SIGNAL PROCESSING AND MACHINE LEARNING TECHNIQUES FOR DOPPLER ULTRASOUND HAEMODYNAMIC MEASUREMENTS

Jack Latham

A thesis submitted to Cardiff University in candidature for the degree of Doctor of Philosophy

School of Engineering
Cardiff University
United Kingdom

August, 2020

Abstract

Haemodynamic monitoring is an invaluable tool for evaluating, diagnosing and treating the cardiovascular system, and is an integral component of intensive care units, obstetrics wards and other medical units. Doppler ultrasound provides a non-invasive, cost-effective and fast means of haemodynamic monitoring, which traditionally necessitates highly invasive methods such as Pulmonary artery catheter or transoesophageal echocardiography. However, Doppler ultrasound scan acquisition requires a highly experienced operator and can be very challenging. Machine learning solutions that quantify and guide the scanning process in an automatic and intelligent manner could overcome these limitations and lead to routine monitoring. Development of such methods is the primary goal of the presented work.

In response to this goal, this thesis proposes a suite of signal processing and machine learning techniques. Among these is a new and real-time method of maximum frequency envelope estimation. This method, which is based on image-processing techniques and is highly adaptive to varying signal quality, was developed to facilitate automatic and consistent extraction of features from Doppler ultrasound measurements. Through a thorough evaluation, this method was demonstrated to be accurate and more stable than alternative state-of-art methods.

Two novel real-time methods of beat segmentation, which operate using the maximum frequency envelope, were developed to enable systematic feature extraction from individual cardiac cycles. These methods do not require any additional hardware, such as an electrocardiogram machine, and are fully automatic, real-time and highly resilient to noise. These qualities are not available in existing methods. Extensive evaluation demonstrated the methods to be highly successful.

A host of machine learning solutions were analysed, designed and evaluated. This led to

a set of novel features being proposed for Doppler ultrasound analysis. In addition, a state-of-the-art image recognition classification method, hitherto undocumented for Doppler ultrasound analysis, was shown to be superior to more traditional modelling approaches.

These contributions facilitated the design of two innovative types of feedback. To reflect beneficial probe movements, which are otherwise difficult to distinguish, a regression model to quantitatively score ultrasound measurements was proposed. This feedback was shown to be highly correlated with an ideal response.

The second type of feedback explicitly predicted beneficial probe movements. This was achieved using classification models with up to five categories, giving a more challenging scenario than those addressed in prior disease classification work. Evaluation of these, for the first time, demonstrated that Doppler scan information can be used to automatically indicate probe position.

Overall, the presented work includes significant contributions for Doppler ultrasound analysis, it proposes valuable new machine learning techniques, and with continued work, could lead to solutions that unlock the full potential of Doppler ultrasound haemodynamic monitoring.

Acknowledgements

The completion of this thesis would not have been possible without the guidance and support provided to me by several people. I would like to briefly mention these people and express my sincere gratitude for their help.

Firstly, I would like to thank my supervisors, Dr Yulia Hicks and Prof. Rossi Setchi. Yulia patiently provided consistent and invaluable guidance and advice throughout this endeavour. Additionally, I would like to thank Prof. Timothy Rainer, who acted in a supervisory role and provided in-depth and valuable clinical knowledge. Crucially, Tim (the 'expert sonographer' referred to in this work) collected essential data that enabled this research. I would also like to express my sincere gratitude to the late Prof. Peter Wells, who during his career was a leader in the field of Doppler ultrasound, and whilst acting in a supervisory role shared some of his knowledge and enthusiasm with me.

Secondly, I would like to thank the Engineering and Physical Sciences Research Council (EPSRC), who funded this project and made this research possible.

Next, I would like to thank my friends and family, who expressed unwavering confidence in my ability to complete this project, and gave loving support for all the challenges that occurred outside of academia. Without this, the motivation, dedication and perseverance needed to finish would have been immeasurably harder to find. These people include my mum (Jo), my dad (Jim), my talented sister (Lille), my beautiful granny (Jennifer) and my best man (Alex).

Finally, I would like to thank my wife, Lydia. During the course of this thesis, we got wed and built a home with our unusual pets. Lydia's strength and love is irreplaceable.

To all of you, thank you.

Contents

List of Figures

List of Tables

Abbreviations

AF	Audio Features
AI	Artificial Intelligence
ANN	Artificial Neural Network
AR	Autoregressive
ARMA	Autoregressive Moving Average
AUC	Area Under the Curve
CHD	Coronary Heart Disease
CO	Cardiac Output
CS	Coefficient Statistics
CSA	Cross-sectional Area
CVD	Cardiovascular Disease
CW	Continuous Wave
DCNN	Deep Convolution Neural Networks
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EB	Envelope Boundary
ECG	Electrocardiogram
EM	Effectiveness Metric
FFT	Fast Fourier Transform
FM	Feature Matrix
FNR	False Negative Rate
FPGA	Field-programmable Gate Arrays
FPR	False Positive Rate

FSG	Falling Slope Gradient
GM	Geometric Method
GTCC	Gammatone Cepstral Coefficient
HOG	Histogram of Oriented Gradients
HR	Heart Rate
IMS	Image Statistics
IP	Image Portion
IPP	Ideal Probe Position
IPS	Integrated Power Spectrum
ISB	Intrinsic Spectral Broadening
LBP	Local Binary Pattern
LPC	Linear Predictive Coefficients
LPCC	Linear Predictive Cepstral Coefficient
LPF	Low-pass Filtered
MA	Moving Average
MF	Median Frequency
MFCC	Mel-frequency Cepstral Coefficients
MFE	Maximum Frequency Envelope
MGM	Modified Geometric Method
MODWT	Maximal Overlap Discrete Wavelet Transform
MSNSI	Modified Signal Noise Slope Intersection
MTCM	Modified Threshold Crossing Method
NE	Novel Envelope
OMM	Otsu Morphological Method
PAC	Pulmonary Artery Catheter
PCA	Principle Component Analysis
PI	Pulsatility Index
PLCC	Pearson Linear Correlation Coefficient
PRF	Pulse Repetition Frequency
PSD	Power Spectral Density
PV	Pulmonary Valve
PW	Pulsed Wave
RI	Pourcelots Resistance Index

RMSE	Root Mean-square-error
ROC	Receiver Operating Characteristic Curve
RSG	Rising Slope Gradient
SE	Sharron Entropy
SMOTE	Synthetic Minority Oversampling Technique
SNR	Signal to Noise Ratio
SROCC	Spearman Rank Order Correlation Coefficient
STD	Standard Deviation
STFT	Short-time Fourier Transform
SV	Stroke Volume
TEE	Transoesophageal Echocardiography
TF	Traditional Features
TNR	True Negative Rate
TPR	True Positive Rate
USCOM	Ultrasonic Cardiac Output Monitor
VTI	Velocity Time Integral
WPD	Wavelet Packet Decomposition
WT	Wavelet Transform

Chapter 1

Introduction

1.1 Motivation

Haemodynamic monitoring is used to assess the health of a person's cardiovascular system. This is achieved through measurements of blood circulation, perfusion and oxygenation of tissues and organ systems [1]; these enable at-risk patients to be identified, and their responses to therapy monitored [2].

Cardiovascular disease is the highest cause of global mortality [3]; considering this, the importance of haemodynamic monitoring is evident. Monitoring can also facilitate diagnoses and treatment of conditions such as sepsis [4], which alone account for 300,000 deaths annually in North America [5]. In the case of sepsis, goal-orientated treatment has been shown to be very beneficial, particularly when treatment is commenced early [4,6–8].

Haemodynamic monitoring can provide multiple desirable outcomes. These include improved cardiac health, reduced mortality and limitations with respect to associated economic burdens, which can be vast [9]. However, such monitoring typically requires invasive techniques, which have inherent risks and are intrinsically linked to increased mortality, morbidity and hospitalisation costs [10]. Furthermore, there are concerns regarding their efficacy and ease of use [2, 11], which make early detection and treatment difficult. These challenges, and the critical importance of early implementation, make non-invasive and efficient methods of haemodynamic monitoring highly desirable.

Doppler ultrasound can be used as a means of non-invasive haemodynamic monitoring. This technology measures blood velocity by analysing changes in the frequency of reflected

ultrasound waves [12]. In turn, this can be used to calculate an array of haemodynamic metrics, such as stroke volume (SV) or cardiac output (CO). This was first demonstrated using Doppler ultrasound over 60 years ago [13], and has over 30 years of safe clinical use. Modern implementation can be achieved with small, inexpensive portable devices, which have the potential to enable haemodynamic assessments in preclinical emergency medicine, such as air rescue, or even on scene [14]. Early treatment using such a device has shown substantial reductions in mortality rates for cases of septic shock [15].

However, the quality and reliability of measurements acquired using Doppler ultrasound are critically linked to operator skill [16], and the necessary experience required to be sufficiently competent is disputed [17, 18]. Reported limitations of the technology include difficulty maintaining and locating the correct probe position [19], variations in interpatient scanning difficulty including deterioration of measurement quality with increased age [16], and further complications such as reduced echogenicity in patients who have undergone cardiac surgery [20]. A combination of these factors have likely contributed to the large range in reported validation accuracies [21]. To realise the clinical potential of portable Doppler ultrasound, these limitations need to be mitigated and the gap between expert and routine assessment needs to be bridged. This is the motivation behind the work presented in this thesis.

This document presents research into methods and solutions that could enable Doppler ultrasound haemodynamic monitoring to be more accessible. The primary goal of this work is the design of automatic intelligent systems that give feedback reflective of a given measurement. To this end, a suite of signal processing methods were designed and evaluated in this work. These methods enabled the automatic extraction and processing of data, and were combined with machine learning procedures to develop feedback models. A selection of novel features were proposed for these models. Furthermore, state-of-the-art image recognition techniques for Doppler ultrasound analysis were proposed and shown to be superior to more traditional methods that employ hand-crafted features. The specific objectives corresponding to this work will now be described.

1.2 Objectives

The primary goal of the presented work is to design methods that can reflect scan quality and guide probe position during Doppler ultrasound haemodynamic scan acquisition. To

realise this goal, research is centred upon six objectives. These objectives are listed in Table 1.1, and will be explicitly detailed in this section. This work is focused on measurements of blood flow through the aortic valve, and subsequent references to Doppler measurements will refer to this.

The solution adopted for this primary goal is to design automatic intelligent feedback models. This is reflected by two thesis objectives. The first objective is to develop a scan quality feedback model. This feedback is a quantitative evaluation of a given measurement, which indicates whether probe movements are beneficial or not. Through a series of measurements, this feedback can indicate the IPP or, for example, be used to locate the best measurements from an examination time-history.

The second objective is to develop a position guidance model. Contrary to the scan quality model, this feedback explicitly predicts beneficial probe manipulations. In this regard, a scan history is not required to hone in on the IPP, and the feedback goes beyond any inference an operator could make using a singular measurement.

Predictive models are developed using features extracted from Doppler ultrasound measurements. As will be described in Section 3.5, the feedback being considered in this work is novel, and previous Doppler classification works are limited. Considering this, the most suitable feature types to construct such models are not evident. To reflect this, the third objective of the work is to develop and propose valuable features for these applications.

Several features are derived using the maximum frequency envelope (MFE). The MFE is a time history of the maximum frequency contained in a measurement, and is proportional to maximum blood velocity; this is described in detail in Section 3.1. The MFE, therefore, captures the cyclic nature of blood as it is pumped around the body; an example of an MFE is provided in Figure 1.1. This cyclic characteristic can be used to perform beat segmentation, which enables features to be extracted from individual cardiac cycles.

In the given application, beat segmentation and feature extraction must continue to function even when applied to data measured away from the IPP. These data can contain high levels of noise and atypical Doppler profiles. Furthermore, to preserve the technologies speed and cost-effectiveness, implemented methods must operate without additional hardware (e.g., an electrocardiogram (ECG)). This means MFE estimation and beat segmentation can be very challenging, and when performed using existing methods, result in incorrectly segmented beats and poor-quality features. These limitations are addressed

using two further objectives.

The fourth objective is to design a stable method of MFE estimation that is capable of preserving the pulsatile shape of blood flow across a wide range of conditions. The fifth objective is to develop an automatic method of beat segmentation, which requires no additional hardware to operate.

Particularly challenging portions of measurements, such as regions that contain erroneous signals (e.g., artefacts from sudden probe movements or external noise), can inevitably lead to incorrect beat segmentation. In offline settings (i.e., training and testing models), these observations can be removed. However, in real-time applications, this is not possible and is detrimental for model performance. Considering this, the sixth objective is to design a means of rejecting incorrectly extracted data. The primary objectives

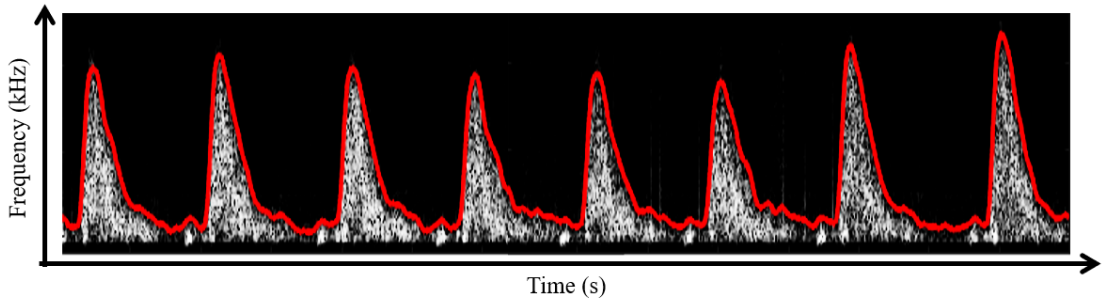


Figure 1.1: Example spectrogram of blood flow through the aortic valve with the MFE displayed in red.

Table 1.1: Summary of thesis objectives

Number	Objective
1	Develop a model to provide automatic scan quality feedback
2	Develop a model to provide automatic probe position feedback
3	Propose existing and novel features for intelligent scan acquisition guidance models
4	Develop a stable and real-time method of MFE estimation
5	Develop an automatic method of beat segmentation that operates using only the Doppler measurements
6	Develop a model to automatically reject poor quality beats

outlined above are summarised in Table 1.1, and discussed with respect to corresponding contributions and thesis structure in the following section.

1.3 Thesis Contributions

The work documented in this thesis centres around the six primary objectives listed in Table 1.1. The resulting contributions associated with each of these are detailed in Table 1.2. This table includes information regarding chapters documenting the associated work. As highlighted, the documented work is presented in a chronological order (e.g., the feature extraction and machine learning tasks proceed the design of the MFE estimation and beat segmentation methods, which were designed first). The thesis content and structure are summarised in the following section.

Table 1.2: Summary of thesis contributions

Obj.[1]	No.[2]	Contribution	Chapter
1	1	A model designed to give continuous feedback was proposed. This gives dynamic feedback to reflect beneficial or detrimental probe movements	7
2	2	Classification models proposed for four classification scenarios that categorise probe position with respect to the IPP	8
2	3	Proposed image recognition method of classifying Doppler signals, this approach was compared to more traditional methods.	
3	4	An analysis of existing and novel Doppler ultrasound features was performed. Spectral features previously not used in Doppler ultrasound analysis, and novel envelope features were suggested.	7, 8, 9
4	5	A real-time stable method of MFE estimation called OMM	5
5	6	A real-time automatic method of beat segmentation, that functions solely using the MFE called the slope-gradient method	6
5	7	A real-time automatic method of beat segmentation, that functions using a characteristic MFE and characteristic low frequency spectral content called the correlation method	6
6	8	A classification model for rejecting poor quality data was proposed	9

1 – Objective number associated with each contribution.
2 – Contribution number.

1.4 Thesis Structure

This thesis consists of 10 chapters, including the introduction. This section provides a brief summary of what is contained within each of these chapters, and how they relate to the objectives outlined in Table 1.1.

Chapter 1 – Introduction: Introduction to thesis topics, detailing thesis motivations,

objectives, and overall structure.

Chapter 2 – Background: This chapter builds on the themes summarised in the introduction to give both contextual and prerequisite information. This provides the reader with a fundamental understanding of the cardiovascular system, current haemodynamic monitoring methods, and Doppler ultrasound technology. The documented limitations of Doppler ultrasound technology are discussed, illustrating the need for the presented work.

Chapter 3 – Literature Review: A critical analysis of the work leading up to, and including, state-of-the-art methods associated with the outlined objectives. Gaps in the literature are identified, providing a logical basis for the proceeding chapters. Promising machine learning methods reported for Doppler ultrasound and other disciplines are analysed.

Chapter 4 – Survey Data: Different data extraction methods and machine learning techniques are investigated throughout this work. This chapter summaries the data used to design and evaluate these and describes the survey methodology. This provides a clear understanding of the challenges involved in the work, and provides context with respect to feedback models designed in proceeding chapters.

Chapter 5 – Envelope Estimation: A novel image-processing method of MFE estimation is presented. The proposed method is evaluated using phantom, simulated, and in-vivo data, and is shown to be highly stable in poor signal qualities. The methods performance is compared to that of three state-of-the-art methods.

Chapter 6 – Beat Segmentation: The chapter presents two novel beat segmentation methods, which require no additional hardware to function. These methods are used in subsequent chapters to extract features from Doppler measurements.

Chapter 7 – Continuous Feedback: This chapter describes the large array of different features extracted in the presented work. The methods used to then train and evaluate models are described. A model capable of producing continuous feedback reflective of probe position is designed and evaluated.

Chapter 8 – Localisation Feedback: The features and methods described in Chapter 7 are used to classify probe positions. Additionally, state-of-the-art image recognition methods are also investigated.

Chapter 9 – Beat Rejection: This chapter details a classification model and associated techniques for rejecting poor data in real-time applications.

Chapter 10 – Conclusions and Future Work: The presented work and its primary outcomes and significance are summarised. Potential future research, based on the progress and areas highlighted in the presented work, are discussed.

1.5 Publication

The design and evaluation of the novel MFE estimation method, which is presented in Chapter 5, formed the basis of the following published journal article:

J. Latham, Y. A. Hicks, X. Yang, R. Setchi, T. Rainer. Stable Automatic Envelope Estimation for Noisy Doppler Ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

Chapter 2

Background

Haemodynamic measurements give crucial insight into cardiovascular function. Within this thesis, methods of processing and extracting information from Doppler ultrasound haemodynamic measurements are proposed and evaluated.

This chapter provides a fundamental understanding of why haemodynamic assessment is essential, and the role that Doppler ultrasound plays in this area. The basic underlying physics and implementation of this technology is given. This details valuable clinical metrics that can be used to assess the cardiovascular system. This system, the system upon which every cell in the body relies, will now be described.

2.1 The Cardiovascular System

With every heartbeat, blood is forced around the body. Vessels guide this blood through a myriad of arteries, veins and capillaries, forming two circuits that start and end at the heart. One circuit oxygenates the blood, and the second traverses the body, supplying cells with vital oxygen and nutrients.

At the core of this system is the heart, a muscular organ that acts as a pump. The heart contains four chambers, as illustrated in Figure 2.1. Each side of the heart has a pair of chambers: an atrium and a ventricle. Ventricles actively contract and eject blood into arteries, whilst atria passively fill with blood returning from veins. The left side receives and pumps deoxygenated blood, whilst the right side receives and pumps oxygenated blood. Cyclic stimulation of the heart muscles, in response to the body's needs, enables

the heart to pump blood around the body [22].

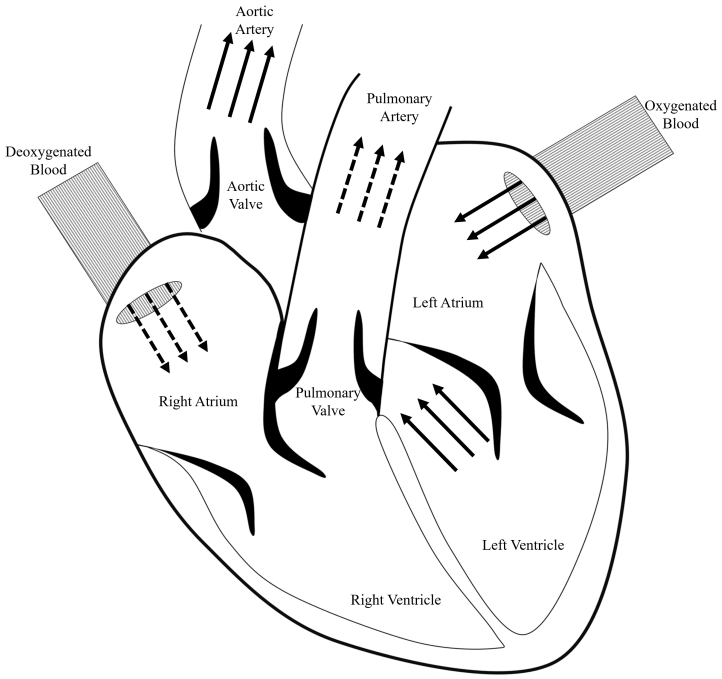


Figure 2.1: Diagram of the heart

2.1.1 The Cardiac Cycle

A series of mechanical and electrical events transpire with each heartbeat; this sequence is called the cardiac cycle [22]. With reference to the heart, events include valve openings and closures, isovolumic contraction and relaxation, and periods of blood ejection and filling. The duration of a given cardiac cycle (T_{cc},) is determined by the person's heart rate (HR).

$$T_{cc}(s) = \frac{60}{HR}; \tag{2.1}$$

These events can be used to divide the cardiac cycle into different phases. This can be performed to varying extents, for example separating the cycle into four [23] or even seven stages [24].

In this thesis, the cardiac cycle is considered to have two distinct phases. During one of these periods, the heart contracts and blood is pumped into the arteries; this period is termed systole. Following this, the heart relaxes and expands. This allows returning blood to flow back into the heart; this period is called diastole.

During systole, the heart ventricles contract and blood is ejected through the pulmonary and aortic valves, on the right and left sides of the heart, respectively. Blood entering the pulmonary artery is deoxygenated and is channelled to the lungs, where carbon dioxide is expelled, and oxygen is collected. Blood entering the aortic artery is oxygenated and is subsequently supplied to a network of vessels that transport it throughout the body. Diastole follows and is a period of ventricular relaxation, during which the ventricles are refilled with blood. The left atrium, and subsequently left ventricle, are filled with oxygenated blood returning from the lungs. On the right side, the deoxygenated blood returns via the superior and inferior vena cava. The process can now repeat. This circulation forms a closed system, which can be analysed and measured at different points and using different technologies.

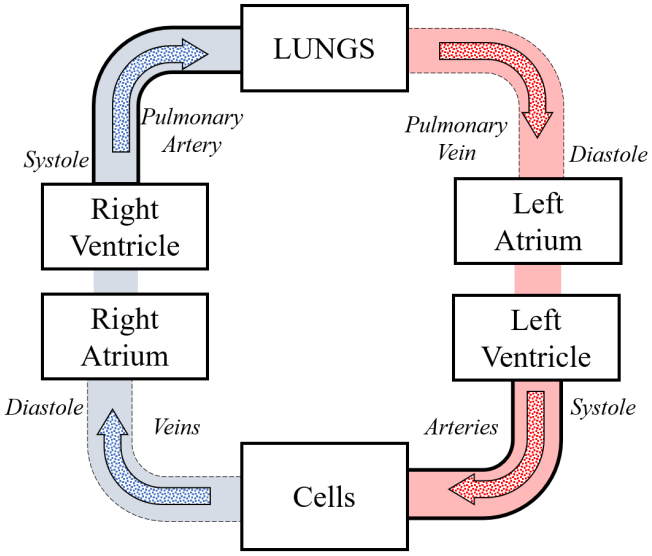


Figure 2.2: Cardiovascular flow diagram

2.2 Cardiovascular Monitoring

Cardiovascular disease (CVD) includes several heart and vessel disorders, for example, coronary and rheumatic heart disease. CVD results in 31% of deaths worldwide, which corresponds to 17.9 million people and makes it the highest global cause of mortality [3]. Clearly it is imperative that people at risk are identified, and appropriate treatments are provided. Cardiovascular management is especially vital for patients undergoing surgery. Only through precise haemodynamic and cardiac monitoring can therapies be guided to treat CVDs [25]. Changes that can occur in response to CVDs include hypertension, hypotension, valve diseases and heart failure [22]. Different methods have been developed to diagnose the cause of such changes, and to monitor aspects of cardiovascular status.

The ECG is a fundamental tool for cardiac monitoring. This familiar device measures electrical activity of the heart. Signals are acquired using electrodes placed at strategic points on the body, including the limbs and the chest. The resulting rhythmic waveform can be used to diagnose a variety of abnormalities, such as atrial or ventricular fibrillation, or tachycardia [24]. Another standard tool for assessing cardiac health is the phonocardiograph. Phonocardiographs monitor heart sounds, which include the iconic 'lub' and 'dub' components. Respectively, these sounds occur as a result of the mitral and tricuspid valves closing, and the aortic and pulmonary valves opening. Additional sounds can be induced – for example due to turbulence – which indicate problems such as regurgitation or valve issues [22]. These occurrences can be assessed quickly using auscultation, or via similar methods using Doppler ultrasound [26].

In addition to these metrics, quantitative measures of blood flow can be used to assess heart performance. In practice, appropriate monitoring will typically necessitate multiple interdependent parameters, such as blood pressure, stroke volume variation and CO. These measures can also be used to enable goal-directed fluid therapy. This approach replaces more traditional and less successful approaches based on a patient's weight and their duration without fluids [27]. Monitoring such metrics is also invaluable for the detection and treatment of sepsis, septic shock and other types of shock or hypertension.

Early goal-directed therapy, which is guided by cardiovascular monitoring, is an effective method of improving outcomes for patients with severe sepsis and septic shock [4]. Sepsis is defined as a systemic inflammatory response in the presence of an infection, in which a number of symptoms can present such as an elevated heart rate or an irregular

body temperature [28]. This dangerous bodily response is a major cause of disability and death across the world, with 300,000 deaths being attributed to the condition annually in North America alone [5]. The danger of sepsis is sadly being illustrated by the current pandemic of COVID-19, in which the inflammatory response is the most common complication. Following this are respiratory issues, and then heart failure and septic shock [29]. These stark examples illustrate the power of clinical information garnered through cardiovascular monitoring, and the need for safe, accurate and practical means of measuring blood flow metrics.

As discussed, a valuable and commonly measured metric is CO. This is the amount of blood pumped by the heart during 1 minute. It is regarded as a principal variable for cardiovascular assessment and is often used to guide treatment protocols [30]. This metric will now be discussed in more detail.

2.3 Cardiac Output

CO is routinely monitored as part of intensive care practice. It is a principal determinant of oxygen delivery and blood pressure [31], and as such can guide treatment and identify patients at risk. The first method of measuring CO, the 'Fick principle', was described by Adolph Fick in 1870 and was first clinically used in 1940. However, measurements of CO only became routine as recently as 1970, following work which detailed catheterisation of the right side of the heart using a balloon-tipped catheter [32]. This led to specially designed pulmonary artery catheters (PAC), and a method of CO estimation using thermodilution curve analysis. There have been hundreds of studies addressing the accuracy of this technique. Typically, these demonstrate PAC thermodilution to be a reliable, and clinically informative method of measuring CO; there are however known inaccuracies with the method [33]

Drawbacks of PACs include their inherently invasive nature, which introduces risk (e.g., arrhythmia, infection or possible pulmonary disruption [34]). This consideration and other factors such as time restraints, cost, and required level of expertise can limit their use; this is particularly true within the emergency department [17]. These issues make alternative and less-invasive methods highly desirable. A number of promising methods have been proposed. However, inaccuracies associated with PACs have led to debate regarding its use as a gold standard for validating novel methods. This can make validating an alternative

method's efficacy challenging [31].

Non-invasive methods of CO measurement include pulse contour analysis, the partial rebreathing technique, transthoracic electric bioimpedance and Doppler ultrasound. These will now be discussed.

2.3.1 Pulse Contour Analysis

Pulse contour analysis was suggested over a century ago [35], however, until recently it could not be used to produce accurate measurements due to technical limitations [36]. The method is based on the principle that through analysing the arterial pressure waveform (obtained from an arterial line), SV can be continuously monitored [37]. The method has been implemented in a commercial device [38], which has been included in some comparison studies. An assessment of its role in the care of critically ill patients still needs to be performed, as well as further validation and improvements to the system [31].

2.3.2 Partial Rebreathing Technique

The partial rebreathing technique is based on the Fick principle discussed above. The principle is based on the conservation of mass, it states that the total uptake or release of a substance by an organ remains relatively constant during the non-breathing and breathing periods, and as such cancel out [36]. Traditional implementation necessitated invasive measurements, however,a recent innovation is capable of evaluating CO_2 elimination, from which pulmonary blood flow can be calculated. This method is less invasive and can be used to estimate CO [31]. However, the technique cannot be applied in certain conditions, such as for infants [39], and is limited due to the need for intubated and ideally ventilated patients [40].

2.3.3 Transthoracic Electrical Bioimpedance

Transthoracic electrical bioimpedance is an interesting method, first developed by NASA and Kubicek in the 1960's [41]. As its name implies, the approach exploits bioimpedance. This is the resistance to an electrical current being transmitted through the body. The path of lowest resistivity is that of blood and plasma (i.e., these fluids have the highest conductivity). An electrical current will flow through the path of least resistance, thus when a current is applied to the body it will be primarily conducted by these fluids. An

increase in blood volume is associated with a proportional increase in conductance; dynamic changes in blood and plasma, therefore, correspond to varying bioimpedance. This is most significant where blood flow is pulsatile, for instance, blood leaving the left side of the heart. Variations in bioimpedance are used to calculate SV, which when combined with heart rate enables CO to be found [42]. The method has been validated with respect to thermodilution techniques [43, 44] and has been shown to perform accurately. However, the technology has a number of limitations, for example, it requires haemodynamic stability and cannot be applied to certain patients [36].

This technique is non-invasive but does require four pairs of probes comprised of transmitting and receiving sensors [45], and an ECG. This limits the viability of the method in certain conditions and prevents quick implementation.

2.3.4 Transoesophageal Echocardiography

Transoesophageal Echocardiography (TEE) is a technique that has been used intraoperatively since the 1980s and has progressively become more popular [46]. Particular complications associated with PACs, such as catheter size, have resulted in the method being common in paediatrics. TEE can measure blood velocities using ultrasound, through exploiting the Doppler effect (i.e., Doppler ultrasound). This method uses a small specialised probe that is inserted into the oesophagus. This probe is positioned so that data is collected from a particular position with known area. This is commonly the left ventricular outflow tract [31]. Measured velocities and area are used to determine CO. Similarly, oesophageal Doppler can be used to measure blood flows within the descending aorta.

TEE can also determine CO in a non-Doppler manner. Using an imaging modality, the volume of the left ventricular chamber during systole and diastole can be measured. The difference between these volumes allows SV to be estimated. However, this is typically impractical for perioperative cases, and imaging difficulties can result in inaccurate measurements [31]. TEE methods, as a whole, necessitate a high level of knowledge and skill to ensure accurate results. Comparatively with transcutaneous approaches, TEE is more invasive, requires more time, and cannot be used to provide continuous monitoring.

2.3.5 Transcutaneous Doppler Ultrasound

Transcutaneous Doppler ultrasound is completely non-invasive, cost-effective and easy to administer. This technique exploits the same physics as TEE methods; SV is calculated using blood velocity measurements and lumen area. The term "transcutaneous" describes a method applied across the depth of the skin, meaning that a transducer need only be placed on the skin and not inserted internally. The ease of applying this technique makes it highly desirable in certain situations, such as in the emergency department where time is highly precious. This technique has been implemented using stand-alone commercial Doppler devices, which are geared towards providing cheap and easy methods of monitoring CO. As will be discussed in Section 2.7, this technology has been widely validated. The operation of such Doppler ultrasound devices are briefly discussed in Section 2.5.

With respect to other non-invasive methods of CO measurement, transcutaneous Doppler ultrasound can be administered very quickly and accurately. Furthermore, the technology can be realised in cheap and compact devices. This lends the technology to challenging scenarios, such as within emergency transport or rural settings. The speed of application also enables repeat measurements, which otherwise would be impractical.

2.4 Doppler Ultrasound

The preceding sections have highlighted various ways in which Doppler ultrasound has been implemented in clinical settings, and specifically, the life-saving information provided by associated blood velocity measurements. In this section, the physical principles of Doppler ultrasound, and how Doppler measurements can be used to extract haemodynamic metrics are described. In the proceeding chapters, Doppler signals are processed using a variety of techniques. The outputs of these are governed by the fundamental principles outlined here.

Ultrasound has been used as an inspective tool since 1912, following the collision of the Titanic with an iceberg [47], after which solutions for safer navigation included ultrasonic technologies. Diagnostic ultrasound research began around 1950, where its imaging potential was demonstrated by Dussik, who succeeded in generating ultrasound images of the human brain [48]. Since Dussik, who has been referred to as the 'Father of Ultrasonic Diagnosis' [49], huge advances in physics and technology have allowed diagnostic

capabilities and applications to grow tremendously [50].

2.4.1 What is Ultrasound?

Ultrasound is the name given to sound waves with frequencies that exceed 20 kHz. As the name suggests, these frequencies are beyond the normal human auditory range. The most common form of an acoustic wave is a longitudinal compressional wave; ultrasound (and sound in general) is of this type. Propagation of longitudinal waves is enabled through particles oscillating parallel to the direction of travel. These particles move locally but do not travel with the wave, allowing energy to be transferred across mediums.

Ultrasound has become an essential tool in medical diagnosis, familiar uses include imaging of foetus' or other organs. In examinations such as these, transcutaneous ultrasound is used to interrogate targets. Resulting reflections are then processed to give interpretable data. The ultrasound signal is generated by means of an electromechanical transducer (usually a piezoelectric transducer), with frequencies typically ranging from 1 to 20 MHz. The relationship between wavelength (λ), frequency (f) and speed of sound is given below.

$$f = \frac{c}{\lambda} \tag{2.2}$$

Frequency is inversely proportional to wavelength. Considering this, higher frequencies can be used to produce data with increased resolution. However, higher frequencies attenuate faster and so cannot penetrate as deeply. For a given application, this trade-off has to be considered when selecting an operating frequency. In practice, this means lower frequencies tend to be used to inspect deeper targets and higher frequencies for more superficial targets [12].

2.4.2 The Doppler Effect

When a vehicle passes you, you may notice that the engine sound, or the music being played within, varies in pitch. This is because the frequency of sound emitted by a source, is determined by the relative velocity of that sound source. This phenomenon is called the Doppler effect. An analogy of this are the waves surrounding a moving ship, those at the bow of the ship bunch together (i.e., have a shorter wavelength, and higher frequency), whilst those at the rear of the ship spread apart (i.e., have a longer wavelength, and

lower frequency). The effect also occurs for other wave types, such as light, and has been exploited for many uses including within speed cameras, or on a grander scale, to show that the universe is expanding. The Doppler effect was first described by Christian Doppler in 1842, who used it to describe the colour of binary stars [51].

The effect is also exploited in ultrasound technologies. When an ultrasound beam is transmitted into the body, a portion of this signal will be scattered and reflected back. If these scatterers have motion, the returned signal will exhibit a Doppler shift. An increase in frequency occurs if this motion is towards the transducer, and a decrease occurs if it is away from the transducer [52]. This Doppler shift can be used to estimate the velocity of the target. The relationship between target velocity (v), and ultrasound frequency is given in equation 2.4.

$$f_d = f_r - f_t \tag{2.3}$$

$$f_d = f_t \frac{v2cos(\theta)}{c} \tag{2.4}$$

Where f_t, f_r and f_d are the transmitted, reflected, and Doppler shift frequencies respectively, c is the speed of sound within the medium, and θ is the insonation angle. This insonation angle and the transmitted and reflected frequencies are illustrated in Figure 2.3.

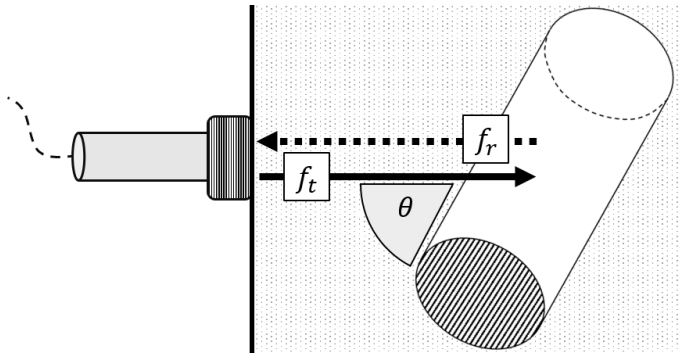


Figure 2.3: Insonation angle associated with Doppler ultrasound measurements

Use of Doppler ultrasound to investigate blood flow was first reported by Satomura in 1959 [53]. Light later demonstrated quantitative blood flow measurements from the ascending aorta using transcutaneous ultrasound [54]. These early examples used continuous

wave (CW) ultrasound. This form of ultrasound operates by continuously transmitting and receiving ultrasound, and therefore exhibits no range resolution. This is in contrast to pulsed wave ultrasound (PW), which transmits and receives ultrasound sequentially, allowing for localisation of signals with respect to distances along the beam [55]. The basic operation of these ultrasound systems is given in Section 2.5.

Medical applications were further illustrated in a 1983 study, within which CO measurements obtained using CW Doppler ultrasound and thermodilution were compared [55]. This study found good correlation between the two methods. With respect to Doppler ultrasound, CO is calculated from the product of SV and heart rate (HR). SV is the product of the velocity-time integral (VTI) and the lumen cross-sectional area (CSA); these relationships are shown in equations 2.5 and 2.6.

VTI is the integral of peak velocity during systole and is equivalent to the distance travelled by the blood column during this time. The measured blood velocity is assumed to be the instantaneous velocity averaged across the whole CSA (i.e., it exhibits plug flow [56]).

$$SV = VTI \times CSA \tag{2.5}$$

$$\mathrm{CO} = SV \times HR \tag{2.6}$$

2.4.3 Intrinsic Spectral Broadening

With respect to the Doppler equation (2.4), estimating the maximum velocity associated with a Doppler signal appears to be straight forward. However, in practice, clinical Doppler measurements can include a host of erroneous signal contributions (for example from tissue movement or nearby blood flows), as well as many other contributing factors such as acquisition errors, systematic quantification or signal processing noise [57].

In addition to these contributions, inherent parameters of the Doppler ultrasound system also gives rise to a phenomenon known as intrinsic spectral broadening (ISB) and manifests as a blurring of the Doppler spectrum. When measuring a single moving target using Doppler ultrasound, the measured frequency content contains a range of frequency shifts and not one singular value. When interrogating blood, many moving targets contribute to the measured Doppler signal, resulting in a smearing of the frequency

spectrum [58]. The presence of ISB is attributed to two spectral contributions: local geometric broadening and transit-time broadening [59].

Geometric broadening occurs due to the finite width of the Doppler beam (i.e., the aperture) [60]. This means a traversing target has a range of incident and reflected ultrasound angles, and as the Doppler shift is proportional to $cos(\theta)$, the measured frequency exhibits a spectrum of values.

Transit-time broadening occurs due to the finite time in which a target traverses the ultrasound beam. The inhomogeneity of an ultrasound beam means the intensity of reflected signals from the target varies as it passes through, and so these reflected signals are amplitude-modulated [58].

The combination of these two effects results in the measured Doppler spectrum being smeared, and sharp variations being smoothed. The practical effect of this is that when inspecting Doppler frequency content, frequencies which correspond to maximum velocity points can be hard to distinguish, and can mean values are overestimated [61]. In clinical situations, diagnoses based upon maximum blood velocities must appreciate the effects of ISB [58].

2.5 Doppler Ultrasound Machines

Doppler ultrasound has been widely researched, and the phenomenon has been exploited for numerous applications. This has resulted in different Doppler ultrasound devices being designed, which enable different aspects of the cardiovascular system to be measured. However, despite this large range of technologies being developed, only a handful of instrument types are used regularly in clinical practice. These can generally be classified as one, or a combination, of the systems described in this section.

2.5.1 Velocity Detecting Systems

Velocity detecting systems produce an output signal that is proportional to the velocity of the targets being interrogated. Such systems tend to be simple stand-alone devices, with a single sample volume. The transducer is handheld, and not connected with a location sensing system (i.e., a system which tracks, and provides feedback on transducer location).

These systems can range from simple arrangements that output audio reflective of

absolute velocity, to more sophisticated systems that produce directional signals, which enable further analysis and multiple envelope types to be extracted [12]. Despite the simplicity of these systems, they can be valuable in detecting the absence or presence of flows, and monitoring changes in these flow characteristics.

These systems lack imaging capabilities, and as such the incident angle of the ultrasound beam and targets cannot be measured. The consequence of this angle is determined by a simple cosine relationship (see equation 2.4), Table 2.1 illustrates the magnitude of this effect. At certain anatomical positions, the angle of incidence can be assumed to be negligible, for example, at the temporal bone when performing cerebral artery measurements. Similarly, measurements of the aortic arch when collected from the suprasternal notch can be considered to have a very small insonation angle, allowing quantitative velocity measurements to be made [54]. As will be described in Section 2.6, this approach is employed when using the primary ultrasound device implemented in this work. Acquired measurements can then be used to calculate information regarding the cardiovascular system [62].

Table 2.1: Relationship between incident angle and calculated velocity

Angle (°)	$\cos(\theta)$	Velocity Reduction (%)
0	1	0.0
10	0.99	1.5
20	0.94	6.0
45	0.71	29.3
60	0.50	50.0
90	0.00	100.0

Velocity detecting systems can be implemented using either CW of PW ultrasound. CW devices transmit and receive ultrasound continuously. To accomplish this, they require two separate crystals, one for transmission and one for receiving. These tend to be contained in the same probe, for example, via two half-moon shaped transducers. As the transmitted signal is continuous, no range resolution can be achieved, only a maximum observable depth due to attenuation of the ultrasound. Attenuation of sound increased with frequency, Table 2.2 illustrate this with respect to ultrasound within the body. This relationship, combined with safe intensity levels, dictates the frequency used to inspect

specific targets. Superficial vessels will tend to be interrogated using frequencies around 8 MHz, whereas for deep targets, frequencies as low as 2 MHz may be used [12]. Lower frequencies are capable of insonifying the aortic valve from the suprasternal notch. Figure 2.4 displays a block diagram illustrating the design of a simple non-directional CW velocity detecting system.

Table 2.2: Approximate ultrasound attenuation in human body with respect to frequency

Transmitted Frequency (MHz)	Attenuation Coefficient (dB cm^{-1})
2	4
4	8
6	12
8	16
10	20

In reference to Figure 2.4, the master oscillator outputs an electrical signal that is amplified and used to drive a transmitting crystal. The transmitting crystal converts the electrical signal into an acoustic one, usually via the piezoelectric effect. Reflected waves are recorded by the receiving crystal and converted back into an electric signal. This is then amplified and passed to a mixer and a low-pass filter [12].

This process allows coherent demodulation to take place (i.e., removal of the carrier frequency to give an output of the Doppler frequency sidebands). The received signal and a reference signal from the master oscillator are passed to the mixer. The low pass filter removes the high-frequency component in this signal, leaving the Doppler frequency. By fortunate coincidence, these Doppler frequencies typically occur in the audible range, and thus, can be passed directly to an audio output to produce feedback for an operator.

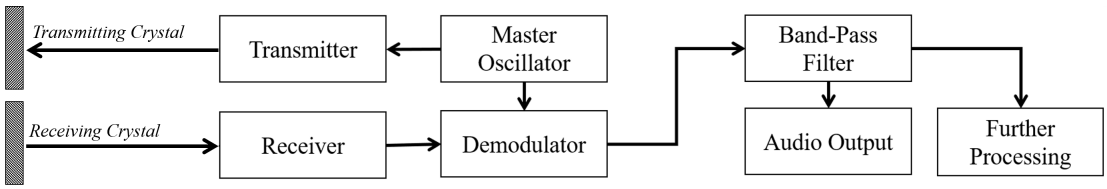


Figure 2.4: Block diagram of a simple non-directional continuous wave Doppler Ultrasound system

Furthermore, this demodulated signal can undergo further processing and analysis.

This straight-forward demodulation technique does not provide directional information (e.g., the maximum positive and negative velocity). In order to acquire directional Doppler information, a more complex demodulation process can be used. Several techniques have been reported for this, but the most common method is quadrature-phase detection or I/Q demodulation [63].

With respect to CW systems, PW systems are advantageous in that they provide both range and velocity information of targets [55]. This is achieved by transmitting short bursts of ultrasound with a specific pulse repetition frequency (PRF), and following a short delay, analysing a portion of the received signal. The period of this delay is user-defined and determines the depth from which signals are gathered. Wells demonstrated this using PW transcutaneous ultrasound measurements of the mitral valve as early as 1969 [64].

However, the maximum measurable Doppler frequency using a PW system is limited to half the PRF, beyond this, aliasing of the Doppler signal occurs. As target depth is increased, so is PRF. This is to account for the greater distance travelled by the ultrasound pulse; such systems are therefore limited to measuring low velocities from deeper targets [65]. Additionally, PW allows adjustable gain with target depth. These features, however, introduce complexity with respect to aiming the device, for example, there is additional user dependency to set range [55].

With respect to CW systems, PW systems require further circuitry to allow transmitted and received signals to be appropriately gated and passed to the demodulator. Although, the transducer arrangement is simplified if the same crystal is used for both transmission and reception [12].

2.5.2 Duplex Systems

Duplex ultrasound systems combine B-mode ultrasound (i.e., a 2D ultrasound image) with Doppler ultrasound. Using the 2D image, the sample volume location used for collecting Doppler recordings can be specified [12]. This allows one to obtain Doppler signals from specific anatomical locations [66]. Duplex systems can also be used to measure anatomical dimensions, such as lumen CSA. Furthermore, these systems can visualise vessel walls, and therefore the ultrasound angle of incidence. This measurement can be automatically

performed and used to produce calibrated velocity calculations [12].

The same transducer array is typically used for Doppler measurements and imaging, this prevents images and Doppler signals being collected simultaneously. The system can, however, be designed to rapidly switch between the two modes. This gives a refresh rate that simulates simultaneous data collection [12]. This device type is used in Chapter 5 to collect phantom data.

2.5.3 Velocity Imaging Systems

As an ultrasound sample volume is moved through the body, the size and strength of resulting Doppler signals will vary. Through mapping the sample volume location on a 2D display, and intensity modulating these positions to correspond with either the frequency or intensity of the Doppler signal, blood flow images can be generated [12]. Construction of simple blood flow images was demonstrated through performing multiple sweeps using a CW transducer connected to position sensing gantry [67]. The resulting image was improved further by colour coding the Doppler quantity [68] (e.g., a temperature scale). This coding corresponds to the peak frequency measured over the cardiac cycle; meaning the sample volume must dwell at each location for a minimum of one cycle.

This time constraint was reduced using a fast-swept linear transducer, capable of sampling each scan line multiple times during a cardiac cycle [12]. Use of a PW system provides depth resolution, allowing an image to be constructed from a desired plane. Such machines are again limited by the long duration needed to acquire a full scan. Attempts were made to overcome this using multichannel devices [69].

Velocity imaging systems evolved further by combining B-scan technology with multigate Doppler imaging technology [70]. This allowed for the B-mode monochrome image to be superimposed with colour coded Doppler information. Since these early systems, real-time colour flow mapping has experienced a number of improvements and become a popular clinical tool. As described in Section 3.5.2, previous works have extracted features from images generated using such systems. These were subsequently used to form linear regression models for diagnostic applications [71].

2.5.4 Ultrasound Research Platforms

Research into novel medical ultrasound techniques can be limited using commercially available clinical machines. This is because such machines offer very little control over signal processing methods employed, and limit the extent of accessible data.

These restrictions can be overcome using high-level research machines. These devices offer high programmability, and access to RF data, however, they tend to be expensive and are not portable [72].

In response to these disadvantages, a number of highly flexible research platforms have been designed and built [72, 73]. The application of modern technologies in these designs enables high programmability with high performing digital signal processing. A common theme through these designs is the use of field-programmable gate arrays. These devices can be used, for example, to synthesise arbitrary waveforms for transmission, perform beamforming, demodulation, filtering and down-sampling [72]. As these processes must be programmed 'from-scratch', the user has full control over their design. These advantages are discussed in Section 8.7.6 and 10 in reference to research into Doppler feedback.

2.5.5 Modern Devices

Modern machines that incorporate Doppler ultrasound range from simplistic foetal heart detectors to complex ultrasonic machines with extensive imaging modalities and probe types. The ability to measure CO transcutaneously, in a simplistic manner from locations such as the suprasternal notch, however, remain a niche application.

The presented work uses a device built specifically for this application, this is the Ultrasonic Cardiac Output Monitor 1A (USCOM). This device currently stands alone in terms of its primary application and ease of use, and has enabled research into a breadth of clinical specialities. The methods designed and investigated throughout this thesis reflect challenges associated with this machine, however, will also be applicable to other devices using the same technologies. The research and reported limitations of the USCOM device are discussed in depth in the remaining two sections of this chapter.

2.6 The Ultrasonic Cardiac Monitor

This thesis describes an investigation into means of automatically generating feedback that reflects Doppler ultrasound measurements. Specifically, this is for measurements of blood flow through the aortic valve, using a CW ultrasound transducer placed in the suprasternal notch.

This section will introduce this technique, and describe its modern implementation and validation. This gives an understanding of the central measurement procedure investigated in this work, and highlights potential sources of error and limitations that the explored solutions seek to mitigate.

2.6.1 Precursor Devices

Using a modified foetal heart detector, in the late 1960s, Light demonstrated that flow measurements could be attained non-invasively from the ascending aorta [54,74]. This was achieved using a 2 MHz CW ultrasound transducer with a slightly divergent beam, and positioned in an intercostal space. It was noted, that given the knowledge of a person's aortic cross-section, one could estimate their CO. Light suggested that the speed, safety and convenience of this technology could be valuable for clinical assessment; a sentiment that is echoed throughout much of the literature explored in this project.

Shortly following this, it was surmised similar flow measurements could be acquired by positioning the transducer in the suprasternal notch [52], an approach which was then demonstrated using both CW and PW ultrasound [75]. This work discussed the need for non-invasive methods, which could replace the more costly, troublesome and dangerous invasive flow measurements using catheters.

This transcutaneous method was later evaluated using a portable CW device called the Ultracom [62]. Using this machine, the ultrasound probe could be placed within the suprasternal notch. This allows the ultrasound beam to become parallel, or close to parallel, with the direction of blood flow. Echocardiography allowed the aortic root diameter to be measured, and subsequently the CSA to be estimated.

The probe was aimed such that resulting signals exhibited a sudden onset and cessation of systolic flow, with minimal diastolic flow. Following data collection, the machine calculated the CO from 12 consecutive beats. This was performed three times and the average of these were used. CO measurements were evaluated against metrics measured

by thermodilution. The Doppler ultrasound approach was shown to be highly correlated with thermodilution, and the authors concluded that trained and experienced technicians could accurately determine CO using the technology [62].

2.6.2 The Ultrasonic Cardiac Output Monitor 1A

Despite the long-established benefits of non-invasive Doppler ultrasound, its use for bedside monitoring has yet to become commonplace. A more recent realisation of this technology has been designed by a company based in Sydney, Australia. This device, the USCOM, has been designed to retain the strong advantages touted by those earlier implementations. The USCOM is a stand-alone system, which is both portable and, with respect to typical ultrasound machines, highly affordable. The device requires no additional hardware to function and can derive an array of haemodynamic metrics on-board. The USCOM is displayed in Figure 2.5.

Figure 2.5: USCOM 1A

The device was made available in 2006, and most variants operate using a 12mm 2.2 MHz CW half-moon transducer. The system claims to employ a unique divergent beam, whose broad characteristics makes it easier to operate, and less user-dependent than its predecessors [76]. The probe's small size allows it to be comfortably positioned in either the suprasternal notch, where it can measure blood flow traversing the aortic valve, or at the parasternal window to measure blood flowing through the pulmonary valve. Rob

Table 2.3: Overview of haemodynamic metrics measured by the USCOM 1A

Metric	Description
CO	Cardiac Output
CI	Cardiac Index
Vpk	Peak Velocity of flow
VTI	Velocity Time Integral
HR	Heart Rate
MD	Minute Distance
ET%	Ejection Time Percent
SV	Stroke Volume
SVI	Stroke Volume Index
SVV	Stroke Volume Variability
SVR	Systemic Vascular Resistance
SVRI	Systemic Vascular Resistance Index
Pmn	Mean Pressure Gradient
FT	Flow Time
FTc	Flow Time Corrected
SW	Stroke Work
CPO	Cardiac Power
BSA[1]	Body Surface Area
OTD[2]	Outflow Tract Diameter

1 – Calculated using user input patient height and weight.
2 – Calculated using user input patient height.

Phillips, the founder of USCOM, demonstrated the feasibility of measuring CO in this manner from the pulmonary valve prior to USCOMs establishment [77].

Using the maximum velocity envelope, CO and a host of additional haemodynamic metrics are calculated in real-time. These metrics are presented in Table 2.3. This is made possible by automatically extracting the MFE using a combination of the signal-noise slope intersection method (discussed in Section 3.1.2), power spectrum smoothing, envelope forcing and artefact removal [78].

Despite the straight-forward manner in which measurements can be collected using this device, concerns have been raised regarding its ease of use and the accuracy of acquired measurements. These considerations contribute to the motivation leading to the work in this thesis, and will be discussed in the following section.

2.7 USCOM Performance and Limitations

The USCOM is a self-contained CW Doppler ultrasound device, capable of providing a swathe of haemodynamic variables through non-invasive scans, which when performed by an experienced operator, can be time-efficient and accurate. In the given work, the USCOM is used to collect blood flow measurements through the aortic valve. Machine learning techniques are later designed that consider scan acquisition from this location. The USCOM is described in more detail in Section 2.6.2.

This section reviews work concerned with validating this device. These works highlight scan acquisition difficulties and potential sources of error associated with acquiring measurements. This provides an understanding of the devices limitations, and gives further context to automatic feedback models designed in later chapters. These limitations have likely contributed to the range in measurement accuracies reported for the USCOM device. These clinical validation studies will now be discussed.

2.7.1 Clinical Validation of USCOM

Since becoming available, a large pool of publications detailing research using the USCOM has amassed. Over 460 publications are claimed to prove the device's accuracy, and its suitability to be considered as a 'gold standard' [79]. However, validation studies exhibit a large range of reported accuracies. The clinically accepted limits of agreement for assessing new techniques have been recommended as $\pm$ 30% [80, 81]. Over the last two decades, several studies have compared USCOM performance to thermodilution methods, the variability in these studies is illustrated in Figure 2.6. The studies displayed in this figure are summarised in Appendix A.

In 44% of the studies displayed in Figure 2.6, a percentage error greater than 30% was recorded. However, there are a number of variables that can make drawing conclusions from these observations difficult.

One variability is the valve from which measurements are acquired. These surveys use measurements from either the aortic valve or the pulmonary valve (PV). In either case, calculated metrics rely upon estimated CSAs using nomograms. The use of which are often cited as a potential source of error [19, 82–85]. The employed nomogram provides an estimate of aortic valve diameter [86], this is then used to calculate CSA. With respect to measurements of pulmonary blood flow, CSA is estimated using an adapted version of

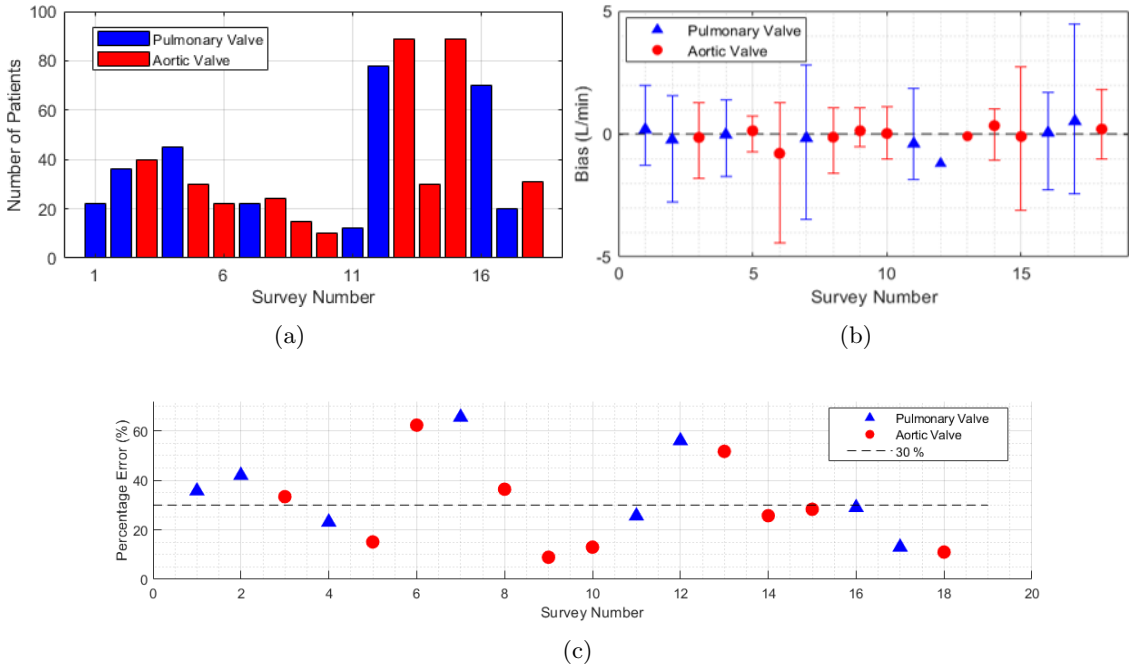


Figure 2.6: USCOM 1A validation meta-analysis, detailing a) patient numbers, b) bias including limits of agreement where available, and c) percentage error. Details regarding individual surveys are provided in Appendix A

the same nomogram, however, no studies are available that consolidate this approach [19]. Although CSA estimation is commonly cited as a potential source of error, replacing these estimates with measured values has demonstrated only minor improvements [19, 83, 87].

The parasternal window most appropriate for a given PV measurement varies and is typically a more difficult location to acquire measurements from [17, 88, 89]. The agreement of metrics using the aortic valve and PV approach has been investigated [90], and differences of less than 5% were found. Interestingly, this is despite measurements not being performed simultaneously, and significant changes in heart rate being reported. It should, however, be noted that measurements from the two locations were performed consecutively, with the more challenging PV scans being performed second. This provides the operator with a baseline value, which could either prolong effort spent aiming the probe, or reduce efforts as values are similar to those from the suprasternal notch. Repositioning of patients from supine has been shown to aid PV data collection [91]; it has been suggested

that this is because when in the supine position, the ultrasound beam is interrupted by the lung edge [89].

Due to the differences in CSA estimation viability and ease of measurement, it seems likely that inherent differences in measurement accuracy between the aortic and PV measurements may occur.

In addition to measurement location, different transducer types have been used (2.2 MHz and 3.3 MHz). As detailed in Section 2.4.1, lower frequencies offer more penetration. Clarification of transducer frequency is not always provided, however, the frequency used will likely impact the ease of performing measurements depending on the patient's physiology. Other factors can influence the ease of interpatient scan acquisition.

It has been shown that decreased accuracy, for adults, is correlated with patients age [16]. This has been linked to physiological changes and aortic calcification which can make adequate insonation challenging. A recent study compared USCOM and TTE measured CO in patients before and after venesection [92]. They found the USCOM to inconsistently detect resulting changes. Conversely, an earlier study reported much better results [93]. A younger population in this study has been suggested as a reason for this disparity [92]. Concerns have also be raised regarding performance in response to arrhythmia, aortic regurgitation or aortic stenosis [19, 94–96]. However, above these technical complications, the most commonly cited source of potential error within studies is operator experience; this is discussed further in the following section.

2.7.2 Operator Experience

Operator experience has been found to be critically linked to the quality of acquired USCOM measurements [16], with results attained during the 'learning phase' of the technology as being highly user-dependent [84,91,97]. Researchers who were unable to demonstrate good performance using the USCOM described the aiming procedure as "... blindly directing a Doppler beam at the aortic and pulmonary valves, and optimising flow traces by minute manipulations...". They suggest that feedback limited to the Doppler profile, which does not provide any positive confirmation, could weaken operator confidence [19]. Additionally, forms of feedback, could in principle guide 'minute manipulations', and provide additional positive confirmation. At present, no additional forms of visual feedback have been described in the literature.

Different levels of training have been recommended to be considered competent. A commonly cited requirement is training with 20 examinations [17], however, doubt has been expressed regarding this [19] and it has been suggested it would be more appropriate to consider this as the starting point of the learning curve [18]. A more recent study found novices with experience of over 30 examinations still produced highly variable measurements [98]. This study compared trainee measurements, which were performed on infants, with those collected by an expert. In addition to age, ethnic physiology has been suggested as a factor which can affect measurement success [16]. These elements, as well as nuances in how novices have been trained, are likely factors in the observed disparity with respect to recommended mandatory experience, and potentially the range of validation performances shown in Figure 2.6.

The reported experience of operators in validation surveys varies, and in some cases is not explicitly stated. Several studies have reported the USCOM under-reporting CO in higher flow rate conditions [84, 91, 99]. This may be attributed to the user assuming an optimal signal is being received based on the measured values appearing good quality or 'normal', preventing perseverance to find a stronger signal. This is a subtly perhaps only a highly experienced user, or additional forms of feedback could overcome.

These difficulties can sometimes contribute to failed scan attempts. The reported proportion of patients in which operators fail to acquire scans varies from 5% [88] to 20% [100, 101]. There is however no consistent condition that constitutes a failed scan, for example, Tan *et al* failed to acquire PV measurements in two patients following 45 minutes of searching [102], whereas Van Den Oever*et al* classed four PV measurements as failed after 10 minutes of examination, stating this was a more realistic time limit for clinical applications [19].

In addition to the challenges highlighted through the discussed validation surveys, concerns have also been raised with respect to the contrasting technologies used in these surveys. Specifically, whether it is correct to treat thermodilution measurements as a 'gold standard', from which alternative methods can be evaluated. These concerns, and the merits of the USCOM device with respect to other technologies, are discussed in the following section.

2.7.3 Validation Hurdles and Alternative Technologies

The USCOM is inarguably safer, quicker and cheaper than thermodilution methods, which were used in the validation surveys displayed in Figure 2.6. Issues raised through these studies include the accuracy of the USCOM device. However, concerns have also been raised regarding the accuracy of thermodilution [11, 103, 104]. A recent study compared variables measured in sheep using PAC thermodilution and the USCOM. The sheep were implanted with aortic flow probes [105]. Measurements using flow probes are highly accurate [106]. However, they are currently limited to non-human studies due to the required implantation. This enabled a fairer means of comparing the two methods. This study found that USCOM outperformed PAC in both accuracy and sensitivity for CO measurements. This is a considerable finding, and if these observations hold true for human measurements, could alter conclusions formed from previous studies. The experience of the operator in this study is not defined. However, the study was performed with the founder and an employee of USCOM, and so their experience was likely considerable.

In addition to these concerns, thermodilution using PACs provide an averaged value with inherent processing delay. Conversely, USCOM provides live, beat-by-beat measures. Therefore comparisons made across these methods use metrics derived from different time periods. This may contribute to some differences observed [107].

A recent study compared the USCOM to two additional non-invasive methods [108]. The methods, in this case, were found to be interchangeable in paediatric critical care. In areas such as paediatrics, use of PACs are rarely feasible due to catheter size [108], and there is additional reluctance to incur associated risks, which are associated even in adult studies. TTE is included in this study, which is a widely used and less invasive method. However, comparably it is slower and requires a higher level of expertise.

The advantages of USCOM also lend to its use in more remote areas, and in clinical settings where accurate assessments and treatment responses are needed [108]. Access to a user-friendly, accurate, fast, non-invasive haemodynamic monitor would result in less complications, morbidity and mortality [108].

Chapter 3

Literature Review

This chapter provides a detailed review of the state-of-the-art with respect to signal processing, analysis and machine learning techniques relating to Doppler ultrasound haemodynamic measurements. This documents the works and techniques that have guided processes in proceeding chapters, as well as the limitations of existing methods.

The content in this chapter goes beyond Doppler ultrasound focused research, and includes a review of advanced machine learning techniques implemented in further disciplines, such as speech recognition and deep learning image recognition. Several leading and outstanding methods highlighted through this are later adopted for the given work.

Guided by these works, a large range of features are extracted. This is facilitated using the MFE, which is used to both extract features and enable beat segmentation.

3.1 Maximum Frequency Estimation

Different approaches can be taken to extract information from clinical Doppler ultrasound measurements. Section 2.3.5 described how the maximum velocity envelope, which is proportional to the MFE, can be used to calculate CO. Further haemodynamic metrics can be derived using this envelope, a selection of these are presented in Section 2.6.2. This section describes the challenges involved with estimating the MFE, and the different techniques proposed to overcome them.

The maximum velocity envelope can help clinicians evaluate and diagnose medical conditions, for example, peak systolic velocities are used to assess the degree of a stenosis

[109], and to determine the need for a carotid endarterectomy [110]. Section 3.4 describes how different features can be calculated from the maximum velocity envelope to facilitate diagnosis. This information enables Doppler ultrasound to assess cardiac health [111], which otherwise would necessitate more invasive and slower traditional methods [105,112]. Considering the clinical use of maximum velocity envelopes, accurate and repeatable means of extracting them are essential.

The maximum velocity envelope is usually estimated from a time-frequency representation of the signal. This allows the frequency content being measured, and therefore the velocities, to be visually inspected. This provides a straightforward method of feedback. An example of this is displayed in Figure 1.1, which displays the MFE and the corresponding spectrogram, whose values were found using the short-time Fourier transform (STFT).

Blood flow measurements produce Doppler signals with a range of spectral content. This is both due to the effects of spectral broadening, which can effectively blur Doppler spectral content (as described in Section 2.4.3), and due to interrogated scatters having a range of velocities.

This means that visually the true maximum frequency is both hard to distinguish from a spectrogram and no longer corresponds to the true maximum blood velocity. In addition to these inherent challenges, erroneous signals such as tissue movement or nearby blood flows can be inadvertently captured, measurements can have poor SNRs, and external factors such as acquisition errors or systematic quantification can additionally complicate envelope estimation [113,114]. These hurdles often contribute to overestimations of blood velocity [115].

Despite the limitations of visually assessing maximum frequencies, MFEs are manually traced in conventional Doppler ultrasound devices. These traced envelopes are then passed to a second computer which calculates haemodynamic properties (e.g., CO) [116]. In addition to concerns regarding accuracy, this procedure is also subject to both inter- and intra-observer variation [117] and is very time costly [118]. These factors, as well as the clinical benefits of averaging measurements [119–121], make automatic methods of envelope estimation desirable.

The valuable information contained within MFEs, and the challenges involved with extracting them, have led to an array of MFE estimation methods being proposed. These techniques will now be described.

3.1.1 Early Methods

A range of methods have been proposed to estimate maximum frequency points from Doppler signals. The typical steps involved in this process are displayed in Figure 3.1 [122]. Early attempts were restricted to analogue methods, for example, systems using banks of narrow-band analogue filters, the outputs of which were then thresholded to give maximum frequency estimates [54, 123].

As technology progressed, digital methods were designed [124, 125]. Approaches included the use of Doppler signal histograms [125], however, assumptions regarding power spectrum density and probability distributions restricted these methods [122]. Furthermore, they were limited in terms of signal-to-noise discrimination.

Prompted by this, D'Alessio proposed a digital threshold crossing method. This method objectively estimated maximum frequency through quantifying noise threshold levels from the power spectrum [126].

In this approach, sequences of power spectrum bins are interrogated from upper to lower frequencies. The maximum frequency is defined as the highest bin at which the spectral power in successive bins exceeds a set threshold.

The threshold is estimated from the tail-end of the Doppler spectrum. As white noise exhibits a flat spectrum, this approach gives information on the level of noise throughout the signal. Shortly following this work a study contrasted the performance of D'Alessio's method, with three other methods [122]. One of these methods, the 'hybrid method' uses the integrated power spectrum (IPS).

3.1.2 Integrated Power Spectrum Methods

Mo *et al* compared the performance of four maximum frequency estimation methods [122]: D'Alessio's threshold crossing method, a modified threshold crossing method (MTCM), a percentile method and a hybrid method. The IPS is used to contrast how each method operates, and to introduce the hybrid method that uses it directly. The hybrid method

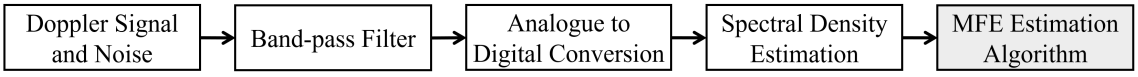


Figure 3.1: Block diagram of general maximum frequency estimation process

identifies the maximum frequency as a point on the IPS curve where the signal transitions to noise. The IPS is calculated using equation 3.1,

$$IPS(f) = \int_{f_L}^{f_H} P(f)df \tag{3.1}$$

where $P(f)$ is the estimated power spectrum [122]. As will be described, a variety of models that use the IPS have been proposed. When applied to a spectrogram, maximum frequency values can be estimated for each time point, thus, producing an MFE. An example of the IPS curve is displayed in Figure 3.2. The IPS methods function based upon assumptions regarding the shape of this curve. The extent to which an IPS resembles its characteristic form is determined by SNR levels, and the presence of erroneous signals.

The primary difference between the MTCM and D'Alessio's method is how the signal to noise threshold is calculated. In the case of the percentile method, a user-defined percentage of the IPS curve is used. The percentage used is chosen with respect to SNR. This introduces a hardcoded (i.e., fixed) value, which does not change in response to evolving levels of noise [122].

The hybrid method uses a geometric approach. Maximum frequency is identified using a straight line that intersects the rising IPS and ends at the maximum frequency bin. The slope of this line determines the intersection point and is calculated using the level of noise in the signal. This intersection point is defined as the maximum Doppler frequency. They

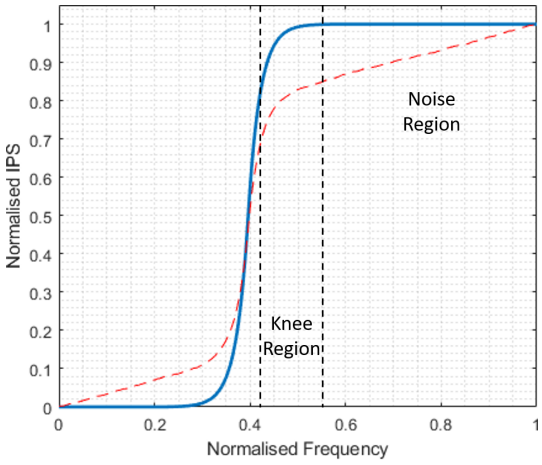


Figure 3.2: Example IPS curves

concluded that the MTMC and hybrid models performed the best, and that performance was dependent upon SNR and the shape of the power spectrum [122].

Moraes *et al* performed a similar study, comparing three MFE estimation algorithms [127]. They compared a simple user-defined threshold method, the MTMC method and a modified version of the geometric method (GM) [128]. As with the hybrid method, GM uses the IPS and determines maximum frequency using a geometric approach. The maximum frequency is identified as the point on the IPS which is at a maximum perpendicular distance from a reference line. This is a straight line plotted from the frequency with peak spectral power, and the end frequency.

The modified GM (MGM) differs to the original with respect to how coordinates are determined for the reference line; this removed inaccuracies incurred due to variable signal gains [127]. Moraes *et al* found that all three methods performed similarly, and suggested algorithm choice should be made with respect to their application.

More recently, Kathpalia *et al* presented a new IPS geometric method. This method is an adapted version of the signal noise slope intersection method [113]. This modified signal noise slope intersection (MSNSI) method incorporates steps from GM [128]. In this method the IPS is divided into three regions; the signal region, the noise region and the knee region. As with GM and MGM, a reference line is used, lines of best fit are calculated for the noise and signal regions separately. The gradients of these lines are then used to model the IPS. This allows the knee region to be inspected, within which the signal transitions to noise. This relationship is shown in equation 7.1.

$$M(x) = m_s x + m_n(1–x) \tag{3.2}$$

Where x is the fractional signal contribution, which ranges from one at the start of the knee region, to zero at the end of the knee region. If Doppler signals contained no broadening, the maximum frequency would occur at an x value of zero. A value of 0.1 was heuristically found to give the most accurate estimations. This allows the MSNSI method to some extent account for spectral broadening. As x was arbitrarily chosen based on observations, it follows that it could be tailored for specific scenarios, leading to more accurate results.

The MSNSI method also provides steps for envelope estimation as a whole. These additional steps aim to make envelope prediction more robust and accurate and are based

on heuristic thresholds, interpolation and smoothing. These are in response to the challenges which can occur during envelope estimation, such as erroneous signals or reductions in SNR. These can be particularly challenging for the IPS methods, which necessitates a characteristic IPS curve to produce accurate estimations. At time points where this is not the case, estimations of maximum frequency can be wildly inaccurate. MSNSI attempts to overcome this by identifying these regions using a dynamic noise threshold. These regions can then be removed, and replaced using interpolation. Where a set number of sequential time points exceed this threshold, the MFE is set to zero. Consequently, noisy signals can potentially result in MFEs with missing regions.

In practice, the resemblance of an IPS to its characteristic form varies across sequential time points. This can produce an erratic envelope with high standard deviation (STD). The MSNSI method employs envelope smoothing to mitigate this.

The methods discussed in this section provide a practical means of estimating MFEs. However, their operation is based upon heuristic factors, are significantly impacted by noise and can only provide approximate estimations. A potential approach, which can theoretically overcome these hurdles, is to use spectral modelling.

3.1.3 Spectrum Modelling Methods

Given enough information about a system, representative models can be designed to mimic them and subsequently garner more information about the underlying processes. Vilkomerson *et al* presented a new method of peak velocity identification using a model of the Doppler power spectrum [129]. This model represented steady parabolic flow and assumed uniform insonation. The resulting spectrum featured a plateau at low frequencies, which would then descend to the noise floor as frequency increased. This model revealed that peak velocity occurred at the midpoint in power on the descending slope of the Doppler power spectrum. To assess a measured spectrum, the spectrum is scanned from high to low frequencies. The mid-point is the location that corresponds to half the power at the end of the plateau region (which is estimated using an iterative process). Vilkomerson *et al* demonstrated high accuracy, and a robustness to SNR using this method.

This method assumes a uniformly insonated vessel and accurately known angle of insonation. This limits the practical application of this method, as both insonation angle

and non-uniform insonation are known primary sources of error for Doppler ultrasound measurements [117, 130]. This is particularly true for applications where measurements need to be attained from a range of angles and signal qualities, as required in the presented work.

3.1.4 Image Processing Methods

Doppler envelopes often need to be traced by a trained technician [131, 132]. This is performed by outlining the perceived edge on the ultrasound machine using a track-ball. This process requires a still frame and is often performed from video recordings post-examination due to time restraints [131]. This process of outlining the MFE using an image has been replicated using software. As with the IPS methods, this can overcome subjective and efficiency limitations of the manual tracing approach.

A limited number of image-processing methods have been described. The methods use a pre-processed image of the spectrogram, and as with the IPS methods must contend with challenges such as speckle noise, wide SNR variation and aliasing [132]. The described approaches use a number of common steps, which include capturing and cropping the video output from an ultrasound machine, converting still images to a binary image, and then extracting the MFE. An example binary image is included in Figure 3.3.

The first identified example was presented in 2001 [131]. This method used image sequences recorded onto videotape, with a time width and refresh rate of 2.5 s and 1 s, respectively. MFE extraction from images requires several steps. These include detecting the image portion displaying the spectrogram image, and ensuring un-associated image sections such as axis are not processed. High-frequency noise is filtered out using a Gaussian low pass filter, with empirically chosen parameters. Using a nonlinear Laplace edge detector, a greyscale image is formed using intensities which correspond to edge strength. A threshold is then derived using pixels close to zero velocity. Noise in the resulting binary image (i.e., pixels that should correspond to black, in fact correspond to signal) are removed, by elimination regions of connected pixels smaller than a set amount. Ideally, following these steps, a black and white image remains in which all white signals correspond to measured blood velocities.

The MFE is extracted from the binary image using a 'biggest-gap' algorithm [131], whereby each image column is inspected separately. The signal pixel that is adjacent to

the largest number of consecutive non-signal pixels is recorded as the maximum velocity point. This process is illustrated in Figure 3.3. The method is made more robust by applying weights to gap sizes depending upon the velocity region in which they occur. This addition, however, imposes assumptions regarding the measured velocities. The resulting MFEs were smoothed using an average filter and compared to manually traced curves. This evaluation was performed using approximately 11 minutes of data, a bias of 0.4 ms^{-1} was reported for peak velocities; it is unclear how the method performed across the whole cardiac cycle.

Several similar approaches have been described. These have been focused on automatically tracing MFEs from the mitral and tricuspid valve [133–135]. High correlations with hand-traced envelopes have been reported. These were calculated using 167 to 467 beats across the methods.

Differentiating the signal portion from the greylevel images during diastole is more challenging, to overcome this the diastolic portion can be effectively set to zero [135]. Evaluations are restricted to systolic flow, and bias with respect to manual tracings of 3.6% to 5.6% were reported [135]. These approaches would be inappropriate in cases where diastolic MFE information is needed. Furthermore, evaluations have been performed using limited recordings and for recordings made with optimal conditions.

Newer image-processing methods have been described that form binary images directly from spectrogram images (as opposed to first applying an edge detector). This is achieved

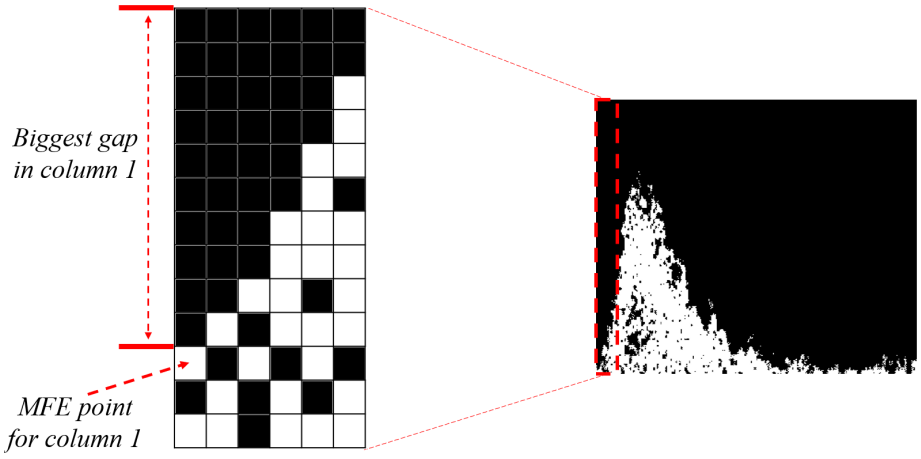


Figure 3.3: Illustrates MFE extraction from binary image using biggest-gaps method

by thresholding the spectrogram image. The image quality, and therefore accuracy of extracted MFEs, are determined by the selection of this threshold. A threshold too high will remove the wanted signal, and give negative bias. Conversely, a threshold too low will produce a binary image with excessive noise, from which the MFE cannot be estimated.

A semi-automatic method was described for transthoracic coronary Doppler flow [136]. The threshold, in this case, was found by inspecting the pixel intensities at three regions within the image. With respect to the inverse cumulative histogram of these pixels, the intensity at which 25% of intensities were greater was chosen as the threshold value. This process assumes a given percentage of image pixels correspond to signal, and is thus dependent upon both device settings and the blood flow being measured. This limitation was highlighted as the implemented percentage had to be manually changed in 10% of beats. The maximum frequency for each column is assumed to be the signal-pixel which occurs before the first non-signal pixel. This would prevent use with CW ultrasound examinations where signal strength is not constant below the maximum frequency, or when using PW ultrasound. The image threshold has also been estimated using the Otsu method [137]. In this case, lots of noise can persist in the binary image, which can be removed by eliminating clusters smaller than a set size. MFE performances were not evaluated in this case.

A recent method used an objective thresholding technique. This applies a range of thresholds to the image, and the number of pixels remaining following each threshold application is recorded. The chosen threshold is taken as that which corresponds to the maximum gradient in these recorded data. This is illustrated in Figure 3.4. This method was again compared with hand traced envelopes and demonstrated good correlation with calculated metrics.

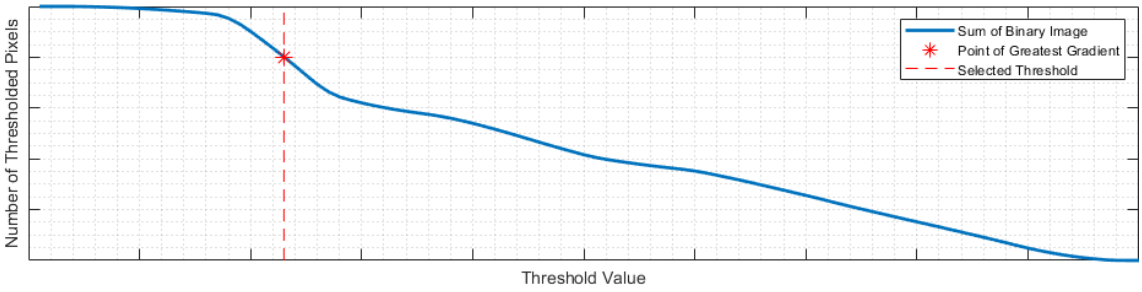


Figure 3.4: Threshold identification process used by Zolgharni *et al*

The image-processing methods described have shown good correlation with hand traced envelopes. These evaluations have, however, been limited to the systolic portion of the signal. Furthermore, the number of beats assessed has been low and restricted to high-quality measurements. The generalizability of these methods is also questionable as images are captured from secondary machines, and thus undergo unknown processing. In no example were the temporal and spectral resolutions of pixels reported, calling into question how transferable the methods are. This is illustrated by the span of cluster sizes used when removing noise, which ranges from 50 [137] to 500 [132].

MFEs in the given project are used to both segment data and extract a range of features across the cardiac cycle. To achieve this, an image-processing method would need to perform well across diastole and a far wider range of signal qualities and types thus far tested. Hitherto, evaluations have not included quantitative assessment using data with known associated maximum velocities.

The discussed MFE estimation methods have been evaluated using a combination of simulated data, and data measured in-vivo and from phantoms. Evaluations using a combination of these data types allow for a thorough interrogation of methods.

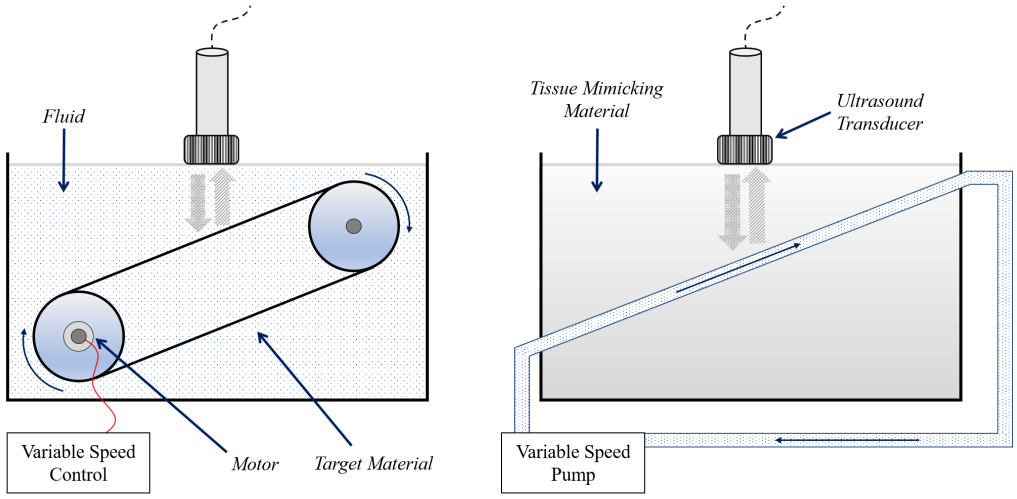
3.1.5 Phantoms for MFE Method Evaluation

Phantoms provide well defined and reproducible testing characteristics. This enables quantitative evaluations of Doppler ultrasound equipment, processing methods and data interpretation. They have been used to calibrate equipment and to negate inaccuracies born from geometric spectral broadening [130], and have highlighted common limitations of Doppler ultrasound machines, including significant overestimations of maximum velocities [138].

These physical models belong to one of two groups: string phantoms and flow phantoms. In the former, the interrogated target is a thin loop of material such as nylon, o-ring rubber or silk [139]. This material is looped over wheels rotating with known and variable velocity. The system is typically submerged in a liquid with similar sound speed to tissue, and precautions are taken to prevent air being introduced [139]. String phantoms can be used to assess the accuracy of maximum velocity estimation, however, they do present some limitations. Notably, in contrast to clinical measurements, at any time there is only one velocity. This can prevent reliable evaluation of extracted waveform indices,

particularly when the degree of intrinsic spectral broadening is low [139].

Flow phantoms allow for a more physiologically accurate model, and models have been designed to generate a wide range of anatomical waveforms [140]. These models typically use fluids with similar ultrasonic scattering properties to human blood, which is pumped through vessels embedded in tissue-mimicking material [141]. As with string phantoms, pumps are chosen to ensure air is not introduced into the system. Flow phantoms can also be used to assess estimated volume flow rates, and spectral broadening indices. Additional applications include comparison between operators and for use as a teaching tool [140]. The general designs of a string and flow phantom are illustrated in Figure 3.5.



(a) String Phantom (b) Fluid Phantom

Figure 3.5: Phantoms block-diagrams

Steinman *et al* designed a flow phantom to investigate sources of error in maximum velocity estimations. They used this system to evaluate their SNSI MFE estimation method using a wide range of insonation angles and steady flow rates. More recently, a string phantom was used in the assessment of the MSNSI MFE estimation method, using steady speeds of 60 and 100 cm/s, and insonation angles of 27° and 58° [114].

Phantoms serve as a vital tool for evaluating ultrasonic hardware, however, can be limited with respect to the range of scenarios that can be investigated (e.g., blood profiles, velocities or insonation angles). When evaluating software, simulations can be used to

provide further and more variable test data.

3.1.6 Simulations for MFE Method Evaluation

Doppler simulations can be used for extensive testing of processing methods. With respect to phantoms, they are cheaper, often faster, and provide complete control each parameter. These features allow quick investigations of different scenarios. In this section, methods of Doppler ultrasound simulation are discussed.

A real-world Doppler signal is both frequency and amplitude modulated, where frequency is directly related to the velocity of the interrogated blood. Doppler signals exhibit stochastic characteristics and can include spectral broadening [142]. Their complex nature makes the process of simulation non-trivial.

The Doppler signal can be considered to be the sum of contributions from numerous individual scatters, which pass through the sample volume. This makes it a Gaussian random process [143]. Several models have been derived by representing Doppler signals as bandlimited Gaussian noise. The random fluctuations that occur in Doppler signals, and the performances of different spectral estimators have been investigated using this technique [143, 144]. These models used a frequency domain approach, involved the summation of many sinusoids and required a power spectrum as a major input. This was either experimentally measured or theoretically derived.

This modelling approach has been used to investigate mean frequency estimators [145], and a selection of MFE estimators using stationary and nonstationary simulated signals [122, 128]. Gaussian noise was added to these simulated signals to yield varying SNRs; allowing the robustness of MFE estimators to be evaluated [122, 128]. As true maximum velocity is known in these scenarios, bias and STD statistics were calculated to evaluate the extracted MFEs.

These statistics have been used to evaluate several MFE estimators using emulated signals [127]. In this case, white noise was sampled and then processed to produce narrowband quadrature signals. The effects of spectral broadening were somewhat incorporated by having the 'maximum frequency' components located in the fall-off region.

In addition to these Gaussian noise models, simulations can be achieved using wave scattering models [146]. Instead of simply replicating the stochastic nature of Doppler signals, scattering models consider the underlying physical mechanisms. This approach

can also include information regarding the Doppler instrument, and allow the effect of varying parameters such as volume size and beam pattern to be included [142]. Bastos *et al* presented a scattering model which combined previously described techniques, and was capable of simulating pulsatile blood flow. Their model considered the signal to be a summation of contributions from subvolumes of red blood cells [146], introduced spectral broadening effects through modelling finite sample volumes [142], and included dynamic velocity profiles using the Womersley model [12,147] The MSNSI method was evaluated in part using a simple simulated Doppler signal [114]. This simulated a parabolic steady flow and used a method similar to that presented by Guidi *et al* [148]. The described simulation methods, and in particular those that have been used to assess MFE estimators, have predominantly simulated simple flow conditions, for example, steady flowrates without spectral broadening. They allow the process to be simplified, however, limit the accuracy of the resulting data.

Field II is a software that performs numeric ultrasound simulation. The software generates raw data similar to that measured by an ultrasound machine, which could enable MFE estimation performance to be rigorously tested on very realistic data. This software, which was developed by Jensen [149, 150], has been well-validated and used extensively for ultrasound research. Applications of Field II include investigations of transducer performance [149], image reconstruction techniques [151] and assessment of Doppler flow algorithms [152]. Jensen et al [153–155] have demonstrated that the simulated data can be used to accurately obtain velocity estimates.

The software, which uses the spatial impulse response method to allow generation and analysis of ultrasound fields, is written in C and uses MATLAB as an interface. This allows a high level of control of the model, including definition of transducer properties (e.g., geometry, apodisation and impulse response), as well as the positions of interrogated scatterers, which are described in 3D space using Cartesian coordinates. Distinctions between different structures such as blood or tissue are possible by assigning different acoustic values to scatterers.

In pulsatile flow, the velocity profile changes as a function of time. Waveforms from different arteries can be synthesised using their Fourier components [12]. As with the above examples, this employs the Womersley model. These time-dependent profiles allow the position of the modelled scatters to change between ultrasound pulses, and thus simulating pulsatile flow.

3.2 Beat Segmentation

Doppler ultrasound, and the extracted MFE, enables straightforward and efficient examinations. However, to calculate quantitative metrics cardiac timing indices are typically required. These indices allow information from particular regions of the signal to be analysed, for example, the systolic portion when calculating stroke volume. This section discusses means of acquiring such timing information.

Traditionally these timing indices are found manually, which clearly prevents instantaneous measurements and is disadvantageous in clinical scenarios [156, 157]. Automatic means of acquiring these indices have been described. These methods either use additional hardware, such as an ECG, or function using the shape of the Doppler MFE. Typically, the former approach is implemented.

Gibbons *et al* presented a system capable of providing real-time PI measurements. This was achieved using a three-lead ECG to act as a trigger. The R-wave associated with the ECG signal is used to provide a gating signal, waveforms during these isolated portions are then used to calculate PI metrics. This approach was later used for automatic carotid artery classification [157], where 20 beats were segmented and averaged; this ensemble average was then used to classify degrees of stenosis. Their method introduced a condition to decrease false detections based on timing variations in succeeding beats. If, with respect to the previous beat, the heartbeat varied greater than 110 ms or the R-R interval differed greater than 33% , the current beat was rejected [157]. A similar condition has been described, where beats exhibiting a heart rate difference greater than 10% , with respect to the overall measured heart rate, are rejected [158].

The ECG signal used in this manner ensures the same portion of the cardiac cycle is always analysed, this approach was used recently to investigate potential means of cancelling intrinsic spectral broadening via ultrafast Doppler ultrasound imaging [159]. Through introducing an ECG trigger delay, waveforms can be segmented at the start of systole. This delay is set by the operator [70]. This was demonstrated as a means of automatically calculating CO [55]. In this example, operator input timing information and the ECG trigger signal were used to integrate measured velocities during systole, giving the stroke volume. This time delay is illustrated in Figure 3.6.

ECG signals can provide a robust means of beat segmentation, however, they do exhibit a number of limitations. Firstly, an ECG requires electrodes to be attached to the

patient, which introduces an additional level of complexity to the examination process [118, 160]. These additional steps and inherent time requirements are particularly restrictive in emergency scenarios, such as on-scene or during air rescue [14]. Secondly, the trigger delay required to capture indices with known locations (such as the start of systole), has to be adjusted with respect to the measurement location [118].

These limitations have led to beat-segmentation methods that function solely using the MFE being developed. Inherent to this approach are a number of challenges. These include a wide variation in MFE shape, slow velocities, artefacts due to turbulence or noise, MFE offsets due to steady flow components, and significant variations in pulse durations in certain conditions [118].

Johnston *et al* described a means of estimating timing indices shortly after the early ECG methods were described [118]. Their method used the extracted MFE and was able to detect the start and endpoints of the waveform, which were then used to calculate PI. The method was reviewed by technicians using approximately 35 minutes of data, which was recorded from four different arteries. This evaluation found that the method

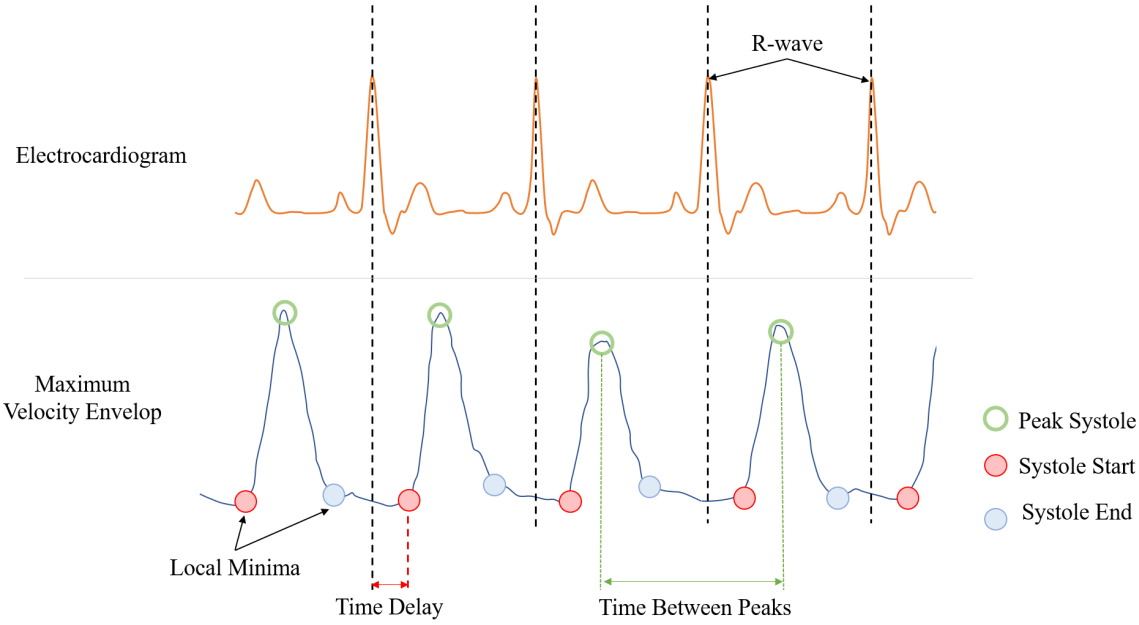


Figure 3.6: Common features used in beat segmentation for ECG and MFE shape approaches

successfully segmented 96% of recordings. Their method processed Doppler audio in 5.12 s segments, from which the MFE is estimated. The five highest peaks with a minimum separation of 200 ms are identified. The minimum distance between identified peaks is then used to find additional peaks, by assuming a maximum pulse interval variation of 20%. This process assumes a heart rate of less than 110 bpm. The first and last beats are rejected if they are incomplete.

The authors partly attribute the high performance of their method to the data used. This is because processed recordings were acquired by 'freezing' the waveform once the operator identified the optimal probe position, and artefacts and noise were at a minimum. They found reasons for failure included heart rates below 50 bpm and greater than 110 bpm, arrhythmias and cases of extreme damping [118]. The minimum heart rate condition makes the method more robust but inherently restricts its application. Rejection of incomplete beats inevitably will result in fewer beats being processed, in modern applications where memory is less restrictive, a sliding window could be implemented. For example, the following 5.12 s of processed audio could be selected to include the start of an incomplete beat in the preceding MFE.

In addition to reducing examination complexity, a shape-based method can be used in scenarios where an ECG cannot. This is the case for foetal work, where maternal signals can interfere with an ECG [160] . Thompson *et al* proposed an autocorrelation method for identifying the start of systole in sequential beats, using MFEs from the umbilical artery [160]; this enabled 7-8 cycles to be averaged. Other waveforms, such as the mean velocity, absolute sum and first moment were considered for extracting timing indices, however, the MFE was found to be the smoothest and least prone to errors. The method required an initial start point to be entered by an operator and was too slow to be implemented for real-time use. A similar method was used to investigate the effect of physical parameters on extracted MFEs and PI [161]. Modern technology would allow real-time implementation of this method, however, the necessity for user inputs prevents it from operating in a fully automatic manner. Tissue Doppler signals offer another means of creating a gating signal, and this has been demonstrated as a means of synchronising 3D ultrasound imaging of foetal hearts [162].

Autocorrelation has also been used to estimate heart rate, and subsequently define a minimum distance between peaks in the MFE [132]. This approach is used to first identify peak systole positions, and assumed a minimum distance of 80% of the estimated

heart rate. The start and end of systole were then found based on heuristic properties of the envelope. This is performed by calculating the first derivative of the velocity curve, and then identifying local maxima and minima in close proximity to the peak of systole. The distance between peaks and the assumed location of minima are illustrated in Figure 3.6. This approach relies upon the envelope remaining close to its characteristic shape, and furthermore, the method was only tested on data deemed to be of good quality by a technician.

The USCOM 1A provides instantaneous haemodynamic measures. The onboard flow tracking software includes algorithms for the aortic and pulmonary valve, which estimate the start and end of systole. The methods employed within this software have been described at a high-level in the corresponding patent [78]. The start of systole for aortic valve signals is found by first identifying approximate start positions. This is done using a velocity threshold, and subsequently a gradient threshold for that portion of the envelope. To make the method more robust, the algorithm ensures the velocity remains above this set threshold for a given period. The identified approximate location and slope gradient is then used to extrapolate the rising edge down to zero; this point is used as the valve opening position. In practice, the USCOM screen is frozen once an optimum trace is found. The positions are then displayed and can be adjusted by the user. Small adjustments are occasionally required due to, for example, noise spikes. An example of this is displayed in Figure 3.7 . In terms of routine examination, these adjustments are not a concern as they are quick to perform. These limitations, however, could be impactful for research applications or when instantaneous metrics from poorer quality scans are needed.

A variety of beat segmentation methods have been described, the evaluation of these methods has however been restricted to optimal recordings and are largely not automatic. In addition, tests have used limited numbers of beats, and methods have not been directly compared. Having acquired cardiac timing information, quantitative features can be derived and used for diagnostic purposes.

3.3 Doppler Ultrasound Diagnosis

The diagnostic potential of ultrasound was fist demonstrated over 70 years ago [48], since then vast advances in physics and technology have led to ultrasound technologies becoming a core clinical tool [50, 163]. Doppler ultrasound can be used to attain an array of

haemodynamic metrics, and to diagnose a range of conditions. The technology can be implemented in a wholly non-invasive manner, which makes it highly desirable for many applications.

As blood flows throughout the body, its characteristics at any point are governed by fundamental physical laws. A basic comprehension of this underlying physics gives a fuller understanding of blood flow measurements performed using Doppler ultrasound, and subsequently gives a degree of diagnostic ability. Bernoulli's principle states that an increase in flow velocity will correspond with a decrease in static pressure [164]. This relationship was used to develop the Gorlin formula, which can be used to calculate stenotic orifice area [165]. These concepts were later used to investigate stenotic aortic valves via non-invasively Doppler ultrasound [166].

The flow rate through a vessel stenosis must equal that through the healthy portion of the vessel. Considering this, the velocity through this diseased section must increase to maintain a constant flow rate [12]. Immediately following a stenosis, a region of turbulent and reverse flow can occur [12]; similar to an eddy one might see following a contraction in a river. These characteristics are illustrated in Figure 3.8. This figure highlights the change in vessel area with respect to the healthy region (A_1) and the diseased region (A_2). Their relationship with respect to flow rate (Q) and average velocities are given below.

$$Q = A_1 v_1 = A_2 v_2 \tag{3.3}$$

Where v_1 and v_2 are the average velocities within the healthy and stenosed vessel

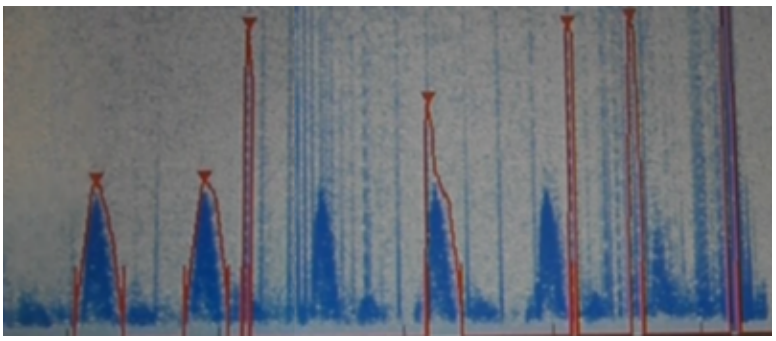


Figure 3.7: USCOM flow tracking in presence of noise, with beat segmentation requiring adjustment due to incorrect peak identification and beat segmentation

regions, respectively.

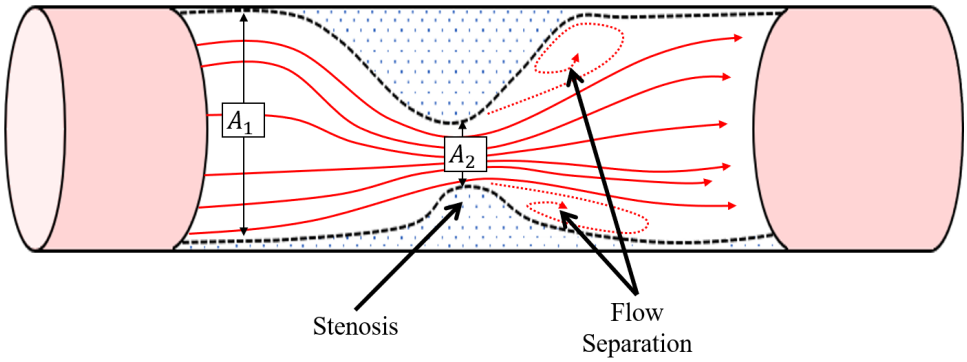


Figure 3.8: Example flow characteristics of stenosed vessel

Spencer *et al* analysed increases of measured blood velocity to quantify the severity of carotid stenosis [167]. This work was followed by several studies that used peak Doppler frequencies to detect carotid stenosis [168–171]. Peak frequencies attributed to a Doppler measurement are found using the associated MFE. To obtain accurate values, the insonation angle must be precisely measured and the effects of intrinsic spectral broadening should be accounted for (these considerations are discussed in Sections 2.4.2 and 2.4.3, respectively).

These techniques enable Doppler ultrasound to join a group of technologies that can be used to diagnose heart disease [172, 173]. Coronary heart disease (CHD) is responsible for around one-third of all deaths in people older than 35 years [174] and is expected to become more common, particularly in developing countries. This reality makes early and accurate diagnosis highly desirable.

To strengthen and extend the diagnostic capabilities of Doppler ultrasound, a variety of features that can be extracted from measurements have been derived. There have been several studies that investigate the use of such features to diagnose stenosis and CHD.

3.4 Diagnostic Features

Doppler heart sounds are collected by placing an ultrasound transducer on a patients chest. The resulting signals contain contributions from blood flow and valve movements, and can

be used to diagnose CHD [175]. This process, and the diagnoses of other conditions, can be strengthened using extracted features.

Early examples mostly used one-dimensional feature vectors, with particular focus on one aspect of the Doppler signal. These were based on the shape of the Doppler profile, or the associated spectral broadening [12]. Single-site normalised indices are a group of features found using the MFE [12], and are popular for applications of CHD detection. These indices use ratios of heights, corresponding to different characteristics within the MFE of a single cardiac cycle. As the numerator and denominator in such ratios both depend on the Doppler insonation angle, inaccuracies associated with incorrectly identifying this angle can be overcome [12]. Figure 3.9 illustrates quantities commonly used to extract such features.

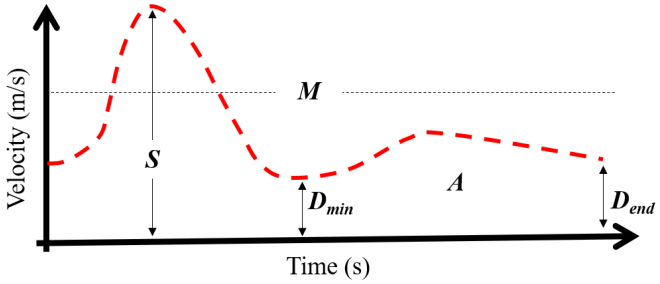


Figure 3.9: Doppler profile characteristics commonly used to derive waveform indices and simple features, where S is the peak systolic frequency, D_{min} is the minimum frequency, D_{end} is end diastolc frequency, A is area below the envelope, and M is the mean envelope value.

The pulsatility index (PI) and Pourcelot's resistance index (RI) are both examples of simple features derived from the MFE. These are calculated using the peak, mean and minimum values associated with an individual cardiac cycle [12].

$$PI = \frac{S - D_{min}}{M} \tag{3.4}$$

$$RI = \frac{S - D_{end}}{S} \tag{3.5}$$

RI is an indicator of circulatory resistance beyond the measurement point [12], and PI describes the pulsatility of a waveform and can aid the diagnosis of peripheral vascular

disease [176]. A variety of similar indices can be extracted to describe changes in waveforms. These have been used for studies in areas such as obstetrics [160, 177, 178], and analysis of supraorbital arteries [179].

However, use of features to assess stenosis, such as PI [180, 181] or Laplace transform coefficients [182, 183], did not achieve widespread use [158]. This is due to them being affected by distal stenoses, and being inaccurate for lower grade stenosis [184–187]. To overcome these limitations, and to unlock further diagnostic capabilities of Doppler ultrasound, machine learning techniques have been investigated.

3.5 Machine Learning with Doppler Ultrasound

Human-machine interfaces can both extend diagnostic capabilities of clinical technologies, and reduce dependencies upon operator expertise and experience. This can be achieved by providing feedback that aids interpretation of measurements and gives diagnostic indications [173]. Recent advances have led to solutions which implement machine learning. This is an interdisciplinary field, which aims to use example data to form predictive algorithms, and via these generate predictions from new data [163, 188]. Advances in machine learning have contributed to advances in computer-aided disease diagnosis, for example in radiology to detect and classify lesions in the breast and liver [163]. Typically constructing models to perform such tasks consists of several common steps, namely feature extraction, feature selection and model training. This fundamental approach has been investigated for over 20 years [163]. However, for any given task, the extracted features and employed methods can vary widely.

To build on the information simple features such as waveform indices provide, and to strengthen predictive capabilities, groups of features can be extracted. Using feature selection, a subset of these features can then be chosen and used for classification tasks.

3.5.1 Doppler Ultrasound Classification

A variety of model types and classification tasks have been described using Doppler ultrasound. Early examples demonstrated promising results using models trained to classify the extent of atherosclerosis in the carotid artery [124, 157, 189]. Using a set of 94 features, derived mostly using the velocity waveform and spectral widths during the cardiac cycle, a

subset of 2-4 features were selected. These selected features allowed a hierarchy of models to be built and to classify the degree of stenosis. There are many examples of Doppler ultrasound being used to investigate stenosis in the carotid artery, the location of narrowing is most commonly the carotid bifurcation point. An example of carotid atherosclerosis at this point is illustrated in Figure 3.10.

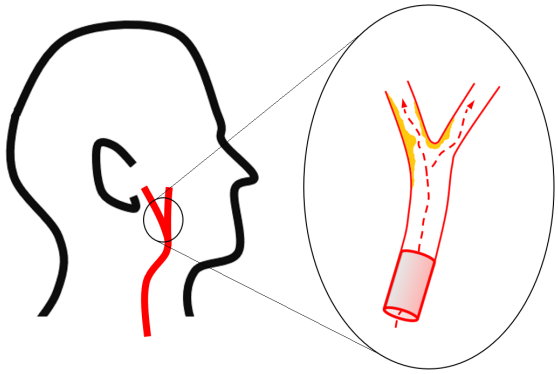


Figure 3.10: Example of carotid artery stenosis, caused by atherosclerotic plaque build-up at carotid bifurcation.

Data from the common femoral artery was used to investigate 20 features for the assessment of aortoiliac disease [190]. Using 19 features, a similar approach was used to classify the severity of stenosis in lower limb arteries [158]. Features were extracted from a spectrogram and were based upon peak systole frequency content, spectral broadening and normalised amplitudes. Following these early studies, the use of artificial neural networks (ANN) for classification tasks has become increasingly popular. ANNs consist of multiple interconnected processing elements, arranged into layers [175]. In this regard, they have been likened to a brain. These models are able to learn nonlinear relationships, as well as the linear discriminant [188].

Blood flow within arteries can be regarded as an inherently nonlinear process [175,191, 192]. ANNs may then be considered advantageous, as they are capable of solving nonlinear tasks and can overcome noise and inaccuracies. These potential benefits with regards to Doppler ultrasound classification have been investigated. Using MFEs as input data, a trained ANN was able to accurately distinguish which of three arteries a Doppler signal was recorded from [175]. Improved detection of abnormal neonatal cerebral haemodynamics

was demonstrated using compensatory fuzzy neural networks [193]. In this case, features included two principal components computed from averaged waveforms using principal component analysis (PCA).

A range of models have been described that aim to classify heart valve Doppler signals as either diseased or normal. These have largely used a combination of MFE and wavelet transform based features (such features are described in Section 3.6.1.2), combined with PCA to give reduce feature sets. A variety of model types have been tested, including the discrete hidden Markov model [194], Least-squared SVM [173, 195], a number of different forms of neural networks [172,195–197], and neural network ensembles [198]. These models have all reported high classification accuracies, of typically greater than 96%. However, the data used within these surveys differ, and the number of observations contained within each data set is limited, ranging from 105 to 215. Furthermore, accuracies have been calculated using one round of cross-validation (e.g., using 75%, and 25% of the data for training and testing respectively [198]), apart from [173] where 3-fold cross-validation was implemented. Considering these limitations, it is difficult to make concrete conclusions as to the best performing models and/or feature combinations.

Similarly, in recent years a range of models and feature types have been investigated for classifying carotid artery atherosclerosis. Feature types explored for heart disease are again used and a number of model types are evaluated, including artificial immune recognition systems [199–201], Least-squared SVMs [202] and ANNs [203]. These again exhibited very high accuracy rates (all bar one achieving 100% classification accuracy). These evaluations implement 10-fold cross-validation, however, the number of observations is more restricted ranging from 78 to 114. As with the heart disease models discussed above, it is difficult to make solid deductions about the best methods used.

These approaches have also been used to perform more informative classifications, using three to four categories for ophthalmic arterial Doppler measurements [204–206], and three categories for carotid arterial Doppler measurements [204–208]. These methods all implemented types of neural networks and demonstrated high classification accuracies. A comparison between a variety of different classification models, including three different neural network types, mixture of experts, a modified mixture of experts and SVMs was evaluated using data from these two locations [205]. Three categories were used for each data set; normal, stenosis and Behcet disease for OA, and normal, stenosis and occlusion for the carotid data. This analysis found that SVM models resulted in the highest

accuracies for both data sets [205].

The models discussed in this section have been concerned with sorting observations into distinct categories. The classifications of stenosis, for example, give limited information regarding its degree. A range of features have been extracted throughout these works, but no thorough feature selection or comparison of the features have been performed. No works have been reported concerning probe position classification.

Labelled data sets can also be used to create regression models, which generate continuous values using a set of observations. This can be useful for predicting subtle changes in a measurement, which otherwise would not be captured using distinct categories; this is explored in the given work in Section 7. Regression models can also be used to predict the probability of a given outcome, for example, the likelihood of an observation belonging to a certain disease category.

3.5.2 Doppler Ultrasound Regression

Regression analysis is an extremely popular approach for drawing inferences from datasets. The fundamental concept is to fit a function so that it approximates trends in a given dataset. The fitted function can then be used with 'unseen' data, to produce useful predictions [209]. Previous works concerned with analysing and processing Doppler ultrasound data using regression models are very limited. Several models have been described that combine Doppler ultrasound based features with other feature types; these will now be briefly discussed.

Ultrasound is often used to inspect and diagnose tumours. With respect to ovarian cancer, B-mode and Doppler ultrasound can be used by an experienced sonographer to differentiate between benign, and malignant tumours [210]. This discrimination is very important, it can allow optimised and individualised treatment [211], and furthermore, prevent unnecessary surgeries [71]. Despite experienced sonographers being adept at identifying malignant tumours, approximately 10% are too challenging to be diagnosed [210]. In cases such as these, and especially in more typical evaluations performed by less experienced sonographers, automatic means of classifying tumours would be advantageous [211].

Doppler ultrasound features have been shown to be useful when performing such a diagnosis, for example, waveform indices PI and RI [212] are generally lower in malignant tumours [213]. Using a large dataset from 14,317 examinations, it was shown that scans

with corresponding RI values below 0.4 corresponded to malignant tumours in 98% of cases [213]. Despite these findings, a large overlap of PI and RI was demonstrated for benign and malignant cases [214, 215].

Flow velocity features were later shown to play an important role in such diagnoses, and gave improved results with respect to PI and RI indices [216, 217]. Extracted features included systolic and diastolic velocity measurements, including peak systolic velocity and time average maximum velocity. Such features provide a basic indication of any potential dangers, however, give no probability information [218]. Through combining Doppler features, with physical features such as patients age and further features attained from the B-mode ultrasound such as tumour diameter, a larger feature set can be formed. A subset of features that give optimal results can then be selected.

This approach was taken to form a logistic regression model, from which probabilities with respect to a tumour being malignant were predicted [218]. This work found time average maximum velocities to be a strong feature in this regard. Predictions using this dataset were later implemented using an artificial neural network, which exhibited improved results [219]. Similar logistic regression models were generated which included menopausal status and levels of a circulating antigen CA 125 as input features. With respect to Doppler ultrasound features, colour score (acquired using a colour Doppler imaging modality, and which indicates vascularity and blood flow) was found to be a strong discriminator [71].

These models demonstrated good performances. However, their evaluations used retrospective testing, and cut-off values chosen with respect to the data. This means their performances were somewhat exaggerated [218]. These logistic regression models were later compared to one another, and with subjective evaluations from experienced sonographer (using B-mode and colour Doppler). They found that typically an experienced sonographer was able to confidently and accurately distinguish malignant tumours, whereas, regression models performed poorly [211]. Similar findings were reported for a more recent analysis [210]. The poor performances exhibited by the logistic regression models were attributed to subtle variations in examination method and the definitions used. Furthermore, the design and testing of these models used the same dataset and so performance was likely emphasized [211]. Given certain features are derived using third party software, variations with respect to blood velocity estimation methods (e.g., to extract the MFE) may attribute to reduced performance. The challenges inherent in this process are

discussed in Section 3.1.

These works illustrate the potential of generating useful feedback using Doppler ultrasound and regression models, however, research in this area is limited. The described regression and Doppler classification tasks share common, long-established features. Machine learning as a whole is a drastically evolving discipline, with new features types and modelling methods continually being developed. More modern machine learning approaches, such as deep learning techniques and advanced computer vision features, can be used to guide and progress research into machine learning solutions for Doppler ultrasound.

3.6 Modern Machine Learning Solutions and Features

Research into Doppler ultrasound and machine learning solutions is a relatively niche area. To broaden avenues of investigation, in terms of predictive models and feature types, works and successes in other disciplines can be explored. This section reviews state-of-the-art feature types and classification techniques. These methods are used for tasks that share similarities to Doppler ultrasound classification, for example, radar micro-Doppler analysis and speech recognition.

3.6.1 Spectral Analysis and Features

The Fourier transform is a classical method of performing spectral analysis. This technique, and other Fourier-related transforms, are extremely common in the study of time-series data. These variants, such as the STFT, are typically more efficient than alternative methods; lending themselves to real-time applications [220]. The STFT is utilised throughout this work, for example, when generating spectrograms for MFE estimation or when extracting audio features (Section 7.3.8).

These transforms are incredibly powerful, however, they do have associated limitations. In particular, the time-frequency resolution of these transforms are uniform across all frequencies [221]. This is because frequency resolution is inversely related to the length of processed data, and subsequently, to time resolution (i.e., improved frequency resolution is at the detriment of poorer time resolution).

3.6.1.1 Parametric Spectral Estimation

Modern methods of spectral estimation have been developed in response to the limitations of classical Fourier based techniques. The most common approaches use autoregressive (AR), and autoregressive moving average (ARMA) models [222]. These methods have been implemented for biomedical applications, for example, in spectral analysis of heart rate variability where AR analysis has been shown to be particularly useful [223].

An AR model predicts future time series values based on past points, the number of past values used is defined by the models order p. Equation 3.6 defines the AR model.

$$x_n = \sum_{i=1}^{p} a_i x(n–i) + \epsilon(n) \tag{3.6}$$

Where a_i are the AR coefficients, and ϵ (n) represents the prediction error [223]. The z-transform can be used to facilitate spectral analysis, and solve this equation. The corresponding transfer function, $H(z)$, can be expressed as,

$$H(z) = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{3.7}$$

Substituting $e^{j\omega T}$ for z, where ω is frequency and T is sampling period, the frequency components of the AR model can be expressed as,

$$P(\omega) = \frac{T\sigma^2}{|1 + \sum_{i=1}^{p} a_i e^{-ji\omega T}|^2} \tag{3.8}$$

Where σ^2 is the variance of driving white noise input [224]. Several methods have been described to estimate σ^2 and the coefficients a_i, allowing power spectral density (PSD) estimations to be obtained. These methods include the covariance method, Burg method, maximum likelihood and Yule-walker algorithms. The performance of these methods and the discrete Fourier transform (DFT) have been compared as a means of estimating ultrasonic backscatter coefficients [225]. The AR methods performed comparably and displayed improved performance with respect to the DFT.

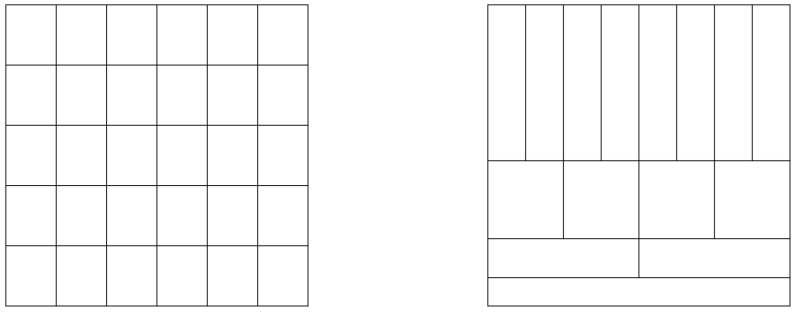
AR methods have demonstrated improved spectral resolution compared to the FFT approaches [226]. Furthermore, AR spectral estimation is more stable [227]. This is of particular significance for short-time spectral analysis, which is needed to analyse rapid

changes in blood flow measurements [222]. AR methods have been used in conjunction with Doppler ultrasound for different applications, including extraction of MFEs [128], spectral broadening indices [224] and as features for classification [194]. Through visual inspection of spectrograms, qualitative improvements have been reported using AR methods compared to FFT methods [224].

AR methods can be combined with a moving average process, to give a modelling method called ARMA [144]. ARMA models can provide better spectral estimations than either AR or moving average methods [128], and have been shown to give improved accuracy in terms of Doppler envelope estimation [128].

3.6.1.2 The Discrete Wavelet Transform

Different means of analysing spectral content have been developed; one example is the wavelet transform (WT). This technique was designed as an alternative to the STFT and is able to overcome the described frequency-time resolution restrictions [221]. The WT was pioneered in the areas of applied mathematics and signal processing in the early 1980's [228]. As with procedures such as the STFT, the WT can be considered as a projection of a signal into a set of basis functions [229]. However, in this case, wavelets, which are finite wave oscillations with zero-mean, are the basis functions. Through scaling and shifting the wavelet, the spectral content in a signal can be visualised.



(a) STFT (b) DWT

Figure 3.11: Difference in time-frequency resolution for STFT and DWT spectral analyses

Following this, the discrete wavelet transform (DWT) was formulated [230, 231]. The DWT is a special case of the WT, which is able to both provide compact representations

of a signal and be computationally efficient [221]. These attributes have inspired extensive research into different time-series studies [232].

The DWT allows analysis of temporal and spectral properties of non-stationary signals, and unlike Fourier analysis, produces a dynamic time-frequency resolution that is optimal in all frequency ranges [173]. This allows for high-frequency resolution at low frequencies, and high time resolution at high frequencies as illustrated in Figure 3.11. This process exhibits similar time-frequency resolution characteristics to the human auditory system [221], as a result, wavelet techniques have been used to perform speech enhancement [233].

An efficient and popular implementation of DWT is using a filter bank structure [229]. This approach is comparable to multi-rate filterbanks, whereby pairs of filters are used to iteratively decompose a given signal. This produces a hierarchy of new time series data [234]. At each level (L), the signal is passed through a lowpass filter and highpass filter [235] and downsampled by a factor of two. This produces detail coefficients (cD_L) and approximation coefficients (cA_L) [229]. In each iteration, the wavelet used is relatively larger or smaller. Longer wavelets are able to capture slower varying changes, whereas compressed wavelets can capture abrupt changes. The choice of mother wavelet and scaling function for this procedure are application dependent [236]. DWT signal decomposition is illustrated in Figure 3.12.

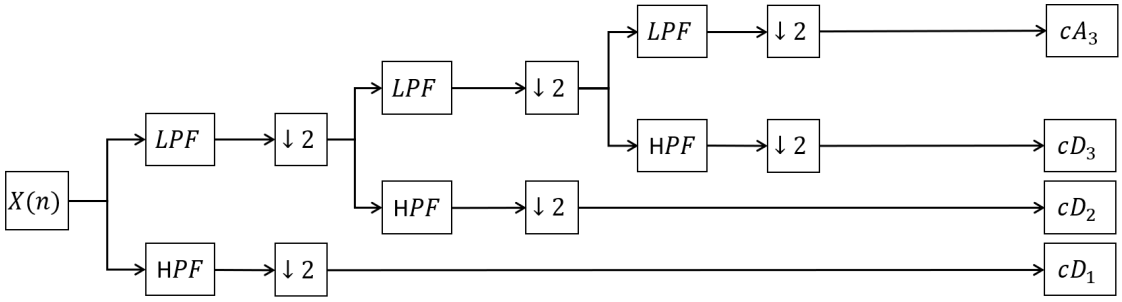


Figure 3.12: DWT decomposition

This approach has been used in many fields to study time series data, with coefficients being used for predictive tasks such as forecasting annual sunspot occurrence using neural networks [234], hydrologic series analysis [237] and classifying signatures from micro-Doppler radar signals [238].

Medical applications include detection of irregular heartbeats, medical image compres-

sion and distinguishing features in mammograms to facilitate breast cancer diagnosis [239]. Wavelet analysis has also been suggested as an appropriate means of diagnosing CHD from heart sounds [239, 240]. Auscultation is the process of listening to sounds of the body using a stethoscope; when placed on a patient's chest a clinician can diagnose heart conditions [240]. In cases of stenosis, characteristics of the sound such as pitch inform the user. Variations of such largely arises due to turbulent flows, which have been shown to be a fractal process. Wavelet-based fractal analysis through signal decomposition can be used to estimate the Hurst exponent, which describes the scaling process. This has been demonstrated as a means of detecting CHD [239].

DWT decomposition has also been used with Doppler heart sounds for disease diagnosis [172,173,195,196,198,204,206] these classification studies have been discussed in Section 3.5.1. The Daubechies wavelet has a fractal structure [206], and has a smoothing characteristic that makes it suitable for analysing variations in arterial Doppler signals [204]. The Daubechies wavelet is the most often used wavelet in cited studies. Daubechies wavelets of various orders have been implemented, however, a comparison across a selection of wavelets using classification accuracies of ophthalmic arterial and internal carotid atrial disease found a Daubechies wavelet of order one (db1) resulted in the highest accuracy [206]. The most commonly cited wavelet for DWT analysis of Doppler signals is db1.

Different means of dimensionality reduction have been implemented on extracted wavelet coefficients, including PCA [196]. However, the most common approach is to calculate statistics from the coefficients, such as absolute mean, maximum value, and STD metrics. An efficient means of signal compression for this application is Shannon entropy, which provides a measure of uncertainty of the wavelet coefficients [241]. This approach has been applied with Doppler heart sounds [173], as well as norm entropy measurements [172, 196, 198]. Shannon entropy is calculated using the following equation.

$$Hx = \sum_{i} P(x_i) log_2(P(x_i)) \tag{3.9}$$

Where $P(x_i)$ is the prior probability of all values of x. The number of decomposition levels varies across these studies, however information of interest is limited to above 40 Hz for Doppler heart sounds [204], which given the data used in these studies typically

corresponds to a maximum level of 7. In addition to these Doppler heart sound studies, Dyadic wavelets have been used to diagnose air embolisms [242].

3.6.1.3 Biologically Inspired Spectral Features

Speech recognition is a field in which enthusiastic development has led to impressive advances, with the technology becoming ubiquitous and installed across platforms as a standard feature [243]. Applications are widespread, examples include enabling voice control of mobile phones or audio whilst driving, or for customer communication to reduce costs [243]. Another important implementation is its use to help people with disabilities accomplishing daily tasks, such as using a computer or even a voice-activated wheelchairs [244].

The development of models to perform these tasks involves extracting useful information from the audio signal. Considerations regarding the non-linear way humans perceive sound has led to certain feature types being developed, and applied to tasks such as speech recognition. A highly successful approach is mel-frequency cepstral coefficients (MFCC) [245]. The use of these features has grown in popularity following their development in the late 1970's [246] .

The MFCC provides a robust and dynamic means of audio feature extraction. The MFCC is computed by first applying the Fourier transform to a windowed portion of the input signal, the resulting power spectrum is then mapped to the mel-scale. The relationship between the mel-frequency scale and true frequency scale are given in equation 3.10. The MFCCs are the spectrum amplitudes of the discrete cosine transform (DCT) of the log of these powers, this is illustrated in Figure 3.13 [247, 248]. This process gives a linear representation of the signal that reflects audio processing in the human brain [248].

$$mel(f) = 2595 \times log10\left(1 + \frac{f}{700}\right) \tag{3.10}$$

The success of these features have spurred researchers to use them in both related and unrelated fields. There are several examples of them being used to interpret emotion from speech [247, 249], an interesting problem which is highly intuitive to humans but challenging to replicate with automatic recognition models. Features are chosen to portray emotionally induced changes in speech, which arise in a predictable manner due to physiological states [247], and include energy, pitch and intensity. Similar applica-

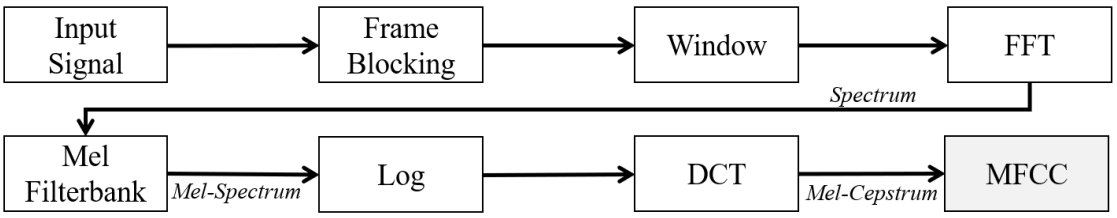


Figure 3.13: MFCC extraction process

tions of MFCCs and speech include biometric authentication [250], and early detection of Parkinson's disease [251].

MFCC features have also been used to analyse audio from music and different acoustic environments. The performance of various classification models were investigated for music, speech and voice analysis [221]. This study compared the accuracy of these models when trained using different feature types; those included were MFCC, DWT and FFT based features. This study found MFCC outperformed for each classification model. Its been shown that combining MFCCs with other feature types can improve classification performance. A cited limitation of MFCCs is degraded performances in poor SNR conditions, and in response to signals with flat spectrums [252].

Models for medical diagnoses have also been investigated using MFCCs. The stethoscope is evocative of clinicians, it can be used quickly and cheaply to examine different internal organs using passive sound; these sounds allow diagnoses of conditions such as heart murmurs. However, the success of these examinations are dictated by user training and aptitude, particularly when faced with infrequent, or rapid quiet sounds [253].

To address these limitations, models have been designed to automate sound diagnoses. A comparison between models using MFCC and STFT features found MFCC features gave best performance [253]. Classification of heart defects from heart sounds was similarly demonstrated using MFCCs and an SVM classifier [254]. As with examples discussed in Section 3.6.1.2, features from the diastolic and systolic portion of the cardiac cycle were considered separately.

Auscultation can also be used to diagnose disease states in the lungs. The ability to automate this using an ANN with combined MFCC and wavelet features was evaluated [255]. This study again found MFCCs produced the best results, interestingly this result extended to measurements even in poor SNR conditions.

A recent study compared the accuracy of SVMs, CNNs and Random forest models trained to classify heart sounds as normal or abnormal. Their accuracies were evaluated when trained using MFCCs, and features based on motifs. This study echoed previous works, finding MFCCs combined with an SVM classifier resulted in the best performance. MFCCs are clearly powerful tools for analysing heart sounds, however no examples were found of their application to Doppler heart sounds, or other Doppler ultrasound measurements.

These studies have all been applied to naturally occurring sounds, however, use of MFCCs are not limited to these cases. Information regarding the heart can also be collected using an ECG. SVM and ANN classifiers have been trained to diagnose cardiac abnormalities, such as arrhythmia, from ECG signals using MFCC features. [256, 257]. When compared to linear predictive coefficients (LPC) and linear predictive cepstral coefficients (LPCC), highest accuracies were achieved using MFCCs [257].

MFCCs have also been used to analyse radar generated micro-Doppler signals [248, 258, 259]. Micro-Doppler signals were discussed in Section 3.6.1.2, interestingly these surveys again demonstrate good performance of SVMs, with improved accuracies over other classifiers [258]. The accuracy achieved using LPC, LPCC and MFCC is again compared, with the latter again corresponding to the best performance.

In addition to MFCCs, another biologically inspired feature type that can be used for audio classification are gammatone cepstral coefficients (GTCC). Their application to speech recognition was first introduced as a means of overcoming noisy signals [260]. These can be extracted using similar steps to MFCCs, and can be considered as a biologically inspired modification which employs gammatone filters with equivalent rectangular bandwidth bands [261]. Gammontone filters are typically applied in the frequency domain, however, time-domain implementations have also been reported as a means of acquiring efficient and good performing features [262].

Evaluations of GTCCs for audio classification are less extensive than MFCCs, however, comparatively these works have all reported increased performances [260–265]. These improved performances can in part be attributed to a superior noise robustness of GTCC features [265]. The combination of both feature types, as opposed to either being implemented singularly, has also been demonstrated as advantageous [263]. Gammatone filters simulate the frequency response of the human auditory system [266], and take their name from the fact that a gamma distribution is used in the derivation of their impulse

response [261].

The methods described in this section provide a range of tools that can be used to thoroughly analyse the spectral content of signals. As described in Section 2.7.2, operators rely heavily upon visual clues contained within the Doppler spectra to perform examinations. To reflect this, further features can be derived from spectral images. As will be described in the following section, different tools are available for this, and these have been used to successfully classify audio signals.

3.6.2 Computer Vision

The field of computer vision is centred around enabling computers to see the world. Progress in this field has led to the technology becoming commonplace, with popular applications including facial recognition or object detection [267]. Descriptors are used to extract information from digital images, these can then be combined with algorithms for analysis or recognition tasks.

Different visual descriptors have been designed. These give information about fundamental components within images, such as shape or texture. Such features have been used in a breadth of fields and classification tasks, with applications from quality evaluation of strawberries [268] to classification of heart sounds [269].

3.6.2.1 Local Binary Patterns

Local binary patterns (LBP) is a popular visual descriptor for classification tasks. The descriptor was designed in response to challenges associated with real-world images, which include non-uniformity due to variable orientation, scale and illumination [270]. LBPs are now considered to be one of the best performing, and computationally efficient texture features available [271].

The process of extracting LBP features involves comparing pixels with their neighbours. This is done with groups of pixels using equation 3.11 [270]. Traditionally groups of nine are used (3 x 3).

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{3.11}$$

$$where,$$

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

Where P represents the sampling points, and R the radius used. The grey value of the centre pixel is given by g_c, and the grey value of the neighbouring pixels is given by g_p. In each application of this equation, the focus is on the central pixel. Circling this central pixel, in either direction, a one or zero is recorded if the neighbouring pixels is greater or smaller than the central pixel respectively. The gives an eight digit binary number, which can be converted to decimal. This procedure is applied to regions of the image, known as cells. In each instance, a histogram can be formed and used as a feature vector. Figure 3.14 illustrates this process using groups of 3 x 3 pixels.

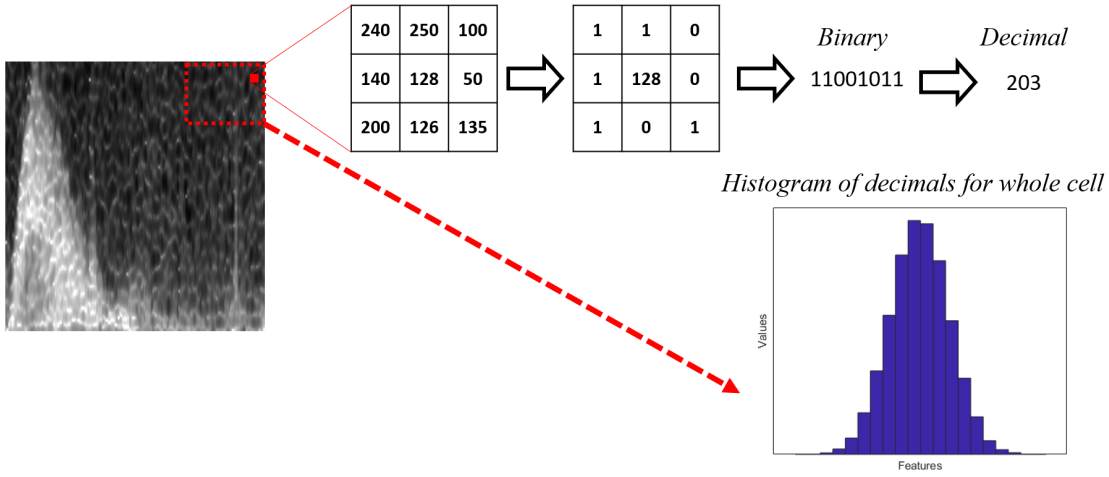


Figure 3.14: Local binary pattern feature extraction

There are many examples of LBP features being combined with SVMs and ANNs for classification tasks. These works include medical applications, an early use in this regard was for diagnosis of mammograms [272]. Newer implementations include diagnosis of coronary artery disease using greyscale echocardiograms [273], and more recently classifying

mitral regurgitation severity using colour Doppler images and SVMs [274]. This used TEE ultrasound images collected at specific times during the cardiac cycle. These applications are somewhat intuitive, as the images used depict physical structures, and are diagnosed visually by clinicians. However, medical applications where this is not the case have been explored.

Lung sounds can be used to diagnose various conditions, for example, pulmonary disorders [275]. Through transforming these sounds to the frequency-time domain, images and subsequently LBP features can be formed. This has been demonstrated using STFT spectrogram images [275] and mel-scale spectrum images [276]. Similarly, images of heart sounds, generated using the wavelet transform, have been used to classify sounds as belonging to specific stages of the cardiac cycle [269]. In this case, a radius of one was used to generate 59 features, which were utilized to train an SVM classifier.

Computer vision algorithms can be strengthened by combing feature types. There are many examples of LBP features being combined with other feature types to give improved performances. This includes the histogram of oriented gradients (HOG), which is a feature descriptor that exhibits a number of similarities to LBP features. The combination of LBP and HOG descriptors have demonstrated promising results for tasks such as human detection [277] and object localisation [278]

3.6.2.2 Histogram of Oriented Gradients

HOG features are a visual descriptor designed for object detection, as with LBP features they are extracted from greyscale images. The descriptors provide a means of expressing information in an image in a compressed form. Dalal and Triggs introduced the descriptor as a means of identifying human detection, and demonstrated improved performances with respect to existing feature sets for this task [279]. As the name suggests, descriptors are constructed using gradients within the image. Gradients describe changes in colour, Figure 3.15 illustrates a sudden change from dark to light, this is an example of a strong positive gradient. Gradients can be found both in the horizontal and vertical direction [280].

Blocks of pixels are analysed at a time. Within a block, the gradients for each pixel are calculated. These values are then used to calculate an associated magnitude and direction, producing a vector. This group of vectors can then be represented using a histogram. A bin width of 20° effectively results in nine features, which represent a compressed version

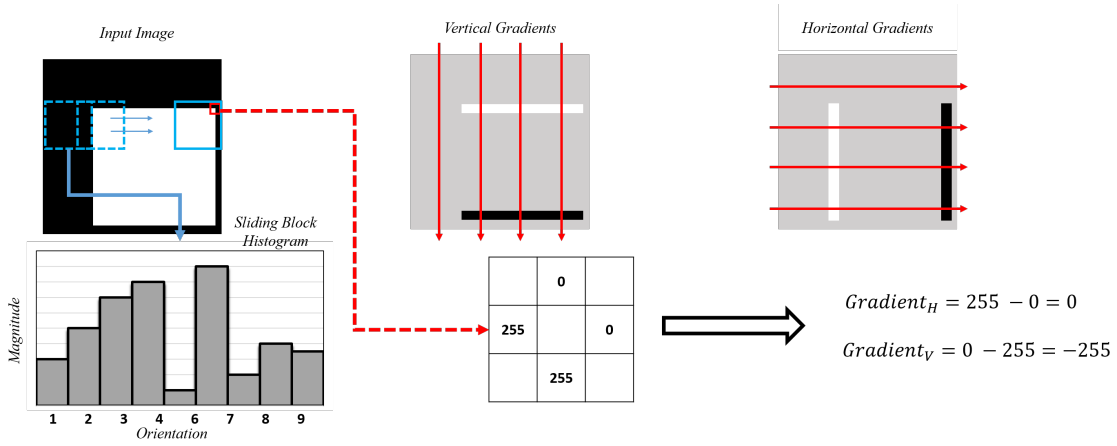


Figure 3.15: Steps involved in extracting HOG features, where Gradient_H denotes horizontal gradients, and Gradient_V denotes vertical gradients.

of that original block.

HOG features have been used for many applications, including widely researched areas such as facial recognition [281], and for more niche challenges such as bird species classification [282]. Several works have investigated using HOG features to perform audio scene classification [283, 284]. To achieve this, audio recordings are transformed into time-frequency representations, from which the descriptors can then be extracted. An example of HOG features extracted from Doppler audio using this approach is displayed in 3.16. This process can be combined with LBP features to classify sounds, this has been done to analyse snoring, which can be used to identify health risks such as apnea [285].

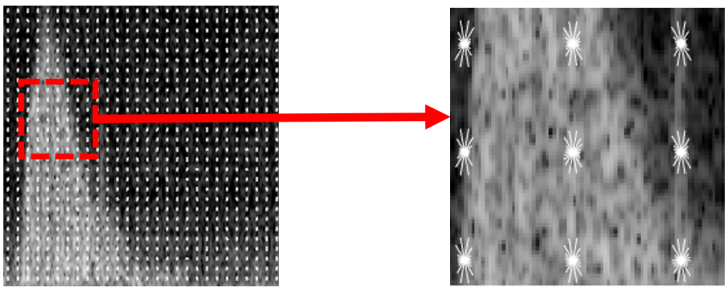


Figure 3.16: Example HOG features extracted from Doppler image

These computer vision descriptors are able to extract large quantities of information from images, however, state-of-the-art technologies are increasingly employing deep learning image recognition techniques. The reasons for this will now be discussed.

3.6.3 Deep Learning Techniques

Classification models are becoming ever more present in modern technology. They have been trained and evaluated for an enormous array of different tasks, and seemingly across all conceivable disciplines. A small sample of which has been used to guide techniques and features used in the presented work; these include familiar tasks like speech recognition [286], and those which are much obscure, such as monitoring frog communities [287].

There are typically two common challenges encountered by automatic classification systems. Firstly, the availability of labelled medical data, without which highly accurate models are impossible to form, and secondly the medical domain knowledge to ascertain which features should be extracted to represent classes [288]. Furthermore, as diagnostic classification models become more advanced and use increasing numbers of classes, similarity across these categories increase. This prevents features from occupying distinct regions of feature space [289]. Deep learning solutions can address these weaknesses.

Biomedical solutions are increasingly employing state of the art deep learning techniques for image classification [290]. This increase in usage follows a recent revival of deep convolution neural networks (DCNN) in the mid-2000's [291, 292], which has been described as the 'deep learning renaissance' and was fuelled by increased processing power and labelled data [293]. As discussed in Section 3.3, a variety of supervised learning approaches, such as SVMs and neural networks, have been used to aid classification of Doppler ultrasound. Neural networks are a subfield of machine learning, which has evolved to include deep learning methods. A deep neural network is a neural network with multiple intermediary layers, linking its input and output [294].

The ImageNet large scale visual recognition challenge is a competition in which the goal is to best predict content of images using designed algorithms [295]. In 2012, a DCNN called AlexNet [296] won this competition and demonstrated considerable improvement with respect to to previous winners; sparking substantial interest in DCNNs [297]. AlexNet is a large network, which was trained to classify 1.2 million images into 1000 different classes. The neural network has five convolution layers, a number of max-pooling

layers, three fully connected layers and a final 1000-way softmax [296]. This results in an impressive 60 million parameters and 650,000 neurons. The unprecedented success of this model, and its successors, is in part due to the vast dataset provided by ImageNet. The architecture of AlexNet is illustrated in Figure 3.17.

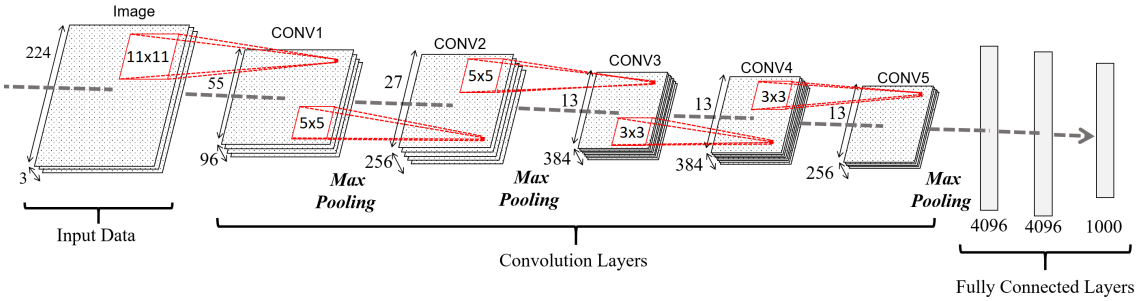


Figure 3.17: AlexNet arhitecture

Diagnostic classification models generally use hand-crafted feature extraction methods, using datasets which are a fraction of the size of that used by AlexNet [288]. A variety of such methods have been described in the previous sections. The relatively small size of these datasets limits the success of newly trained DCNN models.

However, by employing a training method called transfer learning, pre trained DCNNs – like AlexNet – can be repurposed for new classification tasks. This provides an efficient solution for new classification tasks when using limited data [298]. This process aims to transfer knowledge, which was garnered through solving an initial problem, to the new task. This can seem nonsensical in some applications, for example, retraining AlexNet, which was created using natural images (such as fauna and flora) and objects (such as instruments), for the new application of diagnosing grayscale radiographs [290]. All images, however, share common elements such as edges and blobs [290]. This enables DCNNs to be retrained and be used for medical applications.

This has been proven to be a successful approach and has been investigated for numerous tasks. The method has been shown to be a universal learning approach, with high generalisability and robust automatic extraction of useful features. This addresses the limitations of traditional methods discussed above, providing improved classification accuracies with reduced training sets, and negating the need to hand-craft features based on medical knowledge [288].

DCNNs have been used to diagnose many different conditions, as mentioned this includes radiographs. AlexNet and a second successful DCNN, GoogLeNet, were used to classify radiograph images from patients with and without tuberculosis. An ensemble of these methods achieved an area under the curve (AUC) value of 0.99 [290].

Recent examples of biomedical applications using AlexNet type models include diagnosis of lesions in ultrasound images [299], otitis media images (middle ear disease) [288], alcoholism from brain magnetic resonance images [300], and human recognition using images of ears [298]. Despite this increased interest, there is still limited examples of these deep learning techniques being applied to ultrasound images, current examples include classification of thyroid nodules [301], breast tumours [302], and of cirrhosis of the liver [303]. These papers did not include Doppler information, this was noted as an associated limitation [301]. A recent article, however, used Doppler ultrasound colour images to build models which help early recognition of tongue cancer [304]. The images were attained using power Doppler ultrasound.

There were no examples found of Doppler ultrasound spectrograms being classified via DCNNs. However, this approach has been used for classifying micro-Doppler signals [289]. Micro-Doppler signals describe return radar signals from a target undergoing micro-motion dynamics, such as vibration. These contain induced Doppler characteristics [305]. These signals are used in applications such as remote health monitoring, where radar events can be classified as different activities such as walking, falling or wheelchair use [306]; this can allow a person's health to be monitored outside of clinical settings.

The task of classifying such signal has a number of similarities to the work investigated in this project. As with medical Doppler ultrasound, features are extracted from the spectrogram, such as bandwidth or using envelopes [289]. Recent research exploring methods of classification found that implementing deep learning methods drastically improved classification accuracy by 17% (increasing from 72% to 89%). In this example, the comparison method used 127 features with an SVM classifier, whereas the deep learning approach used greyscale spectrogram images as inputs [306]. This work demonstrated that the DCNN approach was able to differentiate signatures within spectrograms that were indistinguishable to the human eye.

An investigation following this used transfer learning to retrain two popular DCNNs that were created using ImageNet [289]. This investigation demonstrated that transfer learning is a good solution, particularly when using limited training data. There are

various related works that have used AlexNet and other deep networks to classify micro-Doppler signals for remote monitoring [307–309], as well as for military applications such as classifying unmanned aerial vehicles [310].

Use of DCNNs has clear benefits, including negating the need to identify and successfully handcraft features capable of discriminating categories. DCNNs are able to differentiate classes with increasing similarity and achieve high classification rates with limited labelled data sets. Implementation of these approaches for classification tasks using medical Doppler ultrasound is currently a gap in the published literature.

The techniques and literature gaps highlighted through this section, and the preceding sections in this Chapter, will now be summarised.

3.7 Summary

This chapter reviewed different aspects of Doppler ultrasound related research. Methods used to analyse and process measurements were discussed, as well as works focused on automatically classifying Doppler ultrasound measurements. In addition to literature centred around Doppler ultrasound, important developments in other fields were detailed. This included a review of different machine learning techniques and feature types that could be extended to Doppler ultrasound research. The discussed research, the gaps within this literature, and the thesis contributions related to these gaps will now be summarised.

Different methods of MFE estimation, and their limitations, were reviewed. These methods predominantly functioned using either IPS or image-processing techniques. The evaluation of such methods have been restricted to data from high-quality measurements or basic simulation. Furthermore, image-processing techniques have not been quantitatively evaluated, or directly compared to IPS techniques. Chapter 5 proposes a new image-processing method of MFE estimation, which overcomes image related limitations of existing methods and is adaptive to different signal qualities. The evaluation of this method is performed using measurements that contain a wide range of signal quality and from poorly aimed measurements, providing an assessment more indicative of conditions encountered in practice. Furthermore, the novel method is compared to two IPS methods (MSNSI and MGM) and one image-processing method (ZIPM) and is shown to give more stable performance.

Documented methods of using extracted MFEs to perform beat segmentation were re-

viewed. Different characteristics restricted these methods, including reliance on additional hardware, user input, MFE shape assumptions, or a combination of these. The evaluation of methods were again restricted to small datasets of high-quality measurements and did not consider the MFE estimation method employed. Chapter 6 presents two novel methods of beat segmentation, that are fully automatic and require no additional hardware. They are evaluated using a large measurement dataset that contains a wide range of signal qualities. Chapter 5 demonstrates that the MFE estimation method can significantly impact the success of beat segmentation.

Previous works devoted to classifying Doppler ultrasound measurements were reviewed. These have operated using high-quality measurements averaged over many cardiac cycles. Machine learning tasks have been limited to classifications with limited categories (a maximum of three), and have considered only distinct conditions of disease. To reflect these gaps, Chapter 7 describes the design of a feedback model that quantifies signal quality in a continuous fashion, and Chapters 8 and 9 present classification models the predict probe position (using up to five classification categories) and reject poor quality data, respectively. These applications are novel, and resulting feedback models operate using a wide range of signal qualities and do not necessitate measurement averaging across many beats.

The previous Doppler specific classification works are not extensive, and therefore, consider only a limited range of different feature types. Various feature types previously un-documented for Doppler analysis, which are valuable in other fields, were described. These features are extracted in the given work and analysed in terms of their value for analysing Doppler ultrasound measurements; all extracted features are described in Chapter 7. In addition to these features, state-of-the-art machine learning techniques were described. These deep learning image recognition solutions have not been previously applied to Doppler ultrasound analysis, and research relating to their application for audio classification is limited. In response to this, Chapter 8 includes AlexNet classification models. The performance of these are contrasted with more traditional approaches that use hand-crafted features.

The design and evaluation of signal processing methods and feedback models proposed in this thesis are facilitated using data collected with the USCOM. The survey methodology for collecting this data, and the inherent challenges associated with building feedback models for Doppler ultrasound, are presented and discussed in the following chapter.

Chapter 4

Survey Data

4.1 Introduction

The primary goal of this thesis is to design automatic intelligent models that generate feedback reflective of a given measurement. Objectives to realise this goal include the design of methods for both MFE estimation and beat segmentation. Combined, these methods enable features and images to be systematically extracted from measurements, and subsequently, regression and classification feedback models to be designed and tested.

The design and evaluation of the MFE estimation method, beat segmentation method and feedback models described throughout this thesis are facilitated using measurements described in this chapter.

In this thesis, automatic feedback models are designed to quantitatively assess the quality of a measurement or to predict a beneficial probe movement in order to locate the correct position. This position is referred to as the ideal probe position (IPP). In order to differentiate between this position and 'off-angles' (i.e., incorrect positions), data were sampled from off-angles and the IPP. As data were collected from the suprasternal notch, the ultrasound probe is very limited in terms of translational movement. This enables off-positions to be defined in terms of angle direction with respect to the IPP, Figure 4.5 illustrates the positions sampled to reflect off-angles. These are later used to form classification models, which are described in Chapter 8.

This chapter details precisely what is meant by the IPP, and how an operator locates it. This gives a fundamental understanding of how USCOM scans are performed, and

crucially, how the models proposed in the following chapters complement the scanning procedure.

A significant challenge of developing feedback models is overcoming inter- and intrapatient scan variability. The acceptability of a scan is traditionally judged using characteristics of the Doppler profile, and can be facilitated using subjective scan scoring systems, such as the Fremantle criterion [17]. However, the variability in measurements can result in the Doppler profile, even when measured from the IPP, deviating far from its idealised form. In practice, this can mean an off-angle measurement from one patient can appear to be of higher quality than an IPP measurement from a second.

The collected data, which are summarised in Section 4.4, illustrate this variability and demonstrate how the ease of acquiring measurements can vary significantly from one patient to the next. These interpatient differences and scan acquisition challenges were discussed in Section 2.7.

The scan acquisition process and associated difficulties are presented in this chapter. This gives further context regarding the design and evaluation of data extraction methods and feedback models. In conclusion, this chapter presents the following information:

- the scanning procedure used by operators to locate the IPP;
- the data collection protocol;
- a summary of the collected data; and
- a discussion highlighting difficulties inherent to processing the data, including intra- and interpatient variability.

4.2 Scanning Procedure

This section details the correct probe position for collecting data from the suprasternal notch (i.e., the IPP), and details the scanning procedure to locate this position. This procedure is that taught by USCOM Ltd., and is used in this work to collect data; this data is summarised in Section 4.4. In addition, the Fremantle scoring system is described. This summarises Doppler profile characteristics commonly used to assess the acceptability of a scan, and to confirm it is from the IPP. As will be discussed, and illustrated in Section 4.4, application of such scoring systems across measurements can be problematic.

4.2.1 An Ideal Probe Position

This thesis is centred upon aortic valve blood flow measurements recorded using the USCOM device, and the design of feedback models that reflect the probe position with respect to the IPP. This section details what is meant by the IPP. These discussions are in reference to the USCOM device specifically, however, are transferable to similar devices.

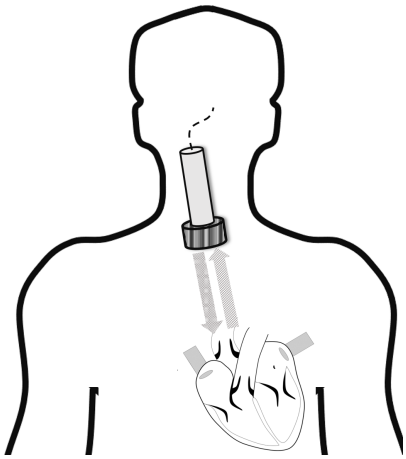


Figure 4.1: Illustrates the aortic valve being targeted from the suprasternal, when aligned this is refferred to as the IPP.

Doppler ultrasound measurements of the aortic valve are acquired using a transducer placed in the suprasternal notch (the dip at the base of the neck, between the two collarbones), the transducer is then manipulated so that the beam becomes aligned with the aortic valve. At this location, the insonation angle is minimised and the ultrasound beam becomes parallel to aortic blood flow [311]. As discussed in Section 2.7, USCOM includes no angle correction, and concerns have been raised regarding successfully aligning the ultrasound beam. Once correctly aligned, inaccuracy due to beam angle is at a minimum. This position is the IPP and is illustrated in Figure 4.1.

The transducer is carefully manipulated in three dimensions. The IPP is identified through the interpretation of visual and audio feedback provided by the device. This process will now be described in more detail. This gives context to the feedback models described in proceeding chapters, which are designed to complement this process.

4.2.2 Targeting the Aortic Valve

This section will describe how the IPP is located. This procedure is that taught by USCOM Ltd., and the described details are given with reference to training documents and training sessions provided by USCOM Ltd [312]. The scanning procedure used to locate the IPP requires manipulating the transducer so that scanning occurs through three planes: the sagittal, coronal and transverse planes. These are illustrated in Figure 4.2.

Probe movements are limited to one plane at a time and begin with the sagittal scan. The probe is slowly adjusted to scan through this arc, during which, the ultrasound beam will approach or traverse the aortic valve. As the beam nears the valve measured blood velocities increase. Once the beam passes the valve, blood velocities decrease. Using this concept, the probe is positioned so that the maximum velocity for the sagittal plane is being received. This procedure is then repeated for the remaining two planes.

It is recommended that during this process, the maximum velocity is monitored by watching the spectral feedback displayed on the USCOM, and not the automatically calculated values. This helps the user focus on the profile shape of each cardiac cycle. The shape of this profile is highly informative and is interpreted during the scan process.

The device uses a broad-beam CW ultrasound transducer, which permits unconstrained scanning. However, it also allows erroneous flows to be captured by the device, which present in the spectral feedback. These could arise, for example, from deeper vessels, as there is no range discrimination for CW ultrasound. Recognising these unwanted

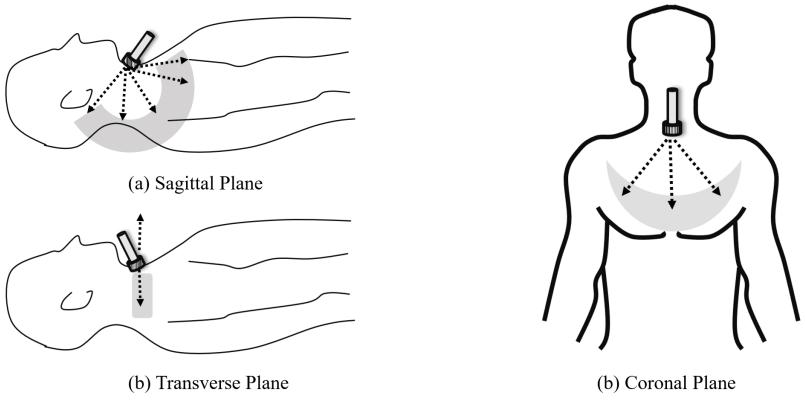
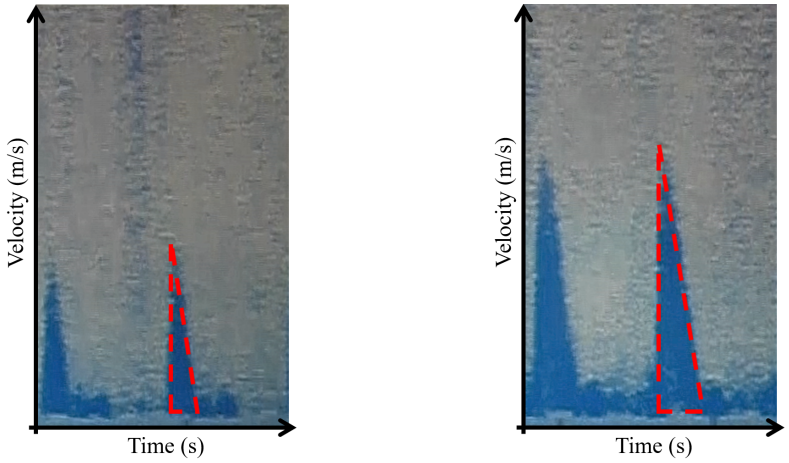


Figure 4.2: Three planes interrogated whilst locating the IPP scan

flows can aid users in identifying the correct signal.

The ascending aorta, for example, results in a profile similar to the aortic valve. The user therefore must be capable of recognising the difference between these two profiles. The ascending aorta has a systolic region that is narrower and shorter with respect to the aortic valve, which contains greater blood velocities [311]. These regions are triangular, an example of this is provided in Figure 4.3.



(a) Ascending Aorta. (b) Aortic Valve.

Figure 4.3: Example of ascending aorta and aortic valve profiles, as displayed on the USCOM, with triangular shapes highlighted using red dotted line.

Once the aortic valve is identified, the probe position can be subtly tweaked to hone in on the IPP. The ideal profile, as described by USCOM Ltd. [312], should have a triangular systolic portion with a full systolic width. The sides should begin at the base and be defined and continuous. The peak of the triangle should be clearly defined.

In practice, subtle transducer movements can cause a received reading to abruptly change. Furthermore, small variations in received measurements can be hard to distinguish; particularly for inexperienced users as reported in the literature (Section 2.7). Due to this, it can be hard for a practitioner to be confident that the IPP has been located. Researchers have implemented offline scoring systems, such as the Fremantle criteria, to validate analysed data.

4.2.3 The Fremantle Criteria

Different criteria have been reported for scoring the quality of scans obtained using the USCOM device, these have been developed by medical researchers [17,313]. These are used to score scans post examination and are based upon the subjective assessment of features present in the Doppler profile. The most commonly used scoring system is the Fremantle criteria [17]. This was presented as a means of assessing the acceptability of obtained measurements and was used to evaluate the proficiency of trainee users. In this example, scans were scored by a more experienced user, however, the criteria are often implemented by more novice users to confirm measurements are of adequate quality. There are potential limitations with respect to such scoring systems, these are discussed in reference to the collected data in Section 4.4.2.

The Fremantle criterion was designed to assess whether an acquired USCOM scan is acceptable or not. Using the criterion listed in Table 4.1, a scan score of up to six can be attributed to a still Doppler spectrum image [17].

Table 4.1: Fremantle criterion to assess USCOM scan image quality

Feature Description[1]	Score[1]
Well-defined image base[2]	1
Well-defined image peak[2]	1
Well-defined commencement of flow or heart sound[2]	1
Appropriate scale used on screen[3]	1
Minimal acoustic interference[4]	1
Total	6

1 – Applied to frozen USCOM scan image; these are 7.5 s long.
2 – Occurs in at least three cardiac cycles.
3 – To maximise frequency region of interest.
4 – Making discrimination between Doppler frequency content and background noise difficult.

Scoring systems such as this give some indication as to the quality of a given scan, however, they cannot indicate how far a given measurement is from the IPP, or whether a poor quality scan is due to the inherent haemodynamics being interrogated. Considering these limitations, it is possible such criterion could give false complacency that a given scan is adequate, or conversely, result in scans being disregarded despite them correctly

portraying the true blood flow. These limitations are explored further using the data collected in the presented work.

4.3 Survey Methodology

The methods used to collect data will now be described. Data included both recordings from the IPP and recordings from off-angles. This enables different feedback models to be trained and data extraction methods to be thoroughly evaluated (e.g., methods of MFE estimation).

The survey included 22 healthy adult volunteers. A minimum of two measurement sets were performed on each participant (a single measurement set is displayed in Table 4.2), and a variety of different data were collected. The proper ethical permission was attained from the School of Engineering Ethics Committee (Cardiff University) and signed consent was obtained from each volunteer. The survey information and consent form provided to participants are included in Appendix C and Appendix D, respectively.

Table 4.2: Example probe positions and recording times for a single measurement set

Probe Position	Recording Duration (s)	Label[1]
IPP	15	G
Patients Left	20	L
IPP	15	G
Patients Right	20	R
IPP	15	G
Patients Head	20	H
IPP	15	G
Patients Feet	20	F
IPP	15	G

1 – Labels used in classification tasks, where G denotes the IPP, L denotes patients left, R denotes patients right, H denotes patients head and F denotes patients feet. All off-angles combined are later referred to as O for practicality in classification algorithms.

Data were collected whilst the probe was held in a number of defined positions. The primary data type collected during this process was the positive Doppler audio signal (one continuous recording per measurement set). This was performed by sampling the

analogue Doppler audio signal using a laptop, limited streaming capabilities of the USCOM restricted this to positive flow. This was recorded using a 44.1 kHz sample rate, and 16-bit bit depth. In addition to this, video and ambient audio data were also recorded. These were included to further assist in labelling the continuous recordings, and as a reference to confirm probe angles and measurement timings. The hardware used is summarised in Figure 4.4.

The probe positions sampled during a measurement set are listed in Table 4.2. This illustrates that the off-angle positions were not arbitrary, rather they were defined relative to the IPP. These positions are illustrated in Figure 4.5. When sampling these positions, the operator kept both probe pressure and translational position constant.

As detailed by Table 4.2, the measurement protocol was: 1) Locate the IPP, 2) sample 15 s at IPP, 3) deviate from IPP, so that probe is angled more towards patients left, 4) sample 20 s at this position, 5) Locate the IPP, 6) sample 15 s at IPP, and so on until each of the five positions have been interrogated. The procedure outlined in Section 4.2.2 was used to locate the IPP.

The sampled off-angles allow feedback models to tackle simplistic classification scenarios, such as IPP or off-angle, and more informative tasks such as where the probe is with respect to the IPP; these classification tasks are explored in Chapter 8.

It can be seen from Table 4.2, prior to each off-angle position the IPP is first sampled. This ensures the off-angles remain relative to IPP. To reflect the more frequent recordings at the IPP, the relative recording times were marginally shorter. Acquiring data in this manner is a laborious task, to ease this process, recordings were performed in a continuous

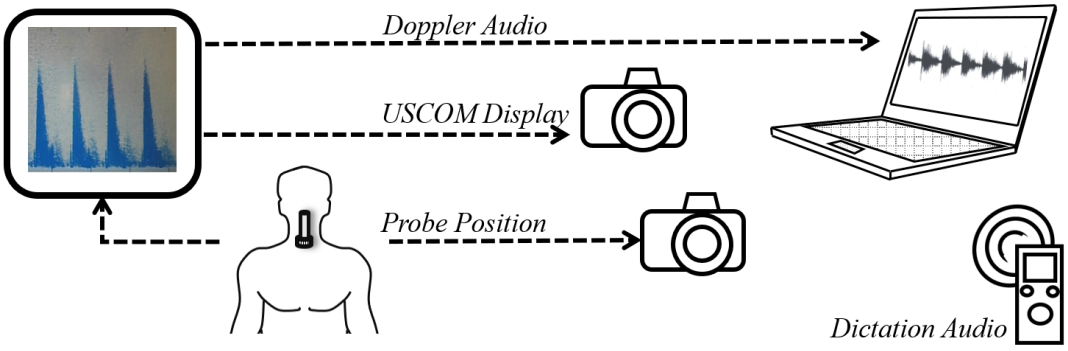
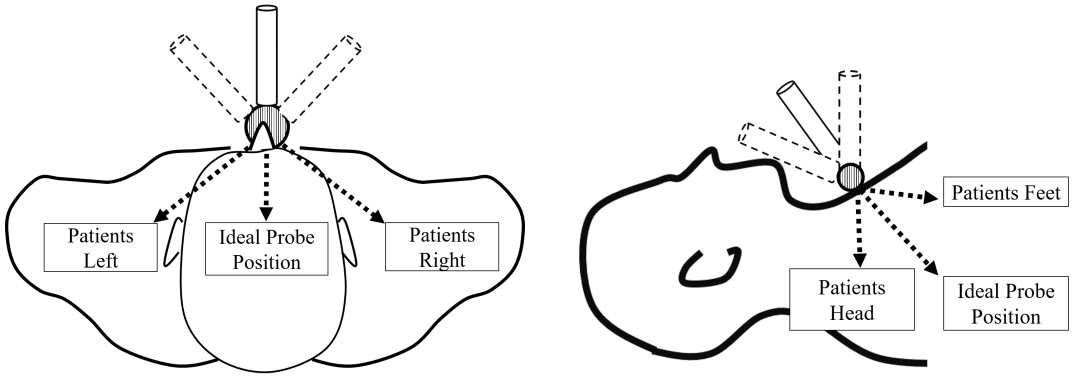


Figure 4.4: Hardware layout and data recorded during survey



(a) Left-Right probe angles

(b) Head-Foot probe angles

Figure 4.5: Off-angle probe positions with respect to IPP, as used for data collection

manner. This is instead of, for example, obtaining measurements from one location and then starting a new measurement, which would be significantly slower.

The video recordings of both the USCOM display and probe position enabled 'sense-checks' to be made with respect to the data (e.g., confirming probe position or the quality of recordings on the Doppler device). Finally, ambient audio was also recorded, providing an additional record of instructions given or dictated during the survey.

In total, data were collected from 22 participants using the outlined procedure. This data is summarised in the following section.

4.4 Survey Summary

This section presents a summary of participant information and the data collected from the outlined survey. The inter- and interpatient variability contained in these measurements are discussed. This gives context to the development and evaluation of feedback models in proceeding chapters, and highlights challenges involved with forming generalised models. The difficulty of obtaining measurements from individuals is also investigated. This is performed through comparing times taken to locate the IPP. This data is employed in Chapter 7 to investigate correlations between model performance and scan difficulty.

4.4.1 Participant and Measurement Summary

This section provides a summary of the participants details, and Doppler data collected during the survey. Data was collected by the author (from 12 different participants), and by an expert sonographer (from 10 different participants). Prior to the author's collection of data, steps were taken to ensure that they were sufficiently competent and had acquired the recommended level of experience. A hands-on training session given by a representative of USCOM was attended, as well as a training session with clinicians developing their skills using the device at Heath hospital (Cardiff, UK), and multiple one-on-one sessions with an expert user. In addition to these training sessions, given the portability of the USCOM device, scans were performed outside of lab conditions for additional experience. This resulted in experience of scanning over 25 different individuals, and many more independent scans (locating the IPP from scratch).

This constituted to more experience than the commonly cited requirement of 20 scans prior to being considered competent [17]. Furthermore, collected data was not restricted to any examination time limits, as can be the case in clinical practice or as imposed in some validation surveys [19]. This meant the examination process could be performed very methodically, ensuring the best scan possible was acquired. The nature of the survey methodology further strengthens this, as the IPP is re-located many times when performing multiple measurement sets (a single set is displayed in Table 4.2). This inherently provides a sense-check of data across the repeat measurements, and any large discrepancies in IPP scan quality can be identified and removed. The details regarding participants are provided in Table 4.3.

Table 4.3: Survey participant details

	Male	Female	Total
Number of Subjects	12	10	22
Age (years)[1]	34.6 ± 10	30.5 ± 11.2	32.7 ± 10.8
Height (cm)[1]	178.3 ± 7.2	166.5 ± 6.3	172.9 ± 9.0
Weight (kg)[1]	93.7 ± 21.7	64.3 ± 5.5	80.4 ± 22
Body Surface Area (m^2)[1]	2.1 ± 0.2	1.7 ± 0.1	1.9 ± 0.3

1 – Presented details are mean values with STD.

The primary data recorded during the survey was the positive flow Doppler audio signal. In total, almost 6 hours worth of data was collected. The parameters of these recordings are detailed in Table 4.4. This includes the corresponding duration for recordings obtained from the IPP, off-angles and whole recordings. This illustrates that the IPP and off-angle measurements exhibited very similar estimated SNR values.

Table 4.4: Details of audio data recorded from all participants

Parameter	**Probe Position**[1]		
	G	*O*	*All*
Position Independent Parameters			
Data Type	Positive Flow Doppler Audio		
Sample Frequency (kHz)	44.1		
Bit-depth	16		
Position Specific Measures			
Duration[2] (s)	2,580	6,222	20,705
SNR[3] (dB)	16.4 ± 5.8	16.2 ± 6	13.9 ± 7.2

1 – Where *G* denotes the IPP, *O* the off-angles, and *All* denotes the complete recording (i.e., *G*, *O* and searching periods).
2 – Total length of time sampled from each position.
3 – Approximated using equation 5.2, this technique and its limitations are discussed in Section 5.5.2.

The audio summarised in Table 4.4 is used to extract a swathe of different features, these are described in Chapter 7, and subsequently used to form feedback models. The extraction of these features necessitate timing indices associated with individual cardiac cycles. These are obtained by combining a novel MFE estimation method (described in Chapter 5) and beat segmentation methods (described in Chapter 6). The continuous recordings inevitably contain a large range of signal qualities, and erroneous contributions (e.g., noise as the probe is lifted, or flash artefacts due to sudden movements [12]), in this respect they reflect realistic conditions that software implemented in practice would encounter. Such portions can lead to poorly segmented data, which are not representative of true cardiac cycles. To ensure feedback models were not trained using such data, images of cardiac cycles (described in Section 8.4.3) were generated and used to visually identify observations corresponding to poorly segmented beats (referred to as fail beats). These

were then removed prior to training and evaluating models. The number of observations and associated fail beats for each probe position analysed is summarised in Table 4.5.

Table 4.5: Cardiac cycles captured

	G	*L*	*R*	*F*	*H*
Total[1]	3,047	1,781	1,828	1,828	1,882
Fail[2]	60	172	104	110	121
Fail Rate (%)	1.9	9.7	5.7	6.0	6.4
Resulting Dataset	2,868	1,609	1,724	1,718	1,761

1 – Total number of cardiac cycles segmented.
2 – Cardiac cycles removed from the extracted dataset due to incorrect beat segmentation.

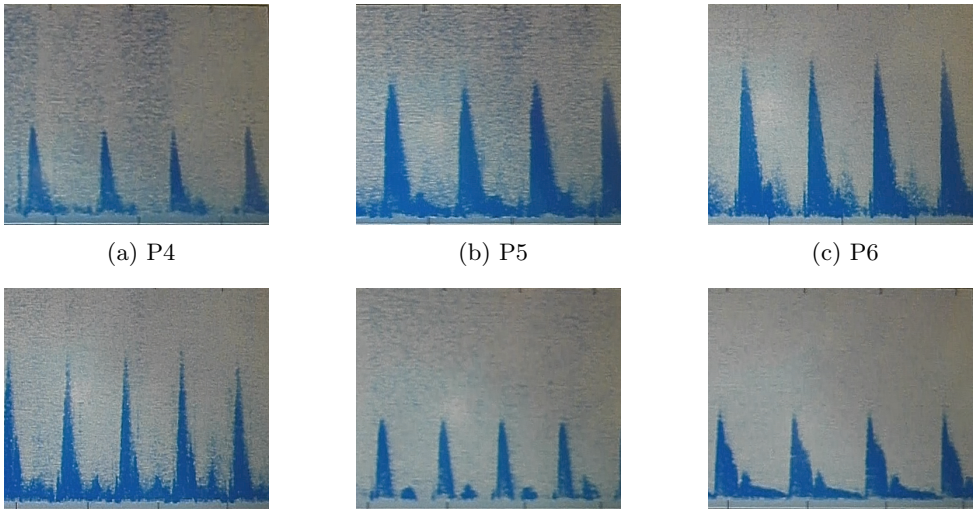
4.4.2 Participant Variability

The ease of performing Doppler ultrasound measurements varies across the population, and can be dictated by factors such as age and body composition [16]. In addition to this, the appearance of measurements can vary substantially across similar individuals, factors that influence this can include a person's health, for example, dehydration leads to reduced CO and increased systemic vascular resistance [314]. Considering the above, it follows that the appearance of Doppler spectra, even at the IPP, can vary significantly.

The interpatient variability present in the collected data (i.e., the variation from one patient to the next) is illustrated in Figure 4.6. This figure displays cropped images of the USCOM display, which is the primary feedback used by an operator. The cropped images were selected from different participants to illustrate the variation in Doppler spectrum at the IPP.

These demonstrate that the IPP Doppler profile can vary considerably across patients. These differences include systolic waveform shape, particularly during the cessation of flow. For example, Figure 4.6f exhibits a 'stepped' slope, which is in contrast to the sharp and more continuous slope belonging to Figure 4.6c. This sharp slope is a feature used to identify the IPP (Section 4.2.1) and is included within the Fremantle criteria.

The diastolic region of the waveforms also exhibit notable differences, for example, Figure 4.6a contains little signal during diastole, whereas Figure 4.6c and Figure 4.6d contain comparatively larger amounts.



(a) P4 (b) P5 (c) P6

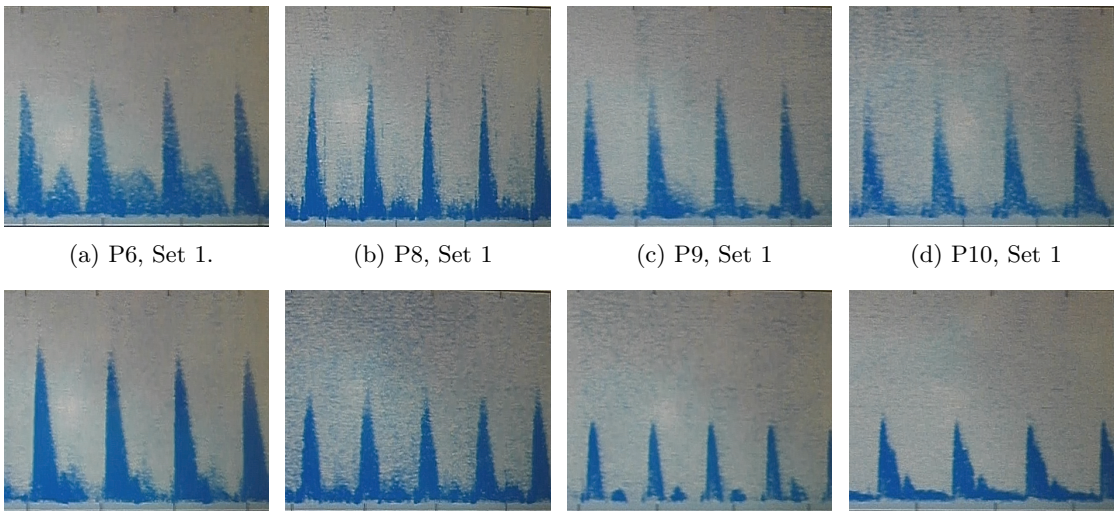
(d) P8 (e) P9 (f) P10

Figure 4.6: Interpatient variability captured by USCOM display at IPP, collected by an expert sonographer across different participants, where 'P' followed by a number indicates an individual participant.

Given the large differences that can present in people's physique and health, the presence of interpatient variation is fairly intuitive. However, it was found that substantial variation could also occur in repeat measurements of the IPP for individual participants. As with Figure 4.6, this was inspected using the captured USCOM display. This found that the greatest variation occurred across sets of measurements, although variation was also present within single measurement sets.

Changes in body position have an impact on measured haemodynamic variables, for example, leg raises have been shown to induce changes in CO, cardiac index and stroke volume [315]. Similarly, changes in heart rate alone can cause variations [316]. Despite all measurements being taken from participants in the supine position, subtle changes in body position or heart rate may account for some intrapatient variability [317]. Figure 4.7 illustrates the intrapatient variability associated with measurements from the IPP (i.e., the variation in measurements from individual patients).

Figure 4.7 illustrates the variation in IPP appearance for individual participants. It can be seen that during the first set of measurements the stepped flow cessation for participant



(a) P6, Set 1. (b) P8, Set 1 (c) P9, Set 1 (d) P10, Set 1

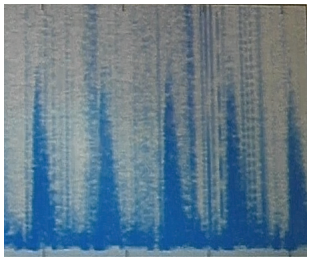
(e) P6, Set 2. (f) P8, Set 2 (g) P9, Set 2 (h) P10, Set 2

Figure 4.7: Intrapatient variability across measurement sets, captured by USCOM display at IPP and collected from participants by an expert sonographer.

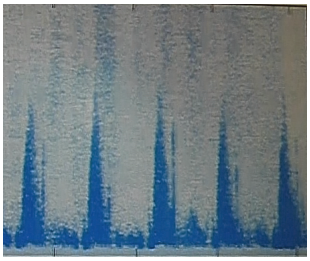
10 was less prevalent, however, overall, the signal was far less distinct. Furthermore, it can be seen that the diastolic signal belonging to patient 6 was far more defined during the first set of measurements. This signal contribution is likely erroneous (i.e., not associated with the target blood flow) and arose due to a nearby blood flow signal.

Another factor that can vary, in both an inter- and intrapatient manner, is the signal to noise level. This is illustrated by scans belonging to P8, which contain a higher level of background noise in Figure 4.7b when compared to Figure 4.7f. A potential cause of this is a discontinuity in the amount of ultrasonic gel used. Insufficient amounts of gel can result in a poorer coupling and increased noise, particularly in the presence of body hair which can trap bubbles of air that reflect and absorb ultrasound [318]. During a measurement scan, the applied gel can be spread out, and require reapplication. There are, however, a number of potential sources of noise [57].

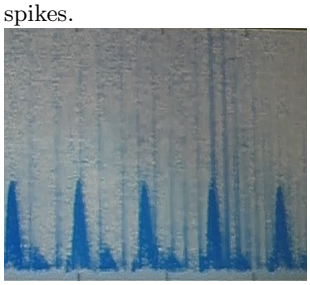
Background noise and erroneous signals can obscure the shape of the Doppler signal, and affect processing stages (e.g., envelope estimation). Erroneous signals can occur both when locating and measuring the IPP. As detailed in Table 4.5, this can lead to fail beats. With respect to real-time feedback models, methods that automatically recognise and



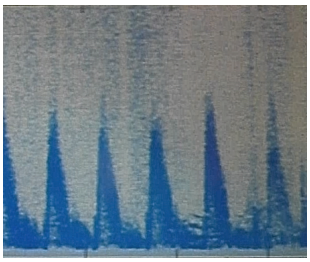
(a) Noise and spectral spikes.



(b) Heart valve clicks.



(c) Spectral spikes.



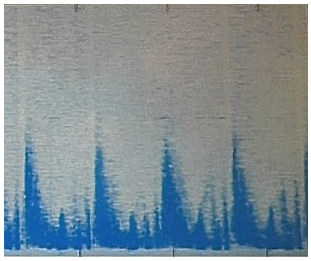
(d) Patient vocalisation between beats four and five.

Figure 4.8: Examples of erroneous signals encountered whilst recording the IPP

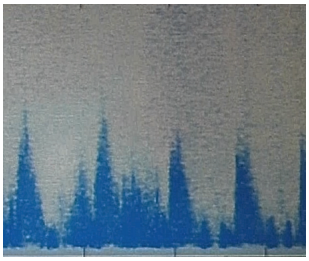
disregard poor data are desirable, Chapter 9 investigates this. The images presented in Figures 4.8 and 4.9 illustrate a variety of erroneous signals and features which can impede post-processing steps.

In Section 4.2.3, it was proposed that criteria used to confirm high-quality USCOM scans may not always be beneficial. Figures 4.6 and 4.7 include a number of IPP examples that would be capped due to certain aspects, such as peak definition in Figure 4.7d. To further illustrate the limitations of applying such criteria to confirm the acquisition of IPP measurements, Figure 4.10 presents data from off-angle and IPP positions. The differences in scan appearance can be very subtle, making aiming the USCOM challenging [19]. Figure 4.10 demonstrates this, and the difficulty of differentiating such positions using a fixed scoring system. The presented examples clearly show that an off-axis measurement from one participant can be scored more highly than an IPP measurement from a second. Furthermore, for individual patients, the score given to an off-axis position may not always be higher than that for the IPP position.

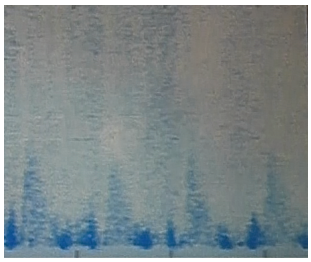
The figures presented in this section give some indication as to the variation in scan



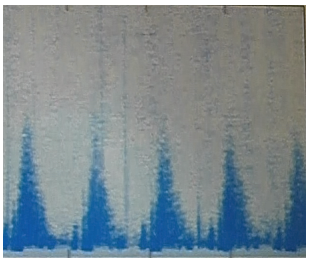
(a) Rhythmic diastolic noise.



(b) Diastolic noise.



(c) Faint signal.

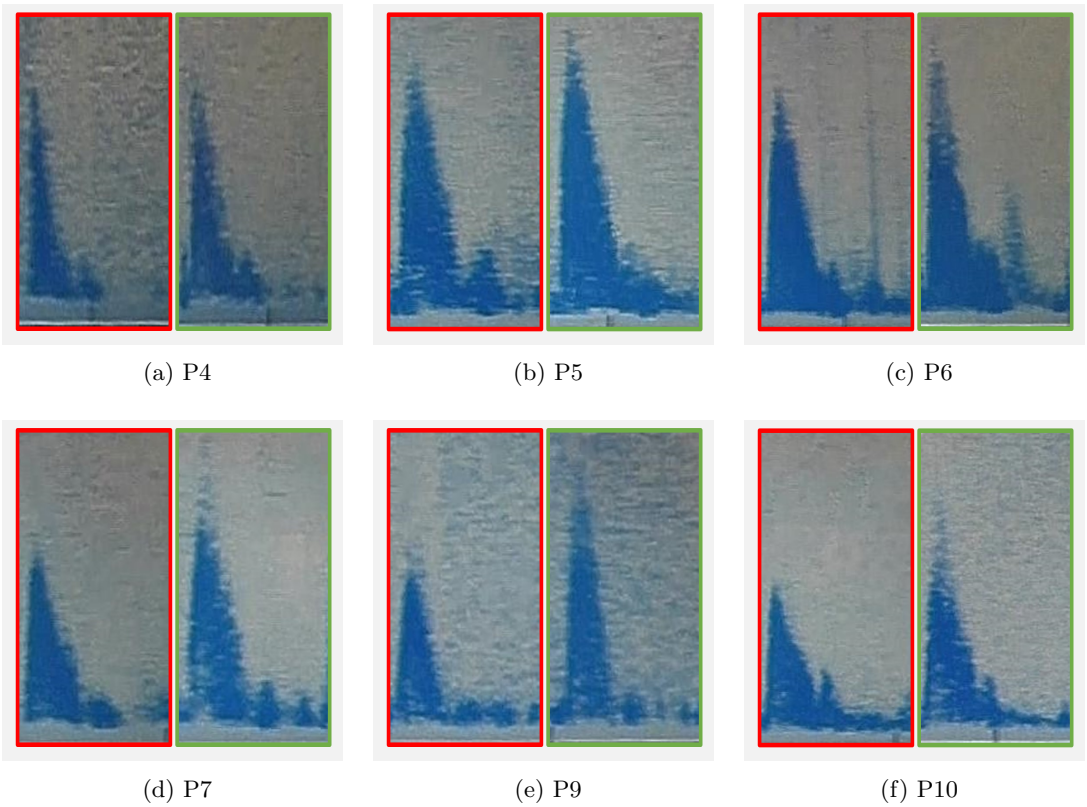


(d) Abnormal profile.

Figure 4.9: Examples of challenging signals encountered whilst locating the IPP

appearances that clinicians can be exposed to. In practice, the Doppler waveforms corresponding to the IPP may differ even further from an idealised waveform. This deviation could be particularly large in clinical measurements from unhealthy patients.

It is intuitive from these observations that sonographers will find some patients harder to obtain measurements from. Considering this, the difficulties of performing individual measurement provides an additional metric upon which trained models in proceeding chapters can be evaluated.



(a) P4 (b) P5 (c) P6

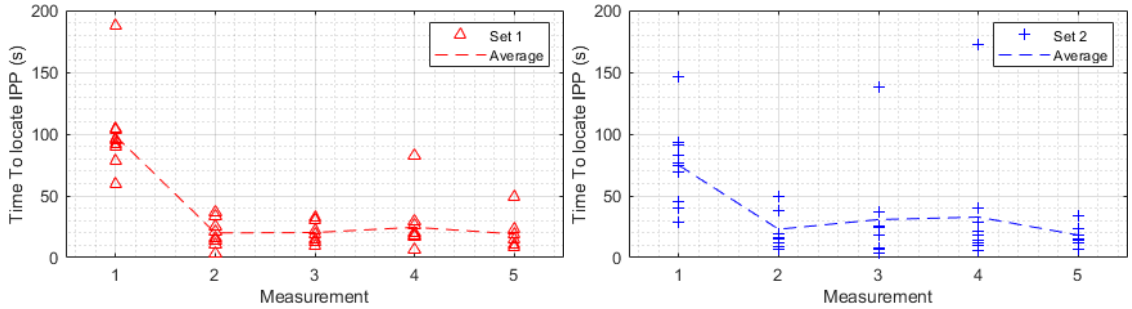
(d) P7 (e) P9 (f) P10

Figure 4.10: Beat examples from off-angles and IPP with similar quality, red borders indicate beats from off-angles, and green borders indicate beats from the ideal probe position.

4.4.3 Participant Difficulty

The ease of acquiring measurements from individual participants has been compared using the times taken to locate the IPP. This provides a quantitative value of scan difficulty for specific measurements and individuals. This provides an additional metric to assess the performance of feedback models. As detailed in Section 4.3, a minimum of two measurements sets were performed on each individual (of which each contains 5 samples of the IPP). The times taken to locate the ideal positions for these two measurement sets are presented in Figure 4.11. This analysis has been limited to the expert measurements, as author collected data acquisition was done in a very slow careful manner, to reflect more

limited experience.



(a) Set 1. (b) Set 2.

Figure 4.11: Times taken to locate IPP by expert sonographer, for all associated participants, and for measurement sets 1 and 2.

It can be seen from Figure 4.11 and Table 4.6, that the IPP is more rapidly located following the first measurement in each measurement set. Interestingly, the average time taken for the second measurement set (Figure 4.11b) is slightly higher for measurements two to five, despite the first measurement of this set being lower. The scanning process can cause slight discomfort for participants due to the pressure of the probe and the need to remain still, slight shifting in body positions during the second set of measurements may attribute to some of these increased scan times. If these measurements are treated as anomalies, the times taken for measurements two to five for both sets are very similar.

As discussed in Section 2.7, differences in a persons age and physique can lead to scan acquisition being more difficult. This could be, for example, due to the IPP being hard to locate or the associated Doppler profile shape being far from the idealised form. A variety of different profile features and artefacts that can occur in USCOM scans have been described [89]. These include characteristics that are not common to all measurements, such as spikes due to valves opening and closing, or faint diastolic signal contributions from ventricles filling. These variations can contribute to Doppler profiles differing significantly across patients, and the IPP being harder to locate and recognise. These aspects can also impede MFE estimation, for example, erroneous blood flows can mask true maximum frequency points and valve clicks can result in sharp peaks being incorrectly captured in the MFE. The following section summarises these difficulties, as well as the content in the preceding sections of this chapter.

Table 4.6: Average times taken for expert sonographer to locate IPP

Participant	Average Times Taken to Locate IPP (s)			
	Measurement 1 - 5[1]		*Measurement 2 - 5*	
	Set 1	Set 2	Set 1	Set 2
1	120	36	44	25
2	43	20	13	15
3	40	28	27	15
4	32	14	14	11
5	30	37	14	27
6	30	70	20	65
7	27	26	15	10
8	42	42	28	35
9	79	60	42	31
10	29	27	13	12

1 - Average times following the initial location of the IPP, to illustrates times taken once operator has become accustomed to the new participant.

4.5 Summary

This chapter detailed precisely how aortic blood flow measurements are acquired using the USCOM device. The correct probe position for this was described (i.e., the IPP), and the procedure performed by an operator to locate this position was outlined. A selection of profile characteristics were described, which can be used to subjectively assess an acquired scan and help identify the IPP.

The methodology used to collect data for this work was then presented. As described, this involved sampling five distinct probe positions. Using this methodology, data was collected from 22 participants. This data is used for work in proceeding chapters. This includes the design and evaluation of a novel MFE estimation method (Chapter 5), novel methods of beat segmentation (Chapter 6), feedback models and an analysis of valuable features for building these types of models (Chapters 7 and 8).

The collected data were summarised, and various intrinsic difficulties associated with subsequent processing were illustrated. These include intra and interpatient variation and artefacts such as valve clicks. These difficulties can impede MFE estimation, which is further exacerbated by low-quality scans. In the given work, data is analysed from dynamic probe movements and positions. This makes consistent extraction of stable MFEs

challenging. The following chapter presents a novel method of MFE estimation, designed to give stable envelopes across a wide range of signal qualities.

Chapter 5

Envelope Estimation

5.1 Introduction

Doppler ultrasound provides an accurate and noninvasive means of haemodynamic monitoring. These measurements contain a wealth of information and accommodate wide diagnostic capabilities [12]; a commonly extracted parameter is the maximum velocity envelope. This can be used to extract a host of different metrics, and can enable valuable clinical diagnosis, for example, the identification and assessment of stenosis [110], or determining the need for a carotid endarterectomy [110]. The maximum velocity envelope is proportional to the corresponding maximum frequencies in the Doppler signal (i.e., the MFE); this is detailed in Section 3.1.

Obtaining clinically useful information from an MFE necessitates a skilled operator. This is both for acquiring measurements, and for tracing or interpreting envelopes. However, time restraints of this manual process can render Doppler ultrasound monitoring impractical and prevent real-time applications. Furthermore, the process of obtaining measurements is subject to inter- and intraobserver variations [117]. These time restraints, as well as clinical benefits of averaging measurements [119–121] make automatic envelope estimation and beat segmentation methods highly desirable. In the context of this thesis, such methods are required to process large datasets, extract features and enable real-time feedback systems.

Automatic methods, for this purpose, must be capable of extracting stable envelopes over extended durations, despite periods of low SNR, variable envelope shapes, and erro-

neous signals (such as tissue movement). Additional external factors, such as acquisition errors and systematic quantification can further hinder MFE extraction [113]. The ability to do this is limited using existing methods of MFE estimation. This chapter presents a novel MFE estimation method, which has been designed to give stable waveforms for both beat segmentation and feature extraction.

The proposed MFE estimation method, the Otsu morphological method (OMM), is based on image-processing techniques. However, it uses Doppler spectrogram images generated directly from the Doppler audio signal, rather than images displayed on the ultrasound machine. This enables processed images to be explicitly defined within the OMM method, thus ensuring image consistency and removing uncertainty regarding their design. The novelty of the proposed method is twofold. First, its ability to define and vary Doppler spectrogram image parameters enable identification of effective threshold values, as explained in detail in Section 5.2.3. Second, OMM operates dynamically with respect to SNR by applying morphological operations in a hierarchical manner. As a result, the OMM method can extract highly stable envelopes from a wide range of signal qualities.

The accuracy of the proposed MFE estimation method is compared to three other methods: two IPS methods (MSNSI and MGM), and one image-processing method (ZIPM). These were discussed in Section 3.1. This is facilitated using in-vivo, simulated and phantom data. Through the addition of noise to these data sets, the MFE estimation methods are evaluated across a range of SNR values. The OMM method is shown to be the most stable in noisy measurements. Furthermore, when combined with an automatic beat segmentation method, MFEs estimated using OMM resulted in 8.2% more beats being correctly segmented than the next best performing method.

To conclude, the main contributions included in this chapter are:

- a method for extracting stable MFEs from low quality Doppler audio signals (Section 5.2); and
- a comparison between the performance of the proposed MFE estimation method with the MSNSI, MGM and ZIPM methods. Comparisons are performed using simulated pulsatile Doppler data, phantom data and over 2 hours of in-vivo data (Section 5.3).

5.2 Proposed MFE Estimation Method

This section describes the OMM method, which consists of three main parts: signal preparation, binary image generation, and hierarchical morphological filtering (Figure 5.1 illustrates this). The proposed method uses the Otsu algorithm [319] to generate binary images, followed by morphological filtering. OMM uses directional Doppler audio signals, Section 2.5 discussed how these signals are generated in an ultrasound device.

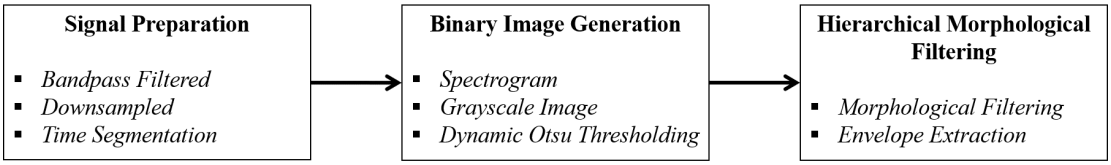


Figure 5.1: Main stages within OMM MFE estimation method

5.2.1 Signal Preparation

The Doppler audio signal is first processed to remove unwanted signals and make the following operations more time-efficient.

A high-pass filter (or 'wall-thump filter') is used to remove extrinsic low-frequency components arising, for example, from vessel walls [320]. A typical cut-off frequency of 200 Hz is used [321–323], which preserves the frequencies of interest. A low-pass filter is applied to remove high frequencies greater than those of interest. A cut-off frequency of 8 kHz is used for recorded data in the presented work. The filtered audio is finally downsampled to 16 kHz (allowing the full frequency range of the filtered audio to be analysed), which removes noise and speeds up the remaining operations. If replicated, the above procedures must be implemented with consideration of the velocities being measured and the hardware used.

5.2.2 Image Generation and Enhancement

The proposed method uses a spectrogram image generated using the Doppler audio signal. This image is also later used for feature extraction (described in Section 7.3), and is referred to as the OMM image. Generating this allows all associated parameter values to

be explicitly defined, giving full control over the characteristics of the processed images. Such an approach contrasts favourably with capturing an image from a machine, where the image generation process and parameter values used are unknown.

The spectrogram is calculated using the STFT of the recorded Doppler audio signal. A 10 ms Hamming window ($W = 10$ ms) is used, with 50% overlap. Using zero padding, a 512-point FFT is calculated. The chosen values ensure that changes in blood velocity are captured [12] and facilitate the application of fine morphological operations to the binary image in the following stage. This process produces image pixels with 5 ms and 31.5 Hz time and frequency resolution , respectively. Further processing of spectrogram images are affected by this time-frequency pixel resolution; to reflect this, this resolution should be replicated when implementing the proposed method.

The resulting matrix of STFT values is converted into decibels with a dynamic range of 60 dB, set with respect to the maximum value in the matrix. This wide range ensures the signal, which can change in intensity, is captured each time.

The matrix is converted into a grayscale image. To aid envelope estimation, high-frequency noise can now be removed using a Gaussian filter [324]. This has been implemented using a 5 x 5 Gaussian kernel in line with previous research [135, 325]. Whilst the previous works do not disclose pixel resolution, in this article a 5 x 5 Gaussian kernel is equivalent to 25 ms by 158 Hz. With respect to these previous works, a smaller STD of 1 has been used to preserve more rapid fluctuations in blood flow. Figure 5.3a and Figure 5.3b provides an example of the spectrogram and final image.

5.2.3 Dynamic Threshold Identification

In this stage, the greyscale image is converted to a binary image by applying a threshold. The purpose of this operation is to separate signal from noise, with the aim of setting all image pixels corresponding to noise to the value of zero and setting all image pixels corresponding to signal to the value of one.

A suitable threshold is identified using the Otsu method [319], which assumes a histogram with bimodal distribution (i.e., signal and noise), and calculates the value which best divides these distributions. This process identifies an optimal threshold using a sequential search, during which the success of each threshold is quantified using Otsu's objective criterion (η). The threshold with the maximum η value is selected. This maximum

η value is referred to as the effectiveness metric (EM).

However, the transition point between signal and noise can be masked in low SNR conditions. To detect this point effectively, a range of images is generated for a variety of window lengths (W). In each case, a threshold and corresponding EM value is calculated using the Otsu method. The values of EM indicate how well an image has been separated into two classes. The W that is most effective at separating the signal from the noise is determined by both SNR and Doppler profile.

Consequently, the best threshold is identified as that which corresponds to the largest EM value. This threshold is then applied to the image generated using the standard $W = 10$ ms. This gives a well thresholded binary image with the specific time and frequency resolution defined in the previous section. This binary image is cropped to remove frequency bins below 200 Hz in response to the high-pass filter described in Section 5.2.1.

In the presented work, 10 window lengths varying linearly from $W = 1$ ms to $W = 0.1$ s were used. This range was chosen empirically, as the best window was found to very rarely exceed it. EM is calculated using the following equation:

$$EM_i = \frac{max(\sigma_{Bi}^2)}{\sigma_{Ti}^2} \tag{5.1}$$

where i varies between 1 and 10 and corresponds to index of window length, σ_B^2 and σ_T^2 are the between-class and total variance within the image [319].

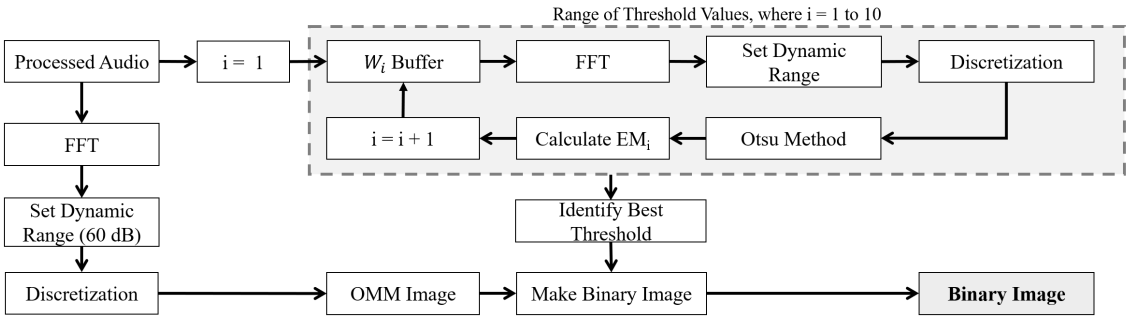
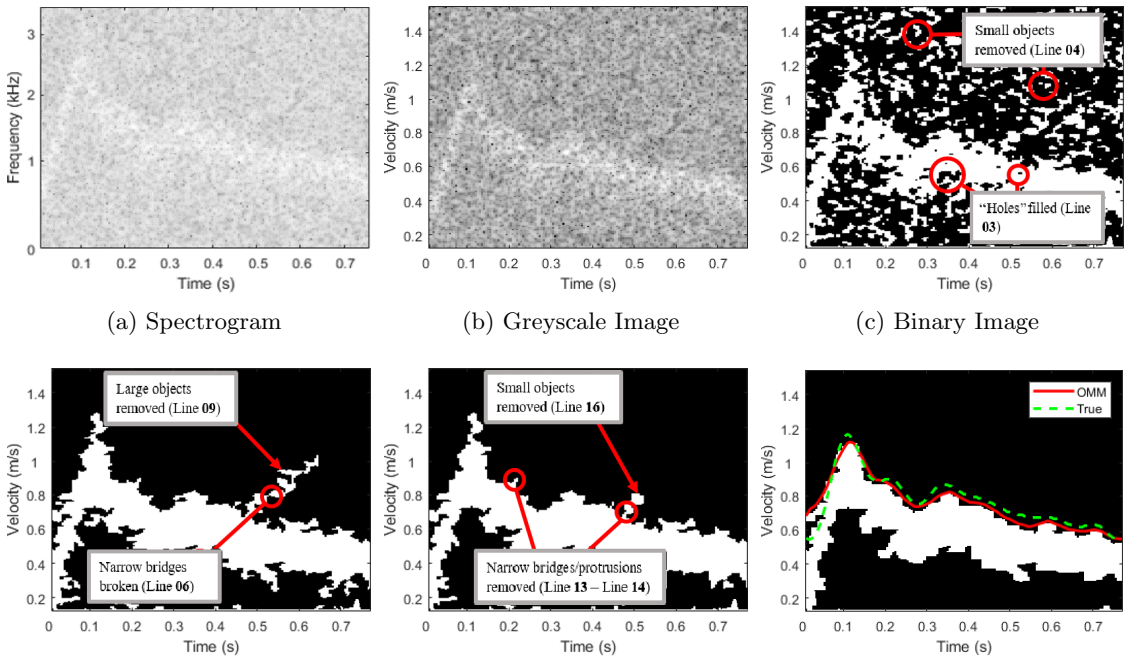


Figure 5.2: Dynamic threshold selection, using variable window lengths (W_i).

5.2.4 Hierarchical Morphological Filtering

The binary image can now be processed and used to estimate the MFE. An example of the binary image at this stage is displayed in Figure 5.3c. This example illustrates how in low SNR conditions, the threshold is unable to fully isolate the signal. Noise, which has been incorrectly identified as signal, will be referred to as noise. In higher SNR conditions, the threshold is better able to isolate the signal producing images more similar to those displayed in Figure 5.3d to 5.3f. To account for the variability in SNR, morphological operations are applied in a hierarchical manner. This prevents images with high SNRs being subject to unnecessary processing. This algorithm is illustrated using pseudocode in Figure 5.4.



(a) Spectrogram (b) Greyscale Image (c) Binary Image

(d) Level One Output (e) Level Two Output (f) Cleaned Output

Figure 5.3: Image stages within OMM, using a simulated common carotid Doppler signal with estimated SNR of -3 dB. Equivalent row velocity is displayed for images 5.3b to 5.3f. Extracted OMM envelope and reference true velocity is displayed in 5.3f. Pixels have a time and frequency resolution of 5 ms and 31 Hz, respectively.

```
[01]    Set First Row To 1
[02]    if # Objects > 1 % Level 1
[03]        Fill Image
[04]        Remove Islands > 300
[05]        if # Objects > 1 % Level 2
[06]            Open Image % vertical line structuring element, length 3
[07]            Set First Row To 1
[08]            Fill Image
[09]            Remove Islands < 500
[10]        end
[11]    end
[···]
                % Clean Image
[···]
[12]    Fill Image
[13]    Open Image % vertical line structuring element, length 5
[14]    Open Image % horizontal line structuring element, length 5
[15]    Set First Row To 1
[16]    Remove Islands < 100
```

Figure 5.4: Pseudocode of the proposed algorithm for hierarchical morphological filtering to extract MFE from binary image.

The binary image is passed to level 1 and possibly level 2 of the algorithm if more than one object is present in the image, i.e., the SNR of the image is low. Objects are defined as clusters of multiple pixels with non-zero values (white pixels) connected either vertically, horizontally or diagonally. Lower SNRs result in more objects and so corresponding binary images undergo further processing. Figure 5.4 illustrates that a number of operations are repeated. These will now be briefly discussed.

Initially, the number of objects is reduced to prevent unnecessary further processing. This is done by setting pixels contained in the lowest frequency row to 1. This is done on line **[01]** and repeated if necessary on lines **[07]** and **[15]**. This is effective at retaining small low-frequency objects, which otherwise would be removed in proceeding steps, for example, within the diastolic portion of Figure 5.6a. This condition, however, does assume that flow is present in the first frequency bin.

Next, the signal regions are strengthened using a flood-fill operation. Objects attributed to noise tend to be smaller, and less homogeneous than those reflective of signal. However, the signal portions can contain 'holes' (as illustrated in Figure 5.3c). This operation 'fills' these holes, and is performed on lines **[03]** and if necessary on line **[08]** and on line **[12]**.

After strengthening the signal, any object with an area smaller than 300 pixels is considered to be noise and is removed in level 1 (line **[04]**), as illustrated in Figure 5.3d. If the image is passed to level 2, the operation is repeated on line **[09]**. Remaining noise at this stage is often contained in larger objects due to the lower SNR of the images reaching this level, and so an area of 500 is used in level 2 (comparison between Figure 5.3d and Figure 5.3e illustrates this). The final area used in the cleaning stage is 100 on line **[16]**.

This assumes little noise and removes small isolated objects, which appeared as a result of opening operations. The choices of these area sizes were guided by previous works and determined through empirical investigation, previous works have used clusters ranging from 50 [137] to 500 [132], however, the pixel resolutions in these cases are not explicitly defined.

Additionally, in level 1 and in the cleaning stage, opening operations are used to break small horizontal and vertical connections respectively. This is done prior to the removal of objects, on lines **[13]**, **[14]** and **[06]**. This isolates weakly connected objects or small protrusions (as illustrated in Figure 5.3e), reduces noise and smoothes the image prior to envelope estimation.

The resulting image can now be used to extract the MFE. The MFE is found for each time point separately, using the column of pixels associated with that time point. The maximum frequency within each column is taken as the white pixel (pixel representative of signal) which is next to the largest number of consecutive black pixels. This is similar to the biggest-gap method, which includes weightings applied to each group of noise pixels based on frequency [135]. The envelope is then smoothed using a 10-point moving average filter. The unsmoothed envelope is also retained for further steps described in Section 6.2. The extracted OMM envelope and reference true velocity is displayed in Figure 5.3f.

In cases where both forward and reverse flow is of interest, the steps described to extract the MFE are repeated for the positive and negative Doppler shifts respectively. This gives an MFE for forward and reverse flow, and an overall MFE can then be found by taking the absolute maximum of each MFE at each time point.

5.3 Evaluation Methods

The performance of the OMM method has been quantitatively assessed using simulation data, phantom data and in-vivo data. The performance of OMM is compared to that of three other MFE estimation methods. The MFE methods chosen to provide comparison are two IPS methods and one image-processing method.

The IPS methods chosen to offer comparison are the MSNSI and MGM methods. MSNSI has been selected as it is focused on envelope estimation as opposed to maximum frequency estimation at specific time points and was shown to perform well [114]. MGM is an older IPS method [127], which has been shown to be reasonably stable in varying SNR.

This is used to provide a further comparison with IPS methods. The image-processing method chosen to offer comparison is ZIPM [132]. ZIPM was selected due to it being a recent method, which demonstrated good correlation with expert tracings. Furthermore, ZIPM is designed for fully automatic tracing (as OMM is), and specifically for aortic Doppler measurements similar to those used in this thesis.

The performance of IPS methods has typically been validated using a combination of simulated data, phantom data, and in-vivo data. Popular simulation approaches model simple constant flow conditions using Gaussian processes to represent Doppler signals [143–145]. For this study, pulsatile flow has been simulated. In the case of phantom and simulated data, the true maximum velocity is known. This allows MFE estimation methods to be quantitatively assessed using statistical measurements; bias and STD are commonly used [113,114,127,128,326,327] and have been implemented in this study. Correlation statistics have also been calculated for these data sets; they provide a measure of the similarity between the extracted MFEs and the true envelope shape. This is included as good correlation is essential for reliable beat segmentation and provides further evaluation with regards to the stability of extracted envelopes.

Image-processing methods have been previously assessed using in-vivo data and thus have not included evaluation with respect to different levels of SNR. Instead, such assessments are based on comparisons made with expertly traced envelopes. Here, we assess an image-processing method using simulated and phantom images; this allows for a quantitative evaluation. The addition of Gaussian noise to signals enable different SNR values to be investigated. This technique has been used with the phantom and simulated datasets, allowing the performance of each MFE estimation method to be assessed with respect to diminishing signal quality.

In-vivo data has been used in the current study to investigate the ability of each MFE estimation method to produce MFEs suitable for beat segmentation. MFEs are extracted and processed using an automatic beat segmentation method that functions solely using the MFE, this is described in Section 6.2. The percentage of beats correctly segmented for each of these traces is then analysed. This data has been processed in 4-second segments.

All processing has been performed using MATLAB R2018a (The Mathworks, Inc., Natick, MA, USA).

5.3.1 OMM Method Implementation

The OMM method has been implemented as described in Section 5.2. Threshold and associated EMs have been calculated using the image-processing toolbox in Matlab.

With respect to simulated data, the positive and negative flow envelopes are calculated separately. These two envelopes are then used to generate the final MFE. At each time point, the positive and negative maximum frequencies are compared, and the maximum absolute frequency is used for the final MFE.

5.3.2 IPS Method Implementation

The performance of the MGM [127] and MSNSI [114] methods has been assessed and compared to that of the OMM method. These have been implemented as described in their publications. Their basic function has been summarised in Section 3.1.2

5.3.3 ZIPM Implementation

The ZIPM method uses images acquired from an ultrasound machine by a frame grabber. In the presented study, the images used were generated using an approach similar to that for the OMM image, as described in Section 5.2. The method differed from OMM by using a dynamic range of 20 dB for the phantom and simulation data, and 40 dB for the in-vivo data.

It was found that varying the dynamic range in this way was necessary to achieve acceptable results across the datasets. These dynamic ranges were chosen by testing values from 15 to 60 dB, and then selecting the best value in terms of STD and bias for the simulation and phantom data (with no added noise), and visually inspecting envelopes produced for in-vivo data. The images were then resized to be more representative of those described in the article [132].

In cases where both positive and negative flow is of interest, the approach implemented by OMM is used (Section 5.3.1). An example of an image generated using this approach is displayed in Figure 5.5.

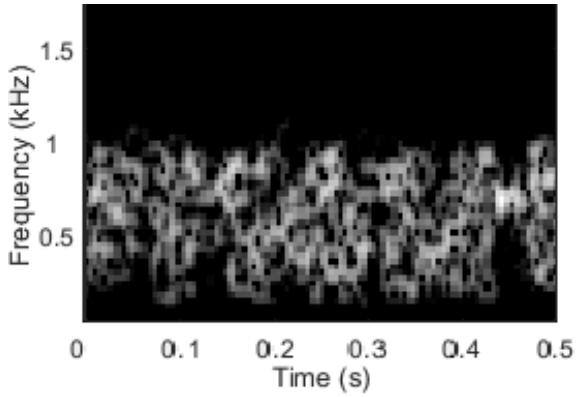


Figure 5.5: Image with dynamic range of 20 dB, generated for ZIPM implementation using Phantom data.

5.3.4 Simulation Data

The performance of MFE estimation methods has previously been investigated using simulated data [114, 127, 128, 143, 327] and [145]. The advantage of using simulations is that the true maximum frequency associated with the modelled scatterers is known. These models have predominantly simulated simple flow conditions, representing steady flow. Such models allow the process to be simplified; however, they limit how realistic resulting data is, for example, they do not take into account ultrasound device parameters such as sample volume and do not represent realistic pulsatile blood flows. The simulated data used within this study represents pulsatile flow and is generated using Field II software, which performs numerical ultrasound simulation and is described in Section 3.1.6.

Field II produces raw data similar to that measured by an ultrasound machine, allowing MFE estimation performance to be rigorously tested on very realistic data. The software was used to simulate pulsed-wave ultrasound interrogating pulsatile flow using insonation angles of 30° and 60°. In pulsatile flow, the velocity profile changes as a function of time. To reflect this, the Womersley model [147] is used to generate realistic flow profiles.

The contributions from all scatterers traversing the sample volume allow the Doppler spectrum to be formed. A range-gate was used to simulate a sample volume positioned at the centre of the lumen and spans its width. An online example was used as a reference to generate a model using a linear array transducer [328]. The settings used to generate the model are displayed in Table 5.1. These were used to generate simulated data for

one complete cardiac cycle, for both femoral and common carotid artery blood flow. An example of simulated data for the femoral artery is presented in Figure 5.6a.

Bias and STD statistics are calculated for the simulated data. These statistics are used to contrast the performance of the tested MFE estimation methods and are calculated using the estimation error at each time point within the MFE [114]. The correlation coefficient of an estimated envelope and the true velocity envelope can also be calculated, providing a numerical measure of the similarity between the two waveforms. These statistics are calculated for the simulated femoral artery data.

In addition to the above performance metrics, waveform indices can be calculated from extracted envelopes and compared to their true values. These indices use minimum and end-diastolic velocity values. Due to the end-diastolic value for the femoral artery being very close to 0 and some MFE methods tending towards 0 in poor noise conditions, simulated data of the carotid artery is used to compare the estimation of these indices.

Table 5.1: Parameters used within Field II to simulate flow

Parameter	Value	Parameter	Value
Scatterer Settings			
Heart Rate	90 bpm[1]	Peak Velocity	1 m/s[1]
	80 bpm[2]		1.2 m/s[2]
Lumen Radius	4 mm	X-range	40 mm
Lumen Centre Depth	40 mm	Y-range	8.8 mm
Number of Scatterers	67,851	Z-range	8.8 mm
Insonation Angle	30°, 60°		
Linear Array Transducer Settings			
Pulse Repetition Frequency	8 kHz	Kerf	0.05 mm
Centre Frequency	2 MHz	Element Pitch	0.44 mm
Sampling Frequency	100 MHz	Speed of Sound	1540 m/s
Cycles in Emitted Pulse	10	Excitation Pulse	Sinusoid
Element Elevation Height	5 mm	No. Elements	64
Element Lateral Width	0.39 mm	Focus vector	[0 0 40] mm

1 – Settings for femoral artery.
2 – Settings for carotid artery.

5.3.5 Phantom Data

Phantom data was collected using a Gammex optimiser 1425A (Gammex Inc., USA). This is a self-contained system fluid phantom, which is capable of generating steady laminar flow rates from 1.7 to 12.5 ml/s. Such systems are described in Section 3.1.5

The 1425A is designed for testing aspects of ultrasound device performance, including the accuracy of measured flow rates. The system uses structures which are ultrasonically similar to human tissue ensuring a realistic platform for research. The embedded vessel can be scanned using an insonation angle of 50°, and has an inner diameter of 4 mm. Data was collected from the phantom using a Toshiba TUS-A500 diagnostic ultrasound system. A 3 MHz probe was used to measure steady flow across the embedded vessel through pulsed-wave ultrasound. An audio output on the machine was employed to record the directional Doppler audio on a laptop, using 44.1 kHz sample rate, and 16-bit depth.

Typical peak blood velocities within the ascending aorta are of the order of 0.7 m/s, but varies between patients [329]. To reflect typical values, data was measured using flowrates of 0.4, 0.8 and 1 m/s; 10 seconds of data was recorded for each flow rate. As with the simulation data, bias and STD statistics were calculated; these were calculated using 1 s segments of data and then averaged. An example of data measured using the Phantom is displayed in Figure 5.6b.

5.3.6 In-Vivo Data

The MFE estimation methods were further evaluated using in-vivo data, an example of in-vivo data is displayed in Figure 5.9. Data was collected by the author from 11 healthy adult volunteers using the USCOM device. This data and the device are described in Sections 2.6.2 and 4, respectively.

The data consists of 229 recordings, totalling over 2 hours of audio. Using equation 5.2, the in-vivo SNR ranges from approximately 10 dB to 30 dB. The true MFEs under these conditions are unknown, meaning performance cannot be investigated using STD, bias or correlation statistics. Instead, in-vivo data has been used to evaluate how well each MFE method produces envelopes suitable for accurate beat segmentation. This is of particular interest for the given work, as cardiac timing indices are required to extract features.

This analysis is done by comparing the percentage of total beats segmented, and the

number of corresponding true positives, false positives and false negatives. These data contain scans exhibiting a range of quality allowing a more realistic investigation of performance with respect to real-world measurements.

The method of beat segmentation used for this is described within Chapter 6, and is referred to as the slope-gradient method. This chapter describes two beat segmentation methods, however, the second method requires additional inputs other than the MFE, namely an example cardiac cycle MFE and the OMM image. The slope-gradient method is based upon the method used in the USCOM device [78], and requires only the MFE to function. For these reasons, it is used within this chapter to compare the MFE estimation methods. This method identifies the starting positions of cardiac cycles by first identifying the peak systole locations. The gradient of the MFE prior to these peaks are used to estimate the starting points.

5.3.7 Evaluation of Performance in Varying SNR

The performance of estimated envelopes has been assessed in response to varying SNR. This has been achieved by adding noise to the simulated and phantom datasets. SNR is estimated from the spectrogram using the following relationship [114]:

$$SNR(dB) = 10log_{10}\left(\frac{\hat{P}_S - \hat{P}_N}{\hat{P}_N}\right) \tag{5.2}$$

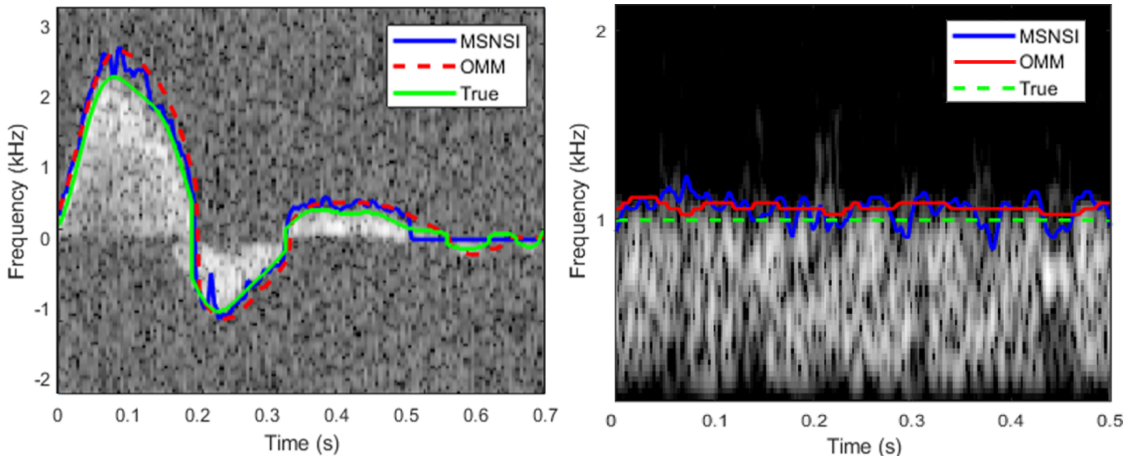
where $\hat{P}_s$ is the mean power contained in the entire spectrogram, and $\hat{P}_N$ is the mean power of a region of the spectrogram which contains only noise. This region is identified as a range of bins which exceed the estimated MFE [114]. In the case of phantom and simulated data the true MFE is known, this enables all bins reflective of noise to be used when calculating $\hat{P}_N$.

5.4 Results

The performance of the four MFE estimation methods has been systematically evaluated using the datasets described in Section 5.3. The results from this analysis are now presented.

5.4.1 Simulation Results

The performance of the MFE estimation methods has been assessed using the simulated pulsatile flow data. An example of this data is displayed in Figure 5.6a.



(a) Simulated Femoral Data

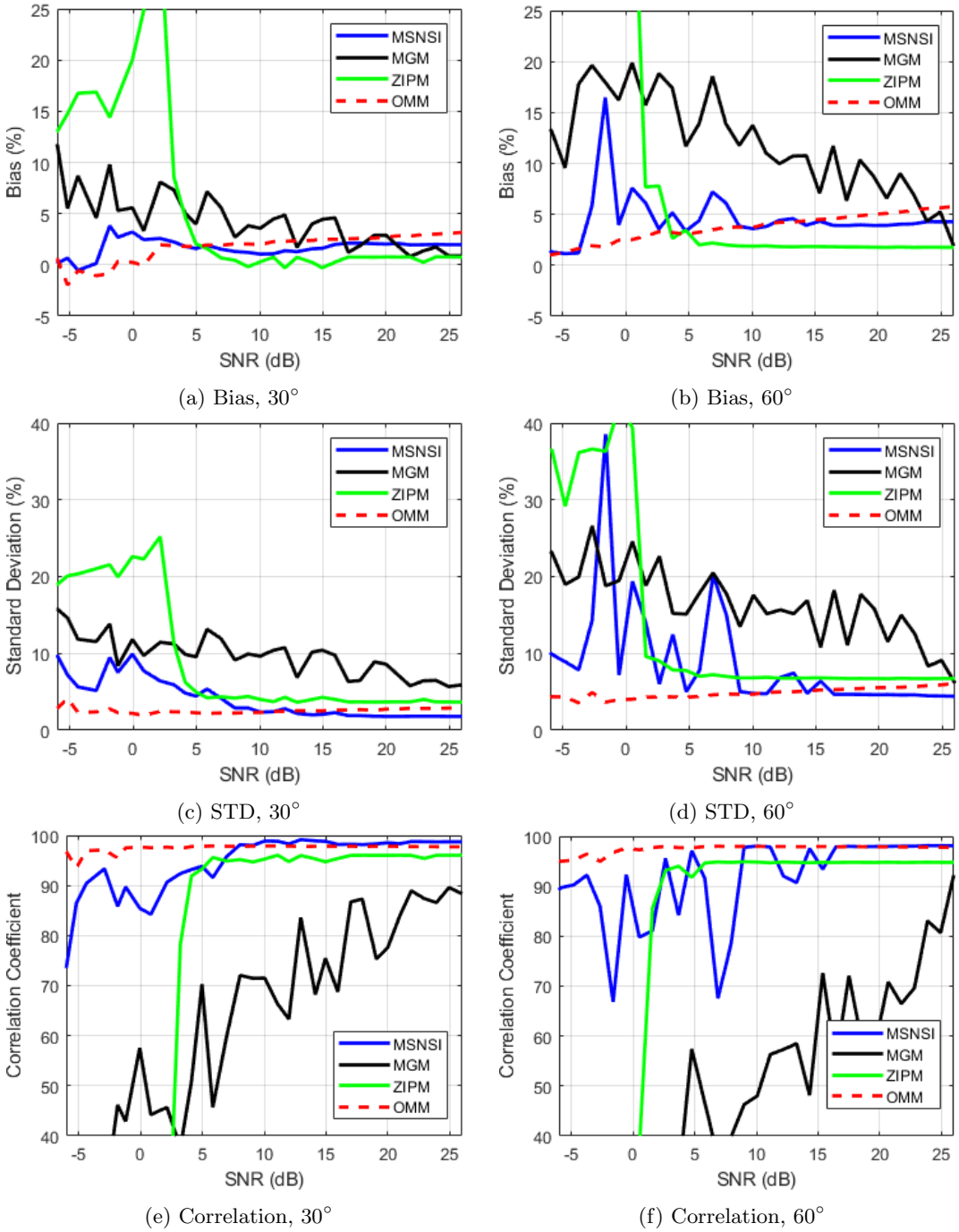
(b) Phantom Data

Figure 5.6: Simulated and phantom data with OMM, MSNSI and true MFEs. Where a) is femoral data with an SNR of 10 dB and insonation angle 60°, displayed with 60 dB dynamic range, and b) phantom data with 0.4 m/s flow rate, displayed using a dynamic range of 40 dB.

Statistics for the simulated data have been acquired at SNR values from -6 dB to 26 dB in steps of 1 dB; this was repeated three times and averaged. The calculated bias and STD of normalised maximum velocity for each method is displayed in Figure 5.7. These have been calculated using absolute envelope values, which prevents bias tending towards zero when an MFE estimation method tends towards zero in low SNR.

At SNRs greater than approximately 5 dB, ZIPM achieved the lowest bias. At SNRs below this, OMM achieved the lowest bias values. OMM achieved the most consistent STD values across the SNR range and the lowest values below 10 dB. Above 10 dB, MSNSI achieved the lowest STD values of approximately 2%.

The correlation coefficient between each method and the true envelope across the investigated range of SNR values, for insonation angles of 30° and 60° are displayed in Figures 5.7e and 5.7f, respectively. This illustrates how similar the extracted envelope is



(a) Bias, 30°

(b) Bias, 60°

(c) STD, 30°

(d) STD, 60°

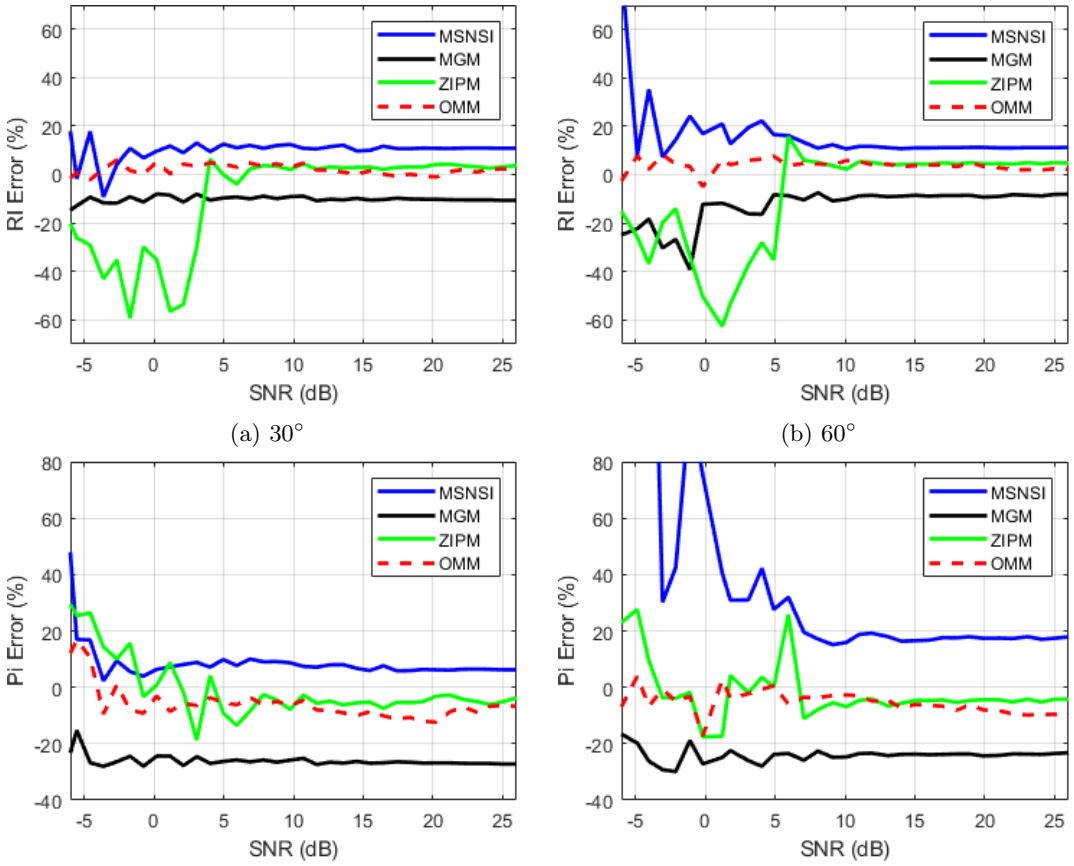
(e) Correlation, 30°

(f) Correlation, 60°

Figure 5.7: Bias, STD and correlation statistics for simulated femoral artery data using insonation angles of 30° and 60° across a range of SNR values.

to the true velocity envelope and the effect SNR and insonation angle has on this quality.

Figures 5.7e and 5.7f demonstrates that overall OMM produces an MFE very similar to the true MFE and remains stable for signals with low SNR values. For both cases of insonation angle, OMM achieved a correlation of greater than 95% at -6 dB. These results are consistent with visual inspections of the extracted MFEs, which demonstrate the OMM envelopes remaining highly stable and consistent across the SNR range, with the envelopes produced by the remaining methods becoming increasingly erratic at SNRs decreasing below approximately 6 dB. This is particularly true for the IPS methods in the



(a) 30° (b) 60°

(c) 30° (d) 60°

Figure 5.8: PI and RI statistics for simulated carotid artery data using insonation angles of 30° and 60° across a range of SNR values.

60° simulation data.

PI and RI indices were calculated for the simulated carotid artery data using equations 3.4 and 3.5. PI describes the degree of damping at different arterial sites, and RI is an indicator of circulatory resistance beyond the measurement point [12]. These features are discussed in Section 3.4, and are based on height ratios of characteristics in the MFE. This means they are less prone to certain errors such as incorrect insonation angle measurement [12]. The calculated RI and PI indices are displayed in Figure 5.8.

OMM achieved the best overall RI percentage error. Above approximately 5 dB, ZIPM and OMM achieved similar results. At high SNRs, ZIPM achieved the lowest PI error of –3.7%, with OMM and MSNSI achieving similar absolute errors for the 30° data (OMM with -6.5%, and MSNSI with 6.3%). Increasing the insonation angle to 60° resulted in OMM decreasing to -10%, and MSNSI increasing to 18% at high SNRs. ZIPM was less effected, reducing to -4.3%.

5.4.2 Phantom Results

The calculated bias and STD statistics are displayed in Table 5.2 for the velocities and SNR ranges investigated using the phantom data. The OMM method consistently resulted in the lowest STD, illustrating the stability of extracted MFEs using this method. No methods consistently performed best with respect to bias measurements. However, with respect to data with SNRs below 10 dB, OMM on average displayed the best performance. Comparing OMMs bias and STD with those from the best performing alternative method at each SNR below 10 dB, on average OMM achieves a bias and STD 0.7% and 3.3% lower, respectively. An example of the recorded sample data and associated OMM and MSNSI MFEs are illustrated in Figure 5.6b.

5.4.3 In-Vivo Data Results

The ability of the MFE estimation methods to perform under challenging conditions is evaluated further using in-vivo data. In addition to a variable SNR, in-vivo data includes erroneous contributions from, for example, tissue movement or other blood flows. The results presented here are for measurements of the aortic valve, in which a number of different features can be present [330]. Resulting envelopes, even in succeeding beats, can display high variability in size and shape.

Table 5.2: Bias and STD statistics for different envelope estimation methods, using phantom data

SNR (dB)	Bias (%)				STD (%)			
	OMM	MSNSI	MGM	ZIPM	OMM	MSNSI	MGM	ZIPM
Constant Flow Velocity of 0.4 m/s								
0	-0.7	3.5	3.7	28.0	0.5	5.0	9.8	8.9
3	0.1	2.0	1.2	28.3	0.4	3.2	4.7	8.3
5	-0.3	1.4	0.5	31.3	0.3	2.1	3.1	8.2
10	0.6	0.9	0.1	2.8	0.3	1.4	1.8	1.3
14	1.3	0.9	0.2	2.2	0.3	1.1	1.7	0.8
Constant Flow Velocity of 0.8 m/s								
0	0.3	3.3	1.1	17.1	0.6	4.9	8.0	10.4
3	0.8	2.5	-0.1	14.7	0.5	4.7	4.8	9.4
5	1.2	1.3	-0.4	21	0.6	4.0	3.1	7.4
10	1.7	1.0	-0.3	4.9	0.5	2.4	2.3	2.2
13	2.2	1.1	-0.3	3.3	0.5	1.9	2.3	0.9
Constant Flow Velocity of 1 m/s								
0	-0.8	2.5	-0.7	8.5	1.0	5.4	7.4	11.1
3	0.1	1.2	-1.8	14.3	0.8	3.7	4.1	8.2
5	0.4	0.8	-1.8	18.4	0.6	2.9	3.3	6.6
10	1.3	0.5	-1.8	4.6	0.5	1.7	2.7	1.7
14	1.4	0.3	-1.7	3.2	0.4	1.2	2.4	0.8

The results using in-vivo data are presented in Table 5.3. They demonstrate that, overall, the OMM method resulted in the lowest percentage error in terms of the total number of beats segmented; producing only 0.4% more beats than the true number of beats. The true number of beats was found by converting each recording into a spectrogram and counting the total number of whole beats present; data in which the total number of beats was hard to distinguish were removed. A whole beat is identified using the start or end of the preceding or proceeding beats, respectively. This allowed the number of beats within each recording to be compared to the number of beats extracted using each MFE method. The number of beats extracted for each MFE estimation method was then summed for all recordings giving a total number of overestimated and underestimated beats; the results

are given in Table 5.3. These values demonstrate that despite the OMM method percentage error being positive, this method resulted in the least number of overestimations. The OMM method also resulted in the least number of underestimations missing 0.8% of beats, compared to 5.8% missed by the next best-performing method, MSNSI.

Table 5.3: Number of Beats Segmented Using Each MFE Method

	OMM	MSNSI	MGM	ZIPM
Beats Segmented	7,908	7,613	7,677	7,491
Percentage Error[1] (%)	0.4	-3.4	-2.5	-4.9
Total Overestimations[2]	100	181	322	117
Total Underestimations[3]	69	445	522	503

1 – Percentage error of number of beats segmented with respect to true number.
2 – Summation of overestimated number of beats segmented from each recording.
3 – Summation of underestimated number of beats segmented from each recording.

The overestimation results in Table 5.3 illustrate that beat segmentation can result in beats incorrectly being segmented from the data; this is in response to erroneous signals or noise. The accuracies associated with performing beat segmentation using each of the MFE methods were further investigated using a sample of the in-vivo data.

A sample size of approximately 12% was used. The sample was attained by using the first four seconds of each of the 229 recordings. The extracted envelope and associated beat timing indices were generated for each audio sample and for each MFE method. Fig. 5.9 provides an example of this for the OMM and MGM envelope. This data allowed the number of false positives (a beat incorrectly segmented, or detected in the audio sample where there was no actual beat), false negatives (a beat present in the audio sample but not detected) and true positives (a beat existing in an audio sample was correctly detected) to be found . In Figure 5.9, the OMM MFE resulted in four true positives, and the MGM MFE resulted in two false positives, one true positive and one false negative. The false positives occur due to the start of systole being incorrectly estimated for the second beat. This is due to there being multiple narrow peaks in the systolic portion of this beat. The false negative occurred due to the final systolic peak not being preserved and therefore identified in the LPF MFE, resulting in the end of the fourth beat not being found. The results from the sample analysis are shown in Table 5.4.

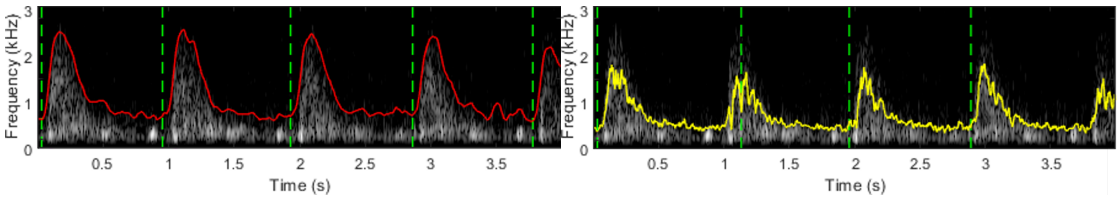
Table 5.4: Beat segmentation performance using each method on sample of in-vivo data

	OMM	MSNSI	MGM	ZIPM
True Positives[1] (%)	99.6	98.4	93.8	98.7
False Positives[1] (%)	0.4	1.6	6.2	1.3
False Negatives[2] (%)	2.2	10.4	35.4	10.4
Beats Segmented[2] (%)	97.8	89.6	64.6	89.6

1 – Percentage of associated beats with respect to number of segmented beats.
2 – Percentage of associated beats with respect to true number of beats.

Table 5.4 illustrates that OMM resulted in the highest percentage of beats being correctly segmented, with OMM achieving 97.8% , followed by MSNSI and ZIPM which both achieved 89.6%. Use of the OMM method also resulted in the lowest percentage of false negatives and false positives, which is in line with the overestimations and underestimations in beat numbers given in Table 5.3.

To evaluate the applicability of the proposed MFE and beat segmentation method for real-time applications, the time taken to run both algorithms for the in-vivo data was recorded. This was done using an Acer G9-592, with an i5-6300HQ CPU 2.3 GHz processor and 8 GB ram. The total processing time was 967 seconds, which corresponds to 0.12 s per beat; accommodating a hypothetical maximum heart rate of up to 490 bpm



(a) OMM (b) MGM

Figure 5.9: Example of in-vivo data containing four whole beats with OMM and MGM MFEs and associated segmented beats, where the green dash indicates the start or ends of identified beats. In this case, OMM segmented four true positives, and MGM segmented two false positives, one true positive and one false negative. Using the OMM MFE, SNR was measured to be 15 dB.

for real-time applications.

Finally, a series of paired t-tests were performed on the results presented in Table 5.3 to confirm the differences in performance. The beat segmentation performance of the OMM method was compared to the remaining three MFE estimation methods independently. The resulting p-values for this analysis are very small ($3 \times 10^{-5}, 1 \times 10^{-3}$ and 4×10^{-14} for the MSNSI, MGM and ZIPM methods respectively), confirming the observed difference in performances was not random.

5.5 Discussion

A series of metrics and scenarios have been analysed to provide an in-depth comparison of the investigated MFE estimation methods. Metrics include STD, correlation and waveform indices. These were used to analyse the ability of the proposed method to extract representative Doppler profiles, from which shape information can be found or further beat segmentation performed.

The implementation of the methods remained constant for each data set, apart from in the case of ZIPM, where this was not possible. This approach simulates real-world, automatic application where true maximum velocities are unknown, or for research scenarios where datasets exhibit high variability (e.g., SNR and flow rates) and thus it is impractical to adjust the implementation of the methods. In certain other applications, for example, where maximum velocities of interest are more restricted, the methods could be tuned to give better performance (e.g., restricting image size for image-processing methods, or adjusting cut-off frequencies to improve the performance of IPS methods), however, this was beyond the scope of the presented work.

5.5.1 Envelope extraction using the Doppler Audio Signal

In the OMM method, the Doppler audio is used to form a spectrogram and from that an image. Thus the parameters used for generation of the spectrogram images are explicitly defined, as discussed in Section 5.3. As described, the applied morphological operations are designed with respect to the resulting pixel resolution of these images. In contrast, the values of the parameters used to generate the images collected from an ultrasound machine are not known, and are device and user-specific. Variations and limitations between ma-

chines in this regard include screen refresh rates, spectrogram parameters, contrast, pixel resolution, zoom and image thresholding. The variations are in part due to the processing required to provide spectrograms as a form of visual feedback. In contrast, Doppler audio tends to naturally occur within a person's hearing range, allowing sonographers to use the audio feedback to guide the probe position during measurements. This means that audio requires less processing prior to being used as a form of feedback in comparison to spectrogram images. The process of sampling audio is relatively straightforward and can be performed using basic hardware (e.g., a laptop). Considering this, implementing an image-processing MFE approach using the Doppler audio signal presents a number of advantages.

For this work, no specific standards with respect to streaming audio from commercial ultrasound machines could be found. The maximum Doppler frequencies, sample rate of any captured audio and the number of channels should, therefore, be considered prior to applying the proposed method. Fortunately, in many applications, Doppler signals are limited to relatively low frequencies and do not require high streaming standards.

5.5.2 Envelope Estimation with Decreasing SNR

The performance of each MFE estimation method was evaluated across a range of SNR values using both simulated and flow phantom data. In each case, SNR was estimated using equation 5.2, which provides a consistent means of comparing each method. This approach has been used in previous studies as it allows the noise to be quantified in real-world measurements [114], however, can give negative estimates in low SNR conditions. This is illustrated in Figure 5.3, in which the Doppler profile is visually discernible at an SNR of -3dB. Evaluations have included such low SNRs to demonstrate the stability of the OMM method, and its potential to be used in automatic research or monitoring scenarios where noise conditions can vary. In terms of bias, no MFE method consistently outperformed across the two data sets. Nonetheless, OMM typically achieved the lowest bias at lower SNRs (e.g., below 5 dB). It can be seen that unlike the other MFE methods, OMM exhibits decreasing bias with decreased SNR. This is due to the automatic thresholding used by OMM, which does not take into account spectral broadening. As more noise is introduced, the lower level spectral broadening is masked and the estimated maximum frequency is closer to the true value.

The OMM method produced very stable results across the SNR range with respect to the simulation correlation statistics and STD statistics for both the simulated and phantom data (Figure 5.7 and Table 5.2 respectively). This is particularly evident in the correlation data, where the OMM method achieves a correlation greater than 95% at -6 dB for both insonation angles.

The IPS methods typically exhibited more erratic behaviour than the image-processing methods. This is due to the fact that IPS methods require the IPS to exhibit its characteristic shape in order to accurately estimate maximum frequency points. This condition is met less consistently (for example, the IPS knee becomes less defined, and the transition from signal to noise more gradual) as SNR deteriorates and the measurements exhibit signal loss and increased variance. The MSNSI method compensates for this by employing only time points with adequate signal strength and then using interpolation and smoothing. If a portion greater than 0.1 s with poor signal strength is identified, that region is set to zero. This is useful in measurements where flow discontinuities occur, in such cases the IPS curve would divert far from its characteristic shape and result in very poor maximum frequency estimations. The OMM method does not include any equivalent conditions, however, given that a global threshold is used to generate the binary images, small discontinuities do not result in incorrectly identified signal contributions. In such cases, the MFE is set to the minimum frequency bin in which flow signal can be detected. If no flow were present at all in a section of processed audio, the OMM method would be unable to detect this, and the resulting MFE would be erratic. The in-vivo data includes a large variety of waveforms and signal qualities, however, specific analysis of how each method performs in response to flow discontinuities was not performed and was beyond the scope of this investigation.

At a certain point, the MSNSI power threshold condition becomes detrimental as more of the envelope is set to zero. This begins to have an effect at SNRs below 10dB, this can be seen in Figure 5.6a, where regions of slow diastolic flow with lower spectral power have been set to 0. As SNR reduces further, the stable performance of MSNSI begins to become more erratic. Overall, the MSNSI method performed better despite MGM producing lower bias values than MSNSI at very high SNRs. This was found to be due to the MGM method performing very poorly during the weaker, diastolic portion of the signal.

The ZIPM method produced stable results at SNRs greater than 5 dB but deteriorated

quickly at lower SNRs. As described in Section 5.3.3, a dynamic range is chosen for the images processed using this method. This dynamic range dictates how successful proceeding thresholding is; as this is fixed, at a certain SNR the signal can no longer be easily distinguished, and the method's performance quickly deteriorates. The dynamic range chosen to display the ZIPM images was chosen based on its performance with respect to the calculated statistics. Choosing the best dynamic range can be challenging when implementing ZIPM for new data. This was found to be the case using the in-vivo data, in which a dynamic range much larger than that used for the simulated and phantom data was required for good performance, and which is discussed further in Section 5.5.5. The dynamic range used for the ZIPM images meant only the strongest portion of the signal was visible (which occurs at the maximum velocity). Consequently, this meant the addition of noise had no effect on the binary image across a large range of the SNRs used. Additionally, the diastolic portion of the simulated femoral data could not be captured. The combination of this factor and slight overestimations with respect to maximum velocity during the systolic portion resulted in very low bias values, but less accurate STD and correlation values. ZIPM was able to produce good RI and PI results, this is partly due to the simulated carotid artery being more consistent. These results are discussed further in Section 5.5.4.

5.5.3 Envelope Estimation with Increasing Insonation Angle

The inherent properties of Doppler ultrasound systems give rise to a phenomenon known as intrinsic spectral broadening, which manifests as a blurring of the Doppler spectrum. A moving target, when measured using Doppler ultrasound, results in spectral content with a range of frequency shifts (and not one singular value). Blood contains many moving targets that contribute to the measured Doppler signal and result in a smearing of the frequency spectrum [12]. The presence of spectral broadening is attributed to two contributions referred to as local geometric broadening and transit-time broadening [59]. The degree of this effect increases with the insonation angle.

Simulated data was generated for insonation angles of 30°, and 60°. This allowed the effect of spectral broadening with respect to MFE estimation performance to be investigated. The bias, STD and correlation statistics show that typically the 60° simulated data resulted in deteriorated performance for all MFE methods. In the case of ZIPM, the

increased insonation angle resulted in a lower bias. This was due to the combination of increased overestimation during systole, and underestimation during diastole as discussed in Section 5.5.2. The ZIPM envelope deviates further from the true envelope within the 60° data, which is illustrated from the STD results.

As discussed in Section 5.5.2, OMM exhibits increased bias at higher SNRs due to spectral broadening. This is more evident in the 60° data as higher insonation angles result in increased levels of spectral broadening. Use of metrics derived from MFEs, such as peak systole, should keep such effects in consideration.

The OMM method was able to generate highly correlated envelopes at low SNR values for both insonation angles. This means that despite spectral broadening, accurate MFE shape can still be extracted allowing successful beat segmentation or accurate waveform features to be obtained. This is illustrated by the RI and PI statistics, which remain on average below 10% for the OMM method, for both angles and low SNRs. OMM and MSNSI resulted in similar absolute PI error at high SNRs for 30°. Increasing the insonation angle resulted in an absolute error increase of 47% and 185% for OMM and MSNSI, respectively. OMM achieved a PI error of -10% for the 60° data, illustrating its potential use for such applications.

5.5.4 PI and RI Estimation

Features can be extracted from MFEs to provide additional means of analysing blood flow. The ability to extract two popular waveform indices, PI and RI, were investigated using each MFE estimation method. These results, shown in Figure 5.8, illustrate that typically PI error was greater than RI error. Through inspecting the corresponding MFEs, it was found that performance was similar to that exhibited in the simulated femoral data. The OMM and MSNSI methods both produced envelopes with low bias. As bias is estimated from each time point, this corresponded to overall good estimates of the envelope mean. However, it was found that MSNSI overestimated peak systole, and increasingly so with SNRs below approximately 5 dB. This results in an overestimation of PI and RI values. Conversely, any inaccuracies in MFE estimation are more consistent across the whole MFE for image-processing methods. This results in more accurate PI and RI estimation. MGM performed similarly to MSNSI, however, underestimated peak systole resulting in negative PI and RI error.

Considering equations 3.4 and 3.5 and the more consistent M values, inaccurate S values are more detrimental for PI estimates. Furthermore, as discussed in Section 5.5.2, the MSNSI method results in more erratic behaviour at lower SNRs and includes a condition that can set portions of the MFE to 0. This increases the likelihood of D_{min} being smaller than D_{end}, and will further detrimentally affect estimates of PI.

It can be seen from Fig 5.8, that ZIPM begins to deteriorate below approximately 5 dB; this is consistent with the bias and STD results in Section 5.4.1. As indicated by the correlation statistics in this Section, the ZIPM MFE rapidly deviates from the true MFE shape. This results in D_{end} increasing relative to S, and estimates of RI decreasing. The whole MFE increases, and although becomes far from the true MFE, does not result in huge RI and PI errors. Comparatively, the MSNSI envelopes resulted in worse PI and RI errors at low SNR, despite them overall being closer in shape to the true MFE. These observations highlight the need to consider different metrics when assessing MFE performance.

5.5.5 Beat Segmentation Performance

The ability to segment beats is essential for automatically extracting and monitoring beat specific measurements, like those discussed in the previous section. It allows for measurement averaging, preventing the practice of calculating values from representative beats, a process which may be a significant factor in test-retest variability [331]. Robust averaging has been shown to be clinically advantageous in certain applications, for example, resynchronising pacemakers [119–121]. Furthermore, it makes analysis of larger datasets more feasible, which would be clinically desirable [332], and could enable research ventures that were previously deemed too time-consuming. In the context of this thesis, good beat segmentation performance is critical. This is to ensure as much information as possible is extracted from Doppler ultrasound datasets, and that it is extracted in a consistent manner. Furthermore, practical solutions investigated in this work for aiding Doppler ultrasound analysis are not helpful unless they can be applied in real-time applications; and thus implemented methods must be reliable in response to a wide array of signal quality and characteristics.

The ability to perform successful beat segmentation was investigated using a large dataset of in-vivo measurements. These measurements inevitably contain more artefacts

than the simulated and phantom data. The measured signals include contributions from tissue movements and erroneous blood flow signals from nearby vessels. Other signals, such as valve clicks, can be present as well as variations in noise due to differences in transducer and tissue coupling and signals due to the transducer moving.

The results demonstrate that combining OMM with the described beat segmentation method can result in a high percentage of beats being correctly segmented, with the OMM method segmenting 8.2% more beats correctly than the next best performing method, MSNSI. Crucially, the sample test indicated that very few false positives were identified using OMM. This is a significant result of this research, as this characteristic is vital for applications which use processed beats to identify abnormalities in measurements, identify certain traits or perform classification tasks (such as classifying heart disease [195]). The difference in performance with respect to true positive and true negatives could be of particular significance when monitoring patients with challenging recordings such as weak CO. The design and testing of the beat segmentation method has been limited to Doppler measurements from the aortic valve. This included data with a wide range of Doppler profiles, including ones far from their characteristic shape (e.g., high-end diastolic velocities). Considering this, the method is expected to also perform well using Doppler measurements from other locations (for example, from the carotid artery), however, this has not been confirmed.

In real-world applications, an automatic method needs to operate at sufficient speeds to extract the envelope, perform beat segmentation and extract information. The proposed method was found to take on average 0.12 s to extract a segmented beat MFE, providing the remaining 0.15 s to extract additional information for a heart rate of 220 bpm. This illustrates that the proposed method could be implemented in real-world applications. Furthermore, the method requires only the MFE to function. This means no additional hardware is required and thus can remain low cost, fast and highly portable. Combining OMM and the proposed beat segmentation software allows real-time, continuous monitoring of a person's blood flow with live cardiac cycle analysis.

It was found that despite the good performance exhibited by ZIPM with respect to the simulation and phantom data, a much higher dynamic range was required to generate the images used in the in-vivo data. The low dynamic range used for the simulated and phantom data allowed the signal to be clearly defined (as shown in Figure 5.5), however, the in-vivo data contains erroneous signals and variable SNR, preventing such

a low dynamic range from being used. The need to select an appropriate dynamic range value for particular datasets stopped the ZIPM method from being truly automatic within this study.

5.6 Conclusion

A new MFE estimation method (OMM) has been presented. OMM is fully automatic, can be implemented in a real-time manner and only requires the Doppler audio signal as an input. The performance of OMM has been systematically evaluated for a wide range of signal qualities using simulated data, phantom data and in-vivo data. The performance has been compared with three other state of the art MFE estimation methods.

It was demonstrated that the proposed OMM method reliably produced envelopes suitable for further beat segmentation. Across a wide SNR range, the OMM method consistently produced the most stable envelopes with good correlation to the true shape. This was further demonstrated using in-vivo data, where it resulted in 8.2% more beats being correctly segmented in comparison to the next best performing method. This is a significant characteristic of the method and demonstrates its potential application for monitoring in clinical scenarios, and automatic processing of large datasets for research purposes.

In the context of this thesis, OMM is used to generate stable envelopes and enable reliable beat segmentation. The method is also combined with MSNSI and MGM to extract a range of MFE based features for machine learning applications, as described in Section 7.3.

Chapter 6

Beat Segmentation

6.1 Introduction

The timing indices of cardiac cycles allow clinically valuable blood flow measurements, or features, to be extracted from Doppler ultrasound measurements. Using these indices to split ultrasound data into individual beats is referred to as beat segmentation. In clinical scenarios, automatic methods can enable real-time haemodynamic measurements and beat averaging, in research applications automatic methods allow large and otherwise overly impractical datasets to be processed.

The work in this thesis necessitated a robust method of beat segmentation for both offline dataset processing and real-time software implementation. In order to preserve desirable characteristics of Doppler ultrasound technology, specifically its affordability and ease of use, this method needs to function without any additional hardware (e.g., ECG). In addition to this requirement, the employed method needs to function well when applied to dynamic MFEs (which occur when the probe is scanning through planes in order to locate the IPP), and MFEs that are far from their characteristic form (as the method needs to function when the operator has not yet located the IPP). This is in contrast to more straightforward and restricted scenarios, such as using data limited to good quality measurements from the IPP. As discussed in Section 3.2, previous methods have been evaluated for these more simplified conditions and necessitate some user input preventing them from being truly automatic, such as entering the first timing index for an MFE containing multiple beats [160].

The research described in this document led to two methods of beat segmentation being developed. The first, the slope-gradient method, is based upon the method used within the USCOM device. Description of this method has been limited to high-level details [78], meaning it could not be replicated identically. The primary concepts were adopted, as well as conditions implemented in previous works. The resulting method uses the slope of systole to identify the starting points of cardiac cycles and requires no input other than the MFE.

In addition to this, a correlation method was later developed. This method uses a heart rate estimate and an ideal MFE profile (MFE_{IPP}) to identify cardiac cycles. The MFE_{IPP}, and correlation measurements, were initially used to quantify the similarity between obtained MFEs and MFEs from the IPP. This characteristic is used as a feature for pattern recognition models (described in Section 7.3.3), and was found to be the number one ranked feature for the beat rejection classification model, which is used to remove poor data for real-time applications (Chapter 9).

The two methods are compared using a sample of data containing 1,324 beats, with varying quality and MFE shape. The two methods were both shown to perform very well, however, the correlation method gave marginally improved results and correctly segmented 99.2% of beats compared to 97.2%. Considering this, and that the correlation method is already used to extract correlation features, this method has been adopted for the feature and image extraction procedures described in Chapters 7 and 8.

Despite the correlation method being adopted as the chosen method of beat segmentation, steps from the slope-gradient method are used to identify the end of systole. This is required to extract certain features and is described in Section 7.3.1. Considering the good performance of both methods and their differences with respect to required inputs, computational cost and complexity, they have both been described in the given work.

The remainder of this chapter contains four sections. The immediate two describe the slope-gradient and correlation methods respectively, this is followed by a section detailing the evaluation methods used, and a final section presents and discusses the results.

6.2 Slope-Gradient Method

The slope-gradient method requires only the MFE as an input. The two main steps involved are the identification of peak systole positions, and subsequently, identification of

the systole start positions. The method shares common techniques to that employed by USCOM [78]; specifically using a low-pass filtered MFE to find temporal indices, and using the rising slope of systole to estimate the start of systole. However, the original method could not be implemented as the relevant document was limited to high-level details [78].

The proposed approach allows the start of systole to be identified, even for MFEs which exhibit unusual behaviour either side of peak systole. This overcomes limitations of assuming that a minima occurs prior to the start of systole [132].

6.2.1 Peak Systole Identification

Peak systole positions are first approximated using the unsmoothed MFE (described in Section 5.2.4). The mean of this envelope is first set to zero, and then it is low-pass filtered (LPF), which removes frequencies above 3.7 Hz, and results in a signal that is more sinusoidal in appearance. This assumes a heart rate of less than 220 bpm, which is well within the normal range for adolescents and adults [333].

The approximate peak systole positions are then found by identifying peaks in the filtered MFE. A 'minimum peak distance' condition is used to make this more reliable (i.e., time between consecutive peaks). A minimum peak distance of $0.8/f_{HR}$ is used, where f_{HR} is the estimated heart rate frequency. This condition assumes the heart rate reduces no more than 20% within the processed envelope, and is a percentage employed by previous methods [118, 132].

The heart rate frequency is estimated from the PSD of the sinusoidal signal, calculated using the Welch method. The frequency corresponding to the maximum value in the PSD is taken as f_{HR}. Final peak systole positions are taken as the peaks in the (smoothed) MFE closest in time to the approximate positions. Figure 6.1 illustrates how peaks found in the LPF MFE are used to identify peak systole in the MFE.

6.2.2 Start of Systole Identification

The next step is to estimate the start of systole. Low-frequency content in the Doppler audio signal can obscure the transition between diastolic and systolic blood flow (Figure 6.2) and thus prevent the start of systole positions from being easily identified.

To overcome this, the rising slope of systole is used. This occurs immediately prior to peak systole. The gradient of this slope is used to plot a line that intersects 0 Hz. This

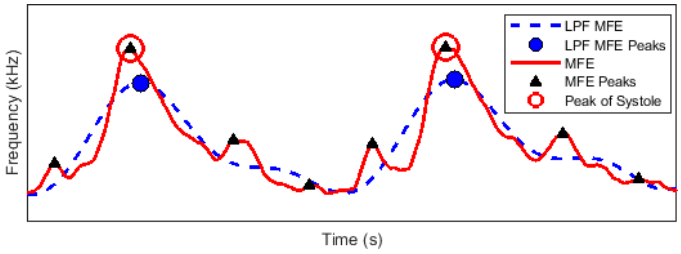


Figure 6.1: Peak systole identification using approximate peaks found from the LPF MFE, illustrated for scenario where MFE contains numerous peaks, for example due to low SNR conditions.

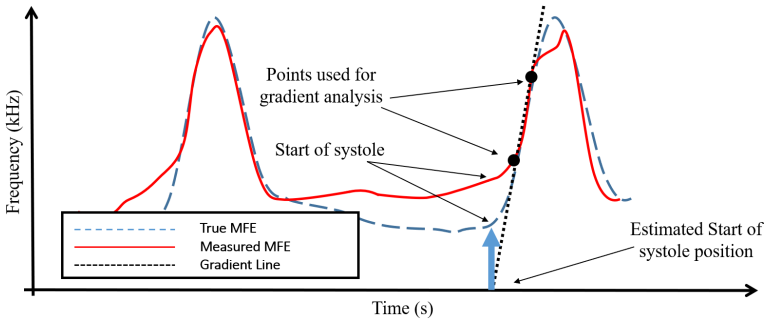


Figure 6.2: Method of estimating start of systole, illustrated for scenario where start of systole and peak systole are obscured, for example due to low SNR conditions.

point is taken as the approximate start of systole. Two points are selected on the rising slope to calculate the gradient and to plot the intersection line. These points correspond to 50%, and 80% of peak systole (the locations of which were found in Section 6.2.1). These percentages were empirically chosen as this region of the envelope typically exhibits a strong signal. Figure 6.2 illustrates how this approach can be implemented in scenarios where peak systole is not clearly defined, for example in poor SNR conditions.

6.3 Correlation Method

The correlation method extracts timing indices by analysing the MFE shape and spectral content at low frequencies. The techniques employed were initially designed to quantify the similarity of a beats MFE with respect to an ideal beats MFE. These features are described

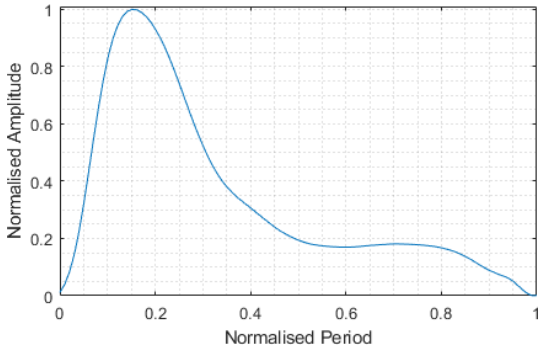


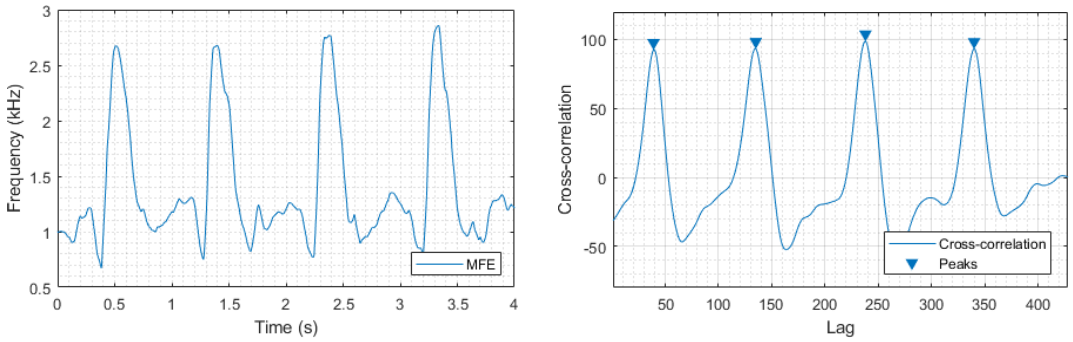
Figure 6.3: MFE_{IPP} used to obtain correlation measures

within Section 7.3.3. Compared to the slope-gradient method, fewer empirically derived variables are required. However, this method does require a reference MFE (MFE_{IPP}), this was formed using the expertly collected data at the IPP. The corresponding beats were normalised and averaged to give the MFE_{IPP}, this is displayed within Figure 6.3.

Previous methods of beat segmentation have included correlation approaches. In these examples, autocorrelation has been used to acquire timing indices. This technique has been used to find approximate peak systole positions [132]; assumptions regarding the envelope shape were then used to identify the start of systole. The start of systole positions have also been found directly using autocorrelation [118]. Identifying the start of systole in this manner is very useful, however, necessitated the first cardiac cycle starting position to be manually entered. The presented correlation method identifies the starts of systole and is fully automatic.

Using the same approach as described in Section 6.2.1, the associated heart rate (HB) is estimated for the processed MFE. This is used to resize the relative length of the MFE_{IPP}, for example, if HB is 60 bpm and the MFE is 4 s long (as used in the given work), the resized MFE_{IPP} will be 0.25 $\times$ MFE duration (i.e., MFE_{IPP} will have an effective length of 1 s). Cross-correlation is then used to identify the starts of systole within the MFE. An example of this is given in Figure 6.4.

Once these systole positions are found, the correlation between individual beats and the MFE_{IPP} are calculated ($Corr$). If all of these correlations are greater than 0.7, the process is complete. Typically, this means at least four beats require a correlation greater than this value, which was found to be empirically indicative of well-segmented beats.



(a) MFE (b) MFE_{IPP} and MFE cross-correlation

Figure 6.4: Identification of start of systole using cross-correlation

If this is not the case, a series of additional HB values (HB_check) are used to re-size the MFE_{IPP} and the process is repeated. HB_check is defined using the values in HB_Buffer (as shown in Figure 6.5), where HB_Buffer is a circular buffer containing up to 10 previous HB estimates. HB_check contains five values linearly spaced from $0.8 \times Average(HB\_Buffer)$ to $1.2 \times Average(HB\_Buffer)$. This is illustrated at a high-level in Figure 6.5.

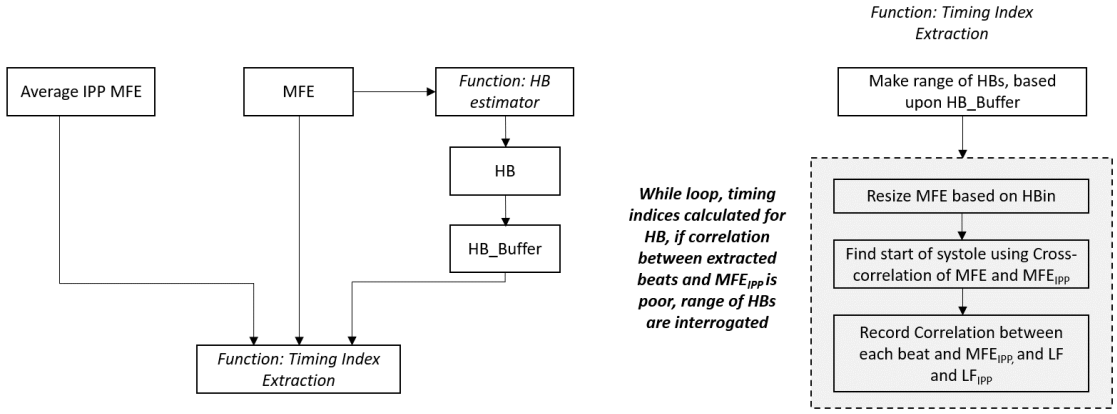
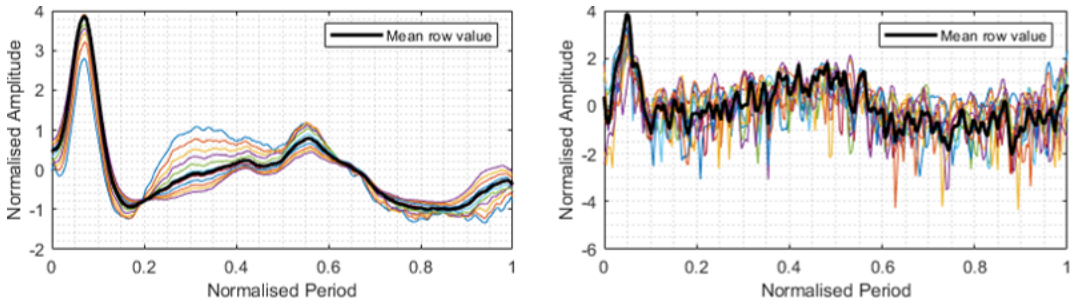


Figure 6.5: High-level block-diagram detailing correlation method of beat segmentation

In each iteration, the mean $Corr$ value is recorded. In addition to this, an image correlation is measured. This uses the lowest 10 rows within the OMM image, which corresponds to low frequencies and captures strong signals such as valves closures, and

is denoted by LF. An ideal LF (LF_{IPP}) was used to attain an additional correlation metric, this was made by averaging images corresponding to the MFEs used to form the MFE_{IPP}; the LF_{IPP} and an example LF are displayed in Figure 6.6. These image correlations ($ImCor$) are included within the feature matrix described in Section 7.3. This additional correlation metric was found to be particularly useful in cases where the MFE shape was far from the MFE_{IPP}. In such cases, $Corr$ values can be less informative than the $ImCor$, as these low-frequency components can be clearly defined even when the MFE is not; this is illustrated in Figures 6.8d and 6.9a.



(a) LF_{IPP} (b) LF

Figure 6.6: Example of low-frequency content in LF_{IPP} and an interrogated cardiac cycle cardiac cycle.

Once these additional HB values have been processed, the timing indices that correspond to the best Cor and $ImCor$ metrics are used. Figures detailing lower-level details regarding this algorithm are provided in Appendix B.

6.4 Evaluation

The performance of both beat segmentation methods were compared using a sample of the recorded data. This sample was created by first constructing a new dataset which contained the first 4 s from each probe position, and searching period (e.g., time between sampling off-angle and IPP) during a measurement set. This produced a dataset with wide variations in MFE shape and quality, including dynamic periods and measurements from each probe position. This also ensured that the sample was not dominated by measurements from individual participants or by data corresponding to searching periods.

Approximately 25 minutes worth of data was then randomly sampled from this new dataset and used to compare the beat segmentation methods. This is illustrated in Figure 6.7.

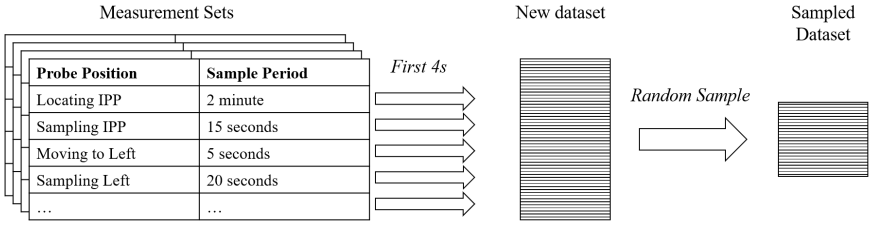


Figure 6.7: Data used to evaluate beat segmentation methods, sampled to ensure balanced and varied data.

For each 4 s data portion, the associated spectrogram, MFE and beat timing indices were saved as an image. This enabled the true number of whole beats to be visually counted for each data portion, as well as the corresponding true positives, false positives and false negatives with respect to the extracted timing indices. To reflect the beat segmentation methods, which are designed to only segment complete cardiac cycles, only complete beats are counted. Cases where beat timing indices where visually ambiguous were not included.

6.5 Results and Discussion

Using the generated spectrogram images the true number of beats present within the sampled data was counted; this found there to be 1,324 beats. The performance metrics for each beat segmentation method is displayed in Table 6.1. This illustrates that both methods performed well, and were able to segment the vast majority of beats. However, the correlation method marginally outperformed the slope-gradient method for each metric.

The data sample used to assess both beat segmentation methods included a wide array of signal qualities and Doppler profile shapes. This provided a thorough comparison of their performances for the given application in this work, which is to operate reliably across different participants and when the probe position is far from the IPP. Figure 6.8 illustrates both methods performing well, despite challenging and dynamic MFEs.

Figure 6.8 illustrates a variety of envelope shapes, from which both beat segmentation methods successfully extracted timing indices. These include profiles more indicative of good measurements (Figure 6.8a), and dynamic profiles which are encountered when the

Table 6.1: Beat segmentation performance for gradient and correlation methods

	Gradient Method (%)	Correlation Method (%)
True Positives[1]	97.5	99.2
False Positives[1]	2.5	0.8
False Negatives[2]	4.7	2.0
Beats Segmented[2]	95.3	98.0

1 – Percentage of associated beats with respect to number of segmented beats.
2 – Percentage of associated beats with respect to true number of beats.

probe is being moved (Figure 6.8b). Figure 6.8c displays a very erratic and noisy profile, and finally Figure 6.8a displays an example where very little flow is captured, potentially due to only a section of the aortic arch being insonified.

Figure 6.9 provides two examples where the correlation method outperformed the

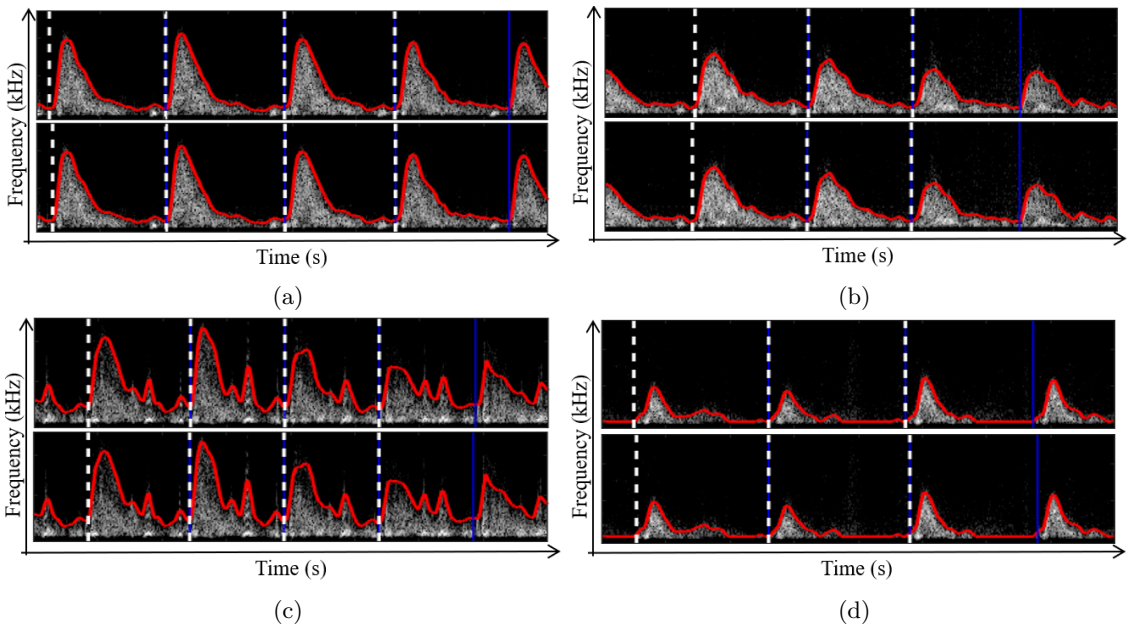
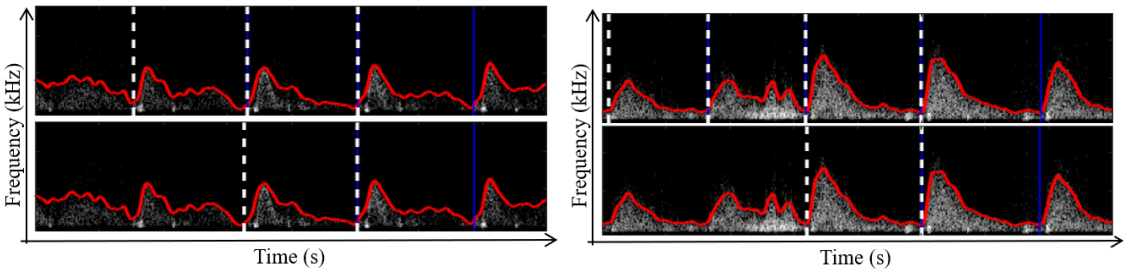


Figure 6.8: Pairs of correct beat segmentation timing indices, for correlation method (top images) and slope-gradient method (bottom images), where white dashed lines indicate start of cardiac cycles, and blue lines indicate end of cardiac cycles.

slope-gradient method. In these cases the slope-gradient method failed to identify the peaks of the missed beats (first beat in Figure 6.9a, and the first two beats in Figure 6.9b). A high diastolic region with many small peaks prior to the first beat in Figure 6.9a likely contributed to this. Within Figure 6.9b the first two beats are smaller than the remaining three and contain two prominent erroneous peaks, which again will have made peak identification challenging. Such cases where sequential beat profiles differ greatly and profiles exhibit high diastolic regions can lead to reductions in slope-gradient performance with respect to the correlation method.



(a) (b)

Figure 6.9: Examples where correlation method (top images) performed better than the slope-gradient method (bottom images), where white dashed lines indicate start of cardiac cycles, and blue lines indicate end of cardiac cycles.

Considering the improved performance of the correlation method, and that these processes enable useful features to be extracted, this method has been adopted for the remaining work as the chosen means of beat segmentation.

Chapter 7

Continuous Feedback

7.1 Introduction

The deterioration of outcomes in medicine and an increasing abundance of clinical data is fuelling and enabling medical applications of artificial intelligence (AI). It is envisioned that almost all clinicians, of all types, will one-day rely upon AI [334]. Within this chapter, pattern recognition methods are used to design a mathematical model that provides continuous and automatic feedback reflective of probe position.

The merits and limitations of Doppler ultrasound have been discussed in depth within Section 2.6. The dependency of measurements with respect to operator competence was highlighted as a major hurdle. This is for both acquiring accurate and consistent performance, and is presented as a potential reason for the technology not being more widespread. This is because a highly skilled user is mandatory for both data acquisition and measurement interpretation [335]. Previous works have investigated machine learning solutions to aid measurement interpretation (e.g., automatic disease diagnoses), these were discussed in Section 3.5. However, automatic systems that assist data acquisition have not been documented.

Human-machine interfaces that enable less experienced operators to collect blood flow data could be hugely impactful. These systems could ease measurement acquisition in clinical settings, allowing more haemodynamic monitoring and management to occur, and would also enable the use of the technology in more rural and developing countries [208]. A sobering example of how such outcomes could be beneficial is the ongoing covid-19

pandemic, in which septic shock presents in serious cases and necessitates management [336].

Data acquisition using Doppler ultrasound requires delicate manipulations of probe angle and pressure. These manipulations can cause subtle changes in the obtained measurements. An inherent difficulty is interpreting these changes, and responding appropriately. To mitigate this difficulty, an SVM regression model that provides continuous feedback has been designed. This feedback increases and decreases throughout a scan to reflect beneficial and detrimental probe movements, respectively. In this regard, the feedback is similar to scan image scores, which are derived offline and were described within Section 4.2.3. No examples of methods that automatically or quantitatively score Doppler scans were found within the literature. With respect to the subjective offline methods, the generated feedback is based on a far larger array of features and is continuous.

The resulting model uses a selection of different features. These include features previously employed for Doppler ultrasound classification tasks, features used in other related disciplines, and novel features. In total, a large feature matrix consisting of 963 different features and 16,553 observations was formed. These are described in-depth in Section 7.3. A thorough hybrid filter-wrapper feature selection method is used to identify the most successful features. Novel features, and those which have not previously been documented with respect to Doppler ultrasound, are shown to give good performance. Specifically, MFCC features, which are commonly used in audio classification, are highly ranked.

In conclusion, the main contributions contained in this chapter include:

- The design of a large feature matrix that includes classic Dopper ultrasound features, novel features and features that have not previously been applied to Doppler ultrasound.
- The design and implementation of a thorough feature selection method, which identifies promising extracted features.
- The development and evaluation of a regression SVM model, designed to quantify scan quality in an automatic and continuous fashion.

7.2 Chapter Contents and Structure

This chapter describes the construction and evaluation of a regression feedback model. There are many techniques and modelling approaches that can be used to build feedback models, however, their construction typically consists of four common steps [205]:

1. Signal Preparation
2. Feature Extraction
3. Feature Selection
4. Model Training

An overview of these steps, with respect to the given work, is displayed in Figure 7.1. This illustrates how different methods designed through the thesis are combined to realise the above steps, and includes footnotes detailing where they are presented. The processes and methods described in this chapter are also used in the proceeding chapters to train classification models.

The methods used to realise steps one to four are described within this chapter. Following this, the corresponding results are presented and discussed. To enable easy navigation of this information, the structure of this chapter with respect to these steps will now be briefly presented.

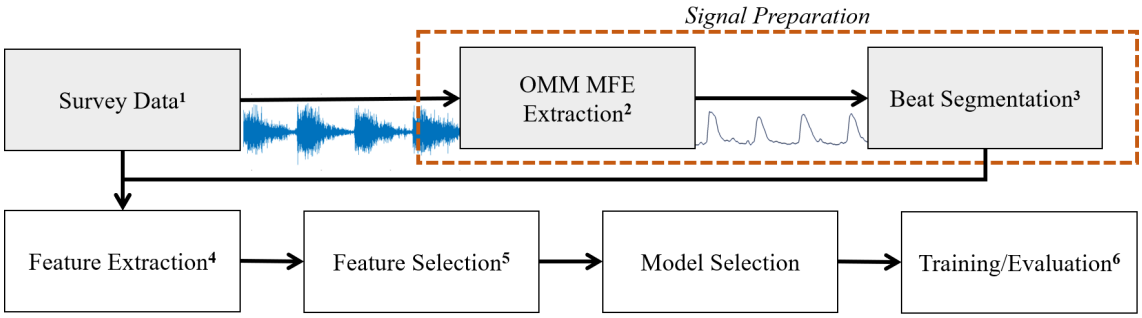


Figure 7.1: Overview of steps used to construct regression model, where methods used in **1** are described in Section 4, **2** in Section 5.2, **3** in Section 6, **4** in Section 7.3, **5** in Section 7.6 and **6** in Section 7.7.1.

7.2.1 Signal Preparation and Feature Extraction Steps

The first step is realised using the MFE estimation and beat segmentation methods described in Chapters 5 and 6. These methods are used to attain timing indices, which allow features to be extracted from individual cardiac cycles.

A large array of features are extracted in this work. These are used to train models in this chapter, and the following two chapters. These procedures result in a variety of beneficial features being identified. These include novel and established features, and is a significant contribution of this work. The features extracted and associated naming conventions are described in Section 7.3.

7.2.2 Feature Selection Step

All models are trained using a subset of the best performing features for that given task. Feature selection is performed using a hybrid method and is detailed in Section 7.6. This process uses labelled data with both filter and wrapper feature selection methods.

In proceeding chapters, discrete known targets are used to label the data. However, for continuous feedback, the choice of labels is less obvious. The labels used, and the justification for this choice are detailed in Section 7.4.

7.2.3 Model Training Steps

Following feature selection, models can be trained and evaluated. The evaluation methods are described in Section 7.7, this includes a method of cross-validation which is also employed in the proceeding two chapters.

Section 7.7 details the metrics used to assess trained continuous feedback models, these results are presented and discussed in Sections 7.8 and 7.9, respectively. These sections also cover the features identified as being the most useful for the continuous feedback task. All extracted features will now be discussed.

7.3 Extracted Features

The features extracted in this work belong to those previously documented for Doppler ultrasound analysis or classification, as well as those which have either not been applied to

Doppler ultrasound previously or have been developed as a part of this work. This section will detail these features, and give context as to why they have been included.

Features have been organised into different types, and where necessary, groups and subgroups. A breakdown of the feature categories used are displayed in Figure 7.2. In many cases, features could be considered as belonging to multiple categories, for example, spectral analysis of envelopes could be considered as belonging to the Waveform (`W`) or Spectral types. These feature categories largely serve to assist in the organisation and referencing of features.

Features are extracted in a beat-by-beat fashion, however, in some cases the data used is limited to the systolic portion (`SYS`) of the cardiac cycle (`CC`). This is to reflect scanning techniques employed when locating the IPP, which focus upon visual and audio feedback within the systolic portion, and previous works which have demonstrated feature extraction from specific regions of the cardiac cycle to be beneficial [337]. Cardiac cycles are segmented using the correlation method described in Section 6.3, and end of systole is identified by applying the slope-gradient method to the falling slope of systole.

The categories displayed in Figure 7.2, as well as the data length (`SYS` or `CC`) and input data, such as audio (`A`), envelope (`E`) or image (`I`), are used to give each feature a unique name. These are constructed using the format '`category_DataLength_InputData_FeatureCode`' where the category is the lowest level category for a given feature (i.e., this will always be `AF` for the Audio features, but can be one of eight subgroups for spectral features). The `FeatureCode` provides specific details regarding the feature, these details are given in the following sections.

An example feature name is '`W_CC_E_NE1_1`', this corresponds to a feature which is of the waveform type (`W`), extracted from the whole cardiac cycle (`CC`) and using an envelope (`E`). As will be described in Section 7.3.3, '`NE`' indicates a novel envelope feature, and the remaining numbers '`1_1`' link to a statistic. These naming approaches make listing and inspecting feature sets more manageable.

The extracted features will now be detailed. Due to the large range of features extracted, these are primarily limited to high-level details. Appendices F to K include further technical details and methods used to extract features where necessary. The principal concepts of previously documented features and examples of their use are described within Section 3.5.

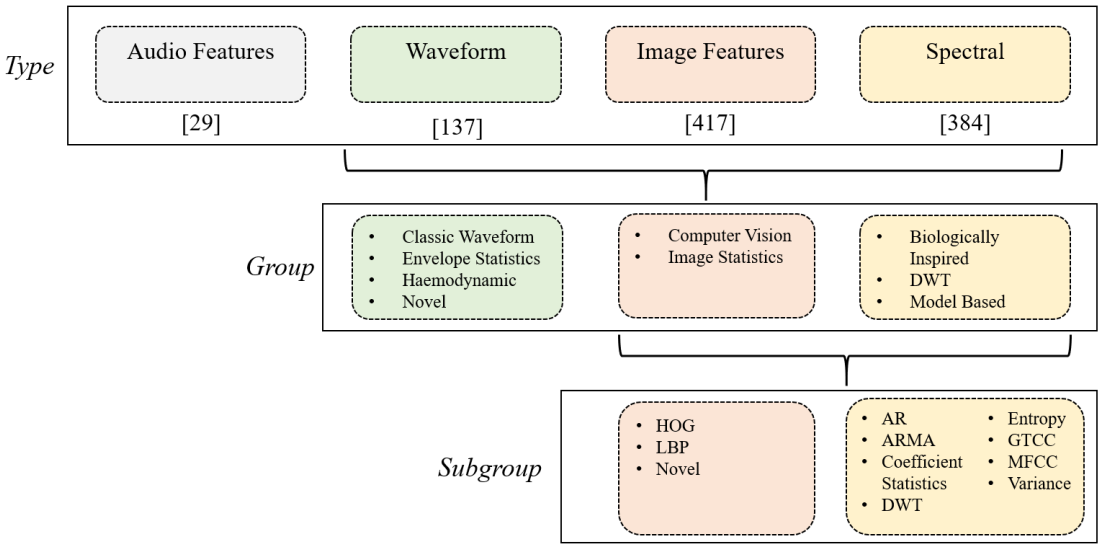


Figure 7.2: Overview of feature types, groups and subgroups used to categorise features

7.3.1 Haemodynamic Metrics and Envelope Statistics

These features consist of basic haemodynamic measurements and envelope statistics. These are derived using the MFE, and associated timing information.

Basic flow measurements such as peak velocity and time-averaged maximum velocity have been used in several Doppler classification tasks, and have been shown to be more discriminatory than other conventional features [216, 217].

Furthermore, when locating the IPP haemodynamic metrics such as peak velocity and CO can help guide data acquisition (Section 4.2.2). CO is derived using stroke volume, which is dependent upon the CSA of the interrogated lumen. This area is estimated using a nomogram, which predicts the outflow tract diameter using the patients height [86]. Related haemodynamic metrics and their uses are discussed in Section 2.3.

Considering the above, the features listed in Table 7.1 have been included in this study. As detailed, the majority of features have been derived using the OMM and MSNSI MFE. Chapter 5 illustrated that MFE estimation methods can function very differently depending upon the analysed data. Different methods are employed during feature selection, providing an investigation into their differences and respective value for pattern recognition. Specific details regarding how each feature is calculated are provided in Appendix

Table 7.1: Haemodynamic features extracted

Statistic Number	Feature Name[1]
1	Heart Rate
2	Ejection Time Percentage
3	Flow Time
4	Velocity Time Integral (VTI)
5	Peak Systolic Velocity (PK)
6	VTI/PK
7	CO
8	CO Forced[2]

1 - Statistics 2 to 8 have been calculate using the OMM and MSNSI MFEs.
2 - Using MFE forced to have triangular systolic region, as detailed in [78].

F.

In total 15 haemodynamic metrics are extracted. These features have been categorised as belonging to the Waveform type, and use envelopes as their data input. Feature names reflect this by beginning with either '`W_CC_E_`' or '`W_SYS_E_`' depending on whether the whole envelope or the systolic portion of the envelope is used. Following this is '`EnvelopeName_H StatisticNumber`', where envelope name is either OMM or MSNSI, '`H`' indicates a haemodynamic statistic and the statistic number is that listed in Table 7.1. Peak systole velocity estimated using the OMM MFE, for example, is named `W_CC_E_OMM_H5`.

The haemodynamic features are all derived using information from the MFE. In addition to these physical quantities, basic statistics can be derived from the MFEs. These include metrics such as STD or gradients. A number of simple statistics have been derived using the OMM and MSNSI MFEs. These are listed in Table 7.2.

This produces 10 envelope statistics (giving a total of 25 features when combined with the haemodynamic features), the naming reference for these are `W_DataLength_E_DataType_S Statistic Number`, where `DataType` is either OMM or MSNSI, and the relative statistic number is listed in Table 7.2.

Table 7.2: Envelope statistics

Feature Name[1]	Statistic number
STD[2]	1
Rising Slope Gradient (RSG)	2
Falling Slope Gradient (FSG)	3
Ratio of RSG and FSG	4

1 - Where RSG denotes rising slope gradient, and FSG denotes falling slope gradient.
2 - Calculated for whole CC and systolic portion.

7.3.2 Traditional Features

Doppler ultrasound was developed as a diagnostic tool over 50 years ago, and revealed a new means of measuring vital haemodynamic metrics. Efforts to unlock further diagnostic capabilities of the technology led to a variety of features being developed. The applications investigated for these are discussed further in Section 3.4. This section presents a selection of these features that have included in the given work, and are referred to as Traditional Features (TF). These are largely simplistic one-dimensional features, which are derived from velocity waveforms and Doppler spectral content.

Many of these were developed to describe the shape of waveforms associated with different vessels. Predominately, they are based on height ratios between different parts of the MFE [176,338]. In addition to the heights of certain characteristics, time relationships related to them have also been used to derive features. Incorporation of time information allows accelerations or decelerations to be quantified [339], as well as features such as systolic decay time index, which is derived using only time information [160].

In addition to shape and the temporal information provided by the MFE, information extracted from the power spectrum may be of diagnostic importance [12]. In particular, it was found that diseased vessels exhibited a degree of spectral broadening [12]. This phenomenon is discussed in Section 2.4.3. In response, a number of indices which attempt to quantify spectral broadening have been described.

The applications and limitations of these traditional features are discussed further in Section 3.4. The estimation of such features is straight-forward, and they have been shown to exhibit diagnostic potential, for these reasons they have been included in the presented

work. These have been calculated using the OMM MFE, and are summarised in Table 7.3.

Table 7.3: Traditional features

TF number	Feature Name
1	Systolic Decay Time Index
2	Pulsatility Index
3	Pourcelots Resistance Index
4	Constant Flow Ratio
5	Height-Width Index
6	Relative Flow Index
7	Early Diastolic Notch Pulsatility Index
8	Acceleration
9	Deceleration
10	Curve Broadening Index
11	Trans-systolic Times
12 - 31	Spectral Broadening Index

The features listed in Table 7.3 are based upon waveform indices, timing information and spectral content. The most commonly used TF features are waveform indices (such as PI). To reflect this, TF features have been categorised as belonging to the Waveform type (and subgroup Classic Waveform), and are referred to as `W_CC_E_TF` followed by the relevant TF number listed in Table 7.3.

The spectral broadening associated with a Doppler signal is related to both the insonation angle and interrogated blood velocities. When locating the IPP, an operator will move the ultrasound beam through different scanning planes. In the given work CW ultrasound is used, this means that aiming the probe inevitably results in different blood flows with individual insonation angles and blood velocities being captured. Therefore, the scanning process will likely give rise to a range of spectral broadening magnitudes. With respect to the aortic valve, once the IPP is located the ultrasound beam should be close to parallel with the direction of blood flow. Off-angles may capture flow from the aorta, but at an increased insonation angle and, therefore, with increased levels of spectral broadening. Considering these observations, metrics which quantify spectral broadening

could be valuable for the given work. This quantity is not straightforward to measure, which is highlighted by the large number of documented SBI features listed in Table 7.3, the associated equations for all features in this table are provided in Appendix F.

The effect of spectral broadening with respect to MFE estimation was discussed in Chapter 5. This work explored various means of MFE estimation, with varying performances with respect to this phenomenon. These observed differences led to several novel envelope features being explored, which aim to exploit these differences.

7.3.3 Novel Envelope Features

In addition to the traditional waveform indices described, a selection of novel envelope (NE) features have also been extracted. Hitherto, documented envelope features have been limited to those derived using one form of MFE estimation method. Furthermore, envelope extraction is often performed using secondary software or hardware, in which methods are either not disclosed or are unknown [193, 340–342] . As discussed in Chapter 5, various methods of MFE extraction have been developed and analysed, including the novel OMM method. The results presented in Chapter 5 illustrate how different MFE methods excel in different respects, for example, the OMM method was shown to be particularly adept at extracting stable envelopes in very noisy signals, whereas the MSNSI method produced MFEs with very low bias at high SNRs.

A variety of novel features were extracted from the Doppler signal. These attempt to capitalise upon the discussed differences, and provide further information for feedback models. The initial pursuit of these features was to provide further quantification of spectral broadening (i.e., in addition to the traditional features). This was investigated by extracting metrics using a combination of envelope types.

These combinations include the MSNSI method and an altered version of MSNSI that does not take into account spectral broadening (MSNSI_{noISB}). The MSNSI method attempts to overcome the effects of spectral broadening using equation 7.1, which is used to estimate the point at which the slope transitions from signal to noise on the IPS.

$$m(x) = m_S x + m_N(1 - x) \tag{7.1}$$

Where x is the fractional contribution to the slope, $m(x)$ is the slope at the signal-noise transition point, m_S is the slope of the signal, and m_N is the slope of the noise. The value

of x is set to 0.1 to accommodate for spectral broadening [114], MSNSI_{noISB} is extracted by setting x to 0. Similarly, the OMM method has no steps which directly address spectral broadening. A range of features were extracted using combinations of these MFEs. An example feature includes the mean difference of the MSNSI and MSNSI_{noISB}, which should increase in response to higher levels of spectral broadening.

MGM envelopes were also included to capture differences between the IPS methods. The envelope combinations used are listed in Table 7.4. The statistics extracted using these combinations are listed in Table 7.5, this includes details about how each feature is derived with respect to the two envelopes of interest (env_1 and env_2). In addition to the MFEs, the median frequency (MF) envelope has also been used to extract further features. This is calculated using each time point of the power spectral density estimate formed when extracting the OMM MFE.

Table 7.4: Envelope combinations used to extract novel envelope statistics

Combination Number	Envelope 1	Envelope 2
1	OMM	MSNSI
2	OMM	MF
3	MSNSI	MF
4	MSNSI	MGM
5	MGM	MF
6	MSNSI	MSNSI_{noISB}
7	MSNSI_{noISB}	MF
8[1]	OMM	OMM_{ideal}
9[1]	OMM	OMM_{seq}
10[1]	OMM_{seq}	OMM_{ideal}

1 - Limited to correlation statistics.

In total, 81 novel envelope statistics are extracted. These features have been categorised as belonging to the Waveform type, are all extracted from whole cardiac cycles and use envelopes as their data input. To reflect this, corresponding feature names begin with '`W_CC_E_NE`' followed by a reference number. The feature reference number is '`Statistic Number_Combination Number`', as listed in Tables 7.4 and 7.5, respectively.

Table 7.5: Novel envelope statistics derived using envelope combinations

Statistic Number	Name	Equation
1	Mean Difference ($Mean_{diff}$)	Mean(env_1 - env_2)
2	Mean Ratio ($Mean_{ratio}$)	Mean(env_1)/Mean(env_2)
3	Maximum Difference	Max(env_1) - Max(env_2)
4	Maximum Ratio (Max_{ratio})	Max(env_1)/Max(env_2)
5	Correlation	Corr(env_1, env_2)
6	STD Ratio (STD_{ratio})	std(env_1)/std(env_2)
7[1]	Normalised STD_{ratio}	STD_{ratio}/max(env_1, env_2)
8[1]	STD Difference (STD_{diff})	std(env_1) - std(env_2)
9[1]	Normalised STD_{diff}	STD_{diff}/max(env_1, env_2)
10[1]	Normalised $Mean_{ratio}$	$Mean_{ratio}$/max(env_1, env_2)
11[1]	Normalised $Mean_{diff}$	$Mean_{diff}$/max(env_1, env_2)
12[1]	Normalised Max_{ratio}	Max_{ratio}/max(env_1, env_2)

1 - Limited to combinations related to spectral broadening differences (1 and 6).

7.3.4 Image Features

The Doppler profile and spectrum are the primary sources of feedback that guide operators during a scan. The features described thus far include those that reflect characteristics of the Doppler profile, such as acceleration and deceleration during systole (Table 7.3). This section details features extracted directly from an image of the Doppler spectrum. These features include ones representative of spectral feedback, as presented in the criteria described in Section 4.2.3, as well as more general descriptors common to image-processing tasks.

As with audio classification, image classification is a very broad field with many applications. Section 3.6.2 presented the highly popular computer vision features LBP and HOGs. In the presented work, LBP and HOG features have been extracted from OMM images.

LBP features have been used to perform classifications using spectral images of clinical audio, including heart and lung sounds [269, 276]. These examples used 8-neighbouring pixels, and a radius of 1; resulting in 59 LBP features. These parameters have been

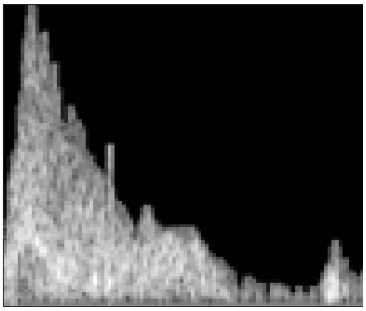


Figure 7.3: Example cardiac cycle image used to extract HOG features

employed for the given work. HOG features have been used in a similar manner with spectral images for audio classification [283–285]. These works used varying cell sizes (16 $\times$ 16 to 32 $\times$ 32) and resized spectral images, this allowed features to be independent of signal length and sample frequency [283]. In the presented work, images were rescaled to a size of 100 $\times$ 100. It was found that images with information limited to the signal portion gave improved results (i.e., the region reflective of blood velocity). These were formed by applying a mask to the images that set pixels outside of the envelope to one, an example image is displayed in Figure 7.3. This was performed using the OMM envelope and resulted in 324 HOG features. LBP and HOG features are named `LBP` and `HOG` respectively, followed by `_CC_I_F` and feature number. Further low-level details regarding the extraction of these features are provided in Appendix G.

Table 7.6: Image statistics

Statistic Number	Statistic Name
1	Contrast
2	Correlation
3	Energy
4	Homogeneity
5[1]	Pixel Range
6[2]	Row Pixel Range
7[2]	Column Pixel Range

1 - Applicable to image portions 1 to 3.
2 - Applicable to image portions 4 and 5.

Table 7.7: Image portions

IP Number	Image Portion	Criteria Characteristics
1	Whole[1]	Global aspects such as noise
2	Envelope Boundary (EB)	MFE boundary
3	Peak EB	MFE peak boundary
4	Peak	Systolic peak definition
5	Base	Systolic base definition

1 - Statistics extracted for whole cardiac cycle, and systolic portion.

A range of more intuitive statistics can also be extracted from images, for example, contrast and homogeneity. Table 7.6 details a range of image statistics extracted in the presented work. These are extracted from both OMM images and secondary images formed from these.

To reflect the Fremantle criteria described within Section 4.2.3, images include those cropped from specific regions of the OMM image. For example, one aspect considered by the Fremantle criteria is the definition of the systolic base, in light of this, statistics have been extracted from an image restricted to this region. This region, and the others listed in Table 7.7 were defined using the OMM MFE; these procedures are given in Appendix H, and correspond to 32 features. Table 7.7 provides details with respect to how each image portion (IP) reflect aspects of the Fremantle criteria. These novel image statistics (IMS) have a naming convention `IMS_DataLength_InputData_`, followed by `IP` with the IP number listed in Table 7.7, and `_S` with statistic number listed in Table 7.6.

The identified peak and base regions were also used to extract two SNR features using the STFT power spectral density estimates. These are listed in Table 7.8.

Table 7.8: Image region SNR features

Feature Name	Feature Equation[1]
`IMS_CC_I_SNR_BASE`	$10log_{10}(P_{sb}/P_n)$
`IMS_CC_I_SNR_PEAK`	$10log_{10}(P_{sp}/P_n)$

1 - Where P_{sb} and P_{sp} is the average power contained within the identified base and peak region respectively, and P_n is the average power in bins greater than the OMM MFE.

7.3.5 DWT Features

An experienced sonographer relies upon both visual and auditory feedback when performing examinations (Section 4.2.2). The features described thus far have been predominantly based on envelopes and visual representations of the Doppler signal. In this regard, they aim to capture, in part, visual information interpreted by an operator.

The remaining features are largely derived from audio directly via different means of spectral analysis. As such, they are more comparable to the auditory information interpreted by an operator during an examination. These features include those previously used for disease classification of the Doppler signal, as well as features implemented in other areas involved with audio classification, such as speech recognition.

Different approaches can be taken to analyse the spectral content of audio signals. Section 3.6.1.2 introduced the DWT technique, and its prior application in classification tasks, including Doppler ultrasound. This technique enables multi-resolution analysis of Doppler signals, and subsequently, various statistics to be acquired.

DWT features were extracted from the audio in a manner previously shown to be useful for Doppler arterial classification tasks [205]. Within this approach, the audio is segmented into individual cardiac cycles and is then decomposed into seven levels using a Daubechies one wavelet (db1). This produces seven detail coefficient (D_{Ln}) arrays of varying lengths, where Ln refers to the associated level number. The dimensionality of each array is reduced by calculating four coefficient statistics (CS): the mean of absolute values, maximum of absolute values, average power and STD. The extracted DWT audio features are summarised in Table 7.9.

Table 7.9: DWT features extracted from audio and envelopes

Group	Number of Levels	Statistic Number	Statistic Equation
DWT	1 to 7	1	Mean($\lvert D_{Ln} \rvert$)
		2	Max($\lvert D_{Ln} \rvert$)
		3	Mean(D_{Ln}^2)
		4	STD(D_{Ln})

The features detailed in Table 7.9 were extracted using audio from complete cardiac cycles, as done within previous works. The previous sections described a range of enve-

lope based features. These have not included spectral analysis of the waveform itself. To investigate whether such information could be beneficial, the features in Table 7.9 were also extracted from OMM and MSNSI MFEs. This resulted in a total of 84 features. The corresponding feature names begin within '`Group_DataLength_InputData`', and are then followed by '$\_D_{Ln}$`_StatisticNumber`' for audio and '`_EnvelopeName_`D_{Ln}`_StatisticNumber`' for envelopes. The first statistic, for the first level detail coefficients and using the OMM MFE, is therefore called `DWT_CC_E_OMM_D1_1`.

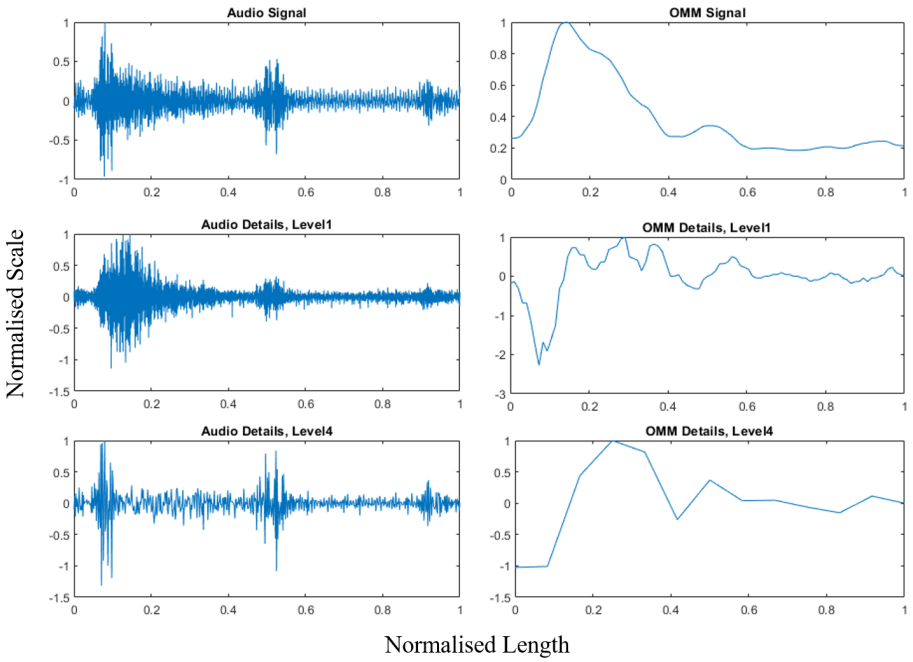


Figure 7.4: Example of original OMM MFE and audio signals for one cardiac cycle, and corresponding DWT detail coefficients D_1 and D_4.

DWT features have been used in countless disciplines for analysis and classification tasks. With respect to Doppler ultrasound, many of these have amassed substantially more research. These works can be drawn from to guide the extraction of further promising DWT related features. ECG analysis is a prime example of this and is an area with similar characteristics to Doppler ultrasound signals. As with Doppler ultrasound, ECG signals can be used to facilitate the diagnoses of heart diseases. The associated high mortality rates of which have driven large amounts of research into methods of classification and

feature types [343]. These have included features similar to those listed in Table 7.9 [344].

Li and Zhou recently demonstrated improved ECG classification accuracies using wavelet packet decomposition (WPD) [343]. This technique also decomposes the detail coefficients, which preserves high-frequency components [343]. Following WPD, Sharron entropy (SE) values were calculated.

The ECG and MFE signal share a number of similarities, for example, both have repeating waveforms whose shape and lengths are determined by processes within the heart. WPD Sharron entropy features were extracted from MFEs and Doppler audio to assess whether they are useful for the given work.

As with prior analysis of DWT, Doppler ultrasound features [206], it was found that Daubechies wavelets resulted in the highest classification accuracies for ECG signals [343]. A db1 wavelet has been used in the presented work. The extracted features are summarised in Table 7.10. These correspond to 80 features in total, whose naming convention is the same as the above DWT features, however, replaces statistic number with `SE` followed by SE number.

A further extension of the DWT is the maximal overlap discreet wavelet transform (MODWT). As with WPD, features derived using MODWT have been shown to be good descriptors for ECG signals [345, 346]. The MODWT produces wavelet coefficients whose lengths are the same as those in the original observation. This allows alignment of coefficients at each level with the original series; therefore providing direct comparisons between the two [347]. This is particularly useful for analyses of variance, which is inherently more restricted for DWT values [348]. The redundancy of MODWT coefficient also increases the effective degrees of freedom, which further strengthens their application to variance analysis [347]. Capitalising upon these characteristics, multiscale wavelet variances (V) have previously been used to classify ECG signals [345]. This technique has been included in the presented study, these features are summarised in Table 7.11. These correspond

Table 7.10: WPD entropy features

Group	Level	Input Data	SE Number
DWT	4	Audio[1], Envelopes[2]	1 to 16

1 - Extracted from systole and complete cardiac cycles.
2 - Extracted using OMM, MSNSI and MF envelopes.

to 44 features in total, whose naming convention is the same as the above DWT features (Table 7.9), however, replaces statistic number with `V` followed by V number. Further details regarding the extraction of all DWT and related features are included in Appendix I.

Works investigating ECG classification has demonstrated that combining DWT with parametric model-based features can result in highly accurate classification models [349]. Such features will now be described.

Table 7.11: MODWT variance features

Group	Input Data	Data Length	V Number[1]
DWT	Audio	Systole	1 to 9
	Audio	Cardiac Cycle	1 to 11
	Envelopes[2]	Cardiac Cycle	1 to 8

1 - Variable for processed data, as determined by data length.
2 - Extracted using OMM, MSNSI and MF envelopes.

7.3.6 Parametric Model Features

Parametric spectral estimation provides another means of analysing signals. The associated coefficients of AR models, of order four, have been shown to be valuable features for describing ECG waveforms [349]; this work classified waveforms associated with singular cardiac cycles. The AR coefficients of a given signal can be estimated using different approaches, the Burg method is an efficient example which has previously been implemented for analysis of Doppler ultrasound [350]. Given the demonstrated success and similarities between ECG and velocity envelopes used in the given work, the same techniques have been implemented to acquire AR coefficient features. The extracted features are listed in Table 7.12. These correspond to a total of 20 features, whose naming convention start with '`AR_DataLength_InputData_C`' , and'`AR_DataLength_InputData_EnvelopeName_C`' for audio and envelopes respectively, and end with AR coefficient number.

AR models have also been used to estimate the power spectral density of Doppler audio signals [194,207,350], these works used models of order 10. Spectral content derived using AR models have been implemented in the given work. As performed in previous

Table 7.12: AR coefficient features

Subgroup	Input Data	Data Length	Coefficient Number
AR	Envelopes[1]	Cardiac Cycle	1 to 4
	Audio	Cardiac Cycle and Systole	1 to 4

1 - Extracted using OMM, MSNSI and MF envelopes.

examples [205, 208, 351], the features are reduced using the same statistics applied to the DWT decompositions.

In addition to these, AR and ARMA power spectral densities and associated statistics have also been extracted. ARMA values were included to reflect reported advantages of using such methods, as discussed in Section 3.6.1. The ARMA coefficients are found using the ARMAsel approach [352], again using a model order of 10.

These features have been derived from complete cardiac cycles, and use the same naming convention as the above AR coefficients, however, `C` and coefficient number are replaced by `S` and statistic number. The extracted features are summarised in Table 7.13, and correspond to 32 features.

Using the ARMA model, AR and MA parameter autocorrelations were also calculated, producing an additional 40 features. The corresponding feature names are the same as those in Table 7.13, however, `S` and statistic number are replaced by `AC` and autocorrelation number. In total, 92 model based features are extracted, further details regarding the extraction of these are provided in Appendix J.

Table 7.13: AR PSD features

Subgroup	Input Data	Statistic Number	Statistic Equation
AR, ARMA	Audio, Envelopes[1]	1	Mean($\|PSD\|$)
		2	Max($\|PSD\|$)
		3	Mean(PSD^2)
		4	STD(PSD)

1 - Extracted using OMM, MSNSI and MF envelopes.

7.3.7 MFCC and GTCC Features

MFCC features have become a staple of audio classification and analysis, and provide a further means of analysing a signals spectral content. The extraction of MFCC and GTCC features, and their widespread applications were introduced in Section 3.6.1.3.

These features mimic aspects of the human auditory system, and as such are categorised as belonging to the 'Biologically Inspired' group illustrated in Figure 7.2. These features, particularly MFCCs, have been widely and successfully used in speech recognition [353]. Spectral analysis of Doppler signals have not previously included features that specifically consider how humans hear them. Considering the success of such features in other applications, and the powerful role Doppler audio can play in guiding probe position for an operator, MFCC and GTCC features have been included in the given work. These features are summarised in Table 7.14.

Table 7.14 details the extracted features and their reference feature numbers. The feature numbers 0 to 12 correspond to the associated coefficient number, and feature number 13 corresponds to log energy. This results in 84 features, the naming convention of which are '`Subgroup_DataLength_InputData_FeatureNumber`'.

Table 7.14: MFCC and GTCC features

Subgroup	Data Length	Feature Number[1]
MFCC	CC	0 to 13
	SYS	0 to 13
GTCC[2]	CC	0 to 13
	SYS	0 to 13

1 - 0 to 12 correspond to coefficient number, and 13 corresponds to log energy.
2 - Extracted using both time and frequency domain filtering, using naming convention `T` followed by feature number to indicate time, and `F` followed by feature number to indicate frequency.

7.3.8 Audio Features

This section presents additional features that are extracted from the Doppler audio. These include established audio descriptors that have been used in several areas of audio analysis, including phsyco-acoustic analysis, musical instrument classification and speech recogni-

tion [247,354]. The chosen features are largely described within the 'Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online' project [354,355]. The goal of which was to design extractors capable of unlocking high-level information from audio signals, from which new applications could arise. The naming convention of these audio features (AF) are '`AF_DataLength_InputData`' followed by the corresponding statistic number given in Table 7.15.

Table 7.15: Audio features

Statistic Number	Description	Feature Name[1]
1	Based upon human hearing, i.e., perceptual	Loudness
2	Based upon spectral shape, using STFT[1]	Spectral Centroid
3		Spectral Decrease
4		Spectral Entropy
5		Spectral Flattness
6		Spectral Flux
7		Spectral Kurtosis
8		Spectral Rolloff Point
9		Spectral Skewness
10		Spectral Slope
11		Spectral Spread
12		Pitch
13	Based upon sinusoidal harmonic modelling[1]	Harmonic Ratio

1 - Extraction and features details given in [354].

Features detailed in Table 7.15 have been extracted from systole and complete cardiac cycles, and correspond 26 features in total. Further details regarding these features are supplied in Appendix K. Also included within this feature type are three SNR features; two of which are derived using equation 5.2, and a third novel measure using audio amplitude. These SNR features are detailed in Table 7.16.

Table 7.16: AF SNR features

Feature Name	Description
AF_CC_E_OMM_SNR	Spectral SNR, estimated using OMM MFE
AF_CC_E_MSNSI_SNR	Spectral SNR, estimated using MSNSI MFE
AF_CC_A_SNR	RMS audio amplitude SNR[1]

1 - SNR measure from audio signal, where diastolic and systolic regions are treated as noise and signal respectively.

7.4 Labelling Data

Labelled datasets form the basis of most predictive models, allowing them to learn patterns associated with certain outcomes. When exposed to unseen observations, these trained models can then make predictions. Given a set of features with known origins it is often straightforward to assign corresponding labels. Models that incorporate information from Doppler signals have been discussed in Section 3.3, these tend to use two goal categories, for example, abnormal or normal arterial signals [356], or benign or malignant adnexal tumours [210]. The 'ground truth' in these situations is known, allowing measurements from each category to be clearly labelled.

This approach is used in Chapter 8, where observations are labelled in terms of probe position. However, in this chapter the goal is to design a model that provides a continuous form of feedback. This is to indicate whether probe adjustments are beneficial. Ideally, measurements that correspond to the probe being at the IPP would consistently share common traits. In certain conditions this seems likely – for example when measurements are limited to a single individual – in which small deviations from the IPP can be visually obvious. However, as discussed in Section 4.4 this is not always the case, and in particular, differences in Doppler profile appearance across participants can be extensive.

This inter- and intrapatient variability is a challenging aspect surrounding the work described in this thesis, and is a contributing factor to the labelling approach used. In this work the regression model has been trained using an 'ideal response', this assigns a value of one to IPP observations and a value of zero to off-angle observations. Several approaches were considered, these will now be briefly discussed, and the reasoning for using an ideal response will be presented.

7.4.1 Scan Quality Labels

Subjective scoring criteria have been developed to assess whether a measured scan using the USCOM is acceptable or not, this is discussed in Section 4.2.3 . Using these criteria, points are awarded in reference to characteristics of the Doppler profile, for example, how well defined the peak of systole is. In turn, these points can be summed to give an overall score. This scan score provides a subjective assessment of a given measurements quality.

An automatic method of replicating these scores was considered as a means of providing continuous feedback. It is intuitive that higher scores correspond to better measurements. However, as illustrated in Section 4.4.2, the off-angle Doppler profiles do not necessarily score worse than those from the IPP, particularly when measurements across participants are considered. The viability of this approach was investigated by comparing scored scans with their corresponding probe positions.

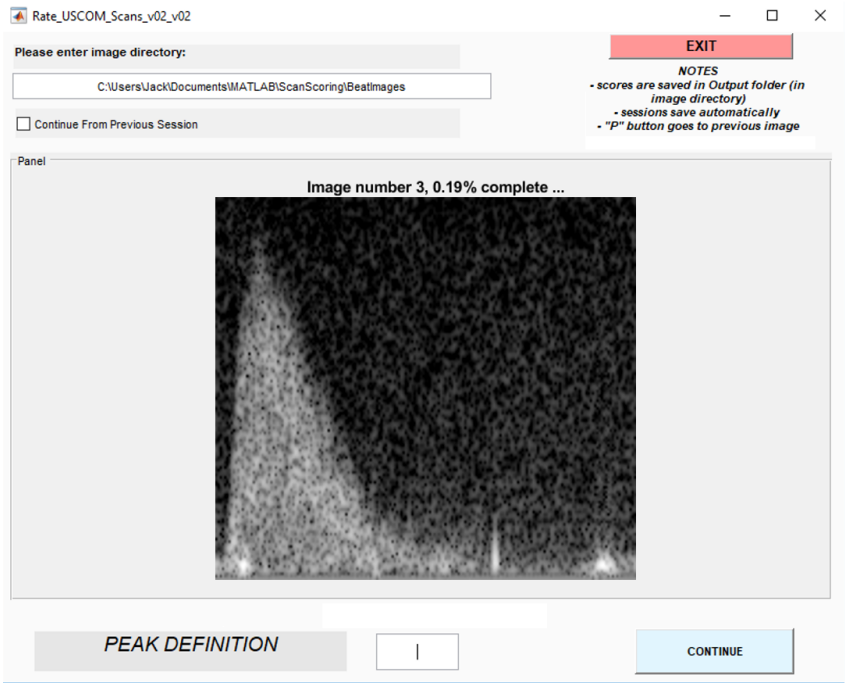


Figure 7.5: GUI used to score beats using Fremantle criteria

To perform this evaluation, the collected Doppler data needed to be scored. Scores were assigned to spectrogram images of individual beats; these were generated automatically

Table 7.17: Desirable features of dataset labels

Label Characteristic
• Highest values at IPP
• Lowest values at off-angles
• Low variability in IPP scores
• Low variability in off-angle scores

using the beat segmentation software described in Section 6. The Fremantle criteria is the most commonly applied method, and includes characteristics considered by alternative methods and used when aiming the USCOM [312, 313]. Considering this, the Fremantle criteria was used to score a sample of scan images. A random sample of approximately 25% of the collected data was used, which corresponded to 2,379 beats.

The scoring process is laborious, to ease this process a GUI was designed in Matlab. This is displayed in Figure 7.5. This also allowed the order of the images to be easily randomised, and for associated scores to be saved. The scans were scored three times, scans with score ranges greater than two were reinspected and re-scored. These were then averaged. The use of these scores as labels for the given task was investigated in terms of intra- and interpatient variability, in both respects the characteristics listed in Table 7.17 are desirable.

The overall distributions of Fremantle scores with respect to probe position are displayed within Figure 7.6. It can be seen from this, that measurements associated with

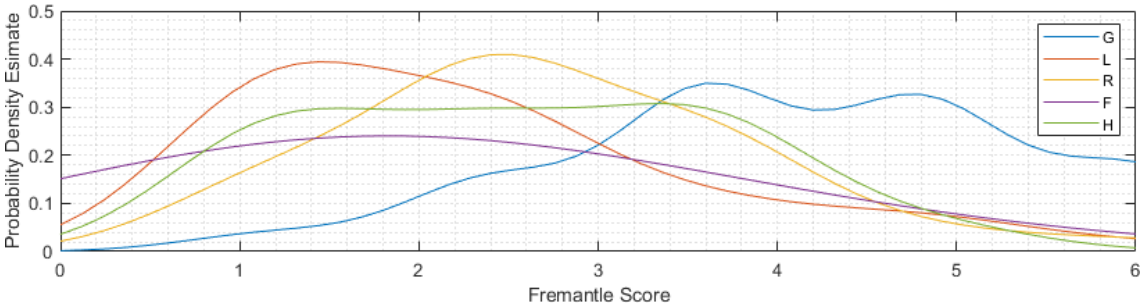


Figure 7.6: Distribution of Fremantle scores for individual probe positions, visualised using probability density estimate based on normal kernel function.

the IPP typically achieve a higher Fremantle score. However, there remains significant overlap with respect to the IPP and the remaining off-angle positions. This is illustrated further in Table 7.18, which show that the mean intrapatient scores for both the IPP and off-angles exhibit an STD of 0.9, and have mean values of 4.4 and 3, respectively.

With respect to the ideal response (which is discussed further in Section 7.4.3), the scores both had a Pearson and Spearman correlation of 0.65. This demonstrates that the labels diverge significantly from the characteristics listed in Table 7.17. These correlation statistics are discussed in Section 7.7.1, and are later used to evaluate the final regression model.

Considering the overlap present in the scores, observations cannot be definitively separated using them alone. Furthermore, predicted scores will include an error percentage, giving further ambiguity as to whether a scan is close to the IPP for an individual. As will be presented in Section 7.8.2, the feedback ultimately generated by the trained model is more highly correlated to the ideal response than the scores discussed here.

Table 7.18: Fremantle mean and STD statistics

Position Label[1]	Mean	Normalised Mean[2] (%)	STD
G	4.4	1.0	0.9
Off-angles	3.0	0.4	0.9

1 - Probe position labels as detailed in Section 4.3.
2 - Mean feedback with respect to mean IPP feedback.

The limitations of subjectively scored observations led to alternative labelling methods being sought. These include labels based on physical haemodynamic values.

7.4.2 Haemodynamic Based Labels

Following data collection, the mean values of certain metrics attributed to IPP observations can be found. A fundamental metric that can be measured using the USCOM is CO; this feature is included in the feature extraction process. A highly informative form of feedback would be the percentage of CO being received (with respect to what is acquired at the IPP). Using the mean CO value during these IPP measurements, this CO percentage value can be calculated. This percentage can then be used to construct a labelled data set.

This approach could provide several advantages. Firstly, it allows quantitative values to be attributed to observations, as opposed to subjectively scored quantities or discrete values. These scores are sensitive and provide participant-specific scaling. This could overcome, in part, the variability in IPP scan quality across participants. Secondly, it unlocks a portion of the dataset which would be impractical to attain labels for through subjective scores or when using labels limited to the IPP or other known angles (i.e., the data captured between defined probe position, such as when the IPP is being located following an off-angle position, can be used to generate labels).

The viability of such an approach, however, is dependent on how consistently a metric such as CO can be calculated across the measurements, and how intertwined the metric and probe position are. The Doppler profile, for example, at an incorrect position may appear to have a wide systolic portion, with low velocities. Despite this profile clearly being far from its ideal form, the CO could be similar to the correct value (due to the increased systolic width). Furthermore, the accuracy of the CO depends on how accurately the MFE has been extracted, and corresponding timing indices estimated. In practice for example, representative beats are selected using the devices screen, and in many cases, metrics calculated from certain beats will be rejected due to inaccuracies in either MFE or timing indices. Clearly, to both function in a real-time manner or to process large datasets, manually checking such inaccuracies is impractical.

In addition to these limitations, the haemodynamic values associated with the IPP can change during a set of measurements. This could be due to, for example, changes in heart rate or small variations in body position [317]. In the given work, haemodynamic percentages were not found to be reliable enough to form valid labels.

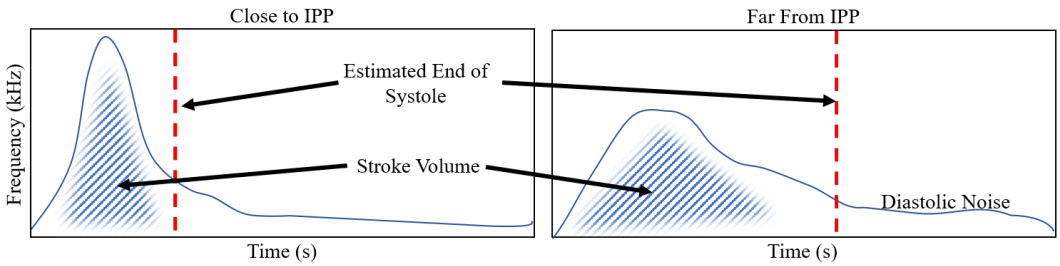


Figure 7.7: Illustration of how stroke volume or CO alone can be a weak indicator of probe position

7.4.3 Ideal Response Labels

Continuous feedback needs to clearly indicate when the probe approaches or leaves the IPP. A simplistic approach to generate labels to reflect this is to assign IPP measurements a value of 1, and measurements from off-angles a value of 0. An example of such labels are illustrated in Figure 7.8.

In contrast to the other approaches, this ensures IPP measurements are given a common label. This allows the approach to be insensitive to variations in scan quality at the IPP across participants. Furthermore, it ensures off-angles are consistently reflected by the labels. This forces the trained model to try and overcome interpatient variability.

With respect to Table 7.17, these labels would fulfil all the desired label characteristics, and as such are referred to as the ideal response. The labels are therefore binary, where all off angles are all considered to be "equally bad", or at an equal distance away from the IPP. When used to train a regression model, positive movements of the probe are reflected by an increase in value, indicating the distance between the current position and the IPP has been reduced. For the given work, this ideal response has been used to form regression models.

An example of a binary ideal response is given in Figure 7.8;

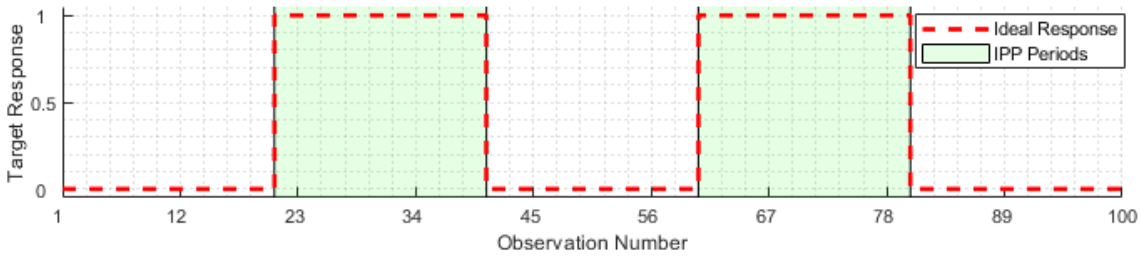


Figure 7.8: Example of ideal response labels, where green regions correspond to observations from the IPP.

7.5 Feature Matrix Conditioning

The extracted features are combined to give a feature matrix (FM), where rows correspond to observations and columns to features. Given the breadth of feature types and the number of observations, this FM is large containing close to 16 million values. Prior

to using it for training and evaluating models, the FM is first processed. This ensures observations with nonsensical data such as infinite or NaN terms are removed. In addition to this, poor observations are removed. These are identified using beat images, and are summarised in Section 4.4. These can arise, for example, due to momentary drops in signal quality or an incorrectly segmented beat.

Finally, the FM is normalised. This standardisation ensures each feature has zero-mean and unit variance. This is a common step for pattern recognition and is performed due to many machine learning techniques being sensitive to the scale of features [357]. This process converts each feature to a STD from the mean, removing any previously associated units. A given feature vector, F, is normalised using equation 7.2.

$$F' = \frac{F - \bar{F}}{\sigma} \tag{7.2}$$

Where F' is the normalised feature, $\bar{F}$ is the mean feature value and σ is the STD of F.

7.6 Feature Selection

The design of classification and regression models, particularly in contemporary applications, are performed using datasets with an abundance of information. Many algorithms have been developed to identify important features contained within these datasets [358]. The dataset used in the presented work includes numerous and varied features, many of which have not been previously used to classify Doppler signals. Rather, their inclusion has been influenced by machine learning and pattern recognition research in other areas. Considering this, a thorough and effective feature selection method is needed.

The presence of redundant or irrelevant features in a dataset can detrimentally affect the accuracy of trained models [359]. The process of feature selection extracts a subset of features from the overall dataset. This reduces the dimensionality of the data and can have a number of benefits: predictive accuracy can be improved, the amount of data needed to train a model can be reduced, and execution time can be decreased [360]. Software which use predictive models in a live setting can also benefit computationally, by reducing the cost of feature extraction and improving the overall throughput [361]. This is of particular importance in the presented work, where a high refresh rate in terms

of feedback is desirable. In addition to these computational and performance-related advantages, feature selection can allow the underlying success of a model to be more comprehensible to humans [360, 362]. This is because concepts that employ many features are hard to understand.

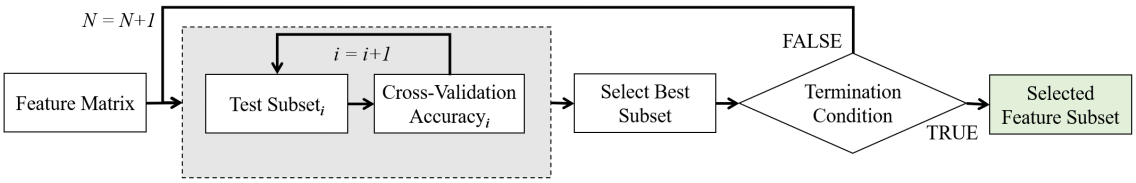


Figure 7.9: Block diagram illustrating Sequential forward feature selection, where i denotes subset, and N subset size.

The goal of a feature selection algorithm is to create a subset of features, within which each feature is a good predictor, either by themselves or when combined with another selected feature. Additionally, a selected feature should not be redundant with respect to the other selected features [362]. For example, a fully correlated feature could be a good predictor, however, would be superfluous when combined with its correlated counterpart.

The task of identifying useful features can be challenging. The way a model relates features to an output are often complex. In such cases, it is near impossible to decrypt the role of a feature and its contribution to an outcome [361]. There have been many proposed methods of feature selection, and further developments are on-going and prevalent. Despite this, research continues to indicate that no single method can be considered universally superior [363], and selecting a model from this growing pool of algorithms is difficult [364]. Feature selection methods can largely be considered as belonging to one of two groups: filter methods and wrapper methods [361,365]. Although a third type, embedded methods, also exists, examples include Lasso [366] or Elastic Net [367].

Filter methods tend to be relatively simple and quick to implement, with searches typically performed once and features considered separately [361]. This approach is often effective at capturing large trends, however, features are selected using the methods objection function, which applies a statistical significance to an inspected feature. This means the requirements of a model are not considered, and so features that fulfil the selection criteria do not necessarily improve predictive accuracy [361].

Wrapper methods, instead 'wrap' the feature selection process around a selected model.

An iterative process is performed, during which subsets of features are repeatedly supplied to the model. The corresponding performances are then used to guide which features are included in the final selection [361]. The resource demands of using the modelling algorithm result in a more computationally expensive operation, and an end feature subset which is biased [358]. This can be further exacerbated by the choice of model, for example in the case of SVMs and neural networks, which can be inherently computationally expensive. However, wrapper methods have been empirically proven to identify the best-performing feature subsets [363, 368].

The two groups of feature selection methods clearly have advantages and disadvantages. An alternative approach is to combine them; forming a hybrid method [364]. Many hybrid methods have been proposed, these seek to combine the best properties of filter and wrapper methods. Using a filter method, the feature set dimension space is first reduced [358]. A wrapper method can then be employed to perform the final feature selection. Considering the feature space being searched during feature selection has size 2^n, where n is the number of features [369], any reduction in dimensionality can be hugely beneficial for more expensive methods. This can give results with the high accuracy exhibited by wrapper methods, with improved efficiency. Furthermore, practically any combination can be used to design a hybrid method [358]. This approach has been used in the presented work.

Several filter methods identify features as being redundant in cases where they are highly correlated. However, unless these features are identical, their inclusion can still provide useful information [368]. RelieF is a popular filter method [370, 371], which is considered to be one of the most successful for measuring feature quality [372], and has demonstrated good performance [363]. This method does not inherently remove correlated features, however, additional steps can be incorporated to remove features deemed redundant.

This approach was adopted by Bins and Draper, who used a k-means algorithm to reject features with correlations greater than 0.97 [373]. In this example, redundant features were removed following the RelieF filter method. This gave a reduced feature matrix, from which a final feature subset was selected using a wrapper method: sequential feature selection [373].

Different variants of sequential feature selection have been proposed, for example, backward elimination [374], or forward selection [375]. Sequential forward selection begins

with an empty set, each feature is then tested individually using the objective function, and the best feature is retained. This process is then repeated to find the next best feature to accompany the first, and so on [376], this is illustrated in Figure 7.9. In sequential backward selection, the method begins with a subset containing all features, and removes each individually, to find the feature which least negatively affects performance. These methods typically assess feature additions using cross-validation [377]. Studies have shown that neither methods consistently outperform [369], however, backward elimination is less feasible with high numbers of features [373].

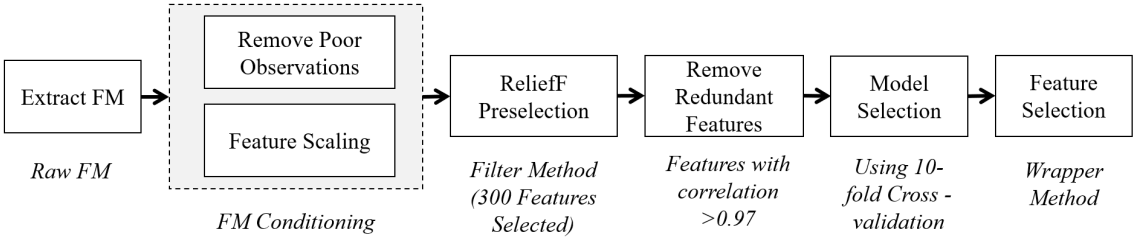


Figure 7.10: High-level block diagram displaying feature selection methods

In the presented work, a hybrid method has been used that combines the ReliefF filter method with correlation feature reduction, and sequential forward feature selection. The ReliefF filter performs feature preselection, retaining 300 features, and features with correlations greater than 0.97 are removed. A variety of model types and predictive goals have been investigated within this work. The described hybrid method provided an accurate means of ranking the extracted features, which was still functional. This is illustrated in Figure 7.10, as illustrated within this, the model type is selected during this process. This is performed using Matlab's Machine learning and Deep learning toolboxes, which enables a variety of models to be trained and compared. This was performed using 10-fold cross-validation, the models types evaluated are illustrated in Table 7.19. This table details those investigated for both the regression and classifications tasks (Chapter 8).

Table 7.19: Regression and classification models investigated

Tree	SVM	Ensemble	Discriminant[1]	KNN[1]	GPR[2]*	Linear Regression[2]
Fine	Linear	Boosted Trees	Linear	Fine	Squared Exponential	Linear
Medium	Quadratic	Bagged Trees	Quadratic	Meduim	Matern 5/2	Interactions
Coarse	Cubic	Subspace Discriminant[1]		Coarse	Exponential	Robust
	Fine Gaussian	Subspace KNN[1]		Cosine	Rational Quadratic	Stepwise
	Medium Gaussian	RUSBoosted Trees[1]		Cubic		
	Coarse Gaussian			Weighted		

1 - Applies to classification only
2 - Applies to regression only
* - Gaussian Process Regression

7.7 Evaluation Methods

This section details the metrics and methods used to evaluate the performance of the trained regression model. This includes cross-validation, which is implemented to give strict and fair evaluation of performance, by ensuring training and testing datasets do not share information from the same individual. Finally, a means of feedback averaging is described.

7.7.1 Regression Models

Previous works involving Doppler ultrasound pattern recognition tasks (discussed in Section 7.3) have classified measurements into distinct groups. Conversely, the models explored within this chapter are regression-based; producing continuous feedback. Considering this, alternative means of evaluation are required.

The goal of the continuous feedback is analogous in many respects to audio-visual multimedia quality assessments. These assessments are concerned with measuring the perceived quality of audio-visual signals, an important factor which can determine many processes for multimedia networks and applications [378]. Different features are used to predict perceived quality, in-turn allowing various aspects of multimedia services to be improved. This presents a similar challenge explored in this work, where an experienced operator uses a variety of visual cues to identify the IPP.

Typical indices used to assess the performance of these models include the Pearson

Linear Correlation Coefficient($PLCC$), the Spearman Rank Order Correlation Coefficient ($SROCC$), and root mean-square-error (RMSE) [379]. These allow the similarity between the predicted quality scores and ground-truth to be assessed. These metrics have been adopted in this work to aid model evaluation.

The $PLCC$ calculates the linear correlation between the predicted feedback and the ideal response, giving a measure of the trained models accuracy [380]. This is calculated using the paired ideal response and feedback at each observation, x_i and y_i respectively, and uses the mean values of these two data sets, $\bar{x}$ and $\bar{y}$. The $PLCC$ is calculated using equation 7.3:

$$PLCC = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}} \tag{7.3}$$

The $SROCC$ quantifies the correlation between feedback and ideal response, and assess how well their relationship can be described using a monotonic function [381]. This indicates the degree to which the feedback and relative magnitudes of the ideal response agree [380]. This is calculated using the rank of x_i and y_i, which are denoted by X_i and Y_i respectively, and the ordered arrays of these values, X' and Y'. $SROCC$ is calculated using equation 7.4.

$$SROCC = \frac{\sum(X_i - X')(Y_i - Y')}{\sqrt{\sum (X_i - X')^2}\sqrt{\sum (Y_i - Y')^2}} \tag{7.4}$$

In addition to these metrics, the mean values and STDs of feedback at specific probe positions are used to further analyse performance.

7.7.2 Cross-validation

In order to perform a fair and representative evaluation of model performances, the way data is divided and evaluated is paramount. Section 3.5.1 discussed a number of examples in which model evaluations had been carried out by separating the data into two groups, one for testing and one for training. In the presented work, a more thorough cross-validation approach is adopted.

Cross-validation provides a robust means of evaluating classification performance. The technique is very common and has been implemented for previous Doppler classification

tasks. These examples mostly have one set of features for a given participant, created using averaged data. In this case, cross-validation provides a good means of assessing a model's generalisability across patients.

The data used in this work has many observations for each participant. This as akin to many ECG classification tasks [382]. This field is abundant with research into classification modelling techniques. However, a notable limitation of much of this work is an adopted intrasubject validation approach [383]. That is to say, the testing and training groups share observations from the same participants. Given the variability in human physique and health, individual participants can exhibit highly unique signatures [383]. Training data sets containing samples of these can, therefore, give heightened measures of a models generalisability, i.e., this approach is not practical for realistic scenarios [343]. A meta-analysis comparing inter- and intrasubject model evaluations found substantially higher accuracies were reported for the intrasubject cases. This is particularly stark in five cases, where the two approaches were directly compared. Across these examples, the reported classification accuracy plummeted from an average of 92.2% to 73.5% [382].

The presented work exhibits these challenges, with the high intersubject variability being illustrated in Section 4.4. Considering this, cross-validation will be performed using an intersubject cross-validation approach. This is achieved using testing data that corresponds to individual participants. This is then repeated for each participant, these results are then combined for an overall evaluation.

This provides a very thorough and transparent means of assessing model performance and generalisability. Figure 7.11 displays a block-diagram illustrating the cross-validation process. Using this approach, the trained model was evaluated using data from 22 participants. This data, and the associated survey methodology, is summarised in Chapter 4. As discussed in Section 4.4.3, scan acquisition difficulty is also investigated using measurements collected by a highly experienced sonographer; the associated measurements for this analysis are from participants one to ten. In this chapter, this information is used to assess regression model performance in the context of scan difficulty.

7.7.3 Moving Average Predictions

Previous Doppler classification studies have extracted features from waveforms averaged across a large number of heartbeats. This can strengthen overall predictions and overcome

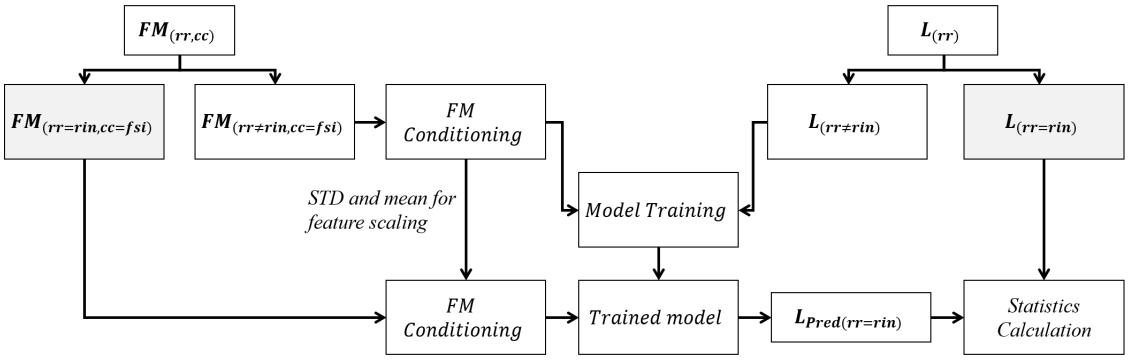


Figure 7.11: Blind testing method to acquire performance statistics for trained models, where rows (rr) in the FM and the Labels (L) correspond to observations, rin corresponds to the row index of a given participant, and cc corresponds to columns in the feature matrix which contain feature vectors.

inaccuracies introduced by unrepresentative beats or momentary noise. These classification tasks were concerned with differentiating unchanging distinct states in a participant, such as a disease. These tasks are also analysed post examination. These aspects allow many consecutive beats to be combined to form ensemble averages (e.g., 30 - 40 beats [193]).

In the presented work, feedback needs to be functional for dynamic probe positions and real-time use. This prevents large numbers of beats being averaged as the inherent delay would render feedback unusable. In addition to this, real-world applications must overcome variable heart rates, erroneous signals or noise, and the potential of missed beats.

Methods of prediction strengthening have been investigated in the presented work. Features are extracted from different data periods, they belong to many different types (i.e., they are not limited to waveform derived features) and include non-linear relationships. To reflect this, predictions from sequential observations are averaged, as opposed to averaging features or input data. This is performed using a time window, with a duration of t_{av}, which identifies a dynamic number observations (i.e., complete cardiac cycles). The corresponding features (F) are used to form predictions, which are then averaged. This process of combining predictions for a given cardiac cycle (n) is displayed in Figure 7.12.

A longer window duration strengthens predictions when the probe is stationary, or slow-moving. However, this also reduces sensitivity in response to more dynamic move-

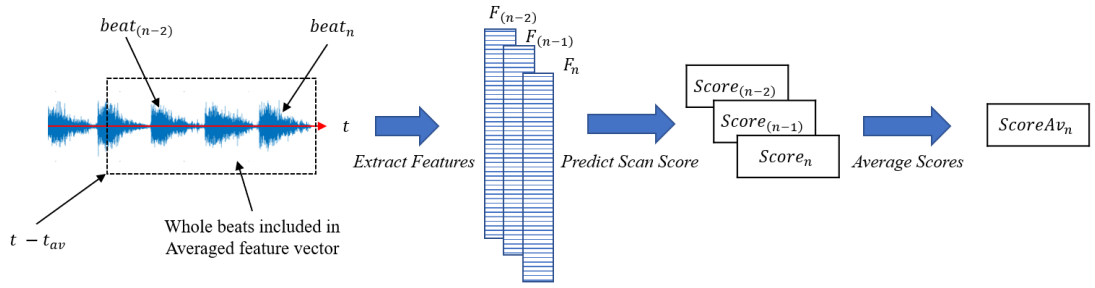


Figure 7.12: Method of combining predictions for regression models using moving time window

ments. This trade-off therefore needs to be considered when averaging predictions. In practice, the window length used could be an operator defined setting determined by their expertise, preference and difficulty of a given examination. In the presented work, a time window of 4 s has been used for all evaluations. This duration was found to strengthen predictions, without introducing significant response delay. For a resting heart rate of 70 bpm, this window duration results in four cardiac cycles being averaged. Results which have been averaged are referred to as moving average (MA) results.

7.8 Results

This section presents the feature subset identified through the feature selection process. This subset was then used to train a regression model, which was then evaluated using the procedure outlined in Figure 7.11. The calculated performance metrics will be presented.

7.8.1 Selected Features

The process outlined in Section 7.6 was used to identify a small subset of well-performing features. Sequential feature selection iteratively assessed increasingly large feature combinations. The loss associated with these enable an ideal subset size to be selected. This is not necessarily the subset size that gives the lowest loss, for example in cases where a very minor reduction in loss necessitates a vast increase in feature size. In such a case, real-time applications could be limited (due to the increased time required to extract features). Furthermore, reducing feature dimensionality can reduce model training time, a models

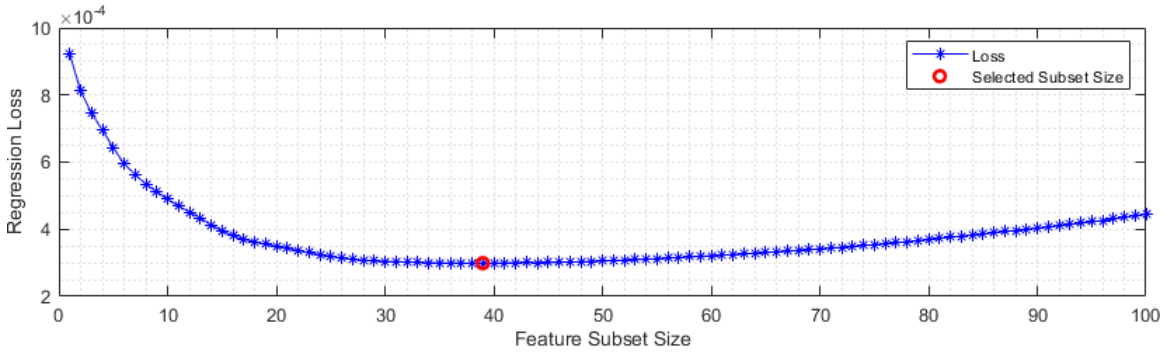


Figure 7.13: Feature subset loss

Table 7.20: Features selected for regression model

Rank	Type[1]	Name	Rank	Type	Name	Rank	Type	Name
1	SF	MFCC_SYS_A2	**14**	W	W_CC_E_NE5_10	**27**	SF	MFCC_SYS_A0
2	SF	DWT_CC_E_OMM_D3_4	**15**	W	W_CC_E_NE1_5	**28**	SF	AR_CC_A_C3
3	SF	AR_CC_E_MSNSI_S2	**16**	SF	ARMA_CC_A_AC9	**29**	SF	DWT_CC_A_D1_1
4	AF	AF_CC_E_OMM_SNR	**17**	SF	DWT_CC_A_SE7	**30**	SF	DWT_CC_E_OMM_D2_2
5	SF	AR_CC_A_S4	**18**	IF	LBP_CC_I_F7	**31**	SF	DWT_CC_A_D1_4
6	SF	MFCC_SYS_A3	**19**	W	W_CC_E_NE4_3	**32**	SF	AR_CC_E_OMM_S4
7	AF	AF_SYS_A1	**20**	SF	AR_CC_A_S1	**33**	SF	ARMA_CC_A_AC6
8	SF	GTCC_CC_A_F0	**21**	IF	IMS_CC_I_IP5_S7	**34**	SF	DWT_CC_E_OMM_D4_2
9	SF	DWT_CC_E_MSNSI_D6_4	**22**	SF	DWT_CC_A_SE9	**35**	SF	DWT_CC_E_OMM_D3_2
10	SF	AR_CC_A_C4	**23**	SF	DWT_CC_E_MSNSI_D4_4	**36**	IF	HOG_CC_I_F247
11	SF	AR_CC_A_S3	**24**	SF	AR_CC_A_S2	**37**	SF	DWT_CC_A_SE6
12	W	W_CC_E_NE7_7	**25**	SF	AR_CC_E_OMM_S2	**38**	SF	MFCC_SYS_A7
13	SF	AR_CC_E_OMM_S1	**26**	IF	IMS_CC_I_SNR_PEAK	**39**	SF	GTCC_SYS_A_F0

1 - Where SF denotes spectral features, AF denotes audio features, W denotes waveform features and IF denotes image features.

complexity and increase its generalisation [364]; given the variability in the training data this is particularly desirable. The regression loss associated with increasing subset size is displayed within Figure 7.13.

The calculated losses detailed in Figure 7.13 were used to select a feature subset that comprised of 39 features. These features are succinctly detailed in Table 7.20, further details are provided in Appendix M. It can be seen from this that the chosen features predominately belong to the spectral type, and are extracted from both audio and envelopes.

The highest-ranked feature is an MFCC feature extracted from the systolic portion of the audio. This is followed by a DWT decomposition statistic extracted from the OMM envelope, a related feature is also ranked ninth. This feature type has previously been extracted from Doppler audio signals. Only four image features were selected, the highest of which is an LBP feature at rank 18. These image features include two novel features, which are ranked 21st and 26th. No haemodynamic features were selected.

The features listed in Table 7.20 belong to a variety of different categories. Figure 7.14 displays a Sankey diagram, this provides a digestible breakdown of the various types of features selected. The feature selection process also identifies the model type which is used; the model types investigated are listed in Table 7.20. This process identified a coarse Gaussian SVM as the best performing model, high-level details regarding this model are included within Appendix L.

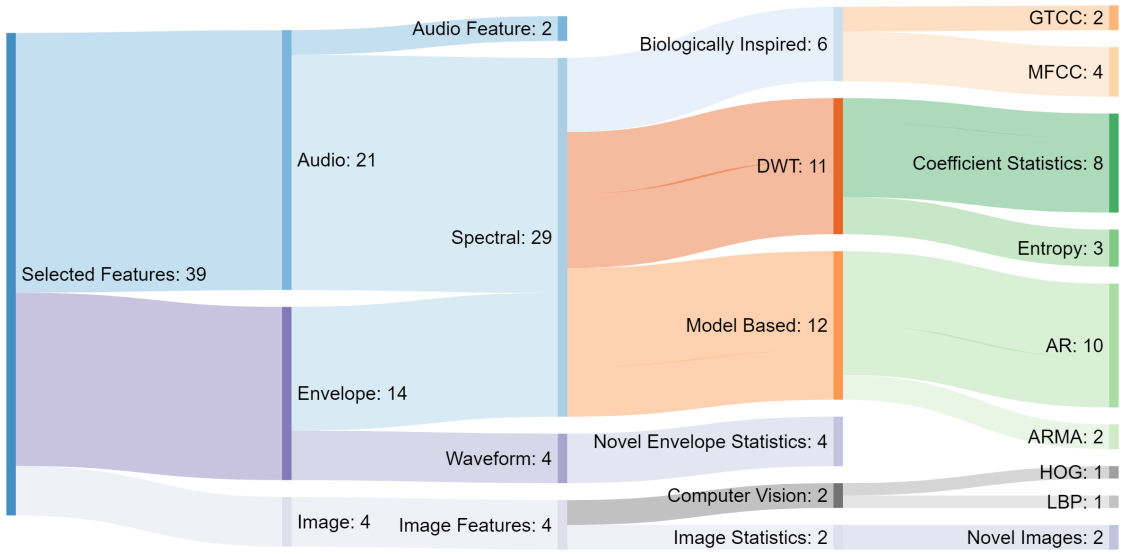


Figure 7.14: Sankey diagram of features selected for regression model

7.8.2 Regression Model

An SVM regression model was trained to provide feedback that reflects a given probe position with respect to the IPP. The trained model realises this by attributing larger values to observations that more closely resemble the IPP. The performance of this model

has been assessed using data collected from 22 participants, this data is presented and discussed in Chapter 4.

Performance is evaluated using correlation statistics. These are calculated with respect to the ideal response, which has been taken to be a binary signal where values of one correspond to the IPP. The overall results from this analysis are given in Table 7.21.

Table 7.21: Overall regression model performance

Feedback	Pearson Correlation	Spearman Correlation	RMSE
Raw[1]	0.82	0.73	0.35
MA	0.86	0.77	0.33

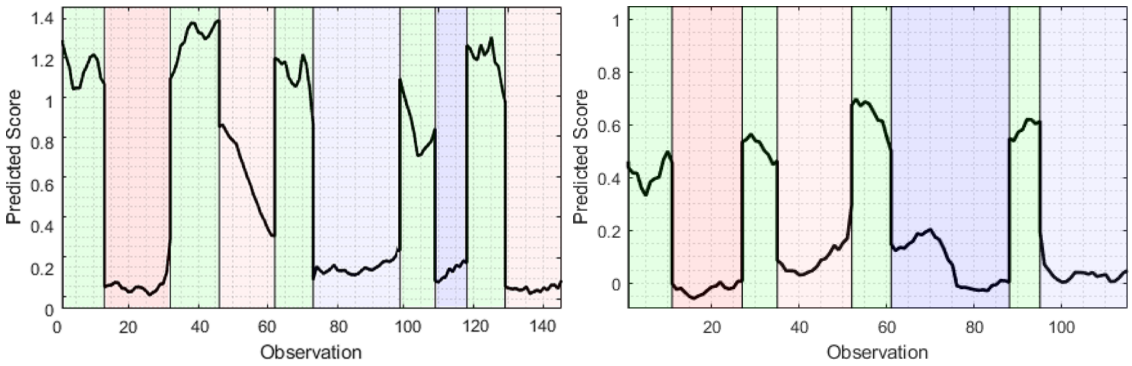
1 - Raw refers to single-beat predictions.

It can be seen from Table 7.21 that the MA results correspond to an increase of 4% for the Pearson (Cor_p) and Spearman correlations (Cor_s); this corresponds to an improvement of approximately 5% . Similarly, the MA results correspond to a reduction in RMSE of approximately 6%.

Chapter 4 illustrated that large interpatient differences are present across the measurement sets. This was discussed as an inhibiting factor when trying to form generalised models. With respect to the regression results, it was found that feedback from different individuals could exhibit similar correlation statistics, but have significant differences with respect to feedback magnitude at the IPP or off-angles. Figure 7.15 provides an example of feedback from the IPP and off-angles, for a single measurement set, from two participants. This illustrates that feedback in both cases is typically largest at the IPP, however, the relative magnitude of feedback at the IPP differs, with it being approximately 1.2 and 0.6 for P1 and P10, respectively.

As detailed in Chapter 4, data was sampled from five distinct probe positions during a measurement set. The variation in feedback, with respect to these positions, have been investigated further. The mean feedback values for each probe position are detailed in Table 7.22. This table also includes the mean percentage value of feedback from specific off-angles with respect to the IPP, illustrating the relative difference in feedback. The variation in feedback is also considered using STD.

It can be seen from Table 7.22, that some off-angles result in feedback more distinct from the IPP than others. For example, feedback associated with the positions L and H



(a) P1 (b) P10

Figure 7.15: Example MA feedback for specific probe positions across measurement sets from two participants, P1 and P10.

Table 7.22: Regression feedback mean and STD statistics

Position Label[1]	Mean	Normalised Mean[2] (%)	STD[3]
G	0.59	1.00	0.13
L	0.05	0.04	0.08
R	0.11	0.19	0.12
F	0.08	0.12	0.08
H	0.15	0.23	0.11
O	0.1	0.15	0.09

1 – Using position labels as defined in Table 4.2, where G denotes IPP, L denotes patients left, R denotes patients right, H denotes patients head, F denotes patients feet and O denotes all combined off-angles.
2 – Mean feedback with respect to mean IPP feedback.
3 – Mean participant STD for given position.

achieved normalised mean values of 0.04 and 0.23, respectively.

In practice, an operator will interrogate a large range of probe positions. As discussed in Chapter 4, this enables them to hone in on the IPP. These searching periods naturally occured whilst collecting data in this work, as well as transitional periods between IPP and off-angle measurements (e.g., due to the period taken to slowly tilt the probe or for the signal to settle in response to noise). The continuous feedback across single measurement

sets from P1, P10 and P5 are presented within Figure 7.16. These illustrate the type of feedback that would be generated in practice (as opposed to feedback limited to distinct positions), and have been annotated to give further context with respect to the scanning process.

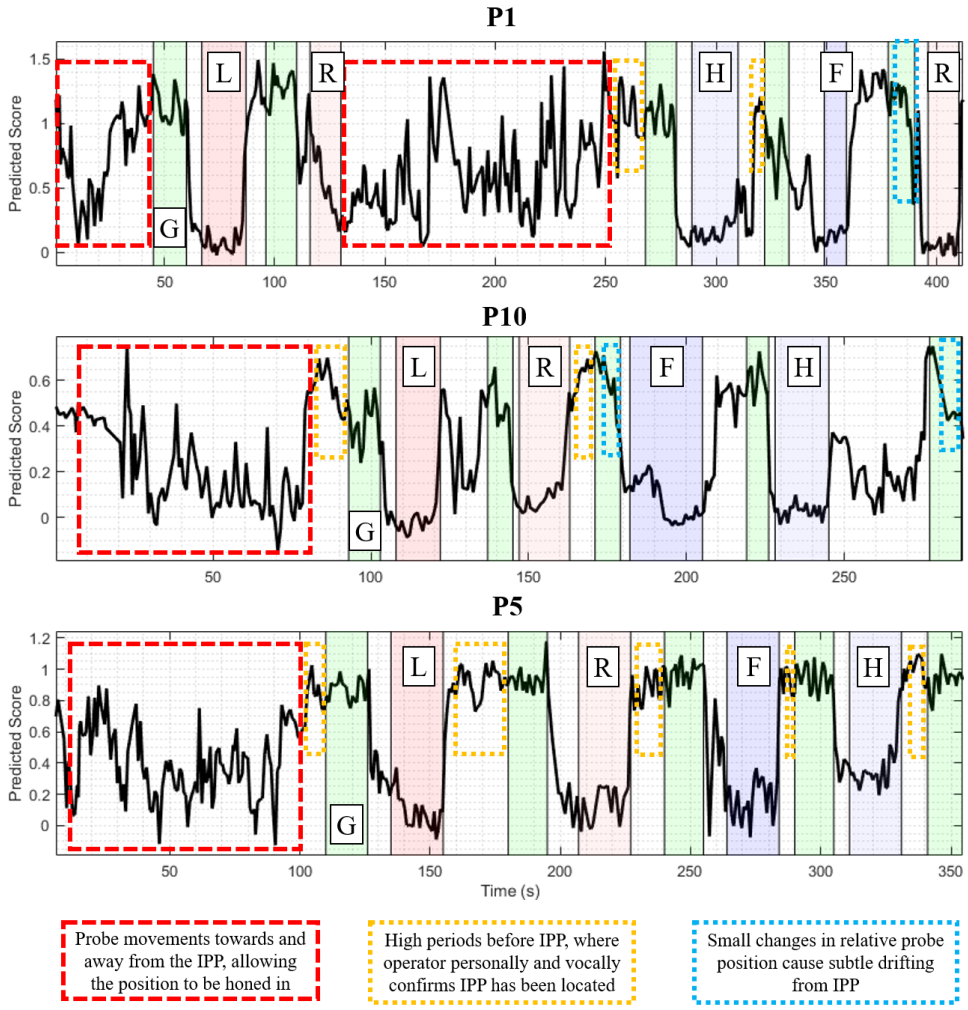
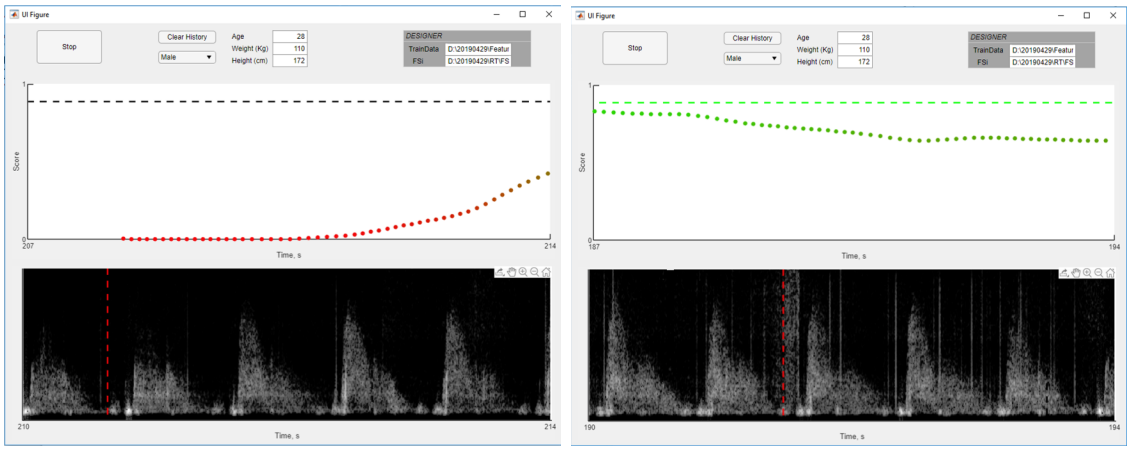


Figure 7.16: Annotated continuous raw feedback for single measurement sets for P1, P10 and P5

Figure 7.16 illustrates that feedback corresponding to measurement sets from different

participants contain common characteristics. These include erratic behaviour prior to initially locating the IPP, reflecting the scanning process whereby feedback indicative of approaching and travelling away from the IPP is used to correctly align the ultrasound beam. The feedback for P1 contains such a section mid-way through the measurement set. In practice, only recent feedback would be used to guide the probe position.



(a) Approaching IPP (b) Leaving IPP

Figure 7.17: Example of how feedback could be incorporated in real-time software

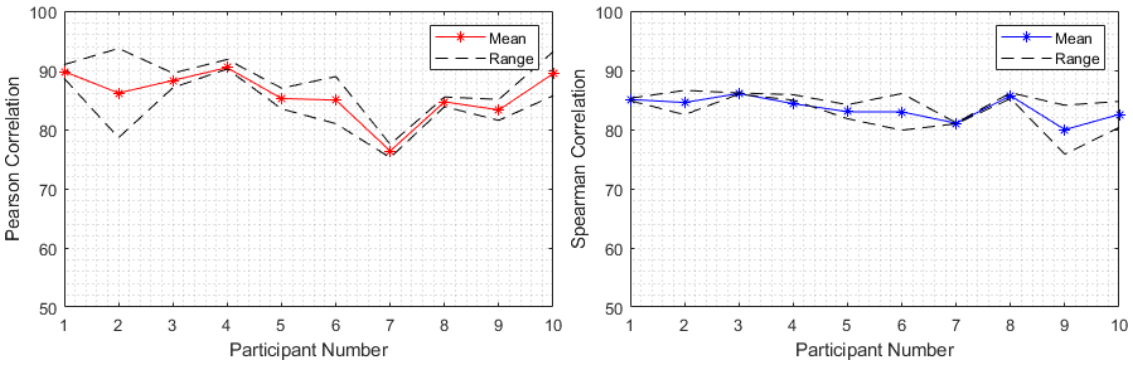
A more specific feature extraction program (i.e., limited to those features selected for the regression model, as opposed to all investigated features) was written and combined with the trained model; this enabled feedback to be generated more efficiently. This served to demonstrate how this feedback could be presented in practice for real-time applications. An example of this software is displayed in Figure 7.17. In this realisation, MA feedback was displayed using an associated colour that transitioned from red to green indicating 'bad' and 'good', respectively. The highest feedback achieved for a given measurement set is illustrated using a horizontal dashed line. This is displayed in green if the best measurement acquired occurred within the previous 15 s (as this is equivalent to the measurement history saved by the USCOM).

As demonstrated above, in Figure 7.15 and Table 7.22, feedback performance varies across participants. This is intuitive given the interpatient variability present in measurements. The times taken for the expert to locate the IPP (T_{IPP}) were detailed in

Section 4.4.3 (and displayed in Table 4.6), as discussed here, this corresponds to measurements from participants one to ten, and serves to indicate scan acquisition difficulty. The interpatient feedback variability, in the context of these times, has been investigated.

The MA correlation statistics for participants one to ten are displayed in Figure 7.18. This includes the performance range for each individual (i.e., the maximum and minimum statistics across the measurements sets).

The correlation between the performances presented in Figure 7.18 with the T_{IPP} values have been calculated. This is to assess whether any relationship between scan difficulty and feedback performance is evident. These correlation values, which are displayed in Table 7.23, indicate that no strong correlation exists between T_{IPP} and feedback



(a) Pearson (b) Spearman

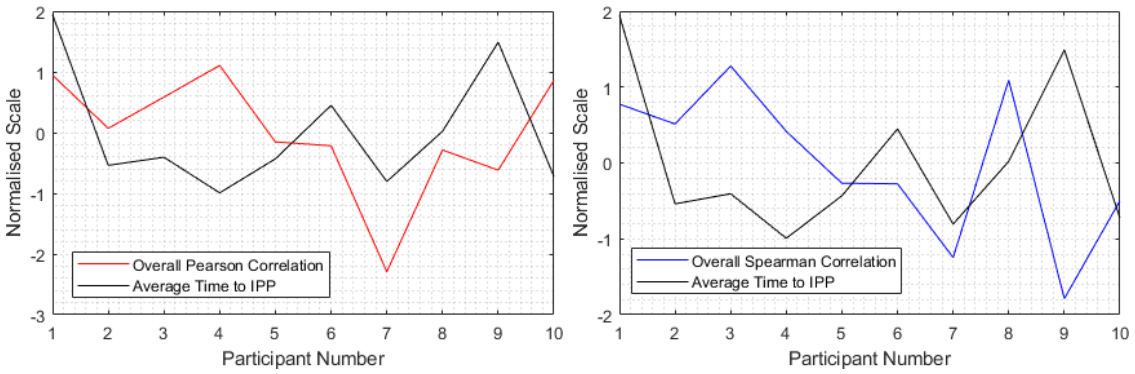
Figure 7.18: Mean correlation statistics for data collected by expert from participants P1 to P10.

Table 7.23: Difficulty of IPP location and SVM performance

Times	Correlation[1]			
	Pearson		Spearman	
	Cor_p	Cor_s	Cor_p	Cor_s
T_{IPP}	0.00	-0.24	-0.11	0.11
$T2^2_{IPP}$	-0.08	-0.20	-0.31	0.01

1 - Pearson and Spearman correlation statistics between regression model correlation statistics (Cor_p and Cor_s) and scan acquisition times.

2 - Average times taken to locate IPP, without time taken to locate first IPP in measurement set.



(a) Pearson (b) Spearman

Figure 7.19: Comparison between times taken to locate IPP and correlation statistics

performance. To visualise any relationship between feedback performance and T_{IPP}, the arrays can be normalised and displayed together. This is illustrated in Figure 7.19, and is consistent with the results in Table 7.23.

7.9 Discussion

The results presented in the previous section, and their implications, will now be discussed.

7.9.1 Selected Features

This chapter detailed an extensive array of features which were then extracted from Doppler ultrasound measurements. These include both novel and traditional descriptors, Figure 7.14 illustrates that the majority of selected features belong to the spectral category. This includes `MFCC_SYS_A2`, which was ranked as the number one feature.

MFCC and GTCC features were included as they have previously been shown to be excellent predictors for audio classification tasks, in total six features belonging to these categories were chosen. These features were predominately calculated using audio limited to the systolic portion. As detailed in Section 4.2.1, the shape and sound of the systolic region are primary sources of feedback for operators locating the IPP. Considering this, it is intuitive that features restricted to this portion may be more informative. Further investigation into potential features for Doppler analysis could be extended to include

those derived specifically from the diastolic region. This could be guided by the findings in the given work, to build on promising features, and reject poorly performing features which otherwise could result in impractically large feature matrices.

Interestingly no traditional or haemodynamic features were selected for the regression model. These features are predominately derived using extracted MFEs. However, a number of novel features derived from envelopes were included in the feature subset; these account for 36% of selected features. This illustrates that, despite the lack of traditional features, the information contained in Doppler envelopes can be very beneficial. The highest-ranking example is `DWT_CC_E_OMM_D3_4`, which is ranked second. This feature is the STD of DWT decomposed coefficients, another five similar features are included within the subset. These types of features have previously been extracted from Doppler audio signals. The selection of these features, and others such as `AR_CC_E_MSNSI_S2` which is the maximum AR PSD value of the MSNSI MFE, illustrate that frequency analysis of envelopes provide informative features. This is consistent with other fields such as ECG waveform classification, in which DWT decomposition [344], and AR analysis [384] have been shown to be powerful techniques for feature generation.

In addition to these spectral based envelope features, four novel envelope features were selected (these were described in Section 7.3.3). These all occur within the top 20 selected features, and as illustrated in Figure 7.13, contribute to the model's performance. These are statistics derived using combinations of envelopes, an approach which has not previously been described; of these selected features the MF envelope is used three times. This may be due to this envelope being less effected by spectral broadening, giving a more consistent paired envelope which still varies in response to blood velocity.

Features derived from images were only selected four times, these include two computer vision features and two novel features. Their inclusion demonstrate that useful information can be garnered directly from images. The following chapter uses images more successfully with DCNNs to perform classifications tasks. Section 8.4.3 details how the images used for this have a large impact with respect to resulting model performances. Considering these findings, it is likely that the value of features derived from images will be similarly dictated by the images used. In the context of this work, image features were derived from the OMM image. Further research into image features for Doppler ultrasound analysis could be extended to include those extracted from different image types, however, this could be computationally challenging given the vast array of descriptors attained using

these methods.

The following chapter includes multiple SVM classification models, and subsequently more discussion regarding features that have been selected. These discussions are also relevant to the regression model features and discuss further details surrounding them. The identification of promising features within this work is a significant contribution that highlights a variety of features, which could also be very beneficial in other areas of Doppler ultrasound analysis, such as disease classification.

7.9.2 Regression Model

An SVM regression model was trained to give continuous feedback to reflect ultrasound probe position. Intra and interpatient variability means the highest achievable subjective scan quality for a given person can vary widely. Binary labels were used to train models in an effort to produce feedback consistent across participants (i.e., similar high feedback values for IPP positions, and lower feedback for off-angle positions). The average Pearson correlation between generated MA feedback and labels was 0.86, demonstrating that the generated feedback and ideal response exhibited a strong linear correlation. This indicates that generated feedback does indeed respond to probe position. Furthermore, it validates the labelling method with respect to subjective scores, which in their raw form exhibited lower correlation statics (with respect to the ideal response) and more similar off-angle values with respect to the IPP position.

However, it was found that interpatient feedback, in terms of mean IPP or off-angle scores, still exhibited high variability. This is seen from Figure 7.16 and 7.15, where IPP regions are approximately 1.2 and 0.6 for P1 and P10, respectively. Considering this, the use of absolute values to determine whether the IPP had been located could be misleading across different patients.

Nevertheless, in terms of intrapatient application, the continuous feedback provides a quantitative and historical score of obtained measurements. When locating the IPP, the experience gained through an examination helps an operator confirm that the IPP has been located. This in response to the subjective quality of current and previously obtained scans. Acknowledging improvements of this, and in particular, mentally contrasting them to earlier obtained measurements can be challenging. The feedback in this respect provides a constant reference, and sense check with respect to current measurements. This aspect

was illustrated using example software in Figure 7.17, here a horizontal line indicates the best score acquired so far within a measurement. In practice, when an operator thinks the IPP has been located the examination is paused, and typically at least three representative beats are selected to generate metrics. With respect to the USCOM device, once paused the operator has access to the previous 15s, from which these beats can be selected. With respect to previous measurements, the feedback could inform a user whether the best measurement thus far was being obtained.

Continuous advances in technology mean retaining more than 15 s of data is easily attainable in modern devices. If a Doppler device retained the data for a whole examination, for example, the historical feedback could be used to quickly locate the highest scored scan regions. In Figure 7.16, P1 exhibits a period of almost 2 minutes midway through the measurement set where the expert sonographer is searching for the IPP. This illustrates that despite the IPP initially being found in under a minute, the location can still become ambiguous. This may have been due to, for example, the participant moving or the coupling between the probe and skin changing due to a reduction in ultrasound gel. Given longer durations of stored data, in normal examination situations, the scan could be paused at this point and the previous highly scored regions relocated using the feedback.

In addition to aiding individual operators, continuous feedback could be useful for trainee sonographers. In this respect, an expert sonographer could perform an initial examination, which would give a target score for trainee users to aim for. In either case, such feedback would be most useful in the more challenging scenarios i.e., patients who are harder to locate the IPP from. In a situation where scan difficulty was strongly linked to model performance, the value of the feedback would be reduced. No correlation between the time taken to locate the IPP and model performance was found in the given work, suggesting this form of feedback is not effected by subjective scan difficulty. The process of designing clinical software that incorporates feedback models are discussed later in Section 8.7.6.

A challenge in this work was the selection of data labels, as discussed in Section 7.4, binary labels were used that assign each off-angle an identical value of 0. Furthermore, off-angles were restricted to four positions. These limitations were necessary to keep the collection and analysis of data practical. Off-angles were measured with respect to the IPP, however, the distance of these off-angles from the IPP was not available, rather they were identified via a subjective reduction in scan quality. Given this approach, and

the variability in measurements, variations in these distances will occur. A potential means of overcoming this would be to use position recording gantry or tracking cameras to accurately record the position of the probe. Labels with respect to probe position could then be formed. Considering the difference in intra- and interpatient scan variability at the IPP, coupled with the additional variations at off-angles and hardware required, such data collection and subsequent processing could be very challenging. Another application of such gantry or cameras could be constructing a simulation from a single participant, enabling a virtual 3D model to be built and used by trainee operators (independently of a participant using the gantry).

7.10 Conclusion

A regression SVM model has been trained using a selection of both previously described and novel features. This model provides continuous feedback that varies dynamically in response to the probe position, with larger values indicating a better measurement. With respect to the ideal response, the generated feedback exhibited a linear correlation of 0.86. No previous quantitative or automatic means of assessing Doppler ultrasound position have previously been described.

The selected features include those previously used in other disciplines, such as speech recognition; specifically, MFCC features were highly ranked. It was found that although traditional envelope features were not selected, a variety of features based upon the spectral content of MFEs proved to be valuable descriptors.

Chapter 8

Localisation Feedback

8.1 Introduction

Doppler ultrasound is a non-invasive, cost-effective and efficient means of performing haemodynamic monitoring. However, the potential of this technology is limited by challenges inherent to acquiring measurements. Currently, an operator requires a high degree of experience and skill to consistently and accurately aim the ultrasound probe. These difficulties are discussed in depth in Section 2.6. This thesis investigates signal processing and machine learning techniques that could mitigate these constraints, and enable less experienced users to operate associated devices.

The preceding chapter explored means of aiding data acquisition using regression models. These models quantified scan quality using continuous feedback, enabling advantageous probe movements to be recognised. This feedback complements current methods of locating the IPP, whereby an operator sweeps an ultrasound beam through different planes (described in Section 4.2.2). Variations in scan quality are then used to locate the IPP. This process mandates a scan history, in which multiple positions are interrogated before deducing the direction of the IPP, or confirming that it has been located. The continuous feedback, therefore, cannot explicitly indicate ideal probe movements, or whether the IPP has been located. Such predictions require discrete localisation feedback. This would further reduce the level of operator experience needed, and make the technology more accessible and faster to implement.

This chapter explores means of providing discrete predictions of probe position, which

are not reliant on historical measurements. These predictions go beyond any inference an operator could make using a single measurement. Classification models are used to categorise observations in terms of the current probe position, giving immediate information with respect to adjustments that need to be made; for example, that the probe needs to be angled to the patients right. This scenario is illustrated in Figure 8.1a.

A variety of different classification tasks have been investigated in this chapter. These are trained and evaluated using the survey data presented in Chapter 4. These data include measurements sampled from five different positions; enabling different target classes to be considered. Models were trained to classify measurements using two classes (IPP or off-angle), using three classes with labels limited to vertical or horizontal directions (e.g., patients left, IPP and patients right), and using all five classes. These classification tasks are summarised in Table 8.1.

Previous works have documented Doppler ultrasound classifiers; these have typically been limited to two distinct categories and are discussed in Section 3.5. The increased number of classes considered in this work, and the high similarity between classes, makes the given task particularly challenging.

A series of models have been trained and evaluated for each classification scenario, these include SVM models trained using the features and feature selection procedures described in Chapter 7. In addition to SVMs, DCNN image recognition models have been created.

DCNNs are the current state-of-the-art for image recognition. They have been purported as a means of overcoming increased intraclass similarity, and reduced observations [289]. In micro-Doppler applications, its been observed that DCNNs are able to distinguish spectrogram characteristics that are undetectable by humans [306]. Furthermore, DCNN image classification of spectrograms has shown improved performances with respect to more traditional methods, where handcrafted features are used to train classifiers (e.g., SVMs). For these reasons, AlexNet DCNNs have been included in the presented work [296].

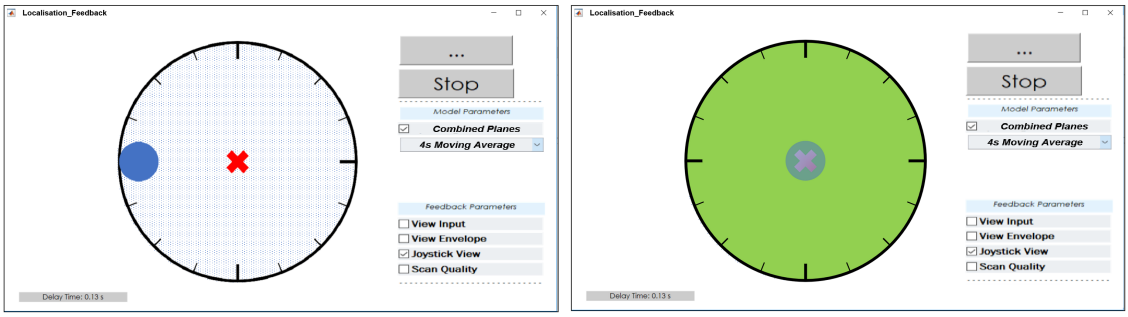
The results in this chapter found that typically DCNNs outperformed the trained SVMs, achieving an overall classification accuracy 5% higher for the 5-class task. This difference is smaller for the less complex 2-class task, where the SVM and AlexNet models achieved overall accuracies of 88.1% and 89.3%, respectively. To conclude, the main contributions contained in this chapter are:

- The identification of promising novel features for Doppler ultrasound classification.
- The design of SVM models for the previously unexplored application of Doppler ultrasound probe position classification.
- The design of image recognition models for probe position classification, a technique previously unreported for Doppler ultrasound analysis.
- An evaluation of different Doppler classification tasks, and an evaluation and comparison of SVMs and DCNNs for this purpose.

8.2 Classification Categories

In practice, the IPP is located by an operator using their prior experience from an examination. This equates to a series of subjective evaluations of received measurements. The previous chapter explored means of quantifying these evaluations, in practice this could both aid less experienced users and give confidence that a given measurement is the best achievable.

In this chapter, observations are classed into definitive categories. Instead of continuous, patient-specific feedback variables, measurements are classified in terms of their



(a) (b)

Figure 8.1: Software designed to display the predicted location of a given measurement. The examples illustrate an example where a) probe is predicted to be angled towards the patients left, and b) the probe is predicted to be at the IPP.

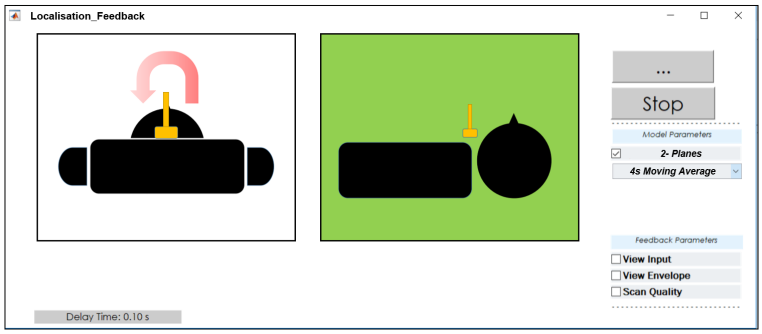


Figure 8.2: Localisation feedback using separate horizontal and vertical axes

location with respect to the IPP. This allows feedback to directly guide the user, as opposed to scoring their probe adjustments. This is enabled using data collected from various probe positions.

Different types of classification task are available given the collected data. Three different labelling methods are explored in this work; these are summarised in Table 8.1. The most challenging task is the 5-class scenario. In this case, each sampled probe position is used as a target category.

With respect to an operator performing a scan, outputs from these trained models could be used to indicate from which location a measurement is being recorded (with respect to the IPP), and thus, which direction a user should move the probe in order to locate the IPP. An example of predictions being used in this manner is displayed in Figure 8.1b. The task can be simplified by reducing the number of categories. The is achieved by training individual models for the vertical and horizontal axes separately. This would allow feedback which complements the current scanning procedure used to identify the IPP, by scanning planes independently. An example of this approach being implemented in software is illustrated in Figure 8.2.

In addition to these models, an overall model which classifies observations as either belonging to the IPP or not can be constructed. Although this offers no localisation feedback, a model of this type could complement the continuous feedback models presented in Chapter 7. In this respect, a user could be guided towards the IPP, and be alerted that they have located this position via the 2-class model. These classification tasks are tackled using DCNNs and SVMs, the following sections will describe these modelling approaches.

Table 8.1: Localisation classification tasks

Scanning Plane	Classification Task	Labels[1]
All	5-class	*G*, *L*, *R*, *H*, *F*
Vertical	3-class	*G*, *L*, *R*
Horizontal	3-class	*G*, *H*, *F*
All	2-class	*G*, *O*

1 – Using position labels as defined in Table 4.2, where *G* denotes IPP, *L* denotes patients left, *R* denotes patients right, *H* denotes patients head, *F* denotes patients feet and *O* denotes all combined off-angles.

8.3 Support Vector Machines

The feature and model selection procedures, described in Section 7.6, considered a variety of different classification algorithms (Table 7.19). Model selection for both regression and classification tasks have identified SVMs as being the most suitable model type. Many machine learning concepts are recent innovations, and support vector algorithms are no exception. Their implementation for a variety of tasks have been discussed in Section 3.5.1, in which they demonstrate good performance and often outperform other algorithms [205]. The fundamental concepts of these algorithms are presented in this section.

Modern SVM implementation was first proposed by Cortes and Vapnik [385], who illustrated their potential for optical character recognition. SVMs were designed as a non-linear solution for classification and regressions tasks, and were developed in response to deficiencies of prior methods such as ANNs, which depended upon large datasets and exhibited poor generalizability [386]. ANNs can often lead to complex models resulting in situations of overfitting, whereby a fitted model can give very low bias with respect to the training set, but high variability when tested using new observations. Where reductions in the complexity of a model is required, regularisation of a selected function can be very helpful. This is the theory of Vapnik-Chervonekis [387], and is a fundamental concept behind SVMs. SVM algorithms draw from three fundamental areas: mathematical optimisation theory, statistics and computer science [388].

A support vector classifier uses boundaries to identify regions of data belonging to a given class. Test samples are then classified as belonging to this class if they fall within this

identified region. These boundaries are created using a core set of points called support vectors, giving the advantage of training data being largely redundant once a boundary is formed [388].

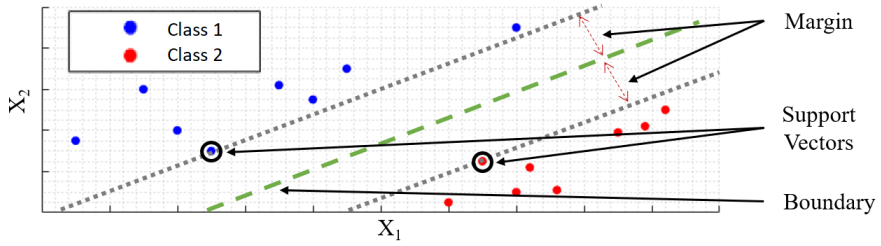


Figure 8.3: A 2D boundary separating two classes, in which margins are maximised.

The minimum distance between an observation and the boundary is called the margin, one choice for a classifier is to use a boundary which gives the largest margin between two classes, this is referred to as the maximal-margin hyperplane. Figure 8.3 illustrates this boundary and the associated margins separating two classes, using a feature space with two features vectors X_1 and X_2. This approach can be very sensitive to top outliers, and in many cases hyperplanes are unable to fully separate classes. Soft margins can be used to allow instances of training noise [389]. Where two classes are not separable, soft margins allow misclassifications to occur. In such cases, a hyperplane is chosen which has the least number of points within the margin. Through cross-validation, an ideal soft margin can be identified. This approach is used in soft margin classifiers, a group to which SVMs belong.

In some instances, data is not linearly separable. SVMs overcome this hurdle by embedding the data into a higher-dimensional space, producing a feature space which is separable and still subject to soft margins [389]. This mapping procedure allows SVMs to tackle nonlinearities in data, and for a support vector classifier that separates classes to be found. This is illustrated in Figure 8.4, in which the one-dimensional data has been squared, allowing a one-dimensional line boundary to isolate the two classes. The process of identifying support vector classifiers in higher dimensional space is achieved using kernel functions, of which there are different types.

A popular kernel function, is the polynomial kernel function, this function is displayed in equation 8.1. Using this, the relationships between pairs of observations can be calcu-

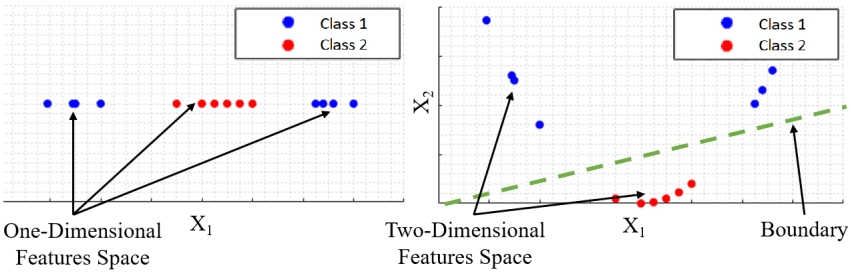


Figure 8.4: Transformation of data to higher-dimensional space, enabling linear boundary to separate classes.

lated.

$$K(x, y) = (x \times y + c)^d \tag{8.1}$$

Where x corresponds to observations of one class, and y to observations of a second, d denotes the degree, and c the coefficient. The example presented in Figure 8.4 used variables where $c = 0.5$ and $d = 2$. Substituting these values into equation 8.1, and re-arranging allows the corresponding dot product and new coordinates to be found.

$$K(x, y) = xy + x^2y^2 + 0.25 \tag{8.2}$$

$$= (x, x^2, 0.5).(y, y^2, 0.5) \tag{8.3}$$

Where x^2 and y^2 are the new coordinates for x and y respectively. In practice, the values of c and d are found using cross-validation. In an example where the degree was set to three, the transformed data would exhibit three dimensions, and a two-dimensional plane could be implemented to separate classes. This procedure is inherently computationally expensive, however, use of kernel functions and the dot product (which is referred to as 'The Kernel Trick') allows the relationships to be found without explicitly transforming the data into a higher-dimensional space [386]. Within this chapter, SVM classifiers are compared with DCNNs.

8.4 Deep Convolution Neural Networks

In recent years, deep learning image recognition methods using DCNNs have been increasingly implemented in clinical scenarios. Such methods are considered to be the state-of-

the-art and have a number of purported advantages compared to more traditional classification methods, which use hand-crafted features. Furthermore, direct comparisons of DCNNs and SVMs are consistent with these claims [306]. The merits of deep learning techniques are discussed in Section 3.6.3.

Considering the potential advantages of DCNNs, their application to Doppler ultrasound has been investigated in this work. DCNNs have been trained for each classification task. The choice of model architecture, training strategy and image formation are summarised within this section.

8.4.1 DCNN Architecture

There are a variety of pre-trained DCNN architectures available. In the presented work two architectures were considered: AlexNet and GoogLeNet. As described in Section 3.6.3, these models have been successfully repurposed for a variety of tasks and have been directly compared. Using the training approach and image types described in the following two sections, AlexNet and GoogLeNet performance was investigated using approximately half of the survey data. This was performed for the 2-class and 5-class classification scenarios. This process found that AlexNet marginally outperformed GoogLeNet, and was substantially faster to train, which is particularly significant considering the expensive cross-validation method and investigation of different image types. Considering these limitations, AlexNet was chosen as the DCNN architecture for the presented work, and was trained using transfer learning. The architecture of AlexNet is discussed further in Section 3.6.3, and displayed in Figure 3.17.

8.4.2 Transfer Learning

A method called transfer learning has been used to train DCNN models in the given work. This section will briefly introduce transfer learning and describe its implementation.

DCNN models can be trained from scratch or using a method called transfer learning. This method is a new solution for pattern recognition, in which a pretrained DCNN is fine-tuned for a new classification task [297]. It is often found that models created using transfer learning out-perform those that were trained from scratch [290, 297, 390, 391]. In these examples, the technique was used to classify medical images.

This technique uses a complex and highly accurate model, which has been trained

using a large collection of natural images, and retrains it for a comparatively simple task with far fewer categories [300]. This is performed by randomly initialising the weights in the fully connected layers, these are then relearnt using new images [290]. This concept is illustrated in Figure 8.5.

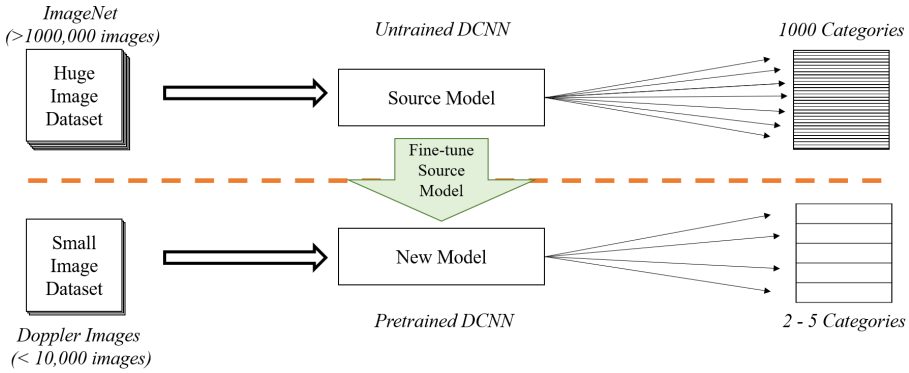


Figure 8.5: Training DCNNs through transferring knowledge via transfer learning

In addition to model performance, the process is computationally more efficient, straightforward, and requires comparatively fewer images and training epochs to give good performance [300]. However, limited training images can result in overfitting. To combat this, and make models more generalised, image augmentation can be performed, for example, by randomly rotating or translating images [290].

8.4.3 Images

Audio can be converted into time-frequency images, which can then be classified using AlexNet. This has been demonstrated for audio classification [392], and similar approaches have been used for micro-Doppler radar classification [307]. These images can be generated using different post-processing and spectral estimation methods, for example using the STFT or the AR method. These methods have been discussed further in Section 3.6.1. AR methods have previously been shown to be advantageous for Doppler analysis, this is discussed in Section 3.6.1.1. Associated spectrograms, with respect to Fourier methods, have been described as qualitatively better [224]. This observation was found to be subjectively true for the given work, and AR images were investigated.

In the presented work, images were generated for individual cardiac cycles. Different

Table 8.2: Selection of image types and variants investigated as inputs for DCNN models

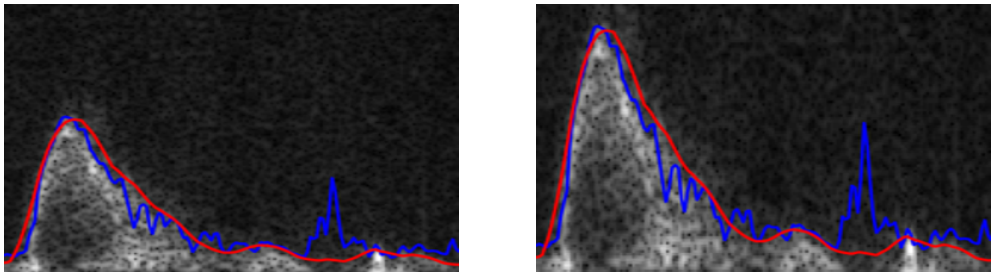
Type	Description
Image Types	
Spectrogram	STFT PSD image
Spectrogram with Envelopes	STFT PSD image, with MFE envelopes
OMM Image	Image generated for OMM MFE extraction
AR Image	Autoregressive PSD image
Image Variants	
Spectrogram	Colour and greyscale
Spectrogram	Exponent contrast limiting [1]
Spectrogram with Envelopes	MFE combinations [2]
OMM Image	Dynamic range[3]
AR	Order choice[4]
AR	Exponent contrast limiting [1]

1 – Compression of dynamic image range, using exponents ranging from 0.1 to 0.5.
2 – OMM and MSNSI MFE combinations, including line colours.
3 – Dynamic ranges from 20 to 60 dB.
4 – Orders from 5 to 80.

methods of images generation were evaluated, these are displayed in Table 8.2. Images used for transfer learning must be the same size as those used for training the original model, in the case of AlexNet these images are 227×227.

In addition to different methods of generating time-frequency images, different means of cropping images were investigated. This is with respect to the extent of frequency content displayed. As with segmenting beats, this was guided using the extracted OMM MFE. The maximum frequency contained within an image was determined using the peak MFE value, this is illustrated in Figure 8.6. Cropping images in this fashion ensures more area is occupied by signal, and preserves associated details when images are resized.

As illustrated in Table 8.2, different parameters can be adjusted when creating these images, such as dynamic range. In addition to this, overlaying extracted MFE envelopes were found to improve performance in some instances; this was the case for the STFT spectrogram images.



(a) (b)

Figure 8.6: Spectrogram Images with OMM and MSNSI envelopes, cropped using maximum OMM envelope frequency, where a) corresponds to $1.8 \times max(MFE_{OMM})$ and b) corresponds to $1.1 \times max(MFE_{OMM})$.

8.5 Evaluation Methods

In total the results from 10 classification models, using two machine learning approaches, have been reported within this chapter. Their implementation and the methods used to contrast their performance will now be described. This has been performed using data collected from 22 individual participants. These measurements, and the associated survey methodology, are presented in Chapter 4.

8.5.1 SVM and DCNN Implementation

SVM models have been trained using Matlab 2019b. Model type and feature selection have been performed as described in Chapter 7.

8.5.2 DCNN Implementation

All models have been constructed using transfer learning within Matlab version R2019a. Pre-trained DCNNs were accessed using the Deep Learning Toolbox. The early layers of a pre-trained model, which learn low-level features such as edges or blobs, are extracted. This leaves three layers, which in the case of AlexNet are configured to 1000 different categories.

The extracted layers are transferred for the given task, and the final layers are replaced with a fully connected layer, softmax layer, and classification output layer; these layers are

displayed in Figure 3.17. The new fully connected layer has the same number of classes as the number of categories in the given classification task [393]. Transfer learning aims to maintain the features of the early stages of the transferred layers. This is achieved by using an initial learning rate which is slow. In order to speed up learning in the final layers, different learning rate factors are applied [393].

Different applications using transfer learning have been described in Section 3.6.3, guided by these works a range of models were trained and evaluated, allowing suitable parameters values to be identified. The parameters used to re-train DCNNs are displayed in Table 8.3.

Table 8.3: Transfer learning training parameters used to retrain DCNN models

Model Type	No. Epochs	Mini batch size	Learn Rate		
			Initial	Weight[1]	Bias[1]
2-Class	6	256	1e-4	20	20
3-Class	10				
5-Class	10				

1 – Applied to fully connected layer.

Images are separated into training and testing sets. Training images are augmented to increase the generalisability of the model, this has the effect of randomly flipping images, and sliding them horizontally or vertically by up to 30 pixels.

8.5.3 Data Partitioning and Averaging

Independent participant cross-validation has been used to thoroughly evaluate each model. This has been described in Section 7.7.2.

In addition to raw predictions (single-beat predictions), predictions using a 4 s time average window are also presented. This is similar to the method used for the regression models described in Section 7.7.3, and associated results are also referred to as MA. However, as this is a classification task, class probabilities ($Prob$) are combined and then used to give an average class prediction ($ProbAv$); this is illustrated in Figure 8.7.

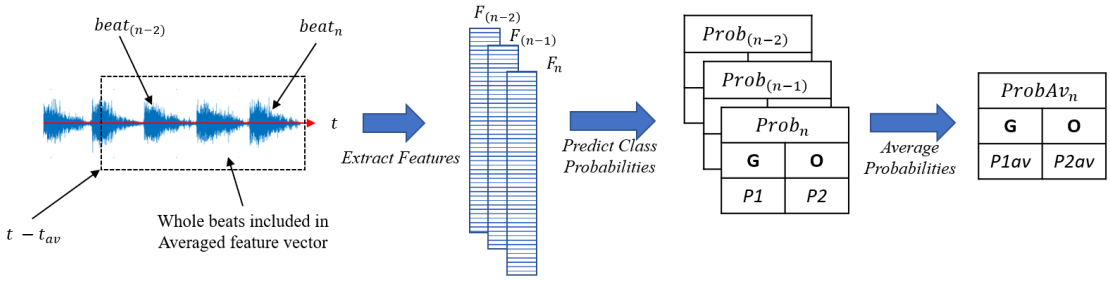


Figure 8.7: Method of combining predictions for classification models using moving time window; the example provided is for the 2-class scenario.

8.5.4 Ensemble Models

Multiple models can be combined to give ensemble models. This has previously been demonstrated using ensemble averages from DCNN networks to classify radiographs [290], or combining posterior probabilities to create neural network ensembles for diagnosing Doppler heart sounds [198]. In these cases, the ensemble methods demonstrated improved performance over singular models. Considering this, ensemble methods have been investigated in the given work as a further means of improving classification performance. This has been investigated for the 5-class scenario by combining classification probabilities from different DCNN models, as well as combing SVM and DCNN models.

8.5.5 Confusion Matrix

The performance of classification models are commonly assessed using the confusion matrix, and associated metrics. The confusion matrix provides an intuitive visualisation of a models behaviour, and is considered to be the best tool for assessing a classifier [394]. The contents of confusion matrices can be used to derive several useful terms, which help understand and explain classification performance. These metrics are very common to classification evaluation, have been used in previous Doppler classification works discussed in Section 3.6.3, and have been adopted for the presented work. Confusion matrices and associated statistics are used in the presented work to contrast SVM and DCNN classification performances. This will be performed for each of the classification tasks (as detailed in Table 8.1).

The confusion matrix is a table which compares the actual classes of a test set to

the predicted ones. Typically, actual classes are detailed in the columns of the table and predicted labels in the rows. The numbers that occur in the diagonal of this matrix correspond to correct predictions, and all other values correspond to incorrect predictions.

The predictions from a classification model can be one of four cases: true positives, false positives, true negatives and false negatives [394]. Given a set of predictions and true labels, a confusion matrix provides a summary of these. The meaning of each case can be intuitively understood when considering a binary classification task. These cases are summarised in Figure 8.8 and Table 8.4, where with respect to a set of predictions, TP denotes the number of true positives, TN the number of true negatives, FN the number of false negatives and FP the number of false positives.

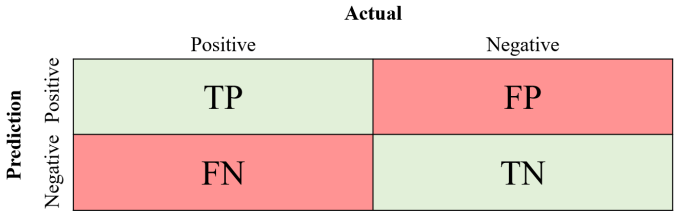


Figure 8.8: Contents of a binary confusion matrix

Table 8.4: Prediction cases for binary example, where positive predictions correspond to 1 and negative predictions correspond to 0

Scanning Plane	Labels
True Positive	Correct prediction of 1
True Negative	Correct prediction of 0
False Positive	Incorrect prediction of 1
False Negative	Incorrect prediction of 0

Sensitivity and specificity are two commonly calculated values and are computed using the four cases listed in Table 8.4. Sensitivity, which is also referred to as true positive rate (TPR), describes the percentage of positive cases which are correctly predicted as being positive. Specificity, which is also referred to as the true negative rate (TNR), describes the percentage of negative cases correctly classed as being negative. These are calculated using equations 8.4 and 8.5 respectively.

$$TPR = TP/(TP + FN) \tag{8.4}$$

$$TNR = TN/(TN + FP) \tag{8.5}$$

A classifiers accuracy is defined as its ability to correctly select and reject cases [394]; this is calculated using equation 8.6. This can be performed for individual classes, or for the confusion matrix as a whole; graphically this is equivalent to the sum of the diagonal divided by the sum of all predictions.

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \tag{8.6}$$

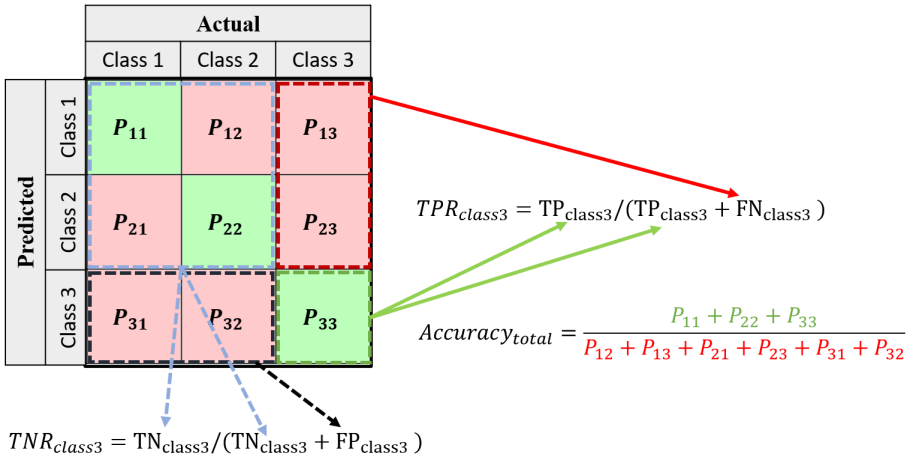


Figure 8.9: Example of multi-class matrix, with TPR, TNR and overall accuracy calculations

In multi-class scenarios there are no single values of TPR and TNR for the entire matrix, instead these are calculated for each class separately; this is illustrated in Figure 8.9. Precision statistics (PPV) can be calculated in the same manner through dividing TP by the total number of positive predictions (i.e., TP and FP combined).

8.5.6 SVM and DCNN Comparisons

The features employed by the SVM models and images used by AlexNet are generated concurrently. Images are also used to visually identify incorrectly segmented cardiac cycles.

These observations are then removed from the data used to train both SVM and AlexNet models, these are summarised in Table 4.5. This allows the two classification approaches to be trained and evaluated using corresponding data and subsequently compared. A high-level representation of this is given in Figure 8.10.

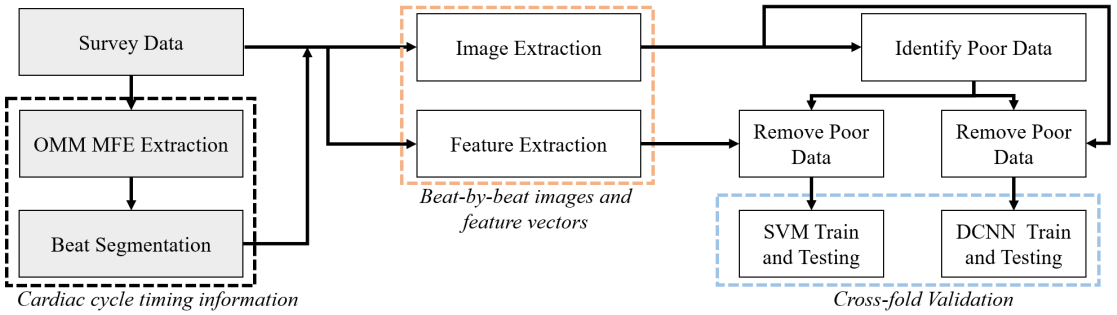


Figure 8.10: Concurrent image and feature extraction for model evaluation and comparison.

8.6 Results

This section presents the features identified as being valuable for classifying Doppler ultrasound probe position. The performance of corresponding SVM models are then compared with AlexNet models for each classification task.

8.6.1 SVM features

The procedures outlined in Section 7.6 were used to both identify a suitable model type and subgroup of features for each classification task. This process identified SVM models as being the most suitable for each of these tasks. Feature selection produced optimised subgroups for each classification scenario. This process identified 46, 60, 63 and 68 features for the 2-class, vertical 3-class, horizontal 3-class and 5-class models, respectively. The complete features lists are included in Appendix M.

As detailed in Section 7.6, the best performing features are selected using sequential forward selection. The number of features then included in subsequent models (i.e., feature subset size) is then determined using cross-validation loss. The minimum number of

features required to give adequate performance is chosen. This helps prevent overfitting and reduces computational costs, which is of particular concern in real-time applications. The loss with respect to feature subgroup size for the 5-class scenario is displayed within Figure 8.11.

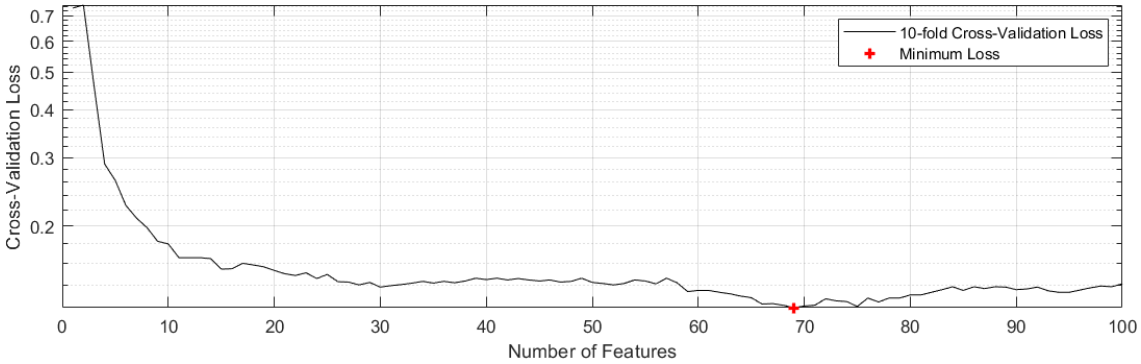


Figure 8.11: Variation in feature subset length with respect to 10-fold cross-validation classification loss for 5-class scenario

The benefits of including additional features tend to reduce as more features are included. It can be seen from Figure 8.11 that the higher-ranked features have the largest impact in terms of improving accuracy, with improvements becoming significantly smaller above feature lengths of approximately 20. These higher ranked features are displayed in Table 8.5, which details the top 20 identified features for each classification task.

A number of common features are selected across the classification tasks, for example, `MFCC_SYS_A4` occurs in the top five selected features for all models bar the horizontal-plane 3-class model. This may indicate that this feature is useful for discriminating between vertical classes, but is less valuable for horizontal cases. Several key features identified across the models are closely related, for example, MFCC and GTCC features. The most common feature types displayed in Table 8.5 for the 2-class model are AR-based, with respect to other models, this feature type is only displayed in two other cases; once for the horizontal model and once for the vertical model. A number of waveform features have been selected across the models, the most common type of waveform features are novel envelope features (as described in Section 7.3.3). Notably, the number one ranked feature for the 3-class horizontal model belongs to this category (`W_CC_E_NE9_2`). This feature is also ranked sixth for the vertical-plane model and corresponds to the ratio of the MF

Table 8.5: Top 20 features identified for each classification task

Rank	2-class	3-class		5-class
		Vertical-plane	*Horizontal-plane*	
1	DWT_CC_E_OMM_D5_3	DWT_CC_E_MSNSI_D5_3	W_CC_E_NE6_7	MFCC_SYS_A2
2	DWT_CC_A_SE11	MFCC_SYS_A2	W_CC_E_OMM_H5	LBP_CC_I_F1
3	AR_CC_A_C2	W_CC_E_NE8_6	AF_CC_A8	AR_CC_A_S3
4	MFCC_SYS_A2	GTCC_SYS_A_F2	GTCC_CC_A_F3	AR_CC_E_OMM_S2
5	GTCC_SYS_A_F0	AR_SYS_A_C3	AR_CC_A_S2	W_CC_E_OMM_H5
6	AR_CC_A_C1	W_CC_E_NE6_7	AR_CC_E_OMM_S2	LBP_CC_I_F58
7	HOG_CC_I_F164	W_CC_E_NE9_6	ARMA_CC_A_AC4	GTCC_CC_A_T2
8	AR_CC_A_C3	DWT_CC_A_SE4	HOG_CC_I_F5	GTCC_CC_A_T6
9	AF_CC_E_OMM_SNR	AR_CC_A_S2	W_SYS_E_OMM_H4	MFCC_CC_A0
10	AR_CC_A_C4	W_CC_E_MSNSI_S1	AF_CC_A13	MFCC_CC_A13
11	W_CC_E_NE6_7	AF_CC_A11	W_CC_E_NE2_3	MFCC_SYS_A3
12	MFCC_CC_A5	DWT_CC_E_OMM_D5_2	MFCC_SYS_A3	IMS_SYS_I_IP1_S1
13	GTCC_CC_A_F13	GTCC_CC_A_F4	AR_CC_E_MSNSI_S4	W_CC_E_TF11
14	ARMA_CC_E_OMM_S1	ARMA_CC_A_AC3	W_CC_E_NE1_3	W_CC_E_NE10_1
15	AR_SYS_A_C4	DWT_SYS_A_SE2	AF_SYS_A1	MFCC_CC_A4
16	MFCC_SYS_A0	DWT_CC_E_OMM_D5_1	DWT_CC_A_D1_4	DWT_SYS_A_V2
17	MFCC_CC_A1	W_SYS_E_MF_CO2	GTCC_SYS_A_T2	HOG_CC_I_F246
18	W_CC_E_TF11	AR_CC_A_S3	ARMA_CC_A_AC7	DWT_SYS_A_V3
19	AF_SYS_A8	MFCC_SYS_A10	AR_CC_A_S3	AF_CC_A2
20	ARMA_CC_E_OMM_S3	GTCC_SYS_A_F0	AR_CC_A_C1	ARMA_CC_A_S1

envelope STD and the MSNSI$_{noISB}$ MFE STD.

In addition to this, the novel envelope features `W_CC_E_NE11_1` and `W_CC_E_NE11_2` are also selected for the vertical model; these correspond to the difference in the STD of the MSNSI and MSNSI$_{noISB}$ MFE, and the normalised difference with respect to the maximum value in these two envelopes. The number one ranked feature for both the 2-class and 3-class vertical-plane models are statistics derived from DWT decomposed MFEs. Overall few image-based features were identified in the top 20 features for each model, however, two LBP features are ranked second and sixth for the 5-class scenario.

Table 8.6: Summary of feature types selected for each model

Feature Type	Contribution (No. \%)			
	2-class	3-class, V	3-class, H	5-class
Audio Based				
GTCC/MFCC	13\28	20\33	19\30	19\28
AR/ARMA	8\17	12\20	12\19	2\3
DWT	5\11	2\3	5\8	14\21
Statistics	6\13	9\15	5\8	13\19
Envelope Based				
AR/ARMA	4\9	0\0	3\5	1\1
DWT	3\7	7\12	3\5	1\1
Novel	3\7	3\5	5\8	3\4
Stats	0\0	1\2	0\0	0\0
Classic/Haemo	1\2	4\7	5\8	3\4
Image Based				
Computer Vision	1\2	2\3	6\10	7\10
Novel	2\4	0\0	0\0	5\7
Total	46	60	63	68

Table 8.6 provides a breakdown of the number of selected features belonging to different groups, for each model. It can be seen from this that in each case approximately 70% of features are derived from the audio directly, 30% of these are either MFCC or GTCC features. Identical features are often extracted from the audio of a whole cardiac cycle, as well as from audio limited to the systolic portion. The number of features selected from each group (where both portions are analysed) is listed in Table 8.7. This table illustrates that features derived using both data portions are valuable, however, those corresponding to whole cardiac cycles are selected more often.

The feature types identified through this selection process are displayed within Figure 8.12. This diagram provides a breakdown of the selected features, with high-level information displayed on the left-hand side (e.g., data type used to extract features), and lower-level details on the right (e.g., more specific feature types). It can be seen from

Table 8.7: Breakdown of data portion used in selected features, where identical features are extracted from whole cardiac cycles and from systole.

Data Portion	Spectral	Image Features	Audio Features
Cardiac Cycle	43	17	11
Systole	31	5	8

this figure that the majority of selected features are extracted from the audio directly (as opposed to from an envelope or image). Only one 'traditional feature' was selected (belonging to the 'Classic Waveform' category in Figure 8.12. This corresponds to trans-systolic times and was selected for each classification task. Features derived from images were the least common, however, those that are selected include novel features derived

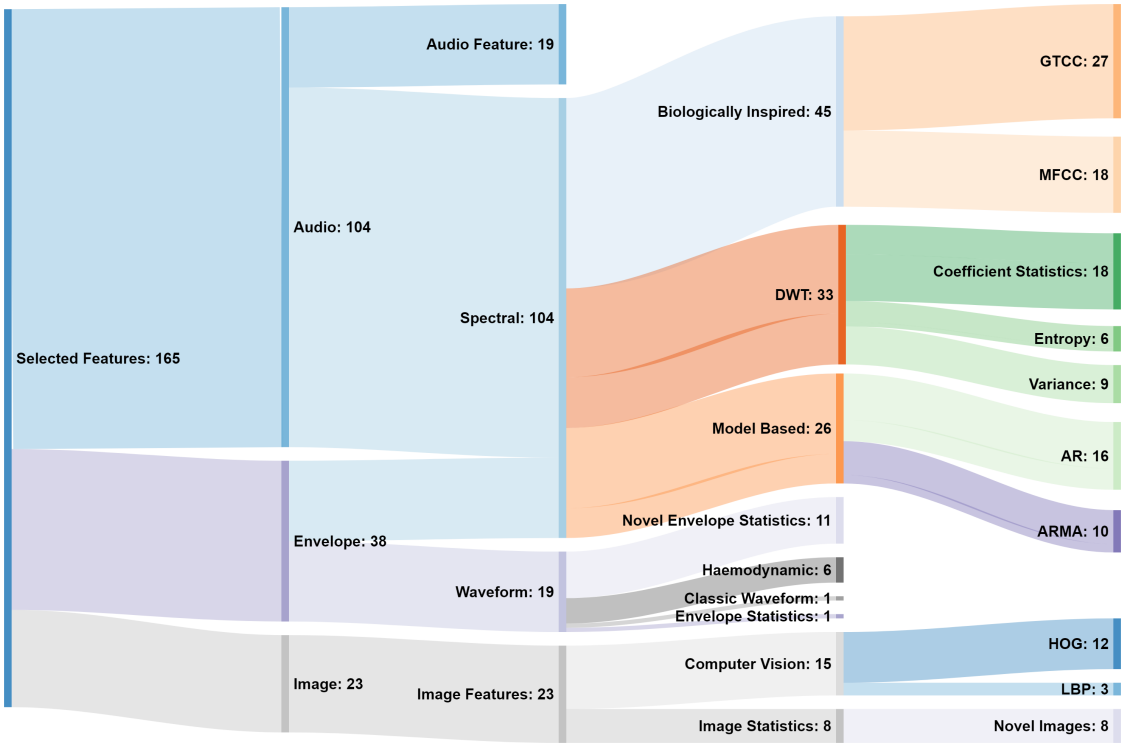


Figure 8.12: Sankey diagram displaying high-level details of all features selected for SVM localisation classification models

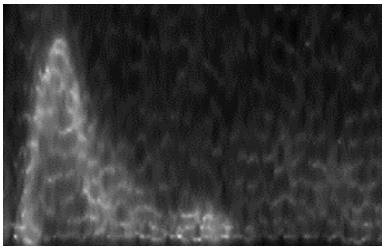


Figure 8.13: AR image type identified as good option for DCNN implementation

from regions of images defined by the MFE. Similarly, a variety of novel envelope features have been shown to be useful. These are derived using combinations of MFEs extracted via different methods.

The process of feature selection also identified the most suitable feature-based classification models. In each case, an SVM with polynomial kernel function was chosen. The 3-class horizontal model used a polynomial order of two, and all others used an order of three. Low-level details regarding these models are included in Appendix L.

8.6.2 AlexNet Images

Different image types were investigated with respect to AlexNet classification performance. This was done for the 2-class and 5-class scenarios using a selection of well-performing image types. With respect to AR images, an order of 40 was found to give the best performance. Model performance was compared using cropped ($AR40_C$) and whole ($AR40_W$) versions of these images, as well as the equivalent versions of OMM images (OMM_C and OMM_W) and spectrogram images with overlain MFEs ($Spect\_MFE_C$ and $Spect\_MFE_W$). The results from these comparisons are detailed in Table 8.8, as illustrated by this, AR40 images corresponded to the highest overall classification accuracies, and OMM images were found to give the second-best performance.

The performance of models using the AR and OMM images were compared for each task. It was found that the cropped AR images, whose dynamic ranges are compressed using an exponent fraction of 0.1, outperformed for each task; an example image is displayed in Figure 8.13. The classification results using these images are presented in the remainder of this chapter.

Having passed an image to a DCNN, the activations of different layers within the net-

Table 8.8: Overall accuracies of 2-class and 5-class AlexNet models for a selection of image types. Accuracies have been found using 10-fold cross validation for data from 10 participants.

	Classification Task	
Image Type	*2-class*	*5-class*
$AR40_C$	87	60
$AR40_W$	86	57
$AR40L^1_C$	84	58
$AR40L^1_W$	80	57
OMM_C	83	49
OMM_W	82	50
$Spect\_MFE_C$	68	50
$Spect\_MFE_W$	70	51

1 – Displayed using log-power.

work can be visualised. This can provide an indication as to which features the network is learning at different points. In shallow layers, this often reveals simple aspects such as shapes and edges, whilst deeper layers learn more complex aspects. As DCNNs are largely comparable to a 'black box', such visualisations can be useful in indicating image characteristics that a DCNN is exploiting. Figure 8.14 displays a number of images corresponding to the strongest activations within each of the five convolution layers. These are presented as greylevel images, where darker regions indicate stronger negative activations, and lighter regions indicate stronger positive activations (i.e., the portions that are used most for classification).

It can be seen from Figure 8.14, that as the convolution layers become deeper, the visualised activations become less recognisable. The images highlight a number of regions that are monitored by operators when performing scans, for example, cessation and commencement of systolic flow. The first image in Figure 8.14 illustrates a strong negative activation on the rising slope of the Doppler profile and more positive activations on the downslope. Conversely, the second strongest activations for the first layer (image two) exhibits strong positive activations on the rising slope and more negative activations on the falling slope. Strong positive activations appear in a number of images on the rising

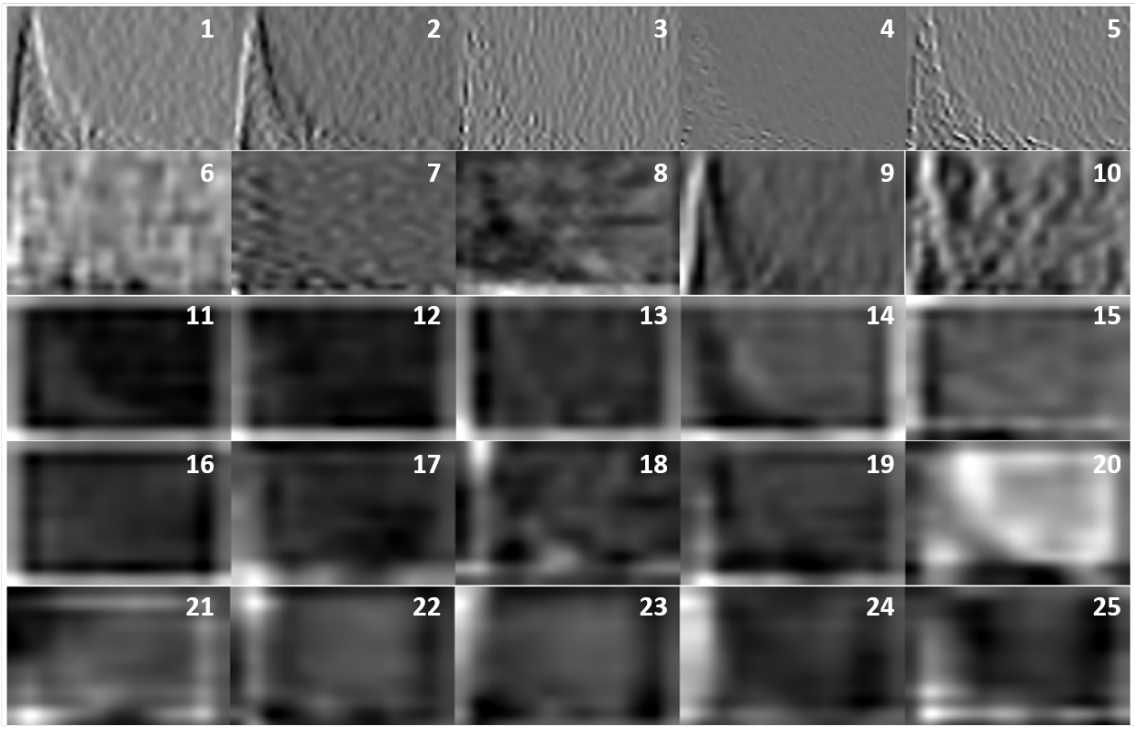


Figure 8.14: Visualisation of the top five strongest activations for each convolution layer within AlexNet using AR40 images, where the top row corresponds to the first layer, and the fifth row corresponds to the final layer.

slope, including images 13, 15, 18, 23 and 24. Images 11 and 25 appear to show positive activations covering the flow portion as a whole, whereas images nine and 20 appear to be more focused on the noise region of the image. Images 19 and 22 display strong activations near the systolic peak, conversely image 21 shows strong negative activations in this region and strong positive activations near the start and end of the cardiac cycle, where spectral content from valve clicks can occur.

8.6.3 2-Class Models

The results for the 2-class scenario will now be presented. The confusion matrices for the raw and MA results are displayed in Figure 8.15, this illustrates that the highest accuracy corresponds to the latter. These matrices also highlight the imbalance between the G and O classes (with an approximate ratio of 2:5), this results in trained models exhibiting a

natural bias towards O class. Section 9.2 describes the synthetic minority oversampling technique (SMOTE). This technique can be used to train an SVM model with reduced bias (SVM_{smote}). This is achieved by introducing artificial observations corresponding to the minority class. This was applied to the 2-class SVM training data, using 100% oversampling; giving a balanced training dataset. The corresponding results are displayed in Figure 8.15.

		Actual	
		G	O
Predicted	G	1953	417
	O	855	6395

(a) SVM raw

		Actual	
		G	O
Predicted	G	1997	329
	O	811	6483

(b) SVM MA

		Actual	
		G	O
Predicted	G	2317	869
	O	491	5943

(c) SVM$_{smote}$ raw

		Actual	
		G	O
Predicted	G	2358	776
	O	450	6036

(d) SVM$_{smote}$ MA

		Actual	
		G	O
Predicted	G	2015	512
	O	793	6300

(e) AlexNet raw

		Actual	
		G	O
Predicted	G	2118	343
	O	690	6469

(f) AlexNet MA

Figure 8.15: Confusion matrices for 2-class classification task using SVM, SVM_{smote} and AlexNet models

The confusion matrices displayed in Figure 8.15 have been used to calculate additional performance metrics. These are summarised in Table 8.9.

As can be seen from Table 8.9, the MA results consistently give a small improvement. The AlexNet model resulted in the highest overall accuracy of 88.5% for the MA results.

Comparatively, the SVM models resulted in a significantly lower G classification accuracy. This was improved using SMOTE oversampling, which resulted in more balanced TPR and TNR values for the SVM model. However, improved TPR is at the cost of

Table 8.9: Performance metrics for 2-class models, for SVM and AlexNet models

Model Type	Results	TPR (%)	TNR (%)	PPV (%)		Accuracy (%)
				G	O	
SVM	Raw	69.6	93.9	82.4	88.2	86.8
	MA	71.1	95.2	85.9	88.9	88.1
SVM $_{SMOTE}$	Raw	82.5	87.2	72.7	92.4	85.9
	MA	84	88.6	75.2	93.1	87.3
AlexNet	Raw	71.8	92.5	79.7	88.8	86.4
	MA	75.4	95	86.1	90.4	89.3

reduced TNR, and a slight reduction in overall accuracy.

Performance is explored further by considering metrics associated with individual participants. The results for the SVM_{smote} and AlexNet models are displayed in Figure 8.16. In this figure, the complementary TNR percentage, false positive rate (FNR), is included. This allows easier comparisons, and is an intuitive value when considering clinical implications, as discussed in Section 8.7.4.

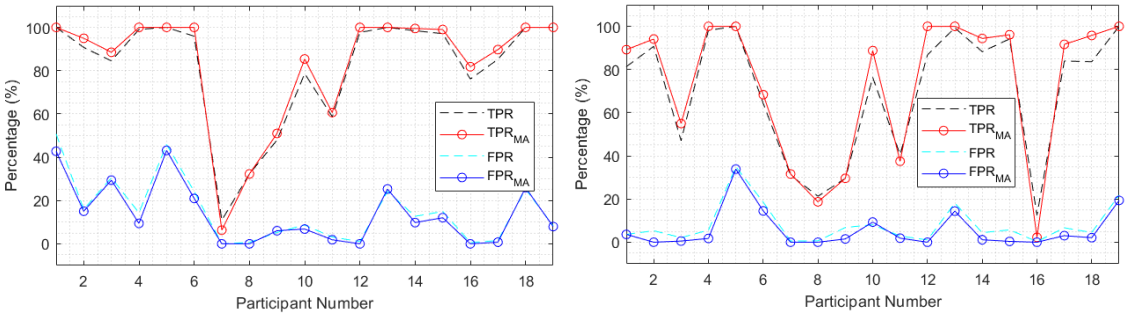


Figure 8.16: TPR and FPR percentages for individual participants for the 2-class scenario, for the a) SVM_{smote} model and b)AlexNet model

8.6.4 3-Class Models

The IPP is found by restricting scanning motions to distinct planes, allowing the operator to systematically hone in on the target signal. To reflect this scanning procedure, two types of 3-class model have been investigated, one for scanning along the vertical plane (from patients head to patients feet), and one for scanning along the horizontal plane (from patients left to patients right), in both cases the IPP is also included to give three distinct classes. These scanning planes are illustrated in Figure 4.5. The results from these models will now be presented. The confusion matrices for each SVM and AlexNet 3-class model are displayed in Figures 8.17 and 8.19, for the vertical and horizontal scanning planes, respectively.

		Actual		
		F	G	H
Predicted	F	1240	203	368
	G	149	2166	299
	H	329	439	1094

(a) SVM raw

		Actual		
		F	G	H
Predicted	F	1302	157	331
	G	123	2266	264
	H	293	385	1166

(b) SVM MA

		Actual		
		F	G	H
Predicted	F	1105	120	282
	G	245	2247	280
	H	368	441	1199

(c) AlexNet raw

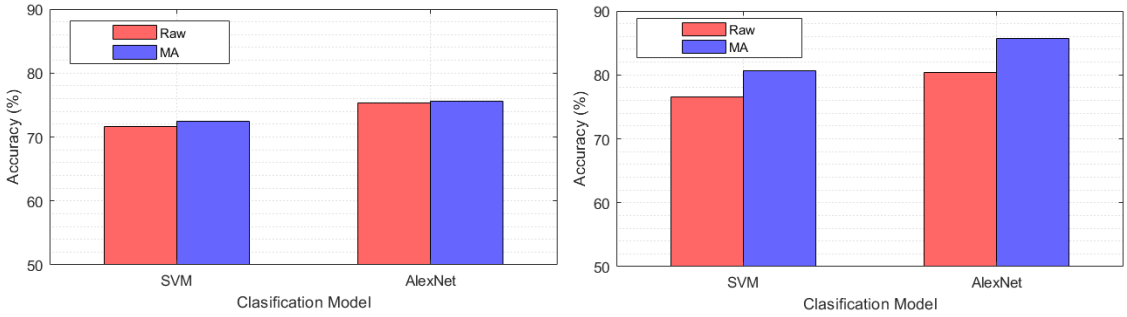
		Actual		
		F	G	H
Predicted	F	1131	61	251
	G	246	2340	227
	H	341	407	1283

(d) AlexNet MA

Figure 8.17: SVM and AlexNet confusion matrices for the 3-class, vertical-scanning-plane classification task

Figure 8.17 demonstrates that overall the SVM and AlexNet models correctly classify a similar number of observations, with them successfully categorising 4,734 and 4,754 cardiac cycles, respectively. However, despite this overall similarity, the number of correct classifications for specific classes exhibit large disparities, for example, the SVM MA correctly classified 171 more F observations than the AlexNet model.

The confusion matrices in Figure 8.19 show that overall the AlexNet model was significantly more successful than the SVM model, correctly classifying 222 more observations for the MA results. This difference is mostly due to the AlexNet model being better at



(a) Vertical scanning plane

(b) Horizontal scanning plane

Figure 8.18: Classification accuracies for the 3-class models

		Actual		
		L	G	R
Predicted	L	1238	223	203
	G	72	2208	268
	R	299	377	1253

(a) SVM raw

		Actual		
		L	G	R
Predicted	L	1257	184	134
	G	63	2334	242
	R	289	290	1348

(b) SVM MA

		Actual		
		L	G	R
Predicted	L	1254	102	229
	G	132	2464	259
	R	223	242	1236

(c) AlexNet raw

		Actual		
		L	G	R
Predicted	L	1327	43	116
	G	98	2575	247
	R	184	190	1361

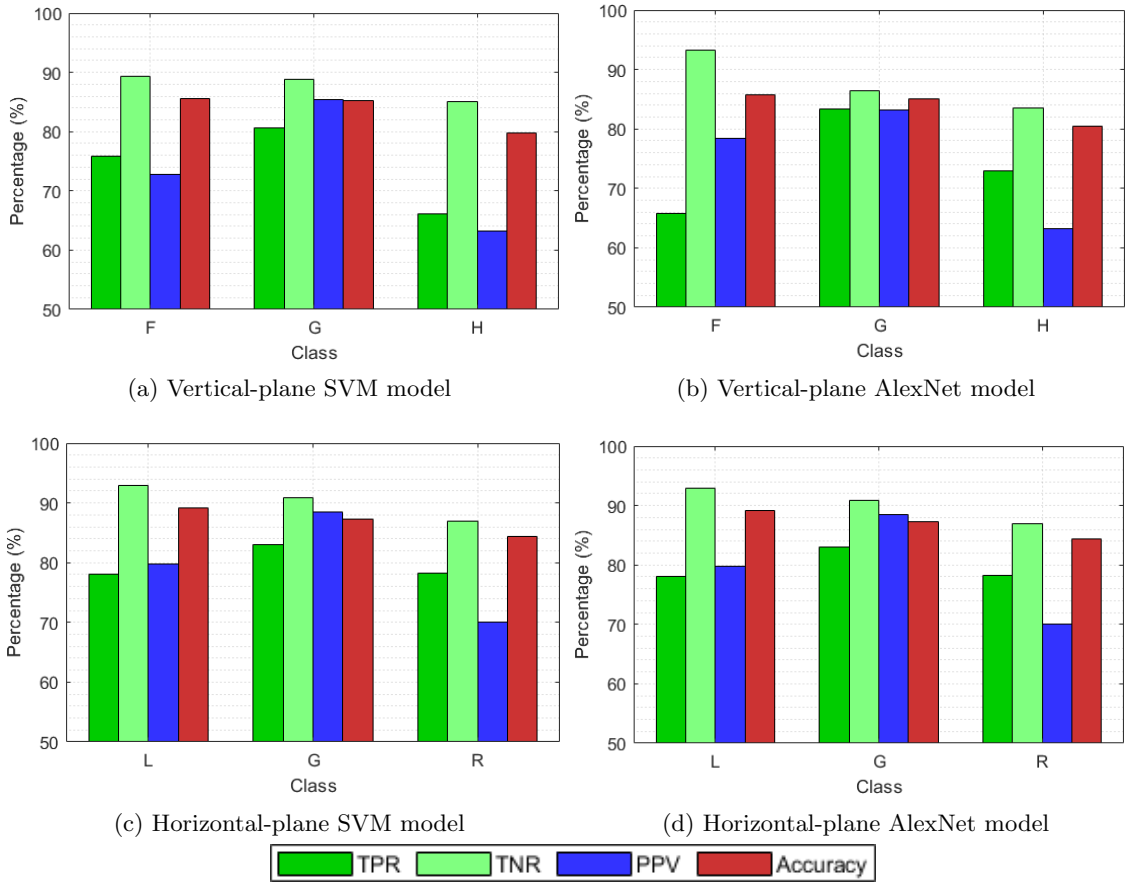
(d) AlexNet MA

Figure 8.19: SVM and AlexNet confusion matrices for the 3-class, horizontal-scanning-plane classification task

classifying G observations. The difference in overall classification accuracies displayed in Figures 8.17 and 8.19 are displayed in Figure 8.18.

It can be seen from Figure 8.18 that the MA predictions are consistently more accurate. Furthermore, AlexNet achieved the highest classification accuracies for each 3-class models, exhibiting an increase of 0.3% and 5% for the vertical and horizontal planes, respectively. The performance metrics for each set of MA predictions, and each scanning plane are displayed in Figure 8.20.

It can be seen from Figure 8.20, that overall the two modelling methods give compa-



(a) Vertical-plane SVM model

(b) Vertical-plane AlexNet model

(c) Horizontal-plane SVM model

(d) Horizontal-plane AlexNet model

Figure 8.20: Performance metrics for 3-class scenario using SVM and AlexNet MA predictions, for vertical and horizontal planes.

rable performance metrics for the vertical scanning plane. As indicated by the confusion matrices, the SVM model, with respect to AlexNet, resulted in TPR values approximately 10% greater for the F class and -6% for the H class. Despite the reduced TPR% for the F class, AlexNet's predictions for this category were more precise compared to the SVM.

With respect to the horizontal-plane, overall AlexNet outperformed the SVM model. In every case, AlexNet achieved higher performance metrics, apart from the PPV and TNR metrics for the G class, where the SVM gave marginally increased values of 0.3% and 1.2%, respectively. The best off-angle performance is achieved by the AlexNet model for classifications of the L class, furthermore, the associated precision of 89% is the highest

observed for all results presented in Figure 8.20. The average precision and accuracy for the off-angles of this model were 10% greater than any other 3-class model.

8.6.5 5-Class Models

The most challenging classification task investigated is the 5-class scenario, where each probe position sampled is included as a target category. This section presents the results for this task. The confusion matrices for the SVM and AlexNet model are displayed in Figure 8.21.

		Actual				
		G	L	R	F	H
Predicted	G	2041	69	202	95	217
	L	98	908	112	210	202
	R	225	172	739	304	533
	F	115	235	209	966	175
	H	329	225	462	143	634

(a) SVM raw

		Actual				
		G	L	R	F	H
Predicted	G	2229	50	228	71	181
	L	71	987	47	209	167
	R	140	157	834	268	597
	F	70	217	157	1056	170
	H	298	198	458	114	646

(b) SVM MA

		Actual				
		G	L	R	F	H
Predicted	G	2247	87	225	219	306
	L	79	1093	129	276	294
	R	173	89	760	195	399
	F	69	176	200	893	210
	H	240	164	410	135	552

(c) AlexNet raw

		Actual				
		G	L	R	F	H
Predicted	G	2422	72	235	242	309
	L	37	1184	67	233	313
	R	155	42	886	181	336
	F	22	166	177	967	197
	H	172	145	359	95	606

(d) AlexNet MA

Figure 8.21: Confusion matrices for the 5-class classification task, for raw and MA predictions using SVM and AlexNet models.

It can be seen from Figures 8.21 that the ability to distinguish specific angles varies, with both model types performing well for the G category, and worse for the F category. The SVM model resulted in more correctly classified observations for the vertical classes H and F.

The 5-class results demonstrate that neither modelling type, AlexNet or SVM, consis-

Table 8.10: Ensemble methods investigated for 5-class scenario

Model Name	Description
$AlexNet_{c+w}$	Combined AlexNet models trained using cropped and whole images
$AlexNet_c + SVM$	Combined AlexNet model, trained using cropped images, and the SVM model
$AlexNet_{c+w} + SVM$	Combination of $AlexNet_{c+w}$ and SVM model

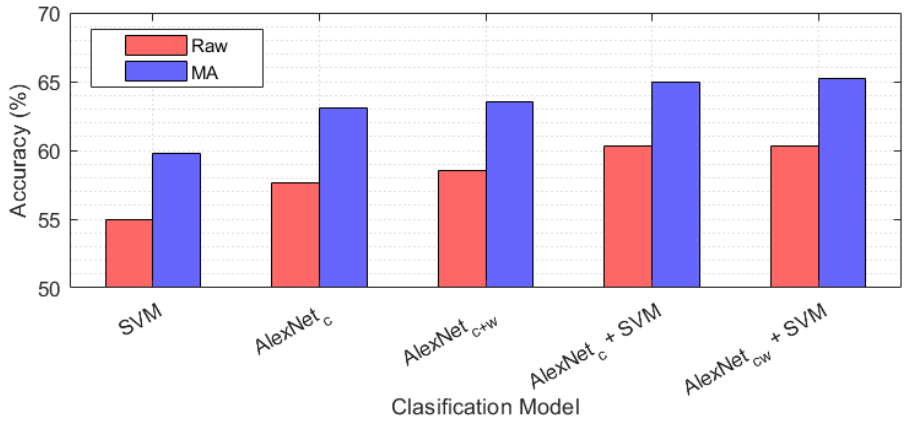


Figure 8.22: Classification accuracies for SVM, AlexNet and ensemble models, for raw and MA results.

tently outperforms the other for each category. Ensemble methods have been investigated as a means of giving an improved and more balanced performance. The results from three different model combinations are presented, the individual models included in each ensemble are detailed in Table 8.10. The overall accuracies for each of these ensemble models are displayed in Figure 8.22.

As with the 2-class and 3-class results, Figure 8.22 illustrates that MA results are consistently more accurate, and AlexNet outperforms the SVM model. Ensemble models were found to improve overall classification accuracy and typically resulted in improved performance with respect to different metrics (Figure 8.24). The highest overall classification accuracy of 65% was achieved by the $AlexNet_{cw} - SVM$ MA ensemble, the confusion matrix for these results are displayed in Figure 8.23.

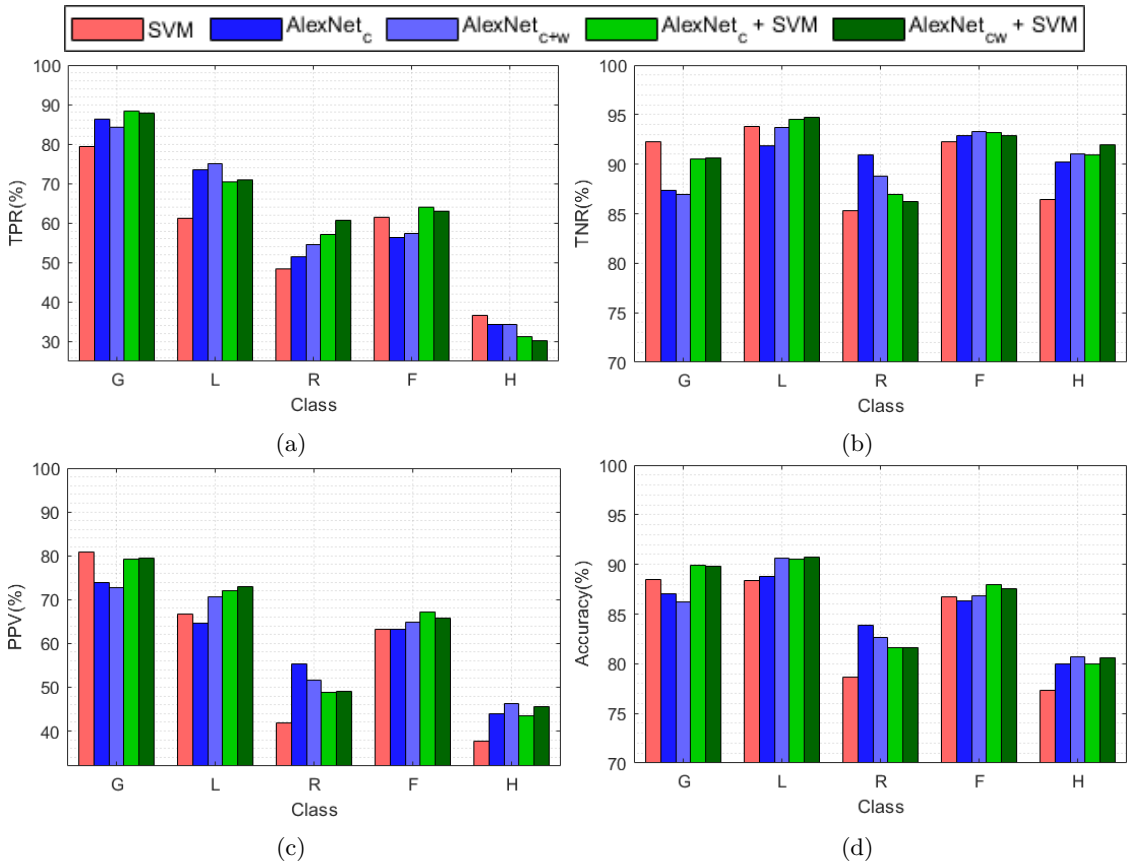
Predicted \ Actual	G	L	R	F	H
G	2310	73	201	154	244
L	67	1033	68	214	201
R	163	134	927	267	590
F	69	201	194	987	182
H	199	168	334	96	544

(a)

Predicted \ Actual	G	L	R	F	H
G	2467	54	206	152	224
L	31	1143	15	195	184
R	116	112	1048	211	649
F	41	166	184	1084	173
H	153	134	271	76	531

(b)

Figure 8.23: $AlexNet_{cw} - SVM$ ensemble confusion matrices for the 5-class classification task for a) raw results and b) 4 s moving-average results



(a) (b) (c) (d)

Figure 8.24: Performance metrics for 5-class scenario MA predictions using SVM, AlexNet and ensemble classification models. Displayed metrics are a) TPR, b) TNR, c) PPV and d) Accuracy.

8.7 Discussion

The results presented in the previous section, and their implications, will now be discussed.

8.7.1 Class Imbalances

Class imbalances are a common occurrence in classification tasks. In the presented work class imbalances have been introduced as a result of three factors. Firstly, the data collection process resulted in more examples of the IPP (i.e., the G class) being collected. As described in Section 4.3, each angle explored is measured in reference to the IPP position, meaning with respect to any other position the IPP is sampled five times more often. In response to this, the time spent at the off angles was marginally longer.

Secondly, the same data has been used for various classification tasks. For example, in the case of the 2-class scenario, G is actually the minority class as all the off-angle categories are combined.

Finally, the likelihood of failed beats due to poorly defined cardiac cycles or noise is not consistent for each category. The IPP is inherently the most likely to result in high-quality measurements, and therefore reduced failures. Conversely, it was found that measurements collected whilst the probe was angled towards the patients left (L class) were more likely to give Doppler profiles leading to incorrectly segmented beats. These are summarised in Table 4.5.

In cases where these imbalances are substantial, methods of oversampling or undersampling observations can be employed to adjust classification bias. This was performed for the 2-class scenario, where the imbalance was most significant. In this case, SMOTE oversampling was employed for the SVM model, and resulted in a more balanced performance with respect to TPR and TNR, however, resulted in an overall reduction in classification accuracy. This tradeoff is discussed further in Section 8.7.4.

8.7.2 SVM Features

Features were selected for each SVM model separately. The selected features were found to predominately belong to novel groups, meaning they either have not been applied to Doppler ultrasound previously or have been designed as part of this work.

Those which have previously been documented include features derived using the

DWT. These are extracted from the Doppler audio and include statistics based upon wavelet decomposition and entropy. These have been shown to be useful for classifying heart valve disease from Doppler heart sounds [172]. These features were selected 12 times across the models, including the second best-ranked feature for the 2-class scenario.

In the presented work, these same techniques were applied to extracted MFEs. These features were also selected 12 times across the models and were ranked first for the 2-class, and 3-class vertical models (`DWT_CC_E_OMM_D5_3` and `DWT_CC_E_MSNSI_D5_3` respectively). In both cases, the same statistic is identified, however, the MFEs used were extracted from different MFE estimation methods (OMM and MSNSI). The feature selection process removes highly correlated features, this process identifies several common features which are extracted using different MFE estimation methods. This is intuitive as the envelopes are often very similar, examples of such features are CO, ejection percentage and flow time. However, `DWT_CC_E_OMM_D5_3` and `DWT_CC_E_MSNSI_D5_3` are not highly correlated, having a correlation coefficient of 0.37; highlighting the difference between these features and that they are not interchangeable. These results illustrate the usefulness of these features for scan position classification. Furthermore, highlighting the importance of MFE estimation method with respect to feature value. These observations are consistent with respect to the selected features for the continuous feedback model.

Section 3.6.1.1 introduced parametric means of spectral estimation, including the AR method. Previous works have shown AR derived PSDs as being useful for diagnoses of Doppler signals [395]. As with the DWT features, statistics were derived from these transformations and included in the given work. These features were ranked particularly high for the 2-class scenario but were found to be much less pivotal for the remaining three models. This may indicate that these features are good at discriminating the G class from all other classes, but are poor at discriminating between off-angle classes. Similar features derived using ARMA models and from MFEs were also shown to be useful. These spectral features were extracted both from the systolic portion of the input data and from the whole cardiac cycle. In some cases, it was found that both portions were valuable, for example, `AR_SYS_A4` and `AR_CC_A4` are both ranked within the top 20 for the 2-class model.

The majority of AR features for the 2-class scenario were extracted from systolic portions of the audio, however, overall selected features were more likely to be derived using the whole cardiac cycle. The inclusion of features from both data regions, and the fact that both regions produce highly ranked features across the SVM models, demonstrate that

extracting features from multiple regions can be advantageous. As discussed in Section 7.9.1, this could be extended to included features derived solely from diastole. Furthermore, feature extraction from bi-directional flow could effectively double the number of available features (as the features from reverse flow could also be analysed). Future work using alternative Doppler devices, and extended feature extraction algorithms would be required to investigate this.

Envelope features were extracted using different types of waveform, and different MFE estimation methods. Those selected include features derived using the OMM and MSNSI method. In cases where different MFE types were used, features derived using OMM MFEs were selected approximately three times more often than the MSNSI method (18 compared to seven). This may be due to the high stability of the OMM MFE, which is illustrated within Chapter 5. In four of the cases where MSNSI were selected, the same features using the OMM MFE were also included. Interestingly, the only CO features included were derived using the MF envelope. This may be due to this envelope being less affected by spectral broadening and spurious signals. The selection of features derived using different envelopes illustrate that envelope estimation methods affect the usefulness of a given feature, and that combining features using different methods can strengthen predictive models. This factor is often overlooked in previous works that include MFE based features, in which either the MFE estimation process is not explicitly defined [197,199,200,202,396], extracted using third party software [193,340–342] or manually traced [356].

The potential benefit of using multiple MFE estimation methods is highlighted by selection of novel envelope features, of which 11 different types were selected across the models. As described in Section 7.3.3, these were included as a means of exploiting the differences in MFE estimation methods.

As with the regression model, MFCC and GTCC features were found to be very valuable. These spectral features have not previously been reported as a means of investigating Doppler ultrasound. The feature selection process resulted in 19 MFCC and 27 GTCC features being selected across the models. Notable examples include `MFCC_SYS_A2`, which occurs in the top five ranked features for four models and was also ranked first for the regression model (Section 7.8.1). These results are significant and could be impactful for other classification tasks, such as disease classification.

8.7.3 AlexNet Images

The input data for the explored DCNNs are images. As these images are passed through the layers of the network, features are extracted at increasingly high levels. Converse to the hand-crafted models, additional steps of feature selection are not required. Previous works detailing DCNN classification of spectral images have been limited to single types of images, namely STFT spectrograms. Furthermore, the parameters used to display these images are not explicitly defined [289].

In the given work, different image types were explored and found to be a significant factor in the resulting performance. This process identified images constructed using AR spectral density estimates as being a good choice for the given classification tasks. This finding is consistent with the selection of SVM feature subgroups, which identified AR features as being valuable.

Different model orders were used to generate AR images. A model order higher than those used to extract AR features was found to give improved performance. These findings could be considered in future works using handcrafted AR features, by investigating features derived using higher-order models. In the given investigation, the orders used when extracting features were guided by previous works.

Having identified AR images as being promising, performance was improved further by cropping images with respect to maximum OMM MFE value. This observation is intuitive as more of the image contains signal, and corresponding details are preserved in subsequent image resizing. In addition to this, the approach used to display spectral estimates was found to impact performance. This is necessary as spectral estimates contain a large dynamic range, and necessitate adequate rescaling to be fully visible. The differences in model performance, with respect to image choice, can be exploited using ensemble models. This was demonstrated for the 5-class scenario, where predicted probabilities using cropped and un-cropped AR images were combined to give a small increase in overall accuracy. It is plausible that given other images or classification tasks, this approach could be very productive.

The evaluation of different image types can be extremely computationally expensive, both in terms of constructing the images, and then training and evaluating resulting models. This is further exacerbated when investigating multiple DCNN models, which may not be consistent with respect to the best performing image type. However, considering the

observed differences with respect to image types, investigations into different methods of displaying and estimating spectral information could be highly beneficial in future works, with respect to both related and unrelated disciplines.

8.7.4 Comparison of Classification Tasks

The overall accuracies associated with each classification task varies as expected, with the 2-class models achieving the highest values, and the 5-class models achieving the lowest. Considering only the best performing models for each task, the reduction in overall accuracy with respect to the 2-class scenario was 3.6%, 13.6% and 24.1% for the horizontal 3-class, vertical 3-class, and 5-class tasks, respectively.

With respect to the 3-class scenarios, the vertical scanning plane was clearly more challenging than the horizontal scanning plane. The results indicate this is partially due to difficulties in classifying the H class. With respect to the other categories, the associated precision for H was the lowest for each model type, with the best performing vertical 3-class task and 5-class task scenarios achieving 63% and 46%, respectively.

Precision and accuracy of G classifications were good for each classification task, achieving percentages greater than 80%. This is most significant in the 5-class scenario, where an accuracy and precision of 90% and 80% were obtained. This is promising for real-world applications, in which automatic detection of the IPP would be particularly useful. Considering the use of such models for detecting the IPP in clinical settings, two aspects are of particular concern with respect to practicality.

Firstly, the FPR of G category classification. The FPR, which is the complementary percentage of TNR, describes how often incorrect classes are classified as the category of interest. In practice, this would be the rate at which an operator incorrectly receives feedback stating the IPP had been found. Clearly this would be undesirable, as a scan may be stopped prematurely, and true values not recorded. The FPR for G ranges from 5% in the 2-class scenario to 14% in the vertical 3-class scenario.

Secondly, the TPR of G category classification, which in practice would represent the rate at which an operator receives feedback correctly stating the IPP had been found. Lower values of TPR would result in operators requiring more examples of the IPP before obtaining confirmatory feedback, or in the worst-case scenario not receiving confirmation at all. Inspecting the TPR rates for individual participants, which form the cross-

validation results presented, there was no instance where G TPR was zero. However, the value of TPR does vary significantly across the participants, exhibiting a STD of 29% for the 2-class scenario.

The extent to which TPR and FPR varies for participants is displayed within Figure 8.16. This provides a more detailed comparison between the SVM_{smote} and AlexNet model. There are clear differences between the models. For example, the AlexNet model performed very poorly in the case of P16 with respect to TPR, whereas SVM_{smote} achieved a value greater than 80%. For approximately 80% of participants, the SVM_{smote} model achieved a TPR greater than 80%. The associated STD, however, is still relatively high at 24%. This is likely a reflection of the variability in IPP scans, as discussed in Section 4.4, limiting the generalisability of models in response to measurements from certain participants. Implementing the SMOTE technique increased accuracy and reduced variability for SVM results.

In clinical settings, the generalisability of models would be a greater factor, as unhealthy patients are more likely to have abnormally weak flow rates or unusual flow patterns. To account for these variations, it is likely a very large labelled data set, for healthy and unhealthy patients would be required (from which more generalised models could be formed).

The 3-class and 5-class results both indicate that the horizontal-plane off-angles were easier to categorise than those in the vertical-plane. This could be due to off-angle measurements being more distinctive for the horizontal case. Considering the ascending aorta, vertical-plane variations are likely to continue capturing aortic blood flow (albeit from a different region such as the aortic arch). Intuitively, horizontal-plane variations are more likely to insonify less of the aorta and capture more significant contributions from other blood flows. Furthermore, the difference in flow contributions may be more significant in the horizontal-plane, as different flows associated with the heart are captured. Further data collection using a duplex ultrasound machine would be needed to investigate and substantiate these potential differences.

The results illustrate that there are scan characteristics that can be used to differentiate probe positions. Features and images have been generated using a very basic directional Doppler signal and from a relatively small pool of participants. A combination of increased data, additional information (e.g., images from bi-directional Doppler signals) and ensemble methods could enable more powerful Doppler classification models.

8.7.5 AlexNet vs SVM

AlexNet outperformed the SVM models for each classification task in terms of overall accuracy. These results clearly demonstrate that state-of-the-art image recognition techniques can be applied to clinical Doppler ultrasound signals. In addition to improved performance, DCNN methods offer a number of benefits.

With respect to the knowledge and skill needed to build large handcrafted feature sets, spectral images can be generated quickly and easily. This enables practical model creation and implementation, and given the ongoing development of DCNN technology, makes this approach very promising for further classification applications (such as diagnosis of heart disease). These findings are a significant outcome of this work, and shed light on a new method of analysing Doppler ultrasound.

The reduced complexity in terms of generating images also means real-time DCNN feedback is more computationally efficient. However, training DCNN models is substantially more costly. This is of particular significance when performing cross-validation, contrasting DCNN models or using larger datasets. SVMs are far quicker to train and evaluate, despite the slow process of feature selection, and hyperparameter optimisation. Having selected features, SVM cross-validation can be performed in a matter of minutes, comparatively the same evaluation could take days using DCNNs.

The results demonstrate that both approaches do not consistently differ in performance across the tasks, for example, the SVM model outperformed AlexNet for the F and H class in the 5-class scenario, despite performing worse overall. These differences can be capitalised upon using ensemble methods. This was demonstrated using a straightforward approach which combined prediction probabilities from various models. The SVM and AlexNet AR40_C models were combined; independently these achieved overall accuracies of 60% and 63%, respectively. Comparatively, the ensemble method achieved 65%, illustrating the benefits of ensemble models. The benefit of these small improvements would need to be considered in the context of computational costs. Further improvements may be feasible through combining different model types or employing more advanced methods of probability combination, for example, by weighting different contributions.

8.7.6 Clinical AI Tools

The results presented in this and the preceding chapter provide quantitative evaluations of designed feedback models. The techniques and features highlighted through this work could ultimately contribute, and lead, to AI tools being designed for clinical implementation. This section briefly discusses the considerations and steps needed to work towards this goal.

Continued research in this field and the design of AI tools would necessitate clinical validations, i.e., clinical verification of their usefulness. The ultimate validation, in this respect, is in terms of added value to patient outcomes [397]. Work towards this goal would need more substantial and varied data collection to prevent implications of spectrum bias. This would be further strengthened through using research ultrasound machines, which would give researchers greater control over software and hardware and be unrestricted in terms of accessible data (such devices are discussed in Section 2.5.4). This is because current models have been trained and evaluated using data limited to healthy adult participants, which does not reflect the spectrum to whom AI tools would be applied to in clinical scenarios [397]. This is of particular concern considering the high variability present in measurements, and that the value of AI tools would be greater in certain conditions (e.g., in patients who are subjectively harder to scan).

Several works have documented clinical AI tool verification, these are largely concerned with automatic image diagnoses, for example, automatic detection of lesions in endoscope images [398]. In this case, the value of the AI tool can be measured in terms of the time taken to manually diagnose images and the number of correctly identified lesions. The value of the tool is measured by comparing the examination results of trainees using the tool, with results from trainees not using the tool. The accuracy of associated measurements is then assessed with respect to a gold standard established by expert examinations. This work also found that the tool's value must be considered in the context of operator experience, with its value being greater for more novice users.

The validation of tools to assist Doppler ultrasound measurements would need to consider these aspects also, and ensure both the number of operators involved and patients examined were broad enough, and large enough, to capture the software's usefulness. This is to overcome the high inter and intrapatient variability, and differences in operator skill level. Furthermore, the effect of repeat measurements would need to be considered, as

monitoring is a valuable application of Doppler ultrasound, and as illustrated in Section 4.4, the ease of acquiring measurements can reduce in repeated measurements.

In a similar manner to the above example, the performance of such software could also be assessed with respect to time (e.g., time taken to locate the IPP). However, subjective measures are also important for the given application. A factor that inhibits current use of the technology is a lack of confidence in acquired measurements, as performing scans can be challenging and current devices cannot give positive confirmation that the IPP has been located; this is discussed in Section 2.7. Considering this, the confidence such software could provide to an operator is a valuable metric for its assessment.

However, reduced examination times, and increased confidence levels have limited value if corresponding measurements are not accurate. The accuracy of measurements, therefore, would need to be validated using reference values. A gold standard, however, cannot be established in a straightforward manner (such as in medical image classification [398]). This is because the haemodynamic characteristics of a patient are non-stationary, and as discussed in Chapter 4, are affected by multiple factors such as body position and heart rate. This means establishing the accuracy of a measurement (e.g., in terms of CO), is limited, even when using sequential measurements with an expert. To reduce error in this regard, multiple 'back-to-back' trainee and expert measurements could be performed, however, the impact of repeat measurements and the considerable time and cost requirements could dictate the practicality of such approaches. In addition to this, trainees would need to be blinded to the 'target' values. Ideally, a secondary continuous means of haemodynamic monitoring would be employed, which would also enable evaluation of operators considered to be more experienced; available technologies are discussed in Section 2.3.

Further determinants of AI tool value include psychometric properties, namely the perceived usefulness and the perceived ease of use [397, 399]. Feedback with respect to these factors would likely be used to inform adjustments to applications and diagnose issues in terms of user acceptance [399]. These, for example, are not necessarily technical, rather could be in terms of how information is presented, interpreted and acted upon; which can significantly impact the ultimate value of the tool [397]. A likely minimum requirement, with respect to ease of use, would be to combine feedback software with the ultrasound hardware (i.e., so they are displayed on a single machine). The constraints of collecting data and combining software with existing commercial hardware makes the use

of research-orientated ultrasound machines more desirable (these are discussed in Section 2.5.4). Such machines would give researchers complete control over data collection and processing, and enable hardware and software to be tailored for the given task.

8.8 Conclusion

A series of models have been trained to predict probe positions for four different classification scenarios of varying difficulty. This goal of providing localisation feedback for Doppler ultrasound has not previously been addressed.

Two primary types of model were investigated. These were state-of-the-art image recognition methods, which have previously not been applied to Doppler ultrasound, and SVM models using both established and novel hand-crafted features. The results demonstrate that trained models are able to distinguish between different probe positions. Furthermore, it was found that the image recognition method, AlexNet, outperformed the SVM models, and that ensemble methods can lead to further improved accuracies. The AlexNet and the SVM model achieved an overall accuracy of 86% and 80%, respectively, when differentiating between a patients left, patients right and the IPP.

Chapter 9

Beat Rejection

9.1 Introduction

The beat segmentation process is automatic, and as illustrated in Section 6, can occasionally produce incorrectly segmented beats (fail beats). These beats are removed prior to training feedback models and can arise due to, for example, excessive noise in recordings, Doppler profiles diverging excessively far from their characteristic shape or portions of recording not containing blood flow. These conditions can occur when the probe is far from the IPP, undergoes sudden movements (i.e., flash artefacts [12]) or is poorly coupled to the patient's skin.

When exposed to fail beats, trained models can generate misleading feedback. Real-time application of the calibration models, therefore, need to automatically replicate the process of identifying and removing fail beats. To address this issue, a classification model has been trained to classify fail beats, which in practice would ensure only correctly segmented beats are used to generate feedback; these beats are referred to as 'pass beats'.

As detailed in Section 4.4, fail beats are identified and removed prior to training models. These identified beats are used to generate a labelled data set, and subsequently, train an SVM classification model. The features selected for this purpose are less populated by spectral based features, and include more image and novel envelope features. To account for class imbalances, which can introduce prediction bias, artificial observations have been introduced into the dataset. The main contributions presented in this chapter are:

- The identification of promising novel features for identifying fail beats.

- The design and evaluation of models trained using these features.
- The evaluation of sampling methods to overcome biases introduced from unbalanced datasets.

9.2 Class Imbalance

In normal instances, the majority of beats are correctly segmented. Due to this, the categories of pass and fail are not equally represented (i.e., there are many more pass beats). This is a common occurrence in real-world data sets, which tend to be composed mostly of 'normal' observations, and a small sample of 'abnormal' observations. This is true for many diagnostic classification tasks, where there is an abundance of observations without a particular disease, and relatively few exhibiting it. Furthermore, misclassifying the abnormal observations as normal is often more costly than the reverse [400].

This imbalance can lead to poorly trained models and misleading performance statistics. This is highlighted by considering a model that always predicts observations as normal. In such a case, an imbalanced data set with 90% normal observations would still give a classification accuracy of 90%. Additional steps of oversampling the minority class, or undersampling the majority class can be used to overcome such class imbalances and remove bias from trained models. A highly successful method of oversampling is the SMOTE technique [400].

9.3 The Synthetic Minority Over-sampling Technique

SMOTE is a popular oversampling method. This technique, as the name suggests, synthetically generates samples using the minority class. It was demonstrated that combining SMOTE oversampling with majority class undersampling can lead to improved results [400]. This technique has been used in the presented work, allowing the performance of the trained model to be tuned with respect to fail and pass classification accuracy.

Over-sampling is performed by introducing synthetic examples at points along lines joining the k nearest neighbours of the minority class within the feature space. The number of neighbours used is determined by the chosen oversampling percentage. For example, if 300% over-sampling is needed, three out of the k nearest neighbours will be randomly

selected. New samples are then created in the direction of each of these neighbours. The points used along these lines are randomly found (i.e., the total distance is multiplied by a random number from 0 to 1).

Undersampling of the majority class is achieved by randomly removing majority class observations. The amount removed depends on the specified undersampling percentage, which is expressed as the resulting percentage of minority classes with respect to majority classes (e.g., following 200% majority class undersampling, the minority class will contain twice as many observations as the majority class).

9.4 Evaluation Methods

This section describes the implemented methods for model construction, class balancing and model evaluation.

9.4.1 Fail Beat Classification Model

The methods outlined in Section 7.3 and 7.6 where used to generate a large feature matrix, perform feature selection and train models to classify fail beats.

9.4.2 Class Balancing

SMOTE is implemented to overcome the inherent limitations of the imbalanced data set. The degree of oversampling and undersampling used determines the performance and bias of resulting models. The balance of these two factors corresponds to a trade-off with respect to true positives (fail beats), and true negatives (pass beats). The percentages used were chosen to be in-line with previous works [400]. This range allowed the 'trade-off relationship' to be captured and used to guide model choice. This range of percentages is summarised in Table 9.1.

9.4.3 Model Evaluation

Models have been trained using the expertly collected data described within Section 4.3. These are then evaluated using TPR and TNR metrics, and with the corresponding Receiver Operating Characteristic (ROC) curve, which illustrates the trade-off between true positive and false positive error rates (AUC values are used to facilitate this) [400].

Table 9.1: Percentages of majority class undersampling and SMOTE oversampling used to identify final class-balancing training parameters

	Percentages (%)
SMOTE	*50 100 200 300 400 500*
Under-sampling	*10 15 25 50 75 100 125 150 175 200 300 400 500 600 700 800 1000 2000*

Statistics were calculated using the same cross-validation approach described in Section 7.7.2. This was performed for each possible under- and oversampling percentage combination.

To further evaluate the usefulness of the resulting model, the probe positions associated with fail and pass classifications are inspected. This allows the detrimental cost of these misclassifications to be assessed, for example, if a pass beat from an off-angle is misclassified as a fail beat, this would be less detrimental than misclassifying a pass beat from the IPP as being a fail beat.

9.5 Results

The selected features and performances of associated models, with respect to rebalanced datasets, will now be presented.

9.5.1 Selected Features

The feature selection process again identified an SVM model as being the best choice. The corresponding low-level details are given in Appendix L. The selected feature subset contains 51 features, these are listed in Table 9.2. Additional low-level details are provided in Appendix M.

9.5.2 Beat Classification Model

The percentages displayed in Table 9.1 were used to train multiple models using SMOTE oversampling, combined with undersampling. It was found that 300% SMOTE oversampling resulted in the best average TPR and TNR, as well as the highest AUC value.

Table 9.2: Features selected for fail beat model

Rank	Name	Rank	Name	Rank	Name
1	W_CC_E_NE5_10	**18**	HOG_SYS_I_F2	**35**	MFCC_SYS_A1
2	HOG_CC_I_F89	**19**	DWT_CC_A_SE6	**36**	MFCC_CC_A1
3	W_CC_E_NE5_4	**20**	DWT_SYS_A_SE6	**37**	DWT_CC_A_SE1
4	HOG_SYS_I_F321	**21**	HOG_CC_I_F201	**38**	GTCC_CC_A_F1
5	AF_SYS_A8	**22**	HOG_SYS_P117	**39**	HOG_SYS_I_F179
6	W_SYS_E_MF_PK1	**23**	MFCC_CC_A8	**40**	HOG_SYS_I_F50
7	W_CC_E_NE5_9	**24**	W_CC_E_MSNSI_S1	**41**	W_CC_E_OMM_H1
8	HOG_CC_I_F101	**25**	W_SYS_E_MF_PK3	**42**	MFCC_CC_A4
9	IMS_CC_CORR2	**26**	DWT_CC_E_MSNSI_noISB_V7	**43**	HOG_CC_I_F316
10	HOG_CC_P1	**27**	HOG_CC_I_F172	**44**	HOG_CC_I_F313
11	HOG_CC_I_F167	**28**	AR_SYS_A_C4	**45**	AF_CC_A3
12	IMS_CC_I_SNR_PEAK	**29**	IMS_SYS_I_IP4_S4	**46**	AF_CC_A7
13	W_CC_E_TF27	**30**	HOG_CC_I_F224	**47**	MFCC_SYS_A2
14	HOG_SYS_I_F86	**31**	AR_SYS_A_C2	**48**	MFCC_CC_A2
15	AR_CC_A_C2	**32**	HOG_CC_I_F79	**49**	HOG_CC_I_F37
16	DWT_CC_A_SE11	**33**	W_CC_E_TF18	**50**	W_SYS_E_MF_PK2
17	HOG_CC_I_F209	**34**	GTCC_SYS_A_F4	**51**	MFCC_SYS_A3

The corresponding accuracies, for each combination of SMOTE 300% and undersampling percentage, are presented in Table 9.3.

It can be seen from Table 9.3 that increases in undersampling percentage correspond to increases in TPR, and a decreases in TNR (i.e., increased fail classification accuracy also results in more pass beats being incorrectly categorised as fail beats). Applying 175% under-sampling was found to give the highest combined TPR and TNR values.

Figure 9.1 illustrates the modified performance, with respect to raw (unbalanced) datasets, using SMOTE minority class oversampling of 300%, and 175% majority class under-sampling. The total TPR is 81% and 95% for the raw and balanced data respectively, whereas the mean TNR is 98% and 93% for the raw and balanced data, respectively. The confusion matrices associated with these two data-sets are presented in Figure 9.1.

The results in Figure 9.2 correspond to a TPR of 81.3% and 94.7% for the raw and balanced datasets respectively. Balancing the classes, therefore, resulted in a 13.4% increase in TPR, and 5.1% reduction in TNR. This trend is also illustrated in Figure 9.1, which gives the change in individual participant classification accuracies. It can be seen

Table 9.3: Classification accuracies for fail and pass classes using 300% SMOTE oversampling

Undersampling (%)	TPR (%)	TNR (%)	Average (%)
75	90.0	95.6	92.8
100	91.9	94.7	93.3
125	93.6	93.6	93.6
150	94.0	93.4	93.7
175	**94.7**	**93.0**	**93.9**
200	95.5	92.0	93.8
300	96.4	90.6	93.5
400	97.1	89.5	93.3
500	96.9	87.5	92.2
600	97.7	87.6	92.7
700	98.3	85.7	92.0
800	98.1	85.9	92.0
1000	98.4	83.5	91.0
2000	99.1	78.5	88.8

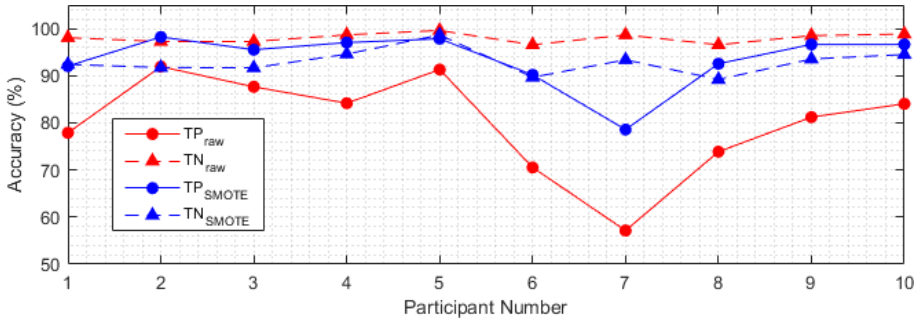


Figure 9.1: Difference in true positive, and true negative accuracies using SMOTE 300% minority class over-sampling, and 175% majority class under-sampling.

from this that the TPR and TNR are more proportional for the balanced data set. The increase in TPR is more notable for certain participants (e.g., P7, whose results exhibited an increase of approximately 23%).

The decrease in TNR is undesirable, the impact of this with respect to combining the model with the feedback models was investigated. The probe angles associated with false positives were recorded. Assessing how regularly the IPP was incorrectly rejected due to

		Actual	
		FAIL	PASS
Predicted	FAIL	929	163
	PASS	213	7837

(a) Raw

		Actual	
		FAIL	PASS
Predicted	FAIL	1081	572
	PASS	61	7428

(b) Balanced

Figure 9.2: Confusions matrices for raw and balanced datasets for fail beat classification models

the fail classification model provides further evaluation of how detrimental this decrease in TNR is (i.e., how often pass beats that correspond to the IPP were rejected). The IPP was incorrectly rejected three times, which corresponds to 0.5% of all false positives. This demonstrates that in practice small decreases in TNR may not be significantly detrimental, as examples from the IPP are far less likely to be rejected.

9.6 Discussion

An SVM model has been trained to classify fail beats. In practice this could be used to reject poorly segmented beats, and prevent misleading feedback. This model was evaluated, the results of which will now be discussed.

9.6.1 Selected Features

The features selected for the fail classification model are notably different from those for the regression and classification models. The features selected for these previous models have been predominantly based upon spectral analysis. Conversely, the most common feature type selected for the fail model are image features (HOG features). Furthermore, three novel envelope features are ranked first, third and seventh; these are all correlation statistics. The highest ranked feature, `W_CC_E_NE5_10`, is a correlation measure using the OMM_{ideal} MFE. The inclusion of these are intuitive, as they provide a direct measure of the MFE with respect to its expected form.

These results highlight the value of different feature types for different classification

tasks within Doppler analysis.

9.6.2 Beat Rejection Model

The results demonstrate that a high percentage of fail beats are correctly identified. The inclusion of this work reflects the challenges involved with combining the methods designed in this thesis, and implementing them in a real-world setting.

In practice, the process of performing an examination commonly results in periods of invalid data. For example, when the probe is lifted from the skin to ease discomfort or add additional ultrasound gel. During these periods, models will continue to provide feedback, which will consequently be erratic and unrepresentative. With respect to the continuous feedback, which could be implemented as a historical reference, this could be very detrimental. Additionally, in certain conditions the likelihood of false beats will increase, for example, with patients from which only a weak signal is attainable. In these cases, informative feedback will be most useful, and subsequently false beats will be the most harmful.

The danger of false beats producing misleading or confusing feedback could be reduced by implementing the models described in this chapter. Furthermore, the sensitivity of models for a given patient could be tailored by an operator, through selecting models trained with different dataset balances.

The demonstrated success of DCNN models in Chapter 8 could guide future design of beat rejection models. This approach could be advantageous, as the same input would be used for different models, as opposed to using SVM models which will require different feature subsets.

9.7 Conclusion

A series of models were trained to classify beats as being incorrectly or correctly segmented. Such systems would be a vital component of real-world Doppler ultrasound feedback software, as poorly segmented data can produce erratic and misleading feedback. Methods of undersampling and oversampling were implemented to address class imbalances, and associated model biases. These approaches enabled a model with 93.4% accuracy to be trained.

Chapter 10

Conclusions and Future Work

The work presented in this thesis will now be summarised, and recommendations regarding future work will be provided.

10.1 Thesis Conclusions

This thesis proposes and evaluates multiple signal processing and machine learning techniques for Doppler ultrasound. These have been designed for the primary goal of automatically generating feedback that reflects probe position. Research focused on this has not previously been documented, and the associated objectives reflected this. To realise this goal, the work in this thesis focused on six primary objectives. An overview of this work and the associated contributions are presented in the following six sections.

10.1.1 A stable and real-time method of MFE estimation

The first objective addressed in this thesis was the design of a real-time stable method of MFE estimation. This was fulfilled through the development of an image-processing method called OMM, which was demonstrated to be highly stable across a wide range of challenging signal conditions. Furthermore, the evaluation of this method resulted in contributions that fill current gaps within the research literature.

The evaluation of OMM was facilitated using a large selection of in-vivo data, phantom data and advanced simulation data. The combination of which enabled a thorough evaluation of the method. The simulation data were more realistic than those implemented

in previous works. Furthermore, the in-vivo dataset was far larger than used in previous works, and scans were not restricted to high-quality measurements from the IPP; giving an assessment more reflective of real-world conditions.

The performance of OMM was contrasted with that from two IPS methods and one image-processing method. Previous works have not documented direct comparisons of IPS and image-processing techniques, or included quantitative evaluations of image-processing methods. Comparatively, OMM was shown to be both more stable than these previous methods, and highly suited for its primary desired application, which was to enable consistent beat segmentation.

10.1.2 An automatic method of beat segmentation

The attributes of OMM MFEs were exploited and enabled work to fulfil a second primary objective. This was to design an automatic method of beat segmentation that requires no additional hardware to function. The research performed through this thesis resulted in two methods of beat segmentation being designed, each of which satisfy this objective. The two proposed methods are the slope-gradient method and the correlation method.

The slope-gradient method can be easily implemented using only an MFE. This lends itself it to replication in future works and can use any MFE estimation method. This facilitated a direct comparison of several MFE estimation methods, and when combined with OMM resulted in 97.8% of beats being correctly segmented.

The correlation method functions using the MFE, an ideal MFE and an image generated during the OMM MFE estimation process. This method was shown to outperform the slope-gradient method using a large sample of in-vivo data, and enabled valuable correlation features to be extracted. This method achieved a true positive segmentation rate of 99.2%, and was adopted for feature extraction. This successfully accomplished the requirement for an automatic beat segmentation method, and enabled features to be consistently extracted.

10.1.3 Propose existing and novel features

The OMM and correlation methods were combined to enable a thorough investigation of features and machine learning techniques. Different feedback models were designed through this thesis. In each case, development consisted of processing and identifying

valuable features.

A thorough means of feature and model selection was described, guided by previous machine learning works. This identified a variety of valuable features previously undocumented for Doppler ultrasound analyses. Notably, MFCC and GTCC features were shown to be powerful descriptors. Frequency analysis of MFEs were also found to be highly beneficial.

Additionally, promising results were demonstrated using state-of-the-art image recognition techniques, whereby spectral images are used instead of features; an approach previously unreported for Doppler ultrasound analysis. These findings. and the novel features identified, add to the arsenal of available techniques for Doppler ultrasound analyses. These could complement future works in this research area or related applications, such as disease classification. Considering this, this objective is considered to be satisfied with several valuable associated contributions.

10.1.4 A model to provide automatic scan quality feedback

The first feedback model proposed in the thesis addressed the objective to develop a model that provides quantitative feedback reflective of scan quality; a previously unreported application. In response to this, a regression model that generates continuous feedback was designed. This dynamic feedback varied in response to probe position; increasing when movements were beneficial and decreasing when they were detrimental.

Stringent evaluation was performed using an intersubject cross-validation approach. The designed model exhibited a high Pearson correlation of 0.86, with respect to the ideal response, and was found not to be detrimentally affected by estimated scan difficulty.

The evaluation of this model, whilst giving promising results, illustrated challenges with respect to generalisability and highlighted areas of interest for continued research. Although high correlations were achieved, absolute values of feedback reflective of the IPP varied across participants. Adoption of this model in clinical settings would have to consider these aspects carefully, and design practices and software to ensure feedback was valuable to medical practitioners.

10.1.5 A model to provide automatic probe position feedback

The second feedback model objective was to design a model that gives automatic feedback indicating probe position. This goes beyond continuous feedback, which compliments current methods of locating the IPP through interrogating multiple positions, and gives explicit predictions with respect to beneficial probe movements.

To address this objective, multiple feature-based and image recognition classification models were developed and evaluated. Previous Doppler ultrasound classification models have not considered guiding probe position, and have been limited to few categories. In the given work, four classification scenarios, with varying difficulty, were considered. These categorised observations as either belonging to the ideal position or not, as belonging to one of three locations when considering either a vertical or horizontal scanning plane separately, or as belonging to one of five positions when considering both the vertical and horizontal scanning planes combined.

These models succeeded in demonstrating that information extracted from Doppler ultrasound measurements can be used to differentiate between observations from various insonation angles, for example, an overall accuracy of 86% was attained when classifying observations from the horizontal plane. The more challenging five-class scenario illustrated that further development would be desirable, as overall accuracies were lower at 64%.

With respect to more traditional methods that employ hand-crafted features, improved classification performance was demonstrated using AlexNet. AlexNet was implemented using spectral images of cardiac cycles; spectral estimation using an autoregressive approach was found to be superior. These results demonstrate an exciting new tool that could be further developed for this application, or extended to additional Doppler ultrasound tasks.

In addition to this, improved performance was further demonstrated using ensemble methods, including combinations of both feature-based and image recognition models.

10.1.6 A model to automatically reject poor quality beats

The final objective was to develop a means of rejecting poor data. In practice, this would be valuable for preventing misleading feedback. In response to this objective, a model designed to identify poor data was designed and evaluated.

SVM classification models were trained and evaluated using additional artificial observations, which were introduced to overcome class imbalances. This technique could be

implemented to train models with varying sensitivity. Using this approach a classification accuracy of 94% was achieved, and true positive rate and false positive rate were closely balanced at 95% and 93%, respectively.

The research described in this thesis resulted in a number of valuable and significant contributions. Furthermore, the successes and limitations of these have highlighted aspects in which continued work would be beneficial, as well as potential avenues for interesting related research. This future work is discussed in the final section.

10.2 Future Work

This work has highlighted a variety of promising and previously undocumented features and machine learning techniques for Doppler ultrasound analysis. These findings could form the basis of continued research, which ultimately could lead to succeeding software being implemented in clinical settings.

An important step towards this goal would be a body of work focused on large scale data collection, of both healthy and unhealthy participants. As illustrated through this thesis, an inherent challenge is the intra- and interpatient variability of measurements. In practice, this challenge would be further exacerbated in clinical settings. This work would more extensively capture these variations and lead to more generalised and useful feedback models.

Future work could include the development of models for the pulmonary window. With respect to the suprasternal notch, measurements from this location can be particularly prone to misleading and erroneous blood flows from nearby vessels. To reflect this, the classification of specific vessels could be another avenue of research, from which resulting models could aid data acquisition. This approach would necessitate data being collected using a machine with imaging capabilities. This is achievable with more expensive (with respect to the USCOM) clinical machines, and can enable more precise and higher quality raw data being collected.

Feedback models could also be implemented using additional hardware. This could, for example, be in the form of robotics that automatically transition a probe through scanning windows (and then locates the IPP with respect to obtained feedback), or by using multiple transducers which could interrogate many different beam angles without being physically manoeuvred. This could speed up Doppler ultrasound examinations and

enable automatic monitoring. However, it could equally negate valuable characteristics of the technology, namely its portability, cost-effectiveness and ease of application.

Previous works detailing Doppler ultrasound classification have been focused on disease classification. The findings and methods in this thesis could equally be applied to these tasks. Similarly, the features and image recognitions techniques could be implemented for other audio research, such as auscultation classification (e.g., heart sounds). DCNNs and image-recognition, in particular, are undergoing rapid development and offer promising, and as of yet little researched, solutions for these tasks.

Appendices

Appendix A

USCOM Validation Surveys

Table A.1: References of USCOM validation surveys

Survey Number	Year	Number of Patients	Valve[1]	Reference
1	2005	22	PV	[91]
2	2005	36	PV	[107]
3	2006	40	AV	[100]
4	2006	45	PV	[100]
5	2007	30	AV	[401]
6	2007	22	AV	[19]
7	2007	22	PV	[19]
8	2008	24	AV	[85]
9	2008	15	AV	[34]
10	2008	10	AV	[94]
11	2008	12	PV	[99]
12	2009	78	PV	[402]
13	2009	89	AV	[88]
14	2009	30	AV	[101]
15	2009	89	AV	[88]
16	2012	70	PV	[403]
17	2012	20	PV	[404]
18	2016	31	AV	[112]

1 - PV denotes pulmonary valve, and AV denotes aortic valve.

Appendix B

Beat Segmentation: Correlation Method

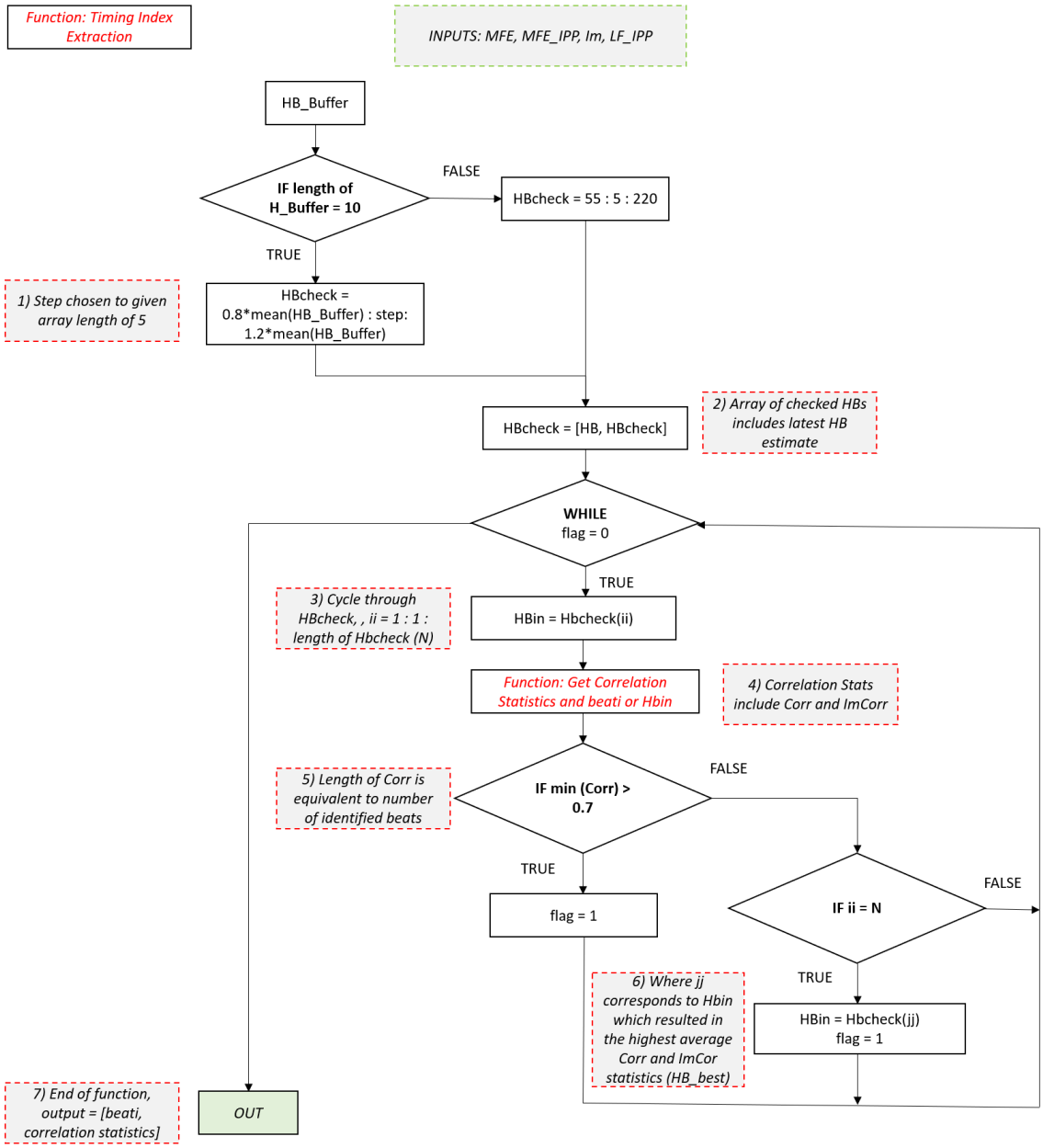


Figure B.1: Annotated block-diagram of correlation beat segmentation method

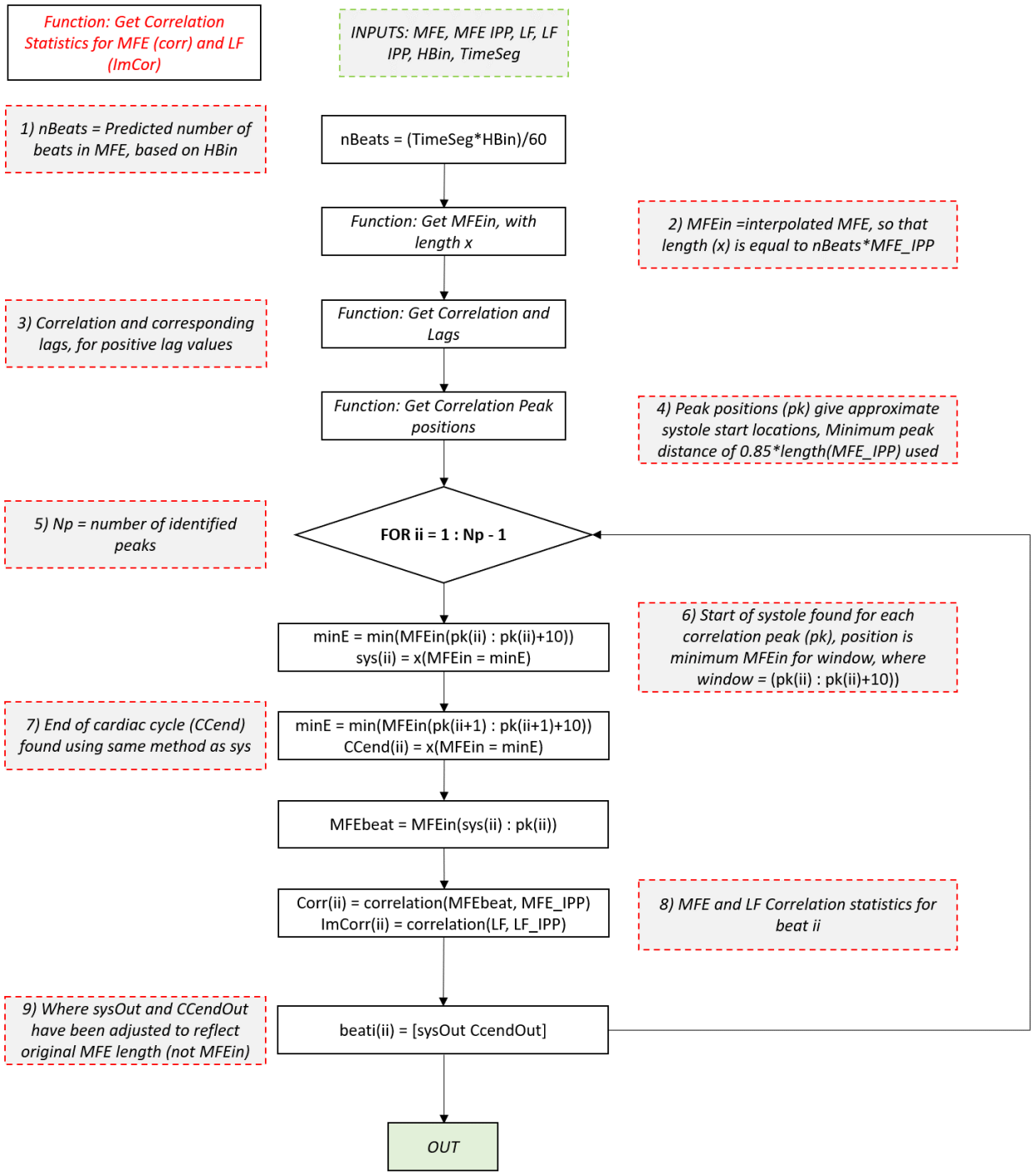


Figure B.2: Annotated block-diagram of function for correlation statistic extraction for correlation beat segmentation method

Appendix C

Survey Participant Information Sheet

Information Sheet for Participants

Study title: Investigation of methods to improve Doppler ultrasound estimated blood velocity acquisition

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

1. What is the purpose of the study?

Cardiac output is routinely monitored as part of intensive care practice, it is a principle determinant of oxygen delivery and blood pressure, and as such can guide treatment and identify patients at risk. Measurement of cardiac output has historically been performed using invasive methods, the invasive nature of these techniques introduces risk (e.g. arrhythmia, infection or possible pulmonary disruption) and results in limited use in the emergency department. These issues lead to alternative non-invasive methods being highly desirable. One such method uses non-invasive, transcutaneous (administered externally, across the skin) ultrasound. By placing an ultrasound device at the base of a patient's neck or on their chest, the reflected signal from moving blood can be analysed to calculate the velocity of that blood. This is possible because the velocity of the blood causes a proportional shift in frequency of the reflected ultrasound, known as a Doppler shift.

This study will collect Doppler frequency data using the above method, by placing an ultrasound device in the participant's suprasternal notch (base of the neck) to measure blood through the aortic valve and/or parasternally (left side of the chest) to measure blood in the pulmonary valve.

This data will be used to develop novel systems to aid users of medical ultrasound devices in performing consistent, accurate measurements. This will include an automated signal quality assessment, a sense check of measured values, and methods of localising the correct signal from the patient.

2. **Why have I been chosen?**

To design user feedback systems for administering medical Doppler ultrasound, data collected from healthy individuals is required, therefore we are seeking healthy volunteers to contribute to this research.

3. Do I have to take part?

The decision to take part is solely up to you. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason

4. What will happen to me if I take part?

- You will be provided with this information sheet and if you decide to take part we will perform Doppler ultrasound measurements on you using the above described method; which will take approximately 30-40 minutes to complete.

- Prior to the start of the experiment, you will be asked to provide information detailing your age, gender and height; this information will be anonymous.

5. **What about confidentiality?**

You do not need to provide your name, address or any other personal data. All the data will be kept in a secure location and will be anonymised using non-identifiable codes before further analysis. The collected data will be given non-identifiable codes and will therefore be anonymised prior to receiving. The researcher for this study will sign confidentiality and non-disclosure agreements for the newly collected data.

6. What do I have to do?

At the beginning of the experiment you would be ask to be in the surpine position (on your back), and to remove any clothing preventing access to your suprasternal notch/parasternal window.

We will then attempt to acquire Doppler ultrasound signals from both positions. This is done by first applying gel to the ultrasound device, and then positioning this until satisfactory signals are measured; this involves slowly moving the device across the skin in both locations and changing its angle and pressure.

7. Are there any risks?

There is no any realistic risk of any participants experiencing either physical or psychological distress or discomfort. You will be supervised by the researchers involved throughout the whole measurement process, ensuring you are safe, comfortable and willing to continue having the measurements performed.

8. **What will happen to the results of the research study?**

The results of the study will be used to design automated methods to aid users of medical Doppler ultrasound devices.

To receive a summary of findings you need to fill in your contact details. At the end of the study the researcher will email or post this summary to you using the personal details you provided, after which your personal details will be destroyed. If you do not wish to receive such a summary you do not have to give your personal details.

9. **Who is organising and funding the research?**

This research is organised by Cardiff University, School of Engineering, and School of medicine. The study will be managed and run by Dr Yulia Hicks, Dr Timothy Rainer, and Mr Jack Latham. The project is funded by the Engineering and Physical Sciences Research Council (EPSRC).

10. **Contact for Further Information**

Please contact me if you require further information.

Contact: Jack Latham, PhD student.
Tel:

School of Engineering, Cardiff University
Queen's Buildings
The Parade
CARDIFF CF24 3AA
Wales, UK.

Appendix D

Survey Participant Consent Form

CONSENT FORM

Title of Project: Investigation of methods to improve Doppler ultrasound estimated blood velocity acquisition

Name of Researcher: **Jack Latham**

Please initial box

1. I confirm that I have read and understand the information sheet dated ..... (version .....) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. ☐

3. I agree to take part in the above study. ☐

Name of Participant	Date	Signature
Name of Person taking consent (if different from researcher)	Date	Signature
Researcher	Date	Signature

Appendix E

Extracted Features Reference

Type[1]	Group	Subgroup	Data	Length	Name
Haemodynamic Metrics and Envelope Statistics					
Waveform	Haemodynamic	N/A	Envelope	CC	W_CC_E_OMM_H1
…	…	…	…	…	…
Waveform	Haemodynamic	N/A	Envelope	SYS	W_SYS_E_OMM_H8
Waveform	Haemodynamic	N/A	Envelope	CC	W_CC_E_MSNSI_H2
…	…	…	…	…	…
Waveform	Haemodynamic	N/A	Envelope	SYS	W_SYS_E_MSNSI_H8
Waveform	Envelope Statistics	N/A	Envelope	CC	W_CC_E_OMM_S1
Waveform	Envelope Statistics	N/A	Envelope	CC	W_SYS_E_OMM_S1
Waveform	Envelope Statistics	N/A	Envelope	SYS	W_SYS_E_OMM_S2
Waveform	Envelope Statistics	N/A	Envelope	SYS	W_SYS_E_OMM_S3
Waveform	Envelope Statistics	N/A	Envelope	SYS	W_SYS_E_OMM_S4
Waveform	Envelope Statistics	N/A	Envelope	CC	W_CC_E_MSNSI_S1
Waveform	Envelope Statistics	N/A	Envelope	CC	W_SYS_E_MSNSI_S1
Waveform	Envelope Statistics	N/A	Envelope	SYS	W_SYS_E_MSNSI_S2
Waveform	Envelope Statistics	N/A	Envelope	SYS	W_SYS_E_MSNSI_S3
Waveform	Envelope Statistics	N/A	Envelope	CC	W_SYS_E_MSNSI_S4
Traditional Features					
Waveform	Classic Waveform	N/A	Envelope	CC	W_CC_E_TF1
…	…	…	…	…	…
Waveform	Classic Waveform	N/A	Envelope	CC	W_CC_E_TF31
Novel Envelope Features					
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE1_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE1_7
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE2_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE2_7
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE3_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE3_7
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE4_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE4_7
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE5_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE5_10
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE6_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE6_7
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE7_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE7_6
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE8_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE8_6
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE9_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE9_6
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE10_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE10_6
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE11_1
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE11_6
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE12_1

Type[1]	Group	Subgroup	Data	Length	Name
…	…	…	…	…	…
Waveform	Novel Envelope Statistics	N/A	Envelope	CC	W_CC_E_NE12_6
		Image Features			
Image Features	Image Statistics	Novel Images	Image	SYS	IMS_SYS_I_IP1_S1
…	…	…	…	…	…
Image Features	Image Statistics	Novel Images	Image	CC	IMS_SYS_I_IP1_S5
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP1_S1
…	…	…	…	…	…
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP1_S5
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP2_S1
…	…	…	…	…	…
Image Features	Image Statistics	Novel Images	Image	SYS	IMS_CC_I_IP2_S5
Image Features	Image Statistics	Novel Images	Image	SYS	IMS_CC_I_IP3_S1
…	…	…	…	…	…
Image Features	Image Statistics	Novel Images	Image	SYS	IMS_CC_I_IP3_S5
Image Features	Image Statistics	Novel Images	Image	SYS	IMS_CC_I_IP4_S1
…	…	…	…	…	…
Image Features	Image Statistics	Novel Images	Image	SYS	IMS_CC_I_IP4_S4
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP4_S6
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP4_S7
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP5_S1
…	…	…	…	…	…
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP5_S4
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP5_S6
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP5_S7
Image Features	Computer Vision	LBP	Image	CC	LBP_CC_I_F1
…	…	…	…	…	…
Image Features	Computer Vision	LBP	Image	CC	LBP_CC_I_F59
Image Features	Computer Vision	HOG	Image	CC	HOG_CC_I_F1
…	…	…	…	…	…
Image Features	Computer Vision	HOG	Image	CC	HOG_CC_I_F324
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_SNR_BASE
Image Features	Image Statistics	Novel Images	Image	CC	IMS_CC_I_SNR_PEAK
		DWT Features			
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D1_1
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D1_2
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D1_3
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D1_4
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D2_1
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D2_2
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D2_3
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D2_4
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D3_1
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D3_2
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D3_3
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D3_4
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D4_1
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D4_2
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D4_3
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D4_4
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D5_1
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D5_2
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D5_3
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D5_4
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D6_1
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D6_2
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D6_3

Type[1]	Group	Subgroup	Data	Length	Name
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D6_4
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D7_1
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D7_2
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D7_3
Spectral	DWT	CS	Audio	CC	DWT_CC_A_D7_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D1_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D1_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D1_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D1_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D2_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D2_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D2_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D2_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D3_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D3_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D3_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D3_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D4_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D4_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D4_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D4_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D6_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D6_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D6_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D6_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D7_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D7_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D7_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D7_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D1_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D1_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D1_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D1_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D2_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D2_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D2_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D2_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D3_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D3_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D3_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D3_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D4_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D4_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D4_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D4_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D5_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D5_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D5_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D5_4
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D6_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D6_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D6_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D6_4

Type[1]	Group	Subgroup	Data	Length	Name
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D7_1
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D7_2
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D7_3
Spectral	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D7_4
Spectral	DWT	Entropy	Audio	SYS	DWT_SYS_A_SE1
…	…	…	…	…	…
Spectral	DWT	Entropy	Audio	SYS	DWT_SYS_A_SE16
Spectral	DWT	Entropy	Audio	CC	DWT_CC_A_SE1
…	…	…	…	…	…
Spectral	DWT	Entropy	Audio	CC	DWT_CC_A_SE16
Spectral	DWT	Entropy	Envelope	CC	DWT_CC_E_OMM_SE1
…	…	…	…	…	…
Spectral	DWT	Entropy	Envelope	CC	DWT_CC_E_OMM_SE16
Spectral	DWT	Entropy	Envelope	CC	DWT_CC_E_MSNSI_SE1
…	…	…	…	…	…
Spectral	DWT	Entropy	Envelope	CC	DWT_CC_E_MSNSI_SE16
Spectral	DWT	Entropy	Envelope	CC	DWT_CC_E_MF_SE1
…	…	…	…	…	…
Spectral	DWT	Entropy	Envelope	CC	DWT_CC_E_MF_SE16
Spectral	DWT	Variance	Audio	SYS	DWT_SYS_A_V1
…	…	…	…	…	…
Spectral	DWT	Variance	Audio	SYS	DWT_SYS_A_V9
Spectral	DWT	Variance	Audio	CC	DWT_CC_A_V1
…	…	…	…	…	…
Spectral	DWT	Variance	Audio	CC	DWT_CC_A_V11
Spectral	DWT	Variance	Envelope	CC	DWT_CC_E_OMM_V1
…	…	…	…	…	…
Spectral	DWT	Variance	Envelope	CC	DWT_CC_E_OMM_V8
Spectral	DWT	Variance	Envelope	CC	DWT_CC_E_MSNSI_V1
…	…	…	…	…	…
Spectral	DWT	Variance	Envelope	CC	DWT_CC_E_MSNSI_V8
Spectral	DWT	Variance	Envelope	CC	DWT_CC_E_MF_V1
…	…	…	…	…	…
Spectral	DWT	Variance	Envelope	CC	DWT_CC_E_MF_V8
		Parametric Model Features			
Spectral	Model Based	AR	Audio	SYS	AR_SYS_A_C1
Spectral	Model Based	AR	Audio	SYS	AR_SYS_A_C2
Spectral	Model Based	AR	Audio	SYS	AR_SYS_A_C3
Spectral	Model Based	AR	Audio	SYS	AR_SYS_A_C4
Spectral	Model Based	AR	Audio	CC	AR_CC_A_C1
Spectral	Model Based	AR	Audio	CC	AR_CC_A_C2
Spectral	Model Based	AR	Audio	CC	AR_CC_A_C3
Spectral	Model Based	AR	Audio	CC	AR_CC_A_C4
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_OMM_C1
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_OMM_C2
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_OMM_C3
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_OMM_C4
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_C1
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_C2
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_C3
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_C4
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MF_C1
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MF_C2
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MF_C3
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MF_C4
Spectral	Model Based	AR	Audio	CC	AR_CC_A_S1
Spectral	Model Based	AR	Audio	CC	AR_CC_A_S2

Type[1]	Group	Subgroup	Data	Length	Name
Spectral	Model Based	AR	Audio	CC	AR_CC_A_S3
Spectral	Model Based	AR	Audio	CC	AR_CC_A_S4
Spectral	Model Based	ARMA	Audio	CC	ARMA_CC_A_S1
Spectral	Model Based	ARMA	Audio	CC	ARMA_CC_A_S2
Spectral	Model Based	ARMA	Audio	CC	ARMA_CC_A_S3
Spectral	Model Based	ARMA	Audio	CC	ARMA_CC_A_S4
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S1
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S2
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S3
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S4
Spectral	Model Based	ARMA	Envelope	CC	ARMA_CC_E_OMM_S1
Spectral	Model Based	ARMA	Envelope	CC	ARMA_CC_E_OMM_S2
Spectral	Model Based	ARMA	Envelope	CC	ARMA_CC_E_OMM_S3
Spectral	Model Based	ARMA	Envelope	CC	ARMA_CC_E_OMM_S4
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S1
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S2
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S3
Spectral	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S4
Spectral	Model Based	ARMA	Envelope	CC	ARMA_CC_E_MSNSI_S1
Spectral	Model Based	ARMA	Envelope	CC	ARMA_CC_E_MSNSI_S2
Spectral	Model Based	ARMA	Envelope	CC	ARMA_CC_E_MSNSI_S3
Spectral	Model Based	ARMA	Envelope	CC	ARMA_CC_E_MSNSI_S4
MFCC and GTCC Features					
Spectral	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A0
…	…	…	…	…	…
Spectral	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A13
Spectral	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A0
…	…	…	…	…	…
Spectral	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A13
Spectral	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F0
…	…	…	…	…	…
Spectral	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F13
Spectral	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F0
…	…	…	…	…	…
Spectral	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F13
Spectral	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T0
…	…	…	…	…	…
Spectral	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T13
Spectral	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T0
…	…	…	…	…	…
Spectral	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T13
Audio Features					
Audio Features	N/A	N/A	Audio	CC	AF_CC_A1
…	…	…	…	…	…
Audio Features	N/A	N/A	Audio	CC	AF_SYS_A13
Audio Features	N/A	N/A	Audio	SYS	AF_SYS_A1
…	…	…	…	…	…
Audio Features	N/A	N/A	Audio	SYS	AF_CC_A13
Audio Features	N/A	N/A	Audio	CC	AF_CC_E_OMM_SNR
Audio Features	N/A	N/A	Audio	CC	AF_CC_E_MSNSI_SNR
Audio Features	N/A	N/A	Audio	CC	AF_CC_A_SNR

1 - Rows containing "…" indicate a series of features with identical details, whose feature names are increasing incrementaly

Appendix F

Haemodynamic and Traditional Feature Equations

Table F.1: Details of haemodynamic and traditional features

Name	Description	Equation*
Systolic Decay Time Index[1] (SDTI)	Ratio of normalised rise and decay slope close to peak systole (S)	$SDTI = \frac{T_r/T_1}{T_d/T_2}$
Pulsatility Index[1] (PI)	Wavefrom feature describing degree of pulse wave dampening	$PI = \frac{MFE(SP) - MFE(SD)}{M}$
Pourcelots Resistance Index[1] (RI)	Wavefrom feature describing degree of pulse wave dampening	$RI = \frac{MFE(SP) - MFE(CE)}{M}$
Constant Flow Ratio[1] (CFR)	…	$CFR = \frac{PI}{RI} - PI$
Height-Width Index[1] (HWI)	Combination of information contained in PI, FT and duration of cardiac cycle (T_b)	$HWI = PI\left(\frac{T_b}{FT}\right)$
Relative Flow Index[1] (RFRI)	Describes the average flow rate prior to systole and in remaining cardiac cycle	$RFRI = \frac{A_1/A_2}{T_1/T_2}$
Early Diastolic Notch Pulsatility Index[1] (NFPI)	Analyses early diastolic notch, originally designed for uterine artery	$NFPI = \frac{100}{M_{ND}^2}\sum_{n=1}^{\infty} a_n^2$, where a_n denotes the amplitude of the i harmonic

Acceleration[1]	Simple measurements, originally used to assess peripheral vascular disease	$Acceleration = \frac{MVE(SP)}{T(SP) - T(SS)}$
Deceleration[1]	…	$Deceleration = \frac{MVE(SP)}{T(SE) - T(SP)}$
Curve Broadening Index[1] (CBI)	Modified version of systolic time decay index	$CBI = T_1/T_R \times T_2/T_D$
Trans-systolic Velcoity[1] (V_{TST})	Previously used to anayse cerebral circulation	$V_{TST} = \frac{MFE(SP) + MFE(CE)}{2}$
Trans-systolic Times[1] (TST)	Duration of V_{TST}	…
Spectral Broadening Index[1] (SB1) 1	Measures of spectral broadening	$SBI1 = 0.47\frac{f_{max} - f_{min}}{f_{max} + f_{min}},$
SBI[1] 2	…	$SBI2 = \frac{f_{max} - f_{min}}{f_{max}}$
SBI[1] 3	…	$SBI3 = \frac{f_{max} - f_{min}}{f_{mean}}$
SBI[1] 4	…	$SBI4 = \frac{f_{max} - f_{mean}}{f_{max}}$
SBI[1] 5	…	$SBI5 = \frac{f_{max} - f_{min}}{f_{median}}$
SBI[1] 6	…	$SBI6 = \frac{f_{min}}{f_{max}}$
SBI[1] 7	…	$SBI7 = \frac{f_{max}}{f_{mean}}$
SBI[1] 8	Coefficient of variation	$SBI8 = \frac{f_{std}}{f_{mean}}$
SBI[1] 9	Coefficient of skewedness	$SBI9 = \frac{m3}{STD_f^3}$
SBI[1] 10	Coefficient kurtosis	$SBI10 = \frac{m4}{STD_f^4} - 3,$
Heart Rate (HR)	Calculated using Welchs PSD estimate as part of OMM method	…
Ejection Time Percentage[2] (ETP)	Percentage of cardiac cycle occupied by systole (\%)	$ETP = \frac{T(SE) - T(SS)}{T(CE) - T(SS)}$
Flow Time[2] (FT)	Duration of systole (s)	$FT = T(SE) - T(SE)$
Velocity Time Integral[2] (VTI)	Integrated velocity during systole, also known as stroke distance (m)	$VTI = \int_{SS}^{SE} MVE, where$ $MVE\ is\ the\ maximum\ velocity\ envelope$
Peak Systolic Velocity (PK)	Peak velocity during systole (m/s)	$PK = MVE(SP), where$ $SP\ is\ peak\ systole\ index$
VTI/PK[3]	Ratio of VTI and PK, providing means of normalisation	…
Cardiac Output[2]	The volume of blood pumped by the heart during a minute (l/min)	$CO = SV \times HR, where$ $SV = VTI \times CSA$

Cardiac Output Forced[2]	Cardiac output using MFE with peak systoles artifically forced based upon assending and descening flow gradients, as implimented in USCOM device	...
Standard Deviation[3]	Standard deviation of MFE of length N	$STD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\|MFE_i - \mu\|^2}, where$ $\mu = \frac{1}{N}\sum_{i=1}^{N} MFE_i$
Rising Slope Gradient[3] (RSG)	Gradient of rising systolic slope using MFE	$RSG = \frac{MFE(SP) - MFE(SS)}{SP - SS}$
Falling Slope Gradient[3] (FSG)	Gradient of falling systolic slope using MFE	$FSG = \frac{MFE(SE) - MFE(SP)}{SE - SP}$
RSG/FSG[3]	Ratio of RSG and FSG	...

1 – As described in [12].
2 – As described in [311].
3 – Explored as a part of this work.

Table F.-1: Variables used to calculate traditional and haemodynamic features

Variables	Defintion	Equation
SE	Systolic end index	…
SS	Systolic start index	…
CE	Cardiac cycle end index	…
SP	Systolic peak index	
T	Duraction of MFE	
T_1	Duration of rise	T(SP) - T(SS)
T_2	Duration of diastole	(T(CE) - T(SS)) - T_1
T_r	Time from 0.75S to S	…
T_d	Time from S to 0.75S	…
SD	Index of minimum value in cardiac cycle	…
M	Mean MFE value	…
A_1	Area below MFE during T_1	…
A_2	Area below MFE during T_2	…
M_{ND}	Mean diastolic MFE value	mean(MFE(SS to CE))
MVE	Maximum Velocity Envelope	
f_{max}	Maximum frequency at peak systole	…
f_{min}	Minimum frequency at peak systole	…
f_{mean}	Mean frequency at peak systole	…
f_{median}	Median frequency at peak systole	…
f_{std}	STD of frequency at peak systole	…
m3	Third moment of frequency at peak systole	…
m3	Fourth moment of frequency at peak systole	…

Appendix G

Computer Vision Feature Extraction Parameters

Table G.1: Parameters used to calculate LBP and HOG features

Setting	Description	Values
	LBP Parameters	
`NumNeighbors`	No. neighbours used to for each pixel to calculate LBP	8
`Radius`	Radius of circular pattern used to select neighbours	1
`Upright`	Determines rotation invariance	TRUE
`Interpolation`	interpolation method used to determine pixel neighbours	Linear
`CellSize`	...	Image Input Size
Matlab Function	...	`extractLBPFeatures`
	HOG Parameters	
`CellSize`	Size of HOG cell	[25 x 25]
`BlockSize`	No. of cells in block	[2 x 2]
`BlockOverlap`	No. overlapping cells between adjacent blocks	[1 x 1]
`NumBins`	No. orientation histogram bins	9
`UseSignedOrientation`	...	FALSE
Matlab Function	...	`extractHOGFeatures`

Appendix H

Image Portion Feature Extraction

ANNOTATED AND SIMPLIFIED CODE TO ILLUSTRATE PRIMARY STEPS IN IMAGE PORTION FEATURE EXTRACTION

```
% SS is the index of systole start for analysed cardiac cycle
% SE is the index of systole end for analysed cardiac cycle
% CE is the index for cardiac cycle end for analysed cardiac cycle
% SP is the index of peak systole for analysed cardiac cycle
% Yref is index of rows in image
% PSD is power spectral density for analysed cardiac cycle

MFE_Beat = MFE(SS:CE); % MFE for analysed cardiac cycle
OMM_Beat_Image = OMM_Image(SS:CE,:); % Image of analysed cardiac cycle
peakP = 0.7; % Percentage of systole peak used to define peak area
baseP = 0.3 % Percentage of systole peak used to define base area
PeakThreshold = peakP*MFE_Beat(SP); % Used to identify MFE_Beat in
 peak area
BaseThreshold = baseP*MFE_Beat(SP); % Used to identify MFE_Beat in
 peak area
```

Boundary Image Statistics

```
WindowSize = 2; % Number of pixels either side of time point used in
 boundary
for ii = 1 : length(MFE_Beat) - WindowSize % Loop through each
 MFE_Beat time point
    % Find indices for image boundary to be analysed, WindowSize of 2
    % equates 3x3 square of pixels
    xi = ii : ii + WindowSize; % Boundary extents in x direction
    yi = MFE_Beat(ii) - (WindowSize/2) :
 MFE_Beat(ii)+(WindowSize/2); % % Boundary extents in y direction

    Image_Boundary = OMM_Beat_Image(yi, xi); % Image Boundary Image,
 cropped from OMM beat image

   % Get statistics for Image_Boundary
   Image_Boundary_Stats = getImageStatistics(Image_Boundary); %
[Contrast, Correlation, Energy, Homogeneity, Pixel_Range]
   % Combine stats from each time point
   StatsToAverage = StatsToAverage + Image_Boundary_Stats;
   count = count + 1; % Number of stats combined

   if max(xi) < SE && min(yi) > PeakThreshold % Check to see if
 Image_Boundary is in peak area
       StatsToAverage_Peak = StatsToAverage_Peak +  [Contrast,
 Correlation, Energy, Homogeneity, Pixel_Range];
       countP = countP + 1; % Number of peak stats combined
   end
end

EnvelopeBoundaryStats = StatsToAverage/count; % average contrast,
 correlation etc for envelope boundary
PeakBoundaryStats = StatsToAverage_Peak/countP; % average contrast,
 correlation etc for peak boundary
```

Base and Peak Image Statistics

```
Peak_Image  = OMM_Beat_Image(Yref > PeakThreshold, 1:SE); % Image of
 peak area
Peak_Image_Stats = getImageStatistics(Peak_Image);

Base_Image = OMM_Beat_Image(Yref < BaseThreshold, 1: SE); % Image of
 base area
Base_Image_Statistics = getImageStatistics(Base_Image);

PSD_peak = PSD(Yref > PeakThreshold, 1:SE); % PSD in peak area
PSD_base = PSD(Yref > PeakThreshold, 1:SE); % PSD in base area
PSD_noise = PSD(Yref > MFE_Beat, 1:SE); % PSD in base area
```

SNR features using identified regions

```
SNR_Base = 10*log10(mean(PSD_peak)/mean(PSD_noise)); % SNR, base
 w.r.t. noise region
SNR_Peak = 10*log10(mean(Ppeakn)/mean(PSD_noise));% SNR, peak w.r.t.
 noise region
```

Published with MATLAB® R2019b

Appendix I

Wavelet Based Feature extraction

ANNOTATED AND SIMPLIFIED CODE TO ILLUSTRATE PRIMARY FUNCTIONS IN DWT FEATURE EXTRACTION

```
% x is input data (e.g. audio)
```

Wavelet Decomposition Features

```
wname = 'db1'; % daubechies level 1 wavelet
dn = 7; % detail level
% Calculate wavelet decompoisiton vector (c) and bookeeping vector
 (l),
[c,l] = wavedec(x ,dn,wname); % MATLAB, inbuilt
% Extract detail coefficients (cd1 to cd7)
[cd1,cd2,cd3, cd4, cd5, cd6, cd7] = detcoef(c,l,[1 2 3 4 5 6 7]); %
 MATLAB, inbuilt
% Calculate feature statistics (WavDec_Features) for detail
 coefficients,
WavDec_Features = get_DWTstats([cd1,cd2,cd3, cd4, cd5, cd6, cd7]);
```

Shannon Entropy Features

```
% Calculate wavelet packet tree (wpt)
[wpt] = modwpt(x, level); % MATLAB, inbuilt
% Calculate Energy of wavelet packet coefficients (E)
E = sum(wpt.^2,2);
% Calculate Probability
Pij = wpt.^2./E;
% Calculate shannon entropy
ShannonEntropy = -sum(Pij.*log(Pij),2)';
```

Variance Features

```
% xin is interpolated input data, to give consistent length
% Calculate maximumal overlap DWT (w)
w = modwt(xin,wname); % MATLAB, inbuilt
% Calculate multiscale variance of w (wvar)
wvar = modwtvar(w,wname); % MATLAB, inbuilt
```

Published with MATLAB® R2019b

Appendix J

Parametric Model Feature extraction

ANNOTATED AND SIMPLIFIED CODE TO ILLUSTRATE PRIMARY FUNCTIONS IN MODEL BASED FEATURE EXTRACTION

```
% x is input data (e.g. audio)
```

AR PSD Features

```
order = 10; % model order
nfft = 256; % Number of DFT points
% Calculate autoregressive power spectral density (ARpxx), length of
 129
ARpsd = pburg(x,order, nfft); % MATLAB, inbuilt
ARL = log10(ARpsd);
% Calculate feature statistics (AR_PSD_FEATURES) for PSD estimate,
AR_PSD_FEATURES = get_PSD_Stats(ARL);
```

AR Coefficient Features

```
orderC = 4;
% Calculate AR parameters
ar =  arburg(x,orderC); % MATLAB, inbuilt
% Select used features
AR_Coefficient_Features = artmp(2:end); % Remove first parameter
 (equals 1)
```

ARMA PSD Features

```
% Calculate AR and MA parts, using ARMASEL method designed by P.M.T.
% Broersen, https://uk.mathworks.com/matlabcentral/fileexchange/1330-
armasa?s_tid=prof_contriblnk
[AR, MA] = armasel(x,order);
% Calculate ARMA power spectral density (ARMApsd), length of 129
ARMApsd=arma2psd(AR,MA, nfft/2); % Part of ARMSEL,
ARMApsd_log = log10(ARMApsd);
% Calculate feature statistics (AR_PSD_FEATURES) for PSD estimate,
ARMA_PSD_FEATURES = get_PSD_Stats(ARMApsd_log);
% Calculate ARMA autocorrelation features
ARMA_Autocorrelations = arma2cor(AR,MA,order); % Part of ARMSEL,
```

Published with MATLAB® R2019b

Appendix K

Audio Feature extraction

Table K.1: Functions and variables used to extract audio features

Feature Name	Function and Inputs[1]
Loudness	`integratedLoudness(x,fs)`
Spectral Centroid	`spectralCentroid(x, fs,'Window', W)`
Spectral Decrease	`spectralDecrease(x, fs,'Window', W)`
Spectral Entropy	`spectralEntropy(x, fs,'Window', W)`
Spectral Flattness	`spectralFlatness(x, fs,'Window', W)`
Spectral Flux	`spectralFlux(x, fs,'Window', W/2)`
Spectral Kurtosis	`spectralKurtosis(x, fs,'Window', W)`
Spectral Rolloff Point	`spectralRolloffPoint(x, fs,'Window', W)`
Spectral Skewness	`spectralSkewness(x, fs,'Window', W)`
Spectral Slope	`spectralSlope(x, fs,'Window', W)`
Spectral Spread	`spectralSpread(x, fs,'Window', W)`
Pitch	`pitch(x, fs,'WindowLength', Window)`
Harmonic Ratio	`harmonicRatio(x, fs,'Window', W)`

1 - Where x is the input audio data, fs is the corresponding sample frequency and W denotes a rectangular window with length equal to x.

Appendix L

Model Training Specifications

Table L.1: Parameters used for SVM regression model

Parameter	Value
Model Type	Coarse Gaussian SVM
Kernal Function	Gaussian
Kernal Scale	27.1
Box Constraint	0.1
Epsilon	0.0026
Optimisation Time (s)	2,972
Training Function[1]	`fitrsvm`

1 – MATLAB function used to train model.

Table L.2: Parameters used for each classification model

Parameter	Classification Task				
	2-Class	*3-class, V*	*3-class, H*	*5-Class*	*Fail*
Model Type	SVM	SVM	SVM	SVM	SVM
Kernal Function	Polynomial	Polynomial	Polynomial	Polynomial	Polynomial
Order	3	3	2	3	3
Box Constraint	0.063	0.018	0.062	1.3	1
Kernel Scale	29.9	1.7	1.5	5	7.2
Coding	N/A	oneVall	oneVone	oneVone	N.A
Optimisation Time (s)	1,233	12,050	10,660	55,029	2,000
Training Function[1]	`fitcecoc`	`fitcecoc`	`fitcecoc`	`fitcecoc`	`fitcsvm`

1 – MATLAB function used to train model.

Appendix M

Selected Features for Feedback Models

Table M.1: Features selected for Regression SVM model

Rank	Type	Group	Subgroup	Data	Length	Name
1	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A2
2	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D3_4
3	S	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S2
4	AF	---	Audio	Envelope	CC	AF_CC_E_OMM_SNR
5	S	Model Based	AR	Audio	CC	AR_CC_A_S4
6	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A3
7	AF	---	---	Audio	SYS	AF_SYS_A1
8	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F0
9	S	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D6_4
10	S	Model Based	AR	Audio	CC	AR_CC_A_C4
11	S	Model Based	AR	Audio	CC	AR_CC_A_S3
12	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE7_7
13	S	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S1
14	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE5_10
15	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE1_5
16	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC9
17	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE7
18	IF	Computer Vision	LBP	Image	CC	LBP_CC_I_F7
19	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE4_3
20	S	Model Based	AR	Audio	CC	AR_CC_A_S1
21	IF	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP5_S7
22	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE9
23	S	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D4_4
24	S	Model Based	AR	Audio	CC	AR_CC_A_S2
25	S	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S2
26	IF	Image Statistics	Novel Images	Image	CC	IMS_CC_I_SNR_PEAK
27	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A0
28	S	Model Based	AR	Audio	CC	AR_CC_A_C3
29	S	DWT	CS	Audio	CC	DWT_CC_A_D1_1
30	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D2_2
31	S	DWT	CS	Audio	CC	DWT_CC_A_D1_4
32	S	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S4
33	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC6
34	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D4_2
35	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D3_2
36	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F247
37	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE6
38	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A7
39	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F0

Where: S - Spectral, W - Waveform, AF - Audio Features, IF - Image Features, CS - Coefficient Statistics, CC - Cardiac Cycle, SYS - Systole

Table M.2: Features selected for 2-class SVM model

Rank	Type	Group	Subgroup	Data	Length	Name
1	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_3
2	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE11
3	S	Model Based	AR	Audio	CC	AR_CC_A_C2
4	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A2
5	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F0
6	S	Model Based	AR	Audio	CC	AR_CC_A_C1
7	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F164
8	S	Model Based	AR	Audio	CC	AR_CC_A_C3
9	AF	---	Audio	Envelope	CC	AF_CC_E_OMM_SNR
10	S	Model Based	AR	Audio	CC	AR_CC_A_C4
11	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE6_7
12	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A5
13	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F13
14	S	Model Based	ARMA	Envelope	CC	ARMA_CC_E_OMM_S1
15	S	Model Based	AR	Audio	SYS	AR_SYS_A_C4
16	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A0
17	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A1
18	W	Classic Waveform	---	Envelope	CC	W_CC_E_TF11
19	AF	---	---	Audio	SYS	AF_SYS_A8
20	S	Model Based	ARMA	Envelope	CC	ARMA_CC_E_OMM_S3
21	IF	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP5_S7
22	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D2_2
23	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE5_10
24	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F13
25	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE9_6
26	S	DWT	CS	Audio	CC	DWT_CC_A_D1_1
27	AF	---	---	Audio	CC	AF_CC_A8
28	AF	---	---	Audio	CC	AF_CC_A10
29	S	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S2
30	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE7
31	S	DWT	CS	Audio	CC	DWT_CC_A_D1_4
32	S	Model Based	AR	Audio	CC	AR_CC_A_S3
33	IF	Image Statistics	Novel Images	Image	SYS	IMS_SYS_I_IP4_S7
34	S	Model Based	AR	Audio	CC	AR_CC_A_S1

35	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A2
36	AF	---	---	Audio	CC	AF_CC_A3
37	S	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S3
38	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A13
39	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_1
40	AF	---	---	Audio	CC	AF_CC_A1
41	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F1
42	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A9
43	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_S3
44	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F2
45	S	DWT	CS	Audio	CC	DWT_CC_A_D1_3
46	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F3

Where: S - Spectral, W - Waveform, AF - Audio Features, IF - Image Features, CS - Coefficient Statistics, CC - Cardiac Cycle, SYS - Systole

Table M.2: Features selected for 3-class vertical-plane SVM model

Rank	Type	Group	Subgroup	Data	Length	Name
1	S	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D5_3
2	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A2
3	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE8_6
4	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F2
5	S	Model Based	AR	Audio	SYS	AR_SYS_A_C3
6	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE6_7
7	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE9_6
8	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE4
9	S	Model Based	AR	Audio	CC	AR_CC_A_S2
10	W	Envelope Statistics	---	Envelope	CC	W_CC_E_MSNSI_S1
11	AF	---	---	Audio	CC	AF_CC_A11
12	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_2
13	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F4
14	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC3
15	S	DWT	Entropy	Audio	SYS	DWT_SYS_A_SE2
16	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_1
17	W	Haemodynamic	---	Envelope	SYS	W_SYS_E_MF_CO2
18	S	Model Based	AR	Audio	CC	AR_CC_A_S3
19	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A10
20	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F0
21	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC2
22	S	Model Based	AR	Audio	CC	AR_CC_A_C1
23	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A2
24	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F3
25	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D5_3
26	S	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D7_4
27	S	DWT	CS	Envelope	CC	DWT_CC_E_MSNSI_D5_1
28	AF	---	---	Audio	SYS	AF_SYS_A3
29	W	Haemodynamic	---	Envelope	SYS	W_CC_E_OMM_H5
30	AF	---	---	Audio	SYS	AF_SYS_A11
31	S	Model Based	AR	Audio	CC	AR_CC_A_C3
32	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T2
33	S	Model Based	AR	Audio	CC	AR_CC_A_S1
34	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_S1

35	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F8
36	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F13
37	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F12
38	W	Haemodynamic	---	Envelope	SYS	W_SYS_E_OMM_H4
39	W	Classic Waveform	---	Envelope	CC	W_CC_E_TF11
40	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F10
41	AF	---	---	Audio	CC	AF_CC_A1
42	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T0
43	AF	---	---	Audio	CC	AF_CC_A2
44	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T2
45	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D7_4
46	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A0
47	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A7
48	AF	---	---	Audio	SYS	AF_SYS_A1
49	AF	---	---	Audio	SYS	AF_SYS_A10
50	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC4
51	AF	---	---	Audio	SYS	AF_CC_A13
52	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A5
53	S	Model Based	AR	Audio	CC	AR_CC_A_C4
54	IF	Computer Vision	LBP	Image	CC	LBP_CC_I_F38
55	AF	---	---	Audio	SYS	AF_SYS_A8
56	S	Model Based	AR	Audio	SYS	AR_SYS_A_C4
57	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F244
58	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A6
59	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F3
60	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F1

Where: S - Spectral, W - Waveform, AF - Audio Features, IF - Image Features, CS - Coefficient Statistics, CC - Cardiac Cycle, SYS - Systole

Table M.2: Features selected for 3-class horizontal-plane SVM model

Rank	Type	Group	Subgroup	Data	Length	Name
1	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE6_7
2	W	Haemodynamic	---	Envelope	SYS	W_CC_E_OMM_H5
3	AF	---	---	Audio	CC	AF_CC_A8
4	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F3
5	S	Model Based	AR	Audio	CC	AR_CC_A_S2
6	S	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S2
7	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC4
8	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F5
9	W	Haemodynamic	---	Envelope	SYS	W_SYS_E_OMM_H4
10	AF	---	---	Audio	SYS	AF_CC_A13
11	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE2_3
12	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A3
13	S	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S4
14	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE1_3
15	AF	---	---	Audio	SYS	AF_SYS_A1
16	S	DWT	CS	Audio	CC	DWT_CC_A_D1_4
17	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T2
18	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC7
19	S	Model Based	AR	Audio	CC	AR_CC_A_S3
20	S	Model Based	AR	Audio	CC	AR_CC_A_C1
21	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D3_2
22	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F8
23	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F13
24	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE1_5
25	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D1_2
26	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F1
27	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F165
28	S	DWT	Entropy	Audio	SYS	DWT_SYS_A_SE2
29	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F5
30	W	Classic Waveform	---	Envelope	CC	W_CC_E_TF11
31	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F12
32	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F278
33	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A0
34	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T0

35	AF	---	---	Audio	CC	AF_SYS_A13
36	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F170
37	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A2
38	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F0
39	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC2
40	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F1
41	S	Model Based	AR	Audio	SYS	AR_SYS_A_C1
42	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F9
43	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F167
44	S	DWT	CS	Audio	CC	DWT_CC_A_D1_1
45	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F2
46	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_S2
47	AF	---	---	Audio	CC	AF_CC_A2
48	S	Model Based	AR	Envelope	CC	AR_CC_E_MSNSI_S1
49	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A0
50	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F3
51	W	Haemodynamic	---	Envelope	SYS	W_SYS_E_MF_CO1
52	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC3
53	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE12
54	S	DWT	CS	Envelope	CC	DWT_CC_E_OMM_D3_4
55	S	Model Based	AR	Audio	SYS	AR_SYS_A_C4
56	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_AC10
57	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE11
58	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A1
59	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F279
60	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE11_6
61	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F13
62	W	Haemodynamic	---	Envelope	CC	W_CC_E_OMM_H1
63	S	Model Based	AR	Audio	CC	AR_CC_A_S4

Where: S - Spectral, W - Waveform, AF - Audio Features, IF - Image Features, CS - Coefficient Statistics, CC - Cardiac Cycle, SYS - Systole

Table M.2: Features selected for 5-class SVM model

Rank	Type	Group	Subgroup	Data	Length	Name
1	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A2
2	IF	Computer Vision	LBP	Image	CC	LBP_CC_I_F1
3	S	Model Based	AR	Audio	CC	AR_CC_A_S3
4	S	Model Based	AR	Envelope	CC	AR_CC_E_OMM_S2
5	W	Haemodynamic	---	Envelope	SYS	W_CC_E_OMM_H5
6	IF	Computer Vision	LBP	Image	CC	LBP_CC_I_F58
7	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T2
8	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T6
9	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A0
10	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A13
11	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A3
12	IF	Image Statistics	Novel Images	Image	SYS	IMS_SYS_I_IP1_S1
13	W	Classic Waveform	---	Envelope	CC	W_CC_E_TF11
14	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE10_1
15	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A4
16	S	DWT	Variance	Audio	SYS	DWT_SYS_A_V2
17	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F246
18	S	DWT	Variance	Audio	SYS	DWT_SYS_A_V3
19	AF	---	---	Audio	CC	AF_CC_A2
20	S	Model Based	ARMA	Audio	CC	ARMA_CC_A_S1
21	S	DWT	Variance	Audio	CC	DWT_CC_A_V1
22	S	DWT	Variance	Audio	SYS	DWT_SYS_A_V1
23	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F1
24	IF	Image Statistics	Novel Images	Image	SYS	IMS_SYS_I_IP4_S3
25	S	DWT	CS	Audio	CC	DWT_CC_A_D7_3
26	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T8
27	S	DWT	CS	Audio	CC	DWT_CC_A_D5_2
28	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE12_1
29	AF	---	---	Audio	SYS	AF_SYS_A2
30	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T7
31	S	DWT	Variance	Audio	SYS	DWT_SYS_A_V7
32	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T1
33	AF	---	---	Audio	CC	AF_SYS_A13
34	AF	---	---	Audio	SYS	AF_CC_A13

35	S	DWT	Variance	Audio	CC	DWT_CC_A_V3
36	S	DWT	Variance	Audio	CC	DWT_CC_A_V7
37	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T0
38	AF	---	---	Audio	SYS	AF_SYS_A7
39	IF	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP2_S3
40	S	DWT	Variance	Audio	CC	DWT_CC_A_V10
41	AF	---	---	Audio	CC	AF_CC_A7
42	AF	---	---	Audio	CC	AF_CC_A8
43	W	Haemodynamic	---	Envelope	CC	W_CC_E_OMM_H2
44	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A6
45	AF	---	---	Audio	CC	AF_CC_A9
46	IF	Computer Vision	HOG	Image	SYS	HOG_SYS_I_F103
47	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T5
48	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F245
49	S	DWT	CS	Audio	CC	DWT_CC_A_D6_3
50	AF	---	---	Audio	CC	AF_CC_A3
51	AF	---	---	Audio	CC	AF_CC_A_SNR
52	AF	---	---	Audio	CC	AF_CC_A6
53	AF	---	---	Audio	SYS	AF_SYS_A1
54	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_T3
55	AF	---	---	Audio	CC	AF_CC_A10
56	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A2
57	S	DWT	CS	Audio	CC	DWT_CC_A_D7_2
58	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F10
59	S	DWT	Entropy	Envelope	CC	DWT_CC_E_MF_SE1
60	IF	Image Statistics	Novel Images	Image	CC	IMS_CC_I_IP1_S3
61	S	DWT	Variance	Audio	CC	DWT_CC_A_V6
62	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F2
63	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T7
64	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE1_1
65	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_T6
66	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE4
67	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F244
68	IF	Image Statistics	Novel Images	Image	SYS	IMS_SYS_I_IP4_S6

Where: S - Spectral, W - Waveform, AF - Audio Features, IF - Image Features, CS - Coefficient Statistics, CC - Cardiac Cycle, SYS - Systole

Table M.2: Features selected for fail beat SVM model

Rank	Type	Group	Subgroup	Data	Length	Name
1	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE5_10
2	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F89
3	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE5_4
4	IF	Computer Vision	HOG	Image	SYS	HOG_SYS_I_F321
5	AF	---	---	Audio	SYS	AF_SYS_A8
6	W	Haemodynamic	---	Envelope	SYS	W_SYS_E_MF_PK1
7	W	Novel Envelope Statistics	---	Envelope	CC	W_CC_E_NE5_9
8	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F101
9	IF	Image Statistics	Novel Images	Image	CC	IMS_CC_CORR2
10	IF	Computer Vision	HOG	Image	CC	HOG_CC_P1
11	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F167
12	IF	Image Statistics	Novel Images	Image	CC	IMS_CC_I_SNR_PEAK
13	W	Classic Waveform	---	Envelope	CC	W_CC_E_TF27
14	IF	Computer Vision	HOG	Image	SYS	HOG_SYS_I_F86
15	S	Model Based	AR	Audio	CC	AR_CC_A_C2
16	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE11
17	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F209
18	IF	Computer Vision	HOG	Image	SYS	HOG_SYS_I_F2
19	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE6
20	S	DWT	Entropy	Audio	SYS	DWT_SYS_A_SE6
21	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F201
22	IF	Computer Vision	HOG	Image	SYS	HOG_SYS_P117
23	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A8
24	W	Envelope Statistics	---	Envelope	CC	W_CC_E_MSNSI_S1
25	W	Haemodynamic	---	Envelope	SYS	W_SYS_E_MF_PK3
26	S	DWT	Variance	Envelope	CC	DWT_CC_E_MSNSI_noISB_V7
27	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F172
28	S	Model Based	AR	Audio	SYS	AR_SYS_A_C4
29	IF	Image Statistics	Novel Images	Image	SYS	IMS_SYS_I_IP4_S4
30	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F224
31	S	Model Based	AR	Audio	SYS	AR_SYS_A_C2
32	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F79
33	W	Classic Waveform	---	Envelope	CC	W_CC_E_TF18

34	S	Biologically Inspired	GTCC	Audio	SYS	GTCC_SYS_A_F4
35	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A1
36	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A1
37	S	DWT	Entropy	Audio	CC	DWT_CC_A_SE1
38	S	Biologically Inspired	GTCC	Audio	CC	GTCC_CC_A_F1
39	IF	Computer Vision	HOG	Image	SYS	HOG_SYS_I_F179
40	IF	Computer Vision	HOG	Image	SYS	HOG_SYS_I_F50
41	W	Haemodynamic	---	Envelope	CC	W_CC_E_OMM_H1
42	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A4
43	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F316
44	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F313
45	AF	---	---	Audio	CC	AF_CC_A3
46	AF	---	---	Audio	CC	AF_CC_A7
47	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A2
48	S	Biologically Inspired	MFCC	Audio	CC	MFCC_CC_A2
49	IF	Computer Vision	HOG	Image	CC	HOG_CC_I_F37
50	W	Haemodynamic	---	Envelope	SYS	W_SYS_E_MF_PK2
51	S	Biologically Inspired	MFCC	Audio	SYS	MFCC_SYS_A3

Where: S - Spectral, W - Waveform, AF - Audio Features, IF - Image Features, CS - Coefficient Statistics, CC - Cardiac Cycle, SYS - Systole

Bibliography

[1] Michele P. West. Medical-surgical equipment in the acute care setting. In *Acute Care Handbook for Physical Therapists: Fourth Edition*, pages 371–408. Elsevier Inc., 2014.

[2] Rocky Tsang. Hemodynamic monitoring in the cardiac intensive care unit. *Congenital Heart Disease*, 8(6):568–575, 2013.

[3] World Health Organisation. Cardiovascular diseases, 2020. `https://www.who.int/health-topics/cardiovascular-diseases/`, Last accessed on 2020-01-01.

[4] Emanuel Rivers, Bryant Nguyen, Suzanne Havstad, Julie Ressler, Alexandria Muzzin, Bernhard Knoblich, Edward Peterson, and Michael Tomlanovich. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England Journal of Medicine*, 2001.

[5] Felicity N E Gavins. *Vascular Responses to Pathogens*. Elsevier Inc., 2016.

[6] Fang Gao, Teresa Melody, Darren F. Daniels, Simon Giles, and Samantha Fox. The impact of compliance with 6-hour and 24-hour sepsis bundles on hospital mortality in patients with severe sepsis: a prospective observational study. *Critical care (London, England)*, 2005.

[7] Scott R. Gunn, Mitchell P. Fink, and Benjamin Wallace. Equipment review: The success of early goal-directed therapy for septic shock prompts evaluation of current approaches for monitoring the adequacy of resuscitation, 2005.

[8] Alan E. Jones, Michael D. Brown, Stephen Trzeciak, Nathan I. Shapiro, John S. Garrett, Alan C. Heffner, and Jeffrey A. Kline. The effect of a quantitative resusci-

tation strategy on mortality in patients with sepsis: A meta-analysis. *Critical Care Medicine*, 36(10):2734–2739, 2008.

[9] Christopher Cook, Graham Cole, Perviz Asaria, Richard Jabbour, and Darrel P. Francis. The annual global economic burden of heart failure. *International Journal of Cardiology*, 171(3):368–376, 2014.

[10] Alfred F. Connors, Theodore Speroff, Neal V. Dawson, Charles Thomas, Frank E. Harrell, Douglas Wagner, Norman Desbiens, Lee Goldman, Albert W. Wu, Robert M. Califf, William J. Fulkerson, Humberto Vidaillet, Steven Broste, Paul Bellamy, Joanne Lynn, and William A. Knaus. The effectiveness of right heart catheterization in the initial care of critically ill patients. *Journal of the American Medical Association*, 276(11):889–897, 1996.

[11] Monica R. Shah, Vic Hasselblad, Lynne W. Stevenson, Cynthia Binanay, Christopher M. O'Connor, George Sopko, and Robert M. Califf. Impact of the pulmonary artery catheter in critically ill patients: Meta-analysis of randomized clinical trials, 2005.

[12] D.H. Evans, W.N. McDicken, R Skidmore, and J.P. Woodcock. *Doppler Ultrasound; Physics, Instrumentation and Clinical Applications*. John Wiley and Sons, 2000.

[13] Dean L. Franklin, William Schlegel, and Robert F. Rushmer. Blood flow measured by Doppler frequency shift of back-scattered ultrasound. *Science*, 132:564–565, 1961.

[14] Karsten Knobloch, Christian Krettek, and Rob Phillips. Novel preclinical non-invasive determination of cardiac output by tissue doppler in air medical service. *Air Medical Journal*, 2005.

[15] Smith B., Phillips R., Madigan V., and West M. Decreased mortality, morbidity and emergency transport in septic shock; a new protocol based on advanced noninvasive haemodynamics and early antibiotics. *Critical Care Medicine*, 2012.

[16] L. Huang and L. A H Critchley. Study to determine the repeatability of supra-sternal doppler (ultrasound cardiac output monitor) during general anaesthesia: Effects of scan quality, flow volume, and increasing age. *British Journal of Anaesthesia*, 111(6):907–915, 2013.

[17] I Dey and P Sprivulis. Emergency physicians can reliably assess emergency department patient cardiac output using the USCOM continuous wave Doppler cardiac output monitor. *Emerg Med Australas*, 17(3):193–199, 2005.

[18] L. Huang and L. A H Critchley. Accuracy and precision of the USCOM: Does a meta-analysis provide the answer? *Anaesthesia*, 68(4):431–432, 2013.

[19] H. L.A. Van Den Oever, E. J. Murphy, and G. A. Christie-Taylor. USCOM (Ultrasonic Cardiac Output Monitors) lacks agreement with thermodilution cardiac output and transoesophageal echocardiography valve measurements. *Anaesthesia and Intensive Care*, 35(6):903–910, 2007.

[20] Lee S. Nguyen and Pierre Squara. Non-invasive monitoring of cardiac output in critical care medicine. *Frontiers in Medicine*, 4(NOV):1–8, 2017.

[21] S. W. Chong and P. J. Peyton. A meta-analysis of the accuracy and precision of the ultrasonic cardiac output monitor (USCOM). *Anaesthesia*, 67(11):1266–1271, 2012.

[22] Ali Ostadfar. *Biofluid Dynamics in Human Organs*. Elsevier Inc., 2016.

[23] Arthur C. Guyton and John E. Hall. *Guyton and Hall Textbook of Medical Physiology Twelfth Edition*. Elsevier Inc., 2017.

[24] Richard E. Klabunde. *Cardiovascular Physiology Concepts Second Edition*. Lippincott Williams & Wilkins, 2012.

[25] Gerard R Jr. Manecke and Timothy M Maus. *Essentials of Cardiac Anesthesia for Noncardiac Surgery: Cardiovascular Monitoring in Noncardiac Surgery*. Elsevier Inc., 2018.

[26] I. Turkoglu, A. Arslan, and E. Ilkay. An expert system for diagnosis of the heart valve diseases. *Expert Systems with Applications*, 23(3):229–236, 2002.

[27] Gerard R Jr. Manecke and Engy T Said. Essentials of Cardiac Anesthesia for Noncardiac Surgery Goal-Directed Fluid Therapy , Perioperative Pain Management , and Enhanced Recovery. In *Essentials of Cardiac Anesthesia for Noncardiac Surgery*, pages 493–508. Elsevier Inc., 2018.

[28] R. C. Bone, R. A. Balk, F. B. Cerra, R. P. Dellinger, A. M. Fein, W. A. Knaus, R. M.H. Schein, and W. J. Sibbald. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest*, 101(6):1644–1655, 1992.

[29] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, Lulu Guan, Yuan Wei, Hui Li, Xudong Wu, Jiuyang Xu, Shengjin Tu, Yi Zhang, Hua Chen, and Bin Cao. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 2020.

[30] M. R. Pinsky. A rose by any other name: Cardiac output, 2001.

[31] Jacob Pugsley and Adam B Lerner. Cardiac output monitoring: is there a gold standard and how do the newer technologies compare? *Seminars in cardiothoracic and vascular anesthesia*, 14(4):274–282, 2010.

[32] H. J.C. Swan, William Ganz, James Forrester, Harold Marcus, George Diamond, and David Chonette. Catheterization of the heart in man with use of a flow-directed balloon-tipped catheter. *New England Journal of Medicine*, 1970.

[33] Jésus Gonzalez, Christian Delafosse, Muriel Fartoukh, André Capderou, Christian Straus, Marc Zelter, Jean-Philippe Derenne, and Thomas Similowski. Comparison of bedside measurement of cardiac output with the thermodilution method and the Fick method in mechanically ventilated patients. *Critical care (London, England)*, 7:171–178, 2003.

[34] Bai Chuan Su, Chih Chung Lin, Chih Wen Su, Yu Ling Hui, Yung Fong Tsai, Ming Wen Yang, and Ping Wing Lui. Ultrasonic cardiac output monitor provides accurate measurement of cardiac output in recipients after liver transplantation. *Acta Anaesthesiologica Taiwanica*, 46(4):171–177, 2008.

[35] Sascha Meyer, David Todd, and Bruce Shadboldt. Assessment of portable continuous wave Doppler ultrasound (ultrasonic cardiac output monitor) for cardiac output measurements in neonates. *Journal of Paediatrics and Child Health*, 45(7-8):464–468, 2009.

[36] Duane J. Funk, Eugene W. Moretti, and Tong J. Gan. Minimally invasive cardiac output monitoring in the perioperative setting. *Anesthesia and Analgesia*, 108(3):887–897, 2009.

[37] J Alhashemi, M Cecconi, and C Hofer. Cardiac output monitoring: an integrative perspective. *Critical care*, 15:214, 2011.

[38] Edwareds Lifesciences. FloTrace system brochure, 2007.

[39] Willem-pieter De Boode, Markus Osypka, Sadaf Soleymani, and Shahab Noori. Assessment of cardiac output in neonates ultrasound. In *Hemodynamics and cardiology: neonatology questions and controversies*, volume 2, pages 237–263. Elsevier, third edit edition, 2018.

[40] Kristen Marshall, Elizabeth Thomovsky, Paula Johnson, and Aimee Brooks. A review of available techniques for cardiac output monitoring. *Topics in Companion Animal Medicine*, 31(3):100–108, 2016.

[41] W G Kubicek, J N Karnegis, R P Patterson, D A Witsoe, and R H Mattson. Development and evaluation of an impedance cardiac output system. *Aerospace medicine*, 37(12):1208–1212, 1966.

[42] David A. Hett and M. M. Jonas. Non-invasive cardiac output monitoring. *Intensive and Critical Care Nursing*, 20(2):103–108, 2004.

[43] Joan Yakimets and Louise Jensen. Evaluation of impedance cardiography: Comparison of NCCOM3-R7 with Fick and thermodilution methods. *Heart and Lung - The Journal of Acute and Critical Care*, 24(3):194–206, 1995.

[44] Arunodaya R. Gujjar, K. Muralidhar, Sanjay Banakal, Ratan Gupta, Talakad N. Sathyaprabha, and P. S. Jairaj. Non-invasive cardiac output by transthoracic electrical bioimpedence in post-cardiac surgery patients: Comparison with thermodilution method. *Journal of Clinical Monitoring and Computing*, 22(3):175–180, 2008.

[45] Lailu Mathews and R K Kalyan Singh. Cardiac output monitoring. *Annals of cardiac anaesthesia*, 11(1):56–68, 2008.

[46] Demetrio Pittarello and Karmi Shafer. Operative Echocardiography. In *Fundamentals of Congenital Minimally Invasive Cardiac Surgery*, pages 31–41. Elsevier Inc., 2018.

[47] Toshio Wagai. Studies on the foundation and development of diagnostic ultrasound. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, 83(8):256–265, 2007.

[48] K.T Dussik. Ultrashallanwendung in der Diagnostik und Therapie der Erkrankung des zentralen Nervensystems. *Ultrashall in der Medicine, Zurich, Hirzel*, 1949.

[49] Adrian M.K. Thomas, Arpan K. Banerjee, and Uwe Busch. *Classic papers in modern diagnostic radiology*. 2005.

[50] P N Wells. Current status and future technical advances of ultrasonic imaging. *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society*, 19:14–20, 2000.

[51] I. M. Coman. Christian Andreas Doppler - The man and his legacy. *European Journal of Echocardiography*, 2005.

[52] R Stuart Mackay. Non-invasive cardiac output. *Microvasc Res.*, 452:438–452, 1972.

[53] Shigeo Satomura. Study of the flow patterns in peripheral arteries by ultrasonics. *The Journal of the Acoustical Society of Japan*, 1959.

[54] L H Light. Transcutaneous observation of blood velocity in the ascending aorta in man. In *Proceedings of the Physiological Society*, volume 204, pages 1P–2P, 1969.

[55] L L Huntsman, D K Stewart, S R Barnes, S B Franklin, J S Colocousis, and E A Hessel. Noninvasive Doppler determination of cardiac output in man. Clinical validation. *Circulation*, 67(3):593–602, 1983.

[56] R. H. Klipstein, D. N. Firmin, S. R. Underwood, R. S. Rees, and D. B. Longmore. Blood flow patterns in the human aorta studied by magnetic resonance. *Heart*, 58(4):316–323, 1987.

[57] Aaron H. Steinman, Jahangir Tavakkoli, Jerry G. Myers, Richard S.C. Cobbold, and K. Wayne Johnston. Sources of error in maximum velocity estimation using linear phased-arraySteinman, A. H., Tavakkoli, J., Myers, J. G., Cobbold, R. S. C., & Johnston, K. W. (2001). Sources of error in maximum velocity estimation using linear phased-array Doppler systems w. *Ultrasound in Medicine and Biology*, 27(5):655–664, 2001.

[58] David H Evans and W Norman McDicken. Doppler Ultrasound: Physics, Instrumentation and Signal Processing. In *Doppler Ultrasound: Physics, Instrumentation and Signal Processing*, pages 134–138. Chichester, 2000.

[59] Gabriele Guidi, Cinzia Licciardello, and Simone Falteri. Intrinsic spectral broadening (ISB) in ultrasound Doppler as a combination of transit time and local geometrical broadening. *Ultrasound in Medicine and Biology*, 26(5):853–862, 2000.

[60] P. R. Hoskins, P. J. Fish, S. D. Pye, and T. Anderson. Finite beam-width ray model for geometric spectral broadening. *Ultrasound in Medicine and Biology*, 25(3):391–404, 1999.

[61] Peter R. Hoskins. Estimation of blood velocity, volumetric flow and wall shear rate using Doppler ultrasound. *Ultrasound*, 19(3):120–129, 2011.

[62] P. Anthony Chandraratna, Michele Nanna, Charles McKay, Ananda Nimalasuriya, Robert Swinney, Uri Elkayam, and Shahbudin H. Rahimtoola. Determination of cardiac output by transcutaneous continuous-wave ultrasonic Doppler computer. *The American Journal of Cardiology*, 53(1):234–237, 1984.

[63] B. A. Coghlan and M. G. Taylor. Directional Doppler techniques for detection of blood velocities. *Ultrasound in Medicine and Biology*, 1976.

[64] Peter N T Wells. A range-gated ultrasonic Doppler system. *Med. & biol. Engng*, 7:641–652, 1969.

[65] DW Baker, SA Rubenstein, and GS Lorch. Pulsed Doppler echocardiography: Principles and applications. *The American journal of medicine*, 63(July):69–80, 1977.

[66] F E Barber, D W Baker, A W Nation, D E Strandness, and J M Reid. Ultrasonic duplex echo-Doppler scanner. *IEEE transactions on bio-medical engineering*, 21(2):109–13, 1974.

[67] JM Reid and MP Spencer. Ultrasonic Doppler technique for imaging blood vessels. *Science*, 176(4040):1235–1236, 1972.

[68] Francis H. Chan and Stephen M. Pizer. An ultrasonogram display system using a natural color scale. *Journal of Clinical Ultrasound*, 1976.

[69] P. J. Fish. Multichannel, direction-resolving Doppler angiography. *2nd European Congress of Ultrasonics in Medicine*, Vol. 72, 1975.

[70] M. K. Eyer, M. A. Brandestini, D. J. Phillips, and D. W. Baker. Color digital echo/Doppler image presentation. *Ultrasound in Medicine and Biology*, 7(1):21–31, 1981.

[71] Dirk Timmerman, Thomas H. Bourne, Anil Tailor, William P. Collins, Herman Verrelst, Kamiel Vandenberghe, and Ignace Vergote. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: The development of a new logistic regression model. *American Journal of Obstetrics and Gynecology*, 181(1):57–65, 1999.

[72] L. Bassi, E. Boni, A. Dallai, F. Guidi, S. Ricci, and P. Tortoli. ULA-OP: A novel ultrasound advanced open platform for experimental research. *Proceedings - IEEE Ultrasonics Symposium*, pages 632–635, 2007.

[73] L. Bassi, E. Boni, A. Cellai, A. Dallai, F. Guidi, S. Ricci, and P. Tortoli. A novel digital ultrasound system for experimental research activities. *2008 11th EUROMICRO Conference on Digital System Design Architectures, Methods and Tools*, pages 413–417, 2008.

[74] L. H. Light. Non-injurious ultrasonic technique for observing flow in the human aorta, 1969.

[75] Lee L Huntsman, Ph D, Curtis C Johnson, and Eugene Fairbanks. Transcutaneous determination vehities in man of aortic blood-flow. *American Heart Journal*, 89(5):605–612, 1975.

[76] USCOM Ltd. Uscom 1a features, 2020. `https://www.uscom.com.au/products/uscom1a/features/`, Last accessed on 2020-02-18.

[77] Robert A. Phillips, Mike J. Dadd, Rob W. Gill, Malcolm J. West, and Darryl J. Burstow. Transcutaneous continuous wave doppler cardiac output monitoring is feasible producing reliable and reproducible signals. *Heart, Lung and Circulation*, 12(2):A67, 2003.

[78] Robert Strand. Automatic flow tracking system and method, U.S. Patent 0137717, June 03, 2003.

[79] USCOM 1A - Evidence, 2020.

[80] Lester A H Critchley and J A J H Critchley. A meta-analysis of studies using bias and precision statistics to compare cardiac output techniques. *Journal of Clinical Monitoring and Computing*, 15(2):85–91, 1999.

[81] Maurizio Cecconi, Andrew Rhodes, Jan Poloniecki, Giorgio Della Rocca, and R Michael Grounds. Bench-to-bedside review: The importance of the precision of the reference technique in method comparison studies–with specific reference to the measurement of cardiac output. *Critical care (London, England)*, 13(1):201, 2009.

[82] A V Varsha, Gladdy Goerge, and Raj Sahajanandan. Lutembacher syndrome: Dilemma of doing a tricuspid annuloplasty abstract. *Annals of Cardiac Anaesthesia*, 22(1):1–5, 2019.

[83] Eva Mulder, Shumalla Basit, Jolien Oben, Sander van Kuijk, Chahinda Ghossein-Doha, and Marc Spaanderman. Accuracy and precision of USCOM versus transthoracic echocardiography before and during pregnancy. *Pregnancy Hypertension*, 17(October 2018):138–143, 2019.

[84] J. G. Mangos, F. Pettit, R. Preece, K. Harris, and M. A. Brown. Repeatability of USCOM measured cardiac output in normotensive non-pregnant and pregnant women. *Pregnancy Hypertension*, 12(August 2017):71–74, 2018.

[85] Walter Knirsch, Oliver Kretschmar, Maren Tomaske, Kathrina Stutz, Nicole Nagdyman, Christian Balmer, Achim Schmitz, Dominique Béttex, Felix Berger, Urs

Bauersfeld, and Markus Weiss. Cardiac output measurement in children: Comparison of the ultrasound cardiac output monitor with thermodilution cardiac output measurement. *Intensive Care Medicine*, 34(6):1060–1064, 2008.

[86] Stefan M. Nidorf, Michael H. Picard, Marco O. Triulzi, James D. Thomas, John Newell, Mary Etta King, and Arthur E. Weyman. New perspectives in the assessment of cardiac chamber dimensions during development and adulthood. *Journal of the American College of Cardiology*, 19(5):983–988, 1992.

[87] Frederik Trinkmann, Christina Doesch, Theano Papavassiliu, Joerg Weissmann, Dariusch Haghi, Joachim Gruettner, Stefano Schoenberg, Martin Borggrefe, Jens J. Kaden, and Joachim Saur. A novel noninvasive ultrasonic cardiac output monitor: Comparison with cardiac magnetic resonance. *Clinical Cardiology*, 33(2):8–14, 2010.

[88] O. Thom, D. M. Taylor, R. E. Wolfe, J. Cade, P. Myles, H. Krum, and R. Wolfe. Comparison of a supra-sternal cardiac output monitor (USCOM) with the pulmonary artery catheter. *British Journal of Anaesthesia*, 103(6):800–804, 2009.

[89] Lester A H Critchley and Li Huang. USCOM - Window to the circulation: Utility of supra-sternal Doppler in an elderly anaesthetized patient for a robotic cystectomy. *Journal of Clinical Monitoring and Computing*, 28(1):83–93, 2014.

[90] Cangel Pui yee Chan, Nandini Agarwal, King keung Sin, Sangeeta Narain, Brendan E. Smith, Colin A. Graham, and Timothy H. Rainer. Age-specific non-invasive transcutaneous Doppler ultrasound derived haemodynamic reference ranges in elderly Chinese adults. *BBA Clinical*, 2:48–55, 2014.

[91] H. L. Tan, M. Pinder, R. Parsons, B. Roberts, and Peter V. van Heerden. Clinical evaluation of USCOM ultrasonic cardiac output monitor in cardiac surgical patients in intensive care unit. *British Journal of Anaesthesia*, 94(3):287–291, 2005.

[92] Benjamin Mothibe Bussmann, William Hulme, Andrew Tang, and Tim Harris. Investigating the ability of non-invasive measures of cardiac output to detect a reduction in blood volume resulting from venesection in spontaneously breathing subjects. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 26(1):1–7, 2018.

[93] E. O'Loughlin, M. Ward, A. Crossley, R. Hughes, A. P. Bremner, and T. Corcoran. Evaluation of the utility of the Vigileo FloTrac???, LiDCO???, USCOM and CardioQ??? to detect hypovolaemia in conscious volunteers: A proof of concept study. *Anaesthesia*, 70(2):142–149, 2015.

[94] Bai Chuan Su, Huang Ping Yu, Ming Wen Yang, Chih Chung Lin, Ming Chang Kao, Chia Hung Chang, and Wei Chen Lee. Reliability of a new ultrasonic cardiac output monitor in recipients of living donor liver transplantation. *Liver Transplantation*, 14(7):1029–1037, 2008.

[95] Shaun D. Gregory, Helena Cooney, Sara Diab, Chris Anstey, Ogilvie Thom, and John F. Fraser. In vitro evaluation of an ultrasonic cardiac output monitoring (USCOM) device. *Journal of Clinical Monitoring and Computing*, 30(1):69–75, 2016.

[96] Yu Wei Cheng, Feng Xu, and Jing Li. Identification of volume parameters monitored with a noninvasive ultrasonic cardiac output monitor for predicting fluid responsiveness in children after congenital heart disease surgery. *Medicine (United States)*, 97(39), 2018.

[97] Lyndon Siu, Adam Tucker, Shashi Kanth Manikappa, and John Monagle. Does patient position influence doppler signal quality from the USCOM ultrasonic cardiac output monitor? *Anesthesia and Analgesia*, 106(6):1798–1802, 2008.

[98] Neil Patel, Melissa Dodsworth, and John F. Mills. Cardiac output measurement in newborn infants using the ultrasonic cardiac output monitor: An assessment of agreement with conventional echocardiography, repeatability and new user experience. *Archives of Disease in Childhood: Fetal and Neonatal Edition*, 96(3), 2011.

[99] Lai Sze Grace Wong, Boon Hun Yong, Karl Kang Young, Lee Sung Lau, King Lik Cheng, Jeff Siu Fan Man, and Michael G. Irwin. Comparison of the USCOM ultrasound cardiac output monitor with pulmonary artery catheter thermodilution in patients undergoing liver transplantation. *Liver Transplantation*, 2008.

[100] Rajesh Chand, Yatin Mehta, and Naresh Trehan. Cardiac output estimation with a new Doppler device after off-pump coronary artery bypass surgery. *Journal of Cardiothoracic and Vascular Anesthesia*, 20(3):315–319, 2006.

[101] Amanda Corley, Adrian G. Barnett, Dan Mullany, and John F. Fraser. Nurse-determined assessment of cardiac output. Comparing a non-invasive cardiac output device and pulmonary artery catheter: A prospective observational study. *International Journal of Nursing Studies*, 46(10):1291–1297, 2009.

[102] H. L. Tan, M. Pinder, R. Parsons, B. Roberts, and Peter V. van Heerden. Clinical evaluation of USCOM ultrasonic cardiac output monitor in cardiac surgical patients in intensive care unit. *British Journal of Anaesthesia*, 94(3):287–291, 2005.

[103] G. F. Maruschak, A. M. Potter, J. F. Schauble, and M. C. Rogers. Overestimation of pediatric cardiac output by thermal indicator loss. *Circulation*, 1982.

[104] A. Van Grondelle, R. V. Ditchey, and B. M. Groves. Thermodilution method overestimates low cardiac output in humans. *American Journal of Physiology - Heart and Circulatory Physiology*, 1983.

[105] Robert A. Phillips, Sally G. Hood, Beverley M. Jacobson, Malcolm J. West, Li Wan, and Clive N. May. Pulmonary artery catheter (PAC) accuracy and efficacy compared with flow probe and transcutaneous doppler (USCOM): An ovine cardiac output validation. *Critical Care Research and Practice*, 2012, 2012.

[106] Saurabh Jain, Alexander Allins, Ali Salim, Amir Vafa, Matthew T. Wilson, and Daniel R. Margulies. Noninvasive Doppler ultrasonography for assessing cardiac function: can it replace the Swan-Ganz catheter? *American Journal of Surgery*, 2008.

[107] Karsten Knobloch, Artur Lichtenberg, Michael Winterhalter, Dagmar Rossner, Max Pichlmaier, and Rob Phillips. Non-invasive cardiac output determination by two-dimensional independent Doppler during and after cardiac surgery. *Annals of Thoracic Surgery*, 80(4):1479–1483, 2005.

[108] Chanapai Chaiyakulsil, Marut Chantra, Poomiporn Katanyuwong, Anant Khositseth, and Nattachai Anantasit. Comparison of three non-invasive hemodynamic monitoring methods in critically ill children. *PLoS ONE*, 13(6):1–12, 2018.

[109] Edward G. Grant, Carol B. Benson, Gregory L. Moneta, Andrei V. Alexandrov, J. Dennis Baker, Edward I. Bluth, Barbara A. Carroll, Michael Eliasziw, John

Gocke, Barbara S. Hertzberg, Sandra Katarick, Laurence Needleman, John Pellerito, Joseph F. Polak, Kenneth S. Rholl, Douglas L. Wooster, and Eugene Zierler. Carotid artery stenosis: grayscale and Doppler ultrasound diagnosis–Society of Radiologists in Ultrasound consensus conference. In *Ultrasound quarterly*, 2003.

[110] J. Golledge, M. Ellis, T. Sabharwal, T. Sikdar, A. H. Davies, and R. M. Greenhalgh. Selection of patients for carotid endarterectomy. *Journal of Vascular Surgery*, 30(1):122–130, 1999.

[111] A. T. Papageorghiou, C. K.H. Yu, S. Cicero, S. Bower, and K. H. Nicolaides. Second-trimester uterine artery Doppler screening in unselected populations: A review. *Journal of Maternal-Fetal and Neonatal Medicine*, 12(2):78–88, 2002.

[112] Fernando Beltramo, Jondavid Menteer, Asma Razavi, Robinder G. Khemani, Jacqueline Szmuszkovicz, Christopher J L Newth, and Patrick A. Ross. Validation of an ultrasound cardiac output monitor as a bedside tool for pediatric patients. *Pediatric Cardiology*, 37(1):177–183, 2016.

[113] Aaron H. Steinman, Jahangir Tavakkoli, Jerry G. Myers, Richard S C Cobbold, and K. Wayne Johnston. Sources of error in maximum velocity estimation using linear phased-array Doppler systems with steady flow. *Ultrasound in Medicine and Biology*, 27(5):655–664, 2001.

[114] Aditi Kathpalia, Yücel Karabiyik, Sturla H Eik-nes, Eva Tegnander, Ingvild Kinn Ekroll, Gabriel Kiss, and Hans Torp. Adaptive spectral envelope estimation for Doppler ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63(11):1825–1838, 2016.

[115] P. R. Hoskins. Accuracy of maximum velocity estimates made using Doppler ultrasound systems. *British Journal of Radiology*, 1996.

[116] Massoud Zolgharni, Niti M. Dhutia, Graham D. Cole, M. Reza Bahmanyar, Siana Jones, S. M Afzal Sohaib, Sarah B. Tai, Keith Willson, Judith A. Finegold, and Darrel P. Francis. Automated aortic doppler flow tracing for reproducible research and clinical measurements. *IEEE Transactions on Medical Imaging*, 33(5):1071–1082, 2014.

[117] Elaine Y.L. Lui, Aaron H. Steinman, Richard S.C. Cobbold, and K. Wayne Johnston. Human factors as a source of error in peak Doppler velocity measurement. *Journal of Vascular Surgery*, 42(5):3–6, 2005.

[118] K. W. Johnston, M. Kassam, and R. S.C. Cobbold. Online identifying and quantifying Doppler ultrasound waveforms. *Medical & Biological Engineering & Computing*, 20(3):336–342, 1982.

[119] S. M.Afzal Sohaib, Zachary I. Whinnett, Kenneth A. Ellenbogen, Christoph Stellbrink, T. Alexander Quinn, Margot D. Bogaard, Pierre Bordachar, Berry M. Van Gelder, Irene E. Van Geldorp, Cecilia Linde, Mathias Meine, Frits W. Prinzen, Robert G. Turcott, Henry M. Spotnitz, Dan Wichterle, and Darrel P. Francis. Cardiac resynchronisation therapy optimisation strategies: Systematic classification, detailed analysis, minimum standards and a roadmap for development and testing. *International Journal of Cardiology*, 170(2):118–131, 2013.

[120] Darrel P. Francis. How to reliably deliver narrow individual-patient error bars for optimization of pacemaker AV or VV delay using a "pick-the-highes" strategy with haemodynamic measurements. *International Journal of Cardiology*, 163(3):221–225, 2013.

[121] Punam A. Pabari, Keith Willson, Berthold Stegemann, Irene E. Van Geldorp, Andreas Kyriacou, Michela Moraldo, Jamil Mayet, Alun D. Hughes, and Darrel P. Francis. When is an optimization not an optimization? Evaluation of clinical implications of information content (signal-to-noise ratio) in optimization of cardiac resynchronization therapy, and how to measure and maximize it. *Heart Failure Reviews*, 16(3):277–290, 2011.

[122] Larry Y.L.Mo, Louis C.M. Yun, and R. S.C. Cobbold. Comparison of four digital maximum frequency estimators for Doppler ultrasound. *Ultrasound in Medicine and Biology*, 14(5):355–363, 1988.

[123] K. W. Johnston, B. C. Maruzzo, and R. S.C. Cobbold. Doppler methods for quantitative measurement and localization of peripheral arterial occlusive disease by analysis of the blood flow velocity waveform. *Ultrasound in Medicine and Biology*, 4(3):209–223, 1978.

[124] F M Greene, K Beach, G Fell, and D J Phillips. Computer based pattern recognition of carotid arterial disease using pulsed Doppler ultrasound. *Ultrasound Med Biol*, 8(2):161–176, 1982.

[125] Andrzej Nowicki, Paweł Karłowicz, Maciej Piechocki, and Wojciech Secomski. Method for the measurement of the maximum doppler frequency. *Ultrasound in Medicine and Biology*, 1985.

[126] T. D'Alessio. 'Objective' algorithm for maximum frequency estimation in Doppler spectral analysers. *Medical & Biological Engineering & Computing*, (January):63–68, 1985.

[127] R Moraes, N Aydin, and DH Evans. The performance of three maximum frequency envelope detection algorithms for Doppler signals. *Journal of Vascular Investigation*, 1(3):126–134, 1995.

[128] Krzysztof Marasek and Andrzej Nowicki. Comparison of the performance of three maximum doppler frequency estimators coupled with different spectral estimation methods. *Ultrasound in Medicine and Biology*, 20(7):629–638, 1994.

[129] David Vilkomerson, Stefano Ricci, and Piero Tortoli. Finding the peak velocity in a flow from its doppler spectrum. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 60(10):2079–2088, 2013.

[130] P. R. Hoskins. Estimation of blood velocity, volumetric flow and wall shear rate using Doppler ultrasound. *Ultrasound*, pages 1–2, 2011.

[131] J Tschirren, R M Lauer, and M Sonka. Automated analysis of Doppler ultrasound velocity flow diagrams. *IEEE Trans Med Imaging*, 2001.

[132] Massoud Zolgharni, Niti M. Dhutia, Graham D. Cole, M. Reza Bahmanyar, Siana Jones, S. M Afzal Sohaib, Sarah B. Tai, Keith Willson, Judith A. Finegold, and Darrel P. Francis. Automated aortic doppler flow tracing for reproducible research and clinical measurements. *IEEE Transactions on Medical Imaging*, 33(5):1071–1082, 2014.

[133] Oron Shechner, Mickey Sheinovitz, Micha Feinberg, and Hayit Greenspan. Image analysis of doppler echocardiography for patients with atrial fibrillation. In *2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano*, 2004.

[134] Oron Shechner, Mickey Scheinowid, Micha S Feinberg, and Huyit Greenspun. Automardiography image analysis. *Proc. 23rd IEEE Convent. Elect. Electron. Eng. Israel*, pages 177–180, 2004.

[135] Hayit Greenspan, Oron Shechner, Mickey Scheinowitz, and Micha S. Feinberg. Doppler echocardiography flow-velocity image analysis for patients with atrial fibrillation. *Ultrasound in Medicine and Biology*, 31(8):1031–1040, 2005.

[136] V. Magagnin, L. Delfino, S. Cerutti, M. Turiel, and E. G. Caiani. Nearly automated analysis of coronary Doppler flow velocity from transthoracic ultrasound images: Validation with manual tracings. *Medical and Biological Engineering and Computing*, 45(5):483–493, 2007.

[137] T. Syeda-Mahmood, P. Turaga, D. Beymer, F. Wang, A. Amir, H. Greenspan, and K. Pohl. Shape-based similarity retrieval of Doppler images for clinical decision support. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 855–862, 2010.

[138] P. R. Hoskins, S. F. Li, and W. N. McDicken. Velocity estimation using duplex scanners. *Ultrasound in Medicine and Biology*, 17(2):195–198, 1991.

[139] P.R. Hoskins. Review of the design and use of flow phantoms. *Testing of Doppler Ultrasound Equipment*, (May):12–29, 1994.

[140] P. R.cobbold pump system for physiological flow simulation Hoskins, T. Anderson, and W. N. McDicken. A computer controlled flow phantom for generation of physiological Doppler waveforms. *Physics in Medicine and Biology*, 34(11):1709–1717, 1989.

[141] A. Goldstein. Performance tests of Doppler ultrasound equipment with a string phantom. *Journal of Ultrasound in Medicine*, 10(3):125–139, 1991.

[142] Fabrice Wendling, Steven A. Jones, and Don P. Giddens. Simulation of Doppler ultrasound signals for a laminar, pulsatile, nonuniform flow. *Ultrasound in Medicine and Biology*, 18(2):179–193, 1992.

[143] Larry Y.L. Mo and R.S.C. Cobbold. "Speckle" in continuous wave doppler ultrasound spectra: a simulation study. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 33(6):747–753, 1986.

[144] PJ Vaitkus and RSC Cobbold. A comparative study and assessment of Doppler ultrasound spectral estimation techniques part I: Estimation methods. *Ultrasound in medicine & biology*, 14(8):661–672, 1988.

[145] G. H. Van Leeuwen, A. P.G. Hoeks, and R. S. Reneman. Simulation of real-time frequency estimators for pulsed doppler systems. *Ultrasonic Imaging*, 1986.

[146] Larry Y.L. Mo and Richard S.C. Cobbold. A unified approach to modeling the backscattered Doppler ultrasound from blood. *IEEE Transactions on Biomedical Engineering*, 39(5):450–461, 1992.

[147] J. R. Womersley. Oscillatory flow in arteries: The constrained elastic tube as a model of arterial flow and pulse transmission. *Physics in Medicine and Biology*, 1957.

[148] Gabriele Guidi, Vernon L. Newhouse, and Piero Tortoli. Doppler spectrum shape analysis based on the summation of flow-line spectra. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 1995.

[149] Jørgen A Jensen and Niels Bruun Svendsen. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 1992.

[150] Jørgen A Jensen. Field: A program for simulating ultrasound systems. *Medical and Biological Engineering and Computing*, 34(1):351–352, 1996.

[151] J Synnevag, A Austeng, and S Holm. Minimum variance adaptive beamforming applied to medical ultrasound imaging. *IEEE Ultrasonics Symposium 2005*, 2005.

[152] Abigail Swillens, Patrick Segers, Hans Torp, and Lasse Løvstakken. Two-dimensional blood velocity estimation with ultrasound: Speckle tracking versus crossed-beam vector doppler based on flow simulations in a carotid bifurcation model. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2010.

[153] Jørgen A Jensen. Focusing along the flow direction I : Theory and simulation. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 50(7):857–872, 2003.

[154] Jørgen A Jensen, Senior Member, and Rasmus Bjerngaard. Along the flow direction II : Experimental investigation. *Development*, 50(7):873–880, 2003.

[155] Jørgen A Jensen. Simulation of advanced ultrasound systems using Field II. *2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (IEEE Cat No. 04EX821)*, 2:636–639, 2004.

[156] D. T. Gibbons, D. H. Evans, W. W. Barrie, and P. S. Cosgriff. Real-time calculation of ultrasonic pulsatility index. *Medical & Biological Engineering & Computing*, 19(1):28–34, 1981.

[157] R. A. Knox, F. M. Greene, K. Beach, D. J. Phillips, P. M. Chikos, and D. E. Strandness. Computer based classification of carotid arterial disease: A prospective assessment. *Stroke*, 13(5):589–594, 1982.

[158] Louis Allard, Yves E. Langlois, Louis Gilles Durand, Ghislaine O. Roederer, Manon Beaudoin, Guy Cloutier, Paul Roy, and Pierre Robillard. Computer analysis and pattern recognition of doppler blood flow spectra for disease classification in the lower limb arteries. *Ultrasound in Medicine and Biology*, 17(3):211–223, 1991.

[159] Bruno Felix Osmanski, Jeremy Bercoff, Gabriel Montaldo, Thanasis Loupas, Mathias Fink, and Mickael Tanter. Cancellation of Doppler intrinsic spectral broadening using ultrafast Doppler imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 61(8):1396–1408, 2014.

[160] R. S. Thompson, B. J. Trudinger, and C. M. Cook. Doppler ultrasound waveforms in the fetal umbilical artery: Quantitative analysis technique. *Ultrasound in Medicine and Biology*, 11(5):707–718, 1985.

[161] P. R. Hoskins, T. Loupas, and W. N. McDicken. An investigation of simulated umbilical artery doppler waveforms. I. The effect of three physical parameters on the maximum frequency envelope and on pulsatility index. *Ultrasound in Medicine and Biology*, 1991.

[162] S. Brekke, E. Tegnander, H. G. Torp, and S. H. Eik-Nes. Tissue Doppler gated (TDOG) dynamic three-dimensional ultrasound imaging of the fetal heart. *Ultrasound in Obstetrics and Gynecology*, 24(2):192–198, 2004.

[163] Laura J. Brattain, Brian A. Telfer, Manish Dhyani, Joseph R. Grajo, and Anthony E. Samir. Machine learning for medical ultrasound: status, methods, and future opportunities. *Abdominal Radiology*, 43(4):786–799, 2018.

[164] Daniel Bernoulli. Hydrodynamica, sive de viribus et motibus fluidorum commentarii. *Opus academicum ab auctore, dum Petropoli ageret, congestum*, 1738.

[165] R. Gorlin and S. G. Gorlin. Hydraulic formula for calculation of the area of the stenotic mitral valve, other cardiac valves, and central circulatory shunts. I. *American Heart Journal*, 41(1):1–29, 1951.

[166] Paul Teirstein, Mark Yeager, Paul G. Yock, and Richard L. Popp. Doppler echocardiographic measurement of aortic valve area in aortic stenosis: A noninvasive application of the Gorlin formula. *Journal of the American College of Cardiology*, 8(5):1059–1065, 1986.

[167] Merrill P. Spencer and John M. Reid. Quantitation of carotid stenosis with continuous-wave (CW) Doppler ultrasound. *Stroke*, 10(3):326–330, 1979.

[168] Pravin Manga, Raja W. Dhurandhart, and Brian Stockard. Doppler frequency ratio and peak frequency in the assessment of carotid artery disease: a comparative study with agiography. *Ultrasound in Medicine & Biology*, 12(7):573–576, 1986.

[169] William J. Zwiebel. Spectrum analysis in carotid sonography. *Ultrasound in Medicine and Biology*, pages 625–36, 1987.

[170] E. I. Bluth, A. T. Stavros, K. W. Marich, S. M. Wetzner, D. Aufrichtig, and J. D. Baker. Carotid duplex sonography: A multicenter recommendation for standardized

imaging and Doppler criteria. *Radiographics : a review publication of the Radiological Society of North America, Inc*, 8(3):487–506, 1988.

[171] K. W. Johnston, P. M. Brown, and M. Kassam. Problems of carotid doppler scanning which can be overcome by using frequency analysis. *Stroke*, 13(5):660–666, 1982.

[172] Ibrahim Turkoglu, Ahmet Arslan, and Erdogan Ilkay. A wavelet neural network for the detection of heart valve diseases. *Expert Systems*, 2003.

[173] Davut Hanbay. An expert system based on least square support vector machines for diagnosis of the valvular heart disease. *Expert Systems with Applications*, 36(3 PART 1):4232–4238, 2009.

[174] Fabian Sanchis-Gomar, Carme Perez-Quilis, Roman Leischik, and Alejandro Lucia. Epidemiology of coronary heart disease and acute coronary syndrome. *Annals of Translational Medicine*, 4(13):1–12, 2016.

[175] Isabel A. Wright, Nigel A.J. Gough, Frank Rakebrandt, Mohamed Wahab, and John P. Woodcock. Neural network analysis of Doppler ultrasound blood flow signals: A pilot study. *Ultrasound in Medicine and Biology*, 1997.

[176] R. G. Gosling and D. H. King. The role of measurement in peripheral vascular surgery: Arterial assessment by Doppler-shift ultrasound. *Journal of the Royal Society of Medicine*, 67:447–449, 1974.

[177] Rosemary S. Thompson, Brian J. Trudinger, and Colleen M. Cook. Doppler ultrasound waveform indices: A/B ratio, pulsatility index and Pourcelot ratio. *BJOG: An International Journal of Obstetrics & Gynaecology*, 95(6):581–588, 1988.

[178] Sarah Bower, Katharina Schuchter, and Stuart Campbell. Doppler ultrasound screening as part of routine antenatal scanning: prediction of pre-eclampsia and intrauterine growth retardation. *BJOG: An International Journal of Obstetrics & Gynaecology*, 100(11):989–994, 1993.

[179] J. J. Baskett, M. G. Beasley, G. J. Murphy, D. E. Hyams, and R. G. Gosling. Screening for carotid junction disease by spectral analysis of doppler signals. *Cardiovascular Research*, 11(2):147–155, 1977.

[180] K. W. Johnston, M. Kassam, J. Koers, R. S.C. Cobbold, and D. MacHattie. Comparative study of four methods for quantifying doppler ultrasound waveforms from the femoral artery. *Ultrasound in Medicine and Biology*, 10(1):1–12, 1984.

[181] R. G. Gosling, G. Dunbar, D. H. King, D. L. Newman, C. D. Side, J. P. Woodcock, D. E. Fitzgerald, J. S. Keates, and D. Macmillan. The quantitative analysis of occlusive peripheral arterial disease by a non-intrusive ultrasonic technique. *Angiology*, 22(1), 1971.

[182] R. Skidmore and J. P. Woodcock. Physiological interpretation of Doppler-shift waveforms-II. Validation of the laplace transform method for characterisation of the common femoral blood-velocity/time waveform. *Ultrasound in Medicine and Biology*, 1980.

[183] R. Skidmore and J. P. Woodcock. Physiological interpretation of Doppler-shift waveforms-I: Theoretical considerations. *Ultrasound in Medicine and Biology*, 6(1):7–10, 1980.

[184] R. Skidmore, J. P. Woodcock, P. N.T. Wells, D. Bird, and R. N. Baird. Physiological interpretation of Doppler-shift waveforms-III. Clinical results. *Ultrasound in Medicine and Biology*, 1980.

[185] Daniel J. Reddy, G. Steven Vincent, Michalene McPharlin, and Calvin B. Ernst. Limitations of the femoral artery pulsatility index with aortoiliac artery stenosis: An experimental study. *Journal of Vascular Surgery*, 4(4):327–332, 1986.

[186] D. H. Evans, W. W. Barrie, M. J. Asher, S. Bentley, and P. R. Bell. The relationship between ultrasonic pulsatility index and proximal arterial stenosis in a canine model. *Circulation Research*, 46(4):470–475, 1980.

[187] J. Dennis Baker, Robert Skidmore, and Susan E.A. Cole. LaPlace transform analysis of femoral artery doppler signals: The state of the art. *Ultrasound in Medicine and Biology*, 1989.

[188] R O Duda, P E Hart, and D G Stork. Pattern classification. *New York: John Wiley, Section*, 2001.

[189] Yves E. Langlois, Francis M. Greene, Ghislaine O. Roederer, Kurt A. Jäger, David J. Phillips, Kirk W. Beach, and D. Eugene Strandness. Computer based pattern recognition of carotid artery Doppler signals for disease classification: Prospective validation. *Ultrasound in Medicine and Biology*, 10(5):581–595, 1984.

[190] W. N.J.C. van Asten, W. J. Beijneveld, B. R. Pieters, H. J.J. van Lier, P. F.F. Wijn, and S. H. Skotnicki. Assessment of aortoiliac obstructive disease by Doppler spectrum analysis of blood flow velocities in the common femoral artery at rest and during reactive hyperemia. *Surgery*, 1991.

[191] Kazuko Imaeda and Frank O. Goodman. Analysis of non-linear pulsatile blood flow in arteries. *Journal of Biomechanics*, 1980.

[192] Wilmer W. Nichols, Michael F. O'Rourke, Charalambos Vlachopoulos, Arnold P. Hoeks, and Robert S. Reneman. *McDonald's blood flow in arteries theoretical, experimental and clinical principles*. CRC Press., 2011.

[193] Huseyin Seker, David H. Evans, Nizamettin Aydin, and Ertugrul Yazgan. Compensatory fuzzy neural networks-based intelligent detection of abnormal neonatal cerebral Doppler ultrasound waveforms. *IEEE Transactions on Information Technology in Biomedicine*, 5(3):187–194, 2001.

[194] Harun Uğuz and Halife Kodaz. Classification of internal carotid artery Doppler signals using fuzzy discrete hidden Markov model. *Expert Systems with Applications*, 38(6):7407–7414, 2011.

[195] Emre Çomak, Ahmet Arslan, and Ibrahim Türkoğlu. A decision support system based on support vector machines for diagnosis of the heart valve diseases. *Computers in Biology and Medicine*, 37(1):21–27, 2007.

[196] Engin Avci and Ibrahim Turkoglu. An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases. *Expert Systems with Applications*, 36(2 PART 2):2873–2878, 2009.

[197] Sadik Kara. Classification of mitral stenosis from Doppler signals using short time Fourier transform and artificial neural networks. *Expert Systems with Applications*, 33(2):468–475, 2007.

[198] Resul Das, Ibrahim Turkoglu, and Abdulkadir Sengur. Diagnosis of valvular heart disease through neural networks ensembles. *Computer Methods and Programs in Biomedicine*, 93(2):185–191, 2009.

[199] Fatma Latifoğlu, Kemal Polat, Sadik Kara, and Salih Güneş. Medical diagnosis of atherosclerosis from carotid artery Doppler signals using principal component analysis (PCA), k-NN based weighting pre-processing and artificial immune recognition system (AIRS). *Journal of Biomedical Informatics*, 41(1):15–23, 2008.

[200] Fatma Latifoglu, Seral Şahan, Sadik Kara, and Salih Güneş. Diagnosis of atherosclerosis from carotid artery Doppler signals as a real-world medical application of artificial immune systems. *Expert Systems with Applications*, 33(3):786–793, 2007.

[201] Seral Özşen, Sadik Kara, Fatma Latifoğlu, and Salih Güneş. A new supervised classification algorithm in artificial immune systems with its application to carotid artery Doppler signals to diagnose atherosclerosis. *Computer Methods and Programs in Biomedicine*, 88(3):246–255, 2007.

[202] Kemal Polat, Sadik Kara, Fatma Latifoğlu, and Salih Güneş. Pattern detection of atherosclerosis from carotid artery Doppler signals using fuzzy weighted preprocessing and Least Square Support Vector Machine (LSSVM). *Annals of Biomedical Engineering*, 35(5):724–732, 2007.

[203] Murat Ceylan, Rahime Ceylan, Fatma Dirgenali, Sadik Kara, and Yüksel Özbay. Classification of carotid artery Doppler signals in the early phase of atherosclerosis using complex-valued artificial neural network. *Computers in Biology and Medicine*, 37(1):28–36, 2007.

[204] Elif Derya Übeyli and Inan Güler. Improving medical diagnostic accuracy of ultrasound Doppler signals by combining neural network models. *Computers in Biology and Medicine*, 35(6):533–554, 2005.

[205] Inan Güler and Elif Derya Übeyli. Automated diagnostic systems with diverse and composite features for Doppler ultrasound signals. *IEEE Transactions on Biomedical Engineering*, 53(10):1934–1942, 2006.

[206] Inan Güler and Elif Derya Übeyli. A recurrent neural network classifier for Doppler ultrasound blood flow signals. *Pattern Recognition Letters*, 27(13):1560–1571, 2006.

[207] Elif Derya Übeyli and Inan Güler. Neural network analysis of internal carotid arterial Doppler signals: Predictions of stenosis and occlusion. *Expert Systems with Applications*, 25(1):1–13, 2003.

[208] Ahmed F. Seddik and Doaa M. Shawky. A low-cost screening method for the detection of the carotid artery diseases. *Knowledge-Based Systems*, 52:236–245, 2013.

[209] Syed Naeem Ahmed. Essential statistics for data analysis. In *Physics and Engineering of Radiation Detection*. 2015.

[210] L. Valentin, L. Ameye, L. Savelli, R. Fruscio, F. P.G. Leone, A. Czekierdowski, A. A. Lissoni, D. Fischerova, S. Guerriero, C. Van Holsbeke, S. Van Huffel, and D. Timmerman. Adnexal masses difficult to classify as benign or malignant using subjective assessment of gray-scale and Doppler ultrasound findings: Logistic regression models do not help. *Ultrasound in Obstetrics and Gynecology*, 38(4):456–465, 2011.

[211] L. Valentin, B. Hagen, S. Tingulstad, and S. Eik-Nes. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: A prospective cross validation. *Ultrasound in Obstetrics and Gynecology*, 18(4):357–365, 2001.

[212] L. Pourcelot. Applications cliniques de l'examen Doppler transcutane. *Velocimetrie ultrasonore Doppler*, 34:780–785, 1974.

[213] A. Kurjak, I. Zalud, and Z. Alfirevic. Evaluation of adnexal masses with transvaginal color ultrasound. *Journal of Ultrasound in Medicine*, 1991.

[214] A. Tekay and P. Jouppila. Validity of pulsatility and resistance indices in classification of adnexal tumors with transvaginal color Doppler ultrasound. *Ultrasound in Obstetrics and Gynecology*, 1992.

[215] Kohkichi Hata, Toshiyuki Hata, Atsushi Manabe, Kazuro Sugimura, and Manabu Kitao. A critical evaluation of transvaginal doppler studies, transvaginal sonography, magnetic resonance imaging, and CA 125 in detecting ovarian cancer. *Obstetrics and Gynecology*, 1992.

[216] H. J. Prömpeler, H. Madjar, W. Sauerbrei, U. Lattermann, and A. Pfleiderer. Quantitative flow measurements for classification of ovarian tumors by transvaginal color Doppler sonography in postmenopausal patients. *Ultrasound in Obstetrics and Gynecology*, 1994.

[217] Kohkichi Hata, Toshiyuki Hata, and Manabu Kitao. Intratumoral peak systolic velocity as a new possible predictor for detection of adnexal malignancy. *American Journal of Obstetrics and Gynecology*, 172(5):1496–1500, 1995.

[218] A. Tailor, D. Jurkovic, T. H. Bourne, W. P. Collins, and S. Campbell. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. *Ultrasound in Obstetrics and Gynecology*, 1997.

[219] Anil Tailor, Davor Jurkovic, Thomas H. Bourne, William P. Collins, and Stuart Campbell. Sonographic prediction of malignancy in adnexal masses using an artificial neural network. *BJOG: An International Journal of Obstetrics and Gynaecology*, 106(1):21–30, 1999.

[220] Amjed S. Al-Fahoum and Ausilah A. Al-Fraihat. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *ISRN Neuroscience*, 2014:1–7, 2014.

[221] G Tzanetakis, G Essl, and P.˜R. Cook. Audio analysis using the discrete wavelet transform. *Proceedings of the WSES International Conference Acoustics and Music: Theory and Applications (AMTA 2001)*, pages 318–323, 2001.

[222] Jean Yves David, Steven A. Jones, and Don P. Giddens. Modem spectral analysis techniques for blood flow velocity and spectral measurements with pulsed Doppler ultrasound. *IEEE Transactions on Biomedical Engineering*, 1991.

[223] Reijo Takalo, Heli Hytti, and Heimo Ihalainen. Tutorial on univariate autoregressive spectral analysis. *Journal of Clinical Monitoring and Computing*, 19(6):401–410, 2005.

[224] P. I J Keeton and F. S. Schlindwein. Spectral broadening of clinical Doppler signals using FFT and autoregressive modelling. *European Journal of Ultrasound*, 7(3):209–218, 1998.

[225] Keith A. Wear, Robert F. Wagner, and Brian S. Garra. A comparison of autoregressive spectral estimation algorithms and order determination methods in ultrasonic tissue characterization. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 42(4):709–716, 1995.

[226] L. Marple. Resolution of conventional Fourier, autoregressive, and special ARMA methods of spectrum analysis. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 1977-May:74–77, 1977.

[227] K. Kaluzynski and T. Palko. Effect of method and parameters of spectral analysis on selected indices of simulated Doppler spectra. *Medical & Biological Engineering & Computing*, 1993.

[228] Alex Grossmann and Jean Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. In *Fundamental Papers in Wavelet Theory*. 1984.

[229] Nasser Kehtarnavaz and Namjin Kim. Digital Signal Processing System-Level Design Using LabVIEW. In *Digital Signal Processing System-Level Design Using LabVIEW*. Elsevier Inc., 2005.

[230] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets II: Variations on a theme. *SIAM Journal on Mathematical Analysis*, 24(2):499–519, 1988.

[231] Stephane G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674 – 693, 1989.

[232] Donald B. Percival and Debashis Mondal. *A Wavelet Variance Primer*, volume 30. Elsevier B.V., 2012.

[233] Yu Shao and Chip Hong Chang. A generalized time-frequency subtraction method for robust speech enhancement based on wavelet filter banks modeling of human auditory system. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(4):877–889, 2007.

[234] Skander Soltani. On the use of the wavelet decomposition for time series prediction. *Neurocomputing*, 48(1-4):267–277, 2002.

[235] Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

[236] N Ahuja, S Lertrattanapanich, and K Bose, N. Properties determining choice of mother wavelet. *IEEE Image Signal Process.*, 152(5):659–664, 2005.

[237] Yan Fang Sang. A practical guide to discrete wavelet decomposition of hydrologic time series. *Water Resources Management*, 26(11):3345–3365, 2012.

[238] T. Thayaparan, S. Abrol, E. Riseborough, L. Stankovic, D. Lamothe, and G. Duff. Analysis of radar micro-Doppler signatures from experimental helicopter and human data. *IET Radar, Sonar and Navigation*, 2007.

[239] Metin Akay. Wavelet applications in medicine. *IEEE Spectrum*, 1997.

[240] Tamer Ölmez and Zümray Dokur. Classification of heart sounds using an artificial neural network. *Pattern Recognition Letters*, 2003.

[241] Ronald R. Coifman and Mladen Victor Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, 1992.

[242] Brent C.B. Chan, Francis H.Y. Chan, K. Lam, Ping Wing Lui, and Paul W.F. Poon. Fast detection of venous air embolism in Doppler heart sound using the wavelet transform. *IEEE Transactions on Biomedical Engineering*, 44(4):237–246, 1997.

[243] Jonathan Lazar, Daniel Goldstein, and Anne Taylor. The history of access technology. In *Ensuring Digital Accessibility Through Process and Policy*. 2015.

[244] Pablo Diez. *Smart Wheelchairs and Brain-computer Interfaces*. Elsevier Inc., second edition, 2018.

[245] Olli Viikki and Kari Laurila. Cepstral domain segmental feature vector normalization for noise robust speech recognition. *Speech Communication*, 25(1-3):133–147, 1998.

[246] Paul Mermelstein. Distance measures for speech recognition, psychological and instrumental. *Pattern recognition and artificial intelligence*, 1976.

[247] Priyanka A. Abhang, Bharti W. Gawali, and Suresh C. Mehrotra. Technical aspects of brain rhythms and speech parameters. In *Introduction to EEG- and Speech-Based Emotion Recognition*. Academic Press, 2016.

[248] Dalila Yessad, Abderrahmane Amrouche, Mohamed Debyeche, and Mustapha Djeddou. Micro-Doppler classification for ground surveillance radar using speech recognition tools. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7042 LNCS:280–287, 2011.

[249] Mehmet Cenk Sezgin, Bilge Gunsel, and Gunes Karabulut Kurt. Perceptual audio features for emotion detection. *Eurasip Journal on Audio, Speech, and Music Processing*, 2012(1):1–21, 2012.

[250] Md Sahidullah and Goutam Saha. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Communication*, 54(4):543–565, 2012.

[251] Harisudha Kuresan, Dhanalakshmi Samiappan, and Sam Masunda. Fusion of WPT and MFCC feature extraction in parkinsons disease diagnosis. *Technology and Health Care*, 27(4):363–372, 2019.

[252] Selina Chu, Shrikanth Narayanan, and C. C.Jay Kuo. Environmental sound recognition with timeFrequency audio features. *IEEE Transactions on Audio, Speech and Language Processing*, 2009.

[253] Sunita Chauhan, Ping Wang, Chu Sing Lim, and V. Anantharaman. A computer-aided MFCC-based HMM system for automatic auscultation. *Computers in Biology and Medicine*, 38(2):221–233, 2008.

[254] P. Lubaib and K.V. Ahammed Muneer. The heart defect analysis based on PCG signals using pattern recognition techniques. *Procedia Technology*, 24:1024–1031, 2016.

[255] Nandini Sengupta, Md Sahidullah, and Goutam Saha. Lung sound classification using cepstral-based statistical features. *Computers in Biology and Medicine*, 75(May):118–129, 2016.

[256] Mohamed Boussaa, Issam Atouf, Mohamed Atibi, and Abdellatif Bennis. ECG signals classification using MFCC coefficients and ANN classifier. *Proceedings of 2016 International Conference on Electrical and Information Technologies, ICEIT 2016*, (November 2017):480–484, 2016.

[257] V Rathikarani, P Dhanalakshmi, Rathikarani V, V Rathikarani, and P Dhanalakshmi. Automatic classification of ECG signal for identifying arrhythmia. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(9):205–211, 2013.

[258] Abdulkadir Eryildirim and Ibrahim Onaran. Pulse Doppler radar target recognition using a two-stage SVM procedure. *IEEE Transactions on Aerospace and Electronic Systems*, 47(2):1450–1457, 2011.

[259] Liang Liu, Mihail Popescu, Marjorie Skubic, Marilyn Rantz, Tarik Yardibi, and Paul Cuddihy. Automatic fall detection based on Doppler radar motion signature. *2011 5th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2011*, pages 222–225, 2011.

[260] Yang Shao, Zhaozhang Jin, Deliang Wang, and Soundararajan Srinivasan. An auditory-based feature for robust speech recognition. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, (1):4625–4628, 2009.

[261] Xavier Valero and Francesc Alias. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. *IEEE Transactions on Multimedia*, 2012.

[262] Jun Qi, Dong Wang, Ji Xu, and Javier Tejedor. Bottleneck features based on gammatone frequency cepstral coefficients. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, (August):1751–1755, 2013.

[263] Jia Ming Liu, Mingyu You, Guo Zheng Li, Zheng Wang, Xianghuai Xu, Zhongmin Qiu, Wenjia Xie, Chao An, and Sili Chen. Cough signal recognition with gammatone cepstral coefficients. *2013 IEEE China Summit and International Conference on*

Signal and Information Processing, ChinaSIP 2013 - Proceedings, (July):160–164, 2013.

[264] Hari Krishna Maganti and Marco Matassoni. Auditory processing-based features for improving speech recognition in adverse acoustic conditions. *Eurasip Journal on Audio, Speech, and Music Processing*, 2014:1–9, 2014.

[265] Xiaojia Zhao and Deliang Wang. Analyzing noise robustness of MFCC and GFCC features in speaker identification. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 7204–7208, 2013.

[266] L. H. Carney and T. C.T. Yin. Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model. *Journal of Neurophysiology*, 1988.

[267] Haowei Liu. Face technologies on mobile devices. In *Facial Detection and Recognition on Mobile Devices*. Elsevier Inc., 2015.

[268] J. H. Cheng, Da Wen Sun, M. Nagata, and J. G. Tallada. Computer Vision Technology for Food Quality Evaluation: Second Edition. 2016.

[269] Andries Meintjes, Andrew Lowe, and Malcolm Legget. Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018-July:409–412, 2018.

[270] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[271] Li Liu, Songyang Lao, Paul W. Fieguth, Yulan Guo, Xiaogang Wang, and Matti Pietikäinen. Median robust extended local binary pattern for texture classification. *IEEE Transactions on Image Processing*, 25(3):1368–1381, 2016.

[272] Arnau Oliver, Xavier Liado, Jordi Freixenet, and Joan Martí. False positive reduction in mammographic mass detection using local binary patterns. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007.

[273] U. Rajendra Acharya, S. Vinitha Sree, M. Muthu Rama Krishnan, N. Krishnananda, Shetty Ranjan, Pai Umesh, and Jasjit S. Suri. Automated classification of patients with coronary artery disease using grayscale features from left ventricle echocardiographic images. *Computer Methods and Programs in Biomedicine*, 112(3):624–632, 2013.

[274] Arun Balodi, R. S. Anand, M. L. Dewal, and Anurag Rawat. Computer-aided classification of the mitral regurgitation using multiresolution local binary pattern. *Neural Computing and Applications*, 3456789, 2019.

[275] Dalal Bardou, Kun Zhang, and Sayed Mohammad Ahmad. Lung sounds classification using convolutional neural networks. *Artificial Intelligence in Medicine*, 88:58–69, 2018.

[276] Nandini Sengupta, Md Sahidullah, and Goutam Saha. Lung sound classification using local binary pattern. *Engineering, Computer Science*, pages 1–21, 2017.

[277] Xiaoyu Wang, Tony X. Han, and Shuicheng Yan. An HOG-LBP human detector with partial occlusion handling. (June):32–39, 2010.

[278] Junge Zhang, Kaiqi Huang, Yinan Yu, and Tieniu Tan. Boosted local structured HOG-LBP for object localization. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1393–1400, 2011.

[279] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. *International Conference on Computer Vision & Pattern Recognition*, pages 886–893, 2005.

[280] Ahmad, Rehan. A take on h.o.g feature descriptor, 2019. `https://medium.com/analytics-vidhya/a-take-on-h-o-g-feature-descriptor-e839ebba1e52/`, Last accessed on 2020-03-03.

[281] Harihara Santosh Dadi and Gopala Krishna Mohan Pillutla. Improved face recognition rate using HOG features and SVM classifier. *IOSR Journal of Electronics and Communication Engineering*, 11(04):34–44, 2016.

[282] Susanto Kumar Ghosh and Rafiqul Islam. *Bird species detection and classification based on HOG feature using convolutional neural network*, volume 1035. Springer Singapore, 2019.

[283] Alain Rakotomamonjy and Gilles Gasso. Histogram of gradients of time-frequency representations for audio scene classification. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 23(1):142–153, 2015.

[284] Victor Bisot, Slim Essid, and Gael Richard. HOG and subband power distribution image features for acoustic scene classification. *2015 23rd European Signal Processing Conference, EUSIPCO 2015*, pages 719–723, 2015.

[285] Fatih Demir, Abdulkadir Sengur, Nicholas Cummins, Shahin Amiriparian, and Bjorn Schuller. Low level texture features for snore sound discrimination. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018-July(October):413–416, 2018.

[286] O. Farooq and S. Datta. Mel filter-like admissible wavelet packet structure for speech recognition. *IEEE Signal Processing Letters*, 2001.

[287] Andrew Taylor, Graeme Watson, Gordon Grigg, and Hamish McCallum. Monitoring frog communities: An application of machine learning. *Proceedings of the 8th Innovative Applications of Artificial Conference*, pages 1564–1569, 1996.

[288] Chuen Kai Shie, Chung Hisang Chuang, Chun Nan Chou, Meng Hsi Wu, and Edward Y. Chang. Transfer representation learning for medical image analysis. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015-Novem:711–714, 2015.

[289] Mehmet Saygin Seyfioglu and Sevgi Zubeyde Gurbuz. Deep neural network initialization methods for micro-Doppler classification with low training sample support. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2462–2466, 2017.

[290] Paras Lakhani and Baskaran Sundaram. Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2):574–582, 2017.

[291] Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006.

[292] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. In *Advances in Neural Information Processing Systems*, 2007.

[293] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review, 2017.

[294] Michael Paluszek and Stephanie Thomas. *MATLAB Machine Learning*. Apress, 2017.

[295] IMAGENET. Large scale visual recognition challenge, 2020. `http://www.image-net.org/challenges/LSVRC/`, Last accessed on 2020-01-02.

[296] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2012.

[297] Hoo Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M. Summers. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016.

[298] Ali Abd Almisreb, Nursuriati Jamil, and N. Md Din. Utilizing AlexNet deep transfer learning for ear recognition. *Proceedings - 2018 4th International Conference on Information Retrieval and Knowledge Management: Diving into Data Sciences, CAMP 2018*, pages 8–12, 2018.

[299] Moi Hoon Yap, Gerard Pons, Joan Martí, Sergi Ganau, Melcior Sentís, Reyer Zwiggelaar, Adrian K. Davison, and Robert Martí. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics*, 22(4):1218–1226, 2018.

[300] Shui Hua Wang, Shipeng Xie, Xianqing Chen, David S. Guttery, Chaosheng Tang, Junding Sun, and Yu Dong Zhang. Alcoholism identification based on an Alexnet transfer learning model. *Frontiers in Psychiatry*, 10(APR):1–13, 2019.

[301] Qing Guan, Yunjun Wang, Jiajun Du, Yu Qin, Hongtao Lu, Jun Xiang, and Fen Wang. Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study. *Annals of Translational Medicine*, 7(7):137–137, 2019.

[302] Jui Ying Chiao, Kuan Yung Chen, Ken Ying Kai Liao, Po Hsin Hsieh, Geoffrey Zhang, and Tzung Chi Huang. Detection and classification the breast tumors using mask R-CNN on sonograms. *Medicine*, 98(19):e15200, 2019.

[303] Xiang Liu, Jialin Song, Shuo Hong Wang, Jingwen Zhao, and Yanqiu Chen. Learning to diagnose cirrhosis with liver capsule guided ultrasound image classification. *Sensors (Switzerland)*, 17(1):1–11, 2017.

[304] Yoshiko Ariji, Motoki Fukuda, Yoshitaka Kise, Michihito Nozawa, Toru Nagao, Atsushi Nakayama, Yoshihiko Sugita, Akitoshi Katumata, and Eiichiro Ariji. A preliminary application of intraoral Doppler ultrasound images to deep learning techniques for predicting late cervical lymph node metastasis in early tongue cancers. *Oral Science International*, (October):1–8, 2019.

[305] Victor C. Chen, Fayin Li, Shen Shyang Ho, and Harry Wechsler. Micro-Doppler effect in radar: Phenomenon, model, and simulation study. *IEEE Transactions on Aerospace and Electronic Systems*, 2006.

[306] Mehmet Saygin Seyfioglu, Sevgi Zubeyde Gurbuz, Ahmet Murat Ozbayoglu, and Melda Yuksel. Deep learning of micro-Doppler features for aided and unaided gait recognition. *2017 IEEE Radar Conference, RadarConf 2017*, pages 1125–1130, 2017.

[307] Ibrahim Alnujaim, Daegun Oh, and Youngwook Kim. Generative adversarial networks to augment micro-Doppler signatures for the classification of human activity. *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 9459–9461, 2019.

[308] Zhaoxi Chen, Gang Li, Francesco Fioranelli, and Hugh Griffiths. Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 2018.

[309] Mehmet S. Seyfioglu, Baris Erol, Sevgi Z. Gurbuz, and Moeness G. Amin. Diversified radar micro-Doppler simulations as training data for deep residual neural networks. In *2018 IEEE Radar Conference, RadarConf 2018*, 2018.

[310] Albert Huizing, Matijs Heiligers, Bastiaan Dekker, Jacco De Wit, Lorenzo Cifola, and Ronny Harmanny. Deep learning for classification of mini-uavs using micro-Doppler spectrograms in cognitive radar. *IEEE Aerospace and Electronic Systems Magazine*, 34(11):46–56, 2019.

[311] Uscom Ltd. USCOM 1A : User Manual, 2013.

[312] Jacobson, Beverley. Uscom 1a targeting training, 2018. Training session provided by USCOM Ltd.

[313] G. N. Cattermole, P.Y.M Leung, and C.O Tang. A new method to score the quality of USCOM scans. *HK J Emerg Med*, 16:288, 2009.

[314] J. Gonzalez-Alonso, R. Mora-Rodriguez, P. R. Below, and E. F. Coyle. Dehydration reduces cardiac output and increases systemic and cutaneous vascular resistance during exercise. *Journal of Applied Physiology*, 79(5):1487–1496, 1995.

[315] Fabio Cavallaro, Claudio Sandroni, Cristina Marano, Giuseppe La Torre, Alice Mannocci, Chiara De Waure, Giuseppe Bello, Riccardo Maviglia, and Massimo Antonelli. Diagnostic accuracy of passive leg raising for prediction of fluid responsiveness in adults: Systematic review and meta-analysis of clinical studies. *Intensive Care Medicine*, 36(9):1475–1483, 2010.

[316] Edward L Melanson and P. S. Freedson. The effect of endurance training on resting heart rate variability in sedentary adult males. *European Journal of Applied Physiology*, 85(5):442–449, 2001.

[317] Y. Wang, R. J. Marshall, and J. T. Shepherd. The effect of changes in posture and of graded exercise on stroke volume in man. *The Journal of clinical investigation*, 1960.

[318] Matthew D.C. Eames, Arik Hananel, John W. Snell, Neal F. Kassell, and Jean Francois Aubry. Trans-cranial focused ultrasound without hair shaving: Feasibility study in an ex vivo cadaver model. *Journal of Therapeutic Ultrasound*, 2014.

[319] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[320] Dev Maulik. Spectral doppler: Basic principles and instrumentation. *Doppler Ultrasound in Obstetrics and Gynecology: 2nd Revised and Enlarged Edition*, pages 19–34, 2005.

[321] Robert W. Gill. Measurement of blood flow by ultrasound: Accuracy and sources of error. *Ultrasound in Medicine and Biology*, 11(4):625–641, 1985.

[322] Yu Zhang, Jose C. Cardoso, Yuanyuan Wang, Peter J. Fish, Carlos A.C. Bastos, and Weiqi Wang. Time-scale removal of "all thum" in Doppler ultrasound signals: A simulation study. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 51(9):1187–1192, 2004.

[323] Peter Fish. *Physics and Instrumentation of Diagnotic Medical Ultrasound*. John Wiley & Sons, Chichester, Sussex, 1990.

[324] Nagashettappa Biradar, Mohan Lal Dewal, and Manoj Kumar Rohit. Comparative analysis of despeckling filters for continuous wave Doppler images. *Biomedical Engineering Letters*, 5(1):33–44, 2015.

[325] Juerg Tchirren, Ronald M Lauer, and MIlan Sonka. Automated analysis of Doppler ultrasound velocity flow diagrams. *IEEE Transactions on Medical Imaging*, 23(1):130–133, 2001.

[326] T. D'Alessio. 'Objective' algorithm for maximum frequency estimation in Doppler spectral analysers. *Medical & Biological Engineering & Computing*, 1985.

[327] Larry Y. L. Mo, Louis C.M. Yun, and R. S.C. Cobbold. Comparison of four digital maximum frequency estimators for Doppler ultrasound. *Ultrasound in Medicine and Biology*, 14(5):355–363, 1988.

[328] Jørgen A Jensen. Calculation of pulsed wave phantom data for arteria femoralis, 2019.

[329] I. T. Gabe, J. H. Gault, J. Ross, D. T. Mason, C. J. Mills, J. P. Schillingford, and E. Braunwald. Measurement of instantaneous blood flow velocity and pressure in conscious man with a catheter-tip velocity probe. *Circulation*, 1969.

[330] Lester A H Critchley and Li Huang. USCOM - Window to the circulation: Utility of supra-sternal Doppler in an elderly anaesthetized patient for a robotic cystectomy. *Journal of Clinical Monitoring and Computing*, 28(1):83–93, 2014.

[331] Judith A. Finegold, Charlotte H. Manisty, Fabrizio Cecaro, Nilesh Sutaria, Jamil Mayet, and Darrel P. Francis. Choosing between velocity-time-integral ratio and peak velocity ratio for calculation of the dimensionless index (or aortic valve area) in serial follow-up of aortic stenosis. *International Journal of Cardiology*, 167(4):1524–1531, 2013.

[332] Charlotte H. Manisty, Ali Al-Hussaini, Beth Unsworth, Resham Baruah, Punam A. Pabari, Jamil Mayet, Alun D. Hughes, Zachary I. Whinnett, and Darrel P. Francis. The acute effects of changes to AV delay on BP and stroke volume potential implications for design of pacemaker optimization protocols. *Circulation: Arrhythmia and Electrophysiology*, 5(1):122–130, 2012.

[333] Anthony D. Jose and D. Collison. The normal range and determinants of the intrinsic heart rate in man. *Cardiovascular Research*, 4(2):160–167, 1970.

[334] Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, 2019.

[335] James B. Seward, Pamela S. Douglas, Raimund Erbel, Richard H. Kerber, Itzhak Kronzon, Harry Rakowski, L. David J. Sahn, Eric J. Sisk, A. Jamil Tajik, and Samuel Wann. Hand-carried cardiac ultrasound (HCU) device: Recommendations regarding new technology. A report from the echocardiography task force on new technology of the nomenclature and standards committee of the American Society of Echocardiography. *Journal of the American Society of Echocardiography*, 15(4):369–373, 2002.

[336] World Health Organization. Clinical management of severe acute respiratory infection when novel coronavirus (nCoV) infection is suspected. (January):12, 2020.

[337] Diogo Marcelo Nogueira, Carlos Abreu Ferreira, Elsa Ferreira Gomes, and Alípio M. Jorge. Classifying heart sounds using images of motifs, MFCC and temporal features. *Journal of Medical Systems*, 43(6), 2019.

[338] P Arbeille, M. Berson, F. Achaibou, S. Bodard, and A. Locatelli. Vascular resistance quantification in high flow resistance areas using the Doppler method. *Ultrasound in Medicine and Biology*, 1995.

[339] Arnost Fronek, Kaj H. Johansen, Ralph B. Dilley, and Eugene F. Bernstein. Noninvasive physiologic tests in the diagnosis and characterization of peripheral arterial occlusive disease. *American Journal of Surgery*, 126(2):205–214, 1973.

[340] P. R. Stone, S. MacPhail, S. B. Bull, and R. J. Morrow. Use of Laplace transform analysis to describe the effect of placental embolisation on umbilical arterial Doppler waveforms in fetal sheep. *Ultrasound in Medicine and Biology*, 1994.

[341] Karen J. Brackley, Margaret M. Ramsay, Fiona Broughton Pipkin, and Peter C. Rubin. A longitudinal study of maternal bloodflow in normal pregnancy and the puerperium: Analysis of Doppler waveforms using Laplace transform techniques. *BJOG: An International Journal of Obstetrics and Gynaecology*, 105(1):68–77, 1998.

[342] Karen J. Brackley, Margaret M. Ramsay, Fiona Broughton Pipkin, and Peter C. Rubin. The maternal cerebral circulation in pre-eclampsia: Investigations using Laplace transform analysis of Doppler waveforms. *British Journal of Obstetrics and Gynaecology*, 107(4):492–500, 2000.

[343] Taiyong Li and Min Zhou. ECG classification using wavelet packet entropy and random forests. *Entropy*, 18(8):1–16, 2016.

[344] Emina Alickovic and Abdulhamit Subasi. Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier. *Journal of Medical Systems*, 40(4):1–12, 2016.

[345] Elizabeth Ann Maharaj and Andrés M. Alonso. Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals. *Computational Statistics and Data Analysis*, 70:67–87, 2014.

[346] Revathi Sundarasekar, M. Thanjaivadivel, Gunasekaran Manogaran, Priyan Malarvizhi Kumar, R. Varatharajan, Naveen Chilamkurti, and Ching Hsien Hsu. Internet of things with maximal overlap discrete wavelet transform for remote health monitoring of abnormal ECG signals. *Journal of Medical Systems*, 42(11):227–229, 2018.

[347] Charles R. Cornish, Christopher S. Bretherton, and Donald B. Percival. Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Boundary-Layer Meteorology*, 119(2):339–374, 2006.

[348] Donald B. Percival and Andrew T. Walden. *Wavelet methods for time series analysis*. Cambridge University Pres, 2000.

[349] Zhao Qibin and Zhang Liqing. ECG feature extraction and classification using wavelet transform and support vector machines. *Proceedings of 2005 International Conference on Neural Networks and Brain Proceedings, ICNNB'05*, 2:1089–1092, 2005.

[350] Elif Derya Ubeyli and Inan Güler. Spectral analysis of internal carotid arterial Doppler signals using FFT, AR, MA, and ARMA methods. *Computers in biology and medicine*, 34(4):293–306, 2004.

[351] Elif Derya Übeyli and Inan Güler. Feature extraction from Doppler ultrasound signals for automated diagnostic systems. *Computers in Biology and Medicine*, 35(9):735–764, 2005.

[352] P. M.T. Broersen. Facts and fiction in spectral analysis. *IEEE Transactions on Instrumentation and Measurement*, 2000.

[353] Honglak Lee, Largman Yan, Peter Pham, and Andrew Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*, pages 1096–1104, 2009.

[354] G Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *Technical report; Institut de Recherche Et Coordination Acoustique/Musique: Paris, France*, 2004.

[355] Hugues Vinet, Perfecto Herrera, and François Pachet. The CUIDADO project. *3rd International Society for Music Information Retrieval (ISMIR) Conference*, (May 2014):197–203, 2002.

[356] Julia H. Smith, Jim Graham, and Robert J. Taylor. The application of an artificial neural network to Doppler ultrasound waveforms for the classification of arterial disease. *International Journal of Clinical Monitoring and Computing*, 13(2):85–91, 1996.

[357] Joel Grus. Working with Data. In *Data science from scratch : First principles with Python*. O'Reilly Media, Inc., 2016.

[358] A. Jović, K. Brkić, and N. Bogunović. A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings*, (May):1200–1205, 2015.

[359] George H. John, Ron Kohavi, and Karl Pfleger. Irrelevant features and the subset selection problem. *Machine Learning Proceedings 1994*, pages 121–129, 1994.

[360] Mark A Hall. *Correlation-based feature selection for machine learning*. PhD thesis, Universiy of Waikato, 1999.

[361] Max Kuhn and Kjell Johnson. Feature Engineering and Selection. In *Feature Engineering and Selection: A Practical Approach For Predictive Models*. CRC Press., 2019.

[362] Sanmay Das. Filters, wrappers and a boosting-based hybrid for feature selection. *Engineering*, pages 74–81, 2001.

[363] Verónica Bolón-Canedo, Noelia Sánchez-Maroño, and Amparo Alonso-Betanzos. A review of feature selection methods on synthetic data. *Knowledge and Information Systems*, 34(3):483–519, 2013.

[364] Beatriz Remeseiro and Veronica Bolon-Canedo. A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, 112(July):103375, 2019.

[365] Noelia Sánchez-Maroño, Amparo Alonso-Betanzos, and María Tombilla-Sanromán. Filter methods for feature selection - A comparative study. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007.

[366] Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996.

[367] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2):301–320, 2005.

[368] Isabelle Iguyon and André Elisseeff. An introduction to variable and feature selection, 2003.

[369] David W. Aha and Richard L. Bankert. A comparative evaluation of sequential feature selection. In *Learning from data: AI and statistics V*, number 0. Springer Verlag, 1996.

[370] Kenji Kira and Larry A. Rendell. Feature selection problem: Traditional methods and a new algorithm. In *Proceedings Tenth National Conference on Artificial Intelligence*, 1992.

[371] Kenji Kira and Larry A. Rendell. A practical approach to feature selection. *Machine Learning Proceedings 1992*, pages 249–256, 1992.

[372] Sun Yijun and Li Jian. Iterative RELIEF for feature weighting. *ACM International Conference Proceeding Series*, 148(6):913–920, 2006.

[373] José Bins and Bruce A. Draper. Feature selection from huge feature sets. *Proceedings of the IEEE International Conference on Computer Vision*, 2:159–165, 2001.

[374] T. Marill and D. M. Green. On the effectiveness of receptors in recognition systems. *IEEE Transactions on Information Theory*, 9(1):11–17, 1963.

[375] J. Kittler. Feature selection and extraction. In *Handbook of pattern recognition and image processing*. Academic Press, 1986.

[376] Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1):16–28, 2014.

[377] David M. Allen. The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, 16(1):125–127, 1974.

[378] Zahid Akhtar and Tiago H. Falk. Audio-visual multimedia quality assessment: A comprehensive survey. *IEEE Access*, 5:21090–21117, 2017.

[379] Yanjiao Chen, Kaishun Wu, and Qian Zhang. From QoS to QoE: A tutorial on video quality assessment. *IEEE Communications Surveys and Tutorials*, 17(2):1126–1165, 2015.

[380] Shyamprasad Chikkerur, Vijay Sundaram, Martin Reisslein, and Lina J. Karam. Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE Transactions on Broadcasting*, 57(2 PART 1):165–182, 2011.

[381] Iñigo Sedano, Kjell Brunnström, Maria Kihl, and Andreas Aurelius. Full-reference video quality metric assisted the development of no-reference bitstream video quality metrics for real-time network monitoring. *Eurasip Journal on Image and Video Processing*, 2014:1–15, 2014.

[382] Eduardo José da S. Luz, William Robson Schwartz, Guillermo Cámara-Chávez, and David Menotti. ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine*, 127:144–164, 2016.

[383] Philip De Chazal, Maria O'Dwyer, and Richard B. Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51(7):1196–1206, 2004.

[384] Chusak Thanawattano and Thaweesak Yingthawornsuk. Cardiac arrhythmia classification using beat-by-beat autoregressive modeling. *International Journal of Computer and Electrical Engineering*, 12:762–765, 2012.

[385] Corinna Cortes and Vladimir Vapnik. Support-Vector Networks. *Machine Learning*, pages 273–297, 1995.

[386] Raoof Gholami and Nikoo Fakhari. Support vector machine: Principles, parameters, and applications. In *Handbook of neural computation*, pages 515–535. Elsevier Inc., 1 edition, 2017.

[387] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, 1996.

[388] Vijay Kotu and Bala Deshpande. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. 2014.

[389] Peter Wittek. Supervised learning and support vector machines. In *Quantum Machine Learning*, pages 73–84. Elsevier Inc., 2014.

[390] Yaniv Bar, Idit Diamant, Lior Wolf, and Hayit Greenspan. Deep learning with non-medical training used for chest pathology identification. In *Medical Imaging 2015: Computer-Aided Diagnosis*, 2015.

[391] Sangheum Hwang, Hyo-Eun Kim, Jihoon Jeong, and Hee-Jin Kim. A novel approach for tuberculosis screening based on deep convolutional neural networks. In *Medical Imaging 2016: Computer-Aided Diagnosis*, 2016.

[392] Shawn Hershey, Sourish Chaudhuri, Daniel P.W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, and Kevin Wilson. CNN architectures for large-scale audio classification. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2017.

[393] MATLAB. Transfer learning using alexnet, 2019. `https://uk.mathworks.com/help/deeplearning/examples/transfer-learning-using-alexnet/`, Last accessed on 2019-10-10.

[394] Vijay Kotu and Bala Deshpande. Model evaluation. In *Predictive analytics and data mining*. Morgan Kaufmann, 2015.

[395] Sadik Kara, Semra Kemaloğlu, and Ayşegül Güven. Detection of femoral artery occlusion from spectral density of Doppler signals using the artificial neural network. *Expert Systems with Applications*, 29(4):945–952, 2005.

[396] Yinghui Chen, Lei Zhang, David Zhang, and Dongyu Zhang. Pattern classification for Doppler ultrasonic wrist pulse signals. *3rd International Conference on Bioinformatics and Biomedical Engineering, iCBBE 2009*, (L):1–4, 2009.

[397] Seong Ho Park and Kyunghwa Han. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction 1 REVIEW: Evaluation of Artificial Intelligence Tools for Diagnostic or Predictive Analysis Park and Han. *radiology.rsna.org n Radiology Radiology*, 286(3—March), 2018.

[398] Tomonori Aoki, Atsuo Yamada, Kazuharu Aoyama, Hiroaki Saito, Gota Fujisawa, Nariaki Odawara, Ryo Kondo, Akiyoshi Tsuboi, Rei Ishibashi, Ayako Nakada, Ryota Niikura, Mitsuhiro Fujishiro, Shiro Oka, Soichiro Ishihara, Tomoki Matsuda, Masato Nakahori, Shinji Tanaka, Kazuhiko Koike, and Tomohiro Tada. Clinical usefulness of a deep learning-based system as the first screening on small-bowel capsule endoscopy reading. *Digestive Endoscopy*, 32(4):585–591, 2020.

[399] Dennis A. Adams, R. Ryan Nelson, and Peter A. Todd. Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly: Management Information Systems*, 1992.

[400] A. Dal Pozzolo, O. Caelen, and G. Bontempi. Comparison of balancing techniques for unbalanced datasets. *Machine Learning Group Universite Libre de Bruxelles Belgium*, 16(1):732–735, 2010.

[401] Dheeraj Arora, Rajesh Chand, Yatin Mehta, and Naresh Trehan. Cardiac output estimation after off-pump coronary artery bypass: a comparison of two different techniques. *Annals of cardiac anaesthesia*, 2007.

[402] Martin Boyle, Liz Steel, Gordon M. Flynn, Margherita Murgo, Lisa Nicholson, Maureen O'Brien, and David Bihari. Assessment of the clinical utility of an ultrasonic

monitor of cardiac output (the USCOM) and agreement with thermodilution measurement. *Critical care and resuscitation : journal of the Australasian Academy of Critical Care Medicine*, 2009.

[403] Sophia Horster, Hans Joachim Stemmler, Nina Strecker, Florian Brettner, Andreas Hausmann, Jitske Cnossen, Klaus G. Parhofer, Thomas Nickel, and Sandra Geiger. Cardiac output measurements in septic patients: Comparing the accuracy of USCOM to PiCCO. *Critical Care Research and Practice*, 2012, 2012.

[404] Sophia Horster, Hans Joachim Stemmler, Jakob Sparrer, Johanna Tischer, Andreas Hausmann, and Sandra Geiger. Mechanical ventilation with positive end-expiratory pressure in critically ill patients: Comparison of CW-Doppler ultrasound cardiac output monitoring (USCOM) and thermodilution (PiCCO). *Acta Cardiologica*, 2012.