# Advancements in Hemodynamic Measurement: Arterial Resonance, Ultrasound, and Machine Learning

Thesis by
Dominic Yurk

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

**Caltech**

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended May 31, 2023

# ACKNOWLEDGEMENTS

sniff test. Dr. Joshua Barrios also provided significant help with orienting me to the dataset and showing me how to interact with and manipulate it, saving me countless hours of muddling through it myself. This project was supported by the Merkin Institute for Translational research through an Innovation Seed Grant.

I am also happy to be able to thank all of my personal supporters who have helped carry me through this journey. My parents have always been an incredible support system, consistently advocating for me and always finding new opportunities to broaden my horizons for both education and the fun parts of life outside of it. My sister Emily has a scientific mind as well but also an incredible artistic ability, and she helps me keep my mind on the world beyond science. The many friends I've made during my time in grad school also made sure that I remembered to have fun and get outside of the Caltech bubble. By far my biggest supporter in grad school has been my wonderful and incredibly talented fiancée, Nitika Yadlapalli. Through late work sessions, wine education nights, convincing me to run a half marathon, and taking me on backpacking trips to get away from it all, Nitika has helped me become a better person inside and outside of the lab. She'll be joining me as a doctor very soon, and I can't wait to see what we both accomplish together in our careers and lives.

Last but not least, this dissertation is dedicated to my adorable cats Percy and Dreamsicle. While they are unlikely to have any scientific accomplishments of their own, their purrs and snuggles will continue to be essential contributors to my own.

# ABSTRACT

This thesis covers two separate projects which both use ultrasound to measure a form of blood pressure in very different ways. The first project focuses on the noninvasive measurement of continuous arterial blood pressure via the previously unstudied phenomenon of arterial resonance. While prior research efforts have attempted many methods of noninvasive blood pressure measurement, none has been able to generate continuous, calibration-free measurements based on a first-principles physical model. This work describes the derivation of this resonance-based model, its *in vitro* validation, and its *in vivo* testing on 60 subjects. This testing resulted in robust resonance detection and accurate calculation of BP in the large majority of evaluated subjects, representing very promising performance for the first test of a new biomedical technology. The second study changes focus to the measurement of blood pressure in the right atrium of the heart, an important clinical indicator in heart disease patients. Rather than developing a new physical approach, this project used machine learning to model the existing assessments made by cardiologists. Comparison to gold standard invasive catheter measurements showed that model predictions were statistically indistinguishable from cardiologist measurements. Both of these projects represent significant advances in expanding precise blood pressure measurements beyond critical care units and expanding access to a much broader population.

# TABLE OF CONTENTS

# LIST OF ILLUSTRATIONS

# LIST OF TABLES

*C h a p t e r 1*

# INTRODUCTION

## 1.1   Ultrasound and Its Clinical Uses

Ever since the invention of the medical X-ray over 100 years ago, engineers have devoted great effort to developing methods of peering inside the human body without breaching the skin. Modern-day X-ray instruments using computational tomography are capable of creating extremely detailed pictures of our bodies' inner workings. However, these machines are not only large and expensive, but they also require exposure of the subject to potentially harmful radiation which must be applied in controlled doses. The drive to develop an alternative to the X-ray led to the creation of ultrasound, which uses audio signals rather than light waves to image the body.

At its core, ultrasound is the measure of acoustic impedances. The measurement begins with the emission of an audio pulse of $\sim 1 - 50$ MHz directed into the body. This range of audio frequencies propagates well in uniform, soft tissue with a relatively low acoustic impedance. However, denser tissues, such as bone, cysts, or the muscular walls of blood vessels and heart chambers have significantly higher acoustic impedance. When the ultrasound pulse hits a boundary with one of these tissues, the impedance mismatch causes a portion of the the pulse's energy to be reflected. Carefully listening for the strength and time delay of these reflected echoes allows for the reconstruction of where these impedance mismatches exist in the tissue.

The key advantage of early ultrasound scanners over X-ray imaging was the lack of harmful radiation. This led to the early adoption of ultrasound technology by obstetricians for prenatal imaging, as even low doses of X-ray exposure could be harmful to a developing fetus. Because early medical ultrasound instruments were large, expensive, and hard to operate, most other physicians initially preferred to stick with X-rays, stethoscopes, and physical exams. Over time, however, the technology gradually improved, yielding higher-quality images from smaller instruments that could be wheeled to a patient's bedside. Physicians began to find a wide variety of diagnostic use cases for ultrasound which could not be easily replicated with X-ray, such as localizing bleeding in trauma patients, looking for signs of cirrhosis in the liver, or distinguishing a malignant tumor from a benign cyst. Modern ultrasound

machines range from high-powered devices which can create detailed 3D renderings of entire organs to ultra-portable devices which are operated from a smartphone. This versatility has made ultrasound a standard part of any doctor's training and an indispensable tool in any hospital.

Among its many uses, ultrasound is particularly well-suited for imaging the cardio-vascular system. The boundary between the muscular walls of arteries and heart chambers and the blood contained within creates a significant acoustic impedance discontinuity, leading to bright and detailed images. The continuous nature of ultrasound imaging allows doctors to investigate the dynamics of the heartbeat, including valve function, cardiac chamber outputs, and artery distention. 3D ultrasound imaging allows for accurate measurement of heart chamber volume. Doppler ultrasound allows for direct visualization of blood flow, providing even deeper insight into system behavior. Recent research efforts have moved beyond simple imaging, harnessing the power of ultrasound in innovative ways to measure previously invisible mechanical properties such as shear wave speed in arterial walls [1] or stress in muscle fibers [2]. This thesis continues along this line of inquiry, using ultrasound to measure one of the most important parameters in the cardiovascular system: blood pressure.

## 1.2 Thesis Outline

The bulk of this thesis focuses on the noninvasive measurement of continuous blood pressure via the previously unstudied phenomenon of arterial resonance. Chapter 2 begins with an overview of blood pressure measurement technology, including both modern clinical standards and the myriad of other research approaches which have attempted to replace these standards, so far without much success. It then provides a first-principles derivation of the arterial resonance approach to blood pressure measurement, as well as an analysis of its feasibility and sensitivity. Chapter 3 describes the prototype device and analysis methods designed conduct arterial resonance measurements, and demonstrates that the physical equations we derived very accurately describe behavior in an idealized *in vitro* model artery. Chapter 4 describes the additional analysis procedures needed to enable real-time blood pressure measurement and shows initial feasibility results demonstrating that resonance can be stimulated and measured in real human arteries. It then presents the results of a 60-subject clinical study, demonstrating that arterial resonance can lead to accurate BP measurements across a wide variety of physiologies and demographics. Chapter 5 describes how arterial resonance also enables the measurement of arterial

stiffness, which is recognized as an important indicator of cardiovascular health but has been difficult to measure in an accurate, localized way.

The final section of the thesis, Chapter 6, changes focus from measuring blood pressure in arteries to measuring pressure in the right atrium of the heart. It begins by describing the clinical importance of this parameter for care management in heart failure patients and how it is currently assessed using ultrasound imaging. Rather than developing a new method of conducting this measurement, we instead leverage a large dataset from the University of California, San Francisco cardiology department which combines ultrasound scans of the heart with cardiologists' measurements of right atrial pressure based on these scans. Using machine learning, we develop a model which can replicate and even improve upon the pressure measurements generated by cardiologists. This model could allow for more robust and accurate measurements of right atrial pressure, particularly in settings without regular access to experienced cardiologists.

*C h a p t e r   2*

# BIOPHYSICS OF BLOOD PRESSURE AND ARTERIAL RESONANCE

Blood pressure (BP) is one of the most important vital metrics for clinicians when assessing patient health. Elevated BP, or hypertension, can be caused by either chronic underlying conditions or acute stress and afflicts an estimated 31% of adults worldwide [3]. It is associated with heart disease as well as a wide spectrum of disorders from dementia to stroke, making it the global leading cause of premature death [3], [4]. Depressed BP, or hypotension, is usually caused by acute events such as shock or trauma, and it can also have severe consequences. Severe hypotension results in insufficient blood circulation throughout the body, potentially leading to organ damage and loss of life [5]. The risks associated with both hypertension and hypotension make BP measurement a regular procedure in any medical setting, from primary care checkups to intensive care units and operating rooms.

Measuring fluid pressure in a pipeline is generally not considered an engineering challenge, as designers of artificial systems can easily place a pressure transducer inside the pipeline during construction. Biological systems, however, do not come with any measurement devices built-in. This has led engineers to explore dozens of methods of BP measurement, but none has matched the standard set by transducers in artifical pipelines: providing real-time, continuous pressure measurements without breaching the arterial wall or performing calibration to an external source. This chapter will provide an overview of current clinical standards for BP measurement as well as the wide variety of research that has attempted to improve upon these standards. It will then provide a theoretical derivation and feasibility analysis of our novel approach to BP measurement: arterial resonance.

## 2.1   Background: Clinical Standards of Blood Pressure Measurement

The gold standard of clinical BP measurement is the arterial catheter, or A-line. The catheter is inserted into the subject's artery of interest (most often the radial artery in the wrist), and BP is continuously recorded via a pressure transducer. Because the measurement is taken directly from inside the artery, this is considered the most accurate way to measure blood pressure. The continuous nature of the measurement means that acute events leading to rapid BP changes can be identified

immediately, which is of particular importance in critical care units and operating rooms. Additionally, the A-line is able to measure the full shape of the blood pressure waveform, rather than just maximum and minimum values. This shape can provide significant clinical value beyond diagnosis of hypertension, such as diagnosis of arterial valve stiffening [6], measurement of cardiac efficiency [7], early identification of hypotensive crises [8], and assessment of vascular health [9].

Despite the value provided by A-lines, they are relatively rarely used due to a number of limitations. Inserting a needle through the muscular wall of an artery is significantly more difficult than the venous insertions typically used for blood draws. The process requires at least two personnel with special training, and even then it can take up to an hour to complete and may fail entirely up to 25% of the time due to muscle spasms in the artery [10]. Successful A-line placements can still cause significant discomfort for the patients, and they carry additional risks of infection, hemorrhage, and ischemia (clot formation) [11]. As a result, A-lines are typically only applied to a high-risk subset of patients in hospitals and are almost never used in outpatient settings.

In almost all other scenarios, BP measurements are performed using an inflatable cuff, or sphygmomanometer, placed on the upper arm over the brachial artery. To perform a measurement the cuff is first inflated to a pressure well above the subject's systolic (maximum) blood pressure, or SBP, which cuts off all blood flow. The pressure is then gradually reduced until blood flow is once again detectable, either manually using a stethoscope or automatically using vibration sensors in the cuff. The highest pressure at which blood flow is intermittently detected is recorded as SBP, and the highest pressure at which blood flow is continuously detected is recorded as diastolic (minimum) blood pressure, or DBP.

BP cuffs have gained widespread usage because they are easy to apply, noninvasive, inexpensive, and generate results within minutes. However, they also come with a variety of drawbacks. The primary limitation is that cuff measurements are intermittent; gradually increasing and decreasing cuff pressure takes significant time, and patients will generally not tolerate frequent re-measurements due to discomfort. This response time may be substantially too slow for patients at risk of rapid health deterioration, such as those in critical care units. Furthermore, these intermittent measurements only provide maximum and minimum values for BP, removing the diagnostic value of analyzing full BP waveform shapes. Finally, cuff use in practice often results in wide measurement variability due to a number of factors, including

cuff size, body positioning, drug and alcohol use, and cold exposure [12]. As determined in a recent meta-analysis by Dankel, Kang, Abe, *et al.* [13], this variation leads to BP cuff measurement errors of up to 15.5mmHg for DBP and 20.2mmHg for SBP when compared to A-line measurements. Patients and doctors would benefit significantly from a technology which could provide the continuous and accurate BP readings of an A-line but with convenience and non-invasiveness of an inflatable cuff.

## 2.2   Background: Prior Work in the Field

The gap in performance and utility left by the two available clinical standards for BP measurement has inspired a wide variety of approaches to the problem of non-invasive blood pressure measurement, or NIBPM. This section will briefly review the body of prior work and examine why none of these methods has achieved widespread clinical acceptance.

The most established method of NIBPM is vascular unloading, which uses an inflatable cuff placed over the subject's fingertip [14]. The cuff continuously measures the level of blood perfusion in the finger on each heartbeat based on infrared light transmission through the fingertip, and a control loop varies pressure in the inflatable finger cuff to keep this perfusion volume constant. The variable pressure required to maintain this constant volume is used to calculate a BP reading. Commercial devices based on vascular unloading have gained regulatory approval and seen some limited use in hospitals, but have not gained widespread acceptance.

The key limitation of vascular unloading is that it does not actually measure pressure in an artery of interest; instead it measures a proxy (in this case perfusion force in the capillary beds of the fingertip) and extrapolates from this value back to a clinically useful BP number. While arterial BP is correlated to this proxy variable, there is not a 1:1 physical relationship between the two, or even an empirical relationship that is consistent across all subjects. Instead, the proxy-BP relationship is generally calibrated for each patient individually based on a standard inflatable BP cuff, and even then periodic re-calibration is often required to correct for drift and can result in data blackouts [15]. These calibration steps and the uncertain proxy-BP relationship create significant potential sources of error, and studies have frequently found errors between finger cuffs and A-lines of over 25 mmHg [16], [17]. Futhermore, commercially available vascular unloading devices typically cost

upwards of $30,000, making them infeasible for widespread deployment beyond well-funded hospitals.

Dozens of research groups have pursued NIBPM techniques using a variety of other physical signals, including (but not limited to) light reflectance from blood (photoplethysmography) [18]–[27], pressure signals at the surface of the skin (tonometry) [28]–[30], electrical conductivity of blood (bioimpedance) [31]–[34], pressure wave velocity (pulse transit time) [21], [33], [35]–[37], reactive forces from cardiac ejection (ballistocardiography) [38]–[40], and millimeter-wave radar for waveform measurement [41]. A more detailed summary of these methods is provided in Appendix Table A1.1. Despite the wide variation in these techniques, they all share the same common limitation as vascular unloading; they are only able to measure a proxy variable that is correlated with, but not directly mapped to, true blood pressure.

Most of the above methods establish the proxy-BP relationship via calibration to an inflatable cuff. However, even once calibration is performed, changes in subject BP may not always lead to consistent changes in proxies. For example, Avolio *et al.* [42] found that BP increases due to cycling stress tests significantly increased pulse transit time, but similar BP increases from ice water immersion or hand grip stress tests did not. Thus, these devices suffer from a compounding of errors due to inherent measurement noise, inaccuracy in the cuff calibration reference, and drift in calibration validity over time. Some methods have attempted to remove the need for direct calibration by applying machine learning models which predict BP from the measured proxies and other factors like subject demographics and heart rate [19], [22]–[25], [39]. However, the accuracy of these empirically trained models is limited by the diversity of training data, which could be especially problematic for patients with extreme BPs or uncommon pathologies [43]. Such accuracy concerns have prevented these various NIBPM devices from gaining significant clinical acceptance.

**Ultrasound-Based Approaches**

In addition to the above work, some research groups have approached NIBPM using ultrasound as a measurement modality [44]–[48]. A significant advantage of ultrasound is that it is able to directly measure properties such as arterial radius, distention, and blood velocity which can only be assessed in relative terms (if at all) by other techniques. Furthermore, while traditional ultrasound devices are bulky and require manual operation, recent advances have led to the design of flexbile,

wearable ultrasound patches that can measure arterial characteristics [47], [49], [50]. By combining measurements of arterial cross-sectional area ($A$) and volumetric blood flow rate ($Q$) over time, it is possible to measure changes in pressure via the "QA" equation: [46]

$$\frac{dP}{dt} = \rho_{blood} \left( \frac{dQ}{dA} \right)^2 A \frac{dA}{dt}.$$

(2.1)

This equation represents a direct physical relation between measured parameters and blood pressure, in contrast to all of the correlated proxies described above. However, it is still only able to measure *relative* changes in pressure over time; the *absolute* baseline pressure must still be determined by calibration to a cuff. Furthermore, since only $dP/dt$ is measured, small errors in each measurement can quickly add up and cause significant drift in inferred $P$ if re-calibration is not performed. As a result, the QA method has not yet seen clinical utility.

An alternative ultrasound-based approach which does attempt to extract absolute BP is described by Zakrzewski, Huang, Zubajlo, *et al.* [51]. With this method, a force gauge is attached to a standard ultrasound probe, and the user images their carotid artery while pressing the probe into their neck with a gradually increasing force over time. The observed distention of the artery and surrounding tissue is then matched up with a pre-computed library of finite element models to extract DBP and SBP. While this measurement technique does have a connection to absolute BP through physics-based modeling, the actual data acquisition method is practically infeasible in a clinical setting, and atypical patients whose anatomies do not match up with the pre-computed model library could cause problems.

## 2.3 Arterial Resonance: A First-Order Approach

The focus this project was developing a model for continuous, absolute BP measurement which calculates pressure directly from observables rather than relying on correlated proxies or trained models. Both clinical standards for BP measurement rely on reference to a known stimulus; the BP cuff measures the collapse of the artery in response to a known external compression, and the arterial catheter measures the reaction of a transducer with a known response function to the pressure exerted by the blood itself. In contrast, almost none of the NIBPM methods described above incorporate a known external stimulus. Instead, they rely on the subject's own heartbeat to perturb the artery and perform passive observation to measure the artery's response. The central problem with such methods is that the magnitude

of pressure exerted by the heartbeat is unknown, so the system is left undercon-strained; we know how it responded but not what exactly it was responding to. As a result, such methods can measure how blood pressure changes relatively over time, but they cannot establish an absolute pressure baseline. It is telling that the only NIBPM method described above which does involve external stimulus application (Zakrzewski, Huang, Zubajlo, *et al.* [51]) is also the only one which can measure baseline absolute BP without inferring it via machine learning, albeit after the application of an empirically tuned tissue model. This suggests that stimulus application is the key missing ingredient in the development of calibration-free NIBPM.

Our novel approach to the NIBPM problem is Controlled Arterial Resonance for Direct Identification of Blood Pressure, or CARDI-BP. The inspiration for this method is the guitar string. If one simply walks up to a guitar that has already been strung, there is no way to tell how much tension any given string is under just by looking at it. However, the tension can be measured by plucking the guitar string and measuring the frequency $f$ at which it resonates. Specifically, because the string is an elastic system its tension $T$ is directly related to its wave velocity $v_w$ and linear mass density $\gamma$ as $T = \gamma v_w^2$. Because the string is fixed at both ends its baseline wavelength will be $\lambda = 2L$, and the wave equation dictates that $v_w = \lambda f$. Putting this all together gives us $T = 4\gamma L^2 f^2$; by measuring the string's density, length, and resonance frequency we can determine the *absolute* amount of tension it is under.

The key to extending this logic to blood vessels is Laplace's Law, as illustrated in Figure 2.1a. This states that for an elastic cylindrical vessel of radius $a$, the absolute internal fluid pressure $P$ and wall tension $T$ are related as

$$P = T/a. \tag{2.2}$$

Passive observation is not capable of measuring an absolute value for $T$ any more than it can measure $P$. However, if we apply an external stimulus and resonate the artery, we can apply guitar string physics to convert resonant frequency to tension and thereby measure pressure. Because the arterial wall is a closed loop, its lowest-order vibrational mode will contain two full wavelengths around its circumference, as illustrated in Figure 2.1b. Substituting $\lambda = \pi a$ and $\gamma = \rho h$ for wall volumetric density $\rho$ and thickness $h$, we get

$$T = \pi^2 \rho h a^2 f^2 \tag{2.3}$$

$$P = \pi^2 \rho h a f^2. \tag{2.4}$$

This equation provides the core of the CARDI-BP method; if we can stimulate the artery to measure its resonant frequency while simultaneously measuring other properties like dimensions and density, we can obtain an absolute measurement of blood pressure without any external calibration or reference.
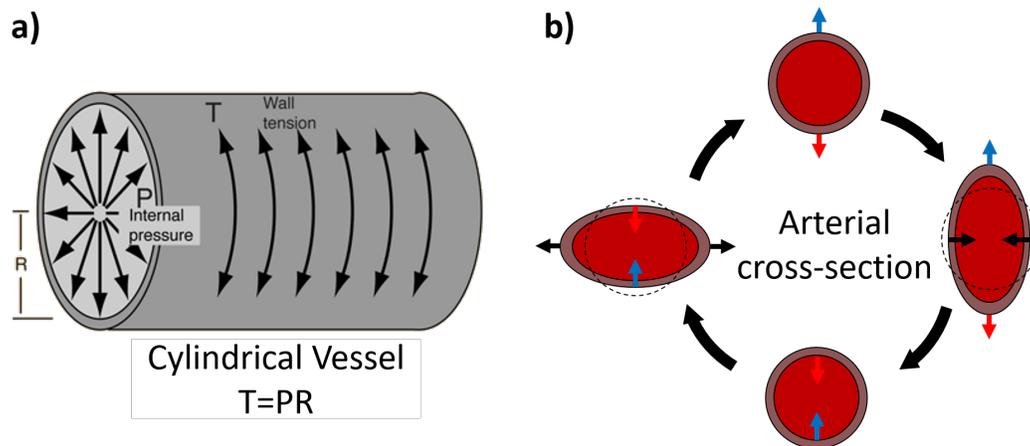


Figure 2.1: Physical concepts underlying arterial resonance. a) An illustration of Laplace's Law, which relates the fluid pressure inside a cylinder ($P$) to the resultant circumferential tension in the wall needed to contain that pressure ($T$). Because blood pressure places the arterial wall under tension, we expect the wall to exhibit resonant behavior. b) The lowest order resonant mode which can be excited around the circumference of an artery, with the scale of deformation exaggerated. Colored arrows represent velocities which would be detectable via Doppler ultrasound if the probe were placed at the top of the page.

## 2.4 Arterial Resonance: A Detailed Approach

The result in Equation 2.4 is illustrative of the system's general behavior, but it is far too simple to be of practical use. Accurate BP measurement requires a more detailed analysis of the system's behavior. The physical model underlying CARDI-BP draws from two disparate lines of analysis: one from aerospace engineering and the other from biomechanics. The first set of analyses deals with vibrational resonance modes in thin-walled cylindrical shells for large-scale industrial applications such as fuel tanks and pipelines [52]. The second set of analyses examines the dynamics of *in vivo* arterial walls by modeling them as long, thin-walled cylindrical shells and using structural and fluid mechanics to calculate how these shells respond to changes in pressure [53], [54]. While these two lines of analysis share fundamental commonalities and assumptions, no work has combined them to create a model of the resonant modes in pressurized arteries. Furthermore, all of these analyses focused on deriving expected responses based on a known applied pressure. We

show that by combining and inverting these relationships, it is possible to calculate *in vivo* arterial blood pressure from measurements of its resonant response.

In order to extend the analysis of Arnold and Warburton [52] to the *in vivo* context, their model must be modified using additional mechanical analyses to account for various physical complexities which had previously only been analyzed in isolation. These include the presence of a pressurized fluid inside the artery [55] and inertial damping due to fluid mass inside and outside the artery [56], [57]. Furthermore, we integrate established biomechanical analysis to account for effects such as the significant distention of the artery as pressure changes and the nonlinear character of its elasticity [53], [54].

**Assumptions**

We begin by making some simplifying assumptions to make analysis of this system tractable:

1. The artery is a cylindrical shell with length $L$ and a constant radius $a$, thickness $h$, density $\rho_S$, Poisson's ratio $\nu$, and circumferential Young's modulus $E$ along its entire length.

2. The artery is long and thin-walled, i.e., $L \gg a$ and $a \gg h$.

3. The artery is surrounded by an incompressible fluid of constant density $\rho_L$ both inside and out.

4. All circumferential tension in the arterial wall is from either the internal blood pressure or the wall's inherent stiffness; no additional tensile or shearing forces are exerted by, e.g., attached tendons or muscle fibers.

5. All induced vibrations can be treated as infinitesimal displacements, with the associated wall displacements much less than $a$. Damping effects due to viscosity of the internal or external media or viscoelasticity of the wall itself can be modeled as a linear effect for the range of displacements induced by the stimulus.

6. The circumferential Young's Modulus of the arterial wall ($E$) behaves in a linearly elastic manner in response to the small radius perturbations induced by the stimulus. Such an assumption does not preclude changes in $E$ over the course of a cardiac cycle; it only asserts that changes in radius induced by

the stimulus are small enough compared to heartbeat-induced radius variation such that the vibrations do not change $E$.

These assumptions invalidate the model in some circumstances, such as when an artery branches or turns sharply. However, they should hold reasonably true for major arteries far from branch points and joints, such as the carotid artery in the neck and the brachial artery in the bicep.

**Mechanics of Cylindrical Shell Resonance**

A cylindrical shell will support many natural modes of wall motion, each composed of a superposition of an axial component consisting of $m/2$ wavelengths along the length of the cylinder and a circumferential component consisting of $n$ wavelengths around the circumference of the cylinder, where $m$ is an integer equal to or greater than 1 and $n$ is an integer greater than 1. In cylindrical coordinates of axial location $z$ and azimuthal angle $\theta$, the radial displacement $w$ of each point on the surface at any given time $t$ can be expressed as a superposition of sinusoidal basis functions given by

$$w(z, \theta, t) = \sum_{m,n} A_{mn} \sin \frac{m\pi z}{L} \cos n\theta \cos 2\pi f t \tag{2.5}$$

for some scalar amplitude $A_{mn}$. The general solutions for the equations of motion of this system are quite complex for arbitrary system and depend on both circumferential wavenumber $n$ and axial wavenumber $\lambda_m = m\pi a/L$. In a system with $L \gg a$, however, we have $n \gg \lambda_m$, so the contributions of the axial modes are greatly suppressed compared to the circumferential ones. Neglecting terms proportional to $\lambda_m$, the resonant frequencies take the form of roots of a cubic polynomial: [55]

$$0 = \kappa^3 - K_2\kappa^2 + K_1\kappa - K_0 \tag{2.6}$$

$$\kappa = \frac{4\pi^2 \rho_s a^2 (1 - \nu^2)}{E} f^2 \tag{2.7}$$

$$K_0 = \frac{h^2(1 - \nu)}{24a^2} \left( n^8 - 2n^6 + n^4 \right) + \frac{Pa}{Eh}\alpha_1 \tag{2.8}$$

$$K_1 = \frac{1 - \nu}{2} \left( n^4 + n^2 \right) + \frac{h^2}{12a^2}\alpha_2 + \frac{Pa}{Eh}\alpha_3 \tag{2.9}$$

$$K_2 = 1 + \frac{3 - \nu}{2}n^2 + \frac{h^2}{12a^2} \left( n^4 + n^2 \right) + \frac{n^2}{1 - \nu^2}\frac{Pa}{Eh} \tag{2.10}$$

$$\alpha_1 = \frac{1-\nu}{2}\left[\left(1 - \frac{h^2}{12a^2}\right)n^6 - n^4\right] \tag{2.11}$$

$$\alpha_2 = \frac{3-\nu}{2}n^6 - \frac{3+\nu}{2}n^4 + n^2 \tag{2.12}$$

$$\alpha_3 = \left(\frac{3-\nu}{2} - \frac{h^2}{12a^2}\right)n^4 - n^2 \tag{2.13}$$

where $\kappa$, $K_i$, and $\alpha_i$ are dimensionless parameters. In a damped system the lowest frequency resonant mode is generally the easiest to excite as first-order damping forces will increase with frequency for a given magnitude of displacement; thus, we focus our attention on the $n = 2$ mode. Finding the smallest real root of Equation 2.6 and converting from $\kappa$ back to $f$ with $n = 2$ yields

$$f_{vac}^2 = \frac{C_0 - \sqrt{C_0^2 - C_1}}{24\pi^2(1-\nu^2)ha^4\rho_S} \tag{2.14}$$

$$C_0 = 5Eh(3a^2 + h^2) + 12a^3P \tag{2.15}$$

$$C_1 = 12Ea^2h(36a^3P - 4ah^2P + 9Eh^3). \tag{2.16}$$

**Inertial Contribution of the Surrounding Medium**

The above resonant frequency is written as $f_{vac}$ because the analysis of Fung, Sechler, and Kaplan [55] is only valid when the shell is in an environment where the mass of the surrounding medium can be neglected, so all inertial contributions come from the shell itself. This makes sense for metal pipelines or tanks containing pressurized gasses, but it is clearly invalid for arteries where the surrounding blood and tissue have a similar density to the arterial wall itself. To account for this, we incorporate the work of Lindholm, Kana, and Abramson [56] and Warburton [57]. These works show that the unstable pressure exerted by the internal and external fluids at the fluid-wall boundary each affect resonant behavior by adding an effective inertial mass term:

$$m_{L,int} = a\rho_L \frac{I_n(\lambda_m)}{\lambda_m I_n'(\lambda_m)} \tag{2.17}$$

$$m_{L,ext} = a\rho_L \frac{K_n(\lambda_m)}{\lambda_m K_n'(\lambda_m)} \tag{2.18}$$

where $I_n$ and $K_n$ are modified Bessel functions of the first and second kind and $I_n'$, $K_n'$ are their derivatives. In general we know that the acceleration of a system under a given force is inversely proportional to its mass, and taking the second derivative of

Equation 2.5 we can see that $\partial^2 w/\partial t^2 \propto f^2$. Thus, the addition of effective mass from the surrounding medium will scale down our result from Equation 2.14:

$$\frac{f_{vac}^2}{f^2} = \frac{\rho_S h + m_{L,int} + m_{L,ext}}{\rho_S h} \tag{2.19}$$

$$= 1 + \frac{\rho_L}{\rho_S}\frac{a}{h}\left[\frac{I_n(\lambda_m)}{\lambda_m I_n'(\lambda_m)} + \frac{K_n(\lambda_m)}{\lambda_m K_n'(\lambda_m)}\right]. \tag{2.20}$$

Applying the previous assumption of $L \gg a$ (and thus $\lambda_m = m\pi a/L \approx 0$) and plugging in $n = 2$, we can take the limit of these Bessel function ratios as $\lambda_m \to 0$, yielding

$$\frac{f_{vac}^2}{f^2} = 1 + \frac{2n}{n^2 + 1}\frac{\rho_L}{\rho_S}\frac{a}{h} \tag{2.21}$$

$$= 1 + \frac{4}{5}\frac{\rho_L}{\rho_S}\frac{a}{h}. \tag{2.22}$$

**Solving for Pressure**

We now have everything we need to solve for pressure. The proper resonant frequency, accounting for inertial mass, can be written as

$$f^2 = \frac{C_0 - \sqrt{C_0^2 - C_1}}{24\pi^2(1 - v^2)a^3\delta} \tag{2.23}$$

$$\delta = ah\rho_S + \frac{4}{5}a^2\rho_L. \tag{2.24}$$

Inverting this equation to solve for pressure yields

$$\boxed{P = \frac{9\alpha^4 - 5\left(3\alpha + \alpha^3\right)D + 3D^2}{-4(9\alpha - \alpha^3) + 12D}E} \tag{2.25}$$

$$\alpha = h/a \tag{2.26}$$

$$D = 4\pi^2(1 - v^2)\frac{\rho a^2 f^2}{E} \tag{2.27}$$

$$\rho = \alpha\rho_S + \frac{4}{5}\rho_L \tag{2.28}$$

where $\alpha$ and $D$ are dimensionless parameters and $\rho$ has units of volumetric mass density. Following Fung, Sechler, and Kaplan [55], Equation 2.6 can be simplified

by neglecting tangential inertial forces in the cylinder, yielding a linear relationship between $\kappa$ and $P$:

$$\kappa \approx \frac{h^2}{12a^2}n^4 + \frac{Pa}{Eh}n^2(1-v^2) \tag{2.29}$$

$$f_{vac}^2 \approx \frac{h^2}{3\pi^2\rho_S(1-v^2)a^4}E + \frac{1}{\pi^2\rho_S a h}P \tag{2.30}$$

$$f^2 \approx \frac{h^3}{3\pi^2\rho(1-v^2)a^5}E + \frac{1}{\pi^2\rho a^2}P \tag{2.31}$$

$$\boxed{P \approx \pi^2\rho a^2 f^2 - \frac{h^3}{3(1-v^2)a^3}E} . \tag{2.32}$$

This linearization exactly recovers our original behavior from Equation 2.4 if $\rho_L = E = 0$, reflecting the fact that this original simplification ignored the inertial mass contribution of the surrounding medium and tensile forces due to the material's inherent stiffness. Equations 2.25 and 2.32 are the key pressure calculation formulae underlying CARDI-BP and will be used throughout the rest of this work, depending on whether working with the full or linearized equation is more tractable.

## 2.5   Parameter Measurement

Using either of our equations for pressure requires determining the values of seven parameters:

1. $f$: Arterial resonant frequency

2. $a$: Arterial radius

3. $h$: Arterial wall thickness

4. $\rho_S$: Density of the arterial wall

5. $\rho_L$: Density of the fluid surrounding the arterial wall

6. $v$: Poisson's ratio of the arterial wall

7. $E$: Young's modulus of the arterial wall.

These can be broken into three categories; those that are directly measured (1-3), those whose values can be assumed constant (4-6), and those which must be inferred based on arterial dynamics (7).

**Directly Measured Parameters**

In order to embody CARDI-BP into a device, we must pick a measurement modality which is able to simultaneously measure arterial radius, thickness, and resonant frequency. The obvious choice for measuring arterial dimensions is ultrasound. While ultrasound is less precise than methods like computational tomography and magnetic resonance imaging, it is far less expensive, portable, and free from harmful radiation. For these reasons, clinical studies which assess arterial dimensions primarily rely on ultrasound to make these measurements, even for small features like wall thickness [58]–[60].

Selecting a method of measuring arterial resonant frequency is less obvious, as such measurements have never been performed. Fortunately, previous work in related areas like shear wave elastography (SWE) has shown that ultrasound is capable of detecting very small displacements in arterial walls generated by an audio stimulus [2], [61]. If we stimulate the artery at various frequencies and measure its response, the resulting magnitude and phase curves should show characteristics of resonance, i.e., a peak in magnitude and sigmoidal transition in phase. Thus, using ultrasound as our sole measurement modality should allow us to simultaneously assess radius, thickness, and resonant frequency. An embodiment of this device is illustrated below in Figure 2.2, which consists of two separate components; a speaker to stimulate arterial resonance and an ultrasound wand to measure both the resonant response and arterial dimensions. Further details on exactly how the device and measurements are implemented are provided in Chapters 3 and 4.

**Assumed Parameters**

Ultrasound cannot be used to directly measure material density or Poisson's ratio. Fortunately, these parameters can be assumed to hold constant across individuals. The IT'IS database of tissue properties shows that arterial walls and blood have fairly stable densities of 1102 kg/m$^3$ and 1050 kg/m$^3$, respectively [62]. The density of the tissue surrounding the artery is a bit less well-defined, as it depends on the artery's location and surrounding physiology. However, most soft tissue types in the IT'IS database have densities between 1000 and 1100 kg/m$^3$, so we assume an average density of 1050 kg/m$^3$ for the surrounding tissue as well. Finally, prior studies have shown that the arterial wall is very nearly incompressible [63]–[65], so we can assume its Poisson's ratio will be $\nu = 0.5$.
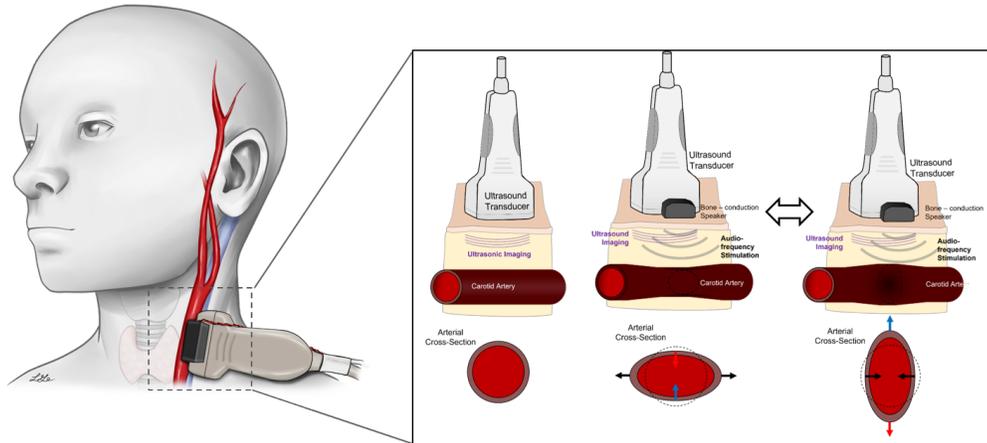
Figure 2.2: An illustration of a device embodiment which uses the CARDI-BP method to measure blood pressure in the carotid artery. The device consists of two attached components; an audio-frequency speaker which stimulates arterial resonance, and an ultrasound probe which measures both the resonant response and arterial dimensions. While this version of the device requires handheld operation, future versions could incorporate the speaker and ultrasound probe into a wearable form factor.

**Young's Modulus Measurement**

The only remaining parameter to be measured is Young's modulus, which is a measure of the stiffness of the arterial wall. The value of this stiffness can vary substantially between individuals due to differences age, pathologies, and vascular muscle tone, so assuming a constant value is not feasible. Furthermore, within a given artery the Young's modulus changes significantly over the course of a heartbeat, with stiffness starting at a minimum in diastole and rapidly increasing during systole [66], [67]. Thus, we need a way to measure stiffness continuously over the course of a heartbeat. Prior studies have used ultrasound to conduct this measurement based on a combination of the Moens-Kortweg [53] (first term) and Bramwell-Hill [54] (last term) equations, which are related but independent measures of pulse wave velocity down the length of the artery: [44]

$$c = \sqrt{\frac{Eh}{2\rho_L a}} = \sqrt{\frac{A}{\rho_L}\frac{dP}{dA}} \qquad (2.33)$$

where $c$ is pulse wave velocity and $A$ is arterial cross-sectional area. Rearranging the latter two terms and substituting $A = \pi a^2$ yields a useful equation for calculating Young's modulus based on changes in pressure:

$$\frac{Eh}{2a} = \frac{\pi a^2}{2\pi a}\frac{dP}{da} \tag{2.34}$$

$$E = \frac{a^2}{h}\frac{dP}{da}. \tag{2.35}$$

Prior studies which utilized this relationship measured $a$ and $h$ with ultrasound, but relied on an external reference such as a cuff or tonometer to measure pressure. We replace this external reference with pressure measurements generated by Equation 2.25. This creates a recursive relationship, as these pressure measurements are themselves dependent on the value of $E$ I'm measuring. This interdependency can be resolved using the Gauss-Seidel method. First, a physiologically reasonable value for $E$ is chosen as a starting point, and $P$ is calculated at all radii based on this value using Equation 2.25. These $P$ values are then used to calculate $E$ using Equation 2.35. By repeating these two steps, both $P$ and $E$ converge on a self-consistent set of values that satisfy both Equations 2.25 and 2.35. Importantly, this method does not require $E$ to be constant at different times. Instead, it provides instantaneous $E$ estimates at the same rate that pressure and radius measurements are generated. A more detailed investigation of the convergence of this method and its results is the focus of Chapter 5.

## 2.6   Feasibility and Sensitivity Analysis

We have now established that it should be possible, in theory, to generate blood pressures based on the measurement of arterial resonance. However, this does not guarantee that such measurements will be practically obtainable in real arteries. This section is focused on investigating these practicalities.

### Expected Resonance Frequencies

The first question to answer is what frequencies real arteries should be expected to resonate at. The parameters of our system dictate a "sweet spot" of frequencies which can reasonably be measured. This sweet spot is bounded on the low end by heart rate, which can range up to ~4Hz in humans. If the resonance is not significantly faster than this heart rate we will get at best an averaged resonance behavior that cannot distinguish systole from diastole. On the high end, the sweet spot is bounded by the rate of measurements attainable with ultrasound. Standard ultrasound techniques for measuring target velocity have a minimum period dictated

by the round-trip time for sound to travel from the probe to the target and back. For artery depths up to 4cm under the surface of the skin, the fastest possible measurement rate is ~20kHz; if the resonance is not significantly slower than this, we will run into undersampling problems.

Table 2.1 shows expected resonance frequencies for three different arteries where BP measurement would be of clinical interest; the carotid artery (in the neck), the brachial artery (in the bicep), and the radial artery (in the wrist). Each artery is analyzed under low-frequency (i.e., low pressure, high radius) and high-frequency (i.e., high pressure, low radius) conditions to get an idea for the range of expected resonance frequencies which may be measured in the artery. Listed parameters are taken from prior literature to represent the range of what might commonly be seen in adult patients under non-extreme conditions [68]–[70]. All of these frequencies fall in the range of 100-1500 Hz, and fortunately this lies squarely within our sweet spot. This indicates that it should be possible to measure arterial resonance with ultrasound and that this resonance should vary quickly enough to track BP changes over the course of a heartbeat.

| Artery | Predicted Resonant Frequency (Hz) | Pressure (mmHg) | Radius (mm) | Thickness (mm) | Young's Modulus (MPa) |
|---|---|---|---|---|---|
| Carotid (Low $f$) | 122 | 40 | 5 | 0.5 | 0.1 |
| Carotid (High $f$) | 459 | 180 | 3 | 0.3 | 1 |
| Brachial (Low $f$) | 245 | 40 | 2.5 | 0.25 | 0.1 |
| Brachial (High $f$) | 918 | 180 | 1.5 | 0.15 | 1 |
| Radial (Low $f$) | 382 | 40 | 1.6 | 0.16 | 0.1 |
| Radial (High $f$) | 1378 | 180 | 1.0 | 0.1 | 1 |

Table 2.1: Expected resonance frequencies for three different arteries that would be likely clinical targets for measurement. Each artery is analyzed in a low-frequency state (i.e., low pressure, high radius) and a high-frequency state (i.e., high pressure, low radius). The predicted range of frequencies falls significantly above human heart rate (~4 Hz) and significantly below the sampling limit of ultrasound (~20 kHz), making CARDI-BP potentially viable on any of these three arteries.

**Sensitivity Analysis**

Another important practical consideration is the sensitivity of the final pressure output to its various input parameters. We begin with the standard formula for propagation of error under the assumption that errors in various parameters are independent and uncorrelated, where $\sigma_i$ represents the standard deviation of parameter $i$ and $|_0$ denotes the measurement reference state:

$$\sigma_P^2 \approx \left(\frac{\partial P}{\partial f}\bigg|_0 \sigma_f\right)^2 + \left(\frac{\partial P}{\partial a}\bigg|_0 \sigma_a\right)^2 + \left(\frac{\partial P}{\partial h}\bigg|_0 \sigma_h\right)^2 + \left(\frac{\partial P}{\partial E}\bigg|_0 \sigma_E\right)^2$$
$$+ \left(\frac{\partial P}{\partial \rho_S}\bigg|_0 \sigma_{\rho_S}\right)^2 + \left(\frac{\partial P}{\partial \rho_L}\bigg|_0 \sigma_{\rho_L}\right)^2 + \left(\frac{\partial P}{\partial v}\bigg|_0 \sigma_v\right)^2 . \quad (2.36)$$

In order to make the derivatives tractable to gain intuition for system behavior, we start with the linearized equation 2.32, yielding

$$\left(\frac{\partial P}{\partial f}\bigg|_0\right)^2 = \frac{P_0^2}{f_0^2}\left[4 + O\left(\alpha^3\right)\right] \tag{2.37}$$

$$\left(\frac{\partial P}{\partial a}\bigg|_0\right)^2 = \frac{P_0^2}{a_0^2}\left[4 + O\left(\alpha\right)\right] \tag{2.38}$$

$$\left(\frac{\partial P}{\partial h}\bigg|_0\right)^2 = \frac{P_0^2}{h_0^2}\left[\frac{25\rho_{S0}^2}{16\rho_{L0}^2}\alpha^2 + O\left(\alpha^3\right)\right] \tag{2.39}$$

$$\left(\frac{\partial P}{\partial E}\bigg|_0\right)^2 = \frac{P_0^2}{E_0^2}\left[\frac{25E_0^2}{144\pi^4(1-v_0^2)^2\rho_{L0}^2 a_0^4 f_0^4}\alpha^6 + O\left(\alpha^7\right)\right] \tag{2.40}$$

$$\left(\frac{\partial P}{\partial \rho_S}\bigg|_0\right)^2 = \frac{P_0^2}{\rho_{S0}^2}\left[\frac{25\rho_{S0}^2}{16\rho_{L0}^2}\alpha^2 + O\left(\alpha^3\right)\right] \tag{2.41}$$

$$\left(\frac{\partial P}{\partial \rho_L}\bigg|_0\right)^2 = \frac{P_0^2}{\rho_{L0}^2}\left[1 + O\left(\alpha\right)\right] \tag{2.42}$$

$$\left(\frac{\partial P}{\partial v}\bigg|_0\right)^2 = \frac{P_0^2}{v_0^2}\left[\frac{25v_0^4 E_0^2}{36\pi^4(1-v_0^2)^2\rho_{L0}^2 a_0^4 f_0^4}\alpha^6 + O\left(\alpha^7\right)\right] . \tag{2.43}$$

Note that we have expanded each error term about small values of the parameter $\alpha = h/a$ to give intuition about the relative scale of the uncertainties, since $\alpha$ is assumed to be a relatively small parameter for our system. The terms $f$, $a$, and $\rho_L$ all have order-unity leading terms, indicating that they will be relatively important in contributing error. In contrast, the terms $h$, $E$, $\rho_S$, and $v$ have order-$\alpha^2$ or higher leading terms, indicating that they will be relatively unimportant in the total error budget.

We can substitute representative values into the above equation to quantify how much each term contributes to the overall error budget. Assigning values to each parameter corresponding to a typical carotid artery ($a_0 = 4$ mm, $f_0 = 270$ Hz, $h_0 = 0.6$ mm, $E_0 = 0.385$ MPa, $\rho_S = 1102$ kg/m$^3$, $\rho_L = 1050$ kg/m$^3$, and $v = 0.5$) gives:

$$\frac{\sigma_P}{P_0} \approx \left[ 4.4 \left( \frac{\sigma_f}{f_0} \right)^2 + 4.4 \left( \frac{\sigma_a}{a_0} \right)^2 + 0.00024 \left( \frac{\sigma_h}{h_0} \right)^2 + 0.0028 \left( \frac{\sigma_E}{E_0} \right)^2 \right.$$
$$\left. +0.030 \left( \frac{\sigma_{\rho S}}{\rho_{S0}} \right)^2 + 0.77 \left( \frac{\sigma_{\rho L}}{\rho_{L0}} \right)^2 + 0.0012 \left( \frac{\sigma_v}{v_0} \right)^2 \right]^{1/2}. \quad (2.44)$$

The numerical coefficients in Equation 2.44 are broadly in line with expectations from leading-order scaling in Equations 2.37 to 2.43, with error terms for $f$, $a$, and $\rho_L$ over 10x more prevalent in the final error budget than error terms for $h$, $E$, $\rho_S$, and $v$.

**Variability of Measured Parameters**

Equations 2.37 to 2.40 indicate that errors in thickness and stiffness will propagate very weakly into our final pressure estimate, but we will be significantly sensitive to errors in resonant frequency and radius. For either of these parameters, a measurement error of 5% (corresponding to 13.5Hz in frequency or 200$\mu$m in radius in the above example) would lead to over a 10% error in pressure, which could be problematic in a clinical setting. As we shall see in Chapter 4, it is practical to measure arterial stimulus response in 20Hz increments, which is not enough precision on its own to yield acceptable accuracy in pressure; getting past this precision limit requires the use of curve fitting techniques. For radius, running ultrasound with pulses at 5MHz (roughly the value used in this work) can yield raw spatial resolution of up to 300$\mu$m when analyzing a single return channel; once again, this is insufficiently precise on its own. Attaining higher resolution requires dynamic focusing of the image using multiple channels of return information as well as image analysis and denoising processes. The techniques we implemented for resonant curve fitting and image analysis will be described in Chapter 3.

**Variability of Assumed Parameters**

Equations 2.41 to 2.43 indicate that error in arterial wall density and Poisson's ratio will propagate very weakly into our final pressure estimate, but we will be sensitive

to errors in fluid density. Since this parameter is given an assumed value, we need to obtain an estimate for what the underlying variance around this assumption might be.

The IT'IS database lists the density of blood as $1050 \pm 17$ kg/m$^3$, so it may be expected to vary by about 1.6%. However, blood only constitutes half of the total inertial mass contribution from the artery's surroundings; the other half comes from the surrounding tissue. This component is harder to quantify, since we do not know exactly what tissue the artery will be embedded in. Most types of soft and connective tissue in the IT'IS database have densities of roughly 1050 kg/m$^3$ as well, which is why we use blood density to represent the total surrounding medium. If we are 5% off in our estimate of external medium density (i.e., the tissue in which the artery is embedded has a density close to 1000 or 1100 kg/m$^3$, towards the extreme of soft tissue densities in IT'IS), this would translate to a roughly 2% error in the final pressure estimate based on equation 2.44. An outlier in soft tissue density is fat deposits, which have a density of only 911 kg/m$^3$. Central arteries such as the carotid are generally not surrounded by fat deposits except in mordibly obese patients with poor muscle tone; for such patients, a special correction to the density term may be necessary.

Even though our pressure equation is not directly very sensitive to Poisson's ratio, at multiple points later in this research we will assume arterial incompressibility (i.e., $\nu = 0.5$), in line with common arterial models in the literature [71]. Thus, it is relevant here to consider how robustly this assumption can be expected to hold. Multiple groups have compressed *ex vivo* arterial tissue and measured its volumetric strain $\Delta V / V$, finding values between 0.2% and 2% [63]–[65]. This implies true values for Poisson's ratio between 0.49 and 0.499 and indicates that our assumption of arterial incompressibility is a safe one.

A final assumed parameter to consider is the speed of sound. While this does not appear explicitly in our pressure equations, it is implicitly involved in the calculation of arterial dimensions. In particular, arterial radius is calculated by taking the time delay in echoes between the top and bottom arterial walls and converting this travel time into a distance using the speed of sound in blood. Ultrasound processing techniques generally assume a constant sound speed of 1540 m/s, regardless of tissue type [72]. This assumption of a constant speed is necessary for generating coherent images without *a priori* knowledge of the spatial distribution of tissue types that will be imaged. However, the speed of sound in blood has been measured as

1578 ± 11 m/s [62]. To account for this, during *in vivo* data analysis all B-mode images were generated using the standard sound speed of 1540 m/s, but all calculated radii were scaled up by a factor of $38/1540 \approx 2.5\%$ to account for the speed of sound difference.

*Chapter 3*

# *IN VITRO* TESTING

The physics model we have derived for the CARDI-BP method has not previously been described or tested. Thus, before applying the model to human arteries it was important to test it on an idealized system which matched the assumptions of Section 2.4 as closely as possible. This chapter will describe the measurement device and data analysis procedures which were used to perform CARDI-BP measurements, as well as the setup and results of our idealized *in vitro* testing.

## 3.1 Measurement Device

The measurements we need to perform in order to obtain pressure can be done with standard ultrasound acquisition modes, namely B-mode (for arterial dimensions) and Doppler (for resonant response). However, performing analysis requires both access to raw data from ultrasound returns and precise synchronization between the ultrasound pulses and our audio stimulus source; neither of these is possible with standard commercial ultrasound devices. Instead, we constructed our own ultrasound system consisting of an off-the-shelf Acuson 6L3 ultrasound imaging probe (Acuson, Mountain View, CA) with a custom backend.

Interfacing between a control computer and the backend is performed via two software-defined radios (USRP N210 from National Instruments, Austin, TX). The first powered the ultrasound transmit and receive chains. The transmit chain used a bipolar high-voltage pulse generator (MAX4940 from Maxim Electronics, San Jose, CA) to generate ultrasound pulses from one of 16 piezoelectric elements on the probe, selected via a multiplexer. Returns from these elements were routed to a low-noise amplifier (AD8336, Analog Devices, Wilmington, MA) via an automatic transmit/receive switch (MD0100 from Microchip, Chandler, AZ) and sampled by the N210's analog-digital converter, which operated at 25 MHz. The receive chain was linked to an additional 16 elements via a separate amplifier for a total of 32 receive elements, of which 2 could be recorded simultaneously. Raw returns from the receive chain were recorded for both real-time and post-processing.

The additional component needed to enable our resonance measurements was an audio stimulus driver, which was controlled by the second software-defined radio.

In principle, we could have used high-intensity focused ultrasound (HIFU) to drive arterial displacements using the ultrasound probe[61]; however, this would have required more complicated engineering in our transmit chain and would have also raised potential safety concerns[73]. Fortunately, the resonant frequencies we expected to drive were in the range of  100-1500 Hz (see section 2.6), which is well within the range of commodity speakers. Standard speakers are designed to broadcast sound into air, which has an acoustic impedance of roughly 400 kg*m$^{-2}$*s$^{-1}$, while skin has an acoustic impedance of roughly $1.6 \times 10^6$ kg*m$^{-2}$*s$^{-1}$ [62]. As a result, audio energy broadcast from standard speakers is almost entirely reflected at the air-skin boundary and very little stimulus reaches the artery. Instead, we generated our audio stimulus using moving-coil drivers (BC-10 from Ortofon, Nakskov, Denmark), which are designed to transmit acoustic energy directly into tissue and typically employed in bone-conduction headphones. For symmetry, one driver was attached to each side of the ultrasound probe via epoxy with the broadcast surfaces leveled such that the probe and both drivers could all be placed in simultaneous contact with the subject's skin. The form factors of the device and *in vitro* target are illustrated below in Figure 3.1.
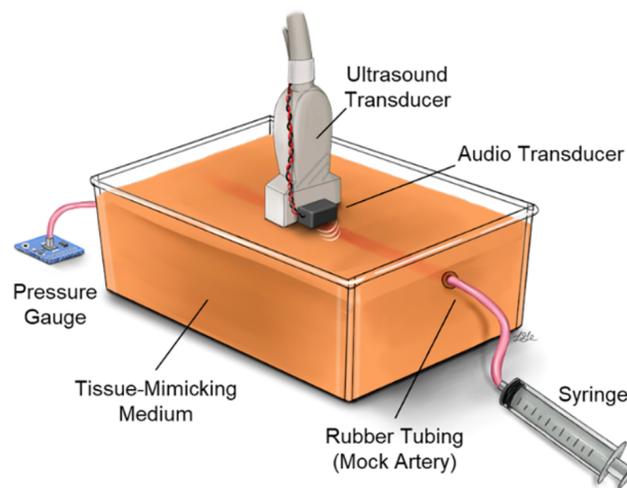


Figure 3.1: An illustration of both the measurement device and the phantom used for *in vitro testing*. The device consists of a commercial ultrasound probe with attached speakers, which simultaneously conduct imaging and stimulate vibrations in the target. The phantom consists of rubber tubing suspended in a tissue-mimicking medium, with a syringe at one end to apply internal fluid pressure and a gauge at the other to precisely measure this pressure.

### 3.2 Signal Processing

**Ultrasound Pulse Chirping**

Forming a 2D cross-sectional image of the volume being assessed by the ultrasound probe, referred to as "B-mode" imaging, is essential for measuring arterial dimensions. An important consideration for this imaging is the shape of the pulse used. When an emitted pulse reflects off of a point source, the receiving element will see the same scaled pulse shape scaled down by some attenuation factor. The location of the reflector is then localized by correlating the received signal with the original pulse. As shown in Figure 3.2a-b, applying this procedure to a single-frequency signal leads to a very wide autocorrelation, which in turn would lead to blurry images after autocorrelation is applied. This behavior can be substantially improved by using a pulse with a linearly time-varying frequency, more commonly referred to as a chirp [74]:

$$x(t) = \sin\left(\frac{\omega_1 - \omega_0}{2T}t^2 + \omega_0 t\right) \tag{3.1}$$

where angular frequency sweeps from $\omega_0$ to $\omega_1$ and $T$ is the total length of the pulse. Figure 3.2c-d shows autocorrelation results for a chirp spanning 3 to 6 MHz, which is the response range of our 6L3 ultrasound probe. This autocorrelation shows a significantly sharper peak, leading to more precise localization of reflection sources.

**Synthetic Aperture Imaging**

Almost all commercial ultrasound instruments use beamforming to focus transmitted ultrasound pulses onto an area of interest; however, this was not possible with our device as we could only transmit from one pixel at a time. Instead, we used synthetic aperture (SA) imaging [74], a family of imaging algorithms which synthesize returns from a series of single-pixel transmissions to create a full image which is focused everywhere. The most basic SA algorithm, referred to as Delay-And-Sum (DAS), begins by converting each individual ultrasound return into an image. This is done by converting each pixel location in the image into a time delay based on the distance from the pixel to the transmit/receive elements and the speed of sound in the medium. Each pixel is then assigned an value equivalent to the intensity of the ultrasound return at the corresponding delay time. This can be summarized mathematically as

$$S_{DAS}(t, j) = \frac{1}{N} \sum_{i=1}^{N} s_i\left(t + \Delta t_{ij}\right) \tag{3.2}$$

$$\Delta t_{ij} = \delta_{ij}/c \tag{3.3}$$
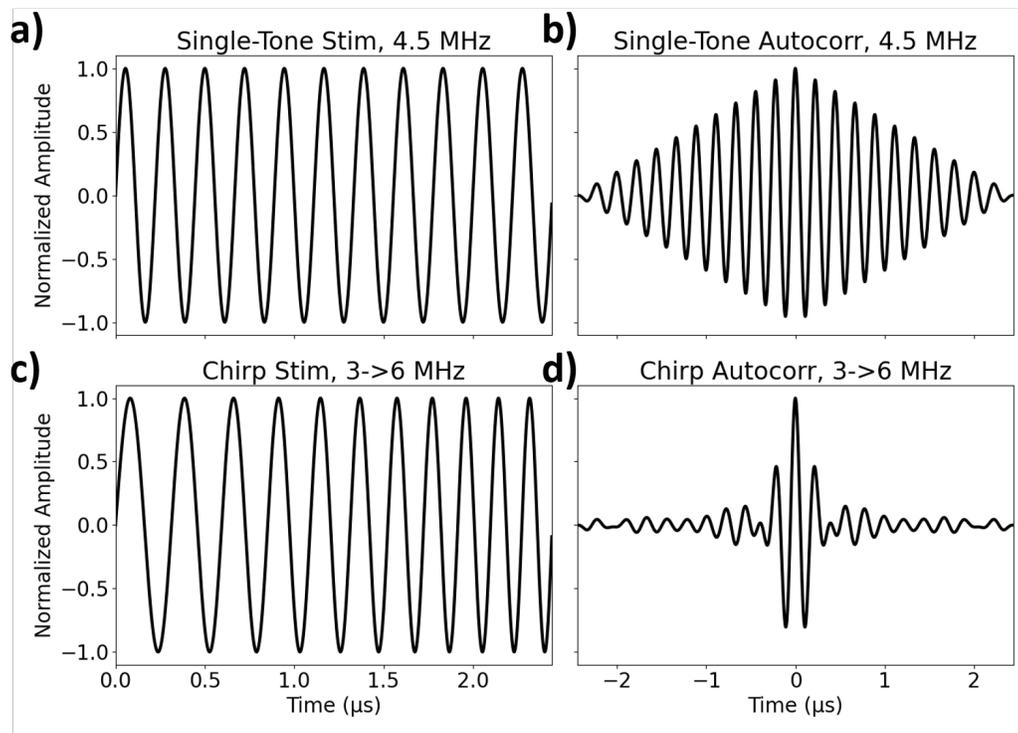
Figure 3.2: The benefits of using a chirped pulse. a) A single-frequency pulse at the center frequency of our transducer. b) The autocorrelation of this single-frequency pulse. The significant spread in energy far from the central peak would lead to blurry images. c) A chirp pulse with frequency linearly ramping across the bandwidth of our transducer. d) The autocorrelation of this chirp pulse. The energy in this autocorrelation is significantly suppressed away from the central peak, leading to sharper images.

where $S_{DAS}$ is pixel intensity, $t$ is the starting time for image construction, $j$ indexes pixel location, $i$ indexes transmit/receive element combination, $\Delta t_{ij}$ is the time delay factor for element combination $i$ and pixel $j$, $s_i(t + \Delta t)$ is the measured ultrasound return from transmit/receive combo $i$ at time $t + \Delta t$, $\delta_{ij}$ is the round-trip distance from the transmit element to pixel $j$ and back to the receive element, and $c$ is the speed of sound in the medium. Since any pixel with the same $\delta_{ij}$ corresponds to the same delay, a point-source reflector will be "imaged" by a single return as an elliptical arc with foci at the transmit and receive elements. When many delay-based images from many transmit-receive pairs are added together (the "sum" part of DAS), the areas of the image with actual reflector sources will constructively reinforce while arcing artifacts in other areas will, on average, cancel each other out.

In practice, DAS imaging requires averaging across many elements to suppress arcing artifacts and attain a useful lateral resolution; this makes it unsuitable for our

device with 16 transmit and 32 receive elements (see Figure 3.3a,b). Multiple improved SA methods have been proposed to increase imaging performance with fewer elements by utilizing phase information from returns in addition to magnitude. The most successful method for our application was CF-DMAS [75], which combines CF-DAS (delay-and-sum with a coherence factor) with DMAS (delay, multiply, and sum):

$$
S_{CF-DMAS}(t, j) = \frac{\left[ \left( \sum_{i=1}^{N} \sqrt{s_i' \left( t + \Delta t_{ij} \right)} \right)^2 - \sum_{i=1}^{N} s_i' \left( t + \Delta t_{ij} \right) \right]^3}{\left( \sum_{i=1}^{N} \left| s_i' \left( t + \Delta t_{ij} \right) \right| \right)^2 - \sum_{i=1}^{N} \left| s_i' \left( t + \Delta t_{ij} \right) \right|^2} \tag{3.4}
$$

$$
s_i' \left( t + \Delta t_{ij} \right) = a_{ij} s_i \left( t + \Delta t_{ij} \right) \tag{3.5}
$$

where $a_{ij}$ is an apodization factor. This factor stems from the fact that piezoelectric elements are most sensitive to signals which arrive normal to their face, and this sensitivity falls off as the angle from normal increases; thus, returns for pixels with extreme angles to a transducer element are suppressed. Rather than the approximate Hanning window used in Jeon, Park, Choi, *et al.* [75], we used a more exact form derived in Selfridge, Kino, and Khuri-Yakub [76]:

$$
a_{ij} = f(\theta_t) f(\theta_r) \tag{3.6}
$$

$$
f(\theta) = \frac{\sin(\pi d/\lambda \sin \theta)}{\pi d/\lambda \sin \theta} \cos \theta \tag{3.7}
$$

where $\theta_t$ and $\theta_r$ are the angles from normal between pixel $j$ and the transmit and receive elements of combination $i$, respectively, $d$ is the element width in the transducer, and $\lambda$ is the average wavelength of the transmitted ultrasound pulse. We also modified the procedure of Jeon, Park, Choi, *et al.* [75] by converting real-valued ultrasound returns into their complex equivalents using a Hilbert transform, generating more accurate phase information. As shown in Figure 3.3c, the CF-DMAS procedure generated significantly better images of point targets than simple DAS; examples of *in vivo* images of arteries using CF-DMAS are shown in Chapter 4. A flowchart summary of the CF-DMAS procedure is provided in Appendix Figure A2.2. Final images were generated with a pixel size of $100\mu$m and a total size of 2.8cm wide by 4cm deep.

**Doppler Velocity Measurement**

The first step in measuring the resonant response of an artery (or arterial phantom) is measuring its wall velocity. This was done using single-element pulsed Doppler

| a) | Ideal | b) | DAS | c) | CF-DMAS |
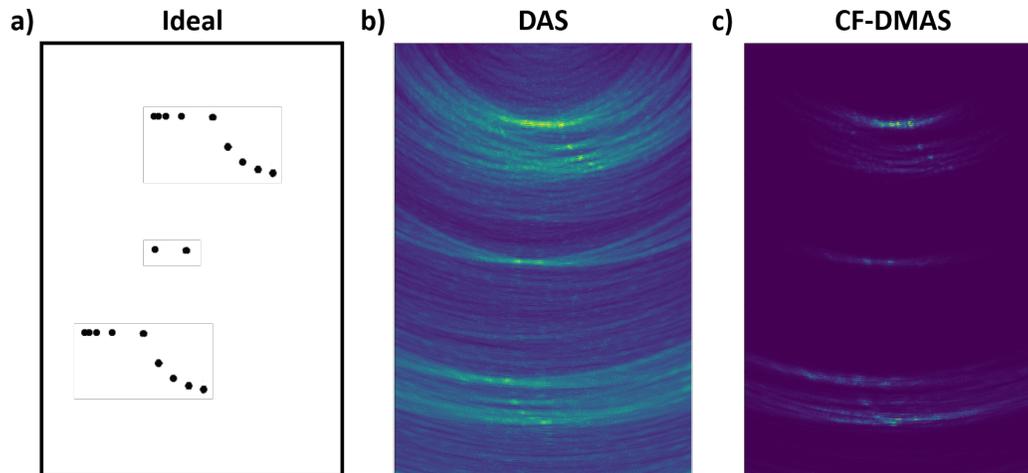


Figure 3.3: The benefits of CF-DMAS image synthesis. a) The layout of a commercial ultrasound phantom used for this test, which uses thin nylon threads to simulate ideal point targets. b) An image from our device with DAS image synthesis. While the point targets can be made out, there are many wide arcing artifacts that significantly clutter the image. c) An image from our device using the same ultrasound parameters but with CF-DMAS image synthesis. While noisy artifacts have not been entirely eliminated, they have been drastically suppressed compared to DAS imaging.

ultrasound [77] which, despite its name, does not actually rely on the Doppler effect. Instead, the target of interest is interrogated at a high rate using repeated ultrasound pulses from the same element. If the target is stationary, these pulses will all generate the same return. However, if the target moves slightly between pulses its return will keep the same shape but shift slightly in time. This shift is measured by taking the difference in return phase (as determined via the Hilbert transform) across consecutive returns; a higher velocity will lead to a larger shift in the return and thus a larger change in phase. The conversion factor between phase change and physical velocity units of m/s depends on factors such as the ultrasound pulse shape, speed of sound, and Doppler pulse repetition frequency. We determined this factor by simulating returns from a perfect point reflector moved by controlled amounts.

Applying this velocity extraction to an entire Doppler return will generate a separate velocity for every return sample at every pulse interval. These velocities were first bandpass-filtered in the time domain using a 4th-order Butterworth filter with cutoffs at 50 and 1000 Hz to suppress signal far from our range of stimulus. To isolate the velocity of an arterial wall, we started with a user-defined window centered on the wall of interest. The darkest 75% of samples in this window (as measured by

magnitude of the chirp-correlated return) were discarded, and the velocities of all remaining samples were averaged with weights proportional to the magnitude of return of each pixel. This procedure was applied separately to the top and bottom wall of the artery, generating a velocity vs. time measurement for each. Summary flowcharts of this process are provided in Appendix Figures A2.3 and A2.4.

**Resonance Extraction**

The $n = 2$ resonance mode we hoped to stimulate and measure is characterized by the top and bottom walls moving opposite from each other (see Figure 2.1b); thus, the velocities of the top and bottom walls were subtracted to generate a differential velocity measure. A digital lock-in amplifier was then applied to this differential velocity to extract the magnitude and phase of its response at each stimulus frequency (Figure 3.4). While the phase of this differential response should follow a sigmoid characteristic of resonance, the magnitude will be determined by both the artery's resonant response and the response function of the audio transducer. To account for this, the differential response magnitude at every frequency was divided by the common (i.e., adding the wall velocities) response magnitude at that frequency (Figure 3.4a). This was based on the assumption that there was no resonant behavior in the common mode response, which meant that variation in magnitude at different frequencies was due solely to the response function of the audio transducer; this assumption was supported by the nearly flat phase response of the common mode signal (Figure 3.4b).

As mentioned in section 2.6, final pressure determination is significantly sensitive to error in resonant frequency, so simply taking the frequency of maximum response was not sufficient. Thus, we performed curve fitting on the response to improve precision beyond the granularity of the individual stimulus frequencies. The functional form of this fitted curve would ideally be determined from physical analysis of the system; however, the unknown nature of damping forces made this intractable. Instead we applied the vector fitting algorithm [78], which models an arbitrary complex frequency response $H(f)$ as a sum of rational functions:

$$H(f) = \Sigma_{m=1}^{N} \frac{r_m}{if - a_m} + d + fe \tag{3.8}$$

where $f$ is frequency, $a_m$ and $r_m$ are complex poles and residues, respectively, and $d$ and $e$ are real linear offset parameters. In particular, a resonant system will have a complex conjugate pair of poles. Vector fitting is an algorithm which uses iterative least-squares fitting to find an optimal set of values for $(r_m, a_m, d, e)$ which best
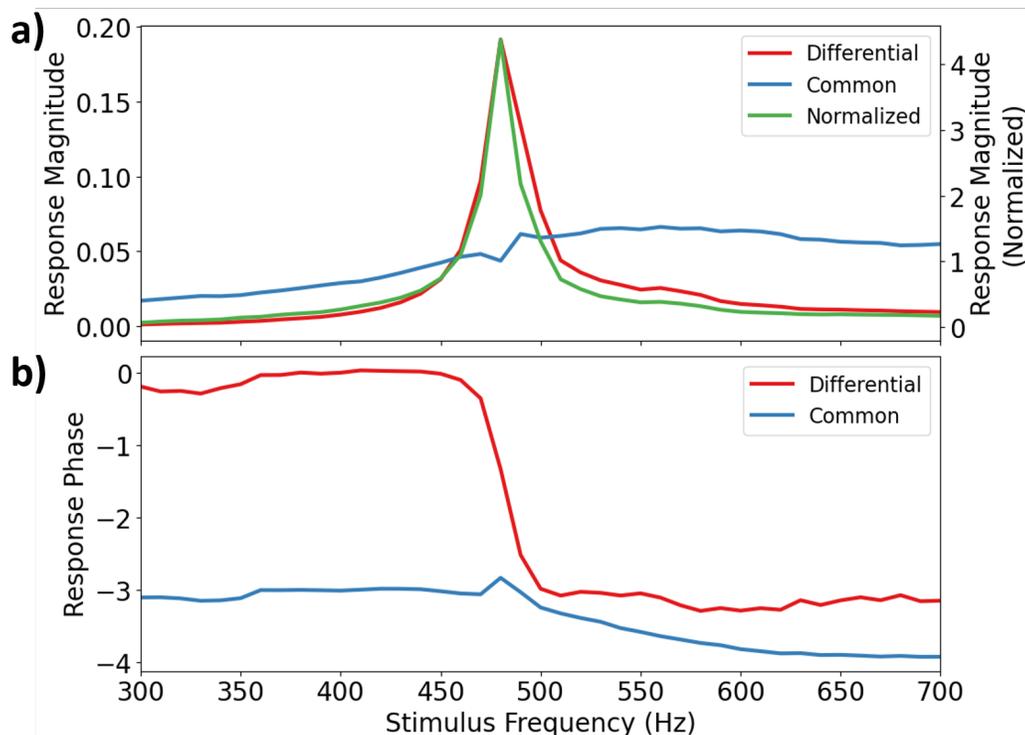
Figure 3.4: Example of a resonant response in the phantom. a) Magnitude response for differential, common, and normalized wall motion in the small phantom at 135 mmHg. Differential magnitude shows a sharp peak characteristic of resonance, while common magnitude shows a gradual change due to non-flat response of the stimulus speaker. To normalize the resonant response we divide differential magnitude by common magnitude, removing the effect of speaker response and leading to a sharper resonant peak. b) Phase response for differential and common wall motion. Differential phase shows a drop of $\pi$ radians centered around the peak in magnitude response, characteristic of resonance. Common phase, in contrast, barely changes, as expected due to lack of common wall motion in the $n = 2$ resonance mode.

match the observed frequency response of the system. The final fitted resonant frequency of the system is represented by the magnitude of our complex conjugate pair of poles; if resonance was not present, the vector fitting algorithm would return a set of purely real poles. In principle this method can match any response function given a sufficient number of poles. For the *in vitro* data only one pairs of pole was necessary, but for the *in vivo* data fit fidelity was improved by adding a second pair of poles, and the pole with maximum response was chosen as the "true" resonance frequency. Examples of the resulting fits are shown in Results, and a summary flowchart of the frequency fitting process is provided in Appendix Figure A2.5.

**Timing**

An important consideration underlying all of the above methods is timing, as measurements must be taken frequently enough to appropriately sample the behavior we are trying to measure. As we discussed above, the sampling rate on our ADC (25 MHz) is fast enough to capture the highest frequency component of our ultrasound pulses (6 MHz). The remaining concerns were imaging frequency and velocity measurement frequency.

The fundamental limit on how quickly ultrasound measurements can be repeated is the travel time of sound, as we need to wait for an emitted pulse to return before moving onto the next pulse. The arteries we wished to image may have bottom walls up to 5cm deep beneath the skin, which means sound waves have to travel about 10cm round trip. Given that the speed of sound in tissue is 1540 m/s [72], the round trip travel time was 130 $\mu$s, implying a maximum possible pulse repitition rate of about 15.4 kHz. In practice, we limited our rate to 10 kHz to allow for a safety buffer.

The pulses we emitted needed to be allocated to provide both B-mode imaging and Doppler velocity measurements. Analyzing frequency components in Doppler response is easiest if measurements are obtained at constant intervals, so we allocated B-mode and Doppler acquisitions to alternating pulses such that each was performed at a constant rate of 5 kHz. For Doppler measurements the same transmit element was used every time, so we obtained velocity measurements at 5 kHz. For larger arteries such as the carotid and brachial we expected resonant frequencies to be under 1 kHz (see Table 2.1), so this sampling rate was more than sufficient; for future device iterations that target smaller arteries such as the radial, it may need to be increased. For B-mode measurements, forming a full image required performing one pulse cycle from each of our 16 transmit elements, so the effective rate of image formation was 312.5 Hz. When imaging a human artery the only factor causing significant image variation will be heartbeats, which occur no faster than 4 Hz; thus, this imaging rate was also more than sufficient.

A further consideration was the maximum velocity measurable using this method, which depends on the conversion factor between phase change and physical units as well as the Doppler pulse repetition frequency. For our particular chirp, a phase change of 1 radian corresponded to physical motion of 27.4 $\mu$m. The maximum shift we could reliably measure was a motion of $\pi$ radians between consecutive Doppler pulses, which occurred at a rate of 5 kHz. This yielded a maximum measurable

velocity of $(27.4 * 10^{-6}$ m$) \times \pi \times (5000$ s$^{-1}) = 0.43$ m/s; system velocities higher than this would result in aliasing. Even for a very fast heartrate of ~4 Hz, the distention of the artery from systolic rise will not take less than 1/40 s and each wall will not move by more than 1 mm; thus, we could be confident that wall velocity due to the heartbeat would not exceed 0.04 m/s, well below our aliasing limit. The frequency of motion from our audio stimulus may be 20 times higher than this (i.e., motion occurring in ~ 1/800 s), but the magnitude of displacement excited by our stimulus was far lower than displacement from the heartbeat. This indicated that the audio stimulus would also not generate wall velocities nearing our aliasing limit.

## 3.3 Experimental Methods

**Phantom Target**

To construct a synthetic artery analogue we used compliant, thin-walled rubber tubing sourced from latex rubber balloons (Qualatex 160Q or similar). Ultrasound phantoms (tissue-mimicking mock-ups) consisting of fluid-filled rubber tubing suspended in gelatin are often used as ultrasound teaching aids as they provide similar imaging properties to blood vessels embedded in tissue [79], [80]. We chose to substitute a water/psyllium fiber (Metamucil) mixture [81] for the gelatin/psyllium fiber mixture, as tubing can disbond from the gelatin as pressure (and thus also the tubing's radius) is changed, leading to air pockets and behavior disparate from real anatomy. Commercial ultrasound tissue models were unsuitable for these experiments as they do not have similar elasticity in their vessel analogues; several vendors we investigated used rigid tubing, and all were flow-only (no pressure simulation).

For our experiments we used two sizes of tubing: 2.18mm radius ("small") and 3.23mm radius ("large"). The wall thicknesses of each were 0.25mm and 0.28mm, with a density of 1930 kg/m$^3$ measured for both. The tubing was submerged to a depth of 2-3 cm in the water/psyllium fiber bath, modeling a depth similar to that of the human carotid artery. Psyllium fiber was used as a tissue-mimicking scattering medium to disperse ultrasound echoes reflected from the edges of the container. The tubing was filled with water and inflated using a syringe to add pressure.

**Experimental Procedure**

Each scan on the small phantom consisted of a stimulus sweep from 200 to 600 Hz in 10 Hz steps with simultaneous measurement using the ultrasound transducer. Five scans were performed at each pressure, and pressure was swept from 60 to

150 mmHg (targeting a physiologically-relevant range) in 5 mmHg increments, for a total of 95 scans. Pressure was held constant for the course of a scan.

The experiment was repeated using the larger diameter tubing to confirm that the model held across different vessel sizes. Pressures were swept from 60 to 150 mmHg in 10 mmHg increments. Above 130mmHg, we found that sections of the tubing entered elastic failure and expanded rapidly in an uneven manner, as normally seen during inflation of a balloon. As this behavior is not seen in healthy arteries, we discarded data above 130mmHg, for a total of 40 scans.

**Data Analysis**

For each individual scan, radius was calculated from the average delay in echo timings between the brightest points in the top and bottom walls, and resonant frequency was calculated using the vector fitting method described above. Because the tubing walls were significantly thinner than those of *in vivo* arteries, thickness could not be determined accurately from our ultrasound imaging due to limited resolution. Instead, we used high-precision calipers to measure the unpressurized radius and thickness of the tubing ($a_0$ and $h_0$). Because the tubing was assumed to be incompressible ($v = 0.5$), a pressure-dependent thickness could be calculated as $h = h_0 * (r_0/r)$. These caliper measurements along with the weight of the tubing were also used to calculate its density.

The Young's modulus of the tubing was calculated by comparing radius and resonant frequency measurements across multiple scans at different pressures using equation 2.35. We assumed that the tubing was linearly elastic, so a single value of $E$ was calculated which minimized the relative error in pressure as determined by Equations 2.25 and 2.35; this value came out to 1.16 MPa. The balloon material was later analyzed with a tensile strength measurement instrument from Instron (Norwood, MA). This test yielded an average stiffness of 1.10 MPa which held nearly constant across our strain range, validating both our calculated value and our linearity assumption.

Measured radii were adjusted for each scan based on this fixed $E$ value to generate agreement with Equation 2.35. For the larger tubing, obtaining alignment with theory required adding $h/2$ to all radii; this would be explained if peak echoes from this system corresponded to the inner rather than average radius of the tube. The measured values for radius, thickness, resonant frequency, and stiffness were

combined with prior values for wall density, fluid density, and wall Poisson ratio in Equation 2.25 to generate the final calculated pressure values.

## 3.4    Results and Discussion

Figure 3.4 shows the measured frequency response of our small phantom at a single pressure, and it demonstrates the classic signatures of resonance, i.e., a peak in magnitude space and a co-located sigmoidal rolloff in phase with height $\pi$ radians. This indicates that our measurement device was able to both stimulate and detect resonance in an artery-like system. The next step was to assess whether or not this resonant behavior varied in the way predicted by the physical model. Figure 3.5 shows the measured response of the small phantom at five different internal pressures from 95 to 135 mmHg, as well as best fit functions obtained from vector fitting. The resonant frequency shifted steadily upwards as pressure was increased, in line with our physical model.
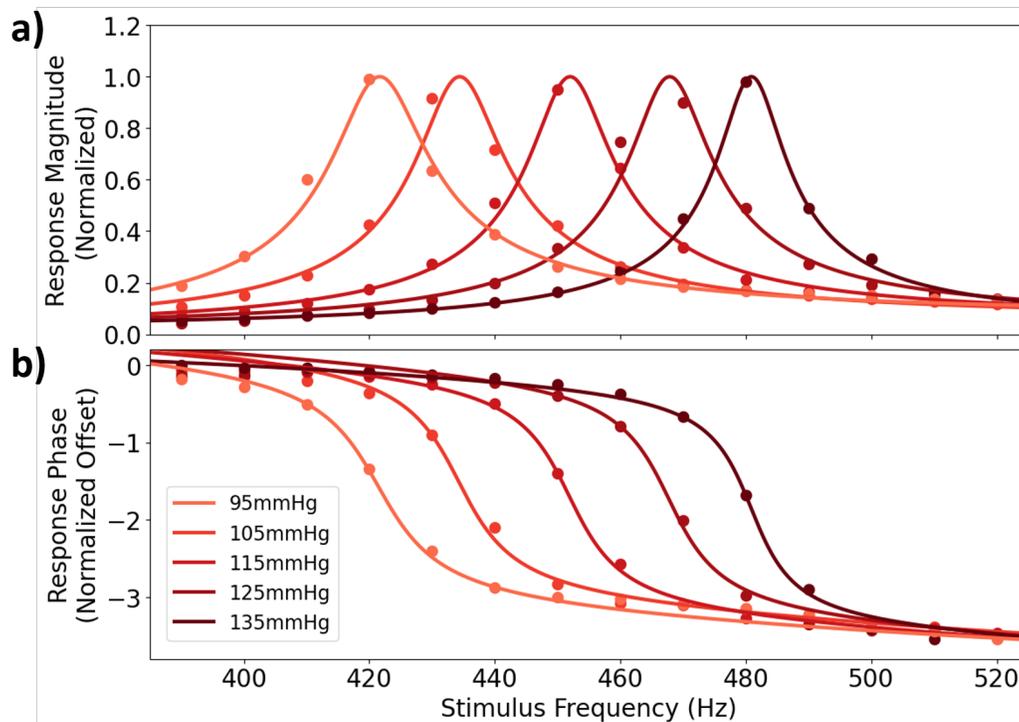


Figure 3.5: Plots showing the (a) magnitude and (b) phase responses of the small phantom as pressure is increased from 95 to 135 mmHg. Points represent individual values (measured in 10 Hz intervals), and lines represent best-fit curves obtained via vector fitting. The resonant frequency clearly moves up as pressure increases, in line with the physical model.

We next combined all of our *in vitro* scans to quantitatively assess the accuracy of our physical model. Because our tubing was inflated with a controlled pressure, we began by using equation 2.23 to predict resonant frequencies and compared with experimental results. Overnight storage of our phantoms at pressure caused some plastic deformation partway through the experiments, so each phantom had a different pressure vs. radius relationship between the two days of data collection. Despite this, the predicted curves very closely matched the observed resonant frequencies across all scans, as shown in figure 3.6a. The measured resonance frequencies were then used to predict internal pressure in the phantoms using equation 2.25, as shown in figure 3.6b.

Comparison of predicted to actual phantom pressure shows that our physical model produced consistently accurate predictions, with a correlation of $r > 0.995$ for both the small and large phantoms. The residual errors are shown in the inset of 3.6b; across all measurements, the mean error was $-1.09 \pm 1.98$ mmHg. The residual errors did not show any obvious pattern when comparing day 1 to day 2 measurements (as indicated in figure 3.6a) despite the plastic deformation, indicating that the change in radius trend was correctly accounted for. Our mean error compares favorably with standards for BP cuff accuracy set by the International Organization for Standardization (ISO), which limit the mean and standard deviation of BP measurement error to 5mmHg and 8mmHg, respectively [82]. These results indicate that, at least in an idealized system, the physical model and measurement device are more than capable of producing BP measurements with clinically useful accuracy and precision over a range of pressure values and artery sizes.
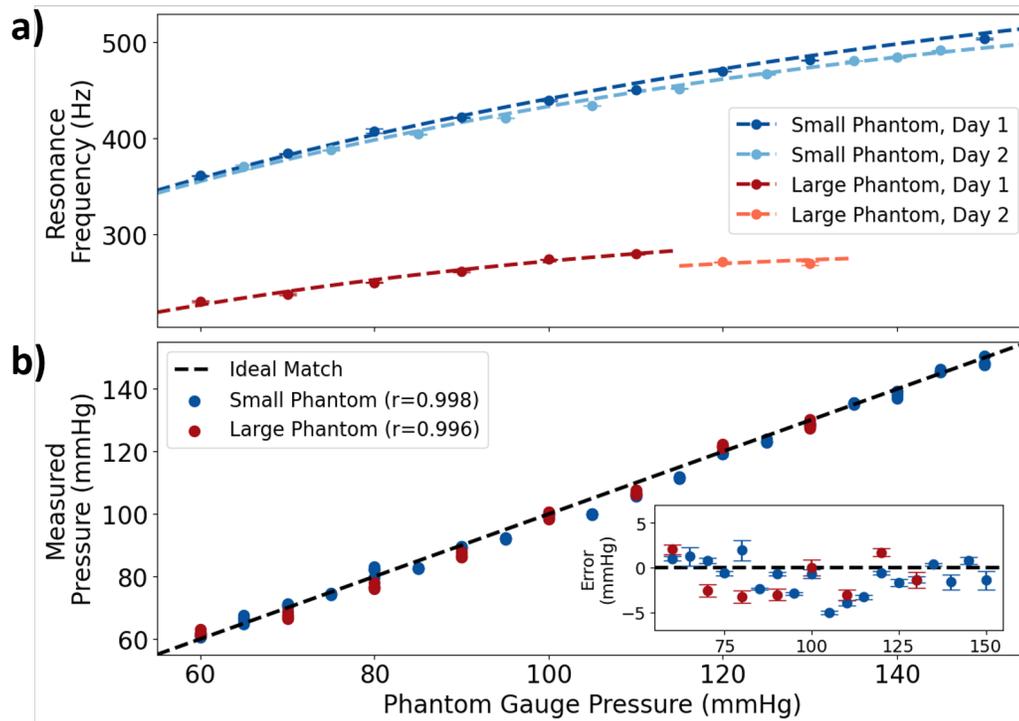
Figure 3.6: *In Vitro* phantom testing results. a) Predicted and measured resonant frequencies for the small and large phantoms across a range of pressures. Measured values (points) were generated via vector fitting on frequency responses, and error bars represent spread across 5 replicate measurements. Physical model predictions (dashed lines) were calculated from measured parameters using equation 2.23. The phantoms experienced some plastic deformation overnight between measurements, leading to two different pressure-frequency curves for each phantom. b) Measured vs. true fluid pressure for the small and large phantoms (main plot), along with residual errors (inset). Measured pressure values were calculated from equation 2.25, and true values were taken directly from the pressure gauge. The close agreement between measurements and true values indicates that our resonance model is an accurate description of the physical system.

*C h a p t e r  4*

## *IN VIVO* TESTING

Chapter 3 demonstrated that our physical model is very accurate at matching resonant behavior to pressure in a static artificial system that matches all of our assumptions. However, real arteries are made up of a multi-layered wall structure embedded in inhomogeneous surrounding tissue, and both the pressure and dimensions change continuously due to heartbeats. This chapter will demonstrate that, despite these complexities, CARDI-BP is able to generate continuous and accurate blood pressure measurements in human arteries. It will first describe the modifications to stimulus and data analysis necessary to provide robust real-time measurements, and then show initial measurements demonstrating that resonance could be measured in a small sample size of human arteries. It will then show the results of our $N = 60$ clinical study, which directly compared the performance of CARDI-BP against a gold-standard arterial catheter to provide a quantitative test of accuracy.

### 4.1   Real-Time Signal Processing
Because our phantoms were maintained at a static pressure for the duration of a scan, measurements could be taken by stimulating the system with one frequency at a time over many steps and compiling the results together afterwards. In addition to simplifying the stimulus protocol, the long measurement time allowed for significant noise suppression through averaging. *In vivo* measurements demand a much faster response time, as pressure changes rapidly due to heartbeats; thus, making these measurements work required modifications to both stimulus and signal processing.

**Multisine Stimulus**
The first step was developing a way to stimulate the artery with many frequencies at once. A multi-frequency stimulus, which we term "multisine," can easily be constructed via linear superposition of many single-frequency signals, and the response of the artery to each individual frequency can be extracted via a Fourier transform. The primary challenge lies in crafting an effective superposition. Any arbitrary

superposition with period $T$ and finite bandwidth can be expressed in the following way:

$$S(t) = \sum_{k=1}^{N} p_k \sin\left(\frac{2\pi kt}{T} + \theta_k\right) \tag{4.1}$$

where $p_k$ is the power of the kth harmonic (normalized such that $\sum p_k = 1$) and $\theta_k$ is its phase angle. Since we were measuring velocity rather than displacement, we set $p_k \propto 1/k$ such that $dS(t)/dt$ had a flat frequency content. Commercial vital sign monitors which process A-line data generally apply a lowpass filter with a cutoff frequency of 12 to 20 Hz [83], so we infered that there is not important BP information at frequency bands above 20 Hz. Thus, we set our period to $T = 50$ms, which led to a 20 Hz discretization in our stimulus frequencies. Based on the analysis in Table 2.1, we set the bandwidth of our stimulus to 140-440 Hz when measuring the carotid artery to capture all but the most extreme cases (this range was shifted as appropriate for measurements on other arteries).

The simplest way to generate a multisine signal with these parameters is to set all $\theta_k = 0$, which leads to the stimulus waveform shown in Figure 4.1a. If our audio transducer and velocity measurements had unlimited dynamic range and sensitivity, this stimulus would have worked fine. However, in reality our audio transducer had a maximum displacement, so the actual output of our stimulus waveform was always normalized to the waveform's peak value. While a single-frequency sine wave has an average output power (defined as the integral of its square) of 0.5 over one period, the normalized waveform of Figure 4.1a has a very low average output power of 0.047. Using this stimulus would have significantly reduced the signal-to-noise ratio of our measurements.

It was clear that we could achieve better average power in our stimulus with appropriate phase shifts in the various components. Solving for the optimum phases to maximize average power is an unsolved problem for the general case. Fortunately, however, Schroeder [84] provides a useful heuristic for making a multisine signal with relatively good average power:

$$\theta_k = \pi \left| \sum_{i=1}^{k-1} (i - k)p_i \right|. \tag{4.2}$$

Applying these phases resulted in the stimulus waveform shown in Figure 4.1b. This stimulus has an average power of 0.312, over 6 times higher than the $\theta_k = 0$
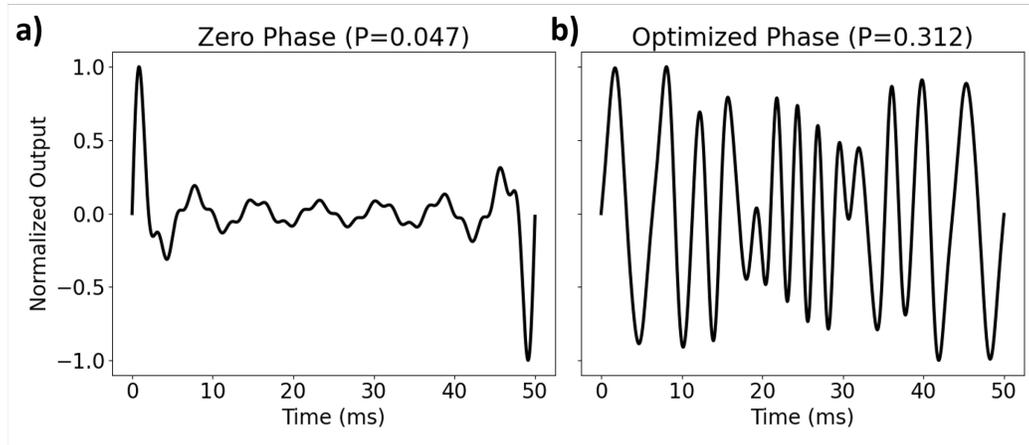
Figure 4.1: The benefits of multisine phase optimization. a) A multisine stimulus covering 140-440Hz in 20Hz steps, with all components starting at 0 phase at time 0. Because the peak of the signal is normalized to 1 (the maximum output displacement of the audio transducer), the total energy contained in the signal is relatively small. b) A multisine stimulus covering the same frequencies, but with component phases optimized according to Equation 4.2. Because the peak is much less prominent, the total energy of the signal is significantly higher.

waveform and approaching the single-frequency limit of 0.5. Equation 4.2 was used to generate stimulus waveforms for all *in vivo* experiments.

The arterial wall velocities produced by the multisine stimulus were analyzed using a spectrogram, which applies a sliding window FFT to generate a complex frequency response function at each time step. Because our stimulus function was periodic and smooth, any time window could be used for analysis; this theoretically capped our response measurement rate at the rate of Doppler velocity acquisitions, or 5 kHz. In order to reduce computational overhead we limited the response measurement rate to 200 Hz, or 10x higher than the highest expected frequency content of the waveform we were trying to measure.

**Artery Identification and Dimension Extraction**

In our *in vitro* experiments we were guaranteed a clean environment with no significant ultrasound returns from anything other than the thin wall of the tubing. Unfortunately, real arteries are embedded in messy environments with other nearby reflectors. This necessitated new image analysis techniques to extract artery radius and wall thickness from our images.

To begin the process, the user was shown a live view of the B-mode images generated by the ultrasound probe using synthetic aperture processing (see Section 3.2). This

allowed the user to find the artery and position the probe to obtain a longitudinal view of both walls of the artery, as shown in Figure 4.2a. The user then drew two region of interest (ROI) boxes, one encompassing the center of each wall (see Figure 4.2b). This ensured that all subsequent analysis was performed on the artery of interest rather than a neighboring structure such as a vein.
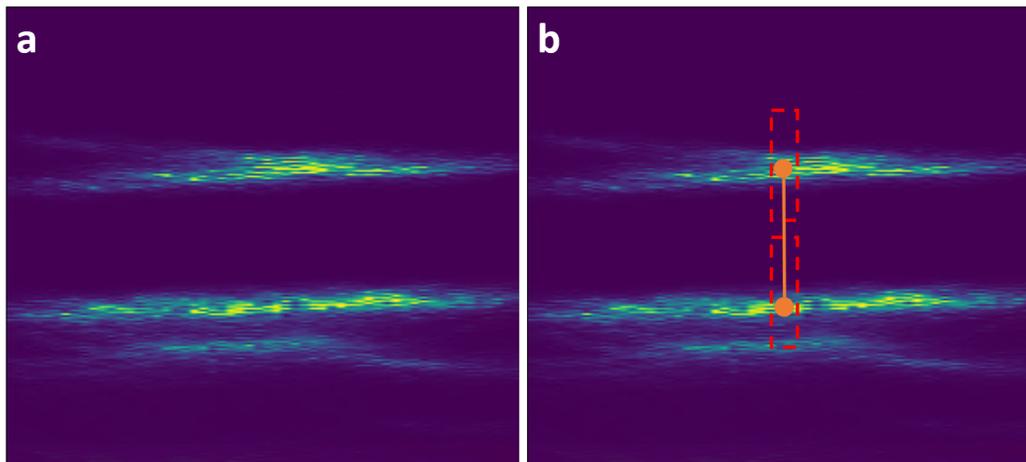


Figure 4.2: Artery identification and annotation. a) An example of an un-annotated artery image. We can see the artery walls as two bright horizontal features, along with some dimmer neighboring features which we want to ignore. b) The same artery image annotated with user-defined top and bottom wall ROIs (red dashed boxes) and computed wall centers of mass (orange dots). The distance between the wall centers of mass (orange line) was used to measure artery radius.

Within each wall ROI, the dimmest 75% of pixels were zeroed, and the remaining brightest pixel locations were used to determine a brightness-weighted center of mass (COM) location for each wall. The radius of the artery was measured as half of the vertical distance between the two wall COMs. A brightness curve was then determined above and below each COM point using cubic interpolation, and the thickness of each wall was measured as the full width at half max of the brightness peak surrounding the COM. The single thickness parameter for the artery as a whole was taken to be the average of the computed top and bottom thicknesses.

An additional correction was made based on the angle of the artery in the image, as a non-horizontal artery would result in over-estimation of radius and wall thickness and under-estimation of wall velocity. This was done by taking the brightness-weighted COM of each pixel column in each wall ROI and then performing a linear regression through these column COMs to find a slope. The total angle of the artery away from horizontal, $\theta$, was measured as the average angle from the two artery

walls, and radius, thickness, and wall velocity were corrected by factors of $\cos\theta$, $\cos\theta$, and $1/\cos\theta$, respectively.

**Unscented Kalman Filter**

The three system parameters we measured in order calculate BP (frequency, radius, and thickness) all had their own noise characteristics. The easiest approach to denoising these measurements would have been to simply apply a separate lowpass filter to each one. However, this would have discarded significant information we had about how the various system parameters evolved in relation to each other. This sort of knowledge about interrelated parameters in a nonlinear system is often integrated via an Unscented Kalman Filter (UKF) [85], [86]. With any type of Kalman filter, the system is paramaterized according to its underlying state $\vec{x}_k$ and a set of observations $\vec{z}_k$ which are somehow related to the underlying state. Defining the filter requires specifying the state transition function $f_k(\vec{x}_{k-1})$, the observation function $h_k(\vec{x}_k)$, the process noise covariance $\mathbf{Q}_k$, and the observation noise covariance $\mathbf{R}_k$ such that

$$\vec{x}_k = f(\vec{x}_{k-1}) + \vec{w}_k \tag{4.3}$$

$$\vec{z}_k = h(\vec{x}_k) + \vec{v}_k \tag{4.4}$$

$$\vec{w}_k \sim \mathcal{N}(0, \mathbf{Q}_k) \tag{4.5}$$

$$\vec{v}_k \sim \mathcal{N}(0, \mathbf{R}_k) \tag{4.6}$$

where $\vec{w}_k$ and $\vec{v}_k$ and process and observation noise parameters drawn from a 0-mean multivariate normal distribution. These equations describe the observations which are expected from a system which is evolving under its own internal dynamics. In a noisy system, observations and expectations will not exactly match; the goal of the Kalman filter is to determine the optimal evolution of state vectors which minimizes the residual between observations and expectations. The UKF is an extension of the original Kalman filter which able to handle nonlinear systems. For this project, the UKF was implemented using the *UnscentedKalmanFilter* module in the *filterpy* package [87].

Applying the UKF to the full nonlinear physics model of Equation 2.25 tended to result in unstable behavior, presumably because the evaluation points used to estimate the system's mean and covariance could stray too close to the poles inherent in the equation's denominator. However, the filter did perform very well with the

approximate version of the physics model from Equation 2.32 (which lacks any poles). The system state and observation vectors were defined as

$$\vec{x} = [a, v, q, E, P] \tag{4.7}$$

$$\vec{z} = [\hat{a}, \hat{v}, \hat{h}, \hat{E}, \hat{f}] \tag{4.8}$$

where $a$ is radius (in m), $v$ is half of total differential wall velocity (in m/s), $q$ is the radius-thickness product (in m$^2$), $E$ is Young's Modulus (in Pa), $P$ is blood pressure (in Pa), $h$ is wall thickness (in m), and $f$ is resonant frequency (in Hz), and the hat is used to denote an observed (as opposed to a system state) quantity. The state transitions were defined as

$$a_k = a_{k-1} + v_{k-1}dt \tag{4.9}$$

$$v_k = v_{k-1} \tag{4.10}$$

$$q_k = q_{k-1} \tag{4.11}$$

$$E_k = E_{k-1} + \left.\frac{dE}{da}\right|_{a_{k-1}} \left.\frac{da}{dt}\right|_{t_{k-1}} dt \tag{4.12}$$

$$= E_{k-1} + \left.\frac{dE}{da}\right|_{a_{k-1}} v_{k-1}dt \tag{4.13}$$

$$P_k = P_{k-1} + \left.\frac{dP}{da}\right|_{a_{k-1}} \left.\frac{da}{dt}\right|_{t_{k-1}} dt \tag{4.14}$$

$$= P_{k-1} + \frac{q_{k-1}E_{k-1}}{a_{k-1}^3} v_{k-1}dt \tag{4.15}$$

where we have used Equation 2.35 to express $dP/da$ in terms of our other state parameters. In Equation 4.10 we know that the actual wall velocity will be far from constant as heartbeats go through, but at any given time step we do not have a good prior for whether the wall's motion will accelerate or decelerate in the next time step. In kinematic KFs this is normally addressed by giving velocity a constant state transition with a relatively high process noise and relatively low measurement noise. The expected observation functions were defined as

$$\hat{a}_k = a_k \tag{4.16}$$

$$\hat{v}_k = v_k \tag{4.17}$$

$$\hat{h}_k = q_k/a_k \tag{4.18}$$

$$\hat{E}_k = E_k \tag{4.19}$$

$$\hat{f}_k = \left[ \frac{1}{\pi^2 \left(q_k \rho_S + \frac{4}{5}\rho_L\right) a_k^2} \left(P + \frac{q_k^3}{3(1 - v^2)a_k^6}E_k\right) \right]^{1/2} . \tag{4.20}$$

The measurement of $E$ and $dE/da$ relied on an iterative convergence process to simultaneously satisfy Equations 2.32 and 2.35; further details are provided in Section 5.3. The process noise matrix was defined as

$$\mathbf{Q}_k = \begin{bmatrix} \frac{dt^2}{40}\overline{v^2} & \frac{dt}{20}\overline{v^2} & 0 & 0 & 0 \\ \frac{dt}{20}\overline{v^2} & \frac{1}{10}\overline{v^2} & 0 & \frac{dE}{da}\Big|_{a_{k-1}}\frac{dt}{20}\overline{v^2} & \frac{q_{k-1}E_{k-1}}{a_{k-1}^3}\frac{dt}{20}\overline{v^2} \\ 0 & 0 & \bar{q}^2/1000 & 0 & 0 \\ 0 & \frac{dE}{da}\Big|_{a_{k-1}}\frac{dt}{20}\overline{v^2} & 0 & \bar{E}^2 & 0 \\ 0 & \frac{q_{k-1}E_{k-1}}{a_{k-1}^3}\frac{dt}{20}\overline{v^2} & 0 & 0 & \bar{P}^2/1000 \end{bmatrix} dt^2 \quad (4.21)$$

where $\bar{x}$ is the mean value of $x$ across the scan. The relative covariance relationship between $a$ and $v$ was taken from the standard first-order kinematic filter setup in Labbe [87], and the process noise associated with this $v_{k-1}dt$ transition term was added into off-diagonal terms for $E$ and $P$ as well. The process noise variances for $q$ and $P$ were set relatively low as we expected them to follow our predicted behavior quite well, while variance for $E$ was set relatively high because its short-term estimates can be relatively noise (see Chapter 5). The observation noise matrix was defined as

$$\mathbf{R}_k = \begin{bmatrix} \bar{a}^2/1000 & 0 & 0 & 0 & 0 \\ 0 & \overline{v^2}/1000 & 0 & 0 & 0 \\ 0 & 0 & \bar{h}^2 & 0 & 0 \\ 0 & 0 & 0 & 10^5 & 0 \\ 0 & 0 & 0 & 0 & 25*\epsilon_f \end{bmatrix} \quad (4.22)$$

where $\epsilon_f$ is an error modification term that was normally 1, but was set to 100 if vector fitting did not converge properly (indicating that the final pole location was unreliable).

The application of a filter that integrates information from multiple sources can produce powerful results. Figure 4.3 shows an example result snippet from our clinical study data (see Section 4.5). The top panels show measurements of resonant frequency and radius, which are the two most sensitive parameters in the BP calculation (see section 2.6). In this case the measured resonant frequency was fairly stable in its range but highly noisy, making it difficult to pick out a pulsatile pattern. The measured radius was more pulsatile but still showed significant noise. As shown in Figure 4.3c, simply combining these inputs using our formula and applying a lowpass filter to the result produced a substantially noisy BP output. However, the UKF was able to incorporate information from other measured parameters, such as

velocity, to denoise the output and produce a BP measurement curve which looked much more physiologically reasonable.
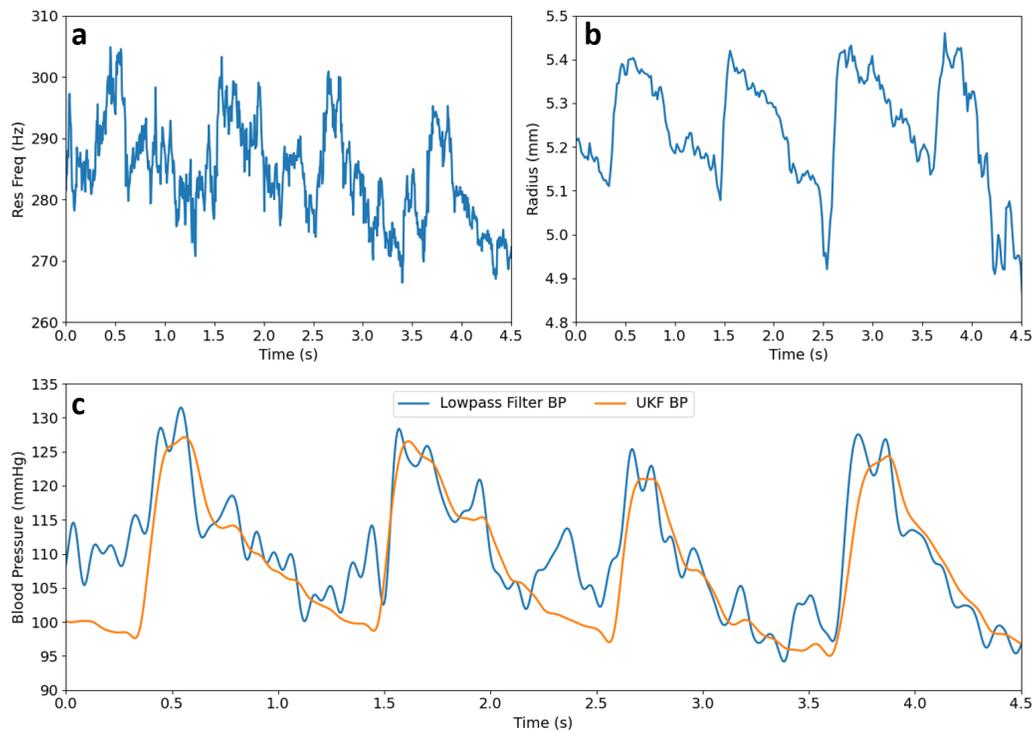


Figure 4.3: An example of unscented Kalman filter (UKF) performance. a) Measured resonant frequency over a ~4 heartbeat period. The values stay in a reasonably stable range but are noisy, showing only a weak pulsatile pattern. b) Measured radius over the same period. The behavior is more pulsatile than that of frequency, but it still shows substantial noise. c) BP over the same period after passing through a 12Hz lowpass filter vs the UKF. After lowpass filtering alone the output BP is still significantly noisy, reflecting the noise in its input parameters. The UKF, however, is able to effectively denoise by synthesizing multiple streams of measurement information, producing a far more physically plausible BP curve with more stable systolic and diastolic values.

The determination of optimal $\mathbf{Q}$ and $\mathbf{R}$ matrices is in general an area of open research, particularly for nonlinear systems. The matrix values listed above were chosen to yield good stability on this dataset via a combination of system knowledge, inspection of noise in some representative scans, and trial and error. Future work may experiment with methods of estimating $\mathbf{Q}$ and $\mathbf{R}$ dynamically from observed data [88].

**Data Collection Software**

All of the signal processing and analysis tasks described above were integrated into a real-time BP measurement interface, as shown in Figure 4.4. This allowed us to get immediate feedback on whether or not we were appropriately imaging the artery and obtaining a resonance signal. The full flow of data through the software pipeline is illustrated in Appendix Figure A2.1. Although our timing setup allowed a maximum B-mode imaging rate of just over 300 Hz, the computation required for SA image generation limited the effective frame rate to roughly 60 Hz. Cubic interpolation was performed on extracted dimensions to upsample them to the 200Hz rate of resonant frequency calculation, allowing for final BP outputs at 200Hz.
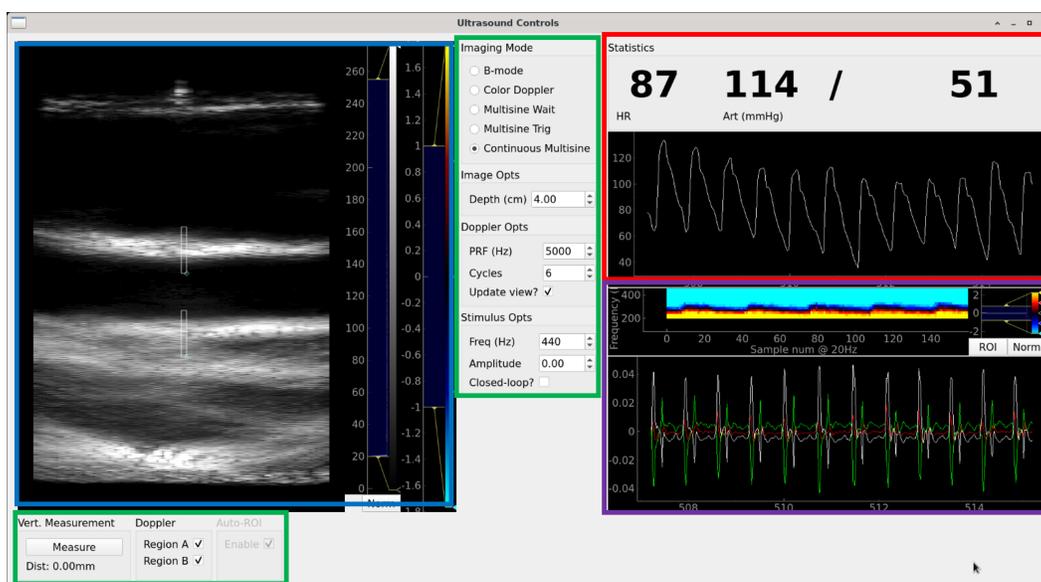


Figure 4.4: A screenshot of the software used for real-time BP data acquisition. Blue box: B-mode image of the target artery with wall ROIs drawn. Green boxes: ultrasound parameter controls. Red box: real-time output of measured BP curve, along with inferred heartrate and systolic/diastolic pressures. The numerical values and y-axis scale were omitted during data collection to avoid bias. Purple box: Real-time output of frequency response and wall velocities to allow for troubleshooting.

## 4.2 Feasibility Study: Methods

An initial feasibility study was conducted on $N = 6$ human subjects to determine whether the resonance underlying the CARDI-BP method could be detected in human arteries. As a first target, we elected to study the common carotid artery in the neck. This artery is relatively large, shallow, and can be easily accessed without any disrobing, making it an ideal target for manually operated ultrasound. The neck is also relatively broad and flat, allowing the ultrasound probe and attached speakers

to simultaneously make good skin contact. Subjects also had their BP measured with a traditional oscillometric cuff which reported diastolic BP (DIA), systolic BP (SYS), and mean arterial pressure (MAP). These data were obtained under guidelines for self-experimentation [89], and written, informed consent was obtained prior to data collection. Data collection from each subject consisted of 6 separate 30-second scans using the prototype device, along with cuff BP measurements before any scans, at the halfway point, and after all scans.

It was observed that successfully obtaining a BP waveform required careful positioning of the probe directly over the center of the artery, and slight shifts from either the operator or subject could easily disrupt the measurement. Furthermore, data were processed without the unscented Kalman filtering described above; instead, a simple low-pass filter was applied. Due to these factors and the relative inexperience of the ultrasound operators, much of the data obtained was not usable. As a screen to identify segments of data likely to contain useful information, we selected for 5-second time spans with a correlation of $r > 0.63$ between measured arterial radius and fitted resonance frequency; if all aspects of the measurement were working properly, both parameters should have varied synchronously with the heartbeat. This limit was determined by inspection to filter out unphysical data while maintaining a reasonable amount of passing data for all subjects. We also rejected any 5-second windows where the pulse pressure (i.e., systolic minus diastolic pressure) exceeded 150 mmHg, a threshold that has been used in prior literature to exclude unfeasible results [90]. These two metrics were found to be sufficient for removing unphysical results resulting from poor image and/or resonance quality. Statistics (e.g., mean arterial pressure) for the present method were computed over the valid time frames and the standard deviations of each statistic was determined using the number of heartbeats as the population size.

We also wished to investigate whether or not the CARDI-BP method could be applied to other arteries beyond the carotid. From various attempts on a single subject, we were able to obtain results from the brachial artery in the bicep, the axillary artery in the shoulder, and the femoral artery in the leg. We also attempted to measure BP in the radial artery in the wrist, but this artery was too small and superficial to be effectively imaged with the 6L3 ultrasound probe.

## 4.3   Feasibility Study:  Results and Discussion

The primary result from the feasibility study was that, despite the complexities of the *in vivo* system, we were able to stimulate and detect the same kind of resonance that was measured in the idealized phantom experiments. An example is shown in Figure 4.5: both the phase and magnitude of the differential wall velocity followed the expected resonance pattern, with phase showing a sigmoidal decrease and magnitude showing a peak near the midpoint of this sigmoid. While the width of these resonance features was significantly wider than the resonance features from the phantom (presumably due to much higher damping in real tissue), we were still able to recover a precise resonance frequency via vector fitting. Importantly, the resonance was not just present, it also shifted quickly to match changes in arterial dimensions and BP due to the heartbeat. This responsiveness was critical for measuring the full shape of the BP waveform rather than just assessing averaged characteristics.

A similar resonant response pattern was measured in all 6 subjects in the study. The measured resonant frequencies were combined with arterial dimensions via Equation 2.25 to generate continuous BP waveforms. An example trace from each subject is shown in Figure 4.6, along with systolic and diastolic BPs measured by the inflatable cuff. In all cases, the CARDI-BP measurements followed a heartbeat-like pattern and were broadly in line with measurements from the cuff. In some cases, the curves exhibited low enough noise that we were able to pick out smaller-scale features such as the dicrotic notch.

As mentioned above, CARDI-BP data from each subject were filtered based on criteria of physically reasonable pulse pressure and strong radius-frequency correlation, and measurements were broken up into individual heartbeats to generate mean and standard deviation statistics (see Table 4.1). While this dataset was too small to generate statistically powerful conclusions regarding accuracy, we can make some general observations. Overall, mean CARDI-BP measurements lined up fairly well with BP measurements from the cuff. This alone represents a significant step forward in the NIBPM field. Every other NIBPM method is effectively guaranteed to produce a biologically normal range of BP values, as they are either directly calibrated against a cuff measurement or indirectly referenced to cuff measurements via a machine learning model. In contrast, the CARDI-BP method has no such inherent guarantee; Equation 2.25 could easily produce pressure measurements of 500 mmHg or even -50 mmHg if given a resonant frequency value that was far from
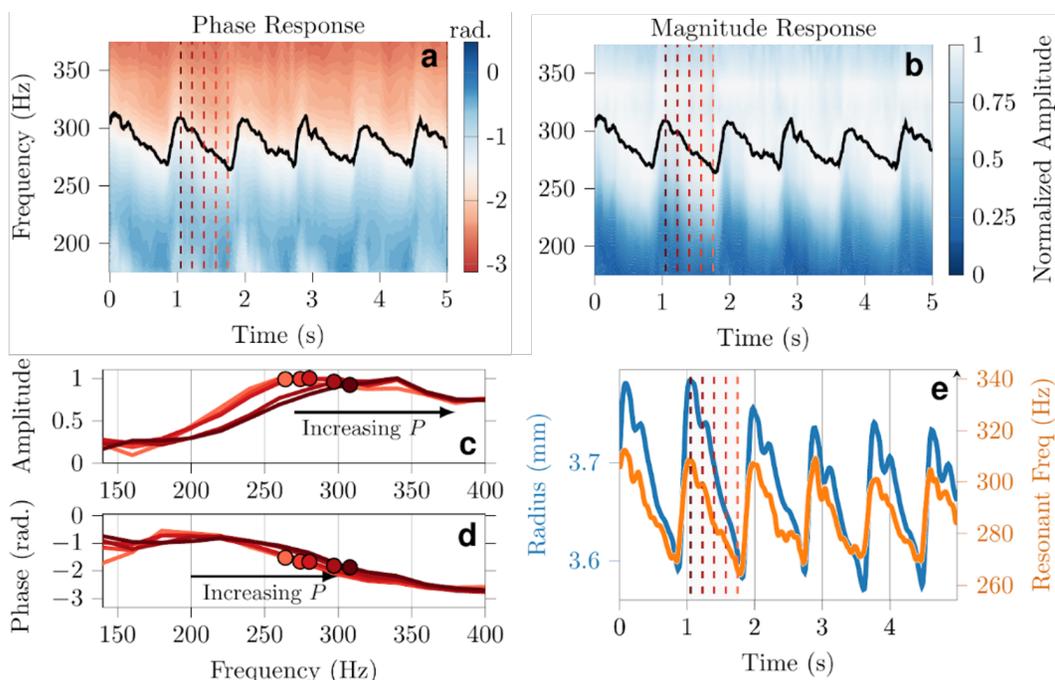
Figure 4.5: An example of resonant behavior observed *in vivo* in a carotid artery. a) The phase component of the spectrogram of differential wall velocity over time. The color map shows a gradient with a center which shifts in time with the heartbeat. The overlaid best fit frequency (black line) confirms that the resonant frequency is moving in a heartbeat pattern. b) The magnitude component of the same spectrogram from (a), showing the same type of synchronized shifts with heartbeat. c, d) Individual point-in-time curves of the magnitude and phase of frequency response; colors correspond to the times indicated by vertical dashed lines in (a) and (b). As $P$ increases, we see the magnitude and phase curves both shift to the right while maintaining roughly the same shape. e) Fitted resonance frequency plotted on top of measured arterial radius. This makes it clear that the frequency and radius are both changing synchronously with the heartbeat, as predicted by our physical model.

expectations. Thus, obtaining BP values roughly in line the cuff directly from our physical model, without the need for any empirical fudge factors, is a strong result that supports the potential viability of the CARDI-BP method.

There are also more specific observations we can make from Table 4.1. First, the number of heartbeats that passed our QC thresholds varied significantly from subject to subject. A primary reason for this was most likely the manual operation of the device. Finding an ultrasound probe position which yielded a clean, longitudinal image in the middle of the carotid artery while maintaining skin contact for the speakers was more challenging for some subjects compared to others. Even once a good spot was found, the ultrasound gel made it easy to slip off of this position.
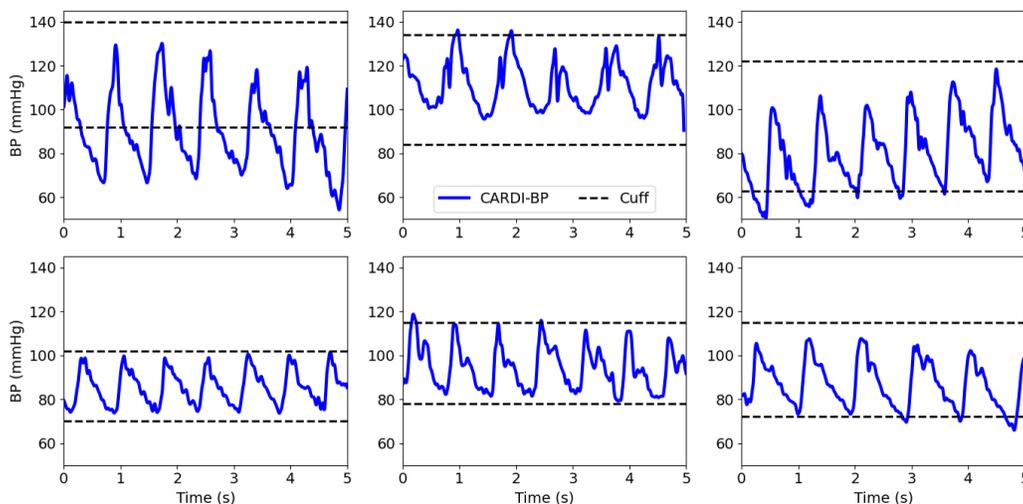
Figure 4.6: A selection of computed BP waveforms from subjects in the initial $N = 6$ feasibility study (solid lines), along with corresponding BP cuff measurements (dashed lines). These BP traces have roughly the expected shape and have systolic and diastolic pressures broadly in line with the cuff readings. In segments with relatively low noise we also have sufficient temporal resolution to pick out smaller-scale waveform features such as the dicrotic notch.

| Value | Method | Subj A | Subj B | Subj C | Subj D | Subj E | Subj F |
|---|---|---|---|---|---|---|---|
| No. | CARDI-BP | 83 | 22 | 145 | 22 | 55 | 141 |
| Samples | Cuff | 6 | 6 | 6 | 6 | 6 | 6 |
| DIA | CARDI-BP | 68 ± 10.5 | 90 ± 3.6 | 63 ± 9.2 | 71 ± 2.9 | 74 ± 10.4 | 68 ± 4.1 |
| (mmHg) | Cuff | 90 ± 6.2 | 83 ± 3.2 | 58 ± 5.7 | 70 ± 6.1 | 74 ± 2.7 | 74 ± 5.0 |
| MAP | CARDI-BP | 86 ± 7.8 | 102 ± 3.6 | 79 ± 10.4 | 85 ± 4.0 | 85 ± 9.7 | 77 ± 4.1 |
| (mmHg) | Cuff | 101 ± 5.0 | 93 ± 3.3 | 73 ± 3.8 | 78 ± 5.3 | 82 ± 1.8 | 83 ± 3.8 |
| SYS | CARDI-BP | 105 ± 10.3 | 116 ± 5.2 | 99 ± 17.4 | 103 ± 15 | 98 ± 9.7 | 88 ± 4.9 |
| (mmHg) | Cuff | 140 ± 3.4 | 130 ± 4.9 | 121 ± 4.0 | 101 ± 4.0 | 110 ± 7.0 | 116 ± 7.7 |

Table 4.1: Summary statistics from the $N = 6$ initial feasibility study, comparing CARDI-BP to the cuff for measurements of diastolic BP (DIA), mean arterial pressure (MAP), and systolic BP (SYS).

Second, some subjects (particularly B, D, and F) had variance in CARDI-BP measurements that was similar to or even lower than variance in cuff measurements, but other subjects had relatively high CARDI-BP measurement variance. This indicated a need for better data QC procedures to identify reliable low-noise measurements, which will be discussed in section 4.4. Finally, while DIA and MAP mean measurements tended to be roughly the same between CARDI-BP and the cuff, SYS measurents from CARDI-BP tend to be systematically lower than those from the cuff. This was expected due to site-specific BP differences between the brachial and carotid arteries [91], and will also be discussed further in section 4.4.

**BP from Other Ateries**

In addition to the carotid artery, we also applied CARDI-BP to three more peripheral arteries on a single subject: the axillary (in the shoulder), brachial (in the bicep), and femoral (in the thigh). Snapshot of results are shown below in Figure 4.7. In all three arteries we were able to stimulate and measure a resonant frequency in the same manner as in the carotid. Because all measurement sites were roughly level with the heart (femoral measurements were taken with the subject lying down), we should expect to obtain BPs close to those of the cuff. Indeed, the systolic and diastolic values in Figure 4.7a,c,e align well with those of the BP cuff. We can also see in Figure 4.7b,d,f that as arterial radius got smaller, resonant frequency increased to compensate, in line with the physical model. This indicates that the CARDI-BP method is not limited only to the carotid artery and can be applied elsewhere around the body. The waveforms themselves are somewhat noisier than those obtained in the carotid due to a combination of smaller arterial dimensions and greater difficulty of probe positioning, particularly for the femoral. Future device implementations targeting these arteries would likely need a combination of a wearable form-factor and higher-frequency ultrasound to obtain more consistent measurements.

## 4.4 Clinical Study: Methods

The initial feasibility study convinced us that arterial resonance, and by extension BP, could be measured in human arteries. However, this study represented a small and fairly demographically homogeneous subject pool with an imperfect reference pressure from an inflatable cuff. To more rigorously assess CARDI-BP's accuracy in a broader set of subjects, we contracted with ARK Clinical Research (Tustin, CA) to conduct a 60-subject clinical study comparing CARDI-BP performance to a gold standard arterial catheter.

**Study Design and Subject Recruitment**

The research protocol was approved by Advarra Institutional Review Board (protocol #: Pro00063289) and all test subjects gave written informed consent for participation. The study design was a prospective observational feasibility study evaluating the correlation, accuracy, and precision of the CARDI-BP prototype device as compared to an arterial catheter. A brachial cuff was used to ensure the consistency of measurements from the arterial catheter. Appendix Figure A3.1 outlines the study population and inclusion/exclusion criteria. Healthy volunteers were recruited by Ark Clinical Research with attention to recruiting a diverse subset of the population
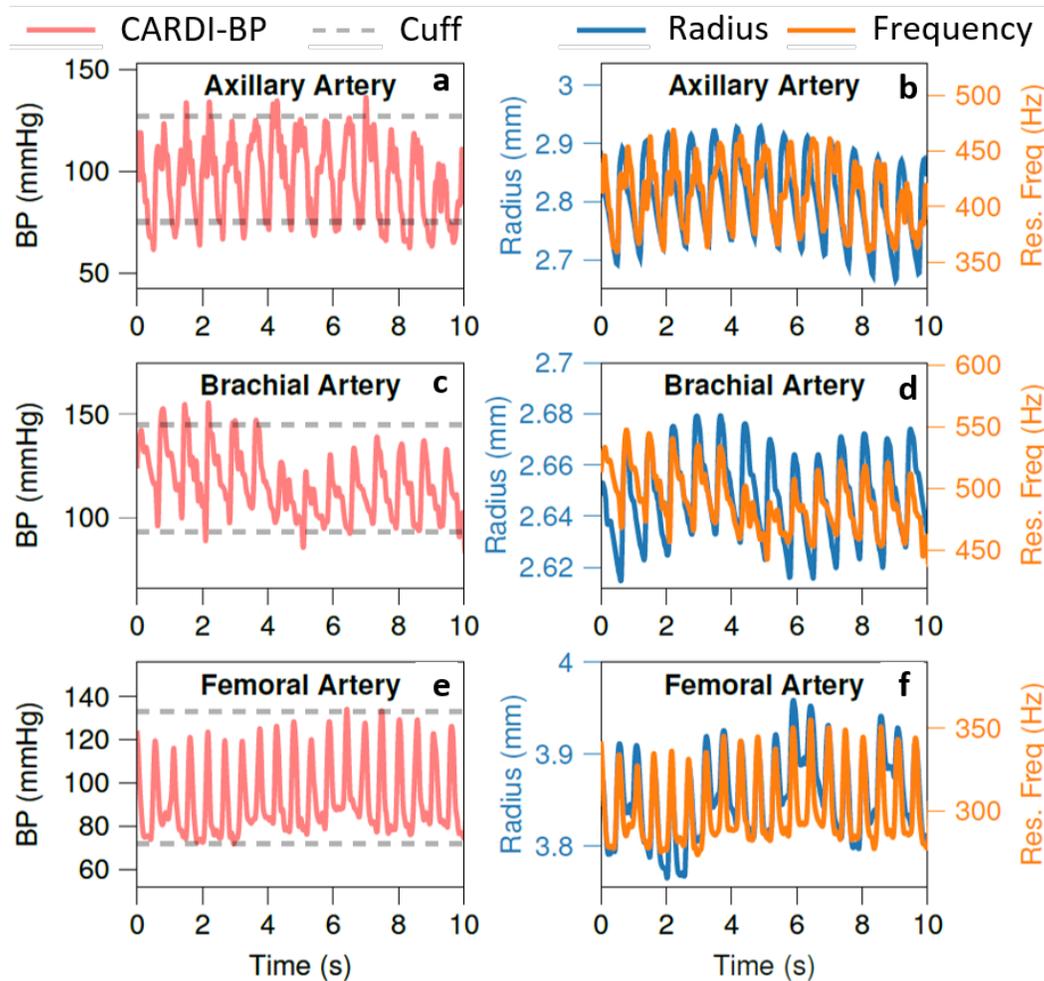
Figure 4.7: Example blood pressure (left) and resonant frequency/radius (right) measurements from three other sites: the axillary (shoulder) artery (a,b), the brachial (bicep) artery (c,d), and the femoral (thigh) artery (e, f). The CARDI-BP prototype device was able to detect resonance and measure BP values in general agreement with those of a cuff in all three arteries, indicating that the method is not just limited to the carotid.

with varying ages, ethnicities, genders, body mass indices (BMI), and underlying health conditions. Because the subjects were recruited from a generally healthy population, critically low and critically high BPs were not evaluated in this study. Sixty-eight subjects were screened and 60 subjects completed the study. A summary of subject demographics is included in Appendix Table A3.1.

**Data Collection**

Subjects were interviewed to assess their medical and surgical histories, allergies, demographics, height, weight, and body mass index (BMI). After subjects had

satisfied all inclusion and exclusion criteria, and signed informed consent, the subject had a 20-G arterial catheter (Arrow NA-04020-1A) inserted into the radial artery. The pressure line was flushed with normal saline, the transducer (Edwards TruWave PX260) was calibrated at the level of the right atrium, and the subject was allowed to choose a seated or reclined position which was kept consistent across scans. The approximate angle of reclination was recorded for each subject. After arterial catheter placement, arterial catheter data and all monitor outputs were recorded continuously throughout the study with a Philips Intellivue MP70 monitor with a M3001A measurement pack (Holland). An oscillometric cuff (Philips M1877A or Orantech BP-50BS) was placed on the contralateral arm to the arterial catheter. To ensure consistency of measurements, catheter diastolic values needed to be within 10 mmHg and catheter systolic values needed to be within 20 mmHg of the cuff measurement. Due to difficulty maintaining a difference of less than 10mmHg in diastolic pressures, the allowable limit of diastolic pressure difference was amended to 20mmHg partway through the study. Additional vital sign data were recorded, including oscillometric cuff, pulse rate, respirations, pulse oximetry, and oral temperature.

After initial setup, 15 one-minute scans were taken with the CARDI-BP prototype device from the subject's carotid artery ipsilateral to the arterial catheter (Figure 4.8). After 15 scans, all vital signs were recorded and then 15 additional scans were taken from the carotid artery for a total of 30 one-minute scans per subject. Vital sign data were then recorded again at the end of the study and the arterial catheter was removed. 16 scans were discarded due to failure of A-line data recording, leaving 1784 total scans available for analysis across all subjects. Researchers were blinded to the magnitude of CARDI-BP prototype outputs for the duration of the study; they could see waveform shape in order to assess whether quality data was being obtained, but could not see the actual BP values (see Figure 4.4). All raw ultrasound data from the CARDI-BP prototype was recorded for post-processing.

**Development and Validation Subjects**

Prior to this study, CARDI-BP had never been tested in a clinical setting across a range of blood pressures and physiologies. Thus, the analysis and quality control algorithms needed to be tuned during the course of the study in order to ensure correct identification of arterial walls for radius and thickness measurements and robust analysis of stimulus response for calculating resonant frequency. In order to avoid overtuning, the data were split into 40 "development" subjects and 20
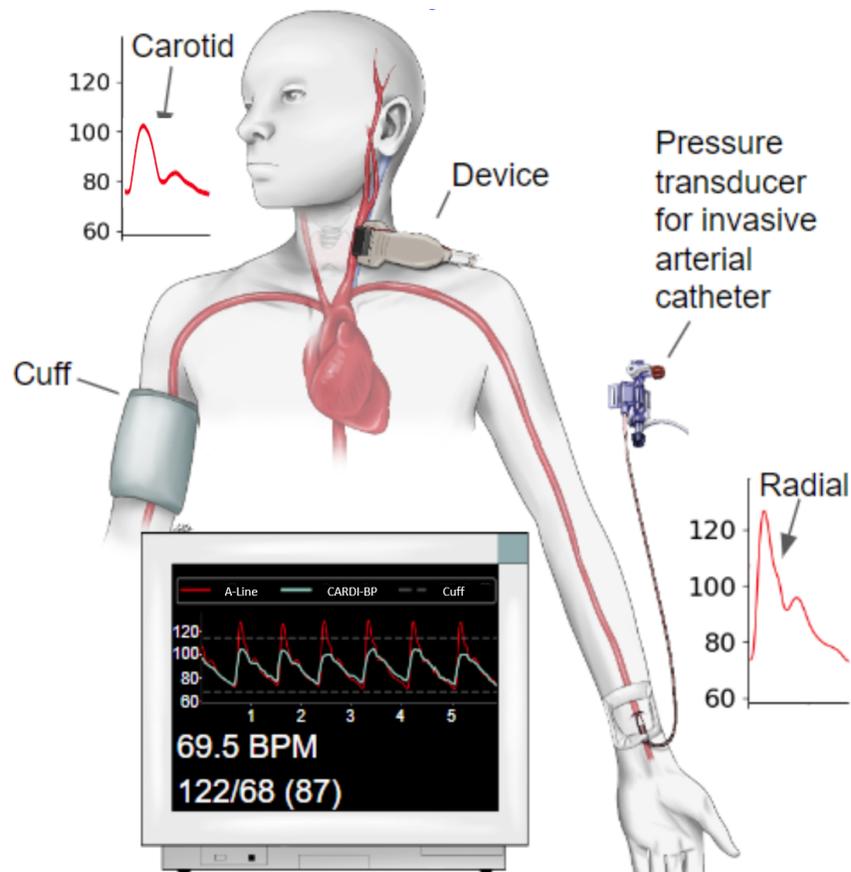
Figure 4.8: An illustration of the data capture setup for the clinical study. BP was simultaneously measured using the CARDI-BP prototype ("Device") on the carotid artery and an arterial catheter on the ipsalateral radial artery, and intermittent measurements were acquired using an oscillometric BP cuff on the contralateral brachial artery. Inset graphs show expected example BP curves from CARDI-BP and the A-line; note that CARDI-BP is expected to produce lower systolic peaks than the A-line, as systolic BP is systematically lower in the carotid artery compared to the radial. All vertical axes are in units of mmHg.

"validation" subjects. This split was determined before any data from validation subjects had been analyzed and was chosen based on demographic data to create roughly equal representation across demographic groups for the two datasets. All algorithm and quality control criteria tuning was performed solely on data from the development subjects. Once tuning was complete the final algorithm was applied to data from the validation subjects.

After data analysis, average CARDI-BP performance was found to be statistically equivalent between the development and validation subjects, indicating that we did

not overfit the training data through algorithm or QC tuning (see Appendix Table A3.2 for details). Thus, all results are presented for the full 60-subject cohort.

**Data Analysis and QC**

BP measurements from CARDI-BP were processed using standard methods from the literature that are consistent with those employed by vital sign monitors [90]. First, measurements were passed through an interquartile range filter and a 12 Hz lowpass filter. To convert continuous measurements into clinically relevant metrics of diastolic blood pressure (DBP), systolic blood pressure (SBP) , and mean arterial pressure (MAP), data were divided into non-overlapping time windows with a length of six seconds, rounded down to the nearest heartbeat interval. For each window, DBP and SBP were calculated as the average of peak minima and maxima, respectively, and MAP was calculated as the average of all BP values. A-line data were analyzed in the same way during the same time windows for comparison.

The current iteration of the device is sensitive to motion-induced operator error, including shifts due to operator fatigue, as well as test subject movement. As a result, signal was lost and these data were deemed unusable and excluded from the final analysis. These exclusions proceeded in the following steps:

1. For each scan, take the difference between the 90th and 10th percentiles of calculated radius. If this span is greater than 0.4mm, discard the scan.

2. For each scan, take the difference between the 90th and 10th percentiles of fitted resonant frequency. If this span is greater than 75 Hz, discard the entire scan.

3. Any individual measurements below 0 mmHg or above 250 mmHg are marked as invalid and cannot be included in a window [90] (this check is performed after our IQR and lowpass filtering)

4. Search data in the scan using a rolling window. Reject the window if MAP is outside of the range (40, 160) or pulse pressure is outside of the range (20, 150) [90]. If a window is rejected, roll forward by 0.1s and check again.

5. Also reject window if the correlation between radius and resonant frequency is below 0.5

6. Any window that makes it this far is accepted, and the rolling window search skips forward so that windows cannot overlap. All accepted windows are included in the final results.

The first two criteria filtered for scan stability. The thresholds of 0.4 mm and 75 Hz were chosen as being higher than the span of radius or frequency seen in any high-quality training scan. Breaking either of these thresholds indicated that there was a significant disruption for a large portion of the scan (likely due to subject or operator motion), leading to loss of arterial wall imaging, resonance signal, or both. The most robust way to deal with such a disruption was to discard the entire scan. Criteria 3 and 4 were taken from the literature as thresholds commonly used to filter out unphysical BP data from healthy individuals. Finally, criterion 5 was used to filter out minor scan disruptions where either arterial wall imaging or resonance signal was briefly lost.

**Site-Specific BP Differences**

One limitation of this study is that our CARDI-BP prototype device measurements were obtained from a central (carotid) artery, while cuff measurements were obtained from a somewhat peripheral (brachial) artery and A-Line measurements were obtained from a very peripheral (radial) artery. Prior studies have shown that BP changes as the pressure pulse wave moves through the arterial tree, meaning that even if CARDI-BP was functioning perfectly it would not be expected to measure the exact same BP as the A-line [91]–[95]. The largest effect comes from pressure wave reflections, which are more prominent in peripheral arteries due to their closer proximity to terminal capillary beds. These reflections tend to amplify systolic pressure, with the upshot that SBP becomes higher and peak features become more pronounced as one moves from central to peripheral arteries. This peripheral systolic amplification is subject-specific and cannot be fully corrected with a simple population-wide scaling factor. On the other hand, both DBP and MAP have been found to stay relatively consistent between central and peripheral arteries. Due to these effects, when comparing BP traces between CARDI-BP and the A-Line we expected that CARDI-BP curves would have significantly lower SBP values, similar DBP and MAP values, and less pronounced waveform features than A-line curves (as illustrated in the inset of Figure 4.8). Thus, our discussion of CARDI-BP performance will primarily focus on comparisons of DBP and MAP, with SBP understood to be non-representative of performance due to large site-specific differences.

Another factor to consider is hydrostatic offset due to gravity. The pressure transducer for the arterial catheter was aligned to the level of the subject's right atrium (in line with clinical practice). However, because the subjects were not laying flat, the CARDI-BP measurement site was higher than the A-line transducer, so the force of gravity over this vertical separation would be expected to lower true BP at the CARDI-BP measurement site. While precise measurements of heart to neck distance were not taken during the study, standard clinical practice assumes an average distance of 7cm when the subject is upright [96], corresponding to a pressure offset of roughly 5.2 mmHg. This pressure was added to all CARDI-BP outputs to correct for hydrostatic effects, scaled by a factor of $\sin \theta$ for subjects who were reclined.

## 4.5    Clinical Study: Results and Discussion

**Overall Performance**

The data QC procedures described above yielded a total of 6794 data windows, including at least one window from 59 out of 60 subjects. Each data window was converted into values of DIA, MAP, and SYS within that timeframe for both CARDI-BP and the A-line. These window values, along with linear regression results, are shown in Figure 4.9a-c. We can see that see that CARDI-BP predictions align fairly well with A-line measurements, showing correlations (as measured by Pearson's $r$) of $r > 0.6$ for all metrics and regression lines close to the 1:1 line for DBP and MAP. Figure 4.9d-f show the same statistics, but grouped by averaging all window BP values for each subject (yielding $N = 59$ points). Performing this averaging decreased overall CARDI-BP measurement noise and increased correlations by roughly 0.07 for each metric.

In isolation, it is difficult to say whether or not these statistics constitute "good" BP measurement performance. A useful point of comparison is the BP cuff measurements that were taken for all subjects, which represent the current standard of BP measurement in hospitals outside of critical care situations. Table 4.2 shows correlation and linear regression slope statistics for CARDI-BP and the cuff, along with two other metrics commonly used to assess the clinical performance of BP devices; mean difference between the measurement and the A-line (accuracy) and standard deviation of differences beween the measurement and the A-line (precision). The latest approved standards for BP measurement device evaluation defined by the International Organization for Standardization, ISO 81060-2:2018 [82], set clinically acceptable limits for accuracy and precision at ±5mmHg and ±8mmHg, respectively.
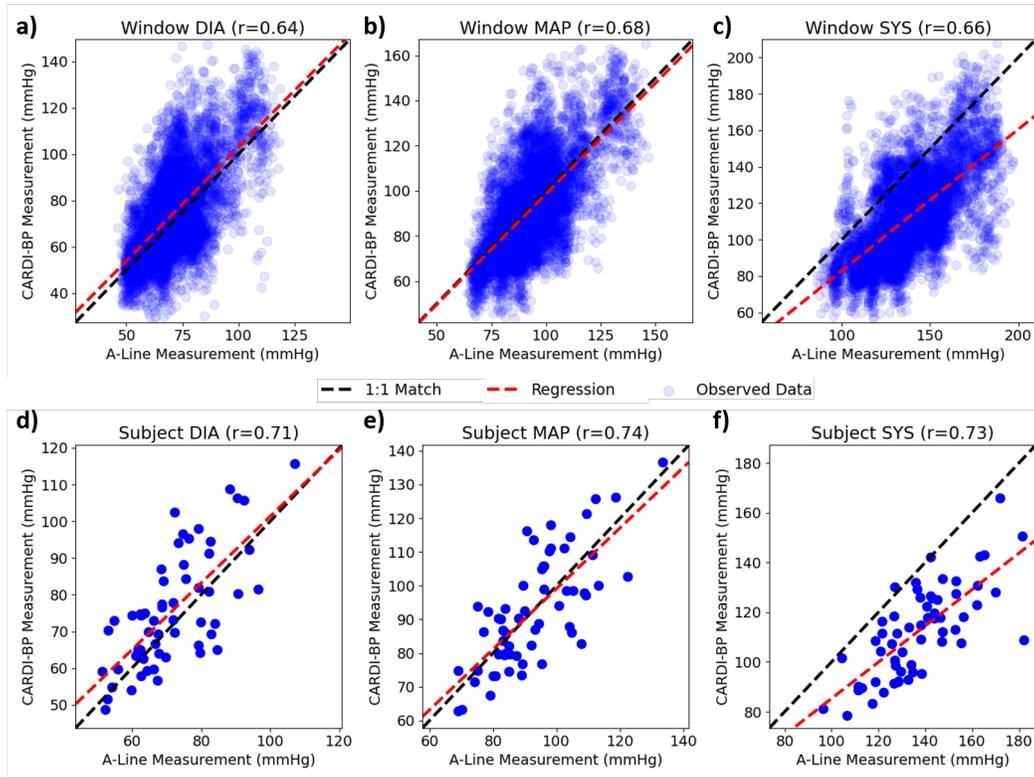
Figure 4.9: Plots showing overall correlation results for CARDI-BP measurements in the clinical study. The top row shows results for a) diastolic pressure, b) mean pressure, and c) systolic pressure across all individual data windows, as well as a linear regression. For both DIA and MAP the regression line lies close to the 1:1 line, and in every case we get a fairly strong correlation of $r > 0.6$. The bottom row shows the same results, but with all windows for each subject averaged together to generate one data point per subject. This averaging suppresses noise and increases correlations to $r > 0.7$.

The results in Table 4.2 show that CARDI-BP measurements met clinical standards for accuracy in DBP and MAP measurements and were not far off of clinical standards for precision, especially when subject averaging was applied. For comparison, the BP cuff did meet clinical standards for precision by a narrow margin and failed to meet the standard for MAP accuracy. Our measured values for cuff performance are in line with those found by a meta-analysis of cuff performance in Dankel, Kang, Abe, *et al.* [13], which found a mean and standard deviation of errors in cuff DBP measurements of 6.2mmHg and 8.0mmHg, respectively; this indicates that the observed cuff errors were not a product of incorrect cuff usage in our study. CARDI-BP also compared favorably to the cuff in terms of regression slope; in particular, the cuff tended to overestimate low DBP values and underestimate high DBP values (a trend also reflected in prior literature [13]), while CARDI-BP showed no such trend

| Device | Measurement | Correlation | Slope | Mean Diff (mmHg) | Std. Dev. of Diff (mmHg) |
|---|---|---|---|---|---|
| **CARDI-BP (Window)** | DIA | 0.64 | 0.98 | 3.49 | 15.33 |
| | MAP | 0.68 | 0.98 | -1.22 | 15.81 |
| | SYS | 0.66 | 0.78 | -25.50 | 19.53 |
| **CARDI-BP (Subject)** | DIA | 0.71 | 0.91 | 3.96 | 10.89 |
| | MAP | 0.74 | 0.90 | -0.14 | 11.17 |
| | SYS | 0.73 | 0.73 | -24.77 | 13.43 |
| **Cuff** | DIA | 0.81 | 0.79 | 3.3 | 7.7 |
| | MAP | 0.85 | 0.73 | -7.4 | 7.4 |
| | SYS | 0.84 | 0.68 | -15.1 | 10.2 |

Table 4.2: An overall statistical performance summary for CARBI-BP (both across all windows and across subject-averaged values), as well as equivalent performance statistics from the BP cuff for comparison. While CARDI-BP does show somewhat lower correlation and higher variance than the cuff, it is not far off, especially when subject averaging is applied. In some statistics, such as slope of the linear fit and bias of MAP measurements, CARDI-BP actually out-performs the cuff.

with a slope very close to 1. Appendix Table A1.2 provides a further comparison of CARDI-BP performance to other NIBPM methods in the literature which have been evaluated against an A-Line reference. This shows that CARDI-BP's precision was better than that of some commercially available NIBPM devices, particularly for devices which also attempted to provide calibration-free measurements. All of these comparisons indicate that, while CARDI-BP is not yet ready for clinical deployment, its performance is quite promising for the first ever test of a novel and unique BP measurement modality.

**Continuous BP Traces**

Aside from overall statistics, another important indicator of CARDI-BP's performance is its ability to measure continuous BP traces rather than simply getting the maximum and minimum. As mentioned above, we expected the height and shape of continuous BP waveforms to differ significantly between the CARDI-BP and A-line measurement locations, so we could not directly compare them to assess concordance. However, there are other ways to indirectly look at continuous BP performance. One example is respiration signal. It has long been established that breathing causes sinusoidal variations in blood pressure [97], which are detectable with A-lines but averaged over and obscured by BP cuffs. Figure 4.10 shows example 30-second BP traces from the A-line and CARDI-BP, along with an overlay of beat-to-beat mean arterial pressure calculated from each. Both the A-line and

CARDI-BP show synchronous oscillations in MAP at a rate consistent with normal resting respiration (period ≈ 12s), indicating that CARDI-BP was able to measure these subtle continuous BP changes.
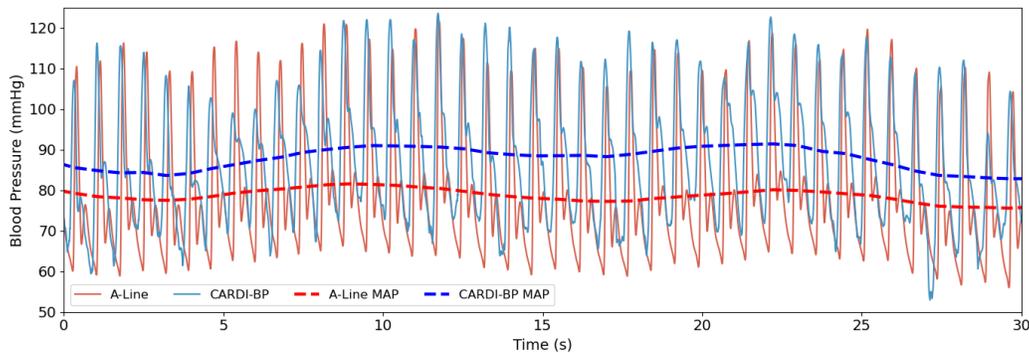


Figure 4.10: Aligned 30s traces of continuous BP from the A-line and CARDI-BP. We see that CARDI-BP is capable of generating stable outputs when the operator and subject remain still. The graph also shows continuous MAP measurements from each device, calculated once per heartbeat. MAP in the A-line shows slow sinusoidal variation at the subject's respiration rate, a pattern expected from prior literature. This pattern is also evident in the CARDI-BP readings, indicating that the method is picking up on BP changes over time.

A much more clinically valuable use of continuous BP information is the identification of waveform features indicative of various diseases. The subject pool recruited for the clinical study was generally healthy, and almost all showed typical arterial BP waveforms. However, the A-line readings for one subject showed a distinctive anacrotic notch, which is a sloped bump just before the systolic peak. This shape difference is evident in Figure 4.11a, which compares an example A-line waveform from this subject with a more standard A-line waveform from another subject without an anacrotic notch. This anacrotic notch is often characteristic of aortic stenosis (AS)[6], i.e., a thickening and narrowing of the aortic valve which prevents blood from flowing normally. While the subject did not report known AS and a definitive diagnosis would require follow-up examination, a consulted physician confirmed that this A-line waveform was a likely signature of moderate AS. As shown in Figure 4.11b, this distinctive anacrotic notch shape was also distinctly observable in the corresponding CARDI-BP waveform. This indicates that CARDI-BP has sufficient time resolution in measuring continuous BP waveforms to allow for the diagnosis of vascular diseases which could not be identified from cuff readings alone.
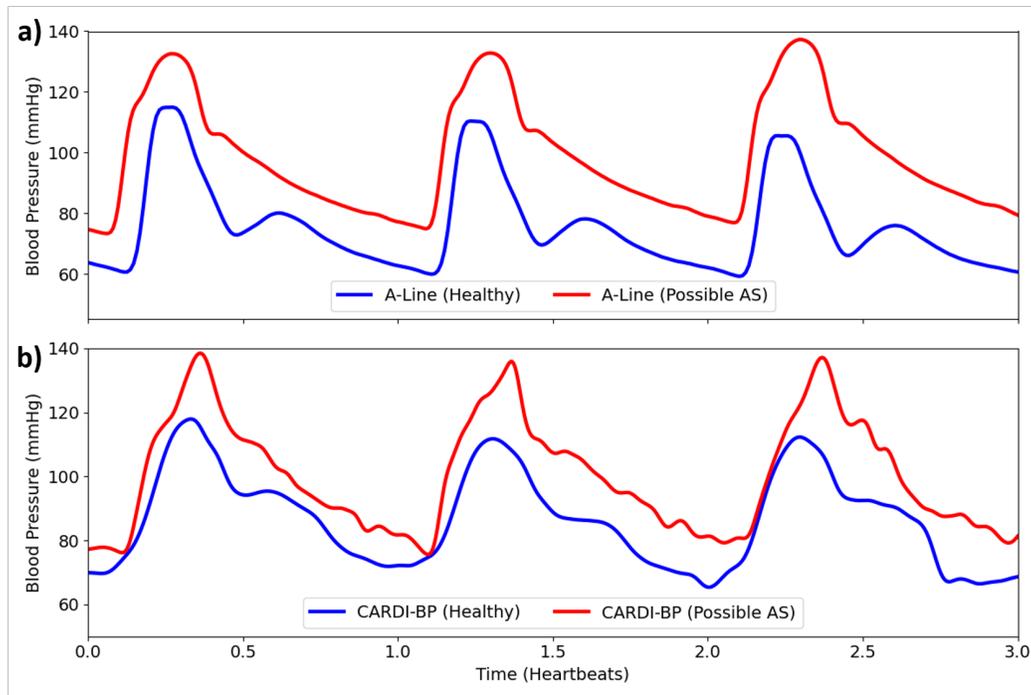
Figure 4.11: Evidence of a cardiovascular disease from the BP waveform shape. a) Example A-line traces from two different subjects, one showing a typical healthy waveform and one showing possible aortic stenosis (AS). A primary indicator of aortic stenosis is a prominent anacrotic notch (a sloped bump just before the systolic peak), which is evident in the red waveform. b) CARDI-BP traces for the same subjects in the same time windows. Just as in the A-line, CARDI-BP waveforms show an anacrotic notch in the red waveform but not in the blue waveform. This indicates that CARDI-BP is capable of measuring subtle features in BP waveform shape, not just maximum and minimum pressure values.

**Availability**

Another factor to consider is how often we were able to successfully obtain CARDI-BP measurements that passed QC, which is often referred to in clinical circles as "availability." Overall, we were able to successfully stimulate resonance and measure blood pressure in 59 out of 60 tested subjects. These subjects covered a range of demographics (see Appendix Table A3.1), indicating that resonance-based measurement is viable across a variety of anatomies. It is worth noting that this 59/60 success rate significantly exceeded that of arterial catheter placement, which failed in 8 out of the 68 screended subjects (see Appendix Figure A3.1). For the subject which failed to pass QC, it was observed that they had a very pronounced jugular vein immediately overlaying their carotid artery, and pulsation in the vein visibly pressed on and moved the artery. While we cannot be sure, it seems likely that

this strong coupling of the artery to a neighboring vessel interfered with resonance measurements.

In addition to subject availability we can consider total availability, or the fraction of all collected data that made it into final analysis. Figure 4.12 presents a flowchart of the QC steps and the fraction of data rejected at each step. We can see that almost half of all scans were rejected in their entirety due to instability. This reflects the difficulty of maintaining the measurement probe in a nearly constant position and orientation for a full 60s scan; subject motion or operator fatigue could easily throw off positioning, especially since ultrasound gel made the probe-skin interface slippery. Out of the remaining scans, on average about 30% of the scan was rejected due to locally poor measurement quality; this was likely due to minor motions, such as the subject swallowing or small shifts by the operator. In total, 35.6% of all data originally collected made it into the final results. Future iterations could significantly improve this rate by removing the need for manual operation.
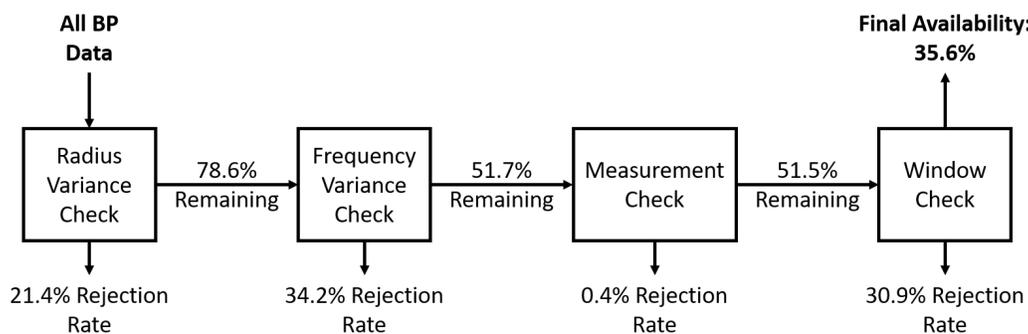


Figure 4.12: A depiction of the data QC chain for the clinical study and how much data was rejected at each step. QC steps 4 and 5 from the text are grouped into the "Window Check" box since both checks wer performed simultaneously during rolling window selection. Overall, 35.6% of all collected data was included in the final results. Almost 50% of data were rejected due to insufficient scan stability, as it was difficult for both the subject and operator to stay sufficiently still for 60 seconds. If a scan passed the variance checks, on average about 70% of that scan was included in the final data.

**Sources of Error**

The sources of error in our measurements can generally be broken down into 3 possible categories:

1. Systematic error in the underlying physical model and/or data processing which affected measurements from all subjects.

2. Systematic error due to subject-specific factors, such as anatomy or positioning, which affected all measurements for a given subject.

3. Various forms of measurement noise which led to varying errors across measurements on the same subject.

We can approximately tease apart the relative contributions of these error types via averaging. Figure 4.13 shows the window statistics from Figure 4.9a-c with overlaid bin averages. The bin for each point was determined based on A-line values with boundaries at even 4mmHg spacing, and the y-values for each bin represent the average of all CARDI-BP measurements in that bin. For both DBP and MAP these bin averages consistently land close to the 1:1 line, with a correlation of $r = 0.98$ for both and no obvious pattern in deviation from the linear fit. Even for systolic measurements, which we expect to be noiser due to variable site-specific differences, the binned averages still achieve a correlation of $r = 0.96$. This indicates that the underlying physical model and data processing algorithms were very accurate when averaged across all subjects, so type 1 error was likely not a significant contributor to our overall variability.
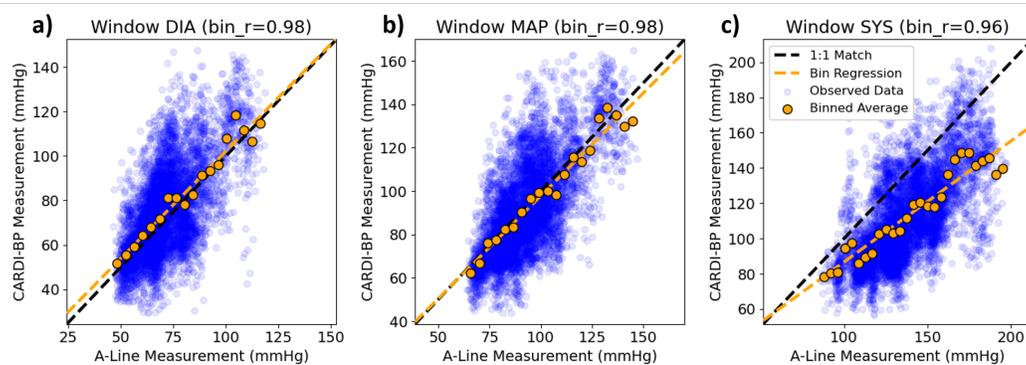


Figure 4.13: Plots with the same window statistics from Figure 4.9, but with binned averages and linear regressions of these binned averages overlaid. Data were divided into even 4mmHg-wide bins based on A-line pressures, and each point represents the aveage of all CARDI-BP measurements in that bin. The fact that the binned averages are well aligned along a linear regression, without an obvious nonlinear pattern, indicates that our underlying physical model and measurement techniques are, on average, working quite well.

Error types 2 and 3 can be roughly separated by comparing variability in window results to variability in subject-averaged results. If we assume that these errors were independent and added in quadrature, the results in Table 4.2 indicate that

both error types had roughly equal contribution. Type 3 errors are relatively easy to explain due to noise in our underlying measurements: artery dimensions and frequency. Because our imaging probe had just 32 active elements and could only transmit from one element at a time, imaging quality was significantly worse than what could be obtained from commercial instruments which can use hundreds of active elements with complex beamforming patterns; this inherently increased noise in radius and thickness measurements. Radius measurements were also sensitive to slight shifts in probe or subject positioning, which could cause the artery to be imaged along a chord rather than along its true diameter. Future device iterations could mitigate these problems using a wearable form factor and a better ultrasound imaging array. For frequency, we used a constant wide stimulus bandwidth of 140-440 Hz for all measurements in order to ensure that all subjects were covered. This spread out stimulus power over many frequency bins, hurting the SNR of any one bin. Future implementations could dynamically narrow the stimulus bandwidth once an approximate resonance is identified, increasing SNR in the relevant range. More sophisticated speakers could also potentially focus the stimulus energy onto the artery rather than broadly insonating the whole tissue volume, which would be particularly helpful for subjects with deeper arteries.

A portion of the type 2 error also came from errors in radius and frequency measurements. Some subjects had very acoustically bright anatomical features just outside of their arterial walls which were often confused with the walls themselves by dimension extraction algorithms, leading to overestimation of arterial radius. This could be addressed in future iterations via better imaging and image analysis algorithms. For resonant frequency, we fit the frequency response with a generalized 2-pole linear response function rather than using a specific functional form derived from the physical model. This caused some degree of error in true resonance estimation, which could have varied by subject based on factors like degree of damping. Further work in biophysical analysis or computational modeling could lead to more accurate resonance analysis. There could have also been some error in the A-line reference due to factors like improper transducer leveling or greater-than-expected site-specific differences in DBP or MAP. Future experimental designs could remove these factors by taking CARDI-BP and reference measurements at the same site.

The trickiest potential source of type 2 error is subject-specific deviation from the physical model due to the underlying anatomy. As described in Chapter 2, our physical model assumes that the artery is an infinitely long cylinder with constant radius,

thickness, and stiffness. Deviations from this ideal, such as tapering, branching, bending, uneven wall thickness and stiffness due to plaque buildup, etc. could all lead to varying degrees of error in model BP outputs. Future work could predict the relative severity of these various non-idealities via computational modeling, which could also potentially lead to empirical corrections for these effects. Resonant behavior could also be affected by neighboring anatomical structures that impinge upon the arterial wall, as was likely the case for the one subject that had no data passing QC. Dealing with such issues could potentially require selecting a different measurement site; fortunately, our initial data suggest that CARDI-BP is applicable to a variety of arteries, so many options should be available.

**Effect of Tuning QC Criteria**

Another way to examine our experimental error is to tune our QC to be more or less permissive and see how it affects our various metrics. The most obvious parameter to adjust is the radius/frequency correlation cutoff, which effectively determined how much type 3 noise was allowable in a measurement window. Figure 4.14 below shows correlation, bias, variance, and number of windows passing QC as the correlation cutoff was varied from 0.05 up to 0.9. We can see that increasing the stringency of this criterion did indeed increase correlation and decrease variance, at the cost of fewer windows of "valid" data making it through. The value of 0.5 used to compile the results presented above seems to represent a reasonable middle ground between performance and availability.

An interesting effect to observe is that tightening radius/frequency correlation QC slightly decreased our bias for both DBP and MAP, indicating that CARDI-BP tended to produce slightly lower BP values for less noisy windows. This could plausibly be explained by bright features next to arterial walls. These features tend to increase noise in radius measurement, as the detected "center" point of the wall may vary between the true wall and the neighboring feature. This type of noise can only increase radius values, and therefore increase measured BP. Even with very tight QC, both DIA and MAP biases still fall within the ±5mmHg ISO limit, indicating that our accuracy was not an accidental by-product of one-sided noise.

Figure 4.15 presents the same results as Figure 4.14, but averaged into per-subject values. In contrast to the window performance, both correlation and variance remain relatively flat for most of the QC range before getting sharply worse at the very tight end of the QC spectrum past a cutoff of 0.8. This flatness indicates that,
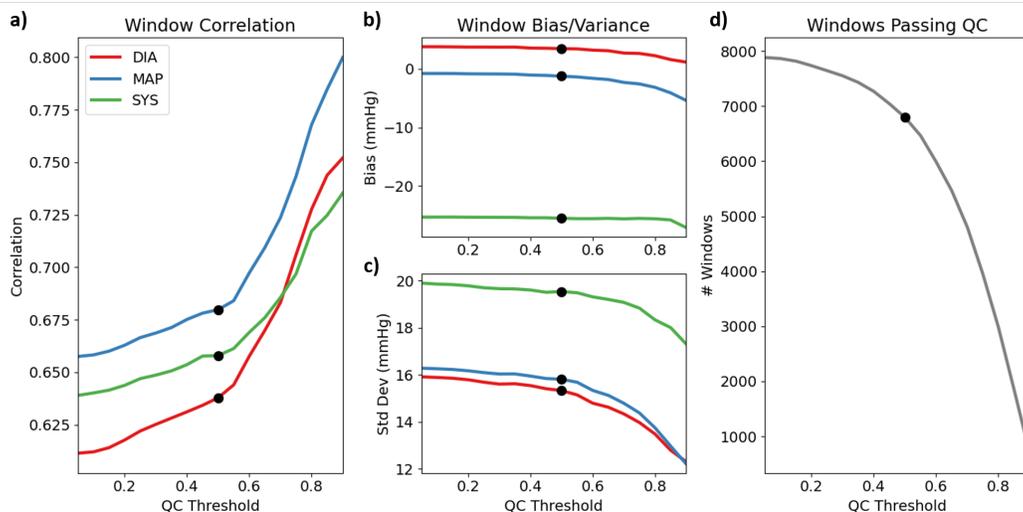
Figure 4.14: Plots of various statistics for CARDI-BP measurements of DIA, MAP, and SYS compared to the A-line as the radius/frequency correlation QC threshold is varied from 0.05 (very permissive) to 0.9 (very strict). Presented statistics are a) Correlation, b) Bias (average of differences betwen CARDI-BP and A-Line), c) Variance (standard deviation of differences betwen CARDI-BP and A-Line), and d) The number of windows passing QC. Black dots represent the QC threshold chosen for results presented above. We can see that tightening QC leads to significant improvement in correlation and variance, at the cost of fewer passing windows.

as expected, the radius-frequency correlation QC suppressed type 3 error but did not substantially affect type 2 error. The drastic shift at the tight end is likely a product of small-number statistics rather than a real effect, as indicated by the fact that by-subject performance for correlation and variance actually gets worse than by-window performance.

## 4.6 Conclusion

Despite the inherent complexities of human physiology and the assumption of a simple geometric configuration, the CARDI-BP model still produced results that were consistent with gold-standard arterial catheter measurements. The strength of this method lies in the small number of observables required: arterial radius, thickness, and resonant frequency. The accuracy of the results generated from these observables over a varied set of subjects suggests the independence of the method to demographics and physiology, as the pressure curves were created utilizing only information collected by the device in real-time, without cuff information, prior training data, or any input of demographic information. Future improvements to device accuracy to meet clinical standards will likely come from a combination
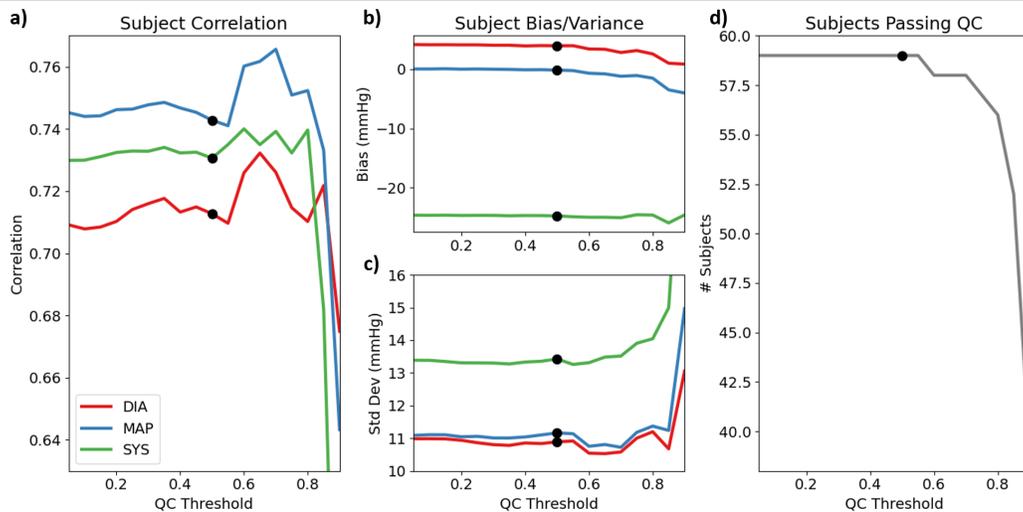
Figure 4.15: A replication of plots from Figure 4.14, but with BP values averaged for each subject. In contrast to the window statistics, subject statistics do not show any consistent improvement as QC is tightened, and indeed sharply worsen at the very strict end of the spectrum. This lack of improvement indicates that this QC parameter is affecting type 3 but not type 2 error. The sharp drop-off at the end is likely an artifact of small-number statistics.

of two factors; integration into a wearable form factor with more sophisticated ultrasound imaging, and integration of blood velocity measurements.

The most important factor for improving clinical usability as well as accuracy is integration into a wearable form factor. The manual operation required by the current prototype iteration introduced error in radius measurements, and continuous manual operation would not be feasible in a clinical setting even if the radius issue was resolved. A logical next iteration would be integration of the system into a wearable strap which could be wrapped around an arm, leg, or wrist. A combination of more stable imaging from a fixed probe along with improvements to image quality from a better ultrasound array should significantly improve the consistency and accuracy of radius measurements. Further in the future, the CARDI-BP method could be combined with recently described advancements that allow the incorporation of an ultrasound imaging array into an adhesive patch [49], [50]. Such a form factor would enable wearable measurement of central BP in the carotid artery, which is not measurable at all using a BP cuff.

Another significant improvement could come from integration of blood velocity measurements, which are commonly performed using ultrasound arrays capable of beam steering. It has been shown that a combination of blood velocity and arterial

radius measurements can be used to calculate pulse pressure in human arteries via the QA relationship (equation 2.1)[46]. This would provide an orthogonal measurement of pressure that is independent of resonance characteristics, which would make our system overspecified; this could allow for error correction if the pulse pressures generated by CARDI-BP and the QA method disagree. Importantly, the only assumed parameter in equation 2.1 is blood density, which is very consistently known (see section 2.6). Thus, comparing CARDI-BP and QA pulse pressures could allow us to correct for inaccuracies in the other assumed parameters underlying CARDI-BP, namely the densities of the arterial wall and its surrounding medium and the Poisson's ratio of the arterial wall.

Integration of CARDI-BP technology into a product will provide significant clinical benefits. In critical care situations where A-lines are currently deployed, replacement of these catheters with a robust, non-invasive monitoring solution would carry the same benefits of immediate identification of cardiac crises without the risks, difficulties, and costs associated with an invasive procedure. However, the greatest benefits would come from expanding continuous BP measurement beyond critical care units, making it available to every patient in a hospital and even those visiting a primary care physician. The analysis of continuous BP waveforms would allow for the early diagnosis of a variety of conditions that cannot be identified from a cuff measurement alone, such as arterial valve stiffening (aortic stenosis), backflow (aortic regurgitation), or thickening of heart muscles (hypertrophic obstructive cardiomyopathy) [6], [98]. CARDI-BP also allows for the direct measurement of central blood pressure, which has been found to have greater clinical value than peripheral BP for the diagnosis of a variety of conditions [99]. Central BP cannot be accessed at all with cuffs, and it is rarely measured even with A-lines due to increased risks from catheterization on more central arteries. CARDI-BP will drastically expand access to non-invasive and continuous central BP monitoring, enabling early outpatient diagnosis of a variety of diseases and improving the overall standard of care for our cardiovascular system.

*Chapter 5*

# ARTERIAL STIFFNESS MEASUREMENT

## 5.1  Background

In addition to blood pressure, the CARDI-BP method produces another measure of arterial dynamics: the circumferential Young's modulus of the arterial wall, more commonly referred to as arterial stiffness. Young people typically have low arterial stiffness, and this stiffness tends to increase naturally as arteries age. If a subject presents an abnormally high arterial stiffness for their age it may be a warning sign of various pathologies. While elevated blood pressure has long been recognized as an indicator of various health risks, arterial stiffness has more recently gained recognition as a valuable prognostic indicator in its own right [100]. In the last 20 years, stiffness measurements of central arteries have been shown to be an indicator for risk of coronary artery disease [101], stroke [102], atherosclerosis [103], vascular damage [104], and chronic kidney disease [105]. As commercial stiffness assessment devices become more widely available, the breadth of these stiffness-related clinical indications will likely continue to expand.

Modern measurements of arterial stiffness almost all rely on the surrogate parameter of pulse wave velocity, or PWV, which measures how quickly a blood pressure pulse moves down the length of the artery [106], [107]. PWV is directly related to arterial stiffness via the Moens-Korteweg equation [53], and it is relatively easy to measure using cheap sensor technology such as ECGs and inflatable cuffs. As a result, there are over a dozen commercial devices which measure PWV in various ways [107]. However, these devices come with some key limitations. Common implementations measure the difference in pulse arrival time between distant sites on the body (such as the carotid and femoral arteries) and assume a constant PWV for the entire path of the pulse wave, ignoring local variation. Even when measuring a single artery these methods can only calculate an averaged PWV over the cardiac cycle, even though arterial stiffness (and thus PWV) is known to vary substantially between systole and diastole [66], [67]. Both of these averaging assumptions have the potential to obscure important information.

Multiple recent methods have been proposed to generate local, continuous measurements of PWV or stiffness [107]. The most prominent is shear wave elastography

(SWE), which uses high intensity focused ultrasound (HIFU) to perturb a section of the arterial wall and generate shear waves down its length [1]. The shear waves are tracked using ultrafast ultrasound imaging at a rate of up to 10 kHz in order to determine their velocity; this can be repeated quickly enough to track PWV as it changes throughout the cardiac cycle. The main limitation of SWE is that it requires expensive, specialized ultrasound equipment capable of both HIFU and ultrafast imaging, limiting its application mainly to research centers. Furthermore, HIFU creates safety risks not typically present with ultrasound, as turning the intensity too high can potentially cause tissue damage [73].

The closest method of stiffness measurement to CARDI-BP is that of Khamdaeng, Luo, Vappou, *et al.* [67], which directly measures circumferential arterial stiffness as a function of BP and radius. Their technique measures continuous changes in arterial radius using ultrasound while simultaneously measuring continuous blood pressure at a neighboring site using an applanation tonometer. In contrast to SWE, this method requires neither HIFU nor ultrafast imaging, making it viable on standard ultrasound systems. However, in addition to ultrasound it also relies on a separate tonometer, which in turn must be calibrated against an inflatable cuff. This extra equipment makes the method inconvenient to apply and introduces significant potential sources of error, including inaccuracies in the tonometer and cuff and site-specific BP differences between the brachial artery and the artery of interest [108].

The CARDI-BP method for BP and stiffness measurement combines the best of both of these methods; it provides a direct, local, real-time measurement of arterial stiffness in a self-contained device without relying on high-end ultrasound equipment. This chapter will begin by presenting an analysis of the reliability of this stiffness estimate, and will then present and discuss stiffness measurements obtained from the clinical study.

## 5.2   Convergence of the Stiffness Estimate

As described in Chapter 2.3, stiffness is not measured directly as an observed parameter. Instead, we define a recursive relationship between blood pressure and stiffness and use Gauss-Seidel iteration to converge on an solution which satisfies the physical equations for both parameters. An important consideration for this approach is that there is no guarantee that Gauss-Seidel iteration will converge to a unique solution for an arbitrary system of equations. This is particularly relevant because the true Young's modulus of the carotid artery may vary by at least an order

of magnitude between a young, healthy artery at diastole and an atherosclerotic artery at systole [67]; thus, we cannot guarantee that a constant initial condition for stiffness will be close to the final answer. Here we show that, despite this wide variation, we do obtain robust convergence to a unique solution from any reasonable initial conditions.

**Empirical Approach**

To do this, we first performed an empirical analysis of convergence on data gathered from the carotid arteries of the 6 initial subjects in the *in vivo* feasibility study (see Chapter 4). Data was randomly sampled at 600 time steps from these subjects (100 from each subject). Prior work has shown that circumferential Young's modulus for the carotid artery varies from 0.1 MPa to 1 MPa in healthy adult subjects [67]. To account for potential variations due to age or pathologies, we extended our analysis by a full order of magnitude in either direction, starting the iterative solver with seven initial values for Young's modulus ranging from 0.01 MPa to 10 MPa in geometric steps of $\sqrt{10}$. Gauss-Seidel iteration was performed for 5 steps from each starting value, and the seven final results for each of blood pressure ($P$) and Young's modulus ($E$) were compiled to compute coefficients of variation (CV, defined as the standard deviation divided by mean) for each sampled time step. Results for all 6 subjects are outlined in Table 5.1; the median $CV$ was less than 0.11% for $E$ and less than 0.01% for $P$ in all cases, indicating robust convergence to a unique solution for any reasonable starting value of Young's modulus.

| Subject | Median Blood Pressure CV | Median Young's Modulus CV |
|---------|--------------------------|---------------------------|
| A | 1.0e-4 | 1.1e-3 |
| B | 5.5e-5 | 5.0e-4 |
| C | 4.3e-5 | 4.9e-4 |
| D | 1.4e-5 | 1.4e-4 |
| E | 2.2e-5 | 2.4e-4 |
| F | 2.8e-5 | 3.9e-4 |

Table 5.1: Summary of the empirical investigation into the convergence of pressure and stiffness estimates from a range of initial conditions. The small coefficients of variation (CVs) for both pressure and stiffness indicate that our iterative method for determining pressure and stiffness robustly converges to a unique answer for a wide range of initial conditions.

Figure 5.1 shows representative plots of how this iterative convergence looked in practice for the various initial values. In all of these examples we see that the iteration converges very nearly to its final answer within 3 Gauss-Seidel steps, even

for far-off initial values for stiffness. The iteration could be caused to diverge with extreme initial values, such as $E < 0$ or $E > 100$ MPa, and unphysical inputs, such as $dP/da < 0$. However, this analysis indicates that convergence should be expected for physically reasonable initial values and parameters.
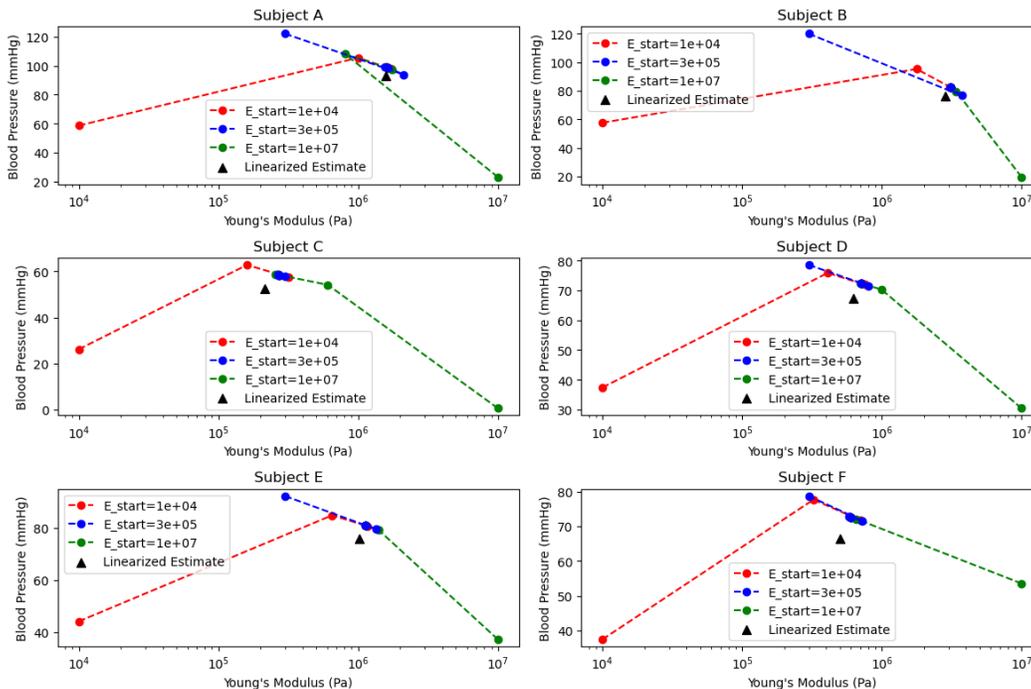


Figure 5.1: Demonstration of convergent behavior in the iterative stiffness estimation approach. Each panel shows pressure/stiffness iteration tracks for a randomly selected time step from a specified subject in the initial $N = 6$ *in vivo* feasibility study (circular scatter points and dashed lines). Even though the iterations begin from initial conditions differing by 3 orders of magnitude in stiffness, they consistently converge to the same final answer after only a few rounds of iteration. This final answer is also consistently close to the initial linearized estimation (triangular points) from Equations 5.13 and 5.14.

**Analytical Approach**

The above empirical analysis shows that our iterative procedure robustly converges for a wide range of initial values for $E$ encompassing the physiologically relevant range, even if the starting point is orders of magnitude off from the final value. However, we do not need to rely on random initial values for $E$; instead, we can use an approximate solution to the system of equations and improve the accuracy of the starting value. The physical model consists of Equations 2.25 and 2.35 (reproduced below as Equations 5.1-5.2), which represent a system of differential equations which we must solve in order to determine $P$ and $E$:

$$P = \frac{9\alpha^4 - 5\left(3\alpha + \alpha^3\right)D + 3D^2}{-4(9\alpha - \alpha^3) + 12D}E \tag{5.1}$$

$$E = \frac{a}{\alpha}\frac{dP}{da}. \tag{5.2}$$

Such systems may, in principle, generate a family of many different solutions, in which case $P$ would not be uniquely determined. We wish to show that, for this particular system, the solution $(P, E)$ which satisfies these equations at any given instant is uniquely determined. From an analytical standpoint, the full equation 5.1 is intractable; however, we can analyze its linearized version from Equation 2.32:

$$P \approx \pi^2 \rho a^2 f^2 - \frac{h^3}{3(1 - v^2)a^3}E \tag{5.3}$$

$$= \pi^2 \rho a^2 f^2 - \frac{\gamma^3}{3(1 - v^2)a^6}E \tag{5.4}$$

$$E = \frac{a^3}{\gamma}\frac{dP}{da}. \tag{5.5}$$

The product $ah$ has been replaced with $\gamma$ because prior studies have found the arterial wall to be very nearly incompressible [63], [64]. This means that for a fixed length of the arterial wall, its cross-sectional area must remain constant even as pressure changes; thus, $\gamma$ is a constant independent of changes in $a$. To analyze this system, we make the common assumption that it behaves smoothly as radius changes without sharp discontinuities in pressure, stiffness, or frequency. Take two consecutive measurements where the radius has changed by a small quantity $\epsilon$. These will generate a system of four equations:

$$P(a) = \pi^2 \left(\gamma \rho_S + \frac{4}{5}a^2 \rho_L\right)f^2 - \frac{\gamma^3}{3(1 - v^2)a^6}E(a) \tag{5.6}$$

$$P(a + \epsilon) = \pi^2 \left(\gamma \rho_S + \frac{4}{5}(a + \epsilon)^2 \rho_L\right)\left(f + \epsilon\frac{df}{da}\right)^2 - \frac{\gamma^3}{3(1 - v^2)(a + \epsilon)^6}E(a + \epsilon) \tag{5.7}$$

$$E(a) = \frac{a^3}{\gamma}\frac{P(a + \epsilon) - P(a)}{\epsilon} \tag{5.8}$$

$$E(a + \epsilon) = \frac{(a + \epsilon)^3}{\gamma}\frac{P(a + \epsilon) - P(a)}{\epsilon} \tag{5.9}$$

where $\frac{df}{da} = \frac{f(a+\epsilon)-f(a)}{\epsilon}$. Note that we have used the smoothness assumption to neglect terms of the order $\epsilon \frac{d^2 f}{da^2}$, $\epsilon \frac{d^2 P}{da^2}$, or $\epsilon \frac{d^2 E}{da^2}$. Given that the device is able to measure $a$ and $f$ (and thus $\epsilon$ and $df/da$) at each time step, the system of 4 equations has only 4 unknowns. Rewriting in linear algebra form, we get

$$
\begin{bmatrix} C_1 \\ C_2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{\gamma^3}{3(1-v^2)a^6} & 0 \\ 0 & 1 & 0 & \frac{\gamma^3}{3(1-v^2)(a+\epsilon)^6} \\ \frac{a^3}{\gamma\epsilon} & \frac{-a^3}{\gamma\epsilon} & 1 & 0 \\ \frac{(a+\epsilon)^3}{\gamma\epsilon} & \frac{-(a+\epsilon)^3}{\gamma\epsilon} & 0 & 1 \end{bmatrix} \begin{bmatrix} P(a) \\ P(a+\epsilon) \\ E(a) \\ E(a+\epsilon) \end{bmatrix} \tag{5.10}
$$

$$
C_1 = \pi^2 \left( \gamma \rho_S + \frac{4}{5} a^2 \rho_L \right) f^2 \tag{5.11}
$$

$$
C_2 = \pi^2 \left( \gamma \rho_S + \frac{4}{5}(a+\epsilon)^2 \rho_L \right) \left( f + \epsilon \frac{df}{da} \right)^2. \tag{5.12}
$$

Solving this system and applying the limit of $\epsilon \to 0$ to simplify yields the unique solution

$$
P(a) = \frac{\pi^2 \rho a^2 f^2}{1 - \alpha^2/(1-v^2)} \left[ 1 - \frac{\alpha^2}{1-v^2} \left( 1 + \frac{2a}{3f}\frac{df}{da} + \frac{8}{15}\frac{\rho_L}{\rho} \right) \right] \tag{5.13}
$$

$$
E(a) = \frac{2}{\alpha} \frac{\pi^2 \rho a^2 f^2}{1 - \alpha^2/(1-v^2)} \left[ \frac{a}{f}\frac{df}{da} + \frac{4}{5}\frac{\rho_L}{\rho} \right]. \tag{5.14}
$$

This solution is not exact, as it was generated from a linearized version of the full equations; however, it should represent a close approximation of the true values of $E$ and $P$. It has been shown that for any twice continuously differentiable nonlinear system, Gauss-Seidel iteration is guaranteed to converge to the ideal solution given initial values reasonably close to this solution [109]. Thus, by using our approximate analytical solution as a starting point we can have even greater confidence that numerically solving the nonlinear system through iteration will converge to the true solution. As shown in Figure 5.1 above, this linearized estimate does indeed consistently come close to the final values for $P$ and $E$.

## 5.3   Methods

In principle, the equations described above could be used to measure stiffness instantaneously at every time step. In practice, however, these instantaneous stiffness estimates are highly sensitive to measurement noise due to their dependence on a

derivative. In particular, if $da \approx 0$ between consecutive time steps the estimates for $df/da$ or $dP/da$ can be huge, throwing off any simple lowpass filtering scheme. Fortunately, even though stiffness can change significantly over the course of a cardiac cycle, stiffness at any given radius within the cardiac cycle should hold approximately constant from heartbeat to heartbeat over a short time frame. This allows us to denoise stiffness estimates by averaging data over multiple heartbeats. The following procedure was applied to obtain stiffness for *in vivo* data:

1. Calculate initial pressure estimates based on an assumed stiffness of 0.4 MPa. For this dataset, this was found to yield stable convergence without needing to apply Equation 5.14; however, for future applications on more unhealthy arteries the initial approximation may be necessary.

2. Pass measured radius, thickness, and pressure values through a 20Hz low-pass filter, and calculate $dP$ and $da$ by taking the differences of consecutive measurements. Use these values to generate a time series of instantaneous stiffness estimates using Equation 5.5.

3. Identify time series points where $|da|$ was in the bottom 20th percentile, and discard all of these $E$ estimates due to derivative instability.

4. Divide all remaining time series points into 20 evenly spaced bins based on radius value. Generate an $E$ estimate for each bin based on the median of all instantaneous estimates with a radius in that bin. This yields 20 points, each with a coordinate $(a, E)$.

5. To further increase stability, fit these binned $(a, E)$ points with a generic sigmoid function. This functional form can be used to get a value for $E$ at every radius.

6. Calculate a new set of pressure values based on $E$ values from this sigmoid function; this constitutes the first pass of Gauss-Seidel iteration. Repeat steps 2-6 for three more iterations.

7. By the final pass, we have obtained fairly stable $E$ estimates. Replace the sigmoid with a cubic spline to allow it to take on non-sigmoidal behavior.

For real-time BP measurement, this process was performed over a time window of the previous 5 seconds of data. For post-processing to generate the results presented

in Chapter 4, the process was performed over all data in the scan which passed the QC check of falling into a physical pressure range. The interpolated stiffness values from our fitted funtion (either sigmoid or spline) were used to generate the $E$ and $dE/da$ values required by out Kalman filter (see Equation 4.14).

## 5.4 Results and Discussion

The clinical study was not designed to measure any form of reference stiffness value, so we cannot compare our stiffness measurements against a ground truth to gauge accuracy. However, we can examine the results for overall trends. Figure 5.2 shows distributions of diastolic and systolic stiffness values, as well as their ratio. Diastolic and systolic values were calculated as the median stiffness in the lower and upper 3 radius bins, respectively. We see that the large majority of stiffness values fell between 0.1 and 1 MPa, roughly matching the range reported in Khamdaeng, Luo, Vappou, *et al.* [67]. As expected, the systolic stiffness distribution is shifted upwards relative to the diastolic distribution, and most scans had a systolic to diastolic stiffness ratio greater than 1.
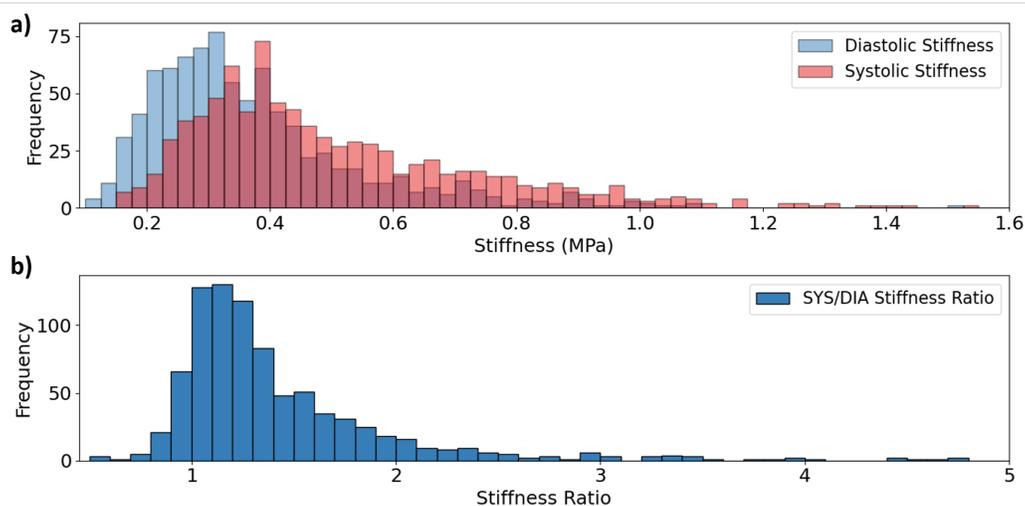


Figure 5.2: Overall distribution of stiffness results. a) Histograms showing the distribution of diastolic and systolic stiffness values. Both values show similarly shaped distributions, with the systolic distribution shifted upwards. b) Histogram showing the distribution of systolic/diastolic stiffness ratios. The large majority of scans showed higher stiffness at systole than at diastole, as expected.

We can also inspect our data for another predicted trend, which is that arterial stiffness tends to increase with age. Figure 5.3 shows the age and averaged diastolic and systolic stiffness for each of the 59 subjects that passed QC, along with a linear

regression for each. For both diastolic and systolic stiffness we do indeed see significant positive correlations of $r = 0.51$ and $r = 0.69$, respectively ($p < 0.001$ for both). The higher correlation for the systolic stiffnesses suggests that age-related arterial stiffening may be dominated by stiffening of collagen networks in arterial walls, as these are thought to be the dominant contributor to the increase in stiffness at systole [71].
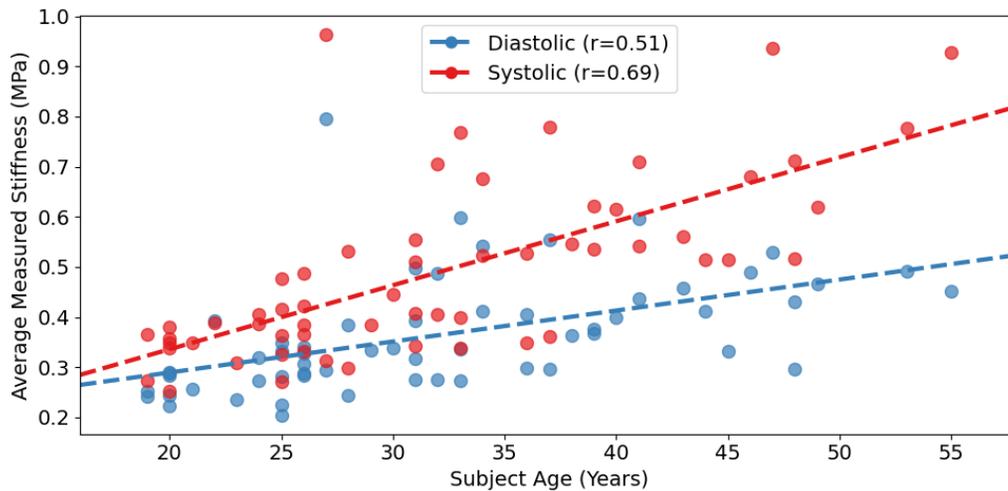


Figure 5.3: Averaged values of diastolic and systolic stiffness for each of the 59 clinical study subjects that passed QC, plotted against the subject's age. Physiologically, we expect vascular stiffness to increase with age. This trend is reflected in the data, with significant positive correlations for both diastolic and systolic stiffness.

A more specific facet to consider is the shape of the stiffness vs. radius curve. Prior studies have come to differing results; Johnson and Tarbell [66] reports that stiffness increases roughly linearly with radius, while Khamdaeng, Luo, Vappou, *et al.* [67] reports that stiffness increases sharply at the early onset of systole but then levels out, forming a sigmoidal curve with respect to radius. These studies were based off of relatively small samples of 19 rabbits and 7 healthy young men, respectively, so they may have each correctly identified behavior in specific sub-populations. The broader subject population covered by the clinical study offers the opportunity to investigate the $E$ vs $r$ relationship more broadly.

Figure 5.4 below shows some examples of $E$ vs $r$ relationships observed in 12 different subjects in the clinical study, which have been manually assigned to different shape categories. The majority of inspected scans roughly followed either a linear or sigmoidal shape, in line with expectations from literature. Even for scans with similar shapes, the magnitude of stiffness change between diastole and systole varied

significantly; for some scans the difference was as low as 20%, while for others the difference was well over 200%. For some other scans stiffness was roughly constant through the whole cardiac cycle, changing by less than 10% from diastole to systole. The final category, irregular shapes, was the least often observed but also the most interesting to consider.
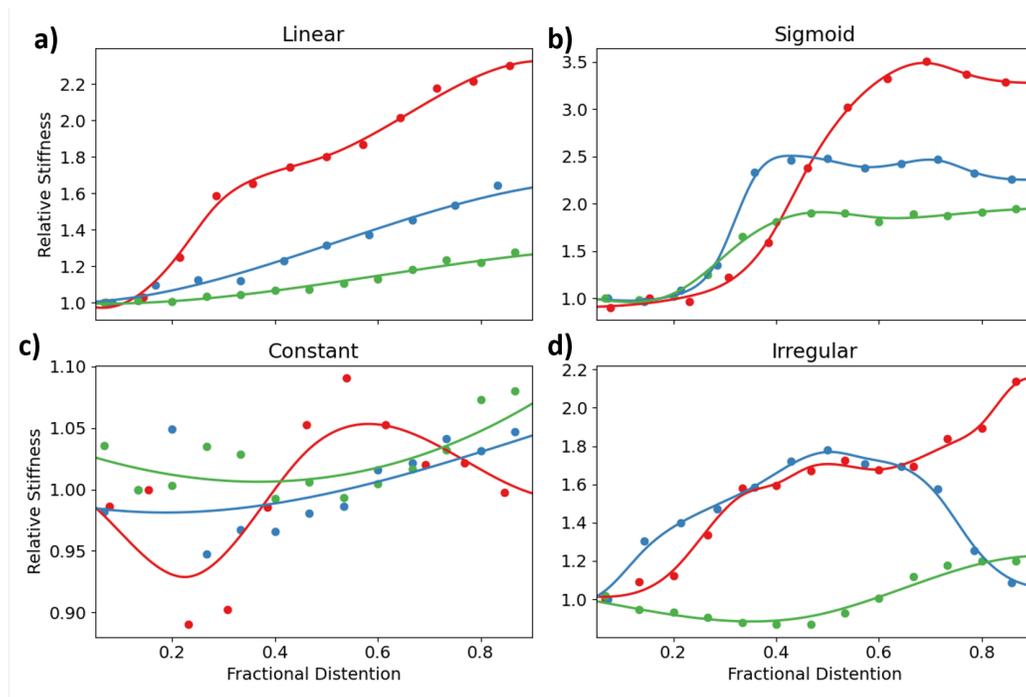


Figure 5.4: Examples of different types of stiffness vs. radius shapes observed in clinical data. For ease of comparison, the x-axis represents the radius of the scan normalized from 0 to 1, and the y-axis represents the stiffness of the scan normalized such that its diastolic value is 1. Each panel shows representative stiffness curves of a given shape class, obtained from 3 different subjects. The most common types observed were linear (a) or sigmoidal (b), which are shapes that have been seen in prior literature. In some scans stiffness was roughly constant (c) and did not vary by more than 10% from diastole to systole. Some scans exhibited more irregular behavior (d), suggesting unexpected arterial dynamics.

Studies of arterial wall stiffness generally model the wall as a combination of low-stiffness elastin and high-stifness collagen fibers [66], [67], [71]. Near diastole the collagen is in a relaxed state, so wall tension is dominated by elastin and stiffness is relatively low. However, as the artery expands collagen becomes engaged, increasing stiffness. Modeling of this form has been used to explain linear [66] and sigmoidal [67] $E$ vs. $r$ relationships, and a constant shape could also be explained if the entire cardiac cycle stays in the elastin-dominated or collagen-dominated regime. However, such models would not easily be able to explain the shapes seen in Figure

5.4d, which include features such as double sigmoids or are non-monotonic. This suggests that some mechanism other than simple elastin-collagen interaction is likely at play, at least in some subset of arteries; one possible contributor is the smooth muscle sheathing the wall, which can has the capability to dynamically relax and contract to change its stiffness [71]. Future studies using CARDI-BP could yield further examples of irregular stress-strain relationships, enabling more detailed investigation into arterial wall mechanics.

## 5.5    Conclusion

Arterial stiffness measurement is a relatively new field, and doctors are still researching new diagnostic uses for it. One factor limiting in this research has been deficiencies in stiffness measurement devices, which often yield one overall value averaged across many major arteries for the entire cardiac cycle. CARDI-BP is the first self-contained method which can provide real-time, local measurement of arterial stiffness as it varies over the course of a heartbeat. This has the potential to enable novel stiffness-based diagnostics that have not been possible with current measurement methods. For example, the results in Figure 5.3 suggest that systolic stiffness may be more indicative of vascular age than overall averaged stiffness due to its higher correlation. The results in Figure 5.4 show that there is not a one-size-fits-all model for the shape of the stiffness vs. radius relationship, and the nature of this shape for a particular patient could also contain as-yet-unknown diagnostic value. Widespread deployment of CARDI-BP would drastically increase our level of knowledge of arterial stiffness dynamics across wide ranges of patients, potentially leading to broad improvements in patient care.

*Chapter 6*

# CLINICAL APPLICATION OF MACHINE LEARNING TO ULTRASOUND DATA

While arterial blood pressure is the most well-known and commonly measured form of BP, there are also other forms of blood pressure which can be useful in various medical contexts. Amongst the myriad of parameters that can affect cardiac function are intracardiac filling pressures, which represent measurements of the diastolic pressure in a each chamber of the heart. Measurement of intracardiac filling pressures provides important quantitative clinical information that can aid in the diagnosis and management of patients with a spectrum of different pathologies including heart failure, a condition that results from a weakened heart muscle's inability to pump enough blood to the rest of the body. This condition is associated with fluid retention and volume overload, a diagnosis that requires inpatient hospital admission and fluid removal via aggressive diuresis. Heart failure afflicts over 20 million people worldwide [110], and related hospitalizations represent one of the greatest burdens to the healthcare system, driving up cost and resource utilization in wealthy and poor economies alike.

This work deals with intracardiac filling pressure in the right atrium, or right atrial pressure (RAP). In many ways, the clinical status of RAP measurement is quite similar to that of arterial BP measurement. Physicians must choose between two measurement methods: a catheter-based approach which is highly accurate but also difficult and invasive, or a non-invasive approach which is prone to inaccuracy. In contrast to the BP cuff, however, even the noninvasive assessment of RAP is relatively difficult to perform, requiring time from specially trained physicians to interpret the results. The goal of this study is not to create an entirely new measurement method as we did for arterial blood pressure, but rather to replicate and potentially improve the existing non-invasive measurement method using machine learning.

## 6.1   Background: Right Atrial Pressure and the Sniff Test

Volume overload (VO) is one of the most common complications in patients with heart failure. VO results from the inability of the heart to adequately circulate blood throughout the body, causing fluid backup and leading to elevation in intracardiac filling pressures. The symptoms of VO associated with heart failure can manifest

as multi-system organ damage, including fluid buildup in the lungs and difficulty breathing, delerium and raised intra-cranial pressure, acute kidney failure, liver damage, and a generally higher risk of mortality [111]. As outlined above, patients with heart failure suffer from frequent exacerbations of their disease, often prompted by dietary indiscretion, medication non-adherence, or progression of their underlying cardiac dysfunction. Early identification of VO allows for effective management of patient care through treatments such as diuretic therapy [112] which, if initiated early in the disease course, can be managed with oral medications and avoid costly hospitalizations. In contrast, if VO goes unidentified, its negative effects on organ systems can progress and may even be aggravated if extra fluid is administered for other concomitant indications such as septic shock [111], [113]. As a result, there is significant clinical interest in methods of easily and accurately identifying volume status.

Overall volume load in the circulatory system is closely related to RAP. The right atrium collects deoxygenated blood from the body, after almost all of the blood pressure originally generated by the heartbeat (i.e., left ventricle contraction) has been dissipated in capillary beds. When the heart is healthy and fluid volume in the body is normal, RAP is generally quite low (~3 mmHg). However, for patients with VO there is an increase in the amount of fluid within the venous system, leading to an elevated RAP in the range of ~15 mmHg [114]. This relationship makes RAP assessment one of the most common methods for screening for VO [111], and it can also be used as a more general indicator of cumulative cardiac burden in patients with heart failure [115].

The gold standard for RAP measurement is right heart catheterization, or RHC. In this procedure, a doctor inserts a flexible catheter with a pressure transducer at its tip into a vein in the neck, groin, or arm. The catheter is then gradually advanced through the venous system until it reaches the right side of the heart. This procedure yields highly accurate pressure measurements in the right atrium, right ventricle, and pulmonary artery. However, the process of RHC is time-consuming, difficult, and costly, requiring a specialized cardiac catheterization laboratory and roughly an hour of time from a team of specialists. The invasive nature of the procedure also carries risk of various side effects, such as bleeding, infection, heart valve damage, or pulmonary artery perforation [116]. As a result, RHC is only conducted when physicians have an appropriate indication and is generally reserved for hospitalized patients who are critically ill. Non-invasive estimation of RAP can be performed via

physical examination of the jugular vein, but this process is qualitative, subjective, and can be sensitive to factors such as operator experience and patient positioning.

A more quantitative way to non-invasively assess cardiac structure and function is the transthoracic echocardiogram, or TTE. A TTE study consists of a series of ultrasound scans of the heart and its surrounding vessels from many different viewpoints in order to gain an overall picture of its behavior and health. Various views can be used for the diagnosis of many different cardiac conditions, from cardiomyopathy to valvular regurgitation or stenosis [117]. The TTE component which is of particular interest for our study is the assessment of the interior vena cava (IVC), which is the major vein that returns blood to the right atrium. The behavior of the IVC is connected to RAP measurement via the "sniff test."

The idea behind the sniff test is that blood pressure in the IVC near its juncture with the right atrium is very similar to pressure within the right atrium. In a healthy patient RAP is quite low (~3 mmHg), so pressure in the IVC is low as well. Compared to arteries, venous blood vessels have thin walls with very little stiffness; the only thing holding them open is a positive pressure differential between the blood they contain and surrounding tissue. Thus, in healthy patients the IVC is relatively narrow, and a mild increase in external pressure (in this case pressure within the thoracic cavity) is sufficient to make the vessel collapse. However, in patients with high RAP (~15 mmHg) there is greater internal pressure in the IVC, leading to a larger IVC diameter and little to no collapse with elevation in intrathoracic pressure. To take advantage of this, sonographers first image the IVC at rest to determine its diameter, as shown in Figure 6.1a,c. They then perform a sniff test by asking the patient to sharply inhale (thereby acutely increasing intrathoracic pressure) while their IVC is being imaged, resulting in a compressive force on the vessel. For patients with a healthly RAP, this pressure spike is usually sufficient to almost entirely collapse the IVC (Figure 6.1b), while in patients with an elevated RAP the sniff may barely collapse the IVC at all (Figure 6.1d). The ratio of decrease in IVC diameter during a sniff to its resting diameter is defined as the collapsibility index (CI); a CI of 100% represents full collapse, while a CI of 0% represents no collapse.

The method of assessing RAP based on the sniff test has been codified by Rudski, Lai, Afilalo, *et al.* [118], and further recommended by the American Society for Echocardiography and European Association of Cardiovascular Imaging in Lang, Badano, Mor-Avi, *et al.* [117]. This standard breaks the assessment into two criteria: 1) is the resting diameter of the IVC below 21 mm, and 2) is the CI greater than
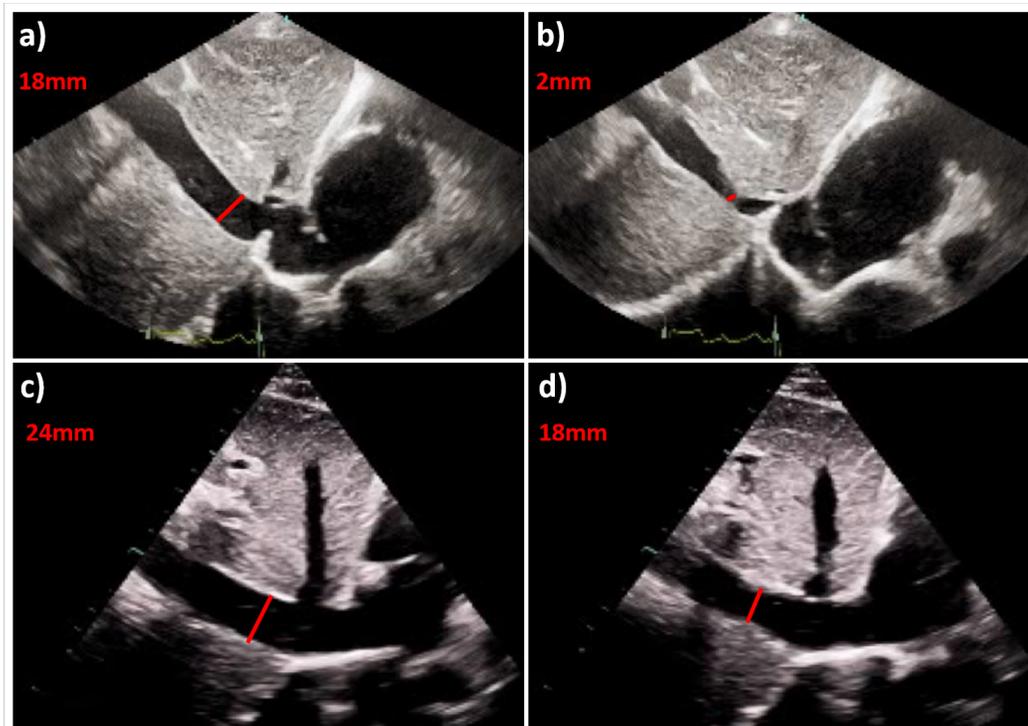
Figure 6.1: Sniff test illustration. a) An ultrasound view of the inferior vena cava (IVC) and right atrium at a resting state. The IVC tends to curve in starting from the top left of the field of view. Towards the bottom right, the IVC expands and flows into the right atrium of the heart. This IVC has a relatively small diameter of 18 mm. b) Another view of the IVC in (a), from a different frame in the same video showing the sniff. The high degree of collapse, combined with the low resting diameter, indicates that this patient likely has a normal RAP. c) A view of another IVC at rest, this one with a relatively large diameter of 24 mm. d) Another frame of the IVC in (c), showing a sniff. The low degree of collapse, combined with the high resting diameter, indicates that this patient likely has an elevated RAP.

50% (see Table 6.1). An answer of "yes" to both criteria gives an RAP estimate of between 0 and 5 mmHg; for convenience, this is generally simply written as a pressure of 3 mmHg. An answer of "no" to both criteria gives an RAP estimate of 10-30 mmHg, commonly written as 15 mmHg. A split of one "yes" and one "no" gives an RAP estimate of 5-10mmHg, commonly written as 8 mmHg. This relatively quick and easy assessment of RAP as either 3, 8, or 15 (low, medium, or high) can be used to guide clinical decision-making. An RAP of 3 indicates that the patient likely does not have VO, while an RAP of 15 indicates that the patient likely does have VO and should be directed to additional treatment. A value of 8 is considered indeterminate, and other metrics may be used in concert with the RAP estimate to determine whether further tests or treatment for VO are appropriate.

| Criterion | IVC Diameter < 21 mm | IVC Diameter > 21 mm |
|---|---|---|
| **IVC CI > 50%** | 3 mmHg | 8 mmHg |
| **IVC CI < 50%** | 8 mmHg | 15 mmHg |

Table 6.1: Summary of the "sniff test" standard for estimating RAP from an ultrasound scan of the IVC. If the IVC has a small resting diameter and significant collapse during a sniff it is labeled as low RAP (~3 mmHg), and if it has a large resting diameter and little collapse during a sniff it is labeled as high RAP (~15 mmHg). Scans with split results between these two criteria are labeled as having an intermediate RAP (~8 mmHg).

While conducting a sniff test according to Table 6.1 may seem straightforward, in practice there are many potential sources of error [119]. For example, operators are advised to measure IVC diameter between 0.5 and 3 cm from the junction with the right atrium [118]. Since the IVC tends to flare as it approaches this junction, the chosen point of measurement can impact the measured diameter significantly. Diameter and CI can also be mis-estimated if the imaging plane is not perfectly in line with the IVC, especially if the sniff causes the IVC to move out of the original imaging plane. In some cases the IVC can even be confused with the abdominal aorta, which is located near the IVC and looks very similar in some patients but does not collapse significantly during a sniff regardless of RAP.

Navigating these uncertainties to obtain robust RAP estimates requires experienced sonographers and interpreters. Indeed, multiple studies have found significant inter-operator variability amongst medical trainees, fellows, and emergency physicians when assessing IVC diameter and collapsibility, even after dozens of hours of training [120]–[122]. The most reliable interpreters are generally considered to be experienced cardiologists, who have spent years evaluating many thousands of TTEs. However, while there is widespread access to ultrasound equipment capable of imaging the IVC for a sniff test, many medical centers do not have 24/7 access to cardiologists capable of making an assessment. Even at large medical centers with significant resources and staffing in cardiology, the availability of expert TTE assessment for volume status is not always immediately feasible.

**Study Goals**

The rapid proliferation in recent years of machine learning (ML) models capable of analyzing video data, and particularly ultrasound video data [123], [124], makes ML an obvious approach to this problem. If an ML model could be trained to

automatically analyze sniff tests with the proficiency of a veteran cardiologist, quick and non-invasive RAP assessment could be expanded from cardiology departments and democratized to anyone with access to an ultrasound machine and a competent sonographer. The goal of this study is to train such an ML model, and to compare its performance to both expert cardiologist assessments and the gold standard of RHC.

Multiple prior studies have attempted to use machine learning to analyze ultrasound scans of the IVC and use it to predict parameters such as RAP and overall fluid responsiveness [125]–[128]. However, these studies only had access to relatively small datasets of between 41 and 175 patients. After most of the data was allocated for model training and validation there was very little left for out-of-sample testing, increasing the chances of overfitting. Furthermore, these datasets were mostly collected under controlled conditions, with all scans obtained using the same ultrasound instrument and assessed by the same interpreter. This would lead to significant generalization concerns if a model trained on such a dataset were applied to data from different hospitals using different ultrasound machines.

Through a collaboration with the cardiology division at the University of California, San Francisco (UCSF) Medical Center, we obtained access to a large database of over 16,000 individual TTE studies and associated 3/8/15 RAP assessments made by expertly trained and board certified UCSF cardiologists. This represents a nearly 100x increase in dataset size over any previous study which used ML to predict RAP, enabling far more robust training and testing. Furthermore, the dataset covers TTEs collected over a period of 4 years by dozens of operators using 4 different models of ultrasound machine, encompassing all of the messiness inherent in real-world medical data. As a result, this study represents a significant opportunity to both train an accurate ML model and robustly evaluate its real-world performance.

## 6.2  Dataset and Data Screening

The database underlying this project covered a significant portion of all echocardiograms performed in the UCSF cardiology division from 2012-2020, along with associated measurements, metadata, and doctors' notes. The high-level unit of data was the TTE study, which consists of up to 200 individual scans which are taken in sequence on the same patient to gain many views of the heart and its surrounding vessels from different angles. Interpretations of these studies, such as RAP estimates, were made for the study as a whole. The information relevant to this research was split among three separate datasets with varying degrees of mutual overlap:

- **Dataset 1:** Raw video data and associated metadata from echocardiogram studies. This covered 182077 studies on 48481 different patients, for a total of over 11 million individual scans. Video data from the scans was already pre-processed to remove patient information, doctor's notes, Doppler velocity traces, etc. so only the actual ultrasound video remained.

- **Dataset 2:** RAP estimates made by cardiologists based on the sniff test. This covered 51324 studies on 35003 different patients.

- **Dataset 3:** Gold-standard RAP measurements from right heart catheterization. This contained 9001 individual readings from 5585 different patients. These catheter measurements were independent of ultrasound measurements, and each patient may or may not have received a TTE.

**Data Exclusion Criteria**

TTE studies were removed from data sets 1 and 2 in their entirety if they met any of the following criteria:

- Multiple different sniff-based RAP estimates were recorded for the same study.

- The recorded study type noted a stress test, pediatric or fetal subject, transesophageal or intracardiac scanning, or patient on a ventilator. None of these study types are representative of how a normal sniff test would be evaluated.

- 99.7% of remaining studies were conducted on one of 4 models of ultrasound machine; the remaining 0.3% of studies were eliminated.

Furthermore, individual scans were removed from dataset 1 if they met any of the following criteria:

- Video pre-processing failed.

- The scan was less than 20 frames long. Such video clips are too short to contain a full sniff.

- Physical pixel size was not recorded in metadata. This made it impossible to measure IVC diameter in real units.

- Physical pixel size was in the lower or upper 5th percentile of pixel scale. This narrowed the total range of pixel scales from (0.002, 5.2) centimeters to (0.074,

0.168) centimeters. Extreme pixel sizes imply extreme ultrasound settings for parameters like scanning frequency and depth, which would generally not be used for viewing the IVC.

- Color Doppler mode was enabled. Sniff scans are rarely taken with Color Doppler enabled, and extra splashes of color in a small subset of data would be likely to confuse ML models.

Finally, RHC measurements were excluded from dataset 3 if they were less than 0 or greater than 30mmHg, as these represent non-physical values that would have resulted from bad setup or calibration of the catheter.

**RAP Estimate Data**

Dataset 2 covered RAP estimates from 2012 to 2020, but these estimates were not always made in the same way. The standard described in Table 6.1 was officially recommended by the American Society of Echocardiography in 2015 [117], so the large majority of RAP estimates in 2016 and later were recorded in the 3/8/15 format. Prior to this, UCSF cardiologists primarily used a different standard to estimate RAP as either 5, 10, 15, or 20 mmHg. Because these two standards were based on different sets of measurements and thresholds, they were not inter-convertible. The frequency of these standards in dataset 2 broke down as follows:

1. Year >= 2016, Pressure in (3, 8, 15): 29863 Studies,

2. Year <= 2015, Pressure in (5, 10, 15, 20): 18116 Studies,

3. Other: 3345 Studies.

We chose to focus exclusively on the first category, as it was both the largest and aligned with modern clinical practice. A study was only useful for ML training if it was associated with both input data (raw ultrasound video) and target data (an RAP value). Thus, we needed to see how many studies were contained in the intersection of our various datasets in order to determine how much useful data was available. Of the 29863 studies for which we had valid RAP estimates, 19047 of them were associated with raw video data. After applying the data exclusion criteria described above, 16823 studies remained; this represented the full breadth of our possible training data.

Out of the 16823 remaining studies, 78.5% had an estimated RAP of 3mmHg, 13.8% had 8mmHg, and 7.8% had 15mmHg. This indicates that a substantial majority of patients assessed with the sniff test had a healthy RAP. The large imbalance between these classes required special treatment when training ML models, as described in section 6.4.

**RHC Measurement Data**

Datasets 1 and 2 were relatively easy to align, as they were both based on the data unit of a TTE study. Right heart catheterization, however, is an entirely separate procedure; patients who receive a TTE study may never receive an RHC (and vice versa), and patients that do receive both measurements will generally not have them taken simultaneously or even on the same day. Thus, we needed to take special care when aligning dataset 3 to the rest of our data.

Out of the 5585 patients represented in dataset 3, 2299 of them had at least one TTE study in dataset 1 (which may have been from any date). We paired up TTE studies and RHC measurements in a 1:1 mapping such that only the closest-in-time RHC measurement for each TTE study was kept; this left a total of 3483 RHC/TTE data pairs. After applying the data exclusion criteria described above, 2586 of these studies remained.

The next factor to consider was time separation between the echo and RHC measurements. Since the two measurements were not taken at the same time, there was some chance that the patient's true RAP changed significantly during the intervening time. If the two measurements were taken on the same day, this was relatively unlikely; if they were taken years apart it was significantly more likely. To narrow the data, we needed to decide on a maximum allowable interval between the RHC and echo measurements in order to be considered a valid data point. Based on analysis of cardiologist accuracy at different time intervals (see section 6.5), we settled on a cutoff of 1 month (30 days), which left a total of 1739 studies which could be possibly used for RHC training data. Of these, 527 were labeled with an RAP estimate by cardiologists. The remaining 1212 were not, which could indicate that either a) a sniff test was not performed, or b) a sniff test was performed, but for unknown reasons a corresponding RAP estimate was not recorded. This will be investigated further in section 6.4.

A final consideration was how to determine the "accuracy" of cardiologist or machine learning RAP estimations (which are categorical) when comparing to RHC

measurements (which are continuous). We could make a direct comparison by binning the RHC values according to the ranges in Lang, Badano, Mor-Avi, *et al.* [117]. However, these ranges have a bit of ambiguity; in principle, an RHC value of 5 could be binned as "3" or "8," and an RHC value of 10 could be binned as "8" or "15." Since RHC measurements are reported as integers rather than floating points, these ambiguous situations were fairly common. Upon consultation with UCSF cardiologists, we elected to put these edge case readings into the lower of the two possible bins. Thus, an RHC value in the range $[0, 5]$ was binned as "3," a value in the range $(5, 10]$ was binned as "8," and a value in the range $(10, 30]$ was binned as "15." Continuous outputs from regression ML models were binned in the same way. This allowed for direct accuracy comparisons between any combination of outputs from RHCs, cardiologist estimations, categorical ML models, and regression ML models.

**Making Use of RHC Data**

Training a model to reproduce cardiologist estimates of RAP can, at best, reproduce the performance of these cardiologists. Prior studies have found that physicians using the sniff test can often make errors in predicting RAP, as compared to gold-standard RHC measurements [119], [129]. These errors must stem from one of three sources:

1. The interpreter made an error in evaluating the sniff test.

2. The sniff scan contained sufficient information to accurately estimate RAP, but the measurements and thresholds from Table 6.1 were not the correct way to perform this estimate.

3. The sniff scan did not contain sufficient information to accurately estimate RAP, no matter what analysis was performed.

Training on a large body of sniff scans paired with cardiologist RAP estimates may partly mitigate type (1) error by averaging operator variability over many interpreters; however, no amount of such data could address type (2) or (3) errors. Making use of RHC data eliminates type (1) and (2) errors by removing human interpretation variability as well as potential issues with the standards of Table 6.1. Thus, we combined training on cardiologist RAP estimates as well as ground-truth RHC values in an attempt to surpass cardiologist performance at predicting true

RAP. Mitigating type (3) error would require bringing in additional information from other TTE views or non-TTE data, which is beyond the scope of this study.

## 6.3 IVC/Sniff Identification

A full TTE study may contain up to 200 individual scans, but in general only 2-4 scans show the IVC and only 1-2 of these contain a sniff. Unfortunately, even though the ultrasound technician knew which scans contained a sniff at the time of acquisition, these labels (or any other labels related to what was being viewed in each scan) were not recorded in dataset 1. Thus, we needed an initial screening process to isolate the sniff scans in each study. The 16823 studies mentioned above contained a total of over 800,000 scans which passed the exclusion criteria and could potentially be sniffs. This was far too many scans to screen manually, so we instead trained a front-end ML model to screen for sniff scans.

Developing an IVC/sniff identification model still required generation of a manually labeled set of scans for training. We began with a set of 1243 scans that had been marked as "subcostal" for a different study; this label included both IVC scans and other non-IVC views of the surrounding area. With guidance from a cardiologist, this first set of scans was reviewed to identify 420 which contained an IVC. Using this dataset, we trained an initial binary classification ML model to identify IVCs using the X3D-M architecture with default hyperparameters [130], which had previously been found to perform well for other ultrasound image classification tasks by other members of the research group. To further build the training set, we used this initial model to classify many additional random scans out of the remaining 800,000 and selected 2757 which were identified by the model (either correctly or incorrectly) as IVCs to add to the dataset.

This set of 4000 scans was manually reviewed and labeled each as either 1) not a view of an IVC (2453 scans), 2) a view of an IVC without an obvious sniff (708 scans), or 3) a view of an IVC with a sniff (839 scans). Because most of the training videos were scans which had been classified as an IVC by the initial model, many of the type (1) scans were views which shared similarities to IVCs that caused them to be mis-classified. The goal of this was to present the final IVC/sniff classification models with many challenging training cases, enhancing their ability to distinguish true IVCs from look-alikes (see examples in Figure 6.2).

To use this larger labeled dataset, we divided the 4000 scans into 2800 for training and 600 each for validation and testing. A binary classification X3D-M architecture
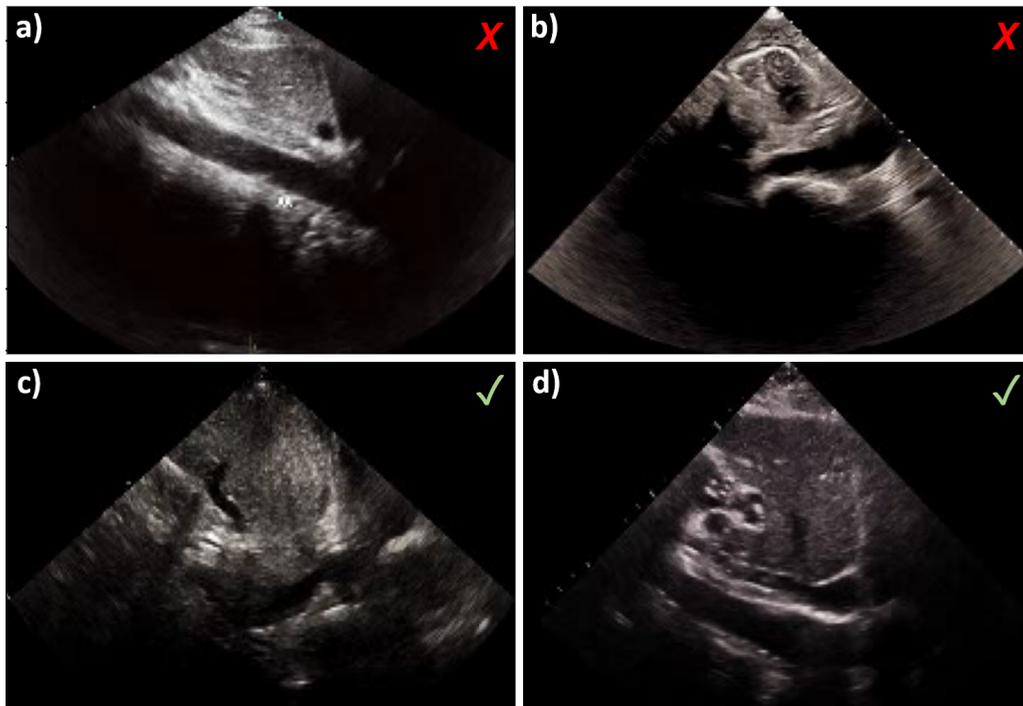
Figure 6.2: Examples of scans that could be confused with an IVC, or that are in fact IVCs but are hard to identify. a) This is a vessel sloping in the correct direction with a darker space to the right, but it is the abdominal aorta rather than the IVC. b) This is from another region of the heart, but the dark area in the middle right happens to resemble an IVC starting to expand into the right atrium. c) This is an IVC, but the scan is quite noisy and a lot of the vessel and atrium are obscured. d) This is an IVC with an artery next to it. This artery would not collapse during a sniff, which could confuse the sniff classifier.

was trained to distinguish type (3) scans from types (1) and (2). Training scans were augmented by random rotation and random scaling of brightness, saturation, and contrast on each epoch. Multiple training runs were conducted while varying the hyperparameters of optimizer type, learning rate, number of sampled frames, degree of augmentation, and dropout rate. The final model was chosen by maximizing validation accuracy. The out-of-sample performance results are shown below in Table 6.2.

Despite the relatively small training set and deliberate selection of challenging training and test cases, the sniff classification model still performed well when presented with out-of-sample scans. However, analysis of the results showed that the model was more likely to classify true sniffs as non-sniff scans when the degree of collapse induced by the sniff was small. This made intuitive sense, as a full collapse of the IVC could be seen very easily while a slight collapse may have been

| Confusion Matrix | Predicted No Sniff | Predicted Sniff |
|:---:|:---:|:---:|
| True No Sniff | 439 | 41 |
| True Sniff | 25 | 95 |
| Accuracy | 89% | |

Table 6.2: Out-of-sample performance of the sniff/no sniff binary classification model. Green/red cells represent the raw count of correct/incorrect predictions.

easier to miss and could potentially have been confused with a simple shift in probe positioning or fidgeting of the subject. While understandable, this bias created a problem; sniffs with low collapse were more likely to come from patients with high RAP, so the model would be biased against identifying sniffs in such patients even though RAP was not part of the training data. Patients with an RAP estimate of 15 mmHg already made up only 7.8% of our total training set, so we could not afford to disproportionately exclude them.

Our solution to this problem was to train a separate binary classification X3D-M model to identify IVC views without regard to whether or not a sniff was present (i.e., distinguishing type (1) from types (2) and (3)). The out-of-sample results of this model were similar to those of the sniff classifier, as shown in Table 6.3.

| Confusion Matrix | Predicted No IVC | Predicted IVC |
|:---:|:---:|:---:|
| True No IVC | 308 | 51 |
| True IVC | 25 | 216 |
| Accuracy | 87.3% | |

Table 6.3: Out-of-sample performance of the IVC/no IVC binary classification model. Green/red cells represent the raw count of correct/incorrect predictions.

The results of these classifier models were combined to identify sniffs in a 2-step procedure:

1. Run the IVC classifier on all scans in a study. Identify a scan as an IVC "candidate" if the classifier's probability output is at least 20%. If no IVC candidates are found, reject the study.

2. Run the Sniff classifier on all IVC candidates from the study. The candidate with the highest probability output from the sniff classifier is identified as the

representative sniff scan for the study, even if that sniff probability is quite low.

Out of the 16823 studies remaining from our initial data selection procedure, 993 were excluded as having no IVC candidates. The remaining 15830 studies were assigned a representative sniff scan and represented the full dataset used for model training and evaluation. Of these studies, 79.1% had an estimated RAP of 3mmHg, 13.4% had 8mmHg, and 7.5% had 15mmHg. This is very similar to the original split of 78.5%/13.8%/7.8%, indicating that we did not introduce a substantial bias towards any particular class of RAP in our sniff identification procedure.

It is worth taking a moment to examine the diversity of this dataset. The 15830 studies include data from 11869 patients using one of 4 different models of ultrasound machine. Patients ranged from 18 to 102 years old and spanned a wide range of medical conditions. Study evaluations were conducted by a total of 45 different physicians, and 20 of these physicians evaluated over 100 studies. This diversity of data is the main strength of this study; rather than training to match the judgement of a single physician reading off of a single ultrasound machine under controlled conditions, our training incorporated the full complexity of real-world medical data, making it far more likely to generalize to future real-world applications.

Inspection of a random subset the 15830 selected "sniff" scans revealed that a substantial majority did appear to be clear IVC scans, but roughly 25% were either low-quality IVC scans (where a significant portion of the vessel was noisy or obscured) or mis-selected scans which were not views of the IVC at all. To get a sense for the impact of these scans on model performance, we also compiled a "high-quality" dataset via the same procedure described above, with the difference that the cutoff for IVC classification probability was 90%. This cutoff eliminated roughly 65% of studies in the original dataset; out of what was left, inspection of a subset found that 98% of scans seemed to be high-quality IVC images. While this rate of data discarding was too high to be useful for training, it was useful for evaluating model test performance on a closer-to-ideal dataset.

### Sniff Selection for RHC Data

The same sniff selection described above was also applied to the 1739 TTE studies with an associated RHC measurement within 30 days of the study date. This yielded 932 studies where were identified as containing an IVC. This relatively low IVC identification rate is likely due to the fact that 1212 out of the original 1739 studies

did not have a recorded echo-based RAP estimate, indicating that many of these studies likely did not contain a sniff. The 932 identified "sniff" scans were manually screened to remove any non-IVC views (IVC views without an obvious sniff were left in), leaving 866 studies remaining. This 93% success rate indicates that our sniff identification process performed fairly robustly. For these studies, 35% had a catheter pressure in the range [0,5], 29% were in (5,10], and 36% were in (10, 30]. This indicates, unsurprisingly, that patients who underwent RHC were significantly more likely to have elevated RAP compared to the broader population that received a TTE.

Of the 866 remaining studies, 319 had a recorded echo-based RAP estimate. These studies studies represented our "golden" data for which we had a matched set of a sniff video, a cardiologist's RAP estimate based on this sniff, and a ground-truth RAP measurement from right heart catheterization. These scans were always kept in the test set for all model training to avoid contamination. The 547 studies with a sniff scan and an RHC measurement but no RAP estimate were used as training and validation data for model fine-tuning, as described below.

## 6.4 Methods

### RAP Classification

The set of 15830 studies with a cardiologist-generated RAP estimate was divided into 12664 training studies, 1583 validation studies, and 1583 test studies. As mentioned above, we ensured that all 319 "golden" studies with RHC measurements were included in the test set; otherwise, allocation was random. All models were constructed with a length-3 output layer and softmax activation to generate probabilities for 3 classes: 3 mmHg, 8 mmHg, and 15 mmHg. Training was performed on an NVIDIA RTX 6000 Ada graphics card.

We began by applying multiple modern video processing ML architectures to the problem, including X3D [130], SlowFast [131], MoViNet [132], TimeSformer [133], STAM [134], and ViViT [135]. The model type which generated the highest validation performance within GPU memory constraints was SlowFast R50, so this architecture was selected for further tuning and evaluation (see Appendix D.1 for discussion of other models and model aggregation). SlowFast works by dividing the flow of data into "slow" and "fast" lanes. The slow lane is only shown a fraction of frames in the full video and has a high number of convolutional channels. The high channel count gives the model a lot of power to identify relatively static features,

while the downsampling in frame count keeps computational load at a tractable level. The fast lane, in contrast, is shown the full set of video frames but has a relatively low number of convolutional channels. Seeing all frames allows this section of the model to detect quicker motions in the video, while the low channel count controls computational load. This architecture is particularly suited for our task, which requires identifying both slow features (IVC location and resting diameter) and fast features (sniff-induced collapse) to make an accurate classification. Model implementation was performed using the *PyTorchVideo* package [136].

When training the model, input pre-processing was applied in the following steps:

1. Standardize data length by selecting 64 frames evenly spaced throughout the length of the video. For a 10 second clip length (which is relatively long for this dataset) this will provide over 6 frames per second, which should guarantee catching a sniff. If the video is less than 64 frames long, pad at the end with empty frames.

2. Rescale each frame in the video such that the physical size of each pixel is 0.122 cm (the median for the whole dataset). If we need to zoom in, center crop. If we need to zoom out, pad the edges with the mean brightness of the image. The final size of all frames was kept at 224x224 pixels.

3. Apply random rotation and brightness/contrast adjustment transforms.

4. Convert the video to grayscale. Since we have excluded color Doppler scans the underlying data is already grayscale, so going from 3 color channels to 1 cuts down on data size without losing information.

5. Normalize each frame in the video to have mean intensity 0 and standard deviation 1.

Regularization was performed via data augmentation as well as the architecture default parameters of batch normalization and 50% dropout in the final classification layer. In addition, we applied label smoothing to the output targets [137]. With this scheme, the targeted probability distribution is no longer 100% on the single correct answer; instead, some percentage is redistributed to the other categories. In noisy systems such as this one, label smoothing discourages the model from assigning unrealistically high certainty to any one answer. Typically, the smoothed probability is distributed evenly to all non-target classes. However, in this system we had some

domain knowledge about which forms of noise were most likely; a scan labeled as 3 may properly have been an 8, but it was almost certainly not a 15. Thus, if label smoothing strength was 0.1, target probability distributions for 3, 8, and 15 were assigned as $(0.9, 0.1, 0.0)$, $(0.05, 0.9, 0.05)$, and $(0.0, 0.1, 0.9)$, respectively.

Hyperparameter optimization was performed over number of frames, frame sampling stride, slow lane downsampling ratio, optimizer type, whether or not to apply grayscale, data augmentation strength, and label smoothing strength. Final parameters selected were: 64 frames, even sampling throughout the video, 8x slow lane downsampling, RAdam optimization [138], grayscale application, 5°/10% data augmentation strength, and 0.1 label smoothing strength. Further discussion of the range of values tested and performance evolution is provided in Appendix D.2.

Models were trained for 100 epochs with categorical cross-entropy loss. Loss for each target class was weighted inversely to the frequency of that class in the dataset; e.g., loss for a video with target class 15 was multiplied by roughly 10 compared to loss for a video with target class 3. This encouraged the model to evenly distribute probabilities across the 3 classes if it was unsure, rather than guessing the most common class; without this scaling, models tended to get stuck in a local minimum of assigning label 3 to everything. The epoch which produced the highest validation accuracy was selected as the best model state, and these validation performances were compared to select the best set of hyperparameters. Out-of-sample performance from the best model was not measured until after hyperparameter optimization was completed.

**Fine-Tuning to Catheter Data**

As mentioned above, training a model to replicate cardiologist estimates of RAP can only hope to match human performance; RHC data gives us the potential to surpass human performance. However, our dataset of 438 train scans and 109 validation scans with associated RHC data was too small to train a new ML model from scratch. Instead, we utilized transfer learning to take advantage of all of the information about processing sniff scans that had already been learned from training to match cardiologist evaluations. To do this, the best classification model from the previous section was modified by replacing its output layer with a 1-dimensional output with no activation in order to convert it into a regression model. This new model was then trained to match RHC measurements. The optimized hyperparameters from

before were kept, but there were two more factors to tweak; layer freezing and loss function.

The first consideration was layer freezing. The goal of transfer learning is to get the model to utilize its prior information about common features in the system to generate answers in new ways. If the entire model is trained normally, instability in the new randomly initialized output layer may propagate backwards and erase feature knowledge in earlier layers. The extreme solution to this is to freeze all pre-existing layers, such that only the weights of the final output layer can be changed. However, in our system which had been pre-trained to replicate the sniff test as defined in Table 6.1, this type of freeze would have prevented any discovery of new useful features not related to existing standards. After testing various compromises between these two extremes, the strategy which produced the best validation error was training the whole model, but assigning a 3x lower learning rate to all pre-existing layers compared to that of the output layer. This allowed the model to learn to make use of existing features without erasing them while maintaining the possibility of morphing these features to better fit the new data.

The second consideration was the loss function. The standard loss function for regression tasks is mean squared error, or L2. However, this loss function was not well suited for our problem. To see why, consider a data point with a target of 11 mmHg and two different predictions; 4 mmHg or 21 mmHg. L2 loss would penalize the second prediction twice as strongly as the first, as the squared error would be 100 mmHg$^2$ compared to 49 mmHg$^2$. However, from a clinical standpoint, the first error is quite significant as someone with a problematic RAP has been given a healthy prediction. The second error, in contrast, is less significant, as both the prediction and target are unhealthly pressures that would likely lead to similar courses of treatment. To address this, we first applied a LeakyReLU transform to soft-cap outputs and targets to the (2, 15) mmHg range; beyond that, being extra-low or extra-high does not have much clinical significance. We then applied a log transform to the outputs and targets before getting L2 loss, reflecting the fact that each incremental increase in pressure is more clinically significant at the lower end of the spectrum.

With these adjustments in place, RHC-based training of the model was performed for 50 epochs, with early stopping based on best validation accuracy when converting predictions and targets back into 3/8/15 bins. To reduce noise in validation

performance, the 109 validation scans were augmented 5x with random rotations and brightness/contrast scaling.

## 6.5 Results and Discussion

**Echo-Only Results**

The first phase of model training and evaluation relied only on TTE videos of sniff tests and cardiologists' interpretations of those sniff tests. The model with the best validation performance, as measured by average classification accuracy across the three target classes, was run on the 1583 test scans to gauge out-of-sample performance. The average model accuracy, weighted by the frequency of each target category in the overall dataset, was 77.3% (see Table 6.4). Looking at the off-target predictions, it is worth noting that mis-classifying a targeted 3 as an 8 was significantly more likely that mis-classifying a targeted 3 as a 15, and similarly the 15→8 mis-call rate was higher than the 15→3 mis-call rate. This indicates that the model was able to learn that category 3 is "closer" to 8 than it is to 15, despite the fact that the model architecture treated all categories equally and independently with no inherent encoding of which ones were closer to each other. A similar pattern persisted even if label smoothing, the only bit of category-asymmetric information in the training procedure, was turned off.

**RAP Prediction Performance: Full Test Set**

| Confusion Matrix | Predicted 3mmHg | Predicted 8mmHg | Predicted 15mmHg |
|---|---|---|---|
| **Cardiologist 3mmHg** | 952 | 167 | 29 |
| **Cardiologist 8mmHg** | 77 | 132 | 37 |
| **Cardiologist 15mmHg** | 31 | 44 | 113 |
| **Accuracy** | **77.3%** | | |
| **False Negative** | **16.5%** | | |

Table 6.4: Out-of-sample performance of the echo-based RAP classification model, evaluated on the full test set of 1583 scans. Green boxes represent correct predictions, and the red box represents a clinically problematic false negative prediction. The overall accuracy of 77.3% represents a weighted average of accuracy in each target class, with weights based on observed frequency of each class across the dataset. False negative rate represents the percentage of patients classified as 15 mmHg by cardiologists (7.5% of the patient population) which the model classified as 3 mmHg.

For performance evaluation, it is also worth considering that not all wrong answers carry the same clinical significance. If a model like this were deployed as a screening tool in a clinical setting, the most problematic error would be a false negative, i.e., analyzing a scan from a patient who truly had an RAP in the 15 mmHg range and placing that patient in the 3 mmHg category. This could result in a patient who needed immediate treatment for volume overload instead being evaluated as healthy and not given follow-up testing or treatment. From this perspective, these results show a false negative rate 16.5%. The ideal screening model would have a higher sensitivity for 15mmHg predictions and a correspondingly high specificity for 3mmHg predictions.

As mentioned above, roughly 25% of this test set consisted of scans in which the IVC was either obscured or not present at all. For obscured IVCs the model may have been able to extract some information, but for scans with no IVC at all the model was essentially forced to guess at the answer. We hypothesized that these guesses could be a significant contributor to model errors, particularly the false negative rate. We tested this hypothesis by evaluating model performance on both a manually trimmed test set, which eliminated these 25% of scans without a clear IVC, and the strictly filtered "high-quality" set (described above) which only kept the 35% of scans with the highest scores from the IVC classification model. Test performance from both of these datasets is shown below in Table 6.5.

Increasing stringency on the test set only slightly increased overall accuracy. However, even an initial trimming to remove the worst data from the test set dropped the clinically problematic false negative rate by almost half (from 16.5% to 9.3%), and applying the strict high-quality IVC threshold dropped this rate even further to 5.6%. Both dataset reduction procedures were blinded to any RAP information (either cardiologist estimates or model predictions), so this should accurately represent the performance we could expect from higher-quality data inputs. Obtaining this higher-quality data would be possible in a clinical setting, as the operator could trigger ML-based RAP classification only once the IVC classification model indicated that a good view of the IVC had been obtained.

Further examination of the confusion matrices in Table 6.5 shows that the extreme false positive rate (true 3mmHg classified as 15mmHg by the model) also fell as data quality increased. As a result, in the high-quality test set only 6% of incorrect predictions were in the 3→15 or 15→3 categories; the remaining 94% were off by 1. Some portion of these "errors" were likely due to human variability in generating

**RAP Prediction Performance: Manually Trimmed Test Set**

| Confusion Matrix | Predicted 3mmHg | Predicted 8mmHg | Predicted 15mmHg |
|---|---|---|---|
| **Cardiologist 3mmHg** | 678 | 130 | 17 |
| **Cardiologist 8mmHg** | 60 | 116 | 35 |
| **Cardiologist 15mmHg** | 14 | 34 | 103 |
| **Accuracy** | **77.5%** | | |
| **False Negative** | **9.3%** | | |

**RAP Prediction Performance: High-Quality IVC Test Set**

| | | | |
|---|---|---|---|
| **Cardiologist 3mmHg** | 328 | 44 | 4 |
| **Cardiologist 8mmHg** | 41 | 58 | 7 |
| **Cardiologist 15mmHg** | 4 | 30 | 38 |
| **Accuracy** | **80.3%** | | |
| **False Negative** | **5.6%** | | |

Table 6.5: Out-of-sample performance of the echo-based RAP classification model, evaluated on the manually trimmed test set of 1187 scans and the high-quality IVC test set of 554 scans. The overall accuracy only increased slightly, from 77.3% originally to 77.5% on the trimmed set and 80.3% on the high-quality set. However, the clinically problematic false negative rate went down substantially, from 16.5% originally to 9.3% on the trimmed set and 5.6% on the high-quality set.

the target estimates rather than underlying deficiencies in the model. For example, as mentioned in the introduction, guidelines give a range of possible locations for measuring the diameter of the IVC, but the IVC generally does not have a constant diameter over this range. If the diameter changed from 20 mm at one end to 22 mm at the other, the "correct" RAP classification based on Table 6.1 is not clear, and different doctors may reasonably come to different conclusions. Indeed, prior studies which had the same IVC scans independently analyzed by multiple trained physicians to assess diameter and CI have found significant inter-operator variability [139], [140], and those that specifically looked at inter-operator agreement rates for RAP assessment found agreement rates of 70-75% [141], [142]. These studies were not performed under identical conditions to those at UCSF, but if we assume that the conditions leading to this variability are broadly similar between different medical

centers it seems plausible that our model performance is nearing the limit set by underlying uncertainty in the target data.

**RHC Results**

Right heart catheter measurements provided us with a means of assessing model performance in a way that was independent of human interpreter variability. One limitation of our RHC data, as discussed above, is that the RHC and TTE measurements were generally not taken on the same day, and the patient's true RAP may have changed in the intervening time. We can get a picture of the effect of this variance by analyzing the accuracy of cardiologists' sniff-based assessments of RAP compared to RHC measurements with different thresholds of allowed time separation between the two measurements, as shown in Figure 6.3. As expected, widening the allowable time window significantly increased the total number of data points, from 85 which occurred on the same day to 596 which occurred within 30 days of each other. Somewhat surprisingly, however, accuracy did not drop very sharply as we widened our time window, going from 49.4% for same-day measurements to 48.2% for all measurements with any time separation. This seems to indicate that variation in true RAP between RHC and TTE measurements was not a significant error contributor in our dataset. Based on consultation with UCSF cariologists about the tradeoff between dataset size and accuracy, we chose a 30-day time window. While the width of this window created some risk of inaccuracy in the gold standard reference, errors should be reflected equally in both cardiologist and model predictions since they were both made based on the same TTE data.

We can also evaluate our dataset in comparison to a prior study which was specifically designed to evaluate the accuracy of the sniff test. In Magnino, Omede, Avenatti, *et al.* [129], 153 patients had their RAP measured by both RHC and a sniff test in quick succession. The sniff test was used to generate an RAP estimate based on multiple different standards which have previously been proposed, including the one in Table 6.1. A summary of the results of this study, compared to RHC and TTE data from UCSF with a 30-day time window applied, is shown below in Table 6.6. We may expect the Magnino et al. dataset to yield better results, as the sniff test and RHC were always conducted in quick succession in a single standardized manner. Suprisingly, however, the UCSF dataset actually showed higher overall accuracy and an effectively equivalent false negative rate, despite the time gap between sniff test and RHC measurements. This could plausibly be due to differences in patient populations or greater experience of UCSF cardiologists in applying the specific
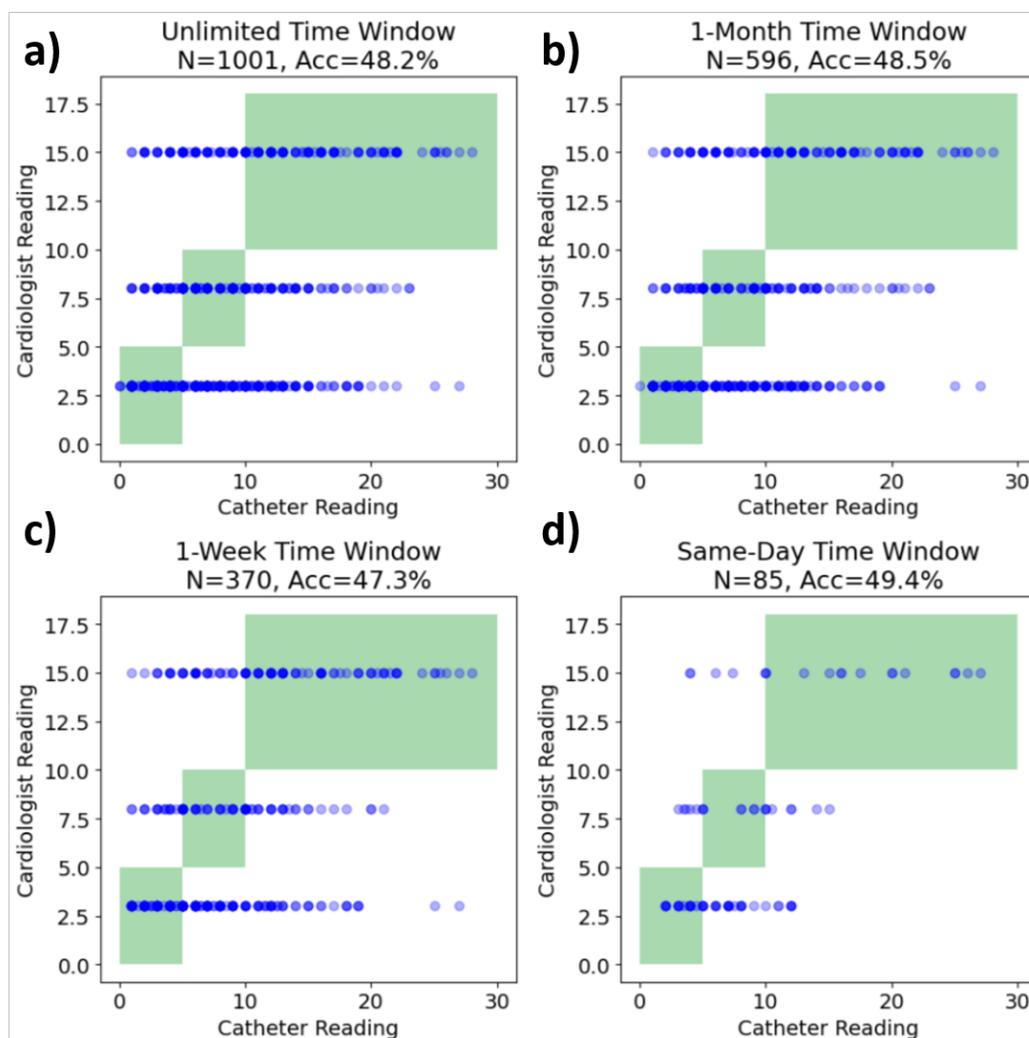
Figure 6.3: Accuracy of cardiologist RAP estimates compared to true RAP values from right heart catheterization, with various time windows allowed between cardiologist and RHC measurements: a) unlimited time, b) 1 month, c) 1 week, or d) same day. Blue circle represent individual measurement pairs, and green boxes are the "correct" prediction ranges. Reducing the time window significantly reduced the amount of data available, but surprisingly only moderately increased prediction accuracy.

sniff test standards from Table 6.1, as compared to physicians in Magnino et al. who were evaluating multiple different sniff test standards. Regardless of the reasons, the favorable accuracy levels of the UCSF dataset compared to prior literature provide further evidence that this dataset can be effectively used for evaluating ML model performance.

Before using the RHC data for further model training, we evaluated the performance of the best echo-only model on our "golden" test set which had echo video data, a

**Data from Magnino et al. 2017**

| RHC Value Range | Cardiologist Estimate | | | |
|---|---|---|---|---|
| | N | 3mmHg | 8mmHg | 15mmHg |
| **[0, 5] mmHg** | 17 | 8 | 6 | 3 |
| **(5, 10] mmHg** | 68 | 31 | 27 | 10 |
| **(10, 30] mmHg** | 68 | 19 | 19 | 30 |
| **Accuracy (Scaled)** | | 43.8% | | |
| **False Negative** | | 27.9% | | |

**UCSF Data, 30-day Window**

| | N | 3mmHg | 8mmHg | 15mmHg |
|---|---|---|---|---|
| **[0, 5] mmHg** | 118 | 84 | 24 | 10 |
| **(5, 10] mmHg** | 104 | 56 | 28 | 20 |
| **(10, 30] mmHg** | 97 | 26 | 27 | 44 |
| **Accuracy (Scaled)** | | 48.9% | | |
| **False Negative** | | 26.8% | | |

Table 6.6: Accuracy of sniff-based estimates of RAP compared to RHC measurements of RAP from Magnino et al. 2017 (left) [129] and our UCSF dataset (right). Scaled accuracy for both datasets calculated as a weighted average of accuracy in each target class, with weights for both classes determined by frequency of each class in the UCSF dataset. Magnino et al. may have been expected to yield better results as the sniff test and RHC were always performed in quick succession, while in the UCSF data they may have been performed up to a month apart. However, we see that the UCSF data actually yielded better overall accuracy and effectively equivalent false negative rate.

sniff-based RAP estimate from a UCSF cardiologist, and an associated RHC measurement within 30 days. This performance is summarized in Table 6.7. We can see that, even though the model only agreed with cardiologist predictions roughly 77% of the time in the original test set (see Table 6.4), the performance of the model with respect to RHC results was effectively equivalent to that of the cardiologists, both in terms of overall accuracy and the false negative rate. Applying a categorical chi-squared test with 8 degrees of freedom to compare the distribution of predictions between the cardiologists and echo-only ML model (using *scipy.stats.chi2_contingency*) yielded $p = 0.98$, indicating that the two distributions

are statistically indistinguishable. These results provide strong evidence that the ML model effectively replicated the performance of cardiologists in analyzing sniff tests.

**RAP Prediction Performance: Echo-Only ML Model**

| RHC Value Range | Predicted 3mmHg | Predicted 8mmHg | Predicted 15mmHg |
|:---:|:---:|:---:|:---:|
| [0, 5] mmHg | 78 | 30 | 10 |
| (5, 10] mmHg | 50 | 35 | 19 |
| (10, 30] mmHg | 27 | 25 | 45 |
| Accuracy | 49.5% | | |
| False Negative | 27.8% | | |

**RAP Prediction Performance: Fine-Tuned ML Model**

| [0, 5] mmHg | 42 | 62 | 14 |
|:---:|:---:|:---:|:---:|
| (5, 10] mmHg | 17 | 59 | 28 |
| (10, 30] mmHg | 7 | 32 | 58 |
| Accuracy | 49.8% | | |
| False Negative | 7.2% | | |

Table 6.7: Accuracy of RAP measurements from ML models vs. RHC. The top table shows performance for the "Echo-Only" model, which was trained to match cardiologist estimates. The performance of this model is effectively equivalent to that of cardiologists, in terms of both overall accuracy and false negative rate. The bottom table shows performance for the "Fine-Tuned" model, which was also trained to match RHC measurements. This model maintains an equivalent overall accuracy, but shows a drastically reduced false negative rate.

We also attempted to push performance further by taking the best echo-only model and fine-tuning it using our smaller set of training data with RHC measurements. The results from this fine-tuned model are shown in the lower half of Table 6.7. We can see that the fine-tuning had little effect on overall accuracy; however, there was a drastic change in the false negative rate, which dropped by over 70% compared to either cardiologist estimates or the echo-only ML model. This led to an overall improvement in the accuracy of classifying patients in the 15 mmHg category, which went from 45.4% with cardiologist estimates to 59.8% with the fine-tuned

ML model. Such an improvement would represent a substantial decrease in the number of patients with elevated RA pressure being falsely given a clean bill of health.

Another way to look at the results is to consider how often each measurement method yielded results of 3, 8, or 15 mmHg, as shown in Table 6.8. The true distribution for the UCSF patient population can be read from the RHC results, which showed a roughly equal proportion across the three classes. We can see that cardiologists tended to underestimate RAP from the sniff test, with over half of patients assessed as 3 mmHg and fewer than a quarter assessed as 15 mmHg. This pattern of underestimation was replicated in the Magnino et al. results, indicating that it is not just an artifact of our dataset. The Echo-Only ML model which was trained to match cardiologist predictions, unsurprisingly, exhibited an equivalent tendency towards underestimation. The Fine-Tuned ML model, in contrast, over-corrected a bit, exhibiting underprediction of 3 mmHg and overprediction of 8 mmHg. If such an ML model were used as an initial patient screening tool upon admission this tendency towards overestimation may be preferable, as an overly high RAP estimate from the model would be resolved after the patient was referred for further testing while an underestimate may go undetected if the patient is deemed healthy and not referred for further testing.

| Measurement Type | 3 mmHg Frequency | 8 mmHg Frequency | 15 mmHg Frequency |
|---|---|---|---|
| RHC | 37.0% | 32.6% | 30.4% |
| UCSF Cardiologist | 52.0% | 24.8% | 23.2% |
| Echo-Only ML | 48.6% | 28.2% | 23.2% |
| Fine-Tuned ML | 20.7% | 48.0% | 31.3% |
| RHC (Magnino et al. 2017) | 11.1% | 44.4% | 44.4% |
| Cardiologist (Magnino et al. 2017) | 37.9% | 34.0% | 28.1% |

Table 6.8: The frequency of different RAP range measurements across different measurement methods, including both the current study (top section) and Magnino et al. (bottom section). Cardiologists tended to underestimate true RAP in both the UCSF and Magnino et al. data, and this pattern was reflected in the echo-only model which was trainined on cardiologist predictions. The fine-tuned model, in contrast, over-corrected a bit and tended to overestimate true RAP.

## 6.6   Conclusion

Our results indicate that our ML model was able to effectively replicate the cardiologist interpretations on which it was trained, and fine-tuning on RHC data may

have even allowed it to surpass human performance in identifying elevated RAP. The diversity of the training and testing datasets used for these models, covering multiple years of scans by many doctors across a wide array of patients, make these results likely to generalize well to future applications. Integration of this tool into existing ultrasound machines could allow for more consistent and rapid assessment of RAP in medical centers with limited access to cardiologists. Further integration of the IVC and Sniff identification models as well could improve the process by automatically identifying high-quality sniff scans in real time, improving model performance and potentially allowing measurements to be performed even by relatively inexperienced ultrasound operators. These advancements would allow for faster and more accurate identification of volume overload in a range of medical settings, improving patient care and reducing the health burden associated with heart failure.

There are multiple avenues to further improve upon this work. Obtaining more RHC data in particular could allow for more extensive training against this gold standard, allowing the model to push further past current levels of human performance. Performance improvements could also potentially be obtained by integrating other sources of data, such as TTE scans of other heart regions or other measurement modalities such as electrocardiograms. Such improvements could reveal new indicators for elevated RAP outside of the sniff test, potentially improving clinical evaluations and patient care even when ML is not applied. Finally, improvements to model analysis and architecture could attempt to make the model's predictions explainable, allowing for easier identification of edge cases which cause errors and increasing physician trust.

*Chapter 7*

# CONCLUSION

The development of ultrasound imaging technology has provided huge benefits to many forms of medical care. No other technology can provide precise, real-time views under a patient's skin in a form that can be carted to the patient's bedside or even, in some modern iterations, operated from an iPad. Throughout its history ultrasound has primarily been viewed as a tool for generating images, relying on human interpretation to translate what is seen into medically useful information. However, recent advances have shown that ultrasound is also capable of much more, providing insight into a variety of biophysical parameters that are invisible to the human eye. My work has advanced this line of inquiry by applying ultrasound across two separate projects to the measurement of one of the most important human vital signs: blood pressure.

The first project focused on the development of CARDI-BP, an entirely new concept for arterial blood pressure measurement. While much prior research has been devoted to continuous non-invasive blood pressure measurement through many different means, all of these efforts have measured some proxy for blood pressure and used an empirically calibrated mapping to create an approximate BP measurement. This is the first work to describe a first-principles physics model which converts observed parameters into an absolute, calibration-free BP measurement. The core of this model is the measurement of circumferential resonance, which had never been observed in real arteries and was not guaranteed to persist in these more complex *in vivo* systems. Despite this, in a clinical study covering a demographically diverse cohort we were able to detect resonance in 59 out of 60 subjects. Our physics model was able to convert these resonance measurements into clinically relevant blood pressure and stiffness measurements, and the BP measurements had a favorable level of accuracy compared to the gold standard invasive catheter. Further work to embody CARDI-BP into a commercially viable form factor will greatly expand access to outpatient diagnosis of cardiovascular problems by revealing continuous, central BP information that has previously only been measured for those in critical care.

The second project focused on the noninvasive measurement of blood pressure in the right atrium (RAP), an important indicator for managing care in patients with heart failure. In this case, there is already a clinically accepted ultrasound-based method of non-invasively assessing RAP, but results are difficult to interpret and can yield inconsistent measurements even from trained physicians. This project applied machine learning techniques to a UCSF database of over 15000 expert cardiologist assessments of RAP, a dataset drastically larger and more diverse than any previously used to address this problem. When compared to gold standard measurements from invasive cathethers, the machine learning model yielded RAP measurements that were statistically indistinguishable from those of UCSF cardiologists. Deployment of this technology could democratize access to expert RAP measurements beyond large medical centers with cardiology departments to anywhere with an ultrasound machine, improving care for tens of millions of people worldwide suffering from heart failure.

The common thread underlying both of these projects is the development of new ways of looking at biomedical measurement through interdisciplinary research. The first project combined biomechanical models of the arterial wall with pipe resonance models from aerospace engineering, leading to a new model of resonance-based pressure measurement unique to the arterial system. The second project combined the medical intuition of physicians with the pattern recognition power of machine learning, leading to model-based measurement ability on par with expert cardiologists. It is my hope that these results prove useful to the scientific community and inspire more research that approaches biological problems from a new angle.

# BIBLIOGRAPHY

[1] M. Couade, M. Pernot, C. Prada, *et al.*, "Quantitative assessment of arterial wall biomechanical properties using shear wave imaging," *Ultrasound in Medicine & Biology*, vol. 36, no. 10, pp. 1662–1676, 2010, ISSN: 0301-5629. DOI: `10.1016/j.ultrasmedbio.2010.07.004`.

[2] Z. Zhang, G.-Y. Li, Y. Jiang, *et al.*, "Noninvasive measurement of local stress inside soft materials with programmed shear waves," *Science Advances*, vol. 9, no. 10, eadd4082, 2023. DOI: `10.1126/sciadv.add4082`.

[3] K. T. Mills, A. Stefanescu, and J. He, "The global epidemiology of hypertension," *Nature Reviews Nephrology*, vol. 16, no. 4, pp. 223–237, 2020, ISSN: 1759-507X. DOI: `10.1038/s41581-019-0244-2`.

[4] F. D. Fuchs and P. K. Whelton, "High blood pressure and cardiovascular disease," *Hypertension*, vol. 75, pp. 285–292, 2020. DOI: `10.1161/HYPERTENSIONAHA.119.14240`.

[5] P. Muntner *et al.*, "Measurement of blood pressure in humans: A scientific statement from the American Heart Association," *Hypertension*, vol. 73, e35–e66, 2019. DOI: `10.1161/HYP.0000000000000087`.

[6] M. F. Fallouh and A. Arabi, "Hemodynamic corner," *Heart Views*, vol. 21, no. 4, pp. 281–283, Oct. 2020, ISSN: 1995-705X. DOI: `10.4103/HEARTVIEWS.HEARTVIEWS_37_20`.

[7] D. Chemla *et al.*, "Subendocardial viability ratio estimated by arterial tonometry: A critical evaluation in elderly hypertensive patients with increased aortic stiffness," *Clinical and Experimental Pharmacology and Physiology*, vol. 35, pp. 909–915, 2008. DOI: `10.1111/j.1440-1681.2008.04927.x`.

[8] G. Moody and L. Lehman, "Predicting acute hypotensive episodes: The 10th annual physionet/computers in cardiology challenge," in *Proceedings of Computers in Cardiology Conference*, vol. 36, IEEE, 2009, pp. 541–544.

[9] P. Boutouyrie *et al.*, "Association between local pulse pressure, mean blood pressure, and large-artery remodeling," *Circulation*, vol. 100, pp. 1387–1393, 1999. DOI: `10.1161/01.CIR.100.13.1387`.

[10] A. G. Miller and A. J. Bardin, "Review of ultrasound-guided radial artery catheter placement," *Respiratory Care*, vol. 61, no. 3, pp. 383–388, 2016, ISSN: 0020-1324. DOI: `10.4187/respcare.04190`.

[11] G. Nuttall *et al.*, "Surgical and patient risk factors for severe arterial line complications in adults," *Anesthesiology*, vol. 124, pp. 590–597, 2016. DOI: `10.1097/ALN.0000000000000967`.

[12] N. Kallioinen, A. Hill, M. S. Horswill, H. E. Ward, and M. O. Watson, "Sources of inaccuracy in the measurement of adult patients' resting blood pressure in clinical settings: A systematic review," *Journal of Hypertension*, vol. 35, pp. 421–441, 2017. DOI: 10.1097/HJH.0000000000001197.

[13] S. J. Dankel, M. Kang, T. Abe, and J. P. Loenneke, "A meta-analysis to determine the validity of taking blood pressure using the indirect cuff method," *Current Hypertension Reports*, vol. 21, no. 1, p. 11, 2019, ISSN: 1534-3111. DOI: 10.1007/s11906-019-0929-8.

[14] B. P. Imholz, W. Wieling, G. A. van Montfrans, and K. H. Wesseling, "Fifteen years experience with finger arterial pressure monitoring: Assessment of the technology," *Cardiovascular Research*, vol. 38, pp. 605–616, 1998. DOI: 10.1016/S0008-6363(98)00067-4.

[15] B. M. Deegan *et al.*, "The effect of blood pressure calibrations and transcranial doppler signal loss on transfer function estimates of cerebral autoregulation," *Medical Engineering & Physics*, vol. 33, pp. 553–562, 2011. DOI: 10.1016/j.medengphy.2010.12.007.

[16] V. Eley, R. Christensen, L. Guy, *et al.*, "Clearsight™ finger cuff versus invasive arterial pressure measurement in patients with body mass index above 45 kg/m$^2$," *BMC Anesthesiology*, vol. 21, p. 152, 1 2021, ISSN: 1471-2253. DOI: 10.1186/s12871-021-01374-x.

[17] A. Bugarini, A. J. Young, C. J. Griessenauer, *et al.*, "Perioperative continuous noninvasive arterial pressure monitoring for neuroendovascular interventions: Prospective study for evaluation of the vascular unloading technique," *World Neurosurgery*, vol. 153, e195–e203, 2021, ISSN: 1878-8750. DOI: 10.1016/j.wneu.2021.06.095.

[18] P. Shaltis, A. Reisner, and H. Asada, "Calibration of the photoplethysmogram to arterial blood pressure: Capabilities and limitations for continuous pressure monitoring," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, IEEE, 2005, pp. 3970–3973. DOI: 10.1109/IEMBS.2005.1615331.

[19] H. Samimi and H. R. Dajani, "Cuffless blood pressure estimation using cardiovascular dynamics," in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, IEEE, 2022, pp. 1–8.

[20] H. Shin and S. D. Min, "Feasibility study for the non-invasive blood pressure estimation based on ppg morphology: Normotensive subject study," *BioMedical Engineering OnLine*, vol. 16, no. 1, p. 10, 2017, ISSN: 1475-925X. DOI: 10.1186/s12938-016-0302-y.

[21] Y. Yoon, J. H. Cho, and G. Yoon, "Non-constrained blood pressure monitoring using ECG and PPG for personal healthcare," *Journal of Medical Systems*, vol. 33, no. 4, pp. 261–266, Aug. 2009, ISSN: 1573-689X. DOI: 10.1007/s10916-008-9186-0.

[22] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A neural network-based method for continuous blood pressure estimation from a PPG signal," in *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2013, pp. 280–283. DOI: `10.1109/I2MTC.2013.6555424`.

[23] S. S. Mousavi, M. Firouzmand, M. Charmi, M. Hemmati, M. Moghadam, and Y. Ghorbani, "Blood pressure estimation from appropriate and inappropriate ppg signals using a whole-based method," *Biomedical Signal Processing and Control*, vol. 47, pp. 196–206, 2019, ISSN: 1746-8094. DOI: `10.1016/j.bspc.2018.08.022`.

[24] Y. Zhang and Z. Feng, "A svm method for continuous blood pressure estimation from a ppg signal," ser. ICMLC 2017, Singapore, Singapore: Association for Computing Machinery, 2017, pp. 128–132, ISBN: 9781450348171. DOI: `10.1145/3055635.3056634`.

[25] M. Panwar, A. Gautam, D. Biswas, and A. Acharyya, "Pp-net: A deep learning framework for ppg-based blood pressure and heart rate estimation," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 000–10 011, 2020. DOI: `10.1109/JSEN.2020.2990864`.

[26] J. Fortin *et al.*, "A novel art of continuous noninvasive blood pressure measurement," *Nature Communications*, vol. 12, p. 1387, 2021. DOI: `10.1038/s41467-021-21271-8`.

[27] C. Pellaton, A. Vybornova, S. Fallet, *et al.*, "Accuracy testing of a new optical device for noninvasive estimation of systolic and diastolic blood pressure compared to intra-arterial measurements," *Blood Pressure Monitoring*, vol. 25, no. 2, pp. 105–109, 2020.

[28] K. Takazawa, H. Kobayashi, N. Shindo, N. Tanaka, and A. Yamashina, "Relationship between radial and central arterial pulse wave and evaluation of central aortic pressure using the radial arterial pulse wave," *Hypertension Research*, vol. 30, pp. 219–228, 2007. DOI: `10.1291/hypres.30.219`.

[29] X. Quan *et al.*, "Advances in non-invasive blood pressure monitoring," *Sensors (Basel)*, vol. 21, 2021. DOI: `10.3390/s21134273`.

[30] H.-M. Cheng, D. Lang, C. Tufanaru, and A. Pearson, "Measurement accuracy of non-invasively obtained central blood pressure by applanation tonometry: A systematic review and meta-analysis," *International Journal of Cardiology*, vol. 167, no. 5, pp. 1867–1876, 2013, ISSN: 0167-5273. DOI: `10.1016/j.ijcard.2012.04.155`.

[31] B. Ibrahim and R. Jafari, "Cuffless blood pressure monitoring from an array of wrist bio-impedance sensors using subject-specific regression models: Proof of concept," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, pp. 1723–1735, 2019. DOI: `10.1109/TBCAS.2019.2946661`.

[32] T. Huynh, R. Jafari, and W.-Y. Chung, "An accurate bioimpedance measurement system for blood pressure monitoring," *Sensors*, vol. 18, no. 7, p. 2095, Jun. 2018, ISSN: 1424-8220. DOI: 10.3390/s18072095.

[33] D. Buxi, J.-M. Redouté, and M. R. Yuce, "Blood pressure estimation using pulse transit time from bioimpedance and continuous wave radar," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 917–927, 2017. DOI: 10.1109/TBME.2016.2582472.

[34] D. Kireev, K. Sel, B. Ibrahim, *et al.*, "Continuous cuffless monitoring of arterial blood pressure via graphene bioimpedance tattoos," *Nature Nanotechnology*, vol. 17, no. 8, pp. 864–870, 2022, ISSN: 1748-3395. DOI: 10.1038/s41565-022-01145-w.

[35] H. Gesche, D. Grosskurth, G. Küchler, and A. Patzak, "Continuous blood pressure measurement by using the pulse transit time: Comparison to a cuff-based method," en, *European Journal of Applied Physiology*, vol. 112, no. 1, pp. 309–315, Jan. 2012, ISSN: 1439-6327. DOI: 10.1007/s00421-011-1983-3.

[36] C. Poon and Y. Zhang, "Cuff-less and noninvasive measurements of arterial blood pressure by pulse transit time," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, pp. 5877–5880. DOI: 10.1109/IEMBS.2005.1615827.

[37] R. Mukkamala *et al.*, "Toward ubiquitous blood pressure monitoring via pulse transit time: Theory and practice," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 1879–1901, 2015. DOI: 10.1109/TBME.2015.2441951.

[38] C.-S. Kim, A. M. Carek, O. T. Inan, R. Mukkamala, and J.-O. Hahn, "Ballistocardiogram-based approach to cuff-less blood pressure monitoring: Proof-of-concept and potential challenges," *IEEE Transactions on Biomedical Engineering*, vol. 65, pp. 2384–2391, 2018. DOI: 10.1109/TBME.2018.2797239.

[39] W. Seok, K. J. Lee, D. Cho, J. Roh, and S. Kim, "Blood pressure monitoring system using a two-channel ballistocardiogram and convolutional neural networks," *Sensors*, vol. 21, no. 7, p. 2303, Mar. 2021, ISSN: 1424-8220. DOI: 10.3390/s21072303.

[40] S.-H. Liu, B.-H. Zhang, W. Chen, C.-H. Su, and C.-L. Chin, "Cuffless and touchless measurement of blood pressure from ballistocardiogram based on a body weight scale," *Nutrients*, vol. 14, no. 12, p. 2552, Jun. 2022, ISSN: 2072-6643. DOI: 10.3390/nu14122552.

[41] C. Liao, O. Shay, E. Gomes, and N. Bikhchandani, "Noninvasive continuous blood pressure measurement with wearable millimeter wave device," in *2021 IEEE 17th International Conference on Wearable and Implantable Body*

*Sensor Networks (BSN)*, 2021, pp. 1–5. DOI: `10.1109/BSN51625.2021.9507020`.

[42] A. Avolio *et al.*, "Challenges presented by cuffless measurement of blood pressure if adopted for diagnosis and treatment of hypertension," *Pulse*, 2022. DOI: `10.1159/000522660`.

[43] V. Volovici, N. L. Syn, A. Ercole, J. J. Zhao, and N. Liu, "Steps to avoid overuse and misuse of machine learning in clinical research," *Nature Medicine*, vol. 28, pp. 1996–1999, 2022. DOI: `10.1038/s41591-022-01961-6`.

[44] B. W. Beulen *et al.*, "Toward noninvasive blood pressure assessment in arteries by using ultrasound," *Ultrasound in Medicine and Biology*, vol. 37, pp. 788–797, 2011. DOI: `10.1016/j.ultrasmedbio.2011.01.020`.

[45] J. Vappou, J. Luo, K. Okajima, M. D. Tullio, and E. E. Konofagou, "Non-invasive measurement of local pulse pressure by pulse wave-based ultrasound manometry (PWUM)," *Physiological Measurement*, vol. 32, 2011. DOI: `10.1088/0967-3334/32/10/012`.

[46] J. Seo, H.-S. Lee, and C. G. Sodini, "Non-invasive evaluation of a carotid arterial pressure waveform using motion-tolerant ultrasound measurements during the valsalva maneuver," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 163–174, 2021. DOI: `10.1109/JBHI.2020.2995344`.

[47] C. Wang *et al.*, "Monitoring of the central blood pressure waveform via a conformal ultrasonic device," *Nature Biomedical Engineering*, vol. 2, pp. 687–695, 2018. DOI: `10.1038/s41551-018-0287-x`.

[48] B. Jana, K. Oswal, S. Mitra, G. Saha, and S. Banerjee, "Windkessel model-based cuffless blood pressure estimation using continuous wave doppler ultrasound system," *IEEE Sensors Journal*, vol. 20, pp. 9989–9999, 2020. DOI: `10.1109/JSEN.2020.2990648`.

[49] H. Hu, H. Huang, M. Li, *et al.*, "A wearable cardiac ultrasound imager," *Nature 2023 613:7945*, vol. 613, no. 7945, pp. 667–675, Jan. 2023, ISSN: 1476-4687. DOI: `10.1038/s41586-022-05498-z`.

[50] C. Wang *et al.*, "Bioadhesive ultrasound for long-term continuous imaging of diverse organs," *Science*, vol. 377, no. 6605, pp. 517–523, Jul. 2022, ISSN: 10959203. DOI: `10.1126/science.abo2542`.

[51] A. M. Zakrzewski, A. Y. Huang, R. Zubajlo, and B. W. Anthony, "Real-time blood pressure estimation from force-measured ultrasound," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 11, pp. 2405–2416, 2018. DOI: `10.1109/TBME.2018.2873297`.

[52] R. N. Arnold and G. B. Warburton, "Flexural vibrations of the walls of thin cylindrical shells having freely supported ends," *Proceedings of the Royal Society A*, vol. 197, pp. 238–256, 1949. DOI: `10.1098/rspa.1949.0061`.

[53] D. J. Korteweg, "Ueber die fortpflanzungsgeschwindigkeit des schalles in elastischen röhren," *Annalen der Physik*, vol. 241, pp. 525–542, 1878. DOI: `10.1002/andp.18782411206`.

[54] J. C. Bramwell and A. V. Hill, "The velocity of pulse wave in man," *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, vol. 93, no. 652, pp. 298–306, 1922.

[55] Y. C. Fung, E. E. Sechler, and A. Kaplan, "On the vibration of thin cylindrical shells under internal pressure," *Journal of the Aerospace Sciences*, vol. 24, pp. 650–660, 1957. DOI: `10.2514/8.3934`.

[56] U. S. Lindholm, D. D. Kana, and H. N. Abramson, "Breathing vibrations of a circular cylindrical shell with an internal liquid," *Journal of the Aerospace Sciences*, vol. 29, pp. 1052–1059, 1962. DOI: `10.2514/8.9693`.

[57] G. B. Warburton, "Vibration of a cylindrical shell in an acoustic medium," *Journal of Mechanical Engineering Science*, vol. 3, pp. 69–79, 1961. DOI: `10.1243/JMES\_JOUR\_1961\_003\_011\_02`.

[58] K. K. Shung, "High frequency ultrasonic imaging," *Journal of Medical Ultrasound*, vol. 17, no. 1, pp. 25–30, 2009, ISSN: 0929-6441. DOI: `10.1016/S0929-6441(09)60012-6`.

[59] M. Johansson, A. Myredal, P. Friberg, and L. M. Gan, "High-resolution ultrasound showing increased intima and media thickness of the radial artery in patients with end-stage renal disease," *Atherosclerosis*, vol. 211, no. 1, pp. 159–163, 2010, ISSN: 0021-9150. DOI: `10.1016/j.atherosclerosis.2010.01.031`.

[60] B.-B. Zhang, Y.-J. Zhou, J. Du, *et al.*, "Comparison of very-high-frequency ultrasound assessment of radial arterial wall layers after first and repeated transradial coronary procedures," *Journal of Geriatric Cardiology*, vol. 14, no. 4, pp. 245–253, Apr. 2017, ISSN: 1671-5411. DOI: `10.11909/j.issn.1671-5411.2017.04.003`.

[61] C. Papadacci, T. Mirault, B. Dizier, M. Tanter, E. Messas, and M. Pernot, "Non-invasive evaluation of aortic stiffness dependence with aortic blood pressure and internal radius by shear wave elastography and ultrafast imaging," *Innovation and Research in Biomedical Engineering*, vol. 39, pp. 9–17, 2018. DOI: `10.1016/j.irbm.2017.12.001`.

[62] P. A. Hasgall *et al.*, *IT'IS database for thermal and electromagnetic parameters of biological tissues*, doi:10.13099/VIP21000-04-1, Feb. 2022.

[63] T. E. Carew, R. N. Vaishnav, and D. J. Patel, "Compressibility of the arterial wall," *Circulation Research*, vol. 23, no. 1, pp. 61–68, DOI: `10.1161/01.RES.23.1.61`.

[64] A. Karimi, T. Sera, S. Kudo, and M. Navidbakhsh, "Experimental verification of the healthy and atherosclerotic coronary arteries incompressibility via digital image correlation," *Artery Research*, vol. 16, pp. 1–7, 2016. DOI: `10.1016/j.artres.2016.08.002`.

[65] P. Skacel and J. Bursa, "Poisson's ratio and compressibility of arterial wall – improved experimental data reject auxetic behaviour," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 131, p. 105 229, 2022, ISSN: 1751-6161. DOI: `10.1016/j.jmbbm.2022.105229`.

[66] M. Johnson and J. M. Tarbell, "A Biphasic, Anisotropic Model of the Aortic Wall," *Journal of Biomechanical Engineering*, vol. 123, no. 1, pp. 52–57, Aug. 2000, ISSN: 0148-0731. DOI: `10.1115/1.1339817`.

[67] T. Khamdaeng, J. Luo, J. Vappou, P. Terdtoon, and E. E. Konofagou, "Arterial stiffness identification of the human carotid artery using the stress-strain relationship in vivo," *Ultrasonics*, vol. 52, no. 3, pp. 402–411, Mar. 2012, ISSN: 1874-9968. DOI: `10.1016/J.ULTRAS.2011.09.006`.

[68] J. Krejza, M. Arkuszewski, S. E. Kasner, *et al.*, "Carotid artery diameter in men and women and the relation to body and neck size," *Stroke*, vol. 37, no. 4, pp. 1103–1105, 2006. DOI: `10.1161/01.STR.0000206440.48756.f7`.

[69] J. M. Arnold, G. E. Marchiori, J. R. Imrie, G. L. Burton, P. W. Pflugfelder, and W. J. Kostuk, "Large artery function in patients with chronic heart failure. studies of brachial artery diameter and hemodynamics.," *Circulation*, vol. 84, no. 6, pp. 2418–2425, 1991. DOI: `10.1161/01.CIR.84.6.2418`.

[70] W. Wahood, S. Ghozy, A. Al-Abdulghani, and D. F. Kallmes, "Radial artery diameter: A comprehensive systematic review of anatomy," *Journal of NeuroInterventional Surgery*, vol. 14, no. 12, pp. 1274–1278, 2022, ISSN: 1759-8478. DOI: `10.1136/neurintsurg-2021-018534`.

[71] C. G. Caro, T. J. Pedley, R. C. Schroter, W. A. Seed, and K. H. Parker, *The Mechanics of the Circulation*, 2nd ed. Cambridge University Press, 2011. DOI: `10.1017/CBO9781139013406`.

[72] T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out*, 2nd ed. Academic Press, 2014. DOI: `10.1016/C2011-0-07261-7`.

[73] G. Fleury, R. Berriet, J. Y. Chapelon, *et al.*, "Safety issues for hifu transducer design," in *AIP Conference Proceedings*, vol. 754, 2005, p. 233. DOI: `10.1063/1.1901645`.

[74] J. A. Jensen, S. I. Nikolov, K. L. Gammelmark, and M. H. Pedersen, "Synthetic aperture ultrasound imaging," *Ultrasonics*, vol. 44, e5–e15, 2006, Proceedings of Ultrasonics International (UI'05) and World Congress on Ultrasonics (WCU), ISSN: 0041-624X. DOI: `10.1016/j.ultras.2006.07.017`.

[75] S. Jeon, E.-Y. Park, W. Choi, R. Managuli, K. jong Lee, and C. Kim, "Real-time delay-multiply-and-sum beamforming with coherence factor for in vivo clinical photoacoustic imaging of humans," *Photoacoustics*, vol. 15, p. 100 136, 2019, ISSN: 2213-5979. DOI: `10.1016/j.pacs.2019.100136`.

[76] A. R. Selfridge, G. S. Kino, and B. T. Khuri-Yakub, "A theory for the radiation pattern of a narrow-strip acoustic transducer," *Applied Physics Letters*, vol. 37, no. 1, p. 35, DOI: `10.1063/1.91692`.

[77] W. R. Hedrick and D. L. Hykes, "Autocorrelation detection in color doppler imaging: A review," *Journal of Diagnostic Medical Sonography*, vol. 11, pp. 16–22, 1995. DOI: `10.1177/875647939501100104`.

[78] B. Gustavsen and A. Semlyen, "Rational approximation of frequency domain responses by vector fitting," *IEEE Transactions on Power Delivery*, vol. 14, pp. 1052–1061, 1999. DOI: `10.1109/61.772353`.

[79] M. Earle, G. D. Portu, and E. DeVos, "Agar ultrasound phantoms for low-cost training without refrigeration," *African Journal of Emergency Medicine*, vol. 6, pp. 18–23, 2016. DOI: `10.1016/j.afjem.2015.09.003`.

[80] C. Richardson, S. Bernard, and V. A. Dinh, "A cost-effective, gelatin-based phantom model for learning ultrasound-guided fine-needle aspiration procedures of the head and neck," *Journal of Ultrasound in Medicine*, vol. 34, pp. 1479–1484, 2015. DOI: `10.7863/ultra.34.8.1479`.

[81] R. O. Bude and R. S. Adler, "An easily made, low-cost, tissue-like ultrasound phantom material," *Journal of Clinical Ultrasound*, vol. 23, pp. 271–273, 1995. DOI: `10.1002/jcu.1870230413`.

[82] "Non-invasive sphygmomanometers — part 2: Clinical investigation of intermittent automated measurement type," International Organization for Standardization, Geneva, CH, Standard, Nov. 2018.

[83] L. T. Hersh, B. Friedman, W. Luczyk, and J. Sesing, "Evaluation of filtering methods for acquiring radial intra-artery blood pressure waveforms," *Journal of Clinical Monitoring and Computing*, vol. 29, pp. 659–669, 2015.

[84] M. Schroeder, "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (corresp.)," *IEEE Transactions on Information Theory*, vol. 16, no. 1, pp. 85–89, 1970. DOI: `10.1109/TIT.1970.1054411`.

[85] S. J. Julier and J. K. Uhlmann, "New extension of the kalman filter to nonlinear systems," in *Signal processing, sensor fusion, and target recognition VI*, Spie, vol. 3068, 1997, pp. 182–193.

[86] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*, Ieee, 2000, pp. 153–158.

[87]  R. Labbe, *Filterpy: Kalman filters and other optimal and non-optimal estimation filters in python*, `https://github.com/rlabbe/filterpy`, 2015.

[88]  M. R. Rajamani, "Data-based techniques to improve state estimation in model predictive control," University of Wisconsin, Madison, 2007.

[89]  B. P. Hanley, W. Bains, and G. Church, "Review of scientific self-experimentation: Ethics history, regulation, scenarios, and views among ethics committees and prominent scientists," *Rejuvenation Research*, vol. 22, no. 1, pp. 31–42, 2019.

[90]  W. Pasma, L. M. Peelen, S. van Buuren, W. A. van Klei, and J. C. de Graaff, "Artifact Processing Methods Influence on Intraoperative Hypotension Quantification and Outcome Effect Estimates," *Anesthesiology*, vol. 132, no. 4, pp. 723–737, Apr. 2020, ISSN: 0003-3022. DOI: `10.1097/ALN.0000000000003131`.

[91]  J. E. Sharman, A. P. Avolio, J. Baulmann, *et al.*, "Validation of non-invasive central blood pressure devices: Artery society task force consensus statement on protocol standardization," *European Heart Journal*, vol. 38, no. 37, pp. 2805–2812, 2017.

[92]  C. G. Caro, T. J. Pedley, R. C. Schroter, W. A. Seed, and K. H. Parker, *The Mechanics of the Circulation*, 2nd ed. Cambridge University Press, 2011. DOI: `10.1017/CBO9781139013406`.

[93]  E. J. Kroeker and E. H. Wood, "Comparison of simultaneously recorded central and peripheral arterial pressure pulses during rest, exercise and tilted position in man," *Circulation Research*, vol. 3, no. 6, pp. 623–632, 1955.

[94]  H. Miyashita, "Clinical assessment of central blood pressure," *Current Hypertension Reviews*, vol. 8, no. 2, pp. 80–90, 2012.

[95]  T. G. Papaioannou, A. D. Protogerou, K. S. Stamatelopoulos, M. Vavuranakis, and C. Stefanadis, "Non-invasive methods and techniques for central blood pressure estimation: Procedures, validation, reproducibility and limitations," *Current Pharmaceutical Design*, vol. 15, no. 3, pp. 245–253, 2009.

[96]  M. Applefeld, "The jugular venous pressure and pulse contour," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, H. Walker, W. Hall, and J. Hurst, Eds., 3rd ed., Boston: Butterworths, 1990, ch. 19. [Online]. Available: `https://www.ncbi.nlm.nih.gov/books/NBK300/`.

[97]  A. Dornhorst, P. Howard, and G. Leathart, "Respiratory variations in blood pressure," *Circulation*, vol. 6, no. 4, pp. 553–558, 1952.

[98]  A. Barison, A. Del Franco, and G. Todiere, "Haemodynamic findings in obstructive hypertrophic cardiomyopathy: Pulsus bisferiens and brockenbrough–braunwald–morrow sign," *Journal of Cardiovascular Medicine*, vol. 17, e154–e155, 2016.

[99]  C. M. McEniery, J. R. Cockcroft, M. J. Roman, S. S. Franklin, and I. B. Wilkinson, "Central blood pressure: Current evidence and clinical importance," *European Heart Journal*, vol. 35, no. 26, pp. 1719–1725, 2014.

[100]  P. Segers, E. R. Rietzschel, and J. A. Chirinos, "How to measure arterial stiffness in humans," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 40, no. 5, pp. 1034–1043, 2020.

[101]  B. Gaszner, Z. Lenkey, M. Illyés, *et al.*, "Comparison of aortic and carotid arterial stiffness parameters in patients with verified coronary artery disease," *Clinical Cardiology*, vol. 35, no. 1, pp. 26–31, DOI: `10.1002/clc.20999`.

[102]  "Aortic stiffness is an independent predictor of fatal stroke in essential hypertension," *Stroke*, vol. 34, no. 5, pp. 1203–1206, DOI: `10.1161/01.STR.0000065428.03209.64`.

[103]  S.-i. Katsuda, K. Takazawa, M. Miyake, D. Kobayashi, M. Kusanagi, and A. Hazama, "Local pulse wave velocity directly reflects increased arterial stiffness in a restricted aortic region with progression of atherosclerotic lesions," *Hypertension Research*, vol. 37, no. 10, pp. 892–900, 2014, ISSN: 1348-4214. DOI: `10.1038/hr.2014.96`.

[104]  C. Huang, Y. Su, H. Zhang, L.-X. Qian, and J. Luo, "Pulse wave velocity measurement in healthy and diseased carotid arteries in vivo," in *2015 IEEE International Ultrasonics Symposium (IUS)*, 2015, pp. 1–4. DOI: `10.1109/ULTSYM.2015.0003`.

[105]  A. P. Guérin, B. Pannier, F. Métivier, S. J. Marchais, and G. M. London, "Assessment and significance of arterial stiffness in patients with chronic kidney disease," en, *Current Opinion in Nephrology and Hypertension*, vol. 17, no. 6, pp. 589–594, 2008, ISSN: 1062-4821. DOI: `10.1097/MNH.0b013e3283105265`.

[106]  E. Messas, M. Pernot, and M. Couade, "Arterial wall elasticity: State of the art and future prospects," *Diagnostic and Interventional Imaging*, vol. 94, no. 5, pp. 561–569, 2013, Ultrasound elastography, ISSN: 2211-5684. DOI: `10.1016/j.diii.2013.01.025`.

[107]  P. M. Nabeel, V. R. Kiran, J. Joseph, V. V. Abhidev, and M. Sivaprakasam, "Local pulse wave velocity: Theory, methods, advancements, and clinical applications," *IEEE Reviews in Biomedical Engineering*, vol. 13, pp. 74–112, 2020. DOI: `10.1109/RBME.2019.2931587`.

[108] S. Hansen and M. Staber, "Oscillometric blood pressure measurement used for calibration of the arterial tonometry method contributes significantly to error," *European Journal of Anaesthesiology*, vol. 23, pp. 781–787, 2006. DOI: 10.1017/S0265021506000688.

[109] M. Vrahatis, G. Magoulas, and V. Plagianakos, "From linear to nonlinear iterative methods," *Applied Numerical Mathematics*, vol. 45, pp. 59–77, 2003. DOI: 10.1016/S0168-9274(02)00235-0.

[110] G. Savarese and L. H. Lund, "Global public health burden of heart failure," *Cardiac Failure Review*, vol. 3, no. 1, p. 7, 2017.

[111] A. Koratala, C. Ronco, and A. Kazory, "Diagnosis of fluid overload: From conventional to contemporary concepts," *Cardiorenal Medicine*, pp. 1–14, 2022.

[112] D. H. Ellison and G. M. Felker, "Diuretic treatment in heart failure," *New England Journal of Medicine*, vol. 377, no. 20, pp. 1964–1975, 2017.

[113] S. Macdonald, "Fluid resuscitation in patients presenting with sepsis: Current insights," *Open Access Emergency Medicine*, pp. 633–638, 2022.

[114] R. E. Hughes and G. J. Magovern, "The relationship between right atrial pressure and blood volume," *AMA Archives of Surgery*, vol. 79, no. 2, pp. 238–243, 1959.

[115] R. Nagata, T. Harada, K. Omote, *et al.*, "Right atrial pressure represents cumulative cardiac burden in heart failure with preserved ejection fraction," *ESC Heart Failure*, vol. 9, no. 2, pp. 1454–1462, 2022.

[116] J. V. Hull, M. R. Padkins, S. El Hajj, *et al.*, "Risks of right heart catheterization and right ventricular biopsy: A 12-year, single-center experience," *Mayo Clinic Proceedings*, vol. 98, no. 3, pp. 419–431, 2023, ISSN: 0025-6196. DOI: 10.1016/j.mayocp.2022.07.025.

[117] R. M. Lang, L. P. Badano, V. Mor-Avi, *et al.*, "Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the american society of echocardiography and the european association of cardiovascular imaging," *Journal of the American Society of Echocardiography*, vol. 28, no. 1, 1–39.e14, 2015, ISSN: 0894-7317. DOI: 10.1016/j.echo.2014.10.003.

[118] L. G. Rudski, W. W. Lai, J. Afilalo, *et al.*, "Guidelines for the echocardiographic assessment of the right heart in adults: A report from the american society of echocardiography: Endorsed by the european association of echocardiography, a registered branch of the european society of cardiology, and the canadian society of echocardiography," *Journal of the American Society of Echocardiography*, vol. 23, no. 7, pp. 685–713, 2010, ISSN: 0894-7317. DOI: 10.1016/j.echo.2010.05.010.

[119] S. J. Millington, "Ultrasound assessment of the inferior vena cava for fluid responsiveness: Easy, fun, but unlikely to be helpful," *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*, vol. 66, no. 6, pp. 633–638, 2019.

[120] K. Rollas, B. Kilicaslan, A. Erden, I. Kilicaslan, E. O. Ersoy, and S. B. Akinci, "The interrater reliability of inferior vena cava ultrasonography performed by intensive care fellows," *Journal of Critical & Intensive Care*, vol. 12, no. 2, 2021.

[121] M. Senthilnathan, P. Kundra, S. Mishra, S. Velayudhan, and A. Pillai, "Competence of intensivists in focused transthoracic echocardiography in intensive care unit: A prospective observational study," *Indian Journal of Critical Care Medicine*, vol. 22, no. 5, 2018.

[122] J. Bowra, V. Uwagboe, A. Goudie, C. Reid, and M. Gillett, "Interrater agreement between expert and novice in measuring inferior vena cava diameter and collapsibility index," *Emergency Medicine Australasia*, vol. 27, no. 4, pp. 295–299, 2015.

[123] Q. Huang, F. Zhang, and X. Li, "Machine learning in ultrasound computer-aided diagnostic systems: A survey," *BioMed Research International*, vol. 2018, 2018.

[124] S. Liu, Y. Wang, X. Yang, *et al.*, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, 2019.

[125] S. Albani, B. Pinamonti, T. Giovinazzo, *et al.*, "Accuracy of right atrial pressure estimation using a multi-parameter approach derived from inferior vena cava semi-automated edge-tracking echocardiography: A pilot study in patients with cardiovascular disorders," *The International Journal of Cardiovascular Imaging*, vol. 36, pp. 1213–1225, 2020.

[126] S. Wshah, B. Xu, J. Bates, and K. Morrissette, "Deep fusion of ultrasound videos for furosemide classification," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2022, pp. 1–4.

[127] M. Blaivas, L. Blaivas, G. Philips, *et al.*, "Development of a deep learning network to classify inferior vena cava collapse to predict fluid responsiveness," *Journal of Ultrasound in Medicine*, vol. 40, no. 8, pp. 1495–1504, 2021. DOI: 10.1002/jum.15527.

[128] L. Mesin, P. Policastro, A. Re, *et al.*, "271 automated real time echocardiographic tool for edge tracking of inferior vena cava and non-invasive estimation of right atrial pressure," *European Heart Journal Supplements*, vol. 24, no. Supplement_K, Dec. 2022, suac121.212, ISSN: 1520-765X. DOI: 10.1093/eurheartjsupp/suac121.212.

[129] C. Magnino, P. Omede, E. Avenatti, *et al.*, "Inaccuracy of right atrial pressure estimates through inferior vena cava indices," *The American Journal of Cardiology*, vol. 120, no. 9, pp. 1667–1673, 2017.

[130] C. Feichtenhofer, "X3d: Expanding architectures for efficient video recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 203–213.

[131] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "Slowfast networks for video recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6202–6211.

[132] D. Kondratyuk, L. Yuan, Y. Li, *et al.*, "Movinets: Mobile video networks for efficient video recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16 020–16 030.

[133] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" In *Proceedings of the International Conference on Machine Learning (ICML)*, Jul. 2021.

[134] G. Sharir, A. Noy, and L. Zelnik-Manor, "An image is worth 16x16 words, what is a video worth?" *arXiv preprint arXiv:2103.13915*, 2021.

[135] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lučić, and C. Schmid, "Vivit: A video vision transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 6836–6846.

[136] H. Fan, T. Murrell, H. Wang, *et al.*, "PyTorchVideo: A deep learning library for video understanding," in *Proceedings of the 29th ACM International Conference on Multimedia*, `https://pytorchvideo.org/`, 2021.

[137] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818–2826.

[138] L. Liu, H. Jiang, P. He, *et al.*, "On the variance of the adaptive learning rate and beyond," in *Proceedings of the Eighth International Conference on Learning Representations (ICLR 2020)*, Apr. 2020.

[139] N. M. Finnerty, A. R. Panchal, C. Boulger, *et al.*, "Inferior vena cava measurement with ultrasound: What is the best view and best mode?" *Western Journal of Emergency Medicine*, vol. 18, no. 3, p. 496, 2017.

[140] A. J. Weekes, H. M. Tassone, A. Babcock, *et al.*, "Comparison of serial qualitative and quantitative assessments of caval index and left ventricular systolic function during early fluid resuscitation of hypotensive emergency department patients," *Academic Emergency Medicine*, vol. 18, no. 9, pp. 912–921, 2011.

[141] J. M. Fields, P. A. Lee, K. Y. Jenq, D. G. Mark, N. L. Panebianco, and A. J. Dean, "The interrater reliability of inferior vena cava ultrasound by bedside clinician sonographers in emergency department patients," *Academic Emergency Medicine*, vol. 18, no. 1, pp. 98–101, 2011.

[142] M. R. Randazzo, E. R. Snoey, M. A. Levitt, and K. Binder, "Accuracy of emergency physician assessment of left ventricular ejection fraction and central venous pressure using echocardiography," *Academic Emergency Medicine*, vol. 10, no. 9, pp. 973–977, 2003.

[143] E.-F. Chou *et al.*, "Clinical validation of a soft wireless continuous blood pressure sensor during surgery," *Frontiers in Digital Health*, vol. 3, 2021. DOI: `10.3389/fdgth.2021.696606`.

[144] J. Harju, A. Vehkaoja, P. Kumpulainen, *et al.*, "Comparison of non-invasive blood pressure monitoring using modified arterial applanation tonometry with intra-arterial measurement," *Journal of Clinical Monitoring and Computing*, vol. 32, pp. 13–22, 2018.

[145] S.-H. Kim *et al.*, "Accuracy and precision of continuous noninvasive arterial pressure monitoring compared with invasive arterial pressure: A systematic review and meta-analysis," *Anesthesiology*, vol. 120, no. 5, pp. 1080–1097, 2014. DOI: `10.1097/ALN.0000000000000226`.

*Appendix A*

# SUMMARY OF PRIOR NIBPM STUDIES

Dozens of prior studies have attempted to produce continuous non-invasive blood pressure measurements using may different methods, but none has yet gained widespread clinical usage in replacing arterial catheters. To see why, we can consider four different aspects that relate to whether or not an NIBPM measurement method can be considered equivalent to an A-line:

- **Is calibration required?** Many NIBPM methods rely on an initial calibration against a reference BP measurement device, most commonly a standard brachial cuff. Variations in underlying physiology, such as the constriction or relaxation of vascular smooth muscle, also necessitate periodic re-calibration for long-term monitoring. This calibration reduces convenience, increases cost, and introduces potential errors from inaccuracy in the reference cuff.

- **Is machine learning used?** Some NIBPM methods attempt to replace direct calibration with indirect calibration via machine learning. This increases convenience, but also introduces significant risks. Limitations in the training data limit the generalization of model performance, particularly for extreme BPs or uncommon pathologies.

- **Is a full BP waveform generated?** One of the advantages of an A-line over a cuff is that variations in the full BP waveform shape can be indicative of cardiovascular pathologies that could not be identified from simple SYS/DIA readings. NIBPM methods that cannot produce full BP waveforms sacrifice this clinically valuable information.

- **Is there a physical relationship between BP and what is measured?** A-lines are considered highly reliable because they directly measure pressure inside of a blood vessel. In contrast, most NIBPM methods measure some form of proxy information which is generally correlated to BP but does not have a direct 1:1 physical relationship. Identifying such a physics-based relationship between observables and outputs would increase reliability of measurements.

Table A1.1 shows the CARDI-BP method in comparison to many other prior NIBPM studies on these four points. CARDI-BP is the only method in the literature that satisfies 1) no calibration, 2) no machine learning, 3) full waveform output, and 4) a full physical model relating observables to BP. Table A1.2 compares the performance of CARDI-BP in our $N = 60$ clinical study with some past studies that have compared NIBPM devices against an A-line. CARDI-BP compares favorably to many of these methods, including some commercially available devices.

| Method | Modality | As presented by... | Calibration source | Full wave-form? | Uses machine learning | Physical model? | Additional drawbacks/remarks |
|---|---|---|---|---|---|---|---|
| CARDI-BP | Ultrasound | This manuscript | **None** | Yes | No | Yes (**fully determined**) | |
| Radius tracking | | Wang *et al.* [47] | Brachial cuff | Yes | No | No | Core exponential equation is empirical |
| Pulse-wave velocity (QA method) | | Seo, Lee, and Sodini [46], Beulen *et al.* [44], Vappou, Luo, Okajima, *et al.* [45] | Finger cuff (Seo, Lee, and Sodini [46]), brachial cuff (others) | Yes | No | Yes | Physical model uses Moens-Korteweg/Bramwell-Hill |
| WFA | | Jana, Oswal, Mitra, *et al.* [48] | None / Brachial cuff (for training) | No | No* | No* | *Wendkessel provides underlying model, but feature extraction relies on linear regression. |
| Force-measured distension | | Zakrzewski, Huang, Zubajlo, *et al.* [51] | None | No | No* | Yes | *Requires steady manual applantation pressure and empirical model matching |

| Volume clamp | Finger Cuff | Imholz, Wieling, Montfrans, et al. [14] | Brachial cuff (optional but recommended) | Yes | No | Yes | Difficult to measure BP on patients with low perfusion in extremities, periodic Physiocal self-calibration results in lower data availability |
|---|---|---|---|---|---|---|---|
| Volume control + PPG | | Fortin et al. [26] | Brachial cuff | Yes | No | No | Difficult to measure BP on patients with low perfusion in extremities |
| WFA | PPG | Shaltis, Reisner, and Asada [18] and Shin and Min [20] | Finger cuff (Shaltis), Wrist Cuff (Shin) | No | No | No* | *Shin has an approximate physical model for a "pressure index" which must be converted to BP via empirical regression |
| WFA + PTT | | Yoon, Cho, and Yoon [21] | Brachial Cuff | No | No | No | |
| WFA + ML | PPG | Samimi and Dajani [19], Kurylyak, Lamonaca, and Grimaldi [22], Mousavi, Firouzmand, Charmi, et al. [23], Zhang and Feng [24], and Panwar, Gautam, Biswas, et al. [25] | None / A-Line or Brachial Cuff* (for training) | No | Yes | No | *BP data obtained from third-party databases of multi-signal vital sign data |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Proprietary WFA | PPG | Pellaton, Vybornova, Fallet, *et al.* [27] | A-Line | No | Unclear | Unclear | 30% of subjects excluded due to low signal quality, including all subjects with darker skin |
| WFA | Tonometry | Takazawa, Kobayashi, Shindo, *et al.* [28] | Brachial cuff | No | No | No | Sensitive to noise and movement artifacts |
| WFA + ML | | Quan *et al.* [29] | Brachial cuff | No | Yes | No | |
| Proprietary WFA | Radar | Liao, Shay, Gomes, *et al.* [41] | Finger Cuff | Yes | Unclear | Unclear | |
| WFA | Bioimpedance | Huynh, Jafari, and Chung [32] | Brachial Cuff | No | No | No | |
| WFA + ML | | Ibrahim and Jafari [31] and Kireev, Sel, Ibrahim, *et al.* [34] | Finger cuff | No | Yes | No | |
| PTT | Bioimpedance + Radar | Buxi, Redouté, and Yuce [33] | Brachial cuff | No | No | No | |
| PTT | ECG + PPG | Gesche, Grosskurth, Küchler, *et al.* [35] | Brachial Cuff | No | No* | No | *Best fit to an empirically determined nonlinear function |
| PTT | ECG + PPG | Poon and Zhang [36] | Brachial Cuff | No | No | Yes | Uses the Moens-Korteweg formula to convert PTT to BP differences |
| Ballistocardiog-raphy | Force Plate | Kim, Carek, Inan, *et al.* [38] | Finger cuff | No | No | No | Requires patient to stand/sit on force plate, not appropriate for ambulatory use |

| Ballistocardiog-raphy + ML | Force Plate | Seok, Lee, Cho, *et al.* [39] | Brachial Cuff | No | Yes | No | Requires patient to stand/sit on force plate, not appropriate for ambulatory use |
|---|---|---|---|---|---|---|---|
| Ballistocardiog-raphy + PTT | Force Plate + PPG | Liu, Zhang, Chen, *et al.* [40] | Brachial Cuff | No | No | No | |

Table A1.1: Comparison of noninvasive blood pressure measurement methods previously described in the literature. Acronyms used are: PPG (photoplethysmography), ECG (electrocardiogram), WFA (waveform feature analysis), and PTT (pulse transit time).

| Study | Method | # Subjects | | Standard Deviation Of Differences vs A-Line (mmHg) | | | Pearson's Correlation Coefficient | | | Commercial Device | Externally Calibrated |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Data Collected | Passed QC | DBP | MAP | SBP | DBP | MAP | SBP | | |
| Present Work | Arterial Resonance | 60 | 59 | 10.9 | 11.2 | 13.4 | 0.71 | 0.74 | 0.73 | No | No |
| Bugarini, Young, Griessenauer, *et al.* [17] | Volume Clamp (ClearSight) | 24 | 24 | 14.6 | 17.9 | 21.5 | 0.29 | 0.35 | 0.47 | Yes | No |
| Eley, Christensen, Guy, *et al.* [16] | Volume Clamp (ClearSight) | 32 | 30 | 10.8 | 10.9 | 14.1 | - | - | - | Yes | No |
| Chou *et al.* [143] | Capacitance (Vena Vitals) | 32 | 17 | 12.2 | - | 12.6 | 0.57 | - | 0.83 | No | Yes |
| Harju, Vehkaoja, Kumpulainen, *et al.* [144] | Tonometry (BPro) | 38 | 28 | 9.4 | 11.6 | 19.9 | 0.72 | 0.64 | 0.61 | Yes | Yes |
| Kim *et al.* [145] (Meta-Analysis) | Various | - | 919 | 8.3 | 8.4 | 12.2 | - | - | - | - | - |
| Kim *et al.* [145] (Meta-Analysis) | Various (Commercially Available) | - | - | 8.6 | 8.7 | 12.4 | - | - | - | Yes | - |

Table A1.2: Comparison between $N = 60$ clinical study statistics and other A-line–to–NIBPM studies in the literature.

# FLOWCHARTS OF DATA ANALYSIS PROCEDURES
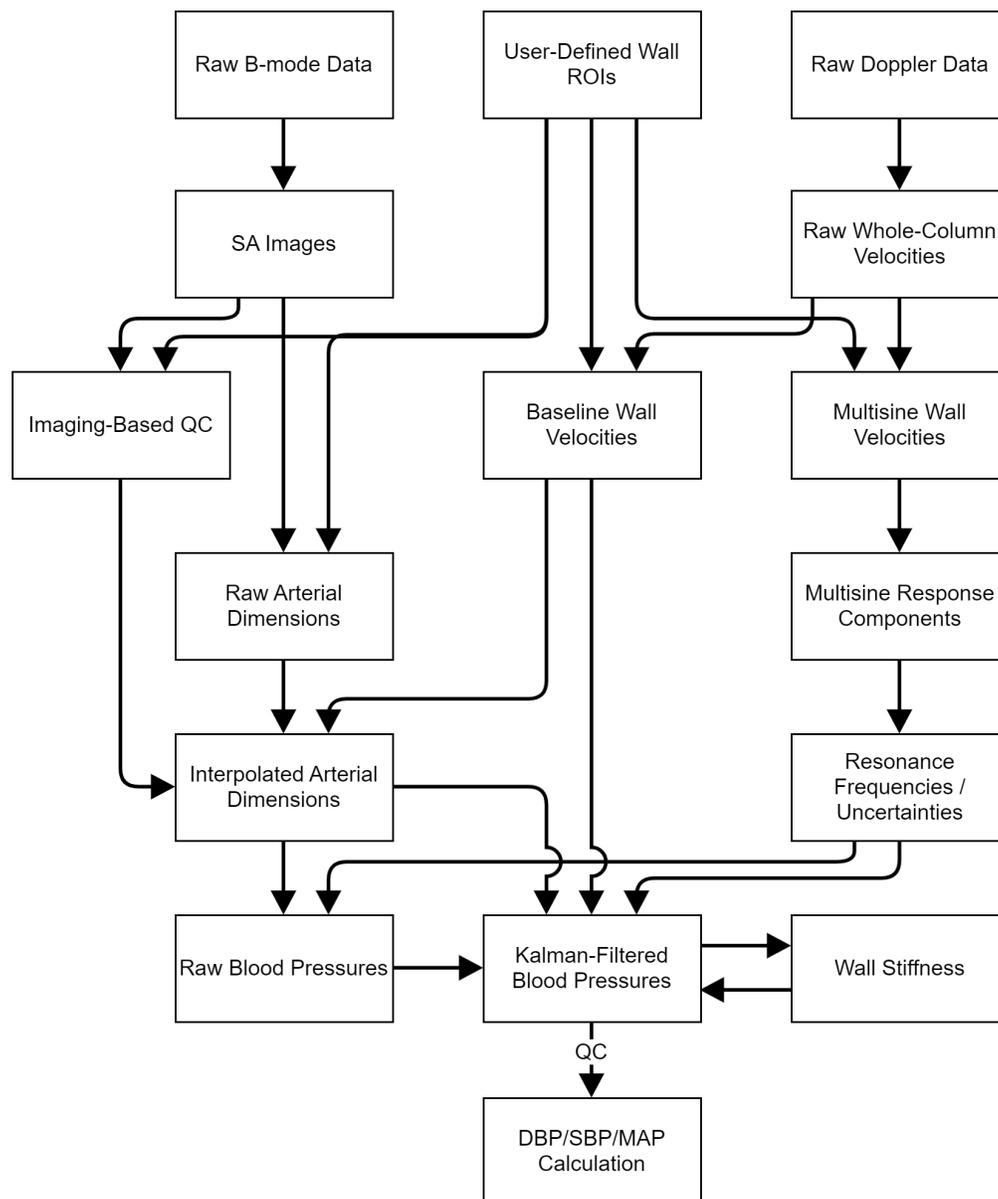


Figure A2.1: Flowchart describing the high-level data flow for BP measurement, from raw ultrasound data acquisition to final outputs.
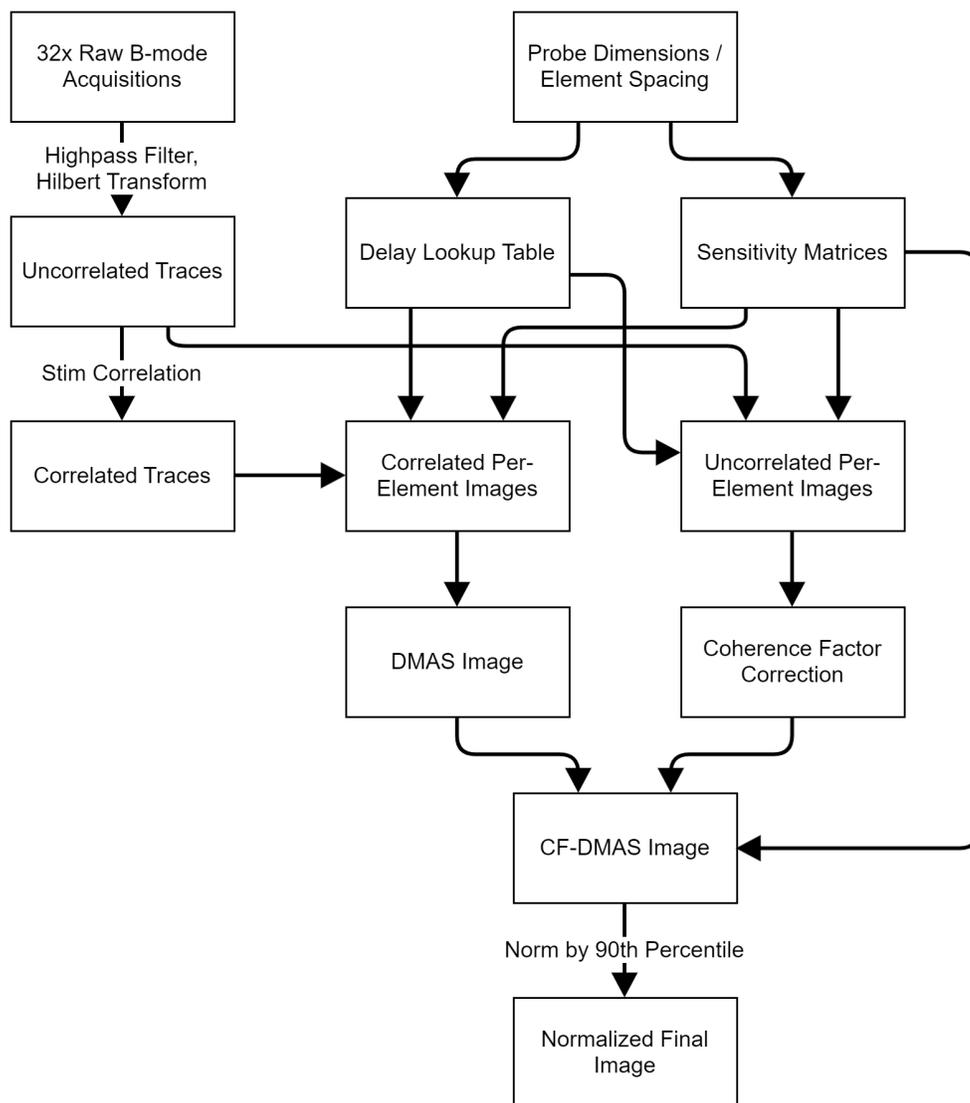
Figure A2.2: Flowchart describing the data flow for generating B-mode images from raw ultrasound returns using the CF-DMAS synthetic aperture algorithm.
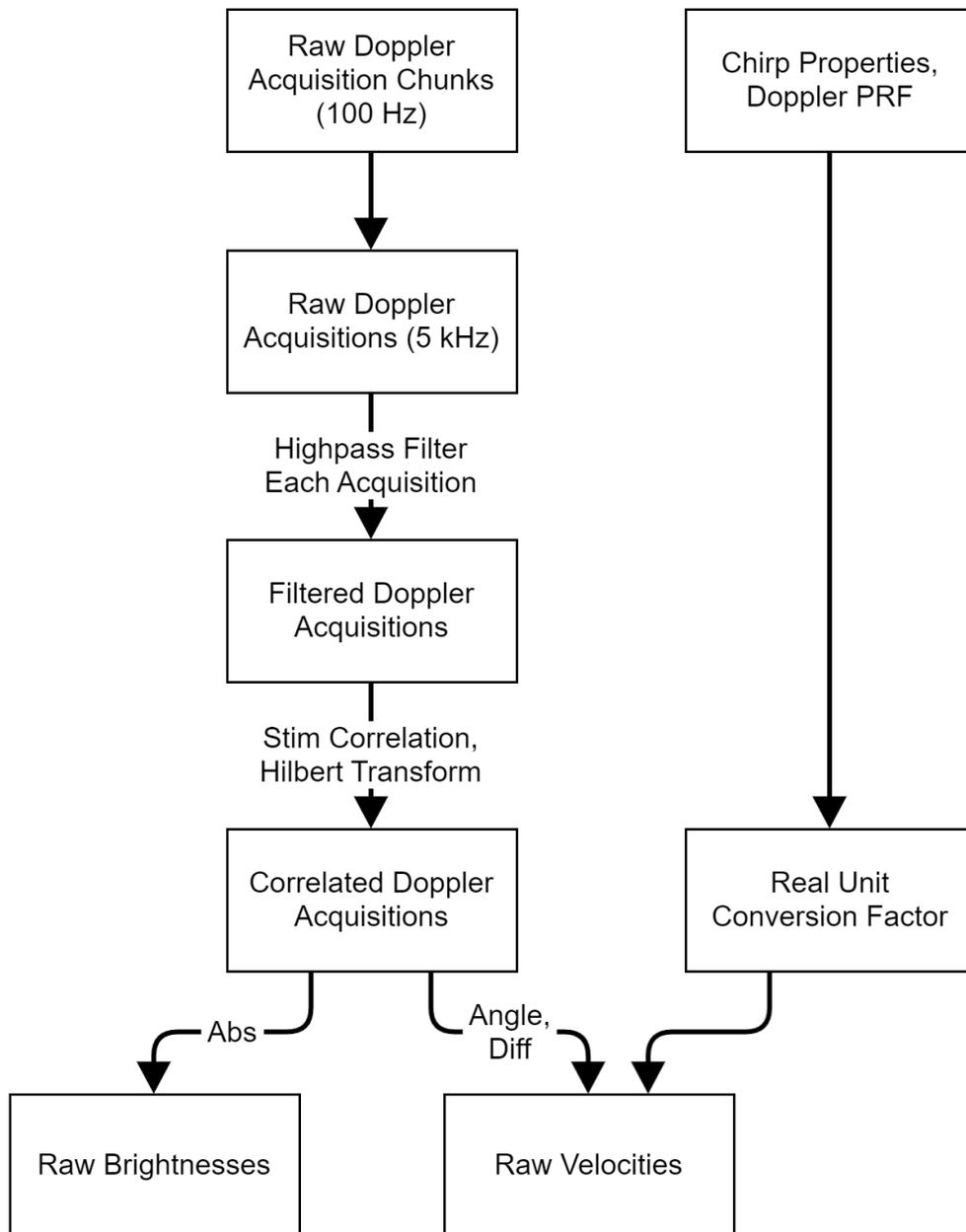
Figure A2.3: Flowchart describing the data flow for converting raw pulsed Doppler ultrasound returns into velocities.
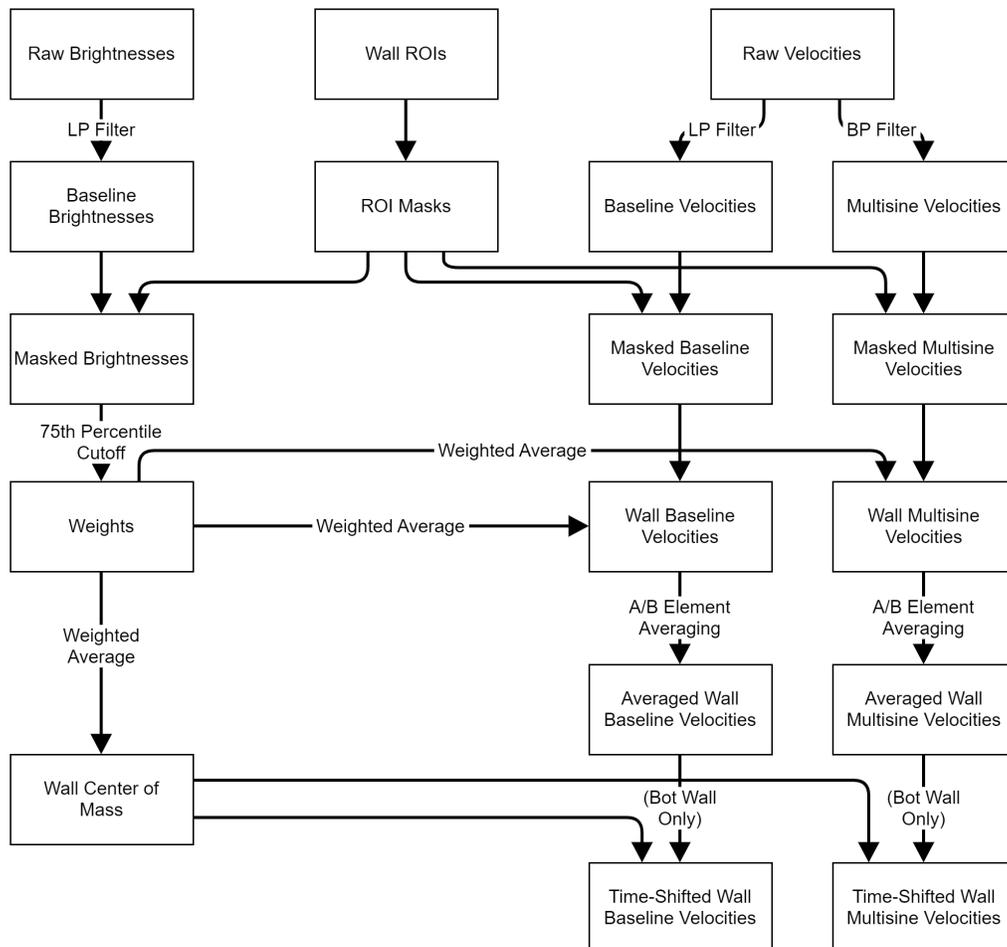
Figure A2.4: Flowchart describing data flow for converting full-column Doppler ultrasound velocities into arterial wall velocities, both "baseline" (due to the heart-beat) and "multisine" (due to the audio stimulus).
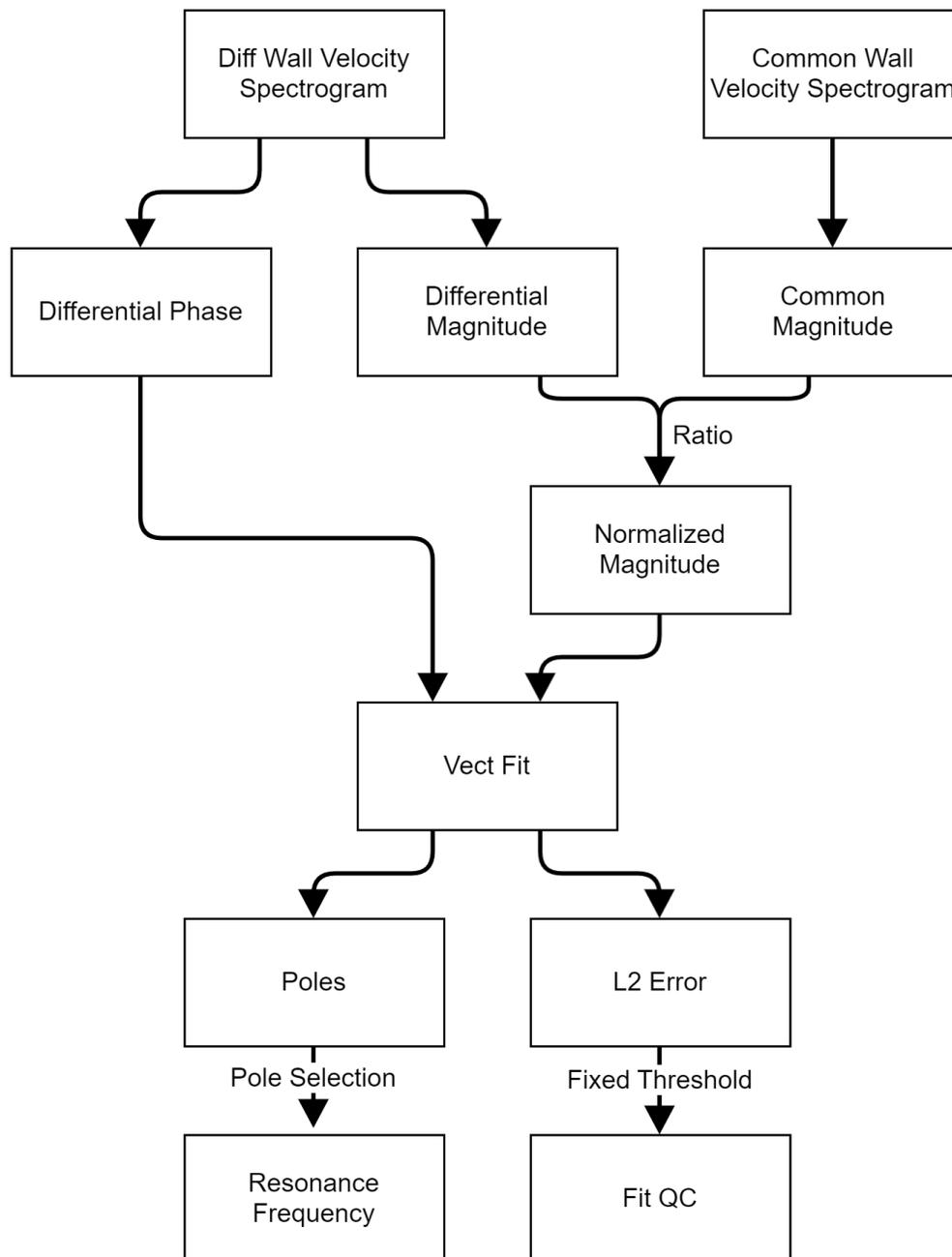
Figure A2.5: Flowchart describing the data flow for converting multisine wall velocities into a frequency response spectrum, and fitting this spectrum to obtain a resonant frequency.
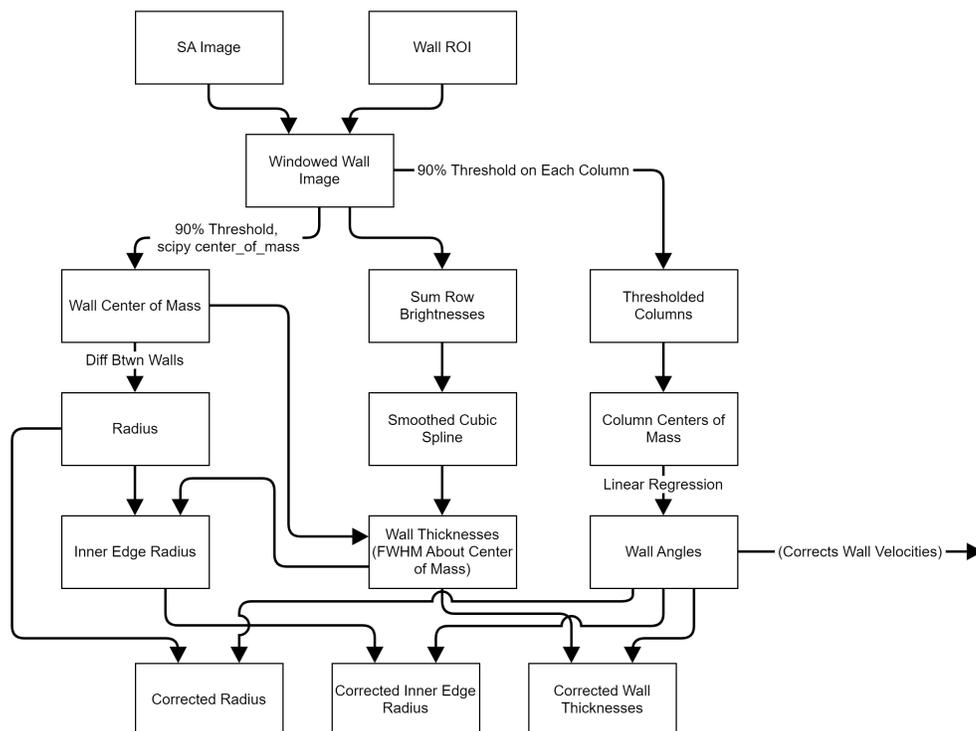
Figure A2.6: Flowchart describing the data flow for combining synthesized B-mode images and user-defined wall ROIs to measure arterial dimensions.
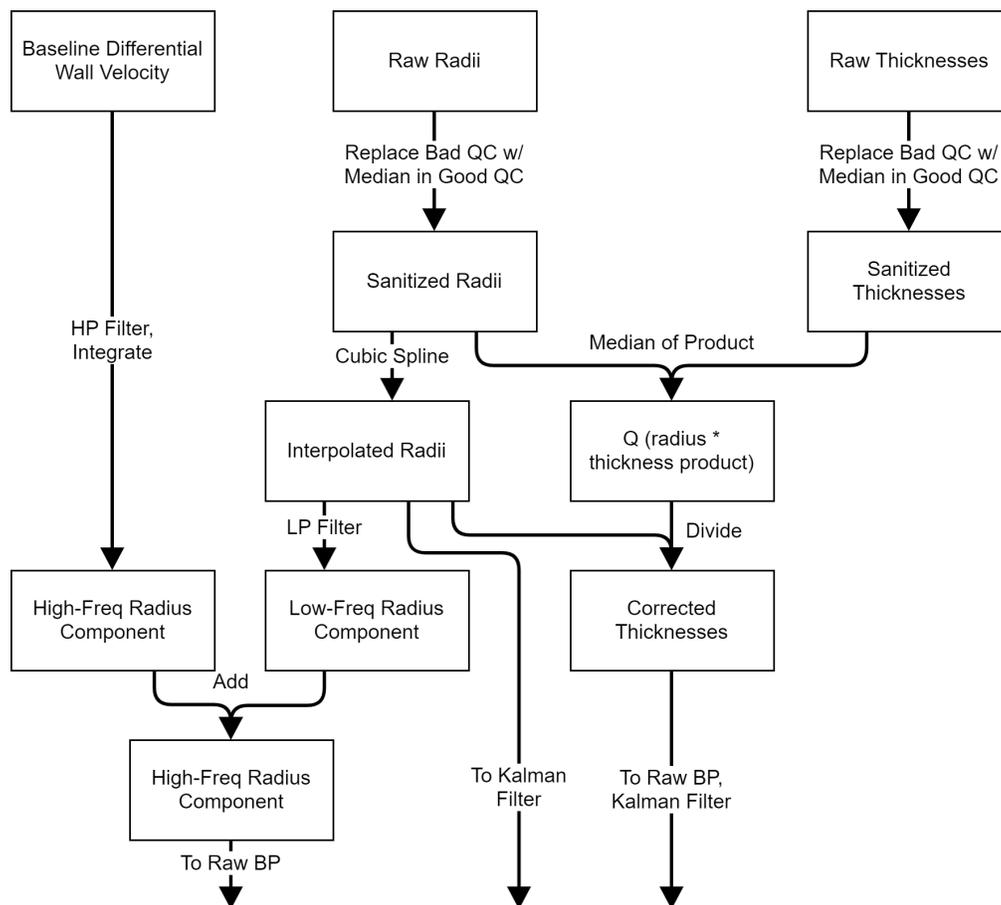
Figure A2.7: Flowchart describing the data flow for interpolating wall dimension measurements and combining with baseline wall velocity measurements to generate values at the appropriate rate for feeding into the Unscented Kalman Filter for BP generation.

*A p p e n d i x   C*

# FURTHER DATA FROM CLINICAL STUDY
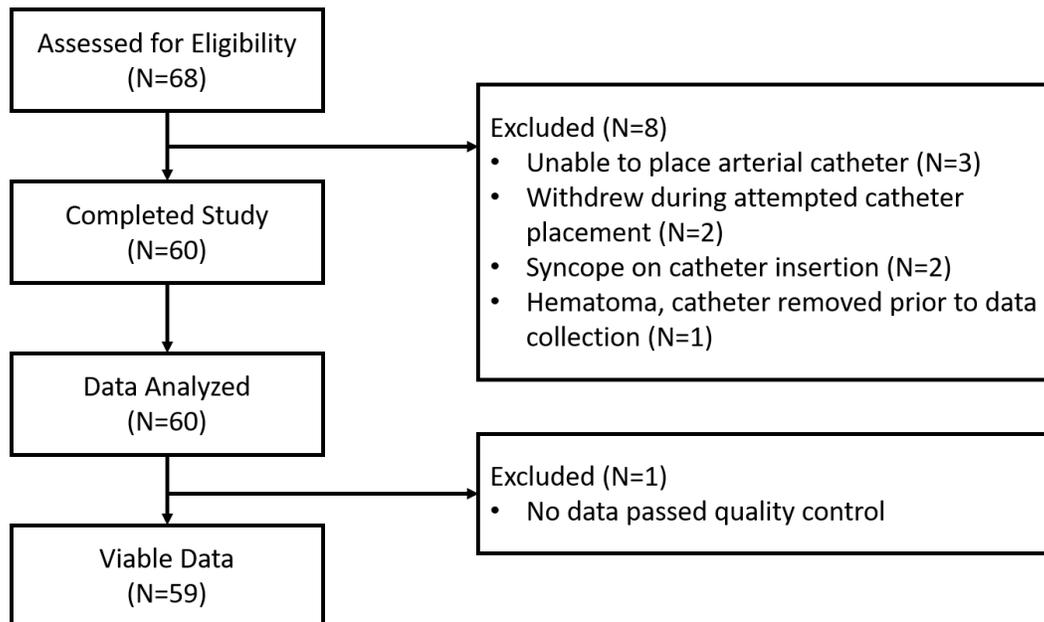


Figure A3.1: Subject inclusion and exclusion for the clinical study. It is worth noting that the rate of subject exclusion due to arterial catheter placement (6/68) significantly exceeded the rate of subject exclusion due to failure of quality control for CARDI-BP data (1/60).

| Characteristic | Combined (N=68) | Excluded (N=8) | QC Included (N=58) | QC Excluded (N=2) | Development (N=40) | Validation (N=20) |
|---|---|---|---|---|---|---|
| **Age** *Mean (range)* | | | | | | |
| | 33 (19-55) | 39 (27-53) | 32 (19-55) | 24 (22-26) | 32 (19-55) | 32 (19-48) |
| **Gender** *n* (%) | | | | | | |
| Female | 28 (41.2%) | 6 (75.0%) | 20 (34.5%) | 2 (100%) | 14 (35%) | 8 (40%) |
| Male | 39 (57.3%) | 2 (25.0% ) | 37 (63.8%) | 0 (0%) | 25 (62.5%) | 12 (60%) |
| Transgender | 1 (1.5%) | 0 (0%) | 1 (1.7%) | 0 (0%) | 1 (2.5%) | 0 (0%) |
| **Race** *n* (%) | | | | | | |
| Latinx White | 33 (48.5%) | 4 (50.0%) | 28 (48.3%) | 1 (50%) | 19 (47.5%) | 10 (50%) |
| Asian | 22 (32.4%) | 4 (50.0%) | 18 (31.0%) | 0 (0%) | 12 (30%) | 6 (30%) |
| Non-Latinx White | 9 (13.2%) | 0 (0%) | 9 (15.5%) | 0 (0%) | 6 (15%) | 3 (15%) |
| Non-Latinx Black | 3 (4.4%) | 0 (0%) | 3 (5.2%) | 0 (0%) | 2 (5%) | 1 (5%) |
| Mutliracial | 1 (1.5%) | 0 (0%) | 0 (0%) | 1 (50%) | 1 (2.5%) | 0 (0%) |
| **BMI** *Mean (range)* | | | | | | |
| | 28.8 (19.3-44.3) | 31.6 (22.8-44.3) | 28.5 (19.3-43.1) | 28.5 (24.1-32.8) | 28.4 (19.3-43.1) | 28.6 (20.4-43.0) |
| **Medical Conditions** *n*(%) | | | | | | |
| Hypertension | 7 (10.3%) | 2 (25%) | 4 (6.9%) | 1 (50.0%) | 3 (7.5%) | 2 (10%) |
| Diabetes Mellitus | 1 (1.5%) | 0 (0%) | 1 (1.7%) | 0 (0%) | 1 (2.5%) | 0 (0%) |
| Hyperlipidemia | 4 (5.9%) | 1 (12.5%) | 3 (5.2%) | 0 (0%) | 2 (5%) | 1 (5%) |
| Heart murmur | 5 (7.4%) | 0 (0) | 4 (6.9%) | 1 (50.0) | 4 (10%) | 1 (5%) |
| Cardiac Arrhythmia | 1 (1.5%) | 0 (0) | 1 (1.7%) | 0 (0%) | 1 (2.5%) | 0 (0%) |

Table A3.1: Demographic data for clinical study population.

| Statistic | Development (N=39) Mean (mmHg) | Std. Dev. (mmHg) | Validation (N=20) Mean (mmHg) | Std. Dev. (mmHg) | *p*-value |
|---|---|---|---|---|---|
| DBP | 2.5 | 10.9 | 6.9 | 10.8 | 0.144 |
| MAP | -1.1 | 10.5 | 1.8 | 12.6 | 0.375 |
| SBP | -23.4 | 11.9 | -27.4 | 16.3 | 0.339 |

Table A3.2: BP statistics for the "development" and "validation" populations. Mean differences between the arterial catheter and present method and standard deviation of differences are computed for DBP, MAP, and SBP for each population. *p*-values from a two-sample Welch's t-test for each population exceed 0.1 in all cases. Statistics are computed per subject, since these represent independent data points.

*A p p e n d i x   D*

# FURTHER DETAILS FROM RAP ESTIMATION STUDY

## D.1   Other Model Architectures and Aggregation

We tested six modern video-processing ML architectures for this study: X3D [130], SlowFast [131], MoViNet [132], TimeSformer [133], STAM [134], and ViViT [135]. The three transformer-based models (TimeSformer, STAM, and ViViT) were relatively computationally intensive did not yield promising results within 72 hours of training time, so they were discarded. The first three models, however, all yielded promising results when trained to match cardiologist RAP estimations. The results of these models when evaluated against the test set of cardiologist estimates and the test set of RHC measurements are shown in Table A4.1. The predictions of the three models were also compiled to create an aggregated prediction based on majority vote (if the three models gave 3 different answers, the consensus answer was considered to be 8). On the left of the table we can see that SlowFast outperformed either of the other models and the aggregrate predictions when gauged against cardiologist predictions. However, when gauged against RHC measurements the three models all performed similarly well, and the aggregate predictions slightly outperformed any individual model. We did not pursue the fine-tuning of other other model architectures and evaluation of fine-tuned aggregate performance in this study due to the required computational time investment. However, these results suggest that model aggregation could be a future avenue of research for further improvement.

|  | Accuracy vs Cardiologist | False Negative vs Cardiologist | Accuracy vs RHC | False Negative vs RHC |
|---|---|---|---|---|
| Cardiologist | - | - | 48.9% | 26.8% |
| SlowFast | 77.3% | 16.5% | 49.5% | 27.8% |
| X3D | 71.2% | 19.7% | 46.7% | 25.8% |
| MoViNet | 72.3% | 16.5% | 49.8% | 25.8% |
| Aggregate | 74.7% | 17.6% | 51.1% | 25.8% |

Table A4.1: Performance from the three promising model architectures as well as their aggregation, gauged against cariologist estimates and RHC measurements. SlowFast alone was the best at matching cardiologist estimates, but model aggregation could provide a future route to improving performance against RHC measurements.

## D.2   Progressive Analysis of Hyperparameter Optimization

During the process of hyperparameter optimization, we identified 6 hyperparameters which increased performance when changed from their default values. In decreasing order of importance, they were: # of sampled video frames, type of optimizer, whether or not to randomly augment data, period of video frame sampling, application of grayscale, and label smoothing. After generating final optimized test results, we reran models with progressive degrees of hyperparameter tuning on the test set to gauge whether or not hyperparameters had been overtuned. The model stages tested were:

- Stage 1:  16 video frames, SGD optimization, no data augmentation, fixed period 2 frame sampling, no grayscale, no label smoothing.

- Stage 2: 64 video frames, RAdam optimization, no data augmentation, fixed period 2 frame sampling, no grayscale, no label smoothing.

- Stage 3: 64 video frames, RAdam optimization, random data augmentation, evenly spaced frame sampling, no grayscale, no label smoothing.

- Final:  64 video frames, RAdam optimization, random data augmentation, evenly spaced frame sampling, grayscale, label smoothing.

Each intermediate model stage was tested against cardiologist estimates on the full untrimmed dataset, tested against RHC measurements, and then fine-tuned and retested against RHC measurements. The results are summarized below in Table A4.2. We can see that test performance against cardiologist predictions steadily increased as hyperparameters were optimized, in line with validation performance. This indicates that we did not overtune our hyperparameters for the initial training phase, which consisted of the bulk of model learning. Performance against RHC measurements showed a similar trend, although it was not strictly monotonic (potentially due to the smaller test set size). The absolute highest accuracy vs RHC came from the stage 1 model without fine-tuning, but was accompanied by a high false negative rate. The absolute best false negative rate (discounting Stage 0 FT, which just made predictions in the 8mmHg range for everything) came from the fine-tuned stage 2 model, but this had a relatively low overall accuracy. The best combination of high overall accuracy and low false negative rate did come from the final fine-tuned model, further indicating that we did not overtune our hyperparameters.

| | **Accuracy vs Cardiologist** | **False Negative vs Cardiologist** | **Accuracy vs RHC** | **False Negative vs RHC** |
|---|---|---|---|---|
| Cardiologist | - | - | 48.9% | 26.8% |
| Stage 1 | 59.1% | 36.7% | 42.3% | 33.0% |
| Stage 1 FT | - | - | 32.0% | 0.0% |
| Stage 2 | 72.9% | 18.6% | 51.5% | 21.6% |
| Stage 2 FT | - | - | 42.6% | 7.2% |
| Stage 3 | 74.2% | 19.7% | 47.6% | 23.7% |
| Stage 3 FT | - | - | 44.8% | 6.2% |
| Final | 77.3% | 16.5% | 49.5% | 27.8% |
| Final FT | - | - | 49.8% | 7.2% |

Table A4.2: Model performance at various stages of hyperparameter optimization, before and after fine-tuning to RHC data (FT models). Overall best test performance is obtained in the final stage, indicating that we did not overtune our hyperparameters.