Jugular Venous Height Measurement Through Contactless Monitoring

Mohamad Hosein Davoodabadi Farahani

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the degree of Master of Applied Science, Biomedical Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Prof. Miodrag Bolic

Abstract

Heart failure remains the leading cause of mortality worldwide, underscoring the urgent need for early diagnosis to improve patient outcomes and alleviate the burden on healthcare systems. One of the key indicators of heart failure is elevated Jugular Venous Pressure (JVP), which reflects increased venous pressure and fluid retention when the heart fails to pump blood effectively. Traditional methods of measuring JVP are often time-consuming and expensive. However, the advent of smartphones equipped with high-quality cameras and advanced processing units present an opportunity for a more efficient and cost-effective approach. This research proposes a novel framework utilizing machine learning and signal processing techniques to measure the JVP height in the neck through smartphone video recordings. The framework enables patients to capture and assess their JVP height by simply using their smartphones, offering a practical alternative to conventional methods. To evaluate the effectiveness of the proposed system, a dataset comprising videos of patients recorded from various angles and distances at a cardiopulmonary clinic was utilized. The algorithm's performance in detecting JVP height was compared with assessments made by medical professionals using ultrasonic methods. The results demonstrate that the proposed system is comparable in terms of accuracy in JVP height measurement with expert's evaluations. This advancement suggests a promising solution for improving the accessibility and efficiency of heart failure diagnostics, potentially transforming patient self-monitoring and reducing healthcare costs.

Acknowledgements

I would like to extend my deepest gratitude to those who have supported and contributed to the completion of this thesis. First and foremost, I want to thank my supervisor, Prof. Bolic, for his unwavering support, guidance, and encouragement throughout this research journey. His expertise and insightful feedback have been invaluable in shaping this work. I extend my heartfelt gratitude to Dr. Roland Sabbagh for kindly granting us permission to record videos at his clinic. His invaluable comments and assistance significantly expedited our research and enriched the quality of our findings. I am also grateful to the people at the clinic who assisted with data acquisition and data labeling, ensuring the accuracy and reliability of our study. Lastly, I wish to express my heartfelt appreciation to my parents, whose continuous support and encouragement have been a source of strength and motivation throughout this process.

Author Contributions

This study presents a smartphone-based framework for the measurement of JVP using deep learning and signal processing. The key contributions of this research are summarized below:

- **Advanced Feature Extraction:** We developed a deep learning network to reliably detect and track key anatomical landmarks, including the ear, clavicle, and sternal notch.
- **Integrated Physiological Metrics:** The network measures Heart Rate (HR) and Respiration Rate (RR) within a specifically defined Region of Interest (ROI) on the neck, providing critical physiological inputs.
- **Novel JVP Assessment Criterion:** By integrating the mentioned measurements, we introduce new, data-driven criteria for assessing JVP height.
- **End-to-End Processing Pipeline & Clinical Validation:** The complete processing pipeline is validated using a comprehensive, clinically annotated dataset, ensuring robustness and practical utility in real-world scenarios.
- **Publications Supporting This Work:**
  - *"Non-Contact Imaging of the Neck for Monitoring Cardiopulmonary Clinic Patients"*: Focused on HR and RR detection based on the automatic ROI detection on the neck.
  - *"Non-contact Heart Rate and Respiratory Rate Estimation from Videos of the Neck"*: Intrdouced a novel way of HR and RR prediction.

Contents

List of Figures

List of Tables

List of Abbreviations

AFib Atrial Fibrillation 15

CCA Common Carotid Artery 22

CHD Coronary Heart Disease 1

CNN Convolutional neural network 28

COPD Chronic Obstructive Pulmonary Disease 16, 17

cVCGI Camera-based vibrocardiography imaging 19, 26

CVD Cardiovascular diseases 1, 2

CVP Central Venous Pressure 13, 18–20, 22, 23, 26

DCNN Deep Convolutional neural network 28

ECG Electrocardiogram 14, 25, 46–48, 80, 92

EJV External Jugular Vein 13, 22, 26, 58

EVM Eulerian Video Magnification 31, 32, 35, 94

FFT Fast Fourier Transform 57

FPS Frames Per Second 50–52, 94

HF Heart Failure 1, 2

HR Heart Rate iv, 5, 25, 26, 53, 66, 75, 77, 78, 80, 82, 99

IJV Internal Jugular Vein 3, 9, 13, 14, 21–24, 26, 58, 59

JVP Jugular Venous Pressure ii, iv–vi, 2–9, 11, 12, 14–17, 21–26, 40, 43, 44, 49, 51–53, 56–59, 65, 66, 84, 91, 101

LED Light Emitting Diode 92

NHANES National Health And Nutrition Examination Survey 2

NIR Near-Infrared 92, 94

PPG Photoplethysmography 4, 14, 23, 24, 26, 47, 75, 92

RAP Right Atrial Pressure 2, 15, 17, 20

ROI Region of Interest iv, 26, 53–55, 84

rPPG Remote Photoplethysmography 24, 25, 46, 51, 77, 78, 80, 82

RR Respiration Rate iv, 5, 26, 56, 66, 75, 78, 82, 99

sMOT Skin-Motion 4, 77, 78, 80, 82, 99

SNR Signal to Noise Ratio 51, 75, 88

TR Tricuspid regurgitation 15

US Ultrasound 19, 22, 23, 26, 44, 98

YOLO You Only Look Once 29, 56, 67

1

Introduction

1.1 Background

Cardiovascular diseases (CVD) are the leading cause of morbidity worldwide, surpassing cancer and chronic lower respiratory diseases in terms of annual fatalities. In the U.S., the prevalence of CVD encompassing Coronary Heart Disease (CHD), Heart Failure (HF),

stroke, and hypertension was 48.6% among adults over the age of 20 in 2020, totalling approximately 127.9 million individuals, with prevalence increasing with age for both genders. Globally, CVD was responsible for an estimated 19.05 million deaths in 2020, marking an 18.71% increase since 2010 [1]. The incidence of HF notably escalates with age, showing up to a fourfold higher prevalence (8.0%–9.1%) in U.S. adults aged 65 and older compared to those under 65. Projections indicate that the number of Americans with HF will rise to 8.5 million by 2030, according to National Health And Nutrition Examination Survey (NHANES) data. There is a unanimous opinion that key global risk factors for HF include advancing age, ischemic heart disease, hypertension, obesity, diabetes, and smoking [2]. From a financial perspective, the economic burden of CVD is substantial, with the average annual direct and indirect costs in the United States estimated at $407.3 billion between 2018 and 2019, and direct costs having doubled from 1997 to 2019. Notably, hospitalizations for HF represent a major contributor, accounting for up to 80% of direct healthcare expenditures [1].

HF can present in diverse ways, with right-sided HF or fluid accumulation in the body often indicating a more severe and potentially life-threatening condition. Fluid overload can be assessed and managed through the measurement of JVP, which serves as an indirect indicator of (Right Atrial Pressure (RAP)). The heart comprises four chambers: the left and right atria and the left and right ventricles. The right atrium is connected to the superior vena cava, a large vein that channels deoxygenated blood from the brain and face into the heart.

The superior vena cava branches into the internal and external jugular veins, which are

utilized for measuring JVP. The internal jugular vein, located approximately 1 to 2 cm below the surface of the neck, is preferred for JVP measurement due to its larger diameter and more direct alignment with the superior vena cava and right atrium. In contrast, the external jugular vein, visible on the skin's surface, has a more complex path with two 90-degree angles between the right atrium and the neck segment. These angles can introduce slight discrepancies in pressure readings, making the external jugular vein less reliable for accurate JVP measurement. The pressure in these veins (i.e., the right internal jugular vein Internal Jugular Vein (IJV)) is closely related to the pressure in the right atrium due to their direct connection.

1.2 Problem Statement

Accurate measurement of JVP is a critical component in the diagnosis and management of heart failure, as it offers a non-invasive estimate of central venous pressure and provides insights into cardiac function and fluid status [3]. Clinicians traditionally assess JVP at the bedside by visually estimating the vertical distance between the sternal angle and the highest point of jugular venous pulsation. However, this method is subjective, dependent on clinical expertise, and often impractical for routine use in outpatient or remote settings [4]. Although implantable hemodynamic monitors have demonstrated potential for early detection of heart failure, they remain invasive and are not widely used outside of specialized care due to procedural risks, infrastructure requirements, and cost-related considerations [5]. Moreover, existing non-invasive techniques for estimating JVP such as

those based on contact Photoplethysmography (PPG), radar sensors, or infrared imaging typically require controlled lighting conditions, specialized hardware, or direct skin contact [6]–[9]. These constraints limit their feasibility for use in real-world, low-resource, or home-based environments. While recent studies have explored non-contact methods for extracting cardiovascular signals from video recordings, there is a lack of research focused specifically on developing an accessible, AI-based framework for measuring JVP height using consumer-grade smartphones. In particular, the potential application of Skin-Motion (sMOT) for neck-based monitoring remains underexplored. Therefore, a gap exists in creating a practical, low-cost, and scalable solution that leverages modern mobile technology, deep learning, and physiological modeling to enable non-contact measurement of JVP height in everyday settings.

1.3 Research Questions

This study aims to develop a smartphone-based system that enables measurement of JVP using deep learning and signal processing techniques. The goal is to create an accessible and clinically meaningful tool for estimating JVP height from video recordings, potentially supporting remote monitoring and telemedicine applications. To achieve this, the study employs methods such as anatomical landmark detection through artificial intelligence, signal extraction for physiological parameters, and integration of medical guidelines into the computational framework. The research addresses the following key questions:

- How accurately can AI-based models detect relevant anatomical landmarks to support JVP height estimation?
- Is the sMOT method a reliable approach for estimating HR and respiratory rate RR from video recordings?
- Can a robust framework be designed that effectively integrates medical knowledge and clinical constraints to enable JVP height measurement from smartphone-captured videos?

1.4 Objectives

The increasing integration of telemedicine into healthcare, along with the widespread use of smartphones equipped with high-resolution cameras and powerful processors, has created new opportunities for non-invasive, remote patient monitoring. These advancements can help reduce the burden on healthcare systems by enabling early detection and continuous monitoring outside clinical settings, especially for patients with cardiovascular conditions. While several frameworks have successfully demonstrated the remote estimation of physiological signals such as heart rate and respiratory rate, little to no research has focused on the measurement of JVP using smartphones. Existing studies often rely on artificial lighting, controlled environments, and specialized sensors (e.g., laser, ultrasonic, microwave, infrared), which limits their practicality in real-world applications. Therefore, a need for a solution that integrates clinical knowledge with widely accessible technology to enable

practical and scalable JVP assessment is needed. The primary objective of this study is to design and implement a smartphone-based framework for the measurement of JVP height, using deep learning and signal processing techniques. The specific goals of the study are:

- To detect key anatomical landmarks on the neck, such as the clavicle, sternal notch, and ear, supporting accurate JVP height estimation.
- To propose and develop an AI-based framework for JVP measurement that is clinically informed and technically feasible.
- To design a near real-time data processing pipeline capable of handling video input from standard smartphone cameras under typical lighting conditions.

1.5 Significance

The benefits of developing a contactless framework for JVP height measurement brings imminent benefits for patients and healthcare systems that are as follows ;

1. **Patient Comfort and Safety**
    - Traditional methods of measuring JVP typically involve direct physical examination, which may cause discomfort for patients and rely heavily on clinician expertise. A contactless approach could eliminate the need for physical contact, thereby enhancing patient comfort and enabling more consistent assessments.

- Reduced Risk of Infection: In the context of infectious diseases (e.g., COVID-19), minimizing physical contact is crucial. Contactless measurement reduces the risk of transmitting infections between patients and healthcare providers,m

2. **Increased Efficiency and Accessibility**

   - Efficient Assessment: Contactless technology can facilitate faster and more efficient JVP measurements. This efficiency is particularly valuable in high-throughput settings like emergency rooms and large clinics, where rapid patient assessment is essential.
   - Wider Accessibility: A contactless approach can be more easily deployed in various settings, including remote or underserved areas. It can also be integrated into telemedicine platforms, allowing for remote monitoring of patients.

3. **Cost-Effectiveness**

   - Reduced Need for Invasive Equipment: By minimizing reliance on invasive tools and physical examinations, contactless methods can lower the costs associated with medical supplies and equipment.
   - Decreased Labor Costs: Automating the measurement process can reduce the need for manual assessments by healthcare providers, allowing them to focus on more complex aspects of patient care.

4. **Continuous Monitoring and Early Detection**

- Compatibility with Wearable Devices: Contactless JVP measurement can be integrated with wearable health devices, which are increasingly popular for remote health monitoring. This integration supports a more comprehensive approach to patient management.
- Data Analytics and Decision Support: Advanced contactless systems can be equipped with data analytics capabilities, providing valuable insights and supporting clinical decision-making through predictive analytics and trend analysis.

5. **Research Contributions**

   The following are key contributions focused on developing methods for assessing cardiopulmonary parameters. The contributions are reflected in multiple publications, including work on heart rate and respiratory rate estimation from neck video data, as well as JVP measurement.

   - **First Author** – *"Non-Contact Imaging of the Neck for Monitoring Cardiopulmonary Clinic Patients"*, accepted for publication in *Biomedical Optics Express*.
   - **First Author** – Manuscript in preparation to be submitted to a journal.
   - **Co-author** – *"Non-contact Heart Rate and Respiratory Rate Estimation from Videos of the Neck"*, published in *Proceedings of the IEEE EMBC* [10]. The author actively contributed to the dataset preparation.

1.6 Scope and Limitations

Having examined the advantages of contactless patient monitoring, it is also important to address and discuss the limitations associated with this approach. Certain physiological and environmental factors, such as variant anatomies, difficulties with visualization in obese individuals or those with thick necks [11], video acquisition tools and illumination can adversely impact the performance of the proposed framework and need to be addressed before proceeding. In short, the mentioned reasons could be summarized into the following points;

- **Anatomical Variations:** The IJV's anatomy can vary significantly among individuals. Variations include the difference in location of multiple veins, different junction points with the subclavian vein, or variations in the vein's course and size. These variations can affect the ability to accurately assess JVP using traditional methods.
- **Obesity and Thick Necks:** In patients with obesity or thick necks, the increased subcutaneous fat or tissue can obscure the IJV, making direct visualization difficult. This can hinder the accurate assessment of JVP, as the proposed framework relies on the detection of the pulsations.
- **Positioning Sensitivity:** To minimize positioning effects, consistent patient positioning is necessary. Standardizing the measurement procedure, such as using a specific angle of head elevation (typically 30 to 45 degrees), helps in achieving con-

sistent results. Regular training and protocol adherence can mitigate the effects of positional changes.

- **Video Acquisition Tool:** The quality of the extracted signal is significantly influenced by the type of RGB sensor used. Unlike previous studies that utilized high-bit-depth cameras (10 or 12-bit), this study employs two commonplace smartphone cameras. This choice offers both advantages and limitations. While smartphone cameras enhance accessibility and affordability, they lack the ability to capture raw, uncompressed video, leading to potential signal degradation due to compression artifacts, ultimately affecting the signal-to-noise ratio.

- **Illumination:** Using a fixed light source on patients can enhance video quality and improve the extracted signal. However, this study relies on ambient room lighting rather than controlled illumination. While this approach increases the system's applicability and repeatability in real-world scenarios, it also introduces variability that may impact the quality of the captured signals.

- **Technical Factors:** A key challenge encountered during data collection was the lack of standardized setup conditions for the recording equipment. Some recordings were affected by noise and motion artifacts due to smartphones being placed directly on a surface rather than being stabilized with a tripod or dedicated holder. Additionally, the need to frequently adjust the camera angle introduced inconsistencies, impacting the repeatability of video acquisition and reducing measurement reliability.

1.7 Outline of the study

This study is organized into five main chapters, each addressing a specific aspect of this research. The first chapter, Introduction, provides an overview of the research topic, including background information and the problem statement. It discusses the significance of noncontact JVP height measurement and outlines the research objectives and limitations. Following the introduction, the Literature Review chapter examines existing research related to JVP measurement, including contactless monitoring technologies, and the use of smartphone cameras in medical diagnostics. This chapter analyzes previous studies, and current methodologies and identifies gaps in existing knowledge that the research aims to address. The Methodology chapter details the research design and methods used to develop and validate the proposed framework. It describes the experimental setup, data collection procedures, and the algorithms or models employed. Additionally, it discusses the criteria used for evaluating the framework's performance and provides a comprehensive description of the framework's architecture. The Results chapter presents the findings from testing the framework which includes both quantitative and qualitative analyses of the framework's performance, compares these results with existing methods, and interprets the findings in relation to the research objectives. The final chapter, Conclusion, highlights the contributions of the study, discusses its limitations, and suggests directions for future research to further enhance the framework or explore related applications.

2

Literature Review

2.1 Introduction

This literature review synthesizes key research on JVP, emphasizing its anatomical features, the significance of its detection, and the methodologies and findings of relevant studies. Inspecting the neck veins is important for two primary reasons: first, to identify

elevated central venous pressure; and second, to detect specific abnormalities in venous waveforms, which can be indicative of certain arrhythmias as well as valvular, pericardial, and myocardial disorders. Central Venous Pressure (CVP) reflects the pressure in the right atrium and can become elevated in certain conditions. An elevated CVP can cause the neck veins to appear abnormally distended and is often associated with conditions such as chronic lung disease, pulmonary arterial hypertension, and pulmonary stenosis. To accurately measure a patient's CVP, clinicians should focus on assessing the veins on the right side of the neck. This is because the right-sided veins have a direct path to the heart, providing a more reliable measurement of venous pressure. In contrast, veins on the left side of the neck must traverse the mediastinum, where they can be compressed by the aorta. This compression can lead to an elevation in left jugular venous pressure, even when the CVP and right venous pressure are within normal ranges [12]. The IJV is generally preferred for central venous pressure measurement due to its direct alignment with the superior vena cava and the right atrium. In contrast, the External Jugular Vein (EJV) is characterized by two 90-degree bends between the right atrium and its course through the neck. These bends can impede blood flow and create discrepancies in pressure readings. As a result, the pressure in the external jugular vein may differ slightly from the pressure in the heart, rendering it a less ideal candidate for accurate pressure measurement [3].

2.2 JVP Characteristics

When the patient is in a supine position, the IJV is fully open and serves as the primary route for blood draining from the brain and other extracerebral areas. However, when the patient changes to an upright position, the jugular venous system experiences significant flow limitations. In this position, the IJV tends to collapse due to a decrease in transmural pressure caused by gravity, which leads to a notable increase in resistance within the vein [13].

Because the carotid artery also pulsates in the neck, clinicians need to differentiate it from the internal jugular vein. Venous pulsations exhibit a pronounced inward or descending movement, while the outward movement is slower and more diffuse. In contrast, arterial pulsations show a prominent ascending or outward movement, with the inward movement being slow and diffuse [12]. As illustrated in Figure 2.1, dashed vertical lines divide the cardiac cycle into two phases: systole and diastole. Each physiological signal is labelled with its corresponding significant waves. For the Electrocardiogram (ECG), the P wave represents atrial depolarization, the QRS complex indicates ventricular depolarization and the T wave denotes ventricular repolarization. In the JVP tracing, the a, c, and v waves are shown in blue, while the x, x', and y descents are in gray, for the PPG signal, 's' marks the systolic peak, and 'd' marks the diastolic peak [6]. The following Table 2.1 also outlines the differences between Cartoid and JVP pulsations, and a sample of JVP signal is depicted in Figure 2.2. To explore the timing of these events in detail on

the JVP trace: The A wave represents right atrial contraction, while the x descent marks right atrial relaxation. Patients with reduced right ventricular (RV) compliance from any cause can have a prominent a wave and it is absent with Atrial Fibrillation (AFib). The x descent reflects the fall in right atrial pressure after the a wave peak and in normal persons, the x descent is the predominant waveform in the jugular venous pulse [14]. The c wave, once thought to be a carotid artifact, actually corresponds to right ventricular contraction and tricuspid valve closure. The x' descent occurs as the A-V valve ring moves downward during right ventricular contraction. The v wave indicates right atrial filling, which increases venous pressure as it overcomes the descent of the base and is smaller than the a wave because of the normally compliant right atrium. The y descent starts with the opening of the tricuspid valve at the beginning of diastole, causing a sharp decline in venous pressure as the atrium empties into the ventricle [12]. In patients with atrial septal defect (ASD), the a and v waves may be of equal height, and the v wave is accentuated. With Tricuspid regurgitation (TR), the v wave will merge with the c wave. Normally, venous pressure drops by at least 3 mm Hg during inspiration. However, a failure to decrease with inspiration (Kussmaul sign) suggests conditions like constrictive pericarditis, restrictive cardiomyopathy, pulmonary embolism, right ventricular infarction, or advanced systolic heart failure. A Kussmaul sign is observed in cases of right-sided volume overload and reduced right ventricular (RV) compliance.

RAP and respiration are closely linked due to changes in intrathoracic pressure during breathing. During inhalation, the diaphragm contracts, creating negative intrathoracic pressure, which lowers RAP. This enhances the pressure gradient between the systemic

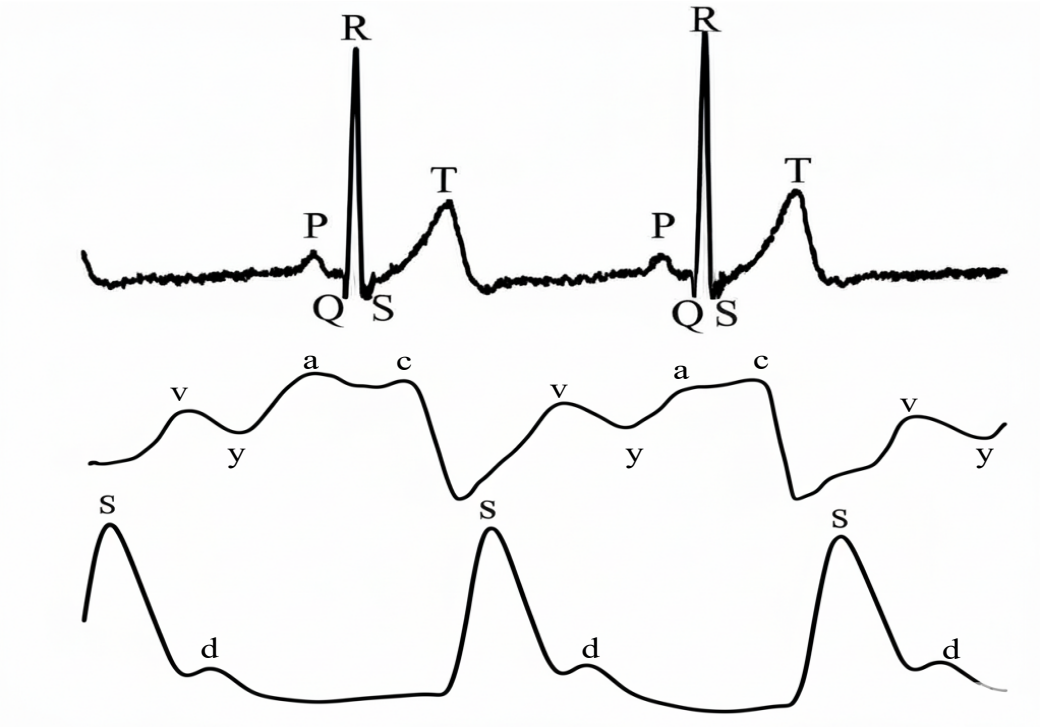

Figure 2.1: ECG, JVP and PPG signals traces (From top to bottom) adapted and redrawn from [6].

veins and the right atrium, improving venous return and increasing right atrial and right ventricular filling. In contrast, during exhalation, intrathoracic pressure becomes less negative or slightly positive, especially during forced exhalation or in conditions like Chronic Obstructive Pulmonary Disease (COPD). Pulmonary artery pulse pressure declines during expiration, especially in patients with low right atrial filling pressure. Increased intrathoracic pressure during expiration can impair both right atrial and right ventricular filling, resulting in a reduced stroke volume. There is a significant interaction between intrathoracic pressure, right atrial filling, and right ventricular output, particularly during

Table 2.1: Distinguishing Internal Jugular Waveforms from Carotid Pulses [12], [14].

Characteristic	***Internal Jugular Vein***	***Carotid Artery***
Character of movement	*Descending movement most prominent*	*Ascending movement most prominent*
Number of pulsations per ventricular systole	*Undulating two troughs and two peaks for every cardiac cycle (biphasic)*	*Single brisk upstroke (monophasic)*
Palpability of pulsations	*Generally not palpable*	*Palpable*
Change with respiration	*During inspiration, pulsations become more prominent and drop lower in the neck*	*No Change*
Change with position	*Pulsations appear lower in neck as patient sits up*	*No Change*
Change with abdominal pressure	*Pulsations may temporarily become more prominent and move higher in neck*	*No Change*
Effect of pressure	*Can be obliterated with gentle pressure at the base of vein/clavicle*	*No Change*

expiration. The study [15] shows that as intrathoracic pressure rises during expiration, pulmonary artery pulse pressure and right ventricular stroke volume decrease, which is linked to impaired venous return and right atrial filling. The decline in pulmonary artery pulse pressure is most pronounced in patients with normal RAP, where high intrathoracic pressure leads to low transmural pressure in the right atrium, significantly impairing venous return. In contrast, COPD patients with higher RAP (often associated with pulmonary hypertension) do not experience as much of a decline in pulmonary artery pulse pressure during expiration. This is because the dilated right atrium acts as a reservoir, maintaining

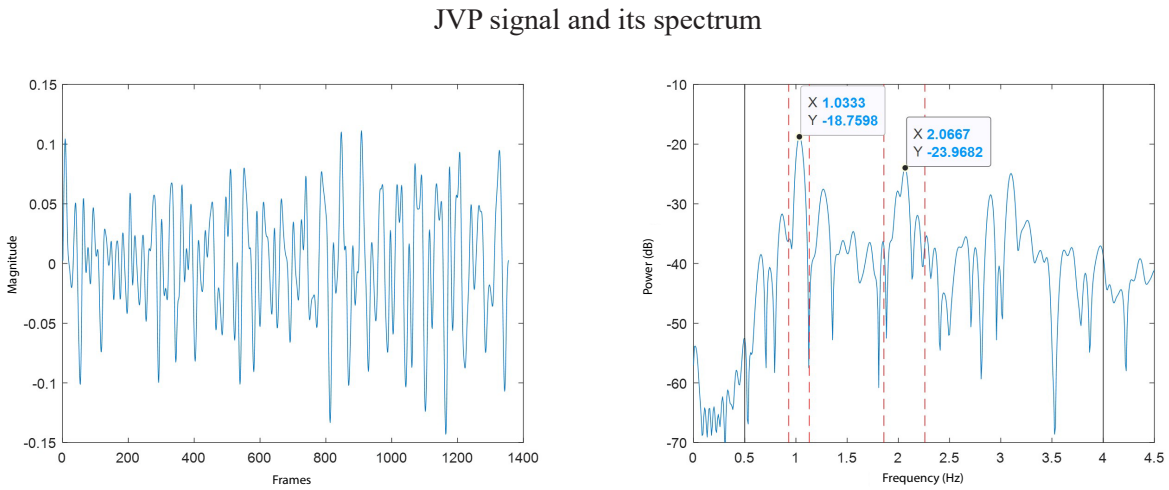

Figure 2.2: Sample of a JVP signal (Left) with its power spectrum (Right).

right ventricular filling despite the increased intrathoracic pressure.

2.3 Background on CVP

Central venous pressure CVP can be measured using three primary approaches, along with emerging methods that will be discussed subsequently.

2.3.1 Clinical Examination

This method estimates CVP by using the jugular veins as manometer tubes to the right atrium. Two reference points are needed to determine the height of the blood column in the jugular veins. The first point is the position of the right atrium, often referred to as the "zero reference point." Since the right atrium cannot be directly located on the thorax, the

angle of Louis, which is at the junction of the manubrium and the body of the sternum, is commonly used as a substitute. This landmark is typically 5 cm above the right atrium, regardless of the patient's position. The second point is the top of the column of blood in the jugular vein, which is identified by visualizing the point of collapse. At this point, the vein shows pulsations related to the cardiac cycle, while it is collapsed above this level and distended and non-pulsatile below it. The vertical distance between the top of the jugular pulse and the angle of Louis is measured. Adding 5 cm to this measurement provides the CVP in centimeters of H_2O [16] as shown in Figure 2.3.

The sternal angle is a more reliable reference point for bedside examination because clinicians can consistently and accurately locate it with greater ease. A metric ruler is positioned vertically at the sternal angle, and the vertical distance from this point to the highest point of the jugular venous pulsation is measured. For the patients who have normal CVP, the neck veins are fully distended when supine and completely collapsed when upright. Therefore, a semi-upright position is employed for measurement [12] which is in line with study [8] that examined different patients' positions and compared the results of Camera-based vibrocardiography imaging (cVCGI) with Ultrasound (US).

Although this method is the most established and commonly used for measuring JVP, its accuracy can be limited. This is due to the reliance on the physician's ability to precisely align the measuring tools both vertically and horizontally. Additionally, obese patients pose a significant challenge because a thick layer of fat can attenuate jugular pulsations, making them harder to detect. Furthermore, when the CVP exceeds 25 cmH_2O and the top of the blood column is positioned above the angle of the mandible, it further

complicates accurate measurement [16]. On the other hand, the non-invasive nature of this method is beneficial as it requires minimal medical equipment and provides a simple way to approximate values [3]. In summary, the estimated height of venous pressure reflects

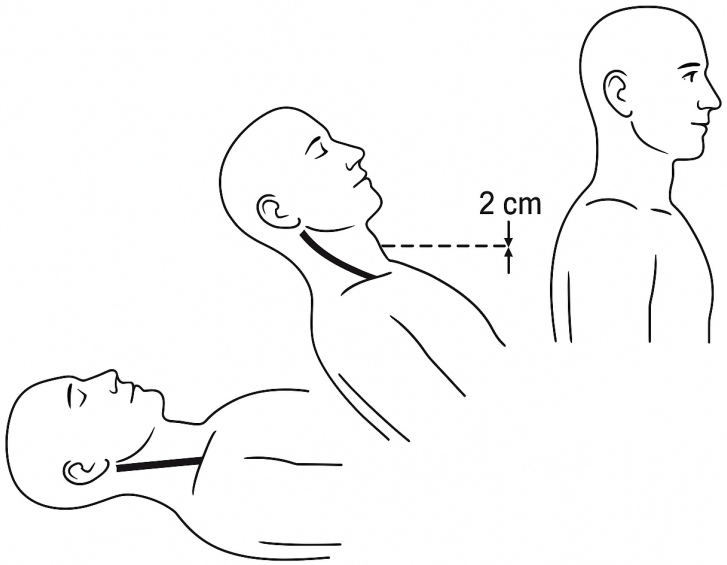

Figure 2.3: Measurement process of venous pressure. The clinician needs to adjust the patient's position until the top of the neck veins becomes visible. In the semi-upright position, the top of the neck veins is 2 cm above the sternal angle, which corresponds to a CVP of 7 cmH_2O (2 cm + 5 cm) adapted and redrawn from [12].

CVP or RAP. While observers often differ in their CVP assessments, recognizing that the pressure is elevated, rather than knowing the exact value, can still aid in diagnosis and treatment decisions. A distance of greater than 3 cm (1.2 inches) in the reclined position is considered abnormal and Venous pulsations above the clavicle with the patient in the sitting position are clearly abnormal because the distance from the right atrium is at least 10 cm (4 inches). The estimated central venous pressure CVP only moderately

correlates with direct measurements. Bedside measurements in centimetres of water must be converted to millimetres of mercury (mmHg) for accurate comparison with catheter-based values, using the conversion factor: 1.36 cm $H_2O = 1.0$ mmHg adapted and redrawn from [12].

2.3.2 Catheter Examination

A less commonly used method for measuring JVP involves direct measurement with a pressure-monitoring catheter. This technique is typically employed in critically ill patients or those undergoing significant fluid shifts, such as during blood transfusions or surgery. Right heart catheterization allows for the measurement of cardiac output, right heart, and pulmonary pressures, enabling the calculation of derived hemodynamic parameters like cardiac work indices and systemic and pulmonary vascular resistance. These key hemodynamic variables provide detailed insights into the patient's physiological condition, which can improve management decisions and support the care of critically ill patients. The first recorded cardiac catheterizations were performed by Hale in 1711, using brass pipes to access the right and left ventricles of a horse's heart via the internal jugular vein and carotid artery.

The procedure begins with the use of ultrasound to locate the IJV, ensuring accurate placement for subsequent pressure analysis. After the patient is induced with general anesthesia, the IJV is accessed with a syringe needle. A guide wire is then advanced through the needle into the vein, followed by the insertion of an intravenous cannula over

the wire, which is done after removing the needle to minimize bleeding at the puncture site. Finally, a pressure-monitoring catheter is inserted through the IJV to provide a direct measurement of JVP.

While this method offers more reliable and consistent results compared to the mentioned non-invasive technique, it is an invasive procedure that carries risks, including catheter malfunction, thrombosis, infection, hemothorax, pneumothorax, and cardiac tamponade. Additionally, improper needle placement can result in damage to nearby arteries or nerves, potentially leading to significant vascular or neurological complications [17].

2.3.3 Ultrasound Measurement

IJV is best examined by real-time US with a high-frequency linear transducer, typically operating between 7 to 9 MHz. In the transverse plane, the IJV appears as a black oval structure on either side of the neck. The Common Carotid Artery (CCA) is also black but is rounder than the adjacent IJV, exhibits more pronounced pulsations in real-time, and is not compressible. During the Valsalva maneuver[1], the size of the IJV increases, whereas the CCA remains unchanged. An alternative approach would be by taking a deep inspiration or passive leg elevation. The EJV is typically not visualized sonographically due to its smaller size and high compressibility. In a sitting position, a patient with normal CVP will have a nearly collapsed internal jugular vein (IJV), which may be barely visible in the transverse

[1]The Valsalva maneuver involves forced expiration against a closed glottis. It simulates various everyday activities, such as straining during defecation, inflating a balloon, or playing a wind instrument like the saxophone

plane. The IJV will briefly distend during forced expiration or the Valsalva maneuver but will quickly collapse with normal breathing. When in a semi-upright position, the top of the blood column will be below the clavicles, and the IJV will collapse mid-neck. It should be noted that a normal CVP would be between 0 to 10 cmH_2O while for an elevated CVP it would be above 10 cmH_2O [16]. In [18], a study is conducted to compare the invasive measurements with ultrasound and visual inspection of JVP on the neck. The correlation of 0.79 between invasive and US measurements is yielded, while the visualization of JVP was possible for only 69% patients with traditional approach compared to 100% for the US measurements leading to this conclusion that the traditional bedside measurements could be enhanced by ultrasound examination. The feasibility of obtaining a periodic signal from the neck that illustrates the a, c and v waves and the x and y descents of the JVP is explored using US B-mode through a sequential variation of the IJV-CSA[2] during the cardiac cycle in [19].

2.3.4 Emerging Approaches

New methods for detecting JVP can be divided into two key areas:

1. Wearable Devices: innovative technologies like PPG sensors and biosensors offer continuous, non-invasive monitoring, enabling real-time tracking of JVP changes. Study [6] demonstrates the feasibility of using a reflectance PPG sensor to extract the JVP waveform from the neck. It was shown in their study that JVP-PPG signals were

[2]Cross Sectional Area

inversely correlated with respect to reference arterial finger PPG. [20] demonstrates the effectiveness of a wearable RF sensor capable of noninvasive acquisition of the JVP features. In their research, they proposed a wearable near-field radio frequency (RF) sensor affixed with a neck collar on the clavicle over the internal jugular vein IJV to enable noninvasive JVP sensing. In a similar study [13], researchers explored the feasibility of a wireless, non-invasive wearable device for acquiring the JVP signal, aiming to validate it as a new tool. The device features a capacitive strain gauge sensor, designed to be worn around the neck like a necklace, to measure the JVP waveform, which later a similar approach was used for exercise monitoring to provide additional parameters on cardiac filling and cerebral venous drainage to the widely used heartbeat rate value [21].

2. Radars: Radars are advancing carotid and JVP detection by offering non-invasive, real-time monitoring. This technology provides precise measurements of blood flow and pressure, enhancing diagnostic accuracy and patient care. In [7] the feasibility of monitoring respiration and carotid pulsation from a distance using impulse-radio ultra-wideband (IR-UWB) radar is investigated. A prototype non-contact device that employs microwave radar 24 GHz and 7 mW power has also been studied for JVP measurement [22] as an alternative to the conventional bedside exam.

3. Cameras: Employing a wide range of cameras enhanced diagnostic precision and allowed for more accurate assessments of cardiovascular health. Various studies have been done on the Remote Photoplethysmography (rPPG) among which [23]–[27] and

[28]–[30] are related to the unsupervised and supervised rPPG waveform detection correspondingly, leading to HR estimation. These algorithms are primarily evaluated on datasets such as COHFACE [31], UBFC [32], and PURE [33], all of which utilize the face as the region of interest for processing. One of the first studies on the neck using a camera was done in [34], where a camera was utilized to capture the image from the neck and locate the whereabouts of the JVP, but the scope of the study was limited due to the number of participants and not analyzing the resulted waveforms. In [16], the feasibility of the remote assessment of the JVP using social media apps was studied. To record the videos accurately, experts helped the patients adjust their smartphones, repositioning their heads and changing the angle of the patient's body. Each patient underwent one bedside JVP assessment and up to four remote assessments conducted by different cardiologists. Right heart catheterization was performed on the same day, following the JVP evaluations. The study found that remote JVP estimates correlated with bedside measurements at a rate of 63.5%. A similar study is done in [35] where a video magnification algorithm is used to help cardiologists with the detection of pulsations better and the results proved a better accuracy compared to non-magnified videos. In [36], the theoretical foundation and initial implementation of a non-contact remote system are presented. The system uses a video camera to capture footage of a target area illuminated by a rectangular LED source, introducing a novel vascular imaging technique known as Spatially Resolved Vascular Imaging (SRVI). This method collected waveform data from healthy volunteers and, through comparison with ECG, identified the waveforms as originat-

ing from the jugular vein and carotid artery. In another similar study, [8], using US to verify the patient's posture and skin-site variability within the neck captured by cVCGI is studied. An optical imaging system is proposed in [9] for quantitatively assessing jugular venous response to CVP and they could confirm their studies with JVP waveforms exhibited biphasic wave properties when time-aligned with an electrocardiogram. In an insightful study [37], a technique to directly measure skin deformations caused by the JV pulse using a very accurate subpixel registration algorithm is proposed that uses a laser and a camera to record the displacement of control points on the skin near the IJV and EJV. the negative correlation of the JVP and Carotid is clearly shown and studied in [38] in which finger PPG was compared with resulted signals from the absorbance leading into a spatial correlation map.

In [10], a novel algorithm was proposed for HR and RR estimation using RGB videos captured from a patient's neck with a smartphone camera. The study employed Laplacian pyramid analysis to detect cardiac signals, where temporal signals were extracted through multiscale analysis and processed in 10×10 blocks, followed by principal component analysis (PCA) to obtain reliable HR and RR measurements. Additionally, [10] relied on a manual method for extracting the ROI, whereas this study introduces a deep learning segmentation network to automatically detect the neck region for further processing.

2.4 Theoretical Foundation

The following section reviews the state-of-the-art methods for the constituent components of the proposed framework.

2.4.1 Object Detection:

Object detection is a key task in computer vision, focused on identifying and locating objects like humans, animals, or cars in images. The two most important metrics for object detection are accuracy (in both classification and localization) and speed. Object detection is a key task in computer vision that supports other tasks like instance segmentation, image captioning, and object tracking. Recent advances in deep learning have led to major breakthroughs in object detection, making it a popular research topic. It is now widely used in real-world applications such as autonomous driving, robot vision, and video surveillance. Over the past two decades, the number of publications on object detection has significantly increased [39].

Traditional object detection methods involve three main phases, each with its drawbacks:

- Selection of region: Objects can appear in different parts of an image with varying sizes and aspect ratios. A multi-scale sliding window approach is used to scan the entire image, but it's computationally expensive and leads to many unnecessary selections.

- Extraction of features: Once an object is located, features are extracted using methods like HOG, Haar-like, and SIFT to represent the object. However, creating a perfect feature descriptor is difficult due to variations in background, lighting, and perspective.

- Classification: Classifiers are used to recognize and categorize objects, making the detection process more structured for visual perception.

Overall, traditional object detection methods are hindered by inefficiencies in region selection, limitations in feature extraction, and classification complexities. Additionally, the sliding window approach for generating bounding boxes is both computationally expensive and ineffective. Deep learning has the potential to overcome some limitations of traditional techniques. Recently, it has become prominent for automatically learning feature representations from data, leading to significant improvements in object detection [40]. With the emergence of Convolutional neural network (CNN) models, deep learning enabled deep-level feature extraction with less bias, revolutionizing object detection. However, for effective results, CNNs require large labelled datasets, which also increases processing time. Deep Convolutional neural network (DCNN) act as the backbone for object detection models. As the network becomes more complex with deeper layers and more parameters, feature representation improves. In DCNN-based object detection, the backbone CNN is responsible for extracting features. Various architectures, like LeNet [41], AlexNet [42], R-CNN [43], Faster R-CNN [44], and GoogLeNet [45], are used as backbones for object detection. Object detection methods are categorized into two types: single-stage

and two-stage detectors [46].

A one-stage detector features a simpler architecture and excels in rapid image feature extraction. Models such as YOLO [47], single-shot multibox detector (SSD) [48], RetinaNet [49], and light-weight and adaptive network for multi-scale object detection (LADet) [50] are examples of one-stage detectors that deliver effective results.

Two-stage detectors identify objects through a two-step process The first stage involves detecting the regions where objects are located. In the second stage, deep learning models are used to classify these detected regions. Examples of such detectors include RCNN, spatial pyramid pooling in deep convolutional networks for visual recognition (SPP) [51], Fast RCNN [52], and Faster RCNN.

2.4.2 YOLO

You Only Look Once (YOLO) has been proved one of the efficient object detection frameworks among others. There are various versions of YOLO from YOLOv1 to YOLOv8 and YOLOv8 is widely acclaimed for its versatility, effectively handling diverse vision tasks such as segmentation, tracking, object detection, classification, and pose estimation.

The relationship between Latency and Accuracy is shown, which highlights the trade-off between inference speed (latency in milliseconds per image on an A100 TensorRT FP16) and accuracy (Right). YOLOv8 also achieves better accuracy with faster inference compared to older YOLO versions. In the following Table 2.2 various YOLOv8 networks and their details are described.

Table 2.2: Different YOLO architectures from the smallest to the largest [53].

Model	size	mAP	Speed(CPU)	Speed (A100)	params(M)	FLOPs(B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

2.4.3 Video Magnification:

This is a technique used to enhance and amplify subtle changes in videos that are typically invisible to the naked eye. It works by highlighting small variations in motion, colour, or intensity, revealing details such as minute vibrations, slight deformations, or even physiological signals like a heartbeat or breathing. Some small differences in videos are often hard for the human eye to detect due to limited spatial perception. In biomedical applications, these subtle changes can reveal important physiological information, such as blood flow, heart rate, and breathing. For example, shifts in skin tone caused by blood flow can indicate heart activity, while slight oscillations in the head due to the carotid artery's pulse can be used to measure heart rate. Similarly, video analysis can amplify motion related to blood pressure in various body parts and respiration, making these processes visible and measurable. Video magnification techniques have various medical applications, such as SpO2 extraction, heart failure monitoring, and surgery. They are also used in laparoscopy, detection of limb ischemia, examining tympanic membrane mobility, assessing vascular function during surgery, free flap monitoring, retinal angiography, and detect-

ing stress, anxiety, and depression through visual micro-motions [54]. This technique is valuable in fields like medicine, where it can monitor vital signs, structural engineering to detect small vibrations or deformations in buildings, and surveillance to track minute movements or hidden details in video footage. Motion magnification techniques act like virtual microscopes, amplifying small, imperceptible motions in videos. Unlike optical microscopes, these software-based methods enhance temporal variations in videos without the need for extra equipment, making them cost-effective and low-maintenance. These techniques provide an affordable alternative to traditional magnification, making them appealing for various research and practical applications [55]. Motion magnification techniques are categorized into Lagrangian and Eulerian methods. Lagrangian methods track specific points or regions in a video over time, making them time-consuming and computationally intensive. In contrast, Eulerian methods measure motion at fixed locations, offering a more efficient, field-based approach. Both methods effectively capture subtle motion signals from video with minimal manual effort. Fluid mechanics studies fluid motion using two main approaches: Lagrangian and Eulerian. The Lagrangian method tracks individual fluid particles over time, similar to following someone's movements, but it can be complex and costly. The Eulerian approach, however, observes fluid properties at fixed points, like watching water flow from a bridge. This method is more practical for technologies requiring fixed locations, such as tracking objects at specific pixels in video processing [55].

Video magnification using optical flow and a Lagrangian approach was first introduced in [56], however, it was too complex and time-consuming for practical use. Eulerian Video

Magnification (EVM) [57] was developed later, which simplifies calculations and amplifies subtle temporal changes in videos. However, EVM introduces noise and artifacts when magnifying larger spatial frequencies. using a phase-based approach and steerable pyramids in [58] improved upon EVM, reducing noise and allowing for greater magnification. A study introduced the EVM-enhanced (E2VM) [59] post-processing technique to amplify motion without increasing noise. E2VM builds on EVM for spatio-temporal motion analysis and uses image warping between original and magnified videos. It allows for greater enlargement and reduces the impact of frame noise by not altering pixel values. However, the process is time-consuming, and some amplification requirements may be missed during image warping. Later a framework [60] using wavelet decomposition and denoising was proposed to address noise and processing time issues, achieving better results than previous methods.

- **Lagrangian Motion Magnification**: Lagrangian approach was first introduced in [56] to amplify subtle motions in video sequences by tracking particle trajectories over time. This method groups pixel clusters based on position, intensity, and motion similarities. A motion similarity measure is used to categorize these motions, allowing users to selectively magnify smaller motions while leaving larger ones unchanged. The approach is simple and effective for enhancing small motions and can be applied in various fields. The approach for motion magnification begins by registering input frames to correct the camera shake and remove small motion artifacts. Initial tracking is performed to detect the motion of feature points, with outliers being discarded. Feature points are then tracked across the sequence, and their trajectories are clus-

tered into motion groups based on similarity, allowing small and large motions to be grouped if they share a common cause. Each pixel is assigned to a motion cluster using motion data, with additional factors like position and proximity considered, and graph cuts are used to ensure optimal assignment. The user can choose a specific motion layer for amplification. Finally, the video is rendered, starting with a constant background layer, copying outlier pixels, and then adding the magnified motion layers. The approach has several limitations. It requires precise motion estimation, which can be difficult and may lead to errors. It is computationally expensive, making it unsuitable for real-time use. Artifacts may appear, particularly with complex motion patterns or occlusion boundaries. The technique also depends on additional processes like motion segmentation and image inpainting, which increase the overall complexity [55].

- **Linear Eulerian video magnification**: The Linear Eulerian video magnification approach is an efficient method for amplifying subtle motions in videos. It involves decomposing the input video into spatial frequency bands using an image pyramid, applying a temporal filter to remove frame noise, and then magnifying the filtered spatial bands. These bands are combined with the original signal, and the spatial bands are collapsed to produce the magnified output video. This approach has been used in applications like visualizing blood flow by enhancing specific motion frequency bands in videos [55].

  Concept : For a 1D signal in translational motion, let $I(x, t)$ be the image intensity

at position x and time t, and $\delta(t)$ be the displacement function. The intensity due to this motion is expressed as $I(x,t) = f(x + \delta(t))$ with $I(x,0) = f(x)$.

$$I(x,t) = f(x + \delta(t)) \tag{2.1}$$

where

$$I(x,0) = f(x) \tag{2.2}$$

To amplify the signal by a factor of α, the target output signal is:

$$\hat{I}(x,t) = f(x + (1+\alpha)\delta(t)) \tag{2.3}$$

Assuming a small displacement $\delta(t)$, a first-order Taylor series expansion yields:

$$I(x,t) \approx f(x) + \delta(t)\frac{\partial f(x)}{\partial x} \tag{2.4}$$

Let $B(x,t)$ be the result of applying a broadband temporal bandpass filter to $I(x,t)$, excluding $f(x)$. Assuming $\delta(t)$ is within the filter's passband, we have:

$$B(x,t) = \delta(t)\frac{\partial f(x)}{\partial x} \tag{2.5}$$

To obtain the magnified signal, amplify $B(x,t)$ by α and add it back to $I(x,t)$:

$$\tilde{I}(x,t) = f(x) + (1+\alpha)\delta(t)\frac{\partial f(x)}{\partial x} \tag{2.6}$$

Using the first-order Taylor expansion for $(1 + \alpha)\delta(t)$:

$$\tilde{I}(x, t) \approx f(x + (1 + \alpha)\delta(t)) \tag{2.7}$$

This final equation represents the magnification of the spatial displacement $\delta(t)$ by a factor of $(1+\alpha)$. Linear motion magnification methods using first-order Taylor expansion struggle with high magnification or large motions, leading to noise amplification. Eulerian video magnification with spatiotemporal filtering [59] is an improved version of EVM, enabling greater amplification and reduced noise impact by mapping and warping pixel-level motions without altering pixel values.

- **Phase-based magnification**: A Phase-based Eulerian approach introduced in [61] to improve motion representation and noise resistance over linear methods. This technique involves decomposing the video into multi-scales with complex steerable pyramids, applying temporal band-pass filtering, and magnifying the phase. However, it is computationally intensive and slower. The input video is decomposed into multi-scales using complex steerable pyramids and into complex wavelets at various positions, scales, and orientations. A temporal band-pass filter is applied to each sub-band image, magnifying the phase while keeping the magnitude unchanged. The final video is then reconstructed

  In the given process, a 1D image intensity undergoing translation over time, $f(x + \delta(t))$, can be modified to represent amplified motion as $f(x + (1 + \alpha)\delta(t))$. The

function $f(x)$ is decomposed into a Fourier series:

$$f(x) = \sum_{w=-\infty}^{\infty} A_w e^{iwx} \tag{2.8}$$

When incorporating translation over time, it becomes:

$$f(x + \delta(t)) = \sum_{w=-\infty}^{\infty} A_w e^{iw(x+\delta(t))} \tag{2.9}$$

A temporal bandpass filter isolates the phase component $w(x + \delta(t))$, and removing the DC component yields:

$$B_w(x, t) = w\delta(t) \tag{2.10}$$

To enhance motion, the bandpassed phase $B_w(x, t)$ is amplified by α, resulting in the amplified sub-band:

$$\hat{S}_w(x, t) := S_w(x, t)e^{i\alpha B_w(x,t)} = A_w e^{iw(x+(1+\alpha)\delta(t))} \tag{2.11}$$

Finally, summing all sub-bands collapses the pyramid, resulting in the amplified sequence $f(x + (1 + \alpha)\delta(t))$. The complex steerable pyramids are able to capture local motions in a video by calculating the local phase across time for each spatial scale and orientation. These phases are temporally band-passed to filter relevant frequency bands and remove the DC component. The amplified motion is obtained by multiplying the band passed phases by a magnification factor [55]. Phase-based

methods for motion magnification are sensitive to noise, complex to implement and require fine-tuning. They also show a weak performance in the presence of large motions and complex deformations. Furthermore, their high computational complexity makes them time-consuming and unsuitable for real-time applications, limiting their effectiveness in diverse scenarios with noisy or poor-quality videos [55].

- **Learning-based motion magnification**: A novel approach using neural networks that learns decomposition filters directly has been introduced in [62]. Their model, trained on a synthetic dataset with two-frame inputs, amplifies the motion (velocity) between frames without relying on temporal filters to separate motions. Their proposed network consists of three main components: an encoder, a manipulator, and a decoder. The encoder extracts shape and texture representations from video frames, similar to phase extraction in traditional pyramids. The manipulator amplifies motion by multiplying the difference between shape representations from two frames with a magnification factor, while the decoder reconstructs the final magnified output. The model uses a fully convolutional architecture with residual blocks to enhance quality. During training, *L1-loss* is employed to ensure a clear separation between shape and texture, and a regularization term controls the influence of intensity changes. The Adam optimizer is used for loss minimization. Additionally, perturbations are applied to differentiate texture and shape representations, and the magnification process focuses solely on motion amplification without distorting intensity values [55]. This motion magnification approach using neural networks has several limitations. Temporal filters, while reducing artifacts, complicate the design

and can degrade performance, especially with slight motions and high magnifications. The model requires large, diverse datasets, which are costly to create and struggle with colour magnification, needing separate models and data. Training is computationally intensive, with long runtimes and overfitting and generalization issues often arise. Additionally, errors from early frames can propagate, distorting the output, and discrepancies between the paper and code add further uncertainty.

2.4.4 Optical Flow:

Motion analysis is a fundamental challenge in computer vision with applications in areas like autonomous driving and robotics. It involves estimating optical flow (2D-pixel motion) and scene flow (3D motion). Traditional methods use variational techniques by utilizing the energy minimization process, while deep learning has recently made significant advancements by learning feature representations directly from data [63].

The evolution of optical flow algorithms began with early methods in the 1980s. The Horn-Schunck method in [64] introduced a global energy minimization approach for dense flow fields, while the Lucas-Kanade method in [65] provided a localized gradient-based technique for small motions. Throughout the 1990s and 2000s, variational and energy-based methods gained prominence, incorporating regularization techniques and multi-scale (pyramidal) implementations to handle larger displacements. By the early 2000s, robust estimators and feature-based approaches like Black and Anandan's method in [66] improved the handling of real-world complexities. The advent of SIFT Flow [67] used dense

descriptors for non-rigid motion estimation. The 2010s saw the rise of deep learning-based approaches, starting with FlowNet [68], which trained CNNs on synthetic data to achieve significant efficiency and accuracy gains. PWC-Net [69] introduced pyramid, warping, and cost volume structures, enhancing accuracy for small motions, followed by RAFT [70], which achieved state-of-the-art results through iterative refinement. Recent advances focus on real-time applications and unsupervised learning methods, driven by the need for efficient, accurate optical flow estimation in real-world tasks like autonomous driving. In the following, a review of dense optical flow estimation is presented, covering three categories: knowledge-driven methods, data-driven methods, and hybrid-driven methods.

- **Knowledge-driven**: The initial framework for optical flow estimation was introduced by Horn and Schunck (HS) [64], utilizing a variational approach to estimate optical flow. Building on this, Lucas and Kanade (LK) [65] presented a local constraint method to estimate a sparse flow field. Numerous enhancements and modifications have since been developed based on the foundational HS and LK methods.

- **Data-driven**: Recently, convolutional neural networks (CNNs), a leading technique in deep learning, have been effectively applied to address the optical flow problem. Data-driven methods leverage labelled data for training, and the existing architectures for these approaches fall into two main categories: U-Net and spatial pyramid networks. Data-driven methods often require ground truth as a supervised signal, which depends heavily on labelled data. However, obtaining dense pixel-level ground truth for real scenes is both complex and time-consuming. CNNs have been intro-

duced for optical flow with two models [68]: FlowNetS and FlowNetC. Both use an encoder-decoder architecture (U-Net). FlowNetS processes two stacked images with successive convolution layers, reducing the feature map scale to 1/64 of the original input. FlowNetC, on the other hand, splits the input into two branches, each with three convolution layers to extract features, and then uses a correlation layer to compute the matching cost between the features.

- **Hybrid-driven**: Hybrid-driven methods utilize unlabelled data for training while incorporating prior knowledge assumptions, such as brightness constancy, gradient constancy, and spatial assumptions, to guide the learning process. Additionally, semi-supervised approaches have been explored, utilizing both labelled and unlabeled data to balance the advantages of supervised and unsupervised learning. However, semi-supervised methods still yield lower accuracy than fully supervised ones on public datasets.

2.5 Research Gaps

A review of the existing literature on JVP measurement reveals several critical gaps in current methodologies. While advancements in ultrasound, radar, and camera-based technologies have improved medical diagnostics, the integration of novel, non-invasive approaches for JVP measurement remains underexplored. This study aims to address these limitations by introducing a more practical and accessible framework for JVP assessment. The key

research gaps that this work seeks to fill are as follows:

- **Use of clinic patient data:** Unlike many studies that rely on healthy volunteers, this research is based on a dataset collected from patients referred to a cardiologist for cardiac disorders. This ensures that the study includes a diverse range of participants with varying underlying health conditions, making the findings more clinically relevant.

- **Room lighting conditions:** Many existing studies conduct video acquisition under controlled artificial lighting to enhance image clarity. In contrast, this study uses normal room lighting to better simulate real-world conditions, improving the applicability of the proposed method in non-clinical settings.

- **Use of smartphone cameras:** While previous research often relies on high-end, specialized cameras, this study demonstrates the feasibility of using commonplace smartphones for JVP measurement. This approach enhances accessibility, making remote monitoring more practical for widespread clinical and home use.

- **Multiple postures and camera angles:** Most studies limit JVP measurement to a single posture, typically a semi-recumbent position. This research expands the scope by capturing data in two different positions (45-degree recline and fully sitting upright) and from two distinct camera angles (direct and tangential). This variation provides a more comprehensive understanding of JVP dynamics across different conditions.

- **Integration of object detection models:** The potential of advanced object detection models in JVP assessment has been largely overlooked. This study incorporates state-of-the-art deep learning models to the workflow, enhancing both efficiency and accuracy in estimating and analyzing JVP features.

- **Semi-automated JVP estimation framework:** Existing research primarily focuses on detecting JVP pulsations without proposing a semi-automated framework to locate and analyze JVP height dynamically. This study introduces a novel framework capable of identifying JVP positioning and measuring its variations, paving the way for a more reliable assessment method.

By addressing these gaps, this research aims to advance the state of the art in JVP monitoring, ultimately contributing to the development of a practical, non-invasive system for remote cardiovascular assessment.

3

Methodology

3.1 Introduction

In this chapter, the methodology of the proposed framework for evaluating the JVP height measurement will be presented. Our approach encompasses three key components: data collection, the development of a proposed algorithm, and the theoretical basis underlying

each of the components of the proposed method. Each of the components is designed to contribute to a comprehensive framework for JVP measurements using smartphones.

Data Collection: We begin with a detailed protocol for collecting JVP height measurements from patients who volunteered to take part at a Cardiopulmonary clinic[1]. This procedure involves positioning patients at various angles and using smartphones in various degrees to standardize the measurement process as well as gaining insights into the results. We will utilize advanced imaging technologies ultrasound method and camera recordings to capture RGB video for analysis. This approach allows us to compare and validate our framework's results with the US method to make sure of its accuracy.

Proposed Algorithm: The algorithm aims to reduce manual measurement errors and the costs associated with the US measurements as well as the time and workload of the healthcare workers. Following data collection, we will develop and apply a novel algorithm designed to analyze the JVP measurements. This algorithm leverages image processing and machine learning techniques to detect jugular venous pulsation and quantify its height with enhanced precision.

Theoretical Basis: Our methodology is grounded in the principles of cardiovascular physiology and signal processing. By integrating these fundamental principles, we aim to evaluate the effectiveness of the proposed approach in comparison with physicians' assessments.

[1]A healthcare facility that specializes in diagnosing, treating, and managing conditions related to the heart (cardio) and lungs (pulmonary) located in Orleans, Ottawa, ON, Canada.

3.2 Data Collection

Informed consent has been obtained from patients for image acquisition and processing as part of a clinically indicated echocardiographic study, and the study was approved in accordance with the guidelines and regulations of the Montfort Hospital ethics committee. The participants were seated in a hospital bed at a 45-degree position and 90-degree (sitting position) for the study duration. Once the subject was relaxed and comfortable, the camera recorded the subject's neck for approximately 30 seconds. Recordings were performed using two Android smartphones (Google Pixel 3a), with autofocus disabled, capturing video at 60 frames per second and 1080p resolution from two perspectives: tangential and front-facing (Figure 3.1). Patients were evaluated by a sonographer who acquired images of the neck vein and carotid artery position by placing a vascular ultrasound probe (Vivid E90 2D cardiovascular ultrasound system by GE Healthcare) on the patient's neck identified the location of the respective vessels, and marked a spot on the neck for further investigation. It is notable that participants who suffered from a variety of medical heart conditions, including congestive heart failure, coronary artery disease, shortness of breath, hypertension, constrictive pericarditis, palpitations, and etc. participated in our study. Various factors, including body mass index (BMI), body surface area (BSA), left ventricular size (LV size), right ventricular basal size (RV size), right atrium volume index and left atrium volume index, have been taken into demographic information. Cardiac disorders are also categorized into:

- Prolapse which includes Mitral Valve Prolapse (MVP)
- Tricuspid Valve Prolapse (TVP)
- Regurgitation including Mitral Valve regurgitation (MVR)
- Tricuspid regurgitation (TR)

Also in terms of Rhythm, it could be categorized into Normal Sinus Rhythm, Atrial Fibrillation, Premature Ventricular Contraction (PVC) and Bradycardia. The acquisition of the neck video is simultaneously recorded as the ECG signal, respirometer (acquisition done with the echo machine) and a belt technology from a Nox portable sleep monitor measures the simultaneous heart rate and respiration of the patient. The two cameras were placed on a tripod at about 1.2 m height, pointing toward the right side of the patient's neck at a distance of about 20 cm - 30 cm with two angles of 90-degree difference from each other as shown in Figure 3.1. The participants had no skin makeup and wore suitable clothing for recording the footage. It is essential to point out that the participants' appearances varied considerably; some had facial hair or wore glasses, and their skin tone ranged from pale white to dark brown. Some participants were removed from further processing due to inappropriate video quality or missing reference data.

The whole process, including the experimental setup of the conducted study and the processing steps, is depicted in Figure 3.2. The camera was zoomed on the neck, recording the neck region while the Vivid E90 2D system was monitoring the vital signs. From the region of interest selected on the neck, the algorithm extracts a large number of rPPG and

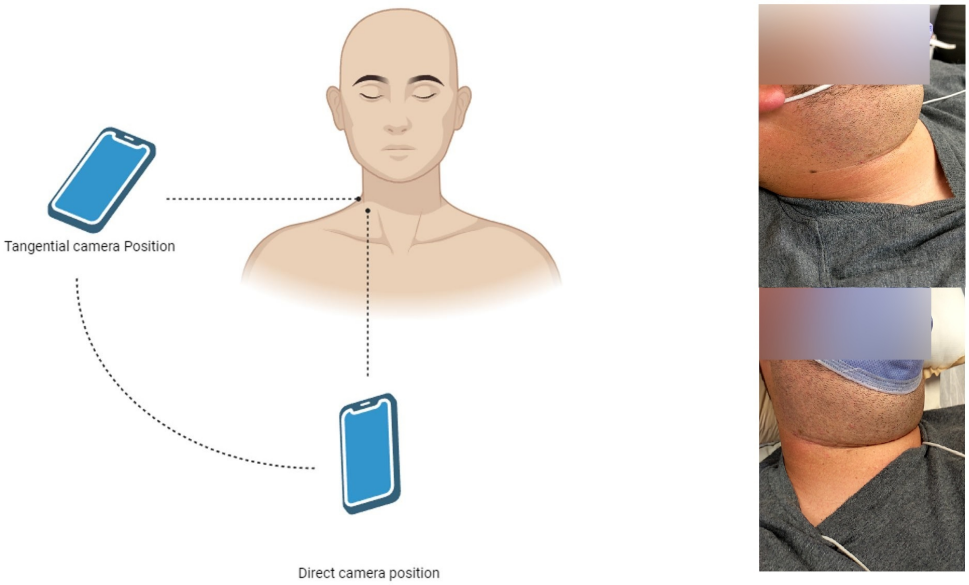

Figure 3.1: Two-view video acquisition.

sMOT time series depending on the region of interest. The data collection process followed a structured approach to ensure consistency and reliability. The detailed steps are outlined below:

1. Patients were asked if they were willing to participate as volunteers for data collection.
2. Participants were provided with a consent form to review, and those agreeing to participate signed the form to indicate their cooperation.
3. Devices were set up to record the patient's ECG, fingertip PPG, and breathing pattern using a respiratory belt.
4. A camera was positioned in the direct view and tangential view to capture video

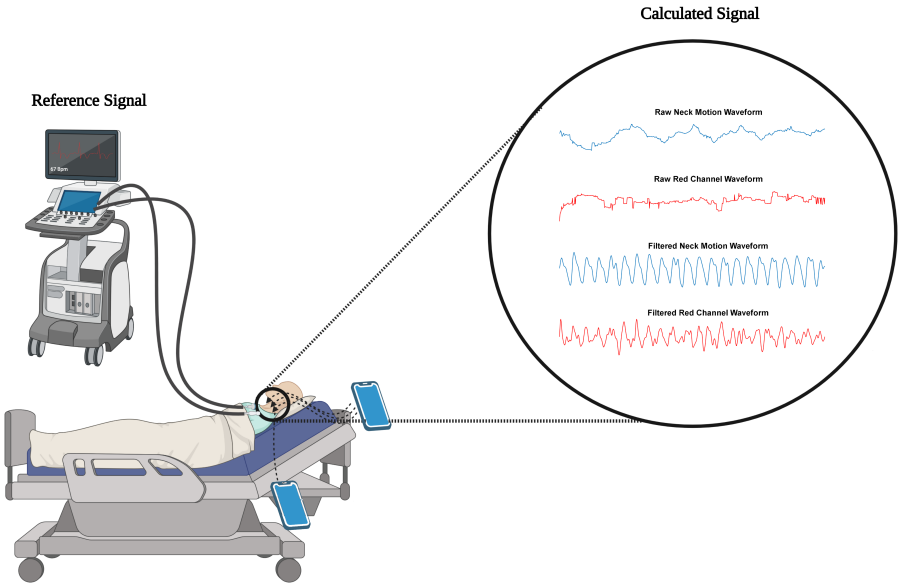

Figure 3.2: The experimental setup for data acquisition.

footage of the patient's neck.

5. The recording process began by first capturing vital signs, followed by video recordings. Three taps were applied to the patient's chest on the ECG pads for later synchronization purposes.

6. The recording lasted for 30 seconds, after which another three taps were applied to the ECG pads to mark the end of the process.

7. The procedure was then repeated with the patient in a sitting position.

8. A sonographer marked key anatomical points on the patient's neck for subsequent analysis and evaluation.

9. Finally, the patient was thanked for their participation and guided out.

3.3 Proposed Framework

The proposed framework for detecting JVP integrates object detection, video magnification, and optical flow analysis. Several preprocessing techniques are employed to ensure video quality before applying these methods to estimate JVP height, as illustrated in Figure 3.3.

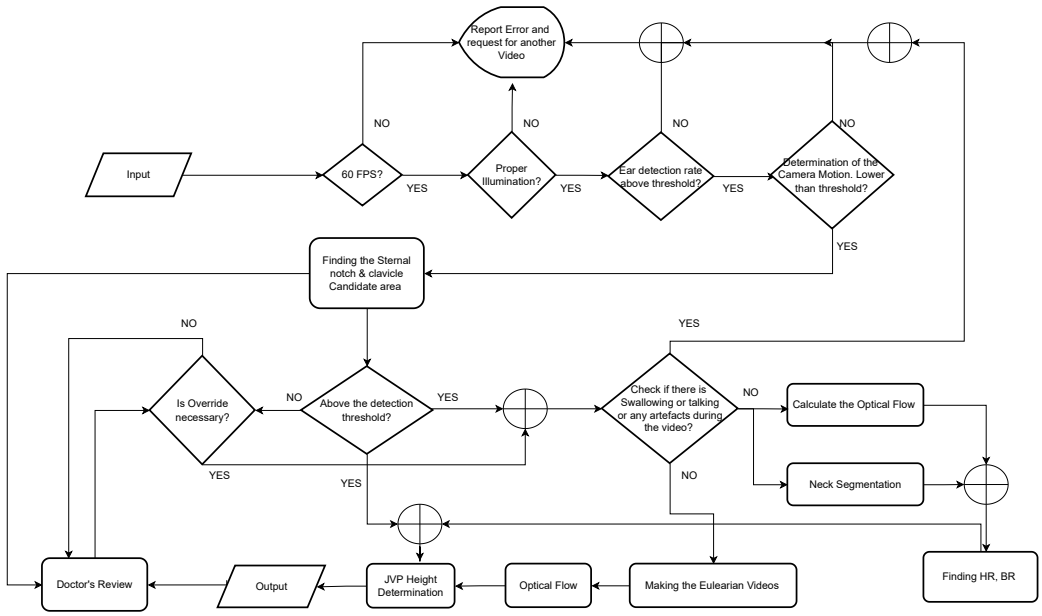

Figure 3.3: Proposed framework for JVP height measurement.

This framework is designed to measure Jugular JVP height, allowing patients to use their own smartphones conveniently from the comfort of their homes. To ensure the videos

are suitable for accurate processing, patients are advised to follow specific guidelines as illustrated in Figure 3.4. These criteria, recommended by physicians and clinicians, aim to enhance the quality and reliability of the captured videos. Before acquiring the video, the patient should make sure that He/She follows the instructions, including:

- Find a well-lit area with a chair that has arm support.
- While sitting upright, use a cushion to support your neck and keep it relaxed.
- Make sure your neck and collarbone are clearly visible by tucking away long hair and removing any jewelry or necklaces.
- Hold your smartphone in your right hand and rest your arm on the armrest while aiming the camera at your neck.
- Keep the camera 20-30 cm from your neck and ensure the image is not blurry for a clear, high-quality video.
- Gently tilt your head to the left to expose the veins. Please avoid talking, chewing, or swallowing during the recording.
- Record the video for 30 seconds and avoid any movement during this time.

Once the video is properly acquired by the patient, the quality of the recording will be assessed, with the primary criterion being a frame rate of 60 Frames Per Second (FPS) with a minimum resolution of 1080p. This enhanced signal fidelity is crucial for accurate

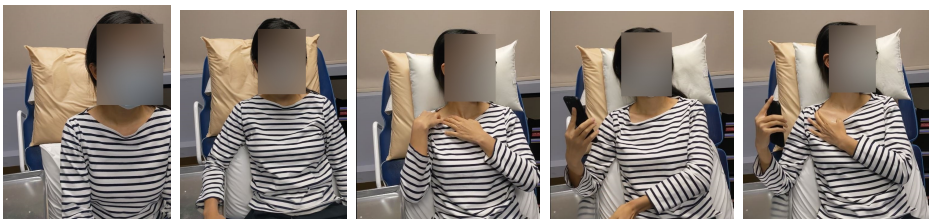

Figure 3.4: Patient's guidelines from left to right.

processing, capturing subtle details that are essential for the system to determine the JVP height effectively. This is in agreement with the findings in [71], which explain how having higher FPS and resolution would be beneficial for processing. Among the various factors influencing the signal's Signal to Noise Ratio (SNR) are bit depth, compression, and brightness. While the first two are beyond the scope of this study, the proposed framework is designed to assess ambient illumination and notify the patient to adjust lighting conditions if they are not suitable for optimal video acquisition. It should be mentioned that compression has a detrimental effect on the yielded signal causing a loss of details, temporal artifacts, and increased noise as studied in [72]. As for bit depth, higher bit depths (e.g., 10, 12, or 14 bits) capture more precise variations in light intensity, which can enhance the detection of subtle changes in skin colour due to blood flow. This improved precision can lead to better rPPG signal extraction, particularly under challenging lighting conditions. Illumination plays a crucial role in the performance of rPPG, as the technique relies on detecting subtle changes in skin colour caused by blood flow. The illumination processing block determines whether there is a sudden change in the brightness of the video by obtaining the average value of the brightness in the frame comparing it with the previous frame, and determining whether the difference between them exceeds the threshold; in

addition to that, it also determines whether the average value of the brightness is below or exceeds the preset threshold, which prevents the video itself from being low or high even if there is no sudden change in the brightness. The advantage of this approach is that it is simple and fast with performance averaging 20 FPS.

3.3.1 Operational Perspective

The primary objective of this study is to develop a comprehensive framework that assists both patients and healthcare professionals in the expedited measurement of JVP height. To achieve this, the proposed algorithm has been modified to ensure its compatibility and effectiveness for this specific application. Figure 3.5 outlines the operational workflow of the algorithm, starting with video acquisition conducted through an application designed for recording patient videos. These videos are subsequently transferred to cloud storage via Sync.com, where a copy is processed using the proposed framework. The algorithm outputs an enhanced version of the video along with the analysis results, which are then forwarded to the cardiopulmonary clinic for detailed review. Upon completion of the review, the patient is notified of the findings.

3.4 Theoretics

In this section, a detailed explanation from a signal processing perspective will be presented. Estimating JVP height has four major parts, which are shown in the following diagram;

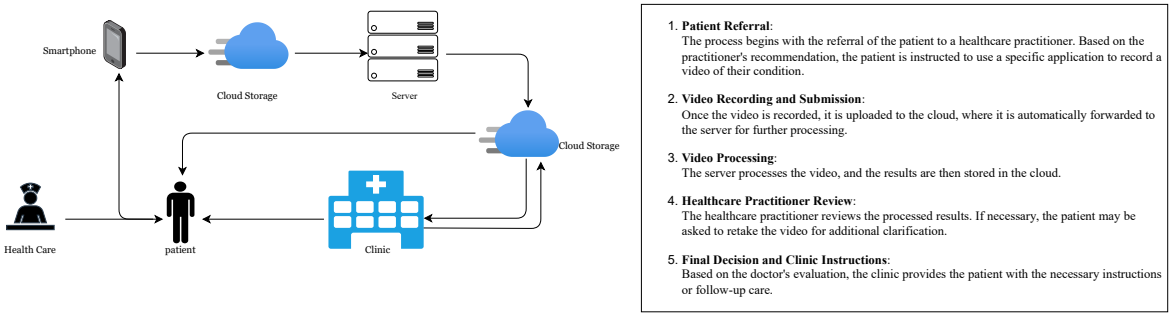

Figure 3.5: Operational workflow of the framework.

1. **Finding HR:** This step is particularly important since we are going to find out the fundamental frequency of the cardiac activities and use it to find the JVP in the next steps. The process begins with the automated segmentation of the neck region from the surrounding areas. For this purpose, a segmentation network [73], pre-trained on the COCO dataset[74], was utilized. This network is designed for both instance and semantic segmentation, leveraging point-based segmentation predictions at selected locations through an iterative subdivision algorithm [73]. To enhance its applicability to our specific context, the pre-trained network was fine-tuned using a dataset of 500 annotated neck images from patients. This fine-tuning process significantly improved the network's robustness and reliability for accurate neck segmentation. The segmented ROI typically includes anatomical features such as the sternocleidomastoid and trapezius muscles, accommodating variations in patient postures.

   Following segmentation, the proposed approach employs dense optical flow, as developed by [75], to extract subtle skin motion displacement signals from the ROI. Dense optical flow is particularly effective for capturing fine-grained motion, making

it an ideal technique for analyzing skin displacement in this context. The complete workflow of the proposed approach is illustrated in Figure 3.6, highlighting the segmentation process, ROI selection, and subsequent extraction of skin motion signals. For the sMOT analysis, the mean value of pixel displacements within a 5×5 block in

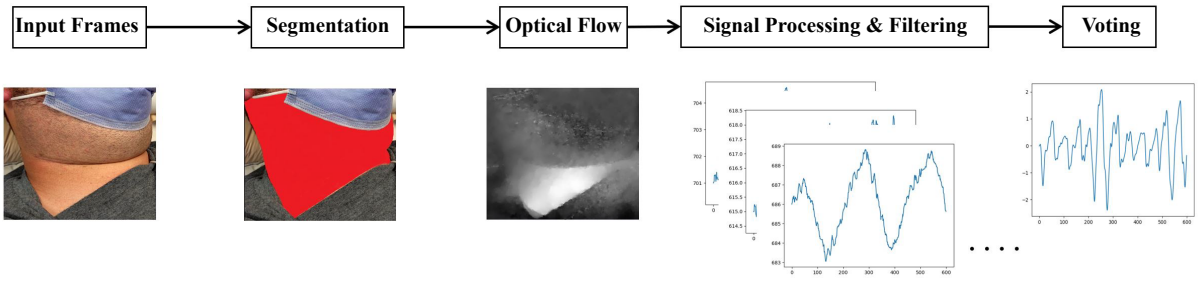

Figure 3.6: Workflow of the HR and RR estimation.

the selected ROI is calculated using the dense optical flow algorithm. This approach aggregates movement data across small neighbourhoods, effectively reducing noise and enhancing the reliability of the resulting sMOT time-series signals. Initially, the ROI contains n individual pixels, which, after processing with the optical flow algorithm, yield $n/25$ sMOT time-series signals, corresponding to the 5×5 pixel blocks. During processing, pixels with unusual trajectories that deviate significantly from expected motion patterns are identified and excluded. Specifically, pixels exhibiting displacement exceeding the maximum distance defined by the most stable pixels are discarded. This is achieved by setting a threshold based on the calculated variance of the output signals and removing points that fall below this threshold. Outlier analysis revealed that fewer than 1% of the total pixels were classified as outliers, ensuring the overall integrity of the data. After initial processing, each pixel

and its corresponding time-series signal are grouped and assigned to their respective block based on spatial positioning. The ROI is divided into non-overlapping square blocks, with each block producing a single combined sMOT waveform. Compared to individual pixel-based signals, block-based analysis offers greater robustness against noise and data distortion, thereby enabling more accurate estimations of HR and RR. Cardiac signals were extracted using cutoff frequencies of 0.7–4.5 Hz, while respiratory signals were filtered using a frequency range of 0.17–0.5 Hz of a third-order Butterworth bandpass filter. These filters effectively removed unwanted noise and retained the frequency components of interest, ensuring the reliability of HR and RR estimations.

The power spectral densities (PSDs) of the processed signals in each block are estimated to identify frequency components corresponding to physiological rhythms. For each block, the frequency of the maximum power (first harmonic) and the second-highest frequency component (second harmonic) are identified. The frequency with the maximum combined power is stored as the heart rate candidate for that block. This step ensures that the HR estimation incorporates both fundamental and harmonic components, improving the accuracy of the frequency selection. To determine a final HR value, a histogram-based approach is applied to the HR candidates derived from all blocks. Each HR candidate contributes to a histogram bin corresponding to its frequency. The frequency associated with the bin containing the maximum number of occurrences is selected as the final HR. This approach offers significant advantages, as it inherently mitigates the impact of outliers by relying on the most

frequent HR candidates across all blocks. By aggregating frequency components in this manner, the method ensures a robust and reliable HR estimation, even in cases where individual blocks may yield noisy or inconsistent signals.

2. **Extracting Respiration Waveform:** This step plays a crucial role in ensuring accurate JVP height measurement. While the RR is estimated in the previous step, the full respiration waveform is necessary for further processing and signal refinement. The clavicle, being a stable anatomical landmark that exhibits clear respiratory-induced motion, serves as the target location on the neck for acquiring the respiration waveform.

   To facilitate this, object detection is performed using YOLO, a state-of-the-art deep learning model for real-time object detection. The model has been trained on an extensive dataset comprising over 5000 images of the ear and 800 images of the sternal notch and clavicle, ensuring robust detection under various lighting conditions, camera angles, and skin tones. The detection threshold is set at 0.75, meaning that only clavicle candidates with confidence scores above this threshold are selected for further processing. This ensures reliability and reduces false detections.

   Once a clavicle candidate is consistently detected, it is used as the primary location for extracting the respiratory signal. Optical flow analysis is then applied to track the subtle vertical motion of the clavicle induced by breathing. To minimize noise and enhance accuracy, the respiratory signal is extracted by averaging the motion values within a 5×5 pixel region centred on the detected clavicle. Following extraction, the

raw respiratory signal undergoes signal filtering to remove artifacts and unwanted noise. A third-order Butterworth bandpass filter with a frequency range of 0.17–0.5 Hz is applied, allowing only relevant breathing frequencies (corresponding to 10–30 breaths per minute) to pass through. This step refines the signal, ensuring that only the true respiratory-induced motion is retained.

To determine the dominant breathing frequency, the Fast Fourier Transform (FFT) is computed on the filtered signals. The frequency component with the highest magnitude is identified as the fundamental breathing frequency. This extracted respiratory signal serves as a critical input for the subsequent processing stages, ensuring precise JVP height estimation in later phases.

3. **Detection of the Line between Ear and Sternal Notch:** During the study, it was determined that the most effective approach for estimating JVP height is to focus on a constrained and well-defined anatomical line connecting the right ear to the sternal notch. This approach minimizes the search space and reduces the likelihood of false detections compared to analyzing the entire segmented neck area from the previous steps. While this method is not entirely free from error and may introduce some inaccuracies, it significantly improves measurement consistency and reliability.

   One of the main reasons for adopting the ear-to-sternal notch line is the inherently diffuse and variable nature of JVP pulsations on the neck surface. These variations in magnitude can introduce inconsistencies, reducing measurement accuracy. Additionally, since the jugular veins exist symmetrically on both sides, a standardized

approach is necessary to ensure repeatability across subjects. Based on expert consultation with a lead cardiologist, it was recommended to focus solely on the right jugular vein, as it provides a more direct pathway to the right atrium, which is essential for accurate JVP assessment. Despite the benefits of this constrained approach, several challenges were encountered. Anatomical variability posed a significant issue, as differences in neck structure, including variations in fat distribution and skin tone, affected the detectability of key landmarks such as the sternal notch. To address these issues, YOLO-based object detection was employed to identify both the ear and sternal notch. However, unlike the ear, which has a rigid and well-defined structure, the sternal notch proved more challenging to detect due to its shape variability and occlusion in individuals with higher body fat (chubby necks). In some cases, the sternal notch was not detected at all due to poor visibility or occlusion. To compensate for instances where the sternal notch could not be reliably detected, an alternative detection strategy was implemented. When the sternal notch was undetectable, the clavicle served as an alternative reference point, and in such cases, the measurement line was extended from the ear to the clavicle instead of the ear-to-sternal notch. While both the ear-to-sternal notch and ear-to-clavicle paths could theoretically be used for JVP height estimation, they differ in clinical significance. The ear-to-sternal notch path measures the height of the IJV, which is the preferred approach for estimating right atrial pressure. In contrast, the ear-to-clavicle path traces the EJV, which is more superficial and less reliable for JVP estimation due to its twists and variable drainage path into the subclavian vein before reaching the

superior vena cava. It should be noted that a feedback loop is provided for doctors to verify detection results, helping to reduce false positives and false negatives.

Although both veins could be used interchangeably in some cases, the IJV is preferred due to its more direct connection to the right atrium, providing a more accurate estimate of central venous pressure. Therefore, unless anatomical constraints necessitate the use of the clavicle, the IJV-based measurement, following the ear-to-sternal notch path, remains the primary focus of this study. This approach balances practicality and accuracy, ensuring robust JVP height estimation across a diverse range of subjects.

4. **JVP height measurement:** In the final stage, the results obtained from the previous steps are integrated to estimate the JVP height. To ensure accurate differentiation between the jugular vein and the carotid artery, two key clinical principles were incorporated into the framework. First, the carotid artery is characterized by a single brisk upstroke per cardiac cycle (i.e., monophasic waveform), whereas the jugular vein exhibits a biphasic waveform, presenting two troughs and two peaks per cardiac cycle. Second, the carotid pulse remains unchanged with respiration, while jugular venous pulsations typically become more prominent and descend lower in the neck during inspiration [12], [14]. These clinical features were carefully considered to guide the subsequent stages of the processing pipeline and to minimize the risk of misidentification between the carotid artery and the jugular vein.

   The proposed framework integrates HR estimation, respiration waveform extraction,

and precise anatomical measurements to estimate JVP height on the neck, as illustrated in Figure 3.7. The first harmonic is identified based on the HR, which corresponds to the fundamental frequency f_h, and the second and third harmonics are denoted as $2f_h$ and $3f_h$, which will contribute to the framework in determining the JVP signal. Let the breathing signal matrix be $B \in \mathbb{R}^{n \times m}$, where n is the number of pixels in the ROI and m is the number of features (i.e, number of frames), and each row $B_y \in \mathbb{R}^{1 \times m}$ corresponds to the breathing signal at vertical position y. If the clavicle reference signal be $B_{\text{clavicle}} \in \mathbb{R}^{1 \times m}$. The correlation between each row of B and the clavicle signal is given by:

$$C_y = \text{Corr}(B_y, B_{\text{clavicle}}), \quad \forall y \in \{1, 2, ..., n\} \tag{3.1}$$

This can be expanded using the Pearson correlation coefficient as:

$$C_{y_n} = \frac{\sum_{i=1}^{m}(B_{y,i} - \bar{B}_y)(B_{\text{clavicle},i} - \bar{B}_{\text{clavicle}})}{\sqrt{\sum_{i=1}^{m}(B_{y,i} - \bar{B}_y)^2} \cdot \sqrt{\sum_{i=1}^{m}(B_{\text{clavicle},i} - \bar{B}_{\text{clavicle}})^2}} \tag{3.2}$$

where $B_{y,i}$ is the i-th sample of the breathing signal at position y, $\bar{B}_y$ is the mean of the signal at row y, $\bar{B}_{\text{clavicle}}$ is the mean of the clavicle signal, and C_{y_n} is the Pearson correlation coefficient between the signal at pixel n and the clavicle signal where $C \in \mathbb{R}^{n \times 1}$. Another criterion to be considered is obtained by looking at the power of

the cardiac signal:

$$P_y = P(2f_h, y) + P(3f_h, y), \quad \forall y \in \{1, 2, ..., n\} \tag{3.3}$$

where $P \in \mathbb{R}^{n\times 1}$ is the resulting column vector containing summed harmonic powers for all vertical positions. Taking the aforementioned clinical principles into account, we define the following criterion for the assessment of JVP height:

$$H_{\text{JVP}} = \arg\max_y \left(\max\left(\|\mathbf{P}_y\| - \|\mathbf{C}_y\|,\ 0\right)\right), \quad \forall y \in \{1, 2, ..., n\} \tag{3.4}$$

where $\|\mathbf{P}_y\|$ and $\|\mathbf{C}_y\|$ denote the min-max normalized power and correlation values at position y, respectively, both scaled to the range $[0, 1]$. Using Eq. 3.4, a single location corresponding to the maximum difference value is initially identified. However, since the JVP is physiologically distributed over a region rather than being confined to a discrete point, we define the upper and lower bounds of the JVP region by including all neighboring positions y for which the difference value $D = (\max(\|\mathbf{P}_y\| - \|\mathbf{C}_y\|,\ 0))$ exceeds 75% of the maximum, i.e., $D_y > 0.75 \cdot \max(D) \quad \forall y \in \{H, ..., H_{JVP}, ..., L\}$. H and L are the higher and lower limits that fall within the criteria to define the range of JVP. This approach enables the identification of a physiologically realistic range for JVP height rather than a single pixel.

The processing workflow for JVP height estimation is shown in Figure 3.7 where the ear (red rectangle) is connected to the sternal notch (green square) by a black line,

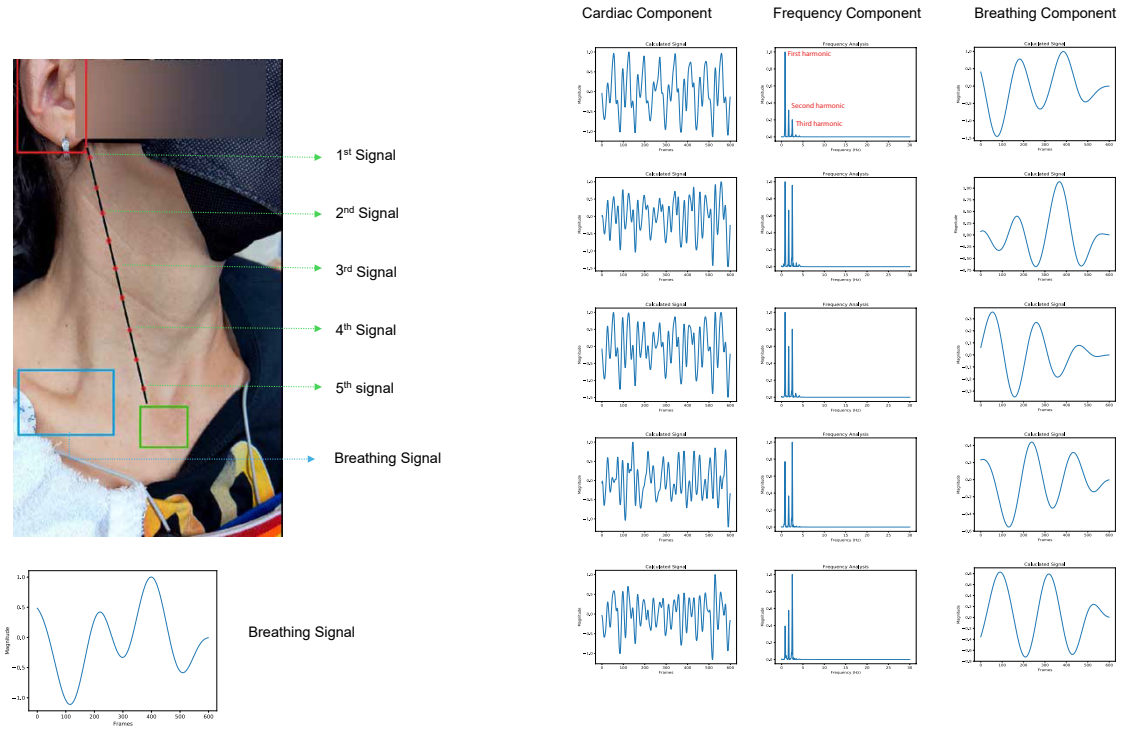

Figure 3.7: Integration of the object detection with signal processing.

marking the ROI. This targeted ROI selection is motivated by the fact that the neck region is susceptible to motion artifacts such as swallowing and contains overlapping signals from both the carotid artery and the jugular vein. By constraining the analysis to this specific region, the approach reduces computational load, improves processing speed, and minimizes noise and outliers that could interfere with accurate JVP estimation. The detected clavicle is highlighted in blue, and the reference respiratory signal is derived from filtering the optical flowing in that ROI. The sample analysis is performed at five points along the ear-to-sternal notch line (left column,

top to bottom). The filtered optical flow signal in the frequency components of 0.7-4.5 Hz for these points is shown in the left column, followed by the PSD illustration of the related signals in the middle column. Also, the respiratory components of the signals by filtering with the bandpass filter in range (0.17–0.5 Hz) is illustrated in the right column. The power summation at the second and third harmonics and the correlation values are normalized between 0 and 1. Figure 3.8 illustrates how the JVP location for the same patient in Figure 3.7 is estimated. H_{jvp} is determined using the Eq.3.4 and D_y, which is the JVP region, determined by looking at the region beginning from higher limit to lower limit. The JVP is treated as distance on the ROI line, and the mean boundary values (i.e., lower limit and higher limit) along the detected line are used for classification.

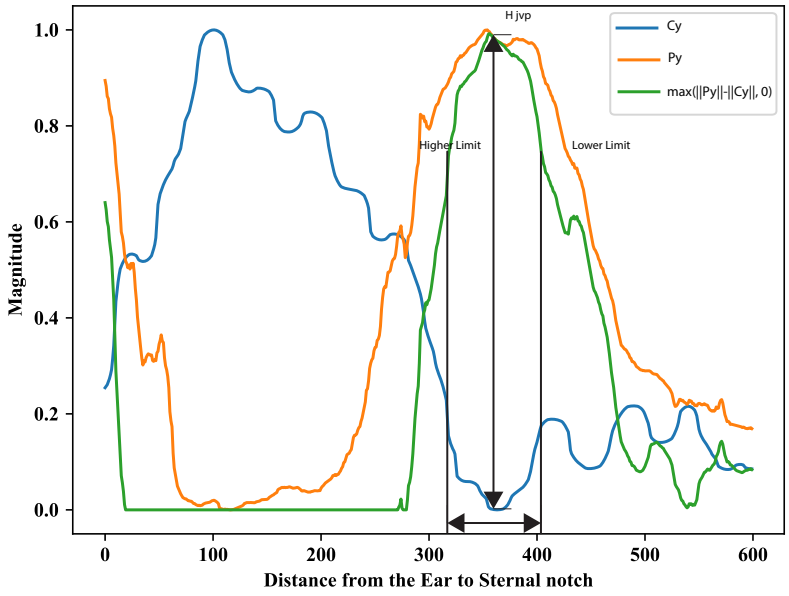

Figure 3.8: Estimation of the JVP on the detected ROI line.

4

Results and Analysis

4.1 Introduction

In this chapter, we present the results obtained from the key components of this study, including object detection, optical flow analysis, and the performance of the integrated framework as a whole for JVP height measurement. YOLO achieved high accuracy in

detecting the ear, with reliable performance for clavicle and sternal detection, though accuracy was slightly reduced due to the complexity of these structures. Video magnification significantly enhanced the visibility of anatomical features, making subtle details more discernible compared to non-magnified videos. This enhancement facilitated more precise detection and analysis. Optical flow analysis further refined our results by effectively tracking dynamic movements, offering a clearer understanding of anatomical changes over time. Together, these methods demonstrate a comprehensive approach to improving object detection and analysis in medical imaging (shown in Figure 4.1). In the following, results

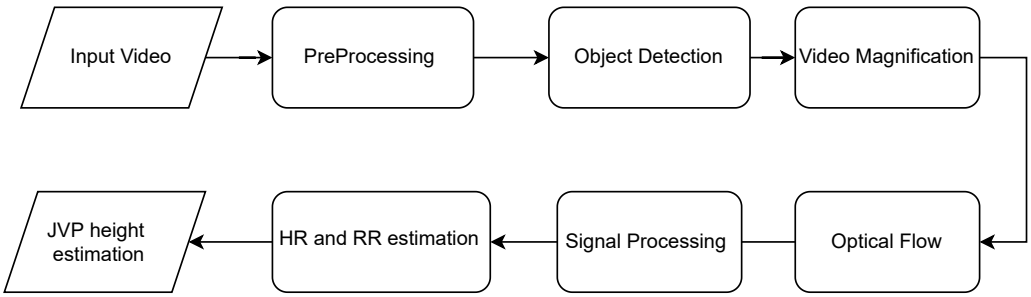

Figure 4.1: High-level illustration of the proposed framework.

of YOLO, video magnification, optical flow, and the signal processing for HR and RR are examined correspondingly. Finally, the results of the Jugular Venous Pressure (JVP) measurement are analyzed and presented.

4.2 Object Detection

For our proposed framework, we have used YOLO which is a state-of-the-art object detection algorithm that processes images in a single pass, offering real-time performance and high accuracy. By dividing images into a grid and predicting bounding boxes and class probabilities, YOLO efficiently detects and classifies objects, making it ideal for dynamic environments.

4.2.1 Ear Detection

As shown in Figure 3.3, camera motion detection relies on the ear detection block, which must produce a value below a predefined threshold set during the experiments, emphasizing the importance of ear detection at this early stage of the process. To train an ear detection network, 5121 images have been labelled and 10% have been used for validation of the pre-trained network of the smallest YOLOv8. The choice of the YOLOv8n is due to its lower amount of parameters while maintaining the accuracy with low inference time. Fine-tuning the network using the pre-trained network over the training network from scratch is a better choice due to having limited labelled data and the fact training the network is much faster. Figure 4.2 illustrates a batch of the training dataset (Left) and the results regarding the validation dataset (Right).

The evaluation of the detector's robustness is conducted through the use of confusion matrices and F1 scores to provide a thorough analysis. True Positives (TP), False Pos-

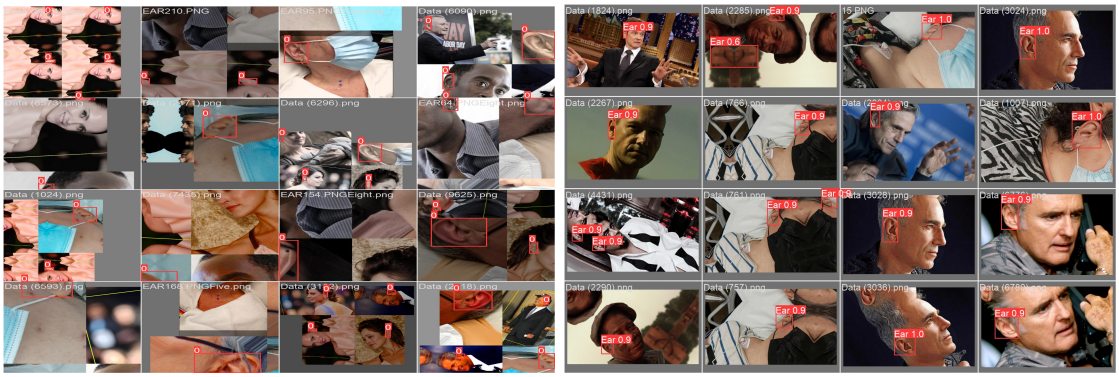

Figure 4.2: Performance comparison of YOLO object detection models for the ear detection.

Table 4.1: Confusion Matrix.

	Predicted Positive	Predicted Negative
Actual Positive	True Positives (TP)	False Negatives (FN)
Actual Negative	False Positives (FP)	True Negatives (TN)

itives (FP), False Negatives (FN), and True Negatives (TN), and Table 4.1) have been established to standardize the criteria. Precision, Recall and F1-score are metrics used to evaluate the performance of classification models. They provide insights into the accuracy and reliability of the model's predictions, particularly in cases where the classes are imbalanced. Precision is the ratio of correctly predicted positive instances to the total predicted positive instances, while Recall (Sensitivity or True Positive Rate) is the ratio of correctly predicted positive instances to the total actual positive instances and The F1-score is the harmonic mean of Precision and Recall, providing a single metric that balances both which is particularly useful when you want a balance between Precision and Recall and are dealing with imbalanced datasets. The mentioned network has been tested against 50 videos from the dataset that were not used for training, and the corresponding results

are presented in Table 4.2. As it is shown, the network's performance is acceptable with a Precision of 0.94, Recall 1 and F1 score of 0.97.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN} \quad F1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.1}$$

Table 4.2: Performance of the Ear Detection Model.

Metric	Value
True Positives (TP)	45
False Positives (FP)	3
True Negatives (FN)	0
False Negatives (TN)	2
Dropped Cases	0
True Positives (TP)	45
Precision	0.94
Recall	1
F1 Score	0.97

In Figure 4.3, the performance of a binary classification model is illustrated using different curves. The F1-Confidence Curve (Top Left) visualizes the trade-off between the F1 score and the confidence threshold used for classifying predictions into positive or negative classes. The precision-confidence curve (Top Right) shows how precision changes with different confidence thresholds in a classification model. It helps evaluate the model's performance by examining how confident the model is in its predictions and the resulting precision at each confidence level. A precision-recall curve (Bottom Left) illustrates the trade-off between precision and recall across different decision thresholds in a binary classification model. High Precision and High Recall indicate a good classifier that captures

most positives with few false positives. Lastly, the recall confidence curve (Bottom Right) visualizes how the model's recall changes as the confidence threshold for classifying positive instances is adjusted. Table 4.2 illustrates the results of using 50 labelled images that the network hasn't been trained on.

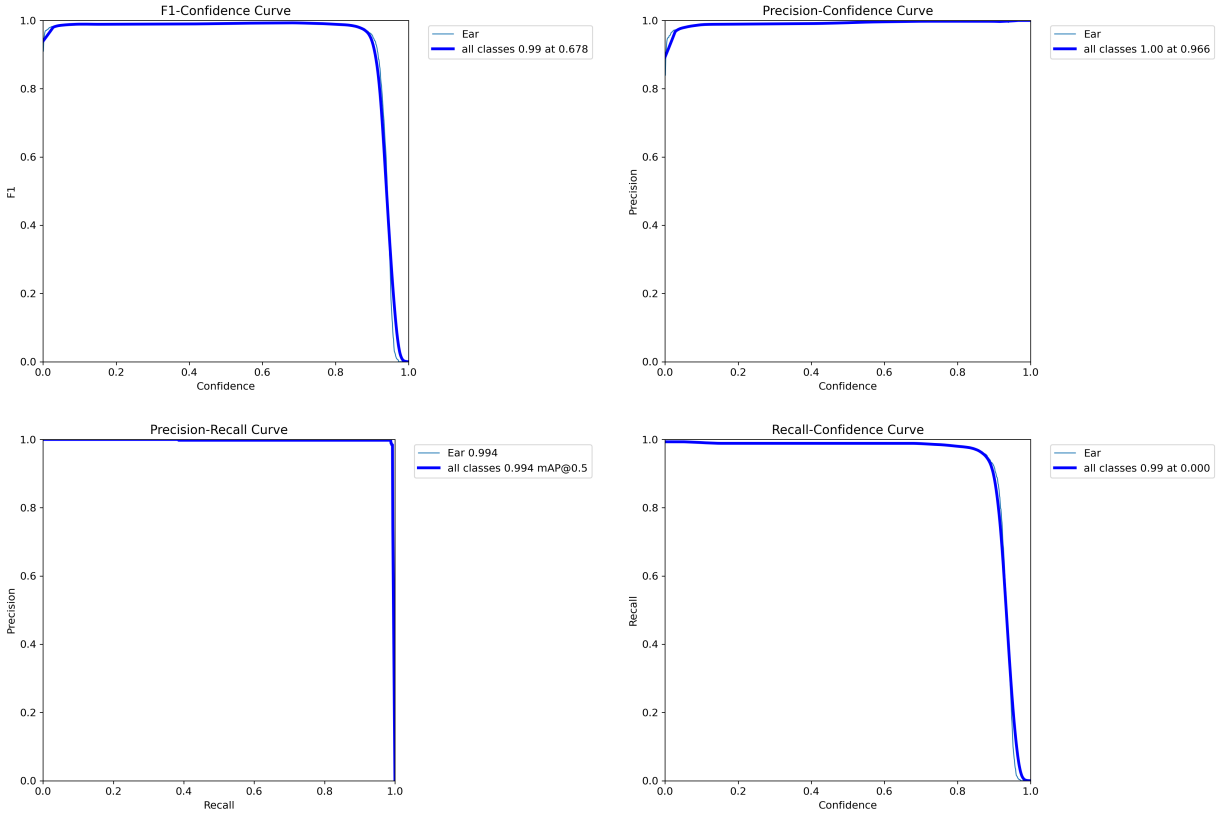

Figure 4.3: Performance comparison of YOLO object detection models for the ear detection.

4.2.2 Sternal Notch and Clavicle Detection

Sternal/Clavicle detection could be considered similar to Ear Detection. However, the primary reason for not combining all three detections into a single network is resource management, specifically minimizing runtime and accommodating the multiple roles that ear detection plays in subsequent processing steps. For this purpose, 800 images have been labelled and annotated for the training process, as shown in Figure 4.4. Labels for this dataset were created under the supervision of a health worker, as identifying the sternal notch and clavicle in some patients can be challenging, especially for individuals with thick necks. The high false positive (FP) values in both clavicle and sternal notch detection negatively affected the overall performance of the framework. In other words, the network incorrectly identifies something other than the clavicle as the clavicle, which could be due to several factors. Firstly, the limited number of training images used for the network has adversely affected its performance. Secondly, opting for the smallest version of the YOLO detection module was prioritized for time efficiency, which also contributes to the higher FP rates. Lastly, the irregular and shapeless nature of the sternal notch poses a critical challenge. Although this point could be seen as the end of the automated workflow, such scenarios have been anticipated. To ensure continuity, a manual override by an expert is proposed (Figure 3.3) when both clavicle and sternal notch detections fall below a defined threshold.

The recall-confidence curve shown in the Figure 4.5 illustrates how recall changes with varying confidence values. As confidence increases, recall decreases, reaching zero at a

Figure 4.4: Performance comparison of YOLO object detection models for the Sternal/Clavicle detection.

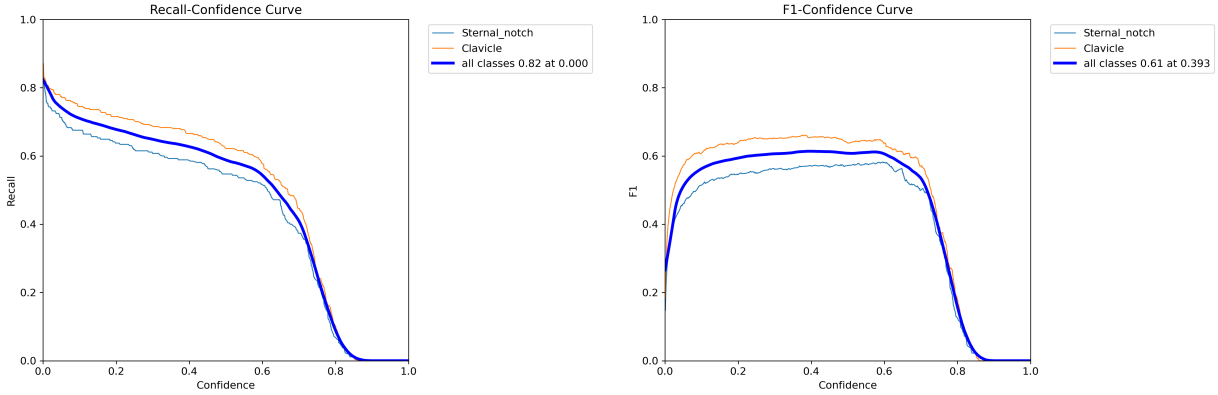

Figure 4.5: Performance comparison of YOLO object detection models for the Sternal/Clavicle detection.

confidence value of one. This indicates that as the classifier becomes more confident in its predictions, it starts missing positive instances, resulting in lower recall. On the right side, the F1 score peaks around a confidence value of 0.5. This suggests that the classifier achieves its optimal balance between precision and recall at this threshold. As confidence increases beyond 0.5, the performance of the binary classification task declines, showing

Table 4.3: Performance of the Sternal/Clavicle Detection Model.

Metric	Sternal Notch	Clavicle
True Positives (TP)	25	27
False Positives (FP)	18	22
True Negatives (FN)	5	0
False Negatives (TN)	1	1
Dropped Cases	1	0
Precision	0.58	0.55
Recall	0.83	1
F1 Score	0.68	0.71

that no object is detected correctly with a confidence value of one. It could be seen that the values of the Table 4.3 and the Figure 4.5 are close enough to show the agreement.

4.3 Optical Flow

Optical flow algorithms are essential in computer vision, facilitating motion estimation between consecutive frames, which is crucial for applications such as object tracking, video stabilization, and action recognition. OpenCV provides a suite of optical flow methods, each designed to address specific challenges in motion estimation, balancing accuracy, noise resilience, and computational efficiency.

Lucas-Kanade Optical Flow [76] is a foundational sparse technique that calculates the motion of feature points based on the image's intensity gradients. It is computationally efficient and suitable for tracking relatively small displacements but is sensitive to noise and can struggle when tracking complex, fast-moving objects due to its sparse approach.

Robust Local Optical Flow (RLOF) [77] is a sparse optical flow algorithm, similar to Lucas-Kanade but designed to be more robust against occlusions and varying lighting conditions. It is particularly useful for handling challenging conditions in real-time applications. Farneback Optical Flow [78] improves upon this by offering a dense estimation method, calculating the flow for all pixels using polynomial expansion. This method captures more comprehensive motion details across the entire frame, making it valuable for applications requiring full motion fields. However, it is computationally more intensive, limiting its use in real-time applications without optimization. PCAFlow [79] leverages principal component analysis to provide dense optical flow with improved robustness in handling complex motions. It delivers more accurate estimations but comes with higher computational demands, making it suitable for detailed offline analysis rather than real-time applications. Brox Optical Flow [80] is known for its high accuracy and robustness, particularly in handling complex motion and capturing fine motion details. This method is computationally intensive but provides high-quality results, making it suitable for applications where precise motion estimation is crucial, such as object tracking, video analysis, and computer vision research. Dual TV-L1 Optical Flow [75] employs energy minimization techniques to achieve a reliable balance between speed and accuracy, allowing real-time capabilities without sacrificing output quality. DeepFlow [81] incorporates deep matching to handle large displacements and complex motion patterns, addressing the limitations of traditional methods in scenes with significant motion. While effective, it is computationally intensive, limiting its practical use in real-time scenarios. Overall, these optical flow methods in OpenCV provide a spectrum of solutions, and the choice of method depends

on the specific application requirements, balancing accuracy, and processing speed. Figure 4.6 presents the average computation time per frame for each optical flow method, calculated using ten videos from the database. The total computation time for each method across all videos was divided by the total number of frames to obtain the average values and lower values indicate faster processing. Based on the values that are shown in Figure 4.6, Pyramid Flow, Farneback, and Dual TV-L1 exhibited faster computation times compared to the other methods. Figure 4.7 shows the relative accuracy of several optical flow algorithms. Ten videos were analyzed, with the region of interest (typically the carotid artery) manually selected in each. Contact PPG was used to estimate the fundamental frequency, and the SNR at this frequency was calculated for each video and each optical flow method. The results are compared in Figure 4.7. The algorithms PCAFlow, RLOF, and Dual TV-L1 outperformed the other tested methods, yielding higher quality cardiac signals for subsequent processing. Based on the performance criteria shown in Figures 4.6 and 4.7, the Dual TV-L1 algorithm [75] demonstrated both high accuracy and efficient computation time and therefore, it was selected for use within the proposed framework.

4.4 Results

4.4.1 sMOT and vital signs

During this study, datasets of different sizes and patients have been used. For the purpose of HR and RR estimation, data were collected from 138 participants (60 females and 78

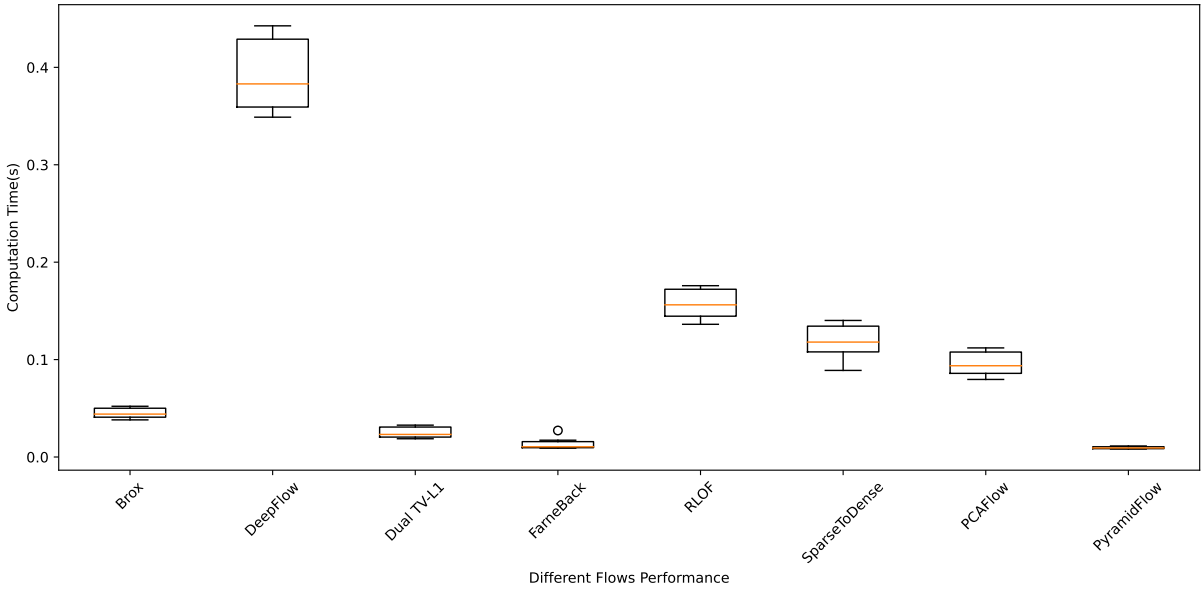

Figure 4.6: Average computation time per frame for different optical flows for ten videos.

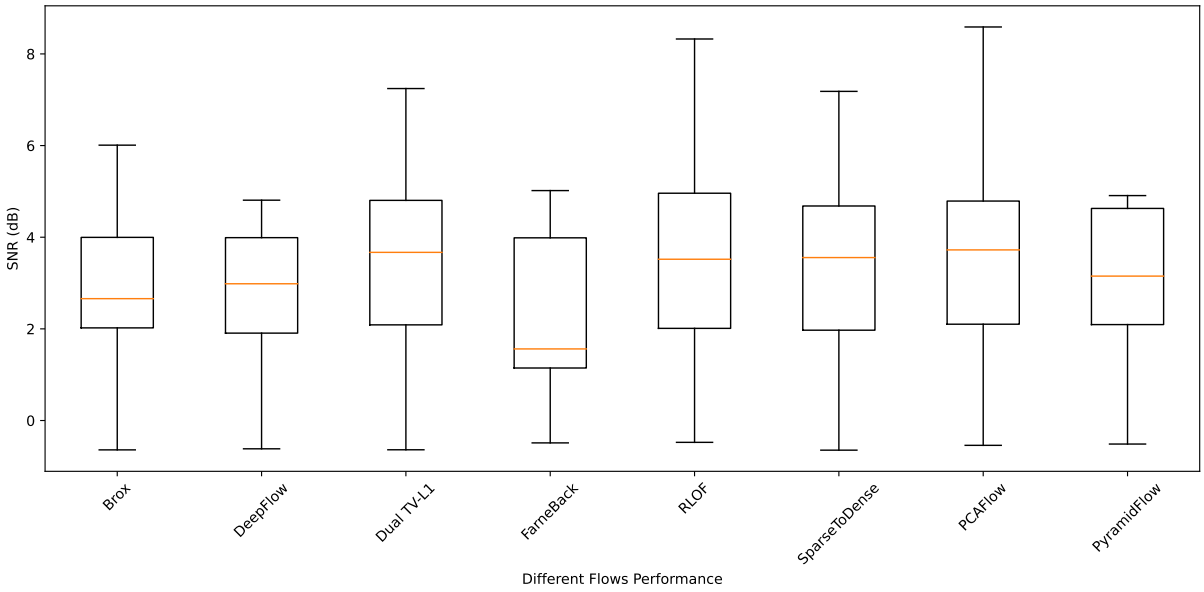

Figure 4.7: SNR(dB) values for different Optical Flows for ten videos.

males) with a mean age of 61 years old (ranging from 17–90 yr) with an average BMI of 28.8 with a physician referral to the cardiac clinic or background cardiac diseases. The following Figure 4.8 illustrates the cardiac disorder distribution as well as their heart rhythm among patients. In the following, Table 4.4 indicates the participants' demographics in detail.

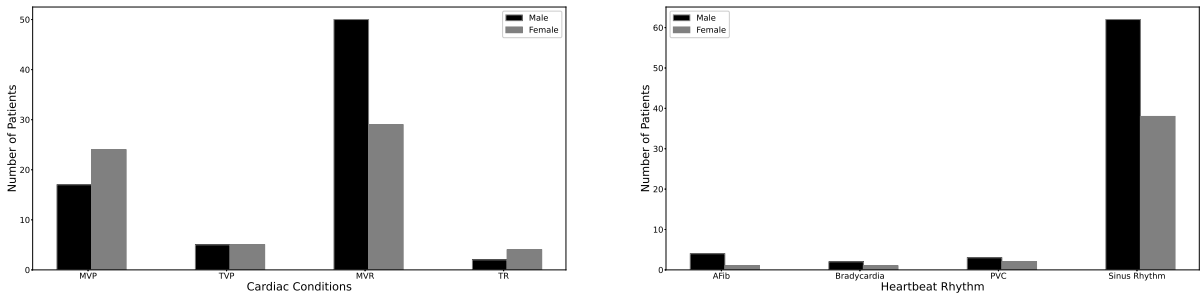

Figure 4.8: Distribution of the cardiac disorders between participants (Left), and Heartbeat Rhythms (Right).

Table 4.4: Demographic of the Study participants.

Participants	60 Female, 78 Male
Age	61 ± 15.21
BMI	28.8 ± 5.94
BSA	1.92 ± 0.22
LV size (mm)	44.1 ± 6.37
RV basal (mm)	38.43 ± 8.55
RA Vol Ind	29.8 ± 16.29
LA Vol Ind	39 ± 15.45

In Figure 4.9, a comparative analysis of sMOT and rPPG using the mentioned approach [24] at the carotid region in both time and frequency domains is presented. rPPG signals are calculated in the same way of sMOT signals as described for finding HR in Section 3.4 with

this exception that average values of pixels are used instead of optical flow results. Since rPPG is only used to show the agreement in findings with sMOT, only the average values of the pixels in the manually chosen area throughout the video frames are taken, and the green channel is used for further processing with the same filtering approach and parameters for sMOT to acquire the rPPG signal at that location. The first and third columns of Figure 4.9 depict the extracted sMOT and rPPG signals in the time domain which are normalized as well. These signals illustrate the temporal dynamics of cardiac activity for each participant. Notably, for some participants, such as the third patient, the sMOT and rPPG signals exhibit a zero-phase shift, indicating strong temporal alignment between the two modalities. The second and fourth columns of Figure 4.9 show the single-sided power spectra of the time-series signals. These spectra reveal the frequency components of the sMOT and rPPG signals. A key observation is that the dominant frequency in both signals aligns closely with the HR, which is approximately 1 Hz. This alignment demonstrates that both sMOT and rPPG methods capture similar frequency components and share the same fundamental frequency, underscoring the consistency and reliability of these two approaches for cardiac signal extraction. This analysis highlights the robustness of sMOT and rPPG techniques in capturing cardiac activity and validates their potential for clinical applications. It should be noted that only sMOT is used for this study and Figure 4.9 is illustrated to demonstrate agreement in both time and frequency domains, these methods provide complementary insights into physiological dynamics, with potential implications for non-invasive monitoring of HR and RR.

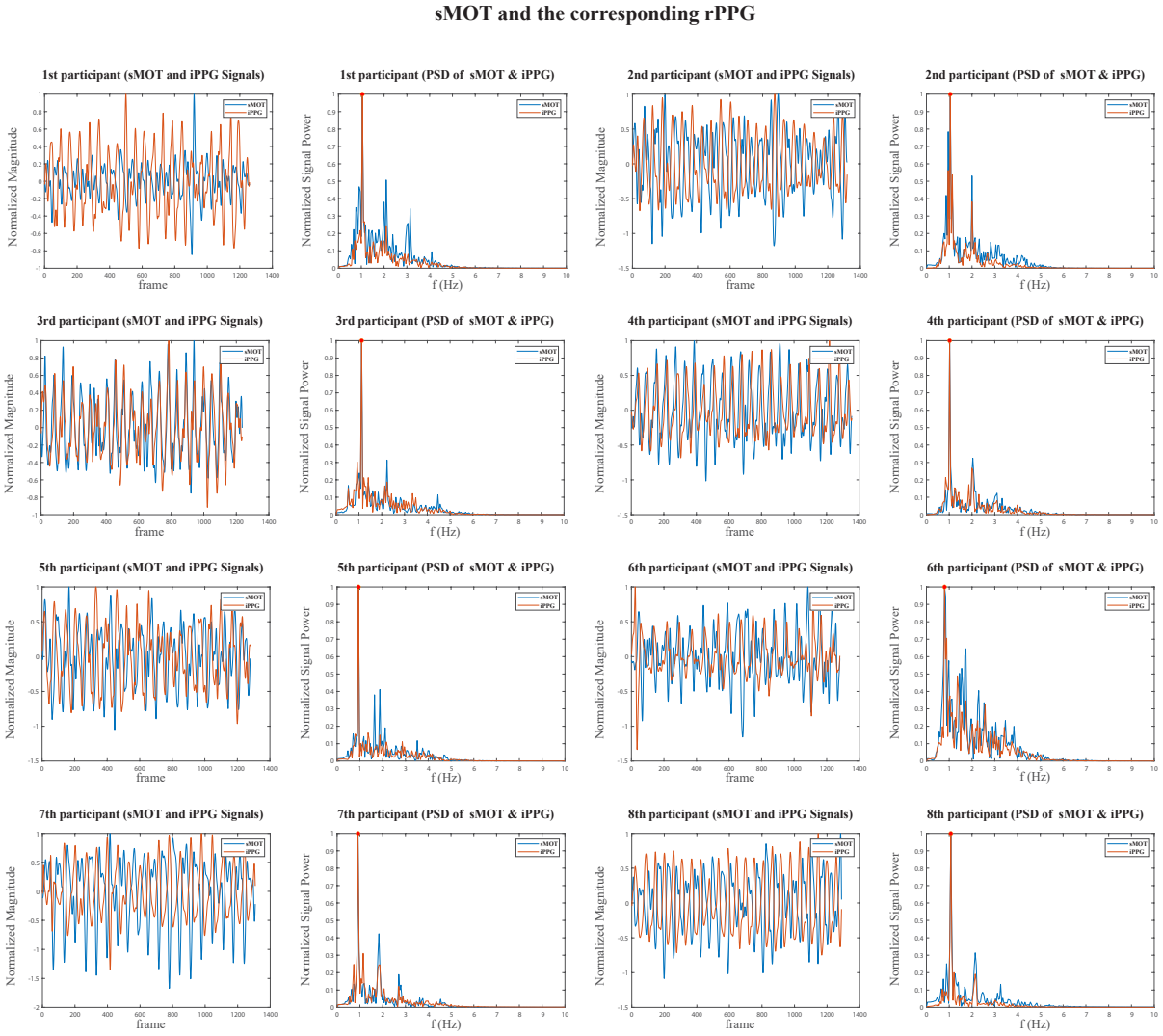

Figure 4.9: Comparison between the sMOT and the corresponding rPPG in time and frequency domain for 8 subjects.

4.4.2 Heart Rate Estimation

The agreement between the estimated HR and the extracted HR from reference ECG is illustrated using Bland-Altman and correlation plots shown in Figure 4.10. For the Direct view with 45-degree and sitting positions, the mean value of the estimated HR is 65.59 and 66.03 bpm, and for the tangential view with 45-degree and sitting positions, the mean value of the estimated HR is 65.19 and 65.97 bpm.

In the Bland-Altman plot, the black dotted lines and the numbers beside them indicate 95% confidence intervals of the mean bias between the reference HR and estimated HR using sMOT or rPPG measurements. As shown in Figure 4.10 1st row (direct view with 45-degree patient position), the correlation coefficient is 0.994 and the corresponding BA plot illustrates near zero bias with mean=-0.82 (bpm), and 95% of the data falls between -3.07 beats per minute and 1.42 beats per minute. The 2nd row illustrates the results for the sitting position with the same camera view with the correlation coefficient 0.992, and the corresponding BA plot illustrates near zero bias with mean=-0.86 (bpm). 3rd and 4th rows are related to the tangential view with patients in 45-degree and sitting positions correspondingly with correlation coefficients 0.994 and 0.991 relatively and bias -1.05 and -0.92 respectively.

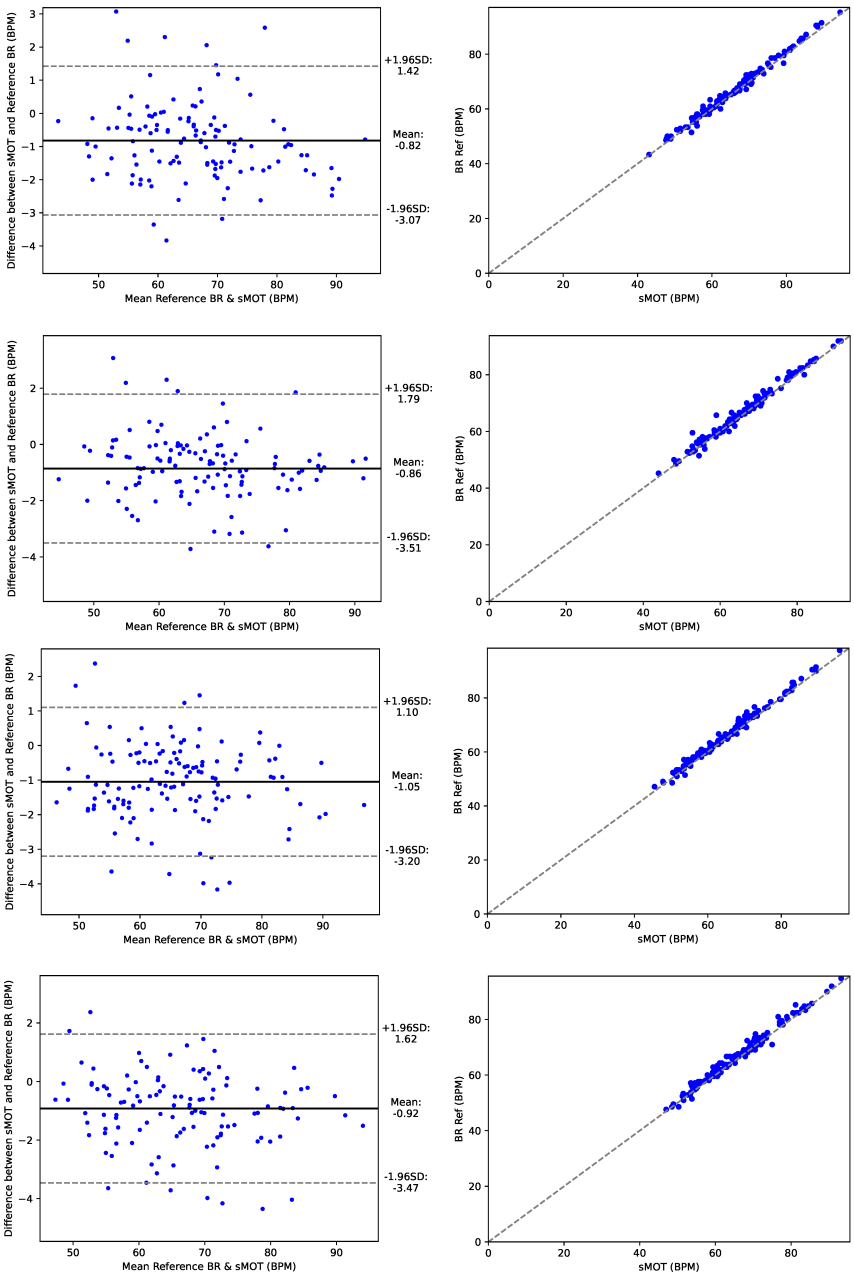

Figure 4.10: Direct view result for sMOT-based HR. 45-degree patient's position (1st row) and sitting (2nd row). Tangential view result for sMOT-based HR. 45-degree patient's position (3rd row) and sitting (4th row).

4.4.3 Respiration Rate Estimation

Our experiments showed that the lower-frequency components, particularly those containing respiration patterns, are more susceptible to motion artifacts, as their frequencies often overlap with those of the motion itself. In addition, the respiration of some subjects is irregular, and the respiration rate is difficult to determine even from the reference respiration belt. Therefore, the agreement between the estimated RR and the reference RR is worse than the agreement between the estimated and reference HR. For the Direct view with 45-degree and sitting positions, the mean value of the estimated HR is 15.53 and 15.84 beats per minute (bpm), and for the tangential view with 45-degree and sitting positions, the mean value of the estimated HR is 15.32 and 15.78 beats per minute (bpm). In the Bland-Altman plot, the black dotted lines and the numbers beside them indicate 95% confidence intervals of the mean bias between the reference HR and estimated HR using sMOT or rPPG measurements. As shown in Figure 4.11 1st row (direct view with 45-degree patient position), the correlation coefficient is 0.847 and the corresponding BA plot illustrates near zero bias with mean=0.3 (bpm), and 95% of the data falls between -3.49 beats per min and 4.8 beats per minute. The 2nd row illustrates the results for the sitting position with the same camera view with the correlation coefficient 0.846, and the corresponding BA plot illustrates near zero bias with mean=0.46 (bpm). 3rd and 4th rows are related to the tangential view with patients in 45-degree and sitting positions correspondingly with correlation coefficients 0.867 and 0.859 relatively and bias 0.2 and 0.48 correspondingly.

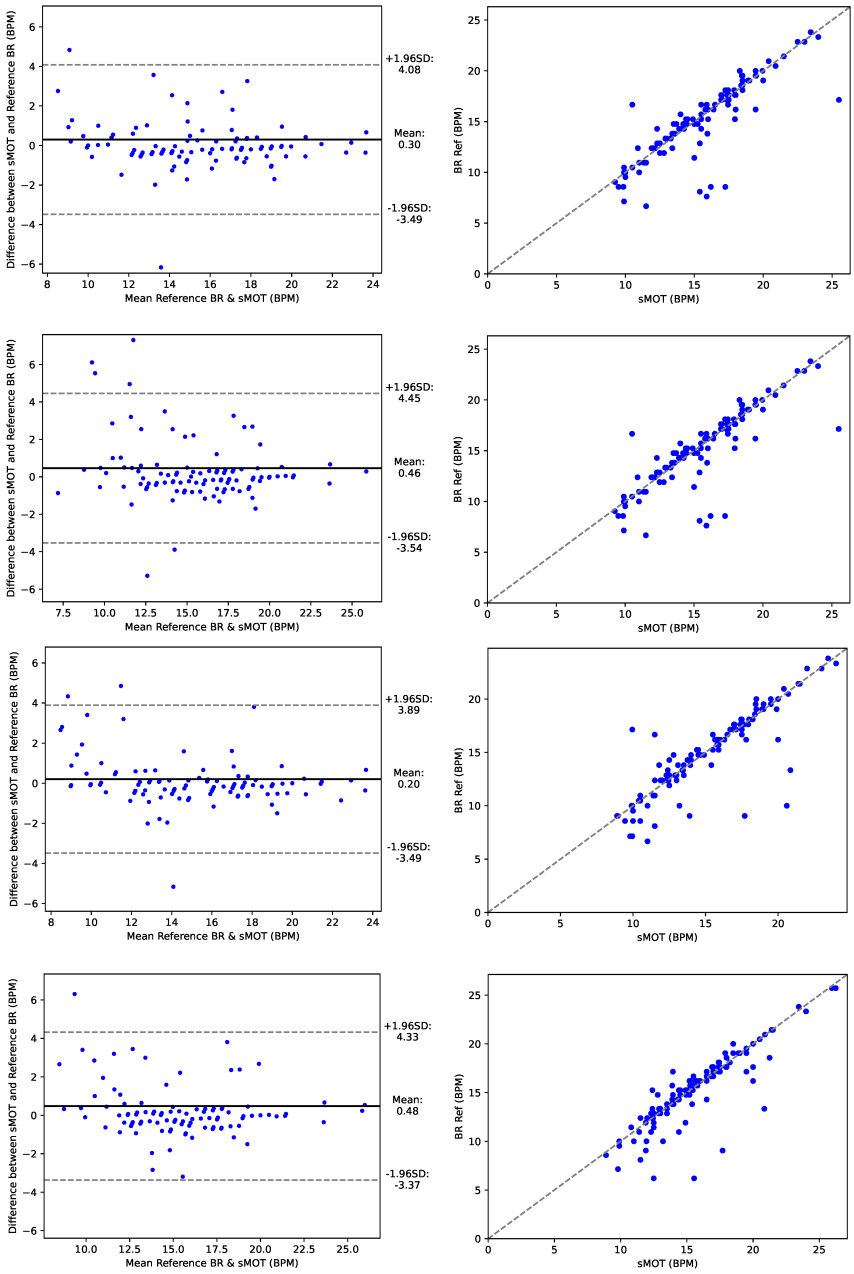

Figure 4.11: Direct view result for sMOT-based RR. 45-degree patient's position (1st row) and sitting (2nd row). Tangential view result for sMOT-based RR. 45-degree patient's position (3rd row) and sitting (4th row).

4.4.4 JVP HEIGHT MEASUREMENT

The final step in the proposed framework is to determine the JVP height on the neck. For this part, 61 participants (25 females and 36 males) with a mean age of 61 years (range: 26–86 years) and an average BMI of 28.75 were chosen for this section. Among these participants, 5 reported smoking more than half a pack daily. In order to exactly locate the pulsation location of the JVP, the ROI is limited to the line that connects the sternal notch to the ear. Limiting the ROI region to a line could be beneficial to reduce the amount of computation and process time as well as remove the unwanted noise and outliers that may interfere with the proposed framework. Based on the expert's opinion, we could use two points alternatively including the sternal notch or clavicle to draw the line to the ear which has been detected in the previous steps in Figure 4.12.

The following Figure 4.13 illustrates the results of the proposed approach for 4 patients in a reclined position (45 degrees) and the cameras in direct and tangential view with the corresponding labels put by the cardiologists and physicians. It should be noted that the JVP location is distended on the skin and it covers a region on the neck and not a single pixel. To ensure a fair comparison of the JVP height values, a line is first drawn from a common reference point to connect the expert's annotations. The measurements are then projected onto a vertical axis and normalized to the range [0, 1]. This process is applied to both the proposed results and the expert labels to digitize the values, enabling a direct and consistent comparison. In terms of height, results can be classified into three groups, **Normal**, **Mild** and **Severe**. This classification has been done for the purpose of

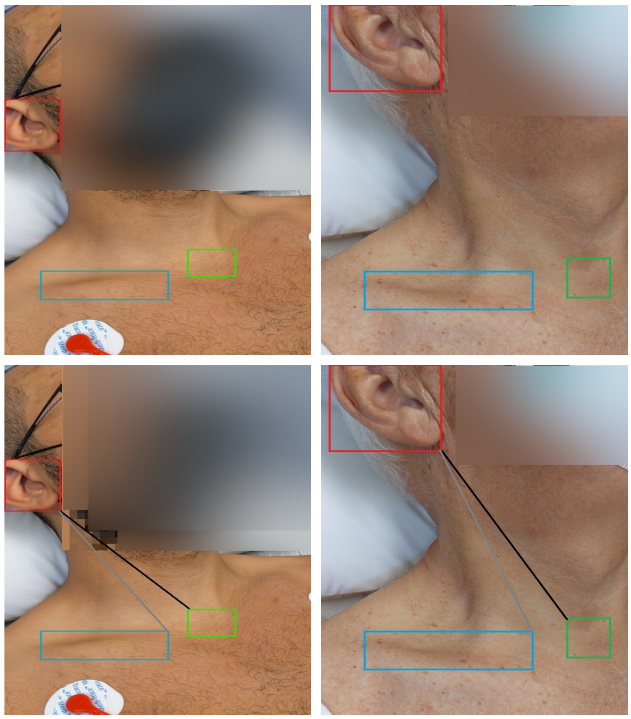

Figure 4.12: Detected Regions of Interest (Ear, Sternal Notch, Clavicle) on the Top Row, and the line connecting them on the Bottom Row for processing.

the diagnosis and to compare the annotated results by the experts with those obtained from the proposed approach.

In Table 4.5, the results for 61 patients are reviewed.

4.5 Analysis

Table 4.5 summarizes the results for each patient along with the corresponding measurements provided by the doctors (shown in the two rightmost columns). Typically, each

Table 4.5: Summary of results across patient positions and camera angles

NAME	45P˚ Direct View			90P˚ Direct View			45P˚ Tangential View			90P˚ Tangential View			Label's Location	Class
Case1	✓	0.4∼0.5	2	✓	0.3∼0.4	1	✓	0.2∼0.4	2	✗	NA	NA	0.3∼0.4	2
Case2	✗	NA	NA	✓	0.25∼0.3	1	✗	NA	NA	✓	0.2∼0.3	1	0.1∼.15	1
Case3	✓	0.1∼0.35	1	✗	NA	NA	✓	0.1∼0.2	1	✓	0.1∼0.2	1	NA	NA
Case4	✓	0∼015	1	✓	0.3∼0.35	1	✓	0∼0.2	1	✗	NA	NA	NA	NA
Case5	✓	0.4∼0.5	2	✗	NA	NA	✗	NA	NA	✗	NA	NA	NA	NA
Case6	✓	0.2∼0.6	2	✗	NA	NA	✗	NA	NA	✓	0.1∼0.25	1	0.1∼0.6	2
Case7	✓	0∼0.1	1	✗	NA	NA	✗	NA	NA	✗	NA	NA	0.1	1
Case8	✓	0.1∼0.3	1	✓	0∼0.1	1	✓	0∼0.3	1	✗	NA	NA	0∼0.3	1
Case9	✓	0.3∼0.4	2	✓	0.2∼0.3	1	✗	NA	NA	✗	NA	NA	0.2	1
Case10	✓	0.1∼0.15	1	✓	0.1∼0.2	1	✗	NA	NA	✗	NA	NA	0.1	1
Case11	✓	0∼0.15	1	✓	0.05∼0.2	1	✗	NA	NA	✗	NA	NA	0.1∼0.2	1
Case12	✓	0.1∼0.2	1	✗	NA	NA	✗	NA	NA	✓	0.1∼0.2	1	0∼0.2	1
Case13	✓	0∼0.2	1	✗	NA	NA	✓	0∼0.2	1	✗	NA	NA	0∼0.3	1
Case14	✓	0.45∼0.6	2	✓	0.4∼0.6	2	✗	NA	NA	✓	0.25	1	0∼0.3	2
Case15	✗	NA	NA	✓	0.1∼0.15	1	✓	0.2∼0.3	1	✓	0.05∼0.15	1	0.1∼.15	1
Case16	✓	0∼0.2	1	✓	0∼0.1	1	✗	NA	NA	✓	0.2∼0.6	1	0.1∼0.2	1
Case17	✓	0.4∼0.5	2	✗	NA	NA	✗	NA	NA	✗	NA	NA	0.1∼0.15	1
Case18	✓	0∼0.1	1	✓	0.2∼0.35	1∼8	✓	0∼0.2	1	✓	0∼0.2	1	0.1∼0.2	1
Case19	✓	0∼0.2	1	✓	0∼0.2	1	✗	NA	NA	✗	NA	NA	0∼0.2	1
Case20	✗	NA	NA	✗	NA	NA	✗	NA	NA	✓	0∼0.3	1	0∼0.1	1
Case21	✗	NA	NA	✗	NA	NA	✓	0.2∼0.55	2	✓	0.05∼0.15	1	0.1∼0.2	1
Case22	✓	0∼0.2	1	✓	0∼0.1	1	✓	0∼0.25	1	✓	0.1∼0.25	1	0.1∼0.2	1
Case23	✓	0.1	1	✓	0∼0.1	1	✓	0.2	1	✓	0.1∼0.2	1	0∼0.1	1
Case24	✓	0∼0.2	1	✗	NA	NA	✓	0.1∼0.2	1	✓	0.1∼0.2	1	0.1∼0.15	1
Case25	✗	NA	NA	✓	0.1∼0.2	1	✓	0∼0.05	1	✓	0∼0.2	1	0∼0.35	1
Case26	✓	0.1∼0.15	1	✗	NA	NA	✗	NA	NA	✗	NA	NA	0.1∼0.15	1
Case27	✓	0.4∼0.5	2	✗	NA	NA	✗	NA	NA	✗	NA	NA	0.1∼0.5	2
Case28	✓	0.7∼0.8	3	✓	0.15∼0.25	1	✓	0.7∼0.8	3	✗	NA	NA	0∼0.1	1
Case29	✓	0∼0.1	1	✗	NA	NA	✗	NA	NA	✗	NA	NA	0∼0.4	2
Case30	✗	NA	NA	✓	0.4∼0.55	1∼8	✗	NA	NA	✗	NA	NA	0.1∼0.6	2
Case31	✗	NA	NA	✓	0.3∼0.6	2	✗	NA	NA	✗	NA	NA	0.1∼0.55	2
Case32	✓	0.35∼0.45	2	✓	0.35	1	✗	0∼0.35	2	✗	NA	NA	0.5	2
Case33	✓	0.4∼0.6	2	✓	0∼0.2	1	✓	0.5∼0.7	2	✓	0.45∼0.55	2	0.4∼0.5	2
Case34	✓	0.45∼0.6	2	✗	NA	NA	✓	0.8∼0.95	3	✗	NA	NA	NA	NA
Case35	✓	0.3	1	✓	0.8∼1	3	✗	NA	NA	✗	0.3	1	0.1∼0.65	2
Case36	✓	0∼0.3	1	✗	NA	NA	✗	NA	NA	✗	NA	NA	0.1	1
Case37	✓	0∼0.2	1	✓	0∼0.2	1	✓	0.05∼0.3	1	✓	0.05∼0.3	1	0.05∼0.1	1
Case38	✓	0.05∼0.1	1	✗	NA	NA	✓	0.05∼0.1	1	✗	NA	NA	NA	NA
Case39	✓	0.05∼0.1	1	✗	NA	NA	✓	0.15∼0.2	1	✗	NA	NA	0.1∼0.2	1
Case40	✗	NA	NA	✗	NA	NA	✓	0∼02	1	✗	NA	NA	NA	NA
Case41	✗	NA	NA	✗	NA	NA	✓	0.5∼0.6	2	✗	NA	NA	0.1	1
Case42	✗	NA	NA	✗	NA	NA	✓	0.1∼0.2	1	✗	NA	NA	0.1∼0.3	1
Case43	✗	NA	NA	✓	0.1∼0.45	2	✓	0.15∼0.3	1	✓	0.15∼.025	1	0.1	1
Case44	✓	0∼0.2	1	✗	NA	NA	✗	NA	NA	✗	NA	NA	0∼0.2	1
Case45	✓	0.2∼0.35	1	✗	NA	NA	✗	NA	NA	✗	NA	NA	0.1	1
Case46	✗	NA	NA	✗	NA	NA	✓	0∼0.2	1	✗	NA	NA	0.1∼0.2	1
Case47	✗	NA	NA	✓	0∼0.2	1	✗	NA	NA	✓	0∼0.2	1	0.1∼0.2	1
Case48	✗	NA	NA	✓	0∼0.2	1	✗	NA	NA	✗	NA	NA	0.1∼0.3	1
Case49	✓	0∼0.6	2	✗	NA	NA	✓	0.5∼0.6	2	✗	NA	NA	0.2	1
Case50	✓	0.8∼1	3	✓	0.8∼1	3	✗	NA	NA	✗	NA	NA	0.1∼0.2	1
Case51	✓	0.4∼0.6	2	✓	0∼0.3	1	✓	0.5	2	✓	0.1∼0.35	1	0.1∼0.45	2
Case52	✓	0.8	3	✗	NA	NA	✓	0.6∼0.8	3	✓	0.4	1	0.2∼0.7	2
Case53	✓	0.7∼0.85	3	✓	0.74∼0.85	3	✓	0∼0.2	1	✓	0∼0.2	1	0.1∼0.5	2
Case54	✓	0∼0.2	1	✓	0∼0.2	1	✗	NA	NA	✗	NA	NA	0.1∼0.15	1
Case55	✓	0.05∼0.15	1	✓	0.05∼0.15	1	✓	0.4∼0.6	2	✗	NA	NA	0.1	1
Case56	✓	0.1∼0.2	1	✗	NA	NA	✓	0.45∼0.65	2	✗	NA	NA	0.2∼0.25	1
Case57	✓	0.4∼0.85	3	✓	0.4∼0.65	2	✓	0∼0.3	1	✓	0∼0.3	1	0.1∼0.25	1
Case58	✓	0.05∼0.1	1	✓	0.05∼0.1	1	✓	0.2∼0.3	1	✓	0.2∼0.3	1	0.1∼0.2	1
Case59	✓	0.9∼1	3	✓	0.9∼1	3	✓	0.8∼1	3	✓	0.8∼1	3	0.1∼0.2	1
Case60	✓	0.9∼1	3	✗	NA	NA	✓	0.5∼0.6	2	✓	0.2∼35	1	0∼0.2	1
Case61	✓	0.1∼0.35	1	✗	NA	NA	✓	0.65∼0.85	3	✓	0.2∼0.45	3∼8	0.1∼0.2	1

Figure 4.13: (Top Row) Patients in a reclined position, (Bottom Row) The maximum and the Minimum of the estimated JVP height on the neck.

patient has two measurements indicating label locations. Check marks and cross marks are used for each position of the patient to show if the results are consistent with the doctor's opinion. One measurement is taken with the patient at a 45-degree reclined position, and the other while sitting upright at 90 degrees. The lower of the two values corresponds to the JVP measurement in the upright position, while the higher value corresponds to the 45-degree position. For analysis, the lower value is compared with the 45-degree position, and the higher value with the 90-degree position. This is because JVP is more visible when the patient is reclined and often collapses when sitting upright, as illustrated in Figure 4.13. (JVP readings are marked with blue dots, and carotid points with red dots.)

If the patient's reading is the same for the reclined and sitting up position, then one value will be used for different positions' analysis. As shown, six "NA" values in the rightmost column are associated with discrepancies between the labels and the physician's video review. These cases have been excluded from further analysis. Each class is assigned a number 1,2,3 representing **Normal**, **Mild** and **Severe** cases similar to what we already have with doctor's labels and annotation. The experiment has been done in four different situation namely 45P˙ Direct View, 90P˙ Direct View, 45P˙ Tangential View, 90P˙ Tangential View. Figure 4.14 presents the unresolved cases for various positions and camera angles. These issues may arise from the inability of optical flow techniques to capture signals, which can be attributed to factors such as the homogeneity of the skin surface, SNR, lighting conditions at different angles, and other factors that are further discussed in the following chapter.

Figure 4.15 shows the confusion matrix for four different situations with an accuracy of 46% for 45P˙ Tangential View, 54% for 45P˙ Direct View, 72% for 90P˙ Tangential View, and 59% for 90P˙ Direct View. This pattern is further visualized in Figure 4.16, which illustrates the mean of the predicted values, calculated as the average of the lower and upper bounds for each patient and each position, plotted against the clinically measured distances for three severity categories. While it provides a statistical comparison between the proposed approach and clinical measurements, it does not explicitly distinguish between true positives, false positives, or false negatives. As such, its interpretation is primarily statistical rather than diagnostic.

As mentioned, Figure 4.16 provides a comparison of the proposed approach's results

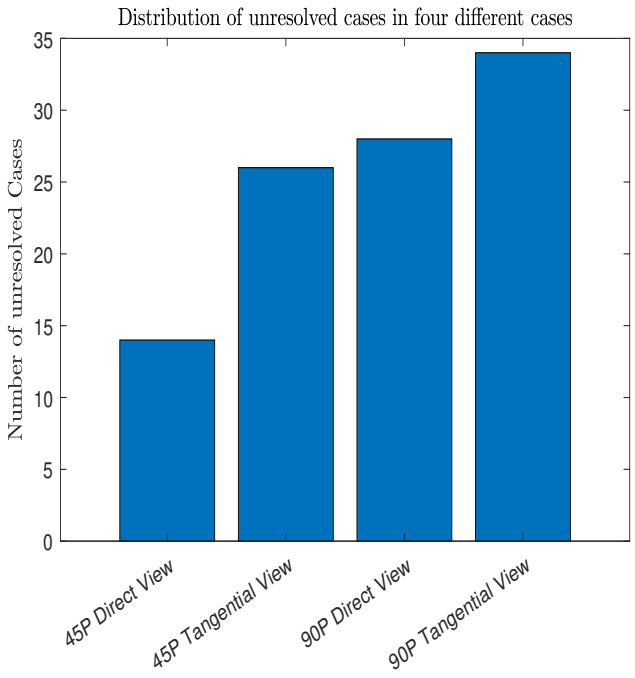

Figure 4.14: Number of unresolved cases for each position and camera view (the lower the better).

with expert measurements, specifically in terms of the JVP distance from the clavicle, which is illustrated as follows: The values in Table 4.6 correspond to those in the figure and are reported as median [$Q1$ $Q3$], with Q1 and Q3 denoting the first and third quartiles.

Table 4.6: Quantitative values of statistical comparison of measured JVP height.

Position	**Clinical**			**Experimental**		
	Normal	Mild	Severe	Normal	Mild	Severe
45° Reclined – Tangential View	0.2 [0.1–0.2]	0.475 [0.4–0.5]	NaN	0.2 [0.2-0.3]	0.6 [0.475–0.6]	0.8 [0.8–0.965]
45° Reclined – Direct View	0.2 [0.125–0.2]	0.5 [0.45–0.6]	NaN	0.2 [0.125–0.2]	0.55 [0.5–0.6]	0.85 [0.8125–1]
90° Upright – Tangential View	0.1 [0.025–0.1]	0.4 [0.4–0.4]	NaN	0.225 [0.2–0.3]	0.425 [0.35–0.55]	1 [1–1]
90° Upright – Direct View	0.1 [0.1–0.1]	0.45 [0.4–0.5]	NaN	0.2 [0.125–0.2]	0.5 [0.375–0.6]	1 [0.925–1]

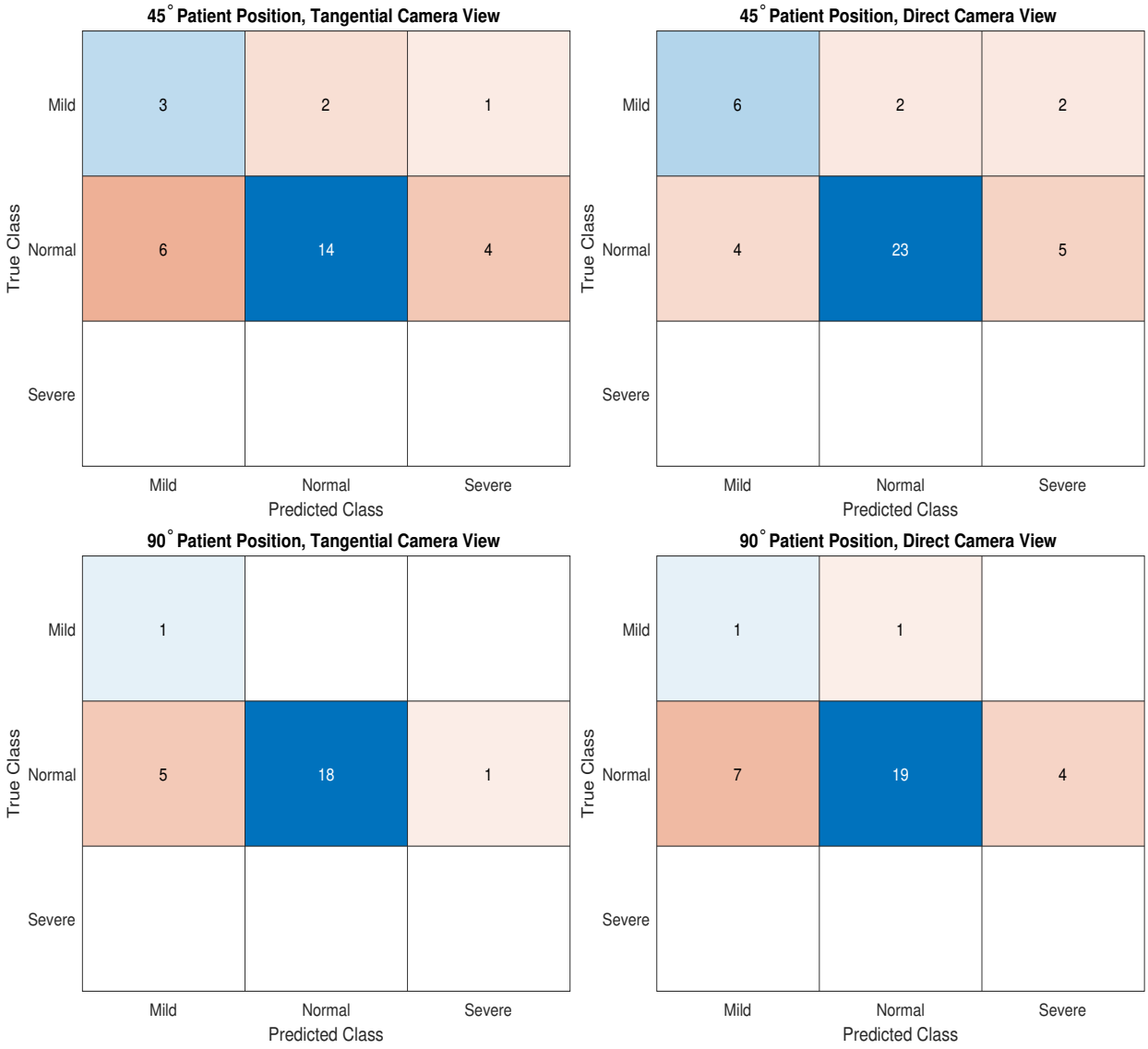

Figure 4.15: Confusion Matrix of the Predicted Classes Vs. True Classes for each Patient's Position and Camera View.

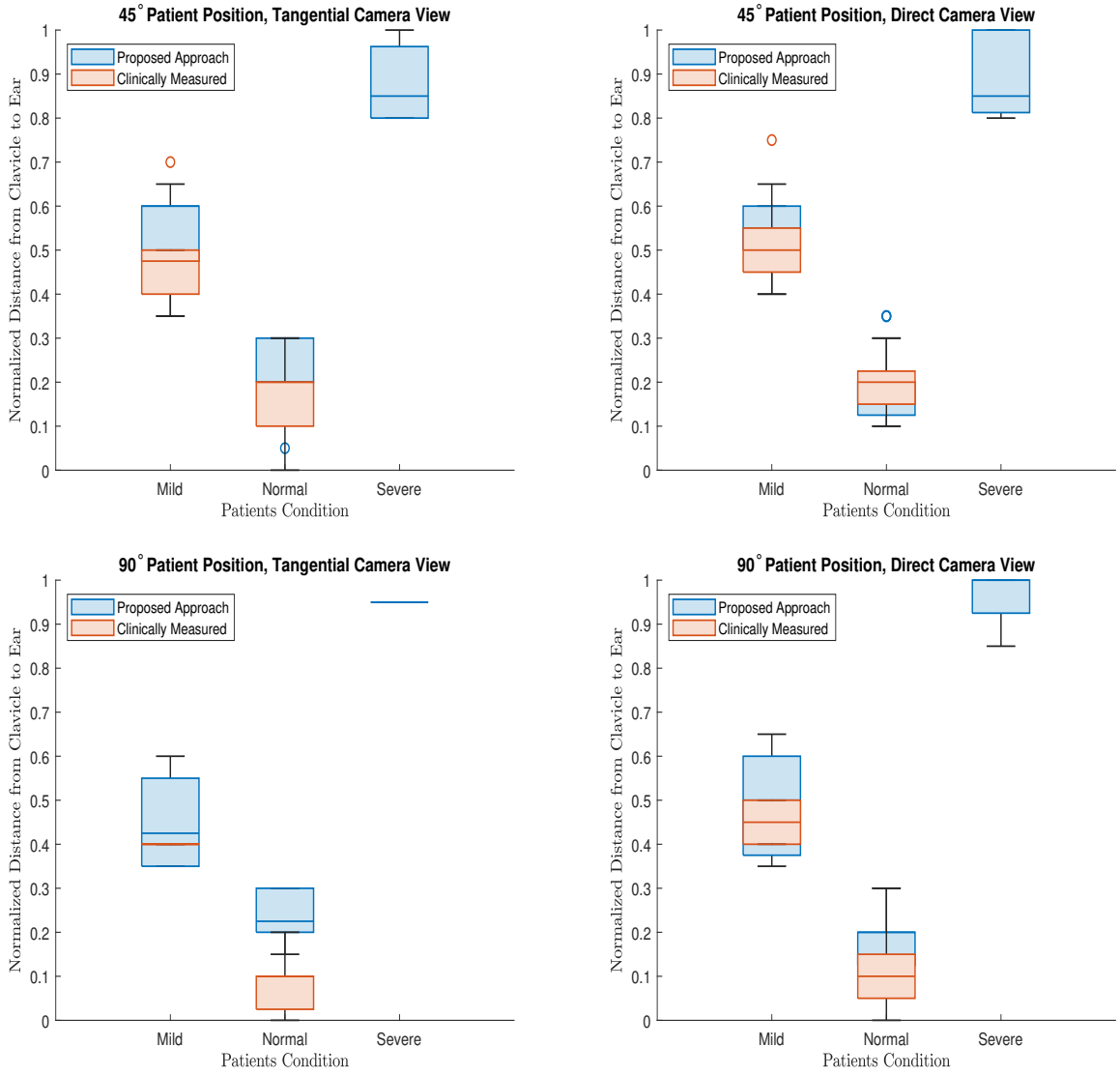

Figure 4.16: T-test results of the clinically measured versus proposed approach for the resolved cases.

4.6 Discussion

Locating the JVP on the neck with a camera involves various challenges stemming from technical, physiological, and environmental considerations. Like any research study, this

investigation also has its limitations, which will be discussed in the following sections.

- **Illumination** is one of the primary challenges in video acquisition. Proper lighting is essential for capturing the subtle details of veins, but excessive light or inconsistent lighting can cast harsh shadows or overexpose the skin, thereby obscuring vein visibility (as shown in Figure 4.17). Ambient light interference can further complicate the situation, making it difficult to distinguish veins without specialized lighting, such as Near-Infrared (NIR). In the majority of research studies focusing on the neck, artificial light sources have been employed, including Light Emitting Diode (LED) arrays in the blue and NIR ranges, fluorescent lamps, halogen lamps, and other unspecified lighting setups. As is described in Chapter 3, usual room lighting (Fluorescent lamp) has been used for the experiment and data collection. However, the effects of the amount of light reflected on the neck will be reviewed in the following section. For this experiment, two different lighting conditions have been studied. For one minute, the patient is recorded under the usual lighting conditions, and for another minute, the lamp is turned off, and only environmental light is used. For each of the patients, ECG and PPG are also available, and the results of the algorithms applied to the videos can be compared fairly, indicating the importance of lighting conditions in video acquisition and related studies. Also, glare from oily or shiny skin can reduce the quality of captured images, while reflections from the environment can interfere with vein detection. These reflections and unwanted highlights can obscure fine details, leading to poor imaging results.

Figure 4.17: (Top Row) Patients in a room with environmental lighting, (Bottom Row) Patients in a room with fluorescent lamp lighting.

- **Specialized cameras** are required for extracting pulsatile signals from the face, such as those related to heart rate or blood flow, by detecting subtle changes in skin colour or reflectance due to blood flow. Camera resolution is critical as well. A high-resolution camera is required to capture the fine details of veins, and the depth of focus needs to be adjusted precisely to avoid blurring. However, ensuring the camera is correctly focused on the surface of the skin while keeping the veins sharp can be tricky. Here are the primary types of cameras used for this purpose, along with their characteristics and roles in capturing pulsatile signals.

  RGB cameras capture images in red, green, and blue colour channels. These can

be used to extract pulsatile signals by detecting subtle colour changes in the skin as blood volume changes with each heartbeat. As demonstrated in [57], standard webcams can be used to extract pulse signals from facial regions by applying EVM, which enhances subtle temporal variations in pixel intensity caused by blood flow. Similarly, [24] showed that heart rate can be estimated using conventional digital cameras by analyzing the green color channel, which corresponds to changes in blood volume.

NIR cameras operate in the infrared spectrum, typically between 700 nm and 1000 nm, making them less sensitive to visible light variations. This characteristic makes NIR cameras suitable for environments with fluctuating lighting conditions, as they provide more consistent measurements. Moreover, NIR cameras can capture signals from deeper skin layers, potentially improving signal quality. Despite these benefits, NIR cameras offer limited colour information compared to RGB cameras, which may impact certain analysis techniques. Additionally, they are generally more expensive than standard RGB cameras.

High-speed cameras, which capture video at very high frame rates often exceeding 100 FPS, provide high temporal resolution essential for precise detection of rapid physiological changes. This high temporal resolution allows for more accurate measurements of heart rate and other physiological signals, and enhances the ability to analyze subtle movements and variations. However, high-speed cameras generate large amounts of data, requiring significant storage and processing power, and they are typically more expensive than standard cameras.

Depth-sensing cameras, such as those using Time-of-Flight (ToF) or Structured Light technologies, provide information about the distance between the camera and the subject. While primarily used for capturing 3D information, depth cameras can complement rPPG by providing additional context for signal extraction. The 3D information helps in isolating the face from the background, reducing noise from irrelevant movements, and improving the accuracy of physiological signal extraction by providing spatial context. However, depth cameras involve more complex data processing requirements and are generally more expensive than standard RGB cameras.

Thermal cameras detect infrared radiation emitted by objects, translating it into temperature data. Although not traditionally used for rPPG, some research explores their potential in capturing physiological signals based on temperature variations related to blood flow. Thermal cameras are advantageous because they operate in the non-visible spectrum, similar to NIR cameras, and provide additional physiological information that can complement rPPG measurements. However, they typically have lower temporal resolution, which may limit the accuracy of physiological measurements, and they are generally more expensive and less commonly available for consumer use.

- **Skin tone** plays a significant role. In darker skin tones, the higher melanin content can make veins harder to detect. Additionally, the texture of the skin—whether it's smooth or features wrinkles, freckles, or hair—adds noise to the image, which

complicates the process of isolating veins from other surface details.

- **Physiological variability** adds further complexity. Individuals have unique vein structures, and some people naturally have more prominent veins on their necks than others. For some, veins may be barely visible or hidden beneath layers of skin, which makes detection difficult. Age also plays a role, as older people may have more visible veins due to thinner skin, while younger individuals might have veins that are less apparent. Another challenge is the depth of veins. Veins that are closer to the skin's surface are easier to detect, but those that are deeper are less visible and harder to capture with a regular camera. Factors like blood pressure, hydration, and body temperature can affect vein prominence, causing them to appear more or less visible depending on the subject's physical state.

- **Motion artifacts** present a significant issue. Even slight movements, such as swallowing or neck rotation, can distort the image and affect vein detection. Additionally, natural muscle movements and the pulse of the veins themselves can cause shifts, making it hard to get a stable view.

In summary, detecting veins on the neck using a camera is a challenging task that requires overcoming lighting issues, skin tone variations, motion artifacts, and resolution limitations, often demanding advanced image processing techniques and specialized equipment.

5

Conclusion

5.1 Introduction

From the beginning of this study, several key questions guided our research efforts. One of the primary challenges was incorporating criteria, such as those outlined in Table 2.1, which are grounded in cardiovascular physiology, into the processing framework. Another major

concern was the development and integration of separate components, including object detection and signal processing modules, into a unified and comprehensive methodology. While multiple approaches were explored, only the final and most effective method is presented in this study.

In the final chapter, the accuracy of the proposed method was evaluated against a dataset obtained with US measurements, which requires significant resources and healthcare personnel. Additionally, the study's limitations, discussed in the relevant section, highlight certain constraints that may have influenced the results. These factors underscore the need for further refinements to enhance the robustness and applicability of the framework in real-world clinical settings.

5.2 Contribution

The findings of this study contribute to several key areas.

Firstly, the integration of an object detection module for identifying key anatomical landmarks namely the ear, clavicle, and sternal notch, introduces a novel approach for automated anatomical feature recognition. This research uniquely addresses the independent detection of both the clavicle and sternal notch, which are critical for accurate JVP measurement. To date, no prior work has specifically tackled the challenge of reliably detecting these anatomical points in a fully automated manner, showing the originality of this framework. While the detection rate still requires further refinement to enhance

robustness across diverse skin tones, lighting conditions, and body types, this study establishes a foundational methodology for future advancements in automated anatomical landmark detection. This challenge was framed as a key research question, investigating the feasibility of detecting anatomical features on the neck in a real-world setting. As demonstrated in Chapter 4, the successful detection of the clavicle and sternal notch significantly contributes to the overall framework for JVP measurement, illustrating its importance in achieving a reliable, non-invasive clinical tool.

Secondly, the study presents a novel methodology for detecting HR and RR on the neck using a combination of neck segmentation and sMOT, a voting-based strategy, and signal processing techniques. The challenge of achieving a robust and reliable measurement of these vital signs was a key aspect of this research and was posed as a second research question. As discussed in Chapter 4, the results demonstrate the feasibility of this method, highlighting its potential for real-world applications in telemedicine, cardiopulmonary assessment, and continuous patient monitoring.

Thirdly, and most importantly, this research represents one of the first attempts to semi-automated JVP height measurement on the neck, stepping toward non-invasive cardiovascular monitoring. Traditional JVP measurement relies on manual clinical assessment, requiring trained professionals to visually estimate the jugular pulsation relative to anatomical reference points. This study bridges the gap between clinical expertise and automation by developing a comprehensive AI-driven framework that eliminates the need for manual intervention. This framework integrates multiple advanced computer vision and signal processing techniques into a seamless pipeline. It begins with object detec-

tion, which automatically identifies critical anatomical landmarks such as the ear, clavicle, and sternal notch, ensuring accurate spatial localization. To extract JVP waveforms with high fidelity, the system employs a combination of optical flow analysis and Eulerian video magnification, enhancing subtle pulsations that would otherwise be imperceptible. By successfully combining these elements into a fully automated manner, this research provides a major step forward in remote cardiovascular assessment. The ability to estimate JVP height non-invasively using only a standard camera opens new possibilities for heart failure monitoring and early cardiovascular disease detection. This work directly addresses the third research question, which explores the feasibility of automating JVP measurement, and the findings presented in Chapter 4 demonstrate the framework's potential to serve as a valuable tool in both clinical and remote healthcare settings.

Finally, one of the primary goals of this study was to develop a real-time framework for JVP height measurement. However, during the evaluation of computational performance, it became evident that the system did not operate in real-time. Specifically, analyzing a one-minute video required over an hour of processing time, with the majority of this delay attributed to the video magnification step. The computational burden of video magnification significantly increased as the number of processing layers expanded, making real-time implementation impractical in its initial form. To address these challenges, the code was restructured and optimized, and CUDA-based GPU acceleration was integrated into the processing pipeline. By leveraging parallel computing capabilities, particularly in optical flow calculations and video magnification processing, the overall runtime was successfully reduced to approximately twenty minutes per one-minute video. This represents a notable

improvement in the system's efficiency and marks a significant advancement toward achieving real-time processing. Despite these optimizations, further performance enhancements are still required to achieve true real-time operation.

5.3 Future Work

To improve the accuracy and reliability of JVP measurements, it is essential to address several key challenges that may negatively impact performance. Future research should focus on implementing strategic advancements to mitigate these limitations and enhance the robustness of the measurement framework.

To improve the accuracy of the results, one suggestion is to combine data from both the direct and tangential views. Additionally, implementing multiple magnification approaches can help ensure that the most optimal version, offering effective enhancement with minimal added noise, is applied to the video. Moreover, training different networks with different sizes helps to reduce the problem of false positives and provides a more reliable framework for object detection.

As discussed earlier, unwanted motions and motion artifacts negatively impact the accuracy of JVP measurement. To address this limitation, implementing an advanced motion registration scheme is crucial for effectively tracking the subtle physiological changes occurring in the neck. A robust registration framework would allow for the precise alignment of consecutive frames, minimizing distortions caused by involuntary head movements,

breathing-induced shifts, or external disturbances. By compensating for these variations, an improved registration approach would enhance the detection of pulsatile signals by ensuring that the jugular vein remains consistently positioned throughout the video sequence. This would allow even the smallest fluctuations in venous pulsation to be accurately captured and analyzed, leading to more reliable and precise JVP assessments.

Illumination proved to be a critical variable in this study, significantly influencing the accuracy and consistency of JVP estimation. Variations in lighting conditions, such as changes in ambient brightness, shadowing effects, and reflections, introduced challenges in reliably capturing venous pulsations. To overcome these limitations, there is a strong need to develop a robust framework capable of adapting to different lighting environments. One potential solution involves recording video samples under diverse lighting conditions to build a more comprehensive dataset that enables the system to generalize effectively. Additionally, integrating an end-to-end deep learning network could significantly enhance the framework's resilience to lighting variations by learning invariant features that are less sensitive to illumination changes could be considered as an alternative.

By focusing on these key areas including advanced registration techniques, and the development of end-to-end neural networks to handle the variety of illumination changes, future work can substantially enhance the performance of the JVP measurement method. These advancements are expected to lead to more accurate cardiovascular assessments and ultimately result in better patient outcomes.

References

[1] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, *et al.*, "Heart disease and stroke statistics—2023 update: A report from the american heart association," *Circulation*, vol. 147, no. 8, e93–e621, 2023.

[2] W. C. MEMBERS, B. Bozkurt, T. Ahmad, *et al.*, "Heart failure epidemiology and outcomes statistics: A report of the heart failure society of america," *Journal of cardiac failure*, vol. 29, no. 10, p. 1412, 2023.

[3] A. S.-T. Leader, T. Moehling-Communicator, K. Harris-BWIG, D. Horn-BSAC, S. Yale, and C. Brace, "Bedside device to measure the jugular venous pressure,"

[4] F. Abnousi, G. Kang, J. Giacomini, *et al.*, "A novel noninvasive method for remote heart failure monitoring: The eulerian video magnification applications in heart failure study (amplify)," *NPJ digital medicine*, vol. 2, no. 1, p. 80, 2019.

[5] A. T. Sandhu, J. D. Goldhaber-Fiebert, D. K. Owens, M. P. Turakhia, D. W. Kaiser, and P. A. Heidenreich, "Cost-effectiveness of implantable pulmonary artery pressure monitoring in chronic heart failure," *JACC: Heart Failure*, vol. 4, no. 5, pp. 368–375, 2016.

[6] I. García-López and E. Rodriguez-Villegas, "Extracting the jugular venous pulse from anterior neck contact photoplethysmography," *Scientific reports*, vol. 10, no. 1, p. 3466, 2020.

[7] J.-Y. Park, Y. Lee, Y.-W. Choi, *et al.*, "Preclinical evaluation of a noncontact simultaneous monitoring method for respiration and carotid pulsation using impulse-radio ultra-wideband radar," *Scientific reports*, vol. 9, no. 1, p. 11 892, 2019.

[8] A. Moço, P. Hamelmann, S. Stuijk, and G. de Haan, "The importance of posture and skin-site selection on remote measurements of neck pulsations: An ultrasonographic study," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2018, pp. 5918–5921.

[9] R. Amelard, A. D. Robertson, C. A. Patterson, H. Heigold, E. Saarikoski, and R. L. Hughson, "Optical hemodynamic imaging of jugular venous dynamics during altered central venous pressure," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 8, pp. 2582–2591, 2021.

[10] T. Zhang, M. Bolic, M. H. Davoodabadi Farahani, T. Zadorsky, and R. Sabbagh, "Non-contact heart rate and respiratory rate estimation from videos of the neck," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4.

[11] L. Wang, J. Harrison, E. Dranow, N. Aliyev, and L. Khor, "Accuracy of ultrasound jugular venous pressure height in predicting central venous congestion," *Annals of internal medicine*, vol. 175, no. 3, pp. 344–351, 2022.

[12] S. McGee, "Chapter 36 - inspection of the neck veins," in *Evidence-Based Physical Diagnosis (Fourth Edition)*, S. McGee, Ed., Fourth Edition, Philadelphia: Elsevier, 2018, 301–314.e3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323392761000366.

[13] A. Proto, D. Conti, E. Menegatti, A. Taibi, and G. Gadda, "Plethysmography system to monitor the jugular venous pulse: A feasibility study," *Diagnostics*, vol. 11, no. 12, p. 2390, 2021.

[14] E. BraunWald and S. Zoneraich, "Heart disease: A textbook of cardiovascular medicine," *JAMA-Journal of the American Medical Association-International Edition*, vol. 277, no. 15, p. 1247, 1997.

[15] B. Boerrigter, P. Trip, H. J. Bogaard, *et al.*, "Right atrial pressure affects the interaction between lung mechanics and right ventricular function in spontaneously breathing copd patients," *PLoS One*, vol. 7, no. 1, e30208, 2012.

[16] B. Lipton, "Estimation of central venous pressure by ultrasound of the internal jugular vein," *The American journal of emergency medicine*, vol. 18, no. 4, pp. 432–434, 2000.

[17] C. E. Bagwell, A. M. Salzberg, R. E. Sonnino, and J. H. Haynes, "Potentially lethal complications of central venous catheter placement," *Journal of Pediatric Surgery*, vol. 35, no. 5, pp. 709–713, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022346800260478.

[18] L. Wang, J. Harrison, E. Dranow, N. Aliyev, and L. Khor, "Accuracy of ultrasound jugular venous pressure height in predicting central venous congestion," *Annals of internal medicine*, vol. 175, no. 3, pp. 344–351, 2022.

[19] F. Sisini, M. Tessari, G. Gadda, *et al.*, "An ultrasonographic technique to assess the jugular venous pulse: A proof of concept," *Ultrasound in medicine & biology*, vol. 41, no. 5, pp. 1334–1341, 2015.

[20] T. B. Conroy, J. Zhou, and E. C. Kan, "Jugular venous pulse waveform extraction from a wearable radio frequency sensor," *IEEE sensors journal*, vol. 23, no. 9, pp. 10 140–10 148, 2023.

[21] E. Menegatti, A. Proto, G. Paternò, *et al.*, "The effect of submaximal exercise on jugular venous pulse assessed by a wearable cervical plethysmography system," *Diagnostics*, vol. 12, no. 10, p. 2407, 2022.

[22] S. Suzuki, M. Hoshiga, K. Kotani, and T. Asao, "Assessment of non-contact measurement using a microwave sensor to jugular venous pulse monitoring," *Journal of Biomedical Science and Engineering*, vol. 14, no. 3, pp. 94–102, 2021.

[23] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE transactions on biomedical engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.

[24] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light.," *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[25] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.," *Optics express*, vol. 18, no. 10, pp. 10 762–10 774, 2010.

[26] W. Wang, A. C. Den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2016.

[27] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3430–3437.

[28] W. Chen and D. McDuff, "Deepphys: Video-based physiological measurement using convolutional attention networks," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 349–365.

[29] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," *arXiv preprint arXiv:1905.02419*, 2019.

[30] X. Liu, J. Fromm, S. Patel, and D. McDuff, "Multi-task temporal shift attention networks for on-device contactless vitals measurement," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19 400–19 411, 2020.

[31] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement. arxiv 2017," *arXiv preprint arXiv:1709.00962*,

[32] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognition Letters*, vol. 124, pp. 82–90, 2019.

[33] R. Stricker, S. Müller, and H.-M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, 2014, pp. 1056–1062.

[34] T. T. Dang, C. Huynh, A. Dinh, and K. Tran, "Recognizing area of pulsations on the neck via video camera systems," in *2015 International Conference on Advanced Technologies for Communications (ATC)*, IEEE, 2015, pp. 139–144.

[35] F. Abnousi, G. Kang, J. Giacomini, *et al.*, "A novel noninvasive method for remote heart failure monitoring: The eulerian video magnification applications in heart failure study (amplify)," *NPJ digital medicine*, vol. 2, no. 1, p. 80, 2019.

[36] T. Burton, G. Saiko, and A. Douplik, "Towards development of specular reflection vascular imaging," *Sensors*, vol. 22, no. 8, p. 2830, 2022.

[37] E. J. Lam Po Tang, A. HajiRassouliha, M. P. Nash, P. M. Nielsen, A. J. Taberner, and Y. O. Cakmak, "Non-contact quantification of jugular venous pulse waveforms from skin displacements," *Scientific reports*, vol. 8, no. 1, p. 17 236, 2018.

[38] R. Amelard, R. L. Hughson, D. K. Greaves, *et al.*, "Non-contact hemodynamic imaging reveals the jugular venous pulse waveform," *Scientific reports*, vol. 7, no. 1, p. 40 150, 2017.

[39] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[40] R. Kaur and S. Singh, "A comprehensive review of object detection with deep learning," *Digital Signal Processing*, vol. 132, p. 103 812, 2023.

[41] Y. LeCun, B. Boser, J. S. Denker, *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[43] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 142–158, 2015.

[44] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[45] C. Szegedy, W. Liu, Y. Jia, *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[46] A. Vijayakumar and S. Vairavasundaram, "Yolo-based object detection models: A review and its applications," *Multimedia Tools and Applications*, pp. 1–40, 2024.

[47] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[48] W. Liu, D. Anguelov, D. Erhan, *et al.*, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, Springer, 2016, pp. 21–37.

[49] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[50] J. Zhou, Y. Tian, W. Li, R. Wang, Z. Luan, and D. Qian, "Ladet: A light-weight and adaptive network for multi-scale object detection," in *Asian Conference on Machine Learning*, PMLR, 2019, pp. 912–923.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[52] R. Girshick, "Fast r-cnn," *arXiv preprint arXiv:1504.08083*, 2015.

[53] Ultralytics, *Yolov8*, `https://docs.ultralytics.com/models/yolov8/UltralyticsYOLODocs`, Sep. 2024. [Online]. Available: `https://docs.ultralytics.com/models/yolov8/`.

[54] M. S. Mohsen, A. B. Fakhri, N. M. Ahmed, M. F. Mahmood, and S. L. Mohammed, "Video magnification techniques: Medical applications and comparison of methods," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 1105, 2021, p. 012 074.

[55] A. M. Ahmed, M. Abdelrazek, S. Aryal, and T. T. Nguyen, "An overview of eulerian video motion magnification methods," *Computers & Graphics*, 2023.

[56] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, "Motion magnification," *ACM transactions on graphics (TOG)*, vol. 24, no. 3, pp. 519–526, 2005.

[57] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM transactions on graphics (TOG)*, vol. 31, no. 4, pp. 1–8, 2012.

[58] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Riesz pyramids for fast phase-based video magnification," in *2014 IEEE International Conference on Computational Photography (ICCP)*, IEEE, 2014, pp. 1–10.

[59] L. Liu, L. Lu, J. Luo, J. Zhang, and X. Chen, "Enhanced eulerian video magnification," in *2014 7th International Congress on Image and Signal Processing*, IEEE, 2014, pp. 50–54.

[60] A. Al-Naji, S.-H. Lee, and J. Chahl, "An efficient motion magnification system for real-time applications," *Machine Vision and Applications*, vol. 29, pp. 585–600, 2018.

[61] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Phase-based video motion processing," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 4, pp. 1–10, 2013.

[62] T.-H. Oh, R. Jaroensri, C. Kim, *et al.*, "Learning-based video motion magnification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 633–648.

[63] M. Zhai, X. Xiang, N. Lv, and X. Kong, "Optical flow and scene flow estimation: A survey," *Pattern Recognition*, vol. 114, p. 107 861, 2021.

[64] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[65] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI'81: 7th international joint conference on Artificial intelligence*, vol. 2, 1981, pp. 674–679.

[66] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer vision and image understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[67] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 5, pp. 978–994, 2010.

[68] A. Dosovitskiy, P. Fischer, E. Ilg, *et al.*, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE International Conference on computer vision*, 2015, pp. 2758–2766.

[69] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8934–8943.

[70] Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, Springer, 2020, pp. 402–419.

[71] C. Arrow, M. Ward, J. Eshraghian, and G. Dwivedi, "Capturing the pulse: A state-of-the-art review on camera-based jugular vein assessment," *Biomedical Optics Express*, vol. 14, no. 12, pp. 6470–6492, 2023.

[72] D. J. McDuff, E. B. Blackford, and J. R. Estepp, "The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, IEEE, 2017, pp. 63–70.

[73] A. Kirillov, Y. Wu, K. He, and R. Girshick, "Pointrend: Image segmentation as rendering," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9799–9808.

[74] T.-Y. Lin, M. Maire, S. Belongie, *et al.*, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, Springer, 2014, pp. 740–755.

[75] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l 1 optical flow," in *Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29*, Springer, 2007, pp. 214–223.

[76] J.-Y. Bouguet *et al.*, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel corporation*, vol. 5, no. 1-10, p. 4, 2001.

[77] T. Senst, V. Eiselein, and T. Sikora, "Robust local optical flow for feature tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 9, pp. 1377–1387, 2012.

[78] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*, Springer, 2003, pp. 363–370.

[79] J. Wulff and M. J. Black, "Efficient sparse-to-dense optical flow estimation using a learned basis and layers," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 120–130.

[80] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV 8*, Springer, 2004, pp. 25–36.

[81] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *IEEE Intenational Conference on Computer Vision (ICCV)*, Sydney, Australia, Dec. 2013. [Online]. Available: `http://hal.inria.fr/hal-00873592`.