Cuffless Blood Pressure Measurement

Stefano Villata

Advisor:
Prof. Gian Alessandro Eros Pasero

Co-Advisor:
Annunziata Paviglianiti

Master Thesis
Biomedical Engineering

DET - Department of Electronics and Telecommunications
Politecnico di Torino, Turin, Italy.
October 5, 2020

Abstract

Continuous vital signals monitoring has gained a huge relevance for disease prevention that afflict a large part of the world population, for this reason the healthcare equipment should be easy-wear and convenient-operate. Nonintrusive and noninvasive detective methods are the basic requirement for the wearable medical devices, especially when the devices are used in sports applications or by the elderly for self-monitoring. The aim of this thesis is to measure continuous arterial blood pressure through a cuffless non-intrusive approach.

The arterial blood pressure is an essential physiological parameter for health monitoring. Most blood measurement devices determine the systolic and diastolic arterial blood pressure through the inflation and the deflation of a cuff. This method is uncomfortable to the user and may cause anxiety which in turns can affect the blood pressure.

The approach utilized in this thesis is based on deep learning techniques: different neural networks are used to infer ABP starting from photoplethysmogram and electrocardiogram. In particular we predicted ABP first utilizing only PPG and then PPG and ECG, we demonstrated that adding ECG improved performance in every configuration achieving, after personalization, a MAE equal to 4.118 mmHg on systolic blood pressure 2.228 mmHg on diastolic blood pressure with a modified ResNet followed by 3 LSTM layers. Results were compliant with the American National Standards of the Association for the Advancement of Medical Instrumentation.

ECG, PPG and blood pressure measurements are extracted from the MIMIC database that contains clinical signal data reflecting real measurements and validates the results on a custom dataset created at Neuronica Lab, Politecnico di Torino.

Dedication

Finalmente ho raggiunto questo traguardo, ringrazio tutti quelli che mi hanno supportato, sopportato e incoraggiato. In modo particolare la mia famiglia e i miei amici Giulia, Luigi, Mattia, Miriam, Ortenzia e Sonia con cui ho condiviso tanti anni di università. Infine, un particolare ringraziamento al professore Pasero per avermi dato l'opportunità di dedicarmi a questo progetto.

Acronyms

ABP arterial blood pressure. I, 1, 3, 7, 13, 47, 55

ANN artificial neural network. 27

ANSI American National Standards Institute. 70

BLSTM Bidirectional Long Short Term Memory. 59

CNAP Continuous noninvasive arterial pressure. 2

CNN convolutional neural network. 35

CVD cardiovascular disease. 1

DBP diastolic blood pressure. I, 3, 8, 11, 13, 48, 55

ECG electrocardiogram. I, 3, 19, 47, 65

FDA Food and Drug Administration. 4

HR heart rate. 10, 17

ICU intensive care unit. 15

LOO Leave One Out. 55, 65

LSTM Long Short Term Memory. 42, 55

MAE mean absolute error. I, 30, 58

MAP mean arterial pressure. 8

ML machine learning. 27

MLP multilayer perceptron. 28

mmHg millimeters of mercury. 8

NN neural network. 27, 31

PAT pulse arrival time. 3

PPG photoplethysmogram. I, 3, 15, 47, 65

PTT pulse transit time. 3, 11, 13, 16

PWV pulse wave velocity. 3

RMSE root mean squared error. 4, 30

RNN recurrent neural network. 39, 41, 55

SBP systolic blood pressure. I, 3, 8, 11, 13, 48, 55

TLU threshold logic unit. 28

WHO World Health Organization. 9, 12

WHT white coat hypertension. 1

XOR Exclusive OR. 28

Contents

1 Introduction **1**
1.1 Related work 2

I Biological Signals **5**

2 Blood Pressure **7**
2.1 Hypertension 9
2.2 Hypotension 10
2.3 Non-Invasive measurements 11

3 Photoplethysmogram **15**
3.1 Physical principles 16

4 Electrocardiogram **19**
4.1 Leads 20

II Neural Network principles **25**

5 Artificial Neural Networks **27**
5.1 Multilayer perceptron 28
5.2 Hyperparameters 29

6 Convolutional Neural Networks **35**
6.1 ResNet 37
6.2 WaveNet 39

7 Recurrent Neural Networks **41**
7.1 LSTM 42

III Methods **45**

8 Dataset **47**
8.1 MIMIC database 47
8.2 Data cleaning 48

9 Tested neural architectures **55**
9.1 Direct SBP/DBP prediction 56
9.2 Entire BP prediction 59

10 Validation **65**
10.1 Leave-One-Out 65
10.2 Experimental setup 68

11 Results **71**
11.1 MIMIC database results 71
11.2 Polito database results 73
11.3 Conclusion 73

1 Introduction

Recent studies have highlighted the clinical relevance of continuous blood pressure monitoring [12]. Arterial blood pressure (ABP) is an indicator of hypertension, which is one of the most important risk factors of cardiovascular disease (CVD). For this reason, its variability is an independent and important risk factor associated with cardiovascular events. Despite the importance of regular monitoring, there aren't proper standard to easily measure blood pressure, indeed cuff-based devices, the actual gold standard for non-invasive measurements, require a fairly strict protocol and can only be used at rest.

In order to measure arterial blood pressure there are two clinical gold standard: the invasive catheters system and the cuff sphygmomanometer [42].

The invasive catheters system is performed through an arterial line: a catheter is inserted into an artery. It is used in intensive care units and anesthesia to directly monitor blood pressure in the most accurate way possible and to obtain samples for arterial blood gas analysis. Only physicians and specialized nurses can do the insertion, also, it is often painful and performed using an anesthetic to make it more tolerable and to help prevent vasospasm [43].

On the other hand, cuff-based devices are the golden standard for indirect measurements and are commonly recommended by physicians. These devices offer the highest measurement accuracy, however, they also have several downsides. Cuff size is usually too small leading to errors in diagnosis [44] and the person using a cuff-based device must follow a relatively strict measuring protocol in order to ensure the measured values are correct. The measuring procedure can be tedious and requires dedicated time and effort, also, physical activity (e.g., exercise) typically does not allow for simultaneous measuring of BP with a cuff [36].

Furthermore, the measuring event itself can cause white coat hypertension (WHT), commonly known as white coat syndrome. It is a condition where a patient's blood pressure is higher when taken in a medical setting, while it is normal during daily activities. It is believed that the phenomenon is due to anxiety experienced during a clinic visit [44]. Continuous noninvasive arterial pressure (CNAP) measurement combines the advantages of the two methods. CNAP systems have different requirements for general public purposes and clinical purposes. In the first case it is sufficient to measure blood pressure changes over time, while in the latter the system must provide not only how it changes, but also absolute blood pressure, physiological rhythms and its pulse waves for quality control, , fig.1.1, [42].

In order to detect pressure inside an artery from the outside several techniques were developed. The starting points usually are volume and flow changes in the artery, these data are easily collectable in the periphery (e.g. in a finger), however, they are not linearly correlated with blood pressure, because of the non-linearity of the elastic components of the arterial wall as well as the non-elastic parts of the smooth muscles

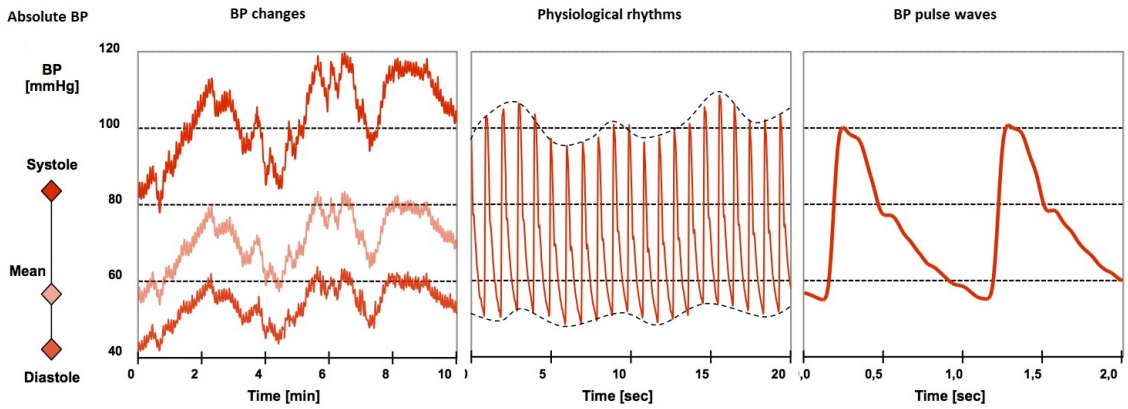


Figure 1.1: Different blood pressure information according to time resolution

of the finger artery [42].

Several techniques involve monitoring the pulse wave, for example the velocity or some other parameter related, indeed pulse wave has a clear relationship with blood pressure when vessels are more relaxed or elastic, the blood travels more slowly and exerts less pressure [36].

However, due to the non-linearity of the problem Neural Networks appear to be an ideal framework, it is the so-called data-driven system identification.
Neural networks are conceptually simple, easy to train and use and can approximate a target function in an excellent way, however, the biggest criticism is that the models produced are completely opaque, fig.1.2. It is therefore very difficult to know what is causing what, to analyse the model, or to compute dynamic characteristics from the model [45].

Figure 1.2: ANN are usually characterized as “black box”

Deep learning is gaining popularity for their ability to achieve state-of-the-art performances in different settings. Deep neural netowrks have been applied to an increasing number of problems spanning different domains of biomedical application [4], such as protein structure classification [31] and prediction [5] [16], medical image classification [26] or genomic sequence analysis [33]. In this study several different configurations to infer blood pressure were used, it is a typical regression task: two configurations, the former with the PPG signal as input and the latter with the ECG and PPG combination as input, are set in the neural network and thus the measurement is obtained at the output.

1.1 Related work

Traditional ABP measurement techniques are either invasive or cuff-based, which are impractical, intermittent, and uncomfortable for patients. For this reason, several alternatives were investigated.

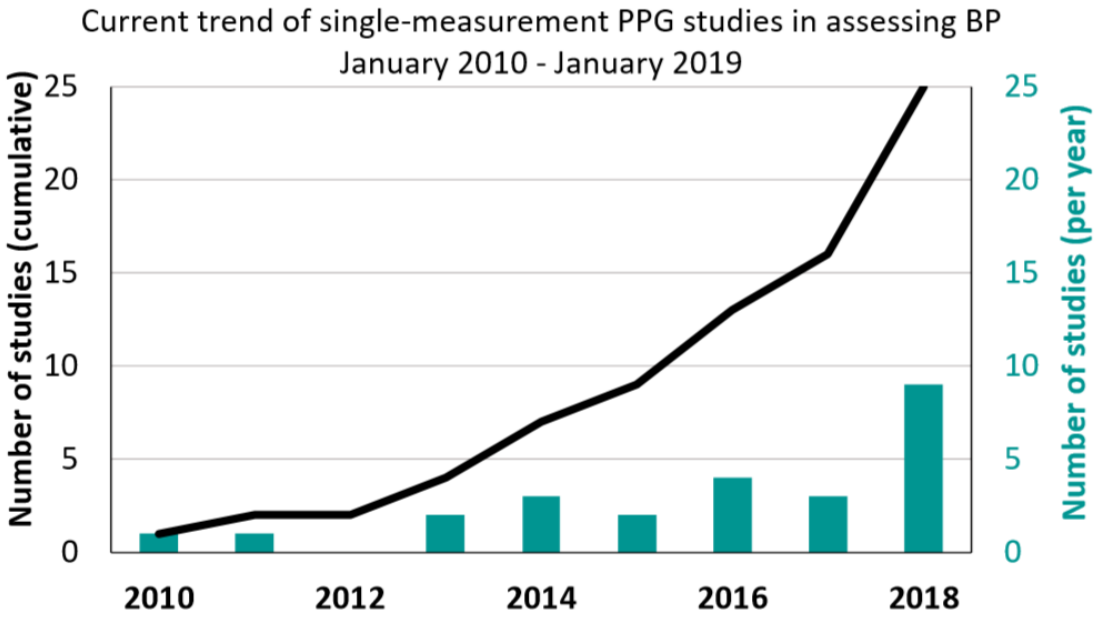


Figure 1.3: Trend of publications on PubMed database regarding single-site measurement PPG to estimate BP

In particular, PPG emerged as a potentially useful signal, fig. 1.3 [23], indeed many studies point out a clear relation between PPG and ABP, also PPG can easily be integrated into wearable devices. Initially, indirect approaches, using features derived from PPG and ECG, were the preferred ones: [19][35] shows a strong negative correlation between ABP and pulse transit time (PTT), but also pulse wave velocity (PWV) [55] and pulse arrival time (PAT) [8] were studied.

[19] established a correlation without trying to predict actual blood pressure, while [55] tried to show a relationship between PWV and BP. Lastly, [8] used the mean error as the evaluation metric between their target BP value and their predicted one, they achieved ±6 and ±4 mmHg respectively for SBP and DBP. However, the mean error is not a suited error metric for regression tasks because positive and negative differences cancel each other out in the overall mean, showing a low overall ME even if individual errors are large.

[22] demontrated neural networks can perform better than linear regression, they extracted a set of feature from PPG recordings, taken out from MIMIC database. The achieved 3.80 ± 3.41 mmHg on SBP and 2.21 ± 2.09 mmHg on DBP on a very small dataset dataset (15000 heartbeats are analysed, which means roughly 4 hours of recordings).

[34] used a complex recurrent neural network on 22 features extracted from PPG and ECG, RMSE is used as metric and achieved great performances: 3.63 on SBP and 1.48 on DBP. However, it is not easy to compare different studies because BP prediction performances are heavily influenced by the dataset, this is crucial because a selected subset of data may have a relatively constant BP, in which case the network will easily achieves low errors.

Nowadays, thanks to the advancements in deep neural techniques new approaches are evaluated, they directly utilize PPG raw signal.

The idea to measure BP using only PPG signals was already investigated in [38], four features from PPG signal in three different setups: rest, exercise and recovery. It

was one of the first study based only on PPG, which shows a good correlation between BP and some features, meaning it is certainly possible to predict BP using only PPG. The obtained results are good, but probably because the experiments are conducted on few healthy people (low ABP variability).

[36] had a huge impact on our research, their preprocessing pipeline was well suited to remove noisy signals in MIMIC and their validation system was the most robust applied on ABP regression task. However, their goal was just to measure ABP starting from PPG and for this reason they did not evaluate if ECG could improve performances. Their approach is one of the very first based on deep learning, they developed a complex neural network, which analysed both temporal and spectral features automatically extracted. [36] used a huge amount of data extracted from MIMIC III and trained their network using a top-notch GPU cluster.

PPG-based approaches for ABP measurement are gaining more and more interests in both academic and industrial fields. The approval in 2019 of the Food and Drug Administration (FDA) first cuffless device [46] is opening a new and interesting market.

Part I

Biological Signals

2 | Blood Pressure

Arterial blood pressure (ABP) is one of the so-called vital signs and is accepted as an index of the circulatory condition. It is the pressure of circulating blood on the walls of blood vessels. Most of this pressure is due to work done by the heart by pumping blood through the circulatory system: when the left ventricle ejects blood into the aorta, the aortic pressure rises. As the left ventricle is relaxing and refilling, the pressure in the aorta falls. ABP is influenced by many factor, for instance, cardiac output, arterial stiffness, but also health and emotional state [47].

Fig. 2.1 shows the pressure over time, the waveform is influenced by many factors: aortic valve conditions, compliance of the aorta, vascular resistence cardiac output, technical considerations of recording and lastly ventricular filling. Ventricular filling is well explained by the Frank/Starling law: the energy of contraction is a function of the length of the muscle fibre. So the greater the filling of the ventricles (which implies more filling time) the stronger the subsequent systolic contraction. Also diastolic valley is influenced by time between heartbeats: faster pace means blood has less time to flow out of the aorta and, therefore, the pressure in the aorta to fall.

Also, the waveform depends on where it is recorded, narrower arteries (usual in the periphery of the body) are less compliant, therefore, the pressure here has different shapes, in particular the anacrotic limb (the ascending part of the wave) is steeper and the SBP is generally higher.

Lastly, the dip in the ABP waveform occurring on the descending part of the wave is referred to as the dicrotic notch. The dicrotic notch in an arterial pressure waveform can correspond to the closure of the aortic valve, however, usually the dicrotic notch and the dicrotic wave that follow it due to a reflected pressure wave.

Blood pressure is usually expressed by two measurements, the systolic blood pres-

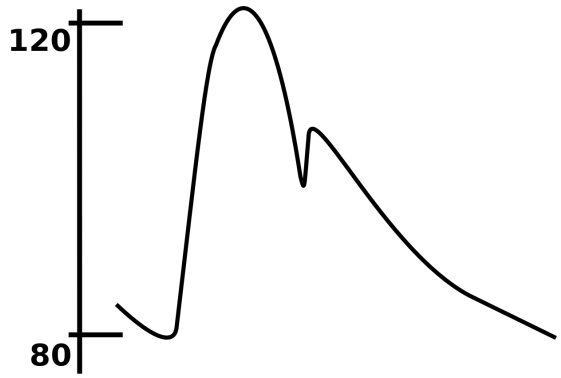


Figure 2.1: Schematic representation of the arterial pressure waveform. The notch in the curve is associated with closing of the aortic valve.

sure (SBP) and diastolic blood pressure (DBP) pressures, which are the maximum and minimum pressures, respectively. Usually, the normal range, at rest, is within 100-130 millimeters of mercury (mmHg) for the systolic one and 60-80 mmHg for the diastolic one [48].

Another index of blood pressure is the mean arterial pressure (MAP), MAP is related with cardiac output and systemic vascular resistence. It is the average arterial pressure during the cardiac cycle and it can be estimated from the systolic and diastolic pressures, equation 2.1.

$$P_{mean} = \frac{P_{sys} + 2P_{dias}}{3} \tag{2.1}$$

Mean arterial pressure is a useful concept because it can be used to calculate overall blood flow, and thus delivery of nutrients to the various organs. It is a good indicator of perfusion pressure (ΔP). The ideal blood pressure adequately perfuses all of the various organ systems without causing damage. Any organ not adequately perfused will suffer ischemic damage and/or be unable to perform adequately [40].

Blood flow is defined by Poiseuille's law, equation 2.2

$$Q = \Delta P \times \frac{\pi r^4}{8NL} \tag{2.2}$$

where Q is the blood flow, ΔP is the pressure difference between the two ends, r is the radius of the vessel, N is the blood viscosity, and L is the length of the vessel. This formula can be restated in a more clinically useful expression, equation 2.3

$$CO = \frac{MAP \times 80}{TPR} \tag{2.3}$$

CO is the cardiac output in liters/minute (clinical equivalent of blood flow Q). MAP in mmHg is used to approximate the pressure gradient ΔP. TPR is the resistance to flow in $dynes \times sec \times cm^{-5}$ and clinically represents $8\frac{NL}{\pi r^4}$. The conversion factor 80 appears in the formula simply to allow use of more conventional units [40].

However, to calculate mean arterial pressure exactly, it must be obtained from the whole time course of arterial pressure. Also, continuous blood pressure measurement is required for certain disease such as sleep disturbance [37].

For this reason, direct and indirect blood pressure monitoring methods are used. Direct monitoring is commonly used during operations, while indirect monitoring is used during physical examination and checkups. Wearable and portable wearable devices are based on indirect monitoring. Every device, both commercial and laboratory-based equipment, are reviewed [37].

Auscultation is generally considered the gold standard for non-invasive blood pressure measurements, while during invasive measurements a thin catheter is inserted into an artery.

Traditionally, blood pressure was measured in a non-invasive way using mercury sphygmomanometer, however, due to the ban on use of mercury devices and seeking to create ease-to-use devices new alternatives are developed. In particular, semi-automated methods have become common, they have good accuracy: according to international standards they achieve an average difference between two standardized reading methods of 5 mm Hg or less. However, these devices are still not designed to perform continous measurements [13].

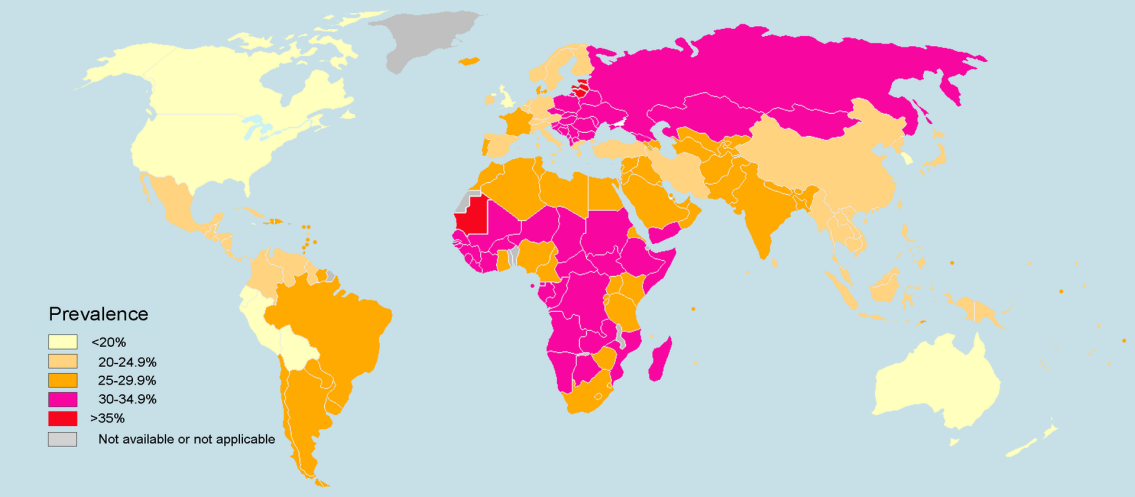


Figure 2.2: High blood pressure affects between 16 and 37% of the population globally, in 2010 hypertension was believed to have been a factor in 18% of all deaths

Continous measurements are of utmost importance, indeed blood pressure fluctuates from minute to minute, also, it have a circadian rhythm over a 24-hour period, with highest readings in the early morning and evenings and lowest readings at night. Loosing this rythm is correlated with future cardiovascular diseases, in particular night-time blood pressure is a stronger predictor of cardiovascular events than day-time blood pressure [47].

Blood pressure that is too low is called hypotension, pressure that is consistently high is called hypertension and normal levels of blood pressure is called normotension. Both hypertension and hypotension have many causes and may be of sudden onset or of long duration. Long-term hypertension is more common than long-term hypotension, which is usually only diagnosed when it causes symptoms.

Hypertension is particularly problematic, indeed it is demonstrated that the risk of cardiovascular disease increases progressively above 115/75 mmHg.

2.1 Hypertension

Hypertension (HTN or HT), also known as high blood pressure (HBP), is a long-term medical condition in which the blood pressure in the arteries is persistently elevated.

High blood pressure typically does not cause symptoms. Long-term high blood pressure, however, is the most important preventable risk factor for premature death worldwide, indeed, it is a major risk factor for coronary artery disease, stroke, heart failure and several other diseases.

According to WHO approximately 22% of the population of the world have hypertension[17], fig. 2.2, and it becomes more common with age. It is a huge cost for national healthcare systems, the direct and indirect costs related to high blood pressure in 2010 are estimated around $76.6 billion in USA [48].

There are two kinds of hypertension, primary or secondary, depending on whether the cause is related to a generic lifestyle or genetic cause or an identifiable cause, such as chronic kidney disease or endocrine disorder. In particular the predominant one is primary hypertension, which covers close to 90% of cases.

It is considered high blood pressure if it is persistently around 130/80 or above. Different numbers apply to children.
Lifestyle changes and medications can lower blood pressure and decrease the risk of health complications improving life expectancy.

A particular risky condition is characterized by hypertensive crises, this situation arises when the blood pressure is equal to or greater than 180/110. There are two different situations when the blood pressure is so high, urgency or emergency, depending on the presence of end-organ damage.
In hypertensive urgency there is no organ damage, therefore, the blood pressure is lowered gradually in 48 hours timespan using oral medications.
In a hypertensive emergency, there is evidence of direct damage to one or more organs and therefore the blood pressure must be reduced more rapidly to stop ongoing organ damage.

Primary hypertension usually is caused by increased resistance to blood flow thus, while cardiac output remains normal. There is evidence that some younger people with prehypertension or 'borderline hypertension' have high cardiac output, an elevated heart rate (HR) and normal peripheral resistance, termed hyperkinetic borderline hypertension. These individuals develop the typical features of established essential hypertension in later life as their cardiac output falls and peripheral resistance rises with age. Whether this pattern is typical of all people who ultimately develop hypertension is disputed. The increased peripheral resistance in established hypertension is mainly attributable to structural narrowing of small arteries and arterioles [48].

2.2 Hypotension

Hypotension is low blood pressure. A systolic blood pressure of less than 90 mmHg or diastolic of less than 60 mmHg is generally considered to be hypotension. Different numbers apply to children. However, in practice, blood pressure is considered too low only if noticeable symptoms are present [49].

It is best understood as a physiological state rather than a disease, even though severely low blood pressure can deprive the brain and other vital organs of oxygen and nutrients, leading to a life-threatening condition called shock.

The earliest clinical manifestation of low blood pressure may be fatigue or shortness of breath on exertion. Further declines in blood pressure may lead to dizziness and fainting, particularly on assuming an upright posture. An easy way to monitor vital organ perfusion is to check urine output, which should never drop below 20 ml/hr if the intrarenal blood pressure is satisfactory. The most common causes of low blood pressure are dehydration or decreased cardiac output.

Hypovelemia, or low blood volume, is the most common cause followed by decreased cardiac output, other causes are hormonal changes, widening of blood vessels, anemia, heart problems, or endocrine problems. Also, some medications and syndromes can also lead to hypotension.

Treatment of hypotension may include the use of intravenous fluids or vasopressors. Poor cardiac output may result in hypotension; a thorough examination of cardiac contractility and heart rate will dictate appropriate therapy. Inadequate contractility may be improved with positive inotropic agents such as digoxin. Abnormal heart rates may require antiarrhythmic agents if too rapid, or vagolytic agents such as at-

ropine if too slow. Occasionally, a pacemaker is required to maintain a satisfactory heart rate and blood pressure [40].

However, hypotension isn't always a disease, indeed for some people who exercise and are in top physical condition, low blood pressure could be normal. A single session of exercise can induce hypotension and water-based exercise can induce a hypotensive response.

2.3 Non-Invasive measurements

Most used techniques are: the auscultatory method (generally identified as sphygmomanometer), the oscillometric method and the unloaded method. Several other techniques exist, however most of them, despite many promising studies, are not widely used or have few approved device. In particular, there is great interest in CNAP systems, because they combine advantages of the two methods: continuous and non-invasive measurements. For instance, PTT is one of the most studied.

Auscultatory method

In 1856 blood pressure was recorded in humans for the first time using a U-shaped manometer tube connected to a brass pipe canula plugged directly into the artery, however since then there were efforts to achieve a non-invasive ABP measurement.

The first instrument to accurately measure ABP was the sphygmomanometer developed by Scipione Riva-Rocci in 1896.

A sphygmomanometer consists of an inflatable cuff to collapse and then release the artery under the cuff in a controlled manner and a measuring unit (the mercury manometer or aneroid gauge). It is based on the auscultatory method.

The auscultatory method is the listening of Korotkoff sounds in the brachial artery. When a cuff is inflated to a level higher than the systolic pressure, the brachial artery is occluded. Thus the artery is completely compressed, there is no blood flow, and no sounds are heard. As the cuff is gradually deflated, blood flow is reestablished and the Korotkoff sounds are first heard with a stethoscope below the cuff. When the sound is first heard the SBP is signified. Korotkoff sounds will continue to be heard as the cuff pressure is further lowered. However, when the cuff pressure reaches DBP, the sounds disappear, fig. 2.3.

This method has always been the gold standard for clinical blood pressure measurement, it is performed by a trained healthcare provider and blood pressure values are obtained from either aneroid or mercury device. In particular, mercury sphygmomanometers were considered the gold standard, however, the WHO banned mercury based device. They were particularly appreciated because they showed blood pressure by affecting the height of a column of mercury, which did not require recalibration. Aneroid sphygmomanometers (mechanical types with a dial) are now in common use, however, they may require calibration checks with a standard pressure monitor.

There are many variables that affect the accuracy of this method, i.e. cuff size, and numerous studies have shown that physicians and healthcare providers rarely follow the established guidelines for taking proper manual blood pressure measurements [37].

When the cuff size is too small the sphygmomanometer will output an higher pressure, while when the cuff is too large it will output lower values. The "ideal" cuff should

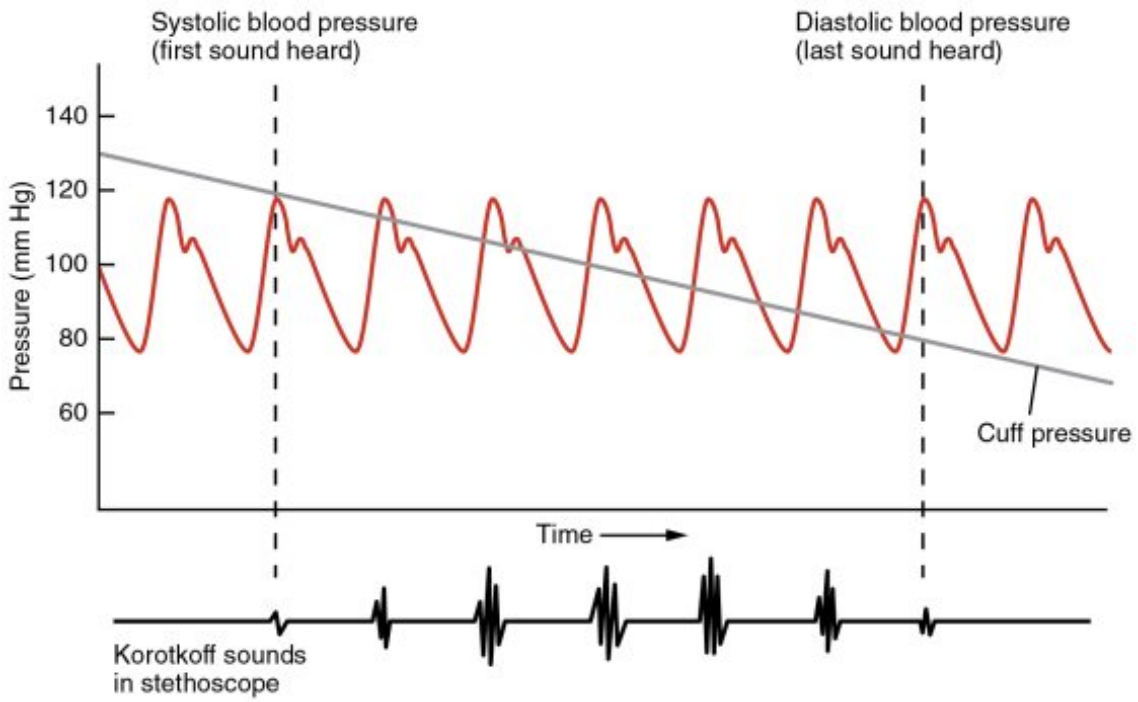


Figure 2.3: Cuff pressure during deflation overimposed on blood pressure. When the cuff pressure is between SBP and DBP Korotkoff sound can be heard.

have a cuff bladder length of 80% and a cuff bladder width of at least 40% of arm circumference. A recent study comparing intra-arterial and auscultatory blood pressure concluded that the error is minimized with a cuff width of 46% of the arm circumference [37].

Furthermore, blood pressure measurement is most commonly made in either the sitting or the supine position, but the two positions give different measurements. Diastolic pressure measured while sitting is higher than when measured supine, while systolic pressure is generally higher, which means in a considerable amount of subjects this is not true.

Also, the position of the arm can affect the measurements. The arm should stay on the same height as the heart, otherwise hydrostatic pressure will chenge the recording, for instance, if the arm is above the heart level, the readings will be lower.

Lastly, the rate of deflation has a significant effect on blood pressure determination. Deflation rates > 2 mmHg per second can lead to a significant underestimation of systolic and overestimation of diastolic blood pressure. Automated devices with a linear deflation rate may have improved accuracy over the more common circumstances in automated devices that have stepwise deflation.

Oscillometric method

The oscillometric method was first introduced in 1876 and involves the observation of oscillations in the sphygmomanometer cuff pressure which are caused by the oscillations of blood flow [50].

Oscillometric measurements are usually used in digital meters, the first fully automated oscillomettric blood pressure cuff was made available in 1981. These devices may use manual or automatic inflation, but both types are electronic, easy to operate

without training, and can be used in noisy environments. However, like auscultatory method the cuff size must be appropriate.

This method uses a sphygmomanometer cuff, like the auscultatory method, but with an electronic pressure sensor (transducer) to observe cuff pressure oscillations. It employs either deformable membranes that are measured using differential capacitance, or differential piezoresistance, and they include a microprocessor to automatically interpret the sensor results. The pressure sensor should be calibrated periodically to maintain accuracy.

They accurately measure mean blood pressure and pulse rate, while systolic and diastolic pressures are obtained less accurately than with manual meters [37].

Initially the cuff is inflated to a pressure higher than SBP and then it is reduced to below DBP over a period of about 30 seconds. The sphygmomanometer records a constant pressure when the blood flow is blocked or unimpeded, while when the blood flow is present, but restricted the recorded pressure will vary periodically in synchrony with the cyclic expansion and contraction of the brachial artery.

Recently, in order to improve ABP estimates the oscillometric method has been supported by algorithms based on machine learning and on PTT [50].

Digital oscillometric monitors may not be advisable for some patients, such as those suffering from arteriosclerosis, arrhythmia, preeclampsia, pulsus alternans, and pulsus paradoxus, as their calculations may not correct for these conditions.

Unloaded method

The unloaded method was first developed by Penaz and works on the principle of the “unloaded arterial wall” and allows continuous BP measurements.

It is performed through a pulse oximeter, in general a photoplethysmograph, fig. 2.4, indeed it can measure finger blood volume changes using light. However, transforming volume changes into pressure is not easy because of the non-linearity of the elastic components in the finger (arterial walls and muscles).

In order to linearize the phenomenon the photoplethysmograph is used with a pressure cuff placed over the finger. The PPG output is used to drive a servo loop, which rapidly changes the cuff pressure to keep blood volume constant, so that the artery is held in a partially opened state. The continuously changing outside pressure that is needed to keep the arterial blood volume constant directly corresponds to the arterial pressure.

This method gives an accurate estimate of the changes of systolic and diastolic pressure, although both may be underestimated when compared with brachial artery pressures.

Commercial applications are already available, i.e. the Finometer (formerly Finapres) and Portapres recorders. They are used in several PPG studies as ground truth [8], however they are cumbersome, costly and their accuracy isn’t always as high as it should be. For these reasons, this method in its current form is not suited to clinical setting.

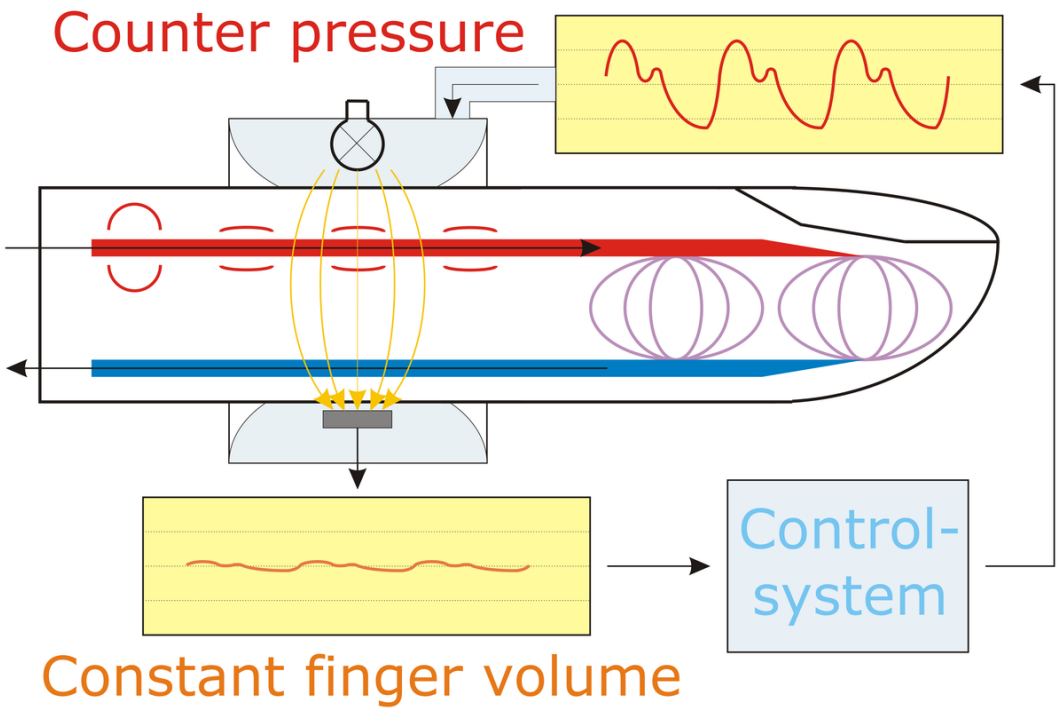


Figure 2.4: Vascular unloading technique scheme

3 | Photoplethysmogram

A photoplethysmogram (PPG) is a method for measuring changes in blood volume in the microvascular bed of tissue. It is a volumetric signal, measuring how big the finger or ear blood vessels (arterial and venous) become with each heartbeat and breath. A PPG is often obtained by using a pulse oximeter which illuminates the skin and measures changes in light absorption [46]. PPG was introduced in 1937 by Alrick Hertzman.

Pulse oximeter is one of the most popular wearable devices, it is mainly used to determine heart and respiratory rates, however, PPG signal, fig.3.1, is almost only displayed ICU applications [3].

PPG signal is composed by two components: a continuous one (DC) attributable to the bulk absorption of the skin tissue, and an alternate one (AC) attributable to variation in blood volume in the skin caused by the pressure pulse of the cardiac cycle [46].

The shape of the PPG waveform differs from subject to subject, and varies with the location and manner in which the pulse oximeter is attached. Usually, pulse oxymeter are worn on the finger, but PPG can be obtained also applying pulse oximeters on the ear, nasal septum and forehead, while wrist-worn PPG devices in clinical use are currently under study.

In addition to the difficulties related to the subject-dependent waveforms, PPG is also difficult to analyze because its prominent feature, the pulse amplitude, is often filtered by device manufacturers using specifically “auto-gain” or “auto-amplification”, this limits its usability by the practicing clinician [3].

The PPG amplitude is the result of of a complex interaction of several factors: stroke volume, vascular compliance, and tissue congestion effects. A large PPG pulse does not imply a high arterial pressure, absurdly, PPG amplitude can decrease during significant increases in blood pressure that are due to increased sympathetic tone, e.g. this phenomenon is usually seen in an incision on a subject under general anesthesia.

PPG has many applications, like mentioned before it is used to monitor depth of anesthesia, oxygenation, but it can, also, precisely measure heart rate variability (HRV). Indeed, the PPG is very sensitive to irregularity of the pulse, i.e. premature ventricular or atrial beats and atrial fibrillation (which is often difficult to diagnose directly from the ECG).

Wrist-worn PPG device are, already, widely used in commercial application, fig 3.2, to show pulse rate, however, the accuracy and validity of commercial devices is largely unknown [37].

Other PPG applications are measuring hypo- and hypervolemia and promising studies used PPG also to monitor arterial blood pressure. Combining PPG with ECG it is possibile to measure the pulse transit time (PTT) which is a parameter related to ABP. It is obtained by measuring how long it takes the PPG to arrive at a distal portion of the body (i.e., finger or toe).

Several approaches are being developed in ABP continuous measurements, and in

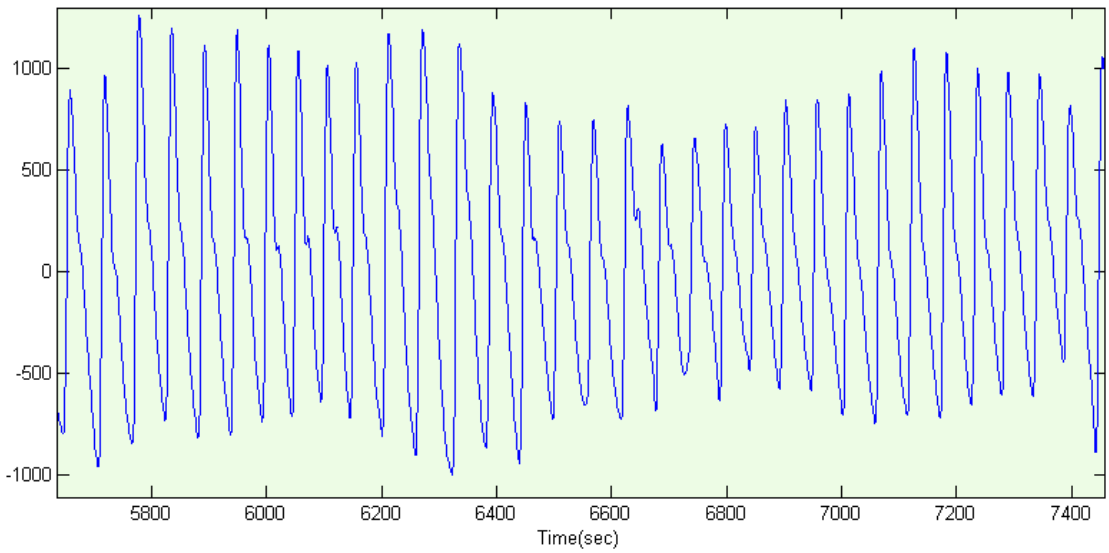


Figure 3.1: Photoplethsmograph obtained from a Nonin pulse oximeter attached to the ear. The secondary peaks are due to venous plexus.

Figure 3.2: Some PPG-based PR monitors available on the market.

late August 2019 the FDA has cleared a photoplethysmography-based cuffless blood pressure monitor [11].

Conventional PPG systems have been limited to single spot and contact measurements, however, it is an emerging technology and many new applications are being studied nowadays, e.g. multi-site photoplethysmogram and noncontact photoplethysmographic imaging systems.

3.1 Physical principles

PPG is obtained illuminating the skin with the light from a LED and then optically detecting changes in the transmitted or reflected light intensity, depending on how it is designed, fig 3.3.

Light traveling through biological tissue can be absorbed by different substances,

including pigments in the skin, bones, and arterial and venous blood. Only blood volumes changes over time, thus observating how much light is absorbed it is possible to understand how blood flow changes. For instance, arteries contain more blood volume during the systolic phase than during the diastolic phase of the cardiac cycle. Venous blood flow is mostly ininfluent and it is, therefore, ignored.

The continuous component of the PPG waveform is due to the structure and to the average blood present in the sensed volume.

The DC component changes slowly with respiration, while the alternate component shows blood volume changes, which occur during the cardiac cycle. The fundamental frequency of the AC component depends on the HR and is added to the DC component.

The interaction of light with biological tissue is quite complex and may involve scattering, absorption, and/or reflection, however, it can be fairly approximated by BeereLambert's law, equation 3.1.

$$A = -\ln \frac{I_{out}}{I_{in}} = l \sum_{i=1}^{N} \epsilon_i c_i \tag{3.1}$$

In practice, for N attenuating species in case of uniform attenuation, the absorbance A depends on the path length of the beam of light through the material sample l, the concentration c and the molar attenuation coefficient ϵ. The absorption clearly depends on the amount of light provided I_{in} and the amount of light which is transmitted through the volume (I_{out}, which refers to both reflected and transmitted light).

ϵ depends on wavelenght, therefore, different wavelengths are absorbed in different ways. The shorter one, like ultraviolets, are harmful and, therefore, must be avoided. Longer infrared are highly absorbed by water, while in the visible spectrum the least absorbed is red. In particular red and near infrared (NIR) are the ones who easily passes through. For this reason, IR wavelengths have been used as a light source in PPG sensors [37].

Traditionally, IR light was used in PPG devices because it can measure blood flow in deep-tissues, however, green light wavelength is becoming increasingly popular because of the large intensity variations in modulation observed during the cardiac cycle for these wavelengths. Green light is highly absorbed by hemoglobin thus it has an higher signal-to-noise ratio, however it can be used only in superficial blood flow measurements.

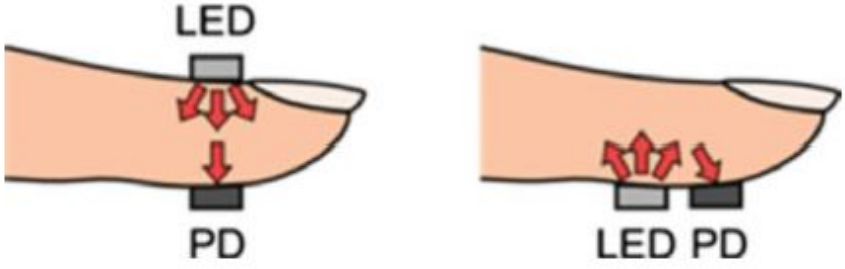


Figure 3.3: On the right there are LED and photodetector placement for transmission mode PPG, while on the left for reflectance mode [37].

4 | Electrocardiogram

An electrocardiogram (ECG) is a graph which shows the electrical activity of the heart, it is obtained using electrodes placed on the skin. Rhythmic cardiac activity is based on repetitive depolarization and repolarization of the entire heart [51]. Depolarization of the heart leads to the contraction of the heart muscles.

An ECG is an indirect indicator of heart muscle contraction and it is measured on the skin surface using electrodes.

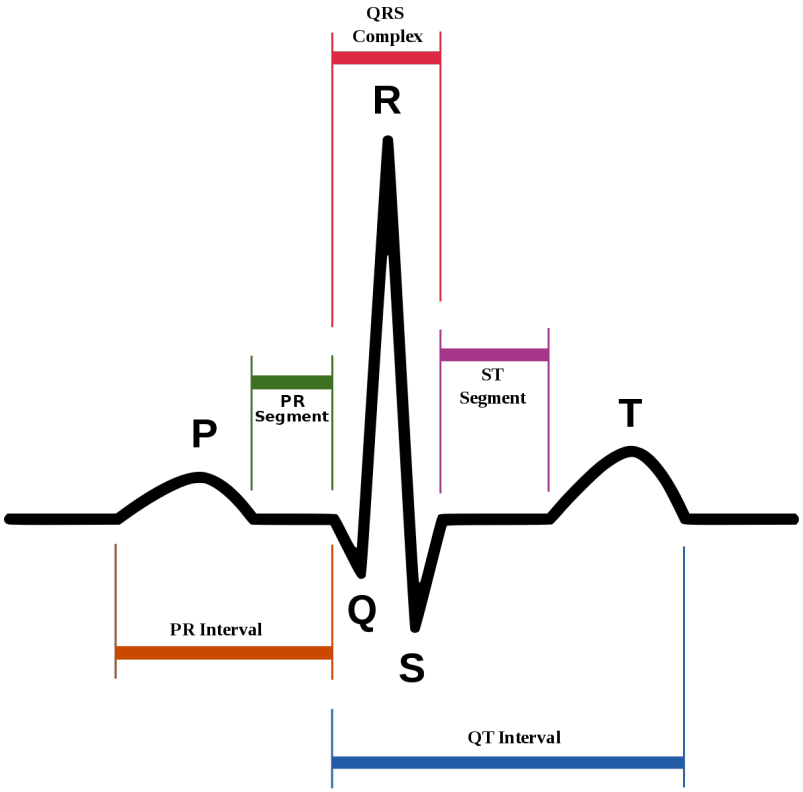


Figure 4.1: Illustration of the Electrocardiogram over an heart cycle

ECG signals can be resolved into heartbeats, each of which represents, from a physiological point of view, a cardiac cycle.

The depolarization starts in the sinoatrial node, located in the upper area of the right atrium, it corresponds to the P wave on the ECG. The action potential generated in the sinoatrial node spreads throughout the atria, then it strikes the atrio-ventricular node wich is situated in the lower portion of the right atrium. After 100 ms the atrioventricular node sends an impulse to start contraction in both ventricles, seen in the QRS wave. At the same time, the atria re-polarize and relax. The amplitude of the R wave is typically much larger than the P wave because ventricular systole involves a

much larger number of muscle cells than atrial systole. The QRS complex has a typical length which spans from 60 ms to 100 ms. Lastly, the ventricles are re-polarized and relaxed, on ECG this is represented by the T wave [27]. Full ECG is represented in fig. 4.1.

The time between the beginning of the P wave and the beginning of the QRS complex is the interval between the beginning of electrical excitation of the atria and the beginning of excitation of the ventricles. This period is called the PQ interval, or PR interval because usually the Q wave is absent. The normal PQ interval is about 0.16 second. From the beginning of the Q wave to the end of the T wave is the so-called QT interval, which usually last 0.35 second. It represents the contraction of the ventricles. The ST segment represents the isoelectric period when the ventricles are in between depolarization and repolarization, it has a duration of 0.005 to 0.150 sec (5 to 150 ms). The PR segment is the flat, usually isoelectric segment between the end of the P wave and the start of the QRS complex, it represents the time delay between atrial and ventricular activation.

All recordings of ECGs are made with appropriate calibration lines on the recording paper, fig. 4.2. The horizontal calibration lines are arranged so that 10 of the small line divisions upward or downward in the standard ECG represent 1 millivolt, with positivity in the upward direction and negativity in the downward direction. The vertical lines on the ECG are time calibration lines. A typical ECG is run at a paper speed of 25 millimeters per second, although faster speeds are sometimes used.

Several cardiac diseases can be recognized observating the ECG tracing (morphology and interval lenght), including abnormal heart rhythm, inadequate coronary artery blood flow and electrolyte disturbances. However, there is no evidence of beneficial use in prevention of ECGs among those without symptoms or at low risk of cardiovascular diease, nevertheless it is still used on persons employed in critical occupations and sometimes on adolescents in order to prevent hypertrophic cardiomyopathy [51].

Various modalities for ECG acquisition have been developed through years. Early ECG machines were constructed with analog electronics and the signal was printed on paper. Today, electrocardiographs use analog-to-digital converters to convert the electrical activity of the heart to a digital signal, usually they are now portable and commonly include a screen, keyboard, and printer. Also smaller device are being developed combining the electrode technologies with microelectromechanical systems (MEMS), integrated circuits, and nanomaterials, the goal is to get continuous ECG measurements in the most comfortable way, i.e. including it in fitness trackers or smart watches.

4.1 Leads

ECG is acquired using electrodes attached to the body, they measure the electrical potential difference between the two corresponding locations of attachment. A pair of electrodes forms a lead.

Leads can also be formed between a physical electrode and a virtual electrode, i.e. the Wilson's central terminal which is a measure of the average potential of the body and can be taken as a reference for the limbs electrodes.

It is possible to get an ECG utilizing only 4 electrodes placed on the limbs, however, in order to have a greater definition of cardiac activity were introduced electrodes

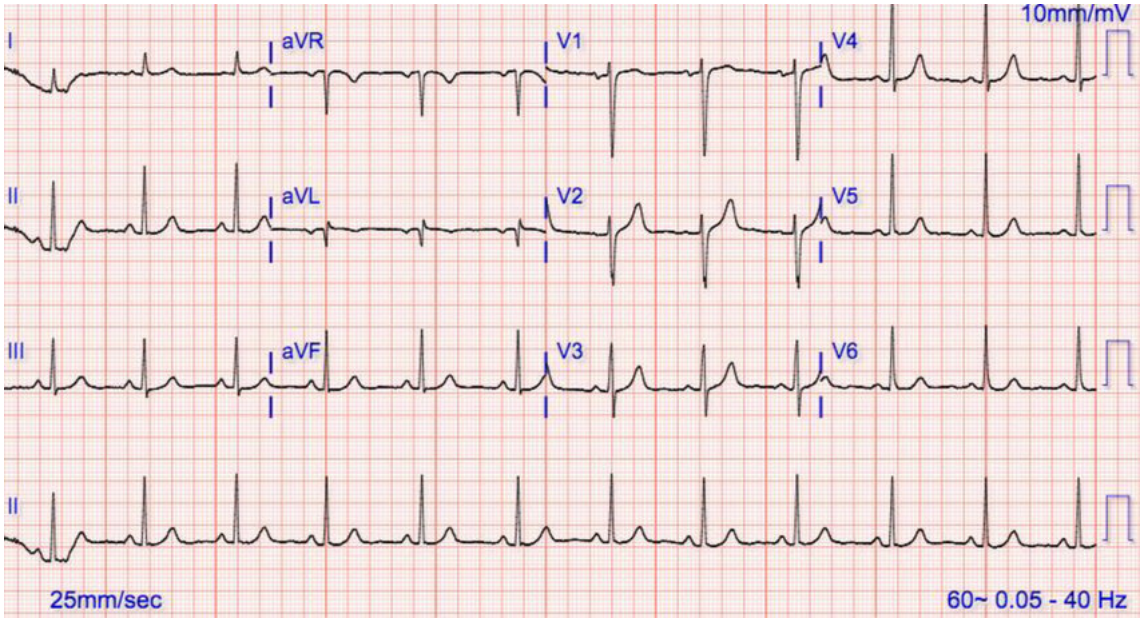


Figure 4.2: A standard 12-lead ECG

closer to the heart. In particular, these new electrodes can identify with more accuracy lesions invisible to standard leads. Also, they grant a full description of the heart's electrical depolarization. [51]

For this reason, standard ECGs have 12 leads, fig. 4.3, grouped in 3 categories and are acquired using 10 electrodes, 4 placed on the limbs and 6 on the chest.

Every electrode has a name and a precise placement:

- Right Arm Electrode (RA): on the right arm;
- Left Arm Electrode (LA): on the left arm;
- Right Leg Electrode (RL): on the right leg;
- Left Leg Electrode (LL): on the left leg;
- V1: in the 4th intercostal space to the right of the sternum;
- V2: in the 4th intercostal space to the left of the sternum;
- V3: midway between V2 and V4;
- V4: in the 5th intercostal space at the midclavicular line;
- V5: on the anterior axillary line at the same level as V4;
- V6: on the midaxillary line at the same level as V4 and V5.

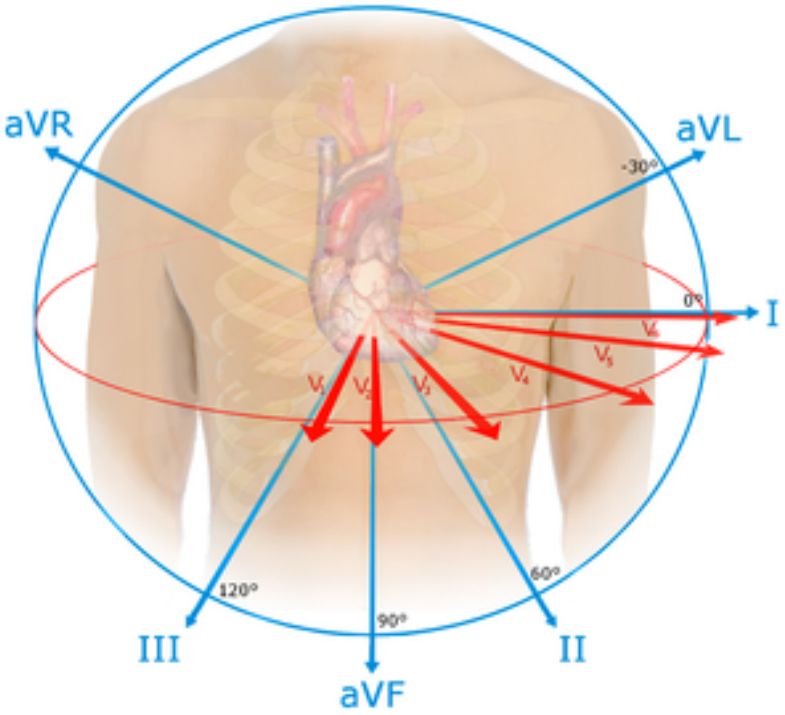


Figure 4.3: ECG leads

Limb leads

$$\begin{aligned} I &= LA - RA \\ II &= LL - RA \\ III &= LL - LA \end{aligned} \tag{4.1}$$

Limb leads are bipolar leads because they are expressed as potential differences between two electrodes. The limb leads form the points of what is known as Einthoven's triangle.

The sum of the voltages around any closed path around the triangle is equal to zero. As a consequence, a virtual ground point can be derived from limb leads: the Wilson's central terminal, which is defined as the average of the vertices of the triangle.

$$Vw = \frac{1}{3}(RA + LA + LL) \tag{4.2}$$

Augmented leads

Augmented limb leads are obtained from the same electrodes which give limb leads. However, they are unipolar leads, which means only one terminal at a time is used to define leads along with Goldberger's central terminal as a reference. Goldberg's terminal is preferred over wilson's one because it allows to measure a signal with a greater amplitude.

Goldberger's central terminal is a combination of inputs from two limb electrodes, with a different combination for each lead. Augmented leads are defined as follow:

$$
\begin{aligned}
aVR &= RA - \frac{1}{2}(LA + LL) = \frac{3}{2}(RA - V_W) \\
aVL &= LA - \frac{1}{2}(RA + LL) = \frac{3}{2}(LA - V_W) \\
aVF &= LL - \frac{1}{2}(RA + LA) = \frac{3}{2}(LL - V_W)
\end{aligned}
\tag{4.3}
$$

Limb leads and augmented limb leads give informations about the electrical activity of the heart in the frontal (vertical) plane.

Precordial leads

The precordial leads, also called unipolar chest leads, lie in the transverse (horizontal) plane. The six precordial electrodes act as the positive poles for the six corresponding precordial leads, while Wilson's central terminal is used as the negative pole.

Because the heart close to the electrodes, relatively minute abnormalities in the ventricles, particularly in the anterior ventricular wall, can cause marked changes in the ECGs recorded from individual chest leads.

Part II

Neural Network principles

5 | Artificial Neural Networks

An artificial neural network (ANN) or simply neural network (NN) is a machine learning (ML) model inspired by the networks of biological neurons first introduced in 1943 McCulloch and Pitts [25]. NNs are becoming more and more important since they frequently outperform other ML techniques on very large and complex problems and modern computers can manage train large networks in a reasonable amount of time.

These systems learn to perform tasks by considering examples. A NN is a network of simple computational units, called neurons, connected by links, called synapsis. In NN implementations, the "signal" at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called edges.

Knowledge is stored in numbers associated to neurons and edges: the weights. The weights are modified by a learning algorithm according to a loss function, which models the observed errors. Adjusting these weights the network can learn and improve the accuracy of its predictions. Learning is complete when examining additional observations does not usefully reduce the error rate.

A Neural network is defined by several elements, the most important are:

- **Neurons:** The neurons are the basic information processing units of a NN, they receive an input, pass it through an activation function, and produce an output using an output function. The activation function or squashing function limits the amplitude of the output of the neuron and adds some non linearity.

- **Architecture:** defines the network structure that is the number of artificial neurons in the network and their interconnectivity.

- **Learning algorithm:** is the procedure used to perform the learning process, it modifies the weights to compensate for each error found during learning. One of the most popular methods is based on gradient descent and it is called Backpropagation or generalized delta rule. For each training instance the backpropagation algorithm first makes a prediction (forward pass), measures the error, then goes through each layer in reverse to measure the error contribution from each connection (reverse pass), and finally tweaks the connection weights to reduce the error (Gradient Descent step). How much the weights are modified in a single step is decided according to the learning rate, which is multiplied to the error.

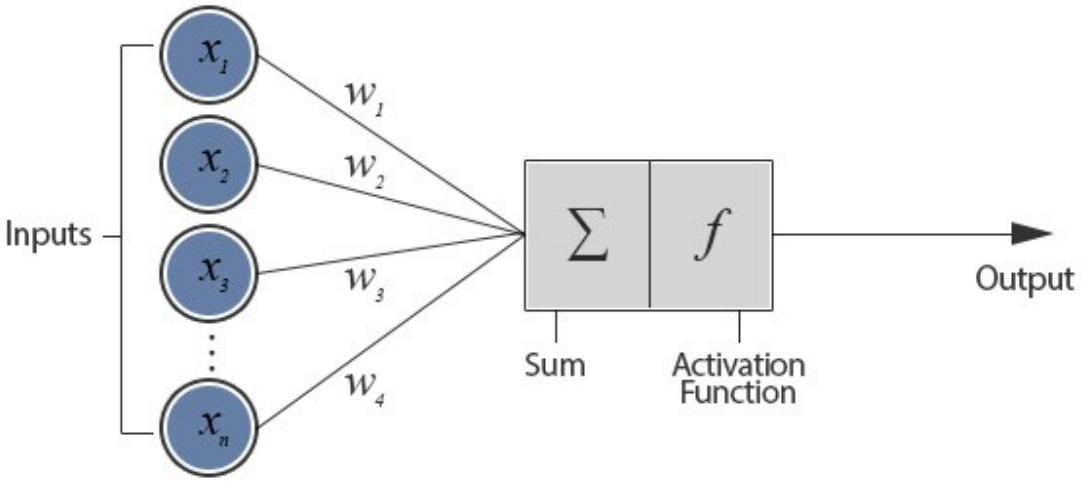


Figure 5.1: TLU, every input is associated with a weight, the weighted input are summed and then an activation function is applied.

5.1 Multilayer perceptron

The Perceptron is one of the simplest ANN architectures, invented in 1957 by Frank Rosenblatt. It is based on the threshold logic unit (TLU), the first artificial neuron developed [52]. The inputs and output are numbers, and each input connection is associated with a weight. The TLU computes a weighted sum of its inputs , then applies the Heaviside step function to that sum and outputs the result, fig. 5.1.

The aim of the perceptron is to classify inputs in one of two classis. Thus , in the case of an elementary perceptron the n-dimensional space is dived by a hyperplane into two decision regions.

However, Perceptron can't deal with problem which are not linearly separable, for instance, they are incapable of solving Exclusive OR (XOR) classification. These problems can be overcome stacking several layers of Perceptrons, creating the so-called multilayer perceptron (MLP).

In this architecture the neurons are organized in an input layer receiving input directly from the environment, one or more hidden layers, and an output layer producing the final output of the network, fig. 5.2. Every layer except the output layer includes a bias neuron and is fully connected to the next layer. The layers close to the input layer are usually called the lower layers, and the ones close to the outputs are usually called the upper layers.

The neurons in the hidden layers detect the features; the weights of the neurons represent the features hidden in the input patterns. With one hidden layer we can represent any continuous function of the input data, and with two hidden layers even discontinuous functions can be represented.

This architecture is an example of a fully connected feedforward neural network, which means all the neurons in a layer are connected to every neuron in the previous layer and the signal flows only in one direction: from the input to the output.

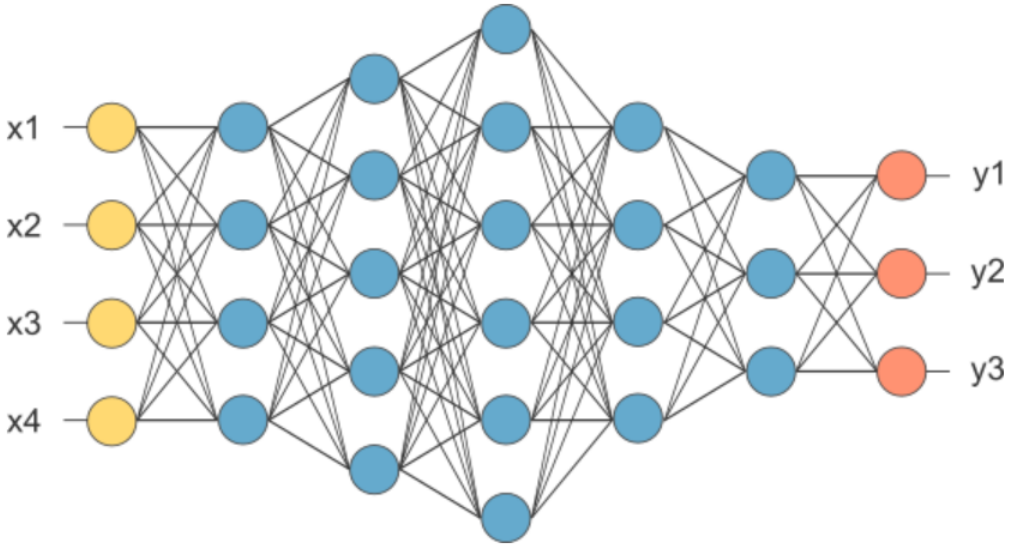


Figure 5.2: A fully connected network with 4 inputs and 3 outputs. The input layer is represented in yellow, hidden layers in blue, output layer in red.

5.2 Hyperparameters

Every machine learning model needs to be tuned for the considered problem, it is the so-called hyperparameter optimization. A hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are learned. In order to achieve the best performances there are a set of options which have to be careful selected, for instance, the network architecture, the learning rate (even though recent optimizers automatically adjust it), activation functions, batch size, etc.

Scaling

Neural networks aims to map an input variable into an output variable, however differences in the scales across input variables may increase the difficulty to model the problem. Indeed large input values can result in a model that consider some input more important than others or that learns large weight values. A model with large weights values is often unstable, meaning that it may suffer from poor performance during learning and sensitivity to input values resulting in higher generalization error.

Also target variable often need rescaling, in fact a target with a large spread of values may result in large error gradient values causing weight values to change dramatically, making the learning process unstable. Furthermore rescaling the target can make the values more easily attainable and therefore faster to optimise for, in fact they are all closer in magnitude to the gradients that are being computed.
Therefore scaling input and output variables is a critical step in using neural network models. [20]

It is nearly always advantageous to apply pre-processing transformations to the input data before it is presented to a network. Similarly, the outputs of the network are often post-processed to give the required output values [6].

Even when the other hyperparameters are tuned, there are still big differences between the results using different scaling methods. Therefore, the scaling method must be considered as a crucial hyperparameter of model.There are two common ways to get all attributes to have the same scale: min-max scaling and standardization.

Min-max, also called normalization, is the simplest: values are shifted and rescaled so that they end up ranging from 0 to 1.

$$y = \frac{x - min(x)}{max(x) - min(x)} \tag{5.1}$$

Standardization, also called whitening: first it subtracts the mean value so standardized values always have a zero mean, and then it divides by the standard deviation so that the resulting distribution has unit variance. Unlike min-max scaling, standardization does not bound values to a specific range, which may be a problem for some algorithms. However, standardization is much less affected by outliers.

$$y = \frac{(x - \mu)}{\delta} \tag{5.2}$$

If the distribution of the quantity is normal, then it should be standardized, otherwise the data should be normalized.

In order to avoid data leakage during model evaluation it is important to fit these scaling techniques always on the training set, so it is important that the training set can give a reliable mean and standard deviation for the standardization or minimum and maximum for the normalization [15].

Loss function

Historically, the most common metric used in regression task is root mean squared error (RMSE), equation 5.3, it tells how big is the difference between the squares of the predicted values and the target values. Since it is a squared difference, it means this metric will give more weight to large errors. It is also called ℓ_2 norm and corresponds to the euclidean norm.

$$RMSE(X,h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2} \tag{5.3}$$

RMSE is, usually, the preferred metric, however, in some context there could be better functions: if there are many outliers mean absolute error (MAE), equation 5.4, could describe the performances with more accuracy.

$$MAE(X,h) = \frac{1}{m}\sum_{i=1}^{m}|h(x^{(i)}) - y^{(i)}| \tag{5.4}$$

MAE (also called Manhattan norm or norm ℓ_1) measures the distance between 2 vectors, the vector of the predicted values and the vector of target values.

There are, also, norm ℓ_0 that gives the number of non-zero elements in a vector and norm ℓ_∞ which gives the maximum absolute value in a vector. In general, it is possible to define any norm ℓ_k, equation 5.5, the higher the index the more the norm will focus on high values, for this reason RMSE weights more outliers and should be preferred when the data have bell-shaped distribution [15].

$$\|v\|_k = (|v_0|^k + |v_1|^k + ... + |v_n|^k)^{\frac{1}{k}} \tag{5.5}$$

However, MAE cannot be utilized as loss function for NN because its gradient is always the same, thus it will be large even for small loss values, for this reason RMSE was the preferred one, fig. 5.3.

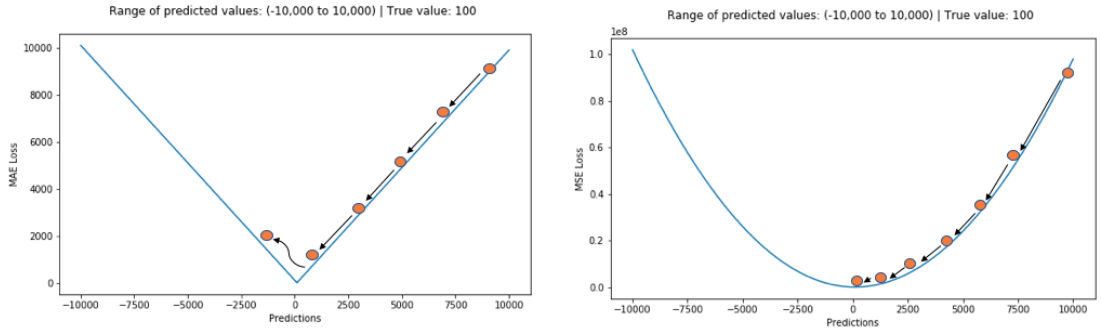


Figure 5.3: Gradient descent on MAE (on the left) and MSE (on the right) [1]

In order to overcome problems regarding MAE, but still have a robust metric unaffected by outliers it was developed the Huber loss, equation 5.6. Like RMSE it is differentiable in zero, but it introduces another hyperparameter that needs to be tuned δ, fig. 5.4 [1].

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{for } |y - f(x)| \leq \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases} \tag{5.6}$$

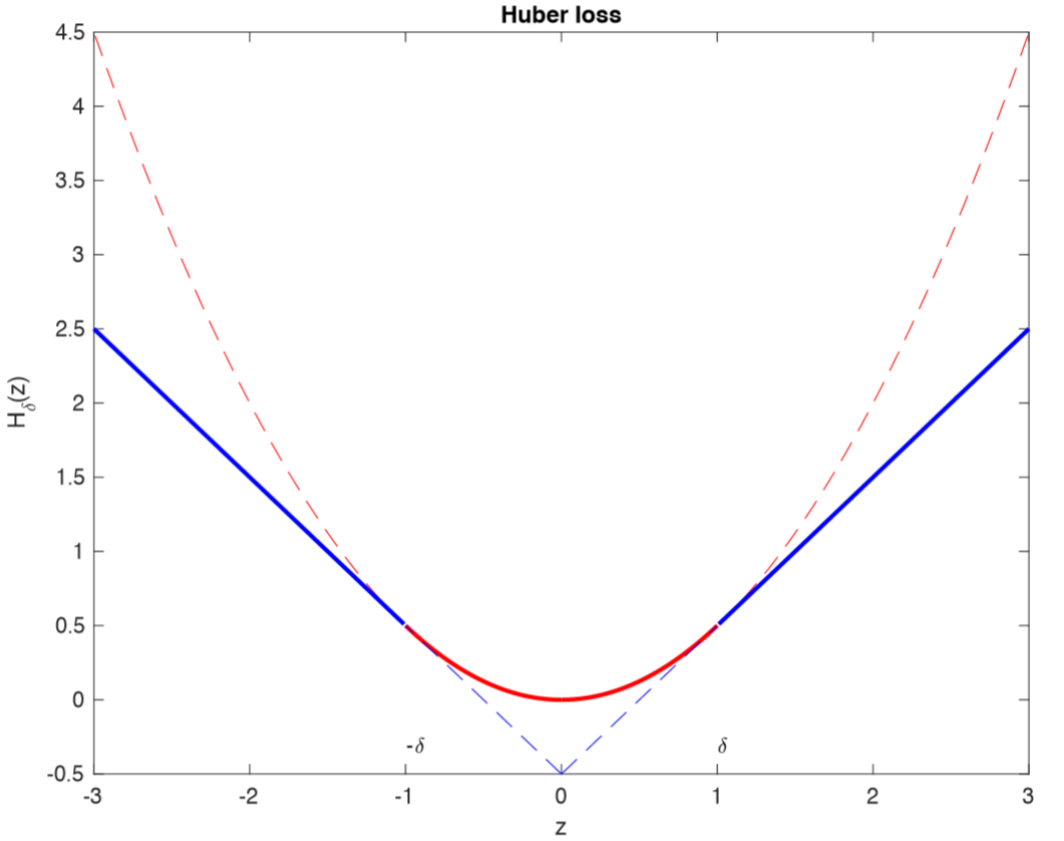


Figure 5.4: Huber loss, red dashed line is MSE, blue dashed line is MAE, $\delta = 1$

Batch size

The batch size can have a significant impact on model's performance and training time. There are three common way to update weights in a neural network using an optimizer like Stochastic Gradient Descent or one of its variants:

- Full gradient: the gradient is being calculated on all the data points at a single shot and the average is taken. Hence we have a smoother version of the gradient takes longer time to learn.
- Stochastic gradient: the gradient is calculated on one-data point at a time hence the gradient is noisy this can be mitigated using Momentum parameters. So there is a chance that your oscillations can make the algorithm not reach a local minimum.
- Batch gradient: the gradient is being calculated on a batch at a time. Bigger batches behaves similar to full gradient computation, smaller batches are similar to stochastic gradient computation.

The main benefit of using large batch sizes is that hardware accelerators like GPUs can process them efficiently, increasing the available computational parallelism. Therefore, many researchers and practitioners recommend using the largest batch size that can fit in GPU RAM. However, computing gradients on big batches average them over potentially a vast amount of information. It takes lots of memory to do that, but the real problem is the batch gradient trajectory land in a saddle point.

Indeed large batch sizes often lead to training instabilities, especially at the beginning of training, and the resulting model may not generalize as well as a model trained with a small batch size [15]. On the other hand updating the gradient after every random instance can make it very noisy. However, the noisiness is exactly what you want in non-convex optimization, because it helps you escape from saddle points or local minima. [14]

The minibatch methodology is a compromise that injects enough noise to each gradient update, while achieving a relative speedy convergence. In particular small batch training has been shown to provide improved generalization performance and allows a significantly smaller memory footprint, which might also be exploited to improve machine throughput. Therefore modern deep neural network training is typically based on mini-batch stochastic gradient optimization and in particular batch sizes should always lie between 2 and 32, fig. 5.5.

Some papers point in the opposite direction, however, showing that it is possible to use very large batch sizes, up to 8192, using various techniques such as warming up the learning rate [9] [32].

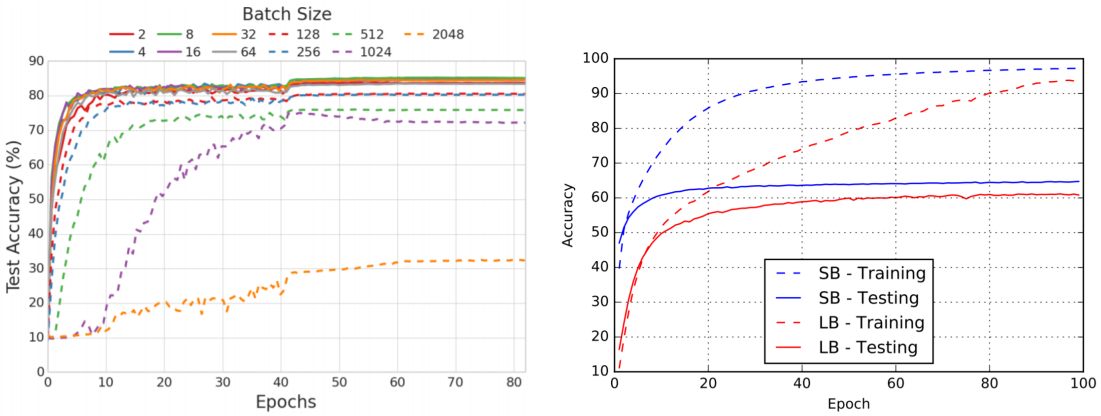


Figure 5.5: Training and testing accuracy for small batch and large batch methods as function of epochs. The graph on the left was obtained using a ResNet-32 on CIFAR10 dataset [24], while the one on the right using a fully connected NN on Timit dataset [21]

6 | Convolutional Neural Networks

Also convolutional neural network (CNN) were inspired by animal's visual cortex and are mainly known for their applications in image recognition even though CNNs are not restricted to visual perception: they are also successful at many other tasks, such as voice recognition, natural language processing and in general time series analysis.

Neurons in the visual cortex usually have a small local receptive field, meaning they react only to visual stimuli located in a limited region of the visual field, also some neurons may have the same receptive field but react to different line orientations. Other neurons have larger receptive fields, and they react to more complex patterns that are combinations of the lower-level patterns. These observations led to the idea that the higher-level neurons are based on the outputs of neighboring lower-level neurons [15].

The core building block of a CNN is the convolutional layer. Neurons in every convolutional layer are not connected to every single input point, but only to those in their receptive fields, fig. 6.2. The shift from one receptive field to the next is called the stride. This architecture allows the network to concentrate on small low-level features in the first hidden layer, then assemble them into larger higher-level features in the next hidden layer, and so on.

Convolutional layers are build on top of the so called called convolutional kernels or filters, during the forward pass each filter is convolved on its input producing a 2-dimensional activation map of that filter. This map is called feature map.

Convolutional layers usually have multiple filters and outputs one feature map per filter. All neurons within a given feature map share the same the same weights and bias term.

A neuron's receptive field is the same as described earlier, but it extends across all the previous layers' feature maps. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input. The

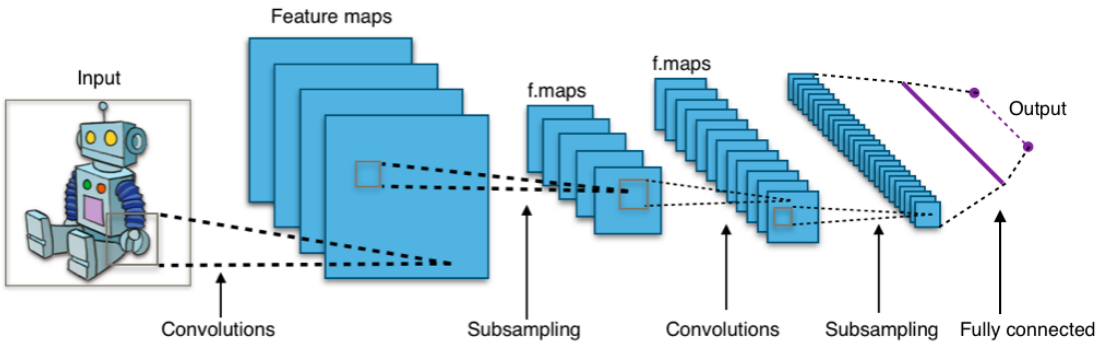


Figure 6.1: Schematic convolutional neural network

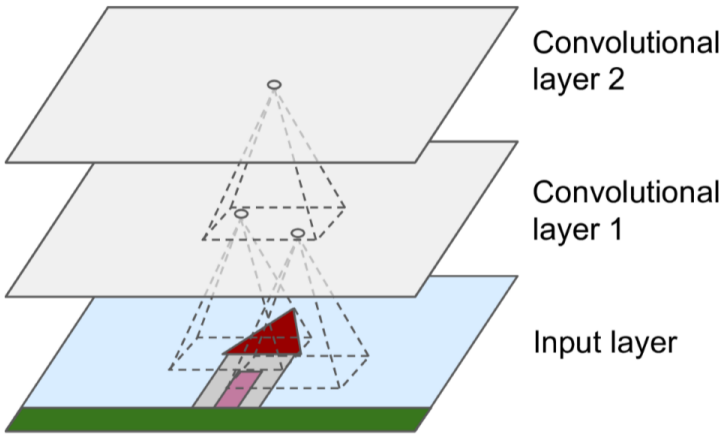


Figure 6.2: Receptive field of each neuron in different layers

fact that all neurons in a feature map share the same parameters dramatically reduces the number of parameters in the model.

Similarly, a 1D convolutional layer slides several kernels across a sequence, producing a 1D feature map per kernel. Each kernel will learn to detect a single very short sequential pattern (no longer than the kernel size). It is possible to use only 1D convolutional layers in a time series analysis. However they shine when combined with recurrent layers in fact they can extract feature from a signal and they can also downsamples the input sequence using the right kernel size, stride and padding. Indeed, the model can learn to preserve the useful information dropping only the unimportant details and shortening the sequences, the convolutional layer may help the following recurrent layers to detect longer patterns [15].

Another important concept of CNNs is pooling, which is a form of non-linear downsampling in order to reduce the computational load, the memory usage, and the number of parameters. Just like convolutional layer it has a receptive field, however it has no weights, it just aggregate the inputs using an aggregation function. There are several non-linear functions to implement pooling among which max pooling is the most common. It partitions the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum [53].

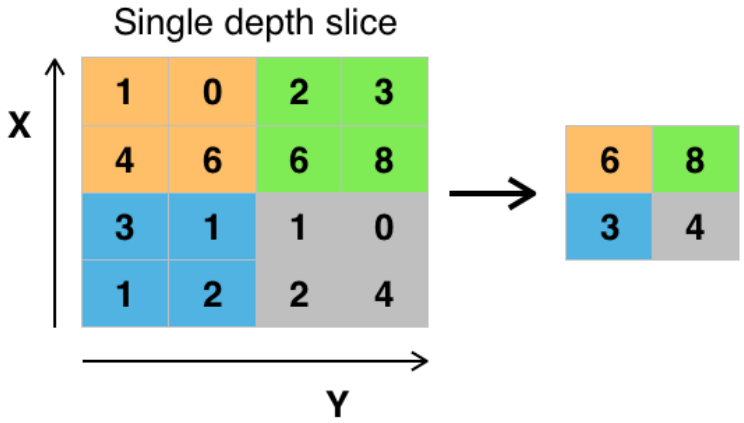


Figure 6.3: Max pooling with a $2x2$ filter and stride = 2

Typical architecture based con convolutional layers are: ResNet and WaveNet.

6.1 ResNet

Residual Network or ResNet is an architecture inspired by pyramidal cells in the cerebral cortex developed by Kaiming He et al and originally used for image classification [18].

The reason why it was created is that there was evidence revealing that network depth is of crucial importance, however, deep neural network training is difficult to perform due to vanishing/exploding gradients and degradation of the accuracy. The first problem was solved normalizing initialization and intermediate layers. The second problem involves saturated accuracy that degrades rapidly. Such degradation is not caused by overfitting, and adding more layers to a suitably deep model leads to higher training error [18].

In order to avoid degradation He introduced skip connections, fig. 6.4 the signal feeding into a layer is also added to the output of a layer located a bit higher up the stack. This new technique allowed to train very deep networks like the original ResNet, a CNN composed composed of 152 layers. There are many variants of this net depending on how much is deep.

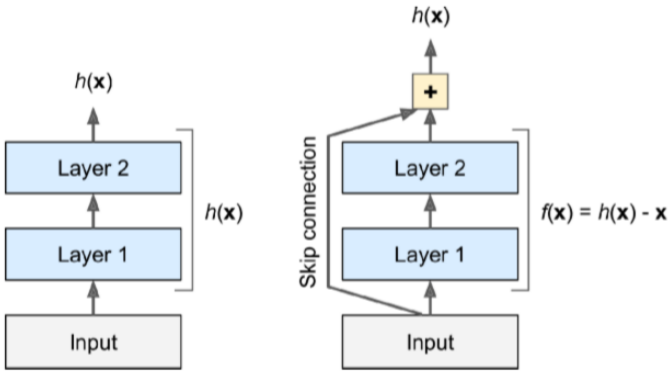


Figure 6.4: Skip connection

Usually when training a neural network, the goal is to make it model a target function $h(x)$, also called underlying mapping, however is difficult to optimize it when the network is deep. For this reason the input x is added to the output of the network forcing the network to learn the so called residual map $f(x) = h(x)–x$ .This is called residual learning. When a regular neural network is initialized its weights are close to zero, so the network just outputs values close to zero. If there is a skip connection, the resulting network just outputs a copy of its inputs; in other words, it initially models the identity function. If the target function is fairly close to the identity function (which is often the case), this will speed up training considerably [15].

Moreover, if there are many skip connections, the network can start making progress even if several layers have not started learning yet, fig. 6.5. Thanks to skip connections, the signal can easily make its way across the whole network. The deep residual network can be seen as a stack of residual units, where each residual unit is a small neural network with a skip connection.

ResNet's architecture, fig. 6.6 starts with a convolutional layer and ends with a fully connected layer, the original one since it was used for a classification task ends with

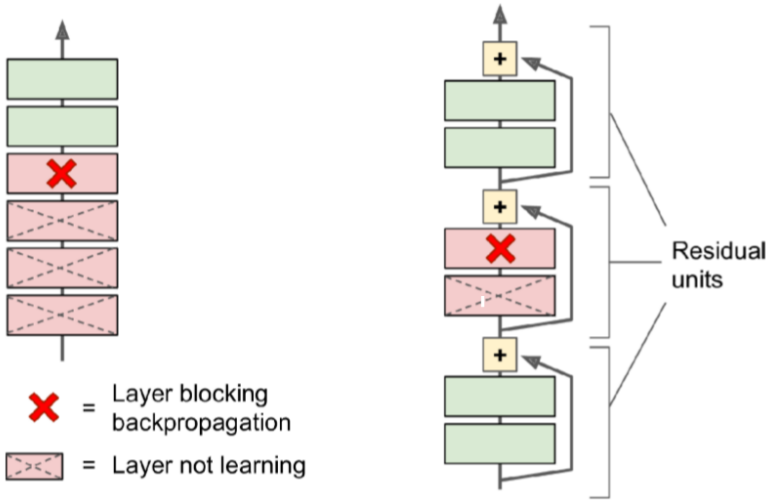


Figure 6.5: Skip connections allow backpropagation to modify weights in deep layers even though superficial layers have not started learning

a softmax as activation function, and in between is just a very deep stack of simple residual units. Each residual unit is composed of two convolutional layers, with Batch Normalization and ReLU activation, using 3×3 kernels and preserving spatial dimensions). There is no need of pooling layers in the residual units because downsampling is performed directly by convolutional layers that have a stride of 2.

The convolutional layers mostly have 3×3 filters and follow two simple design rules: for the same output feature map size, the layers have the same number of filters; and if the feature map size is halved (using a convolutional layer with stride 2), the number of filters is doubled so as to preserve the time complexity per layer. When the height and width are halved the inputs cannot be added directly to the outputs of the residual unit because they don't have the same shape. To solve this problem, the inputs are passed through a 1×1 convolutional layer with stride 2 and the right number of output feature maps [15].

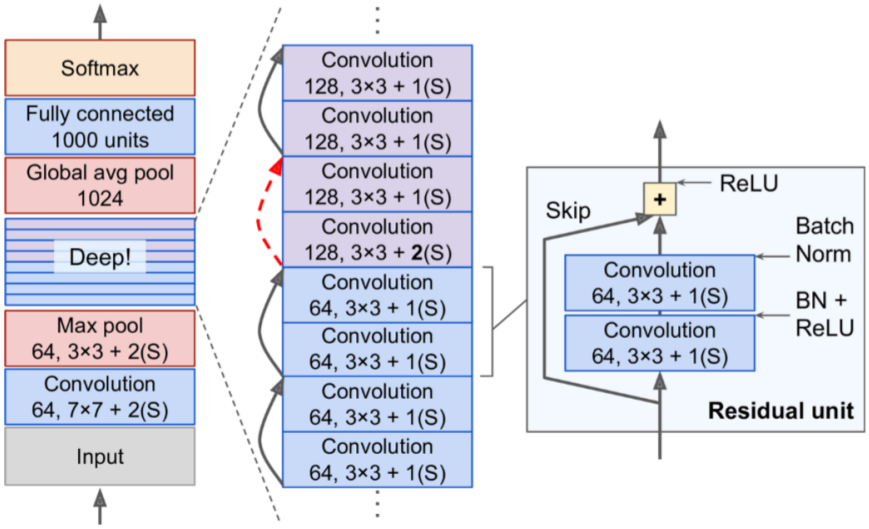


Figure 6.6: Schematic ResNet architecture

6.2 WaveNet

An architecture developed by Aaron van den Oord et al. in 2016 [30], it is a model originally designed to operate directly on the raw audio waveform. In its simplest variant it is just a stack of convolutional layers without pooling layers and with a particular type of padding: causal. This padding allows the output to have the same time dimensionality as the input.

Since this model does not require recurrent connections, it is typically faster to train than RNN, especially when applied to very long sequences. However, one of the problems of causal convolutions is that they require many layers, or large filters to increase the receptive field [30]. In order to solve this problem the WaveNet utilizes a dilation rate, fig. 6.7, which represents how spread apart each neuron's inputs are.

A dilated convolution is a convolution where the filter is applied over an area larger than its length by skipping input values with a certain step. This way, the lower layers learn short-term patterns, while the higher layers learn long-term patterns. Thanks to the doubling dilation rate, the network can process extremely large sequences very efficiently [15].

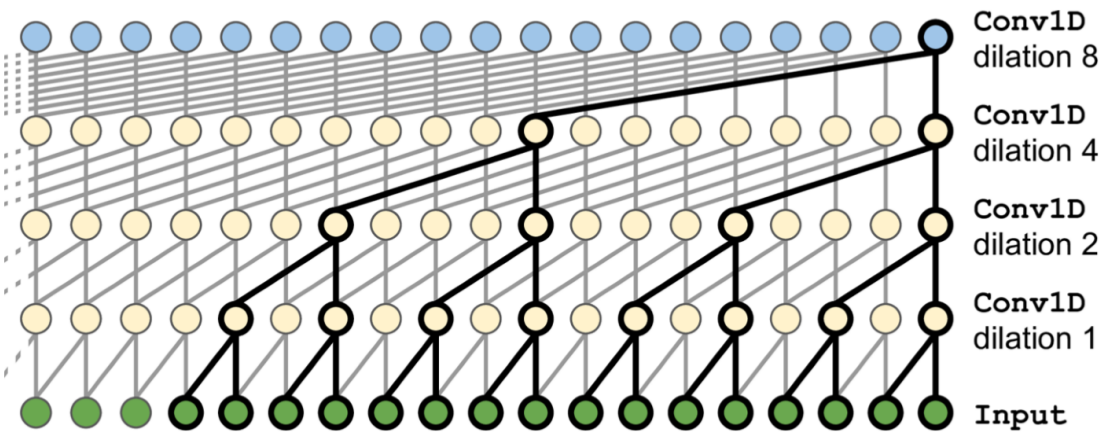


Figure 6.7: Dilated convolution

The original WaveNet has 10 stacked convolutional layer with a dilation rate up to 512. A single stack of 10 convolutional layers with these dilation rates act like a super-efficient convolutional layer with a kernel of size 1024, but is significantly more efficient. In particular, the original WaveNet is a a deep neural network for generating raw audio waveforms which output a categorical distribution over the next expected sample, since every sample x_t is conditioned by previous timesteps.

At training time, the conditional predictions for all timesteps can be made in parallel because all timesteps of ground truth x are known. When generating with the model, the predictions are sequential: after each sample is predicted, it is fed back into the network to predict the next sample.

The network utilizeds softmax to model the conditional distribution over the individual audio samples, it performs well because categorical distribution is more flexible and can more easily model arbitrary distributions because it makes no assumptions about their shape [30].

Because raw audio is typically stored as a sequence of 16-bit integer values (one per timestep), a softmax layer would need to output 65536 probabilities per timestep

to model all possible values. To make this more tractable, the authors applied a μ-law companding transformation to the data, and then quantized it to 256 possible values.

7 | Recurrent Neural Networks

Unlike humans, traditional neural networks restart thinking from scratch every second, they do not have memory. This is crucial in certain tasks like reading where the meaning of each word is based on the previous ones. Also, feedforward NN accept a fixed-sized vector as input and produce a fixed-sized vector as output, which is incompatible with time series analysis.

For these reasons, recurrent neural network (RNN) were introduced in the late 80's. They are networks with loops in them, fig. 7.1, allowing information to persist.

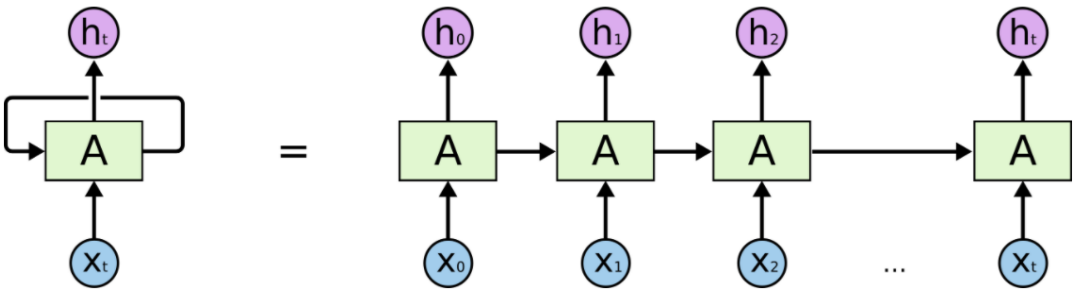


Figure 7.1: RNN unrolled through time

Many different architectures can be built with RNNs [15]:

- Sequence-to-sequence: it takes a sequence of inputs and produce a sequence of outputs, mostly used in time series predicion, i.e. stock prices
- Sequence-to-vector: it takes a sequence of inputs and produce a one output (every outputs except the last one are ignored), it is used in sentiment analysis
- Vector-to-sequence: it takes the same input repetitively at each timesteps and produce a sequence, it is used to automatically create image captions
- Encoder-decoder: it is composed by two blocks, the encoder, which is a sequence-to-vector RNN, and a decoder, which is a vector-to-sequence RNN. It is used in deep learning translator.

Despite the chosen architecture, at each time step every RNN receives inputs, it produces an output and then it sends the output back to itself. The network will use the last output together the next input to produce a new output. A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor. Since the output of a recurrent neuron at time step is a function of all the inputs from previous time steps it has a form of memory. A part of a neural network that preserves some state across time steps is called a memory cell.

In particular, at each timestep t every neurons takes as input the current input $x(t)$ and the last output $y(t-1)$, it combines the entries with their respective weights (W_x and W_y) and then it produce the output $y(t)$, equation 7.1.

$$y_{(t)} = \phi(W_x^\top x_{(t)} + W_y^\top y_{(t-1)} + b) \tag{7.1}$$

The most common activation functions used in RNN modules are sigmoid, tanh and ReLU. In this kind of networks, however, the tanh is the preferred one because it is less prone to exploding/vanishing gradients, which are a big problem in this kind of network. Indeed, it is difficult to capture long term dependencies because of multiplicative gradient that can be exponentially decreasing/increasing with respect to the number of layers.

Also, the loss function $\mathscr{L}$ of all time steps is defined based on the loss at every timestep, equation 7.2. $\hat{y}$ represents the predicted output, y is the expected output.

$$\mathscr{L}(\hat{y}, y) = \sum_{t=1}^{T_y} \mathscr{L}(\hat{y}_{(t)}, y_{(t)}) \tag{7.2}$$

RNNs are trained using the backpropagation through time, which means backpropagation is done at each point in time, equation 7.3, however when the input sequence is long the unrolled network becomes deep, fig. 7.1. Thus, like every deep NN it suffers from unstable gradients, moreover it may forget the first input of the sequence. In addition to tanh activation function, to avoid exploding/vanishing gradients in this kind of network is often used the gradient clipping, which simply caps the maximum value for the gradient.

$$\frac{\partial \mathscr{L}^{(T)}}{\partial W} = \sum_{t=1}^{T} \left.\frac{\partial \mathscr{L}^{(T)}}{\partial W}\right|_{(t)} \tag{7.3}$$

Due to these issues, several types of memory cells were studied. The most famous is probably the Long Short Term Memory (LSTM) cell.

7.1 LSTM

LSTMs are explicitly designed to avoid the long-term dependency problem.

This kind of cells have a long-term state, at every iteration the network learn what to store and what to read from it.

The cell regulates its state using the gates, fig. 7.2, first there is a forget gate where some memories are dropped, then the memories are replaced with new ones selected by the input gate. One copy of the new state is sent to the next iteration; the other one is passed through a tanh function and filtered by the output gate. This is combined with the current inputs and the previous outputs to create the new output. The input gate recognizes important inputs and store them in the long-term state, while the forget gate deletes input that are no longer needed, the output gate decide when to extract a specific input from the long-term state.

The current input and the previous output, also called short-term state, are fed to four different fully connected layers.

The ones controlled by a sigmoid function are the layers that control the gates, their outputs is between 0 and 1 and are fed to element-wise multiplication operations, so if they output 0s they close the gate, and if they output 1s they open it. The forget gate

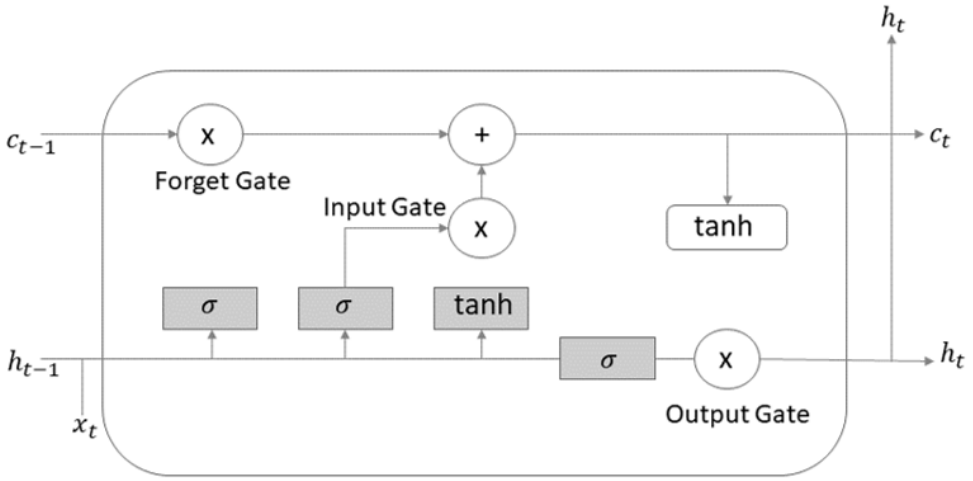


Figure 7.2: LSTM cell

controls which parts of the long-term state should be erased, the input gate controls which new memories should be added to the long-term state, the output gate controls which parts of the long-term state should be read and output at this time step. The new memories are calculated in the layer controlled by the tanh function.

Part III

Methods

8 Dataset

8.1 MIMIC database

To evaluate how PPG, ECG and ABP are related the MIMIC database was used. It was chosen because it should be representative of the full range of pathophysiologies that result in sudden blood pressure changes [29] [2]. This database consists of different physiological signals recorded from 121 ICU patients; however, in its first release, the one used in this thesis, only 72 patients were available.

The data include signals and periodic measurements obtained from a bedside monitor as well as clinical data obtained from the patient's medical record. The recordings vary in length from 1 to 80 hours depending on patients. The data obtained from the bedside monitors are divided into files each containing 10 minutes of recorded signals, which can then be assembled without gaps to form a continuous recording [28].

The data were written in ten-minute segments in order to limit possible loss of data from power interruptions. The ECG, PPG and ABP signals are sampled at 125 Hz with 12-bit precision and negligible jitter [22].

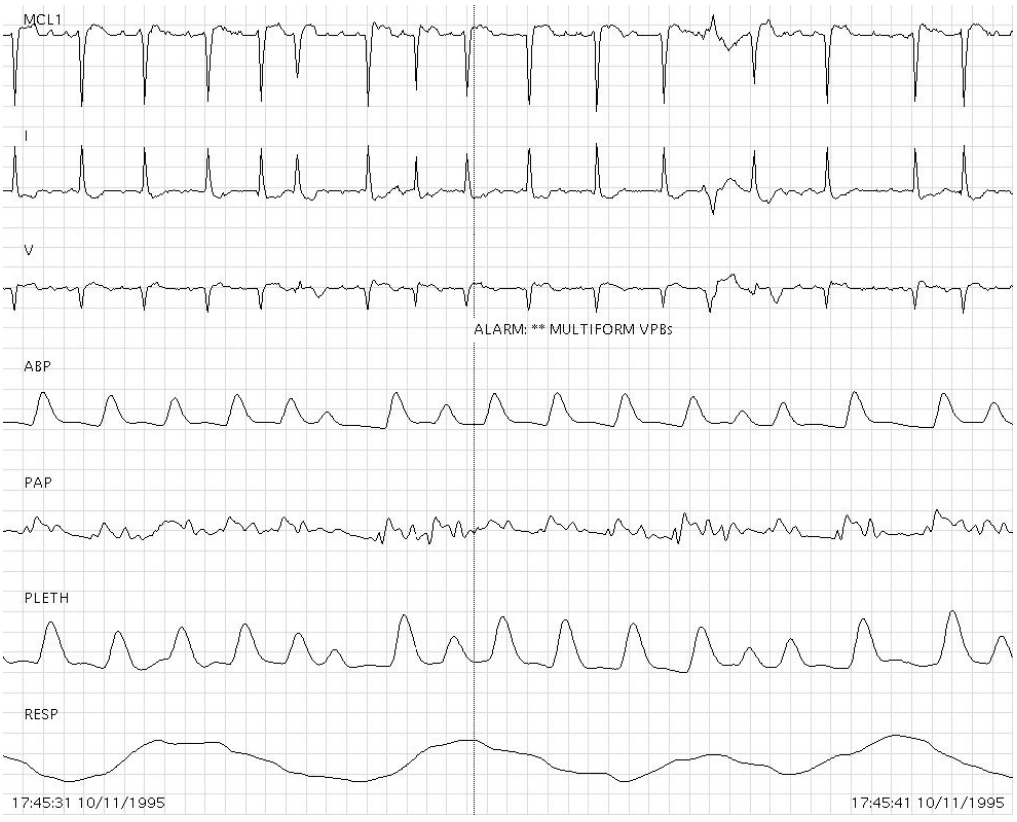


Figure 8.1: Patient 248 visualized using WAVE, the official physionet tool to open MIMIC dataset

8.2 Data cleaning

MIMIC was extracted via WFDB [41], python library supported by physionet [39], and then every record without the requested signals were discarded.

Preprocessing pipeline is based on the histogram of SBP and DBP distributions, fig. 8.2. In order to extract the SBP and DBP values, it was applied on the raw dataset the algorithm developed by [10].

From the graph several problems emerged: first of all, there are negative BP values, some astonishing high values and a unusual peak at 180 mmHg. Finally, it is possible to make some considerations on the distributions: both distributions are heavily skewed towards physiological values, but SBP appears to have much larger support.

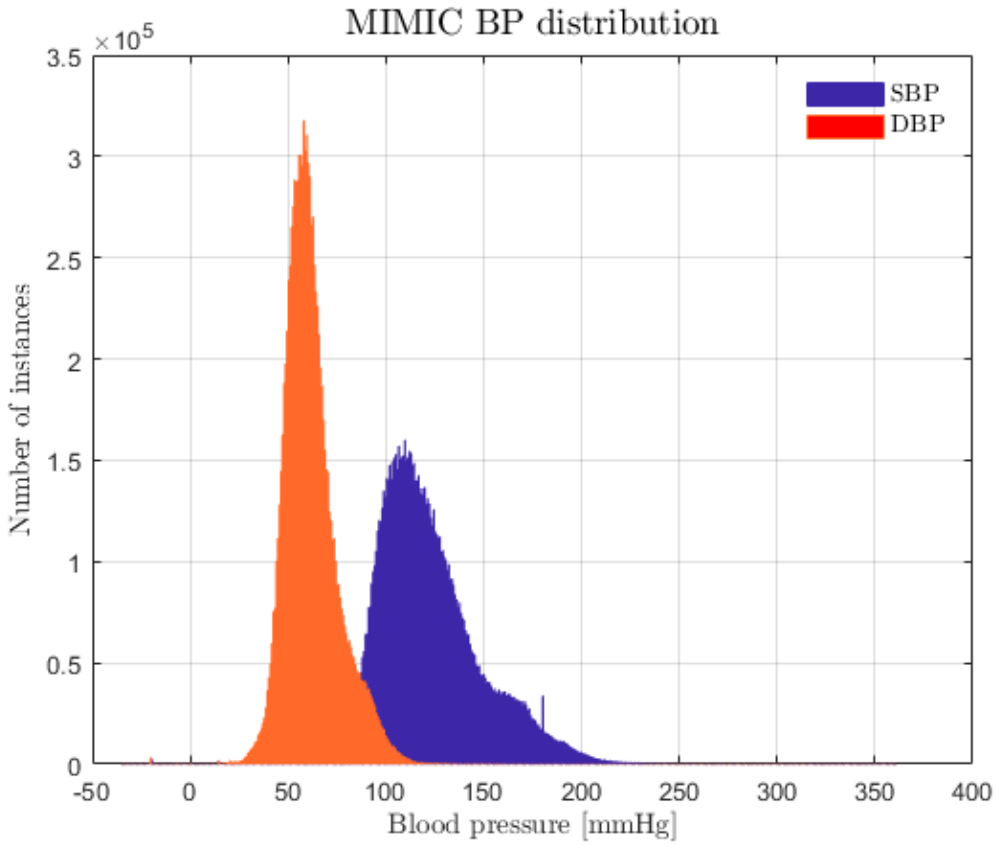


Figure 8.2: MIMIC SBP and DBP distributions

Since MIMIC is organized in 10 minutes recordings, in order to maintain consistency the following preprocessing pipeline, fig. 8.3, was applied on 10-minutes blocks of recordings.

- Searching for NaN values and replacing them with the closest recorded data available;
- Deleting records with *flat lines* in the ABP or PPG;
- Deleting where more than 5% of the ABP or PPG peaks were *flat peaks*;
- Removal of records containing abnormal ABP or PPG;

- PPG filtering using a 4th order Butterworth filter;
- PPG and ABP outlier removal through Hampel filter;
- Removal of patients with less than 3 hours of recording time;
- Standardization of the PPG, ABP normalization.

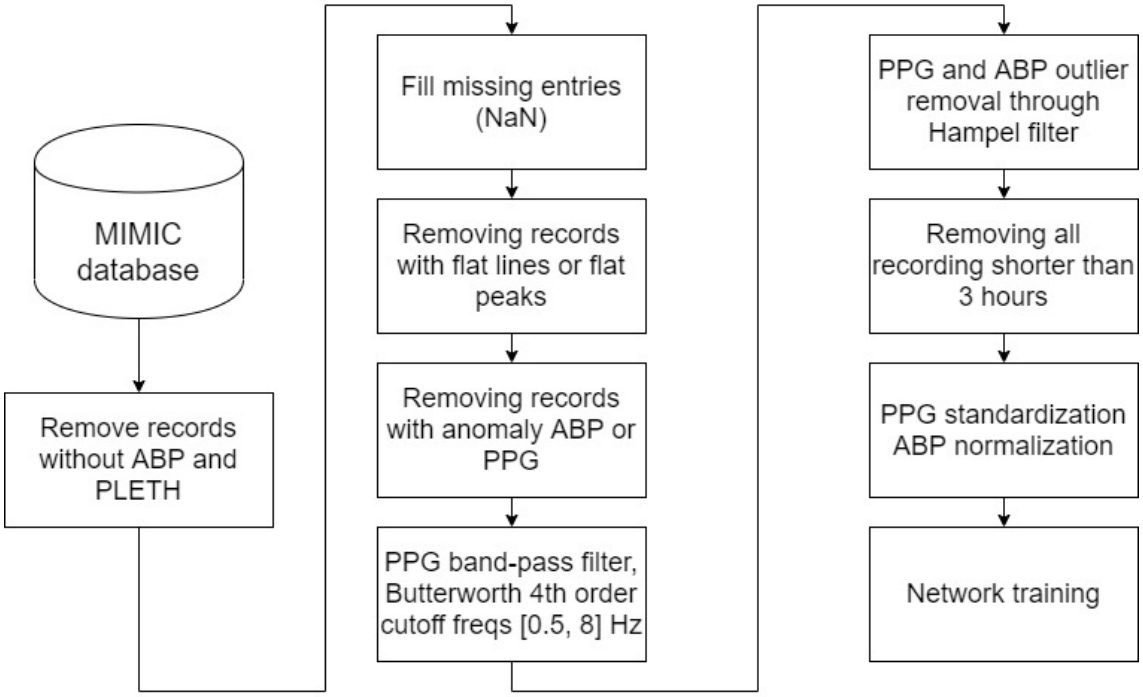


Figure 8.3: Preprocessing scheme

After deleting recordings without the requested signals (ECG, PPG, ABP), NaN values were managed replacing them with the first available value.

The NaN values were simply replaced with the closest value because in some cases there were long stretches with missing values, so it was impossible to reconstruct the missing signal. Moreover, replacing the NaNs with the nearest value was not a problem because they were associated to *flat lines*, which are managed in the next step of the pipeline.

Then, it was necessary to exclude low quality recordings, i.e. those containing the so-called *flat lines* and *flat peaks*, which are recording errors mostly due to sensor problems, for example, a simple disconnection. Flat lines, fig. 8.4 a, are long periods of time where the same value is always detected, while flat peaks, fig. 8.4 b, are peaks with a flattened tip.

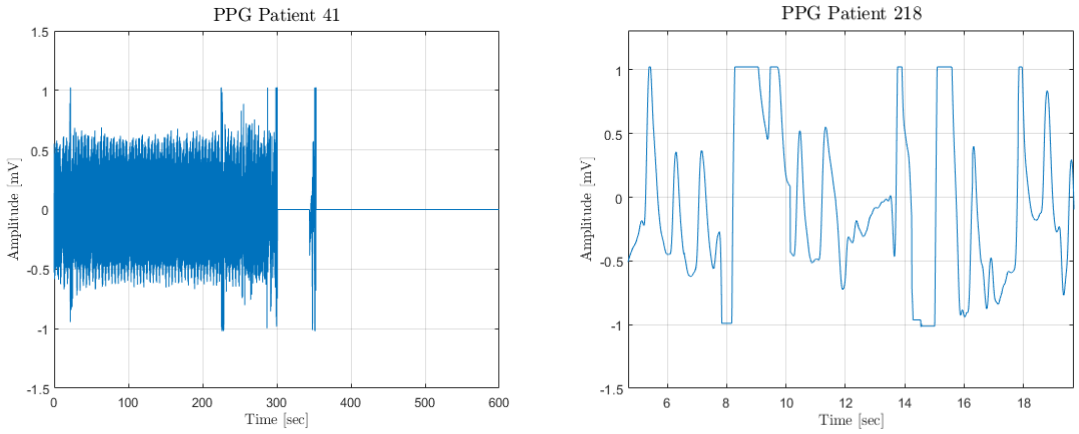
Subsequently, anomalies in ABP signal were managed: within the 10-minute recording the ABP signal should always be between a minimum of 15 and a maximum of 300mmHg, fig. 8.5 a. In addition, a control on the pressure and plethysmography signal derivatives has been introduced, in particular the recordings that had first derivative always more than zero, or always less than zero, for more than 170 samples have been deleted, fig. 8.5 b. In practice, recordings in which the trend was increasing monotonous or decreasing monotonous for at least 1.36 seconds were eliminated.

The remaining PPG recordings were then filtered through a band-pass 4th order Butterworth filter with a bandwidth between 0.5 and 8 Hz, fig. 8.6, and then both PPG and ABP were filtered with a Hampel filter according to [36]. Those frequencies were chosen because anything below 0.5 Hz is due to baseline wandering, while over 8 Hz is high-frequency noise. [36]

For computational reasons, patients with less than 3.10 hours of recordings were discarded, while it was taken only the first 3.10 hours from those with longer recordings. Out of 72 patients in MIMIC dataset, only 61 had at least both PPG and PPG. Following preprocessing pipeline the dataset was built extracting exactly 3.10 hours of recording from each of the 50 patients remained.

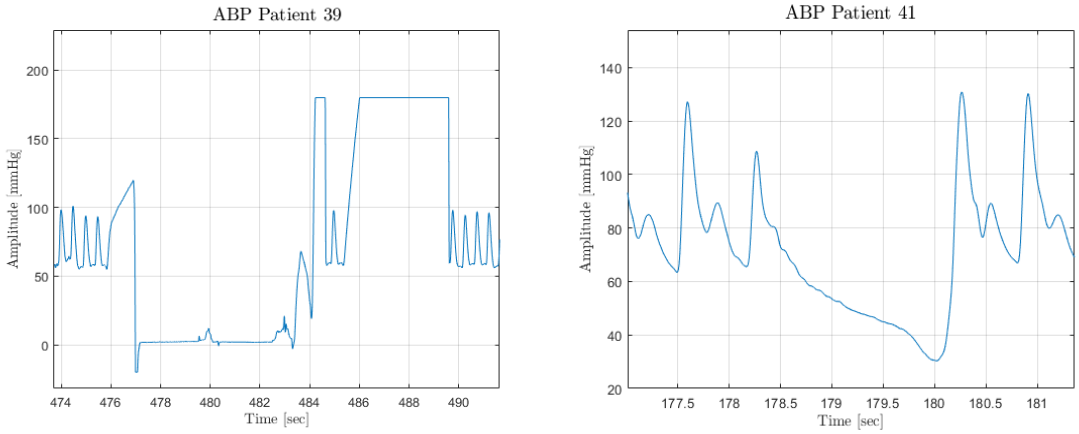
Also in this case, systolic and diastolic BP distribution were plotted, fig. 8.7, and it is possible to notice that they are heavily skewed toward physiological values.

Lastly, PPG signal was standardized and ABP was normalized, according to 5.2. Since the output is normalized, predictions made by the networks need to be de-normalized utilizing minimum and maximum values calculated on training set.



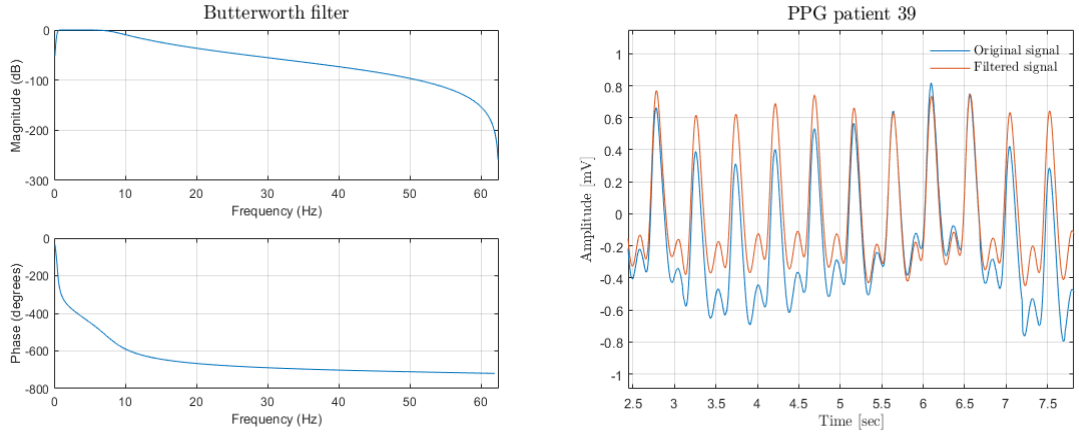
(a) (b)

Figure 8.4: Frequent PPG anomalies: flat lines (a) and flat peaks (b)



(a) (b)

Figure 8.5: ABP anomalies: negative BP followed by flat peaks (a) and no heartbeat for almost 2 seconds (b)



(a) (b)

Figure 8.6: Betterworth filter: Frequency response (a), comparison between original signal and filtered signal (b).

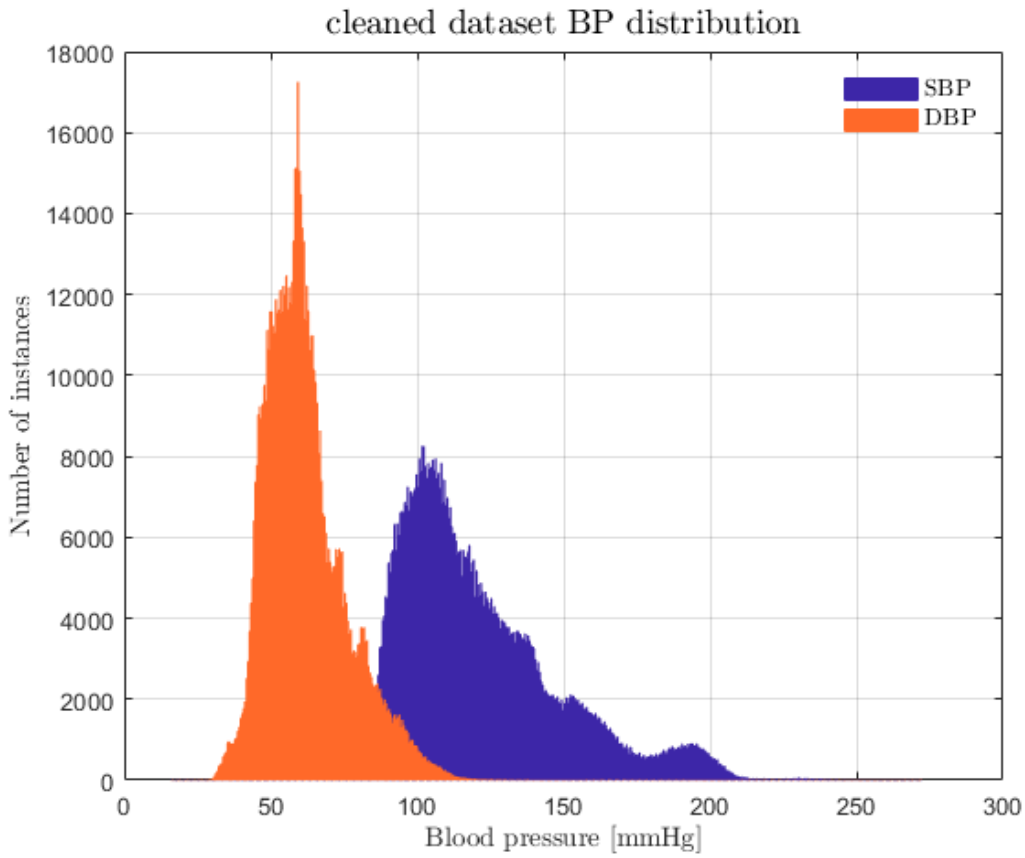


Figure 8.7: Dataset, with only PPG, SBP and DBP distributions

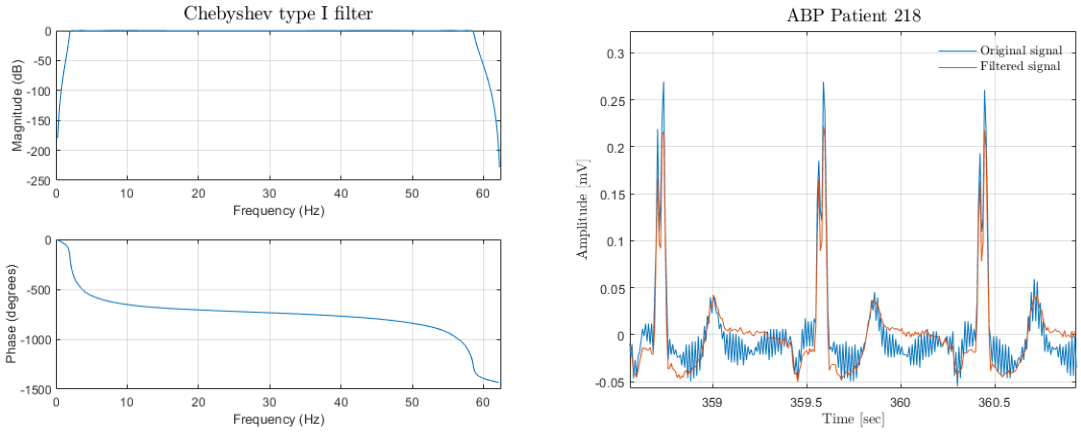
Dataset with PPG and ECG

ABP is strictly related to ECG, indeed several techniques used to predict ABP were developed starting from ECG and PPG (for instance, Pulse Transit Time). Thus a second dataset was created to study if using also ECG could help deep learning approaches.

In order to get the biggest dataset possible it was used ECG lead V, which represent the most frequent ECG lead recorded in the MIMIC database.

The preprocessing pipeline was the same used before extended also to ECG. On ECG signal it was used a 8° order passand Chebyshev type 1 filter, fig. 8.8, with cutoff frequency of 2 and 59 Hz in order to avoid motion artifacts and alternating current artifacts.

Out of 72 patients in MIMIC dataset only 51 had at least both PPG, ECG lead V and PPG. Following preprocessing pipeline the dataset was built extracting exactly 3.10 hours of recording from 40 patients. SBP and DBP distributions of the dataset with both PPG and ECG signals, also in this dataset, are skewed towards physiological values, fig. 8.9.



(a) (b)

Figure 8.8: Chebyshev filter: Frequency response (a), comparison between original signal and filtered signal (b).

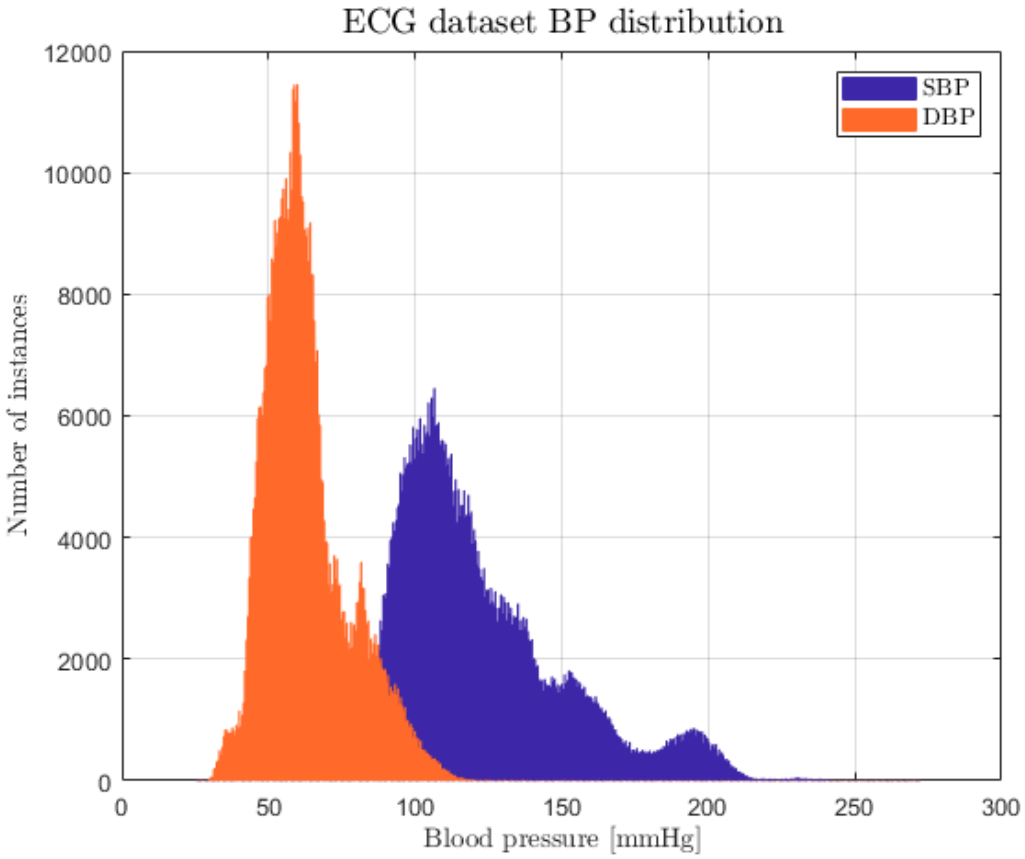


Figure 8.9: Dataset, with ECG and PPG, SBP and DBP distributions

9 | Tested neural architectures

In order to evaluate the best neural network architecture two different setup were implemented.

- *direct SBP/DBP prediction*: the network analyses 5 seconds of recording and then directly outputs a single value for SBP (peak) and another one for DBP (valley).
- *entire ABP signal prediction*: the network predicts the entire ABP in real time.

Predicting the entire signal would be better for clinical application, however, for commercial healthcare device implementation only systolic and diastolic values are predicted.

Every presented neural network for both setups was trained with both datasets and evaluated on a validation set, and then the best performing networks were cross-validated using Leave One Out (LOO) since it is the most robust approach in terms of generalization performance [36].

ANNs were trained utilizing Adam optimizer, learning rate $\eta = 0.001$, Huber loss and mini-batch training, also everything was implemented in Tensorflow 1.15 and training graphs are visualized through Tensorboard (official Tensorflow visualization tool). Since Adam is an adaptive learning rate algorithm, it didn't require a lot of tuning and therefore the default learning rate was used, while Huber loss was chosen because it is a robust metric, unaffected by outliers, considering that the dataset did not have bell-shaped distribution.

Lastly, samples of recorded PPG have different dimensions between the two setups, indeed, samples in *direct SBP/DBP prediction* are 5 seconds long, while in *entire BP prediction* are 2 seconds long. This difference is due to LSTM, indeed this cell have problems managing sequences too long, even though it performs better than classic RNN neuron. LSTMs are used also in the first setup, however in this case it was possible to downsample the input, through convolutional layers, since it wasn't necessary to output a value for every input.

In *direct SBP/DBP prediction* recordings were divided in 5 seconds chunks and then on the samples was applied the algorithm developed by [10] to extract SBP and DBP values. Since in 5 seconds usually there are between 4 and 6 caridiac cycles, it was taken the mean SBP and mean DBP as target value.

9.1 Direct SBP/DBP prediction

ResNet

The first attempt, fig. 9.1, was done using a ResNet-18 and testing different batch sizes.

Smaller batch allowed a faster training and achieved better results probably because they did not stuck in some local minimum, therefore also in regression task mini-batch training is the better way to train a neural network.

In particular three settings were tried, in the first one 650 samples per batch were used, maximum size permitted by google colab GPU, before performing backpropagation, in the second one 128 sample and in the third one 32 sample. In every setup the number of training steps is always the same, equation 9.1, this is important because it is the number of times the weights are updated, thus the networks are comparable only if their weights are updated the same number of times.

$$Training\ steps = epochs * \frac{Number\ of\ samples}{Batch\ size} \tag{9.1}$$

Classical feature selection is here automated by convolutional layers and through skip connections it is possible to stack layers creating a deep neural network, which can better analyse input data.

Once the best batch size was chosen it was trained a network also on dataset composed of PPG and ECG, fig. 9.1 c and d.

ResNet and LSTM

Subsequently, it was tried a ResNet like the previous one, followed by 3 LSTM layers each one composed by 128 neurons. Also, the first LSTM layer was bidirectional.

Convolutional layers achieve best performances when combined with recurrent layers, in fact they can extract features from a signal and they can also downsample the input sequence using the right kernel size, stride and padding. Indeed, the model can learn to preserve the useful information dropping only the unimportant details and shortening the sequences, the convolutional layer may help the following recurrent layers to detect longer patterns.

This network was the best performing network to directly predict SBP/DBP values for both datasets. Training results are showed in 9.2, since the two datasets have different number of samples the networks were trained for a different number of training steps.

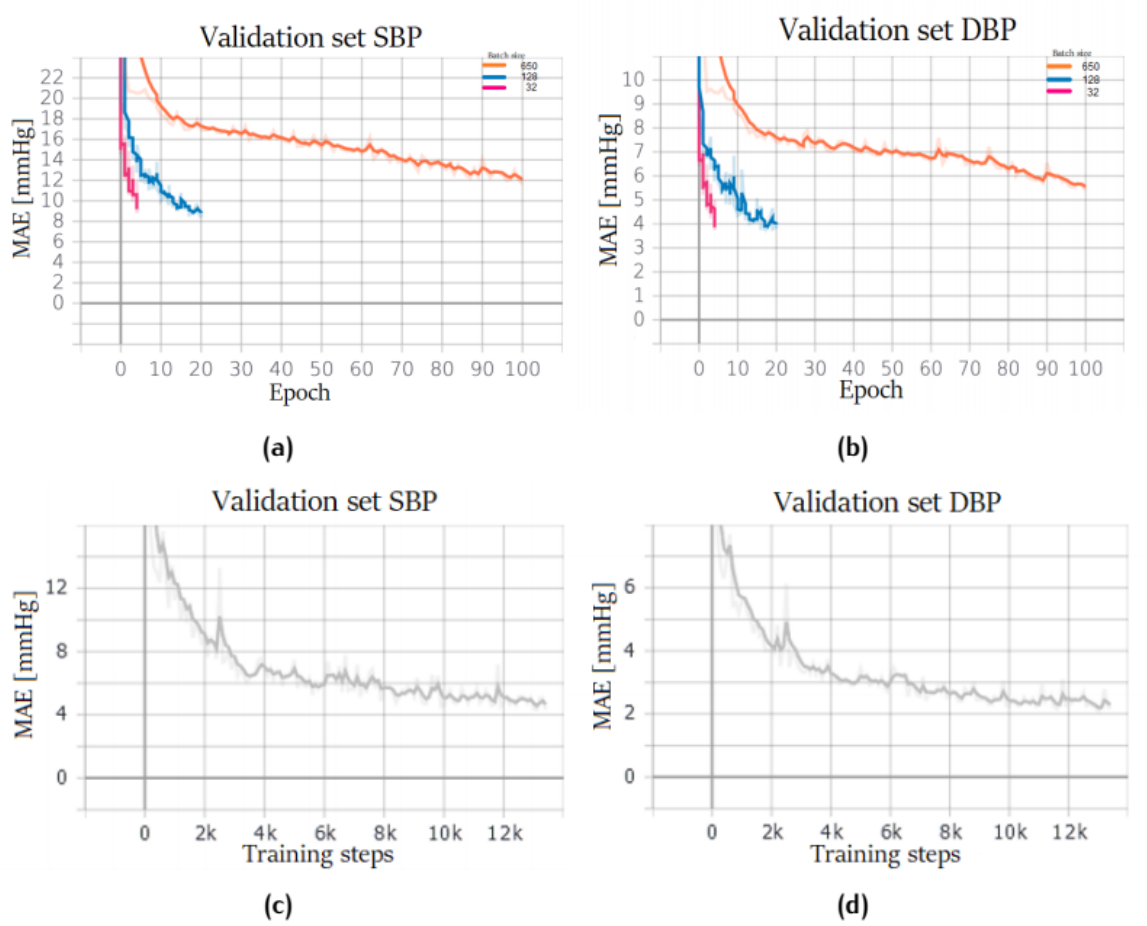


Figure 9.1: ResNet training: (a) and (b) the network is trained on PPG dataset with different batch sizes (orange 650, blue 128, magenta 32), (c) and (d) is trained on PPG + ECG dataset, batch size equal to 32

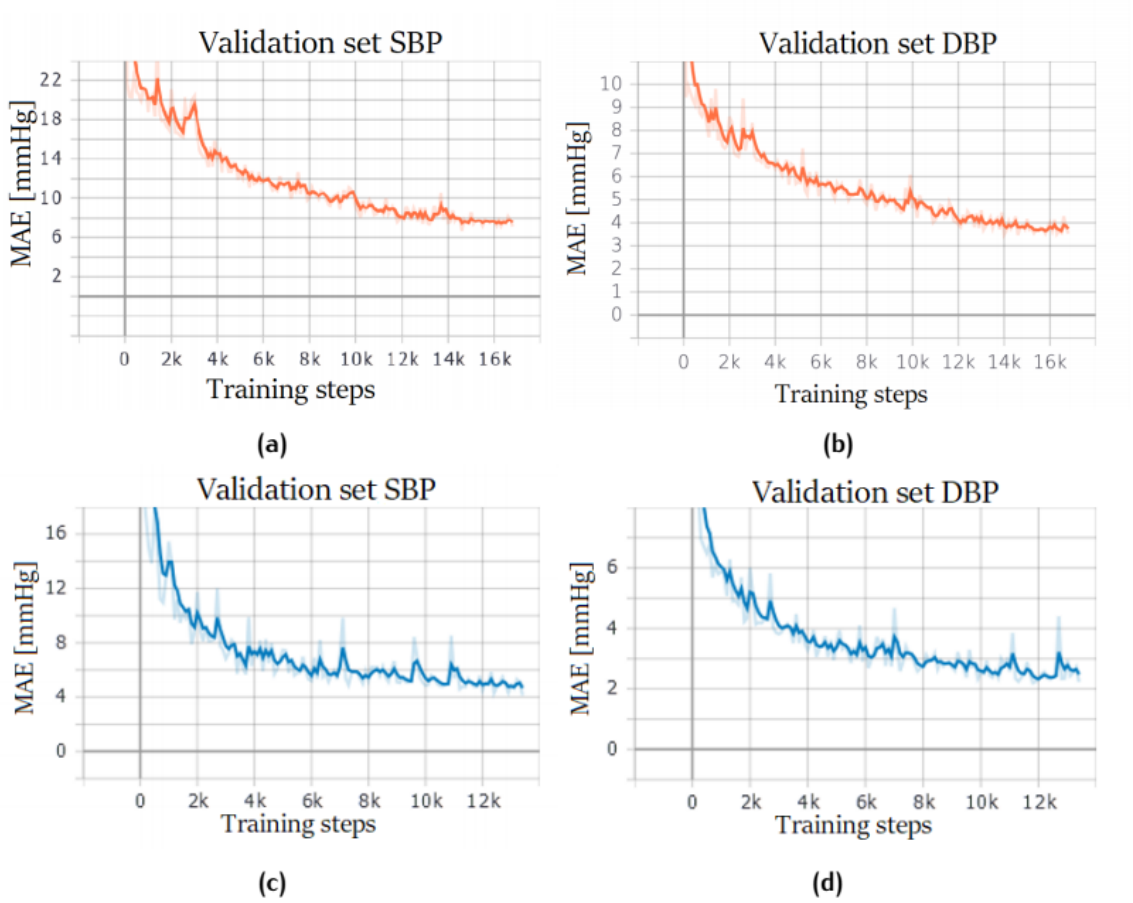


Figure 9.2: ResNet + LSTM: (a) and (b) the network is trained on PPG dataset, (c) and (d) on PPG + ECG dataset

9.2 Entire BP prediction

Fully connected network

Since there is a non-linear correlation between PPG and ABP it was tried to predict ABP signal through a simple fully connected neural network.

Different architecture were tried, however due to the simplicity of the model it wasn't possible to achieve good results. Deeper models appear to converge faster, but still maintain high errors, fig. fig. 9.3 a. PPG + ECG dataset were trained only on the deepest model, however results did not improve significantly, fig. 9.3 b.

LSTM - Long Short-Term Memory

The network is composed by three stacked LSTM layers each one with 128 cells. The first one is bidirectional, while the output layer is a fully connected neuron without any activation function. Training results are showed in fig. 9.4.

Bidirectional Long Short Term Memory (BLSTM) looks for contextual features both forward and backward, this is usefull because it is not known where is the feature the network want to forget. It is an approach used also by humans every day: sounds, words, and even whole sentences that at first mean nothing are found to make sense in the light of future context, in practice they are used to increase the amount of input information available to the network. It is an approach widely used in natural language processing; however, it was succesfully used also in BP prediction by several researchers like [34].

BLSTM usually are placed as first layer of the network because they have access to a much larger-scale context of input sequence. However, they heavily increase the computational cost, for this reason it is reasonable to use only one bidirectional layer.

Every sample is composed by 2 seconds of recording, this length was set because LSTMs have problems managing longer sequences. It is hard to remember lont term pattern if the sequence is too long, also, long sequences create deep unrolled network making really hard computing the gradient through time.

WaveNet

Another approach is based on a simplified version of the WaveNet composed of two blocks each one with 4 convolutional layer. Dilation rate is double in every convolutional layer inside a block: from 1 to 8. The output layer is a fully connected neuron without any activation function.

Since the network is composed only by convolutional layers it converges fast and thanks to the doubling dilation rate, the network can process extremely large sequences very efficiently.

Then a second network was built stacking 3 LSTM layers, each composed by 128 neurons and the first one bidirectional, on top of the simplified wavenet presented, convolutional layers extracts feature which are then analyzed by LSTM layers.

Training results are showed in fig. 9.5

ResNet and LSTM

Lastly, since using LSTMs layers on top of convolutional layers was proven a good approach, it was done an attempt using a deeper ANN: a modified ResNet followed by 3 LSTM layers, the first one bidirectional.

This network is different from the one presented in section 9.1 because it doesn't use maxpooling layers and because convolutional layers here have causal padding like WaveNet. This is a crucial step: in order to predict the entire signal it was necessary to output a sequence of the same length as the input sequence.

This network achieved the best performance in predicting the entire signal, fig. 9.6-9.8.

Every ResNet in this thesis is composed by 4 ResNet blocks. Convolutional layers have kernels equal to 3 and strides equal to 2, while the number of filters rises up in every block starting from 64 up to 512. In particular this ResNet is then followed by 3 LSTM layers, the first one bidirectional. Every layer is composed of 128 cells. Network graph is showed in fig. 9.7.

Although by default, Keras uses Glorot initialization with a uniform distribution to reduce the risk of exploding/vanishing gradients at the beginning of training, it doesn't guarantee that they won't come back during training. For this reason every convolutional operation is here followed by batch normalization, which zero-centers and normalizes each input, then scales and shifts the result using two new parameter vectors per layer: one for scaling, the other for shifting. In other words, the operation lets the model learn the optimal scale and mean of each of the layer's inputs. [15]

This network was the best performing network to predict the entire BP signal for both datasets.

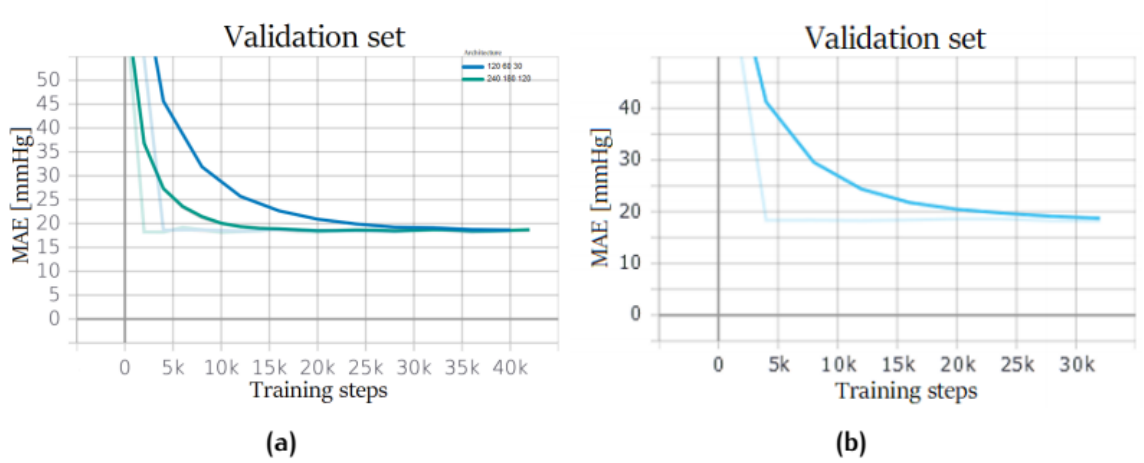


Figure 9.3: Fully Connected: Network trained on PPG dataset with different number of neurons (120-60-30, 240-180-120) (a), trained on PPG + ECG dataset (240-180-120) (b)

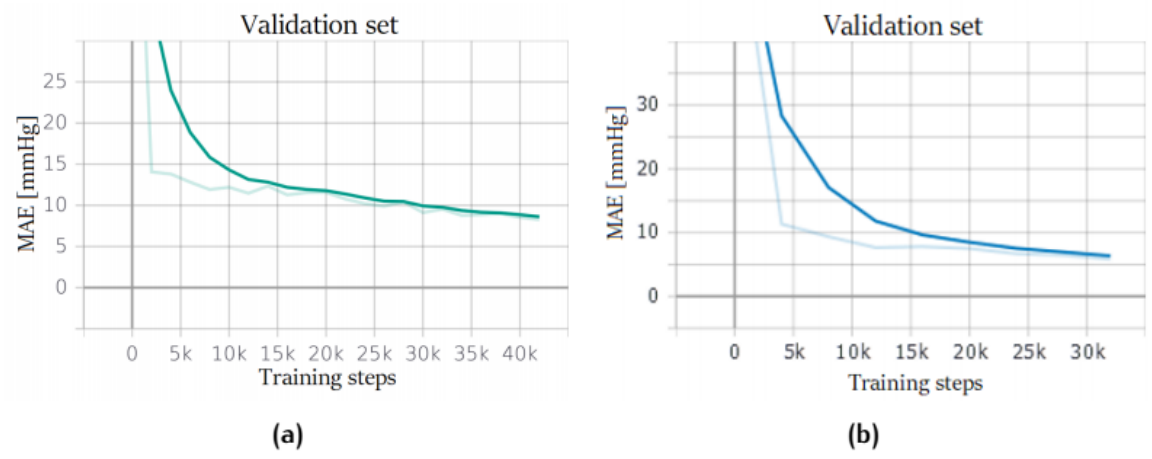


Figure 9.4: LSTM stack: Network trained on PPG dataset (a), PPG + ECG (b)

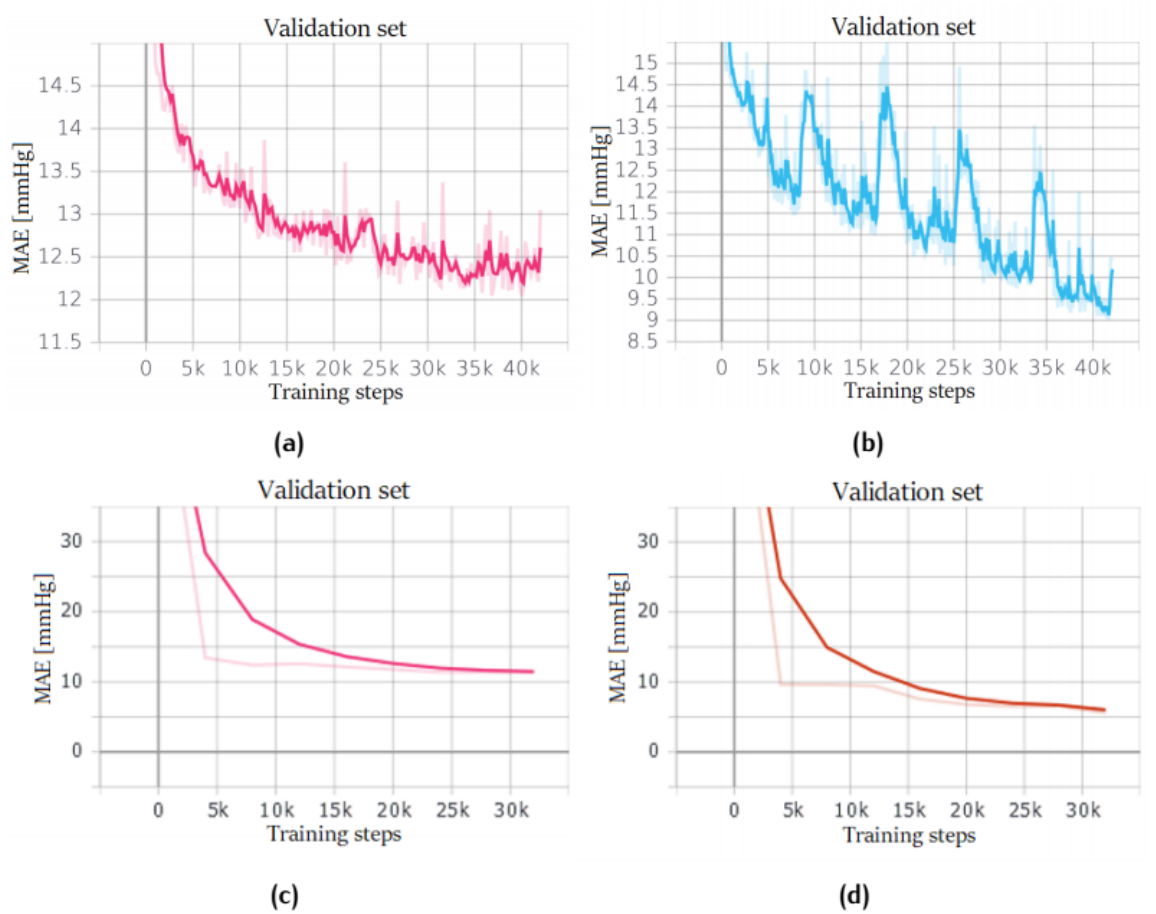


Figure 9.5: WaveNet (a) and WaveNet+LSTM (b) trained on PPG dataset, WaveNet (c) and WaveNet + LSTM (d) trained on PPG + ECG dataset

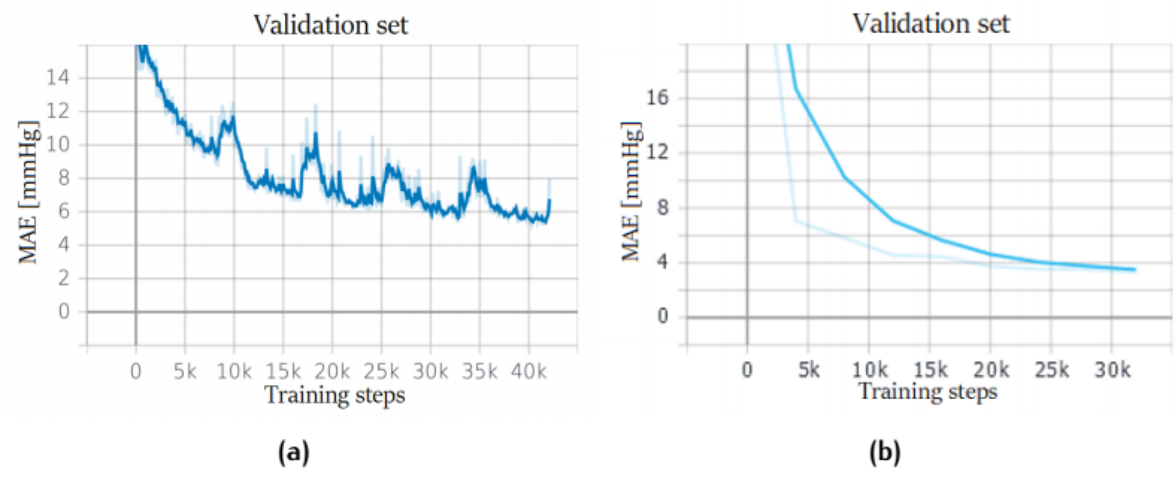
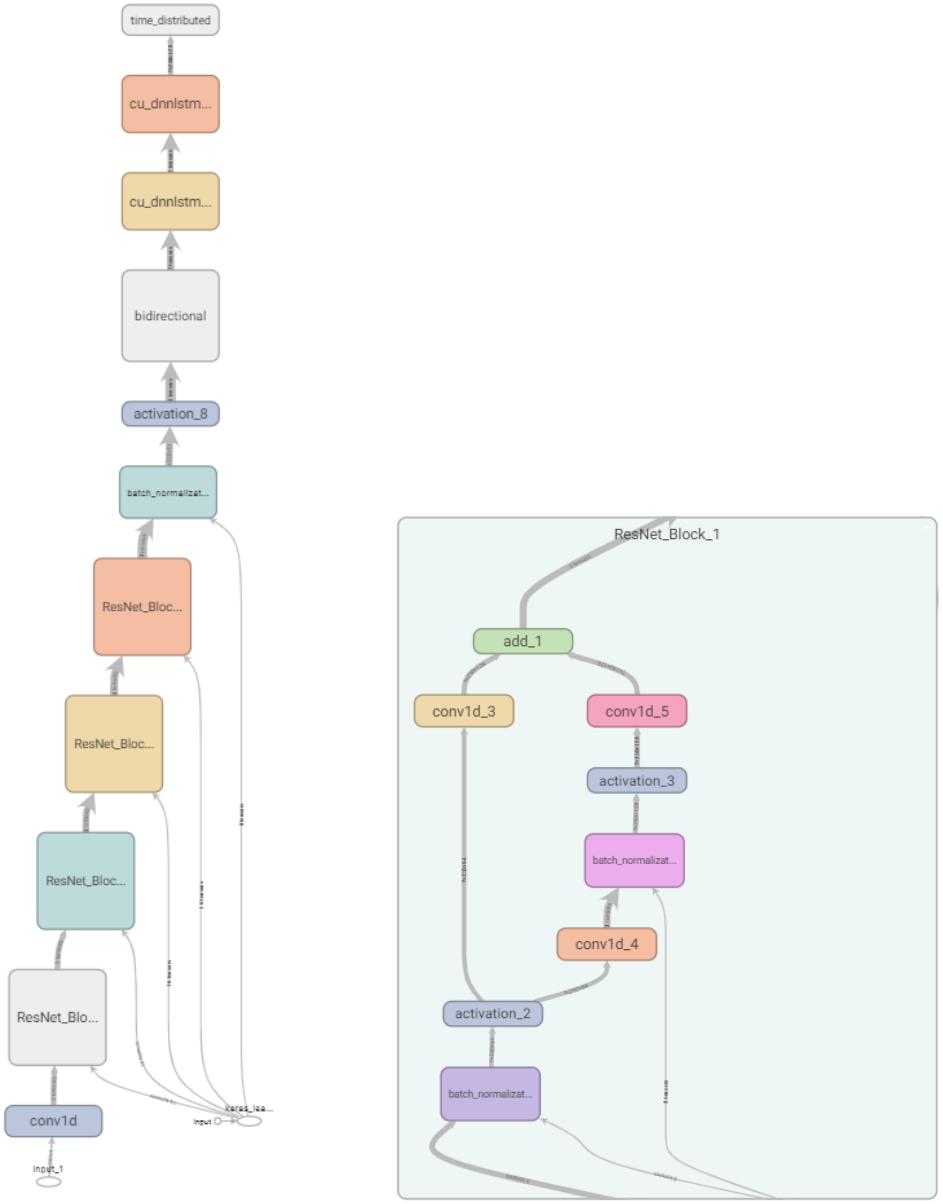


Figure 9.6: ResNet + LSTM: trained on PPG dataset (a), PPG + ECG dataset (b)



(a) (b)

Figure 9.7: ResNet graph visualized with Tensorboard, ResNet+LSTMl graph (a), resnet block (b). Every ResNet in this paper is composed of 4 blocks and then depending on the architecture (ResNet or ResNet+LSTM) it may be followed by LSTM layers

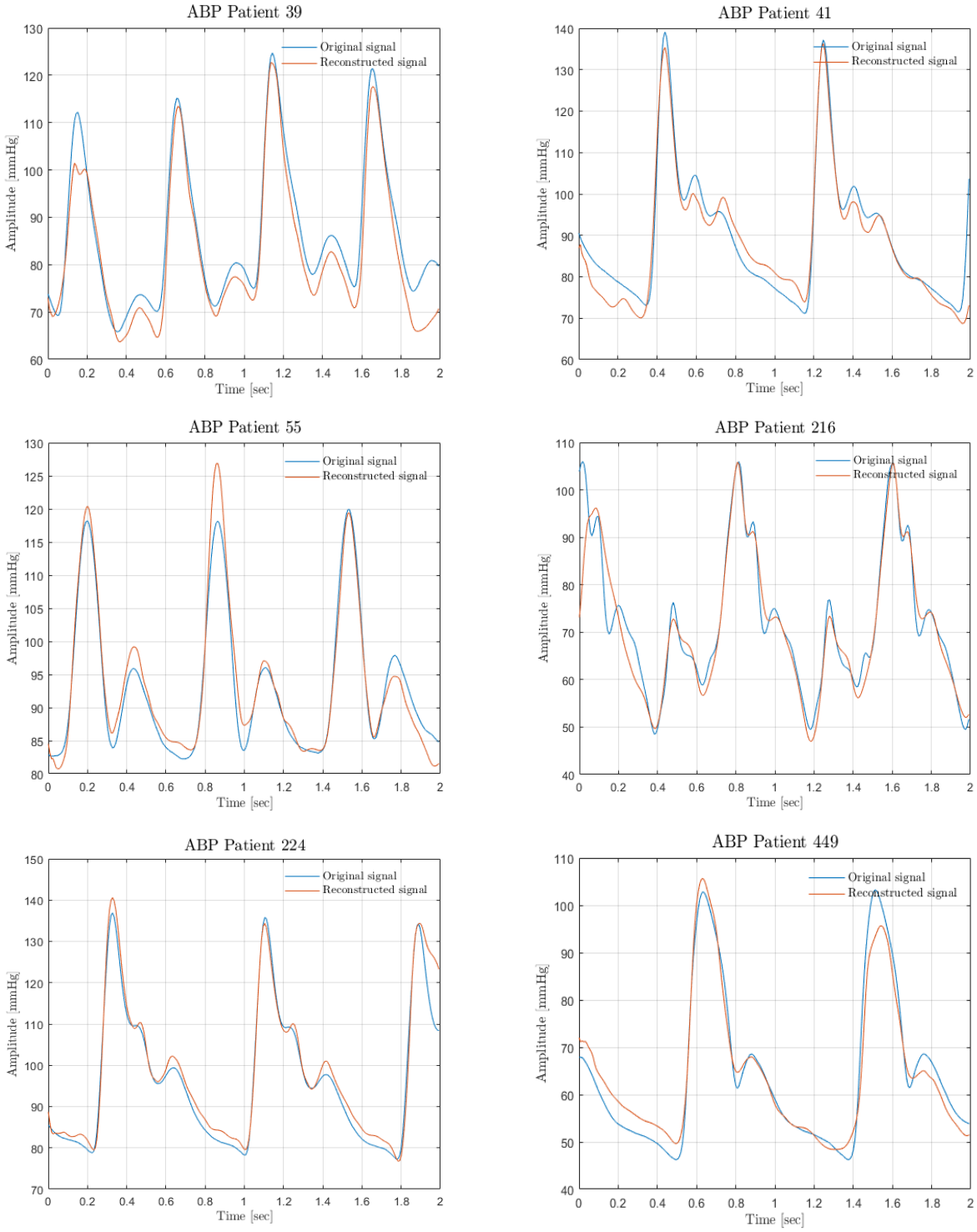


Figure 9.8: Predicted ABP signals on different patients using ResNet+LSTMs trained with PPG dataset

10 | Validation

10.1 Leave-One-Out

In order to understand how results will generalize, a Leave One Out (LOO) cross-validation was conducted on the better networks, Resnets followed by LSTM, for both datasets: the one built using only PPG and the one built using PPG and ECG. This method was chosen because it is exhaustive, which means it tests every possible way to divide the original sample into training and validation, and it has lower computational cost compared to its alternative. [54]

The overall errors were computed as the average of individual MAEs in each LOO iteration. The results were worse than when the network was trained and tested on the same patients, which means personalization for sure boosts the predictions.

There is a correlation between mean ABP and the errors, fig .10.1; indeed, the dataset have a majority of physiological ABP, for this reason when the network is trained with a great majority of healthy BP and then it is used to predict an unhealthy BP the error is greater than what it should be.

Also, there are some long patterns in some PPG, fig. 10.2, these patterns can not be recognized by the network because they are longer than 2 seconds, which is the length of training samples.

Lastly, using ECG improved network performance both on validation set and on LOO cross-validation and made the predictions less dependent from mean ABP, fig. 10.1.

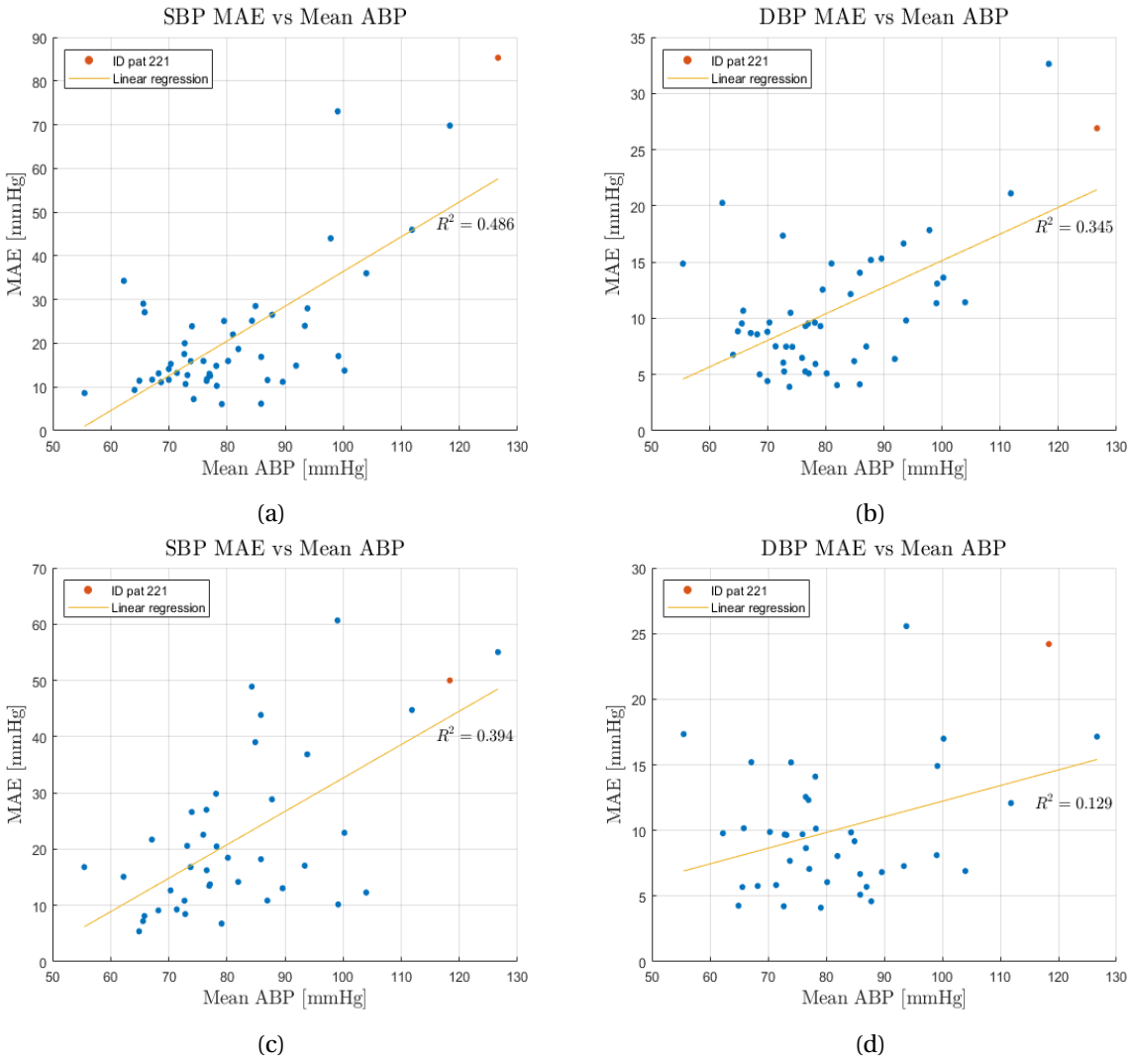


Figure 10.1: "Entire BP prediction" LOO error for different patients depending mean ABP, (a) and (b) refer to dataset with only PPG, (c) and (d) to dataset with PPG and ECG.

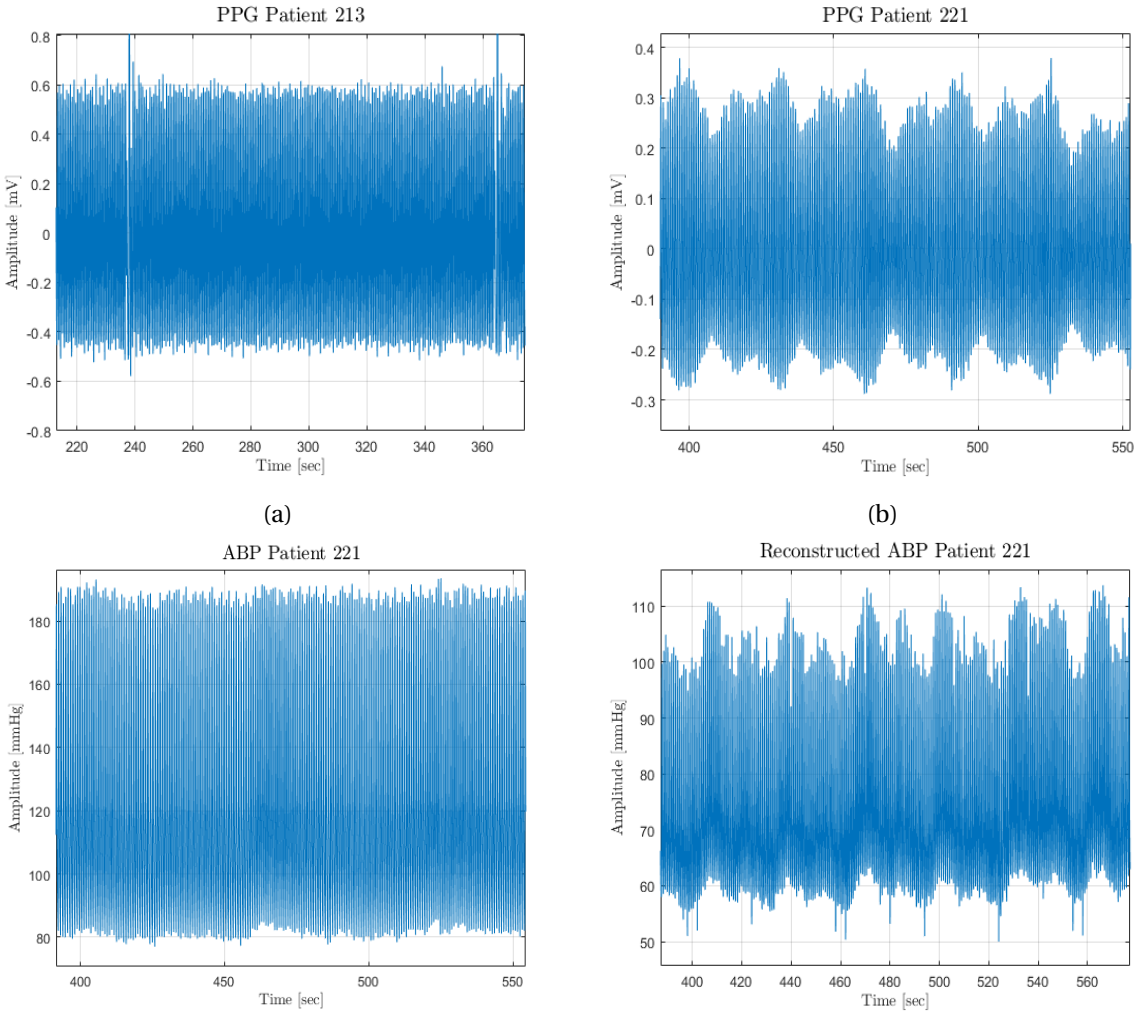


Figure 10.2: NN tends to predict an ABP with similar pattern to those in PPG, while real ABP does not have them

10.2 Experimental setup

Lastly, it was built a custom dataset at Neuronica Labs (Politecnico di Torino) to test our algorithm. Nine healthy students (5 males, 4 females, aged 22.84 ± 1.07 years) were recruited to participate in the experiment of PPG, ECG and BP signal acquisitions.

The recordings were gathered using a GE Healthcare B125 patient monitor, which is a clinical device generally appreciated for its intuitivity and reliability in a variety of acuties. The monitor delivers proven NIBP technology, utilizing GE-patented "smart cuff" pressure control to improve measurement time, patient comfort, and artifact rejection. It meets the requirements expected by AAMI ISO81060-2 and IEC 80601-2-30. More technical specifications at [7].

Time of day and ambient temperature were not controlled for, although most recordings were made in the morning. The students were seated and put at ease so that the commitments of everyday life did not affect the recordings, after which PPG, ECG and ABP were measured 3 times using the following recording protocol. First the PPG and ECG were recorded simultaneously, then ABP was measured using a sphygmomanometer. The PPG and ECG recordings were 15 seconds long, after which there were a BP measurement, however PPG was sampled at 300 Hz, while ECG at 100 Hz, thus the signals were resampled both at 125 Hz. A sphignanometer was used because it wasn't available a CNAP system, while invasive methods can only be performed by trained personnel.

As far as the ECG is concerned, lead I was recorded because the algorithm developed is designed to be then implemented in a wearable device and these typically only measure lead I. For this reason a new dataset was later created from MIMIC, this time using PPG, ECG lead I and ABP.

The previous dataset created using lead V served to demonstrate how the ECG could actually improve the performance of a neural network that has to predict ABP without having to deal with a small dataset. However, the networks trained with that one could not be reused on the Polito data because their weights were not trained for lead I, thus it was necessary to create a new dataset starting from MIMIC in order to train a network which will be then used on Polito dataset.

The new MIMIC dataset consisted of 12 patients on which it was applied the described preprocessing pipeline. Lastly, on this dataset was trained the best performing NN: "direct SBP/DBP prediction" ResNet+LSTM. The trained network was then applied on the polito dataset.

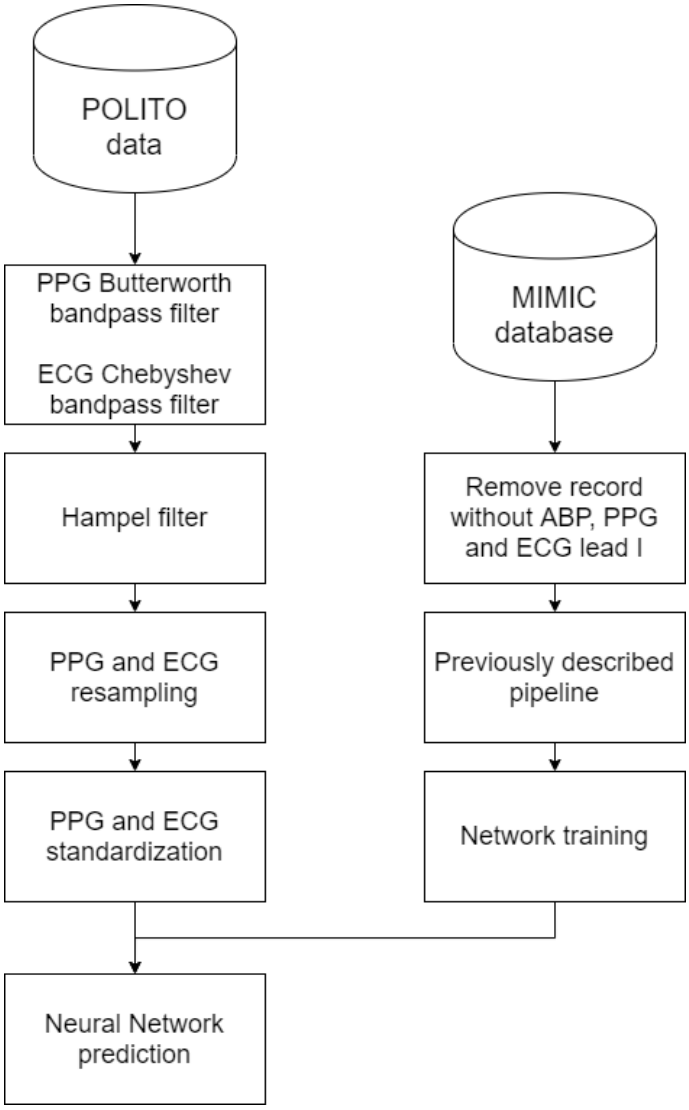


Figure 10.3: Pipeline used to process never seen data (Polito data)

11 Results

11.1 MIMIC database results

Table 11.1 summarizes performances on SBP and DBP prediction obtained with the two different setups and networks.

Using PPG and ECG in combination improved the performance in every setup. In particular, the best network was the ResNet+LSTM which directly predict SBP and DBP. The network overall MAE on the validation set were 4.118 and 2.228 mmHg. Errors were lower on DBP because it had lower variability relative to SBP.

As expected direct SBP/DBP prediction seems to be the best approach if the goal is just to output SBP and DBP values because the networks are geared for this purpose, while when the networks have to infer the entire signal they have to learn pieces of information that will not be used.

Another advantage of direct SBP/DBP approach is the possibility to analyse longer sequences, thus recognizing longer patterns. Finally, applying the algorithm [10] on a predicted ABP signal may introduce further errors. Nevertheless, entire BP prediction is an interesting approach for its clinical application and its results are fully showed in table 11.2.

Finally, it was performed LOO cross-validation on the best performing networks for both setups. From table 11.1 it is clear the best network is ResNet+LSTM in both cases. LOO was performed two times on PPG trained network because as explained in 8.2 there are two different datasets, in particular, the dataset created using only PPG had 50 patients, while the dataset created using PPG and ECG had only 40 patients. During training it was important to have access to as much data as possible, thus every data available was used, however, to compare performance it was useful to have the same dataset. ECG improved also generalization, as showed in table 11.3, however, errors were higher than when the networks were trained and tested on the same patients (different recordings). This phenomenon appears in several other types of research, i.e. [36][8], and it is generally called personalization.

With individual calibration, PPG and ECG can be used to directly estimate SBP and DBP on new data obtained from the same individual. According to the American National Standards Institute (ANSI) for the "Development of Medical Instrumentation", in order to validate a new device, there should be an average difference of 5± 8mmHg between the standard and the developed device [34]. The square root of the performance error the estimated SBP and the actual SBP values is the square root of the error probability (RMSE) 5.682, and the error performance estimated DBP and the actual DBP values has been RMSE 2.986.

Table 11.1: Errors (mmHg) on SBP and DBP prediction for different setups

Neural network (training dataset)	MAE		RMSE	
	SBP	DBP	SBP	DBP
Direct SBP/DBP prediction				
ResNet (PPG)	9,556	4,217	13,572	6,012
ResNet (PPG+ECG)	4,667	2,445	6,227	3,042
ResNet + LSTM (PPG)	7,122	3,534	11,214	5,029
ResNet + LSTM (PPG+ECG)	4,118	2,228	5,682	2,986
Entire BP prediction				
Fully Connected (PPG)	36,559	10,602	45,013	13,417
Fully Connected (PPG+ECG)	29,753	12,759	39,330	15,198
LSTM (PPG)	12,118	5,018	17,875	6,890
LSTM (PPG+ECG)	7,603	3,688	11,846	5,320
WaveNet (PPG)	18,539	8,154	26,638	11,441
WaveNet (PPG+ECG)	14,501	7,224	22,922	10,477
WaveNet + LSTM (PPG)	14,353	6,311	21,323	9,150
WaveNet + LSTM (PPG+ECG)	8,812	3,471	12,967	4,864
ResNet + LSTM (PPG)	8,660	3,843	13,439	5,718
ResNet + LSTM (PPG+ECG)	4,507	2,209	6,414	3,101

Table 11.2: Errors (mmHg) entire BP prediction

Neural network	PPG		PPG + ECG	
	MAE	RMSE	MAE	RMSE
Fully Connected	18,547	27,214	18,329	25,740
LSTM	8,591	13,306	5,897	9,321
WaveNet	12,292	18,297	11,338	17,518
WaveNet + LSTM	10,009	15,610	5,658	8,919
ResNet + LSTM	6,230	8,883	3,282	5,010

Table 11.3: LOO results, since direct SBP/DBP prediction didn't predict the entire signal the first 2 columns are empty. Errors are expressed in mmHg, S stands for SBP, D for DBP

NN (Dataset)	MAE	RMSE	MAE S	MAE D	RMSE S	RMSE D
Direct SBP/DBP prediction						
PPG (50 pat)			23,5976	10,7459	27,6430	12,3444
PPG (40 pat)			24,2227	11,1056	28,2470	12,6419
ECG (40 pat)			20,3667	9,5484	23,0699	10,8475
Entire BP prediction						
PPG (50 pat)	15,3419	19,1549	21,4666	10,6841	25,3825	12,3489
PPG (40 pat)	15,6788	19,5598	22,4095	10,8180	26,2460	12,4111
ECG (40 pat)	14,6093	18,0184	22,0995	10,1053	24,5865	11,5292

Table 11.4: Errors (mmHg) on SBP and DBP prediction using ResNet+LSTM trained on MIMIC dataset built using PPG and ECG lead I and tested on different set

Tested set	MAE SBP	MAE DBP	RMSE SBP	RMSE DBP
	PPG			
Validation set	7,409	3,706	9,875	4,833
Leave-One-Out	15,706	7,251	17,792	8,171
Polito dataset	9,916	5,905	11,879	7,273
	PPG + ECG			
Validation set	4,546	2,515	5,766	2,982
Leave-One-Out	16,128	6,743	17,875	7,902
Polito dataset	12,435	8,567	14,082	10,211

11.2 Polito database results

The best performing NN trained on PPG and ECG lead I was then used to predict SBP and DBP on Polito students achieving MAE equal to 12.435 mmHg on SBP and 8.567 mmHg on DBP, results in table 11.4. The network was trained, also, using only PPG achieving MAE equal to 9.916 mmHg on SBP and 5.905 on DBP. In this case, ECG improved the performance on data extracted from the MIMIC database but did not change generalization, furthermore, it negatively influenced the results on the Polito. The reason is probably due to the small training set, indeed only 12 patients had ECG lead I in MIMIC database.

Also, the unexpected results on the Polito dataset may be due to different pressure acquisition method. In this case, for technical reasons the pressure was not acquired with an invasive method, but measured with a sphygmomanometer, this can introduce an epistemic uncertainty. Furthermore, this instrument has an uncertainty of 5 mmHg which has therefore introduced additional noise to the measurements, the so-called aleatoric uncertainty.

11.3 Conclusion

PPG-based techniques allow continuous and automated ABP measurements, also they are well tolerated by patients, cheap and portable. These techniques are based on direct detection of the blood volume in the arteries under the cuff. ECG improves the performances in every setup and allows the network to generalize better. It is important to collect also this data in deep learning approaches.

This system represents a non-invasive, easy technique for blood pressure measurements, the measurements were carried out on a subset of patients from MIMIC database and on a dataset consisting of Polito students. Within-subject validation is compliant with ANSI normative. In particular, the best performing network achieved a MAE equal to 4.118 mmHg on SBP and 2.228 mmHg on DBP.

The selected network was tested, also, on a custom dataset, created at Neuronica Labs (Politecnico di Torino), with never seen patients achieving better performance than in MIMIC LOO cross-validation. This is, probably, due to the fact our dataset

was smaller and, therefore, had lower variance. Indeed, patients were all young and healthy students, also MIMIC dataset is a particularly difficult dataset because its patients show a huge variety of pathophysiologies that result in sudden blood pressure changes. Furthermore, ABP, PPG and ECG in MIMIC were probably collected with different measurement devices.

The proposed method can be embedded in wearable portable devices to perform continuous healthcare monitoring of arterial blood pressure to prevent the onset of irreversible damages, such us cardiovascular diseases and hypertension.

Bibliography

[1] *5 Regression Loss Functions All Machine Learners Should Know*. URL: https://heartbeat.fritz.ai/5-regression-loss-functions-all-machine-learners-should-know-4fb140e9d4b0.

[2] L. Goldberger et al. "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." In: *Circulation* 101.23 (2000). DOI: 10.1161/01.cir.101.23.e215.

[3] AA Alian and KH Shelley. "Photoplethysmography". In: *Best Practice & Research. Clinical Anaesthesiology*. 28 (2014), pp. 395–406. DOI: https://doi.org/10.1016/j.bpa.2014.08.006.

[4] Filippo Amato et al. "Artificial neural networks in medical diagnosis". In: *J Appl Biomed* 11 (Dec. 2013), pp. 47–58. DOI: 10.2478/v10136-012-0031-x.

[5] Mohammad Reza Bakhtiarizadeh et al. "Neural network and SVM classifiers accurately predict lipid binding proteins, irrespective of sequence homology". In: *Journal of Theoretical Biology* 356 (2014), pp. 213–222. ISSN: 0022-5193. DOI: https://doi.org/10.1016/j.jtbi.2014.04.040. URL: http://www.sciencedirect.com/science/article/pii/S0022519314002768.

[6] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1996. ISBN: 978-0198538646. URL: https://www.amazon.it/Neural-Networks-Pattern-Recognition-Christopher/dp/0198538642.

[7] *BP 125 patient monitor*. URL: https://www.deveaconseil.fr/wp-content/uploads/2019/07/Moniteur_-B125-.pdf.

[8] C. P. Chua and C. Heneghan. "Continuous Blood Pressure Monitoring using ECG and Finger Photoplethysmogram". In: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. 2006, pp. 5117–5120.

[9] Daniel Soudry Elad Hoffer Itay Hubara. "Train longer, generalize better: closing the generalization gap in large batch training of neural networks". In: (2017). URL: https://arxiv.org/abs/1705.08741.

[10] M. Elgendi et al. "Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions". In: *PLoS ONE* 8 (2013), pp. 113–115. DOI: DOI:10.1371/journal.pone.0076585.

[11] *FDA clears Biobeat's cuffless BP wearable*. URL: https://www.massdevice.com/fda-clears-biobeats-cuffless-bp-wearable/.

[12] M. Valentini G. Parati. "Prognostic relevance of blood pressure variability". In: *Hypertension* (2006).

[13] Thomas Pickering Gbenga Ogedegbe. “Principles and techniques of blood pressure measurement”. In: *Cardiology Clinics* (2010), pp. 571–586.

[14] Rong Ge et al. “Escaping From Saddle Points - Online Stochastic Gradient for Tensor Decomposition”. In: *CoRR* abs/1503.02101 (2015). arXiv: 1503.02101. URL: http://arxiv.org/abs/1503.02101.

[15] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent*. O’Really, 2019. ISBN: 978-1492032649. URL: https://www.amazon.it/Hands-Machine-Learning-Scikit-learn-Tensorflow/dp/1492032646.

[16] Darisz Przybylski et al. Gianluca Pollastri. “Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles”. In: *Proteins* 42 (2002), pp. 228–235. DOI: https://doi.org/10.1002/prot.10082. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.10082.

[17] *Global Health Observatory (GHO) data*. URL: https://www.who.int/gho/ncd/risk_factors/blood_pressure_text/en/.

[18] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385.

[19] X. He, R. A. Goubran, and X. P. Liu. “Evaluation of the correlation between blood pressure and pulse transit time”. In: *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 2013, pp. 17–20.

[20] *How to use Data Scaling Improve Deep Learning Model Stability and Performance*. URL: https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/.

[21] Nitish Shirish Keskar et al. “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima”. In: *CoRR* abs/1609.04836 (2016). arXiv: 1609.04836. URL: http://arxiv.org/abs/1609.04836.

[22] Y. Kurylyak, F. Lamonaca, and D. Grimaldi. “A Neural Network-based method for continuous blood pressure estimation from a PPG signal”. In: *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. 2013, pp. 280–283.

[23] et al. Manish Hosanee Gabriel Chan. “Cuffless Single-Site Photoplethysmography for Blood Pressure Monitoring”. In: *Journal of Clinical Medicine* (2020).

[24] Dominic Masters and Carlo Luschi. “Revisiting Small Batch Training for Deep Neural Networks”. In: *CoRR* abs/1804.07612 (2018). arXiv: 1804.07612. URL: http://arxiv.org/abs/1804.07612.

[25] Warren S. McCulloch and Walter Pitts. “A Logical Calculus of the Ideas Immanent in Nervous Activity”. In: *The Bulletin of Mathematical Biology* 5 (1943), pp. 113–115. DOI: https://doi.org/10.1007/BF02478259.

[26] *Medical Diagnosis with a Convolutional Neural Network*. URL: %5C%5Chttps://towardsdatascience.com/medical-diagnosis-with-a-convolutional-neural-network-ab0b6b455a20.

[27] Luca Mesin. *Introduction to biomedical signal processing*. 2017. ISBN: 9788892332485. URL: https://ilmiolibro.kataweb.it/libro/didattica-e-dispense/314585/introduction-to-biomedical-signal-processing/.

[28] *MIMIC Database*. URL: https://physionet.org/content/mimicdb/1.0.0/.

[29] G. B. Moody and R. G. Mark. "A database to support development and evaluation of intelligent intensive care monitoring". In: *Comput. Cardiol.* 0.0 (1996), pp. 657–660. DOI: 10.1109/cic.1996.542622. URL: https://physionet.org/physiobank/database/mimicdb/mimic-cic96/mimic.html.

[30] Aäron van den Oord et al. "WaveNet: A Generative Model for Raw Audio". In: *CoRR* abs/1609.03499 (2016). arXiv: 1609.03499. URL: http://arxiv.org/abs/1609.03499.

[31] Jonathan D.Hirst Pooja Jain Jonathan M.Garibaldi. "Supervised machine learning algorithms for protein structure classification". In: *Computational biology and chemistry* 33 (2009), pp. 216–223. DOI: https://doi.org/10.1016/j.compbiolchem.2009.04.004. URL: https://www.sciencedirect.com/science/article/abs/pii/S1476927109000279.

[32] Kaiming He et al. Priya Goyal Piotr Dollár. "Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour". In: (2017). URL: https://arxiv.org/abs/1706.02677.

[33] Martin G Reese. "Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome". In: *Computers & Chemistry* 26 (2001), pp. 51–56. DOI: https://doi.org/10.1016/S0097-8485(01)00099-7.

[34] Ü. Şentürk, I. Yücedağ, and K. Polat. "Repetitive neural network (RNN) based blood pressure estimation using PPG and ECG signals". In: *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. 2018, pp. 1–4.

[35] R. Shriram et al. "Continuous cuffless blood pressure monitoring based on PTT". In: *2010 International Conference on Bioinformatics and Biomedical Technology*. 2010, pp. 51–55.

[36] G. Slapničar, N. Mlakar, and M. Luštrek. "Blood Pressure Estimation from Photoplethysmogram Using a Spectro-Temporal Deep Neural". In: *Sensors (Basel)* (2019).

[37] Toshiyo Tamura and Yuka Maeda. "Seamless Healthcare Monitoring: Advancements in Wearable, Attachable, and Invisible Devices". In: ed. by Toshiyo Tamura and Wenxi Chen. Cham: Springer International Publishing, 2018. DOI: 10.1007/978-3-319-69362-0_6. URL: https://doi.org/10.1007/978-3-319-69362-0_6.

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

[39] *The WFDB Software Package*. URL: https://archive.physionet.org/physiotools/wfdb.shtml.

[40] Brzezinski WA. “Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition.” In: (1990). URL: https://www.ncbi.nlm.nih.gov/books/NBK268/.

[41] *WFDB python*. URL: https://github.com/MIT-LCP/wfdb-python.

[42] *Continuous noninvasive arterial pressure*. URL: https://en.wikipedia.org/wiki/Continuous_noninvasive_arterial_pressure.

[43] *Arterial line*. URL: https://en.wikipedia.org/wiki/Arterial_line.

[44] *White coat hypertension*. URL: https://en.wikipedia.org/wiki/White_coat_hypertension.

[45] *Nonlinear system identification*. URL: https://en.wikipedia.org/wiki/Nonlinear_system_identification#NARMAX_methods.

[46] *Blood pressure*. URL: https://en.wikipedia.org/wiki/Photoplethysmogram.

[47] *Blood pressure*. URL: https://en.wikipedia.org/wiki/Blood_pressure.

[48] *Hypertension*. URL: https://en.wikipedia.org/wiki/Hypertension.

[49] *Hypotension*. URL: https://en.wikipedia.org/wiki/Hypotension.

[50] *Blood pressure measurement*. URL: https://en.wikipedia.org/wiki/Blood_pressure_measurement.

[51] *Electrocardiography*. URL: https://en.wikipedia.org/wiki/Electrocardiography.

[52] *Artificial neuron*. URL: https://en.wikipedia.org/wiki/Artificial_neuron.

[53] *Convolutional neural network*. URL: https://en.wikipedia.org/wiki/Convolutional_neural_network#Convolutional.

[54] *Cross-validation (statistics)*. URL: https://en.wikipedia.org/wiki/Cross-validation_(statistics)#Leave-one-out_cross-validation.

[55] Jungil Choi Yinji Ma. “Relation between blood pressure and pulse wave velocity for human arteries”. In: *Applied physical science* (2018).