thanks to the mathematical approach used, is able to predict BP quite accurately in few seconds and with a low computational cost. The use of the mentioned signals allows continuous monitoring of blood pressure thanks to the fact that they can be acquired in a non-invasive and non-expensive way; thanks to continuous monitoring of a large portion of the world population (especially people who are generally healthy, remember that hypertension is also called the "silent killer") a detection of hypertension is possible and thus a reduction of incidence of Cardio Vascular Diseases. The algorithm was trained and tested on data downloaded from MIMIC3 an on-line public database containing signals recorded from patients in Intensive Care Unit. The results using a non-beat-to-beat approach are compliant with the international requirements for validated medical devices.

The approach used requires a calibration on data recorded from the patient to predict its blood pressure and this is a disadvantage. Indeed the calibration needs to be performed with the help of qualified personnel (e.g. IT-technicians or a physician). Moreover, since the mathematical approach relies on the assumption of constant stiffness of the arterial walls, there may be a need to calibrate the algorithms more the once. This possibility couldn't be investigated due to the lack of data on a single patient and could represent a huge issue when using the algorithm on data recorded from a specific device. This happens because in the current work the error associated with the acquisition of the signal was neglected since the data downloaded from MIMIC are recorded using very invasive methods but, in a real case, it is expected to be significant (it is not feasible to acquire data with invasive techniques on the target consumers).

Other critical issues arise because PPG is heavily affected by artifacts, especially motion artifacts (as mentioned in [67]), which will impact the performance of the algorithm. Moreover, data recorded from MIMIC are not representative of the

target population, since they are recorded from patients in ICU: they're older than the average target, are in critical health conditions, are administered with drugs of various types. As a consequence, as well as the impossibility to well define the time delay constants in the algorithm, there is the issue that healthy people are not subject to dramatic changes in BP as the ones present in MIMIC but have a more stable trend in blood pressure.

The algorithm was written with the purpose of using it as a start for the device developed in the SINTEC project, of which the prototype is still not present. The future work will be to acquire new data with the prototype and use them as input for the algorithm. There will surely be a need to modify the code, with particular focus on the filtering part, due to the expected high noise. Moreover, there will be the need to state the recurrence of the calibration phase and the validity of the used formula and approach varying the acquisition conditions, i.e. the patients keeping his posture sitting, standing, walking or during physical activity, or changing posture during time. Lastly, considering the discussion in Section 4.4, it would be interesting to consider the potentiality of an algorithm based on classification rather than regression since it would overcome both the question related to the mathematical definition of the problem and the errors due to second order effects. Classes could be linked to certain ranges of blood pressure, for example $class1 \rightarrow SBP \in [100 - 105]$, $class2 \rightarrow SBP \in [105 - 110]$, and so on (the same approach can be used for DBP).

Bibliography

[1] World health Organization https://www.who.int/news-room/fact-sheets/detail/hypertension

[2] Wikipedia - Hypertension https://en.wikipedia.org/wiki/Hypertension

[3] Eurostat https://ec.europa.eu/eurostat/statistics-explained/index.php/Cardiovascular_diseases_statistics

[4] Kario K. Management of Hypertension in the Digital Era: Small Wearable Monitoring Devices for Remote Blood Pressure Monitoring. Hypertension. 2020 Sep;76(3):640-650. doi: 10.1161/HYPERTENSIONAHA.120.14742. Epub 2020 Aug 3. PMID: 32755418; PMCID: PMC7418935.

[5] Allied Market Research https://www.alliedmarketresearch.com/smartwatch-market

[6] Wikipedia - Circulatory system https://en.wikipedia.org/wiki/Circulatory_system

[7] Panerati A.,Pasero G.,Paviglianiti A.,Randazzo V.(2020). Reti neurali ricorrenti per la misura non invasiva della pressione arteriosa,Politecnico di Torino, Ingegneria biomedica.

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

[9] Tiloca A., Demarchi D., Pagana G.(2020). A machine learning approach for non-invasive blood pressure estimation, Politecnico di Torino, Biomedical Engineering.

[10] MyHealth.Alberta.ca https://myhealth.alberta.ca/Health/Pages/conditions.aspx?hwid=tx4097abc

[11] Texas Heart https://www.texasheart.org/heart-health/heart-information-center/topics/heart-anatomy/

[12] Pediatric Heart Specialists `https://pediatricheartspecialists.com/heart-education/14-normal/152-normal-heart-anatomy-and-blood-flow`

[13] Wikipedia - Heart `https://en.wikipedia.org/wiki/Heart#Heart_wall`

[14] Medical Education for Undergraduate MD Students `https://mdmedicine.wordpress.com/2011/04/24/heart-conduction-system/`

[15] Villata S., Pasero G., Paviglianiti A. (2020). Cuffless Blood Pressure measurement, Politecnico di Torino, Biomedical Engineering.

[16] Wikipedia - Blood Pressure `https://en.wikipedia.org/wiki/Blood_pressure`

[17] Chambers, D., Huang, C.,& Matthews, G. (2019). Cardiovascular Physiology. In Basic Physiology for Anaesthetists (pp. 111-188). Cambridge: Cambridge University Press.

[18] Hopkins Medicine `https://www.hopkinsmedicine.org/health/conditions-and-diseases/`

[19] Pulse Pressure Calculation Explained `https://www.healthline.com/health/pulse-pressure`

[20] Wikipedia - Mean Arterial Pressure `https://en.wikipedia.org/wiki/Mean_arterial_pressure`

[21] Delong C, Sharma S. Physiology, Peripheral Vascular Resistance. [Updated 2020 Jul 14]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK538308/

[22] Wikipedia - Vascular resistance `https://en.wikipedia.org/wiki/Vascular_resistance`

[23] Omron - Blood Pressure monitors `https://omronhealthcare.com/service-and-support/faq/blood-pressure-monitors/`

[24] BioNinja `https://ib.bioninja.com.au/options/option-d-human-physiology/d4-the-heart/heart-measurements.html`

[25] Arterial Line `https://acuclinic.com.au/pocit/ArtLine.htm`

[26] Ogedegbe G, Pickering T. Principles and techniques of blood pressure measurement. Cardiol Clin. 2010 Nov;28(4):571-86. doi: 10.1016/j.ccl.2010.07.006. PMID: 20937442; PMCID: PMC3639494.

[27] Wikipedia - Blood pressure measurement https://en.wikipedia.org/wiki/Blood_pressure_measurement#Ambulatory_and_home_monitoring

[28] Nguyen Y, Bora V. Arterial Pressure Monitoring. [Updated 2020 Mar 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK556127/

[29] Wikipedia - Ambulatory blood pressure https://en.wikipedia.org/wiki/Ambulatory_blood_pressure

[30] Stergiou, GS, Alpert, B, Mieke, S, Asmar, R, Atkins, N, Eckert, S, Frick, G, Friedman, B, Grassl, T, Ichikawa, T, Ioannidis, JP, Lacy, P, McManus, R, Murray, A, Myers, M, Palatini, P, Parati, G, Quinn, D, Sarkis, J, Shennan, A, Usuda, T, Wang, J, Wu, CO, and O'Brien, E. "A Universal Standard for the Validation of Blood Pressure Measuring Devices." (2018). Web.

[31] Nelson D, Kennedy B, Regnerus C, Schweinle A. Accuracy of automated blood pressure monitors. J Dent Hyg. 2008 Summer;82(4):35. Epub 2008 Jul 1. PMID: 18755068.

[32] Rotch AL, Dean JO, Kendrach MG, Wright SG, Woolley TW. Blood pressure monitoring with home monitors versus mercury sphygmomanometer. Ann Pharmacother. 2001 Jul-Aug;35(7-8):817-22. doi: 10.1345/aph.10270. PMID: 11485126.

[33] UC Berkley school of public health - Blood Pressure Monitors: Wrist vs. Arm https://www.healthandwellnessalerts.berkeley.edu/topics/hypertension-stroke/blood-pressure-monitors-wrist-vs-arm/

[34] Beurer - Blood Pressure monitors https://www.beurer.com/web/gb/products/medical/blood-presure/upper-arm-blood-pressure-monitors/bm-95.php

[35] Harvard Medical School - Some home blood pressure monitors aren't accurate (2014) https://www.health.harvard.edu/blog/home-blood-pressure-monitors-arent-accurate-201410297494

[36] News Medical Lifescience (2019) https://www.news-medical.net/health/Are-Blood-Pressure-Monitors-Accurate.aspx

[37] Goodwin J, Bilous M, Winship S, Finn P, Jones SC. Validation of the Oscar 2 oscillometric 24-h ambulatory blood pressure monitor according to the British Hypertension Society protocol. Blood Press Monit. 2007 Apr;12(2):113-7. doi: 10.1097/MBP.0b013e3280acab1b. PMID: 17353655.

[38] Hodgkinson, James A, Lee, Mei-Man, Milner, Siobhan, Bradburn, Peter, Stevens, Richard, Hobbs, Fd Richard, Koshiaris, Constantinos, Grant, Sabrina, Mant, Jonathan, and McManus, Richard J.(2020). Accuracy of Blood-pressure Monitors Owned by Patients with Hypertension (ACCU-RATE Study): A Cross-sectional, Observational Study in Central England. Web.

[39] Medel - Blood Pressure monitors https://www.medelinternational.com/en/blood-pressure-monitoring/

[40] Withings - Blood Pressure monitors https://www.withings.com/it/en/blood-pressure-monitors

[41] Wikipedia - Hypertension https://en.wikipedia.org/wiki/Hypertension#Primary_hypertension

[42] Wikipedia - Elettrocardiogramma https://it.wikipedia.org/wiki/Elettrocardiogramma

[43] Wikipedia - Einthoven's triangle https://en.wikipedia.org/wiki/Einthoven%27s_triangle

[44] Khunti, K. (2014). Accurate interpretation of the 12-lead ECG electrode placement: A systematic review. Health Education Journal, 73(5), 610–623. https://doi.org/10.1177/0017896912472328

[45] Solomon, David & Ferenchick, Gary. (2005). Sources of Measurement Error in an ECG Examination: Implications for Performance-Based Assessments. Advances in Health Sciences Education. 9. 283-290. 10.1007/s10459-005-4844-1.

[46] Nurse24 https://www.nurse24.it/studenti/indagini-diagnostiche/elettrocardiogramma-interpretazione-infermiere.html

[47] Wikipedia - Photoplethysmogram https://en.wikipedia.org/wiki/Photoplethysmogram

[48] Vera Lucia Da Silveira Nantes Button(2015).Principles of Measurement and Transduction of Biomedical Variables,Academic Press, ISBN 9780128007747

[49] Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. (2018). A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron. 4(4):195-202. doi: 10.15406/ijbsbe.2018.04.00125. Epub 2018 Aug 6. PMID: 30906922; PMCID: PMC6426305.

[50] V. Ferrer-Mileo, F. Guede-Fernandez, M. Fernández-Chimeno, J. Ramos-Castro and M. A. García-González (2015). Accuracy of heart rate variability estimation by photoplethysmography using an smartphone: Processing optimization and fiducial point selection. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 5700-5703, doi: 10.1109/EMBC.2015.7319686.

[51] Kavsaoğlu, Ahmet & Polat, Kemal & Bozkurt, Mehmet. (2016). An innovative peak detection algorithm for photoplethysmography signals: An adaptive segmentation method. TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES. 24. 1782-1796. 10.3906/elk-1310-177.

[52] Using Reflectometry for a PPG Waveform `https://www.maximintegrated.com/en/design/technical-documents/app-notes/6/6547.html`

[53] Elgendi M. (2012). On the analysis of fingertip photoplethysmogram signals. Current cardiology reviews, 8(1), 14–25. https://doi.org/10.2174/157340312801215782

[54] Yousef, Q., Reaz, M. B. I.,& Ali, M. A. M. (2012). The Analysis of PPG Morphology: Investigating the Effects of Aging on Arterial Compliance, Measurement Science Review, 12(6), 266-271. doi: https://doi.org/10.2478/v10048-012-0036-3

[55] Bolanos M, Nazeran H, Haltiwanger E.(2006). Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals. Conf Proc IEEE Eng Med Biol Soc.;2006:4289-94. doi: 10.1109/IEMBS.2006.260607. PMID: 17946618.

[56] Julia Pietilä, Saeed Mehrang, Johanna Tolonen, Elina Helander, Holly Jimison, Misha Pavel, Ilkka Korhonen (2018). Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities. Springler Singapore,EMBEC & NBC 2017, isbn 978-981-10-5122-7.

[57] Apple `https://www.apple.com/it/watch/`

[58] Bent, B., Goldstein, B.A., Kibbe, W.A. et al. (2020).Investigating sources of inaccuracy in wearable optical heart rate sensors. npj Digit. Med. 3, 18. https://doi.org/10.1038/s41746-020-0226-6

[59] Mccarthy, Brian & O'Flynn, Brendan & Mathewson, Alan. (2011). An Investigation of Pulse Transit Time as a Non-Invasive Blood Pressure Measurement Method. Journal of Physics: Conference Series. 307. 012060. 10.1088/1742-6596/307/1/012060.

[60] Younhee Choi, Qiao Zhang, Seokbum Ko, Noninvasive cuffless blood pressure estimation using pulse transit time and Hilbert–Huang transform, Computers & Electrical Engineering, 2013, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2012.09.005. (https://www.sciencedirect.com/science/article/pii/S0045790612001711)

[61] PhysioNet `https://archive.physionet.org/`

[62] Kevin P. Murphy. 2012. Machine Learning: A Probabilistic Perspective. The MIT Press.

[63] Fortmann-Roe, S. (2012). Understanding the Bias-Variance Tradeoff. `http://scott.fortmann-roe.com/docs/BiasVariance.html`

[64] Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. Introduction to Linear Regression Analysis. Fourth edition. Hoboken, N.J: Wiley-Interscience, 2006. Print.

[65] 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 Sep;38(5):569-74. doi: 10.1007/BF02345755. PMID: 11094816.

[66] C. El-Hajj, P.A. Kyriacou, A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure, Biomedical Signal Processing and Control, Volume 58, 2020, 101870, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2020.101870.

[67] Sharma, M.; Barbosa, K.; Ho, V.; Griggs, D.; Ghirmai, T.; Krishnan, S.K.; Hsiai, T.K.; Chiao, J.-C.; Cao, H. Cuff-Less and Continuous Blood Pressure Monitoring: A Methodological Review. Technologies 2017, 5, 21.

[68] Kachuee M, Kiani MM, Mohammadzade H, Shabany M. Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring. IEEE Trans Biomed Eng. 2017 Apr;64(4):859-869. doi: 10.1109/TBME.2016.2580904. Epub 2016 Jun 14. PMID: 27323356.

[69] Ghosh, Shrimanti & Banerjee, Ankur & Ray, Nilanjan & Wood, Peter & Boulanger, Pierre & Raj, Padwal. (2016). Continuous blood pressure prediction from pulse transit time using ECG and PPG signals. 188-191. 10.1109/HIC.2016.7797728.

[70] Liang Y, Chen Z, Ward R, Elgendi M. Hypertension Assessment via ECG and PPG Signals: An Evaluation Using MIMIC Database. Diagnostics (Basel). 2018 Sep 10;8(3):65. doi: 10.3390/diagnostics8030065. PMID: 30201887; PMCID: PMC6163274.

[71] Khalid, Syed Ghufran & Zhang, Jufen & Chen, Fei & Zheng, Dingchang. (2018). Blood Pressure Estimation Using Photoplethysmography Only: Comparison between Different Machine Learning Approaches. Journal of Healthcare Engineering. 2018. 10.1155/2018/1548647.

[72] Lazazzera R, Belhaj Y, Carrault G. A New Wearable Device for Blood Pressure Estimation Using Photoplethysmogram. Sensors (Basel). 2019 Jun 4;19(11):2557. doi: 10.3390/s19112557. PMID: 31167514; PMCID: PMC6603632.

[73] Ding, Xiaorong & Zhang, Yuan-Ting. (2019). Pulse transit time technique for cuffless unobtrusive blood pressure measurement: from theory to algorithm. Biomedical Engineering Letters. 9. 10.1007/s13534-019-00096-x.

[74] Chen, Shuo & Ji, Zhong & Wu, Haiyan & Xu, Yingchao. (2019). A Non-Invasive Continuous Blood Pressure Estimation Approach Based on Machine Learning. Sensors (Basel, Switzerland). 19. 10.3390/s19112585.

[75] Zhang B, Ren H, Huang G, Cheng Y, Hu C. Predicting blood pressure from physiological index data using the SVR algorithm. BMC Bioinformatics. 2019 Feb 28;20(1):109. doi: 10.1186/s12859-019-2667-y. PMID: 30819090; PMCID: PMC6396542.

[76] SINTEC project `https://www.sintec-project.eu/`