Politecnico di Torino

DEPARTMENT OF MECHANICAL AND AEROSPACE ENGINEERING

Master's degree in Biomedical Engineering

Systolic Blood Pressure Estimation from PPG Signal Using Artificial Neural Networks

Supervisor
Prof. Gabriella OLMO

Co-supervisor
Eng. Alessandro GUMIERO

Candidate
Benedetta Caterina CASADEI

Academic year
2020/2021

List of Acronyms

3G Third Generation

4G Fourth Generation

5G Fifth Generation

AAMI Association for the Advancement of Medical Instrumentation

AAMI Association for the Advancement of Medical Instrumentation

ABP Arterial Blood Pressure

AC Alternating Current

AF Activation Function

AI Artificial Intelligence

ANN Artificial Neural Network

APG Accelerated Plethysmography

AV Atrioventricular

B3G Beyond Third Generation

BCU Body Control Unit

BHS British Hypertension Society

BLE Bluetooth Low Energy

BP Blood Pressure

bpm beats per minute

CM Confusion Matrix

CP Cardiac Period

CPU Central Processing Unit

CVD Cardiovascular Disease

CVDs Cardiovascular Diseases

DC Direct Current

DP Diastolic Blood Pressure

DT Diastolic Time

DW Diastolic Width

ECG Electrocardiogram

FN False Negative

FP False Positive

GD Gradient Descent

Hb Deoxygenated Haemoglobin

HbO_2 Oxygenated Haemoglobin

HR Heart Rate

HRV Heart Rate Variability

IBI Inter Beat Interval

IBP Invasive Blood Pressure

ICT Information and Communication Technology

ICU Intensive Care Unit

IoT Internet of Things

IR Infrared

LED Light Emitting Diodes

LSB Least Significant Bit

MAC Multiply-Accumulate Operation

MAE Mean Absolute Error

MAP Mean Arterial Pressure

MCU Microcontroller Unit

MIMIC Multi-parameter Intelligent Monitoring Intensive Care

ML Machine Learning

MLP Multilayer Perceptron

NG Next Generation

NIBP Non-Invasive Blood Pressure

NIR Near-Infrared

NN Neural Network

NTC Negative Temperature Coefficient

PCB Printed Circuit Board

PD Photo-Detector

PPG Photoplethysmography

PTT Pulse Transit Time

PWB Pulse Wave Begin

PWD Pulse Wave Duration

PWDP Pulse Wave Diastolic peak

PWE Pulse Wave End

PWSP Pulse Wave Systolic peak

PWV Pulse Wave Velocity

RAM Random Access Memory

ReLu Rectified Linear Unit

RMSE Root Mean Squared Error

ROM Read Only Memory

SA Sinoatrial

SD Standard Deviation

SGD Stochastic Gradient Descent

SMOTE Synthetic Minority Oversampling Technique

SP Systolic Blood Pressure

SpO2 Peripheral Oxygen Saturation

SUT Systolic Upstroke Time

SW Systolic Width

TLM Telemedicine

TP True Positive

WBAN Wireless Body Area Network

WBAS Wearable Body Area Sensors

WFDB WaveForm DataBase

WHO World Health Organization

WIoT Wearable IoT

WLAN Wireless Local Area Network

WMAN Wireless Metropolitan Area Network

WPAN Wireless Personal Area Network

Contents

List of Figures

List of Tables

Abstract

High blood pressure is one of the most important precursors for cardiovascular diseases (CVDs), and the most common cause of death in 2015, as reported by the World Health Organization (WHO). For this reason, continuous and non-invasive blood pressure remote monitoring would be a great revolution, especially for hypertensive patients.

The existing cuff-based methods cannot provide continuous blood pressure readings; actual cuff-less methods are based on Pulse Transit Time[1] (PTT) or Pulse Wave Velocity[2] (PWV). They require both the electrocardiogram (ECG) and the photoplethysmogram (PPG) signal, positioning multiple sensors in different body sites.

Our work has been inspired by the newest approaches based on the PPG signal only, which has been used to continuously estimate systolic blood pressure[3] (SP), using artificial neural networks (ANN), as PPG sensor would allow the realization of more compact and wearable devices.

Our first database has been derived from the PhysioNet resource; we have extracted PPG and arterial blood pressure (ABP) signals, collected at a sampling frequency of 125 Hz, in a hospital environment. It consists of 249,672 PPG periods and the relative SP values. The second database has been collected at STMicroelectronics S.r.l, in Agrate Brianza, using the MORFEA3 wearable device and a digital cuff-based sphygmomanometer, as reference.

The pre-processing phase, in order to remove noise and motion artifacts and to segment the signal into periods, has been carried out on Matlab R2019b.
Different solutions have been implemented to choose the input features that best represent the period morphology. The first database has been used to train the multilayer feed-forward neural network with a back-propagation model; whereas the second one has been used to test it.

The results obtained from the PhysioNet dataset show a mean absolute error of 2.49 mmHg and a standard deviation of ± 2.84 mmHg over the test set. The results for the everyday-life dataset show a mean absolute error of 3.85 mmHg and a standard deviation of ± 4.29 mmHg.

[1]The time that the pressure wave takes to travel between two arterial sites.
[2]The blood flow velocity in the circulatory system.
[3]It is a more potent predictor for coronary heart disease, heart failure and mortality, at age 50.

Chapter 1

Introduction

Project aim

The purpose of this project is to implement a machine learning algorithm, based on ANNs, able to estimate SP starting from the PPG signal.
The implemented ANN model investigates the non-linear relation between blood pressure and PPG, in order to accurately estimate SP value.

The first part of this work has been inspired by the previous thesis, whose aim was to implement a multiclass classification model; we have divided the SP range considered[1], following the division adopted by the American Heart Association for adults (Table 1.1).

<90
90 - 119
120 - 139
140 - 159
160 - 180

Table 1.1: SP range division

In the second part we have implemented a regression model, whose aim was to predict the SP numerical value.

The ultimate goal would be to embed this model within a *STM32* MCU of a STMicroelectronics wearable device. The final wearable device should allow real-time, beat-to-beat, blood pressure prediction.
Continuous blood pressure monitoring aims to prevent and diagnose heart disease, due to high blood pressure, such as stroke, heart attack, coronary artery disease.

[1]It begins with 80 mmHg and ends with 180 mmHg.

1.1 Telemedicine

The number of people in need of health care, such as post-hospital patient monitoring, home-care monitoring, patient education, and continuing medical education, has grown considerably in recent decades, from the children to the elderly. The cost and the overload on the healthcare service providers have increased significantly [35]. For this reason, a reduction in healthcare costs appears necessary, and the wide adoption of telemedicine (TLM) helps to meet this requirement.
By definition, TLM utilizes information and communication technologies (ICT) to securely transfer medical information and monitoring data, to a remote site, for prevention, diagnosis, therapy and education. It aims to provide expert-based medical care to any place where health care is needed [59].
TLM can be divided into two basic modes of operation: real-time and store-and-forward mode. In the first interactive mode all relevant information (data, graphics, images, etc.) are sent to a remote terminal immediately after acquisition. In this case the capability to transmit a large amount of information in a short time is mandatory. In the second case the response does not have to be immediate, and, in most cases, a few hours or days will elapse before receiving the specialist's report. This is less demanding in terms of speed and bandwidth resources.
TLM does not replace the traditional healthcare service in the doctor-patient relationship, although it can improve its effectiveness and efficiency.

1.1.1 Main components of TLM system

Thanks to the telecommunication technologies evolution, and to the robustness and the reliability that these systems provide, the use of TLM has spread rapidly and is now becoming integrated into the ongoing operation of hospitals, as well as patient's homes.
Telecommunication infrastructures play a fundamental role in TLM services, as they ensure the communication between the user and the medical center [33].
A possible communication protocol can be summarized as follows:

- The user sends health information to the medical center;
- The medical center transmits the service results;
- The service center manages the health information generated by the user, in order to reach the medical center of the health care service;
- The service center manages the results of the service to be transmitted by the medical center to the User;

- The service center maintains biomedical devices in remote sites like the patient's home for acquiring, processing signals and communication between patients and doctors.

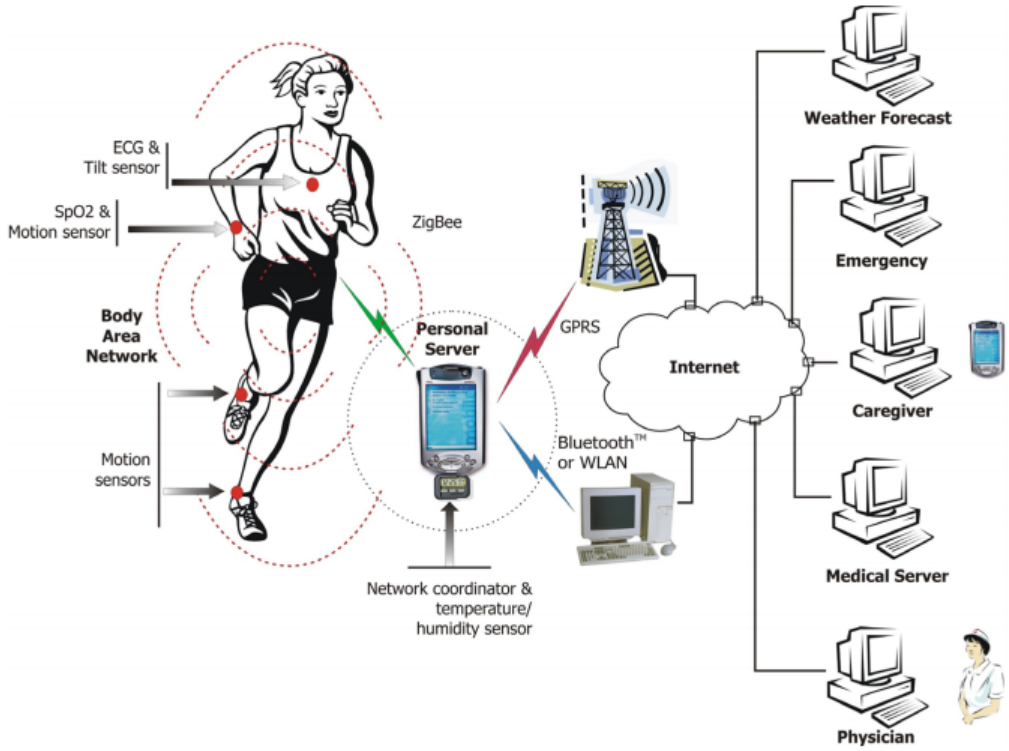

Figure 1.1: A Wireless Body Area Network of intelligent motion sensors for computer assisted physical rehabilitation [59]

The first TLM services were based on cable communication technologies, such as plain old telephone network (POTN) and integrated services digital network (ISDN). The continuous development of TLM relies on the continuous innovation of network communication technology, allowing communication in a short (WPAN), medium (WLAN) and long range (WMAN).
Usually, wireless TLM systems consist of:

- wearable medical devices;
- wireless communications networks.

In recent decades we have witnessed the emergence of various low-cost wireless broadband solutions.
The wireless communication systems involved in healthcare domain are [80]:

- *Wireless personal area network* or WPAN: it allows a last metre connection by the introduction of IrDA, RFID, Bluetooth, ZigBee and ultra-wide band.

- *Wireless local area network* or WLAN: it is the first and the oldest wireless technology used in medical applications; it provides moderate to high-speed data communication. The standards of WLAN was first introduced in 1997, namely IEEE 802.11. Many extensions of 802.11 were released, (IEEE 802.11a/ b/g); it allows users to access a data network at high speeds up to 54 Mbps working on 2.4 GHz band as long as users are located within a relatively short range from the access point (typically 30 m indoors and 100 m outdoors) [58]. By using this transmission media, patient's data are easily transferred around the hospital.

- *Wireless metropolitan area networks* or WMAN: it has enabled to develop the IEEE 802.16 and IEEE 802.20 standards, which provide strong-security wireless data transmission over long distance, with better quality-of-service (QoS) support than Wi-Fi. The integrated network of IEEE 802.11/WLAN and IEEE 802.16/WiMAX can bring a synergetic improvement to the telemedicine services on coverage, data rates, and QoS provisioning to mobile users. This can improve the availability of the health services. However, to integrate the WiMAX network in the WLAN one, many challenging problems have to be addressed, as explained in [80].

- *Wireless body area network* or WBAN: it relies on the connection of independent sensors situated on the body or implanted into a patient's body, through a wireless communication channel. It offers many promising applications in remote health monitoring thanks to the advantage that the patient does not have to stay in bed, improving the quality of life for the patient and reducing hospital costs.

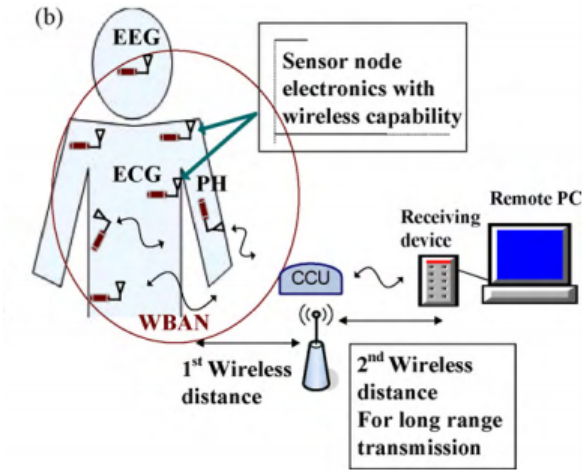

Figure 1.2: Medical sensor network targeted by WBAN [79]

- *Third-generation* or 3G: it enhances the delivery of healthcare, ensuring the possibility to transfer a great number of voice and non-voice data, at higher rate.

- *Beyond-3G* or B3G and the *Fourth-generation* or 4G: they provide the integration of existing wireless technologies, and other newly developed wireless technologies, into a seamless platform. B3G and 4G are based on mobile triple-play (voice, video, data) with guaranteed QoS to support real-time multimedia applications such as citizen-centric healthcare systems.

- *Next-generation* or NG: *Fifth-generation* or 5G offers data transfer rate up to 10 Gbps; super-low latency (delay in the data transmission-response system); connectivity and capacity up to 100%, fundamental in higher density areas and in areas where macro site signals can be blocked by buildings, walls and terrain; high bandwidth and durability per unit area.

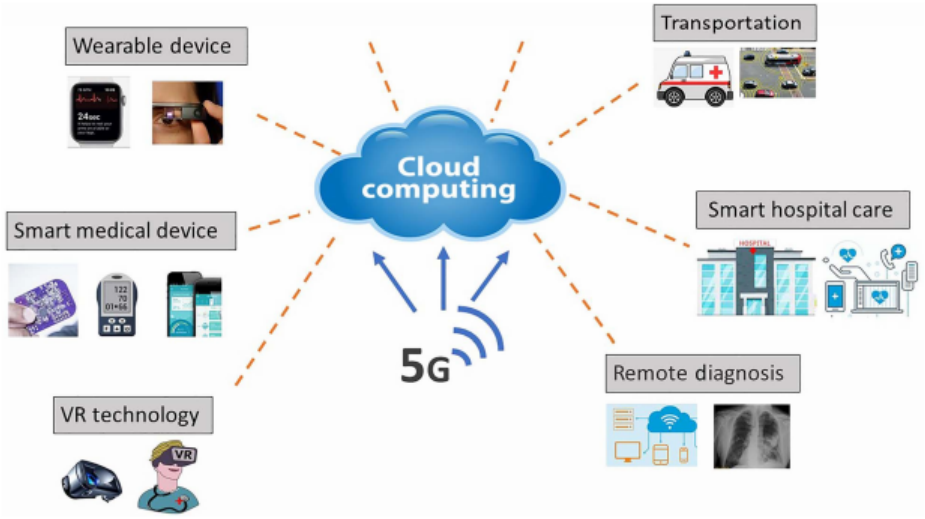

Figure 1.3: Applications of 5G technology in healthcare [47]

Nowadays, the telecommunication aim is to enlarge the transmission bandwidth, so that more and heavier files can be sent in the time unit.

1.1.2 Types of TLM

Technological development has allowed TLM to spread in the following areas:

- *Teleconsultation*: it is the most widely used among telemedical procedures, using about 35% of TLM networks [30]. It consists in the communication at distance, through ICT, between two or more carers, to obtain a second opinion about a patient's case, and between one or more carers and a patient in order to perform a diagnose or treat a patient in a remote location. *Teleconsultation* provides easy and affordable access to medical services for everyone, in particular for people living in rural and remote areas, and for individuals with physical disabilities, who have greater difficulty organizing hospital visits. Therefore, it helps to save time and reduce transport costs for patients [32].

- *Tele-education*: the online information available over the Internet, academic study or public education through the web, offer low cost and easy access to clinical educational material.

- *Telesurgery*: it can be practiced in two ways. *Telementoring* consists of giving, through video and audio connection, the assistance by an experienced surgeon, to another surgeon, engaged in a surgical practice, at a remote site. In *Telepresence Surgery*, the surgeon guides robotic arms to perform remote surgical procedures [30].

- *Telemonitoring*: it provides the acquisition and the management of the patients' vital parameters from a remote location, allowing them to conduct a normal life. Most of telemonitoring systems are composed as follows:

  - *data acquisition system*, consisting of electronic system with an embedded sensors;
  - *data transmission system* based on telecommunication technologies, from patient to physician;
  - *data integration*, to recreate the patient's history and correctly assess the patient's status;
  - *data examination*, to take action in case of an abnormal situation;
  - *data storage*, which can be either on the cloud or in a local machine.

Remote patient monitoring

Nowadays population is increasing rapidly, as well as diseases and healthcare costs.
Our work focuses on remote patient monitoring, more specifically on continuous SP remote monitoring. With wearable devices becomes vital to create a Wearable IoT (WIoT) ecosystem in which body-worn sensors perfectly synchronize data to the cloud services through the IoT infrastructure.

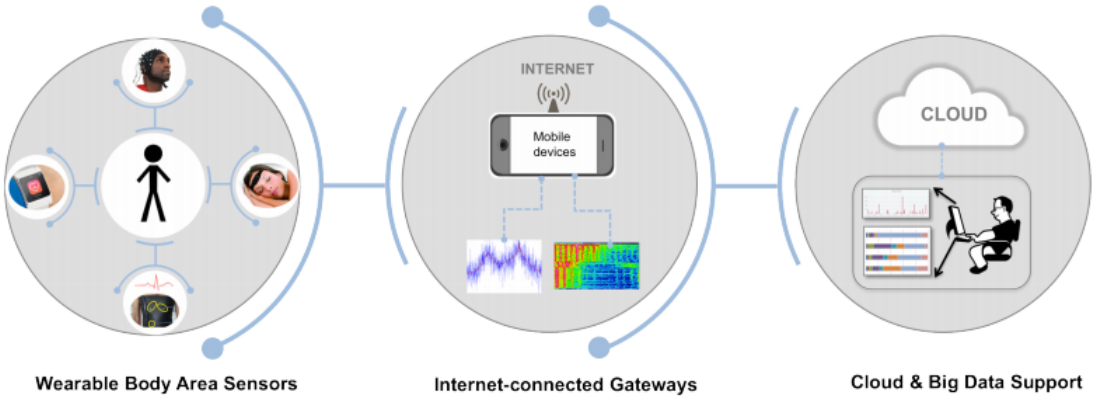

Figure 1.4: Architectural elements of WIoT [40]

Figure 1.4 shows all the components of a WIoT system.

- *Wearable body area sensor* or WBAS: it collects physiological and motion data without restricting the patient's movement and prepare the data for remote transmission for analysis and decision support.
- *Internet-connected gateways*: sensors usually have a limited computing and memory resources; for these reasons the information are transmitted through the body control unit (BCU).
- *Cloud & big data support*: it facilitates the management of wearable data.

Sensor data can be electrical (for example in electrocardiography applications), mechanical (such as accelerometers) or optical (like in photoplethysmography applications).
In our work an optical sensor has been used to acquire the photoplethysmographic signal.

1.2 Conceptual framework

1.2.1 Cardiovascular system

The cardiovascular system is composed of the heart and a closed system of blood vessels. Through blood transport it delivers oxygen, nutrients and hormones to the whole human body, it regulates body temperature, pH, water content of cells and it removes the waste products.
Blood vessels can be divided into:

- **Arteries**: they carry oxygen and nutrients away from the heart, to the tissues of the body.
- **Veins**: they take oxygen-poor blood back to the heart.
- **Capillaries**: they are responsible of the transfer of oxygen and other nutrients from the bloodstream to the tissues.

There are two types of circulation: pulmonary circulation and systemic circulation. *Pulmonary circulation* transports deoxygenated blood between the heart and the lungs, in order to absorb oxygen and release carbon dioxide.
Systemic circulation transports oxygenated blood away from the heart to the rest of the body; and returns deoxygenated blood back to the right atrium of the heart [20].

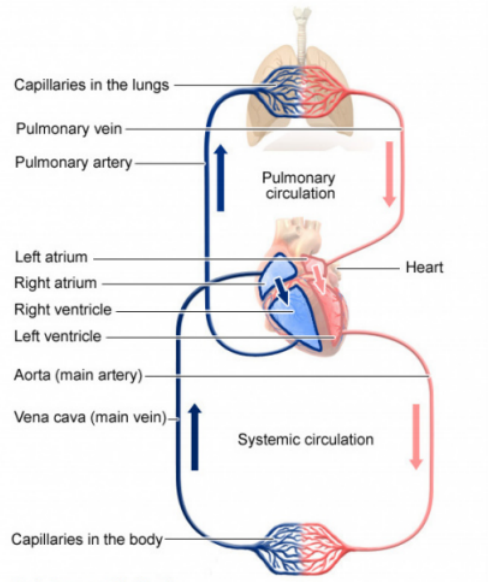

Figure 1.5: Pulmonary and systemic circulation [20]

The heart is the key organ of the cardiovascular system, as it enables the continuous blood flow. The wall of the heart is composed of three layers: the *epicardium* (the outer layer), the *myocardium* (the middle layer) and the *endocardium* (the inner layer). The *myocardium* is the muscle of the heart, accountable for its pumping action. The heart cavity is made up of two portions, left and right, separated by a inner wall called septum. On the right side, atrium and ventricle are divided by the tricuspid valve, on the left side are divided by the mitral valve.

1.2.2 The heart physiology and the cardiac cycle

The heart accomplishes its role in the cardiovascular system thanks to the *conduction system*: it initiates and coordinates the synchronized contraction of the atria and the ventricles [29].
The first electrical impulse comes from the so-called *heart's natural pacemaker*, the sinoatrial (SA) node, placed in the right atrium, able to generate action potential at a specific rhythm [11]. Then, the electric current travels through the right atrium and through the atrioventricular (AV) node, located in the interatrial septum, causing the contraction of atria that pushes the blood into the ventricles. Hence, it reaches the Bundle of His, divided into two bundle branches, the left and right bundles. These two bundles give rise to the Purkinje fibers which, spreading the wave impulses along the ventricles, allow an organized contraction of the left and right ventricles.

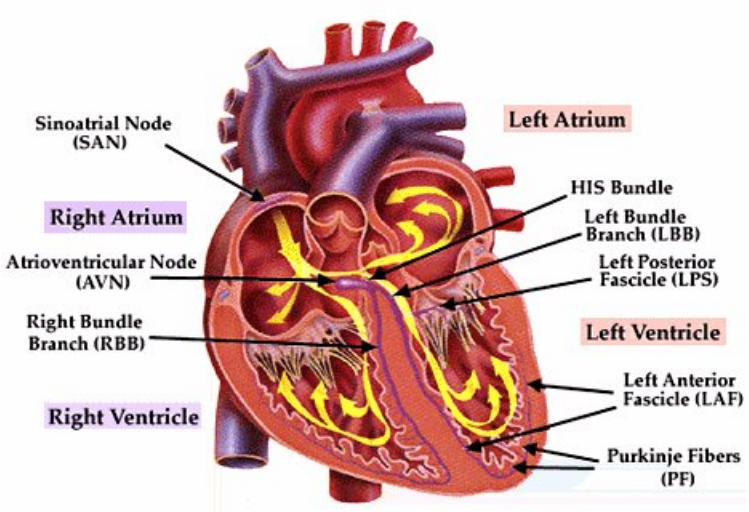

Figure 1.6: Cardiac conduction system

The cardiac action potential has five phases [60]:

- Phase 4: The resting phase;
- Phase 0: Depolarization;
- Phase 1: Early repolarization;
- Phase 2: The plateau phase;
- Phase 3: Repolarization. SA node produces 60-100 action potentials per minute.

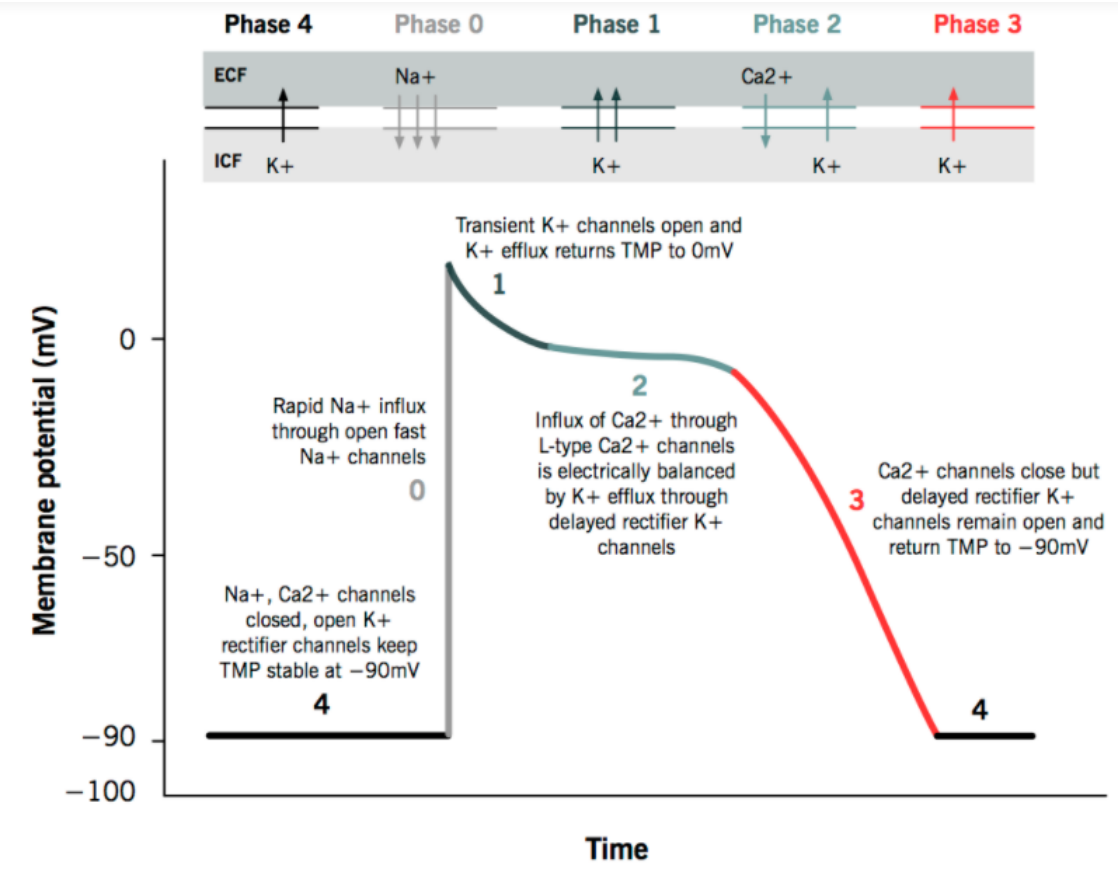

Figure 1.7: Action potential of cardiac muscles

The cardiac cycle is defined as the time interval between the ending of one heartbeat to the beginning of the next one.
It consists of two periods:

- Diastole: the heart muscle relaxes and the heart chambers fill with blood.
- Systole: the heart muscle contracts, generates a pressure higher than the arterial tree, and pushes the blood from the heart to the large blood vessels.

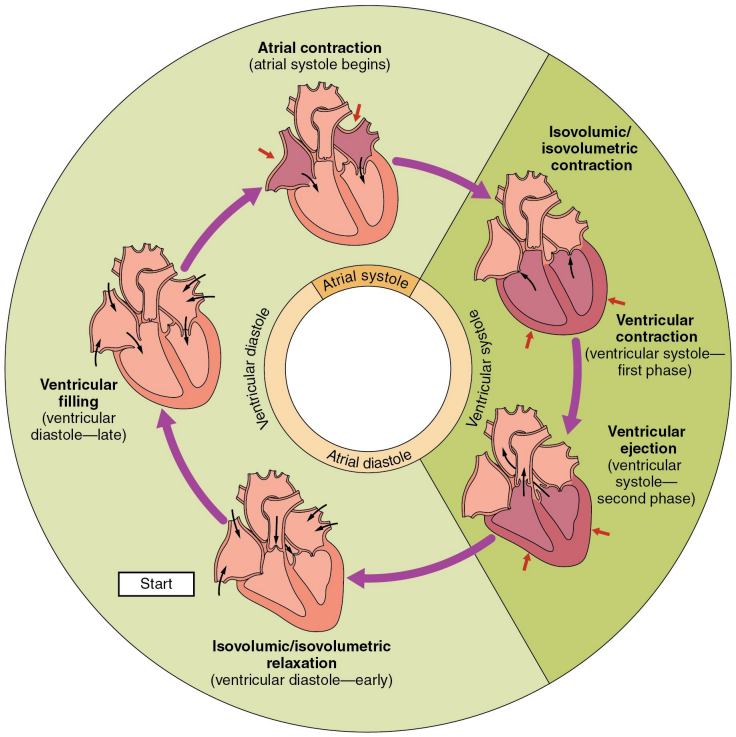

Figure 1.8: Cycle diagram [4]

1.2.3 Biological signals

Biological signals, or biosignals, represent all the measurements of the physiological activity in the human body. They provide useful information about the patients' medical condition.
Biosignals can be, for instance [65]:

- Electrical: generated by nerves and muscles, such as the ECG.
- Optical: generated by the amount of light absorbed or scattered by the biological tissue, for example the PPG.

Electrocardiogram

Cardiovascular parameters are extremely useful to investigate patient condition.

The ECG is one of the most common heart tests used in the assessment and diagnosis of CVDs [67]. The ECG intervals can provide information about the rate, the rhythm and the electrical activity of the heart.

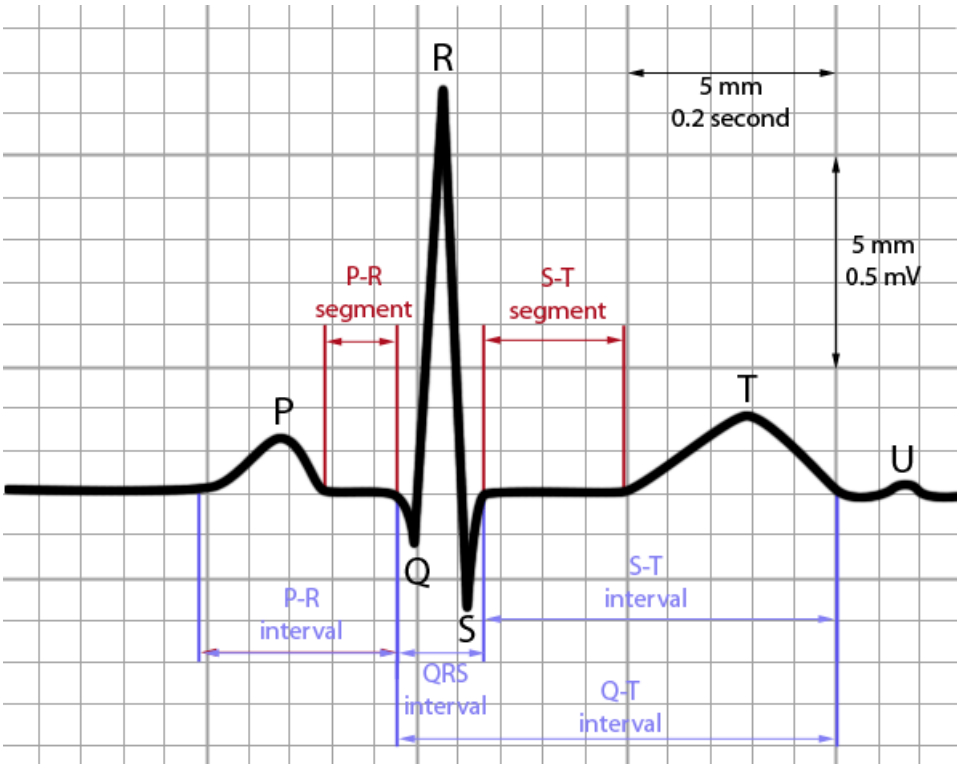

Figure 1.9: ECG Signal with characteristic intervals and segments [21]

Figure 1.9 represents an ECG typical waveform, recorded by placing electrodes on the surface of the human body [64] and characterized by:

- The P-wave: it represents the depolarisation of the atria.
- The QRS complex: it represents the depolarisation of the ventricles (most prominent wave in ECG).
- The T-wave: it represents the repolarization of the ventricles.
- The PR interval: it represents the conduction of the impulse from atrium to ventricles.
- The ST segment: it represents the beginning of the ventricular repolarization[21].

Photoplethysmogram

A less common approach to monitor cardiovascular parameter is the photoplethysmography, whose adoption has been growing significantly.
It is a simple and low-cost optical technology able to detect changes in light absorption, due to blood volume fluctuations.

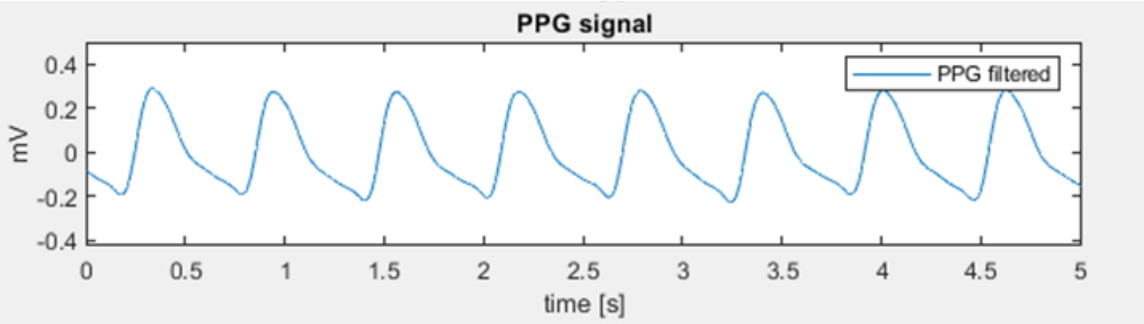

Figure 1.10: PPG Signal extracted from MIMIC III database

Starting with ECG and PPG different cardiovascular parameters can be calculated.

1.2.4 Parameters derived from biosignals

Heart rate

The **heart rate** (HR) can be determined by the interval between two consecutive QRS complexes. It is measured in beats per minute (bpm) and it is a strong medical diagnostic parameter.

Heart rate variability

The **heart rate variability** (HRV) is represented by the variance in the beat-to-beat time interval [76]; HRV increases during relaxing moments and decreases during stress. A low HRV is considered to be associated with an increased risk of CVDs [78]. Arrhythmia as well can be detected by the ECG: it is characterized by an abnormal rhythm of the heartbeat and is generally caused by valve disorders, injury from a heart attack, cardiomyopathy, high blood pressure.

Blood pressure

Another cardiovascular parameter important in determining the state of the heart health is the pressure exerted on the vessels walls by the blood flow: the **blood pressure**.

Blood oxygen saturation

An additional vital parameter is the *peripheral oxygen saturation* (SpO2); it measures the amount of oxygen bound to hemoglobin in the blood. Normal values, for adults, are within the range [94%-100%] [45]; with values lower than 90%, external oxygen supplementation is needed.

1.3 Photoplethysmography

The word photoplethysmography, introduced in the 1930s by Hertzman and colleagues, indicates the use of light to detect changes in blood volume in the microvascular bed of tissue. Photoplethysmography (also referred to as PPG) is a simple, inexpensive and non-invasive method which makes use of optical sensors.

Physiological blood volume fluctuations, at each heart beat, are responsible of the pulsatile component (AC) of PPG waveform; instead, the slow varying baseline (DC) is imputable to respiration, thermoregulation and sympathetic nervous system activity. The study of the pulsating behaviour of arterial blood volume has great clinical potential. Therefore, PPG is now widely used and investigated in clinical applications, although the origins of its different components are still not fully understood [22].
PPG technique makes only use of a light source and a photo-detector, resulting in a simpler and more comfortable continuous and long-term monitoring system.

1.3.1 Light propagation principles

Usually, when an incident light interacts with a medium, in this case the tissues, it can be reflected, refracted and absorbed.
Transmitted light in Figure 1.11 represents the incident light not reflected, not scattered and not absorbed.

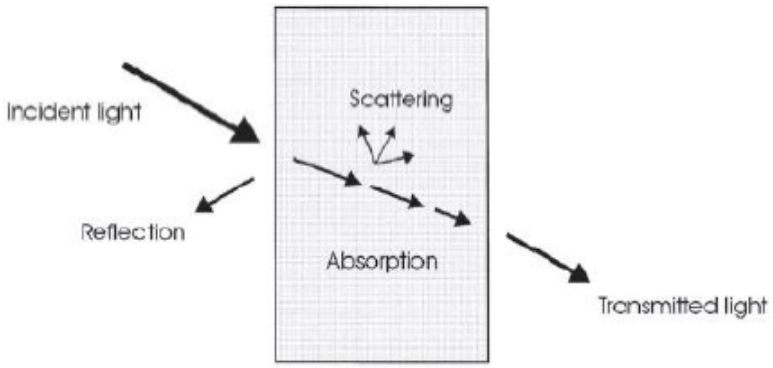

Figure 1.11: Light-matter interaction [51]

Reflection:

Reflection is the phenomenon that occurs when the light wave is bounced away from the surface.

- *Specular reflection*: takes place when the wavelength of the incident radiation compared to the surface's discontinuities is larger. Thereby the reflected and the incident angle are equal (θ=θ')

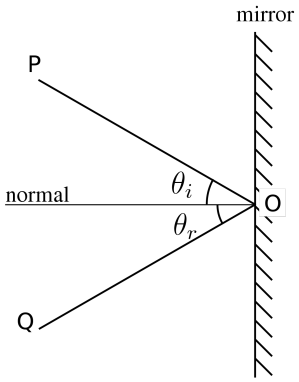

Figure 1.12: Specular reflection [15]

- *Diffuse reflection*: on the contrary, it occurs when the incident wavelength is smaller than the irregularities of the surface. The beam is re-emitted in several directions.

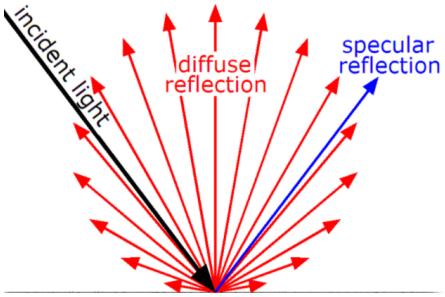

Figure 1.13: Difference between specular reflection and diffuse reflection [7]

Refraction:

Refraction is the change in direction of a light wave, due to the change in the incident wave speed in different mediums. It is the most commonly phenomenon and is governed by Snell's law:

$$\frac{sin\theta}{sin\theta''} = \frac{v}{v'} \tag{1.1}$$

where θ is the angle of incidence, θ'' is the angle of refraction, v and v' are the light velocities in the two media.

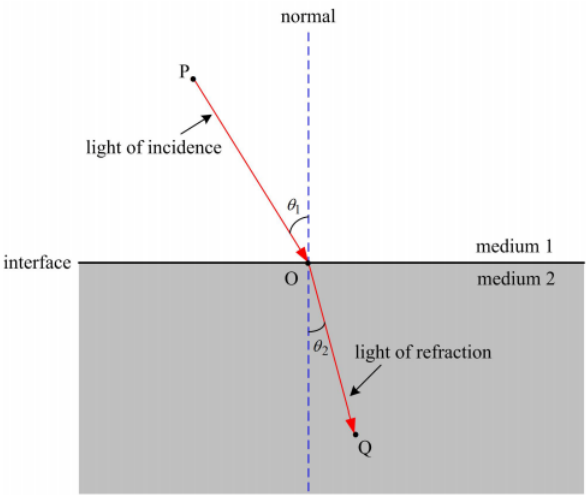

Figure 1.14: Principles of refraction of light [49]

Absorption:

Absorption occurs when a portion of incident light is retained by different biological tissues, and converted into other forms of energy. To characterize this phenomenon, it is necessary to introduce the absorption coefficient α.
According to the Lambert-Beer's Law, light intensity, in a homogeneous medium, decays exponentially as a function of the path length (l) and of the light absorption coefficient (α):

$$I = I_0 e^{-\alpha l} \tag{1.2}$$

I represents the intensity of the transmitted light through the medium and I_0 the intensity of the incident light.

1.3.2 Light-skin interaction

Since PPG sensors are placed on the skin surface, its properties play a major role in affecting the sensing outcome.
Human skin is composed of three layers:

- The *epidermis*: it is 0.01-0.15 mm thick, including the stratum corneum. It is the outermost of the three layers, provides protection to the body and is bloodless hence it depends on the dermis to be nourished. It acts mostly as an absorbent medium, minimal scattering occurs in this layer.
- The *dermis*: it is 1-4 mm thick, its blood vessels provide nourishment and waste removal for both dermal and epidermal cells [6].
- The *hypodermis*: also known as *subcutaneous*, it is 1-6 mm thick, it is the lowermost layer and is mainly composed of fat, connective tissue and pulsating arterioles.

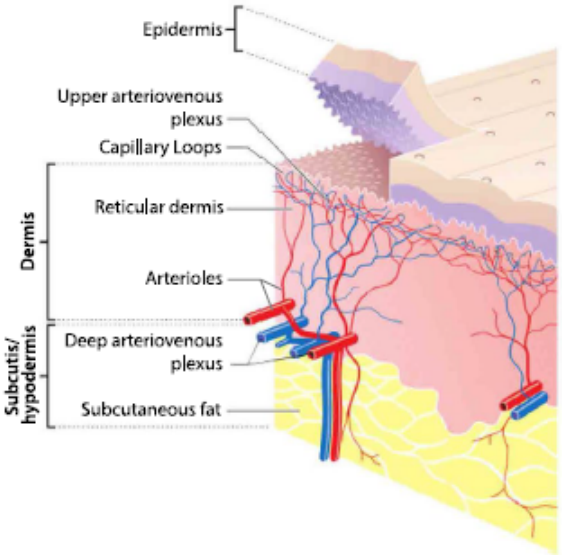

Figure 1.15: Skin structure: epidermis, dermis, hypodermis [55]

The light detected by the PPG photo-detector (reflected or transmitted, depending on the device configuration), provides important information about the performance of the vascular system. However, each layer's optical behaviour depends on the biological elements it is composed of.
Water, Melanin and Hemoglobin are some of the substances that affect the optical properties and the absorption spectrum of the skin layers. The absorption spectrum shown in Figure 1.16 represents the optical behaviour in terms of coefficient of absorption of different skin components.

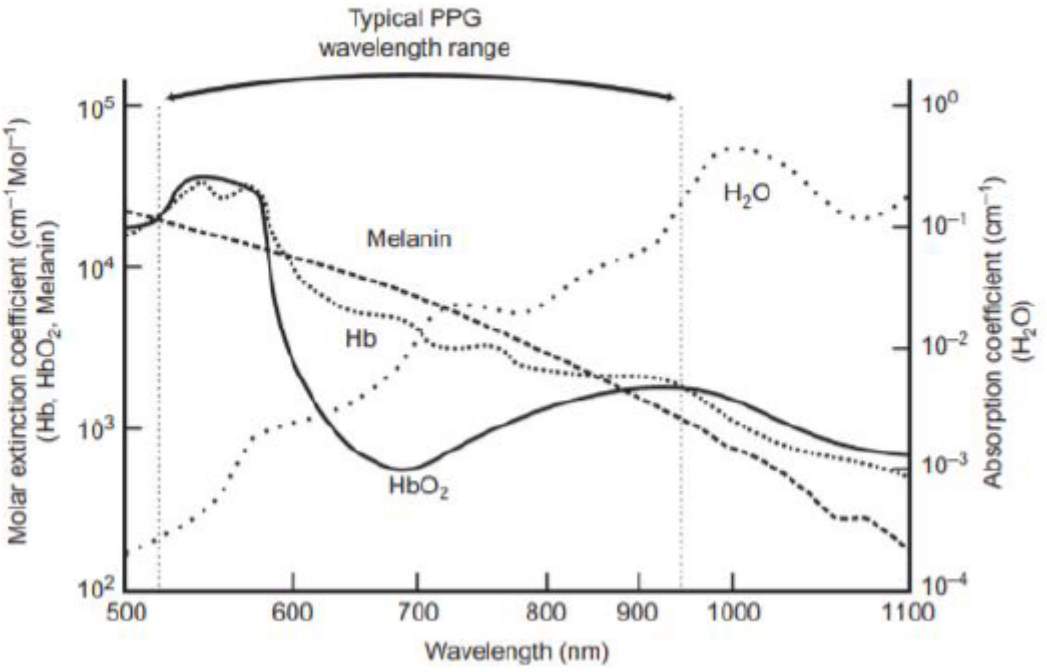

Figure 1.16: Absorption spectrum of the three main chromophores of skin [31]

- *Water* composes approximately 70% of tissues and allows an efficient transmission for wavelengths shorter than 950 nm; at a wavelength of 900 nm α is less than 0.1.
- *Melanin* is responsible of the colour of the skin and its absorption curve decreases from the *blue* zone to the *red* one. However, being restricted to a very thin layer, its effect on light propagation is negligible.
- *Hemoglobin* represents the main component of blood; it can be dysfunctional (not able to bind reversibly with oxygen) or functional, which can be fully saturated with oxygen (HbO_2), or not (Hb). Oxygenated or deoxygenated hemoglobin influences the absorption of light, depending on the wavelength adopted.

As a result the absorption spectrum of the tissue's components has a fundamental role in the choice of the wavelength of the light source.

1.3.3 The PPG light wavelengths

As known from literature [48], frequencies in the range 620 - 920 nm, corresponding respectively to *red* and *infrared* (IR) light, can penetrate deep, and reach the arteries in the epidermal layer. However, the light that reaches greater depths is more affected by movement artifacts.

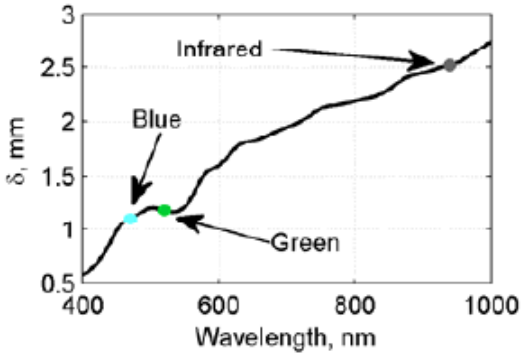

Figure 1.17: Different penetration depths [46]

On the contrary, *green* light (530 nm), according to the Beer–Lambert Law, has a small penetration depth, just enough to perceive blood volume fluctuations, although in an indirect way [46].
The morphology of the PPG signal collected with a green LED, in accordance with [42], is associated to the change in capillarity density, due to the transmural pulsatile pressure exerted by the arteries.

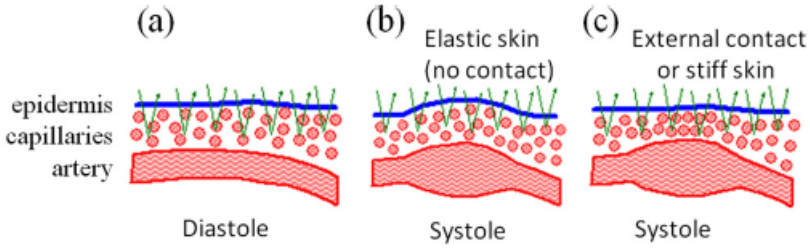

Figure 1.18: New model of light interaction with living tissue [42]

Reflected green light variation, having green LED a higher absorption coefficient for both Hb and HbO_2, is greater than reflected IR one. Moreover, the signal obtained with green light has been shown to be less susceptible to motion artifacts.

1.3.4 The PPG sensor

The working principle of PPG sensors is the optical detection of blood volume fluctuations, by monitoring the light intensity absorbed or reflected by the tissue.
The sensor system is composed of:

- *Light source*: in clinical application, light emitting diodes (LED) are used to illuminate biological tissues [73]. Conventionally, PPG sensors operate with IR or NIR light; but at present, because of its relative freedom from artifacts, green light is often used in ambulatory monitoring applications.
  However, the wavelength choice depends on the application and must be considered both susceptibility to motion artifacts and sensitivity in poor skin perfusion.

- *Photo-detector* (PD): it collects and then converts light intensity variations into electric current [23]; its spectral characteristics need to be chosen coherently with that of the light source.

The wavelength selection also depends on the configuration in which PPG device is used. Wearable PPG sensors have two operating modes:

- *Transmittance* mode: it represents the traditional configuration, in which light source and photo-detector have diametrically opposed positions. This requirement imposes limited measurement sites, with limited blood perfusion, such as fingertip and earlobe. However, this technique allows to isolate in a better way the sensor from the environmental light, so as to collect higher quality signals [73].

- *Reflectance* mode: the photo-detector is placed on the same side of the LED and aims to detect back-scattered or reflected light. Not requiring special conditions, it can be applied to thicker sites of the body, for instance: wrist, forehead and chest. On the other hand its high sensitivity to motion artifacts and to the presence of ambient light, limits the accuracy of the physiological measure. Moreover, an opaque shield is needed between the two opto-electronic elements to avoid artifacts generated by the light emitted directly from the source [73].

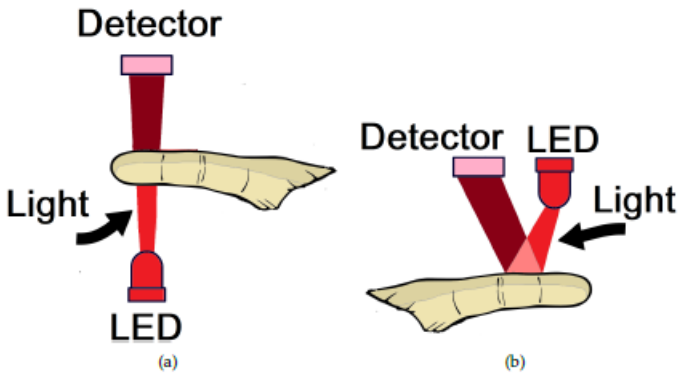

Figure 1.19: Light source and photo-detector placement in transmittance (a) and reflectance (b) mode [56]

1.3.5 The PPG waveform

An example of a typical photoplethysmographic waveform is shown in Figure 1.20.

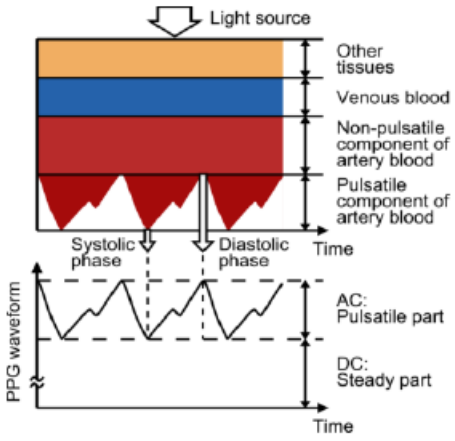

Figure 1.20: An example of the photoplethysmographic waveform [73]

It encompasses two different components: alternating current (**AC**) and direct current (**DC**) components. The latter depends on respiration, thermoregulation, vasomotor activity and sympathetic nervous system activity; it represents the optical signal transmitted or reflected by the underlying tissue and is linked to the average blood volume [73].
The former, superimposed to the DC component, is related to the cardiac cycle periodicity and stems from blood volume variations between systolic and diastolic phases. The analysis of the AC component is very important for heart activity monitoring, and it has been shown to be maximal at the green light wavelength, because of its stronger absorption by erythrocytes [42].

The pulsatile component (Figure 1.20) is made of two phases:

- *Anacrotic* phase: it represents the rising edge of the pulse and is mainly related to the contraction of the heart during the systolic phase;
- *Catacrotic* phase: it represents the descending edge of the pulse, and is identified with the diastolic phase.

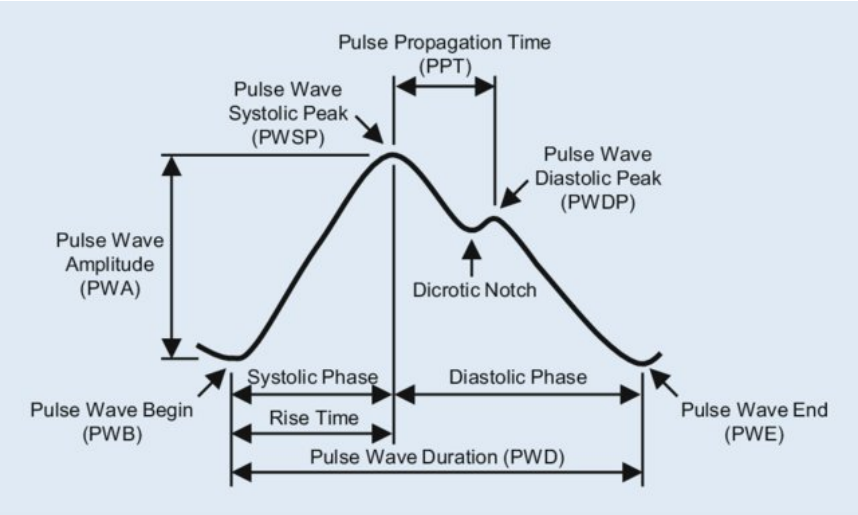

Figure 1.21: Characteristic points of the PPG waveform [37]

The systolic phase starts with the the pulse wave begin (PWB) and ends with the pulse wave systolic peak (PWSP); the pulse wave end (PWE) states the end of the diastolic phase and the time elapsing between PWB e PWE represents the pulse wave duration (PWD).
The characteristic points of the PPG are the pulse wave systolic peak (PWSP), the dicrotic notch, due to a reflected pressure wave caused by the aortic valve closure, and the pulse wave diastolic peak (PWDP). The time elapsing between two consecutive PWSP, usually expressed in ms, is called inter-beat interval (IBI) and has high correlation with the inter-beat R-R interval of the ECG signal.

As can be seen from Figure 1.20 and 1.22, PPG waveform, by convention, is inverted, so that light intensity registered by the photo-detector results proportional to the blood volume change in the tissue [42].

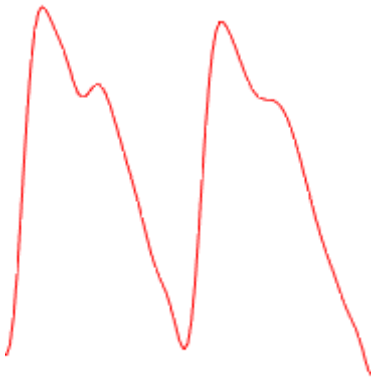

Figure 1.22: PPG signal acquired with the STMicroelectronics device

Photoplethysmography applications

PPG has been widely adopted in different clinical contexts. One of the most significant applications is pulse oximetry, which measures HR and SpO2. PPG can be even used in physiological parameter monitoring (e.g heart rate, cardiac output, respiration and blood pressure), in vascular assessment (arterial compliance, endothelial function), and in autonomic function assessment (thermoregulation function, heart rate variability and neurology assessment).

1.3.6 Blood pressure estimation

As mentioned in Section 1.1.2, our work aims to continuously monitor the SP value.
Blood pressure (BP) is one of the principal vital signs. Indeed hypertension, i.e a high value of BP, can lead to coronary heart disease, stroke, heart failure.
BP is measured in units of millimeters of mercury (mmHg) and its ideal values depend on the vessels in which BP is measured.

- **Systemic venous pressure** is much lower than the arterial one; it can be measured at different sites: vena cava, jugular vein and portal vein. Its value fluctuates in the range of 8-12 mmHg and for this reason it is not commonly used as health status indicator; however it plays an important role in intensive care medicine.
- **Systemic arterial pressure** is much higher than in the veins, and represents the value to which BP, without further specification, usually refers to.

Arterial blood pressure

The arterial blood pressure represents the pressure exerted by the blood within an arterial vessel and is composed of two values:

- The **Systolic blood pressure**: it denotes the pressure exerted against the wall while the heart is beating, that is the maximal aortic pressure [63]. It depends on the left ventricular ejection rate, arterial stiffness and peripheral resistance.
- The **Diastolic blood pressure** (DP): it represents the pressure value when the heart is at rest, it is always lower then the SP, specifically the lowest pressure value. It is influenced by peripheral resistance and arterial stiffness.

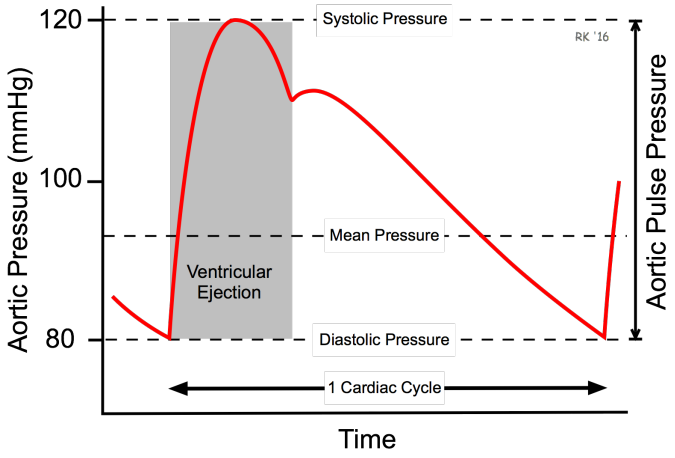

Figure 1.23: Time-domain representation of the aortic pressure curve [63]

The aortic pressure waveform in Figure 1.23 depends on the blood volume trend: a greater blood volume results in a greater pressure against the wall [63]. The first part of the cardiac cycle, the systole, corresponds to the contraction of the heart and consequently to an increase in blood volume and then to an increase in blood pressure, whose maximum is the SP. Normal SP value is <120 mmHg; normal DP is <80 mmHg, as shown in Figure 1.23.

The ABP is made of two components: steady and pulsatile. The **mean blood pressure**, the steady component, is the average pressure during the cardiac cycle [63]. It represents a better indicator of perfusion to vital organs than SP, and depends on cardiac output and peripheral resistance.
The pulsatile component is represented by the **pulse pressure**, defined as the difference between systolic and diastolic blood pressure. The pulse pressure value depends proportionally on factors such as stroke volume, fraction of blood ejected from the left ventricle during pump action and inversely proportional to the compliance of the aorta [19].

Classification of blood pressure

BP abnormalities are important predictors for many CVDs, such as heart attack, stroke, heart failure. Continuous BP monitoring improves accuracy in diagnosis of chronic hypertension [61]. The development of a continuous non-invasive BP measurement device would be a great revolution for hypertension diagnosis and treatment.

Our work focuses on **Photoplethysmogram** analysis in order to develop a non-obtrusive method for continuous BP monitoring.
Table 1.2 shows the SP and DP range division adopted by The American Heart Association for adults, and used in our project.

Arterial blood pressure ranges			
Category	**Systolic** (mmHg)	**and/or**	**Diastolic** (mmHg)
Hypotension	< 90	and	< 60
Desired	90-119	and	60-79
Pre-Hypertension	120 - 139	and	80 - 89
Stage 1 Hypertension	140-159	or	90-99
Stage 2 Hypertension	160-179	or	100-109
Hypertensive Crisis	>180	or	>110

Table 1.2: Table showing the ABP classification adopted by The American Heart Association for adults

1.4 State of the art

There exist several methods for (BP) measurement, usually divided in two categories: *invasive blood pressure* (IBP) monitoring and *non-invasive blood pressure* (NIBP) monitoring.

1.4.1 Invasive methods

IBP is commonly used in the intensive care unit (ICU) and avoided in hypertensive patients. This technique involves the use of a cannula inserted in a suitable artery (radial, femoral or brachial). The main advantages are:

- the continuous monitoring (beat-to-beat) of the patient's blood pressure, which may undergo rapid changes (e.g vascular surgery);
- the displaying of the corresponding waveform, to better understand the clinical scenario;
- the accurate measurement at very low pressure values (e.g shocked patients) and in people not suitable for NIBP (e.g obese patients) [12].

Generic intra-arterial monitoring systems consist of:

- the measuring apparatus: it is in turn formed by a cannula connected to tubing, containing a saline column conducting the pressure waveform to the transducer [12];
- the transducer: it converts the physical input into an electrical output;
- the monitor: it displays, in real time, the amplified and filtered signal.

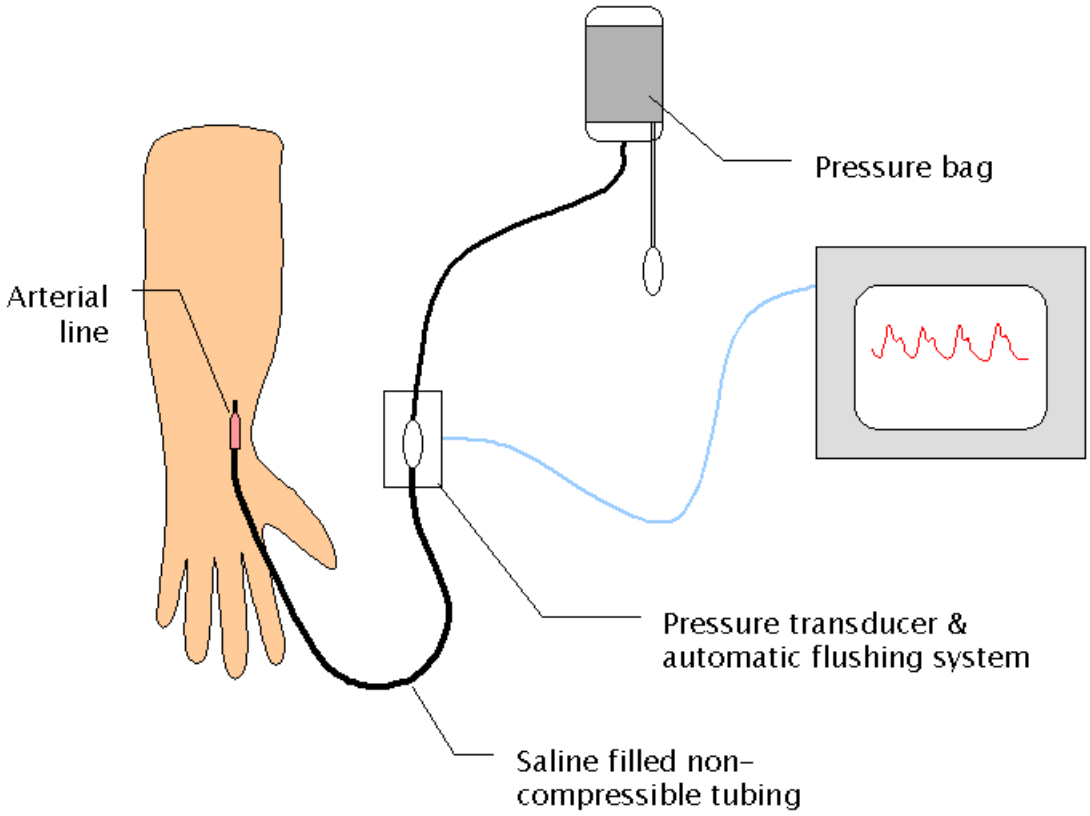

Figure 1.24: Intra-arterial pressure monitoring system [5]

However, invasive monitoring has its disadvantages: it may cause serious complications when compared to noninvasive techniques, such as hematoma, infections due to the use of the cannula and possible occurrence of thrombi, resulting in arterial occlusion.

1.4.2 Non-invasive cuff-based methods

An invasive set up is not appropriate in routine clinical practice; non-invasive methods are of common use in ambulatory measurements.

Auscultatory method

The gold standard is represented by the *auscultatory* method, based on the detection of *Korotkoff* sounds.
Measurements based on this non-invasive method are performed using a mercury sphygmomanometer and a stethoscope. The cuff pressure, positioned around the upper portion of the arm, is brought up to a level higher than the SP, up to the total compression of the artery, determined by the palpatory method in the distal side of the hand [66]. Once the cuff is gradually deflated, with the stethoscope, it is possible to hear the appearance of the Korotkoff sounds, used to determine SP and its disappearance to determine DP.

New sphygmomanometers, known as hybrid, combine electronic and auscultatory features: the mercury column is replaced by an electronic pressure gauge, but the systolic and diastolic pressure is detected likewise by an observer using a stethoscope to listen to the Korotkoff sounds [74].

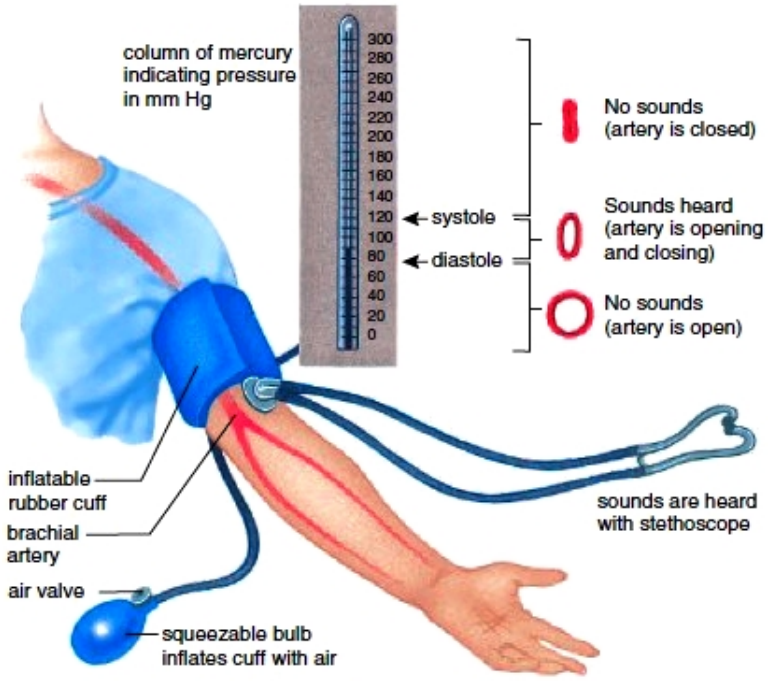

Figure 1.25: Auscultatory blood pressure measurement method [14]

Oscillometric method

The *oscillometric* method is based on the demonstration made by Marey in 1876: when the oscillation, measured during the gradual deflation and caused by blood flow oscillation, reaches its maximum, mean arterial pressure (MAP) is determined.
Starting from the MAP, systolic and diastolic pressure can be estimated applying some mathematical algorithms. These empirical algorithms, however, change according to physiological conditions, such as arterial stiffness [26].

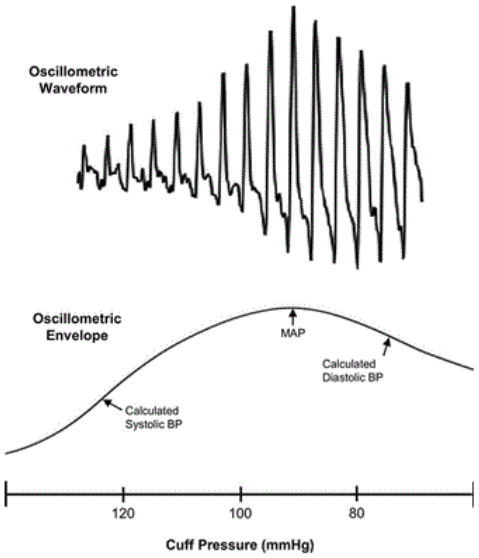

Figure 1.26: The oscillometric pressure waveform [26]

This method uses a sphygmomanometer occlusive cuff wrapped around the patient's biceps with a transducer to detect blood pressure value. Oscillometry technique is commonly used in clinical and ambulatorial application. Moreover, not requiring trained staff, automated or semi-automated devices, based on the oscillometric method, can be used at home by the patients themselves.
The main disadvantages of the auscultatory and oscillometric methods are the intermittent monitoring of the BP values and the pain or discomfort caused by the inflation of the cuff.

1.4.3 Non-invasive cuff-based continuous method

In the last few years, a new technique for continuous BP monitoring has been proposed. In the following paragraph the volume clamp method will be described.

Volume clamp method

Volume clamp method has been introduced by the Czech physiologist Jan Penáz in 1973. One of the devices based on this method is known as Finapres. It uses a small inflatable cuff around the finger, whose pressure aims to keep the finger artery diameter constant, and a photoplethysmograph, to detect blood volume. If the artery diameter is constant, and therefore also the blood volume, the arterial pressure is equal to the pressure exerted by the finger cuff [53]. Finapres Medical System BV offers beat-to-beat pressure measurement; the disadvantage remains the discomfort caused by the inflatable cuff.

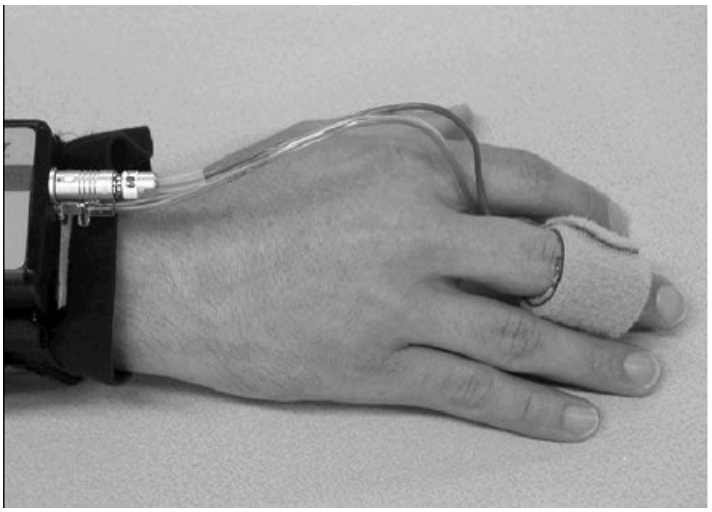

Figure 1.27: Finapres Medical System: an example of implementation of volume clamp technique

1.4.4 Non-invasive continuous cuff-less methods

Recently, cuff-less BP measurement methods have attracted a huge attention, being considered as an efficient non-invasive approach for continuous home monitoring.
The most widely used are:

- the applanation arterial tonometry;
- the PWV analysis.

Tonometric method

Tonometric method was first introduced by Pressman e Newgard in 1963 [66]. It consists on exerting a pressure over a superficial artery, with a flat bone underneath. Then, with a transducer attached to the skin, the vertical displacement of the tonometer, proportional to the pressure exerted by the blood flow, is measured.

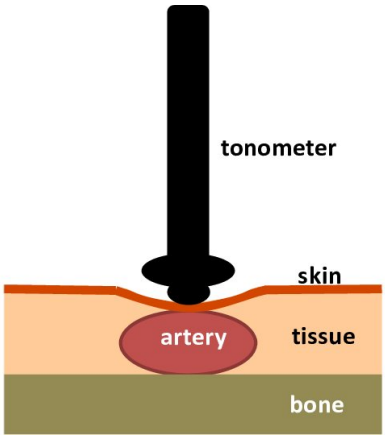

Figure 1.28: Principle of work of tonometric method

However, even the tonometric method has its drawbacks: the first is the sensor positioning, especially in the home use. Indeed, it is necessary to accurately determine the artery location; few mm are sufficient to incorrectly estimate BP value. The second disadvantage is its high sensitivity to motion: this monitoring system is suitable only if the patient is immobile, also in relation to the previous problem [62].

PWV-based methods

The most common method uses the **Pulse Wave Velocity** (PWV) [41], the blood flow velocity in the circulatory system.

In [54] **PTT** and **PWV**, calculated from the ECG and PPG signals, have been correlated with BP readings through a standard linear curve fitting algorithm.

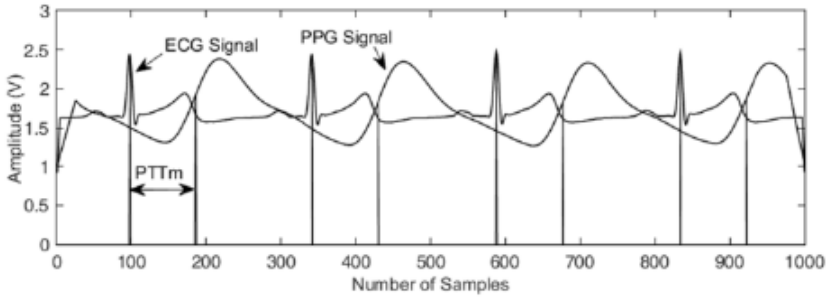

Figure 1.29: PTT extracted by ECG and PPG

In [77] PTT is computed using the ECG and the PPG signals, and is employed with other information, such as gender, age, height and weight in order to compute PWV:

$$PWV = \frac{d}{PTT} \tag{1.3}$$

d represents the distance from the PPG acquisition site and the heart. Then, the features selected are used to train the deep neural network model [77]. In literature there are countless works demonstrating the huge PTT potential to enable convenient cuff-less and BP measurement. However, methods using PWV and/or PTT, require a calibration stage and the use of two sensors placed at a known distance.

Pure PPG-based methods

In order to overcome the above mentioned drawbacks and to achieve better results, the idea to use pure **Photoplethysmographic** signal based methods has been explored. Some studies have tried to characterize the relationship between the ABP and the PPG waveforms, given their similar morphology, but it is very difficult, because there is no linear correlation between PPG morphology parameters and blood pressure.

Therefore, in [44] an ANN is used to estimate systolic and diastolic blood pressure values, from twenty-one time domain features, extracted from 15,000 PPG periods.

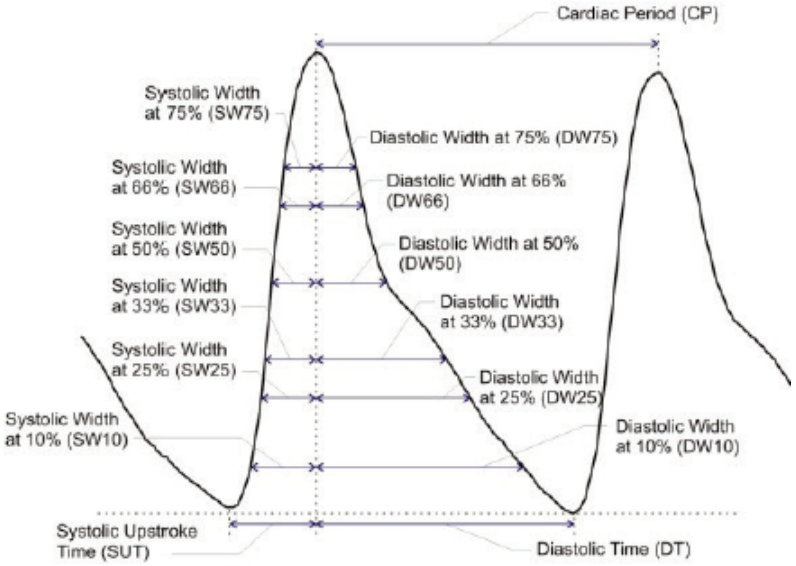

Figure 1.30: Time domain features extracted in [44]

The results obtained fall within the standards imposed by the American Association for the Advancement of Medical Instrumentation (AAMI).

In [69], in addition to the time-domain features, frequency and complexity-analysis domain features were computed, so as to describe PPG periods morphology, complexity, amplitude and phase in the frequency domain, and the arterial stiffness through the *Large artery stiffed index*. After a dimensional reduction through a feature selection algorithm, relevant features are used to predict BP values. The best results have been obtained with the *Ensemble* learning algorithm: a mean absolute error (MAE) of 7.87 mmHg $\pm$ 7.47 mmHg for SP and 3.84 mmHg $\pm$ 3.63 mmHg for DP.

In [68], a Photoplethysmography signal, a velocity plethysmogram (first derivative of PPG) and an accelerated plethysmogram (APG) signal (second derivative of PPG and an important indicator of arteriosclerosis), are fed into an autoencoder. An autoencoder is an ANN used for unsupervised learning processes, that extracts complex features helpful in BP estimation.

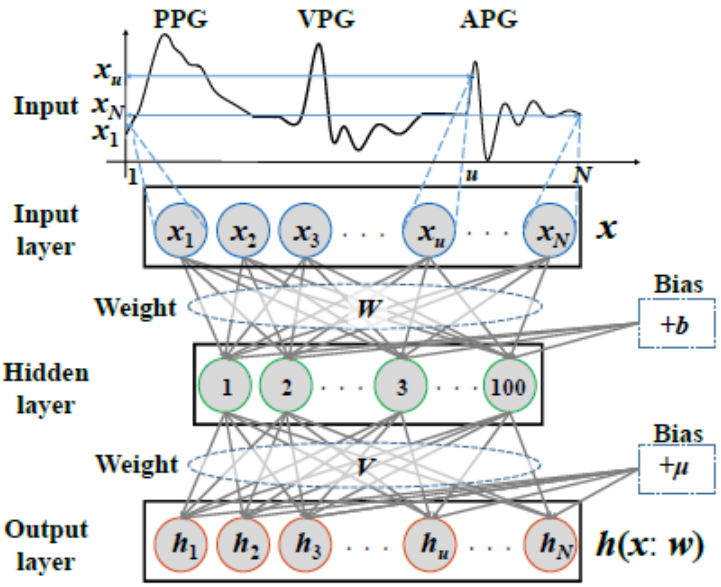

Figure 1.31: The autoencoder used for feature extraction in [68]

The output features of the autoencoder are different from the conventional ones, and effective for the BP estimation.

1.5 Artificial neural network

Referring to the introduction of [34], *artificial intelligence (AI) is the science and engineering of making intelligent machines.*
Artificial neural networks is a class of machine learning (ML) algorithms, one of the applications of AI. ANN is a mathematical model for predicting system performance (i.e., system output) inspired by the networks of neurons in the human brain.

1.5.1 Human neurons physiology

The human brain is composed of approximately 86 billion neurons, its basic functional unit.

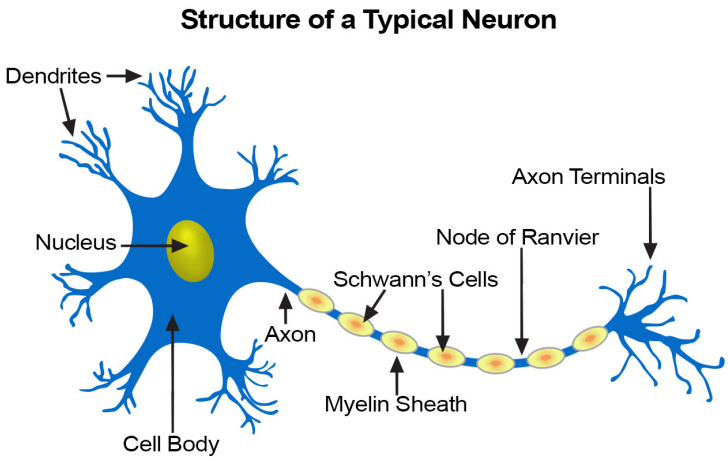

Figure 1.32: Human neuron structure [13]

The neuron is a particular type of human cell, with a limited regeneration capability after injury and whose function is to transmit nerve impulses.
Each neuron consists of three basic parts:

- *Cell body*: also called *soma*; it contains the *nucleus*. It elaborates the incoming information received from other neurons and produces the output signal.
- *Dendrites*: they represent the gateway through which excitatory or inhibitory impulses reach the soma. The dendrites take part to the afferent process.
- *Axon*: it arises from the cell body, it is covered by the *Myelin Sheath*, an insulating material, and acts as action potential propagator. The *axon terminal* is accounted for transmitting the output signal to the afferent neuron. The axon takes part to the efferent process.

Communication between neurons through electrical events, known as action potential, occurs at the junctions, the *synapses*. The postsynaptic neuron can receive stimuli from different input neurons at the same time, so as to be summed spatially, or from the same neuron at different instants of time, so as to be summed temporally. Action potentials come about only when the sum value exceeds the action potential **threshold**; at that point the receiving cell is *fired*.

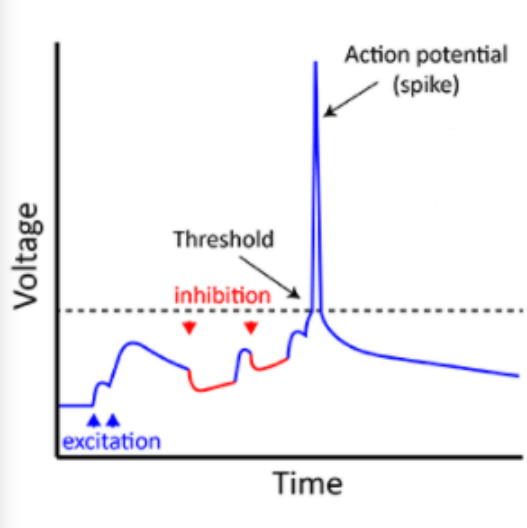

Figure 1.33: Neuron action potential [2]

Although the brain information processing remains mostly unknown, the action potential propagation across thousands of fired neurons leads to the brain **thinking**.
ANN attempt to simulate brain functioning, so as to learn things from experience and make decisions.

1.5.2 Artificial neurons

ANNs processing unit is represented by the *artificial neuron*.
An example of artificial neuron is represented in Figure 1.34; it is a mathematical function that transforms the input data (x_1, x_2, x_3, x_4) in the output. As the input vector enters the node, it is multiplied by the weight vector (w_1, w_2, w_3, w_4) and a possible bias value is added, before passing the sum through the activation function.

$$in_i = bias + \sum w_{j,i} * x_j \tag{1.4}$$

The *activation function*, represented by the unit step function in Figure 1.34, determines the *perceptron* (single unit of ANN) behaviour.

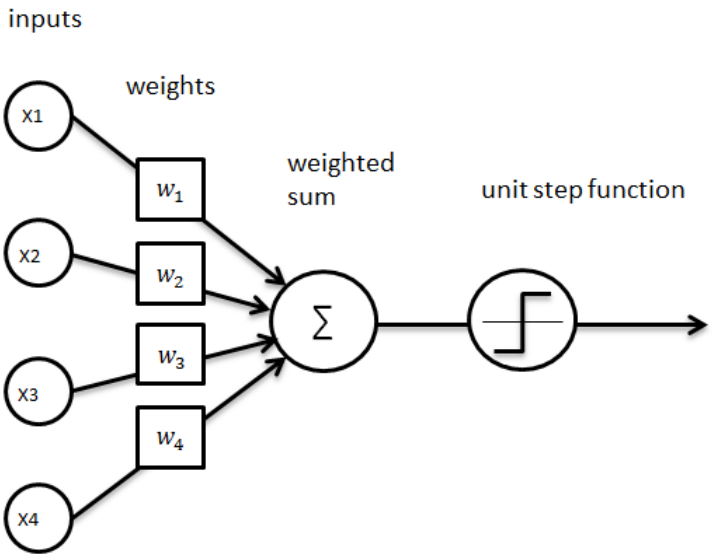

Figure 1.34: An example of artificial neuron [9]

Activation function:

The activation function (AF) establishes whether the node should be activated or not, and how much active, if a certain threshold value has been exceeded by the input. The AFs introduce non-linearity in the node, in order to allow such network to compute highly complicated tasks, for instance pattern recognition, classification/clustering, prediction. Several AFs exist, the simplest one, reported in the artificial neuron in Figure 1.34, is the *Step Function*: if the input value reaches a certain threshold, the node is activated and its output is 1, otherwise the output value is 0. In this case it is not possible to map multiple classes, and it is used only to create binary classifiers [17] [9].

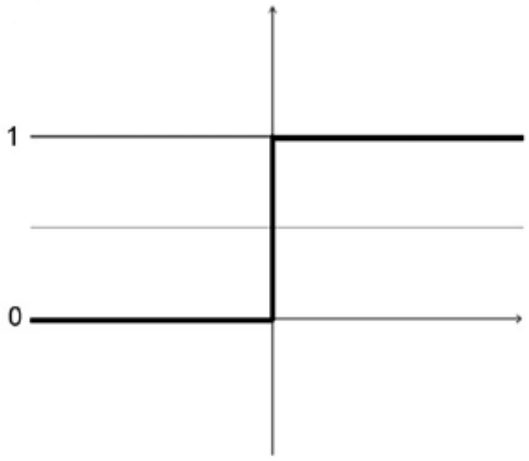

Figure 1.35: Step function [28]

The *Sigmoid Function* is one of most widely used non-linear AFs. Thanks to its properties (non-linearity, range values between [0,1] and steepness), it enables the node to take any values in its range. It is used in predicting probability-based outputs [1].

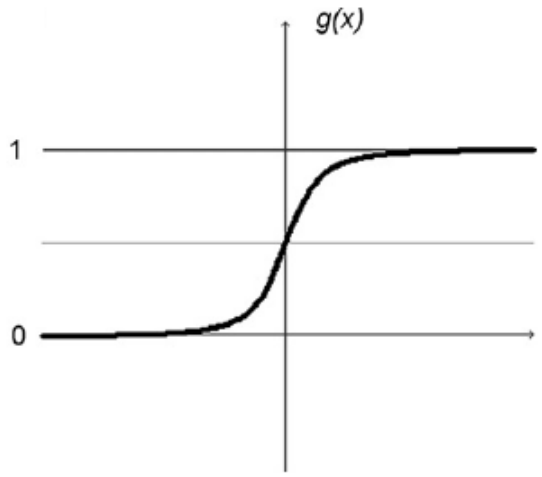

Figure 1.36: Sigmoid function [28]

The *Softmax Function* is commonly used in the output layer of multiclass classification models, only when classes are mutually exclusive. Its output vector returns the membership probability for each class of every input.

The *Rectified Linear Unit Function* (ReLu) is the most used at present [1]. Its output is defined to be 0 for all negative inputs, making the model unable to map negative values properly.

Many other AFs exist; however, there is no systematic method to choose the most suitable AF; it depends on the ANN application.

Architecture:

ANNs are frequently made up of a large number of nodes, usually divided into layers:

- *Input layer*: it represents the initial layer and has the purpose to transfer the input data from the external environment, into the subsequent layers of the system. Its neuron number matches with the number of the input patterns.
- *Hidden layer*: in ANNs there could be zero or more hidden layers. They are placed between input and output layers, their function is to transform the input data, applying non linear AFs;.
- *Output layer*: it produces and presents the final output of the neural network, and its nature depends on the problem considered. The output can be a numerical prediction, in regression task (Figure 1.37 a)), and categorical, or discrete, in classification task. Each output neuron (Figure 1.37 b)) represents a different class, and that with the highest value (the green one in Figure 1.37 b)) indicates the predicted class.

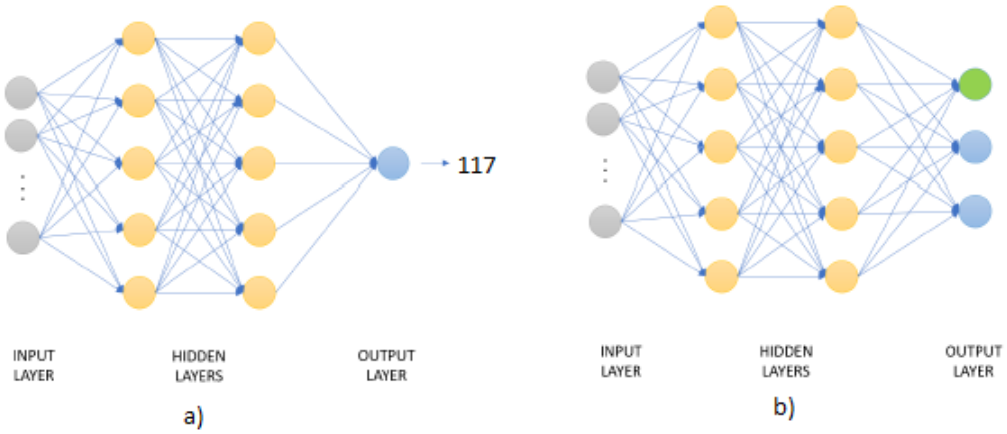

Figure 1.37: Two different ANN architectures used for a) Regression and b) Classification problem

Increasing both depth (number of hidden layers) and width (number of neurons comprising each hidden layer), it is possible to improve the accuracy or not, depending on the problem we are trying to solve. However, there is no theory that helps choosing the optimal number of hidden layers and hidden neurons. Not enough neurons will lead to a high training error; on the other hand, too many hidden neurons will result in low training error but low generalization capability.
An ANN with at least two hidden layers and where each neuron is fully connected only with all the neurons of the previous and the subsequent layer is a multilayer perceptron (MLP).

According to the neurons' weighted connection, different ANN architectures exist. If every node is connected with each other, we are dealing with *Fully connected networks* [52]. Based on the weighted connection direction, ANN architecture can be divided into:

- *Feed-forward networks*: the input information is passed only in one direction, from input to output layer;
- *Recurrent networks*: weighted connection are allowed in both forward and backward directions.

Learning process:

The learning process in ANN consists in improving the model performance, by adjusting the weights and bias levels in an iterative way. The learning process, or training algorithm, based on the data available, can be:

- *Supervised*: it uses labeled training sets, so as to know the distance between prediction and ground truth. The latter is used to modify the internal parameters, in order to achieve better results.
- *Unsupervised*: it uses unlabeled training sets; it is the neural network's duty to find structure or correlation between patterns which fully describe the data. In this case it is difficult to assess the model's accuracy.

The main difference between supervised and unsupervised training is that using the first one it is possible to compare the network outputs with the available desired values.

The *Training process* is one of the key concepts of ANN, during which weights and biases are updated, in order to reduce the prediction or classification error. The original dataset, containing all available samples, is portioned in three subsets: the most popular choice is 70% training set, 15% validation set and 15% test set.

Each iteration, beginning with the first portion of the training set fed into the input layer and leading to neuron weights adjustment, is called *epoch*. Each portion of the training set has the same dimension, also called *batch size*. This is useful to avoid poor generalization capability of neural network, and to have faster convergence to good solutions, even if a small batch size does not guarantee the achievement of the global optima [8].

The cost function, calculated over the entire training set, allows to evaluate the *loss* of the model, to be minimized. However, as the weight space size increases, the cost function computation and minimization become very heavy.
In order to overcome this drawback, *optimization algorithms* are used to change the weights and the learning rate of the neural network in order to reduce the loss and to achieve the most accurate result [18].

Optimization algorithm

The most optimization algorithm used in ML is the *Gradient descent* (GD), and most of the optimization algorithms are essentially variations of GD.

- In the *Gradient descent algorithm*, starting from a random initial weight, the purpose is to find the lowest error value in the cost function.

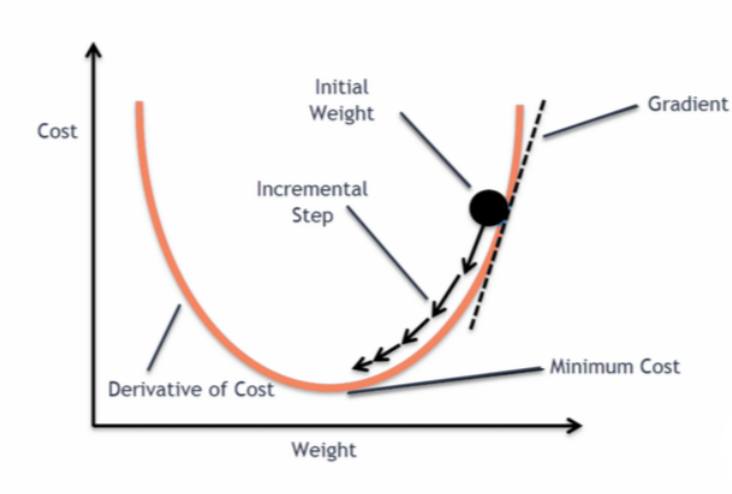

Figure 1.38: The cost function plotted against one weight

The negative value of the partial derivative of the loss function points to the direction we need to change the weight to reach a minimum.

$$w_{i,j} = w_{i,j} - \nabla \frac{dJ(W)}{dw_{i,j}} \tag{1.5}$$

In order to update the weight value, according to the Equation 1.4, the *Back-propagation* algorithm is applied. The hyper-parameter *learning rate* tells us how much to change the weight along the direction given by $dJ \backslash \mathrm{dw}$[2].

[2]Leibniz's notation.

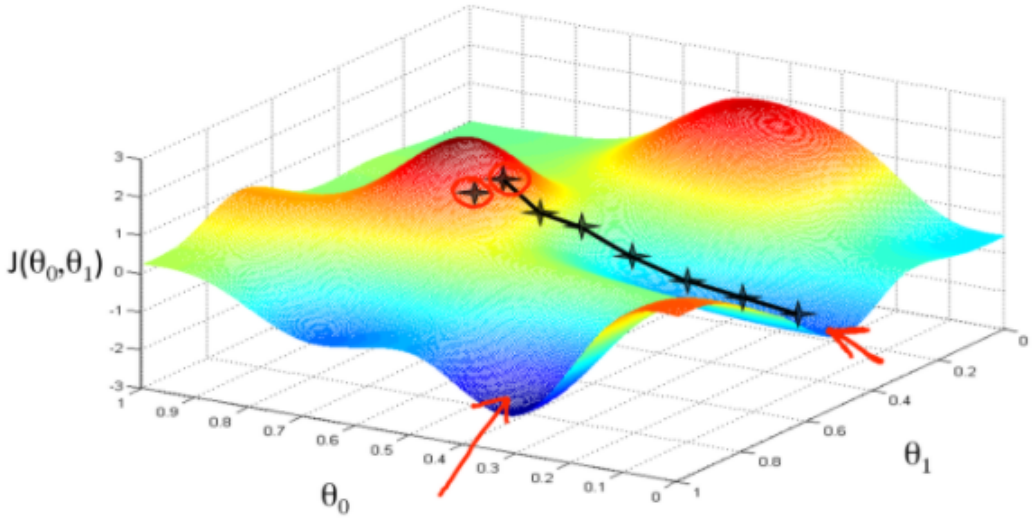

Figure 1.39: Working principle of GD [10]

- The *stochastic GD* is a variant of GD and consists in updating the weight and the learning rate using a *subset* of the training set, instead of using all the samples as in the GD.

- The *Adam* optimization algorithm is based on the method of adaptive learning rates, by assigning individual values to each parameter [3].

Back-propagation algorithm:

The forward phase consists in propagating the input data, from the input layer to the output layer, passing through the hidden layers.

Once the network response is compared to the desired output, being a supervised learning process, the corresponding error is used to adjust the weight of all the neurons. *Back-propagation* algorithm is used to train ANNs.

Before the backward propagation is implemented, the cost function ($J(w)$) and its partial derivative with respect to the weight space ($\frac{\partial J}{\partial w}$) and with respect to the bias ($\frac{\partial J}{\partial b}$) are calculated one layer at a time. We consider a Neural Network with m input nodes, N layers, a loss matrix $\Delta_{i,j}^{(n)}$ composed of a specific value for each neuron j in each layer n at each training example i.

The backward phase, hence, consists in the back-propagation of the error to the first hidden layer in order to alter the ANN's parameters.

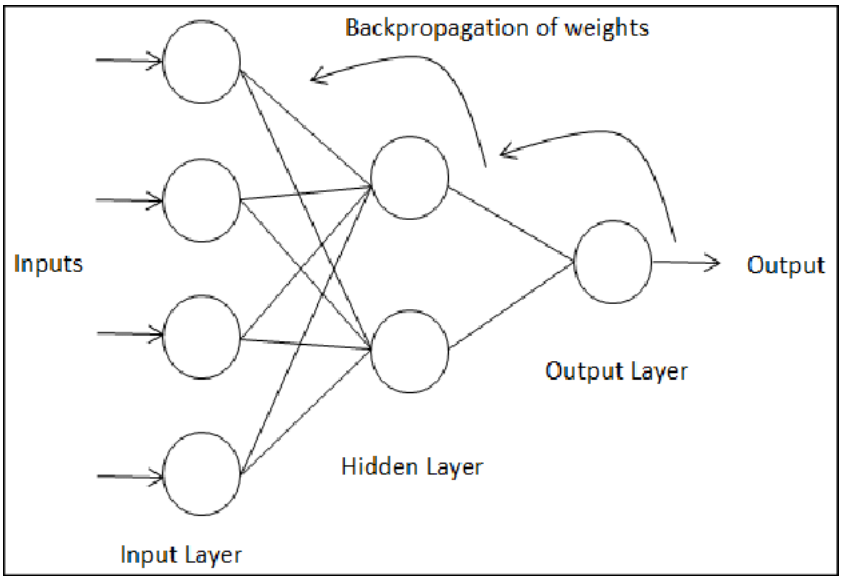

Figure 1.40: Back-propagation algorithm [39]

Multilayer perceptron

MLP is a class of feed-forward ANN and it is the mostly used model to solve complex tasks, and we, therefore, decided to use it in this work as well.

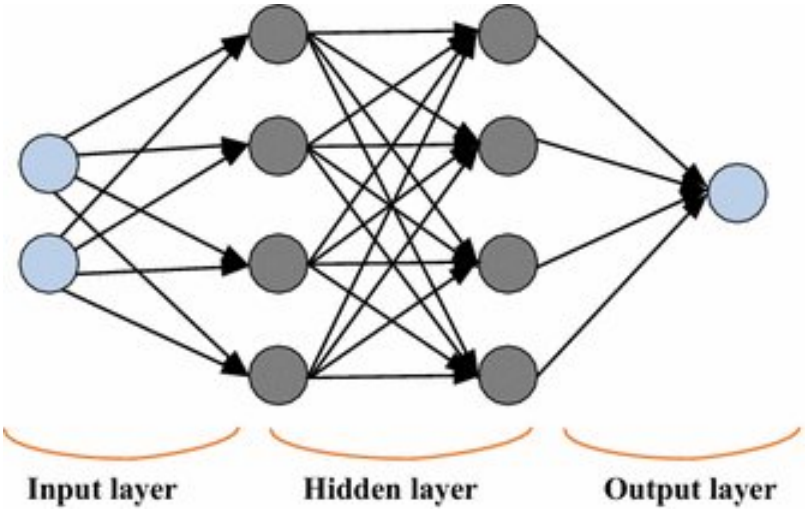

Figure 1.41: A simple example of a multilayer perceptron ANN [57]

In order to obtain good prediction performance, *multilayer perceptron* NN must be trained with the *back-propagation* method. During the training process, as mentioned above:

- the training set is fed into the input network;
- the output obtained is compared to the desired one;
- the error is back-propagated through the hidden layers in order to adjust the network's parameters and to improve its performance.

Once the training process is completed, the network's prediction capability is evaluated over the test set. If the results achieved are similar it is possible to assert that the MLP network generalizes well. Otherwise, if the training set performance is much better than the test set ones, *overfitting* occurs. *Overfitting* is the network's inability to generalize, and it occurs when the model learns specific patterns of the training set, that turn out to be irrelevant for the new unseen data.
To verify overfitting during the training process the validation set is used, so that it is possible to interrupt the learning process as soon as different performance results occur. The test set, instead, is used at the end of the training phase, once the validation performance has proved to be good, in order to demonstrate the prediction or classification precision of the model on a new set of data.

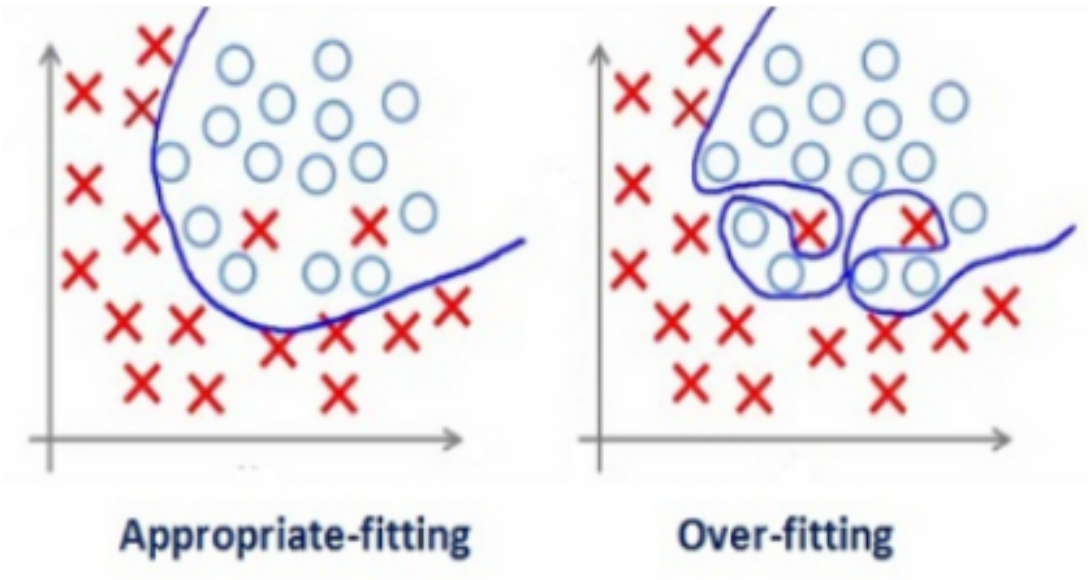

Figure 1.42: A good generalization and an overfitting example [16]

Chapter 2

Materials and Methods

The aim of this study is to create a non-invasive wearable system for continuous SP monitoring. It has been developed at STMicroelectronics s.r.l in Agrate Brianza, in the *Remote Monitoring* division.
This project is divided into two parts:

- in the first part publicly available data has been used;
- in the second part signals acquired by the *MORFEA3* have been used.

2.1 Data Collection

Our first database has been derived from *Multi-parameter Intelligent Monitoring Intensive Care* (MIMIC) III Waveform Database Matched Subset of the PhysioNet organization [38]. MIMIC III Waveform Database Matched Subset contains records (ECG, ABP and PPG, heart rate, oxygen saturation, etc.) for 10,282 different patients, collected in ICUs. In order to carry out our project, the PhysioNet *WaveForm DataBase* (WFDB) Software Package was used to download PPG and ABP signals.

The first step was to select a list of records containing at least simultaneous PPG and ABP signals:

Figure 2.1: PhysioNet record search tool

Then, with the WFDB tool, we have extracted PPG and ABP signals, at sampling frequency of 125 Hz, from 47 patients in ICU.

Displaying the signals before downloading them, allowed us to discard low quality segments (Figure 2.2). This second step has been executed on *Google Colaboratory*[1] platform.

[1]It is a free cloud service used to execute Python code and to train our ML model on CPUs, GPUs, and TPUs.

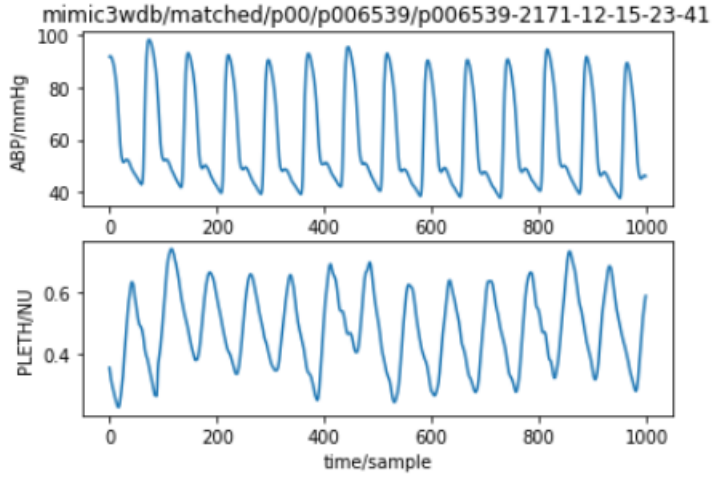

Figure 2.2: Example of raw ABP and PPG signals extracted from MIMIC III

2.2 Signal pre-processing

Although we have tried to select only good quality signals, it has been necessary to perform a pre-processing stage in *Matlab R2019a*.

The first step was to plot both signals entirely for each patient in order to find further distorted portions of ABP or PPG, and to delete the same segment from both signals. Even after this trial, the amount of high quality periods remained significant.

To pre-process the ABP signal, a low-pass filter, with a cutoff frequency of 6.6 Hz, has been used to remove high-frequency noise. Low-frequency components, on the contrary, are maintained so as not to change SP values. Then, the *Hampel filter* has been applied to the ABP signals to detect and remove the outliers.

To build our Database, only SP values are needed; the *findpeaks* function has been used in *Matlab R2019a* to detect the ABP waveform peaks, and the indices at which the peaks occur.

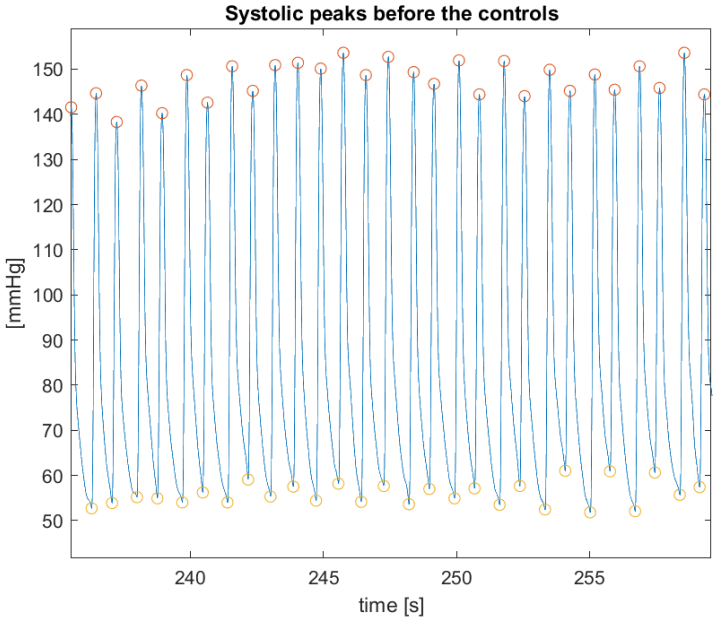

Figure 2.3: Findpeaks function in ABP segment

Each SP value, after PPG signals pre-processing, has been associated with the PPG corresponding period.

To pre-process the PPG signal a 4^{th} order Butterworth band-pass filter at 0.5 Hz and 8 Hz, has been used. In contrast to ABP filtering, in PPG pre-processing it is essential to remove the trend.
Since each SP value corresponds to a single PPG period, a segmentation algorithm is performed.

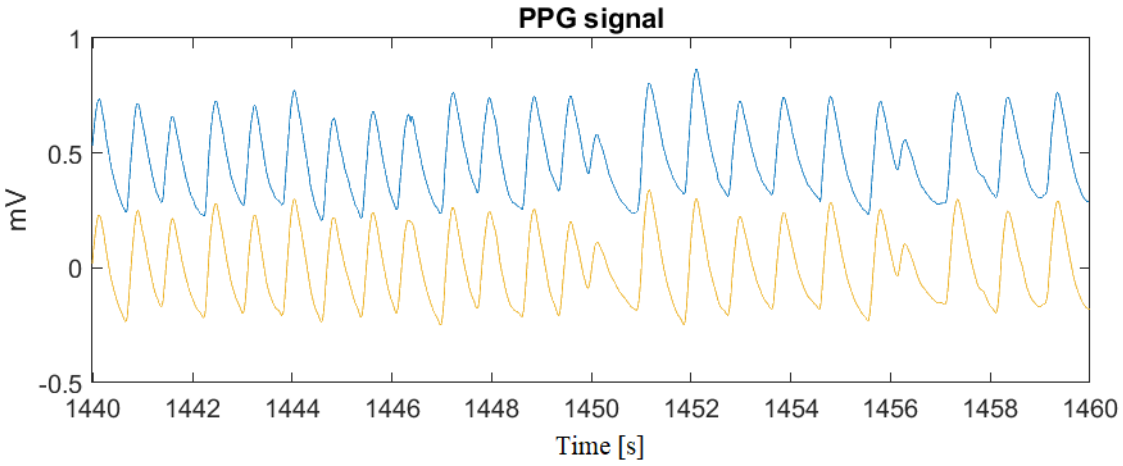

Figure 2.4: PPG segment before (blue) and after (yellow) filtering

1. The PPG period is the portion of signal between two minima, so the first part of the algorithm consists in the detection of local maxima (PWSP) and minima (PWB and PWE).
   As shown in Figure 2.5, it is possible to detect wrong periods, because of some remaining artifact. In order to avoid undesired acquisitions, the algorithm in Figure (2.8) has been implemented.

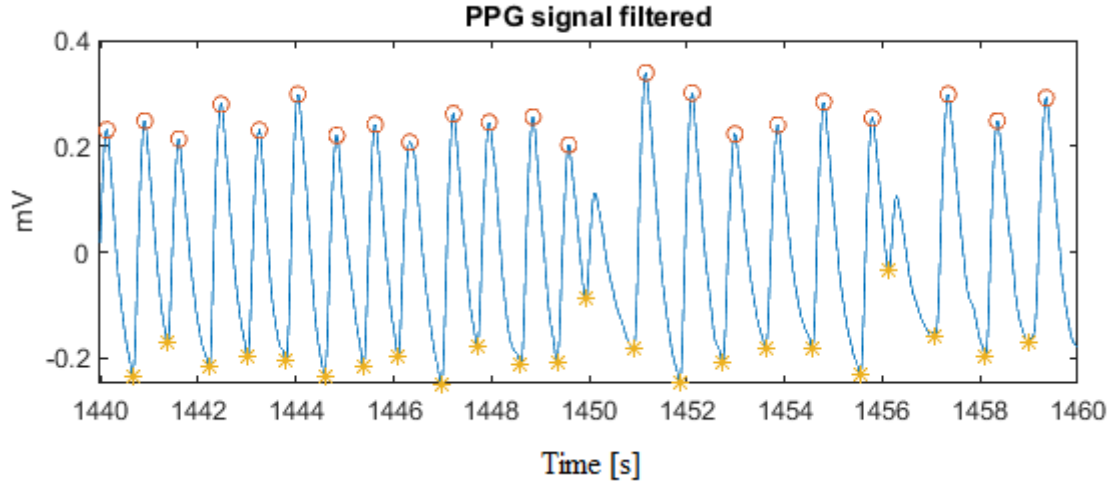

Figure 2.5: Local maxima and minima detection in *Matlab R2019a*

2. The second part is the identification and the removal of all periods shorter or longer than two threshold values[2] defined experimentally for each patient. This periods, indeed, represent respectively portions of PPG periods or two adjacent PPG periods not recognized by the findpeaks function.
   The remaining PPG periods have been plotted overlaid in order to check the segmentation quality. An example is reported in Figure (2.6)

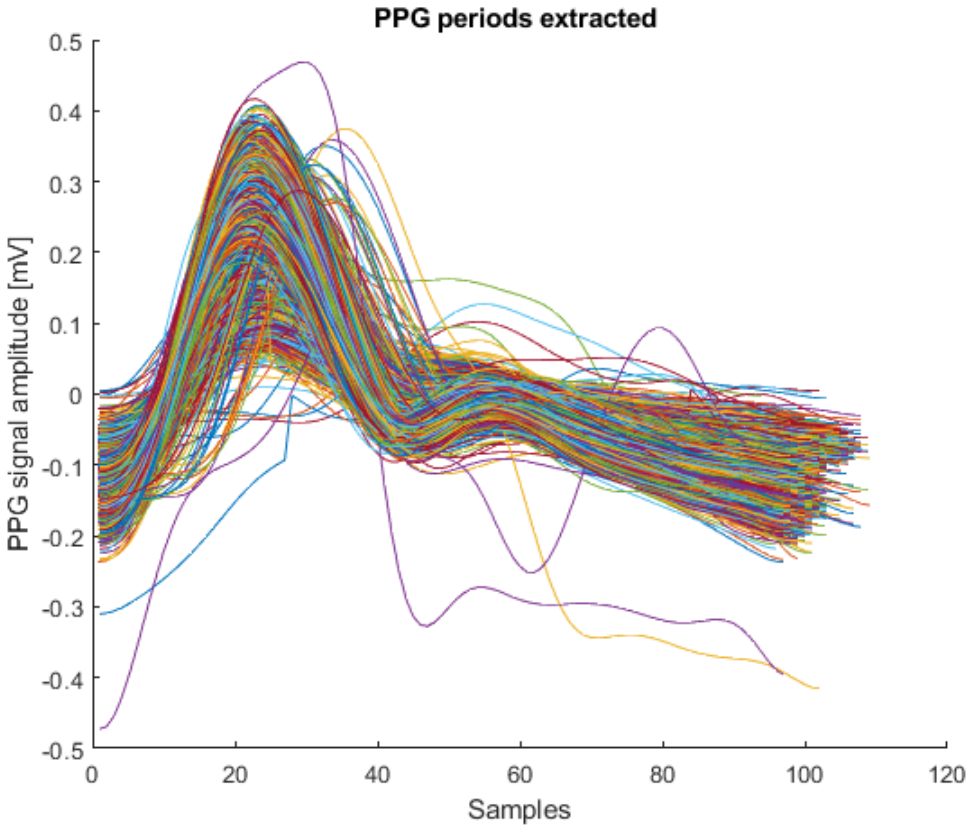

Figure 2.6: PPG period extracted after shorter and longer periods removal

[2]The two threshold values are respectively the minimum and maximum lengths accepted: the 83.3% and the 120% of the average length of the patient periods.

3. Then, the third step is applied: it consists in the detection and the removal of all the periods with artifact peaks, by calculating the first and the second derivative[3]. An example of the segmentation result is reported in Figure 2.7; the PPG periods in figure are collected in a matrix with the corresponding SP value.

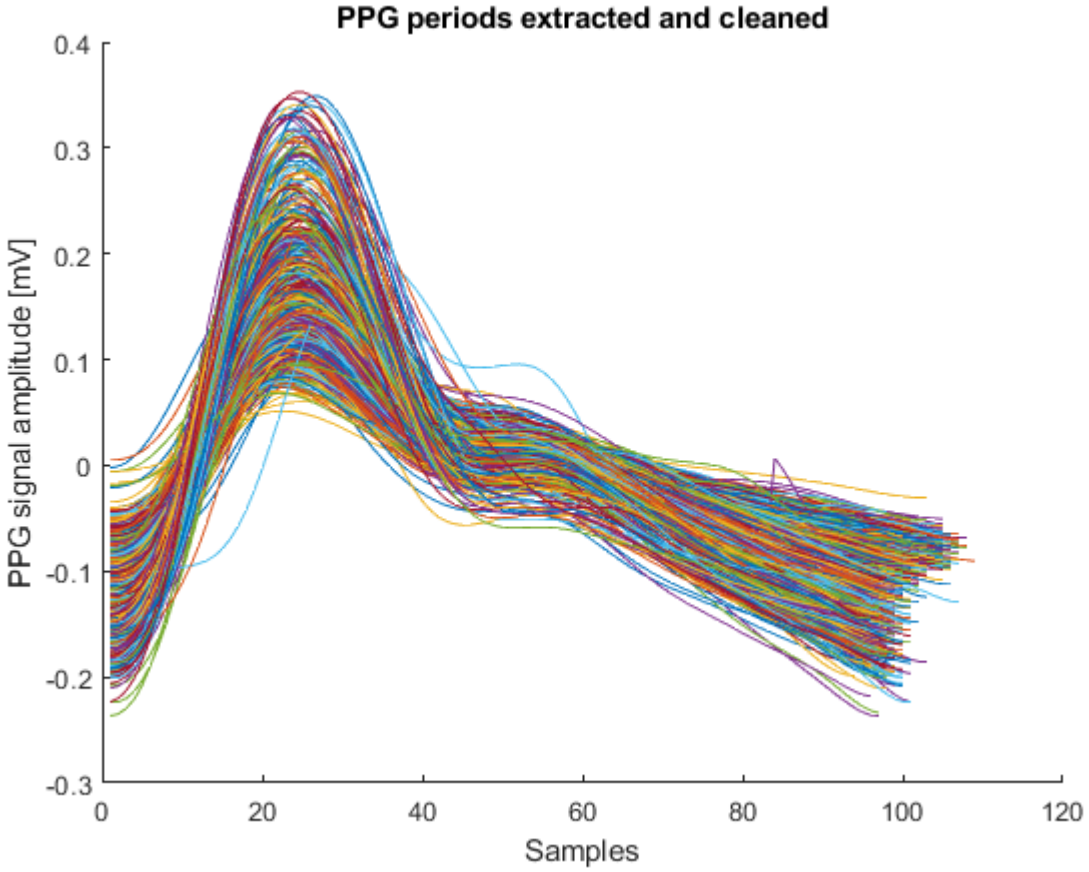

Figure 2.7: PPG period extracted and cleaned

[3]If the first derivative is 0 and the second derivative is positive, there is a local minimum. If the first derivative is 0 and the second derivative is negative, there is a local maximum

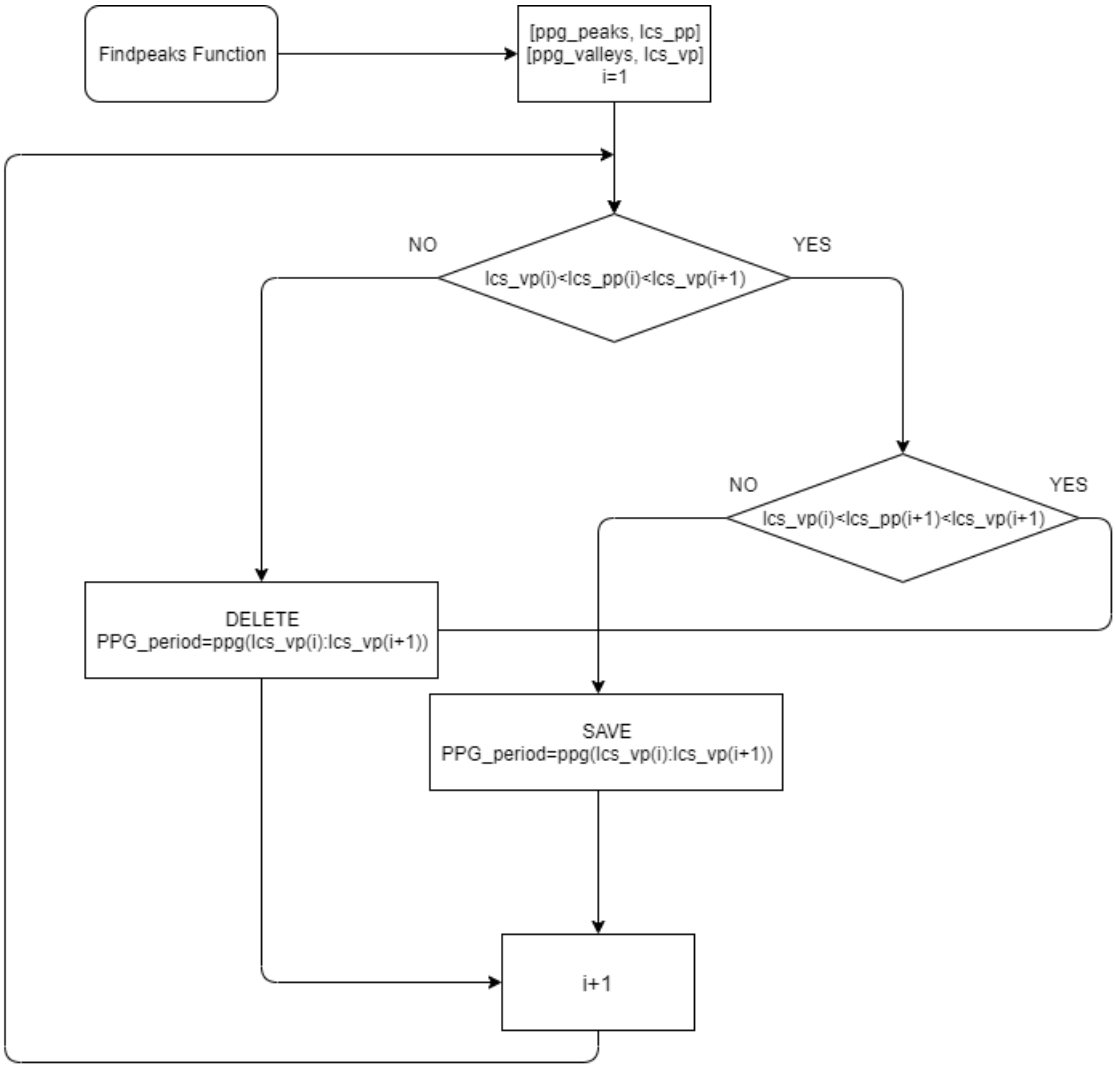

Figure 2.8: Peak detection control algorithm

2.3 Dataset construction

In the first part of our work several datasets have been built.

Thirty-seven patients' records have been used. Each dataset contains a set of different features, extracted for each PPG periods, to which the corresponding SP value is associated.

	FEATURES								TARGET
	F1	F2	F3	F4	F5	F6	...	FN	SP VALUE
1									
2									
3									
4									
...									
M									

Figure 2.9: An example of our dataset template

The number of features (N) depends on the dataset used and the number of periods (M) equals 249.672. Features are calculated through the feature extraction process.

1. The first dataset has been built on the basis of [44]. Fifteen temporal variation features have been extracted from each PPG periods (Figure 2.10):

   - Cardiac period (CP), systolic upstroke time (SUT), diastolic time (DT).
   - Systolic width (SW) in terms of time at 10%, 25%, 33%, 50%, 66% and 75%.
   - Diastolic width (DW) in terms of time at 10%, 25%, 33%, 50%, 66% and 75%.

2. In the second dataset, we have added to the previous fifteen features, four additional features extracted from the second derivative of the PPG waveform, the APG.

   - The first step aims to identify the *a, b, c, d* and *e* peaks (Figure 2.11), which are respectively the early systolic positive wave, the early systolic negative wave, the late systolic reincreasing wave, the late systolic redecreasing wave and the early diastolic positive wave [50].

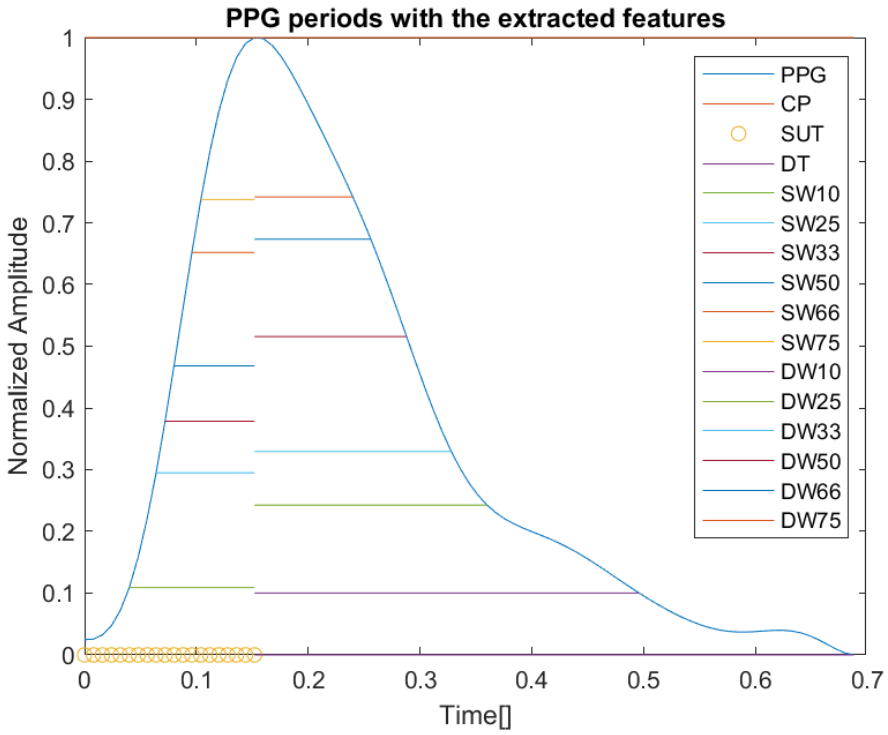

Figure 2.10: An example of a PPG period with the extracted features

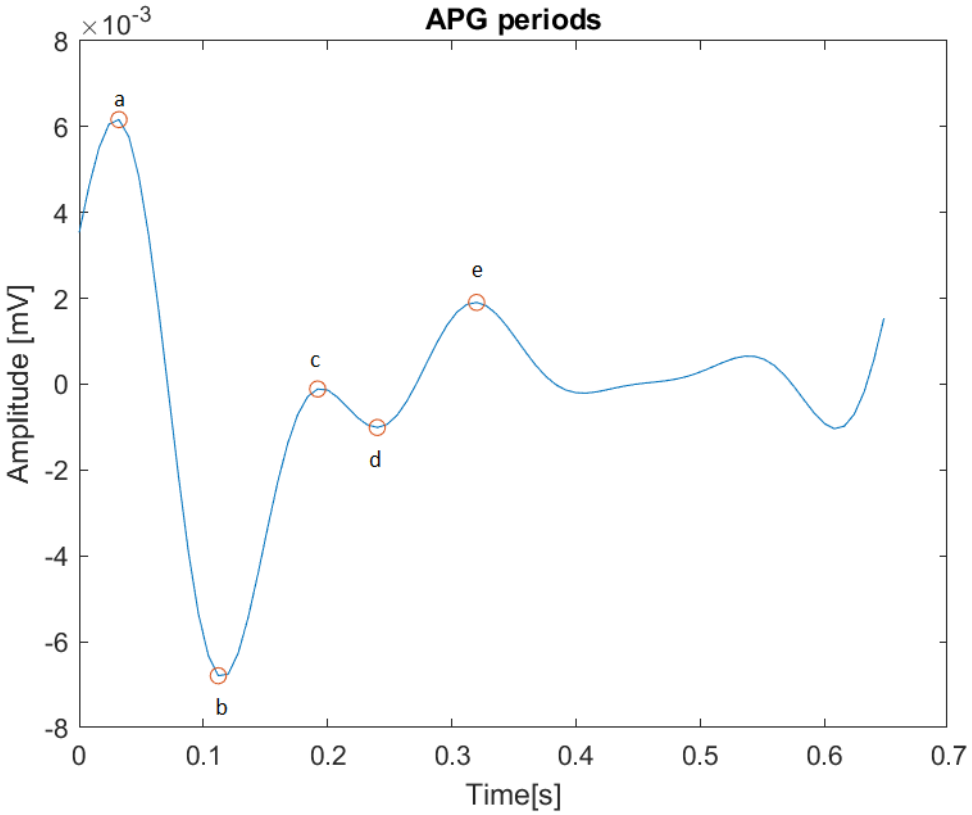

Figure 2.11: An example of an APG period with the first-mentioned peaks

- The second step is the computation of the APG features: the ratios between the intensities of the identified peaks: $b\backslash a,\ c\backslash a,\ d\backslash a,\ e\backslash a$ [50].

3. The third dataset was made up of twenty features: the previous ones, plus the PTT. In order to build this dataset, it has also been necessary to download the ECG signals of the same patients. In this work PTT is intended as the time between the ECG R-peak and the PPG Systolic peak (Figure 2.12).

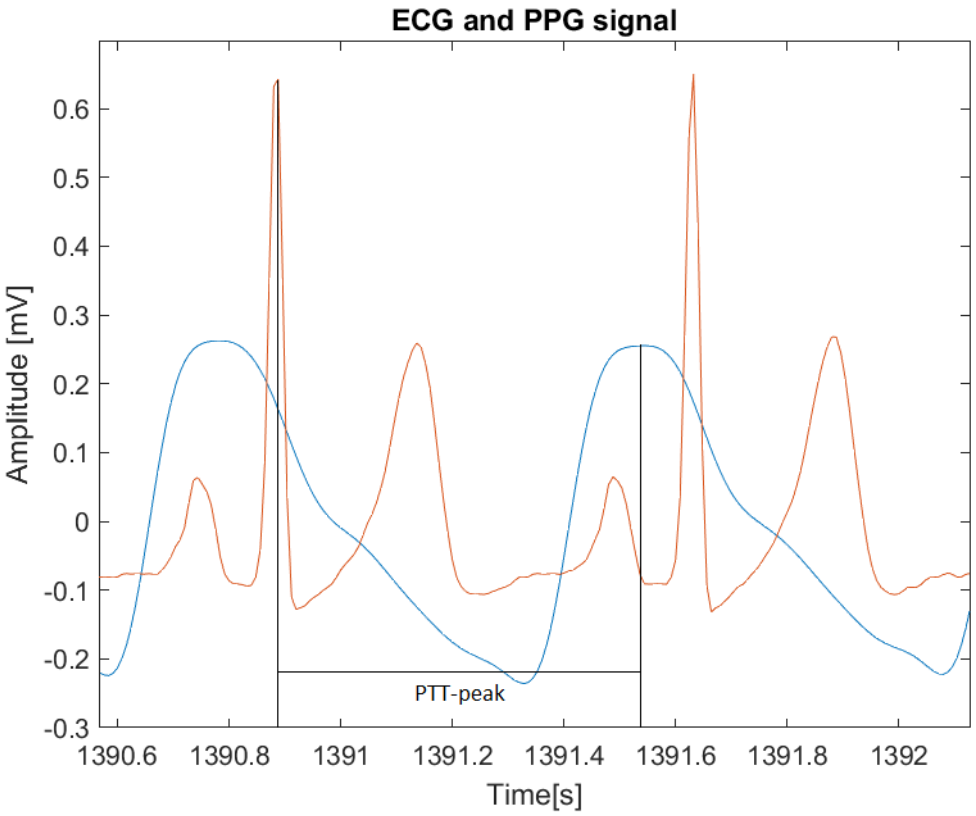

Figure 2.12: An example of PTT computation

4. The fourth and last dataset has been built with all the period samples; but since each period is made up of a different number of samples, it has been necessary to zero-pad all the periods, so as to have one hundred samples each.
   This dataset has been then chosen to carry out our work (Figure 2.13)

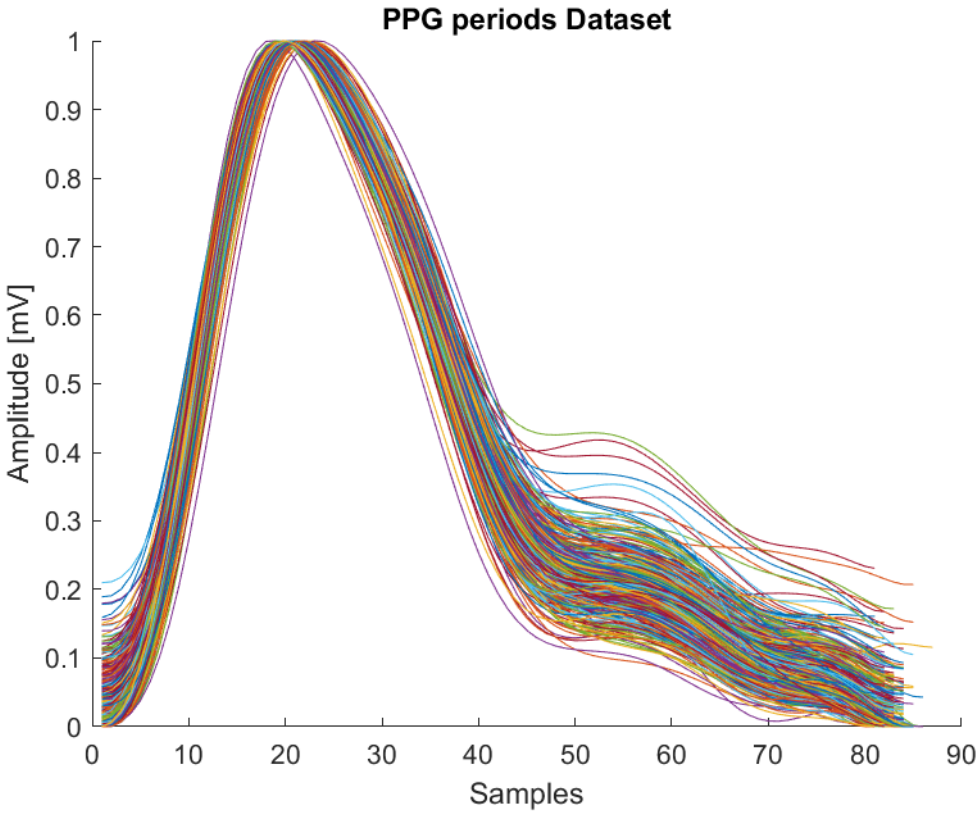

Figure 2.13: A portion of PPG periods dataset

2.3.1 Dataset preparation

Once the datasets have been created in *Matlab R2019a*, they could not be used straight away for our project.

Dataset preparation helps the model to fit the data well. In order to prepare the data and make them more suitable for the model, the feature engineering process has been carried out on *Google Colaboratory*.

Both features and PPG periods datasets need to be converted into pandas DataFrames, in order to be used in *Python*, since our ANN algorithm has been implemented in *Keras*[4]. Pandas DataFrames are data structures containing numeric values organized in rows and columns, in which every column is labeled.

[4]Keras is an open-source library, written in Python, used to build and evaluate ML models, reducing cognitive load.

In the first three DataFrames each column has the corresponding feature's name (Figure 2.14); in the last one the labels represents the sample numbers (Figure 2.15). The last column elements, the SP values, are called *targets*.

	CP	SUT	DT	SW10	SW25	SW33	SW50	SW66	SW75	DW10	DW25	DW33
171927	0.104014	0.584843	-0.139671	0.609685	0.894523	0.517470	0.598149	0.641684	0.821952	0.694760	0.863136	0.351071
58535	0.104014	-0.086274	0.134642	-0.175576	0.043529	0.076011	0.118092	0.110976	0.247252	0.994382	0.457949	0.050927
36141	-0.326651	0.584843	-0.551141	0.217055	0.043529	0.076011	0.118092	0.110976	0.247252	-0.803348	-0.082300	-0.249217
80900	0.391124	-0.757391	0.683268	-0.568207	-0.807464	-0.806906	-0.842023	-0.419731	-0.327448	-0.054294	-0.757611	-0.699434
243995	-0.183096	2.262635	-1.099768	2.572838	2.596510	2.283305	2.518378	2.233805	2.546052	-1.552402	-1.027735	-0.849506
...	...	...	...	...	...	...	...	...	...	...	...	...	...
175638	1.970228	1.255960	1.369051	1.394946	1.320020	1.400388	1.558263	1.172391	1.396652	1.893247	1.808571	1.851793
95816	-0.613761	-0.421832	-0.413985	-0.175576	-0.381968	-0.365448	-0.361966	-0.419731	-0.327448	-0.653537	-1.162797	-0.999578
203245	-0.613761	0.584843	-0.825454	0.609685	0.469026	0.517470	0.598149	0.641684	0.821952	-0.054294	0.457949	0.651216
100879	-0.757316	-0.086274	-0.688298	-0.175576	-0.381968	-0.365448	-0.361966	-0.419731	-0.327448	-1.852024	-1.567984	-1.149650
89256	0.104014	-1.092949	0.546111	-1.353468	-1.232961	-1.248365	-1.322080	-0.950438	-0.902147	-0.204105	-0.892673	-0.699434

249672 rows × 21 columns

Figure 2.14: A portion of the twenty-feature DataFrame on Google Colaboratory notebook

	1	2	3	4	5	6	7	8	9	10	11	12
129074	0.334699	0.334600	0.338038	0.345672	0.358479	0.377811	0.405800	0.443665	0.479705	0.435553	0.238652	0.058449
57485	0.326348	0.330275	0.331479	0.328068	0.317361	0.294756	0.251119	0.166131	-0.005749	-0.303335	-0.539710	-0.593549
61497	0.293532	0.286982	0.281824	0.278816	0.278722	0.282594	0.291560	0.305549	0.314948	0.267847	0.127961	0.009379
1033	0.386881	0.389546	0.387975	0.380139	0.362790	0.330203	0.271303	0.161278	-0.053508	-0.411088	-0.675831	-0.719178
237470	0.383270	0.385221	0.383337	0.376579	0.363275	0.340103	0.299940	0.226078	0.078349	-0.186779	-0.418136	-0.487798
...	...	...	...	...	...	...	...	...	...	...	...	...
89633	0.376138	0.370791	0.368862	0.372375	0.384184	0.408675	0.453214	0.529623	0.645309	0.720696	0.579416	0.380501
181990	0.439605	0.441807	0.440629	0.433990	0.418565	0.388323	0.331054	0.219176	-0.009703	-0.411693	-0.738402	-0.821843
229390	0.711121	0.706933	0.704507	0.705224	0.711109	0.725038	0.751076	0.792440	0.832275	0.748376	0.430384	0.151689
203425	-2.451924	-2.424512	-2.396128	-2.366915	-2.335286	-2.296047	-2.233909	-2.100952	-1.736643	-0.765641	0.563797	1.337048
38063	0.255208	0.250063	0.238164	0.217102	0.182746	0.128033	0.039152	-0.112197	-0.377766	-0.747022	-0.909206	-0.836053

249672 rows × 101 columns

Figure 2.15: A portion of the PPG periods DataFrame on Google Colaboratory notebook

Once the DataFrame has been created, its values have to be scaled, since this helps the model to achieve better performance. Two of the most common scaling methods are normalization and standardization.

Scikit-learn, one of the most useful library for ML in Python [75], used in our project, has two main scalers:

- *MinMaxScaler*: it scales all the values in the range [0,1], [-1,1], [-2,2], according to the chosen range.
- *StandardScaler*: it follows the Standard Normal Distribution, hence it removes the mean (mean = 0) and scales the values to unit variance.

2.3.2 Targets

In our study only periods with 80 mmHg<SP<180 mmHg have been considered. At first the aim of our project, referring to the previous thesis, was to create a SP values classifier, based on the ANN approach. It should have been able to classify the input (features or PPG period) within the right class.
SP values have been divided in six classes:

Class Label	**SP Range**
Hypotension	<90
Desired	90 - 119
Pre-Hypertension	120 - 139
Stage 1 Hypertension	140 - 159
Stage 2 Hypertension	160 - 180

Table 2.1: SP classes

The range width, as shown in Table 2.1, can be 10 mmHg or 30 mmHg. In order to perform the classification task, the target column had to be discretized[5] and then converted into a categorical value.

However, when working with ANN algorithm, it is necessary to encode categorical data. In our work we use the *One-Hot encoding* scheme, in which each bit represents a class. The position of the **1** bit represents the class to which the input belongs.

SP Value	**Label**	**Categorical**	**One-Hot Encoded**
83,7	Hypotension	0	10000
107,4	Desired	1	01000
132,7	Pre-Hypertension	2	00100
151,8	Stage 1 Hypertension	3	00010
173,1	Stage 2 Hypertension	4	00001

Table 2.2: Column target transformation

[5]It is the process of transforming the values in the range in the corresponding class label.

2.3.3 Dataset balancing

Before building the model, we have checked our dataset distribution (Figure 2.16).

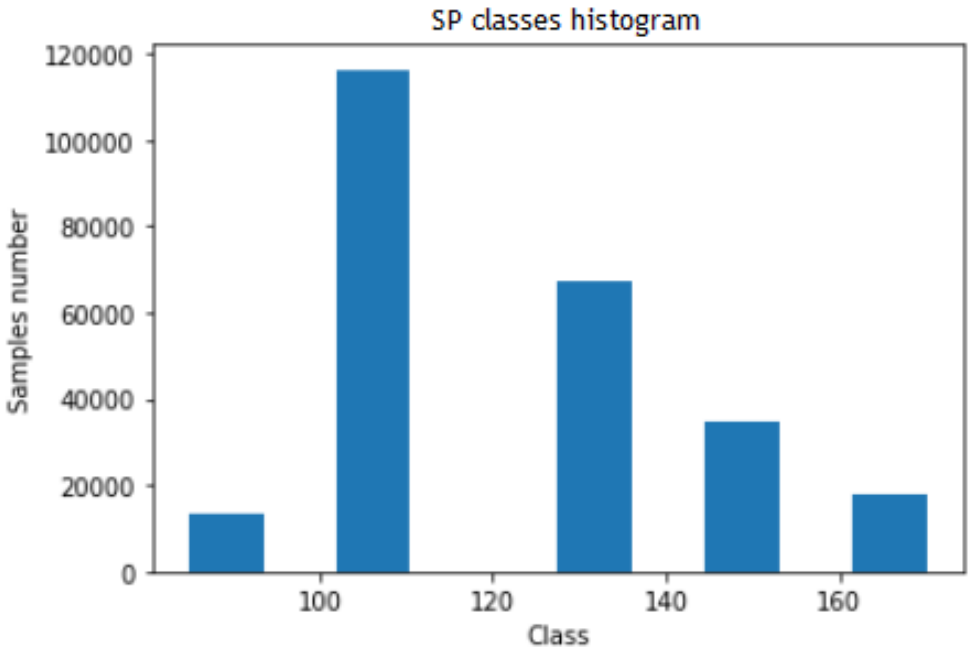

Figure 2.16: Our unbalanced dataset distribution

Even though our dataset is highly unbalanced, it is representative of the reality. However, models trained on an unbalanced dataset have, usually, poor generalization capability. There are few approaches to balance our classes distribution:

- *Undersampling* and *Upweighting* the majority classes;
- *Oversampling* the minority classes.

The *Synthetic Minority Oversampling Technique* (SMOTE) is the most commonly used technique to oversample a dataset, creating synthetical observation of the minority classes. We have applied the SMOTE technique (Figure 2.17) after splitting the dataset into training, validation and test set, in order to have only original samples in the testing and validation sets and all the synthetic samples in the training set.

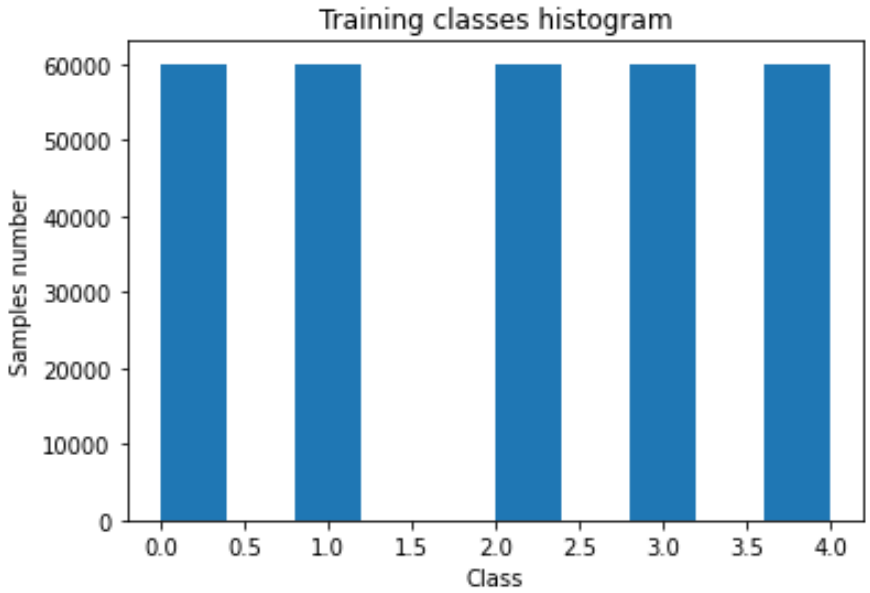

Figure 2.17: Training set distribution after SMOTE strategy has been applied

As mentioned above, to avoid overfitting, we have performed the dataset splitting, using the *train_test_split* function twice. The first time to split the dataset into training (70%) and test (30%) set and the second to split the training set into training (80%) and validation (20%) set.
The resulting data are:

1. **300,000** training set samples;
2. **34,954** validation set samples;
3. **74,902** test set samples.

2.4 ANN classification model

Our ANN model has been developed with *Keras* and trained in Google Colab, in order to take advantage of its computational power.
The way we have chosen to build our Keras model is the easiest one, the *Sequential* [24]; it enables to build models layer-by-layer, each of which can have one input tensor and one output tensor [36]. To create our MLP feed-forward NN, as mentioned in Section 1.5.2, we have used the *.add()* method:

- The first layer is the input layer; its size depends on the input features[6] and represents the input data.
- *Dense* layers have been added as hidden layers; as their name suggests, all their neurons are fully connected to those in the next layer.
- The output layer, since we have used the One-Hot encoding scheme, has five neurons, one per class. Their AF is the *Softmax*, necessary in multi-class classification tasks, which provides, in our case, a 5-element vector, whose sum is 1. Each value represents the probability of the input to belong to the respective class.
- In order to reduce overfitting and to make nodes more robust to the inputs, *Dropout* regularization has been used between layers [27].

[6] 15 input neurons for the first dataset, 19 for the second, 20 for the third and 100 for the last dataset.

Starting from the first model (Figure 2.18), hyperparameter optimization has been implemented.

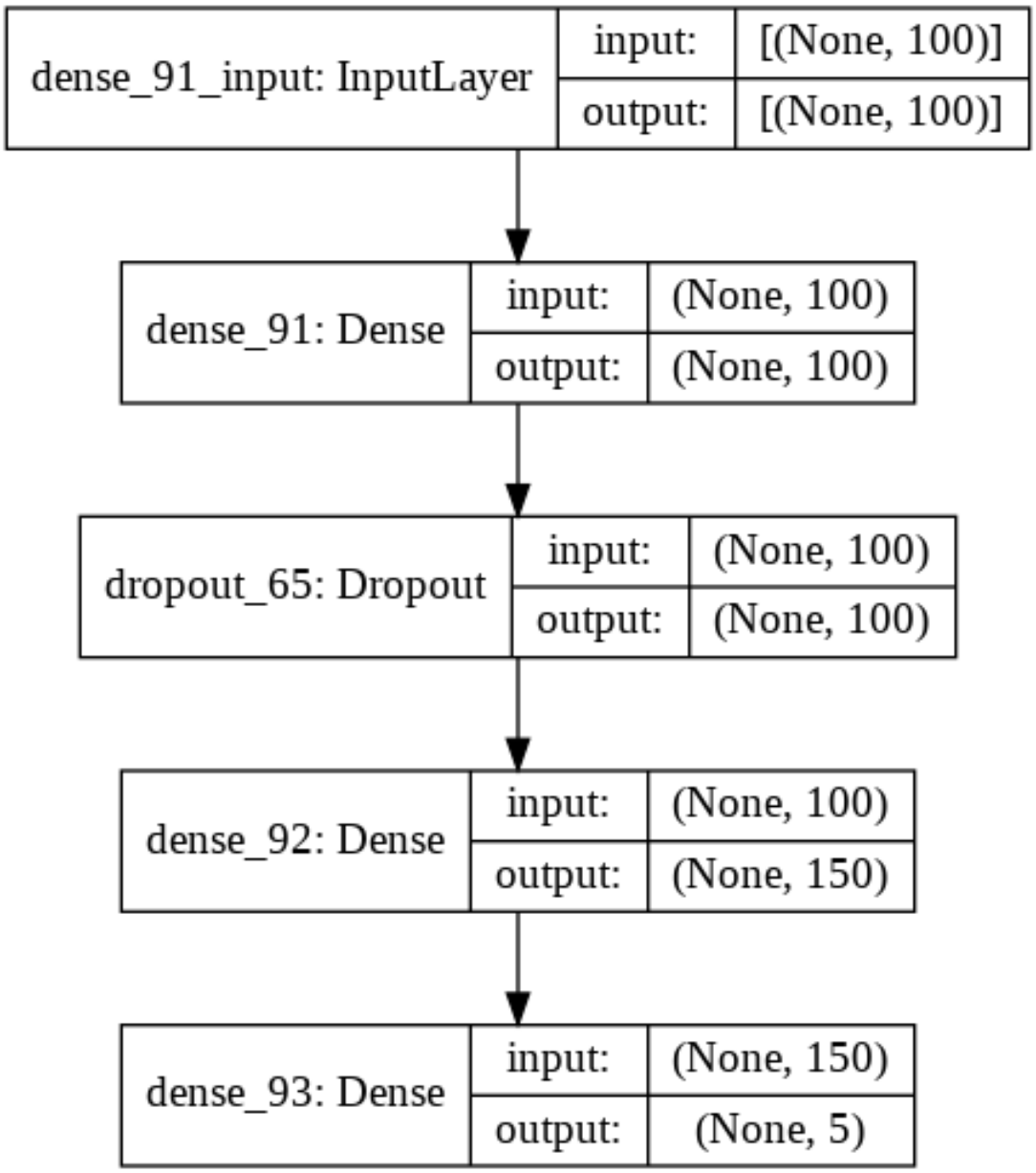

Figure 2.18: First ANN model

The first training has been performed with the model shown above (Figure 2.18) and with the following hyperparameters, inspired by the previous work of thesis [25]:

Parameters	**Values**
Optimization algorithm	Adam
Learning rate	0.007
Initialization mode	glorot_uniform
Activation Function	sigmoid
Dropout rate	0.013
Batch size	1024
Epochs	1500

Table 2.3: Our first model hyperparameters

In order to reduce overfitting, besides using *Dropout*[7], one *Early stopping* has been used. It stops the training process when there is no improvement in model performance.

Although this implementation does not yield the best results, it has been a good starting point for the architecture and parameters tuning process.

2.4.1 Architecture and hyperparameters tuning

The first thing was to explore the prediction performance of deeper and wider architectures. The result has been to improve the predictive power of the model, increasing both depth and width, until overfitting occurs.

Once we have found the right architecture, we have used the *Grid Search* method to tune the model hyperparameters. It compares prediction performance using a provided subset of hyperparameters, building a different model for each possible parameter combination.

[7]It is a technique where units are dropped-out in a random way.

Parameters	Values
Optimization algorithm	Adam, SGD, Nadam, AdaDelta
Learning rate	0.0001, 0.0005, 0.001, 0.005, 0.007, 0.01
Initialization mode	glorot_uniform, glorot_normal, lecun_uniform
Activation Function	sigmoid, softsign, ReLu
Dropout rate	0.05, 0.2, 0.3, 0.5
Batch size	256, 512, 1024, 2048
Epochs	500, 750, 1000

Table 2.4: Hyperparameter subset

2.4.2 Performance metrics for classification problems

In order to calculate the final model classification performance, *Precision*, *Recall* and the *F1-Score* have been used.

Metrics	Formula
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	2* $\frac{Precision*Recall}{Precision+Recall}$

Table 2.5: Performance metrics

The *Confusion Matrix* (CM) represents an intuitive metric used to evaluate algorithm classification accuracy. It is a n x n matrix, where n is the number of target classes. It contains in rows the predicted values and in columns the real ones[8]. All the performance metrics are based on the numbers represented in the CM. Before going into the metrics details, the True positive, False positive and False negative concepts need to be explained:

- *True positive* (TP): it represents the values being correctly predicted in the positive class;

[8]The Scikit-learn convention, used in our project, places predicted values in columns an actual values in rows.

- *False positive* (FP): it represents the values being incorrectly predicted as positive;
- *False negative* (FN): it represents the values being incorrectly predicted as negative;

Precision ranges in [0,1], and represents the portion of examples being predicted as positive, that actually are positive.
Recall is within the range [0,1] as well, and measures the portion of truly positives, predicted as positive. In short, Recall tells us how many positives *we missed* and Precision how many positives *we caught* [72].

In our case, for example, if the Stage 1 Hypertension class has a Precision percentage of 20%, it means that our model tends to label the samples in the Stage 1 Hypertension class even if they belong to other classes. Otherwise, if the Stage 1 Hypertension has a Precision ratio of 94%, it means that the model is able to correctly predict the Stage 1 Hypertension samples in most cases.
A Recall percentage of 25% for the Stage 2 Hypertension, instead, means that most of its samples are predicted as belonging to other classes. A Recall percentage of 91% means that a few samples belonging to Stage 2 Hypertension class are misclassified.
F1-Score represents the Precision and Recall harmonic mean, giving equal importance to both, and it is particularly useful in unbalanced multiclass classification tasks.

In this kind of problems, such as medical data prediction, it is better to classify all the *Desired* cases, as *Hypertensive*, rather than not identifying actual *Hypertensive* patients. This said, what should be minimised are FN cases, so as to maximise Recall percentage. However, a high value of Recall leads to a useless model, if Precision is extremely low.

2.5 ANN regression model

In parallel, we have decided to build a regression model. It provides a continuous output variable, instead of a discrete output variable, as in the classification task.
To switch between models, it was sufficient to change the output layer of the classifier: the number of neurons is 1 with a *linear* AF. The linear AF directly provides the value of the weighted sum of the input.

In order to evaluate the regression predictive model performance, the metrics listed in Section 2.4.2 are not appropriate. In such cases the most widely used are:

- The *Root mean squared error* (RMSE): it measures the square root of the average of the difference between predicted values and the actual values.

$$RMSE = \sqrt{\frac{\sum (y - y')^2}{N}} \tag{2.1}$$

  In Equation 2.1 y represents the actual, y' the predicted value, and N the number of sample points.

- The *Mean absolute error* (MAE): it represents the absolute value of the average of the error[9].

$$MAE = \frac{\sum |y - y'|}{N} \tag{2.2}$$

In our project, in order to evaluate the regression model performance, we have adopted the MAE. It is more intuitive, since its values are expressed in the same unit of measurement of the input, and it is less sensitive to outliers than the RMSE.

[9]The difference between real and predicted value.

2.6 Computational complexity of the model

Since the ultimate goal of this study is to embed the final model in a MCU, it is necessary to evaluate its computational cost in terms of memory capacity required.
In order to do this, the *STM32CubeMX* graphical tool has been used. In STM32CubeMX, the *X-CUBE-AI* software pack and the *STM32L4* firmware package have been installed. The X-CUBE-AI expansion package supports model trained with *Keras, TensorFlow™ Lite, Caffe, ConvNetJs*[10]. The model analysis performed with the AI package installed provides:

1. information about the memory (RAM and ROM) required to store the pre-trained NN in a MCU;
2. the number of Multiply-Accumulate Operation (MAC) needed (CPU load);
3. a list of MCUs that meet the specific requirements in terms of memory;
4. the CPU load and the ROM required by each layer;
5. a 8-bit quantized network in order to reduce the computational cost, the required Flash memory[11] and to improve the inference time. This has been done without compromising the model accuracy [70];
6. an automatic conversion of the pre-trained NN model into a C-compiled optimized code that can run on STMicroelectronics MCUs.

The MCU used as a reference in this project belongs to the *STM32L4+*[12] series. It has up to 2 MB of Flash memory and 640 KB of RAM. In particular the MCU used in this project is the STM32L4R7 and has:

- **640 KB** of RAM;
- **2048 KB** of Flash memory.

Input and output nodes[13], and the activations are stored in the RAM; weights are stored in the ROM.

Once the values are known, in order to embed the ANN model on the MCU, the RAM and the Flash memory required must be lower than the available ones.
If the required memory is larger than the usable one, it is possible to compress the weights with a compression ratio of 4 or 8. However, after this process, it is necessary to evaluate

[10]The most common AI libraries.
[11]It is a type of Electronically Erasable Programmable ROM.
[12]The ultra-low power MCUs family.
[13]They occupy 4 B of the RAM each.

the performance of the obtained model, in order to verify that they are not impaired. Another important parameter to assess is the *inference time*[14]. As the goal of this study is the implementation of a real-time beat-to-beat SP monitoring system, the inference time needs to be shorter than the cardiac cycle duration. The normal heart rate at rest for adults ranges from 60 to 100 bpm. Therefore, if the intention is to obtain an SP value for each PPG period, the inference time should be lower than 600 ms.

[14]It is the time taken by the model to calculate the output, starting from live data points.

2.7 Data acquisition

In order to test the best classification and regression models, PPG signal and NIBP have been acquired at STMicroelectronics S.r.l (Agrate Brianza) on 8 healthy subjects (5 male and 3 female), during resting position.

BP has been measured simultaneously using a sphygmomanometer wrapped around the left upper arm, and noted down on an Excel file, with the acquisition instant. PPG signal has been acquired from right hand finger using *MORFEA3* (Figure 2.19) at Fs of 62.5 Hz, using a spectrometer. It uses a white LED as light source, instead of a photodiode, and it is able to separate the beam of white light into its components, in order to measure green wavelength contribution.

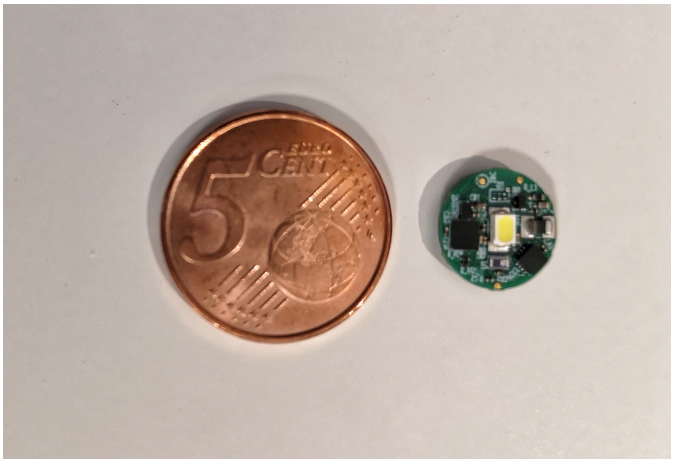

Figure 2.19: MORFEA3 board compared to a 5 euro cent coin

2.7.1 Acquisition system

MORFEA3 is a wearable device, build with a printed circuit board (PCB), with a 11 mm diameter and a 0.8 mm thickness, without considering the electronic components, and a 3 mm thickness with the electronic components. It embeds a certain number of sensors, managed by a MCU:

- A *3-axis accelerometer* to detect the activity, with a resolution of 4mg/LSB[15]. It uses ± 2g as full scale. It communicates with the MCU via I^2C.
- A *negative temperature coefficient* (NTC) Thermistor to monitor the temperature in the range [-40 °C; 150 °C], with a resolution of 0.1 °C. The thermistor resistance is read by the ADC[16].

[15]Least Significant Bit.
[16]Analog to Digital Converter.

- Two little *spectrometers* to detect the PPG signal at different wavelengths (from 360 to 1050 nm). They use the I^2C connection to communicate with the MCU.

The MORFEA3 device interfaces with the outside with a Bluetooth connection.

In order to acquire the signals, MORFEA3 is placed within a suitable case, where it is possible to insert the right hand index fingertip, so as to keep the finger-device contact stable.
After wearing the sphygmomanometer cuff, turning on the device and positioning the right hand finger inside the case, the *Bio2Bit* Windows application has been launched (Figure 2.20).

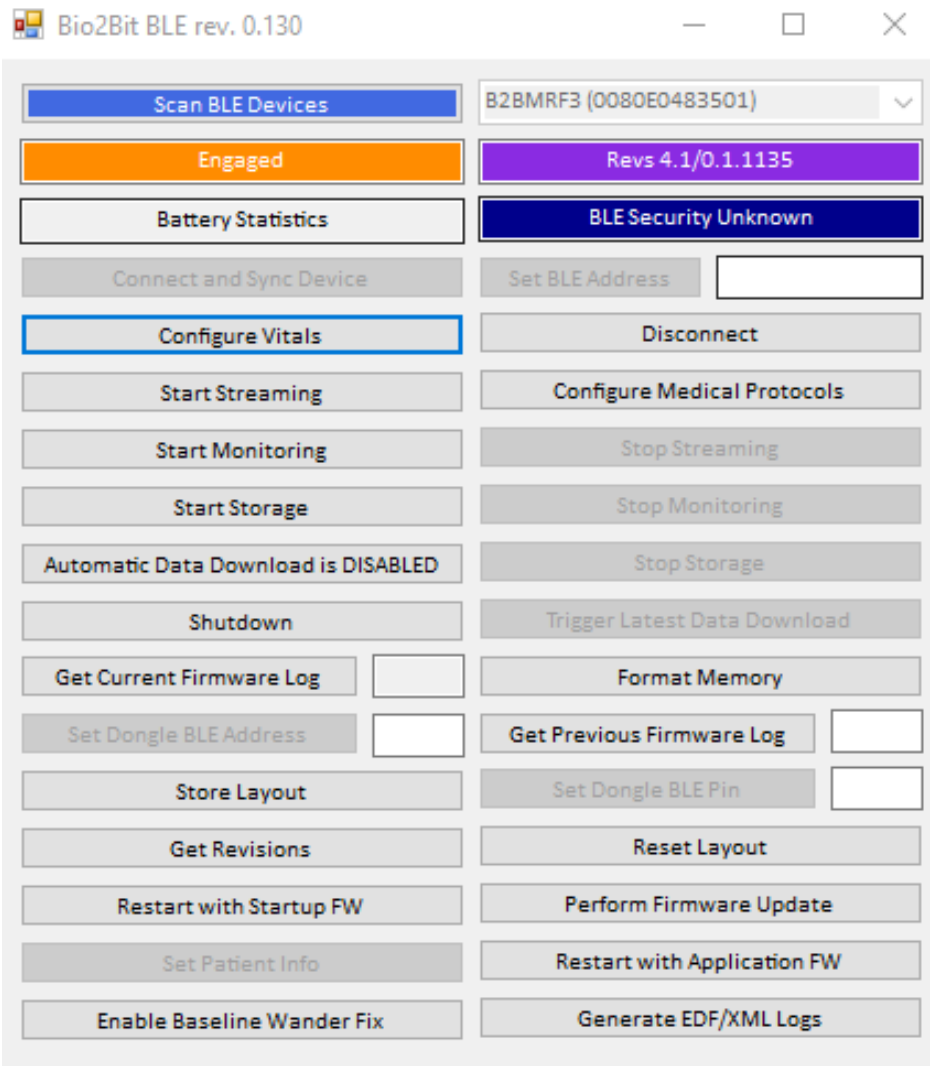

Figure 2.20: *Bio2Bit* Windows application screen, once the MORFEA3 board has been connected

Then, the following steps have been executed:

1. The *Scanning* of the Bluetooth Low Energy (BLE) devices nearby.
2. The *Selection* of the desired device from the dropdown menu.
3. The *Connection* and the *Synchronization* of the selected device.
4. The *Configuration* of the parameters, opening the *Configure Vitals* section and setting the PPG mode[17].
5. The *Start* of the acquisition, pressing the *Start Streaming* button. Simultaneously data are logged in binary format.

At the same time, as said before, the SP has been acquired using an automatic sphygmomanometer, carrying out a record approximately every minute.

2.7.2 Data pre-processing

In the following figures an example of raw (Figure 2.21) and filtered (Figure 2.22) PPG signal acquired with the MORFEA3 device are represented.

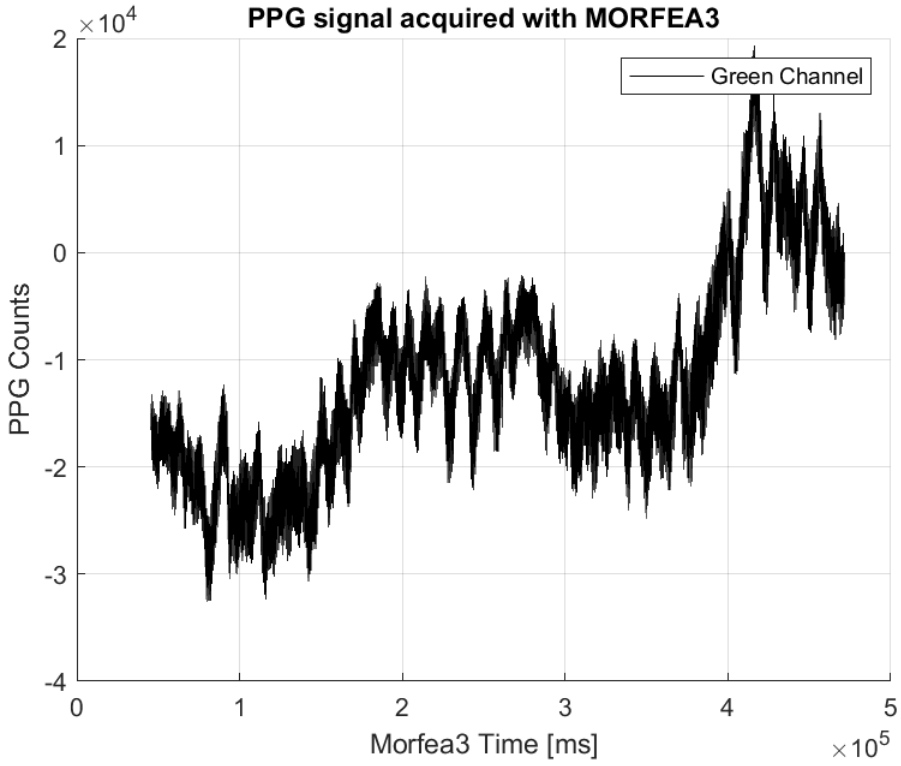

Figure 2.21: Raw PPG MORFEA3 Signal

[17]In order to save the option, it is necessary to click again the Configure Vitals button.

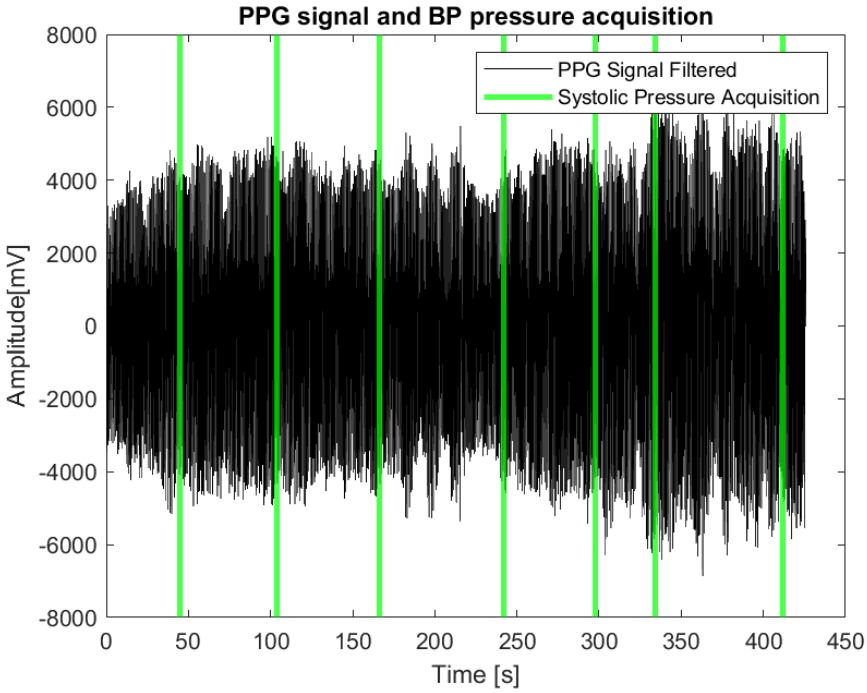

Figure 2.22: Filtered PPG MORFEA3 Signal with, superimposed, the SP acquisition instants

The signal has been filtered with the same 4^{th} order Butterworth filter used in the MIMIC III PPG signal pre-processing.
Then, in order to make the MORFEA3 PPG signal suitable for the chosen ANN model, it must be resampled at 125 Hz. This is because the number of input neurons of a model cannot be varied, and if we had segmented our signal without resampling, we would have obtained a shorter average length of the periods.
Once it has been resampled, the PPG signal has been segmented into periods[18], each associated to an SP value.

[18]Carrying out all the steps executed in the MIMIC PPG signal pre-processing phase and reported in Section 2.2.

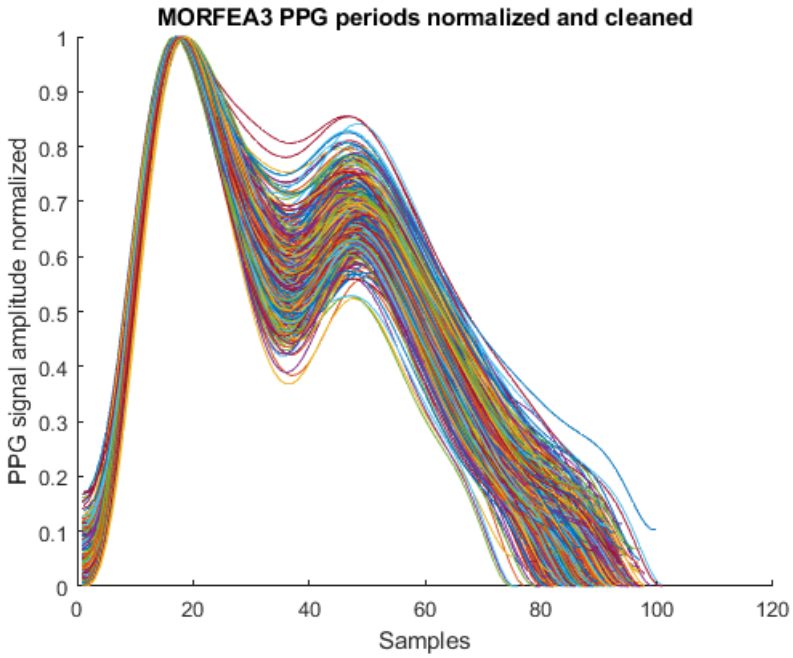

Figure 2.23: The MORFEA3 PPG periods normalized

Our reference is not a continuous signal, as it was in the case of ABP signal. Therefore, in order to obtain a number of SP values equal to the number of PPG periods acquired, two consecutive SP records have been averaged between two acquisition instants.

At the end of our work the dataset used to test the final model was made up as follow:

	1	2	3	4	5	6	7	8	9	10	11	12
2543	-0.583573	-0.771558	-0.988395	-1.110920	-1.081815	-0.957833	-0.793410	-0.601496	-0.367630	-0.056604	0.375119	0.868738
3783	0.029442	-0.241706	-0.461510	-0.536666	-0.457728	-0.299578	-0.114846	0.082996	0.295492	0.523801	0.745656	0.863492
1563	0.226694	0.064790	-0.127848	-0.343975	-0.534737	-0.682603	-0.798612	-0.890654	-0.950269	-0.940382	-0.766235	-0.299739
900	-0.583573	-0.537088	-0.399850	-0.197241	-0.012376	0.111564	0.187366	0.242780	0.306522	0.413207	0.601169	0.848654
3174	1.400370	1.124428	0.692028	0.225369	-0.093162	-0.215618	-0.191897	-0.067752	0.137143	0.417125	0.745656	0.975603
...	...	...	...	...	...	...	...	...	...	...	...	...
3046	0.855481	0.730874	0.362235	-0.157558	-0.583483	-0.806646	-0.865965	-0.822535	-0.721906	-0.596515	-0.461134	-0.292545
1725	-0.583573	-0.885852	-1.231088	-1.460845	-1.493410	-1.403609	-1.270354	-1.119513	-0.939425	-0.686319	-0.279501	0.313277
4079	-0.225694	-0.016464	0.291456	0.623297	0.856128	0.965564	0.988119	0.950992	0.858874	0.694265	0.420197	0.042740
2254	-0.189710	0.058281	0.234187	0.209278	0.012455	-0.231379	-0.444122	-0.594805	-0.659744	-0.595100	-0.318233	0.216603
2915	1.239423	1.024161	0.715365	0.368558	0.119902	0.025172	0.056096	0.173706	0.348867	0.555260	0.745656	0.808785

6460 rows × 101 columns

Figure 2.24: The MORFEA3 PPG periods normalized dataset

It was composed of 6,460 periods, with one hundred samples each, and, in the last column, the corresponding target.
After discretizing the target column, we have realised that low and high SP values are missing in our dataset.

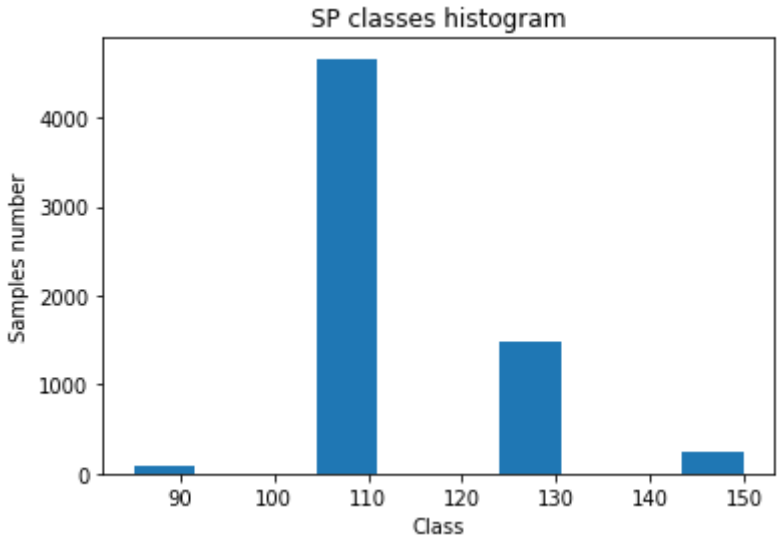

Figure 2.25: The MORFEA3 dataset SP values distribution

Chapter 3

Results

3.1 Dataset selection

In order to find the most suitable network architecture, we have made assumptions on the appropriate number of hidden layers and nodes. Our aim was to reach a good trade off between complexity and performance.
First we have tried to use a basic NN model, so as to increase the number of hidden nodes in every step, until the best performance has been reached.

The first model used, as mentioned in Section 2.4, has been inspired by [25] and is represented in Figure 2.18. The other hyperparameters are: learning rate equal to 0.007, sigmoid as AF, Adam as optimization algorithm, glorot_uniform as initialization mode, Dropout rate equal to 0.2, batch size equal to 512 and number of epochs equal to 500.

The first task we have done was to test the first model on the four datasets described in Section 2.3, in order to select the most performing one.

1. Working on the test set made up of the fifteen features, an average accuracy of **85.41%** has been achieved. The performance, in terms of *Recall*, achieved on the fifteen-feature dataset is shown in Table 3.1.

		Predicted Values				
		Hypotension	Desired	Pre-Hypertension	Stage 1 Hypertension	Stage 2 Hypertension
Actual Values	Hypotension	**81.40 %**	18.48 %	0.12 %	0 %	0 %
	Desired	3.79 %	**88.62 %**	7.34 %	0.20 %	0.06 %
	Pre-Hypertension	0.02 %	7.43 %	**82.52 %**	9.89 %	0.14 %
	Stage 1 Hypertension	0.02 %	0.6 %	11.04 %	**80.43 %**	7.91 %
	Stage 2 Hypertension	0.02 %	0.17 %	0.32 %	14.26 %	**85.23 %**

Table 3.1: Fifteen-feature dataset CM

2. On the nineteen-feature dataset, we have obtained an average accuracy equal to **86.92%**. The *Recall* rate is reported in Table 3.2.

		Predicted Values				
		Hypotension	Desired	Pre-Hypertension	Stage 1 Hypertension	Stage 2 Hypertension
Actual Values	Hypotension	**85.19 %**	14.7 %	0.09 %	0 %	0 %
	Desired	3.52 %	**89.83 %**	6.41 %	0.22 %	0.02 %
	Pre-Hypertension	0 %	6.71 %	**84.56 %**	8.66 %	0.07 %
	Stage 1 Hypertension	0.02 %	0.27 %	10.65 %	**81.15 %**	7.99 %
	Stage 2 Hypertension	0 %	0.09 %	0.38 %	12.23 %	**87.29 %**

Table 3.2: Nineteen-feature dataset CM

3. The third dataset, composed of the previous nineteen features plus the PTT, has led to an average accuracy of **87.16%** and to the CM reported in Table 3.3.

4. The last dataset, which consists of one hundred points for each PPG periods, has led to the best result: an average accuracy over the test set of **88.43%** and the CM in Table 3.4.

Given the results shown in Tables 3.1, 3.2, 3.3 and 3.4, we have decided to continue our work using the last dataset, made up of all the samples composing the PPG period[1].
Even if a higher number of input neurons has been required, as well as a higher RAM memory[2], the feature extraction process has been removed.

[1]Where required, zero-padding has been performed in order to have the same number of points to give in input to the ANN.

[2]The input and output neurons are stored in the RAM memory and require 4 Bytes each.

		Predicted Values				
		Hypotension	Desired	Pre-Hypertension	Stage 1 Hypertension	Stage 2 Hypertension
Actual Values	Hypotension	**84.41 %**	15.53 %	0.06 %	0 %	0 %
	Desired	3.41 %	**90.26 %**	6.19 %	0.11 %	0.02 %
	Pre-Hypertension	0.01 %	6.67 %	**85.14 %**	8.11 %	0.07 %
	Stage 1 Hypertension	0 %	0.14 %	11.82 %	**80.51 %**	7.52 %
	Stage 2 Hypertension	0 %	0.17 %	0.25 %	12.68 %	**86.88 %**

Table 3.3: Twenty-feature dataset CM

		Predicted Values				
		Hypotension	Desired	Pre-Hypertension	Stage 1 Hypertension	Stage 2 Hypertension
Actual Values	Hypotension	**87.19 %**	12.76 %	0.06 %	0 %	0 %
	Desired	2.7 %	**92.50 %**	4.67 %	0.1 %	0.03 %
	Pre-Hypertension	0 %	7.89 %	**83.77 %**	8.34 %	0 %
	Stage 1 Hypertension	0 %	0.39 %	8.89 %	**82.99 %**	7.74 %
	Stage 2 Hypertension	0 %	0.02 %	0.12 %	9.19 %	**90.67 %**

Table 3.4: One hundred-sample dataset CM

3.2 Model selection

In order to improve the model performance, we have started increasing the number of hidden layers.

	Model Architecture	Average Accuracy	Average Loss
1	[10]	0.7974	0.4797
2	[100]	0.8438	0.3587
3	[300]	0.8498	0.3395
4	[100 40]	0.8575	0.3234
5	[100 100]	0.8622	0.3188
6	[100 40 80]	0.8651	0.3056
7	[100 80 100]	0.8680	0.3018
8	[100 80 100 60]	0.8671	0.3004
9	[100 80 100 60 80 100]	0.8581	0.3152
10	[200 180 200 160 180 200]	0.8688	0.3006
11	[100 80 90 60 80 50 80]	0.8630	0.3137
12	[100 80 90 60 80 50 80 100 60]	0.8576	0.3239
13	[400 200 300 200 100 150 180 200 160]	0.8695	0.2992

Table 3.5: Performance of different models in terms of average accuracy and average loss

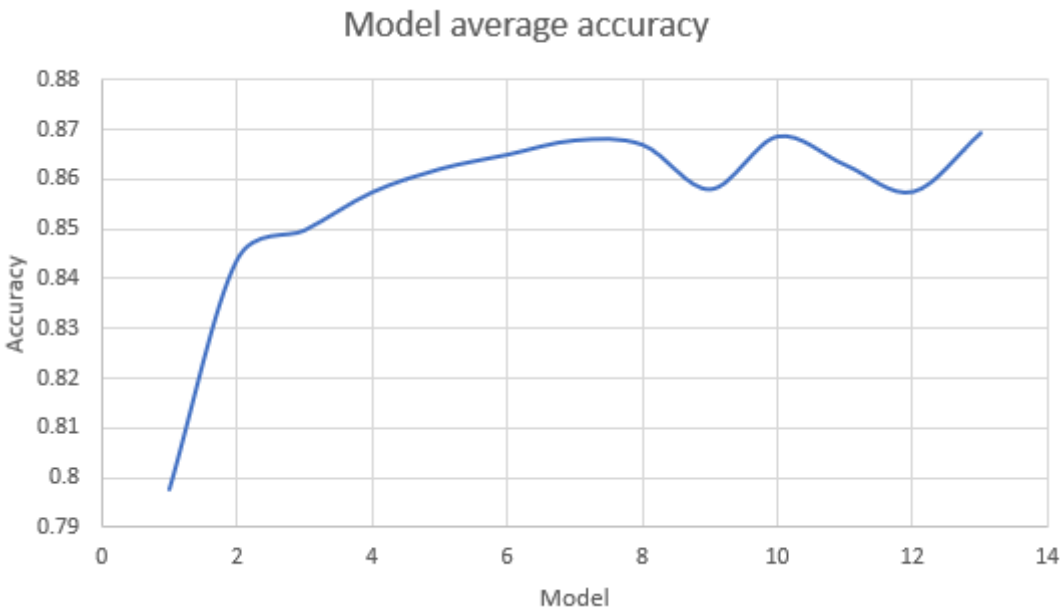

Figure 3.1: Accuracy according to the different model architectures

As it can be seen in Figure 3.1, starting from the simplest models, up to the analysis of more complex models, the accuracy over the test set initially improves, until it reaches the highest accuracy achievable, and then decreases, probably due to the overfitting phenomenon.

Figure 3.2: Loss according to the different model architectures

Similarly, the average loss (Figure 3.2) decreases until it reaches the minimum value, and then increases. Both maximum average accuracy and minimum loss have been obtained with the model n.8 (Table 3.5).

As anticipated in Section 3.1, our aim was to find a good compromise between accuracy and the amount of memory occupied by the model.

	Model Architecture	**Flash memory**	**RAM**
1	[10]	4.16 KB	480 B
2	[100]	41.43 KB	840 B
3	[300]	124.24 KB	1.60 KB
4	[100 40]	56.04 KB	980 B
5	[100 100]	80.88 KB	1.19 KB
6	[100 40 80]	69.63 KB	980 B
7	[100 80 100]	104.63 KB	1.11 KB
8	[100 80 100 60]	127.52 KB	1.11 KB
9	[100 80 100 60 80 100]	179 KB	1.11 KB
10	[200 180 200 160 180 200]	745 KB	1.89 KB
11	[100 80 90 60 80 50 80]	173 KB	1.11 KB
12	[100 80 90 60 80 50 80 100 60]	228.14 KB	1.11 KB
13	[400 200 300 200 100 150 180 200 160]	1.42 MB	2.75 KB

Table 3.6: Models complexity in terms of flash memory and RAM

Comparing the results of the model n.8 in Table 3.6 with the STM32L4R7 MCU memories (reported in Section 2.6), the values are suitable to embed the model into the STMicroelectronics MCU.

3.3 Hyperparameters tuning

Once the network architecture with the best generalization properties has been identified, the hyperparameters have been tuned and the model performance over the test set[3] has been evaluated. The various models explored have been created using the Grid search method, with the purpose of finding the optimal combination of hyperparameters.

3.3.1 Classification model

The first model considered has been the classification model, and its first investigated hyperparameter has been the AF.

In order to evaluate the model accuracy, the other hyperparameters have been manually set[4].

Activation Function	**Average Accuracy**
Sigmoid	**88.75 %**
Hardsigmoid	**88.61 %**
Tanh	**87.97 %**
Softsign	**88.02 %**
Rectified Linear Unit	**88.47 %**

Table 3.7: AFs performance

We have obtained the highest average accuracy value with the Sigmoid AF. ReLu, however, has been further investigated with deeper ANN architecture.

The second hyperparameter tuned has been the optimization algorithm:

Optimization Algorithm	**Average Accuracy**
Adam	**88.96 %**
SGD	**78.89 %**
Nadam	**88.75 %**
Adadelta	**65.28 %**
Adagrad	**71.92 %**
RMSprop	**88.11 %**

Table 3.8: Optimization algorithms performance

[3]The 30 % of the starting training set of the MIMIC III database.

[4]Learning rate is equal to 0.007, the optimization algorithm is the Adam, the initialization mode is the glorot_uniform, the dropout rate is equal to 0.2 and the batch size is equal to 512.

From the results, it can be appreciated that the Adadelta and Adagrad optimizers are not suited to our dataset. The Adam and Nadam optimization algorithms have yielded better results; therefore, we have decided to further investigate them, by varying the learning rate.

Learning rate	**Adam Average Accuracy**	**Nadam Average Accuracy**
0.0001	**84.31 %**	**85.31 %**
0.0005	**86.25 %**	**86.82 %**
0.001	**86.67 %**	**87.48 %**
0.005	**88.01 %**	**88.01 %**
0.01	**87.97 %**	**88.02 %**
0.006	**89.14 %**	**89.12 %**
0.009	**89.04 %**	**89.35 %**
0.007	**89.05 %**	**88.96 %**

Table 3.9: Learning rate performance

The best performance has been achieved with a learning rate equal to 0.009.

Then we have varied the training set batch size, using a test set batch size of 32:

Batch size	**Average Accuracy**
256	**88.24 %**
512	**89.58 %**
1024	**89.67 %**
2048	**88.97 %**

Table 3.10: Batch size performance

Once we have found the best batch size, dropout regularization has been explored in order to reduce overfitting.

Dropout rate	**Average Accuracy**
0.05	**88.60 %**
0.2	**89.35 %**
0.3	**89.01 %**
0.5	**88.37 %**

Table 3.11: Dropout rate performance

Looking at the difference between the average accuracy reached over the training set (**91.73%**) and the accuracy reached over the test set, we have been able to state that using a dropout rate of 0.05, overfitting has occurred.

The last hyperparameter tested has been the initialization mode:

Initialization mode	**Average Accuracy**
glorot_uniform	**89.45 %**
glorot_normal	**89.00 %**
lecun_uniform	**88.88 %**

Table 3.12: Initialization mode performance

At the end of the Grid search hyperparameter optimization, our final classification model was composed as follows:

- Architecture: [100 80 100 60];
- Activation function: Sigmoid;
- Optimization algorithm: Nadam;
- Learning rate: 0.009;
- Batch size: 1024;
- Dropout rate: 0.2;
- Initialization mode: glorot_uniform;

3.3.2 Regression model

In our everyday life dataset, SP values in the range of Hypotension, Stage 1 Hypertension and Stage 2 Hypertension have been not represented, so it has been difficult to evaluate the classification model performance on our dataset. Hence, we have developed the idea of implementing a regression algorithm so as to predict the SP numerical value. The regression model architecture used has been obtained from the final classification model, by varying the output layer (Section 2.5).

However the best hyperparameter combination depends on the task and on the dataset; another hyperparameter tuning has been performed.
The best performance in terms of MAE, on the MIMIC III test set, has been achieved with:

- Architecture: [100 80 100 60];
- Activation function: Softsign;
- Optimization algorithm: SGD;
- Learning rate: 0.007;
- Batch size: 1024;
- Dropout rate: 0.2;
- Initialization mode: glorot_uniform;

3.4 Models results

- The Classification model performance, expressed as a percentage, on the MIMIC III test set is shown in the following CM (Table 3.13).

		Predicted Values				
		Hypotension	Desired	Pre-Hypertension	Stage 1 Hypertension	Stage 2 Hypertension
Actual Values	Hypotension	**87.15 %**	12.62 %	0.14 %	0.08 %	0 %
	Desired	2.7 %	**91.84 %**	5.35 %	0.07 %	0.03 %
	Pre-Hypertension	0 %	6.93 %	**86.9 %**	6.19 %	0 %
	Stage 1 Hypertension	0 %	0.23 %	9.05 %	**85.47 %**	6.25 %
	Stage 2 Hypertension	0 %	0.04 %	0.21 %	9.19 %	**90.55 %**

Table 3.13: CM obtained with the final classification model

Its computational cost, obtained using the STM32Cube.AI software pack, is presented in Table 3.14.

	No compr	**Compr by 4**	**Compr by 8**
Input	400 B	400 B	400 B
Output	20 B	20 B	20B
MAC	35,775	35,775	35,775
Weights	127.52 KB	37.77 KB	18.39 KB
Activations	720 B	720 B	720 B
Flash (total)	127.52 KB	37.77 (-70.38%) KB	18.39 KB (-85.58%)
RAM (total)	1.11 KB	1.11 KB	1.11 KB
Cross correlation accuracy	100%	100%	90%

Table 3.14: The computational cost of the ANN classification final model in terms of MACs, flash and RAM, and the accuracy in reproducing original model performance in the cases of no compression (no compr), compression by 4 (compr by 4) and compression by 8 (compr by 8)

Elapsed time	Original	No compr	Compr by 4	Compr by 8
Running model	0.224 s	0.120 s	0.018 s	0.108
Validate (10 samples)		3.16 s	3.05 s	2.71 s
Single inference		0.316 s	0.305 s	0.271 s

Table 3.15: Validation results of inference time. The elapsed time of the single inference time is lower than 600 ms (Section 2.6)

However this final classification model has shown low performance on our everyday-life dataset.

- The Regression model performance on the MIMIC III test set and on the MORFEA3 dataset is presented in Table 3.16; the model computational cost is resumed in Table 3.17.

	MIMIC III Test set	**MORFEA3** Test set
MAE	2.99 mmHg	3.85 mmHg
SD	3.37 mmHg	4.29 mmHg

Table 3.16: The regression model results, in terms of MAE over the MIMIC III and the MORFEA3 test set

	No compr	**Compr by 4**	**Compr by 8**
Input	400 B	400 B	400 B
Output	4 B	4 B	4 B
MAC	47,500	47,500	47,500
Weights	177.43 KB	52.18 KB	24.68 KB
Activations	720 B	720 B	720 B
Flash (total)	177.43 KB	52.18 KB (-70,59%)	24.68 KB (-86.09%)
RAM (total)	1.10 KB	1.10 KB	1.10 KB
MAE	0.00	0.13	1.53

Table 3.17: The computational cost of the ANN regression final model in terms of MACs, flash and RAM, and the precision in reproducing original model performance in the cases of no compression (no compr), compression by 4 (compr by 4) and compression by 8 (compr by 8)

Elapsed time	Original	No compr	Compr by 4	Compr by 8
Running model	0.218 s	0.118 s	0.109 s	0.105
Validate (10 samples)		2.47 s	2.71 s	2.48 s
Single inference		0.247 s	0.271 s	0.248 s

Table 3.18: Validation results of inference time. The elapsed time of the single inference time is lower than 600 ms (Section 2.6)

Figure 3.3 and Figure 3.4 show, respectively, the real SP values and the regression model predicted SP values on the MIMIC III test set and on the MORFEA3 test set.

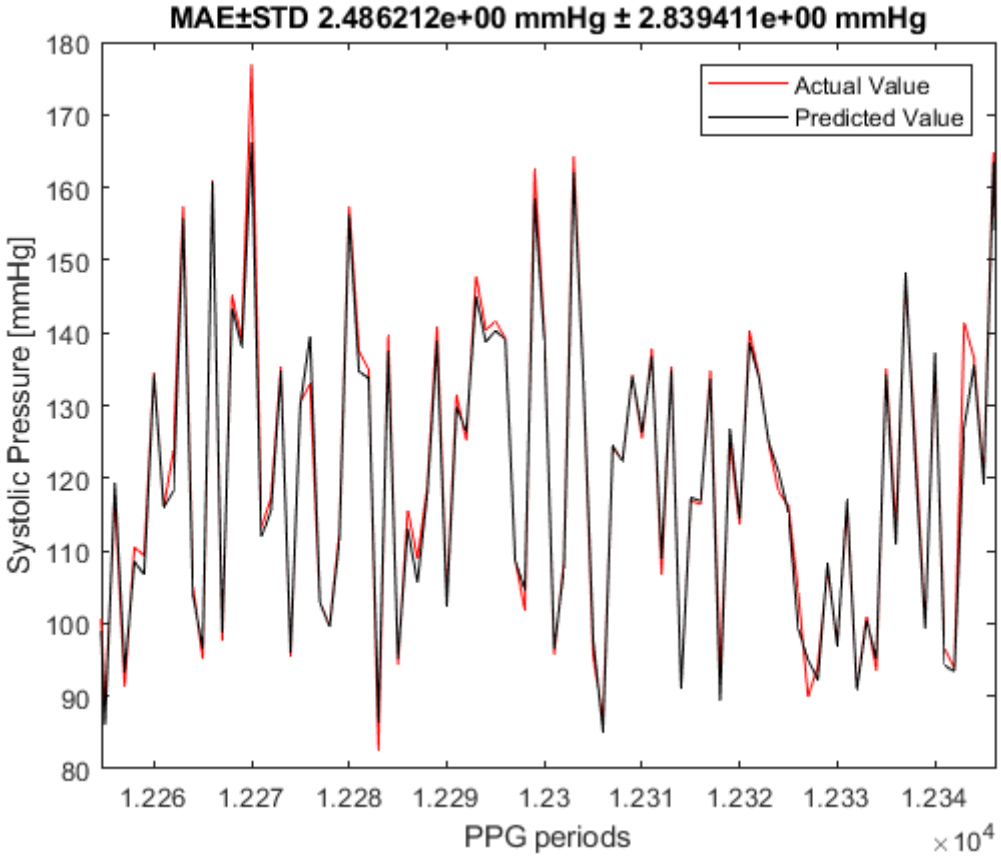

Figure 3.3: An example of actual and predicted SP values obtained on the MIMIC III test set

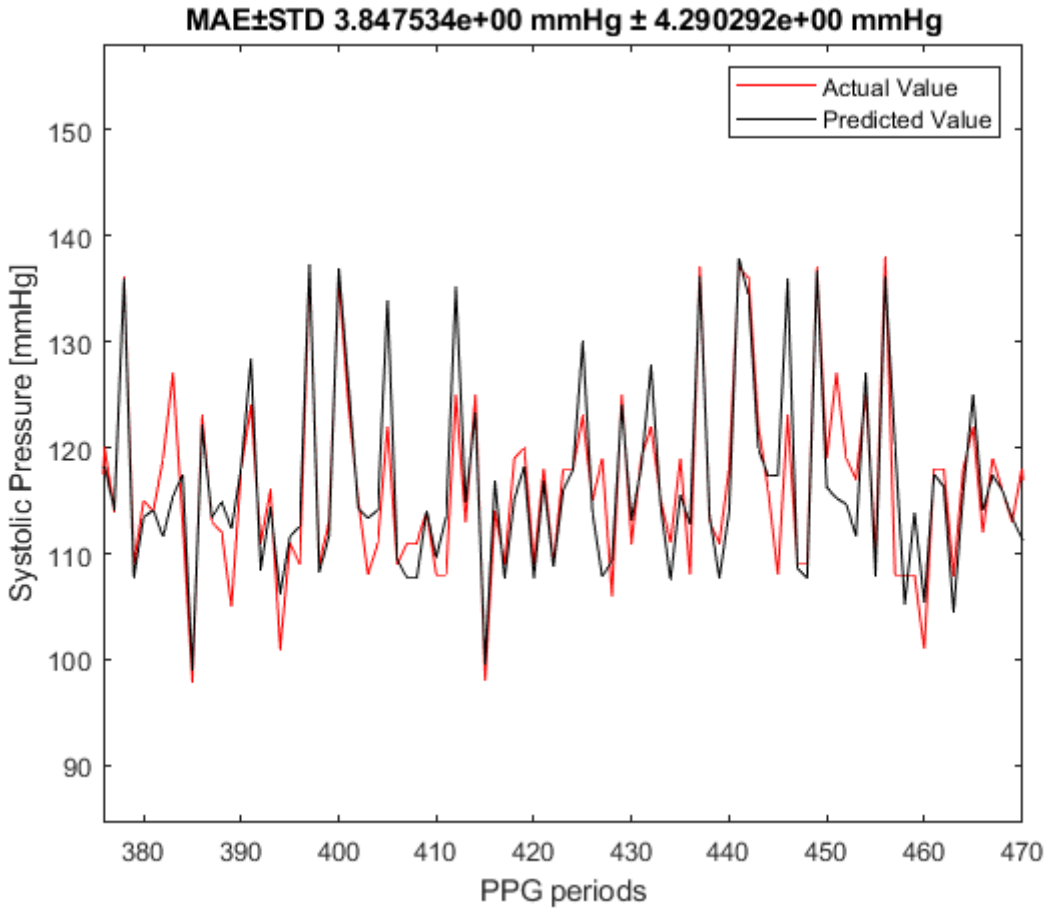

Figure 3.4: An example of actual and predicted SP values obtained on the MORFEA3 test set

3.5 Discussion

1. Given the final classification ANN model results, reported in Table 3.13, we can state that the performance is a good starting point for continuous blood pressure monitoring.

   - The most critical class, *Stage 2 Hypertension*, has been correctly classified in the **90.55%** of the actual cases; however, almost all the other instances have been classified in the *Stage 1 Hypertension*, which still generates an alert status.
   - The *Stage 1 Hypertension* instances, instead, have been predicted as *Pre-Hypertension* in the **6.19%** of the cases. These events are considered FN, and can lead to a no-alarm status, whereas they should in fact raise an alarm. The **9.19%** of *Stage 1 Hypertension* cases have been classified in the *Stage 2 Hypertension* class; this condition (FP elements) is less dangerous than the previous one. It is better to classify the input PPG period at a more critical stage rather than at a less critical one.
   - The highest percentage of *Pre-Hypertension* misclassification has been predicted as *Stage 1 Hypertension*; as mentioned above, this represents a preferable approximation.
   - The *Desired* classification yields the highest accuracy; the **91.84%** of the cases have been correctly classified.
   - The **87.15%** of the *Hypotension* cases have been correctly identified.

   However, the approximation system inherent in this model does not differentiate between minimal and large deviations from the actual SP value[5].
   As a consequence we have chosen to use a regression model.

2. The results obtained with the regression model, presented in Table 3.16, have been shown more in detail in the following table (Table 3.19).

	Error <5 mmHg	**Error <10 mmHg**	**Error <15 mmHg**
MIMIC III Test set	83.48 %	96.34 %	98.88 %
MORFEA3 Test set	74.58 %	90.55 %	97.27 %

Table 3.19: Percentage of absolute difference between actual and predicted SP values (mmHg)

[5]For example both a deviation of 2 mmHg and one of 35 mmHg will be similarly misclassified by this classification model.

MIMIC III Test set consists of 74,902 PPG periods.

- 83.48% of SP values have been predicted with an absolute error lower than 5 mmHg;
- 96.34% have been predicted with an absolute error lower than 10 mmHg;
- 98.88% hev been predicted with an absolute error lower than 15 mmHg.

MORFEA3 Test set consists of 6,460 instances.

- 74.58% of SP values have been predicted with an absolute error lower than 5 mmHg;
- 90.55% have been predicted with an absolute error lower than 10 mmHg;
- 97.27% have been predicted with an absolute error lower than 15 mmHg.

Chapter 4

Conclusions and Future Works

4.1 Conclusions

In recent decades there has been a strong development in TLM and wearable systems. These allow continuous patients remote monitoring, without restricting patient's freedom of movement, and communication at distance between patient and medical staff.

In particular, our study aims to continuous monitoring blood pressure, which is highly desirable in hypertension patients, since high blood pressure is considered as the main cause of CVDs.
In literature photoplethysmography is considered a promising candidate for non-invasive and continuous BP monitoring [43].

In this project the MORFEA3 device has been used to acquire, among other signals[1], the PPG signal in order to provide a real-time SP estimation, using an Artificial Neural Network.
Two ANN models have been developed in this project, but the regression model, with one hundred input neurons and one output neuron, has yielded more stable results; it allows to evaluate the difference between actual SP values and predicted SP values.
This model has the advantage of predicting the SP numerical value, instead of classifying the input PPG period in one of the five classes in which we have divided the SP range considered.

The results obtained on the MORFEA3 test set, in terms of MAE (3.85 mmHg $\pm$ 4.29 mmHg), satisfy the Association for the Advancement of Medical Instrumentation[2] (AAMI) and the British Hypertension Society[3] (BHS) standards [71], as evidenced in Section 3.5 in tables 3.16 and 3.19.

However, in the AAMI and BHS protocols for General Population studies[4] 85 subjects are required, whereas in our work only 8 subjects have been tested.

[1] Accelerometric and temperature signals.

[2] MAE $\leq$5 mmHg with SD$\leq$8 mmHg for SP and BP.

[3] Absolute difference between standard and test device $\leq$5 mmHg in the 60% of the readings, $\leq$10 mmHg in the 85% of the readings and $\leq$15 mmHg in the 95% of the readings.

[4] They should include 85 subjects over 12 years; $\geq$30% males and $\geq$30% females and shall have $\geq$5% of SP values $\leq$100 mmHg, $\geq$5% with $\geq$160 mmHg, and $\geq$20% with $\geq$140 mmHg.

4.2 Future works

In order to improve the model performance, several adjustments can be implemented:

- Increase the number of patients of the MIMIC III dataset, in order to increase the significance of the study;
- Try to build a balanced dataset, without using the SMOTE algorithm;
- Increase the MORFEA3 dataset, in order to satisfy the General Population studies criteria;
- Try to estimate also diastolic blood pressure values.

Bibliography

[1] 6 Types of Activation Function in Neural Networks You Need to Know.

[2] Action potentials and synapses.

[3] Adam — latest trends in deep learning optimization.

[4] Cardiac cycle.

[5] CVS monitoring.

[6] Dermis.

[7] Diffuse reflection.

[8] Effect of batch size on training dynamics.

[9] Everything you need to know about Neural Networks and Backpropagation — Machine Learning Easy and Fun.

[10] Gradient Descent: All You Need to Know.

[11] Heart Anatomy.

[12] Invasive blood pressure.

[13] Nerve Tissue. *National Cancer Institute*.

[14] Physiology of Circulation.

[15] Specular Reflection.

[16] Underfitting and Overfitting in Machine Learning.

[17] Understanding Activation Functions in Neural Network.

[18] Various Optimization Algorithms For Training Neural Network.

[19] Wikipedia-Blood pressure measurement.

[20] How does the blood circulatory system work?, 2012.

[21] Basics of ECG-Interpretation of waves and intervals. Technical report, epomedicine, 2016.

[22] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):1–39, 2007.

[23] John Allen. Photoplethysmography and its application in clinicalphysiological measurement. *PHYSIOLOGICAL MEASUREMENT*, 2007.

[24] Eijaz Allibhai. Building A Deep Learning Model using Keras, 2018.

[25] Filippo Augusti. Systolic blood pressure evaluation from PPG signal using Artificial Neural Networks. 2019.

[26] Charles F. Babbs. Oscillometric measurement of systolic and diastolic blood pressures validated in a physiologic mathematical model. *BioMedical Engineering Online*, 11:1–22, 2012.

[27] Jason Brownlee. How to Reduce Overfitting With Dropout Regularization in Keras, 2018.

[28] Serena H. Chen, Anthony J. Jakeman, and John P. Norton. Artificial Intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation*, 78(2-3):379–400, 2008.

[29] Vincent M. Christoffels and Antoon F.M. Moorman. Development of the cardiac conduction system why are some regions of the heart more arrhythmogenic than others? *Circulation: Arrhythmia and Electrophysiology*, 2(2):195–207, 2009.

[30] Enrico Coiera. Essentials of Telemedicine and Telecare. *Bmj*, 324(7345):1104, 2002.

[31] Jan Sørensen Dam. *Optical analysis of biological media - continuous wave diffuse spectroscopy*. Number November. 2000.

[32] Kolsoum Deldar, Kambiz Bahaadinbeigy, and Seyed Mahmood Tara. Teleconsultation and clinical decision making: A systematic review. *Acta Informatica Medica*, 24(4):286–292, 2016.

[33] Ministero della salute. Ministero della salute TELEMEDICINA Linee di indirizzo nazionali. *Esenzioni*, 2011:3–12, 2013.

[34] Andrew Entwistle. What Is Artificial Intelligence? *Engineering materials and design*, 32(3):1–14, 1988.

[35] Ahmed Fadhil. Beyond Patient Monitoring: Conversational Agents Role in Telemedicine & Healthcare Support For Home-Living Elderly Individuals. *arXiv*, 2018.

[36] Fchollet. The Sequential model, 2020.

[37] C. Fischer, M. Glos, T. Penzel, and I. Fietze. Erweiterter Algorithmus zur Echtzeit-Pulswellensegmentierung und Artefakterkennung in Photoplethysmogrammen. *Somnologie*, 21(2):110–120, 2017.

[38] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[39] Yana Mazwin Mohmad Hassim and Rozaida Ghazali. Training a Functional Link Neural Network Using an Artificial Bee Colony for Solving a Classification Problems. (December), 2012.

[40] Shivayogi Hiremath, Geng Yang, and Kunal Mankodiya. Wearable Internet of Things: Concept, Architectural Components and Promises for Person-Centered Healthcare. pages 304–307, 2014.

[41] Mohammad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring. *IEEE Transactions on Biomedical Engineering*, 64(4):859–869, 2017.

[42] Alexei A. Kamshilin and Nikita B. Margaryants. Origin of Photoplethysmographic Waveform at Green Light. *Physics Procedia*, 86(June 2015):72–80, 2017.

[43] Hyeong-Ju Kim, Woo Hyun Kang, Hyeonseung Lee, and N. Kim. Continuous monitoring of blood pressure with evidential regression. *ArXiv*, abs/2102.03542, 2021.

[44] Yuriy Kurylyak, Francesco Lamonaca, and Domenico Grimaldi. A Neural Network-based method for continuous blood pressure estimation from a PPG signal. *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, pages 280–283, 2013.

[45] Ross Langley and Steve Cunningham. How Should Oxygen Supplementation Be Guided by Pulse Oximetry in Children: Do We Know the Level? *Frontiers in Pediatrics*, 4(January), 2017.

[46] Jihyoung Lee, Kenta Matsumura, Ken Ichi Yamakoshi, Peter Rolfe, Shinobu Tanaka, and Takehiro Yamakoshi. Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 1724–1727, 2013.

[47] Dong Li. 5G and intelligence medicine—how the next generation of wireless technology will reconstruct healthcare? *Precision Clinical Medicine*, 2(4):205–208, 2019.

[48] Jing Liu, Bryan Ping-yen Yan, Wen-xuan Dai, Xiao-rong Ding, Yuan-ting Zhang, Ni Zhao, W B Baker, A B Parthasarathy, D R Busch, R C Mesquita, J H Greenberg, and A G Yodh. Multi-wavelength photoplethysmography method for skin arterial pulse extraction Optical radiation transfer in the human skin and applications in in vivo remittance spectroscopy, " in. *Electron. Biol. Eng. Clin. Phys. Physiol. Meas. BIOMEDICAL OPTICS EXPRESS Journal Phys. Med. Biol. Phys. Med. Biol. Lasers and Current Optical Techniques in Biology*, 2(16):145–154, 1964.

[49] Wen Long, Tiebin Wu, Shaohong Cai, Ximing Liang, Jianjun Jiao, and Ming Xu. A Novel Grey Wolf Optimizer Algorithm with Refraction Learning. *IEEE Access*, 7:57805–57819, 2019.

[50] Buddhishan Manamperi and Charith Chitraranjan. A robust neural network-based method to estimate arterial blood pressure using photoplethysmography. *Proceedings - 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE 2019*, pages 681–685, 2019.

[51] Markolf H Niemz, J. Langer and Markolf H Niemz. *Laser-tissue interactions: fundamentals and applications.* 2007.

[52] Kishan Mehrotra, Chilukuri K. Mohan, and Sanjay Ranka. *Elements of Arficial Neural Networks*. 1996.

[53] Agnes S. Meidert and Bernd Saugel. Techniques for non-invasive monitoring of arterial blood pressure. *Frontiers in Medicine*, 4(JAN):1–6, 2017.

[54] Biswajit Mishra and Nilesh Thakkar. Cuffless blood pressure monitoring using PTT and PWV methods. *International Conference on Recent Innovations in Signal Processing and Embedded Systems, RISE 2017*, 2018-January:395–401, 2018.

[55] Andreia V. Moço, Sander Stuijk, and Gerard De Haan. New insights into the origin of remote PPG signals in visible light and infrared. *Scientific Reports*, 8(1):1–15, 2018.

[56] Jermana L. Moraes, Matheus X. Rocha, Glauber G. Vasconcelos, José E. Vasconcelos Filho, Victor Hugo C. de Albuquerque, and Auzuir R. Alexandria. Advances in photopletysmography signal analysis for biomedical applications. *Sensors (Switzerland)*, 18(6):1–26, 2018.

[57] Gamze Nalcaci, Ayse Özmen, and Gerhard Wilhelm Weber. Long-term load forecasting: models based on MARS, ANN and LR methods. *Central European Journal of Operations Research*, 27(4):1033–1049, 2019.

[58] Hock Soon Ng, Moh Lim Sim, Chor Min Tan, and Chin Chin Wong. Wireless technologies for telemedicine. *BT Technology Journal*, 24(2):130–137, 2006.

[59] C. S. Pattichis, E. Kyriacou, S. Voskarides, M. S. Pattichis, R. Istepanian, and C. N. Schizas. Wireless telemedicine systems: An overview. *IEEE Antennas and Propagation Magazine*, 44(2):143–153, 2002.

[60] Jeremy Pinnell, Simon Turner, and Simon Howell. Cardiac muscle physiology. *Continuing Education in Anaesthesia, Critical Care and Pain*, 7(3):85–88, 2007.

[61] Malikeh Pour Ebrahim, Fatemeh Heydari, Taiyang Wu, Katherine Walker, Keith Joe, Jean Michel Redoute, and Mehmet Rasit Yuce. Blood Pressure Estimation Using On-body Continuous Wave Radar and Photoplethysmogram in Various Posture and Exercise Conditions. *Scientific Reports*, 9(1):1–13, 2019.

[62] S E E Profile. Continuous non - invasive blood pressure estimation. (April), 2015.

[63] Richard E. Klabunde. Blood Volume. In *Cardiovascular Physiology Concepts*.

[64] Michael Sampson and Anthony McGrath. Understanding the ECG. Part 1: Anatomy and physiology. *British Journal of Cardiac Nursing*, 10(11):548–554, 2015.

[65] Albrecht Schmidt. BIOSIGNALS IN HUMAN-COMPUTER INTERACTION, 2021.

[66] D. H.T. Scott. Non-invasive blood pressure measurement displays. *Anaesthesia*, 73(10):1299, 2018.

[67] Mohamed Adel Serhani, Hadeel T. El Kassabi, Heba Ismail, and Alramzana Nujum Navaz. ECG monitoring systems: Review, architecture, processes, and key challenges. *Sensors (Switzerland)*, 20(6), 2020.

[68] Shota Shimazaki, Shoaib Bhuiyan, Haruki Kawanaka, and Koji Oguri. Features Extraction for Cuffless Blood Pressure Estimation by Autoencoder from Photoplethysmography. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018-July:2857–2860, 2018.

[69] Gašper Slapničar, Mitja Luštrek, and Matej Marinko. Continuous blood pressure estimation from PPG signal. *Informatica (Slovenia)*, 42(1):33–42, 2018.

[70] ST Microelectronics Inc. X-CUBE-AI Data brief Artificial intelligence (AI) software expansion for STM32Cube X-CUBE-AI. (June), 2019.

[71] George S. Stergiou, Bruce Alpert, Stephan Mieke, Roland Asmar, Neil Atkins, Siegfried Eckert, Gerhard Frick, Bruce Friedman, Thomas Graßl, Tsutomu Ichikawa, John P. Ioannidis, Peter Lacy, Richard McManus, Alan Murray, Martin Myers, Paolo Palatini, Gianfranco Parati, David Quinn, Josh Sarkis, Andrew Shennan, Takashi Usuda, Jiguang Wang, Colin O. Wu, and Eoin O'Brien. A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement. *Hypertension*, 71(3):368–374, 2018.

[72] Mohammed Sunasra. Performance Metrics for Classification problems in Machine Learning.

[73] Toshiyo Tamura, Yuka Maeda, Masaki Sekine, and Masaki Yoshida. Wearable photoplethysmographic sensors—past and present. *Electronics*, 3(2):282–302, 2014.

[74] 2005 Tang et al. NIH Public Access. *Bone*, 23(1):1–7, 2008.

[75] Tutorialspoint. Scikit Learn - Introduction.

[76] Wikipedia. Heart rate variability.

[77] Yang Xu, Weigong Zhang, Dong Wang, and Ping. A calibration method for cuffless continue blood pressure measurement using gaussian normalized pulse transit time. *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–5.

[78] Hayley A. Young and David Benton. Heart-rate variability: A biomarker to study the influence of nutrition on physiological and psychological health? *Behavioural Pharmacology*, 29:140–151, 2018.

[79] Mehmet R. Yuce. Implementation of wireless body area networks for healthcare systems. *Sensors and Actuators, A: Physical*, 162(1):116–129, 2010.

[80] Yan Zhang, Nirwan Ansari, and Hiroshi Tsunoda. Wireless telemedicine services over integrated IEEE 802.11/WLAN and IEEE 802.16/WiMAX networks. *IEEE Wireless Communications*, 17(1):30–36, 2010.