Application of Artificial Intelligence Strategies for Continuous Cuffless Arterial Hypertension Risk Assessment

Jesús Cano Serrano

TUTOR: José Joaquín Rieta Ibañez

Valencia, Junio 2025
Doctoral Thesis

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Abstract

Hypertension, a leading risk factor for cardiovascular diseases, poses a significant global health challenge. Early detection is critical, as it enables timely intervention and reduces the incidence of stroke, myocardial infarction, and other serious complications. Traditional blood pressure (BP) monitoring methods rely on intermittent cuff-based measurements, which, although accurate, are uncomfortable and unsuitable for continuous tracking. In contrast, photoplethysmography (PPG) is a non-invasive, low-cost optical technique that supports wearable-based monitoring and facilitates continuous data acquisition in everyday settings. This doctoral thesis investigates the application of artificial intelligence (AI) methods—specifically, machine learning (ML) and deep learning (DL)—to assess hypertension risk from PPG signals, with the long-term objective of enabling cuffless, continuous, and personalized BP monitoring.

The research is based on retrospective data from the publicly available MIMIC database and focuses on the binary classification of individuals into hypertensive and non-hypertensive groups. Two main approaches are explored: an ML pipeline using hand-crafted features, such as Pulse Arrival Time (PAT), extracted from PPG and ECG signals; and a DL approach relying exclusively on PPG signals, where continuous wavelet transform (CWT) is used to obtain time–frequency representations that are fed into pretrained convolutional neural networks such as GoogleNet and ResNet. The DL-based approach demonstrated key advantages in terms of automation, scalability, and independence from additional biosignals, making it more practical for real-world deployment.

Beyond performance evaluation, the thesis addresses critical aspects of clinical applicability. The effect of subject-specific model calibration using reference BP values was investigated, revealing that even a single reference measurement significantly enhances classification accuracy, particularly for unseen subjects. Further improvements were observed with repeated or temporally close calibrations, emphasizing the value of periodic calibration in long-term use. Moreover, downsampling the PPG signals—simulating the lower sampling rates typical of commercial wearables—did not degrade classification performance, reinforcing the feasibility of integration into existing technologies.

A distinctive contribution of this work lies in its focus on hypertension risk classification instead of direct BP estimation. This shift aligns with preventive and personalized medicine paradigms by enabling the identification of individuals in prehypertensive or early hypertensive stages. The binary classification framework also ensures that each individual is assigned a risk label, minimizing diagnostic uncertainty and improving clinical interpretability.

In summary, the results of this thesis demonstrate the potential of AI-driven systems for continuous hypertension risk assessment based solely on PPG signals. While most current wearable devices are not yet certified for clinical BP estimation, this work provides a foundation for future cuffless, non-invasive, and patient-friendly monitoring solutions with real-world applicability.

Resumen

La hipertensión, uno de los principales factores de riesgo para enfermedades cardiovasculares, representa un importante desafío para la salud pública a nivel global. Su detección precoz es fundamental, ya que permite una intervención oportuna y reduce significativamente la incidencia de ictus, infarto de miocardio y otras complicaciones graves. Los métodos tradicionales de monitorización de la presión arterial (PA) se basan en mediciones intermitentes mediante manguito, que si bien son precisas, resultan incómodas y no son adecuadas para un seguimiento continuo. En cambio, la fotopletismografía (PPG) es una técnica óptica no invasiva y de bajo coste que permite la monitorización mediante dispositivos portátiles, facilitando así la adquisición continua de datos en la vida diaria. Esta tesis doctoral investiga la aplicación de métodos de inteligencia artificial (IA), en particular aprendizaje automático (ML) y aprendizaje profundo (DL), para la estimación del riesgo de hipertensión a partir de señales PPG, con el objetivo a largo plazo de habilitar una monitorización continua, personalizada y sin manguito.

El estudio se basa en datos retrospectivos procedentes de la base de datos pública MIMIC, y se centra en la clasificación binaria de individuos en grupos hipertensos y no hipertensos. Se abordan dos enfoques principales: una estrategia basada en ML con características extraídas manualment, como el tiempo de llegada del pulso (PAT), a partir de señales de PPG y ECG; y un enfoque DL basado exclusivamente en señales PPG, en el que se aplica la transformada continua de wavelet (CWT) para obtener representaciones tiempo-frecuencia que se utilizan como entrada de redes neuronales convolucionales preentrenadas como GoogleNet y ResNet. Este segundo enfoque mostró ventajas destacables en términos de automatización, escalabilidad e independencia de otras señales fisiológicas, lo que lo hace especialmente prometedor para su aplicación práctica.

Además del análisis del rendimiento de los modelos, la tesis aborda aspectos clave para su aplicabilidad clínica. Se estudió el impacto de la calibración del modelo mediante valores de referencia de PA, observándose que incluso una única medida de referencia mejora significativamente la precisión de la clasificación, especialmente en sujetos no incluidos durante el entrenamiento. Estas mejoras se acentúan cuando la calibración se repite o se realiza en ventanas

temporales próximas, subrayando la importancia de la calibración periódica en contextos de uso prolongado. Por otro lado, la reducción de la frecuencia de muestreo de las señales PPG, simulando las limitaciones típicas de los dispositivos comerciales, no degradó el rendimiento del sistema, lo que refuerza su viabilidad de implementación.

Una aportación distintiva de este trabajo es su enfoque en la clasificación del riesgo de hipertensión, en lugar de la estimación directa de valores de PA. Este cambio se alinea con las tendencias actuales en medicina preventiva y personalizada, al permitir la identificación de individuos en estados prehipertensivos o con hipertensión incipiente. Además, el marco de clasificación binaria garantiza que cada individuo reciba una etiqueta de riesgo, lo que reduce la ambigüedad diagnóstica y mejora la utilidad clínica del sistema.

En resumen, los resultados de esta tesis demuestran el potencial de los sistemas basados en IA para la evaluación continua del riesgo de hipertensión a partir exclusivamente de señales PPG. Aunque muchos de los dispositivos llevables actuales no están aún certificados para la monitorización clínica de la PA, este trabajo contribuye de forma significativa al desarrollo de soluciones futuras que sean cómodas, escalables y adecuadas para una monitorización no invasiva y sin manguito.

Resum

La hipertensió, un dels principals factors de risc per a diverses malalties cardiovasculars, representa un repte important per a la salut pública a escala global. La seua detecció precoç és fonamental, ja que permet una intervenció oportuna i redueix significativament la incidència d'ictus, infarts de miocardi i altres complicacions greus. Els mètodes tradicionals de monitoratge de la pressió arterial (PA) es basen en mesuraments intermitents amb maneguet, que, tot i ser precisos, resulten incòmodes i no són adequats per a una vigilància contínua. En canvi, la fotopletismografia (PPG) és una tècnica òptica no invasiva i de baix cost que permet el monitoratge mitjançant dispositius portàtils, facilitant així la recollida contínua de dades en la vida quotidiana. Aquesta tesi doctoral investiga l'ús de mètodes d'intel·ligència artificial (IA), concretament d'aprenentatge automàtic (ML) i aprenentatge profund (DL), per a l'avaluació del risc d'hipertensió a partir de senyals PPG, amb l'objectiu a llarg termini d'aconseguir un sistema de monitoratge continu, personalitzat i sense maneguet.

L'estudi es basa en dades retrospectives obtingudes de la base de dades pública MIMIC, i se centra en la classificació binària dels subjectes en grups hipertensos i no hipertensos. Es van explorar dos enfocaments principals: un sistema basat en ML utilitzant característiques extretes manualment, com ara el temps d'arribada del pols (PAT), a partir de senyals de PPG i ECG; i un mètode basat en DL que utilitza únicament senyals PPG, en el qual s'aplica la transformada wavelet contínua (CWT) per generar representacions temps-freqüència, emprades per entrenar xarxes neuronals convolucionals preentrenades com GoogleNet i ResNet. Aquest segon enfocament presenta avantatges clars en termes d'automatització, escalabilitat i independència d'altres senyals fisiològiques, cosa que el fa especialment adequat per a aplicacions pràctiques reals.

A més d'analitzar el rendiment dels models, la tesi aborda aspectes fonamentals per a la seua aplicabilitat clínica. En particular, s'ha avaluat l'impacte de la calibració dels models mitjançant valors de referència de PA, observant-se que fins i tot una sola mesura de referència millora notablement la precisió en la classificació, especialment en subjectes no inclosos durant l'entrenament. Quan la calibració es repeteix o es realitza en una finestra temporal curta, el rendiment encara millora més, destacant la importància d'una calibració periòdica per a un

ús fiable a llarg termini. A més, la reducció de la freqüència de mostreig de les senyals PPG —simulant les condicions dels dispositius comercials portàtils— no afecta negativament el rendiment, cosa que reforça la seua viabilitat.

Una altra contribució destacada d'aquest treball és l'enfocament en la classificació del risc d'hipertensió en lloc de la predicció directa dels valors de PA. Aquest canvi s'alinea amb les tendències actuals en medicina preventiva i personalitzada, en permetre la identificació d'individus en fases prehipertensives o d'hipertensió primerenca. A més, l'ús d'un esquema de classificació binària assegura que cada subjecte reba una etiqueta de risc clara, la qual cosa redueix l'ambigüitat diagnòstica i augmenta la utilitat clínica del sistema.

En conclusió, els resultats d'aquesta tesi demostren el potencial dels sistemes potenciats per IA per al monitoratge continu del risc d'hipertensió a partir exclusivament de senyals PPG. Tot i que molts dels dispositius portàtils actuals encara no estan certificats per a la monitorització clínica de la PA, aquest treball suposa una contribució important cap al desenvolupament de solucions futures que siguen còmodes, escalables i adequades per a una vigilància no invasiva i sense maneguet.

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que me han apoyado durante el desarrollo de esta tesis.

En primer lugar a mi director de tesis José Joaquín Rieta por su gran ayuda y por iniciarme en la carrera investigadora.

También a mis compañeros de Dawako y amigos de Valencia y Manzanares.

Por último pero no menos importante, a Tamara, mis padres y mi hermano por su apoyo constante a lo largo de estos años.

Gracias a todos por haber hecho posibe este trabajo.

Contents

Chapter 1

Motivation, Hypothesis and Objectives

1.1 Clinical motivation

Hypertension is a primary risk factor for a wide range of cardiovascular diseases, including stroke, ischemic heart disease, heart failure, and peripheral arterial disease. Its early detection has been shown to prevent more deaths than any other major risk factor modification. Early identification, diagnosis, and treatment of hypertensive individuals are crucial before achieving effective blood pressure (BP) control [1]. Regular BP monitoring is essential for the entire population, with special emphasis on individuals previously diagnosed with hypertension, as they are particularly vulnerable to elevated BP. Shockingly, more than 60% of hypertensive patients continue to have uncontrolled hypertension. Out-of-office monitoring offers the advantage of obtaining multiple readings over time, enabling early detection of asymptomatic individuals without acute target organ damage and facilitating hypertension management [2].

Evidence suggests that early detection and screening for hypertension significantly reduce morbidity and mortality by facilitating timely treatment and promoting healthier lifestyle choices. For instance, a community-based program in

Canada reported a 9% reduction in hospitalizations due to cardiovascular conditions following the introduction of regular BP screenings in pharmacies, combined with educational initiatives targeting both patients and healthcare professionals [3]. However, the accuracy of such screening is critical, as false diagnosis can lead to unnecessary psychological effects and waste of resources for both patients and the healthcare system [4].

For intermittent non-invasive BP measurement, the use of an occluding upper arm cuff is the gold standard due to its high accuracy. BP values can be obtained manually, either by palpation or auscultation of Korotkoff sounds, or automatically using oscillometry methods based on pressure sensors. Continuous non-invasive BP monitoring can be achieved using techniques such as arterial applanation tonometry or the volume clamp method [5]. However, cuff-based devices are uncomfortable and unsuitable for continuous measurements during daily activities or exercise, as they require the user to keep their arm immobile and possess knowledge of measurement procedures [6].

Recent advances in digital technology have led to the development of robust wearable BP monitoring sensors capable of capturing biomedical signals. The ideal wearable device should be non-invasive, inconspicuous, compatible with continuous use over periods ranging from minutes to months, lightweight, energy-efficient, and adaptable to various activities and locations [7].

Given the challenges associated with BP measurement, there is a clear clinical need for user-friendly devices capable of continuously and non-invasively estimating BP. Such technologies would significantly improve both the early detection of potential conditions in individuals with elevated BP who do not yet exhibit hypertensive pathologies and the ongoing management of patients with diagnosed hypertension. Continuous BP monitoring during daily activities is essential in both cases, as it enables healthcare providers to assess the impact of treatments effectively.

1.2 Initial hypothesis

Photoplethysmography (PPG) is an optical measurement technique that has been used to create small, cost-effective sensors capable of monitoring changes in blood volume within the microvascular bed of tissues. PPG waveforms provide valuable cardiovascular information for clinical physiological monitoring, including blood oxygen saturation, blood pressure, heart rate variability, and pulse wave velocity [8]. PPG sensors typically consist of red and infrared light-emitting diodes (LEDs) and a photodetector. These sensors detect changes in reflected light intensity, which correspond to blood volume changes in the tissue, thereby providing cardiovascular information [9]. PPG is considered a promising and economical technique, demonstrating a high correlation with arterial BP in both the frequency and time domains [10].

In recent years, artificial intelligence (AI) has been applied to estimate or discriminate between BP levels. On the one hand, machine learning (ML) techniques combine electrocardiography (ECG) and PPG signals, utilizing propagation theory and parameters such as pulse transit time (PTT), pulse arrival time (PAT), and pulse wave velocity (PWV) to assess how BP affects the cardiovascular system [11]. However, the need to record both biosignals with two different devices in separate locations makes it inefficient and expensive. Recent studies combine these propagation parameters with PPG morphological feature extraction, using them as inputs for regression methods [12], support vector machines (SVM) [13], or neural networks for BP determination [14].

On the other hand, deep learning (DL), a promising ML subfield, has significantly advanced in terms of its utilization and efficiency. These advancements are attributable to the availability of extensive data, the development of neural network models, and the progress in computing hardware, including central processing units (CPUs) and graphics processing units (GPUs) [15]. Currently, DL techniques are considered state-of-the-art for image classification, exemplified by the fact that all the winning entries in ImageNet Large Scale Visual Recognition Competition utilized deep convolutional neural networks (CNNs) [16].

As a result, recent studies have employed DL techniques for BP classification based on PPG signals, utilizing image representations generated through techniques such as continuous wavelet transform (CWT) [17] and short-time Fourier transform (STFT) [18, 19]. Notably, one of the primary advantages over traditional ML methods is that DL eliminates the need for manual feature extraction from PPG signals, as the most relevant features are automatically extracted from the input images, rendering ECG signals unnecessary. This streamlines the practical implementation of the model in wearable devices.

The incorporation of these algorithms into wearable devices has the potential to provide reliable insights into the BP status of monitored individuals. This contribution can play a pivotal role in the prevention, early diagnosis, and ongoing management of hypertension and related cardiovascular conditions. The growing prevalence of wearable devices, smartphones, and biosensors has ushered in a medical revolution, enabling the integration of AI tools to address complex medical challenges. These technologies are rapidly becoming the cornerstone of vital sign monitoring and are pivotal in achieving optimal diagnosis and treatment follow-up, while empowering patients to take a more active role in their healthcare journey [9].

Therefore, the hypothesis of the present thesis is that the combination of biomedical signals, principally PPG signals, with advanced AI tools like DL and ML, could provide effective solutions for rapid and easily accessible hypertension detection and monitoring when implemented in wearable devices.

Furthermore, these types of systems, designed to obtain BP state, do not consistently account for realistic sources of noise. This limitation may impact the system's ability to generalize and adapt to real-world conditions, where the pres-

ence of uncontrolled noise can significantly influence the obtained results. For example, PPG measurement can be affected by source of noise or variance from individual human population characteristics (skin tone, obesity, age, and gender), physiological factors (respiration, venous pulsation, body site of measurement, and body temperature), and external influences affecting the device (motion artifact, ambient light, and applied pressure to the skin) [20].

Hence, this work also studies the relevance of calibrating new subjects who have not been used to train the classification CNN. Thus, reference BP values and the simultaneous behavior of the PPG signal are recorded. This approach ensures model adaptation to the individual subject's characteristics for achieving more precise classification. Furthermore, it is proposed that the calibration process be conducted periodically, optimizing its long-term performance for variations in the subject's health or environment that may impact BP levels.

Finally, this study also seeks to identify hypertension pathology independent of absolute BP values, as BP is known to fluctuate throughout the day, influenced by circadian rhythms and various physical and mental stimuli. By focusing on these fluctuations rather than static BP measurements, this approach seeks to improve the detection of episodic hypertensive events that traditional methods might miss, potentially contributing to a more accurate assessment of long-term cardiovascular risk.

1.3 Objectives

The main objective of this research is to develop AI methodologies with the ability to discriminate between normotensive (NTS) and hypertensive (HTS) subjects. Furthermore, the need and relevance of per-subject calibration and the hypertension pathology detection, not absolute values, is evaluated. For this purpose, this thesis focuses on the development of both ML and DL artificial intelligence methods. More concretely, this research has the following specific objectives:

- Analysis and processing of simultaneously obtained PPG and ECG recordings for the extraction of propagation features, such as PTT and PAT, combined with other morphological features from PPG waveform and its first and second order derivatives.
- Select only those features with relevant information for solving the classification problem optimally applying ReliefF algorithm and analyzing the correlation matrix from the extracted features.
- Testing machine learning classification strategies such as SVM, k-nearest neighbors (KNN) and bagging ensemble for hypertension risk assessment.
- Employ PPG recordings transformed into scalograms through CWT as input images for ResNet and GoogleNet pretrained CNNs, eliminating the

need for simultaneous ECG recordings or manual feature extraction, allowing the DL models to automatically extract deep features from PPG signals.

- Evaluation of the necessity and effectiveness of previous per-subject calibration to improve classification results with different time intervals between PPG segments.
- Identify HTS patients from non-HTS recordings and NTS subjects from non-NTS recordings when they are affected by BP fluctuations.
- Evaluation of classification performance with statistical tests for accuracy (*Acc*), sensitivity (*Se*), specificity (*Sp*), and *F1-Score*.

1.4 Thesis structure

This document is structured in the following chapters:

- **Chapter 1. Motivation, Hypothesis and Objectives.**
  In this chapter, the clinical motivations for performing this work have been explained. Moreover, the initial hypothesis and the main proposed objectives have been commented.
- **Chapter 2. Blood Pressure, Hypertension and Photoplethysmography.**
  This chapter reviews some significant concepts to obtain the proper theoretical framework to understand the work aim, starting with general concepts about BP and the importance of detecting hypertension to continue explaining the main ways to measure it. Moreover, the main signals used in this work, PPG and ECG, are defined.
- **Chapter 3. Artificial Intelligence Background.**
  In this chapter, general principles of AI and the main algorithms for ML and DL classification are explained, with a focus on their application in BP monitoring.
- **Chapter 4. Materials and Methods.**
  This chapter describes the methodology applied in this study. Firstly, the analyzed databases are introduced. Next, PPG, ECG and ABP signals processing is explained. Then, the three main approaches are described: hypertension risk assessment, relevance of calibration, and detection of hypertensive pathology.
- **Chapter 5. Results.**
  In this chapter, the results of the studies introduced in Chapter 4 are systematically presented.
- **Chapter 6. Discussion.**
  This chapter discusses the results presented in Chapter 5 compared with

previous works, as well as the limitations of the study and their possible interpretations and clinical implications.

- **Chapter 7. Conclusions, Future Lines and Contributions.**
  The conclusions of this work are presented in this chapter. Moreover, some possible future research lines are suggested. Finally, the contributions in scientific journals and conferences are enumerated.

Chapter 2

Blood Pressure, Hypertension and Photoplethysmography

This chapter provides a general view of BP physiology, its fluctuations, and the clinical relevance of hypertension in Section 2.1 and Section 2.2. It then explores different BP measurement techniques, emphasizing the advantages and challenges of non-invasive methods in Section 2.3. Additionally, the chapter describes the ECG signal and its relationship with BP assessment in Section 2.4. Finally, a detailed analysis of photoplethysmography is presented, covering signal acquisition, waveform characteristics, and the underlying principles of wave propagation in Section 2.5.

2.1 Introduction to blood pressure

2.1.1 Blood pressure physiology

Arterial blood pressure (ABP) is defined as the pressure measured within the principal arterial network of the human body. It is a critical physiological parameter influenced by the interaction of two primary factors: cardiac output (CO) and total peripheral vascular resistance (PVR). CO (= stroke volume x heart rate), in turn, depends on myocardial contractility and the vascular properties. On the other hand, PVR is determined by the tone of the arteries and arterioles and the structural properties of the arterial wall [21].

The pulsatile conditions of blood flow in the arterial system are a result of the cyclic nature of the heart pump, which injects blood into the arterial tree during the ventricular systole phase. During diastole, the heart is in a rest state, and there is no active ejection of blood. Pulsatile flow creates dynamic pressure fluctuations that vary in different parts of the arterial system. Notably, the aorta and large arteries possess distensible characteristics, allowing them to temporarily store a portion of the blood received during systole within their expanded regions. This stored blood is subsequently released back into circulation during diastole. This phenomenon leads to the continued flow of blood within the arteries during diastole, despite the absence of active ejection by the heart during this phase [22].

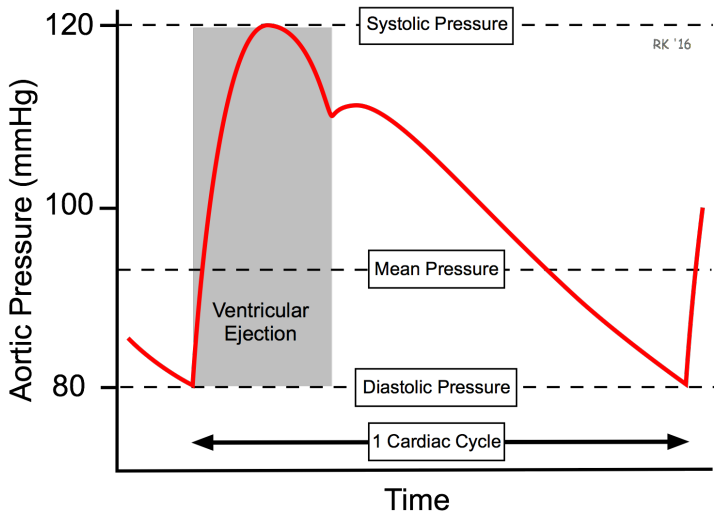


Figure 2.1: Arterial blood pressure pulse of a cardiac cycle. When the left ventricle ejects blood into the aorta, the aortic pressure rises. Systolic pressure reflects the maximal aortic pressure following ventricular ejection. As it is relaxing, the pressure in the aorta falls. The lowest pressure in the aorta is termed the diastolic pressure and occurs just before the next ventricular ejection [23].

Both the arteries and the heart undergo structural adaptations in response to increased load, such as that caused by hypertension. Elevated ABP imposes greater stress on the walls of thin-walled chambers or tubes, leading to an increase in wall tension, as described by Laplace's law [21].

As seen in Figure 2.1, the aortic blood pressure signal waveform is conventionally separated into systolic (SBP) and diastolic (DBP) blood pressure. SBP is the maximum value of pressure during systole and depends fundamentally on cardiac output and the distensibility of the aorta and great arteries, the latter being expressed by the retrograde pulse wave. DBP is the minimum value during diastole and depends fundamentally on peripheral resistance [23].

2.1.2 Blood pressure fluctuation

SBP and DBP are dynamic, changing from one pulse to another and varying throughout the day due to the circadian rhythm. BP variability is more pronounced in hypertensive patients compared to normotensive individuals, with the variability being proportional to the increase in mean BP [24]. External factors such as exercise, stress, medications, and nutrition significantly influence BP levels [25].

Two important concepts associated with office BP variability are white-coat hypertension and masked hypertension. White-coat hypertension is characterized by elevated office BP in normotensive individuals, often due to anxiety or the clinical setting, while masked hypertension involves normal office BP in hypertensive patients, obscuring the true hypertensive condition [24].

These phenomena underscore the complexity of accurately diagnosing and managing hypertension, as single-point measurements may not reflect an individual's true BP profile. Nowadays, most BP monitoring devices provide only punctual information, and artificial intelligence methods estimate BP values based on these isolated measurements [26].

The risk of misdiagnosing hypertension could potentially imply early diagnosis in hypertensive subjects or mislabeling healthy individuals as hypertensive. Such misdiagnosis can lead to unnecessary anxiety, additional medical tests, and inappropriate treatments. Therefore, continuous and unobtrusive BP monitoring devices are essential for accurate hypertension screening. These devices enable the capture of BP fluctuations over extended periods, providing a more comprehensive and reliable assessment of an individual's BP status. The present study has investigated whether adequate subject diagnosis can be provided despite altered BP conditions, emphasizing the importance of advanced monitoring techniques for accurate BP assessment.

2.2 Hypertension

According to European and international guidelines [27], hypertension is characterized by persistent office SBP values equal to or exceeding 140 mmHg and/or DBP 90 mmHg. However, there is a continuous relationship between BP and cardiovascular (CV) or renal adverse events, which begins at an office SBP exceeding 115 mmHg and a DBP surpassing 75 mmHg [28]. Table 2.1 will be used as a reference, where the BP levels are reflected according to the report of The Sixth Report of the Joint National Committee on the Prevention, Detection, Evaluation, and Treatment of High Blood Pressure [29].

Table 2.1: Blood pressure classification in adults.

BP classification	SBP (mmHg)	DBP (mmHg)
Normotensive (NTS)	$<$120	and $<$80
Prehypertensive (PHT)	120-139	or 80-89
Stage 1 hypertensive (HTS)	140-159	or 90-99
Stage 2 hypertensive (HTS)	$\geq$160	or $\geq$100

Only a minority within the hypertensive population experiences an isolated increase in BP, with the majority displaying additional CV risk factors. Moreover, when these risk factors coexist, BP and other CV elements may amplify each other, resulting in a cumulative CV risk that surpasses the combined impact of its individual components [30]. The increase of BP produces structural changes in the arterial system that increase CV risk and affect organs such as the brain, heart and kidneys, determining the main complications of arterial hypertension as risk of strokes, heart failure and progression of chronic kidney disease [31].

Concerning its symptoms, hypertension is typically characterized as a silent disease. Initial symptoms often do not manifest for many years, despite underlying clinical implications. It is not uncommon for the first noticeable symptoms to be severe cardiac issues such as heart attacks or strokes. While some patients may experience mild precursor symptoms, such as tinnitus or occasional headaches, it is generally observed that when a patient presents a more expressive symptomatology, they are already in an advanced stage of hypertension, leading to significant cardiovascular complications [29].

Hypertension is the most prevalent CV disorder in the world and according to the World Health Organization, it affects 1.28 billion adults aged 30 to 79 years worldwide, two-thirds living in low-income and middle-income countries. Furthermore, the global age-standardized average prevalence of hypertension in adults aged 30 to 79 years was reported to be 34% in men and 32% in women [32]. These percentages match those of Spain, representing ≈ 10 million people. Among them, only 68.5% were diagnosed as HTS (≈ 6.78 million), 57.4% received drug therapy (≈ 5.68 million), and 32.7% (≈ 3.24 million) reached controlled values after therapy [33].

The asymptomatic nature of hypertension, observed in 31.5% of HTS subjects, highlights the importance of early prevention and regular monitoring to enable timely diagnosis and appropriate management. This is particularly critical for individuals already diagnosed with hypertension, those over 40 years of age with a family history of the condition, and patients with additional risk factors such as obesity [34]. In Spain, more than two-thirds of hypertensive patients remain uncontrolled, highlighting the need for improved detection and intervention strategies. Continuous BP monitoring emerges as a promising approach to reduce the risk of cardiovascular events associated with undiagnosed or poorly managed hypertension.

2.3 Blood pressure measurement

Blood pressure monitoring plays a pivotal role in both the treatment and early detection of hypertension, a condition that can lead to serious health consequences if not properly managed. In clinical practice, two primary methods are employed for measuring blood pressure: invasive and non-invasive techniques.

2.3.1 Invasive methods

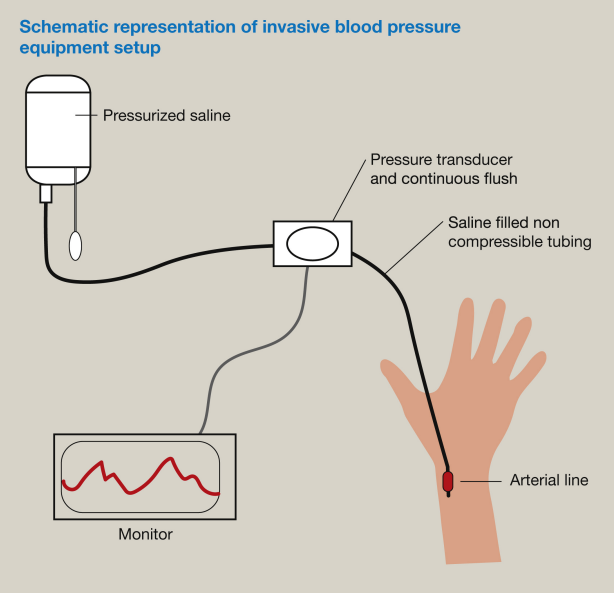


Figure 2.2: Representation of invasive blood pressure monitoring that comprises the use of an indwelling catheter in an artery and an arterial line system. The kinetic energy generated by arterial pressure induces pulsations in the column of saline, which is then transmitted to the transducer [35].

Invasive arterial BP monitoring offers high accuracy and serves as the gold standard in intensive care medicine and high-risk surgeries. Generally, it comprises an indwelling cannula inserted in the artery (radial, femoral or brachial) and an arterial line system, as represented in Figure 2.2. Then, a transducer is used to convert mechanical energy into an electrical signal whose waveform provides BP information. However, they come with several drawbacks, including the time-consuming nature of the procedure, the need for highly trained personnel to perform it, and the potential for complications such as infections, embolism, vessel or nerve damage, and ischemia [5].

2.3.2 Non-invasive methods

As a result of invasive methods limitations, BP measurement is predominantly carried out using non-invasive techniques, whether intermittently or continuously. Although in critically ill patients arterial catheter measurement is still recommended, the choice between intermittent and continuous BP measurement methods depends on the specific needs of patients. Stable, low-risk individuals may find intermittent measurement sufficient, whereas clinical patients are often recommended for continuous non-invasive monitoring [5]. Figure 2.3 represents oscillometry, volume-clamp and applanation tonometry, three of the main non-invasive monitoring techniques, which will be explained below.

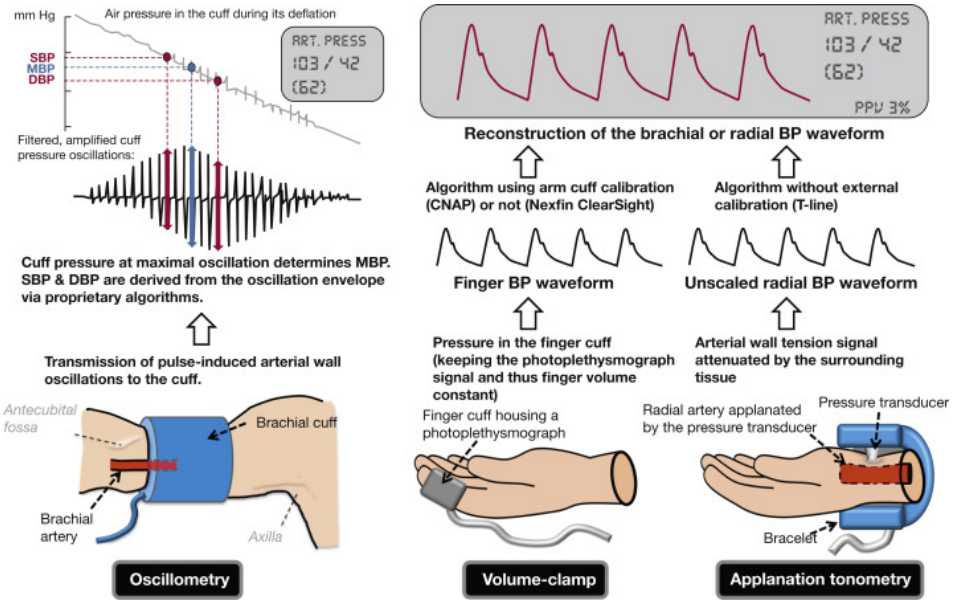


Figure 2.3: Representation of the main non-invasive blood pressure monitoring techniques: oscillometry, volume clamp and applanation tonometry [36].

Intermittent monitoring

Intermittent BP measurement typically involves the use of an occlusion cuff for either manual or automatic measurement. The manual auscultatory method, widely accepted in clinical settings, entails the inflation of a cuff to occlude an artery and the detection of Korotkoff sounds through a stethoscope and manometer during cuff deflation. The first Korotkoff sound signifies the onset of turbulent blood flow, corresponding to SBP, while the fifth sound marks the return to laminar flow and DBP. Alternatively, the palpatory method involves a physician feeling the radial pulse and noting the pressure at which it disappears during cuff deflation, which corresponds to SBP [37].

Additionally, oscillometry is the most widely used automatic and non-invasive technique for measuring BP. It employs a cuff equipped with a pressure sensor to detect small pressure oscillations caused by pulsatile blood flow within the artery. BP values are then estimated from the amplitude and pattern of these oscillations [38]. The main advantages of oscillometry include the provision of reasonably accurate estimates of mean arterial pressure within normal ranges and the ability to automate BP measurements at preset intervals. However, compared to auscultation, oscillometry tends to overestimate low values and underestimate high values, making them less reliable for individuals with cardiac arrhythmias or stiff arteries and measurements. Furthermore, it can be falsified by factors such as patient movement or improper arm positioning. A common disadvantage across all BP measurement techniques is their intermittent nature [5].

Continuous monitoring

Recently, continuous non-invasive monitoring techniques have enabled real-time pulse waveform and numerical SBP value, offering valuable information for the diagnosis and treatment of hypertension, including beat-to-beat variation data. The main techniques employed are arterial applanation, tonometry and the volume clamp method.

Arterial applanation tonometry involves applying an external transducer to flatten a superficial artery against a bone, enabling the acquisition of the arterial pulse wave, SBP and DBP. Alternatively, the volume clamp method utilizes an inflatable finger cuff to measure the diameter of the finger artery with a photodiode, adjusting the pressure to maintain constant diameter measurements [39].

The pulse wave product of applanation tonometry can provide more information than SBP and DBP alone, being useful in cardiology. However, it cannot provide continuous measurements throughout the day, as it must be handheld by the examiner. For its part, volume clamp can be worn for extended periods, making it ideal for long-term monitoring. Their main disadvantages are that peripheral vascular disease can make it challenging to obtain valid waveforms using finger cuffs and some patients may experience discomfort due to congestion where the

cuff is placed, necessitating periodic changes of the cuff to another finger. Finally, all continuous non-invasive devices are sensitive to patient movement and are more expensive than intermittent devices [5].

2.3.3 Proposed monitoring technique

Following a critical analysis of the advantages and limitations of both invasive and non-invasive blood pressure monitoring techniques, current research highlights the growing need for solutions that are wearable, continuous, non-invasive, and unobtrusive. The integration of such monitoring technologies with AI-driven methods offers a promising pathway to enhance hypertension risk classification and support proactive cardiovascular care.

These advanced technologies enable real-time monitoring in daily conditions, allowing the early identification of individuals at elevated cardiovascular risk and supporting personalized intervention strategies. By facilitating continuous assessment outside clinical settings, AI-enhanced monitoring systems can capture transient events and subtle variations in blood pressure patterns that are often missed in conventional episodic measurements.

Furthermore, the non-invasive and transparent nature of these techniques significantly reduces patient discomfort and enhances compliance with long-term monitoring protocols, a crucial factor for the management of chronic conditions like hypertension.

2.4 The electrocardiogram

The electrocardiogram (ECG) is a time-varying signal that reflects the summation of ionic currents driving the contraction and subsequent relaxation of cardiac fibers. Surface ECG recordings are obtained by measuring the potential difference between electrodes placed on the skin, capturing the electrical activity that propagates through the heart during each cardiac cycle. A single normal ECG cycle represents the sequential depolarization and repolarization of the atria and ventricles, triggered by electrical impulses originating from the pacemaker cells in the sinoatrial node [40].

The ECG provides clinically valuable information about the heart's electrical function, including heart rhythm and rate, the integrity of conduction pathways, the condition of the atrial and ventricular chambers, and the presence and extent of ischemia or infarction. Due to its non-invasive nature and diagnostic utility, the ECG remains a cornerstone of cardiovascular assessment in both acute and chronic settings.

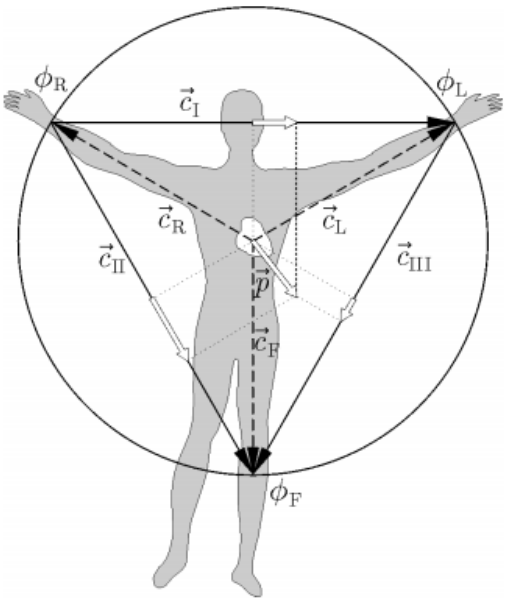


Figure 2.4: Standard limb leads and Einthoven's triangle. It is assumed that the right arm, the left arm and the left leg approximate the vertices of an equilateral triangle with the heart at its center [42].

In 1908, Willem Einthoven introduced the concept of the limb leads, defined as the potential differences between the vertices of an imaginary triangle formed by the two arms and the left leg, as depicted in Figure 2.4. This configuration became the basis for the standard 12-lead ECG system used today.

For this study, only Lead II of the standard ECG is used, where the positive electrode is placed on the left leg and the negative electrode on the right arm. The Lead II voltage V_{II} is calculated as:

$$V_{II} = \Phi_F - \Phi_R \tag{2.1}$$

where Φ_F and Φ_R represent the potentials of the left leg and right arm, respectively [41].

2.4.1 ECG Waves and Intervals

The ECG signal comprises distinct waveforms representing electrical events in different parts of the heart during each cardiac cycle, as illustrated in Figure 2.5. These waves are labeled alphabetically, beginning with the P wave, followed by the QRS complex and the T wave [43].

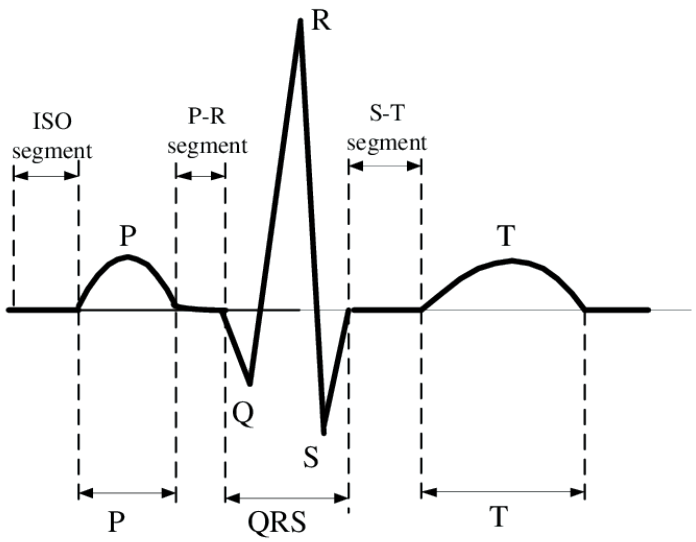


Figure 2.5: ECG typical beat composed of specific waveforms and intervals: P wave, P-R segment, QRS complex, S-T segment and T subwave [44].

- **P Wave**: Represents the depolarization of the right atrium, followed by the left atrium. It appears as a low-amplitude positive deflection preceding the QRS complex, signaling the onset of atrial contraction.

- **PR Interval**: Includes atrial depolarization, conduction through the atrioventricular (AV) node, and propagation through the His-Purkinje system. Its duration varies with heart rate, shortening at higher rates due to improved AV nodal conduction and lengthening at slower rates because of reduced AV nodal conduction.

- **QRS Complex**: Corresponds to ventricular depolarization. The complex begins with a negative deflection (Q wave), followed by a positive deflection (R wave), which predominantly reflects left ventricular depolarization due to its greater mass. A subsequent negative deflection (S wave) represents terminal depolarization of the ventricular walls.

- **ST Segment**: Follows ventricular depolarization and precedes repolarization, representing an isoelectric period in healthy individuals where the ECG typically shows no electrical activity.

- **T Wave**: Denotes ventricular repolarization. The T wave is broader than the QRS complex due to the slower repolarization process, exhibiting a gradual rise, a peak, and a return to the isoelectric baseline.

- **QT Interval**: Encompasses the QRS complex and the ST interval, representing the total duration of ventricular depolarization and repolarization.

In this study, the ECG signal is employed to calculate the time interval between the R wave and the generation of deformations in the PPG signal, referred to as the pulse arrival time. This parameter is widely used due to its strong correlation with BP levels.

2.5 Photoplethysmography analysis

Photoplethysmography (PPG) stands as an optical measurement technique that has been used for the development of small, wearable, and cost-effective sensors. These sensors are designed to detect alterations in blood volume within the microvascular bed of tissues resulting from cardiac pumping. The microvascular tissue network comprises blood vessels with diameters of less than 10 μm and plays an essential role in distributing blood throughout the body's tissues. PPG waveforms contain a wealth of cardiovascular circulatory information [8]. Numerous studies have demonstrated their efficacy in measuring and evaluating parameters such as heart rate [45], blood oxygen saturation [46], blood pressure [47], cardiac output [48], arteriosclerosis [49], and vascular aging [50].

2.5.1 Signal acquisition

The fundamental operation of PPG relies on illuminating the skin and monitoring changes in light absorption. This is typically achieved using LEDs emitting red and infrared light to illuminate the skin, coupled with a photodetector to measure the quantity of light either transmitted through or reflected from the skin [51]. Variations in tissue light absorption are primarily governed by the dynamic circulation of blood within the body, a process driven by the rhythmic systole and diastole of the heart. These fluctuations in absorption are further influenced by the hemodynamic and physiological conditions that stem from arterial properties. These effects manifest as distortions in the PPG waveform profiles.

The visible spectrum of light emitted by the LEDs comprises three components: red, green, and blue (RGB), each exhibiting distinct interactions with tissues, as illustrated in Figure 2.6. The blue component has a relatively weak signal that does not penetrate the epidermis substantially, limiting its informational value. Conversely, the red component, while more penetrating, experiences attenuation as it traverses various tissues, which makes it suitable primarily for measuring blood oxygen levels. Meanwhile, the green component effectively penetrates the epidermis and reaches the capillaries, where it detects blood flow without extensive propagation into larger vessels. This distinct property makes

the green signal the preferred choice in single-channel PPG systems due to its clarity and pronounced variations compared to the other components [9].

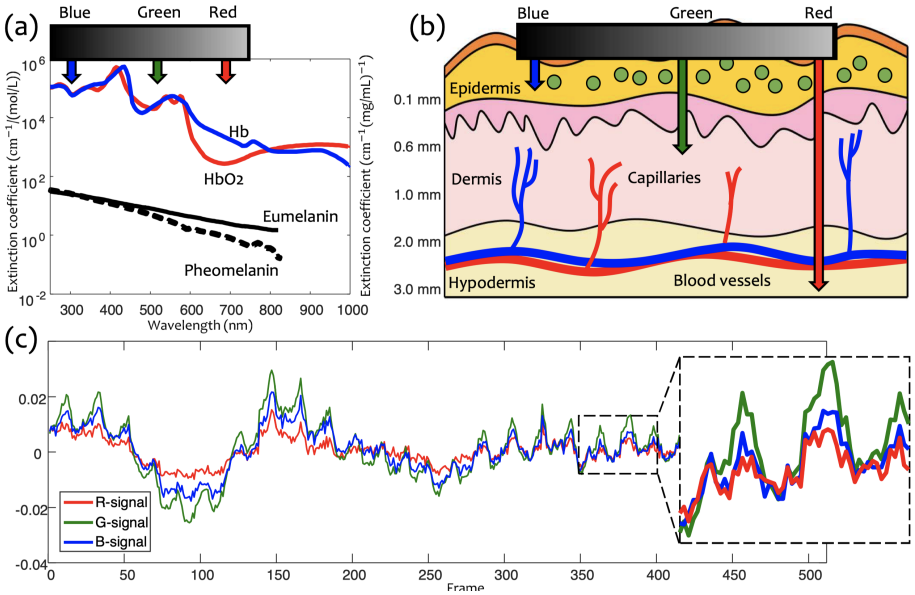


Figure 2.6: Interaction of the RGB components in the skin. (a) Shows the optical absorption spectrum of the PPG. (b) Illustrates the depth of penetration of the LED components in the skin. (c) Represent the pulsatil signal of the three components where it can be seen that the green signal has the highest amplitude [52].

2.5.2 Signal waveform

PPG signal consists of two main components: direct current (DC) and alternating current (AC). DC, known as non-pulsatile component, corresponds to the detected optical signal transmitted or reflected from the tissue, and depends on the tissue structure and the average blood volume of arterial and venous blood. It's noteworthy that the DC component changes slowly with respiration [9].

In contrast, AC, often referred to as the pulsatile component, reflects alterations in blood volume as a result of the heart muscle's contraction during systole, augmenting blood flow in the capillaries near the surface. This, in turn, leads to an elevation in light absorption detected by the photodetector. Conversely, during diastole, as the heart relaxes to fill with blood, BP decreases, causing a reduction in blood flow and light absorption at the photodetector. These BP variations result in a periodic signal representing the cardiac cycle with distinct systolic peaks (representing the blood passage through the sensor as it is ejected from the heart during systole) and more subtle diastolic peaks (corresponding to blood flowing back from the periphery toward the heart during diastole) [53].

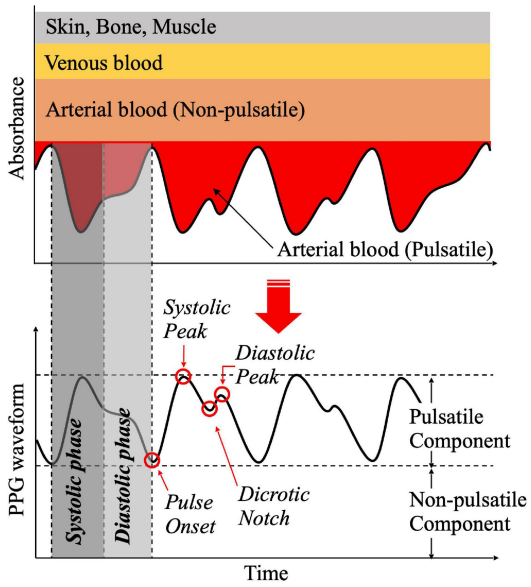


Figure 2.7: Principle of PPG waveform generation, obtained by measuring the variation in light absorption. The signal is derived by inverting the light intensity detected by a photodetector. Characteristic feature points are marked, including the pulse onset and systolic peak during the systolic phase, the diastolic peak during the diastolic phase, and the dicrotic notch, which marks the transition between the two phases [54].

Each cycle comprises a solitary systolic peak, easily identifiable and employed to determine heart rate based on the interval between two consecutive systolic peaks and a single diastolic peak [45]. It is this AC component the one that will be used throughout this work.

PPG waveform and its relation to the amount of light absorption recorded with the photodetector is illustrated in Figure 2.7. The main feature points represented in its pulse waveform are the systolic peak, representing the maximum pulse blood volume, and the pulse onset, representing the lowest blood volume before the systolic peak. Furthermore, fluctuations in the PPG waveform occur during diastole when blood volume experiences a temporary increase due to the presence of a pressure gradient opposing the blood flow, just before the closure of the aortic valve. At this time, the minimum peak is identified as a dicrotic notch, and the point at which the first derivative approaches zero following the systolic peak is designated as a diastolic peak [54].

Given that the PPG waveform itself is relatively simple and provides limited information, derivatives of the signal are also utilized to enhance the evaluation of signal changes attributable to BP variations. These derivatives manifest as the velocity plethysmogram (VPG) and the acceleration plethysmogram (APG) and are derived from one and two consecutive derivatives of the PPG signal, respectively [55], as represented in Figure 2.8.

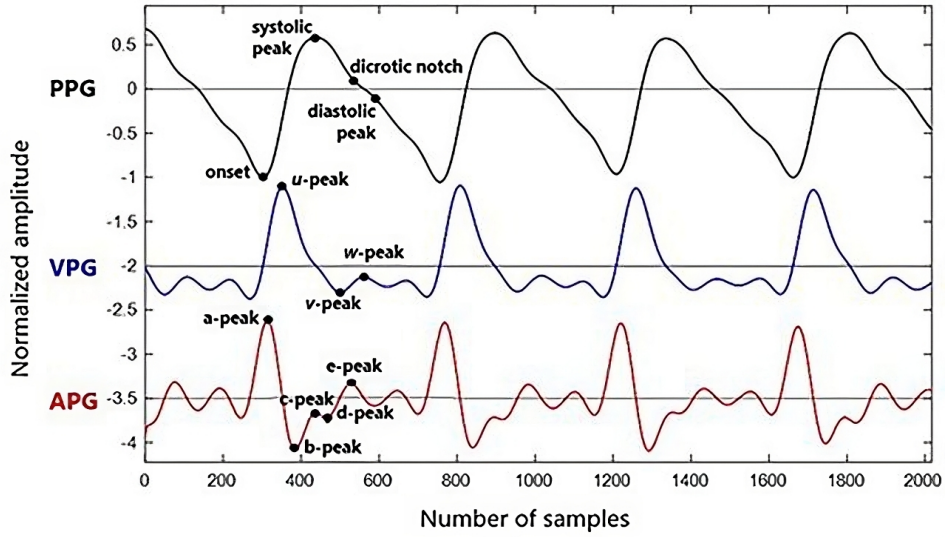


Figure 2.8: PPG, VPG and APG signals waveform and their main fiducial points [56].

2.5.3 Wave propagation theory

The combination of PPG waveform features and features based on wave propagation theory has demonstrated high correlation with BP levels [57]. This theory focuses on the pulse wave propagation between two measurement points, with particular attention given to the analysis of pulse transit time (PTT) and pulse arrival time (PAT), represented in Figure 2.9.

Pulse Transit Time (PTT) from ABP and PPG

The pressure waves generated within the heart propagate through the arteries at a specific velocity termed as the pulse wave velocity (PWV). This velocity is dependent of the elastic characteristics of both the arteries and the blood, and is quantified by the Moens-Korteweg equation [59]

$$PWV = \frac{L}{\text{PTT}} = \sqrt{\frac{E \cdot h}{\rho \cdot 2R}}, \tag{2.2}$$

where PWV is the pulse wave velocity, L is the length of the vessel between two arterial sites, PTT is defined as the time taken for a pressure pulse to transmit between two arterial sites, ρ is the fluid density, R is the inner radius of the vessel, E is the modulus of elasticity of the wall (Young's modulus), and h is the thickness of the vessel.

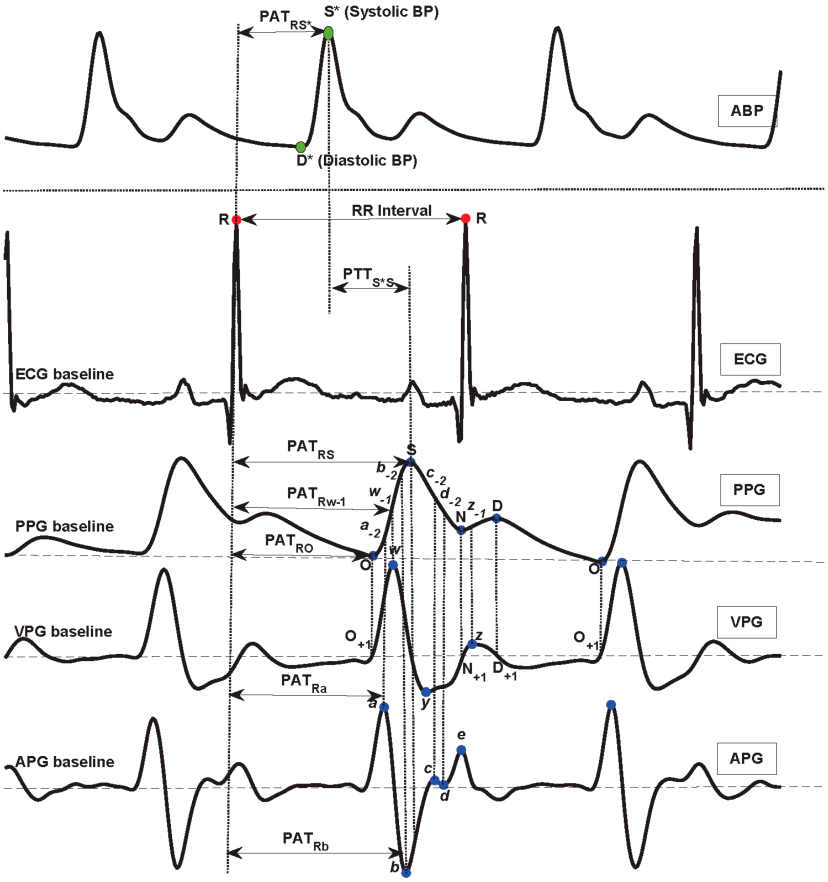


Figure 2.9: Representation of PTT and PAT in PPG, ABP and ECG signals [58].

Furthermore, for an elastic vessel, there is an empirical exponential relationship between E and the fluid pressure P. In this way, the relationship between BP and PTT is demonstrated:

$$E = E_0 e^{\alpha(P-P_0)}, \tag{2.3}$$

where E_0 and P_0 are nominal values of Young's modulus and pressure respectively and α is a constant.

The selection of the two measurement points required to calculate PTT varies depending on the type of study and the sensors to be used. Some studies have utilized two PPG sensors, one proximal and the other distal, or two sensors placed in different distal locations [47,60]. However, in this study, where only a PPG signal from the fingertip is provided, this measurement is used as the distal point, and the ABP signal measured from the upper arm is used as the proximal point [58]

Pulse Arrival Time (PAT) from ECG and PPG

PAT is defined as the time delay between the electrical activity of the heart and the relative changes in blood pressure recorded with distal PPG signal. The electrical activity is recorded with the ECG signal, and for PAT extraction, R-peak is employed [60]. PPG waveform feature points selected for PAT identification are the onset, the W-peak of VPG, and the systolic peak, corresponding to PAT_{foot}, $PAT_{derivate}$ and PAT_{peak}, respectively.

PAT and PTT are often used interchangeably, however, as shown in Figure 2.9, PTT interval is included in PAT interval. Specifically, PAT is the sum of PTT interval and the pre-ejection period (PEP), defined as the delay between the electrical left ventricle depolarization and the initiation of the mechanical ejection of blood from the ventricle [47].

$$PAT = PEP + PTT \tag{2.4}$$

An obvious disadvantage associated with the extraction of PTT and PAT features is the requirement for two sensors to calculate them. For PTT, a proximal PPG or ABP signal and a distal PPG signal are needed, whereas for PAT, both an ECG and a PPG signal are required. The current trend is toward wearable devices that are unobtrusive and comfortable for users, minimizing any inconvenience or discomfort. Requiring multiple sensors complicates the design and usability of such devices, potentially limiting their acceptance for continuous BP monitoring. The need for multiple sensors not only increases the complexity and cost but also reduces the practicality of long-term, everyday use, as it may interfere with the user's normal activities.

Chapter 3

Artificial Intelligence Background

This chapter provides an overview of concepts and methodologies related to AI that are essential for understanding the approaches developed in this thesis. Initially, Section 3.1 introduces the general principles and foundations of AI. Following, Section 3.2 presents a review of ML techniques, including supervised learning methods and feature selection strategies, which form the basis of the ML models evaluated in this study.

Subsequently, Section 3.3 discusses DL approaches, focusing particularly on CNNs due to their proven efficacy in capturing complex, non-linear relationships inherent in biomedical signals such as the PPG. Lastly, Section 3.4 specifically addresses the application of AI to BP monitoring. This section highlights the rationale for integrating AI methodologies into BP assessment, detailing the advantages these techniques offer over traditional measurement and monitoring approaches.

3.1 General Principles of Artificial Intelligence

AI denotes a broad set of computational techniques that enable machines to perform tasks that traditionally require human cognitive functions, such problem solving and learning. It represents a shift in healthcare, driven by the increasing availability of healthcare data and the rapid growth of analytical techniques. Within this domain, two of the most widely used paradigms are supervised learning and unsupervised learning. These paradigms differ primarily in the type of information provided to the model during training.

Supervised classification, the one used in this study, involves the construction of models that are trained on inputs with certain features that are associated with a particular label, so the desired output is known. In this way, classification models can be able to predict the label of new samples from the calculation of the same features selected during training. In contrast, unsupervised learning deals with unlabeled data and aims to uncover hidden patterns or intrinsic structures without predefined outputs. [61].

In this thesis, it is applied to a binary classification task, distinguishing between healthy and hypertensive subjects. By extracting specific features from PPG signals and associating them with known BP categories, the models are trained to identify characteristic patterns indicative of cardiovascular risk. Once trained, these classifiers can be deployed to predict the risk category of new patients based solely on signal-derived features.

3.2 Machine learning algorithms

Machine Learning algorithms are typically classified according to the type of learning process they employ. In this section, it is presented a concise overview of the most relevant algorithm families used in biomedical signal classification, especially those applied in the present work. These paradigms include regression-based models, instance-based methods, tree-based algorithms, ensemble learning strategies, probabilistic approaches, and kernel-based methods.

3.2.1 Regression-Based Models

Regression models aim to model the relationship between input features and a target variable, typically predicting continuous outputs. However, in classification tasks, **logistic regression** is a widely used adaptation of linear regression that enables binary classification [62]. It has been applied in this work as a baseline model due to its simplicity, ease of interpretation, and well-established performance in biomedical applications for disease risk prediction, diagnosis, and treatment outcome assessment [63].

This technique provides a mechanism for applying linear regression principles to classification problems by transforming the output using the logistic (sigmoid) function. As a result, the output of the model is a continuous value between 0 and 1, which can be interpreted as the probability that the given instance belongs to the positive class. The final classification decision is made by applying a threshold, commonly set at 0.5 [64].

3.2.2 Tree-Based Models

Decision tree classifiers use a graphical tree structure to represent classification processes, outlining alternative paths, outcomes, and final class predictions. The algorithm starts from a labeled training set and recursively splits it into subsets based on feature values, aiming to increase the purity of each subset relative to its parent. Internal nodes represent feature-based tests, branches correspond to test outcomes, and leaf nodes indicate the final class labels [65].

This structure enables interpretable and hierarchical decision-making, allowing clinicians to trace the reasoning behind a classification. As represented in Figure 3.1, in this study, a **Coarse Tree** model was employed, with a maximum number of 4 splits and Gini's diversity index as the split criteria to simplify the model and reduce computational complexity [66].

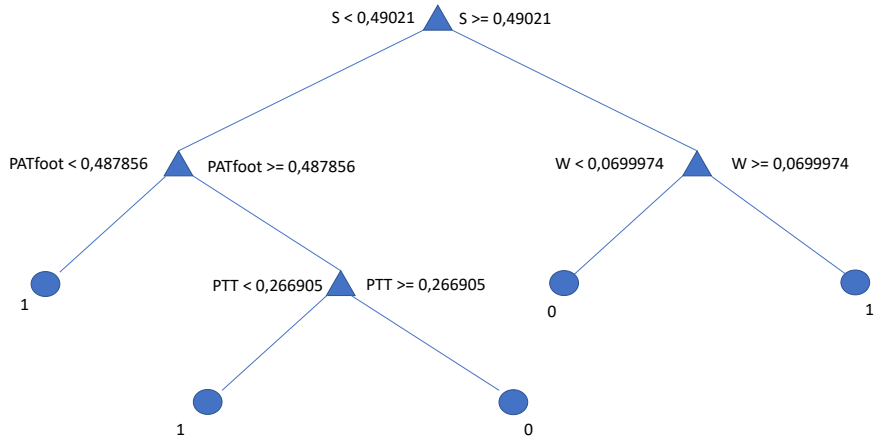


Figure 3.1: Representation of the Coarse Tree model used in this work. It was trained using features extracted from the PPG signal: S-amplitude, PAT_{foot}, W-amplitude, and PPT.

3.2.3 Instance-Based Learning

Instance-based learning (IBL) is a foundational approach in ML where classification decisions are made by directly comparing new instances to stored examples, rather than generating explicit models during training. These algorithms, often referred to as lazy learners, do not construct a general model in advance; instead, they store training instances in memory and defer generalization until classification is required. The underlying principle is that similar instances are likely to share the same class label [67].

The **K-Nearest Neighbor (KNN)** algorithm stands as the most representative IBL method and has been employed in this work. It relies on a k parameter, which defines the number of nearest neighbors to consider when classifying a new instance. Classification is performed by identifying the k closest training samples based on a distance metric, typically Euclidean distance, and assigning the majority class among them [68], as represented in Figure 3.2.

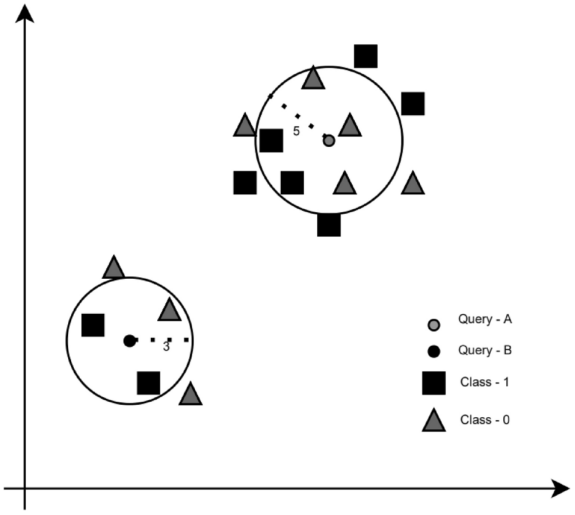


Figure 3.2: Representation of KNN algorithm. As k is 5 for Query A, it searches the 5 nearest neighbors. Then it uses the majority voting rule to classify its class 0 (two of class 1 and three of class 0). Similarly, as k is 3 for Query B, it classifies its class as 1 because there are a greater number of neighbors of this class [68].

3.2.4 Probabilistic Models

Probabilistic models provide a principled framework for modeling uncertainty in machine learning. Unlike deterministic approaches, which output a single prediction, probabilistic models represent knowledge in terms of probability distributions. This allows them not only to make predictions but also to quantify the confidence in those predictions [69].

Naïve Bayes is one of the most widely used probabilistic classifiers due to its simplicity and effectiveness. It can be interpreted as a simplified form of a Bayesian network in which each feature (attribute) is assumed to be conditionally independent of the others, given the class label. In probabilistic terms, this means that the model includes a single parent node (the class) and several independent child nodes (the features), which greatly reduces computational complexity. The classification process relies on the application of Bayes' theorem, where the probability of a class given an instance is estimated by multiplying the conditional probabilities of each feature given the class [65].

3.2.5 Kernel-Based Models

Kernel-based methods are powerful classifiers that implicitly map data from the input space into higher-dimensional feature spaces, allowing for more flexible decision boundaries. This mapping is accomplished through a kernel function, which measures the similarity between pairs of data points. By using kernel functions, computations are performed in the high-dimensional space without explicitly representing the feature vectors, enabling algorithms to work efficiently [70].

Support Vector Machines (SVM) operate by identifying an optimal separating hyperplane that maximizes the margin between two classes, as shown in Figure 3.3. This margin corresponds to the distance between the hyperplane and the nearest data points from each class, known as support vectors. Maximizing this margin helps improve the generalization capability of the classifier.

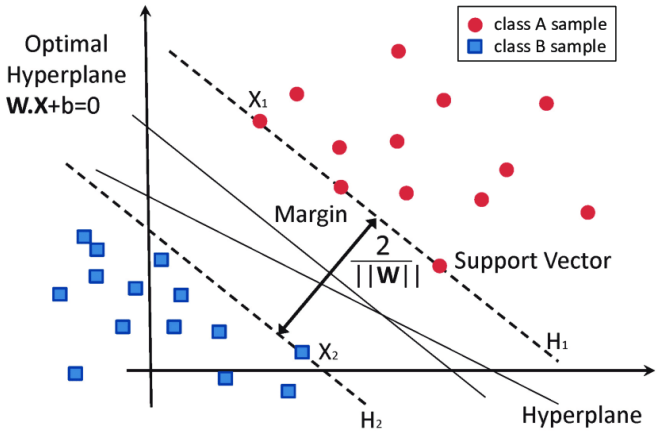


Figure 3.3: Classification of data using SVM. The optimal hyperplane separates classes A and B while maximizing the margin between them. The marginal hyperplanes H_1 and H_2 are shown parallel to the optimal hyperplane and pass through the nearest support vectors of each class, X_1 and X_2 [71].

In this work, **Quadratic** and **Cubic** nonlinear polynomial kernels are proposed. The quadratic kernel transforms the input space to consider second-degree interactions between features. The cubic kernel further extends this idea by considering third-degree interactions, thus allowing more complex boundaries. In both cases, although the decision surface in the original space is nonlinear, the principle of maximizing the margin between classes is preserved in the transformed feature space [72].

3.2.6 Ensemble Learning

Ensemble methods have emerged as a powerful strategy to improve predictive performance and generalization across a wide range of applications. Rather than relying on a single model, these methods combine the outputs of multiple base learners to form a stronger predictive model. In classification problems, the final decision is typically obtained through a majority vote.

Two widely used ensemble approaches are bagging and boosting, both used in this work. Bagging reduces variance by training each model independently on different bootstrap samples of the data and aggregating their predictions. **Bagged Trees** are decision trees trained on resampled datasets help stabilize predictions and mitigate overfitting. Boosting, on the other hand, reduces bias by sequentially training models that focus on correcting the errors of their predecessors. **AdaBoost** improves the performance of weak classifiers by sequentially combining them into a stronger predictive model. Initially, all instances are weighted equally. However, in each boosting round, the algorithm increases the weights of the misclassified examples, making them more influential in the next iteration, while decreasing the weights of the correctly classified ones [73].

3.3 Deep learning algorithms

Deep Learning (DL) is a subfield of ML that leverages neural network architectures composed of multiple layers to automatically learn hierarchical feature representations from data. In this section, a brief overview of the DL models employed in this work is provided, focusing on CNNs and pretrained deep architectures such as GoogLeNet and ResNet.

3.3.1 Convolutional Neural Network models

Neural Networks

Deep Learning, also referred to as deep neural networks, is a subset of machine learning based on artificial neural networks (ANNs) with multiple layers. These

architectures are inspired by the structure and function of the visual cortex in biological systems and are capable of learning complex representations from large amounts of data [74]. The fundamental unit of a neural network is the neuron, which can be represented as a simple classifier model. It takes a bias term ω_0 and a weight vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$ as parameters to compute a nonlinear activation function $h(x)$, which is used to model a decision [75].

$$f(x) = h(\boldsymbol{\omega}^T x + \omega_0) \tag{3.1}$$

where $\boldsymbol{\omega}^T x$ represents the weighted sum of the input features, and ω_0 is the bias term.

The most well-known architecture of a traditional ANN is the feedforward neural network, also referred to as a multilayer perceptron. In this structure, neurons are organized into layers, and the data flows in one direction, from input to output, as represented in Figure 3.4.

Usually, the data is loaded in a multidimensional vector to the input layer, which will distribute it to the hidden layers. Then, the hidden layers will make decisions from the previous layer, transforming the data as it flows according to whether it detriments or improves the final result. This is known as the learning process. In classification models, the output layer has the same number of neurons as the number of classes to predict. Deep neural network is considered when a neural network contains multiple hidden layers; consequently, the term deep learning [76].

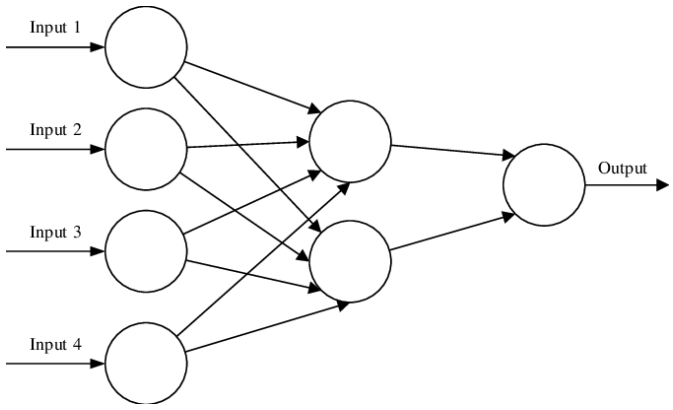


Figure 3.4: A simple three-layer feedforward or multilayer perceptrons neural network comprised of an input layer, a hidden layer and an output layer [76].

Convolutional Neural Networks

Convolutional Neural Network (CNN) is the state of the art for image classification. They are analogous to traditional ANN as both use neurons that receive an input and carry out a nonlinear function to obtain a final output of the class score and use weight vectors. Nevertheless, CNNs are principally used in pattern recognition tasks with images, as they allow for encoding specific features of the images in the architecture [76].

The main advantages of CNNs are the reduction in the number of parameters using large models to solve complex tasks. Besides, CNN do not need to have spatially dependent features, so they are detected regardless of their position in the image. Another relevant advantage is the fact that abstract features are obtained while the input image is propagating toward deeper layers. In this way, the first layers are used to detect the edges, then simple shapes, and in the final ones, high-level features [77]. Before the appearance of CNN, these features were obtained by hand, and now they are automatically learned from image data using the feature detectors of CNNs.

For effective training of a CNN, it is crucial to optimize hyperparameters such as the batch size, defined as the number of images used in each training epoch, and the learning rate or step size at each iteration. An algorithm with an excessively small learning rate hyperparameter will converge slowly, while an overly large one may cause the training process to diverge [79]. Moreover, a batch size that is too small can result in the algorithm oscillating without achieving acceptable performance, while excessively high batch sizes can lead to convergence without meaningful improvements in accuracy [80].

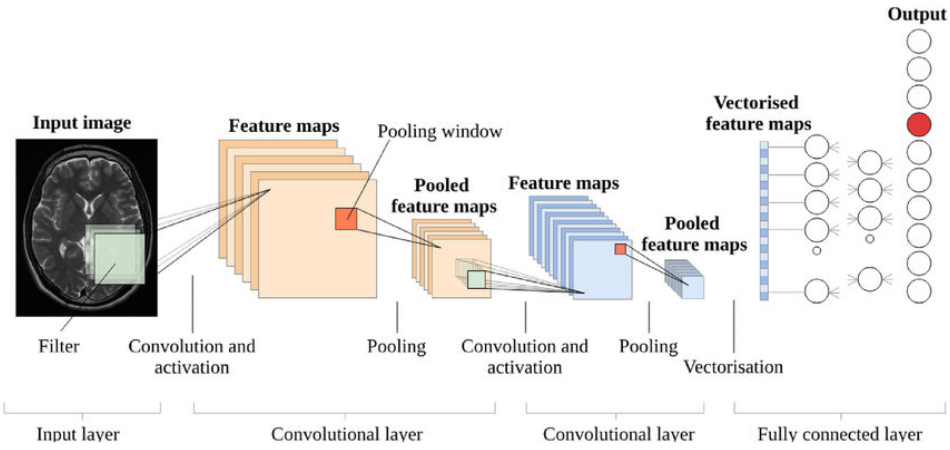


Figure 3.5: Typical CNN building blocks. In this example, the input image is processed by two convolutional layers for feature extraction. The output of the last convolutional layer forms a vectorised feature map, which is passed to a fully connected network. Finally, the neuron with the highest activation in the output layer determines the network's prediction [78].

In contrast to the previous ANN feedforward architecture (Figure 3.4), which is excessively inefficient to connect all nodes from one layer to all nodes in next, they are only connected to a small region of the previous layer to improve the performance. The input image is organized in a grid structure and fed with layers that operate on a small region of the previous layer. CNN has multiple convolutional and activation layers that are usually interspersed with pooling layers. Finally, the final layers are fully connected layers that compute the final outputs of classification [78]. This organization is represented in Figure 3.5, and the main layers are defined below:

- **Convolutional layer.** After the input layer conveys the input images, it is processed by the convolutional layer. It extracts key features using multiple convolution operations from the input feature map to obtain an output feature map. A filter or convolution kernel, described by width, height, and depth, is used to perform the weighted sum in the convolution process. Each layer has different convolution kernels that extract different features as colors or shapes. As illustrated in Figure 3.6, the convolution window acts in a portion of the original feature map called the receptive field.

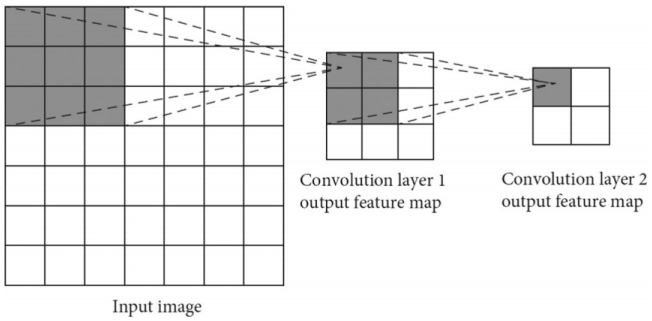


Figure 3.6: Receptive field in the input image and convolutional layers. The 3x3 size receipt field in the input image corresponds to a 3x3 convolutional kernel on the first layer. The same happens with sizes of 2x2 in the following layer [15].

In this way, neurons cannot identify all the information features from the previous layer, only in the receptive field, so the larger the field, the larger the area where the kernel acts on the original image. After the horizontal and vertical displacement of the kernel, at each step, the weight matrix is multiplied by the elements of the input feature map, and at the end, the output feature values are obtained by deducting the offset [15].

- **Activation layer.** The layer after convolution is non-linearity, feeding the input feature map with non-linear activation functions. The Rectified Linear Unit (ReLu) is the most common activation function. This layer filters the output features of the convolution layer to guarantee that the information has been correctly transmitted. With the introduction of nonlinear fac-

tors to the neural network, DL network improves the ability to fit nonlinear features.

- **Pooling layer.** The objective of pooling is down-sampling to gradually reduce the dimensionality and computational complexity for deeper layers, reducing the number of feature values and extracting representative features. One of the most common methods for pooling is Max-pooling, where the image is divided into rectangular sub-regions and only takes the maximum value of the region. Normally, a 2x2 kernel with a stride of 2 is used. This method should be applied when the spatial information is not important, as it is not preserved [77]. Similarly, the same strategy is followed, but taking the average value of the rectangular sub-region in the average pooling. Both methods are represented in Figure 3.7. Another method used is overlapping pooling, where the kernel size is 3 and the stride is set to 2 [76].

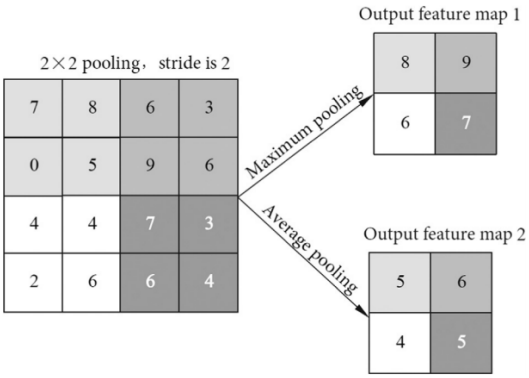


Figure 3.7: Maximum and average pooling methods [15].

- **Fully connected layer.** This layer works similarly to the neurons in traditional networks in the way that each node of the layer is connected to every node of the previous layer, normally the last frames of the pooling layer and the next layer. It takes a long time to train as it includes a lot of parameters that need complex computation. It is possible to remove nodes and connections to reduce this problem with the dropout technique.

3.3.2 Pretrained Models

GoogLeNet

Google proposed an efficient deep convolutional neural network architecture for computer vision named Inception that achieved top results for classification and detection in ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14), becoming one of the best neural networks at the state-of-art [81].

The main problem Inception tries to solve is choosing the right kernel size for the convolution operation, neural network size and computational resources. The important parts of the image vary and require a larger kernel to obtain information of objects distributed in the picture and a smaller kernel for less distributed objects. Moreover, having a bigger and deeper neural network would bring overfitting for having a larger number of parameters and would increase computational resources consumption and inefficiency.

To solve these problems, Inception module forms a "wider" network rather than "deeper". In the naive inception module, 3 different sizes of kernels are used: 1x1, 3x3, 5x5, and a 3x3 max pooling on input from previous layers to perform convolutions, and then the outputs are concatenated and sent to the next inception module. To reduce more the resources of computation, dimensionality reduction is applied by adding 1x1 convolutions before 3x3 and 5x5 ones. The final result is represented in Figure 3.8.

With the introduction of the Inception module, GoogLeNet was built as a 27-layer deep convolutional network using rectified linear unit activations.

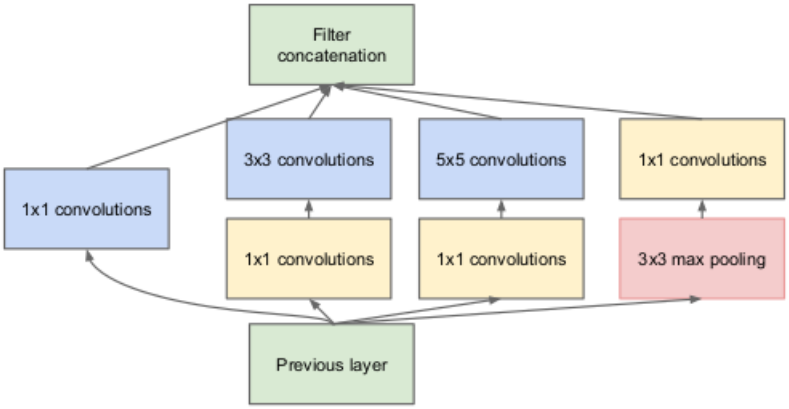


Figure 3.8: Inception module with dimensionality reduction. This module allows for multi-scale feature extraction by applying parallel convolutional filters of different sizes, while 1×1 convolutions are used to reduce dimensionality and computational cost [81].

ResNet

ResNet is a deep CNN that uses a residual learning framework to simplify the training of very deep networks. It achieved state-of-the-art performance and won the ILSVRC15 classification challenge [82]. Contrary to theoretical expectations, increasing the number of layers in a neural network does not always improve accuracy. Instead, it can lead to performance saturation and rapid degradation due to the gradient problem.

The solution proposed for this problem was the residual learning, consisting of adding the input of the hidden layers to their output, so the network only has to learn the residual mapping as seen at the building block of Figure 3.9.

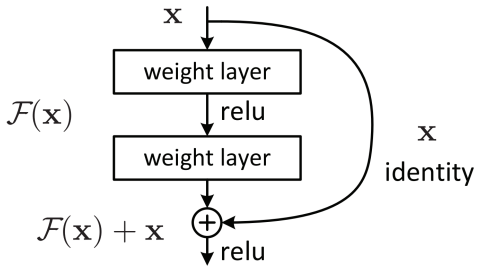


Figure 3.9: Building block of residual learning [82]. x represents the input identify mapping and $F(x)$ the residual function.

In this way, if normally $H(x)$ is the underlying mapping and the layers aim to learn a function $F(x)$ that equals $H(x)$, in ResNets the layers learn the residual function $F(x) = H(x) - x$ and the input x is added afterwards to the result as a shortcut connection since is easier to optimize the residual function than the original one ($H(x)$).

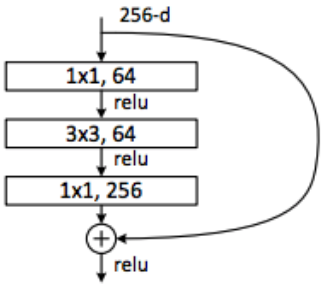


Figure 3.10: Building block of ResNet-50 residual function [82].

In this work, the ResNet-50 CNN with 50 layers deep is used. Although for the authors going deeper reaches better accuracy results, in this study, ResNet-152 did not improve it considerably and consumed much more computational time. In this network, each residual function uses a stack of 3 layers: first a 1x1 convolution layer to reduce dimensionality, then 3x3 convolution layer, and finally a 1x1 convolution to recover original dimensions, as represented in Figure 3.10.

3.4 AI Applications in Blood Pressure Monitoring

The integration of AI into blood pressure monitoring systems has revolutionized how cardiovascular health is assessed. By enabling continuous and adaptive evaluation, AI-enhanced devices with continuous assessment can adapt to each patient's unique cardiovascular profile. This personalization leads to more accurate diagnostics and timely adjustments in treatment plans, significantly improving patient outcomes.

However, one of the major challenges using PPG signals in non-invasive monitoring is the susceptibility to noise caused by factors such as patient movement, variations in ambient lighting, or inconsistencies in sensor placement. Integrating AI techniques, particularly ML and DL models, offers a powerful solution to this problem by effectively filtering out noise and enhancing signal clarity [83,84]. These AI-driven algorithms can be trained to distinguish true physiological signals from artifacts, thereby increasing the accuracy of blood pressure estimations and ensuring reliable monitoring in real-world conditions, where motion and other noise sources are often unavoidable.

Beyond noise reduction, AI integration also addresses other critical challenges in PPG signal analysis, such as variability interindividual and intraindividual. Given that PPG signals can vary significantly between individuals due to factors like age, skin tone, physical condition, and vascular morphology [85, 86], AI-based models excel in adapting to these differences.

Moreover, AI enables dynamic recalibration, allowing models to adjust over time by incorporating new patient-specific measurements. This recalibration ensures that the system can adapt to physiological variations, such as fluctuations in blood pressure due to stress, medication, or circadian rhythms. This adaptability is particularly valuable in wearable blood pressure monitors, where real-time classification can learn from previous readings, improving accuracy and reliability. By continuously refining predictions based on evolving physiological states, these AI-driven systems can reduce diagnostic errors, enhance non-ambulatory monitoring, support early hypertension detection, and optimize treatment adjustments, ultimately leading to better clinical outcomes.

Another significant challenge lies in the non-linear relationship between PPG features and blood pressure. Traditional signal analysis techniques often struggle to capture these complex interactions. However, AI approaches, including neural networks and decision-tree-based models, are adept at handling non-linear patterns, allowing for a more nuanced understanding of how physiological changes influence blood pressure readings [87].

Finally, AI's predictive capabilities extend to the early detection of cardiovascular events [88]. By analyzing large datasets of PPG signals, AI algorithms can identify subtle changes that precede serious cardiovascular incidents, such as hy-

pertensive crises or arrhythmia. This allows for timely intervention, significantly reducing the risk of adverse outcomes in high-risk patients.

In summary, the application of AI in blood pressure monitoring not only overcomes critical challenges such as noise reduction, variability adaptation, and nonlinear signal interpretation but also enhances device accuracy through continuous calibration and predictive analytics. These advancements are transforming the field, offering more reliable, precise, and personalized tools for cardiovascular care.

Chapter 4

Materials and Methods

This chapter provides a detailed description of the materials and methods employed in this study. First, the dataset used for analysis is introduced in Section 4.1, followed by an overview of the signal processing techniques applied to the simultaneous PPG, ECG, and ABP recordings in Section 4.2. The methodology for hypertension risk assessment is then outlined, considering both non-calibrated (Section 4.3) and calibrated (Section 4.4) approaches. Additionally, in Section 4.5, an alternative classification framework is presented, focusing on the detection of hypertension as a pathological condition rather than estimating absolute blood pressure values. Finally, the statistical indices used for performance evaluation are defined in Section 4.6.

4.1 Dataset

The research utilized data from the Medical Information Mart for Intensive Care (MIMIC) database, a freely accessible hospital database containing de-identified data from numerous patients admitted to the Intensive Care Unit (ICU) at the Beth Israel Deaconess Medical Center in Boston, Massachusetts [74]. This dataset was specifically selected due to its simultaneous recording of multiple digital vital signs, including ECG, PPG, and ABP. These signals were acquired at a sampling frequency of 125 Hz, with a resolution ranging from 8 to 10 bits. The duration of the recordings varied across patients, depending on their length of stay in the ICU and the monitoring period.

The first phase of the methodology involved isolating segments from each patient's overall records where the three studied signals were simultaneously recorded. Given the critical condition of patients within the ICU, the monitoring of vital signs is not always simultaneous and varies based on treatment and the attending physician's considerations. Afterwards, an accurate selection process was employed to identify and extract segments devoid of noise or artifacts.

Across the different approaches proposed to achieve the final objective of discriminating subjects based on the PPG signal, both the number of subjects and the volume of measurements analyzed have progressively increased. Additionally, the criteria used to label subjects as healthy or hypertensive have been adjusted in an effort to identify a classification scheme that better aligns with real-world clinical scenarios, as represented in Table 4.1.

Table 4.1: Number of recordings, hypertension labels depending on SBP values and grouping of labels for discrimination in each approach studied.

AI Method	Study	Subjects	Recordings	Hypertension Labels	Comparisons
Machine Learning	No Calibration	86	86 (60s)	35 NT (<120 mmHg) 26 PH (120-140 mmHG) 25 HT (>140 mmHg)	NT vs PH NT vs HT NT+PH vs HT NT vs PH+HT
	Calibration	69	913 (120s)	45 NT (<130 mmHg) 24 HT (>130 mmHg)	NT vs HT
	Pathology Detection	51	668 (120s)	33 NT (<120 mmHg) 18 HT (>140 mmHg)	NT vs HT
Deep Learning	No Calibration	50	635 (120s)	260 NT (<120 mmHg) 174 PH (120-140 mmHG) 201 HT (>140 mmHg)	NT+PH vs HT NT VS PH+HT
	Calibration	69	974 (120s)	45 NT (<130 mmHg) 24 HT (>130 mmHg)	NT vs HT

Since signals were acquired using commercial PPG and ECG devices and an intra-arterial catheter in the radial artery for ABP measurement, significant artifacts were present. They were primarily caused by sensor malfunctions, patient movement, and interference from other physiological signals, such as respiratory activity. Figure 4.1 illustrates typical artifacts identified in the MIMIC database.

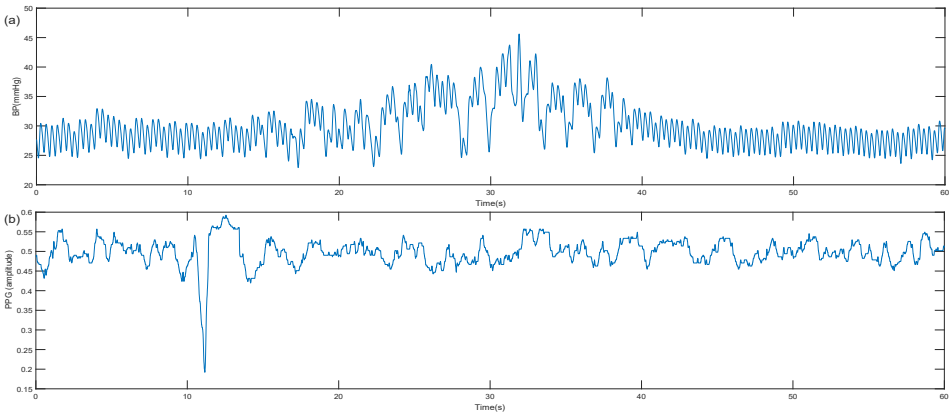


Figure 4.1: Example of ABP and PPG signals artifacts in MIMIC database. (a) ABP signals presenting systolic values below 75 mmHg and exhibiting fluctuations between seconds 20 and 40. (b) PPG signals affected by noise and artifacts, resulting in the absence of a distinguishable systolic and diastolic wave pattern.

The main artifacts identified are:

- Physiologically implausible SBP values outside the range of 75-170 mmHg.
- Unstable SBP values observed within a 120-second window, exhibiting varying blood pressure labels. Similarly, intervals bordering two classifications (120 and 140 mmHg) were excluded due to insufficient representation of a distinct classification group.
- Significant noise and fluctuations in PPG, ECG or BP signals.
- Unusual waveform morphology that lacks differentiation between systolic and diastolic waves.
- Flatlines in both PPG and ABP signals when consecutive samples maintain a constant value.
- Flat peaks in ABP signal, indicating identical SBP values across consecutive samples, possibly due to sensor-related issues.

4.2 Signal Processing

The experiment was executed under MATLAB (MathWorks, Natick, MA, USA), a scientific and engineering computing software, running on a computer equipped with an Intel i7-8700 CPU @ 3.2 GHz, 16 GB of memory.

Raw PPG signals were processed using a fourth-order Chebyshev Type II bandpass filter with cutoff frequencies between 0.5 and 10 Hz, aiming to suppress

minor noise and artifacts caused by factors such as poor sensor contact, patient movement, or interference from other physiological activities, particularly respiratory activity, and to improve the signal quality index [89]. These artifacts were not significant enough to lead to signal dismissal during the prior selection stage. Furthermore, the mean value of the filtered PPG was removed to prevent drifts and to allow a better comparison between different signals.

Since the waveform of the PPG signal itself is rather simple and not very informative, the derivatives of the signal were also used to better assess the changes in the signals caused by BP. VPG represents the velocity waveform of PPG signals, and APG represents the acceleration waveform of PPG signals and was obtained by applying the first and second order derivatives, respectively, to the processed PPG signal [90].

In contrast, the ABP signals, reflecting BP changes across the cardiac cycle, were clear and required no additional processing. For its part, standard preprocessing was applied to each ECG [91]. Thus, they were high-pass filtered with cutoff frequency of 0.5 Hz to remove the baseline, and then low-pass filtered with a cutoff frequency of 50 Hz to reduce high-frequency muscle noise and power line interference, which in this case was 60 Hz [91].

4.3 Hypertension Risk Assessment

Hypertension classification is essential for early detection and risk stratification, supporting timely intervention and contributing to the prevention of cardiovascular complications. In this study, two distinct computational approaches were explored for hypertension risk classification. Machine learning approach relies on manually extracted features derived from PPG and ECG signals, requiring prior identification of fiducial points. In contrast, deep learning approach eliminates the need for manual feature extraction by directly learning representations from raw PPG signals using CNNs. The following subsections describe these methodologies in detail, highlighting their implementation, advantages, and performance in hypertension risk classification.

4.3.1 Machine Learning Approach

Machine learning approach often involves the initial step of manually extracting signal features to gather numerical data. This process is followed by rigorous statistical data preparation and meticulous feature selection. Subsequently, classification models are utilized to distinguish between individuals with hypertension and those who are healthy. As a graphical summary of this approach, Figure 4.2 shows the different stages in the process of creating models for the hypertension risk classification based on ECG and PPG signals employing ML models.

In this initial approach, 86 records, each lasting 60 seconds, were collected from different subjects, distributed across 35 normotensive, 26 prehypertensive, and 25 hypertensive individuals.

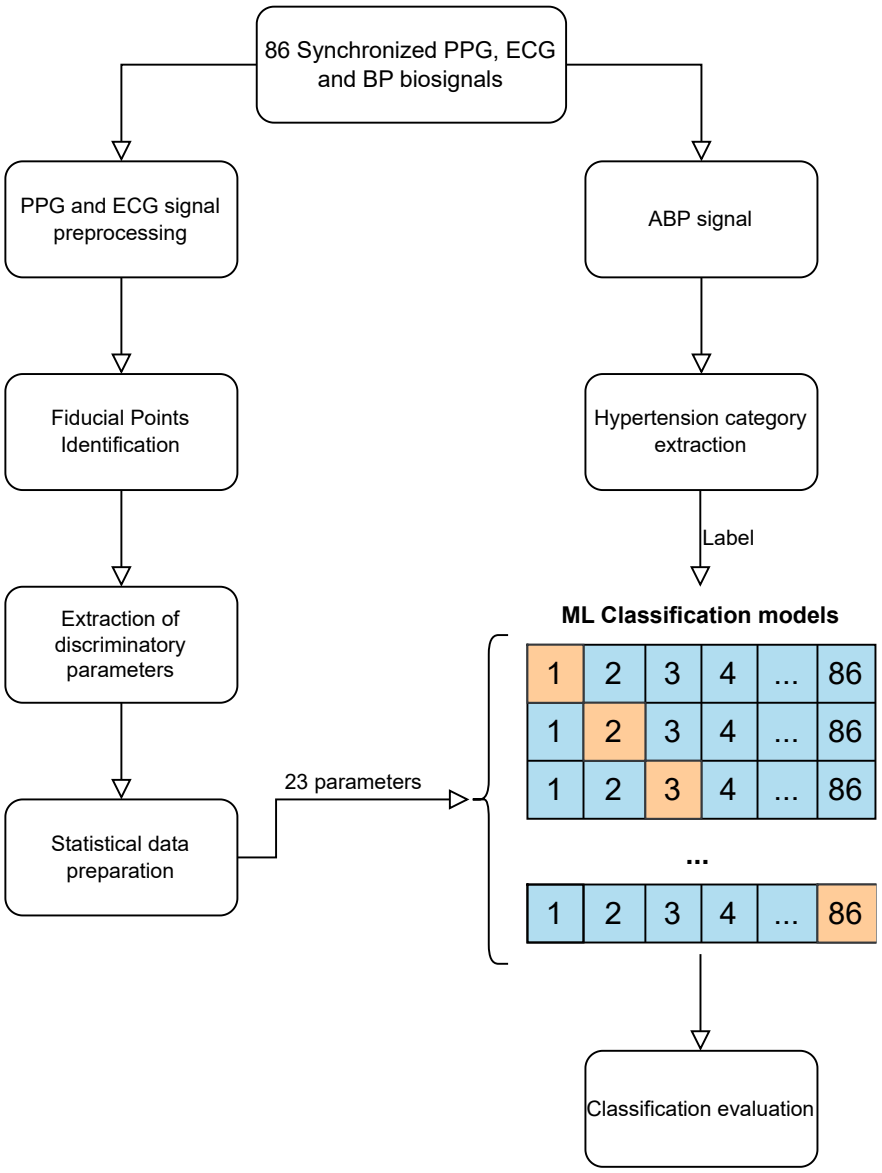


Figure 4.2: Machine learning classification process. PPG and ECG signals were used to extract discriminatory parameters following the identification of fiducial points, while the hypertension risk category was determined from the synchronized ABP signal. Finally, ML classification models were evaluated using the constructed feature matrix and the corresponding classification labels as input, employing a leave-one-out cross-validation strategy to assess model performance.

Fiducial Points Identification

PPG and ECG signals are characterized by their rich physiological information. However, the effective classification of these signals hinges on the accurate extraction and selection of feature points, a critical first stage in ML-based analysis. This process involves identifying and isolating key points within the PPG and ECG waveform that encapsulate the most relevant information, thereby enhancing the performance of subsequent classification algorithms. Optimizing the selection of these feature points can significantly enhance the accuracy, computational efficiency, and interpretability of ML models, making it a fundamental step towards developing robust and reliable systems.

PPG onset indicates the beginning of the pulse in the systolic phase, where oxygenated blood begins to flow at the measurement site. Then, the increasing pressure in the artery is reflected in the ascending edge from the onset to the systolic peak or maximum peak of the pulse. The dicrotic notch and diastolic peaks correspond to wave reflections from the periphery and are modulated by factors such as arterial stiffness, vascular resistance, and compliance. The diastolic point indicates the end of the ejection phase, coinciding with the closure of the aortic valve. With the aging process, there is an observed attenuation of both the notch and diastolic components within the signal, so they are not detected in the study. APG feature points analysis is more habitual than VPG, where only the systolic peak is studied. From APG, four waves are identified from the systolic phase and an early diastolic wave, which corresponds to a dicrotic notch [56].

For ABP signals, the extraction of SBP involved identifying the amplitude of the maximum point in each ABP pulse. This parameter was employed to categorize each PPG segment into distinct BP category labels, a process made feasible through the simultaneous acquisition of PPG and ABP signals.

Subsequently, an R-peak detector, implementing the Pan-Tompkins algorithm, was employed to determine the position of each beat in the ECG signal [92]. Widely recognized for its robustness, the Pan-Tompkins algorithm manages to minimize both false negatives and false positives.

In summary, the feature points extracted to PPG, VPG, APG, ABP and ECG signals are illustrated in Figure 4.3 and listed below:

- **PPG**: Onset (O), Systolic peak (S).
- **VPG**: Maximum peak in systole (W).
- **APG**: Maximum peak in systole (a), maximum peak after a-peak (b), first positive peak after b-peak (c), first negative peak after c-peak (d), beginning of the diastolic component (e).
- **ABP**: Systolic blood pressure (SBP).
- **ECG**: R-peak.

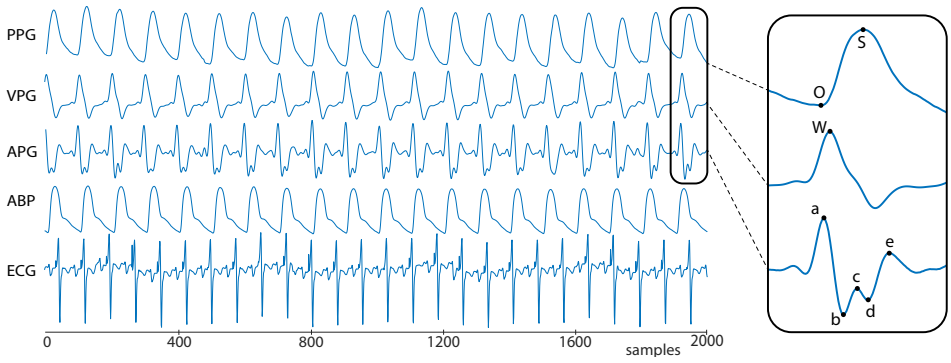


Figure 4.3: Fragment of 2000 samples illustrating signals morphology, together with the representation of the characteristic points of the defined PPG, VPG and APG signals.

Extraction of discriminatory parameters

Following feature points identification, the next critical step in the ML classification process is the extraction of discriminatory parameters. These parameters, derived from the selected feature points, are essential for distinguishing between normotensive and hypertensive stages, as PPG, VPG and APG morphology parameters are closely associated with BP estimation [90]. This process involves transforming raw feature points into a set of quantifiable metrics that capture the underlying patterns and variations within the PPG signal. Identifying parameters with high discriminatory power enables classification models to more effectively distinguish between classes, thereby improving classification accuracy.

Among these parameters, those based on wave propagation theory as Pulse Transit Time (PTT) and Pulse Arrival Time (PAT) are particularly noteworthy, as they provide essential insights into cardiovascular dynamics, as extensively discussed in Subsection 2.5.3. PTT offers information about the time taken for a pulse wave to travel between two arterial sites, thereby reflecting arterial stiffness and vascular health [47]. In this case, PTT is extracted as the time delay between SBP peak in the ABP signal and the S-peak in the PPG signal. PAT, on the other hand, defines the time difference between the onset of electrical activity in the heart, measured with ECG, and the subsequent changes in the PPG waveform. This can be classified into PAT at the foot of the waveform (PAT_{foot}), the first derivative ($PAT_{derivative}$), and the systolic peak (PAT_{peak}), each offering different insights into cardiovascular function [11, 60].

Regarding PPG waveform morphology parameters, PPG systolic peak (S) is commonly used as it serves as an indicator of pulsatile changes in blood volume caused by arterial flow around the measurement site, demonstrating a direct correlation with local vascular distensibility across a broad range of cardiac output [93]. The rising time or crest time, defined as the time taken to reach S-

peak from the onset of the waveform, reflects the rate of ventricular ejection and arterial stiffness, thus being longer in hypertensive patients [54].

Time peak to peak (TPP) and time pulse interval (TPI) are obtained by the elapsed time between two consecutive S-peaks and O-notches, respectively. They correspond to the R-R interval of the ECG, as they represent a complete cardiac cycle, allowing heart rate extraction from the PPG signal without the need for an ECG signal [54].

Additionally, PPG pulse width, measured at half the height of S-peak, is correlated with systemic vascular resistance. The pulse area, defined as the integral of the pulse wave, can be divided into systolic and diastolic sections based on the dicrotic notch. Inflection point area (IPA) or ratio between these areas offers valuable insights into total peripheral resistance, which can be used to differentiate between healthy and hypertensive individuals [54].

In the VPG signal, as it is hardly used in literature, only w-peak amplitude is extracted, reflecting the velocity of the amplitude changes in PPG signal, thus offering insights into the speed and intensity of blood ejection from the heart [56].

For its part, the APG signal provides many features that reflect the blood acceleration at the PPG measurement. These features are based on the relationships between the amplitudes of their waves, which will be positive or negative depending on the baseline [93]. Firstly, the ratios involving the a-wave constitute pivotal features in classification models. The b/a ratio reflects the increase in arterial stiffness, exhibiting an age-associated increase. Conversely, c/a, d/a and e/a ratios have an opposite physiological meaning, reflecting a decrease in arterial stiffness that corresponds to a decrease with age. Furthermore, the $(b-c-d-e)/a$ ratio increases with age and may be useful for the assessment of vascular aging and arteriosclerosis detection. In instances where certain waves are not well-defined, alternative ratios such as $(b-e)/a$ and $(b-c-d)/a$ may be employed [93]. Evdochim *et al.* [94] proposed a more comprehensive ratio: $(c+d-b)/a$ and distinguished seven categories of APG pulse based on the patient's circulatory system state. Furthermore, in the same way as $S-S$ interval in PPG signal, the APG $a-a$ interval is employed as an alternative to obtain heart rate, as it correlates with a complete cardiac cycle.

Finally, a total of 23 classification parameters were extracted from the studied biosignals [90, 95, 96]. These parameters encompass a comprehensive range of temporal, morphological, and wave propagation metrics, each selected for their discriminatory power, allowing for a detailed analysis of both normotensive and hypertensive conditions. These discriminatory parameters are listed below and represented in Figure 4.4.

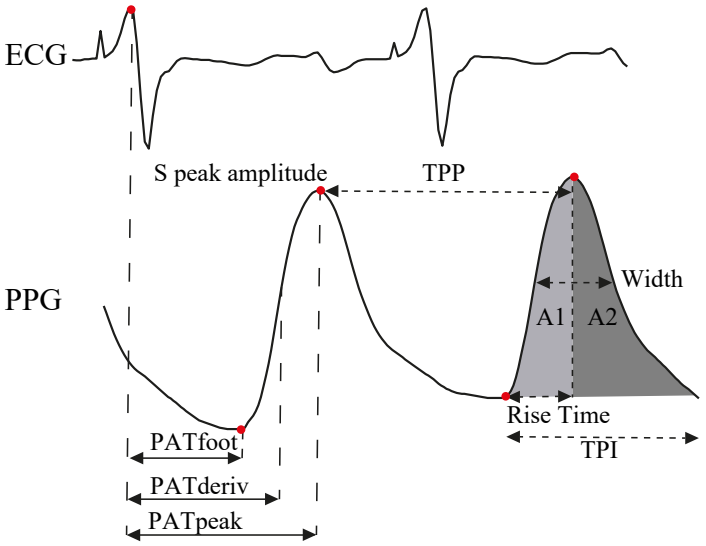


Figure 4.4: Representation of the main discriminatory parameters extracted from synchronized ECG and PPG signals.

- **PAT:** Time interval between R peak in ECG signal to: the O-notche (PAT_{foot}), the maximum slope of PPG signal or W peak of VPG signal ($PAT_{derivate}$) and S-peak (PAT_{peak}).
- **PTT** Time interval between SBP peak in ABP signal to S-peak.
- **Sistolic peak amplitude in PPG:** Amplitude from the baseline to S-peaks.
- **Sistolic peak amplitude in VPG:** Amplitude from the baseline to W-peaks.
- **TPP:** Time interval between two consecutive S-peaks.
- **TPI:** Time interval between two consecutive O-notches.
- **Rising Time:**:Time interval between O-notch and S-peak.
- **Width:** PPG Pulse width at half of S-peak height.
- **Pulse area:** Trapezoidal integration of PPG signal between two consecutive O-notches.
- **Area 1:** Trapezoidal integration of PPG signal from O-notche to S-peak.
- **Area 2:** Trapezoidal integration of PPG signal from S-peak to O-notche.
- **IPA:** Ratio of both areas (A2/A1).
- **a-a:** Time interval between two consecutive a-peaks.
- **Ratios between APG waves with the a-wave:** b/a, c/a, d/a, e/a.
- **Complex APG ratios:** (b-c-d-e)/a, (b-e)/a, (b-c-d)/a, (c+d-b)/a.

Statistical data preparation

After obtaining the values for each discriminatory parameter in all data records, it is imperative to conduct statistical data preparation to address missing values and outliers. Missing values reduce the available data for analysis and compromise its reliability, introducing bias into the results and corrupting the efficiency of the input matrix [97]. To mitigate this, values identified as not a number (NaN) due to incomplete extraction of a predictive parameter in a record will be substituted with the median of the values for that parameter within the same hypertension classification.

Outliers can distort statistical estimations of average and standard deviation, leading to either overestimation or underestimation of values. These outliers, representing atypical values potentially stemming from issues in the detection method or signal defects, require careful handling. To this end, a logical array is generated to identify outliers using the absolute deviation from the median (MAD) method [98], which detects as outliers those values that are more than three scaled mean absolute deviations away from the median. Similarly, to handle missing values, outliers identified in the dataset will be addressed by replacing them with the median of the feature values from patients sharing the same classification label as the outlier. This approach ensures that outliers are treated within the context of the specific subgroup they belong to, maintaining the integrity of the data analysis process

Implementation details

Following data preparation, a total of 23 classification parameters were extracted from the synchronized PPG, ECG, and BP biosignals. These features served as the input for training multiple ML classifiers. The selection of parameters is crucial to the model's performance, as they directly influence the algorithm's ability to generalize across subjects and accurately classify individuals into the appropriate blood pressure categories.

Given the dataset's composition, which includes three distinct patient groups, it was adopted a binary classification approach that allowed for comprehensive, paired comparisons between these groups. Four main binary classification scenarios were defined: (1) NTS vs. PHT, (2) NTS vs. HTS, (3) NTS and PHT combined vs. HTS, and (4) NTS vs. PHT and HTS combined. This binary classification framework allowed for a more granular assessment of the classifiers' ability to distinguish between different stages of cardiovascular health.

To evaluate the models' performance robustly, a leave-one-out cross-validation strategy was implemented. In this validation framework, the classification algorithm was applied once for each individual record, treating that record as the test set and using all other records in the dataset as the training set. This method ensures that each record is tested exactly once, providing a robust performance

assessment while minimizing bias by utilizing nearly the entire dataset for training in each iteration [99].

The initial selection of ML classification models comprised Logistic Regression, AdaBoost Tree, KNN, and Bagged Tree, chosen based on their proven reliability and effectiveness in hypertension risk classification [57]. These models represent a diverse set of fundamental classification methodologies for a preliminary comparative analysis of performance and interpretability.

To further optimize classification performance and explore the influence of algorithmic diversity, the analysis was expanded to include a total of 37 classification strategies. This extended set incorporated models from key families such as logistic regression, Naïve Bayes, discriminant analysis (both linear and quadratic), SVM with various kernels, KNN with different distance metrics, ensemble classifiers (boosting and bagging), and multiple decision tree variants [100].

This exhaustive exploration aimed to identify models capable of effectively discriminating between hypertensive and non-hypertensive subjects using features extracted from PPG and ECG signals, while also evaluating their performance under varying labeling criteria.

4.3.2 Deep Learning Approach

In contrast to traditional ML-based approaches, which rely on the manual extraction of signal features, DL techniques offer a distinct advantage by automatically learning and extracting the most relevant features from generated images. Moreover, DL-based methods provide a significant improvement by eliminating the dependency on ECG signals, simplifying the classification process.

To enhance the number of analyzed segments, the raw 120-second PPG signals were initially divided into 5-second segments. This segmentation strategy is necessary to increase the amount of training data available, which is crucial for DL models to learn robust and generalizable patterns from the input signals. These segmented signals were then transformed into images using continuous wavelet transform (CWT), applied to both the original and downsampled signals. Finally, the resulting images were used as input for pretrained deep neural networks, which have demonstrated exceptional performance in image classification tasks. This paradigm shift underscores the potential of DL methodologies to revolutionize the analysis and interpretation of medical data, particularly in healthcare diagnosis field.

For this approach, a total of 635 recordings from 50 patients were selected, each with simultaneous and stable ABP and PPG signals of 120 seconds in length and sampled at 125 Hz. As a graphical summary of this sub-chapter approach, a flowchart is presented in Figure 4.5, showing the different stages in the process of creating models for the PPG signals classification in DL approach.

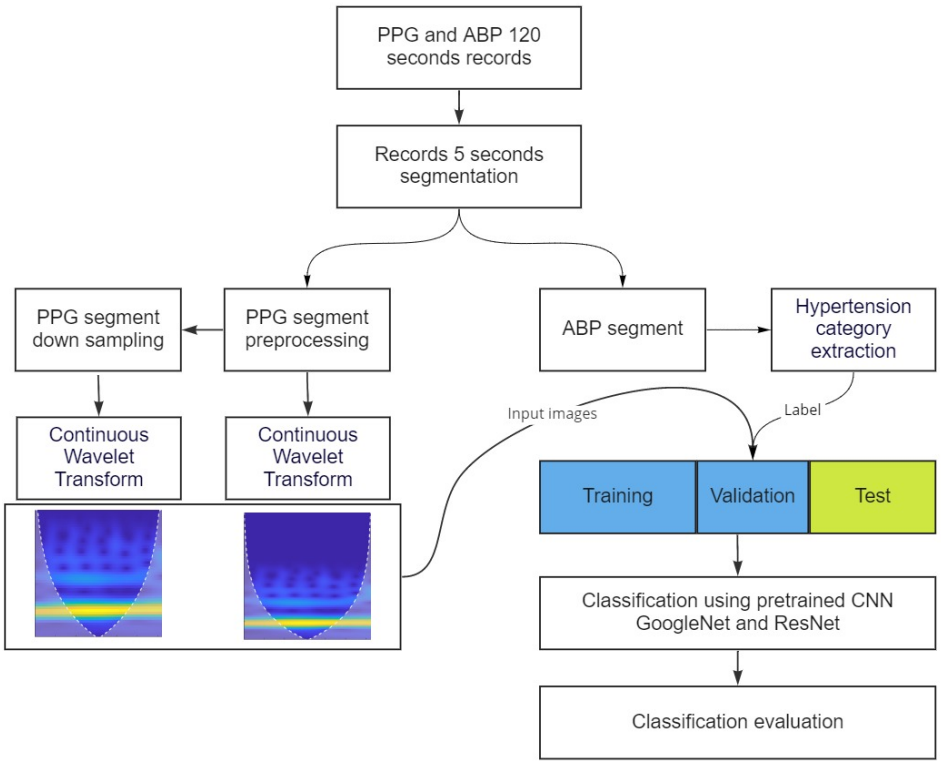


Figure 4.5: Deep learning classification process. Raw and downsampling 5-second PPG signal segments were processed with CWT to feed pretrained DL models. Hypertension risk category was determined from the synchronized ABP signal.

Pretrained CNN Architecture

For the problem of hypertension risk assessment, pretrained CNNs ResNet and GoogLeNet were employed. To retrain the pretrained networks, the last three layers of the original architectures were replaced in order to adapt them to the new image dataset. These layers use image features and information from the previous convolutional layers to classify the input images. The original fully connected layers were replaced with new fully connected layers, with the same number of outputs as the number of classes, in our case, two. Similarly, the original softmax layer and the original classification layer were replaced with two new layers of the same type. Figure 4.6 illustrates the block diagram of the modified architectures of Googlenet and ResNet CNNs.

Training options established a minimum batch size of 128 and an initial learning rate of 1×10^{-4}. The validation frequency was adjusted based on the number of training images and the ratio between training images and the minimum batch size. Additionally, the effectiveness of the classification from the models

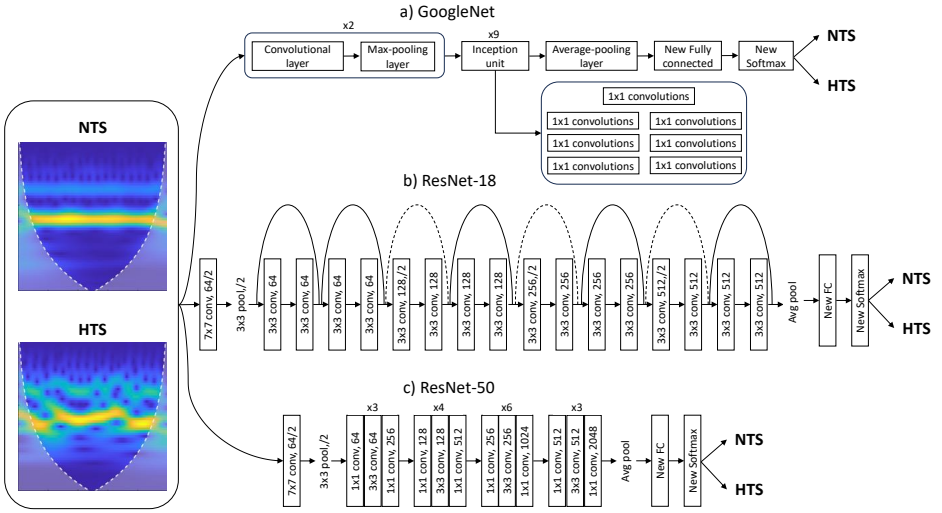


Figure 4.6: Block diagram illustrating the modified architectures of (**a**) GoogleNet, (**b**) ResNet-18, and (**c**) ResNet-50. To align these models with the new task of assessing hypertension risk using PPG recordings, the original fully connected, softmax, and classification layers have been replaced. Additionally, input images with both labels have been resized to 224 $\times$ 224 $\times$ 3 to meet the specific requirements of the studied models.

was compared using two different training solvers, Adam [101] and SGDM [102], with the aim of finding the best training hyperparameter.

In supervised learning, overfitting can be a significant issue when the model struggles to generalize from the training set to the testing set. This issue was mitigated in this work through the use of early stopping, a technique that automatically halts training when the validation error starts to increase [103].

Image Representation of PPG Signal Using CWT

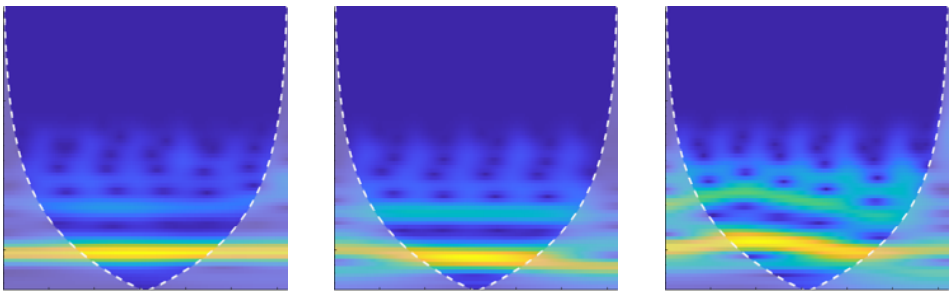
Since both ResNet and GoogLeNet CNNs require inputs of size 224 $\times$ 224 $\times$ 3 in RGB format, PPG segments were processed using CWT [104] and transformed into a scalogram, which provided the absolute value of CWT coefficients of the PPG signal. These scalograms were then resized to 224 $\times$ 224 $\times$ 3 to be fed into the training models. The absolute value of the wavelet coefficients was obtained using the analytic Morse (3, 60) wavelet, with the Voices per Octave set to 12. The choice of CWT and scalograms was based on their effectiveness in providing time-frequency information and useful learning features for CNNs [17].

Additionally, white lines were added to the scalogram to mark the cone of influence. This marks areas in the scalogram that may be affected by border distortions due to missing values at the beginning and end of the transform. The un-

shaded region guarantees that the information is an accurate time-frequency representation of the PPG signal.

Other recent studies have employed spectrograms computed using the STFT and transforming the PPG segment into the instantaneous frequency and spectral entropy [18, 19]. Thus, it is obtained using a window function that moves along the time axis of the signals and provides at a particular instant the time-frequency information. This representation has obtained worse classification results than using CWT as the identification of the low-frequency or fast-changing frequency component of the signal is carried out better with scalograms.

In Figure 4.7 is represented an image example for each hypertension category. It can be observed that the upper portion of the images lacks significant frequency information.



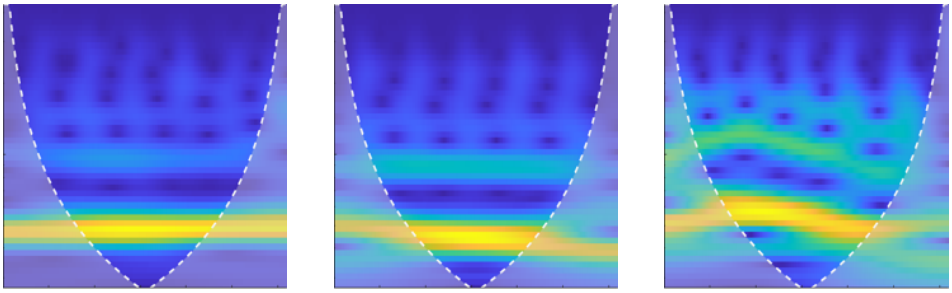
(a) Normotensive record. **(b)** Prehypertensive record. **(c)** Hypertensive record.

Figure 4.7: Scalograms representing CWT with the cone of influence of (a) normotensive, (b) prehypertensive, and (c) hypertensive PPG segments.

To address this lack of information, the original PPG signals were downsampled by a factor of 5, resulting in final signals with a sampling rate of 25 Hz. This downsampling was implemented to evaluate whether it could enhance the classification performance of their CWT representations. A higher downsampling factor was avoided to preserve critical information in the images.

Moreover, the impact of downsampling the PPG signals was analyzed considering potential applications in mobile health. It offers several practical advantages, such as decreased power consumption and lower computational load, which are particularly beneficial for the implementation of cuff-less blood pressure monitoring devices. Additionally, minimizing the size of transmitted data reduces database storage requirements, making the approach more efficient for storing and visualizing patient information via mobile applications [105].

Figure 4.8 represents the effect of downsampling the PPG signals to 25 Hz in the scalograms. This adjustment ensures that almost all pixels in the image contain significant information, which is expected to enhance the performance of future classification models



(a) Normotensive record. **(b)** Prehypertensive record. **(c)** Hypertensive record.

Figure 4.8: Scalograms representing CWT with the cone of influence of (a) normotensive, (b) prehypertensive, and (c) hypertensive downsampled PPG segments.

Implementation details

This approach aims to identify the most effective method for hypertension risk assessment using scalograms of PPG signals as inputs to train pretrained CNN models, specifically GoogleNet and ResNet. Unlike traditional ML methods, this DL approach does not rely on ECG recordings or manual feature extraction for classification.

The PPG signals were first segmented into consecutive 5-second windows. Each segment was then transformed into a scalogram, resulting in a total dataset of 15,240 images. These were split into training and test sets following an 80%-20% partitioning strategy, yielding 12,192 training images from 508 recordings and 3,048 test images from 127 recordings. To ensure subject independence, recordings from individual patients were uniquely assigned to either the training or the test set.

Additionally, the training dataset was further split into 80% for training and 20% for validation to prevent overfitting during model training. It is important to note that this division was applied to the images rather than the recordings, thus, segments derived from the same recording could appear in both the training and validation subsets. While this may introduce some dependency, the separation of test data at the patient level ensured reliable generalization assessment.

The first phase of the study examined the impact of downsampling the PPG signals to 25 Hz by comparing classification performance against models trained with the original 125 Hz signals. Classification performance was evaluated under two binary classification schemes: (1) NTS vs. PHT and HTS combined, and (2) NTS and PHT combined vs. HTS.

Based on initial results, the optimal sampling rate and classification labels were identified and used to refine the training process. Further analysis compared the performance of two optimizers: Adam [101] and SGDM [102], to determine the most effective hyperparameter configuration for model training.

4.4 The Relevance of Calibration in Hypertension Risk Assessment

It has been demonstrated that the relationship between the PPG-based propagation parameters and blood pressure is influenced by multiple physiological factors, including arterial walls thickness and elasticity, age, gender, posture and risk factors of CVDs. Thus, calibration becomes essential when applying automated classification methods to estimate BP levels in new subjects [85]. Moreover, calibration before measurement is essential to adapt the algorithms to the variations on PPG waveforms, as they are easily corrupted by fluctuations in blood circulation state, affecting the connection between BP and peripheral pulses [86]. The following subsections describe the methodologies to develop a classification system for hypertension risk discrimination and to evaluate the need and relevance of per-subject calibration employing both ML and DL based classification models.

4.4.1 Machine Learning Approach

With the aim of improving the approach of machine learning based discrimination between healthy and hypertensive subjects analyzed in 4.3.1, a variation in the implementation was proposed, introducing a context that is closer to reality. To this end, the relevance of the calibration when classifying new subjects that were not included in the training phase of the classification model was studied.

To study the impact of calibration, both the number of subjects and segments per subject were increased, as distant segments of each subject were required. After labeling MIMIC recordings based on SBP values, it was observed that several subjects exhibited stable signal stretches with differing hypertension labels. A possible explanation for these fluctuations in SBP values was that all patients were in an ICU, potentially receiving treatment or medication that significantly affected their SBP levels. Additionally, some subjects had temporally distant segments with SBP values near the classification threshold, leading to inconsistent labels over time despite only minor SBP variations.

To mitigate the effects of these time-related inconsistencies, subjects with substantial SBP fluctuations—whose labels switched between NTS and HTS across different time points—were excluded. These subjects were not suitable for training a classification model aimed at assessing the risk of hypertension. As a result, subjects maintaining consistent labels throughout the recording period were selected. Finally, a total of 913 recordings from 69 subjects, 45 categorized as NTS and 24 as HTS, each lasting 120 seconds and with acceptable signal quality conditions, were selected from the MIMIC database.

To further refine the classification approach, two sequential calibration strategies were implemented. Each strategy incrementally incorporated additional features extracted from new signal segments into a base model. This base model

was initially trained using all segments of the subjects except the one being analyzed. Then, in each iteration, additional PPG-extracted features were integrated, allowing the model to progressively refine its predictions with newly available information.

The first strategy employed sequential validation using continuous segments, where training was conducted on initial segments, and validation was performed on immediately subsequent segments. This method assumes that successive segments will be highly similar, leading to high classification accuracy in validation.

The second strategy investigated the impact of temporal separation between calibration measurements. Different separation intervals were explored, ranging from less than an hour to several hours or even days. This approach aimed to assess how classification performance is affected by increasing temporal gaps, providing insights into the model's ability to generalize hypertension risk assessment over extended periods.

Furthermore, a feature selection stage was introduced after the obtention of the discrimination features from all data recordings. The aim was to select only those features, from the original 23 discriminating parameters, that presented relevant information for solving the classification problem optimally.

Feature Selection

Firstly, since all the features were continuous quantitative variables, it was necessary to carry out a normalization, since each one could take on a different range of values, and more weight would be given to the variables with higher values, not necessarily being more important. The normalization was carried out using "zscore" centering the variables so that they had zero mean and scaling so that they had unit standard deviation, as represented in the following equation:

$$z = \frac{x - \overline{X}}{S}, \tag{4.1}$$

where x is a concrete value of a given feature, $\overline{X}$ is the mean of all values of that feature and S the standard deviation.

Once the variables were normalized, *ReliefF* algorithm was applied to rank predictors by importance, determining which ones had the best discriminatory power. The key idea of this method is to estimate the quality of predictors according to how well instances near to each other are distinguished, rewarding predictors that give different values to neighbors of a different class [106]. Furthermore, by means of positive and negative correlation, the independence between pairs of variables was analyzed in order to discard those that did not provide new information for the classification task. Figure 4.9 illustrates the matrix whose entries

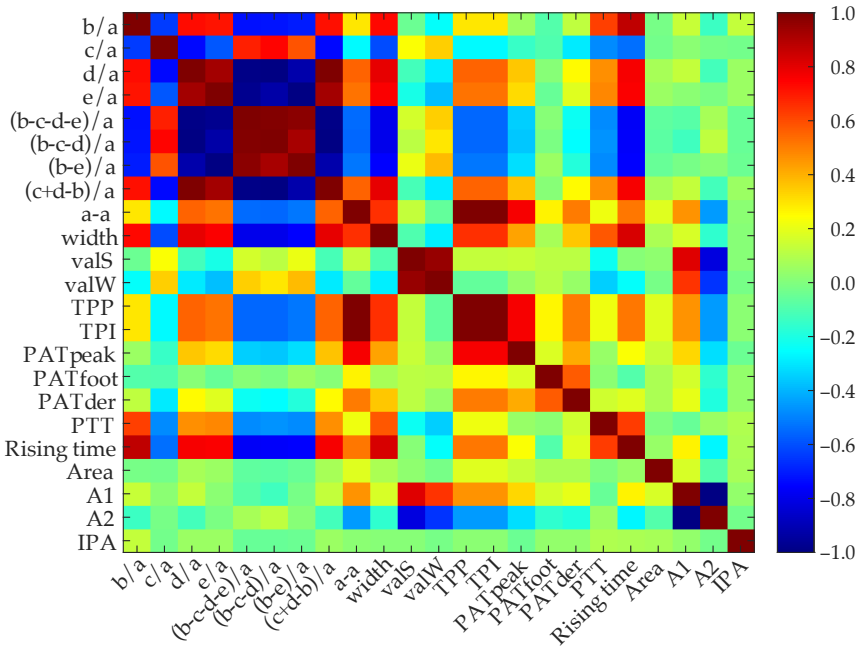


Figure 4.9: Correlation matrix of the 23 initial discriminatory features extracted from PPG and ECG signals. Dark red values represent higher correlation coefficients and dark blue values represent lower correlation coefficients.

are the correlation coefficients obtained by matching pairs of variables, so that highly correlated features can be discarded.

After analyzing the correlation matrix and *ReliefF* results, it was decided to remove three complex APG ratios $(b-c-d-e)/a$, $(b-e)/a$ and $(c+d-b)/a$ as they had high correlation coefficients with other features, as well as the last three *ReliefF* ranked features (TPP, TPI and pulse area), as the deletion of more features worsened classification performance. Finally, after the feature selection, 17 normalized features were obtained, which will be used as inputs to train the classification models with ML techniques.

Implementation Details

After the feature selection stage, the most discriminative features were extracted from the preprocessed PPG and ECG signals and used as inputs to train classification models based on ML techniques. To enable a fair and consistent comparison with the results presented in Section 4.3.1, the same set of classification algorithms was employed, encompassing both fundamental and alternative models to ensure methodological coherence.

In this study, calibration was defined as the inclusion of at least one previous measurement from the subject under analysis in the training set. Aimed at studying the importance of calibration in classifying new subjects as NTS or HTS, three different approaches were explored:

1. **Classification without prior subject-based calibration**. In this approach, the performance of new subjects without prior subject-based calibration was studied. For this purpose, the analyzed segments of the new subject were only used for validation, while the training set consisted only of segments from other subjects.

2. **Effectiveness of calibration for short-term classification**. This approach examined whether incorporating to training set prior measurements from the same subject, taken shortly before, could improve classification accuracy. The procedure followed these steps:

   a. Signal segments with a duration of 2 minutes were divided into 12 sub-segments, each lasting 10 seconds.

   b. The sub-segments of all subjects except the one to be analyzed and the first sub-segment of the analyzed subject, acting as the calibration measurement, were used as the training set and the next sub-segment of the same subject was used for validation.

   c. After its classification, this second sub-segment was introduced in the training set, using the next sub-segment for validation. This step was repeated until the 12 consecutive sub-segments of the subject were processed.

   d. This procedure was repeated for all 2-minute segments across all subjects in the database.

   By performing this sequential calibration and validation, we aimed to analyze how classification performance improved as the model incrementally incorporated subject-specific data taken in close temporal proximity.

3. **Effectiveness of calibration for long-term classification**. This approach investigated whether prior measurements from the same subject, but taken hours or even days apart, could enhance classification accuracy. To control temporal distance between measurements, groups of segments of the same patient that were less than 1h, between 1h and 6h, between 6h and 24h and more than one day apart were selected. A sequential validation process similar to the one used for short-term calibration was also followed in this approach. This allowed us to study whether classification results improved when the model incorporated previous measurements from the same patient taken at increasingly distant time points.

4.4.2 Deep Learning Approach

The DL approach was designed to evaluate the impact of calibration when classifying new subjects as NTS or HTS. In contrast to traditional ML methods, this strategy eliminates the need for manual feature extraction and does not rely on synchronized ECG signals, thereby streamlining the implementation process and enhancing its feasibility for real-world clinical applications.

In this study, a binary classification approach was adopted to discriminate between NTS and HTS subjects, with an SBP threshold of 130 mmHg. Given that prehypertensive patients (120 to 140 mmHg) with stable SBP values within this range are relatively uncommon, the primary objective of this study was to develop tools capable of detecting hypertension or its initial symptoms, i.e., when a patient's SBP exceeds 130 mmHg. Additionally, subjects displaying significant fluctuations in their SBP values between NTS and HTS measurements over time were excluded from the dataset.

Ultimately, a total of 974 recordings from 69 subjects were selected from the MIMIC database, with 45 subjects classified as NTS and 24 as HTS. PPG and ABP signals were recorded simultaneously, with a duration of 120 s and a common sampling frequency of 125 Hz. As discussed in 4.3.2, PPG signals were downsampled to 25 Hz to optimize data efficiency, reducing transmission and storage demands, an essential consideration for real-time processing in wearable health monitoring systems [105].

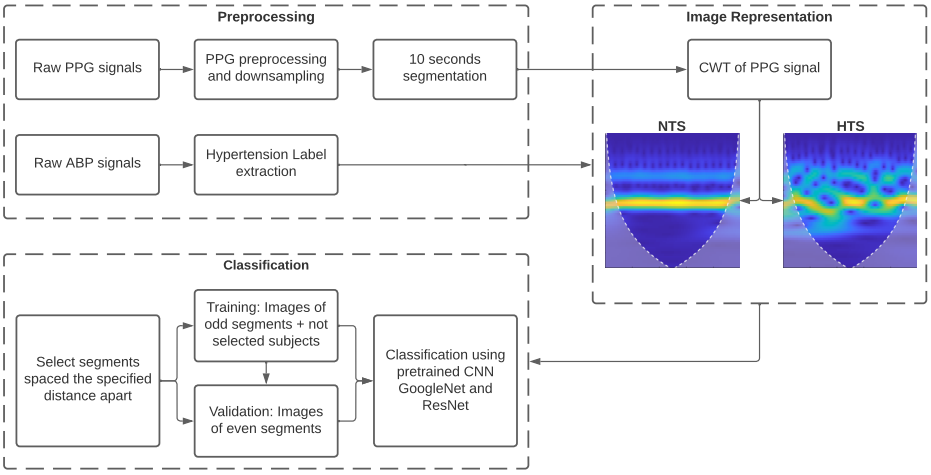


Figure 4.10: Block diagram illustrating the DL classification methodology for assessing the need of calibration in hypertension risk assessment. 10-second segments of downsampled PPG signals were processed using CWT and fed into pretrained DL models. Hypertension risk category was determined from the synchronized ABP signal. The resulting images were divided into training and validation sets to evaluate the impact of calibration across different time intervals.

Figure 4.10 presents a block diagram outlining the methodology developed to evaluate the role of calibration in hypertension classification using DL pretrained models. The process began with basic preprocessing of the raw PPG signals to enhance signal quality and remove noise. To address the need for large datasets in DL, the 120-second PPG and ABP recordings were divided into 10-second subsegments, resulting in a total of 11,688 samples.

Each segment was then transformed into a time–frequency image using the CWT, which captures both temporal and spectral characteristics of the signal. The absolute values of the CWT coefficients were used to generate scalograms, which were subsequently resized to 224×224 pixels and formatted as RGB images to meet the input requirements of the selected pretrained CNN models: ResNet-18, ResNet-50, and GoogLeNet.

Implementation Details

In the previous investigation of the importance of calibration in hypertension risk assessment using ML methods, experiments were conducted where calibration measurements were taken at various time intervals from the test measurements. In the calibration approach with closely spaced measurements, sequential validation was performed with the twelve sub-segments into which each 120-second segment was divided. This approach was not feasible with DL-based methods due to their high computational complexity and long training times. Consequently, the model would require more training time than the time difference between calibration and test measurements. Furthermore, it would not be practical to perform a calibration and predict hypertension risk only a few seconds later, as the image transformations of both measurements would be very similar.

In this study, we examined the effectiveness of calibration with different time intervals between PPG segments of different subjects. The intervals included less than 1 h, between 1 h and 6 h, between 6 h and 24 h, and more than 24 h apart. For this purpose, CNN classification models were trained following the procedure outlined below:

1. Selection of 120-s PPG segments obtained for each time interval between measurements.
2. The images of odd segments (1st segment, 3rd segment, etc., spaced at each interval) were used for training, and the images of even segments were used for validation. Additionally, images corresponding to PPG signals from subjects who did not have segments separated by the specified time interval were added to the training dataset. This increased the training dataset with subject segments that were not part of the validation dataset, enhancing the model's robustness. In this way, the calibration stage was simulated using training segments, with validation segments between two calibration measurements for classification.

3. These images were used as inputs for GoogLeNet, ResNet-18, and ResNet-50 pretrained CNN models, and the classification problem involved discriminating between normotensive and hypertensive subjects using PPG signal image representations.

Table 4.2 displays the number of images used in the training and validation tasks for each segment interval. For measurement intervals less than 24 h, the number of images from subjects with segments spaced at these intervals was higher than from unselected subjects. However, the opposite is observed when the interval between measurements exceeds 24 h, as the number of subjects with PPG recordings of more than 1 day is limited, and most of them do not belong to this group.

Table 4.2: Division of PPG signals' image representation in training and validation datasets for each measurement interval.

	Training Images		Validation Images
	Odd Segments	**Not Selected Subjects**	**Even Segments**
Below 1 h	2136	408	1752
Between 1 h and 6 h	3168	84	2784
Between 6 h and 24 h	1728	732	1356
Above 24 h	288	8868	192

4.5 Detection of Hypertensive Pathology

This section aims to investigate whether PPG recordings can be correctly classified even when an alteration in BP from the analyzed subject occurs. This approach emphasizes the detection of pathological hypertensive conditions rather than a direct classification based solely on BP levels.

To achieve this, a dataset was constructed from 51 subjects, grouped into two categories: 33 subjects with predominant BP values below 120 mmHg, and 18 subjects with predominant BP values over 140 mmHg, labeled as NTS and HTS, respectively. From this dataset, a total of 668 segments with coherent BP values were extracted, corresponding to high BP readings for HTS subjects and normal BP readings for NTS subjects. These segments represent the expected BP behavior for each group. Additionally, 135 incoherent BP segments were identified, where BP levels deviated from the typical group classification (normal BP readings for HTS subjects or high BP readings for NTS subjects). These incoherent segments represent recordings with alterations or unexpected BP behavior.

To enhance the robustness of the ML models, each recording was further divided into 12 sub-segments of 10 seconds each. This segmentation aimed to increase the size of the training and validation datasets while maintaining temporal consistency.

The implementation follows the standard ML procedure developed across the thesis, where discriminatory features were extracted from synchronized PPG, ECG, and ABP biosignals in order to train classification models. A total of 17 discriminative features were selected, identical to those utilized in the ML approach with calibration in Section 4.4.1.

The key distinction of this approach lies in the train-validation split strategy. Instead of randomly partitioning the dataset, discriminative features extracted from coherent segments were exclusively used for training the classification models, ensuring that the model learned from well-defined BP conditions. Conversely, discriminative features from incoherent segments were employed for validation. This approach simulated real-world scenarios where the model encounters unexpected or altered BP behavior, thus evaluating its ability to generalize beyond the conditions observed during training.

By focusing on the detection of pathological alterations rather than specific BP levels, this methodology highlights the potential of integrating PPG, ECG, and ABP signals for robust classification of cardiovascular conditions. This approach also provides a framework for analyzing unexpected BP behaviors, which could improve the early detection of hypertension-related pathologies and other abnormalities.

4.6 Statistical Analysis

Statistical tests for accuracy (*Acc*), sensitivity (*Se*), specificity (*Sp*), and *F1-Score* were employed for evaluating the classification performance. *Acc* represents the percentage of correctly classified PPG segments. *Se* and *Sp* denote the ability to detect HTS subjects as positive and the ability to detect NTS healthy subjects as negative, respectively. *F1-Score* is the harmonic mean of *Se* and *Acc*. These statistical tests were mathematically computed as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

$$Se = \frac{TP}{TP + FN} \tag{4.3}$$

$$Sp = \frac{TN}{TN + FP} \tag{4.4}$$

$$\textit{F1-Score} = \frac{2 \cdot Se \cdot Acc}{Se + Acc} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{4.5}$$

Here, TN represents the number of correctly classified NTS segments, TP the number of correctly classified HTS segments, FN the number of incorrectly predicted HTS segments, and FP the number of incorrectly predicted NTS segments.

Chapter 5

Results

In this chapter, the results of the studies described in Chapter 4 are presented. First, the results of hypertension risk assessment for ML and DL approaches (described in Section 4.3.1 and Section 4.3.2) are presented in Section 5.1. Next, the relevance of calibration for hypertension risk assessment results are presented in Section 5.2, for both ML and DL approaches described in 4.4.1 and 4.4.2. Finally, Section 5.3 presents the results of hypertensive pathology detection rather than absolute BP values, described in 4.5.

5.1 Hypertension Risk Assessment

This section presents the classification results obtained from both traditional ML and DL applied to hypertension risk assessment. Section 5.1.1 reports the performance of conventional ML classifiers using previously validated and newly proposed predictive parameters across different patient group comparisons. Section 5.1.2 focuses on the evaluation of DL architectures using different PPG signals, including training-validation and test results, as well as a comparison between optimizers and model configurations. The objective is to assess the effectiveness and generalization capabilities of each approach in distinguishing between normotensive, prehypertensive, and hypertensive individuals.

5.1.1 Machine Learning Approach

Since three different groups of patients were available, the four most logical paired comparisons were made (NTS vs PHT, NTS vs HTS, NTS + PHT vs HTS, NTS vs PHT + HTS). The statistical analysis commenced by employing the predic-

Table 5.1: Performance of four classification models with characteristic parameters strongly correlated with BP levels in a previous work.

Models	Comparison	Sensitivity	Specificity	F1-score
AdaBoost	NTS vs PHT	61.54%	74.29%	62.75%
	NTS vs HTS	80.00%	74.29%	74.07%
	NTS+PHT vs HTS	88.00%	81.97%	75.86%
	NTS vs PHT+HTS	76.47%	71.43%	78.00%
Logistic Regression	NTS vs PHT	57.69%	68.57%	57.69%
	NTS vs HTS	52.00%	77.14%	56.52%
	NTS+PHT vs HTS	36.00%	90.16%	45.00%
	NTS vs PHT+HTS	60.78%	34.29%	59.05%
KNN	NTS vs PHT	42.31%	80.00%	50.00%
	NTS vs HTS	40.00%	94.29%	54.05%
	NTS+PHT vs HTS	36.00%	100.00%	52.94%
	NTS vs PHT+HTS	62.75%	68.57%	68.09%
Bagged Trees	NTS vs PHT	46.15%	74.29%	51.06%
	NTS vs HTS	76.00%	82.86%	76.00%
	NTS+PHT vs HTS	52.00%	90.16%	59.09%
	NTS vs PHt+HTS	72.55%	57.14%	71.84%

tive parameters that have demonstrated the highest correlation with BP levels in previous studies [57], with their respective performance metrics summarized in Table 5.1. This initial evaluation aimed to assess the effectiveness of conventional classification models when utilizing well-established predictive features.

AdaBoost exhibited competitive classification performance, particularly in the NTS vs. PHT + HTS scenario, achieving an F1-score of 78%. Bagged Trees performed similarly or better in several comparisons, highlighting that no single model consistently outperformed the others across all groupings. However, the remaining models struggled to achieve a balanced classification. A noticeable trend among these models was a strong tendency to classify a high proportion of individuals as NTS, irrespective of their actual health condition. This issue is particularly evident in the imbalance between sensitivity and specificity, as many models favored specificity at the expense of sensitivity, leading to a significant misclassification of hypertensive individuals. Such behavior underscores the limitations of these models in accurately identifying patients at risk when relying solely on traditional predictive parameters.

In contrast, Table 5.2 presents the classification results using the 23 newly proposed predictive parameters, tailored to better capture the underlying physiolog-

Table 5.2: Performance of the four alternative classification models using the 23 new characteristic parameters.

Models	Comparison	Sensitivity	Specificity	F1-score
Naïve Bayes	NTS vs PHT	57.69%	91.43%	68.18%
	NTS vs HTS	64.00%	85.71%	69.57%
	NTS+PHT vs HTS	64.00%	91.80%	69.57%
	NTS vs PHT+HTS	82.35%	82.86%	84.85%
SVM Cubic	NTS vs PHT	57.69%	74.29%	60.00%
	NTS vs HTS	68.00%	82.86%	70.83%
	NTS+PHT vs HTS	64.00%	91.80%	69.57%
	NTS vs PHT+HTS	54.90%	65.71%	61.54%
SVM Quadratic	NTS vs PHT	61.54%	91.43%	71.11%
	NTS vs HT	60.00%	88.57%	68.18%
	NTS+PHT vs HTS	68.00%	93.44%	73.91%
	NTS vs PHT+HTS	76.47%	80.00%	80.41%
Coarse Tree	NTS vs PHT	50.00%	80.00%	56.52%
	NTS vs HTS	40.00%	80.00%	47.62%
	NTS+PHT vs HT	48.00%	77.05%	47.06%
	NTS vs PHT+HT	86.27%	77.14%	85.44%

ical patterns associated with BP fluctuations, along with alternative classification models specifically designed to enhance discrimination between BP categories. Notably, both the Naïve Bayes and Coarse Tree models demonstrated substantial improvements compared to previous results. In particular, the Coarse Tree model achieved an F1-score of 85.44%, while Naïve Bayes attained its best result with 84.85%, both in the NTS vs. PHT + HTS comparison. These results underscore the benefit of using the newly proposed predictive parameters, even if only one model surpassed the 84% threshold.

5.1.2 Deep Learning Approach

The results of the deep learning approach, as detailed in 4.3.2, are summarized below. Tables 5.3 and 5.4 present the classification efficiency when the dataset was randomly divided into training and validation with the original sampling rate of 125 Hz and downsampled to 25 Hz, respectively.

Table 5.3: Training-Validation performance from PPG signals with 125 Hz sampling rate.

Models	Comparison	Sensitivity	Specificity	F1-score
GoogleNet	NTS vs PHS+HTS	95.44%	83.08%	92.10%
	NTS+PHT vs HTS	84.45%	90.99%	84.16%
ResNet	NTS vs PHT+HTS	94.37%	90.52%	93.91%
	NTS+PHT vs HTS	87.67%	93.87%	88.23%

Table 5.4: Training-Validation performance from PPG signals with 25 Hz sampling rate.

Models	Comparison	Sensitivity	Specificity	F1-score
GoogleNet	NTS vs PHS+HTS	94.00%	83.83%	91.54%
	NTS+PHT vs HTS	82.49%	91.87%	83.64%
ResNet	NTS vs PHT+HTS	94.56%	89.29%	93.58%
	NTS+PHT vs HTS	88.94%	92.83%	88.08%

Overall, classification models performed better when distinguishing NTS from PHT and HTS combined, achieving F1 scores above 91.5% compared to NTS and PHT combined vs. HTS, which resulted in F1 scores below 88.5%.

Tables 5.5 and 5.6 present test performance metrics using previously trained models on a test dataset composed of PPG signals from new patients. Despite promising training-validation results, test performance showed a notable decrease, particularly in the NTS + PHT vs. HTS classification task, where F1 scores dropped to around 10%, indicating a tendency to classify most new records as negative.

Table 5.5: Test results using GoogleNet and ResNet models from 125 Hz sampling rate PPG signals.

Models	Comparison	Sensitivity	Specificity	F1-score
GoogleNet	NTS vs PHS+HTS	68.13%	48.03%	65.95%
	NTS+PHT vs HTS	9.05%	87.67%	10.07%
ResNet	NTS vs PHT+HTS	69.67%	49.58%	67.31%
	NTS+PHT vs HTS	16.11%	83.70%	15.38%

Table 5.6: Test results using GoogleNet and ResNet models from 25 Hz sampling rate PPG signals.

Models	Comparison	Sensitivity	Specificity	F1-score
GoogleNet	NTS vs PHT+HTS	59.70%	51.16%	60.97%
	NTS+PHT vs HTS	4.61%	89.47%	5.60%
ResNet	NTS vs PHT+HTS	67.75%	45.68%	65.17%
	NTS+PHT vs HTS	20.18%	82.33%	18.29%

After determining the optimal sampling rate (25 Hz) and classification labels (NTS vs. PHT + HTS), the second phase of analysis was conducted.

Table 5.7: Training performance comparison of GoogleNet, ResNet-18 and ResNet-50 models with SGDM and Adam optimizers in the first and last epoch.

Model	Optimizer	Epoch	Train Loss	Valid Loss	Valid Acc
GoogLeNet	SGDM	1	1.0579	0.8402	57.49%
		25	0.2534	0.3594	89.84%
	Adam	1	1.6798	2.8843	58.72%
		11	0.0850	0.2191	93.37%
ResNet-18	SGDM	1	0.8830	0.8403	48.08%
		25	0.0331	0.2066	92.96%
	Adam	1	0.9322	0.7584	61.06%
		10	0.0146	0.2155	94.84%
ResNet-50	SGDM	1	0.9429	0.8619	58.76%
		22	0.0147	0.2174	92.96%
	Adam	1	0.8279	1.0282	54.42%
		8	0.0398	0.2131	93.98%

Table 5.7 summarizes the training performance of GoogleNet, ResNet-18, and ResNet-50, including training loss, validation loss, and validation accuracy at the first and last training epochs. Although it was specified in the training options that the maximum number of epochs was 25, models trained with the Adam optimizer required fewer epochs (8-11) compared to those trained with SGDM, demonstrating the effectiveness of early stopping in preventing overfitting.

Table 5.8 shows the classification efficiency when the dataset is randomly divided in training and validation in terms of sensitivity, specificity, F1-score, and AUC. The highest classification performance was achieved using ResNet-18 with the Adam optimizer, yielding an F1-score of 95.61%, sensitivity of 95.68%, specificity of 93.65%, and AUC of 98.77%. However, all models demonstrated high performance, with F1 scores exceeding 91%.

Table 5.8: Classification results comparison of GoogleNet, ResNet-18 and ResNet-50 models with SGDM and Adam optimizers.

Model	Optimizer	Sensitivity	Specificity	F1-score	AUC (%)
GoogLeNet	SGDM	89.38%	90.17%	91.25%	95.88%
	Adam	94.42%	91.87%	94.36%	97.51%
ResNet-18	SGDM	94.70%	90.48%	94.04%	97.72%
	Adam	95.68%	93.65%	95.61%	98.77%
ResNet-50	SGDM	94.56%	90.67%	94.04%	97.71%
	Adam	94.56%	93.15%	94.86%	98.40%

Table 5.9 shows models performance when tested on a subset representing 20% of the data. A significant reduction in classification performance was observed, with the highest F1-score (68.51%) obtained using ResNet-18 with the SGDM optimizer, alongside a sensitivity of 70.83%, specificity of 51.39%, and AUC of 62.36%.

This noticeable decline in performance highlights a critical limitation of non-calibrated DL models when applied to unseen subjects. One of the main contributing factors is the high inter-subject variability in PPG morphology, which is influenced by individual physiological differences such as arterial stiffness, vascular tone, age, gender, and the presence of cardiovascular risk factors. Since the training data did not include any samples from the test subjects, the model could not effectively generalize to their specific signal characteristics. This situation underscores the importance of calibration procedures, which can adapt the model to account for subject-specific traits, as represented in the next section.

Table 5.9: Test results comparison of GoogleNet, ResNet-18 and ResNet-50 proposed models

Model	Optimizer	Sensitivity	Specificity	F1-score	AUC (%)
GoogLeNet	SGDM	48.97%	60.34%	54.93%	57.04%
	Adam	62.84%	41.13%	60.90%	54.98%
ResNet-18	SGDM	70.83%	51.39%	68.51%	62.36%
	Adam	59.59%	55.86%	62.00%	61.45%
ResNet-50	SGDM	69.92%	44.06%	66.18%	57.56%
	Adam	59.25%	55.86%	61.75%	59.67%

5.2 Relevance of Calibration

This section evaluates the impact of subject-specific calibration on the performance of hypertension risk assessment models. Section 5.2.1 explores ML strategies to assess the classification improvement obtained through short- and long-term calibration. The results, based on the KNN classifier, demonstrate that incorporating segments from the same subject, even when distant in time, significantly enhances classification accuracy. Section 5.2.2 extends this analysis to DL models. The classification performance of CNN-based architectures is reported for various calibration intervals, showing a clear performance degradation as the temporal distance between calibration and test increases, while confirming the effectiveness of early stopping to prevent overfitting.

5.2.1 Machine Learning Approach

As described in Section 4.4.1, three different approaches were explored to investigate the relevance of calibration for the hypertension risk assessment of new subjects combining PPG and ECG signals with ML strategies. KNN model was chosen since it obtained the most significant improvements throughout the approaches carried out in this section.

1. **Classification without prior subject-based calibration**.

   Firstly, the segments were classified without any previous calibration, in other words, with the training dataset only consisting of segments from other subjects, with each analyzed segment from the subject under study being tested for validation. In this case, classification accuracy with no previous calibration was 51.48%. This proved that the hypertension risk assessment of subjects without a prior calibration provided low accuracy results, as also reported in previous studies [85, 86].

2. **Effectiveness of calibration for short-term classification**.

   Next, applying the second approach aimed at demonstrating whether poor classification results could be improved with calibration, Figure 5.1 shows the classification accuracy between NTS and HTS individuals, in the form of box-and-whisker plots, employing sequential validation of consecutive sub-segments. Moreover, the Figure indicates in each square the mean accuracy for all selected subjects according to the number of consecutive sub-segments in the training dataset acting as subject calibration.

   It can be seen that, with the sole incorporation of one prior close in time sub-segment in the training dataset for calibration, the classification performance increased by 30% with respect to the case without calibration. Furthermore, accuracy improved progressively until it was stabilized above 96%, when more than six prior and close in time sub-segments from the same subject were present in the training dataset.

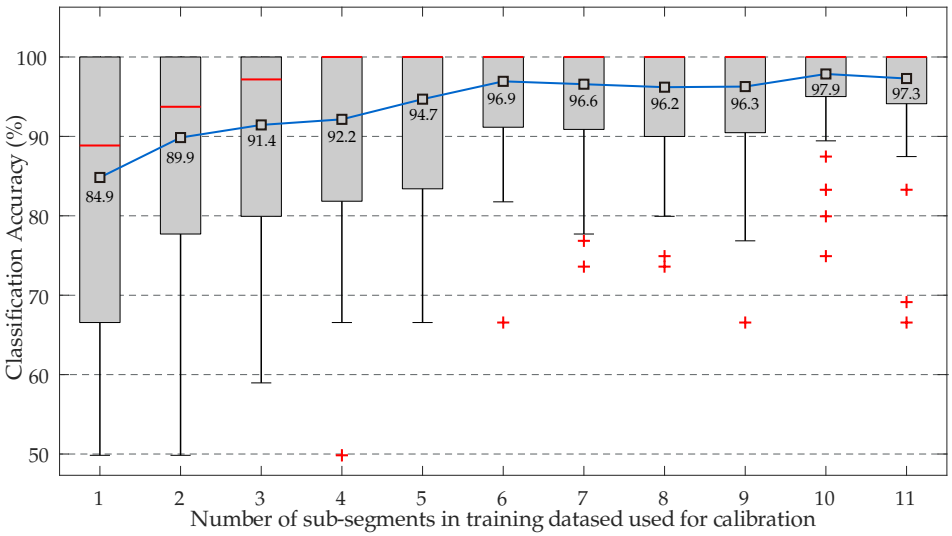


Figure 5.1: Classification performance provided by KNN classifier discriminating between NTS and HTS individuals. Results obtained for sequential validation of consecutive sub-segments for each of the selected subject segments. In each box, the red line indicates the median, and the bottom and top edges indicate the 15th and 85th percentiles, respectively. The whiskers cover the most extreme data points not considered outliers, and the red symbol (+) stands for outliers. Black squares inside each box indicate mean accuracies.

3. **Effectiveness of calibration for long-term classification**.

   Finally, for the third calibration strategy described in Section 4.4.1, which involves performing calibration at a temporal distance from the measurement segments, Figure 5.2 presents the classification results obtained through sequential validation with varying time gaps between the calibration and test segments.

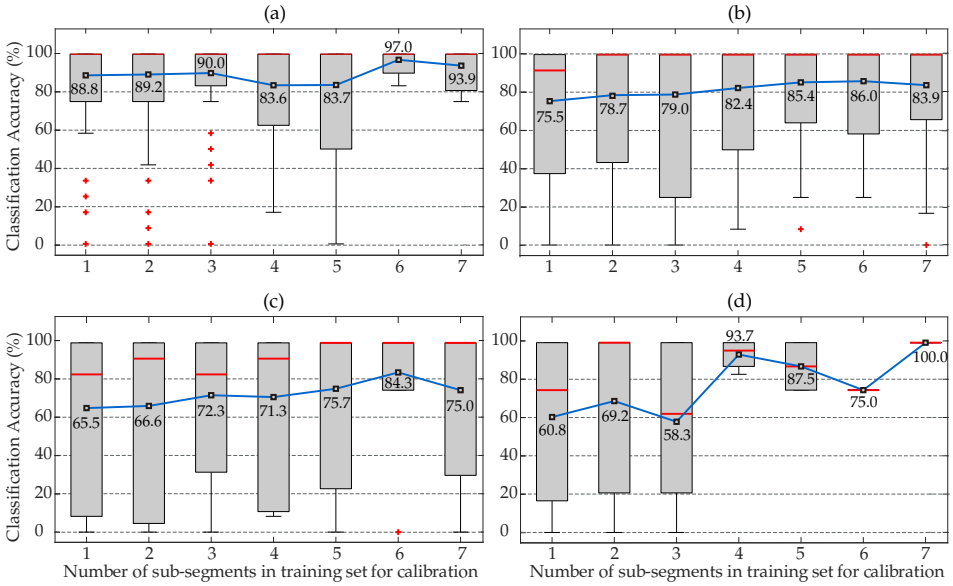


Figure 5.2: Classification accuracy to distinguish between NTS and HTS individuals using KNN sequential validation of segments with calibration distant from test measurements. (**a**) Distance below 1 hour. (**b**) Distance between 1 and 6 hours. (**c**) Distance between 6 and 24 h. (**d**) Distance above 24 h. In each box, the red line indicates the median, and the bottom and top edges indicate the 15th and 85th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the red symbol (+) stands for outliers. Black squares inside each box indicate mean accuracies.

It was demonstrated that calibration improved the classification task discriminating between NTS and HTS subjects because, as the number of measurements of the same subject in the model increased, so did the accuracy rate. It also shows that with calibration and test separated by less than 1 hour, the model was able to classify with an accuracy beyond 94% from the sixth calibration measurement onwards. As expected, these outcomes decreased as the distance between calibrations and test measurement increased, thus requiring up to five calibration measurements with distances between 6 h and 24 h to obtain classification accuracies above 75%.

In any case, the need to perform several calibration measurements to achieve adequate classification accuracy with test measurements, which could be many hours or even days away from calibration, does not seem to be a serious limitation. On the other hand, it is worth mentioning that only five subjects had recording lengths longer than five days, so that less stable results in Figure 5.2 (d) can be considered as normal, because any misclassification would significantly affect the final accuracy. Although the number of subjects was not quite elevated here, the obtained results demonstrate that classification performance was good enough even with distances between calibration and test measurement of several days, which is very promising for real-world applications.

5.2.2 Deep Learning Approach

As outlined in Section 4.4.2, this section evaluates the effectiveness of calibration in the context of DL models trained on CWT-based representations of PPG signals. Specifically, the analysis focuses on the impact of varying the temporal interval between calibration and test segments, comparing short intervals (less than 1 hour) with long intervals (greater than 24 hours), to assess the robustness and generalization capacity of the proposed approach.

The training performance, in terms of the number of epochs after early stopping, training time, and statistical results of classification performance achieved by the proposed CNNs for discriminating between NTS and HTS segments, is presented in Table 5.10. Although the maximum number of epochs was set to 20, no model reached this limit, indicating the effectiveness of the early stopping technique in preventing overtraining.

All models achieved their highest performance when calibration and measurements were separated by less than 1 h, resulting in accuracy exceeding 90%. Among them, ResNet-18 exhibited the best classification results, achieving an accuracy of 93.32%, sensitivity of 84.09%, specificity of 97.30%, and an F1-score

Table 5.10: Classification performance for distinguishing between NTS and HTS individuals using GoogLeNet, ResNet-18, and ResNet-50 pretrained CNNs with calibration distances below 1 h, between 1 and 6 h, between 6 and 24 h, and above 24 h.

Model	Epochs	Train Time	Acc	SE	SP	F1-Score
Below 1 h						
GoogLeNet	13	64 min	90.18%	82.77%	93.38%	83.56%
ResNet-18	6	28 min	93.32%	84.09%	97.30%	88.36%
ResNet-50	8	115 min	92.52%	84.09%	96.16%	87.14%
Between 1 h and 6 h						
GoogLeNet	9	61 min	82.87%	73.12%	88.40%	75.55%
ResNet-18	15	107 min	89.04%	81.15%	93.52%	84.29%
ResNet-50	8	152 min	88.11%	80.16%	92.62%	83.00%
Between 6 h and 24 h						
GoogLeNet	14	66 min	79.57%	70.57%	85.98%	74.18%
ResNet-18	6	29 min	80.16%	73.23%	85.10%	75.43%
ResNet-50	19	99 min	80.68%	76.60%	83.59%	76.73%
Above 24 h						
GoogLeNet	7	57 min	73.37%	38.73%	93.68%	51.80%
ResNet-18	7	32 min	79.17%	68.14%	85.63%	70.74%
ResNet-50	19	221 min	71.56%	56.86%	80.17%	59.64%

of 88.36%. Furthermore, this model demonstrated efficiency, requiring only 6 epochs and a training time of 28 minutes. As expected, the classification performance decreased as the time interval between calibration and test measurements increased, with accuracy falling below 81% for distances above 6 h.

5.3 Detection of Hypertensive Pathology

Finally, the effectiveness of the studied machine learning models based on PPG signals for hypertension risk classification, even in altered blood pressure conditions, is studied.

Table 5.11 presents the statistical results of this study. It can be seen that the classification model that provided the best outcomes was KNN, with F1-Score of 88.30%. Cubic SVM and Bagging Ensemble classifiers worsened the results in the four statistical tests studied. High specificity results in the three classification models reflected correctly identified normotensive individuals when BP reached prehipertensive or hypertensive values.

Table 5.11: Performance of the three classification models analyzed to screen hypertension under Non-Hypertensive BP recordings and normotension under hypertensive BP recordings.

Model	Accuracy	Sensitivity	Specificity	F1-score
KNN	83.70%	79.81%	96.77%	88.30%
SVM	75.56%	72.12%	87.10%	81.97%
Ensemble	79.26%	75.00%	93.55%	84.78%

Chapter 6

Discussion

In this chapter, the main findings obtained from the doctoral research are critically examined and interpreted within the context of current knowledge and existing literature. In Section 6.1, the results achieved in this study are compared with those reported in previous works, emphasizing the unique strengths and novel aspects associated with the methodological approaches proposed here.

Subsequently, in Section 6.2, an in-depth evaluation of the performance of these approaches is presented. Initially, clinical implications and the significance of the models in hypertension risk assessment are discussed, highlighting the importance of early detection of pre-hypertensive states. Additionally, the relevance and practical feasibility of the calibration strategies are examined, emphasizing how personalized calibration can significantly enhance predictive accuracy in real-world applications. Finally, the capability of the developed models to detect hypertensive pathology independently from absolute BP values is explored, underlining their potential clinical utility beyond mere BP classification.

Following, the limitations inherent to the study are discussed in Section 6.3, addressing both methodological constraints and practical challenges that could influence clinical adoption. Finally, in Section 6.4, the clinical and technical implications of the presented work are synthesized. These directions aim to further refine and validate the approaches proposed, facilitating their practical deployment and enhancing their contribution to continuous, non-invasive BP monitoring in both clinical and home-care settings.

6.1 Comparison with previous works

The continuous measurement of BP is of great importance as it facilitates the early detection and prevention of hypertension, being the main risk factor for many cardiovascular diseases. In recent years, the rapid emergence of the Internet of Things [107] and the development of cuff-less devices capable of continuously measuring and processing physiological signals using AI techniques, such as ML and DL, have led to the proposal of alternative methods to traditional cuff-based single-time BP measurement. The main signal used in related studies has been the PPG, as its morphological variations are related to the heart's activity and vascular walls condition, being similar to BP morphology both in frequency and time domains [10]. Furthermore, the PPG signal can be acquired by non-invasive low-cost devices as smart watches, obtaining a continuous and real-time measurement.

BP monitoring through PPG has mainly been studied by two different approaches: addressing the problem of monitoring BP as a regression task, estimating systolic and diastolic values; and addressing the problem of detecting hypertensive subjects as a classification task. In this study, the second approach has been developed, as estimations from the first approach still have serious limitations, so that it is more clinically beneficial to alert hypertensive subjects, acting as a support for clinical decision making.

Tjahjadi et al. [86] proposed the use of KNN technique and PPG signal without ECG, requiring the extraction of 2100 PPG feature points from 2.1 seconds of data. Their classification results achieved an F1-score of 100% for NTS and PHT patients and 90.80% for HTS patients. Although the authors affirm that this method achieved higher classification performance than other ML and DL methods, obtaining 2100 PPG feature points in such a short period of time required a sampling frequency of 1 kHz, which is a serious drawback for embedding this method in wearable devices, as it significantly increases the amount of data sampled, saved and transmitted which, unavoidably, will involve a considerably high power consumption.

Most studies for hypertension risk classification use both PPG and ECG signals, as PAT value is directly related to BP value. Previous works have studied the efficiency of employing PAT as the only parameter to estimate BP [58, 108], or combined with other propagation parameters as PTT or PWD [11]. However, Liang et al. [57] reported a higher correlation with BP levels by combining PAT with additional PPG features. Dividing the dataset into 70% for training and 30% for validation, the KNN classification model obtained the best performance compared to bagged tree, logistic regression and AdaBoost tree. The F1 scores comparing NTS vs. PHT, NTS vs. HTS and NTS + PHT vs. HTS were 84.34%, 94.84% and 88.49%, respectively. As a consequence, the hypertension risk discrimination performance between NTS and HTS was similar to the one achieved in the present study, where F1 score with the KNN classification model was 91.93% employing a leave-one-out cross-validation strategy.

Therefore, both in previous studies and in the present work, the KNN classifier has been the best model to assess hypertension risk, combining PPG recordings and ML techniques. However, until now, there is no agreement about the discriminant features to be used, since it depends considerably on patient selection and database, mode of acquisition and signal quality [12].

Building on the initial hypertension risk assessment, an important aspect to consider is the role of calibration in improving classification accuracy. Calibration has been identified in previous studies as a key factor that significantly enhances hypertension risk assessment performance. However, calibration has primarily been explored in BP estimation studies that integrate PPG signals with additional patient-specific data such as age [109], arterial characteristics (e.g., distance and area between measurement sites), or physiological factors like exercise and posture changes [108].

Recently, Schlesinger et al. [18] used convolutional neural networks and PPG signals for BP estimation, achieving a reduction in mean absolute difference of 2.54 mmHg after calibration, using a single 30-second window of PPG signal and the associated BP reading. Mukkamala et al. [110] studied the age factor in calibration, predicting a maximum calibration interval of 1 year for subjects of 30 years of age, which declined linearly to 6 months for subjects at the age of 70, using the PTT as a discriminatory feature.

In recent years, DL approaches have obtained outstanding performance extracting information from image representations [111, 112]. Prior research in this field has explored the application of DL models for hypertension risk stratification by employing image transforms of PPG signals, such as the Hilbert–Huang Transform [113] and CWT [114]. However, these studies did not adhere to a rigorous classification procedure, as signal segmentation was performed prior to dividing the dataset into training and validation sets. This resulted in consecutive signal segments being present in both subsets, potentially causing an overestimation of classification performance. For example, Liang et al. [17] employed CWT-transformed PPG signals and CNNs for BP classification, dividing the dataset in 80% for training and 20% for testing. The F1 scores for the binary classification comparing NTS vs. PHT were 80.52%, NTS vs. HTS were 92.55% and NTS + PHT vs. HTS trials were 82.95%.

The alternative calibration method proposed in this study significantly enhances these strategies, providing a classification approach that aligns with the decision-making process of a model that can be easily integrated into wearable devices. The main disadvantages of DL approaches are the requirement of a high computational cost, the extra duration of the training stage and the need for a large number of recordings.

An important advancement of this research compared to previous studies is the integration of DL approaches combined explicitly with tailored calibration strategies and advanced feature extraction methodologies. Unlike earlier inves-

tigations, where regression-based methods predominated, the present study prioritized classification strategies due to their greater direct clinical applicability.

Another relevant contribution of this work, not explicitly addressed in related studies, is the study of subjects with stable BP labels. Although BP levels naturally fluctuate throughout the day due to individual activities and various physiological or environmental factors, each subject must ultimately be assigned a single, consistent diagnostic category. For instance, it would not be clinically coherent for a subject to be classified as HTS at certain times, PHT at others, and NTS at yet other moments. This issue is particularly relevant when using databases such as MIMIC, which contain recordings from ICU patients who, due to their unstable condition or the administration of medication, may exhibit highly variable and non-representative systolic BP values.

In addition, most conventional devices provide punctual BP measurements, while artificial intelligence methods typically estimate continuous BP values [26]. Such approaches may not adequately capture BP variations induced by physical or mental stimuli, which can lead to inaccurate diagnoses. To address this limitation, the present study investigates whether an adequate subject diagnosis is provided under altered BP conditions by combining ML models with features extracted from ECG and PPG signals

Some studies have employed PPG signals to analyze BP fluctuations. For instance, Radha et al. [115] estimated nocturnal BP dips by combining PPG sensors with deep neural network architectures, including dense networks, random forests, and linear regression models. Similarly, Finnegan et al. [116] investigated the presence of circadian rhythm in PAT and their applications to classify nocturnal BP dips or rises, an important indicator for cardiovascular diseases. However, prior studies on hypertension risk assessment have predominantly focused on classification based on stable BP labels rather than on detection during fluctuating BP states.

6.2 Performance of studied approaches

This section evaluates in detail the performance of the proposed ML and DL methods, emphasizing three critical aspects: the clinical implications of hypertension risk assessment, the effectiveness of calibration strategies, and the detection of hypertensive pathology independent of absolute BP values. These three dimensions were specifically chosen to provide a comprehensive analysis of the methodologies' robustness, clinical relevance, and real-world applicability, thus offering a clearer picture of their potential impact in practical healthcare settings.

6.2.1 Hypertension Risk Assessment

Early identification of pre-hypertensive states has significant clinical implications, as timely detection allows for prompt preventive interventions, potentially reducing the incidence of cardiovascular complications. The models presented in this study effectively detect subtle physiological changes associated with emerging hypertension, demonstrating their potential to act as an early warning system. Clearly articulating these clinical implications, this study provides strong support for integrating the proposed methodologies into routine patient monitoring systems, particularly in primary care or remote monitoring environments, where early intervention strategies can dramatically improve patient outcomes.

Machine Learning Models

The development of ML models for hypertension risk assessment from PPG signals faces several ongoing challenges: variability in acquisition protocols, lack of standardization in feature extraction and model selection, and heterogeneity in patient populations. For these reasons, this study offers two key contributions: the proposal of novel physiologically inspired features, and a systematic evaluation of their impact on classification performance across clinically relevant blood pressure categories.

One of the most relevant findings of this work is that conventional parameters, even when strongly correlated with blood pressure levels, do not necessarily translate into effective predictors for classification tasks. The newly proposed features in this study were specifically designed to enhance discriminatory power, reflecting vascular stiffness, autonomic function, or pulse propagation alterations. The fact that models trained on these features showed a more balanced sensitivity-specificity profile suggests that they are better suited for risk stratification tasks. This has important implications for early detection and screening strategies, particularly in primary care or wearable health-monitoring contexts.

Moreover, the binary discrimination of normotensive versus “at-risk” group (combining prehypertensive and hypertensive patients) obtained the most robust results. Prehypertension is not a benign intermediate state, but rather a dynamic phase often leading to sustained hypertension. Importantly, elevated blood pressure in its early stages typically presents no symptoms, which contributes to delayed diagnosis and missed intervention. Therefore, identifying prehypertensive individuals as part of an at-risk group enhances the potential for early detection and preventive action. Thus, this grouping is not only clinically justified but also operationally advantageous, as both subgroups may benefit from closer follow-up or further diagnostic evaluation. Moreover, this strategy avoids ambiguous classifications in the “grey zone” of BP values, which often suffer from poor reproducibility in both signal quality and diagnostic labeling.

The performance improvements observed with certain ML models demonstrate that, when combined with well-designed and physiologically meaningful features, relatively simple and interpretable classifiers can achieve promising results. However, these approaches may have some limitations when it comes to capturing subtle or complex patterns present in the PPG waveform, particularly in more diverse or noisy datasets. This raises the question of whether DL models, with their ability to extract features directly from raw or minimally processed signals, might offer further advantages. The next section explores this possibility, evaluating whether these architectures can improve classification performance and generalizability in the context of hypertension risk assessment.

Deep Learning Models

The application of DL techniques to classify hypertension risk from PPG signals offers an alternative to traditional ML approaches, particularly when the goal is to develop wearable, real-time, and non-invasive monitoring solutions. Unlike classical ML methods, which typically require handcrafted feature extraction and additional signals such as ECG to compute pulse arrival time, DL models can learn hierarchical representations directly from raw or preprocessed PPG signals.

The results obtained in this work support the feasibility and effectiveness of CNN models for this task. High validation F1-scores were achieved for both pretrained architectures (GoogleNet and ResNet) and sampling rates (125 Hz and 25 Hz). These results confirm the potential of deep models to extract robust features from time-frequency representations of PPG signals, such as CWT. In fact, the comparable performance between 125 Hz and 25 Hz sampling suggests that lower-resolution inputs are sufficient for accurate classification, an important finding for future implementation in wearable devices.

Consistently, the classification task distinguishing normotensive individuals from prehypertensive and hypertensive, obtained the highest performance. This aligns with previous observations in the ML-based analysis and strengthens the hypothesis that prehypertensive subjects already exhibit waveform alterations similar to those of hypertensive individuals. Thus, the ability of CNNs to detect such patterns could be valuable for timely intervention in at-risk populations, particularly relevant given the silent nature of early-stage hypertension.

Nonetheless, the generalization capacity of the DL models proved to be a limiting factor. While training and validation metrics were strong, performance dropped significantly when tested on unseen subjects. F1-scores on the test set dropped to around 65%, with particularly poor results in the NTS + PHT vs. HTS classification task. This suggests that PPG-BP dynamics are highly subject-specific and trained models may have difficulty generalizing across individuals without incorporating subject-specific calibration. The same PPG waveform morphology does not necessarily correspond to the same BP level in different subjects [117], making it difficult for models to identify consistent patterns. In prac-

tice, this implies that accurate classification may require some level of subject-specific information to be present across training, validation, and test phases, highlighting a fundamental challenge in achieving fully generalized, cuff-less hypertension screening from PPG signals.

Furthermore, while all architectures tested achieved strong results, ResNet-18 combined with the Adam optimizer consistently outperformed other configurations in terms of accuracy, sensitivity, specificity, and training efficiency. The use of early stopping allowed for significantly shorter training cycles, reducing computational costs and preventing overfitting, an essential consideration for practical deployment.

The use of downsampled signals, the inclusion of prehypertensive subjects in the "diseased" group, and the detailed comparison of optimizers collectively represent key contributions of this work, outlining a practical pathway toward the deployment of DL based hypertension screening tools in wearable devices. However, the performance gap observed on unseen patients highlights the need for strategies that improve individual-level generalizability. In this context, incorporating calibration mechanisms appears essential to avoid the gap between research and clinical applicability. These perspectives will be further explored in the next section, where the role of calibration in improving model robustness across individuals is discussed in depth.

6.2.2 Relevance of Calibration

Practical implementation of personalized calibration is critical for sustained accuracy of predictive models. To address previous ambiguities regarding operational feasibility, this study proposes clear recommendations for the frequency and duration of recalibration needed in clinical settings. A brief recalibration period using some reference measurements could maintain predictive performance over extended periods, ensuring clinical relevance without imposing excessive burden on healthcare providers or patients.

Machine Learning Models

Building on the challenge of poor generalization in subject-independent scenarios, this study explored calibration strategies to enhance hypertension risk classification using traditional ML models, where the accuracy was 51.48%. Two complementary approaches were examined, the first investigated sequential validation, assuming high similarity between consecutive signal segments from the same individual. Results showed that incorporating even a single calibration sample led to a substantial improvement in classification accuracy—by more than 30%, which was enhanced even more as the number of calibration measure-

ments raised. This highlights the strong intra-subject consistency of PPG-derived features over short time intervals.

The second approach evaluated calibration efficacy across varying time gaps, from less than one hour to over 24 hours. This way, it was considered if PPG signal properties from the same patient were kept across time or changed along the day or week. Even with only one calibration measurement, accuracy increased by approximately 30% for intervals under one hour. The improvement of hypertension risk classification decreased slightly as the distance between calibration and measurement increased, although the use of calibration always improved classification results compared to classifying a new uncalibrated subject. Thus, after the fifth calibration, all the experiments provided high accuracy. These findings reinforce the temporal variability of PPG signals and underscore the importance of periodic recalibration to maintain reliable classification over time.

These approaches have demonstrated that the properties of each patient's PPG features were variable over time, as worse results were obtained with measurements distant from calibration than with those very close in time to calibration. Therefore, in order to ensure high classification accuracy, several recalibrations performed at distant recording times and, if possible, in different situations are recommended to accurately assess hypertension risk with PPG and ECG recordings, which can be obtained in a simple way through wearable devices.

Deep Learning Models

The final phase of this work explored whether deep learning classifiers could benefit from subject-specific calibration strategies, analogous to those previously tested with ML models. Pretrained models as GoogLeNet, ResNet-18, and ResNet-50 were trained on CWT-based scalograms extracted from PPG signals. This approach allowed the automatic extraction of temporal and frequency-domain features [118], marking a departure from traditional ML-based approaches that rely on handcrafted features from PPG signals and ECG data.

Classification performance was evaluated across four calibration intervals, ranging from less than 1 hour to over 24 hours between calibration and test recordings. Results confirmed the critical role of proximity in calibration, as the highest accuracy was observed when calibration and measurement were separated by less than 1 hour, with ResNet-18 achieving an F1-score of 88.36% and a specificity of 97.30%. As the temporal distance increased, classification performance decreased gradually, yet remained above 71% even for the longest intervals, indicating a robust generalization capacity over time.

These results validate the added value of personalized calibration in DL based models, while also demonstrating greater resilience compared to traditional ML approaches in the presence of temporal variability. This improved resilience may stem from the ability of DL models to extract robust, time-invariant features di-

rectly from PPG signals, without relying on auxiliary inputs such as ECG-derived PAT, which have been shown to vary over time as it depends on the preejection period [108]. This makes DL-based methods particularly well-suited for real-world applications where physiological conditions evolve over time.

Overall, this study presents a significant contribution to the field of hypertension risk discrimination, demonstrating the potential of DL and calibration methods to improve the accuracy of hypertension risk discrimination. The incorporation of these methods into wearable devices has the potential to provide reliable insights into the blood pressure status of monitored individuals, playing a pivotal role in the prevention, early diagnosis, and ongoing management of hypertension and related cardiovascular conditions. This study highlights the importance of individualized calibration for accurate hypertension risk discrimination, demonstrating the potential of this approach to achieve optimal diagnoses and treatment follow-up, while empowering patients to take a more active role in their self healthcare.

6.2.3 Detection of Hypertensive Pathology

This study explored a strategy for detecting hypertensive pathology by training models on PPG segments corresponding to stable BP intervals, while validating them on segments with known BP variability. This approach aimed to assess whether classification models could accurately identify hypertension risk regardless of absolute BP values.

Among the tested classifiers, Fine KNN yielded the highest overall performance, with an F1-score of 88.30%. The specificity reached an excellent 96.77%, while sensitivity was notably lower, suggesting that hypertensive individuals were more likely to be misclassified than normotensive ones. The observed asymmetry between sensitivity and specificity in detecting hypertensive pathology likely arises from several underlying factors. Variability in signal quality due to physiological differences among subjects, subtle irregularities in PPG waveforms, and differences in clinical importance between false-positive and false-negative results may contribute to this discrepancy. To enhance interpretability and clinical value, future studies should incorporate targeted signal-quality enhancement protocols and explore methodologies specifically designed to improve balanced performance metrics. Addressing these factors will enhance the robustness and clinical utility of the developed pathology detection models.

The results demonstrate that classification models trained on temporally coherent segments can effectively detect hypertension-related patterns, even when applied to segments affected by physiological BP variability. This is clinically relevant, as it suggests a degree of robustness to everyday stimuli and fluctuations, which is critical in ambulatory or real-world settings.

6.3 Limitations of the study

Finally, this study has certain limitations that are worth considering. Although over 900 recordings were analyzed, the number of distinct patients was relatively limited, which may restrict their generalization. Additionally, the dataset lacked essential demographic and clinical metadata such as age, sex, body mass index, or comorbidities, preventing the analysis of potential confounding factors and limiting the ability to stratify model performance across population subgroups. Prior work by Mukkamala et al. [110] highlighted the importance of such factors, demonstrating, for example, that the optimal calibration interval for blood pressure models declines with age, from one year in 30-year-old individuals to six months at age 70.

Moreover, the dataset characteristics also constrained the development and evaluation of more complex model architectures. While the PPG signals were of acceptable quality, the relatively low sampling frequency (125 Hz) and the limited recording duration may have hindered the extraction of more detailed fiducial points and waveform descriptors, which could otherwise enhance machine learning models performance.

Despite the numerous benefits of AI, it is essential to consider the limitations and challenges associated with its application in the context of vital sign monitoring and medical diagnosis. Moral dilemmas and communication barriers between physicians and patients may arise. Furthermore, privacy issues related to sensitive health data, the potential for re-identification through AI processes, and concerns about data breaches are prominent. The ethical discourse surrounding AI in healthcare, where large datasets are required, should address these privacy concerns to ensure responsible and secure implementation [119]. To this respect, the use of DL models entails the processing of sensitive health data, including PPG signals and blood pressure measurements. Ensuring the confidentiality and security of this information is paramount to maintain patient trust and comply with data protection regulations. Striking a balance between leveraging the potential benefits of DL for medical diagnosis and safeguarding patient privacy remains a critical challenge [120].

Moreover, the interpretability of AI models poses ethical challenges. The inherent complexity of neural networks makes it challenging to provide clear explanations for the decisions made by these models. In a healthcare setting, where transparency and interpretability are crucial, addressing the "black box" nature of DL algorithms becomes imperative [121]. Understanding how these models arrive at specific hypertension risk assessments is essential for physicians, patients, and regulatory bodies.

As we consider the integration of the developed models into wearable devices for real-time monitoring, additional ethical considerations come to the forefront. Wearables offer the potential for continuous, unobtrusive monitoring of individuals' cardiovascular health. This continuous monitoring, while advanta-

geous for early detection and intervention, raises concerns about the continuous surveillance of individuals' health data. Striking a balance between the benefits of proactive healthcare and the right to privacy is a delicate ethical challenge. Furthermore, while wearables offer the promise of timely health insights, issues such as device accuracy, user adherence, and data transmission security need careful consideration [122, 123]. Ensuring that wearable devices are reliable and provide accurate hypertension risk assessments is crucial for their successful integration into routine healthcare practices.

Other limitations commonly encountered when applying DL in the medical field include the requirement for large datasets to attain optimal performance, challenges related to overfitting where the model trained on data may struggle to generalize effectively with new, unseen data, the computational demands associated with DL models and issues arising from imbalanced datasets where there is a significant disparity between positive and negative samples.

6.4 Clinical Implications

The present work not only introduces promising technical advances for blood pressure estimation, but also opens pathways toward their real-world clinical adoption. This section discusses the broader clinical significance of the developed models, highlighting how their integration into wearable systems could enhance early detection, improve patient adherence to monitoring protocols, support clinical decision-making, and inform the future design and certification of medical-grade devices. Finally, the implications of using a classification-based framework for risk assessment are also examined in light of clinical needs and operational constraints.

- **Early Detection and Timely Intervention.**

  Hypertension is often referred to as a "silent killer" because it typically remains asymptomatic until advanced stages. The classification models developed in this work showed high accuracy in identifying subjects at risk. This is especially relevant for detecting prehypertensive individuals, a group that is frequently overlooked in clinical settings but could greatly benefit from early therapeutic strategies. By enabling timely intervention, the proposed approach could contribute to reducing long-term cardiovascular complications.

- **Continuous and User-Friendly Monitoring.**

  The use of PPG signals, which can be acquired using wearable devices, offers a non-invasive, low-cost, and user-friendly alternative to traditional cuff-based systems. This facilitates out-of-office, long-term blood pressure monitoring with minimal discomfort, encouraging higher adherence and better tracking of blood pressure variability in daily life.

- **Support for Clinical Decision-Making** The models developed can serve as clinical decision support tools by complementing traditional blood pressure measurement. Alerts generated in ambulatory or home-based settings could prompt patients to seek medical attention sooner and assist healthcare professionals in optimizing treatment and follow-up plans.

- **Future Medical Certification of Wearable Devices.**

  Although current commercial smartwatches provide blood pressure estimations, these devices are not yet certified for clinical use. This thesis contributes to bridging this gap by demonstrating that AI-based classification using PPG signals can provide clinically relevant information. Moreover, the study supports the inclusion of recalibration protocols with occasional reference measurements to maintain model performance over time. Additionally, the finding that lower sampling frequencies did not degrade performance would lead to more energy-efficient wearable solutions.

- **Classification Over Regression**

  Unlike regression-based approaches that estimate absolute blood pressure values, the classification-based framework adopted in this work assigns a hypertensive risk level to each subject, regardless of specific blood pressure values. This strategy reflects a clinically valuable perspective by prioritizing risk stratification over exact numerical estimation, which is often influenced by short-term physiological fluctuations and device variability.

Chapter 7

Conclusions, Future Lines and Contributions

This final chapter presents in Section 7.1 the main conclusions drawn from the research conducted throughout this doctoral thesis. In addition, given that the accomplishment of a research work opens new avenues for further exploration, potential future lines of research are proposed in Section 7.2. Finally, the chapter concludes in Section 7.3 by summarizing the scientific contributions resulting from the work.

7.1 Global conclusions

This doctoral thesis has explored the potential of PPG signals obtained through wearable devices and combined with AI techniques as ML and DL to support the detection and monitoring of hypertension in a non-invasive and continuous manner. The main conclusions derived from the research are summarized below.

Firstly, the results demonstrate that AI-based methods can effectively classify hypertensive risk using features derived from PPG signals alone, without the need for traditional cuff-based measurements or complementary signals as ECG. This finding supports the feasibility of integrating such methods into wearable technologies for continuous BP monitoring.

Secondly, the incorporation of reference BP measurements into the models, serving as a form of calibration, significantly improves classification accuracy, especially for unseen subjects not included in the training phase. Notably, repeated calibrations yield better performance; however, even a single measurement or calibrations separated by more than 24 hours provide a meaningful performance boost when compared to non-calibrated approaches.

Thirdly, this thesis emphasizes the clinical relevance of detecting hypertension as a pathological condition rather than predicting absolute BP values, a shift that aligns with current trends in preventive and personalized medicine. Classifying subjects based on risk groups—particularly identifying prehypertensive individuals (systolic BP between 120–140 mmHg) has proven to be an optimal strategy both from a performance and a clinical standpoint, serving as an early warning system for timely medical intervention.

Throughout this work, different modeling strategies have been compared. ML algorithms that rely on hand-crafted features such as Pulse Arrival Time, extracted from PPG and ECG signals, were evaluated alongside pretrained DL models such as GoogleNet and ResNet, which automatically learn features from the CWT representation of the PPG signal. While both approaches yield competitive results, DL models have shown greater generalization capabilities and require less signal preprocessing, making them more suitable for deployment in real-world applications.

Another important contribution is the analysis of signal preprocessing. The study has shown that downsampling PPG signals does not significantly degrade classification performance. This insight plays a key role in supporting its application to wearable devices, as it enables the use of lower sampling rates, reducing computational load and power consumption without compromising accuracy.

In summary, this thesis supports the integration of wearable PPG sensors and AI-based models for non-invasive hypertension risk assessment. It highlights the importance of calibration strategies, appropriate risk group definitions, and the suitability of DL models for real-time implementation.

Compared to existing cuffless blood pressure monitoring systems, the proposed approach offers a calibrated, ML- and DL-based alternative that balances accuracy, interpretability, and computational efficiency. These attributes represent a meaningful step forward in enabling clinically deployable cardiovascular monitoring technologies.

7.2 Future lines of research

Building upon the findings and methodologies developed in this thesis, several promising research directions can be pursued to enhance the performance, interpretability, and clinical applicability of hypertension risk assessment using PPG signals and artificial intelligence.

Firstly, increasing the number of subjects analyzed would contribute to obtaining more robust and generalizable results. A larger and more diverse dataset would improve the ability of the models to learn relevant patterns associated with different patient profiles, thus enhancing predictive performance. In addition, ensuring a balanced distribution of subjects across hypertension risk categories would help avoid biases in classification and strengthen the reliability of model outputs. These subjects could be obtained from the MIMIC database or through new data acquisition protocols.

Another important direction involves improving the preprocessing stage. Currently, the selection of clean PPG segments is performed manually based on visual inspection. In the future, automatic methods for signal quality assessment could be implemented to identify optimal segments for analysis, thus reducing human bias and improving reproducibility. These methods could be based on signal quality indices or machine learning algorithms trained to detect usable portions of the signal.

Moreover, incorporating clinical metadata such as age, sex, cardiovascular risk factors, medication use, and comorbidities would allow the development of personalized models that are more reflective of real-world patient heterogeneity. This integration could help reduce confounding effects and improve both the accuracy and clinical relevance of predictions.

In terms of model development, although this work has focused on CNN architectures such as GoogLeNet and ResNet, it would be worthwhile to evaluate other pretrained models or even design custom architectures tailored to PPG signals. Additionally, hybrid approaches that extract deep features automatically and use them as input to classical machine learning classifiers offer a compelling strategy to reduce computational cost while maintaining high performance.

Future work should also address the need for explainability in AI-based medical systems. Incorporating explainable AI methods, such as Grad-CAM or SHAP values, would help interpret model decisions and build trust among clinicians.

Similarly, estimating model uncertainty could allow systems to communicate confidence in their predictions, which is particularly relevant in clinical decision-making.

Finally, once these models are validated and refined, their implementation in wearable devices such as smartwatches would represent a major step towards practical, continuous blood pressure monitoring. Although current consumer-grade wearables offer rough blood pressure estimations, they are not certified medical devices and often require regular calibration with standard sphygmomanometers. Future research should aim to develop models and systems that can function reliably without frequent calibration and meet clinical validation standards, much like current wearable heart rate monitors. In this regard, classification of blood pressure levels could serve as a first clinically meaningful milestone before tackling precise estimation.

Overall, these lines of future work aim to bridge the gap between experimental research and real-world healthcare applications, contributing to more accurate, accessible, and personalized hypertension monitoring and management.

7.3 Scientific contributions

The work carried out in this thesis has been published in several scientific journals of international scope and has been presented in national and international conferences. Next, these scientific contributions are enumerated.

The study of the relevance of calibration in ML-based hypertension risk assessment has been published in the journal *Biosensors*:

- **Jesús Cano**, Lorenzo Fácila, Juan M. Gracia-Baena, Roberto Zangróniz, Raúl Alcaraz and José J. Rieta. "The Relevance of Calibration in Machine Learning-Based Hypertension Risk Assessment Combining Photoplethysmography and Electrocardiography". *Biosensors*, vol. 12(5), 289, 2022.

A subsequent study, focusing on the same objective but using DL models, was published in the journal *Bioengineering*

- **Jesús Cano**, Vicente Bertomeu-González, Lorenzo Fácila, Fernando Hornero, Raúl Alcaraz and José J. Rieta. "Improved Hypertension Risk Assessment with Photoplethysmographic Recordings Combining Deep Learning and Calibration". *Bioengineering*, vol. 10(12), 1439, 2023.

In addition, some of the main contributions of this thesis for hypertension risk detection employing ML and DL strategies have been presented in international conferences:

- **Jesús Cano**, Aurelio Quesada, Flavia Ravelli, Roberto Zangróniz, Raúl Alcaraz and José J. Rieta. "Novel Photoplethysmographic and Electrocardiographic Features for Enhanced Detection of Hypertensive Individuals". *International Conference on e-Health and Bioengineering (EHB)*, pp. 1-4, 2021.

- **Jesús Cano**, Lorenzo Fácila, Philip Langley, Roberto Zangróniz, Raúl Alcaraz and José J. Rieta. "Application of Deep Neural Network Models for Blood Pressure Classification based on Photoplethysmograpic Recordings". *International Conference on e-Health and Bioengineering (EHB)*, pp. 1-4, 2021.

- **Jesús Cano**, Fernando Hornero, Aurelio Quesada, Arturo Martínez-Rodrigo, Raúl Alcaraz and José J. Rieta. "Improved Discrimination Between Healthy and Hypertensive Individuals Combining Photoplethysmography and Electrocardiography". *Computing in Cardiology (CinC)*, vol.48, pp. 1-4, 2021.

- **Jesús Cano**, Vicente Bertomeu-González, Lorenzo Fácila, Roberto Zangróniz, Raúl Alcaraz and José J. Rieta. "Hypertension Risk Assessment from Photoplethysmographic Recordings Using Deep Learning Classifiers". *Computing in Cardiology (CinC)*, vol.48, pp. 1-4, 2021.

- **Jesús Cano**, Lorenzo Fácia, Fernando Hornero, Philip Langley, Raúl Alcaraz and José J. Rieta. "Cuffless Hypertension Risk Assessment and the Significance of Calibration". *Computing in Cardiology (CinC)*, vol. 498, pp. 1-4, 2022.

- **Jesús Cano**, Vicente Bertomeu-González, Lorenzo Fácia, José Moreno-Arribas, Raúl Alcaraz and José J. Rieta. "ECG and PPG-Based Hypertension Screening Under Non-Hypertensive Blood Pressure Recordings". *Computing in Cardiology (CinC)*, vol. 498, pp. 1-4, 2022.

and in national conferences:

- **Jesús Cano**, Raúl Alcaraz and José J. Rieta. "Discriminación Mejorada Entre Personas Sanas e Hipertensas Combinando Fotopletismografía y Electrocardiografía". *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB)*, 2020.

- **Jesús Cano**, Vicente Bertomeu-González, Lorenzo Fácia, Roberto Zangróniz, Raúl Alcaraz and José J. Rieta. "Evaluación del Riesgo de Hipertensión a partir de Registros Fotopletismográficos utilizando Clasificadores de Deep Learning". *XXXIX Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB)*, 2021.

Bibliography

[1] R. M. Carey, P. Muntner, H. B. Bosworth, and P. K. Whelton. "Prevention and control of hypertension". *Journal of the American College of Cardiology*, 72:1278–1293, 2018. [1]

[2] R. Gauer. "Severe asymptomatic hypertension: Evaluation and treatment". *American family physician*, 95:492–500, 2017. [1]

[3] A. H. Krist, K. W. Davidson, C. M. Mangione, M. Cabana, A. B. Caughey, E. M. Davis, K. E. Donahue, C. A. Doubeni, M. Kubik, L. Li, G. Ogedegbe, L. Pbert, M. Silverstein, J. Stevermer, C.-W. Tseng, and J. B. Wong. "Screening for hypertension in adults". *JAMA*, 325:1650, 2021. [2]

[4] B.-M. Schmidt, S. Durao, I. Toews, C. M. Bavuma, A. Hohlfeld, E. Nury, J. J. Meerpohl, and T. Kredo. "Screening strategies for hypertension". *Cochrane Database of Systematic Reviews*, 2020, 2020. [2]

[5] A. S. Meidert and B. Saugel. "Techniques for non-invasive monitoring of arterial blood pressure". *Frontiers in Medicine*, 4:231, 2018. [2, 12, 13, 14]

[6] E. M. Frese, A. Fick, and S. H. Sadowsky. "Blood pressure measurement guidelines for physical therapists". *Cardiopulmonary Physical Therapy Journal*, 22:5–12, 2011. [2]

[7] P. A. Shaltis, A. Reisner, and H. H. Asada. "Wearable, cuff-less ppg-based blood pressure monitor with novel height sensor". In "2006 International Conference of the IEEE Engineering in Medicine and Biology Society", pages 908–911. IEEE, 2006. [2]

[8] J. Allen. "Photoplethysmography and its application in clinical physiological measurement". *Physiological Measurement*, 28:R1–R39, 2007. [2, 17]

[9] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida. "Wearable photoplethysmographic sensors-past and present". *Electronics*, 3:282–302, 2014. [2, 3, 18]

[10] G. Martínez, N. Howard, D. Abbott, K. Lim, R. Ward, and M. Elgendi. "Can photoplethysmography replace arterial blood pressure in the assessment of blood pressure?" *Journal of Clinical Medicine*, 7:316, 2018. [2, 74]

[11] Y.-Z. Yoon, J. M. Kang, Y. Kwon, S. Park, S. Noh, Y. Kim, J. Park, and S. W. Hwang. "Cuff-less blood pressure estimation using pulse waveform analysis and pulse arrival time". *IEEE Journal of Biomedical and Health Informatics*, 22:1068–1074, 2018. [3, 43, 74]

[12] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany. "Cuffless blood pressure estimation algorithms for continuous health-care monitoring". *IEEE Transactions on Biomedical Engineering*, 64:859–869, 2017. [3, 75]

[13] Y. Zhang and Z. Feng. "A svm method for continuous blood pressure estimation from a ppg signal". In "Proceedings of the 9th International Conference on Machine Learning and Computing", volume Part F1283, pages 128–132. ACM, 2017. [3]

[14] U. Senturk, I. Yucedag, 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)", pages 1–4. IEEE, 2018. [3]

[15] X. Liang. Theoretical basis, pages 1–40. Elsevier, 2020. [3, 31, 32, 104]

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. "Imagenet large scale visual recognition challenge". *International Journal of Computer Vision*, 115:211–252, 2015. [3]

[17] Y. Liang, Z. Chen, R. Ward, and M. Elgendi. "Photoplethysmography and deep learning: Enhancing hypertension risk stratification". *Biosensors*, 8, 2018. [3, 49, 75]

[18] O. Schlesinger, N. Vigderhouse, D. Eytan, and Y. Moshe. "Blood pressure estimation from PPG signals using convolutional neural networks and siamese network". In "ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", volume 2020-May, pages 1135–1139. IEEE, 2020. [3, 50, 75]

[19] H. Tjahjadi, K. Ramli, and H. Murfi. "Noninvasive classification of blood pressure based on photoplethysmography signals using bidirectional long short-term memory and time-frequency analysis". *IEEE Access*, 8:20735–20748, 2020. [3, 50]

[20] J. Fine, K. L. Branan, A. J. Rodriguez, T. Boonya-ananta, Ajmal, J. C. Ramella-Roman, M. J. McShane, and G. L. Coté. "Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring". *Biosensors*, 11:126, 2021. [4]

[21] J. Mayet. "Cardiac and vascular pathophysiology in hypertension". *Heart*, 89:1104–1109, 2003. [8, 9]

[22] D. N. Ku. "Blood flow in arteries". *Annual Review of Fluid Mechanics*, 29:399–434, 1997. [8]

[23] R. Klabunde. Cardiovascular physiology concepts. Lippincott Williams & Wilkins, 2011. [8, 9, 103]

[24] V. M. Chadachan, M. T. Ye, J. C. Tay, K. Subramaniam, and S. Setia. "Understanding short-term blood-pressure-variability phenotypes: from concept to clinical practice". *International Journal of General Medicine*, Volume 11:241–254, 2018. [9]

[25] G. Fortino and V. Giampa. "PPG-based methods for non invasive and continuous blood pressure measurement: an overview and development issues in body sensor networks". In "2010 IEEE International Workshop on Medical Measurements and Applications", pages 10–13. IEEE, 2010. [9]

[26] C. El-Hajj and P. Kyriacou. "A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure". *Biomedical Signal Processing and Control*, 58:101870, 2020. [9, 76]

[27] T. Unger, C. Borghi, F. Charchar, N. A. Khan, N. R. Poulter, D. Prabhakaran, A. Ramirez, M. Schlaich, G. S. Stergiou, M. Tomaszewski, R. D. Wainford, B. Williams, and A. E. Schutte. "2020 international society of hypertension global hypertension practice guidelines". *Hypertension*, 75:1334–1357, 2020. [10]

[28] P. S. Collaboration et al. "Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies". *The Lancet*, 360:1903–1913, 2002. [10]

[29] A. V. Chobanian, G. L. Bakris, H. R. Black, W. C. Cushman, L. A. Green, J. L. Izzo, D. W. Jones, B. J. Materson, S. Oparil, J. T. Wright, and E. J. Roccella. "Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure". *Hypertension*, 42:1206–1252, 2003. [10]

[30] S. E. Kjeldsen. "Hypertension and cardiovascular risk: General aspects". *Pharmacological Research*, 129:95–99, 2018. [10]

[31] G. Mancia, R. Kreutz, M. Brunström, M. Burnier, G. Grassi, A. Januszewicz, M. L. Muiesan, K. Tsioufis, E. Agabiti-Rosei, E. A. E. Algharably, M. Azizi, A. Benetos, C. Borghi, J. B. Hitij, R. Cifkova, A. Coca, V. Cornelissen, J. K. Cruickshank, P. G. Cunha, A. J. Danser, R. M. de Pinho, C. Delles, A. F. Dominiczak, M. Dorobantu, M. Doumas, M. S. Fernández-Alfonso, J.-M. Halimi, Z. Járai, B. Jelaković, J. Jordan, T. Kuznetsova, S. Laurent, D. Lovic, E. Lurbe, F. Mahfoud, A. Manolis, M. Miglinas, K. Narkiewicz, T. Niiranen,

P. Palatini, G. Parati, A. Pathak, A. Persu, J. Polonia, J. Redon, P. Sarafidis, R. Schmieder, B. Spronck, S. Stabouli, G. Stergiou, S. Taddei, C. Thomopoulos, M. Tomaszewski, P. V. de Borne, C. Wanner, T. Weber, B. Williams, Z.-Y. Zhang, and S. E. Kjeldsen. "2023 esh guidelines for the management of arterial hypertension the task force for the management of arterial hypertension of the european society of hypertension". *Journal of Hypertension*, 41:1874–2071, 2023. [10]

[32] N. R. F. C. (NCD-RisC). "Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants." *Lancet (London, England)*, 398:957–980, 2021. [10]

[33] J. R. Banegas, M. Sánchez-Martínez, T. Gijón-Conde, E. López-García, A. Graciani, P. Guallar-Castillón, J. García-Puig, and F. Rodríguez-Artalejo. "Numerical values and impact of hypertension in spain". *Revista Española de Cardiología (English Edition)*, 2024. [10]

[34] "World health organization. a global brief on hypertension: silent killer, global public health crisis: World health day 2013", 2013. [11]

[35] C. T. Goodman and G. B. Kitchen. "Measuring arterial blood pressure". *Anaesthesia & Intensive Care Medicine*, 24:785–789, 2023. [11, 103]

[36] K. Lakhal, S. Ehrmann, and T. Boulain. "Noninvasive bp monitoring in the critically ill". *Chest*, 153:1023–1039, 2018. [12, 103]

[37] D. Perloff, C. Grim, J. Flack, E. D. Frohlich, M. Hill, M. McDonald, and B. Z. Morgenstern. "Human blood pressure determination by sphygmomanometry." *Circulation*, 88:2460–2470, 1993. [13]

[38] B. S. Alpert, D. Quinn, and D. Gallick. "Oscillometric blood pressure: a review for clinicians". *Journal of the American Society of Hypertension*, 8:930–938, 2014. [13]

[39] X.-R. Ding, N. Zhao, G.-Z. Yang, R. I. Pettigrew, B. Lo, F. Miao, Y. Li, J. Liu, and Y.-T. Zhang. "Continuous blood pressure measurement from invasive to unobtrusive: Celebration of 200th birth anniversary of carl ludwig". *IEEE Journal of Biomedical and Health Informatics*, 20:1455–1465, 2016. [13]

[40] N. Peters, M. A. Gatzoulis, and R. Vecht. ECG Diagnosis in Clinical Practice. Springer London, 2009. [14]

[41] A. U, J. A. Spaan, J. S. Suri, and S. M. R. Krishnan. Advances in Cardiac Signal Processing. Springer Berlin Heidelberg, 2007. [15]

[42] J. J. Rieta and R. Alcaraz. The Genesis of the Electrocardiogram (ECG), pages 1–15. Wiley, 2017. [15, 103]

[43] A. L. Goldberger. Basic ECG Waves, pages 7–20. Elsevier, 2006. [15]

[44] W. Liang, Y. Zhang, J. Tan, and Y. Li. "A novel approach to ecg classification based upon two-layered hmms in body sensor networks". *Sensors*, 14:5994–6011, 2014. [16, 103]

[45] Pankaj, A. Kumar, R. Komaragiri, and M. Kumar. "A review on computation methods used in photoplethysmography signal analysis for heart rate estimation". *Archives of Computational Methods in Engineering*, 29:921–940, 2022. [17, 19]

[46] T. Tamura. "Current progress of photoplethysmography and spo2 for health monitoring". *Biomedical Engineering Letters*, 9:21–36, 2019. [17]

[47] M. Elgendi, R. Fletcher, Y. Liang, N. Howard, N. H. Lovell, D. Abbott, K. Lim, and R. Ward. "The use of photoplethysmography for assessing hypertension". *npj Digital Medicine*, 2:60, 2019. [17, 21, 22, 43]

[48] Q. Y. Lee, S. J. Redmond, G. S. Chan, P. M. Middleton, E. Steel, P. Malouf, C. Critoph, G. Flynn, E. O'Lone, and N. H. Lovell. "Estimation of cardiac output and systemic vascular resistance using a multivariate regression model with features selected from the finger photoplethysmogram and routine cardiovascular measurements". *BioMedical Engineering OnLine*, 12:19, 2013. [17]

[49] M.-T. Wu, I.-F. Liu, Y.-H. Tzeng, and L. Wang. "Modified photoplethysmography signal processing and analysis procedure for obtaining reliable stiffness index reflecting arteriosclerosis severity". *Physiological Measurement*, 43:085001, 2022. [17]

[50] P. H. Charlton, B. Paliakaitė, K. Pilt, M. Bachler, S. Zanelli, D. Kulin, J. Allen, M. Hallab, E. Bianchini, C. C. Mayer, D. Terentes-Printzios, V. Dittrich, B. Hametner, D. Veerasingam, D. Žikić, and V. Marozas. "Assessing hemodynamics from the photoplethysmogram to gain insights into vascular age: a review from vascagenet." *American journal of physiology. Heart and circulatory physiology*, 322:H493–H522, 2022. [17]

[51] I. Lee, N. Park, H. Lee, C. Hwang, J. H. Kim, and S. Park. "Systematic review on human skin-compatible wearable photoplethysmography sensors". *Applied Sciences*, 11:2313, 2021. [17]

[52] W. Wang. "Robust and automatic remote photoplethysmography", 2017. [18, 103]

[53] M. Nitzan and Z. Ovadia-Blechman. Physical and physiological interpretations of the PPG signal, pages 319–340. Elsevier, 2022. [18]

[54] J. Park, H. S. Seok, S.-S. Kim, and H. Shin. "Photoplethysmogram analysis and applications: An integrative review". *Frontiers in Physiology*, 12, 2022. [19, 44, 104]

[55] M. Elgendi, Y. Liang, and R. Ward. "Toward generating more diagnostic features from photoplethysmogram waveforms". *Diseases*, 6:20, 2018. [19]

[56] M. Z. Suboh, R. Jaafar, N. A. Nayan, N. H. Harun, and M. S. F. Mohamad. "Analysis on four derivative waveforms of photoplethysmogram (ppg) for fiducial point detection". *Frontiers in Public Health*, 10, 2022. [20, 42, 44, 104]

[57] Y. Liang, Z. Chen, R. Ward, and M. Elgendi. "Hypertension assessment via ecg and ppg signals: An evaluation using mimic database". *Diagnostics*, 8:65, 2018. [20, 47, 63, 74]

[58] Y. Liang, D. Abbott, N. Howard, K. Lim, R. Ward, and M. Elgendi. "How effective is pulse arrival time for evaluating blood pressure? challenges and recommendations from a study using the mimic database". *Journal of Clinical Medicine*, 8:337, 2019. [21, 74, 104]

[59] F. S. Cattivelli and H. Garudadri. "Noninvasive cuffless estimation of blood pressure from pulse arrival time and heart rate with adaptive calibration". In "2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks", pages 114–119. IEEE, 2009. [20]

[60] J. M. Bote, J. Recas, and R. Hermida. "Evaluation of blood pressure estimation models based on pulse arrival time". *Computers & Electrical Engineering*, 84:106616, 2020. [21, 22, 43]

[61] P. Ongsulee. "Artificial intelligence, machine learning and deep learning". In "2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)", pages 1–6. IEEE, 2017. [24]

[62] Q. Dong, X. Chen, and B. Huang. "Logistic regression". *Data Analysis in Pavement Engineering*, pages 141–152, 2024. [24]

[63] Q. Ma. "Recent applications and perspectives of logistic regression modelling in healthcare". *Theoretical and Natural Science*, 36:185–190, 2024. [24]

[64] M. C. Guerrero, J. S. Parada, and H. E. Espitia. "EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks". *Heliyon*, 7:e07258, 2021. [25]

[65] Q. An, S. Rahman, J. Zhou, and J. J. Kang. "A comprehensive review on machine learning in healthcare industry: Classification, restrictions, opportunities and challenges". *Sensors*, 23:4178, 2023. [25, 27]

[66] C. Capra, P. Marti-Puig, D. Moreno, L. Llunas, S. Nicolaou, C. Schmidt, and J. Solé-Casals. "Preliminary results on the use of classification trees to predict non-suicidal self-injury with data collected through a mobile app". In "Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies", pages 278–282. SCITEPRESS - Science and Technology Publications, 2022. [25]

[67] D. W. Aha, D. Kibler, and M. K. Albert. "Instance-based learning algorithms". *Machine Learning*, 6:37–66, 1991. [26]

[68] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide. "Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction". *Scientific Reports*, 12:1–11, 2022. [26, 104]

[69] Z. Ghahramani. "Probabilistic machine learning and artificial intelligence". *Nature*, 521:452–459, 2015. [26]

[70] X. Liang, S. Du, Y. Duan, Y. Chen, K. Zhu, Y. Yin, and L. Jian. "Kernel-based algorithms for image classification: A review", 2023. [27]

[71] E. García-Gonzalo, Z. Fernández-Muñiz, P. G. Nieto, A. B. Sánchez, and M. M. Fernández. "Hard-rock stability analysis for span design in entry-type excavations with learning classifiers". *Materials*, 9:531, 2016. [27, 104]

[72] E. M. dos Santos and H. M. Gomes. A Comparative Study of Polynomial Kernel SVM Applied to Appearance-Based Object Recognition, volume 2388, pages 408–418. Springer, Berlin, Heidelberg, 2002. [28]

[73] A. Ganiyu, I. Darvishi, R. Addo-Quaye, A. Yeboah-Ofori, B. T. Asare, and O. Oguntoyinbo. "Classification algorithms using ensemble methods". In "2024 11th International Conference on Future Internet of Things and Cloud (FiCloud)", pages 168–175. IEEE, 2024. [28]

[74] D. H. Hubel and T. N. Wiesel. "Receptive fields and functional architecture of monkey striate cortex". *The Journal of Physiology*, 195:215–243, 1968. [29, 38]

[75] A. Maier, C. Syben, T. Lasser, and C. Riess. "A gentle introduction to deep learning in medical image processing". *Zeitschrift für Medizinische Physik*, 29:86–101, 2019. [29]

[76] K. O'Shea and R. Nash. "An introduction to convolutional neural networks". *arXiv preprint arXiv:1511.08458*, 2015. [29, 30, 32, 104]

[77] S. Albawi, T. A. Mohammed, and S. Al-Zawi. "Understanding of a convolutional neural network". In "2017 International Conference on Engineering and Technology (ICET)", volume 2018-Janua, pages 1–6. IEEE, 2017. [30, 32]

[78] A. S. Lundervold and A. Lundervold. "An overview of deep learning in medical imaging focusing on mri". *Zeitschrift für Medizinische Physik*, 29:102–127, 2019. [30, 31, 104]

[79] L. N. Smith. "Cyclical learning rates for training neural networks". In "2017 IEEE Winter Conference on Applications of Computer Vision (WACV)", pages 464–472. IEEE, 2017. [30]

[80] I. Kandel and M. Castelli. "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset". *ICT Express*, 6:312–315, 2020. [30]

[81] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. "Going deeper with convolutions". In "2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", volume 07-12-June, pages 1–9. IEEE, 2015. [32, 33, 105]

[82] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition". In "2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", volume 2016-Decem, pages 770–778. IEEE, 2016. [33, 34, 105]

[83] M. S. Roy, R. Gupta, J. K. Chandra, K. D. Sharma, and A. Talukdar. "Improving photoplethysmographic measurements under motion artifacts using artificial neural network for personal healthcare". *IEEE Transactions on Instrumentation and Measurement*, 67:2820–2829, 2018. [35]

[84] R. Ahmed, A. Mehmood, M. M. U. Rahman, and O. A. Dobre. "A deep learning and fast wavelet transform-based hybrid approach for denoising of ppg signals". *IEEE Sensors Letters*, 7, 2023. [35]

[85] A. Attarpour, A. Mahnam, A. Aminitabar, and H. Samani. "Cuff-less continuous measurement of blood pressure using wrist and fingertip photoplethysmograms: Evaluation and feature analysis". *Biomedical Signal Processing and Control*, 49:212–220, 2019. [35, 52, 67]

[86] Tjahjadi and Ramli. "Noninvasive blood pressure classification based on photoplethysmography using k-nearest neighbors algorithm: A feasibility study". *Information*, 11:93, 2020. [35, 52, 67, 74]

[87] M. M. R. K. Mamun and A. T. Alouani. "Cuffless blood pressure measurement using linear and nonlinear optimized feature selection". *Diagnostics 2022, Vol. 12, Page 408*, 12:408, 2022. [35]

[88] A. Gudigar, N. A. Kadri, U. Raghavendra, J. Samanth, M. Maithri, M. A. Inamdar, M. A. Prabhu, A. Hegde, M. Salvi, C. H. Yeong, P. D. Barua, F. Molinari, and U. R. Acharya. "Automatic identification of hypertension and assessment of its secondary effects using artificial intelligence: A systematic review (2013–2023)". *Computers in Biology and Medicine*, 172:108207, 2024. [35]

[89] Y. Liang, M. Elgendi, Z. Chen, and R. Ward. "An optimal filter for short photoplethysmogram signals". *Scientific Data*, 5:180076, 2018. [40]

[90] M. Elgendi. "On the analysis of fingertip photoplethysmogram signals". *Current Cardiology Reviews*, 8:14–25, 2012. [40, 43, 44]

[91] L. Sörnmo and P. Laguna. Bioelectrical Signal Processing in Cardiac and Neurological Applications. Elsevier, 2005. [40]

[92] J. Pan and W. J. Tompkins. "A real-time qrs detection algorithm". *IEEE Transactions on Biomedical Engineering*, BME-32:230–236, 1985. [42]

[93] K. Takazawa, N. Tanaka, M. Fujita, O. Matsuoka, T. Saiki, M. Aikawa, S. Tamura, and C. Ibukiyama. "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform". *Hypertension*, 32:365–370, 1998. [43, 44]

[94] L. Evdochim, D. Dobrescu, and L. Dobrescu. "Advanced electro-optical analysis of photoplethysmogram signal". In "2019 IEEE 31st International Conference on Microelectronics (MIEL)", pages 259–262. IEEE, 2019. [44]

[95] S. S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, and Y. Ghorbani. "Blood pressure estimation from appropriate and inappropriate ppg signals using a whole-based method". *Biomedical Signal Processing and Control*, 47:196–206, 2019. [44]

[96] M. S. Tanveer and M. K. Hasan. "Cuffless blood pressure estimation from electrocardiogram and photoplethysmogram using waveform based ann-lstm network". *Biomedical Signal Processing and Control*, 51:382–392, 2019. [44]

[97] S. K. Kwak and J. H. Kim. "Statistical data preparation: management of missing values and outliers". *Korean Journal of Anesthesiology*, 70:407, 2017. [46]

[98] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata. "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median". *Journal of Experimental Social Psychology*, 49:764–766, 2013. [46]

[99] S. Arlot and A. Celisse. "A survey of cross-validation procedures for model selection". *Statistics Surveys*, 4, 2010. [47]

[100] S. Shalev-Shwartz and S. Ben-David. Understanding Machine Learning, volume 9781107057. Cambridge University Press, 2014. [47]

[101] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2014. [49, 51]

[102] I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning, pages 1139–1147. PMLR, 2013. [49, 51]

[103] X. Ying. "An overview of overfitting and its solutions". *Journal of Physics: Conference Series*, 1168:022022, 2019. [49]

[104] L. Aguiar-Conraria and M. J. Soares. "The continuous wavelet transform: Moving beyond uni- and bivariate analysis". *Journal of Economic Surveys*, 28:344–375, 2014. [49]

[105] D. Fujita and A. Suzuki. "Evaluation of the possible use of ppg waveform features measured at low sampling rate". *IEEE Access*, 7:58361–58367, 2019. [50, 56]

[106] I. Kononenko. Estimating attributes: Analysis and extensions of RELIEF, volume 784 LNCS, pages 171–182. Springer, Berlin, Heidelberg, 1994. [53]

[107] H. Gao, B. Qiu, R. J. D. Barroso, W. Hussain, Y. Xu, and X. Wang. "Tsmae: A novel anomaly detection approach for internet of things time series data using memory-augmented autoencoder". *IEEE Transactions on Network Science and Engineering*, pages 1–1, 2022. [74]

[108] R. Mukkamala, J.-O. Hahn, O. T. Inan, L. K. Mestha, C.-S. Kim, H. Toreyin, and S. Kyal. "Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice". *IEEE Transactions on Biomedical Engineering*, 62:1879–1901, 2015. [74, 75, 81]

[109] S. Suzuki and K. Oguri. "Cuffless and non-invasive systolic blood pressure estimation for aged class by using a photoplethysmograph." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 2008:1327–30, 2008. [75]

[110] R. Mukkamala and J.-O. Hahn. "Toward ubiquitous blood pressure monitoring via pulse transit time: Predictions on maximum calibration period and acceptable error limits". *IEEE Transactions on Biomedical Engineering*, 65:1410–1420, 2018. [75, 82]

[111] H. Gao, K. Xu, M. Cao, J. Xiao, Q. Xu, and Y. Yin. "The deep features and attention mechanism-based method to dish healthcare under social iot systems: An empirical study with a hand-deep local–global net". *IEEE Transactions on Computational Social Systems*, 9:336–347, 2022. [75]

[112] H. Gao, J. Xiao, Y. Yin, T. Liu, and J. Shi. "A mutually supervised graph attention network for few-shot segmentation: The perspective of fully utilizing limited samples". *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–13, 2022. [75]

[113] X. Sun, L. Zhou, S. Chang, and Z. Liu. "Using cnn and hht to predict blood pressure level based on photoplethysmography and its derivatives". *Biosensors*, 11:120, 2021. [75]

[114] J. Wu, H. Liang, C. Ding, X. Huang, J. Huang, and Q. Peng. "Improving the accuracy in classification of blood pressure from photoplethysmography using continuous wavelet transform and deep learning". *International Journal of Hypertension*, 2021:1–9, 2021. [75]

[115] M. Radha, K. de Groot, N. Rajani, C. C. P. Wong, N. Kobold, V. Vos, P. Fonseca, N. Mastellos, P. A. Wark, N. Velthoven, R. Haakma, and R. M. Aarts. "Estimating blood pressure trends and the nocturnal dip from photoplethysmography". *Physiological Measurement*, 40:025006, 2019. [76]

[116] E. Finnegan, S. Davidson, M. Harford, J. Jorge, and M. Villarroel. "The presence of a circadian rhythm in pulse arrival time and its application for

classifying blood pressure night-time dip". *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 2021:488–491, 2021. [76]

[117] J. Esmaelpoor, M. H. Moradi, and A. Kadkhodamohammadi. "A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals". *Computers in Biology and Medicine*, 120:103719, 2020. [78]

[118] A. A. Fahoum, A. A. Omari, G. A. Omari, and A. Zyout. "Development of a novel light-sensitive ppg model using ppg scalograms and ppg-net learning for non-invasive hypertension monitoring". *Heliyon*, 10, 2024. [80]

[119] P. Esmaeilzadeh. "Use of ai-based tools for healthcare purposes: A survey study from consumers' perspectives". *BMC Medical Informatics and Decision Making*, 20:1–19, 2020. [82]

[120] N. M. Safdar, J. D. Banja, and C. C. Meltzer. "Ethical considerations in artificial intelligence". *European Journal of Radiology*, 122:108768, 2020. [82]

[121] A. I. F. Poon and J. J. Y. Sung. "Opening the black box of ai-medicine". *Journal of Gastroenterology and Hepatology*, 36:581–584, 2021. [82]

[122] M. Canim, M. Kantarcioglu, and B. Malin. "Secure management of biomedical data with cryptographic hardware". *IEEE Transactions on Information Technology in Biomedicine*, 16:166–175, 2012. [83]

[123] L. Ismail, H. Materwala, A. P. Karduck, and A. Adem. "Requirements of health data management systems for biomedical care and research: Scoping review". *Journal of Medical Internet Research*, 22:e17508, 2020. [83]

List of Figures

List of Tables

List of Acronyms

ABP	Arterial blood pressure
AC	Alternating current
Acc	Accuracy
APG	Acceleration Plethysmogram
AI	Artificial intelligence
ANN	Artificial neural networks
AV	Atrioventricular
BP	Blood pressure
CNN	Convolutional Neural Network
CO	Cardiac Output
CPU	Central processing unit
CV	Cardiovascular
CWT	Continuous wavelet transform
DBP	Diastolic blood pressure
DC	Direct current
DL	Deep Learning
ECG	Electrocardiography
FN	False Negative
FP	False Positive
GPU	Graphics processing unit
HTS	Hypertensive
IBL	Instance-based learning
ICU	Intensive Care Unit
IPA	Inflexion Point Area
KNN	K-Nearest-Neighbor
LED	Light-emitting diodes

MAD	Median Absolute Deviation
MIMIC	Medical Information Mart for Intensive Care
ML	Machine learning
NaN	Not a number
NTS	Normotensive
PAT	Pulse arrival time
PEP	Pre-ejection period
PHT	Prehypertension
PPG	Photoplethysmography
PTT	Pulse transit time
PVR	Peripheral vascular resistance
PWV	Pulse wave velocity
RGB	Red-Green-Blue
SBP	Systolic blood pressure
SE	Sensitivity
SP	Specifity
STFT	Short time Fourier transform
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
TPI	Time pulse interval
TPP	Time peak to peak
VPG	Velocity Plethysmogram