RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

ÉCOLE NATIONALE POLYTECHNIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Institut national de recherche en informatique et en automatique

Département d'électronique

End of Studies Project

Submitted for the fulfillment of State Engineer Degree in Electronics

---

PPG signals-based Arterial Stiffness estimation using visibility graphs Image Representation

---

DJERAOUI Houda & MALEK Mohamed Sidali

Under the supervision of **Prof. LALEG Taous Meriem** and **Prof.BOUSBIA-SALAH Hicham**

Presented and publicly defended on : **(21/06/2025)**

Jury Members :

President :	Pr. ADNANE Mourad	ENP
Supervisor :	Pr.LALEG Taous Meriem	INRIA-Saclay, Paris
Co-Supervisor :	Pr.BOUSBIA SALAH Hicham	ENP
Examiner :	Pr. TOUHAMI Rachida	ENP

ENP 2025

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

ÉCOLE NATIONALE POLYTECHNIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Institut national de recherche en informatique et en automatique

Département d'électronique

End of Studies Project

Submitted for the fulfillment of State Engineer Degree in Electronics

---

PPG signals-based Arterial Stiffness estimation using visibility graphs Image Representation

---

DJERAOUI Houda & MALEK Mohamed Sidali

Under the supervision of **Prof. LALEG Taous Meriem** and **Prof. BOUSBIA SALAH Hicham**

Presented and publicly defended on : **(21/06/2025)**

Jury Members :

President :	Pr. ADNANE Mourad	ENP
Supervisor :	Pr.LALEG Taous Meriem	INRIA-Saclay, Paris
Co-Supervisor :	Pr.BOUSBIA SALAH Hicham	ENP
Examiner :	Pr. TOUHAMI Rachida	ENP

ENP 2025

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

ÉCOLE NATIONALE POLYTECHNIQUE

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Institut national de recherche en informatique et en automatique

Département d'électronique

Mémoire de projet de fin d'études

Pour l'obtention du diplôme d'ingénieur d'état en électronique

---

Estimation de la rigidité artérielle à partir de signaux PPG via des représentations en image basées sur les graphes de visibilité

DJERAOUI Houda & MALEK Mohamed Sidali

Sous la direction de **Prof. LALEG Taous Meriem** et **Prof. BOUSBIA SALAH Hicham** ENP

Présenté et soutenu publiquement le : **(21/06/2025)**

Composition du jury:

Président:	Pr. ADNANE Mourad	ENP
Promotrice:	Pr.LALEG Taous Meriem	INRIA-Saclay, Paris
Co-Promoteur:	Pr.BOUSBIA SALAH Hicham	ENP
Examinatrice:	Pr. TOUHAMI Rachida	ENP

ENP 2025

ملخص

يقدم هذا العمل طريقة غير جراحية لتقدير سرعة موجة النبض (PWV) باستخدام إشارات (PPG) من خلال تحويل مخطط الرؤية والتعلم الآلي. يتم استخراج خصائص معقدة واختيارها وفقًا لنهج متعدد المعايير. تعتمد الدراسة على بيانات محاكاة (In-Silico) وبيانات سريرية حقيقية (VitalDB) . تسمح الجمع بين أنواع مختلفة من الخصائص بتمثيل أكثر شمولاً لبنية إشارات (PPG) . وتُظهر النتائج قدرة الطريقة على تقدير صلابة الشرايين بدقة وموثوقية.

الكلمات المفتاحية : سرعة موجة النبض، مخطط الرؤية، التحليل الضوئي للتغيرات الحجمية للدم ، معالجة الإشارة، التعلم الآلي، معالجة الصور.

Résumé

Ce travail propose une méthode non invasive d'estimation de la vitesse de l'onde de pouls (PWV) à partir des signaux PPG, en utilisant la transformation en graphe de visibilité et l'apprentissage automatique. Des caractéristiques complexes sont extraites et sélectionnées via une approche multi-critères. L'étude repose sur des données simulées (In-Silico) et des données cliniques réelles (VitalDB). La combinaison de plusieurs types de caractéristiques permet de représenter de manière plus complète la structure des signaux PPG. Les résultats obtenus montrent la capacité de la méthode à estimer la rigidité artérielle de manière précise et robuste.

Mots clés : Vitesse de l'onde de pouls, Graphe de visibilité, Photopléthysmogramme, Traitement du signal, Apprentissage automatique, Traitement d'images.

Abstract

This work presents a non-invasive method for estimating Pulse Wave Velocity (PWV) from PPG signals using visibility graph transformation and machine learning. Complex features are extracted and selected through a multi-criteria approach. The study is based on simulated data (In-Silico) and real clinical data (VitalDB). The combination of different types of features allows a more comprehensive representation of the PPG signal structure. The results demonstrate the method's ability to estimate arterial stiffness accurately and robustly.

Keywords : Pulse Wave Velocity, Visibility Graph, Photoplethysmogram, Signal processing, Machine-Learning, Image Processing.

Acknowledgments

Before anything at all, we would love to express our sincere gratitude towards our project supervisor Pr.Taous Meriem Laleg, Professor and permanent researcher with INRIA-Saclay, for her constant guidance and support, for her infinite kindness, gentleness and presence at all times, Her wealth of knowledge and infectious enthusiasm for engineering and mathematics consis tently inspire admiration and motivation, The experience we acquired under your supervision in research is invaluable. we are incredibly fortunate to have had the privilege of learning under your guidance.

We would like to express once again our sincere gratitude towards our Co-supervisor Pr.BOUSBIA SALAH Hicham, Professor with Ecole Nationale Polytechnique for his supervision, flexibility and constant love and support towards this projects. One of the wisest individuals we have ever met in our life. The amount of knowledge and kindess he shared us. All of that won't be forgettable.

We would like to thank the President of the jury Pr. ADNANE Mourad, Professor with Ecole Nationale Polytechnique and the Examiner Pr. TOUHAMI Rachida, Professor Researcher with Ecole Nationale Polytechnique for agreeing to be members of the reading committee and for their constructive analysis of the present work. A special thanks is dedicated to them for their time and teaching and support for the last three years.

DJERAOUI Houda and MALEK Mohamed Sidali.

Dedication

To my mother and my sister *— the strongest pillars of my life. Mom, for your infinite love, your constant light, your boundless heart. And you, my sister, my gentle ally through every season — quiet or radiant, always present. Thank you for your strength, your laughter, your presence, your gaze that comforts me. You are the roots of my courage, my greatest sources of inspiration. I admire you deeply, and every day I hope to become, even just a little, like you.*

To my father *— who is no longer here, yet lives on within me in every decision, every effort, every accomplishment. This thesis is also a quiet tribute to your memory, a flame I carry in my heart.*

To my childhood friends: Yousra, Chiraz, and Nouha *— golden laughter, shared tears, a dance in rhythm with the years. You are the first colors on the canvas of my friendships — the oldest witnesses of every version I've ever been. You've seen me grow, fall, evolve, and rebuild — and you've stayed.*

To my university friends: *Asma, Ines, Inès, Bouchra — your voices, your glances, your shoulders carried me through long study days and endless nights of doubt. You gave true meaning to the word sisterhood.*

To my Electronics Class of 2025: *Adel, Serine, Malek, Ilhem, Ouarda — my comrades in chaos and triumph. Together we turned sleepless nights into memories, and even the most infernal lab sessions into shared jokes. We got through it, hand in hand. This journey was ours.*

To my cat Winnie *— my little whiskered nightlight. You curled up beside me through every exam, every deadline. In the silence of those nights, your warmth soothed me when I couldn't find the words. Thank you for purring through my worries.*

— Houda

Dedication

To my beloved parents, for their unconditional love, support, and sacrifices that have shaped the person I am today.
To my dear brothers and sisters, for always believing in me and standing by my side.
To my classmates in electronics, for their friendship and the unforgettable journey we shared during these academic years — especially the Eurobot team, Bras-del and my partner in the final academic project DJERAOUI.
And to my colleagues at the company, for their collaboration, encouragement, and trust throughout our professional experience together.

— *MALEK*

Contents

List of Tables

List of Figures

List of Acronyms

- **AIx** : Augmentation Index
- **baPWV** : Brachial–Ankle Pulse Wave Velocity
- **cfPWV** : Carotid–Femoral Pulse Wave Velocity
- **CVDs** : Cardiovascular Diseases
- **EBM** : Explainable Boosting Machine
- **LPVG** : Limited Penetrable Visibility Graph
- **PPG** : Photoplethysmography
- **PTT** : Pulse Transit Time
- **PWV** : Pulse Wave Velocity
- **SCSA** : Semi-Classical Signal Analysis
- **TAS** : Texture Analysis Statistics
- **VG** : Visibility Graph
- **WPD** : Wavelet Packet Decomposition
- **WVG** : Weighted Visibility Graph

General Introduction

"To understand is to perceive patterns."
— Isaiah Berlin

Cardiovascular diseases are the leading cause of death globally, accounting for a significant proportion of premature mortality and chronic disability. As healthcare systems increasingly emphasize preventive care and early diagnosis, the need for effective, non-invasive tools to monitor cardiovascular health has become more urgent. Among the many indicators of vascular health, arterial stiffness has emerged as a key biomarker, offering valuable insight into vascular aging and the risk of cardiovascular events such as myocardial infarction, stroke, and heart failure.

Arterial stiffness refers to the reduced elasticity of arterial walls, which impairs the ability of blood vessels to expand and contract in response to pressure changes. This phenomenon is primarily driven by aging but is also exacerbated by conditions such as hypertension, diabetes, and chronic inflammation. Clinically, increased arterial stiffness is associated with elevated systolic blood pressure, early return of reflected pulse waves, and increased cardiac workload, all of which contribute to end-organ damage.

The most widely accepted non-invasive measure of arterial stiffness is the pulse wave velocity (PWV), which quantifies the speed at which the pressure wave generated by the heartbeat travels through the arterial system. Higher PWV values are indicative of stiffer arteries and are strongly correlated with increased cardiovascular risk. Another widely used index is the augmentation index (AIx), which quantifies the contribution of reflected waves to the central aortic pressure waveform.

Despite their diagnostic value, traditional measurement techniques for PWV often require specialized equipment and trained personnel, limiting their accessibility in large-scale clinical practice. This has led to increasing interest in alternative, more widely deployable technologies.

In this context, photoplethysmography (PPG) has gained considerable attention as a convenient, non-invasive, and low-cost technique for cardiovascular monitoring. PPG is an optical method that measures variations in blood volume using peripheral sensors such as those found in wearable devices. The morphology of the PPG waveform encodes valuable information about arterial compliance, wave reflection, and vascular aging.

Nevertheless, direct estimation of PWV from PPG signals remains a challenging task due to various sources of variability, including noise, inter-individual physiological differences, and limited ability of traditional analysis methods to capture the complexity of the underlying signals. To overcome these challenges, advanced computational approaches that combine signal processing, visibility graph theory, and machine learning have been explored.

Main Contribution

This project makes several significant contributions to the field of biomedical signal processing, particularly in the non-invasive estimation of arterial stiffness using real clinical photoplethysmography (PPG) signals. The main added value of this research lies in the innovative application and validation of visibility graph-based feature extraction combined with machine learning models on real-world clinical data from the VitalDB database, an aspect that has received very limited attention in prior work. The contribution relies on several key points:

- The integration and utilization of the **VitalDB real-world clinical database**.
- The transformation of PPG time-series signals into visibility graph images.
- Evaluation for the performance.

Organization of the Thesis:

- **Chapter 1** introduces the fundamental biomedical concepts related to cardiovascular physiology, arterial stiffness, pulse wave dynamics, and the role of PPG in vascular assessment.
- **Chapter 2** presents a comprehensive review of existing methods for PWV estimation from PPG signals, including classical time-domain, frequency-domain, entropy-based methods, and recent advances leveraging machine learning and graph-based signal transformations.
- **Chapter 3** details the proposed methodology, describing the datasets used (In-Silico and VitalDB), the visibility graph transformation, feature extraction techniques, feature selection pipeline, and machine learning modeling strategy.
- **Chapter 4** presents and discusses the experimental results obtained across different training configurations, highlighting the improvements achieved through mixed datasets and augmentation strategies.

Finally, we end this project with a general conclusion and future perspectives.

Chapter 1

Physiological Foundations of Arterial Stiffness Assessment

1.1 Introduction

Understanding the physiological and mechanical principles that govern the cardiovascular system is essential for evaluating arterial function and stiffness. This chapter introduces the fundamental concepts necessary to contextualize our work on non-invasive estimation of arterial stiffness. We begin with an overview of cardiovascular anatomy and hemodynamics, followed by a detailed look at blood pressure and pulse wave propagation—key elements in the assessment of vascular health. These concepts serve as the scientific backbone of the estimation techniques discussed in the next chapters.

1.2 Cardiovascular Anatomy and Hemodynamics

1.2.1 Overview of the Cardiovascular System

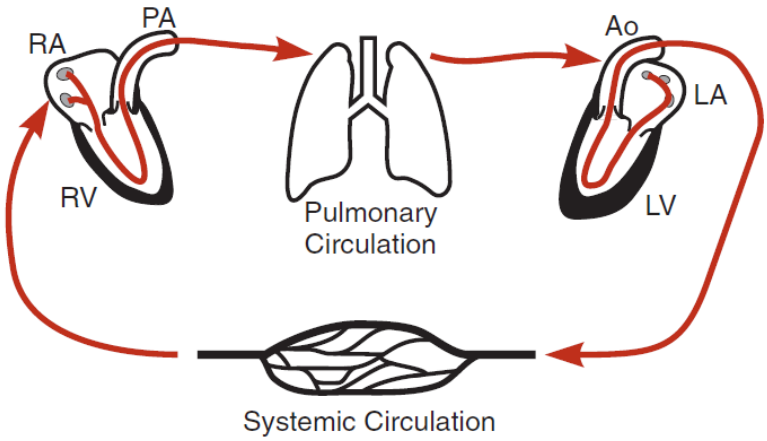


Figure 1.1: Overview of the cardiovascular system. The right side of the heart, pulmonary circulation, left side of the heart, and systemic circulation are arranged in series. RA: right atrium; RV: right ventricle; PA: pulmonary artery; Ao: aorta; LA: left atrium; LV: left ventricle [1].

The cardiovascular system is a closed-loop circuit that includes the heart and a complex network of blood vessels—arteries, veins, and capillaries. Its main role is to transport oxygen, nutrients, hormones, and waste products efficiently throughout the body to maintain homeostasis [2].
At the core of this system lies the heart, a muscular organ located in the thoracic cavity. It consists of four chambers: two atria and two ventricles. The right side of the heart receives deoxygenated blood from the systemic circulation and pumps it to the lungs via the pulmonary arteries for oxygenation. Conversely, the left side receives oxygenated blood from the lungs and propels it into the systemic circulation through the aorta.

The overall flow of blood through the pulmonary and systemic circuits is illustrated in Figure 1.1.

1.2.2 Structural Features of Blood Vessels

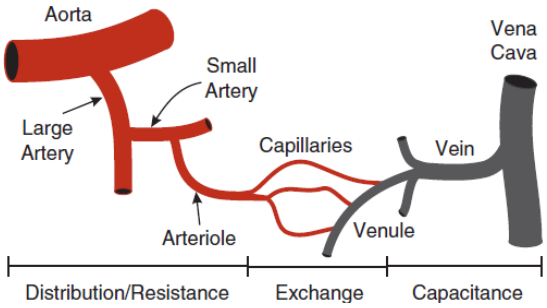


Figure 1.2: Major types of blood vessels found within the circulation.[1]

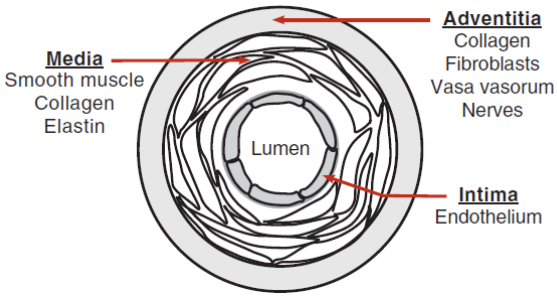


Figure 1.3: Blood vessel components. Blood vessels, except capillaries and small postcapillary venules, are composed of three layers: intima, media, and adventitia.[1]

Blood vessels, as illustrated in Figure 1.2 [1], form an intricate network composed of arteries, veins, and capillaries, each with specific structural and functional characteristics. Arteries, shown in red, are thick-walled and elastic vessels that transport oxygen-rich blood away from the heart under high pressure. Their layered structure—comprising the intima, media, and adventitia as shown in Figure 1.3[1]—provides both strength and flexibility to withstand pulsatile blood flow. As arteries branch into smaller arterioles and eventually into capillaries, the vessel walls become progressively thinner. Capillaries, the smallest blood vessels, consist of a single layer of endothelial cells, allowing for efficient exchange of oxygen, carbon dioxide, nutrients, and metabolic waste between the blood and surrounding tissues. Once this exchange occurs, the blood—now deoxygenated—is collected by venules, which converge into veins. These vessels, shown in blue, have thinner walls and larger lumens than arteries, enabling them to act as low-pressure reservoirs. Many veins, particularly in the limbs, are equipped with valves that prevent the backflow of blood, facilitating its return to the heart despite gravitational challenges.

1.3 Blood Pressure: Definition and Clinical Significance

Having reviewed the structural composition of blood vessels, we now turn our attention to one of the most crucial functional parameters of the cardiovascular system: blood pressure.

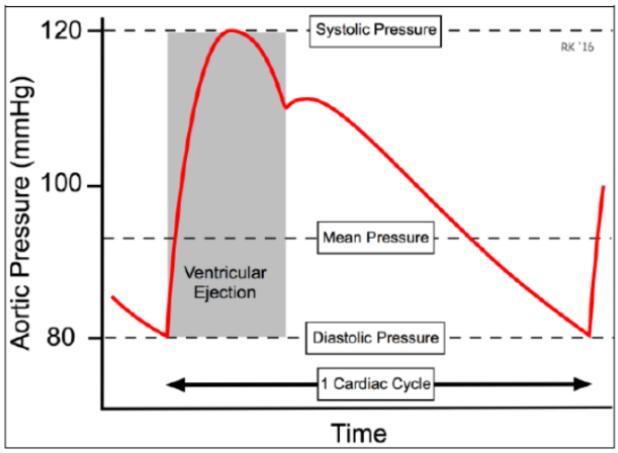


Figure 1.4: Diagram of aortic pressure wave[3]

Blood pressure (BP) is the mechanical force exerted by circulating blood on the inner walls of blood vessels. It is a fundamental physiological parameter that reflects the dynamic interaction between cardiac output and vascular resistance, serving as a vital indicator of cardiovascular health. BP is typically measured in millimeters of mercury (mmHg) using either a sphygmomanometer or automated devices.

BP fluctuates throughout the cardiac cycle, which generates two primary pressure phases: systolic blood pressure (SBP) and diastolic blood pressure (DBP). SBP represents the peak arterial pressure during ventricular contraction (systole), when the left ventricle ejects blood into the aorta, indicating the maximum force exerted on arterial walls. Conversely, DBP reflects the minimum arterial pressure during ventricular relaxation (diastole), when the heart refills with blood, maintaining baseline pressure that ensures continuous blood flow between beats.

A typical BP value is around 120/80 mmHg, where the first number corresponds to SBP and the second to DBP. Deviations from this range—particularly elevated BP or hypertension—are clinically significant. Hypertension is a major risk factor for cardiovascular diseases such as stroke, myocardial infarction, and heart failure, and it contributes to the pathogenesis of arterial stiffness [4].

Figure 1.4[3] illustrates the aortic pressure waveform during one cardiac cycle, highlighting the systolic, diastolic, and mean pressures, as well as the ventricular ejection phase.

1.4 Pulse Wave Dynamics

Blood pressure is not merely a static measure; it also propagates as dynamic signals known as pulse waves. Understanding these waves helps reveal how the heart interacts with the arterial network.

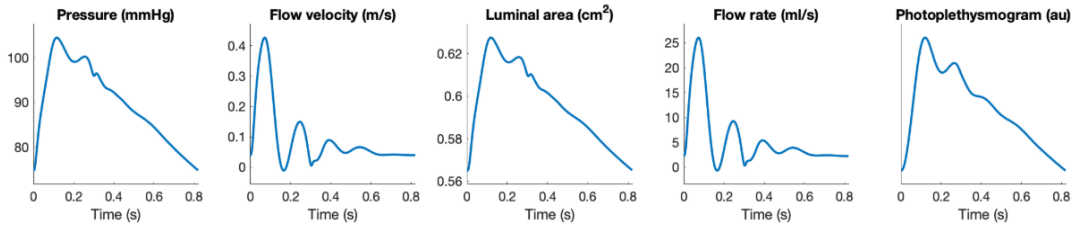


Figure 1.5: Different types of pulse waves: Pressure, flow velocity, luminal area, flow rate and photoplethysmogram (PPG) pulse.[5]

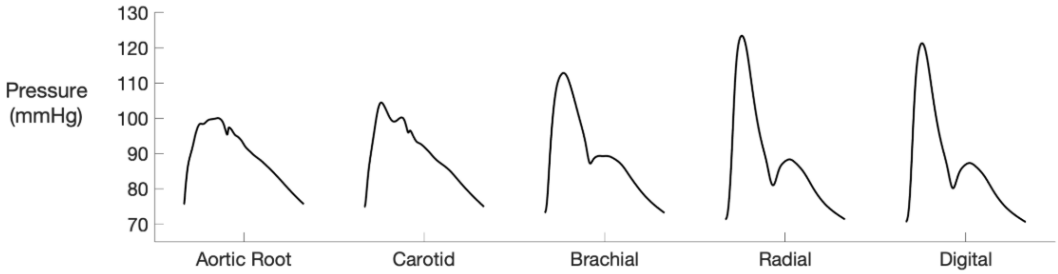


Figure 1.6: Pulse waves at different anatomical sites.[5]

After the heart contracts during systole, the sudden rise in blood pressure pushes a pulse wave into the arteries. This wave is created when blood is ejected from the heart into the aorta and moves much faster than the blood itself. Unlike steady blood flow, the pulse wave is a pressure signal that shows how the heart's output interacts with the flexibility of the arteries.

The way this wave moves through the arteries depends on how elastic, wide, and shaped the arteries are. In healthy and flexible arteries, some of the wave's energy is absorbed, making the wave smoother and delaying the reflections. In stiff arteries, the wave moves faster and the reflected waves come back sooner, which changes the shape of the pressure wave. These reflections usually come from points where the artery branches or where the wall stiffness changes [6].

Figure 1.5[5] shows different types of pulse waveforms such as pressure, flow velocity, luminal area, flow rate, and the photoplethysmogram (PPG). These help us understand how the wave behaves in the body. Figure 1.6[5] shows how the pressure wave shape changes from the central arteries (like the aorta) to more distant ones (like the digital arteries). This happens because of the wave reflections and increasing stiffness as the wave travels. Together, these images help us see how pulse waves move and change through the arterial system.

1.5 Arterial Stiffness: Mechanisms and Clinical Relevance

Arterial stiffness refers to the loss of arterial wall elasticity, commonly associated with aging, hypertension, and metabolic disorders. As arteries stiffen, they lose their ability to buffer the pressure generated by cardiac ejection, which leads to increased systolic blood pressure and earlier wave reflections. Over time, this contributes to vascular damage, left ventricular hypertrophy, and higher cardiovascular risk [7].

Arterial stiffness is a crucial marker of vascular health and is commonly assessed using hemodynamic indices that reflect arterial wall properties. Among these indices, the **Pulse Wave Velocity (PWV)** and the **Augmentation Index (AIx)** are the most established and clinically utilized measures.

1.5.1 Pulse Wave Velocity (PWV)

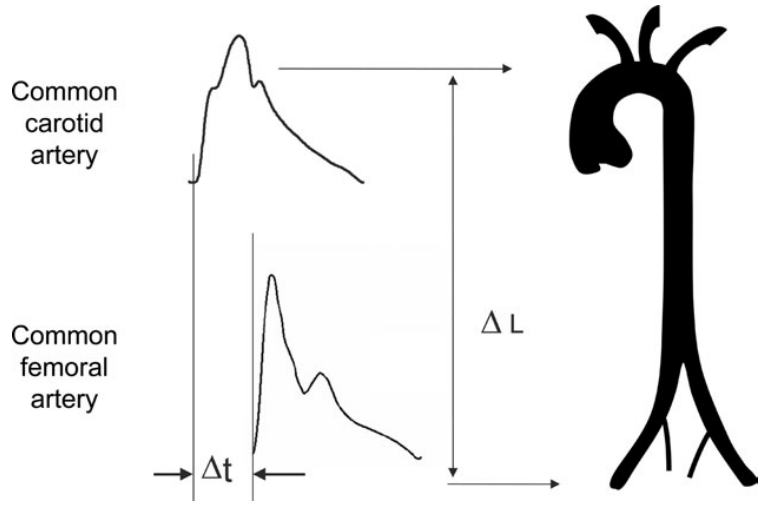


Figure 1.7: Measurement of carotid-femoral PWV using the foot-to-foot method [7]

Pulse Wave Velocity (PWV) quantifies the speed at which the pressure wave generated by the heartbeat travels along the arterial tree, typically measured between two arterial sites such as the carotid and femoral arteries. It is widely recognized as the gold standard for evaluating regional (aortic) arterial stiffness [7], [8].

Physiologically, PWV provides a direct estimate of arterial wall elasticity: increased arterial stiffness results in faster propagation of the pulse wave; thus, higher PWV values correspond to reduced arterial compliance.

As shown in Figure 1.7 [7], the measurement of carotid–femoral PWV is performed using the foot-to-foot method. This method involves recording pulse waveforms at the carotid and femoral arteries and calculating the time delay between the arrival of the pulse wave's "foot" at these two sites. The distance D between the recording points is measured over the body surface to approximate the arterial path.

Mathematically, PWV is calculated using Equation (1.1), where D is the distance between the two recording sites, and Δt is the pulse transit time (the time it takes for the pressure wave to travel from the proximal to the distal site)[7]

$$\text{PWV} = \frac{D}{\Delta t} \tag{1.1}$$

PWV is a clinically valuable biomarker because elevated values are associated with increased cardiovascular risk, including hypertension, atherosclerosis, and adverse cardiovascular events [8].

1.5.2 Augmentation Index (AIx) and Wave Reflection

Another important parameter reflecting arterial stiffness and vascular function is the **Augmentation Index (AIx)**, which is derived from the analysis of central blood pressure waveforms.

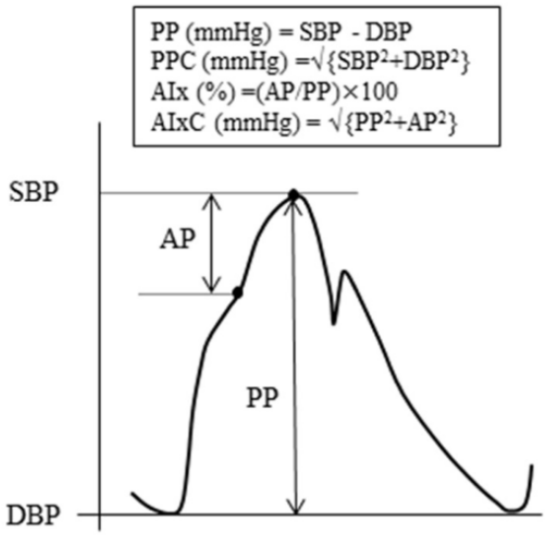


Figure 1.8: Central blood pressure waveform illustrating systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse pressure (PP), and augmentation pressure (AP) used to compute the Augmentation Index (AIx).[9]

Figure 1.8[9] illustrates a typical central arterial pressure waveform featuring two main systolic peaks:

The *first peak* corresponds to the initial pressure wave generated by left ventricular ejection during systole. The *second peak* arises from the reflected pressure wave returning from peripheral sites such as smaller arteries and arterioles.

The vertical difference between these two peaks is termed the **augmentation pressure** (AP), which reflects the magnitude of wave reflection.

Pulse pressure (PP), defined as the difference between systolic blood pressure (SBP) and diastolic blood pressure (DBP), serves as the baseline for normalization.

The Augmentation Index (AIx) quantifies the relative contribution of the reflected wave to the overall pulse pressure and is calculated using the equation (1.2) [7]:

$$\text{AIx} = \frac{AP}{PP} \times 100 \tag{1.2}$$

Expressed as a percentage, a higher AIx indicates greater wave reflection and generally corresponds to increased arterial stiffness or impaired vascular function.

1.6 Measurement Techniques for Pulse Wave Velocity (PWV)

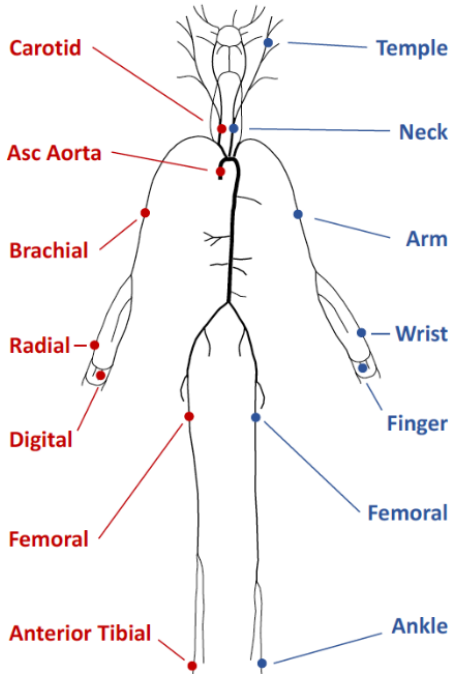


Figure 1.9: Major arterial sites used for pulse wave velocity (PWV) measurement. Carotid and femoral arteries are used for Carotid–Femoral PWV (cfPWV), brachial and ankle arteries for Brachial–Ankle PWV (baPWV), and peripheral arteries such as radial and digital sites are used in Photoplethysmography-based PWV (PPG-PWV) methods [5].

PWV can be measured through both invasive and non-invasive techniques.

Invasive assessment involves direct intraarterial pressure measurements, typically through catheterization of central arteries such as the ascending aorta. This approach provides high-fidelity waveform data and is considered the reference standard. However, due to their complexity, risks and cost, invasive methods are limited to specialized settings and impractical for routine use [7].

To overcome these limitations, non-invasive methods have been developed for safer and more accessible evaluation of arterial stiffness. Pulse wave velocity (PWV) is measured at several arterial sites, as illustrated in the figure. The carotid and femoral arteries, located in the neck and groin, respectively, are used for Carotid–Femoral PWV (cfPWV), which assesses central arterial stiffness by measuring pulse wave transit through the aorta. Elevated cfPWV is strongly correlated with cardiovascular risk [8].

The figure 1.9 [5] also shows the brachial artery at the upper arm and the ankle artery, which define the pathway for Brachial–Ankle PWV (baPWV). This method reflects the stiffness of both the central and peripheral arteries and is widely used in clinical screenings by oscillometric devices, although its specificity for aortic stiffness is somewhat limited [10], [11].

Peripheral arteries such as the radial and digital arteries, also depicted in the figure, are employed in Photoplethysmography-based PWV (PPG-PWV). These non-invasive methods use infrared light to monitor changes in microvascular blood volume. Despite challenges such as motion artifacts, advances in signal processing have improved their accuracy and practical applicability [12], [13].

While cfPWV remains the clinical gold standard and baPWV is widely accessible, their equipment requirements restrict use outside clinics. PPG-PWV, with its wearable optical sensors, offers a real-time, comfortable, and cost-effective alternative well-suited for novel arterial stiffness estimation methods like visibility graph analysis.

While cfPWV remains the clinical gold standard and baPWV is widely accessible, their equipment requirements restrict use outside clinics. PPG-PWV, with its wearable optical sensors, offers a real-time, comfortable, and cost-effective alternative well-suited for novel arterial stiffness estimation methods like visibility graph analysis.

1.7 Photoplethysmography (PPG) for Vascular Assessment

Photoplethysmography (PPG) is a widely used noninvasive optical technique for monitoring cardiovascular parameters by detecting volumetric changes in blood circulation. The method involves illuminating tissue with a light source, typically in the red or near-infrared spectrum, and measuring variations in the intensity of light transmitted through or reflected from the tissue microvasculature [14], [15]. These intensity changes correspond to the pulsatile blood flow induced by cardiac cycles, enabling extraction of critical cardiovascular metrics including heart rate, blood oxygen saturation (SpO_2), and more recently, vascular stiffness indices [16].

Photoplethysmography measurements are commonly performed at peripheral sites such as the fingertip or earlobe, where the microvascular bed is accessible. Figure 1.10[17] illustrates the typical setup of a PPG sensor in transmission mode, where light from red and infrared LEDs penetrates tissue and is partially absorbed by blood before being detected by a photodiode positioned opposite the LEDs. The variations in light intensity detected correspond to changes in pulsatile blood volume synchronized with the cardiac cycle [16].

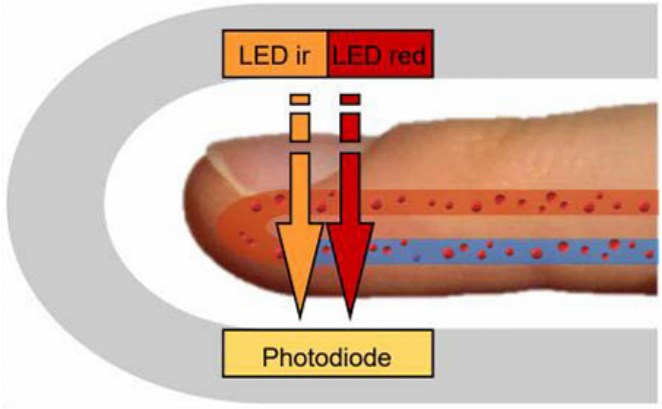


Figure 1.10: Schematic of two LEDs transmitting light through a finger, which is then captured by the photodiode[17]

1.7.1 PPG Signal Morphology and Physiological Significance

The PPG waveform is characterized by periodic pulses representing the cardiac cycle and is composed of two main components: a pulsatile (AC) component related to arterial blood volume changes, and a non-pulsatile (DC) component arising from venous blood, tissues, and other static absorbers [18]. The morphology of each pulse contains distinct features corresponding to physiological events (see Figure 1.11):

- **Systolic foot:** The onset of the pulse that indicates ventricular ejection.
- **Systolic peak:** The highest amplitude indicating maximal arterial expansion.
- **Dicrotic notch:** Reflects aortic valve closure and the beginning of diastole.
- **Diastolic peak:** A secondary peak caused by reflected pressure waves from peripheral sites.

These features enable detailed vascular assessment, where the relative timing and amplitude provide insights into arterial compliance and peripheral resistance. Notably, diminished or altered features such as the disappearance of the dicrotic notch are associated with increased arterial stiffness and cardiovascular pathology [16], [19].

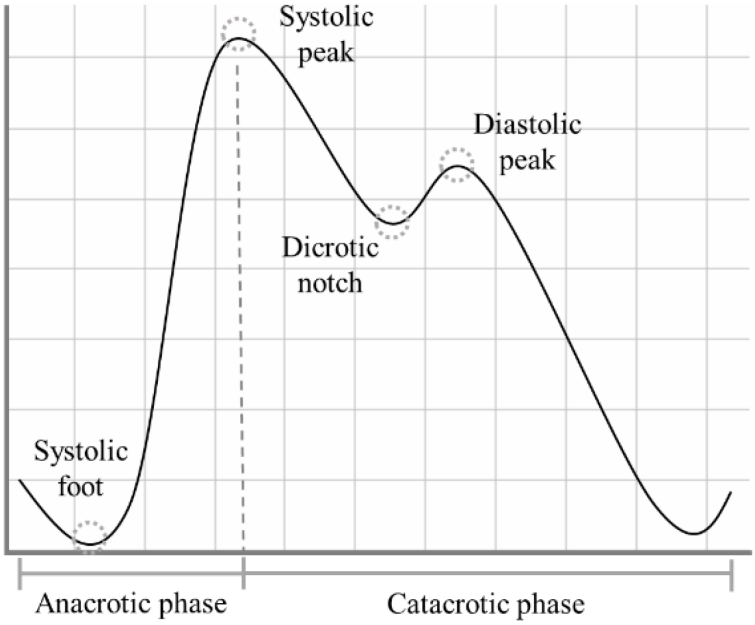


Figure 1.11: PPG pulse morphology showing key features of the cardiac cycle. Adapted from [18].

1.7.2 Instrumentation and Signal Processing

PPG systems employ light-emitting diodes (LEDs) as the light source and photodiodes (PD) as detectors. The photodiode converts incident light into a photocurrent, which is then amplified, typically via a transimpedance amplifier, to produce a voltage signal proportional to blood volume changes [16]. Figure 1.12 [16] illustrates the fundamental electronic architecture including filtering stages that isolate the pulsatile AC component from the DC baseline, and signal conditioning circuits to improve signal-to-noise ratio.

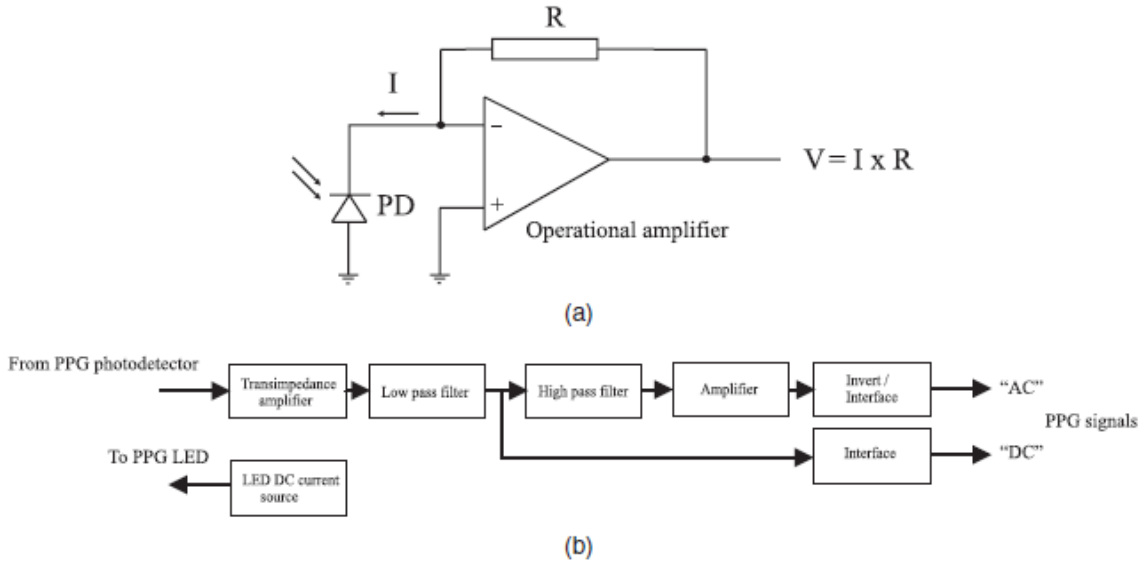


Figure 1.12: Block diagram of typical PPG electronic instrumentation: (a) photodiode with transimpedance amplifier; (b) signal conditioning and separation of AC and DC components.[16].

Due to susceptibility to motion artifacts, ambient light interference, and physiological variability, advanced signal processing techniques such as adaptive filtering, wavelet decomposition, and machine learning-based noise reduction have been incorporated in modern PPG devices, especially wearable sensors [20], [21].

1.7.3 PPG in Vascular Stiffness Assessment

Recent studies have shown that vascular indices derived from PPG signals are strongly correlated with gold-standard PWV measurements. This supports the growing use of PPG as a noninvasive tool for the early detection of vascular aging and for the stratification of cardiovascular risk [11]. In addition, the integration of artificial intelligence techniques into PPG waveform analysis has significantly improved the accuracy and robustness of arterial stiffness estimation in ambulatory and wearable settings [15], [21].

Furthermore, advanced signal analysis of the PPG waveform has made it possible to assess additional vascular parameters, such as endothelial function and microvascular health, offering a more comprehensive evaluation of cardiovascular status [15], [19]. These advanced analytical approaches and their implementation in arterial stiffness estimation will be explored in greater detail in the next chapter.

1.8 Conclusion of Chapter 1

This chapter laid the essential groundwork for understanding how arterial properties influence cardiovascular dynamics. By clarifying the roles of blood pressure and pulse waves, we established a framework for analyzing vascular behavior beyond basic hemodynamic measures.

These principles now provide the necessary context to critically examine current methods used to assess arterial stiffness. The following chapter reviews these techniques and highlights the advances shaping the field today.

Chapter 2

State-of-the-Art methods on PWV estimation using PPG signal

2.1 Introduction

As discussed in the previous chapter, arterial stiffness is a major determinant of cardiovascular risk, and its early assessment is crucial for preventive medicine. While traditional hemodynamic measures such as blood pressure provide indirect insights, Pulse Wave Velocity (PWV) has emerged as the most reliable and biomarker to quantify arterial stiffness.

In particular, carotid-femoral PWV (cf-PWV) is considered the clinical gold standard. It has been validated as a strong and independent predictor of adverse health outcomes, including cardiovascular disease, renal dysfunction, cognitive decline, and all-cause mortality. A large-scale meta-analysis involving over 17,000 individuals showed that each one-standard-deviation increase in cf-PWV is associated with a 20–37% increase in the risk of these conditions[22].

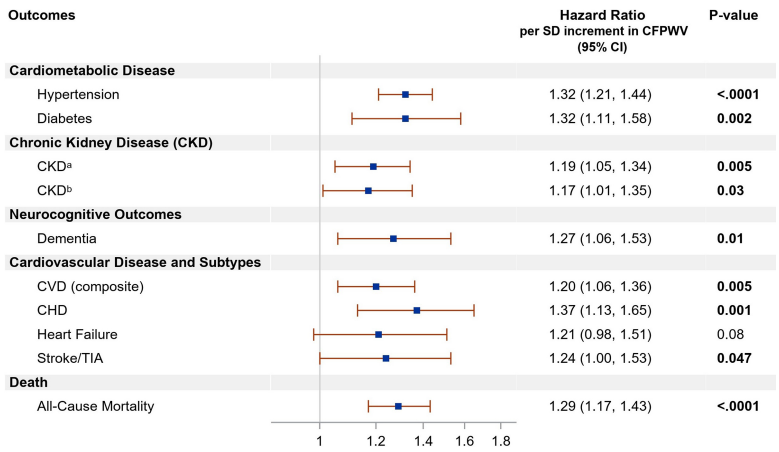


Figure 2.1: Relations of carotid-femoral pulse wave velocity (cf-PWV) and incidence of outcome events dichotomized by median age [22].

However, cf-PWV measurement still relies on specialized equipment and trained personnel, limiting its use in routine or large-scale settings. This limitation has prompted interest in photoplethysmography (PPG)-based methods, which offer a more accessible, wearable, and cost-effective alternative. Yet, extracting accurate and physiologically meaningful information from PPG signals remains a complex task due to their nonlinear and noise-prone nature. This has driven the exploration of richer signal representations and adaptive modeling techniques.

This chapter critically reviews existing approaches for estimating arterial stiffness from PPG signals. It moves from conventional time- and frequency-domain methods to recent techniques involving mathematical transformations and machine learning. It sets the stage for the visibility-graph-based method introduced in the next chapter.

2.2 Traditional Methods for PWV Estimation from PPG Signals

Traditional methods for estimating Pulse Wave Velocity (PWV) from photoplethysmography (PPG) signals mainly fall into two categories: time-domain methods and frequency-domain methods. These approaches are based on easily identifiable features in the PPG waveform and have been widely used due to their simplicity and physiological relevance.

2.2.1 Time-Domain Methods

Time-domain methods focus on detecting fiducial points such as the systolic peak, diastolic foot, and dicrotic notch on pulsatile waveforms. A key metric derived from these landmarks is the Pulse Transit Time (PTT), which represents the time it takes for the pressure wave to travel between two arterial sites. PTT can be estimated using various signal combinations, such as the delay between the ECG R-wave and a point on the PPG waveform [16], as shown in Figure 2.2, or between two peripheral signals like carotid and femoral pressure waveforms. Regardless of the modality, PTT is inversely related to Pulse Wave Velocity (PWV), and therefore serves as an important surrogate marker of arterial stiffness, as discussed in **Chapter 1**.

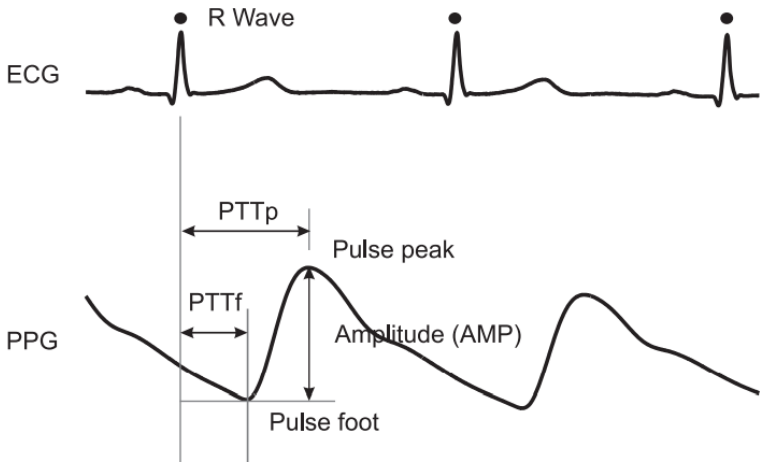


Figure 2.2: Measurement of PTT from ECG R-wave to foot and peak of PPG waveform. [16]

To better identify these fiducial points, various studies apply the first, second, or even third derivatives of the PPG signal [23].
Figure 2.3 shows how these derivatives help in detecting the precise timing of waveform features.

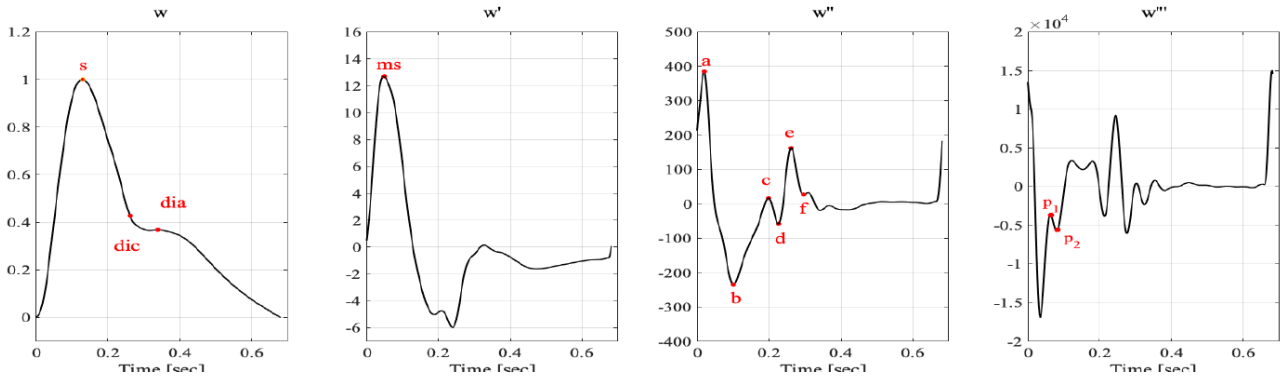


Figure 2.3: Fiducial point detection using raw and derivative PPG signals[23].

Although these methods are relatively easy to use and interpret, they have limitations. They often require synchronized ECG and PPG recordings, are sensitive to noise and motion, and reduce complex waveforms to a few scalar values that may not capture the full signal dynamics.

2.2.2 Frequency-Domain Methods

In frequency-domain analysis, the PPG signal is studied by examining how its energy is distributed across different frequencies. This is done using tools like Power Spectral Density (PSD) or Fourier Transform, which convert the signal into a frequency spectrum. The goal is to identify rhythms that reflect the autonomic nervous system's regulation of blood vessels.

The two main frequency bands commonly studied are:

- **Low-Frequency (LF) band:** 0.04–0.15 Hz, associated with sympathetic activity.
- **High-Frequency (HF) band:** 0.15–0.4 Hz, associated with parasympathetic activity.

As shown in Figure 2.4, a typical Power Spectral Density (PSD) plot illustrates how signal energy is distributed across very low frequency (VLF), LF, and HF bands for ECG, PPG, and remote PPG (rPPG) signals. These distributions reflect the activity of the autonomic nervous system.

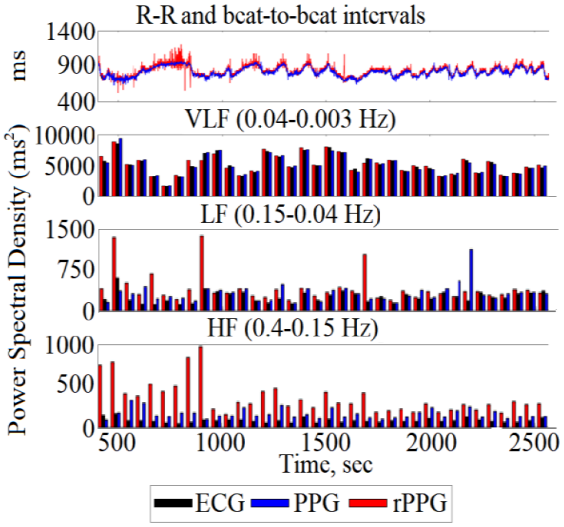


Figure 2.4: Power Spectral Density (PSD) of ECG, PPG, and rPPG signals across VLF, LF, and HF bands [24].

Although these features are not direct measures of PWV, studies have shown that arterial stiffness subtly alters the frequency profile of PPG signals. For example, a study examining frequency domain analysis of PPG and arterial blood pressure (ABP) in postoperative cardiac patients [25] demonstrated changes in spectral power within the high-frequency band, suggesting reduced variability in stiffer arteries pmc.ncbi.nlm.nih.gov .

These frequency-based descriptors can be used in machine learning models to estimate PWV indirectly. They serve as low-dimensional biomarkers of vascular health that reflect stiffness-related changes.

However, frequency-domain analysis assumes signal stationarity—an assumption that often fails in real-world physiological recordings. This limitation can cause important non-linear and time-varying features to be overlooked, reducing the method's accuracy in estimating PWV.

2.3 Modern Methods: Machine Learning and Signal Transformation

To address the shortcomings of handcrafted time or frequency features, recent studies have turned to machine learning (ML) and advanced signal transformations. These methods aim to extract deeper insights from PPG signals—either by learning non-obvious patterns or by re-representing the signal in more expressive forms.

2.3.1 Spectrograms and Scalograms

One effective strategy for analyzing PPG signals is to convert them into 2D time–frequency representations, such as spectrograms. These images capture how the signal's frequency content evolves over time using short-time Fourier transforms, revealing patterns that are difficult to observe in the raw waveform. This transformation is especially useful for identifying subtle features related to arterial stiffness.

Figure 2.5 [26] illustrates a typical pipeline, where PPG and blood pressure signals are processed into spectrograms, from which various features: statistical, energy-based, or derived from semi-classical signal analysis (SCSA)—are extracted. These features are then used to train machine learning models for estimating cf-PWV [26].

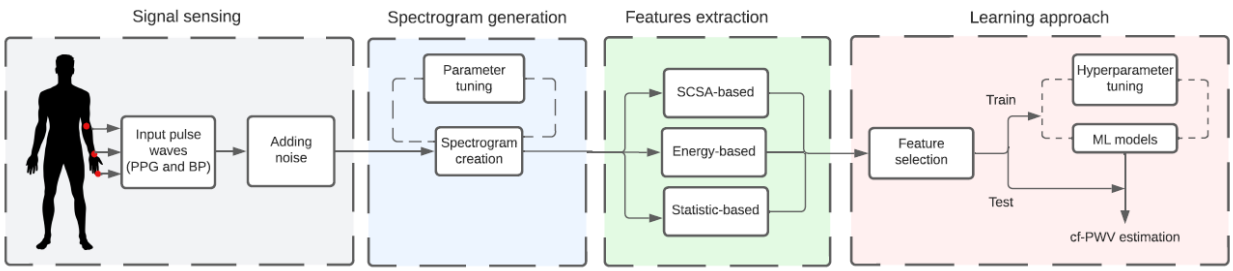


Figure 2.5: Schematic illustration of the estimation pipeline of the carotid-to-femoral pulse wave velocity based on pulse wave images. PPG indicates the photoplethysmography signals, BP the blood pressure signals and ML indicates machine learning models. [26]

2.3.2 Entropy and Wavelet-Based Features

Another set of methods focuses on analyzing the complexity of the PPG signal using nonlinear dynamics. The idea is that healthy arteries produce signals that are more irregular and adaptive, while stiff or aging arteries generate signals that are smoother and more predictable. To capture this, researchers use entropy measures—like sample entropy or multiscale entropy—which calculate how unpredictable the signal is over time. A higher entropy usually means more variability, which is a sign of healthy vascular function. Additionally, wavelet packet decomposition helps analyze different frequency components of the signal while preserving time information, even when the signal is not stable. These features can indirectly reflect changes in arterial stiffness and are useful for estimating PWV in a non-traditional but informative way [27], [28].

2.3.3 Machine Learning Models

Once these features—whether derived from entropy, wavelet analysis, or signal derivatives—are extracted, they are often used to train machine learning models that can predict Pulse Wave Velocity (PWV). These models, which include algorithms like Random Forests, Support Vector Machines, and deep Multilayer Perceptrons (MLPs), learn to identify patterns in the data associated with arterial stiffness. For example, Bahloul et al. [29] used an MLP trained on features such as the velocity and acceleration of the PPG signal (i.e., its first and second derivatives). Similarly, Hellqvist et al. [30] developed models using engineered amplitude ratios and precise timing of fiducial points, feeding them into LASSO and Random Forest models. These ML approaches represent a powerful extension of traditional feature extraction, offering higher performance but often at the cost of interpretability.

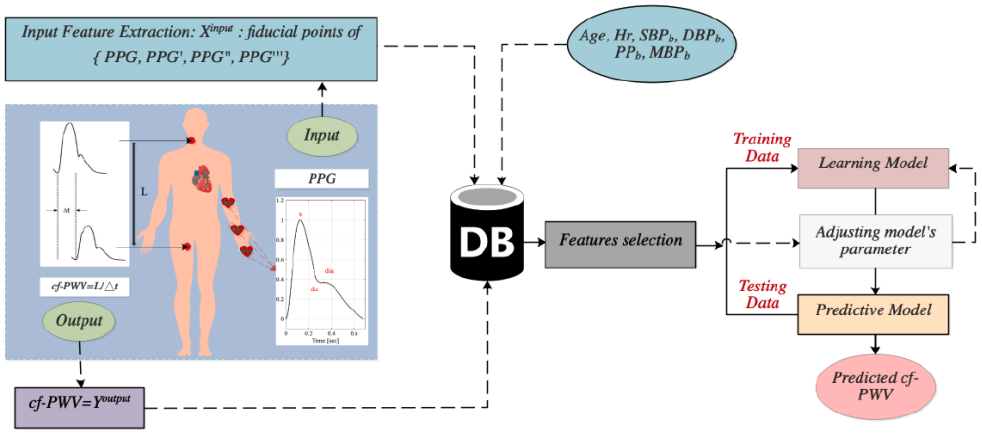


Figure 2.6: Schematic illustration of the estimation pipeline of the carotid-to-femoral pulse wave velocity. The fiducial points from the PPG and its derivatives signals (first derivative: PPG0,second derivative:PPG", and third derivative: PPG000), age, heart rate (Hr), the brachial systolic blood pressure (SBPb), brachial diastolic blood pressure (DBPb), brachial pulse blood pressure (PPb) and brachial mean blood pressure (MBPb) were used as features [29]

Despite their growing success, current approaches often fail to explicitly capture the evolving structure of PPG waveforms in a transparent and interpretable way. This limitation opens the door to alternative frameworks that better preserve temporal dynamics while enabling meaningful feature extraction.

2.4 Visibility Graph Based Modeling of PPG Signals

2.4.1 Introduction to Visibility Graphs

In recent years, visibility graphs have emerged as an innovative tool to analyze complex time series like photoplethysmogram (PPG) signals. The core idea is to convert a time series into a graph by connecting data points that are "visible" to each other based on a geometric rule: two points are connected if a straight line drawn between them does not intersect any intermediate value. This transformation provides a way to retain the signal's temporal structure while encoding its complexity into a topological form.

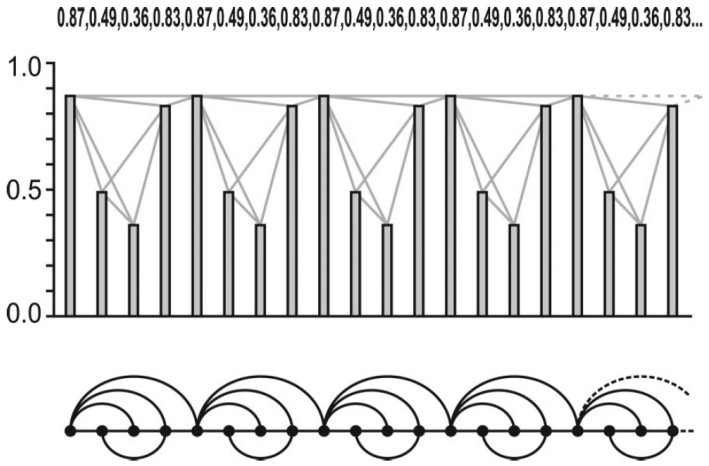


Figure 2.7: Example of a time series and its associated visibility graph. Each node represents a point in the signal, and edges represent visibility relationships according to the visibility graph algorithm [31].

Figure 2.7 illustrates this process using a synthetic time series: each data point becomes a node, and edges are drawn between nodes that satisfy the visibility condition. The resulting graph captures the underlying dynamics of the signal, and standard graph-theoretic tools—such as degree distribution, clustering coefficients, or motif analysis—can then be applied to extract informative features. This makes visibility graphs particularly appealing for biomedical signals, where both structured patterns and irregular fluctuations coexist.

The original algorithm [31] proposes that the mapping retains the temporal structure and encodes the complexity of the signal into topological properties like node degree, clustering, and centrality. The resulting graphs are undirected, connected, and invariant under affine transformations, such as vertical or horizontal rescaling, translation, or trend addition, as illustrated in Figure 2.8

This technique allows the study of different classes of time series:

- Periodic signals become regular graphs,
- Random signals produce graphs with exponential degree distributions,
- Fractal or self-similar signals result in scale-free networks.

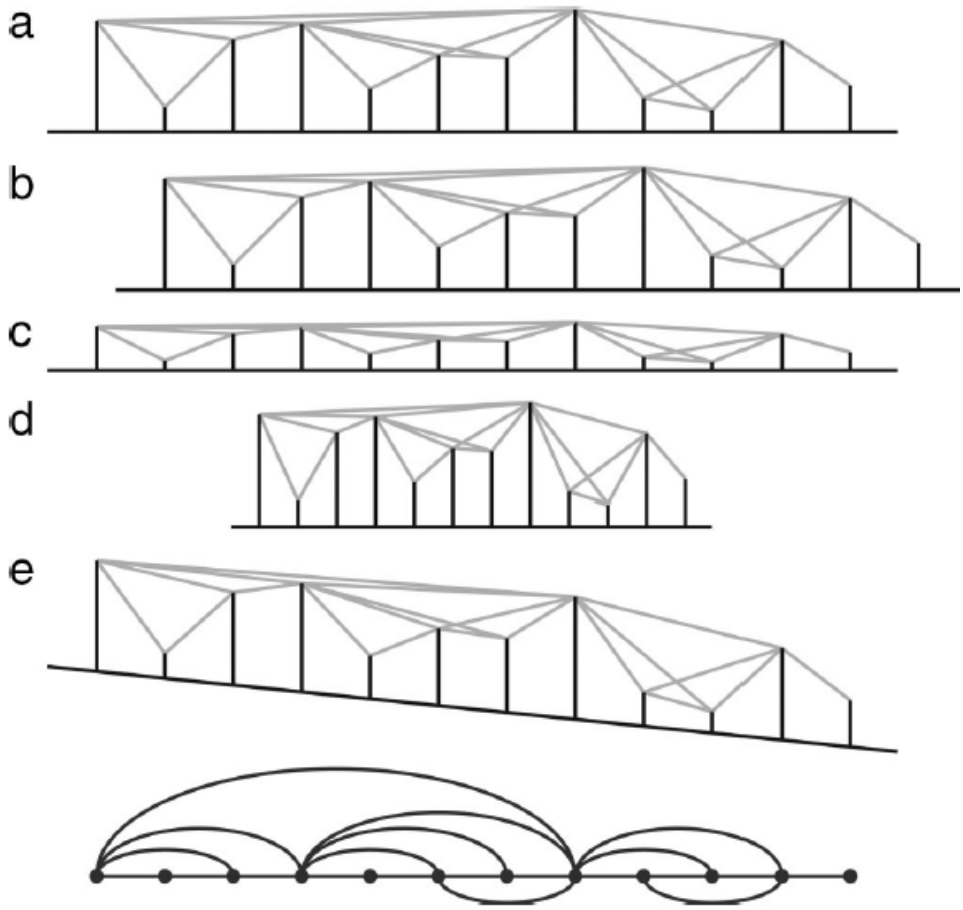


Figure 2.8: Example of a time series and its associated visibility graph. Each node represents a point in the signal, and edges represent visibility relationships according to the visibility graph algorithm. Adapted from Lacasa et al. [31].

This technique has been shown to enhance the accuracy of PWV estimation compared to conventional signal-based approaches. For example, Vargas et al. [32] demonstrated that convolutional neural networks (CNNs) trained on binary adjacency matrices derived from visibility graphs outperformed models based on handcrafted features. In parallel, other studies have explored alternative representations: Mansouri et al. [26] employed CNNs on spectrogram-based images, while Hellqvist et al. [30] used engineered fiducial features combined with classical machine learning models such as LASSO and Random Forest. The visibility graph approach, however, introduces a novel structural perspective to PPG signal modeling. Together, these efforts reflect a broader movement toward explainable and structure-aware AI in biomedical signal analysis.

2.4.2 Applications to PPG Signals

fPPG signals transformed into VG allow extraction of structural metrics like node degree, motif frequency, and shortest path length. These features help characterize the complexity and internal organization of the signal. Figure 2.9 and Figure 2.10 illustrate this transformation process.

Among physiological signals, PPG is particularly suitable for graph-based modeling due to its accessibility and morphological richness. In recent studies, PPG signals have been transformed into visibility graphs, allowing researchers to extract structural features such as node degree, motif frequency, and shortest path length—quantities that reflect the complexity and organization of the underlying signal.

For instance, Vargas et al. [32] applied the limited penetrable visibility graph (LPVG) algorithm to PPG signals, showing that the topology of the resulting graphs encodes information related to cardiovascular health. In their work, each PPG cycle is first segmented and then transformed into a visibility graph. As illustrated in Figure 2.9, each point on the PPG waveform is connected to others based on line-of-sight criteria, resulting in a graph structure that retains temporal information.

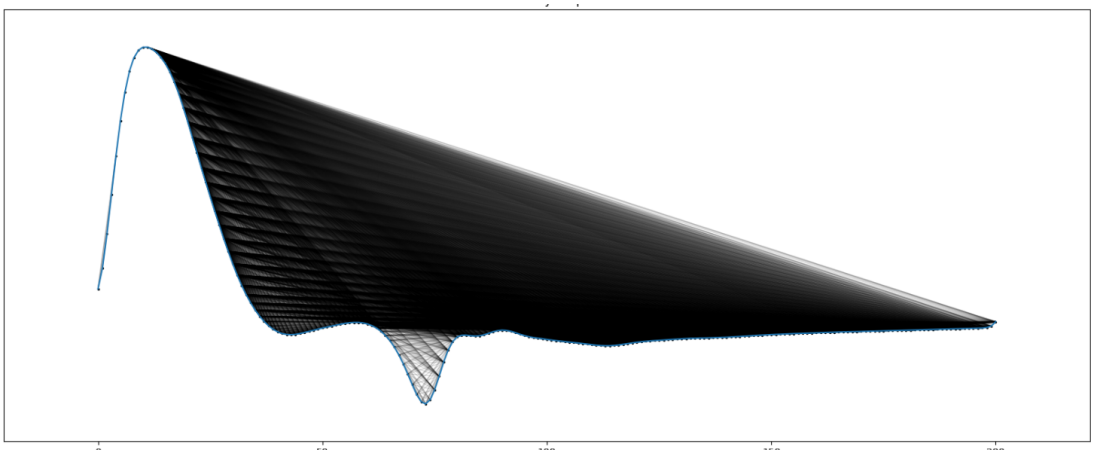


Figure 2.9: Visibility connections projected onto a PPG cycle from the Insilico database, illustrating how graph edges reflect temporal structure[33] .

These visibility graphs can also be represented as adjacency matrices, capturing visibility relations between all data points. These matrices can then be converted into binary images, which are analyzed using convolutional neural networks (CNNs). This hybrid approach merges the structural interpretability of graph theory with the pattern recognition strength of deep learning. Figure 2.9[33] shows an example of such a binary matrix used as CNN input for PWV estimation.

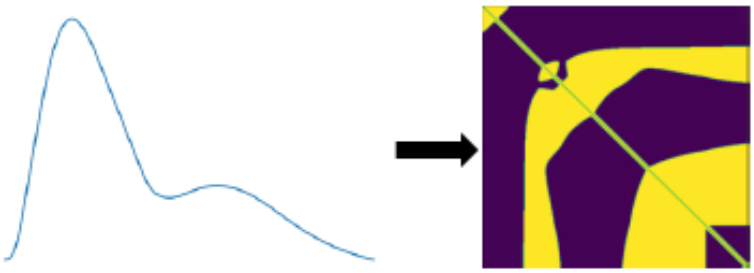


Figure 2.10: Binary adjacency matrix of a visibility graph from a PPG signal, used as CNN input. [32]

This technique has been shown to improve the estimation of PWV compared to traditional signal-based approaches. Vargas et al. [32] demonstrated that CNNs trained on these binary adjacency matrices outperformed conventional feature-based models. While other studies, such as Mansouri et al. [26], explored CNNs applied to spectrogram images, and Hellqvist et al [30] employed fiducial features with classical ML models, the visibility-graph-based approach introduces a novel structural perspective to PPG signal analysis. These methods collectively reflect a broader shift toward explainable and structure-aware AI in biomedical signal processing.

2.5 Conclusion

In this chapter, we explored the main categories of methods used to estimate Pulse Wave Velocity (PWV) from PPG signals, ranging from time- and frequency-domain techniques to more advanced nonlinear and graph-based approaches. While traditional methods such as Pulse Transit Time offer simplicity and physiological interpretability, they suffer from limitations in noise robustness and reliance on additional signals like ECG. Frequency and entropy-based analyses provide richer representations but often rely on assumptions like signal stationarity.

Among all approaches, visibility graphs stand out for their ability to capture both morphological and temporal dynamics of the PPG signal within a graph-theoretic framework. This allows the extraction of structured features that reflect the underlying cardiovascular state, and their compatibility with machine learning pipelines—especially when combined with CNNs—makes them particularly promising for non-invasive PWV estimation.

The next chapter builds upon this foundation by detailing the methodology of our proposed pipeline, which leverages visibility graphs to extract informative features and integrates them into an explainable machine learning framework.

Chapter 3

Proposed Methodology

3.1 Introduction

This chapter outlines the methodology adopted for estimating the carotid-femoral pulse wave velocity (cf-PWV) from PPG signals. As illustrated in Figure 3.1, the process begins with two datasets: Insilico and VitalDB, both containing raw PPG recordings.

The first step involves preprocessing the PPG signals to ensure quality and consistency. These signals are then transformed into Visibility Graph representations, allowing the capture of temporal dynamics in a graph-based format. From these representations, a set of features is extracted to characterize the signal's structure and content. These include spectral features, frequency-based features, morphological features, and shape-based features.

The resulting feature vectors are then fed into a machine learning model, trained and optimized through hyperparameter tuning, and ultimately evaluated on test data to provide cf-PWV predictions.

This end-to-end pipeline is summarized in Figure 3.1, which illustrates each key stage of the workflow.

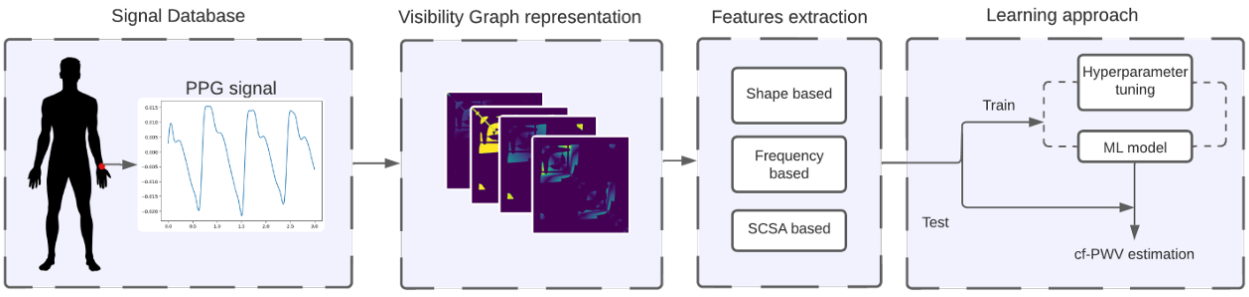


Figure 3.1: Overview of the proposed pipeline for cf-PWV estimation [34].

3.2 Datasets used

3.2.1 In-Silico Dataset

In the context of this study, one of the core approaches involves the use of a simulated photoplethysmogram (PPG) dataset—commonly referred to as an in-silico dataset—to evaluate the effectiveness of the proposed methodology for estimating Pulse Wave Velocity (PWV). Unlike clinical datasets derived from human recordings, in-silico signals are generated computationally using mathematical models. Specifically, this dataset was produced using a one-dimensional hemodynamic model that simulates arterial pulse wave propagation at multiple peripheral vascular sites under physiologically plausible conditions.

The simulated dataset replicates key features of cardiovascular dynamics while maintaining a high level of experimental control. It comprises data from 4,374 healthy virtual subjects ranging in age from 25 to 75 years, evenly distributed across six ten-year age groups. Each profile incorporates age-dependent changes in arterial stiffness, vascular tone, and blood pressure, reflecting the natural progression of cardiovascular aging. The PPG signals were sampled at 500 Hz and recorded at three anatomical sites: the radial, brachial, and digital arteries [35].

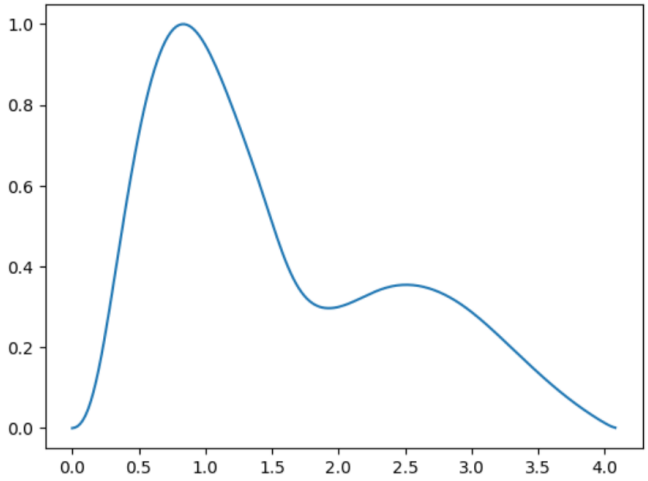


Figure 3.2: Example of a Simulated Photoplethysmogram (PPG) Waveform from the In-Silico Dataset

An example of a clean, noise-free PPG waveform from this dataset is shown in Figure 3.2. These artifact-free signals are well suited for benchmarking signal processing methods and exploring the theoretical limits of machine learning-based PWV estimation. The controlled environment eliminates confounding variables such as motion artifacts, sensor misplacement, and comorbidities—factors that often obscure physiological interpretation in real clinical recordings.

In this study, the in-silico dataset played a dual role. First, it served as a clean testbed to examine the performance of the Visibility Graph (VG) transformation and the feature extraction pipeline. Second, and more importantly, it was used as the sole source of data for both training and testing the PWV estimation model, enabling a rigorous evaluation of the methodology in a fully controlled setting. However, due to its synthetic nature, this dataset cannot fully capture the complexity and variability of real-world physiological signals. For this reason, we complement this analysis with an additional evaluation using real intraoperative recordings from the VitalDB dataset.

3.2.2 The VitalDB Dataset

To evaluate the real-world applicability and generalizability of our proposed methodology, we employed the VitalDB dataset—an open-access, high-resolution database developed by the research team at Seoul National University Hospital. Unlike the in-silico dataset used in the earlier stages of our study, VitalDB consists of real physiological signals collected from over 6,000 patients during surgical procedures.

These recordings include photoplethysmogram (PPG), arterial blood pressure (ART), electrocardiogram (ECG), and femoral pressure (FEM) signals, among others, all acquired using the SNUADC (Seoul National University Anesthesia Data Collection) system. A summary of the key signals available in VitalDB is presented in Table 3.1 [36].. Although the dataset provides a wide range of physiological parameters, this table highlights only a subset of them, focusing on commonly used waveforms and numeric values.

VitalDB Overview

Parameter	Description	Type/Hz	Unit
SNUADC/ART	Arterial pressure wave	W/500	mmHg
SNUADC/CVP	Central venous pressure wave	W/500	mmHg
SNUADC/ECG_II	ECG lead II wave	W/500	mV
SNUADC/ECG_V5	ECG lead V5 wave	W/500	mV
SNUADC/FEM	Femoral arterial pressure wave	W/500	mmHg
SNUADC/PLETH	Plethysmography wave	W/500	unitless
Solar8000/ART_DBP	Diastolic arterial pressure	N	mmHg
Solar8000/ART_MBP	Mean arterial pressure	N	mmHg
Solar8000/ART_SBP	Systolic arterial pressure	N	mmHg
Solar8000/BT	Body temperature	N	℃
Solar8000/CVP	Central venous pressure	N	mmHg
Solar8000/ETCO2	End-tidal CO2	N	mmHg
Solar8000/FEM_DBP	Femoral diastolic arterial pressure	N	mmHg
Solar8000/FEM_MBP	Femoral mean arterial pressure	N	mmHg
Solar8000/FEM_SBP	Femoral systolic arterial pressure	N	mmHg
Solar8000/FEO2	Fraction of expired O2	N	%
Solar8000/FIO2	Fraction of inspired O2	N	%
Solar8000/GAS2_EXPIRED	Expiratory volatile concentration	N	%
Solar8000/GAS2_INSPIRED	Inspiratory volatile concentration	N	%
Solar8000/HR	Heart rate	N	/min
Solar8000/INCO2	Inspiratory CO2	N	mmHg
Solar8000/NIBP_DBP	Non-invasive diastolic arterial pressure	N	mmHg

Table 3.1: Overview of selected physiological parameters recorded in the VitalDB dataset, including waveform (W) and numeric (N) signals, their sampling frequencies, and measurement units [36].

VitalDB stands out due to its synchronized, time-aligned signal acquisition across multiple anesthesia devices and monitors. This synchronization is made possible through the use of Vital Recorder, a specialized Windows-based software that captures high-frequency waveform data (up to 500 Hz) from more than 20 different medical instruments.

Vital Recorder also enables real-time visualization of ongoing signal acquisition, ensures data consistency, and supports remote access and monitoring. According to Lee et al. [37], the system achieves a 98.5% success rate in signal collection, making it highly reliable for research involving intraoperative physiological dynamics.

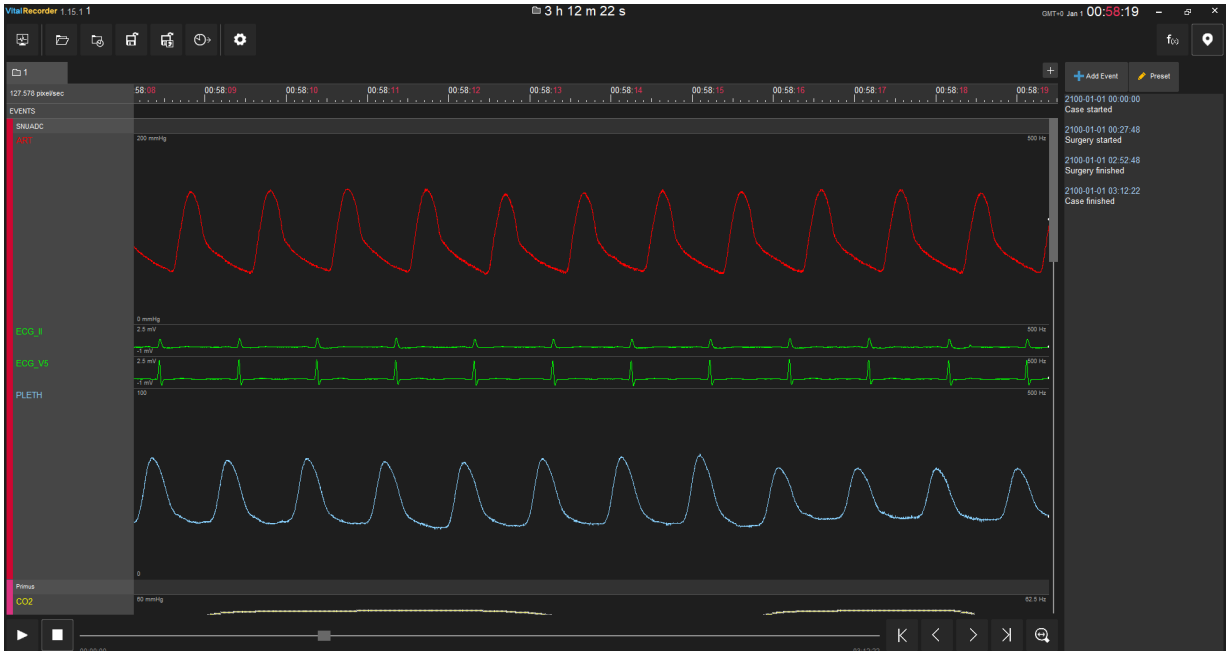


Figure 3.3: Graphical interface of Vital Recorder showing synchronized recordings of ART (red), ECG (green), and PLETH (blue) during surgery. The signals are continuous and span multiple cardiac cycles, illustrating the multicyclic nature of the VitalDB dataset.

3.2.2.1 Challenges and Limitations of VitalDB

Despite its strengths, VitalDB presents several challenges for pulse wave velocity (PWV) estimation:

- **Absence of Ground Truth PWV:** Unlike the in-silico dataset, VitalDB lacks predefined PWV values, requiring us to estimate them manually using pulse transit time (PTT).
- **Incomplete Signal Coverage:** A significant limitation of the VitalDB dataset is that not all patients have simultaneous recordings of central arterial pressure (ART), photoplethysmogram (PPG or pseudo-PPG), and femoral artery pressure (FEM) signals. This incompleteness restricts our ability to perform synchronized multi-signal analyses—especially for Pulse Wave Velocity (PWV) estimation using Pulse Transit Time (PTT), which requires two well-aligned signals (e.g., ART and FEM). Moreover, without the concurrent presence of a PPG-like waveform, it becomes impossible to evaluate the performance of our visibility graph-based estimation method against a physiological ground truth.
- **Imperfect Synchronization:** Although high-resolution acquisition is available, slight temporal misalignments between signals (e.g., ART and FEM) can significantly affect PTT-based PWV calculations, especially when timing differences are on the order of milliseconds. An example of such signal misalignment is illustrated in Figure 3.4, where even manually aligned cycles from the same subject show slight phase differences that can influence foot-to-foot time delay calculations.

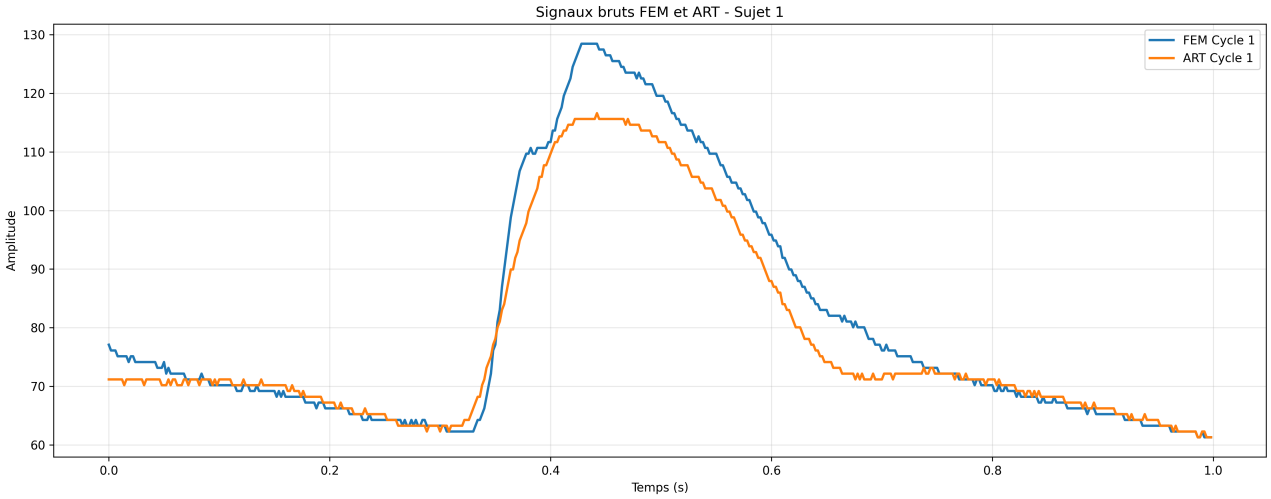


Figure 3.4: Comparison of ART (orange) and FEM (blue) signals over a single cardiac cycle for one subject. Note the visible delay between the foot and peak of the ART and FEM signals, which is critical for PTT estimation.

3.2.2.2 Our Validation Strategy and Signal Processing Pipeline

To overcome inconsistencies and incomplete data in the VitalDB database, we developed a precise validation strategy and an optimized signal processing pipeline focused on ensuring accurate estimation of pulse wave velocity (PWV).

1. Patient Selection and Reference PWV Estimation

We began by selecting only patients with both central arterial blood pressure (ART) and femoral artery (FEM) signals. These two signals are essential for calculating pulse transit time (PTT), a proxy measurement for estimating PWV.

However, the majority of patients in the VitalDB dataset did not have both signals recorded simultaneously or for a sufficient duration. In many cases, one of the two waveforms was missing, noisy, or not recorded in sync with the other, rendering accurate PTT estimation impossible. This lack of synchronized dual-site recordings significantly limited the pool of usable data. From the original cohort of 6,388 surgical patients in the VitalDB dataset, we identified 92 patients who had both femoral (FEM) and central arterial (ART) waveforms. Not all of these signals were usable due to issues such as temporal misalignment, short recordings, or poor waveform quality.

As illustrated in Figure 3.5, one common reason for exclusion was an inverted arrival time of pressure waves, where the femoral signal appeared before the central (aortic) one. This contradicts physiological expectations and points to synchronization or labeling issues. Such anomalies were systematically excluded to ensure valid PWV computation.

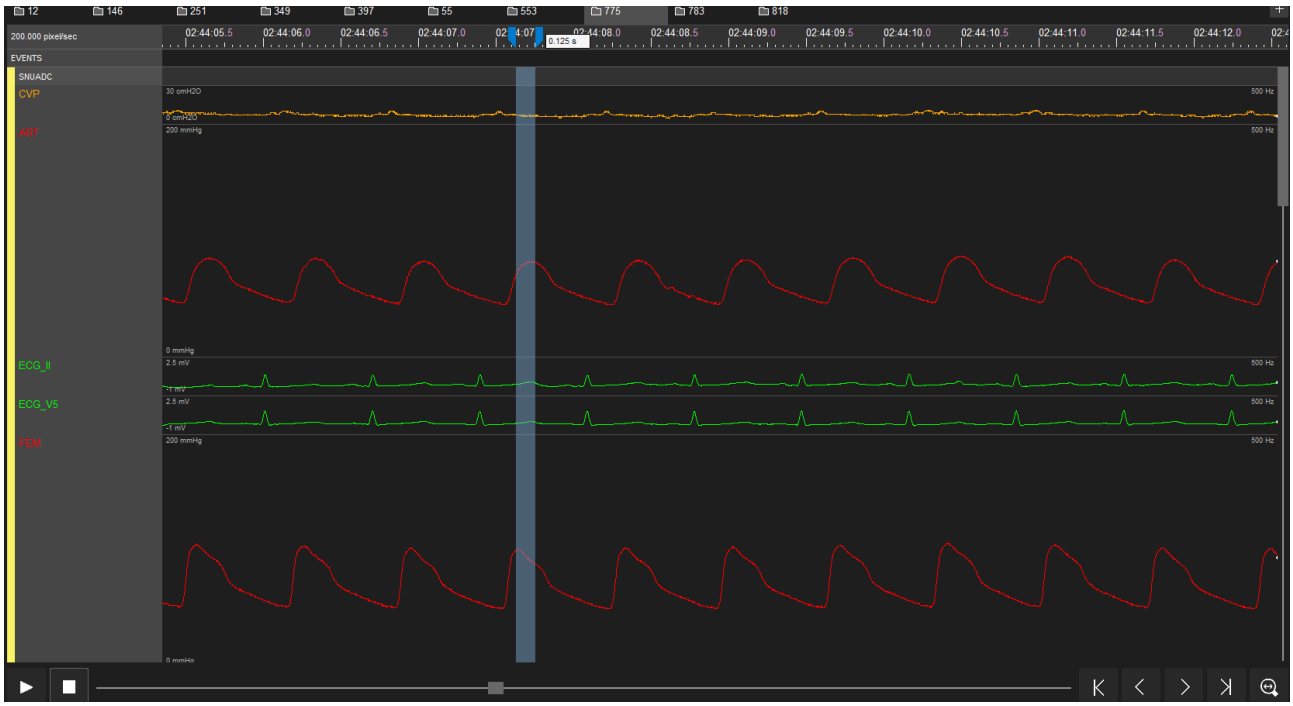


Figure 3.5: Example of a rejected case: the femoral waveform (FEM, bottom trace) precedes the central aortic waveform (ART, second trace).

After applying a strict quality control process focused on temporal alignment and signal integrity, we retained a final subset of 71 patients with high-quality, simultaneous ART and FEM recordings. This curated group served as a reliable foundation for accurate PTT and PWV estimation, minimizing signal-related artifacts and improving measurement consistency.

PWV was computed using the formula:

$$\text{PWV} = \frac{\text{distance}}{\text{PTT}} \tag{3.1}$$

The distance between the central (ART) and peripheral (FEM) arteries was approximated as 0.6,m. This estimate follows standard practice in PWV studies, where carotid–femoral path length is typically taken as 80

2. Signal Refinement: Monocycle Extraction

We transitioned to a monocycle-based strategy for greater reliability and alignment with the structure of VitalDB. Instead of using multiple cardiac cycles, we focused on extracting a single, representative cycle from each patient's signals (ART and FEM).

This involved:

- Detecting systolic and diastolic peaks in the ART and FEM signals.
- Extracting a clean segment that captured a full cardiac cycle between these two peaks.
- Ensuring proper synchronization and alignment across signals to minimize time estimation errors.

3. Pseudo-PPG Generation from Arterial Pressure (ART)

A major limitation of the VitalDB dataset was the absence of photoplethysmography (PPG) recordings for the majority of patients. Since our methodology depends on the morphological structure of PPG waveforms—particularly for visibility graph–based feature extraction—this posed a critical barrier to extending our pipeline across a larger patient cohort.

To understand the physiological basis of our workaround, it is helpful to compare how PPG and arterial signals are acquired and how they reflect cardiovascular dynamics. Figure **??** illustrates the typical acquisition setup: invasive arterial pressure (referred to as ABP in the figure, and equivalent to the ART signal in our dataset) is measured via a catheter in a central artery, whereas PPG is recorded non-invasively using an optical sensor placed on the fingertip. Despite these differences, both signals capture the rhythmic pulsatility induced by cardiac activity and exhibit strong morphological correlation.

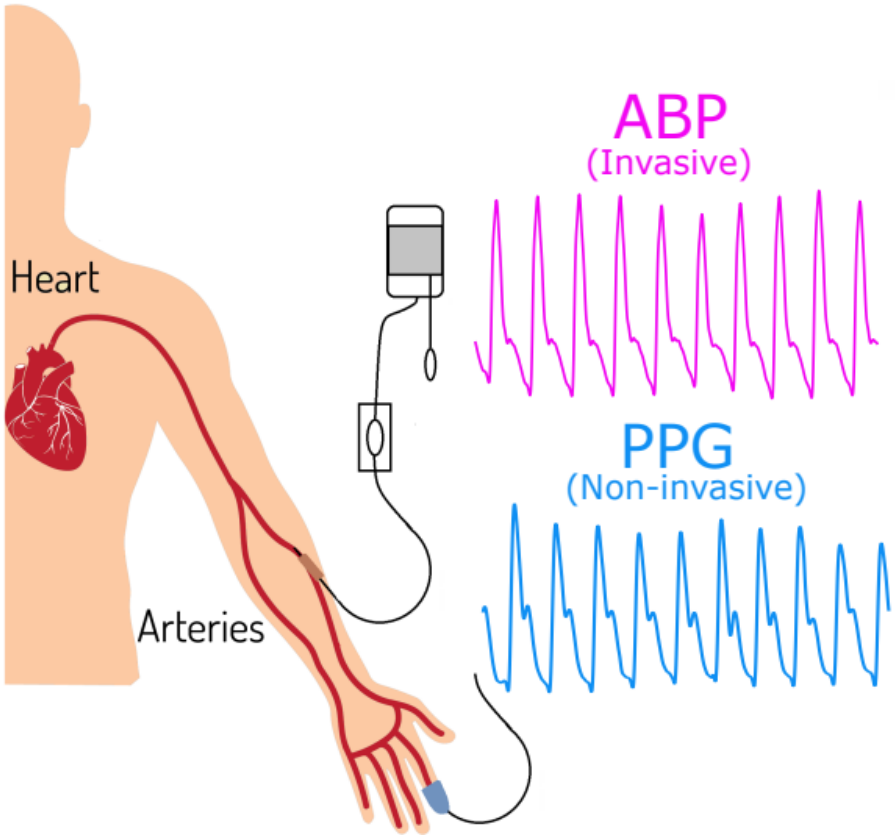


Figure 3.6: Simultaneously collected Arterial Blood Pressure (ABP) and Photoplethysmogram (PPG) signals. Note, ABP was measured invasively while PPG was measured non-invasively[38]

To address this, we leveraged the morphological similarity between invasive arterial pressure (ART) signals and non-invasive PPG waveforms. Although ART and PPG are acquired through fundamentally different methods—catheter-based pressure sensing vs. optical volume changes—they both capture the same underlying physiological process: the pulsatile propagation of blood through the arterial system. Previous studies have confirmed that, when appropriately processed, ART signals can approximate the shape and timing of PPG signals with reasonable fidelity [39].

Rationale: Morphological Compatibility

PPG and ART waveforms share several key features:

- A steep systolic upstroke,
- A dicrotic notch corresponding to aortic valve closure,
- A gradual diastolic decay,
- Cycle-to-cycle regularity under resting conditions.

These shared morphological characteristics, especially when signals are filtered and normalized, support the feasibility of using ART as a structural surrogate for PPG.

Figure 3.7 [38] further illustrates this compatibility across normotensive, prehypertensive, and hypertensive patients. Despite pressure-level differences, the ART (ABP) and PPG waveforms remain in-phase and exhibit strong correlation coefficients ($r > 0.9$ in all cases), reinforcing the physiological interchangeability of their morphological features for tasks like pulse wave analysis [38].

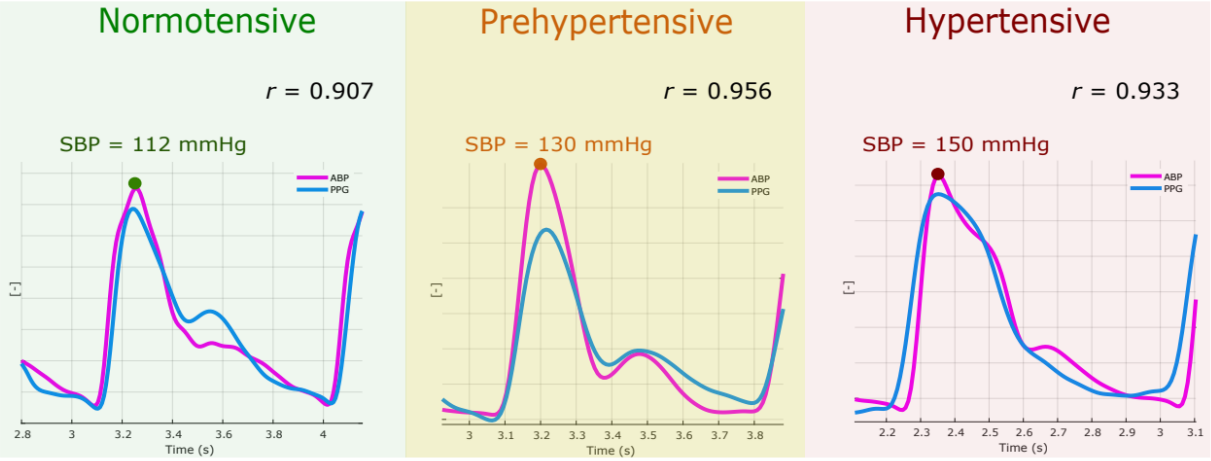
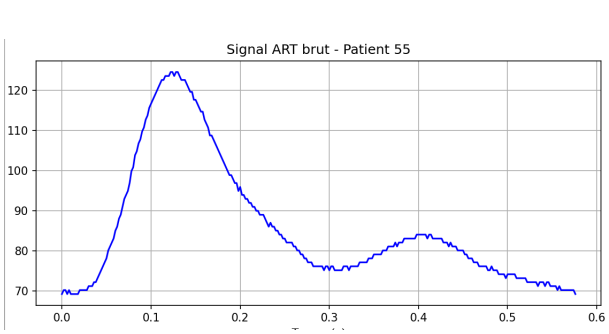


Figure 3.7: Examples of in-phase analysis across varying systolic blood pressures (SBP). The plots show photoplethysmogram (PPG) and arterial pressure (ABP) waveforms. Note that in our context, ABP is equivalent to the ART signal. The correlation coefficient r quantifies the morphological similarity between both signals [38].

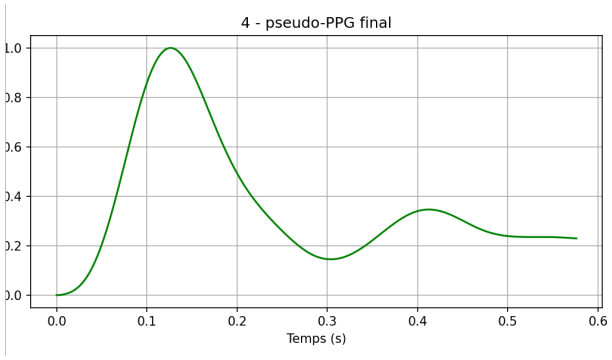
Signal Transformation Pipeline

We developed a robust four-stage signal processing pipeline to convert ART into pseudo-PPG:

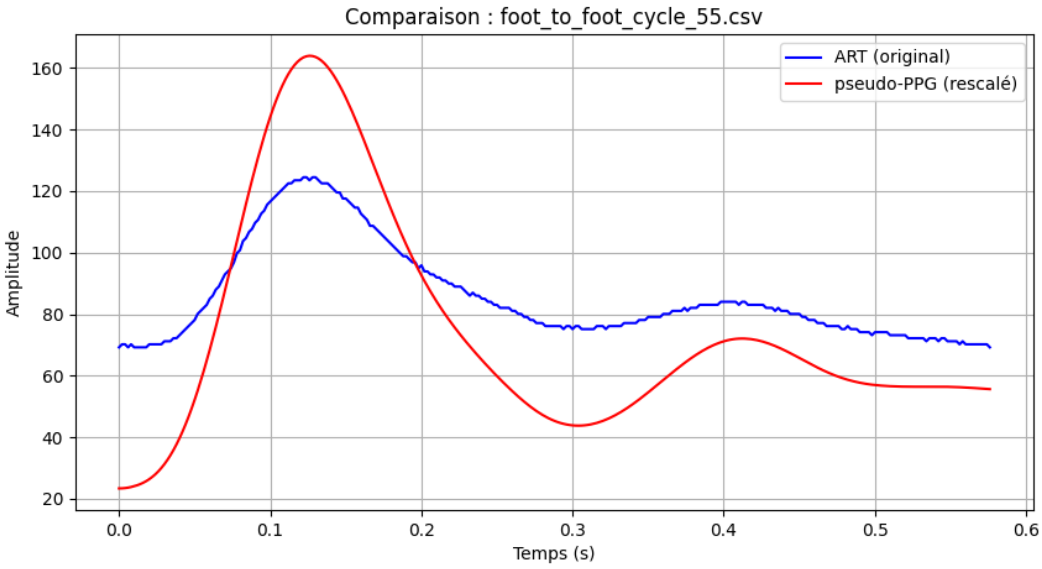
1. **Global Min–Max Normalization:** To ensure consistency across patients, we computed a global minimum and maximum from all ART signals in the dataset and scaled all samples to the $[0, 1]$ range.
2. **Band-Pass Filtering:** We applied a 2^{nd}-order Butterworth filter with cut-off frequencies at 0.5 Hz and 10 Hz to eliminate low-frequency drift and high-frequency noise.
3. **Smoothing:** A Savitzky–Golay filter (window size = 11, polynomial order = 2) was used to reduce noise while preserving waveform shape.
4. **Final Normalization:** A second normalization step was applied post-smoothing to re-standardize the signal range across all patients.



(a) Raw arterial pressure (ART) signal.



(b) Pseudo-PPG signal after processing.



(c) Overlay comparison between ART and pseudo-PPG signals. The generated pseudo-PPG preserves major morphological features such as systolic peaks and dicrotic notches.

Figure 3.8: Stepwise transformation of an arterial pressure (ART) signal into a pseudo-PPG waveform. The pipeline includes normalization, filtering, and smoothing to generate a surrogate signal suitable for visibility graph-based analysis.

Validation Strategy

We estimated PWV using pseudo-PPG signals derived from arterial pressure (ART) waveforms. To validate these estimates, we compared them to reference PWV values computed using the classical pulse transit time (PTT) method — defined as the time delay between central (ART) and femoral (FEM) pressure waveforms, divided by an assumed path length.

While both values represent derived quantities and not direct ground-truth measurements PTT based PWV is a well-established physiological proxy used in both clinical and research contexts. It is based on synchronized invasive recordings and serves as a practical benchmark in the absence of gold-standard modalities such as MRI or carotid-femoral tonometry.

This limitation reflects a broader, clinically recognized challenge: in real-world surgical or ICU environments, complete multimodal vascular data including true PPG recordings or direct PWV measurements are often unavailable. To overcome this, we leveraged accessible arterial signals and derived surrogate PPG waveforms, allowing us to significantly increase the number of usable patient cases without compromising analytical rigor.

Importantly, this validation procedure not only assesses the internal consistency of our pseudo-PPG–based method, but also supports the external validity of models trained on synthetic (in silico) data. By demonstrating agreement between pseudo-PPG–derived PWV estimates and real-world PTT-based PWV values, we establish a physiological bridge between simulation-based training and clinical signal testing.

This cross-validation sets the stage for the next phase of our analysis, in which we combine both datasets in silico and VitalDB to build a hybrid framework that capitalizes on the strengths of each data source.

3.2.3 Merged In-Silico and VitalDB Dataset

Beyond isolated training and validation phases, we pursued a third, integrated approach by merging the in-silico and VitalDB datasets into a unified training and evaluation framework. This strategy was motivated by the need to bridge the gap between idealized, noise-free synthetic data and the inherently noisy, artifact-prone signals encountered in clinical environments.

Each dataset contributes complementary advantages. The in-silico dataset provides high-fidelity waveforms with perfectly defined PWV labels, which are essential for initial model learning and benchmarking. In contrast, the VitalDB dataset offers real patient signals with complex morphology, natural variability, and realistic acquisition imperfections—making it highly valuable for stress-testing model robustness and generalization.

By merging these two domains, we sought to create a training environment that balances clean supervision with clinical realism. This hybrid dataset exposes the model to both controlled and uncontrolled variability in signal quality, amplitude, and temporal dynamics. Such exposure is crucial for mitigating domain shift—the phenomenon where a model trained solely on one data type performs poorly when deployed on another.

The hybrid dataset was used in a cross-validated training and testing scheme. During each fold, data from both sources were sampled and mixed to ensure diversity within training sets and reliability in testing sets. Care was taken to avoid patient overlap and leakage between folds. This approach simulated real-world deployment conditions, where models must interpret unseen, noisy signals while relying on features learned from idealized data.

Table 3.2: Comparison of in-silico and VitalDB datasets

Property	**In-Silico**	**VitalDB**
Signal Source	Simulated	Clinical (ICU)
Signal Quality	Clean, noise-free	Noisy, artifact-prone
Label Accuracy	Perfect (ground truth)	Derived (from PTT)
Morphological Variability	Limited	High
Acquisition Device	Synthetic generator	Real monitors

Summary:

Merging in-silico and VitalDB data into a unified training regime enhances the model's resilience to signal variability, improves its capacity to generalize across domains, and better reflects the diversity of conditions encountered in clinical signal processing. This three-pronged strategy—(1) model development on in-silico data, (2) validation using real VitalDB cases, and (3) training with a merged dataset—constitutes a comprehensive and pragmatic pipeline for building robust, data-driven models for pulse wave velocity estimation.

3.3 Signal to Graph Transformation Using the Visibility Graph

As discussed in the previous chapter, visibility graphs (VGs) have shown significant promise in representing the complex dynamics of physiological time series, particularly photoplethysmographic (PPG) signals. In this section, we detail how the VG framework is implemented in our study to convert 1D PPG signals into structured 2D representations, following and extending the approach proposed by Lacasa et al. [31] and more recently adapted by Vargas et al. [34] for biomedical applications. This transformation forms a core component of our methodology, providing a geometrically motivated bridge between time-domain waveforms and graph-based image representations.

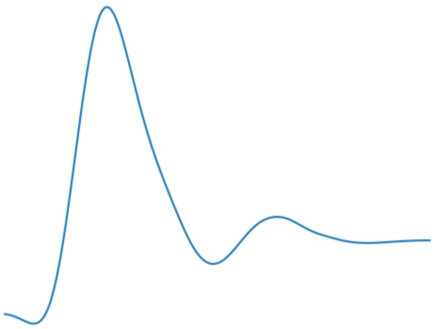
3.3.1 Natural Visibility Graph Construction

Once the signal is preprocessed, the core transformation into a visibility graph is applied. Let $y = [y_1, y_2, \ldots, y_T]$ represent the amplitude values of a PPG signal sampled at uniform time intervals. Each point (t_i, y_i) is treated as a node in the graph. An edge is drawn between two nodes (t_a, y_a) and (t_b, y_b), with $t_a < t_b$, if the following visibility criterion is satisfied for all intermediate points (t_c, y_c) where $t_a < t_c < t_b$:

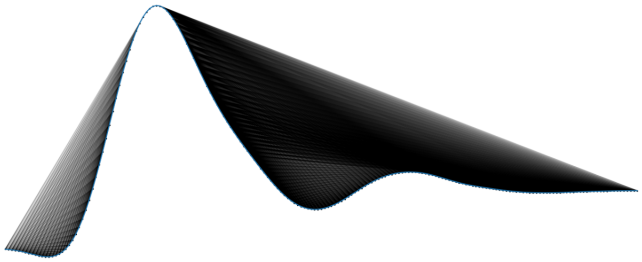
$$y_c < y_a + \frac{y_b - y_a}{t_b - t_a}(t_c - t_a) \tag{3.2}$$

This condition ensures that the straight line connecting the two nodes is not obstructed by any intermediate point, analogous to a line-of-sight in physical space. The resulting graph is undirected, invariant under affine transformations, and encodes both local and global properties of the signal [31].

Figure 3.9a shows an example of a raw PPG segment extracted from the VitalDB dataset. Its corresponding visibility graph structure, constructed using the natural VG criterion, is visualized in Figure 3.9b.



(a) Raw PPG waveform segment from VitalDB.



(b) Natural visibility graph (VG) representation.

Figure 3.9: Transformation of a PPG signal segment into its corresponding visibility graph representation.

3.3.2 Limited Penetrable Visibility Graph (LPVG)

To improve the robustness of the VG approach, especially in the presence of noise or slight waveform variability, we utilize the *Limited Penetrable Visibility Graph (LPVG)* model as introduced by Han [40] and applied by Vargas et al. [34]. In LPVG, the visibility criterion is relaxed by allowing up to p intermediate points to violate the inequality, thus enabling connections between nodes that would otherwise be obstructed by noise.

This penetration parameter p provides a means to balance sensitivity and robustness. When $p = 0$, the traditional VG is recovered; higher values of p enable the graph to better capture long-range dependencies, particularly important when analyzing multicycle signals such as PPG.

3.3.3 Weighted Visibility Graph (WVG)

While VG and LPVG use binary edges (connected or not), the *Weighted Visibility Graph (WVG)* enriches the representation by assigning a real-valued weight to each visible connection. These weights encode additional geometric or temporal information about the signal. We apply five different weighting schemes, each designed to capture a specific aspect of the PPG signal:

- **Vertical distance**:

$$w_{i,j} = |y_j - y_i|$$

  Captures amplitude variation.

- **Horizontal distance**:

$$w_{i,j} = |t_j - t_i|$$

  Reflects temporal separation.

- **Slope (absolute)**:

$$w_{i,j} = \left| \frac{y_j - y_i}{t_j - t_i} \right|$$

  Highlights rapid morphological changes.

- **Euclidean distance**:

$$w_{i,j} = \sqrt{(y_j - y_i)^2 + (t_j - t_i)^2}$$

Combines time and amplitude in a single measure.

- **Squared Euclidean distance**:

$$w_{i,j} = (y_j - y_i)^2 + (t_j - t_i)^2$$

Emphasizes long-range separations.

These metrics are applied not only to the **original PPG waveform**, but also to its **first** and **second temporal derivatives**, providing a multiscale and multidimensional view of the underlying hemodynamics.

3.3.4 Adjacency Matrix as an Image Representation

Once the visibility graph (VG, LPVG, or WVG) is constructed from a PPG signal of length T, it is encoded into an *adjacency matrix* $A \in \mathbb{R}^{T \times T}$. Each element $A_{i,j}$ represents the presence and strength of the connection (or visibility) between the i-th and j-th signal points.

- In **binary visibility graphs**, $A_{i,j} \in \{0, 1\}$: a value of 1 indicates visibility between nodes i and j, and 0 otherwise.
- In **weighted visibility graphs**, $A_{i,j} \in \mathbb{R}^+$ reflects a continuous-valued metric such as vertical distance, slope, or Euclidean distance between the connected nodes.

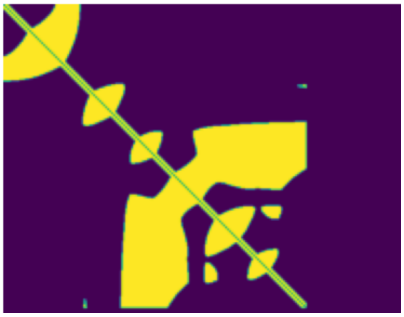
By construction, the adjacency matrix is always square, with size $T \times T$, since it encodes pairwise relationships among all T points in the signal.

To enable consistent image-based processing and comparison across signals of varying lengths, each adjacency matrix is resized to a fixed resolution, typically 35×35. This resizing is performed using bilinear interpolation or anti-aliasing filters, which preserve the structural patterns of the original matrix while reducing its dimensionality.

The resulting matrix is interpreted as a grayscale image, where:

- Each pixel intensity reflects the presence or strength of connection between two points in the signal.
- Values near the main diagonal represent short-term, local relationships (e.g., within one cycle).
- Off-diagonal values capture longer-range dependencies (e.g., across multiple PPG cycles).

Figure 3.10 presents three VG-based image examples: one from the original signal (Figure 3.10a), one from its first derivative (Figure 3.10b), and one from the second derivative (Figure 3.10c).



(a) VG image from original PPG

(b) VG image from first derivative

(c) VG image from second derivative

Figure 3.10: Examples of VG image representations obtained from PPG signals and their derivatives, extracted from the VitalDB dataset.

This transformation from signal to matrix to image provides a compact yet informative 2D representation of the original waveform, encoding both temporal structure and morphological features. These standardized VG images are then used as inputs for downstream feature extraction tasks.

3.4 Feature Extraction

In this study, feature extraction is a key step that translates the complex visual and structural content of PPG-derived visibility graph (VG) images into numerical representations. These features serve as predictors for the PWV estimation task. Four categories of features are extracted: shape-based, morphological, frequency-based, and spectral features. Each group targets a different aspect of the signal's structure and its physiological implications.

3.4.1 Shape-Based Features

Shape-based features capture the overall geometric configuration of the VG images derived from PPG and its derivatives. These features are important because changes in arterial stiffness alter the morphology of the PPG waveform, which in turn modifies the topology of the VG image.

Hu's Invariant Moments Hu's moments are a set of seven mathematical expressions derived from central moments of the image. Each moment captures different aspects of the image's global structure—such as symmetry, orientation, and spread [41].

$$I_1 = \eta_{20} + \eta_{02} \tag{3.3}$$
$$I_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{3.4}$$
$$I_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{3.5}$$
$$I_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{3.6}$$

These moments are invariant to translation, scale, and rotation, making them suitable for comparing PPG images across individuals and conditions.

Zernike Moments Zernike moments are advanced shape descriptors derived from complex orthogonal polynomials defined over the unit disk [42]. These moments are particularly well-suited for capturing detailed local and global shape characteristics of images derived from physiological signals. The Zernike moment of order n and repetition m for a grayscale image function $f(x, y)$ is mathematically defined as:

$$Z_{nm} = \frac{n+1}{\pi} \int\int_{x^2+y^2 \leq 1} f(x, y)\, V_{nm}^*(x, y)\, dx\, dy \tag{3.7}$$

where $V_{nm}(x, y)$ is the Zernike polynomial, expressed in polar coordinates as $V_{nm}(x, y) = R_{nm}(\rho)e^{jm\theta}$, with $\rho = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(y/x)$. The radial polynomial $R_{nm}(\rho)$ is defined by:

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2} - s\right)!\left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \tag{3.8}$$

This formulation enables the decomposition of an image into components that reflect its geometric structure. The magnitude of the moment $|Z_{nm}|$ indicates the strength of the feature corresponding to the specific polynomial pattern, while the phase captures its orientation. Because Zernike polynomials are orthogonal, each moment provides independent and non-redundant information, which is advantageous for machine learning models.

Applied to visibility graph (VG) images constructed from PPG signals, Zernike moments offer a way to detect both coarse and fine structural variations. Low-order moments are sensitive to general shape characteristics such as circularity and symmetry, while high-order moments are more responsive to subtle waveform features like the dicrotic notch or inflections caused by arterial stiffening. Importantly, Zernike moments are invariant to rotation and resilient to noise, making them reliable tools for capturing waveform deformation patterns associated with elevated Pulse Wave Velocity (PWV). As such, they serve as informative and robust biomarkers in the task of non-invasive arterial stiffness estimation.

Robustness and Use of Zernike Moments

Zernike moments are well-suited for analyzing biomedical images due to their key properties:

- **Rotation, scale, and translation invariance**
- **Noise resilience**, especially in low-order moments
- **Multi-resolution analysis**, capturing both global and fine details
- **Clear interpretability**, as each moment corresponds to a known spatial pattern

In our study, Zernike moments are extracted from masked visibility graph (VG) images derived from the PPG signal and its derivatives. The analysis is restricted to a circular region to match the unit-disk definition of the polynomials.

Low-order moments capture general shape, while higher orders reflect subtle waveform features like notches or inflections. These characteristics, often linked to arterial stiffness, make Zernike moments valuable features for non-invasive PWV estimation.

3.4.2 Morphological Features – Threshold Adjacency Statistics (TAS)

Threshold Adjacency Statistics (TAS) is a technique used to extract morphological features by analyzing local pixel structures within the VG image [43]. The method begins by converting the grayscale image into six binary images using different intensity thresholds. In each binary version, pixels with intensities above the threshold are marked as active (white).

For every thresholded image, TAS examines 3×3 pixel neighborhoods and counts how often each of nine specific adjacency patterns occurs—such as isolated pixels, straight lines, corners, or compact clusters. This results in $6 \times 9 = 54$ features, each encoding a particular type of local texture.

These features are physiologically meaningful: higher arterial stiffness leads to sharper PPG waveforms, which translate into denser and more structured patterns in the VG. Softer arteries produce smoother curves, resulting in more scattered textures. Thus, TAS provides a localized morphological description that indirectly reflects the vascular condition and is valuable for machine learning-based PWV prediction.

3.4.3 Frequency-Based Features – Wavelet Packet Decomposition (WPD)

Wavelet Packet Decomposition (WPD) is a method used to analyze how the energy of a VG image is distributed across both spatial and frequency dimensions [44]. Unlike the Fourier Transform, which only provides frequency information, wavelets offer a more flexible analysis by localizing patterns in both time (or space) and frequency. This makes WPD particularly useful for detecting subtle structural changes in physiological signals like PPG.

The decomposition is performed by recursively splitting the image into smaller parts called sub-bands. Each sub-band represents a specific range of spatial frequencies and positions. At every level of decomposition, statistical features—such as the mean and standard deviation of pixel intensities—are computed from each sub-band. These features capture important image textures: lower levels reflect global structures, while deeper levels highlight finer details.

Mathematically, the decomposition applies combinations of low-pass h and high-pass g filters across rows and columns. For example:

$$C_{4k}^{p+1}(i,j) = \sum_{m,n} h(m)h(n)C_k^p(m+2i, n+2j) \tag{3.9}$$

$$C_{4k+1}^{p+1}(i,j) = \sum_{m,n} h(m)g(n)C_k^p(m+2i, n+2j) \tag{3.10}$$

This recursive filtering generates a rich set of coefficients that describe how image energy is organized. Since arterial stiffness affects the sharpness and speed of pulse waves, these changes produce specific frequency patterns in the VG image. WPD captures these patterns effectively, offering informative features for estimating PWV.

3.4.4 Spectral Features – Semi-Classical Signal Analysis (SCSA)

The Semi-Classical Signal Analysis (SCSA) method is a signal decomposition technique based on quantum mechanics, where a real-valued signal or image is interpreted as a potential energy function and analyzed via the Schrödinger operator [45]. This method has been successfully extended to image analysis and demonstrated strong performance as a feature extraction tool for Pulse Wave Velocity (PWV) estimation [46].

Let $I(x, y)$ be a real, positive-valued image (e.g., the VG image). In the SCSA framework, the image is interpreted as a potential in the two-dimensional semi-classical Schrödinger operator:

$$\mathcal{H}_{2,h}(I)\psi(x,y) = -h^2\left(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2}\right) - I(x,y)\psi(x,y), \tag{3.11}$$

where $h > 0$ is the semi-classical parameter, and $\psi(x, y)$ denotes the eigenfunctions associated with the negative eigenvalues $\lambda_{mh} < 0$. These spectral quantities reflect the local and global structure of the input image.

The SCSA-based image reconstruction is then defined by:

$$I_{2h}(x,y) = \left[\frac{h^2}{L_{2,\gamma}^{cl}}\sum_{m=1}^{M_h}(-\lambda_{mh})^{\gamma}\psi_{mh}^2(x,y)\right]^{\frac{1}{\gamma+1}} \tag{3.12}$$

where: - h is the semi-classical (smoothing) parameter, - γ controls the contrast in the reconstructed image, - M_h is the number of negative eigenvalues λ_{mh}, - $\psi_{mh}(x, y)$ are the corresponding eigenfunctions.

From this spectral decomposition, a set of discriminative features is extracted:

$$\text{INV1} = 4h\sum_{m,n}\kappa_h[m,n] \tag{3.13}$$

$$\text{INV2} = \frac{16h}{3}\sum_{m,n}\kappa_h[m,n]^3 \tag{3.14}$$

$$\text{INV3} = \frac{128h}{7}\sum_{m,n}\kappa_h[m,n]^7 \tag{3.15}$$

These INV features represent global energy descriptors. INV1 gives a linear energy measure, INV2 emphasizes stronger spectral components, and INV3 captures highly peaked contributions — all reflecting different levels of detail in vascular signal variability.

$$K_n = \frac{1}{M_h}\sum_{m}\kappa_h[m,n] \tag{3.16}$$

$$E_n = \frac{1}{M_h}\sum_{m}\kappa_h[m,n]^{1/\gamma} \tag{3.17}$$

K_n is the average local energy per eigenmode, and E_n normalizes this energy based on contrast γ. These are useful for capturing smoothness and sharpness in image structure.

$$R_h = \frac{\sum_m \lambda_{mh}^2}{\left(\sum_m \lambda_{mh}\right)^2} \tag{3.18}$$

$$MR_h = \text{median}(\kappa_h) \tag{3.19}$$

R_h reflects the sharpness of the energy spectrum (higher values indicate concentration in fewer modes), while MR_h gives a robust estimate of central spectral activity.

3.5 Feature Selection

In this study, feature selection was a crucial step to enhance the robustness and performance of the PWV estimation model. Given the large number of extracted features from the VG images, it was essential to identify and retain the most informative ones while discarding redundant or irrelevant variables. To achieve this, a rigorous multi-criteria pipeline was employed, combining statistical dependency analysis, multicollinearity reduction, and ranking-based selection.

First, we computed the statistical dependence between each feature and the target PWV values using four correlation metrics:

- **Pearson** — This coefficient evaluates linear dependence and assumes that the data follow a normal distribution.
- **Spearman** — A non-parametric rank-based measure suitable for detecting monotonic but non-linear relationships.
- **Kendall** — Another non-parametric method that measures the concordance between data pairs.
- **XICOR** — A more recent metric based on adjacent differences in ranked data, effective at capturing non-monotonic or complex non-linear associations.

For each feature, we computed correlation coefficients and associated p-values using all four methods. Features with statistically significant associations (p-value < 0.05) were retained. If this threshold excluded too many candidates, we retained the top features based on absolute correlation strength to ensure sufficient representation.

To reduce redundancy and prevent multicollinearity, we applied a filtering step based on pairwise Spearman correlations. When two features were highly correlated (correlation > 0.9), only one was preserved to maintain diversity among predictors.

Next, the remaining features were ranked under each correlation method. A combined score was computed for each feature by integrating its rank and correlation magnitude across all metrics. The final selection included those with the highest composite scores, ensuring that multiple types of statistical dependencies were considered.

This ensemble-based selection strategy balances interpretability, reduces overfitting risk, and improves generalization, making it well-suited for predicting PWV from visibility graph features.

The final selected features, resulting from the multi-step filtering and ranking process, are shown in Figure 3.11. These features were used as input variables for the machine learning model predicting PWV.

```
Features sélectionnées:
 1. first_derivative_abs_slope_std_sums
 2. second_derivative_abs_v_distance_TAS39
 3. second_derivative_abs_v_distance_TAS36
 4. first_derivative_abs_slope_Zernikes_Moments_radius_9_8
 5. original_abs_slope_WP_coif1_dd_std
 6. second_derivative_h_distance_TAS48
 7. second_derivative_abs_slope_TAS46
 8. second_derivative_abs_v_distance_TAS48
 9. second_derivative_distance_TAS18
10. original_None_Hu_Moment_0
11. second_derivative_abs_v_distance_TAS19
12. second_derivative_sq_distance_TAS48
13. Subject_ID
14. second_derivative_distance_TAS28
15. second_derivative_abs_v_distance_TAS10
16. second_derivative_abs_v_distance_TAS4
17. first_derivative_sq_distance_TAS18
18. second_derivative_abs_slope_INV1_sums
19. second_derivative_abs_slope_TAS22
20. second_derivative_abs_slope_Zernikes_Moments_radius_9_1.2
21. first_derivative_abs_slope_Zernikes_Moments_radius_9_5
```

Figure 3.11: Final selected features from the visibility graph-based feature extraction pipeline. Only the most informative features, selected via a multi-criteria strategy (correlation, redundancy reduction, and ranking), are shown.

3.6 Machine Learning Models

Following the feature selection process, we developed a supervised learning pipeline to estimate the carotid-femoral Pulse Wave Velocity (PWV) using the selected visibility-graph-based features. The model chosen for this task was the Explainable Boosting Machine (EBM), which combines ensemble learning techniques with interpretable additive modeling, making it highly appropriate for medical applications that require both performance and transparency.

3.6.1 Model Architecture: Explainable Boosting Machine (EBM)

EBM is part of the Generalized Additive Models (GAMs) family, enhanced with ensemble techniques such as bagging and gradient boosting. Its predictive function is additive and can be written as [47]:

$$\hat{y} = \beta_0 + \sum_{i=1}^{p} f_i(x_i) + \sum_{i<j} f_{ij}(x_i, x_j) \tag{3.20}$$

where:

- β_0 is a bias term (intercept),
- $f_i(x_i)$ are one-dimensional functions representing the effect of each individual feature,
- $f_{ij}(x_i, x_j)$ are optional pairwise interaction terms included only if they significantly improve performance.

As shown in Equation (3.20) [47], EBM combines additive modeling with feature-specific boosting to produce both accurate and interpretable predictions.

Each function f_i is trained using gradient boosting over shallow decision trees. The model is built iteratively: for each feature, a separate weak learner (decision tree) is trained to minimize the residual error of the current ensemble. This process is repeated over several iterations to gradually improve accuracy while maintaining interpretability.

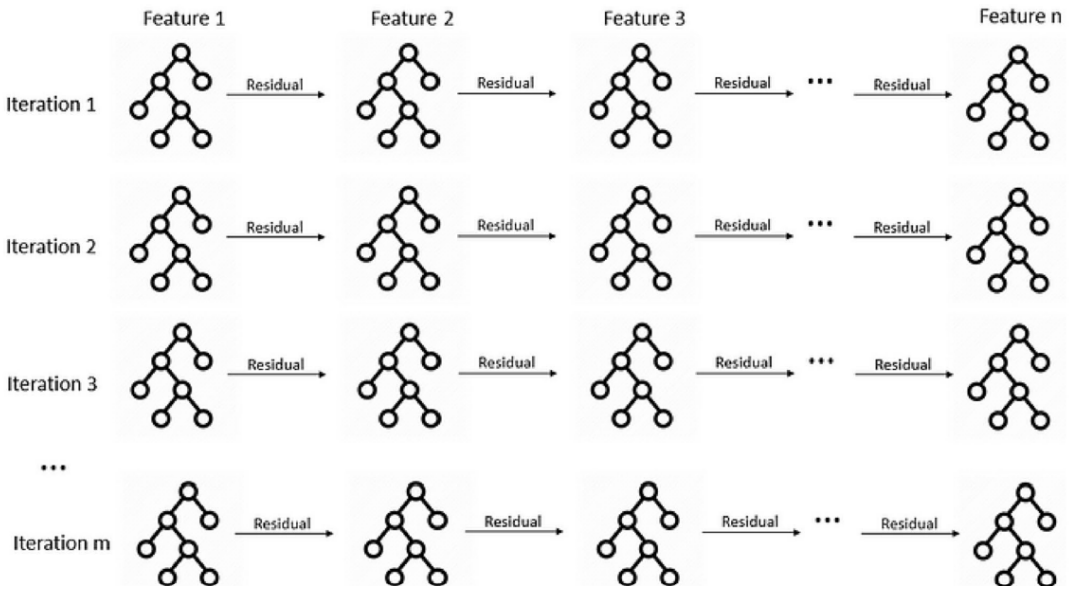


Figure 3.12: EBM learning structure: at each iteration, one shallow tree is trained per feature to fit the residuals [48].

As illustrated in Figure 3.12[48], the EBM model constructs separate trees for each feature at every iteration. Each tree learns to predict the residuals from the previous iteration, allowing the model to refine its predictions feature by feature. This architecture ensures that the final model remains both powerful and fully interpretable, a crucial advantage in clinical applications.

3.6.2 Hyperparameter Optimization using Optuna

To fine-tune the EBM model, we used the Optuna framework [49], which applies a Tree-structured Parzen Estimator (TPE) to search the hyperparameter space in an efficient, guided manner. This optimization process focused on the following parameters:

- `learning_rate`: step size for gradient updates, searched in the range $[10^{-3}, 10^{-1}]$,
- `max_leaves`: maximum number of leaves in each decision tree (3 to 7),
- `min_samples_leaf`: minimum number of samples needed to form a leaf node (5 to 20),
- `max_bins` and `max_interaction_bins`: number of bins for histogram-based feature discretization (32–128 and 16–64).

The performance of each hyperparameter set was evaluated using k-fold stratified cross-validation on the training data. The objective function minimized during tuning was the negative mean squared error (MSE):

$$\mathcal{L}_{\text{Optuna}} = -\frac{1}{k}\sum_{i=1}^{k} \text{MSE}(y_{\text{val},i}, \hat{y}_i) \tag{3.21}$$

This process was repeated across multiple trials. For each one, a model was trained and validated using the proposed configuration. The hyperparameter combination that produced the lowest average error was selected for final model training.

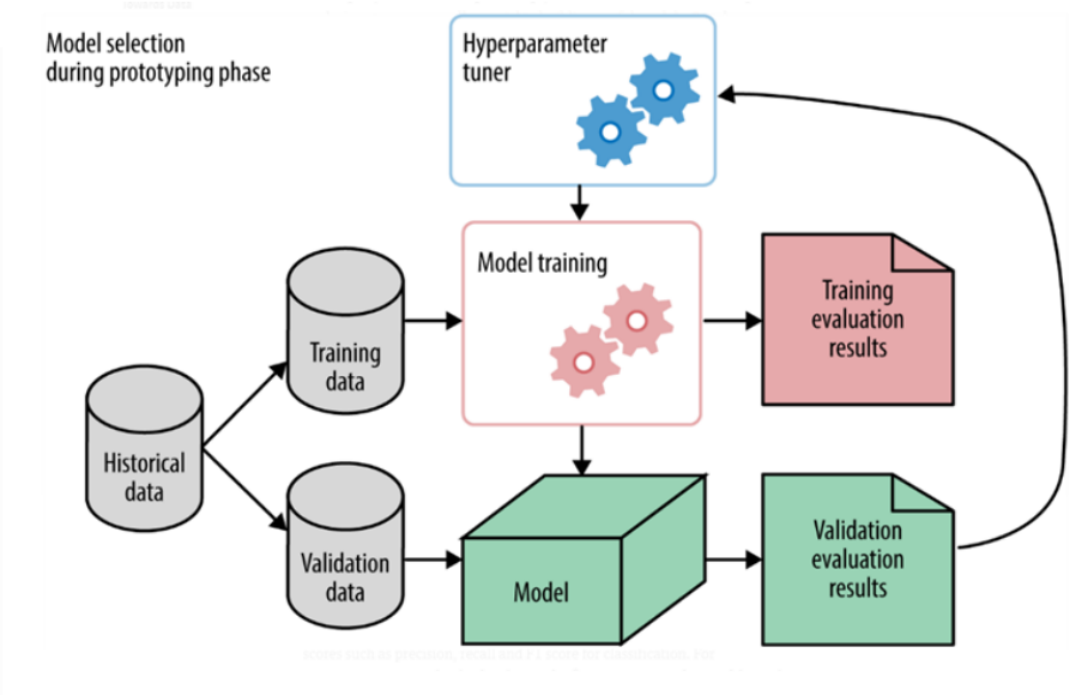


Figure 3.13: Hyperparameter tuning and model selection workflow [50].

Figure 3.13 illustrates the iterative nature of the Optuna-based hyperparameter optimization workflow. The process begins with the preparation of the dataset, typically split into training and validation sets. An objective function in our case, the negative mean squared error (MSE) is defined to guide the tuning process.

During each trial, Optuna proposes a new set of hyperparameters using the Tree-structured Parzen Estimator (TPE), which models the likelihood of promising parameter regions based on past trials. The EBM model is then trained on the training subset with the suggested hyperparameters and evaluated on the validation subset. The resulting performance is reported back to the tuner.

This loop continues across multiple trials, allowing Optuna to iteratively refine its search and concentrate on more promising configurations. At the end of the process, the hyperparameter set that yields the lowest validation error is selected for final model training. This workflow offers an efficient and scalable alternative to manual or grid search, and helps improve model generalization by using validation-based performance feedback.

Final Training and Prediction

Once the best hyperparameters were identified, a final EBM model was instantiated and trained on the full training set. The model was then used to predict PWV values on the independent test set. This step evaluates how well the model generalizes to unseen data and ensures that the results are not biased by the cross-validation procedure.

3.7 Conclusion

In this chapter, we presented the methodological framework developed to estimate arterial stiffness from photoplethysmographic (PPG) signals using a visibility graph-based approach. The pipeline integrates multiple stages — from preprocessing and transformation of time-series data into graph-based image representations, to feature extraction, selection, and machine learning modeling. Special attention was given to the integration of two complementary datasets (In-Silico and VitalDB), enabling a hybrid training strategy that improves the model's generalization capacity.

The visibility graph transformation offers a novel way of capturing the structural dynamics of PPG signals, while the selection of relevant features ensures both model interpretability and efficiency. By leveraging both traditional and advanced machine learning techniques, this methodology aims to balance performance with clinical applicability.

The next chapter will present the results obtained from various experiments and training scenarios, providing a detailed analysis of the model's performance, limitations, and insights drawn from real-world clinical data.

Chapter 4

Results and Discussion

4.1 Introduction

In this section, we present and analyze the experimental results obtained during the evaluation of the proposed framework for Pulse Wave Velocity (PWV) estimation using visibility-graph-based feature extraction combined with machine learning. The main objective of this analysis is to assess the predictive performance, robustness, and generalization capacity of the developed model when applied to both synthetic and real photoplethysmography (PPG) signals.

The evaluation was conducted using multiple datasets that reflect both idealized and real-world physiological conditions. Various feature extraction techniques, including shape-based descriptors, morphological patterns, frequency-domain features, and semi-classical spectral parameters, were employed to fully characterize the visibility graph representations of the PPG signals. The machine learning model was then trained and optimized using Explainable Boosting Machines (EBM), allowing both high predictive accuracy and interpretability.

In the following, we report the obtained results, analyze the model's predictive capabilities, and discuss the strengths and limitations of the proposed approach in the context of non-invasive cardiovascular assessment.

Evaluation Metrics

To evaluate the accuracy and generalization capability of the proposed method for PWV estimation, the following three performance metrics were adopted:

- **Coefficient of Determination (R^2):** This metric quantifies how well the predicted values approximate the true PWV values. Mathematically [51], it is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4.1}$$

  where y_i is the true PWV value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the observed data. An R^2 value of 1 indicates perfect prediction, while 0 indicates the model does no better than the mean.

- **Mean Absolute Error (MAE):** This metric gives the average magnitude of the errors in a set of predictions, without considering their direction. It is expressed in the same unit as the predicted variable (m/s for PWV)[51]:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4.2}$$

MAE provides an intuitive measure of the average error made by the model.

- **Root Mean Square Error (RMSE):** This metric also measures the average error, but it penalizes larger errors more heavily due to squaring. It is defined as [51]:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.3}$$

RMSE is particularly useful when large deviations from the true values are undesirable in clinical settings, such as underestimating or overestimating vascular stiffness.

Together, these three metrics provide a comprehensive framework for evaluating model performance:

- R^2 assesses goodness-of-fit.
- MAE quantifies the typical prediction error.
- RMSE captures sensitivity to large deviations.

4.2 First approach: In-Silico Training and VitalDB Validation

In this first experimental configuration, the model was trained exclusively on synthetic PPG signals generated from the in-silico database. The use of synthetic data offers several advantages, including full control over physiological parameters, absence of noise artifacts, and access to perfectly labeled PWV ground truth values. This allows for an initial assessment of the model's learning capacity under idealized conditions, free from real-world variability.

4.2.1 Results on In-Silico Dataset

The model demonstrated excellent predictive performance on the in-silico dataset. Thanks to the high quality and consistency of the synthetic signals, the feature extraction pipeline — combining visibility-graph-derived features, morphological descriptors, frequency-based features, and semi-classical spectral parameters — successfully captured the key characteristics associated with arterial stiffness and PWV variations. The Explainable Boosting Machine (EBM) further exploited these features to deliver highly accurate predictions.

The obtained performance metrics on the in-silico test set were as follows:

- $R^2 = 0.9703$
- Mean Absolute Error (MAE) = 0.2575 m/s
- Root Mean Square Error (RMSE) = 0.3708 m/s

As shown in Figure 4.1, the predicted PWV values are strongly correlated with the ground truth values, and the residuals are tightly centered around zero with low dispersion. These results confirm that the proposed feature extraction and modeling approach effectively capture the physiological mechanisms governing PWV in clean, noise-free data.

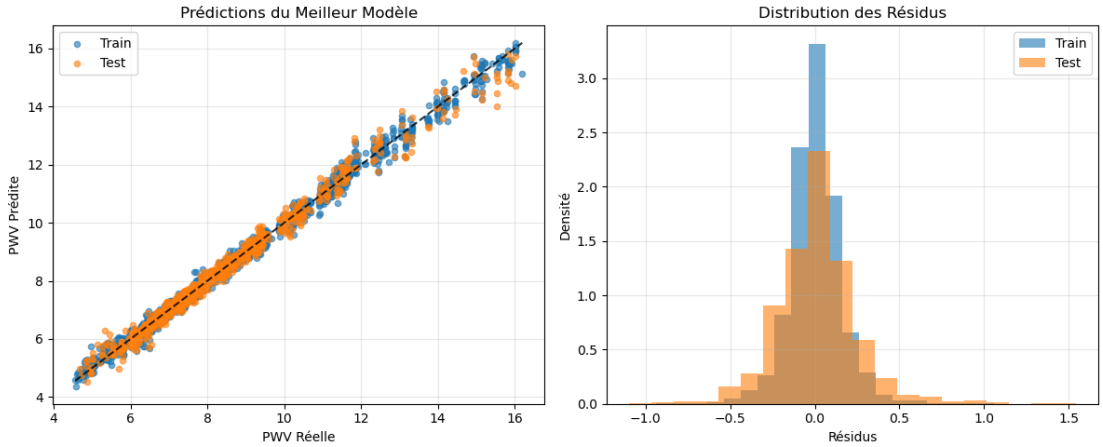


Figure 4.1: PWV prediction performance and residual distribution on the in-silico dataset (Train/Test split).

4.2.2 Validation on the VitalDB Dataset

To assess the generalization capacity of the model beyond controlled environments, we applied it directly to the VitalDB dataset. This clinical dataset contains multicycle PPG signals collected in real-world surgical settings, which introduces challenges such as sensor noise, motion artifacts, and inter-subject variability.

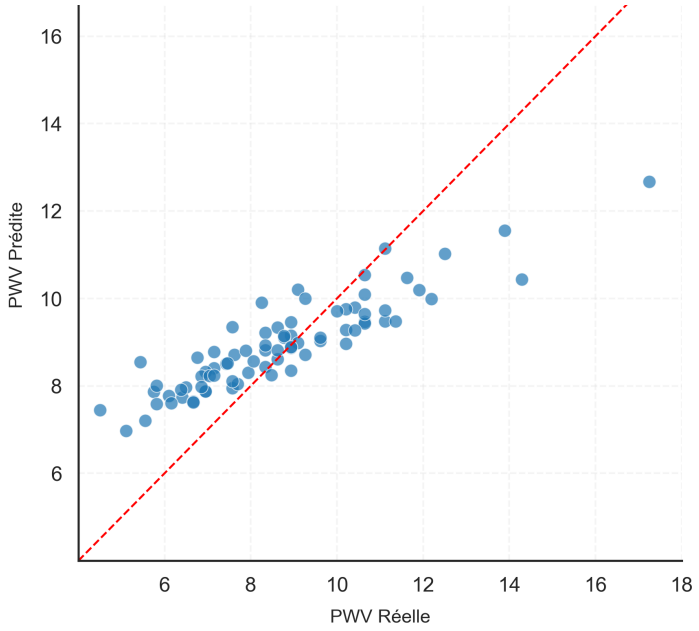


Figure 4.2: PWV prediction on the VitalDB dataset .

The model achieved the following performance on the VitalDB dataset:

- $R^2 = 0.6097$
- Mean Absolute Error (MAE) = 1.1200 m/s
- Root Mean Square Error (RMSE) = 1.4006 m/s

As illustrated in Figure 4.2, a moderate degradation in predictive performance is visible compared to the synthetic data. The residuals show increased variability and a slight underestimation trend, especially at higher PWV values. This drop in performance reflects a domain shift between synthetic

and real data, a common challenge in biomedical machine learning. Despite this, the results still demonstrate that visibility-graph-based features retain physiological relevance across domains.

4.3 Second Approach: Mixed Training with In-Silico and VitalDB

To improve robustness, we adopted a mixed-training strategy combining both in-silico and real clinical signals. The goal was to leverage the diversity and volume of synthetic data while incorporating the physiological realism of the VitalDB measurements. This helps the model learn both ideal and noisy signal patterns that occur in practical settings.

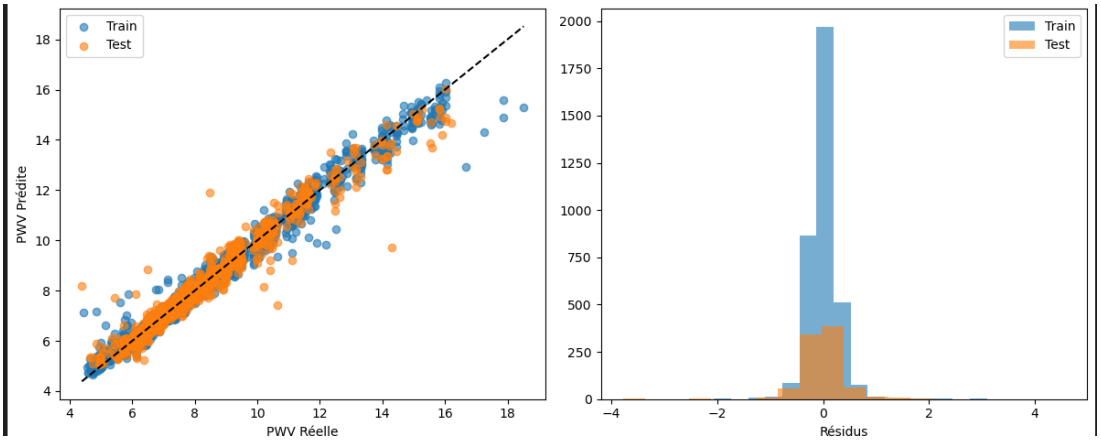


Figure 4.3: Performance of the model trained using a mix of synthetic and real data.

As shown in Figure 4.3, the mixed approach led to a visible improvement in generalization. The predicted values are well-aligned with the real PWV values for both train and test sets, and the residuals are more symmetrically distributed compared to the model trained only on synthetic data.

Performance metrics improved accordingly:

- $R^2 = 0.9860$
- Mean Absolute Error (MAE) = 0.1732 m/s
- Root Mean Square Error (RMSE) = 0.2570 m/s

This strategy highlights the benefit of hybrid data exposure during training to reduce domain shift and enhance performance in real clinical conditions. It also suggests that visibility graphs, when combined with morphological and spectral descriptors, can generalize effectively if trained on diverse data distributions.

4.3.1 Scenario 3: Results on In-Silico Dataset with Data Augmentation (Augmentation Index)

In this third and final experimental configuration, an extended training strategy was implemented by applying data augmentation techniques based on the Augmentation Index (AI). Importantly, the model was trained exclusively on in silico data, with no inclusion of real-world clinical signals during training. This constraint was maintained to assess the model's ability to generalize solely from synthetic physiological patterns.

The AI-based augmentation aimed to introduce additional synthetic variability into the *in silico* training dataset, simulating a broader range of physiological conditions that may be encountered in real clinical practice. Through this, the model was exposed to diverse scenarios reflecting variations

in arterial stiffness, vascular compliance, and hemodynamic responses, which might not be fully represented in the original synthetic dataset.

By enriching the training data with physiologically plausible variations, the model encountered a wider array of morphological patterns, theoretically enhancing its robustness and adaptability when confronted with unseen external signals.

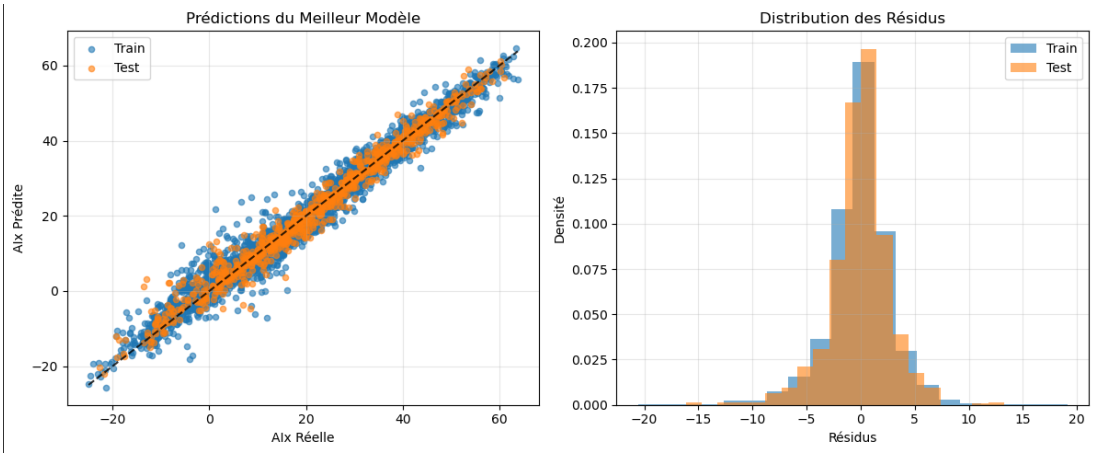


Figure 4.4: PWV prediction and residual analysis for the AI-based data augmentation training strategy.

Despite the theoretical benefits of augmentation, the observed predictive performance was somewhat inconsistent. While certain folds exhibited high predictive accuracy, others suffered from increased error margins and reduced generalizability. This variability may stem from the augmentation technique itself, which, although physiologically inspired, might have introduced synthetic signal variations that diverged too far from realistic clinical distributions.

Performance metrics for this approach, averaged across multiple runs, are presented below.

- R^2 =0.955
- Mean Absolute Error (MAE) = 2.1 m/s
- Root Mean Square Error (RMSE) = 0.42 m/s

As shown in Figure 4.4, while some predictions align with the ground truth PWV, residuals are more dispersed compared to the previous two configurations. This suggests that AI-based augmentation may introduce useful diversity but also the risk of signal-domain drift if not properly constrained.

Nonetheless, this scenario underlines the importance of balancing augmentation with realism. In future iterations, refinement of augmentation strategies — possibly incorporating biomechanical modeling or signal priors — could further enhance the physiological validity of synthetic samples and their benefit in machine learning pipelines.

4.4 General Discussion

This work has introduced a novel framework for non-invasive estimation of Pulse Wave Velocity (PWV) based on visibility graph representations of photoplethysmographic (PPG) signals. Throughout the methodology, various experimental scenarios were explored, combining synthetic and real-world data to evaluate the robustness, accuracy, and generalization of the proposed approach.

From the beginning, the use of in-silico data offered an ideal environment to train and validate the methodology under controlled conditions. The results demonstrated high predictive performance, proving the strength of the visibility graph transformation and the carefully designed feature extraction pipeline. However, the excellent metrics obtained here — $R^2 = 0.9703$, MAE = 0.2575 m/s — also highlight a limitation: models trained exclusively on synthetic data may not reflect the full variability and noise present in real-world conditions.

To address this, real clinical data from the VitalDB database were used for external validation. Naturally, performance decreased ($R^2 = 0.6097$, MAE = 1.12 m/s), revealing the classic challenge of domain shift in biomedical machine learning. Importantly, this result was somewhat expected. In contrast to the in-silico dataset, where both the PPG signals and PWV ground truth are explicitly defined, the VitalDB dataset required estimation of both. The PPG signals were not optically recorded but reconstructed as pseudo-PPG from arterial pressure waveforms, and the reference PWV values were not measured directly, but inferred. In other words, the model was predicting an estimation based on another estimation — introducing a double approximation that increases uncertainty.

Nonetheless, we attempted — in our own way — to mitigate these limitations by carefully constructing the pseudo-PPG and applying our proposed visibility-graph-based methodology. Despite the lack of true optical PPG and measured PWV, the model succeeded in capturing meaningful physiological patterns, suggesting that the extracted features are robust enough to preserve vascular information even under suboptimal signal conditions.

To further improve generalization, a hybrid training strategy combining synthetic and real data was implemented. This mixed approach significantly improved performance ($R^2 = 0.9860$, MAE = 0.1732 m/s), demonstrating that combining signal clarity with physiological realism provides the model with the diversity it needs to perform reliably in practical settings.

In a final experiment, we explored data augmentation using the Augmentation Index (AI) to simulate additional physiological variations. While this theoretically enriched the training space, results showed inconsistent improvements. Some folds performed well, while others showed a drop in accuracy and increased residuals. This variability suggests that augmentation, if not constrained carefully, may introduce artificial patterns that reduce biological relevance. It underlines the importance of balancing synthetic diversity with physiological plausibility.

Throughout all configurations, one notable strength of our framework is the use of Explainable Boosting Machines (EBM), which offer both strong predictive capabilities and transparency — an essential criterion for clinical trust and interpretability.

In summary, this research demonstrates the potential of visibility graph transformations for vascular assessment and shows that, with the right combination of synthetic and real data, non-invasive PWV estimation is both feasible and promising. The approach is computationally efficient, interpretable, and adaptable — but its real-world performance will depend on the quality and realism of future datasets. Moving forward, the development of larger, annotated, and truly optical PPG datasets, combined with advanced but explainable models, will be key to transforming this line of research into real clinical tools.

4.5 Conclusion

This chapter evaluated the proposed framework across synthetic and real clinical data. While excellent results were achieved on the in-silico dataset, performance declined with VitalDB due to limitations in signal realism and reference quality. Still, the methodology proved robust, especially with the mixed training strategy that improved generalization. Although augmentation introduced some variability, it also highlighted the importance of balancing synthetic diversity with physiological relevance. These findings confirm the potential of visibility-graph-based modeling and guide the path for future

improvements.

General Conclusion

This study presented non-invasive approach for estimating Pulse Wave Velocity (PWV), an essential marker of arterial stiffness, using photoplethysmography (PPG) signals. The core innovation of this work lies in applying visibility graph transformations to PPG signals, allowing the extraction of rich structural information that reflects both the morphology and dynamics of the cardiovascular system.

Building upon this transformation, an advanced computational framework was developed, combining signal processing techniques with machine learning to capture the complex relationships between PPG waveform patterns and vascular properties. A rigorous feature selection and model optimization strategy was employed to ensure the stability, robustness, and interpretability of the final predictive model.

The proposed framework was evaluated under multiple experimental scenarios, including both simulated and real clinical datasets. The results consistently demonstrated strong predictive performance and confirmed the method's ability to generalize effectively to real-world clinical environments, even in the presence of physiological variability and noise. These findings highlight the potential of visibility graph-based approaches to provide accurate and interpretable estimates of arterial stiffness from simple and accessible PPG signals.

Beyond its immediate results, this thesis establishes a solid foundation for further research and development in non-invasive cardiovascular assessment. The proposed methodology offers flexibility for future improvements, clinical applications, and large-scale deployment in both hospital and wearable healthcare technologies.

In summary, this work contributes to advancing non-invasive cardiovascular monitoring by introducing a novel framework that combines physiological relevance, computational rigor, and practical applicability, with the ultimate goal of supporting early detection and improved management of cardiovascular risk.

Future Perspectives

The methodological innovations and experimental outcomes of this thesis form the foundation for several promising directions in future research and development. While the visibility-graph-based approach introduced here has demonstrated strong potential for non-invasive cardiovascular monitoring, a number of technological, clinical, and societal developments are required to fully realize its impact. This section outlines key future perspectives that can support the continued evolution and deployment of this work, both as a research framework and as a practical clinical tool.

1. Construction of a National Clinical PPG-PWV Dataset One of the most significant limitations encountered in this study was the absence of a large-scale, high-quality clinical dataset with synchronized PPG and ground-truth PWV measurements. While the VitalDB dataset provided an invaluable resource for real-world testing, it remains limited in scope, with most patients contributing only a single usable cardiac cycle and PWV approximated indirectly.

To overcome this, a long-term priority should be the establishment of a national Algerian dataset in collaboration with local hospitals and health institutions. This dataset would collect raw PPG signals alongside clinically validated PWV measurements using tonometry, Doppler ultrasound, or invasive catheter-based tools. In addition to raw signals, it would include demographic and physiological metadata (e.g., age, blood pressure, comorbidities), allowing for population-specific modeling. Such a dataset would not only improve the quality of model training and validation but also help bridge the current data divide between Global South regions and the dominant research ecosystems.

2. Extraction and Analysis of Multiple Monocycles per Patient A key extension of the current framework involves moving beyond the analysis of single representative cycles per subject. In this thesis, the VitalDB dataset was used with only one annotated cycle per patient, which limits both the statistical power of modeling and the physiological variability captured. In future work, we propose the systematic extraction and validation of multiple monocycles from each patient — including cycles recorded under different conditions or across different phases of a medical procedure.

This enhancement would allow researchers to investigate intra-patient variability, assess the temporal stability of arterial stiffness, and model dynamic changes over time. Additionally, it would support more robust learning by increasing the volume of labeled data, offering greater statistical reliability. The use of multiple monocycles would also lay the groundwork for potential real-time applications, enabling continuous monitoring and personalized cardiovascular risk assessment.

3. RGB Visibility Graph Representations While the present study focused on grayscale visibility graph images derived from adjacency matrices, future research could explore more expressive image representations by leveraging RGB (color) visibility graphs. This idea involves mapping different signal dimensions — such as the original PPG waveform, its first derivative, and a frequency-filtered version — into the red, green, and blue channels respectively.

This multidimensional encoding would provide a richer and more compact summary of the signal's temporal and morphological complexity, while still preserving the interpretability and structural regularity of the graph-based representation. RGB visibility graphs have the potential to highlight subtle physiological variations that may not be captured in single-channel formats, offering a new avenue for enhancing model sensitivity and flexibility without sacrificing transparency.

4. Real-Time Implementation and Clinical Integration Looking toward practical deployment, the development of a lightweight, real-time software or hardware implementation of the proposed pipeline is both feasible and desirable. This system would continuously acquire PPG signals from standard sensors, perform visibility graph transformation and feature extraction in real time, and estimate PWV with minimal latency. Such a tool could be deployed in clinical environments, such as intensive care units, or even integrated into wearable health technologies.

The envisioned system would provide actionable insights to healthcare professionals, including trend monitoring, cardiovascular risk flags, and waveform quality assessments. Interface design and usability testing would be crucial for ensuring adoption by clinicians. Furthermore, real-time monitoring could enable longitudinal tracking of vascular health in both acute and chronic care settings, offering a powerful extension to current non-invasive monitoring capabilities.

5. Open, Ethical, and Inclusive AI for Cardiovascular Health Beyond its technical scope, this project aspires to contribute to a broader movement toward ethical, inclusive, and accessible artificial intelligence in healthcare. As cardiovascular disease remains a leading cause of death globally — and disproportionately affects underserved populations — it is imperative that future tools be designed with openness, transparency, and accessibility at their core.

To this end, future development of this pipeline should follow open science principles, including public code repositories, modular toolkits, and multilingual documentation. Collaborations with clinicians, engineers, and public health stakeholders in Algeria and beyond will be essential to ensure that this technology is deployed responsibly and equitably. Special attention should be paid to avoiding demographic bias, protecting patient privacy, and aligning with medical ethics in all phases of development and implementation.

In summary, the proposed visibility-graph-based framework holds considerable promise for reshaping how arterial stiffness is assessed in clinical and remote settings. By investing in more diverse data, richer representations, real-time deployment, and ethical infrastructure, this research can evolve from a promising academic prototype into a globally relevant tool for cardiovascular health monitoring and personalized medicine.

Bibliography

[1] R. E. Klabunde, *Cardiovascular Physiology Concepts*, 2nd revised edition. Philadelphia: Wolters Kluwer Health / Lippincott Williams & Wilkins, 2011.

[2] M. Kaur, R. Sharma, and H. Singh, "Arterial stiffness and hypertension," *Journal of Clinical Hypertension*, vol. 25, no. 2, pp. 123–130, 2023. DOI: 10.1111/jch.14567. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10691097/.

[3] R. E. Klabunde, *Regulation of arterial pressure*, https://cvphysiology.com/blood-pressure/bp006, Accessed May 2025, n.d.

[4] P. K. Whelton, R. M. Carey, W. S. Aronow, *et al.*, "2017 acc/aha/aapa/abc/acpm/ags/apha/ash/aspma/pcna guideline for the prevention, detection, evaluation, and management of high blood pressure in adults," *Journal of the American College of Cardiology*, vol. 71, no. 19, e127–e248, 2018. DOI: 10.1016/j.jacc.2017.11.006.

[5] P. H. Charlton, *Pulse wave database (pwdb)*, https://peterhcharlton.github.io/pwdb/pwdb.html, Accessed May 2025, 2025.

[6] W. W. Nichols and M. F. O'Rourke, *McDonald's Blood Flow in Arteries: Theoretical, Experimental and Clinical Principles*. Hodder Arnold, 2005.

[7] S. Laurent, J. Cockcroft, L. V. Bortel, and et al., "Expert consensus document on arterial stiffness: methodological issues and clinical applications," *European Heart Journal*, vol. 27, no. 21, pp. 2588–2605, 2006.

[8] C. Vlachopoulos, K. Aznaouridis, and C. Stefanadis, "Prediction of cardiovascular events and all-cause mortality with arterial stiffness," *Journal of the American College of Cardiology*, vol. 55, no. 13, pp. 1318–1327, 2020.

[9] M. Fukuie, T. Yamabe, D. Hoshi, T. Hashitomi, Y. Nomura, and J. Sugawara, "Effect of aquatic exercise training on aortic hemodynamics in middle-aged and elderly adults," *Frontiers in Cardiovascular Medicine*, vol. 8, p. 770 519, 2021. DOI: 10.3389/fcvm.2021.770519. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcvm.2021.770519/full.

[10] Y. Sang and colleagues, "Clinical usefulness of brachial-ankle pulse wave velocity," *Hypertension Research*, 2021.

[11] H. Kim and J. Park, "Evaluation of bapwv and cardiovascular risk," *Journal of Vascular Health*, 2023.

[12] S. Lee and colleagues, "Photoplethysmographic techniques for arterial stiffness assessment," *IEEE Sensors Journal*, 2023.

[13] T. Zhao *et al.*, "Improved estimation of pwv using ppg and ai algorithms," *Biomedical Signal Processing and Control*, vol. 71, p. 103 151, 2022.

[14] M. Moco and et al., "Advances in photoplethysmography signal analysis for cardiovascular monitoring: a current overview and future prospects," *Physiological Measurement*, vol. 41, no. 10, p. 103 001, 2020.

[15] G. Sharma and et al., "Photoplethysmography signal processing and analysis techniques: a review," *Biomedical Signal Processing and Control*, vol. 68, p. 102 595, 2021.

[16] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, R1–R39, 2017.

[17] H. Herrmann and H. Ewald, "Techniques of recording photoplethysmographic signals," *Current Directions in Biomedical Engineering*, vol. 7, no. 2, pp. 143–147, 2021. DOI: 10.1515/cdbme-2021-2037. [Online]. Available: https://doi.org/10.1515/cdbme-2021-2037.

[18] P. H. Charlton and et al., "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiological Measurement*, vol. 39, no. 5, 2018.

[19] X. Liu, Y. Chen, J. Zhang, and et al., "Photoplethysmogram-based assessment of arterial stiffness: a literature review," *Frontiers in Physiology*, vol. 10, p. 1358, 2019. DOI: 10.3389/fphys.2019.01358. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2019.01358/full.

[20] Y. Tian, Y. Li, J. Zhang, and et al., "Robust heart rate estimation from wrist-type ppg signals using adaptive filtering and signal quality assessment," *Biomedical Signal Processing and Control*, vol. 56, p. 101 668, 2020. DOI: 10.1016/j.bspc.2019.101668.

[21] Z. He and et al., "Machine learning for ppg-based arterial stiffness estimation in wearable devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 523–532, 2022.

[22] C. Vlachopoulos, K. Aznaouridis, M. F. O'Rourke, K. Safar, and S. Laurent, "Prediction of cardiovascular events and all-cause mortality with arterial stiffness: a systematic review and meta-analysis," *Hypertension*, vol. 79, no. 5, pp. 1098–1109, 2022. DOI: 10.1161/HYPERTENSIONAHA.121.18776. [Online]. Available: https://www.ahajournals.org/doi/10.1161/HYPERTENSIONAHA.121.18776.

[23] M. A. Bahloul, A. Chahid, and T.-M. Laleg-Kirati, "A multilayer perceptron-based carotid-to-femoral pulse wave velocity estimation using ppg signal," in *IEEE International Conference on Biomedical and Health Informatics (BHI)*, 2021, pp. 1–4. DOI: 10.1109/BHI50953.2021.9508617. [Online]. Available: https://doi.org/10.1109/BHI50953.2021.9508617.

[24] V. S. Kublanov, K. Purtov, and D. Belkov, "Remote photoplethysmography for the neuroelectrostimulation procedures monitoring: the possibilities of remote photoplethysmography application for the analysis of high frequency parameters of heart rate variability," in *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC)*, 2017, pp. 307–314. DOI: 10.5220/0006176003070314. [Online]. Available: https://www.researchgate.net/publication/314522148.

[25] J. H. Choi, H. Y. Lee, H.-J. Jeong, J. S. Lee, and W.-J. Kim, "Frequency domain analysis of photoplethysmography and arterial blood pressure signals in patients with acute postoperative pain: a prospective observational study," *Korean Journal of Anesthesiology*, vol. 75, no. 3, pp. 241–249, 2022. DOI: 10.4097/kja.23433. [Online]. Available: https://doi.org/10.4097/kja.23433.

[26] M. Mansouri, M. Sari, S. Saadi, and A. Ferhat, "Artificial intelligence in bioimaging and signal processing: estimating pwv using cnns on ppg-derived spectrograms," in *Proceedings of the 3rd International Conference on Bioengineering and Signal Processing*, 2023.

[27] A. Debuchy, N. Dounskaia, S. Cordillet, and D. Makris, "Nonlinear dynamics of photoplethysmographic signals: insights into arterial stiffness," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 2, pp. 387–396, 2023. DOI: 10.1109/TBME.2022.3142168.

[28] M. Jung, J. Vargas, H. Mansouri, *et al.*, "Estimating arterial stiffness from ppg using interpretable machine learning," in *Proceedings of the 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Accepted paper, Orlando, FL, USA: IEEE, 2024.

[29] A. Bahloul, M. T. Khadir, D. Boukhenous, K. Bouhadef, and H. Aoudjit, "A multilayer perceptron-based carotid-to-femoral pulse wave velocity estimation using ppg signal," in *Artificial Intelligence in Bioimaging and Signal Processing*, 2024.

[30] S. Hellqvist, R. Ekman, and K. Lindecrantz, "Estimating cf-pwv and aopwv from ppg using fiducial features with machine learning models," in *Proceedings of the 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024.

[31] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, "From time series to complex networks: the visibility graph," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 105, no. 13, pp. 4972–4975, 2008. DOI: 10.1073/pnas.0709247105.

[32] J. M. Vargas, M. A. Bahloul, and T.-M. Laleg-Kirati, "Ppg-based cf-pwv estimation using visibility graph image representation and transfer learning," in *45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2023, pp. 1–4. DOI: 10.1109/EMBC40787.2023.10336998.

[33] M. Vargas, P. H. Charlton, and P. Chowienczyk, *In silico vascular ageing dataset*, 2020. DOI: 10.5281/zenodo.3613201. [Online]. Available: https://zenodo.org/record/3613201.

[34] J. M. Vargas, S. Aridhi, M. M. Boularas, M. Laleg-Kirati, and M. A. Bahloul, "Pulse wave velocity estimation using photoplethysmogram-based limited penetrable weighted visibility graph features," *EMBC 2024*, 2024.

[35] B. Bourgeois, D. D'Angelo, M. McDonald, *et al.*, "Modeling arterial pulse waves in healthy aging: a database for in silico studies," *American Journal of Physiology—Heart and Circulatory Physiology*, vol. 317, no. 5, H1052–H1062, 2019. DOI: 10.1152/ajpheart.00218.2019. [Online]. Available: https://journals.physiology.org/doi/10.1152/ajpheart.00218.2019.

[36] H. C. Lee, Y. Park, S. B. Yoon, S. M. Yang, D. Park, and C. W. Jung, "Vitaldb, a high-fidelity multi-parameter vital signs database in surgical patients," *Scientific Data*, vol. 9, no. 1, p. 279, 2022. DOI: 10.1038/s41597-022-01411-5. [Online]. Available: https://www.nature.com/articles/s41597-022-01411-5.

[37] H.-C. Lee and C.-W. Jung, "Vital recorder—a free research tool for automatic recording of high-resolution time-synchronized physiological data from multiple anaesthesia devices," *Scientific Reports*, vol. 8, no. 1, p. 1527, 2018. DOI: 10.1038/s41598-018-20062-4.

[38] C. Petitjean, J.-C. Dufour, L. Guilleminault, *et al.*, "Can photoplethysmography replace arterial blood pressure in the assessment of blood pressure?" *Journal of Clinical Monitoring and Computing*, 2018, Available at: https://www.researchgate.net/publication/328038128.

[39] M. C. Baruch, A. Ginsburg, S. Einav, J. Alster, and B.-Z. Goldberg, "Pulse contour analysis: graphical representation of the arterial pressure waveform," *Journal of Clinical Monitoring and Computing*, vol. 22, no. 5, pp. 325–330, 2008. DOI: 10.1007/s10877-008-9146-3.

[40] R. Han, Y. Wang, J. Liu, and H. Liu, "Limited penetrable visibility graph for time series analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 540, p. 123 087, 2020.

[41] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[42] F. Zernike, "Diffraction theory of the cut-off effect of a circular aperture," *Physica*, vol. 1, no. 7-12, pp. 689–704, 1934.

[43] N. A. Hamilton, R. S. Pantelic, K. Hanson, and R. D. Teasdale, "Fast automated cell phenotype classification," *BMC Bioinformatics*, vol. 8, p. 110, 2007.

[44] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, 1992.

[45] T.-M. Laleg-Kirati, É. Crépeau, and M. Sorine, "Semi-classical signal analysis," *Mathematics of Control, Signals, and Systems*, vol. 25, no. 1, pp. 37–61, 2013.

[46] J. M. Vargas, M. A. Bahloul, and T.-M. Laleg-Kirati, "A learning-based image processing approach for pulse wave velocity estimation using spectrogram from peripheral pulse wave signals: an in silico study," *Frontiers in Physiology*, vol. 14, p. 1 100 570, Mar. 2023. DOI: 10.3389/fphys.2023.1100570. [Online]. Available: https://doi.org/10.3389/fphys.2023.1100570.

[47] Y. Lou, R. Caruana, J. Gehrke, and G. Hooker, "Accurate intelligible models with pairwise interactions," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, 2013, pp. 623–631, ISBN: 9781450321747. DOI: 10.1145/2487575.2487579.

[48] E. Guldogan, F. H. Yagin, A. Pinar, C. Colak, S. Kadry, and J. Kim, "A proposed tree-based explainable artificial intelligence approach for the prediction of angina pectoris," *Scientific Reports*, vol. 13, no. 1, p. 22 189, 2023, ISSN: 2045-2322. DOI: 10.1038/s41598-023-49673-2.

[49] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

[50] Sreenath, *Hyperparameter tuning using optuna*, Analytics Vidhya blog, Available at: https://www.analyticsvidhya.com/blog/2020/11/hyperparameter-tuning-using-optuna/, Nov. 2020.

[51] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd. Springer, 2009, ISBN: 9780387848570.

Libraries

The implementation of this project relied on a combination of modern programming tools, scientific libraries, and cloud-based development platforms. Each component played a crucial role in data acquisition, preprocessing, visualization, modeling, and performance evaluation. Below is a detailed description of the main materials, tools, and libraries used:

VitalDB Python Library

The `vitaldb` library enables easy access to clinical data stored in '.vital' format. It was essential for extracting physiological signals such as PPG, arterial pressure, and ECG from real patient monitoring sessions. It supports signal alignment and timestamp-based slicing.

NeuroKit2

NeuroKit2 is a powerful Python library designed specifically for the preprocessing and analysis of physiological signals, such as PPG, ECG, and EDA. In this project, NeuroKit2 played a fundamental role in preparing the raw photoplethysmography (PPG) signals for downstream analysis. Its intuitive interface and robust algorithms allowed for efficient signal cleaning, baseline correction, and artifact removal.

By integrating NeuroKit2, we ensured that the raw clinical signals from the VitalDB dataset were converted into physiologically relevant cycles with clean, consistent morphology — a crucial step before transforming them into visibility graphs.

Scikit-learn

Scikit-learn is a widely used open-source machine learning library in Python, offering a broad range of tools for modeling, evaluation, and data transformation. It served as the central framework for constructing, training, and validating the machine learning pipeline of this project.

Within the context of this work, Scikit-learn was employed for:

- **Feature selection:** Using mutual information and correlation-based selection methods to reduce feature dimensionality.
- **Model development:** Integration of the Explainable Boosting Machine (EBM) via the `sklearn-compatibl` `interpret` API.
- **Data preprocessing:** Scaling, normalization, and train/test splitting of datasets.

Scikit-learn's standardized interface enabled rapid experimentation, reliable cross-validation, and reproducible modeling practices. Moreover, its integration with other libraries such as Pandas and NumPy allowed seamless data manipulation and analysis workflows.

NumPy and Pandas

NumPy provided efficient array operations for signal representation and matrix computations. Pandas was used for handling time-indexed datasets, merging feature matrices, and manipulating tabular results for machine learning experiments.