Detection of Prognostically Significant Coronary Artery Disease in Stress Echocardiography using Artificial Intelligence

Ross Upton
BSc (Hons), MSc

Thesis presented for the degree of
Doctor of Philosophy in Cardiovascular Medicine

Supervisors: Professor Paul Leeson,
Professor Hugh Watkins

October 2019
Approximate word count: ~30,000 words



St Hilda's College, University of Oxford
&
Cardiovascular Clinical Research Facility
Division of Cardiovascular Medicine
Radcliffe Department of Medicine
University of Oxford

Aknowledgements

I am hugely grateful to all those that have supported me throughout these years and enabled me to carry out this body of work. Firstly, to my incredible wife Hannah who provides me continuous support and is a rock for our family, I could not have done this without you. Secondly to my daughter Imogen, you are my pride and joy, you make all the hard work, stressful times, early starts, long commutes worth it – this is all for you. To my parents, sister and the bears, thank you for all of your support over the years.

Thank you to both Cardiocity and Ultromics for funding this research. Thank you to my supervisor (and co-founder) Paul Leeson who has been instrumental in my professional development and guiding my research long before I started my DPhil and it continues to this day as co-founders of Ultromics, you have made all of this possible and I cannot thank you enough. Finally, I would like to thank CCRF and Ultromics for all of your support and contributions to this work over the years.

Dedications

For Imogen and OXCF001

Abbreviations

AI	Artificial Intelligence
AP2	Apical 2 chamber view
AP4	Apical 4 chamber view
AUC	Area Under Curve
BMI	Body mass index
BPM	Beats per minute
CAD	Coronary Artery Disease
CMR	Cardiac Magnetic Resonance
CNN	Convolutional Neural Network
CT	Computed Tomography
CV	Cross Validation
DICE	Sorenson-Dice coefficient
DBP	Diastolic blood pressure
Echo	Echocardiography
EDV	End Diastolic Volume
EMR	Electronic Medical Record
ESV	End systolic volume
EF	Ejection fraction
GLS	Global Longitudinal Strain
HR	Heart Rate
LR	Logistic Regression
LV	Left ventricle
LVH	Left ventricular hypertrophy
ML	Machine Learning
MLP	Multi-layer perceptron
NICE	National Institute for Health and Care Excellence
PCI	Percutaneous Coronary Intervention
PET	Positron Emission Tomography
RF	Random Forest
ROC	Receiver Operator Characteristic Curve
RMSE	Root Mean Squared Error
RWMA	Regional Wall Motion Abnormalities
SAX	Short Axis View
SD	Standard deviation
SPECT	Single Photon Emission Computed Tomography
SVM	Support Vector Machine
WMS	Wall Motion Score

Abstract

Coronary artery disease (CAD) is a major cause of morbidity and mortality worldwide, affecting 2.3 million people in the UK. Whilst CAD can be diagnosed through a variety of imaging modalities, stress echocardiography is the most commonly used technique, used in over 50% of hospitals in the UK. However, the current clinical standard for stress echo is very subjective and is only 81 % sensitive and 82% specific. Therefore, there is interest in the development of robust imaging biomarkers which can reliably and reproducibly provide a predictive index for prognostically significant CAD.

The aim of this thesis is to develop and test methods that improve clinical diagnosis of coronary artery disease in stress echocardiograms using machine learning for 1) image contouring 2) producing features from the contours and 3) predicting coronary artery disease.

A framework for auto contouring of stress echocardiograms was developed, initially using two and four chamber cardiac views for both contrast enhanced and non-contrast echocardiograms, which was later expanded to short-axis (SAX) views. Images were obtained from the prospective multisite EVAREST trial imaging database. The contouring technology was based on a U-net convolutional neural network (CNN) and was trained using contours of the endocardial border created manually, comprised of 5692 and 2182 frames for the contrast and non-contrast application, respectively. The autocontouring achieved a dice coefficient score of 93% for contrast images and 92% for non-contrast images.

Following autocontouring, features were generated from the 2-chamber (AP2) and 4-chamber (AP4) contour output. The feature generation methodology was validated by performing a retrospective study in 146 patients to compare the automated outputs of ejection fraction (EF) and global longitudinal strain (GLS) against equivalent measures generated using standard TomTec analysis software. EF demonstrated a root mean square error (RMSE) of 7.36. GLS demonstrated a RMSE of 2.76.

Following validations of contouring and feature calculation these features and additional novel features were inputted into a machine learning classifier to produce a binary high/low risk prediction for CAD. The training dataset consisted of 633 patients. The final trained model was an ensemble. The area under a Receiver Operator Characteristic Curve (AUROC) of 0.93 was obtained on the training data using a nested cross validation. The model was then validated in an entirely independent dataset and a similar AUROC of 0.93 was obtained. These results demonstrate the utility of artificial intelligence for automated echocardiography analysis and, for the first time, an ability to predict clinical outcome directly from echo images.

List of Publications and Patents

Abstract arising from work in this thesis (manuscript in preparation)

Mumith A, Parker A, Beqiri A, Woodward G, Dockerill C, Woodward W, Alsharqi M, O'Driscoll J, Monteiro C, McCourt A, Sarwar R, **Upton R** & Leeson P. Prediction of CAD from Contrast and Non-Contrast Stress-Echocardiograms using Artificial Intelligence. Late-breaking Science Communication. EuroEcho Imaging conference 2019, Vienna, Austria.

Patents arising from work in this thesis

Three patent families have been filed based on work arising from this DPhil. Details of these families are shown in the table below.

Family	Priority*/Filing Date	Title	Jurisdiction	Application Number
ULT-001	UK: 13/06/2016* PCT: 13/06/2017	Image-based diagnostic systems	UK	GB1610269.1
			PCT	PCT/GB2017/051720
			Australia	AU2017286613
			Brazil	BR 11 2018 075961 9
			China	CN201780041969.5
			European Patent Office	EP17732542.0
			India	IN201817047167
			Japan	JP 2019-517180
			New Zealand	NZ 749228
			South Korea	KR 10-2019-7000713
			Singapore	SG11201811125X
			USA	US16/309,432
ULT-002	UK: 13/12/2017* PCT TBC	Diagnostic modelling method and apparatus	UK	GB1720789.5
			PCT	Pending
ULT-003	UK: 13/12/2017* PCT TBC	Method and apparatus for analysing images	UK	GB1720791.1
			PCT	Pending

Related publications during DPhil studies

Stebbing, R. V, Namburete, A. I. L., **Upton, R**., Leeson, P., & Noble, J. A. (2015). Data-driven shape parameterization for segmentation of the right ventricle from 3D+t echocardiography. *Medical Image Analysis*, *21*(1), 29–39. https://doi.org/10.1016/j.media.2014.12.002

Aye, C., Davis, E., **Upton, R**., Lewandowski, A. J., & Leeson, P. (2014). Assessment of cardiac function from fetal to adult life with myocardial deformation imaging. *Ultrasound in Obstetrics & Gynecology : The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, *43*(6), 605–608. https://doi.org/10.1002/uog.13397

Aye, C. Y. L., Lewandowski, A. J., Lamata, P., **Upton, R**., Davis, E., Ohuma, E. O., … Leeson, P. (2017). Disproportionate cardiac hypertrophy during early postnatal development in infants born preterm. *Pediatric Research*, *82*(1), 36–46. https://doi.org/10.1038/pr.2017.96

Aye, C. Y. L., Lewandowski, A. J., Ohuma, E. O., **Upton, R**., Packham, A., Kenworthy, Y., … Leeson, P. (2018). Two-Dimensional Echocardiography Estimates of Fetal Ventricular Mass throughout Gestation. *Fetal Diagnosis and Therapy*, *44*(1), 18–27. https://doi.org/10.1159/000477964

Omar, H. A., Domingos, J. S., Patra, A., **Upton, R**., Leeson, P., & Noble, J. A. (2018). Quantification of cardiac bull's-eye map based on principal strain analysis for myocardial wall motion assessment in stress echocardiography. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 1195–1198). https://doi.org/10.1109/ISBI.2018.8363785

Aye, C. Y. L., Lewandowski, A. J., Oster, J., **Upton, R**., Davis, E., Kenworthy, Y., … Leeson, P. (2018). Neonatal autonomic function after pregnancy complications and early cardiovascular development. *Pediatric Research*, *84*(1), 85–91. https://doi.org/10.1038/s41390-018-0021-0

Alsharqi, M., **Upton, R**., Mumith, A., & Leeson, P. (2018). Artificial intelligence: a new clinical support tool for stress echocardiography. *Expert Review of Medical Devices*, *15*(8), 513–515. https://doi.org/10.1080/17434440.2018.1497482

Alsharqi, M., Woodward, W. J., Mumith, J. A., Markham, D. C., **Upton, R**., & Leeson, P. (2018). Artificial intelligence and echocardiography. *Echo Research and Practice*, *5*(4), R115–R125. https://doi.org/10.1530/ERP-18-0056

Related abstracts during DPhil studies

Upton, R., Levelt, E., Gamble, J., Clarke, K., Betts, T., Neubauer, S., Leeson, P. (2014). Abstract 12183: Strain Dispersion is an Early Subclinical Manifestation of Diabetic Cardiomyopathy Assessed by 3D Echocardiography. *Circulation*, *130*(suppl_2), A12183–A12183. https://doi.org/10.1161/circ.130.suppl_2.12183

Lewandowski, A., Aye, C., Lamata, P., **Upton, R**., Neubauer, S., & Leeson, P. (2015). B Postnatal Cardiac Remodelling after Preterm Birth: Why Preterm-Born Individuals have a Unique Functional and Structural Phenotype in Young Adulthood? *Heart*, *101*(Suppl 4), A125 LP-A126. https://doi.org/10.1136/heartjnl-2015-308066.231

Upton, R., Levelt, E., Ariga, R., Mahmod, M., Brophy, C., Mangual, J., … Leeson, P. (2015). 123 Myocardial and Fluid Mechanics by Echocardiography Detect Subclinical Changes in Type 2 Diabetes Mellitus. *Heart*, *101*(Suppl 4), A70 LP-A71. https://doi.org/10.1136/heartjnl-2015-308066.123

Related IP during DPhil studies

In addition to the filed patents, additional pieces of intellectual property have been generated related to echocardiography analysis and are summarised in the below table (Fig. 1).

1. **Stress echocardiogram automated contouring for two and four-chamber views (general)**

This invention automatically detects the endocardial walls using a neural network and a contour is then drawn along it with no manual, human input. The invention works for echocardiograms acquired in the two and four chamber standard views both with and without contrast enhancement (a separate neural network is used in each case). This has been piloted on several different ultrasound vendors.

2. **Contrast-enhanced stress echocardiogram automated contouring for two & four-chamber views**

This is able to robustly contour contrast enhanced image data, not previously possible by other autocontouring technologies.

3. **Classifier to detect whether echocardiogram images are contrast-enhanced or not**

This invention first detects whether an image has been enhanced with microbubble contrast and then feeds the images to an optimised auto-contouring framework. This detection is done with a simple binary classifier neural network trained on ~118,000 image frames and is very accurate in its classifications.

4. **Short-axis view auto-contouring model**

An auto-contouring method for images acquired in the short-axis view of the heart. This has been trained using both contrast and non-contrast enhanced data and so generalises for both image types.

5. **Contour point allocation methods (two/four-chamber views)**

The output mask from the auto-contouring neural networks must be converted to Euclidean contour coordinates. This is done via a geometric method for the two and four-chamber auto-contouring. The points are finally temporally smoothed.

6. **Contour point allocation methods (short-axis view)**

This invention comprises a method of distributing the points around the edge mask at regular angles with polar coordinates.

7. **Automated end-diastolic and end-systolic frame detection**

An individual cardiac cycle must be selected from the sequence of image frames/contours produced from the auto-contouring pipeline. This is performed by assessing the areas enclosed within the contours and their relative temporal location in the video sequence

8. **Features used as biomarkers for Coronary Artery Disease prediction model**

Novel imaging features have been developed for the quantification of cardiac disease, some of which are detailed in this thesis.

9. **Clinician performance augmentation using EchoGo**

The machine learning classifier is able to improve the prediction of CAD by including prior knowledge in a Bayesian network.

10. **Automated image quality control**

This invention automates the process of echocardiography quality assurance assessment, previously conducted by manual operators.

Table of Contents

Aknowledgements 2

Dedications 3

Abbreviations 4

Abstract 5

List of Publications and Patents 6

Abstract arising from work in this thesis (manuscript in preparation) 6

Patents arising from work in this thesis 6

Related publications during DPhil studies 7

Related abstracts during DPhil studies 8

Related IP during DPhil studies 9

Table of Contents 12

List of Figures 18

List of Tables 21

1. Introduction 23

1.1 Echocardiography and Artificial Intelligence 24

1.2 What is Artificial Intelligence? 25

1.3 The need for AI in Echocardiography 27

1.4 Future and Potential Applications 29

1.4.1 Image Recognition 30

1.4.2 Classification of Pathological Patterns 30

1.4.3 Automated quantification 31

1.4.4 Quantification of wall motion abnormalities 32

1.5 Steps to building a predictive supervised machine learning model 33

1.5.1 Imputation 33

1.5.2 Normalisation 34
1.5.3 Sampling 34
1.5.4 Feature selection method 36
1.5.5 Classifiers 36
1.6 Coronary Artery Disease 40
1.6.1 Overview 40
1.6.2 Pathophysiology 41
1.6.3 Ischaemic Cascade 43
1.6.4 Non-Invasive Diagnostic Investigations 44
1.7 Stress Echocardiography 46
1.7.1 Overview 46
1.7.2 Current Clinical Standard 47
1.8 Thesis Hypothesis and Objectives 49
2.0 Study Population 51
2.1 Overview 52
2.2 Eligibility Criteria 54
2.3 Recruitment 54
2.4 Informed Consent 55
2.5 Data Collection 56
2.5.1 Medical History 56
2.5.2 Stress Echocardiogram 56
2.5.3 Echocardiogram protocol 57
2.5.4 Image processing 58
2.5.5 Blood Sample Collection & Storage 58
2.5.6 Measurement of Microvesicles 59

2.5.7 Imaging Biomarker Measurement 59
2.5.8 One-Year Follow-Up 60
2.5.9 Outcome Adjudication 60
2.5.10 Discontinuation/Withdrawal of Participants from Study 61
2.5.11 Definition of End of Study 62
2.6 Statistical Analysis 62
2.7 Patient Demographics 63
2.8 Stress Echocardiogram Data 65
2.9 Outcome Data 66
2.10 Stress Echocardiography as a Predictor of Outcome 67
2.11 Discussion 68
3.0 Method Development for Autocontouring Stress Echocardiogram Images 71
3.1 Introduction 72
3.2 Methods 74
3.2.1 Dataset 74
3.2.2 Contrast/Non-Contrast Binary Classifier 74
3.2.3 Autocontouring framework for 4C and 2C 75
3.2.4 Short-Axis Auto-Contouring Model 80
3.2.5 Calculating the cardiac cycle 81
3.2.6 Feature Calculation 82
3.2.7 Statistical Analysis 85
3.3 Results 86
3.3.1 Contrast/Non-contrast classifier 86
3.3.2 Autocontouring 86
3.4 Discussion 88

4.0 Comparison of Auto Ejection Fraction and Global Longitudinal Strain Against the Clinical Standard: A Validation Study 93
4.1 Introduction 94
4.2 Methods 95
4.2.1 Study Population 95
4.2.2 Data Selection 95
4.2.3 Data Handling and Transfer 96
4.2.4 Quality Control Process 97
4.2.5 Calculation of Global Longitudinal Strain and Ejection Fraction 97
4.2.6 Accuracy Assessment 98
4.2.7 Repeatability and Inter/Intra Observer Assessment 99
4.3 Results 101
4.3.1 Demographics 101
4.3.2 Quality assurance exclusions 101
4.3.3 Accuracy Results 103
4.3.4 Precision Results 107
4.4 Discussion 109
5.0 Training a Machine Learning Classifier to Predict Coronary Artery Disease 113
5.1 Introduction 114
5.2 Methodology 116
5.2.1 Study Population 116
5.2.2 Outcomes Definition 116
5.2.3 Auto-contouring of left ventricular and cycle extraction 117
5.2.4 Biomarkers and feature engineering 117
5.2.5 Building the Machine Learning Classifier 130

5.2.6 Validating the Models 132
5.3 Results 134
5.3.1 Patient Demographics 134
5.3.1 Optimised Classifier hyper-parameters 135
5.3.2 Features selected 135
5.3.2 Model performance on training data 139
5.4 Discussion 141
6. Predicting Coronary Artery Disease: A Validation Study 145
6.1 Introduction 146
6.2 Methods 147
6.2.1 Study Database 147
6.2.2 Medical Record Review 150
6.2.3 Assessment of Patient Outcomes 150
6.2.4 Quality Assessment 151
6.2.5 Data Processing by EchoGo Pro 152
6.3 Results 152
6.3.1 Patient Demographics 152
6.3.2 Accuracy of Classifiers 153
6.3.3 Training versus Validation Data 155
6.4 Discussion 157
7. Future Developments 161
7.1 Autocontouring 162
7.2 Feature Generation 163
7.3 Future Validation Studies 164
8. References 166

9. Appendices 183

9.1 Appendix A – EVAREST Clinical Trial Protocol 183

9.2 Appendix B- Rainier Clinical Trial Protocol 215

9.3 Appendix C – Contributions to Work 246

List of Figures

Figure 1 Steps to using AI in Medical Imaging 29

Figure 2 Balancing a data set 35

Figure 3 Schematic of Random Forest 37

Figure 4 Schematic of Support Vector Machine 38

Figure 5 Schematic of Logistic Regression 38

Figure 6 Schematic of MLP 39

Figure 7 Atherosclerosis and plaque formation 42

Figure 8 Ischaemic Cascade, sequence of events 43

Figure 9 Biphasic response during a dobutamine stress echocardiogram 47

Figure 10 Range of sensitivity and specificity 48

Figure 11 Study design and workflow 53

Figure 12 EVAREST recruitment sites and phases 55

Figure 13 Decision tree demonstrating the outcome adjudication process 61

Figure 14 Diastolic 2-chamber image with corresponding contour 75

Figure 15 Diastolic 2-chamber image frame with corresponding mask 76

Figure 16 Modified U-net architecture used for auto-contouring 77

Figure 17 Overall auto-contouring pipeline 80

Figure 18 Process for generating contour points from SAX auto-contouring 81

Figure 19 An example disc used when calculating volume 84

Figure 20 Example shown is p^D (red) and p^S (black) 85

Figure 21 Series of image frames for a non-contrast dataset 87

Figure 22 Results of the contours on contrast and non-contrast images 88

Figure 23 A mask used for training in which the lower edge of the mask has been gradually down-weighted 90

Figure 24 Taking the same dataset as in Figure 21 and applying the contrast model yields the incorrect result 91

Figure 25 Short-axis chamber walls are not fully present 92

Figure 26 Diagrammatic overview of operators and datasets 102

Figure 27 Q-Q plots of GLS and EF measurements calculated on EchoGo 104

Figure 28 Q-Q plots of GLS and EF measurements calculated on TomTec Arena 104

Figure 29 Bland-Altman analysis of EchoGo Accuracy 105

Figure 30 Bland-Altman analysis of all data 106

Figure 31 Bland-Altman analysis of data obtained by GE 107

Figure 32 Bland-Altman analysis of data obtained by Philips 108

Figure 33 Bland-Altman analysis of EchoGo Repeatability 109

Figure 34 Decision tree demonstrating the outcome adjudication process 118

Figure 35 Example shown is $\boxed{A^D}$ 120

Figure 36 Example shown is $\boxed{\{c_7^D, c_8^D, c_9^D, c_{10}^D, c_{11}^D, c_{11}^S, c_{10}^S, c_9^S, c_8^S, c_7^S\}}$ 121

Figure 37 Example shown is for diastole 122

Figure 38 Example shown is for diastole 123

Figure 39 Example shown is for the set of points $\{c_7^D, c_{11}^D, c_{15}^D\}$ 125

Figure 40 Example shown is for the set of points $\{\boxed{c_{11}^D, c_{12}^D, c_{13}^D, c_{14}^D, c_{15}^D}\}$. 126

Figure 41 Example for end-systole 127

Figure 42 Example shown is eccentricity 130

Figure 43 Leesonarity examples 131

Figure 44 Nested cross fold validation 135

Figure 45 Box, swarm, violin plots. The feature values are normalised using Z-score 139

Figure 46 Scatter plot showing 3 features 140

Figure 47 Box and Violin plot of resting features 140

Figure 48 ROC curves of different classifiers on training data 142

Figure 49 ROC curve of the ensemble model 142

Figure 50 ROC curve with model trained on just regional solidity 143

Figure 51 ROC curves for each classifier on validation dataset 155

Figure 52 ROC curve for EchoGo Pro 155

List of Tables

Table 1 CAD Disease Deaths ranked by country burden 40

Table 2 Patient demographics 63

Table 3 Patient risk factors at time of stress echocardiogram 64

Table 4 Breakdown of medications 65

Table 5 Stress echocardiography sensitivity and specificity 68

Table 6 DC results for contrast 2 and 4 chamber 87

Table 7 Demographics of patient population 102

Table 8 Summary of the rejected scans 103

Table 9 Summary of the EchoGo scans that passed quality assurance 103

Table 10 GLS and EF measurements comparison 105

Table 11 Summary of Accuracy Results 106

Table 12 Accuracy results of data obtained by GE 106

Table 13 Accuracy results of data obtained by Philips 107

Table 14 EchoGo repeatability results 108

Table 15 Patient Demographics 135

Table 16 Sex, Previous History, Outcome, SE type 136

Table 17 Number of times features were selected 137

Table 18 Table shows how many times each feature was selected 138

Table 19 Model results from nested CV on training data 141

Table 20 Patient Demographics 154

Table 21 Sex, Previous History, Outcome, SE type 154

Table 22 Results of each classifier on validation set 154

Table 23 Age Difference 156

Table 24 Sex Difference 156

Table 25 Height Difference 157

Table 26 Weight Difference 157

Table 27 BMI Difference 157

Table 28 Previous CAD 158

Table 29 Previous MI 158

Table 30 Previous CABG 158

Table 31 Dobutamine Stress Echo 158

1. Introduction

1.1 Echocardiography and Artificial Intelligence

Echocardiography is one of the most widely used imaging modalities for cardiovascular disease and plays a critical role in the diagnosis and management of patients (1,2). Echocardiography allows instant detection of cardiac abnormalities due to its ability for real-time imaging unlike most other imaging modalities. (3,4). For clinical diagnosis and to help guide the correct treatment pathway, visualisation and accurate quantification of cardiac structure and function is essential. However, 2 dimensional echocardiography is limited due to the subjectivity of interpretation and inter-operator variability of acquisition protocols, image quality and measurements despite the abundance of guidelines and is therefore prone to error (5,6). This variability is a well-known and long-standing issue within echocardiography. Nevertheless, echocardiography usage is still growing year on year and so improving usability and accuracy of echocardiography are important goals that artificial intelligence (AI) could potentially help to address (7).

Despite the current surge of interest in the research and use of AI in medical imaging, we have been able to practice AI as early as the 1950's (8). One such AI technique known as machine learning (ML) can be applied to echocardiograms. This method is particularly suitable because it can make use of a wide variety of patterns present in images such as edges, contrast changes and even individual pixel variation within an image. These image-related features can be associated through machine learning with additional patient specific information derived from clinical metadata (9). Models can be trained to identify important features within an image that can distinguish between different images, identify regions of interest within an image, or associate features with different diseases or disease phenotypes (10).

The ability to move beyond just the information contained within the images themselves and combine metadata, or clinical decision making, with image features in machine learning models has the potential to improve the accuracy, inter/intra operator variability and provide critical insights that might be missed by the interpreting physician (11). Therefore, machine learning models show potential as clinical tools for the quick, accurate and precise analysis of cardiac structure and function. I will now consider what AI is, reasons why it should be applied to echocardiography and give examples of the most recent applications in echocardiography.

1.2 What is Artificial Intelligence?

AI is defined as the ability of computer systems to perform tasks that would usually require human levels of intelligence (12). Machine learning is a subfield of AI which enables teaching of a computer to analyse large amounts of data in a quick, precise and efficient process by utilising sophisticated statistical and computer algorithms (13). These algorithms have the ability to recognise patterns from existing data and learn which patterns have the highest predictive accuracy. By having this ability, trained machine learning algorithms have the ability to make predictions on new datasets (14).

Machine learning can be categorised into three groupings: unsupervised, supervised, and reinforcement learning (15). Supervised machine learning is where the machine is ‘taught’ to classify datapoints by providing it with a training dataset of labelled data to create a model (16). During this process, the algorithm learns

patterns within the dataset by analysing independent variables and compares these with known outcomes as the dependent variable. Once the machine has been trained it outputs a fixed model which is then tested using an unseen validation dataset. The validation dataset provides a representation of the accuracy, sensitivity and specificity of the model. (10,17). Examples of supervised learning methods include support vector machines, random forests (RF) and convolutional neural networks (CNN) (18). Unsupervised learning techniques in contrast use unlabelled data to group the input data into multiple clusters based on patterns and similarities between the input data (17,19). Reinforcement learning is based on interactions with an environment. The learner, also called the agent, attempts to discover the most successful actions to achieve maximum reward by learning from trial and error. This leads to an agent suitable for dynamic environments (20,21). These methods of machine learning are not mutually exclusive as for example, Deep learning, a type of neural network capable of handling large complex datasets can be supervised, unsupervised or semi-supervised (10). Furthermore, these subfields within machine learning have been integrated to create more powerful classification techniques for example deep learning can be used in combination with supervised learning and/or reinforcement learning (22,23).

Model training is an essential part of all types of machine learning. Training a model entails feeding a machine learning classifier to learn a variety of features within the data provided and it uses these to learn how to generate output predictions (24). Using echocardiography as an example, a machine learning model could analyse a number of features such as left ventricular ejection fraction and left ventricular wall thickness to predict whether a patient has a certain disease. However, it is important

to note that the inclusion of irrelevant features in small datasets can lead to over-fitting the model, which renders it less accurate when presented with new datasets (25). An example of an echocardiographic feature that could lead to overfitting is regional strain as studies have shown this to be not reproducible in both inter and intra operator variability (26). This highlights the importance of having a training dataset that is representative of the population you want to apply the model to. Information gain analysis is used to select the optimal features to include in the model to ensure that it has the highest predictive accuracy, yet can still be applied to other populations (17,27). However, this does not necessarily de-risk overfitting the model. Once a trained model is finalised, it can be tested on validation datasets to determine the predictive accuracy before being applied elsewhere.

Machine learning has been applied to many modalities in the medical imaging field, for example calcium scoring in CT coronary angiograms, detection of diabetic retinopathy in retinal photography, and skin cancer classification in dermatology (28–30).

1.3 The need for AI in Echocardiography

Echocardiography is an essential diagnostic tool for a wide range of cardiovascular disease. As a result numerous guidelines have been developed to ensure precise quantification and clinical interpretation (31) but the final diagnosis is reliant on the operator having the training, experience and knowledge to acquire good quality images and adhere to these guidelines. However, it could be feasible to overcome this limitation by using machine learning models. One such example is where

guidelines recommend quantitative measures of chambers and valves during assessment to aid clinical diagnosis of disease (5,32). However, in busy healthcare settings such as acute chest pain, quantitative analysis may not be ideal because of the time required for manual tracing of contours (33). Therefore, visual estimation of cardiac structure and function remains the standard in many areas of clinical practice; although this requires significant experience and training in echocardiography (34,35). Application of machine learning models to either highlight need for a specific type of quantification or provide fully automated measurements rapidly could therefore overcome this problem and improve diagnosis (36).

In busy clinics, there is also usually a time-dependent window in which to provide a comprehensive analysis of the results of the scan and, therefore, not enough time to analyse complex datasets (37). During a standard transthoracic echocardiogram, a large volume of data is captured most of which could be diagnostically useful and with the advent of multi-dimensional imaging modalities, such as speckle tacking and three-dimensional echocardiography, the volume of the data acquired is expected to increase on the current clinical standard (38,39). Most of the data obtained however remains under-utilised (11,40) but machine learning algorithms provide the capability to simultaneously interpret multiple datasets extracted from echocardiograms in a rapid and automated manner (17). Furthermore, machine learning models can link available clinical meta data from electronic health records with image data, therefore providing clinicians with more information, and empowering them to make more informed decisions about the care of their patients(19).

One of the greatest advantages of using machine learning within the diagnostic process is that models can also combine data to act as predictive tools with high levels of accuracy and precision (41). Following training, the machine learning model should be able to determine different cardiac structural and functional patterns that, if subtle, could be missed during the diagnosis by the clinician. As mentioned previously, these data are predicted by comparing features from new data to a model fitted on features extracted from the training data (42–44).

1.4 Future and Potential Applications

Applications of machine learning to cardiovascular imaging are at a relatively early stage, however, several machine learning classifiers have already been developed to aid diagnosis. These methods cover classification of disease, image recognition and automated quantification (15). Figure one summarises the steps and applications of AI in medical imaging.

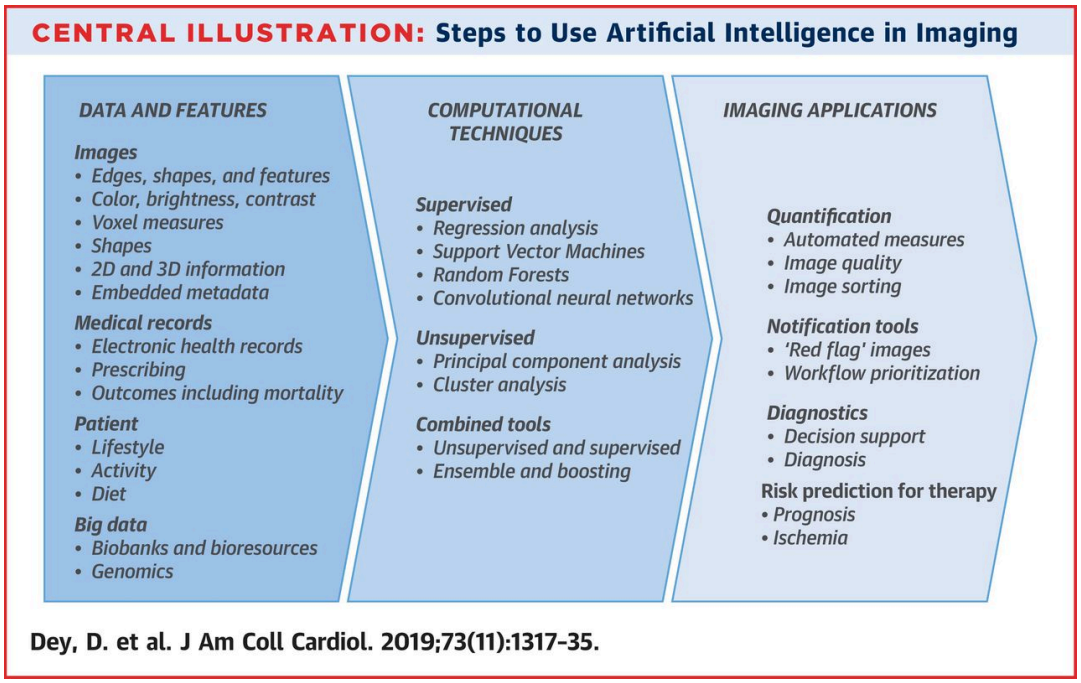


Figure 1: Steps to using AI in Medical Imaging (45).

1.4.1 Image Recognition

The first step in the echocardiogram pipeline to ensure accurate assessment of echocardiograms is to correctly identify views, videos and modality such as continuous and pulsed wave Doppler traces (46). This is an ideal problem for machine learning classification and has proved relatively straightforward. A convolutional neural network, has been trained to detect specific features in each echocardiogram view, regardless of resolution (47). In the trial, a large range of randomly selected retrospective echocardiograms, including normal variants, as well as a range of disease and image quality were used. The deep learning model was able to correctly classify 15 major echocardiography views with an accuracy of 97.8% (48). Another study used machine learning to identify apical four, two, and three chamber views. Despite the views being very similar in terms of recognisable features, the model was able to correctly recognise each view with an accuracy of 95% (49).

1.4.2 Classification of Pathological Patterns

Different pathological and physiological conditions can share similar characteristics that prove difficult for clinicians to differentiate without detailed clinical experience. For example, left ventricular hypertrophy is often seen in athletes, but is also prevalent in diseases such as hypertrophic cardiomyopathy (50). Given the risk of sudden cardiac death in hypertrophic cardiomyopathy accurate differentiation is important. Narula et al. developed an ensemble machine learning model consisting of multiple classifiers including support vector machines, random forests, and

convolutional neural networks to differentiate between hypertrophic cardiomyopathy and athletes, with a sensitivity of 96% (11). Furthermore, machine learning classifiers have also shown ability to differentiate between constrictive pericarditis and restrictive cardiomyopathy which also have similar phenotypes. These pathologies are so similar that there is no single measurement in echocardiography that can differentiate between them (51). Therefore, the diagnosis can be difficult even for the most experienced clinicians. However, a machine learning model was able differentiate between these diseases, with an accuracy of 90%, using multiple echocardiography measurements (37,40).

1.4.3 Automated quantification

Diagnosis in echocardiography relies heavily on the use of quantitative measures. Image processing techniques in combination with machine learning classifiers have shown promise for efficient identification of cardiac structures and quantification of related parameters. Analysis of left ventricular function was one of the first applications of machine learning to reduce error and operator variability (52–55). Methods have however improved so that, recently, Knackstedt *et al.* demonstrated that left ventricular ejection fraction and longitudinal strain could be analysed in less than 10 seconds using machine learning models (34). Within three-dimensional echocardiography, machine learning models such as random forests have been used to identify borders and provide an accurate identification of left and right ventricular cavities so that derived left and right ventricular volumes are equivalent to those measured by cardiac magnetic resonance which is the gold standard (38,56,57). Furthermore, machine learning classifiers have been shown to assist the assessment of valve disease, for example, mitral valve disease (31,58).

Automated assessments of three-dimensional transoesophageal echocardiograms of the mitral valve provided more reproducible measurements of the mitral valve annulus size and morphology than clinical interpretation (59,60).

1.4.4 Quantification of wall motion abnormalities

One of the most important uses of echocardiography is in the diagnosis and care of heart failure related to coronary artery disease or other cardiac pathology (61). Diagnosis of systolic heart failure relies on detection of wall motion abnormalities. Quantitative assessment of wall motion abnormalities is also critical in stress echocardiography to identify patients with prognostically significant coronary artery disease (62). Echocardiography allows real-time visualisation of the contraction of the myocardium during stress and by comparing left ventricular wall motion between resting images and intermediate, peak or post stress images, it is possible to diagnose functionally significant coronary ischaemia (63). Usually, contractility is assessed visually by an experienced clinician and a meta-analysis of stress echocardiography studies demonstrated a wide variation in sensitivities and specificities for dobutamine stress echocardiography (64). Sensitivity ranged from 33% to 98%, whilst the specificity ranged from 38% to 97% resulting in average sensitivity and specificity for dobutamine stress echocardiography of 81 and 82%, respectively. Whilst this is comparable to other functional assessments of coronary artery disease, it still means that approximately one in every five patients could potentially be misdiagnosed (64). Therefore, to increase the accuracy of stress echocardiography diagnosis, machine learning classifiers have been applied to identify inducible wall motion abnormalities (65–68). In one study, Omar *et al.* used imaging derived models of 3D motion at rest and stress within support vector

machines, random forests and a convolutional neural network. The results showed that the convolutional neural network provided the most sensitive model, with a sensitivity of 81.1% in a training dataset compared with expert operator interpretation (67). In another study, an unsupervised machine learning model was used to detect features for linear discrimination, which could differentiate between patients with obstructive coronary disease and normal responses through use of a new coronary artery disease risk index (66). One of the issues with the majority of studies to date is that they have been on relatively small datasets, without adequate testing and prospective validation datasets or have only compared against expert readers rather than outcome which presents a problem as expert readers are only 81% sensitive and 82% specific (64). Nevertheless, they show promise that machine learning models provide potential to support clinical decision-making for stress echocardiography; one of the most commonly used functional imaging tests for coronary heart disease.

1.5 Steps to building a predictive supervised machine learning model

When building a predictive machine learning model the following 5 steps should be considered so that the model is as accurate and robust as possible and to ensure overfitting is avoided (69).

1.5.1 Imputation

The first step is preparing or pre-processing the data. A fundamental part of this is to handle missing values in the dataset of features, known as imputation, as most supervised learning techniques are not capable of dealing with an incomplete dataset (70). The appropriate imputation method depends on the type of data. For

example, a simple parametric method such as mean feature values can be used if the features are normally distributed. In the case of the present dataset, this does not always hold true, therefore a non-parametric method 'K-nearest neighbours' is adopted. This has the advantage that it does not make any prior assumptions about the dataset (71).

1.5.2 Normalisation

While ML algorithms such as random forest does not require normalisation of feature values as it is a rule-based approach, most others such as support vector machine and logistic regression find it harder to correctly classify data without normalisation, due to varying range of values (72). Methods used for ML applications include z-score and min-max scaling (73). In this thesis, the range of feature values is sufficiently similar that it did not require normalisation.

1.5.3 Sampling

For binary classification, a classifier works well when the dataset is balanced, so that there are sufficient examples of both classes that allow patterns in the data to be identified that differentiate the two classes (74). In medical diagnosis the number of examples for each class tends to be skewed in favour of the negative class (75). To counteract this problem over-sampling methods is used, where the minority class is over-sampled to create additional instances in the positive class. One such method is the Synthetic Minority Over-sampling Technique (SMOTE) (76), which uses the K-nearest neighbours technique to create 'synthetic' examples of the minority class to balance out the data. This sampling technique allows better generalisation of the patterns in the data, than simply over-sampling with replacement which does not add

any additional information to the classifier (76). While SMOTE is effective at addressing the imbalance of the classes, the 'synthetic' examples created can be in regions where the differentiation of the classes are blurred, and so this can make it more difficult for a classifier to learn to correctly classify the classes, as shown in Figure 2. A suggested solution to this problem is a combination of over- and under-sampling methods, such as SMOTE + Tomek links or SMOTE + ENN (77). Both Tomek link and ENN (Edited Nearest Neighbours) are under-sampling techniques, that under sample both the minority and majority class, when it works in combination with an over-sampling method. The effect of this can be seen when comparing Figure 2 (c) and (d). This approach clearly allows better distinction between the classes (78). In this thesis, SMOTE+ENN is used as ENN tends to remove more examples than Tomek links, allowing for better separation between classes.

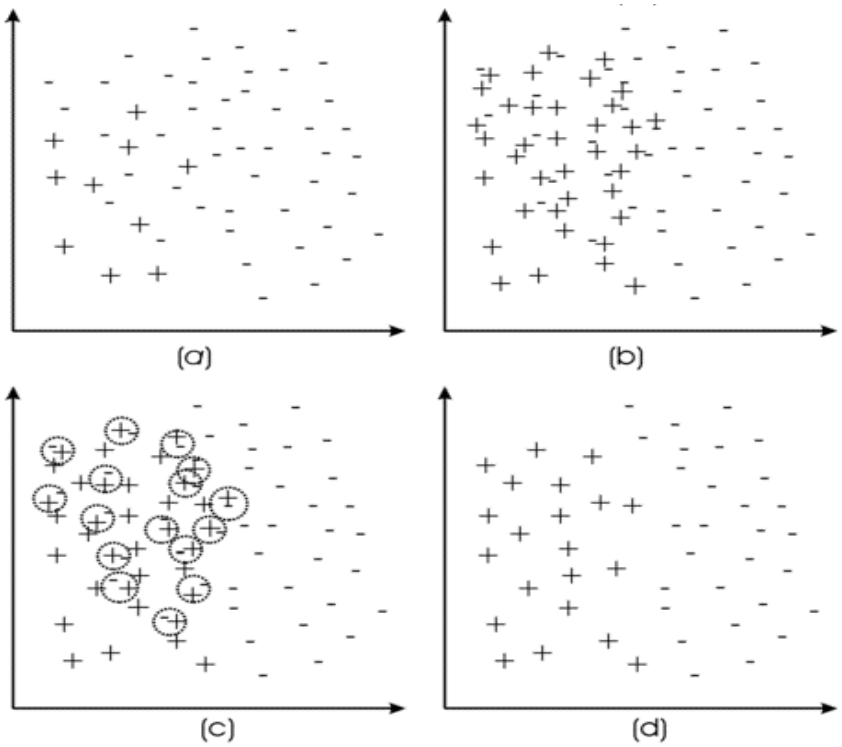


Figure 2: Diagrammatic representation of over and under sampling techniques to balance a data set: (a) original data set; (b) over- sampled data set; (c) over and under sampled dataset; and (d) borderline and noise examples removal from over and under sampled dataset (77).

1.5.4 Feature selection method

While features described in this thesis are believed to be clinically relevant for the prediction of CAD, based on the literature and clinical knowledge, it is necessary to empirically test this hypothesis. It is common practice to start with many features, and then apply a feature selection method to reduce the feature set to the most relevant (79). In this case, the feature selection method chosen is a multivariate approach called the Boruta method (80), which identifies all relevant features. This has the advantage over univariate methods, which is only useful if all features are independent, and employs arbitrary cut off points when selecting the number of features (80). This method is an iterative method, which determines if the features perform better than pure chance, by comparing the feature's importance to a shadow randomised feature set. If it does, then the feature is considered a hit. If a feature has not been a hit for a certain number of iterations it is considered unimportant and removed from the list. The Boruta method selects a feature as important if it is a hit and the calculated p-value for that feature is less than a specified value (default in the Python implementation is 0.05) (81). In the case of predicting CAD, we found the Boruta method dentified the most relevant features in the mid and apical regions of the LV, these are highlighted further in chapter 5.

1.5.5 Classifiers

Decision trees create a set of rules about the cut-off point of feature values, moving down the tree to determine what conditions are required for classification. In a RF instead of one tree, many trees are generated (82). Each tree is created using a sub-sample of the training dataset and votes for a class when making a prediction. Majority voting of the trees is then adopted for the overall class prediction of the random forest, as shown in Figure 3. This technique is perhaps the most widely used technique in healthcare, as it scales up well, handling potentially thousands of features, and runs efficiently on large datasets (83). Also, it requires less data preparation than other classifiers, as it does not require scaling of the features. On the other hand, RF is a nonparametric model, therefore the degree of freedom of the model to learn on the dataset is not limited beforehand, and therefore RF is prone to overfitting the training dataset if it is not regularised (84).

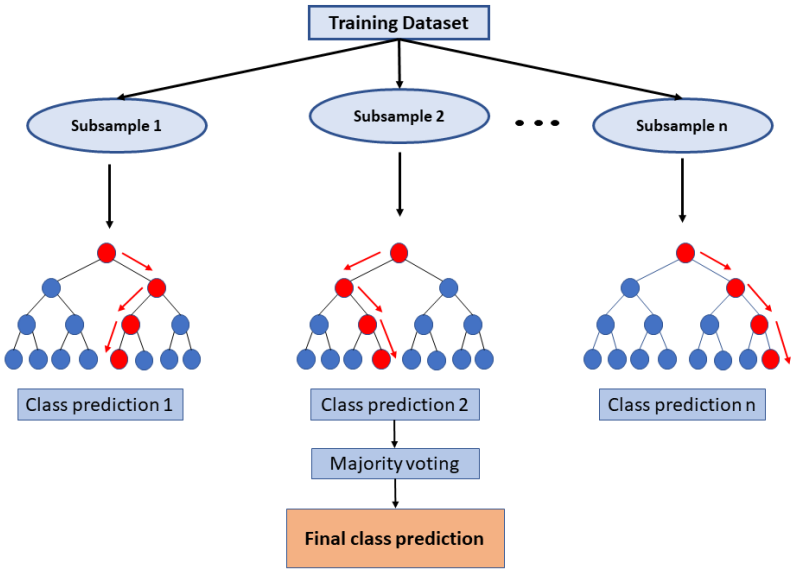


Figure 3: Schematic of Random Forest

Support vector machines (SVM) works by creating line(s) known as hyperplanes that divide the dataset into distinct classes, attempting to optimise the margin between the classes, as shown in Figure 4 (85). With SVM it is possible to use diverse types of kernels to create the hyperplanes, such as linear (shown in Figure 4), polynomial, and radial basis function. It is for this reason that SVM can be applied to both linear and nonlinear problems (86). SVM is also very good at fitting to the data so if the SVM is not trained correctly, overfitting can occur, and the data can be divided unnecessarily into smaller divisions (87).

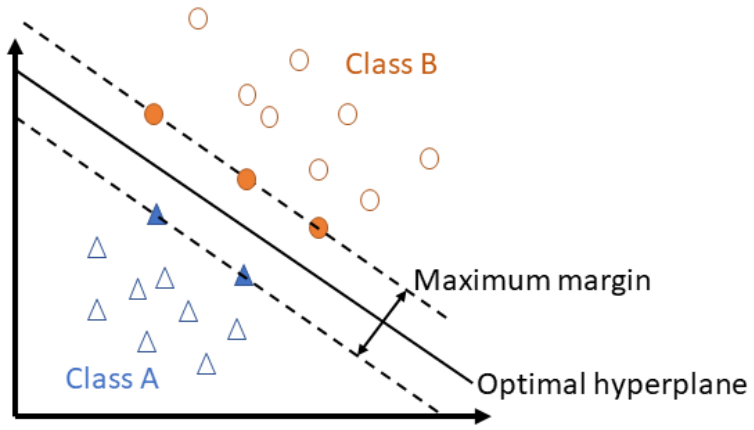


Figure 4: Schematic of Support Vector Machine separating data by linear hyperplane.

Linear logistic regression (Figure 5) estimates the probability that an instance belongs to a class. If the probability that it belongs to a class is greater than 50%, then the model predicts that it belongs to the positive class, otherwise it belongs to the negative class (88). This method works very similar to linear regression, but instead computes logistic of the weighted sum of features (88). This method allows easy update of the model when new data is available and provides an easy

probabilistic interpretation of the results, although this method is not flexible enough to capture very complex relationships (89).

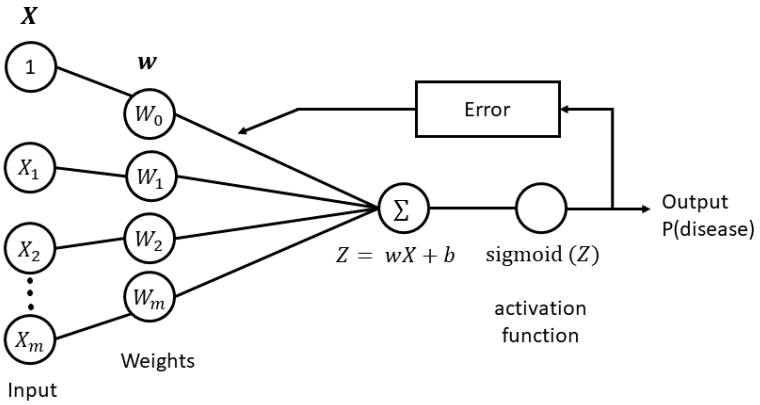


Figure 5: Schematic of Logistic Regression

Multi-layer perceptron is a type of artificial neural network, where there is an input, hidden, and output layers. The signal from the input moves through the hidden layers, where linear and nonlinear transformations occur, and where signals are combined as shown in Figure 6 (90). To map the input to the output, backpropagation of the error occurs from the output to the input layer, where an attempt is made to minimise the error by changing the weights of the nodes in the hidden layers (90).

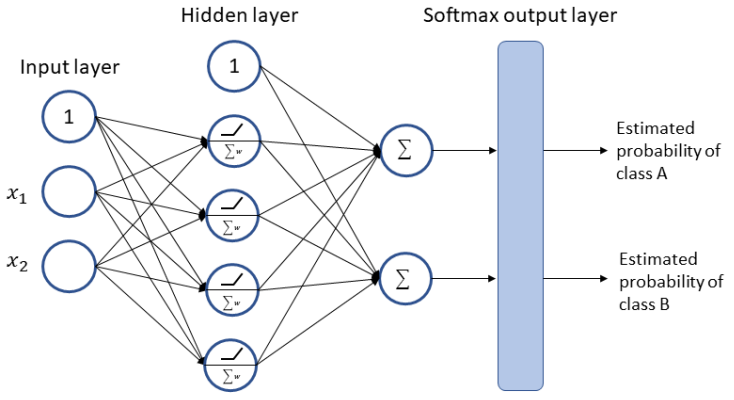


Figure 6: Schematic of MLP using a Relu as the activation function

The four classifiers described above are commonly used for medical diagnosis (91). These classifiers are considered in the current work because of the different algorithms adopted, each having its own set of advantages/disadvantages. This is ideal for ensemble learning, as they are more likely to make different errors, thus improving the performance when a prediction is made with the combined decision from each classifier (84).

1.6 Coronary Artery Disease

1.6.1 Overview

CAD is the most common cause of death worldwide (Table 1) and affects an estimated 1.9 million people in England (92). As such, the early and accurate detection of CAD is key to reducing morbidity and mortality. Whilst there are a wide-range of tests available to examine for CAD, stress echocardiography is a firmly established non-invasive 2D and 3D functional imaging technique used to examine left ventricular regional wall motion during an increase in myocardial oxygen demand, brought about by the use of pharmacological or exercise stress (93,94).

Table. Burden of CAD Deaths From Most Recent Year of Available Data

High-Income Countries		Upper-Middle–Income Countries		Lower-Middle–Income Countries		Low-Income Countries	
Country	**Deaths**	**Country**	**Deaths**	**Country**	**Deaths**	**Country**	**Deaths**
United States	366 801	Russian Federation	529 825*	Ukraine	278 714	Syrian Arab Republic	14 241*
Germany	128 230	Brazil	111 849	Philippines	64 729*		
Italy	73 172	Mexico	85 961	Republic of Moldova	14 280		
Japan	71 673	Turkey	64 691	Kyrgyzstan	11 415		
United Kingdom	69 783	Romania	52 709	Sri Lanka	6641*		
Poland	39 359	Belarus	48 977*	Guatemala	5579		
Canada	33 884*	Colombia	36 197	Nicaragua	3120		
Spain	33 769	Venezuela	24 086*	El Salvador	2260*		
Hungary	33 007	Argentina	21 918	Tunisia	1319*		

France	32 727*	Cuba	17 022	Morocco	807*		
Czech Republic	26 659	Kazakhstan	14 644*	Honduras	286*		
Australia	19 777	Bulgaria	12 652*				
Lithuania	15 566	Thailand	12 163*				
Austria	14 905	Serbia	9567				
Slovakia	13 338*	Turkmenistan	7232				
Greece	12 296	Ecuador	5064				
Sweden	12 121	Dominican Republic	5043				
Croatia	11 509	Georgia	4880				
Netherlands	8982	Peru	3365				
Chile	8391	Costa Rica	2969*				
Latvia	8034	Paraguay	2278*				
Belgium	7845	Jordan	2010*				
Portugal	7456*	Panama	1491				
Switzerland	7373	Mauritius	1199				
New Zealand	5030*	Jamaica	1085*				
Ireland	4283*	Fiji	754*				
Norway	4214	Guyana	605*				
Hong Kong SAR	4123	Suriname	248*				
Israel	3915	Belize	128				
Denmark	3779	Maldives	113*				
Estonia	3404*	Saint Vincent and Grenadines	91				
Puerto Rico	3215	Saint Lucia	82*				
Singapore	3099	Grenada	66				
Uruguay	2373	Dominica	37				
Saudi Arabia	2368*						
Trinidad and Tobago	1383*						
Kuwait	1313*						
Malta	732						
Cyprus	669						
Oman	354*						
Iceland	350*						
Luxembourg	308						
Barbados	202*						
Bahamas	197*						
Brunei Darussalam	78						
Bermuda	70						
Virgin Islands (United States)	65						
Antigua and Barbuda	44						
Seychelles	31						
Saint Kitts and Nevis	30						
Andorra	17						
Turks and Caicos Islands	16*						
Cayman Islands	11*						
British Virgin Islands	9*						
San Marino	8						

Table 1: CAD Deaths ranked by country burden. Data from 2015 except where marked * where data are 2014 or earlier (95).

1.6.2 Pathophysiology

Ischaemic heart disease is usually caused by disease of the coronary artery due to an atheromatous narrowing and subsequent occlusion of the vessel, (96). Atheromas can be present from young adulthood onwards (97). A mature plaque consists of two elements, each from a different type of cell. The lipid core is mainly constructed from necrotic foam cells (monocyte derived macrophages), which move into the intima and ingest lipids (98). The other being the connective tissue matrix which is derived from smooth muscle cells, which transition from the media into the intima, where they proliferate and alter their phenotype to construct a fibrous capsule around the lipid core (99) (Figure 7).

Usually when a plaque produces a >50-70% stenosis, blood flow is reduced through the coronary artery during stress or exertion which can cause myocardial ischaemia and angina (96). Acute coronary disease such as myocardial infarction usually occurs when thrombus formation follows disruption of a plaque. In acute myocardial infarction, occlusion is more complete than in unstable angina, where arterial occlusion is usually subtotal (100). Downstream embolism of thrombus may also produce microinfarcts (101).

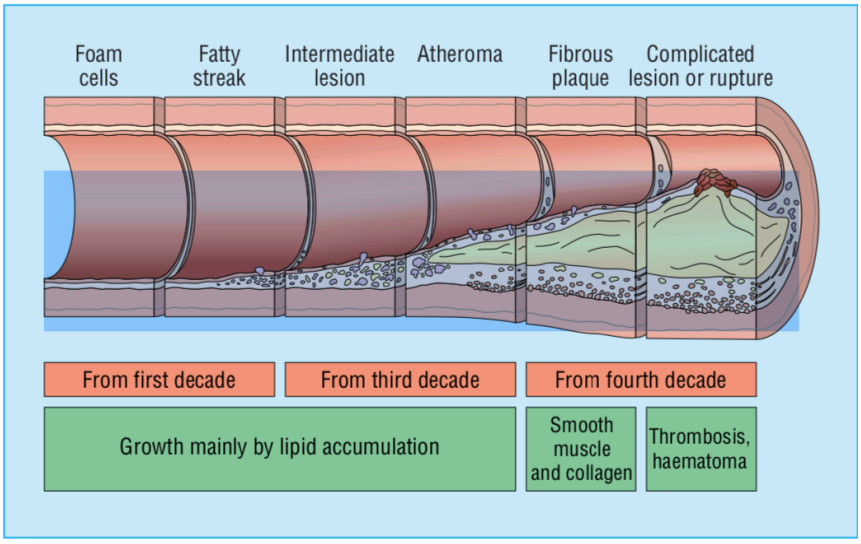


Figure 7: Atherosclerosis and plaque formation (96)

1.6.3 Ischaemic Cascade

Myocardial ischaemia results in a 'cascade' of events in which some biomarkers are ranked in an ordered sequence (Figure 8) (102). Flow heterogeneity, between subendocardial and subepicardial perfusion, is the first indication of ischaemia, metabolic changes arising from mismatch of oxygen demand and supply causing a decrease in mitochondrial oxidative metabolism causing an energy deficient state in the myocardium are next, followed by changes in regional cardiac function and finally changes on ECG and physical symptoms (102). This cascade is clinically evident through sequential changes in clinical markers of ischaemia. The different approaches to assess ischaemia in the clinical setting have varying degrees of sensitivity (102).

Coronary flow reserve reduction is the most common pathophysiological mechanism. Regardless of the stress used, ischaemia tends to propagate centrifugally with respect to the left ventricular cavity: (103,104) it primarily involves the subendocardial layer, compared to the subepicardial layer which is affected at a later stage (103). Interestingly, extravascular pressure is higher in the subendocardial than in the subepicardial layer which causes an increased metabolic demand and a higher resistance to coronary flow. If coronary artery disease is absent, coronary flow reserve can be decreased in the presence of microvascular disease (e.g. in syndrome X) or left ventricular hypertrophy (LVH) (e.g. arterial hypertension). In this situation, angina with ST depression can present with regional perfusion changes, usually in the absence of any regional wall motion abnormalities during a stress

echocardiogram. It has been proven, however that wall motion abnormalities are more specific than coronary flow reserve and/or perfusion changes for the diagnosis of coronary artery disease (105–109).

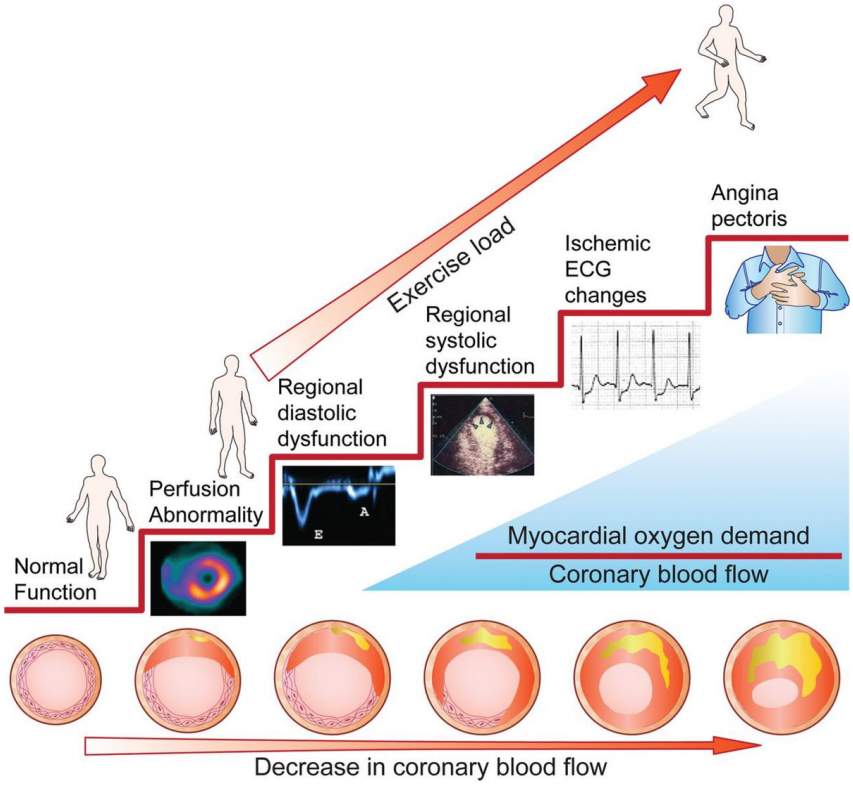


Figure 8: Ischaemic Cascade, sequence of events (110)

1.6.4 Non-Invasive Diagnostic Investigations

There are numerous diagnostic investigations that play a role in the detection of myocardial ischaemia. Recent advances in Computed Tomography (CT) technology have placed coronary CT angiography (CCTA) at the forefront of the non-invasive diagnosis of CAD (17,111,112). The diagnostic accuracy of CCTA has been systematically reviewed with a sensitivity (98%) and specificity (89%) and negative

predictive value of almost 100% (113); therefore, according to the appropriate use criteria, CCTA is now recommended to be the first line modality for the exclusion of CAD in patients of low and intermediate risk (113,114).

However, despite the accuracy of CCTA for coronary anatomy assessment and coronary plaque evaluation, it is limited in detecting the haemodynamic significance of coronary artery disease. Many studies have highlighted that CCTA was not accurate in predicting myocardial ischaemia, and that revascularisation would be better guided by functional imaging, as determined by the FAME (Fractional Flow Reserve versus Angiography for Multivessel Evaluation) and COURAGE (Clinical Outcomes Utilizing Revascularisation and Aggressive Drug Evaluation) studies (115,116). These studies suggest that non-invasive functional imaging techniques such as stress echocardiography, stress cardiac magnetic resonance (CMR), positron emission tomography (PET) and single photon emission CT (SPECT) were shown to be more accurate in detecting perfusion defects (117).

Some studies have demonstrated dobutamine stress CMR to be superior to stress echocardiography in terms of specificity (87.5% vs 72.9%), negative predictive value (80.8% vs 67.3%) and overall diagnostic accuracy (80.4% vs 72%) (118). However, this is not the case when comparing to figures from the last meta-analysis of stress echocardiography studies (64). Adenosine stress perfusion CMR was reported to have a sensitivity of 88.7% and a specificity of 83.5% for the detection of CAD when compared to invasive coronary angiography (ICA), as reported in the CE-MARC trial (119), and higher sensitivity compared to SPECT (75% vs 59%; $p = 0.03$), as reported in the MR-IMPACT II trial (120). Nuclear imaging such as PET and SPECT are also robust modalities for the diagnosis of CAD. PET-MPI has demonstrated

similar accuracies for CAD detection compared to SPECT- MPI and Stress Echocardiography, as recently shown in a large meta-analysis (121).

1.7 Stress Echocardiography

1.7.1 Overview

Stress echocardiography has many advantages over other imaging techniques in the assessment of ischaemic heart disease, such as being relatively inexpensive (in comparison to invasive coronary angiography, for example), portable and free of ionising radiation, whilst maintaining similar accuracy levels as myocardial perfusion scintigraphy (122) and stress cardiac magnetic resonance imaging (123). However, as with all imaging modalities, the accuracy of stress echocardiography is dependent on operator experience mainly due to the inherent subjectivity associated with interpreting myocardial wall motion (7). It has been demonstrated that accurate diagnosis requires a high level of training to achieve accurate interpretation (124). This forms the basis of the recommendation of the British Society of Echocardiography that operators should perform at least 120 stress echocardiograms per year in order to maintain competence (125). Figure 9 demonstrates a stress echocardiogram quad screen view of the mid short axis with positive findings for coronary artery disease.

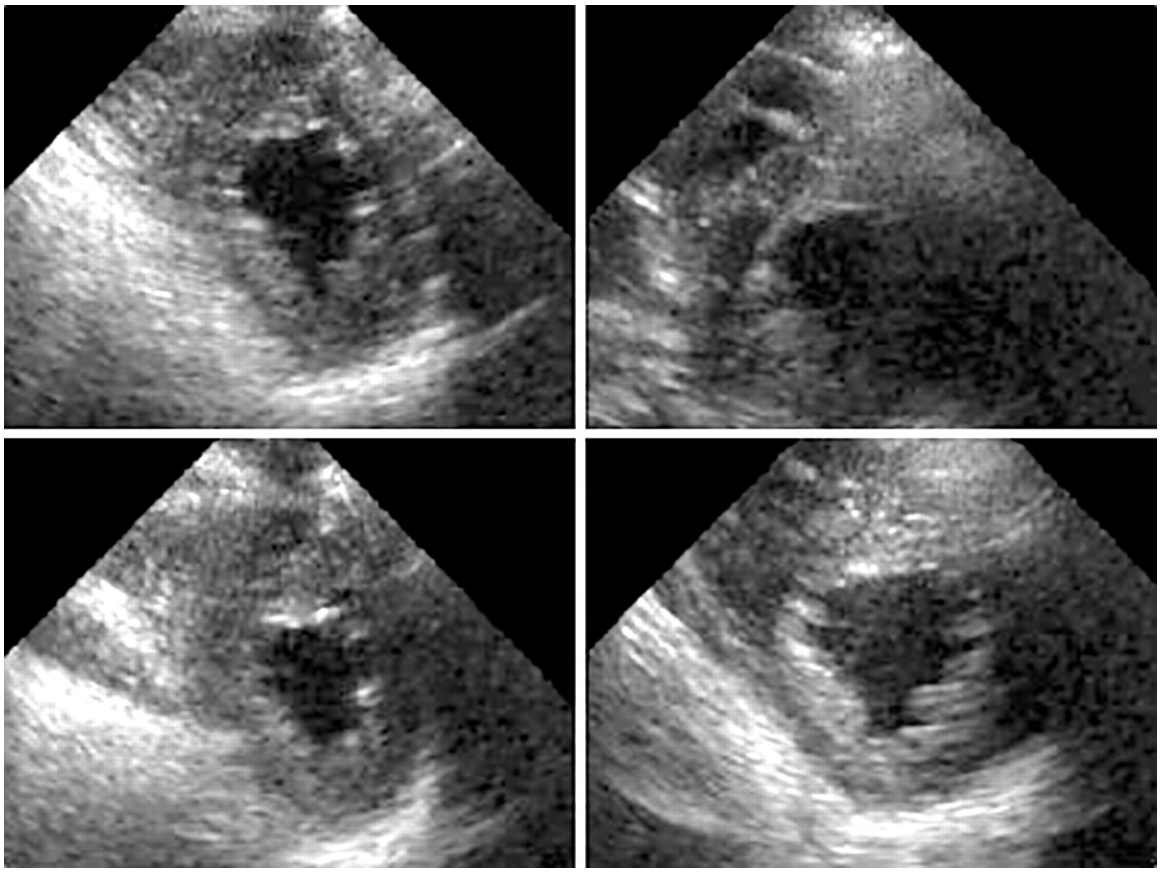


Figure 9: Biphasic response during a dobutamine stress echocardiogram. Upper left (rest), upper right and lower left (low doses of dobutamine), and lower right (high doses of dobutamine). Visualisation of wall motion is not possible in 2D still images but note the initial decrease in size of the cavity at end-systole and then subsequent increase in cavity suggestive of diffuse ischaemia at stress (126).

1.7.2 Current Clinical Standard

The impact of operator-dependence on the accuracy of stress echocardiography is clearly demonstrated in a meta-analysis by Geleijnse et al. of 62 studies investigating dobutamine stress echocardiography (64). The mean sensitivity and specificity values for detecting prognostically significant disease, (defined as >70% stenosis on an angiogram within one year of the stress echo), were calculated at 82% and 81%, respectively, which is in agreement with the values quoted by the

National Institute for Health and Care Excellence (NICE) in their most recent guidelines (127). However, it should be noted that the sensitivity and specificity values from the individual studies included in this meta-analysis ranged from 33-96% and 38-97%, respectively (Figure 10). These results show that there is a wide variation in accuracy and that there is considerable room for improvement as, on average, one in every five patients could be misdiagnosed. However, in some studies this number could have been even higher. This finding is corroborated by the meta-analysis carried out by NICE as part of their guideline development for assessment of stable coronary artery disease. Analysis of heterogeneity revealed an I^2 value of 64% in the assessment of sensitivity values and 65% for stress echocardiography specificity. This suggests that the average sensitivity and specificity values should be interpreted with caution due to the large heterogeneity between the studies.

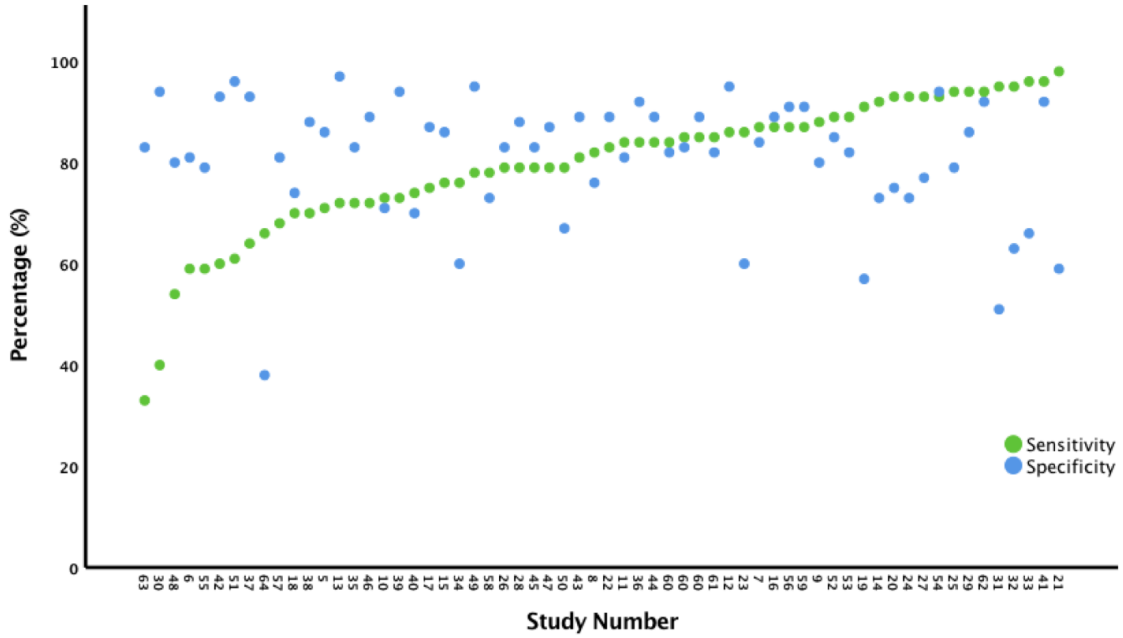


Figure 10: Range of sensitivity and specificity (64)

Whilst these findings describe the accuracy of stress echocardiography used in carefully controlled clinical research trials, usually from a single-centre, there is a paucity of evidence for the use of stress echocardiography in routine clinical practice from multiple centres spread across the country. We therefore set up EVAREST study to investigate the use of stress echocardiography and whether combing the analysis of imaging-based biomarkers with AI can be used to improve its accuracy to predict prognostically significant disease.

1.8 Thesis Hypothesis and Objectives

The aim of this thesis was to develop methodology to improve the clinical diagnosis of coronary artery disease by stress echocardiography using machine learning. To enable this, I aimed to develop technology to address the following steps 1) image processing 2) image quantification and 3) prediction of patient outcomes.

My specific hypotheses were:

1. Using deep learning techniques, I can develop an image processing framework that will set the foundation for extracting measurements and features from images, by providing accurate and robust contours of the left ventricle.

2. Using the above contours, I will be able to extract key clinical and functional measurements that will assist in the diagnosis of coronary artery disease.

3. Utilising both of the previous steps will enable extraction of additional, novel features specific to coronary artery disease and I will be able to feed these into a machine learning classifier which will enable an accurate prediction of patient outcomes in coronary artery disease.

2.0 Study Population

2.1 Overview

This thesis primarily employs data collected from the EVAREST study. EVAREST was set-up by myself and Dr Lisa Ayers and conducted under the auspices of principle investigator Prof Paul Leeson. EVAREST is planned to recruit 8000 patients referred for a stress echocardiogram to evaluate the presence of ischaemic heart disease at one of the 30 participating NHS hospitals throughout England. Participants were recruited into one of two study groups (see Figure 11). Participants who were recruited to group one had venous blood samples obtained to allow for the assessment of blood biomarkers during the stress echocardiogram, however these were not considered for this thesis. Group two participants did not provide blood samples, instead, only their stress echocardiogram images were collected for analysis in conjunction with those images obtained from participants in group one to identify imaging biomarkers.

All study procedures were carried out during the clinical visit of the participant to the cardiology department at the various sites. The stress echocardiogram was carried out by the standard clinical team to which the patient was referred, and the research participation did not affect the clinical care pathway of the patient. The additional study-specific components, including taking venous blood samples (group one only) and the recording of data (such as demographic information, relevant medical history and data about the stress echo) was performed by trained study investigators, identified on the delegation log.

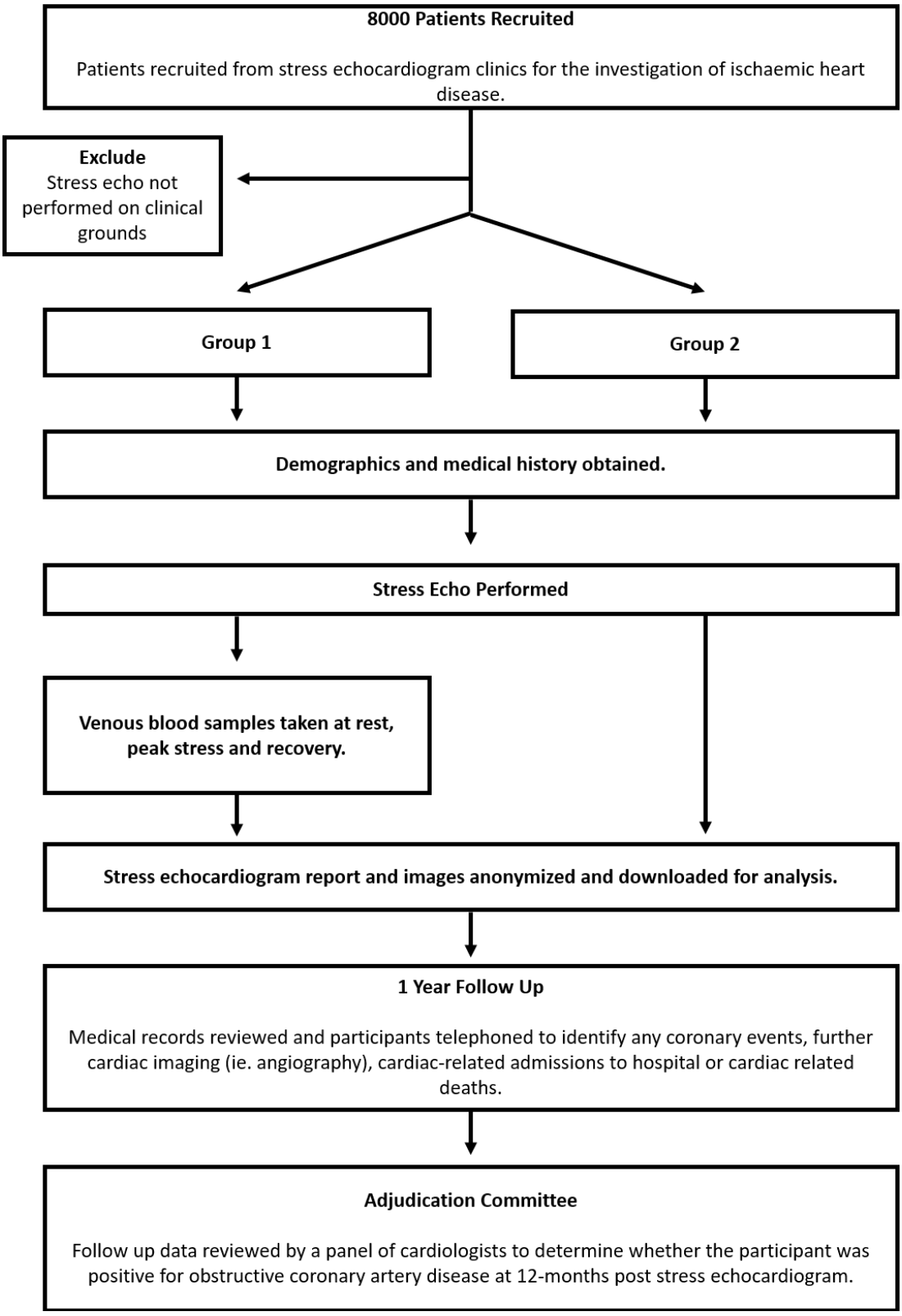


Figure 11: Study design and workflow.

2.2 Eligibility Criteria

Eligible participants had to fulfil the following:

- Be willing and able to give informed consent for participation in the study.
- Be male or female, aged 18 years or above.
- Be being investigated with stress echocardiography for suspected ischaemic heart disease.

The participant could not enter the study if the following applied:

- Unable or unwilling to give valid consent for participation in the study.

2.3 Recruitment

Patients who were currently under investigation or follow up for ischaemic heart disease were recruited. Potential participants that fitted the inclusion/exclusion criteria were identified by the clinical care team from the waiting lists for stress echocardiograms and were sent a letter of invitation and participant information leaflet at least one week prior to their appointment. Patients who were having a stress echocardiogram to assess for valve disease, left ventricular outflow tract obstruction or tissue viability studies were not eligible for recruitment. Figure 12 shows the recruitment sites and phases across the UK.

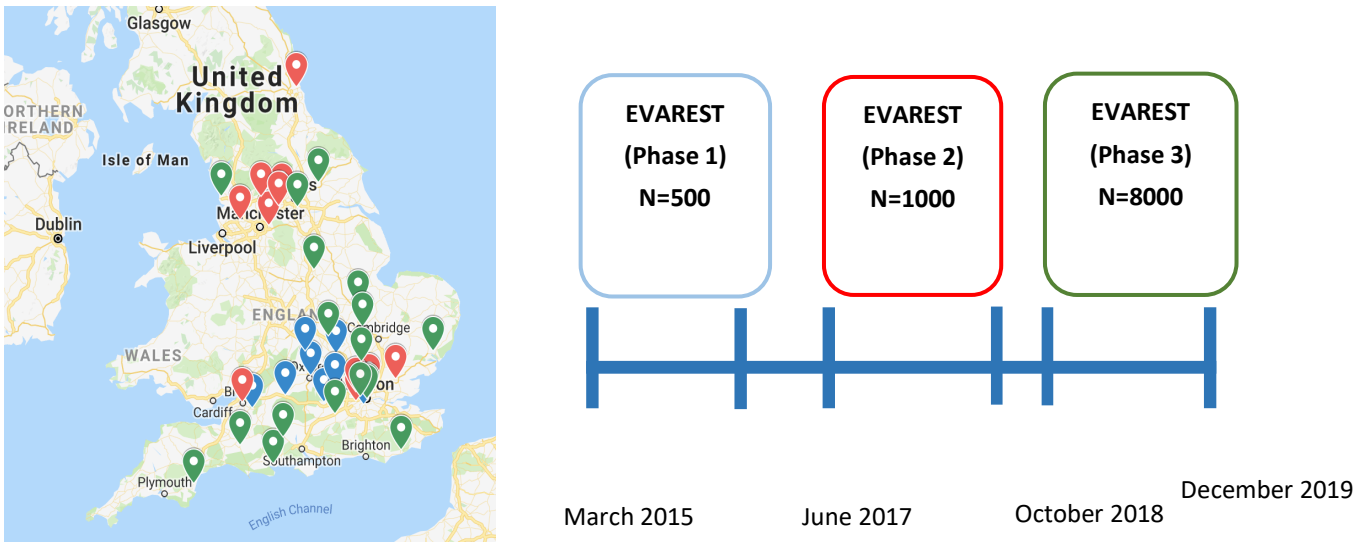


Figure 12: EVAREST recruitment sites and phases

2.4 Informed Consent

Before any study-specific procedures were carried out, written informed consent was obtained by means of a dated signature from the participant. The person who obtained the consent was suitably qualified and experienced and had been authorised to do so by the Chief/ Principal Investigator. A copy of the signed informed consent form was given to the participant, a further copy placed in the medical notes and the original kept by the study site.

In most cases, when the patient was referred for a stress echocardiogram, the participant was given at least 24 hours to consider the information and decide whether they wish to participate in the study. There were occasions, due to the nature of certain clinics, such as rapid access chest pain clinics, where it was not possible to allow 24 hours to read the patient information leaflet (PIL) before the stress echocardiogram. In these situations, the (PIL) was given to the patient as soon as possible after referral for a stress echocardiogram, and the participant given

as much time as possible to consider the study information. The minimum time period between referral from the rapid access clinic to stress echo procedure was one hour. It was emphasized to the participant that they could decide to withdraw from the study at any time. The participant did not have to state a reason to withdraw from the study.

2.5 Data Collection

2.5.1 Medical History

A detailed medical history was obtained for each patient from their medical records including:

- Patient demographics
- Previous and current medical conditions
- Cardiac risk factors
- Cardiac medical history (including details of any previous cardiac procedures or tests)
- Current medications

Following the stress echocardiogram, the interpreting clinician's report was obtained, detailing the protocol used, adequacy of the stress and whether the test was positive or negative for inducible ischaemia. Additionally, the stress echocardiogram images were also downloaded, de-identified and stored for further analysis.

2.5.2 Stress Echocardiogram

The study aimed to capture data on the different protocols used in stress echocardiography across the country. As such, the study does not specify any

particular stress echocardiogram protocol, instead each site used their standard protocol. All methods of stress (pharmacological, exercise and paced stress) was accepted as part of the study and the use of any contrast agent was at the operator's discretion. Group 1 participants had blood samples obtained at rest, peak stress and during recovery.

2.5.3 Echocardiogram protocol

Participants scheduled for DSE were asked to avoid agents that could antagonise the effects of dobutamine (e.g. beta-blockers) in the 24 hours prior to the test. Dobutamine infusion was commenced intravenously at 10μg/kg/min and increased by 10μg/kg/min at 3-minute intervals to a maximum of 40μg/kg/min, or until 85% of maximal age-predicted heart rate was achieved. If peak maximal heart rate was not reached with dobutamine, and the patient had no contraindication, atropine was administered (maximum 1.2mg). The test was terminated early if the patient developed significant ischaemia or an arrhythmia, and, if necessary, for clinical stability, IV metoprolol or verapamil was administered as an antagonist. If exercise was used as a stressor, patients were exercised according to the Bruce protocol for treadmill exercise or the world health organisation protocol for bicycle ergometer.

Echocardiography was performed using Phillips iE33 or EpiQ-7C systems, with an X5-1 transducer (Philips Medical Systems, Zoetermeer, The Netherlands) or a GE Vivid E95, with M5Sc and 4V-D transducers (GE Healthcare, Chicago, USA).

Image quality was assessed prior to the start of the test and, if endocardial definition was poor in two adjacent myocardial segments, a SonoVue contrast infusion was

started at 0.7ml/min to improve left ventricular opacification (Bracco Imaging, Milan, Italy). Two-dimensional echocardiographic images were obtained of the left ventricular apical four, three and two chamber views as well as the parasternal long and short (at the level of the mid-ventricle) views at baseline, at an intermediate level of stress and then at peak stress. Wall motion was assessed off-line, based on systolic wall thickening and endocardial movement, by two experienced operators who evaluated the whole myocardium split into a 16-segment model. Scores were given for wall motion as 1 = normal; 2 = hypokinetic; 3 = akinetic and 4 = dyskinetic. A deterioration in segment score with increasing dobutamine was used as the indicator of ischaemia and each participant defined as having a 'positive' i.e. evidence of inducible ischaemia, or 'negative' stress echocardiogram.

2.5.4 Image processing

All images were anonymised prior to their transfer to Oxford Research Echocardiography Core Laboratory (ORECL) and uploaded into Intellispace Cardiovascular (ISCV) (Philips Medical Systems, Zoetermeer, Netherlands) for quality assurance checking (by BSE certified echocardiographers) and processing.

2.5.5 Blood Sample Collection & Storage

For extracellular vesicle (EV) measurement, citrate blood samples were obtained peripherally before, immediately following and one-hour post-stress echocardiogram in all patients. Citrate blood samples were centrifuged at 3000g for 15 minutes within 10 minutes of the blood collection. Plasma was then centrifuged again at 13000g for 2 minutes to produce platelet-poor-plasma. Aliquots of platelet-poor-plasma were frozen immediately and stored at -80°C. For other biomarker measurement, EDTA,

heparin and serum blood samples were obtained. Samples were centrifuged at 3000g for 15 minutes within 10 minutes of the blood collection. Plasma and serum were frozen immediately and stored at -80°C. The samples of plasma, serum and DNA were stored for at least 10 years and may be used in future ethically approved studies as our understanding of blood vessel function grows.

2.5.6 Measurement of Microvesicles

Aliquots of platelet-poor-plasma were thawed at room temperature, and washed twice at 18,000g for 30 minutes. Levels of microvesicles (MVs) derived from platelets, red blood cells, white blood cells and endothelial cells were assessed by staining with monoclonal antibodies and acquiring data by flow cytometry. Platelet-poor-plasma sample was incubated with the appropriate monoclonal antibody for 30 minutes at room temperature, protected from light, followed by addition of 900µl PBS-Calcium. Data was acquired using a BD FACSCalibur® (Becton Dickinson, Oxford UK).

Levels of circulating EVs were also measured by commercial coagulation assays to define the functional capability of the vesicles, and by novel technologies including Nanoparticle Tracking Analysis (NTA) and Tunable Resistive Pulse Sensing (TRPS) to accurately assess the size of the vesicles being measured.

2.5.7 Imaging Biomarker Measurement

Imaging biomarkers relevant to myocardial ischaemia, such as LV geometry and function, were extracted and analysed from the baseline and peak stress images. Extraction of biomarkers increased throughout the study as technology developed

and novel biomarkers were discovered. All biomarker analysis was performed blind to the patient's one-year follow-up data. Details of biomarkers extracted are described in subsequent chapters.

2.5.8 One-Year Follow-Up

All participants were asked to consent to follow-up for one year following their stress echocardiogram. One year after the stress echocardiogram, the participant's medical records were reviewed to identify whether the participant had any coronary events, further cardiac imaging studies (such as coronary angiograms or repeat stress echocardiograms), cardiac-related admissions to hospital or cardiac-related death. Coronary angiography was used as the gold-standard to identify clinically significant coronary artery disease. Where appropriate, participants were contacted via telephone to ensure that the information obtained from their medical notes is accurate and that any further events or tests carried out at a different hospital that were not documented in the recruiting hospital's medical records were captured. This dual follow-up approach ensured that the outcome data obtained was accurate and complete.

2.5.9 Outcome Adjudication

Following the collection of one-year outcome data, an EVAREST Adjudication Committee at the Oxford Research Echocardiography Core Laboratory, University of Oxford reviewed all outcomes. This committee comprised of a panel of two cardiologists, the Principal Investigator and an independent cardiologist, who reviewed the follow-up data and assigned outcomes for each participant based on their clinical status (see Figure 13). The Adjudication Committee was blinded to the

result of the stress echocardiogram. A positive outcome indicated the presence of clinically significant coronary artery disease (i.e. a stenosis > 70% luminal diameter in one of the main epicardial vessels or > 50% in the Left Main Stem, an FFR < 0.85, decision to revascularize, acute coronary event or cardiac-related death).

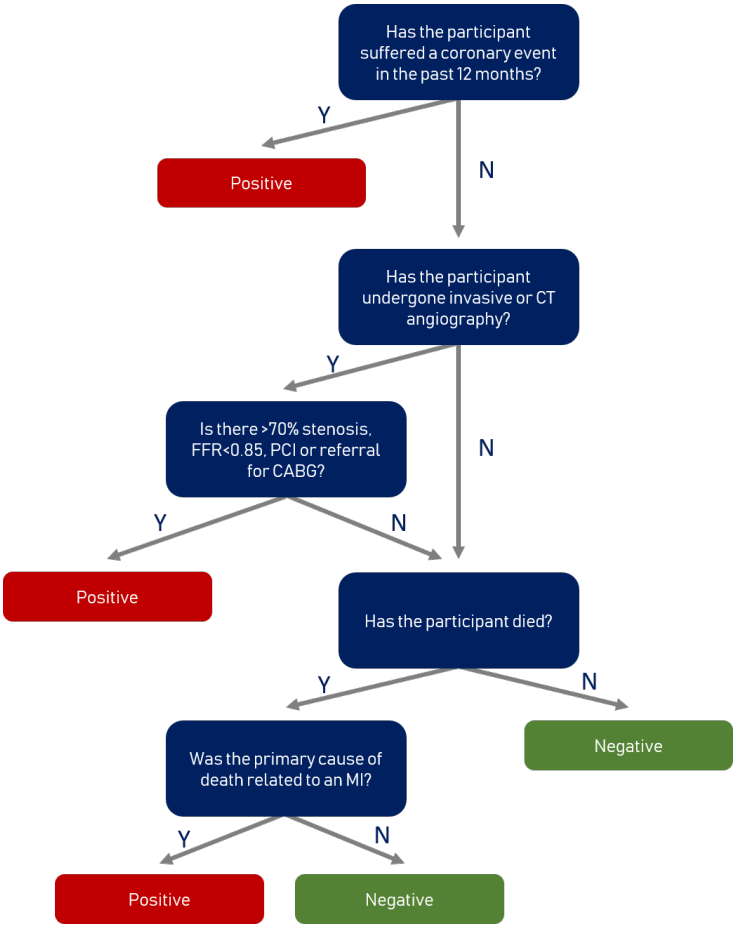


Figure 13: Decision tree demonstrating the outcome adjudication process.

2.5.10 Discontinuation/Withdrawal of Participants from Study

Each participant had the right to withdraw from the study at any time. In addition, the Investigator could discontinue a participant from the study at any time if the Investigator considered it necessary for any reason including:

• Ineligibility (either arising during the study or retrospectively if not known at time of screening).

• Withdrawal of Consent.

• Loss to follow-up.

If the participant stated a reason for withdrawal, the reason was recorded in the e-CRF. If a participant wished to withdraw from the study, they were asked whether they would give consent for the study team to make use of the data they already hold about the participant.

2.5.11 Definition of End of Study

The end of the trial was determined to be when the last analysis of the last blood sample has been carried out or when the final follow-up phone call has been made, whichever is latest.

2.6 Statistical Analysis

Preliminary analysis of baseline data collected has been carried out on 4000 participants, recruited from March 2015 to June 2019. Data were analysed using STATA IC 15 (STATA Corp. LLC, Texas, USA). Comparisons between sub-groups were made using Chi-squared analysis, with a p-value of ≤0.05 used to indicate statistical significance. Figures are presented as mean ± standard deviation. Details of the demographic profile of the EVAREST population, stress echo protocols and outcomes based on this analysis are presented below.

2.7 Patient Demographics

The mean age of patients undergoing stress echocardiography was 65 ± 11.8 years. 54.4% of the population were male. At the time of their stress echocardiogram, patients had a mean body mass index of 28.9 ± 5.6 Kg/m^2, with 75% of patients being categorized as overweight, obese, severely obese or morbidly obese (see Table 2). 1646 patients (47.7%) had never smoked, 1336 (38.7%) were ex-smokers, whilst 472 (13.7%) were current smokers.

	Mean	Standard Deviation
N = 3604 **Missing Data = 396**		
Sex	Male – 1959 (54.4%) Female – 1645 (45.6%)	-
Age	65.0	11.8
Height (m)	1.68	0.97
Weight (Kg)	81.8	17.4
BMI (Kg/m^2)	28.9	5.6

Table 2: Patient demographics

The most common risk-factor for cardiovascular disease was hypertension, affecting 1233 (34%) of the study population, followed by 1161 (32%) with hypercholesterolaemia. 74 patients (2.1%) had pre-existing peripheral vascular disease. 25 patients (0.7%) had type I diabetes, 416 (11.5%) had type II diabetes, whilst 223 (6.2%) had diabetes (type not specified). 286 patients (7.9%) had a positive family history of premature coronary artery disease. On average, patients had at least one risk factor for cardiovascular disease (see Table 3). Furthermore,

1262 patients (35%) had documented pre-existing coronary artery disease. Of this sub-group of patients 560 (15.5%) had suffered a previous myocardial infarction. 1011 (28.1%) had previously undergone percutaneous coronary intervention, whilst 275 (7.6%) had undergone coronary artery bypass grafting.

Risk Factor	Frequency	Percentage (%)
N = 3604		
Smoking Status		
- **Ex-Smoker**	1336	37
- **Current Smoker**	472	13.1
Hypertension	1233	34.2
Hypercholesterolaemia	1161	32.2
Peripheral Vascular Disease	74	2.1
Diabetes		
- **Type I**	25	0.69
- **Type II**	416	11.5
- **Unspecified**	223	6.2
Positive Family History	286	7.9
Previous MI	560	15.5
Previous Revascularisation		
- **Previous PCI**	1011	28.1
- **Previous CABG**	275	7.6

Table 3: Patient risk factors at time of stress echocardiogram.

1374 (38.1%) of patients had non-cardiac co-morbidities. The most common co-morbidities were respiratory conditions, affecting 463 patients (12.8%). Other common co-morbidities included musculo-skeletal conditions (affecting 11% of patients) and gastro-intestinal conditions (affecting 9% of patients). 3174 (88.1%) patients were currently taking medications when they attended hospital for their stress echocardiogram (See Table 4).

Drug Class	Frequency	Percentage (%)
ACE Inhibitor	145	4
Angiotensin Receptor Blocker	69	1.9
Anti-Anginal	110	3.1
Antithrombotic/Anticoagulant	507	14.1
Aspirin	383	10.6
Beta Blocker	192	5.3
Calcium Channel Blocker	224	6.2
Diuretic	232	6.4
Nitrates	580	16.1
Statin	1070	29.7
Steroid	134	3.7
Other Anti-hypertensive	270	7.5
Other Medications	520	14.5

Table 4: Breakdown of medications taken by patients at the time of their stress echocardiogram.

2.8 Stress Echocardiogram Data

The most common method of stress was dobutamine infusion, which was used in 2493 (69%) cases and exercise stress was used in 983 (27.2) cases, there were missing data on 3.8%. No cases of vasodilator stress were recorded.

As per protocol, dobutamine infusion was increased from 5 mcg/kg/min to 10 mcg/kg/min up to a maximum of 40 mcg/kg/min, if required, to achieve target heart rate (85% of maximum predicted HR). Target heart rate was achieved using 5 mcg/kg/min in 28 cases (0.78%), 10 mcg/kg/min in 118 cases (3.3%), 20 mcg/kg/min in 719 cases (20.5%), 30 mcg/kg/min in 1106 cases (30.1) and 40 mcg/kg/min in 339 cases (9.4). Atropine was used to supplement the dobutamine infusion in 1299 cases (36%). The mean dosage administered was 466 ± 300 mcg.

The level of stress achieved was adequate (i.e. greater than 85% of the age-predicted maximal heart rate) in 77% of stress echocardiograms. Therefore 23% did not achieve adequate stress due to symptoms or arrythmias. The mean heart rate achieved was 139 ± 16.4 bpm. This corresponds to 89.5 ± 12.4% of the age-predicted maximum heart rate. The use of left ventricular opacification agents was at the operator's discretion. Contrast (SonoVue, Luminity or Optison) was used to enhance endocardial definition in 82.1% of scans.

Whilst resting wall motion was normal in the majority of cases, resting regional wall motion abnormalities were present in 322 (18.2%) of cases. At peak stress, 356 (14.5%) scans were positive for inducible ischaemia as defined by the clinicians report.

1986 (81.1%) were negative for inducible ischaemia. 108 tests (4.4%) were abandoned. Reasons for early termination included inability to reach target heart rate (despite maximal doses of dobutamine and atropine), sub-optimal image quality, arrhythmia and vaso-vagal reaction to dobutamine.

2.9 Outcome Data

Of the first 4000 patients recruited into the study, 1000 patients have been followed-up to determine whether they had any further cardiac imaging studies, admissions to hospital, coronary events or cardiac-related death. 65 patients were deemed to have positive outcomes by the EVAREST Adjudication Committee. Of these 65, 3 patients suffered cardiac-related death, 15 suffered an acute coronary event and 47 had a positive angiogram. Of the 47 patients with positive angiography, 16 underwent PCI

and 6 were referred for Coronary Artery Bypass Graft (CABG). The remainder were managed with optimal medical therapy.

2.10 Stress Echocardiography as a Predictor of Outcome

Analysis of the first 1000 patients with one-year follow-up outcomes has allowed for the assessment of the accuracy of stress echocardiography (see Table 5). 47 patients were classed as true positives, whilst 826 were classed as true negatives. 109 patients were classed as false positives. Of this sub-group, 42 patients underwent angiography which did not demonstrate any significant flow-limiting lesion. Two patients underwent further assessment with myocardial perfusion scan (MPS), one with a repeat stress echocardiogram and a further two by CMR. However, our analysis also revealed 18 patients who underwent a false negative stress echocardiogram. Of this sub-group, 10 patients suffered an acute coronary event, two patients died of a cardiac-related cause and six patients had a positive angiogram.

Sensitivity and specificity values were calculated at 72.3% and 88.3%, respectively. Stress echocardiography had a positive predictive value of 30.1% and a negative predictive value of 97.9%. Interestingly there was no statistically significant difference between accuracy when taking into account the type of stressor used ($p=0.960$) or the use of LV contrast agents ($p=0.306$) (see Table 5).

	n	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	
Stress Echo Overall	998	72.3	88.3	30.1	97.9	
Exercise Stress Echo	118	92.3	86.7	41.2	86.7	
Dobutamine Stress Echo	862	66.0	88.4	26.9	97.7	
Unrecorded	18					
Contrast Echo	765	72.0	87.8	29.3	97.8	
Non-contrast Echo	200	71.4	90.9	37.0	97.7	
Unrecorded	33					

Table 5: Stress echocardiography sensitivity and specificity analysis, broken down by type of stress and use of contrast agent.

2.11 Discussion

The recruitment of 8000 participants from 30 sites across England will allow for an in-depth analysis of patient profiles, image quality, variation in protocol and use of stress echocardiography by hospitals throughout the country. To our knowledge, this will be the largest prospective stress echocardiography study in the world. By following up each patient for one year and obtaining outcome data, we will be able to determine the diagnostic accuracy of the stress echocardiogram in the detection of prognostically significant coronary artery disease by comparing the clinical outcome with the clinician's interpretation. This will allow for an up-to-date assessment of accuracy, since data used in previous large-scale meta-analyses (64,128) were recorded between the early 1990s and early 2000s. As ultrasound technology is continually advancing, we will be able to assess stress echocardiography accuracy using more up-to-date, commercially-available ultrasound machines and contrast agents currently used in the NHS.

As the exact stress echocardiogram protocol used is left to the operator's discretion at each hospital, the EVAREST study will be uniquely placed to report on how stress echocardiography is performed across the country and the site-to-site variation in protocol. The preliminary results demonstrate that pharmacological stress is the most commonly used form of stress, despite the most recent guidelines from the BSE stating that exercise is the preferred form of stress (129). This propensity to use pharmacological stress over exercise stress may be accounted for by it being considered easier to perform since it does not require a treadmill or bicycle ergometer and it does not require patients to have to exercise, potentially beyond what they would usually do, to ensure that target heart rate is achieved. This is particularly important as it is well-known that the accuracy of exercise stress echocardiography is reduced at sub-maximal heart rates (130).

Analysis comparing the accuracy of exercise stress echocardiography with dobutamine stress echocardiography revealed no statistically significant difference between the two forms of stress. This finding contradicts the findings of other studies, which have previously demonstrated the superiority of exercise stress, compared to pharmacological agents in the detection of prognostically-significant inducible ischaemia (131). However, as demonstrated in Table 5, exercise stress echocardiography was, in fact, more sensitive (92.3% compared to 66.0%). The lack of statistical significance between the difference in accuracy between these two forms of stress may be explained by the considerably smaller number of exercise stress echocardiograms, in relation to the number of dobutamine studies. As the study progresses and the sample of exercise echocardiograms increases, it will be interesting to re-examine how the two forms of stress compare.

Preliminary analysis of the overall accuracy of stress echocardiography has been carried out on the first 1000 patients, yielding a sensitivity and specificity of 72.3% and 88.3%, respectively. Whilst the specificity is in agreement with other studies (64,127), the sensitivity is slightly lower than expected. This may reflect the 'real-world' use of stress echocardiography, in a clinical setting, compared to data obtained purely for research purposes or may reflect the nature and timeframe for the 'real world' outcome being used to assess accuracy.

The EVAREST study provides a unique comprehensive assessment of stress echocardiography across England. The preliminary results obtained from the data collected so far have provided a fascinating insight into the current state of stress echocardiography across six centres in England. However, it remains to be seen whether this is representative of the rest of the 30 centres taking part in the EVAREST study. The analysis of imaging biomarkers in combination with machine learning provides an exciting opportunity to examine whether additional prognostic information can be obtained during a stress echocardiogram that may increase the accuracy of the test and therefore ensure that patients are managed appropriately.

3.0 Method Development for Autocontouring Stress Echocardiogram Images

3.1 Introduction

In this chapter I will describe the methodology for developing an automated contour of the left ventricle (LV) using stress echocardiogram images.

Conventional methods for automated contouring have previously been mainly comprised of using active contour models (132–134), such as contour snakes and level sets, which only use the image intensity in order to generate contours. In non-controlled image cases, these generally do not work so well, particularly for noisy data with poorly delineated edges as seen in ultrasound images (135). If ultrasound images are acquired with a high frame rate and only incorporate smooth motion, these methods can eventually yield good results (136). However, such conditions are usually not present in cardiac ultrasound and particularly not in stress echocardiograms (137).

Speckle tracking can be used to track the endocardial walls with relatively good efficacy (138). This involves using the fact that the background speckle pattern in ultrasound images is comparatively stable so the variations in the speckle are solely due to motion. By tracking segments of the image, the overall motion of the endocardial walls can be captured. However this is only effective in non-contrast enhanced images as the speckle pattern is destroyed when using contrast enhancement micro-bubbles (139). Speckle tracking is more effective in data acquired at rest and often requires manual adjustment (140).

Convolutional neural networks (CNNs) have become increasingly adept at solving image problems (141). The introduction of very deep networks trained on extremely

large datasets has yielded ever increasing performance particularly for classification problems (142). Image segmentation problems are now being tackled more and more using CNNs in a variety of scientific fields (143,144).

U-net is currently the best CNN for image segmentation of biomedical images (145) and has previously been used in cardiac echo to delineate different regions of the heart in 2D echo with resting images (138). Its efficacy is achieved by the concatenation of feature maps across two paths of a CNN thus allowing high- and low-resolution image feature structures to be combined and contextual information to be propagated to the high-resolution image features. One side of the network consists of a contracting path which increases the number of feature channels at each layer, as in a conventional CNN. This is followed by an expansive path in which the convolutional steps also perform up-sampling of the feature maps. At each level, the corresponding feature map from the contracting path is concatenated to the expansive path.

In this work, labels were only available for end-diastolic and end-systolic image frames. However, Lieman-Sifry (2017) shows that by training a neural network simply on end-systolic and end-diastolic labels, the network generalised well to all image frames for cardiac MRI image segmentation (147). The available labels were used to train a U-net type network and construct an auto-contouring framework that worked on rest and stress data, with or without contrast for the 2 and 4-chamber and short-axis views of the heart.

The aim of this work was to produce a framework to 1. Determine the type of stress echocardiogram image (contrast enhanced or non-contrast) 2. Automatically contour in the two and four chamber cardiac views and 3. Develop a separate model for automatic contouring of the short-axis (SAX) view. 4. Generate clinically useful features from the contours.

3.2 Methods

3.2.1 Dataset

Images collected as part of the EVAREST study described in chapter 2 were used for this work. For the first classifier that differentiates between contrast and non-contrast images 124,201 image frames were used. For the second classifier that produced the contours for 2 and 4 chamber views 5692 frames were used for contrast enhanced images and 2182 frames for the non-contrast images. For the third classifier 2747 frames were used to produce contours for the SAX view. In each case the data was split into 80% training and 20% validation.

3.2.2 Contrast/Non-Contrast Binary Classifier

A binary classifier Visual Geometry Group neural network (VGG) was constructed to address this problem. This comprised a basic CNN with five 2D convolutional layers which takes a grayscale image as input and outputs a value between 0 and 1 (as the final activation layer is a sigmoid function, this is a non-binary output), where 0 corresponds to a non-contrast image and 1 to a contrast.

During training, binary cross-entropy was used as the loss function. Class-weighting of a factor of 3 was applied to the non-contrast data to match the uneven distribution of the training data. Data was split into frames for training and frames for validation. Initially a set of 86,941 frames was used for testing, 24,840 for validation and 12,420 for testing. This was instead split into just 80% training and 20% validation sets when it was discovered that the model performance was very effective and generalised directly to the testing set.

3.2.3 Autocontouring framework for 4C and 2C

Next an auto-contouring framework was developed using a CNN. The CNN was trained using contours of the endocardial border manually drawn by trained operators. End-systolic and end-diastolic image frames from 985 subjects acquired in the 2 and 4-chamber views at rest and peak stress were contoured using the Ultromics Echuity v0.5.0 proprietary software platform by 3 experienced contours. An example of a contoured image frame is shown in Figure 14.

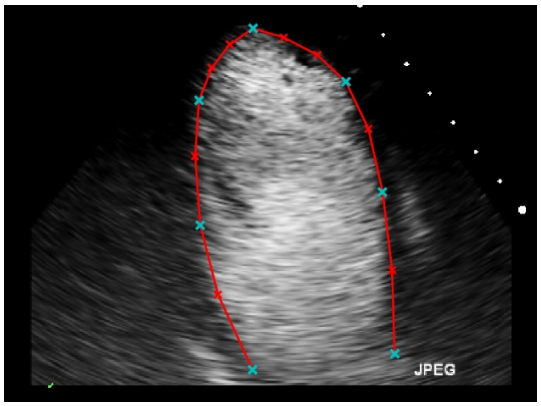


Figure 14: Diastolic 2-chamber image with corresponding contour drawn by a trained operator. The blue crosses indicate the required points which delineate the

segments of the heart and the apex. The red crosses are points added to smooth the contour and ensure it follows the endocardial border.

3.2.3.1. Neural network and training

The contours where then fed into the CNN in order to train the neural network, the problem was set up as an image segmentation task. This required the contours to be converted into binary masks as shown in Figure 15.

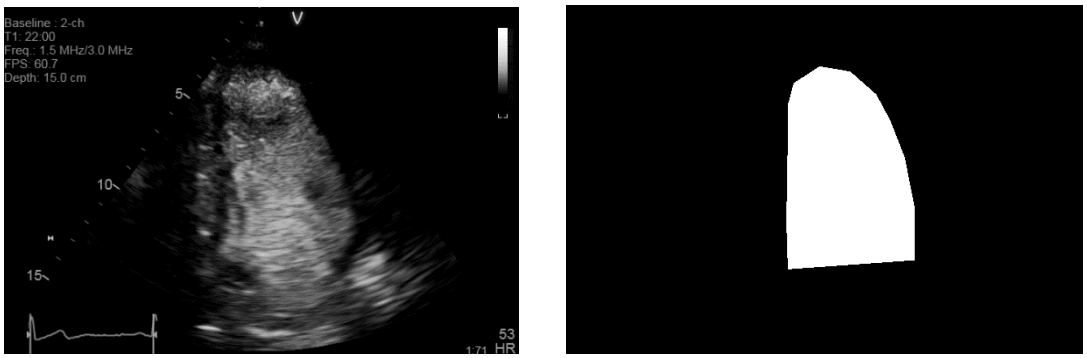


Figure 15: Diastolic 2-chamber image frame acquired at rest (left) and its corresponding mask (right) generated from contours manually drawn by trained operators.

The CNN setup was created in *Python 3.5* using *Keras 2.2.4* (148) with a *Google Tensorflow* (149) backend and all computations were run on a 3XS Deep Learning G10 PC with an Intel Core i9 Processor, 64GB RAM and 2 NVIDIA GeForce GTX 1080 Ti GPUs each with 11GB of memory.

A version of the U-net (145) architecture shown in Figure 16 was used as the neural network. This has been found to be highly effective in biomedical image segmentation problems. For this work, the original U-net framework was slightly

modified so the feature depth of the layers was reduced by halving the number of channels at each level. This appeared to help ensure the network was not so tuned to fine edges and sharp image features; which are less prevalent in noisy ultrasound images.

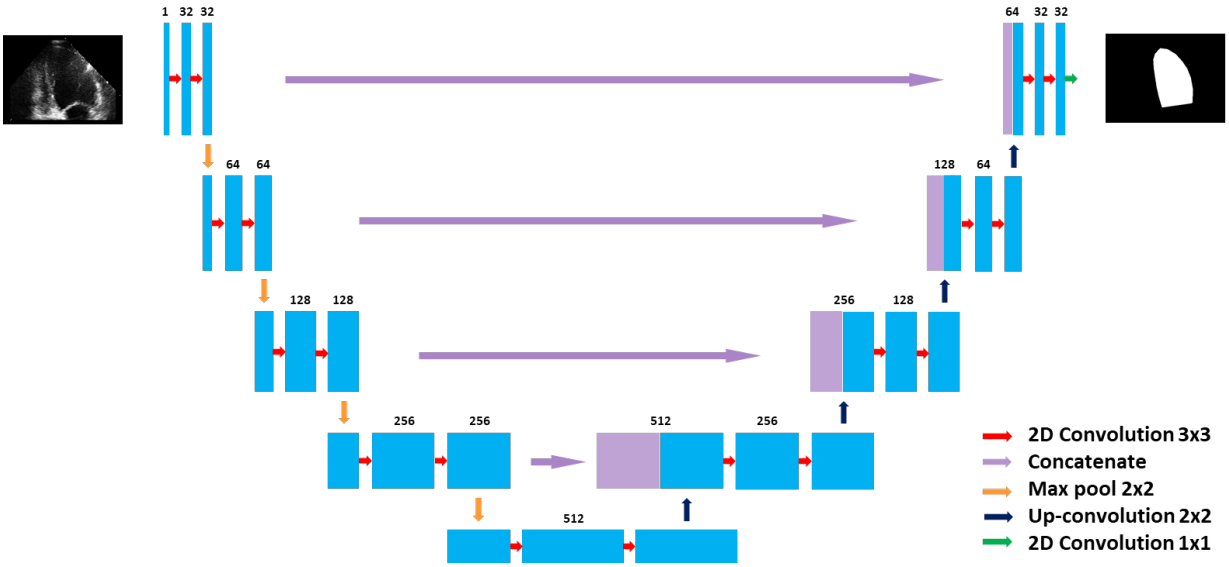


Figure 16: Modified U-net architecture used for auto-contouring. The feature depth of the layers is substantially smaller than those used in the original U-net as the convolutions were set to have half the number of filters. This was found to yield less structured output masks and training runs could be iterated more quickly.

The images and corresponding masks were resized/reshaped to 192x192 pixels, filtered with a Gaussian kernel to reduce speckle noise (also achieved to some degree by the down-sizing of the image) and Z-score normalisation was applied to the training dataset as a whole which was split into training and validation sets. The images were then fed into the U-net CNN framework.

3.2.3.2 Data Augmentation

Real-time data augmentation was performed to increase the accuracy of the model. The parameters for this included:

- Random rotations of up to 30°
- Horizontal shifts of up to distance = 0.2 x image width
- Vertical shifts of up to distance = 0.2 x image height
- Image zoom of up to 0.3 x image size
- Image shear of up to 10°
- Horizontal image flips

The trained network was then applied to all frames (147) in the acquired datasets for the data taking care to account for image scaling and conversion between pixel and real-world coordinates.

3.2.3.3 Generating a Contour from CNN Output

As the CNN framework is set up as an image segmentation problem, generating an actual contour from this mask requires some post-processing steps. This is performed using the flowing steps:

1. The edge is found from the segmentation mask using the *measure* function from the *Python skimage* package.
2. Sometimes noise creates multiple segmentations in the mask. The code then selects the contour of maximum area that lies within central region – defined as the central 50% of the image window.

3. The contour is then aligned with the vertical y-axis. The point on the contour closest to the origin is taken as one basal point and the other basal point is found in the same way by flipping the contour about the y-axis and finding the other coordinate closest to the origin.
4. The highest point in the y-direction on this aligned contour is denoted as the apical point.
5. Then 10 points between each basal point and the apex are distributed evenly to produce an open contour of 21 coordinate points for each image frame.
6. This is performed for each frame in an image sequence through time. Each x and y-coordinate through time is then interpolated using a quadratic spline onto a lower temporal resolution using the *scipy interp1d Python* package.
7. The points are resampled back up to the original time resolution and the movement of the contour points is smoothed through time.

Performing these post-processing steps in this way maximises the chance of producing contours which track the endocardial walls smoothly through time. Figure 17 shows an outline of the overall processing steps required to generate a contour from an input image using this framework.

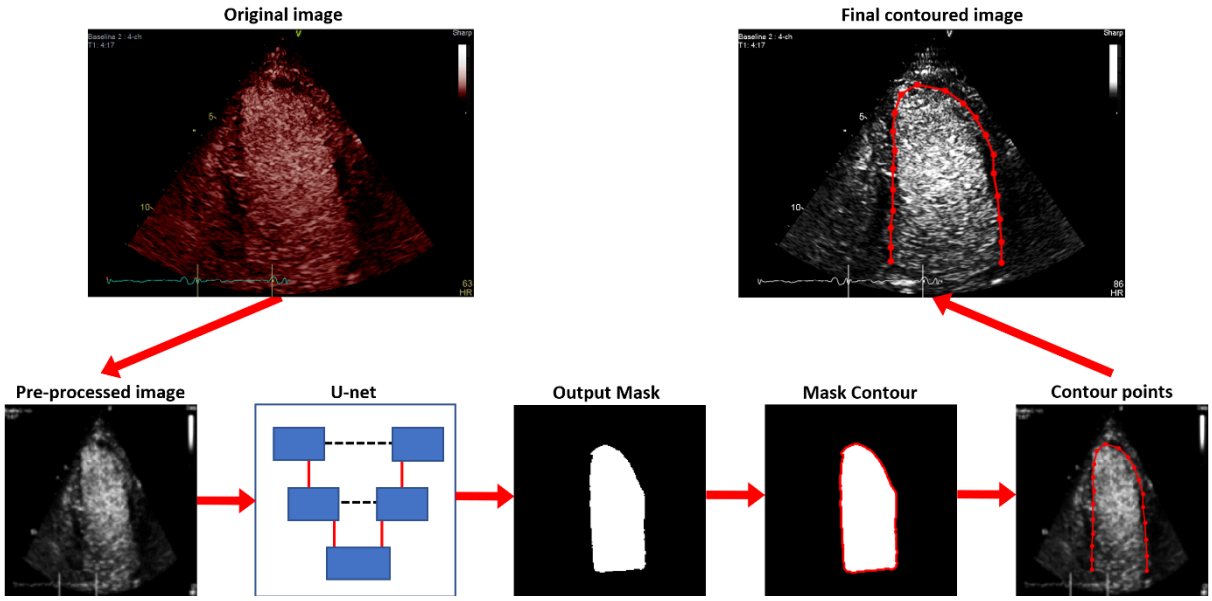


Figure 17: Overall auto-contouring pipeline described for a single image.

The input image is first pre-processed by converting to grayscale, resizing to 192x192 pixels and blurring with a 1-pixel kernel Gaussian filter. This is fed into the U-net neural network which outputs a segmentation mask. The edge of the mask is found, and a contour of 21-points is constructed from this as described in the methods. The contour can then be rescaled back into the correct size for the original image.

3.2.4 Short-Axis Auto-Contouring Model

While the two and four chamber views are relatively similar in appearance and structure, the short-axis (SAX) view is very different. Consequently, a separate U-net based model was trained to contour the short-axis view data.

The structure of the CNN was exactly the same as for the other views, but the input data and post-processing steps were modified. The input data was derived from the

same dataset as for the other model but there were a number of studies missing the SAX view, resulting in a smaller training dataset for this model of 2197 frames for training and 550 for validation. The smaller dataset also meant that it would likely be detrimental to the performance of the model to separate the training data into contrast and non-contrast data as for the two/four chamber model, although this has yet to be fully explored.

The point distribution steps for the SAX view differs in steps 3-5 in the process for the two/four chamber view contouring. Instead a set of 18 points is distributed at 20 degree intervals around 360 degrees centred about the centre of mass of the initial contour as shown in Figure 18. The points are then temporally smoothed in the same way as for the two/four chamber view models.

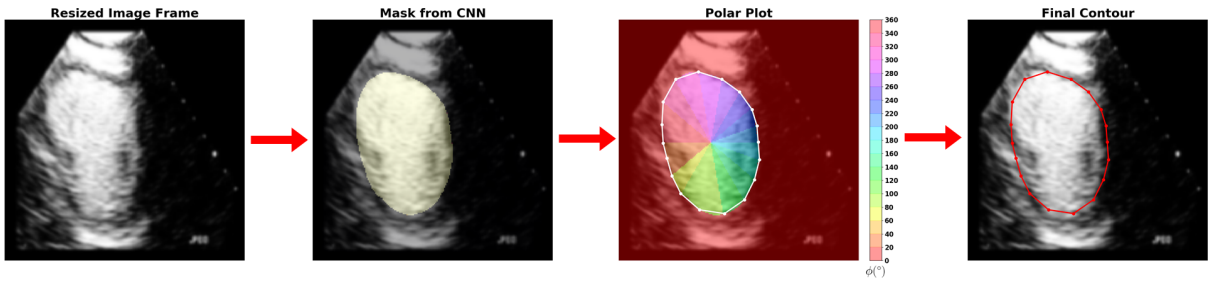


Figure 18: Process for generating contour points from SAX auto-contouring CNN output mask shown for an example individual image frame.

3.2.5 Calculating the cardiac cycle

As the auto-contouring models contour all frames of an image, it is necessary to first identify the cardiac cycle between end-diastole and end-systole, before computing biomarkers that are potentially clinically relevant to CAD. An algorithm has been

developed to perform this task automatically, calculating area from the contours in each frame and then adopting the following method:

1. Find candidates for start frame and end frame,
2. Filter candidates to only consider those with a reasonable area difference and area ratio,
3. Choose the start frame and end frame corresponding to the largest area difference.

This was implemented as a classifier and integrated as a step function after the autocontouring models in the above framework.

3.2.6 Feature Calculation

The 2-chamber (AC2) and 4-chamber (AC4) contour is made up of 21 points. Seven of these points correspond to the segments of the LV and are termed labelled points. The labelled points have indices, 1, 4, 7, 11, 15, 18, 21 in the set of 21 points.

- Points 1 and 21 are the basal points
- Point 11 is the apical point

A short axis view contour is made up of 18 points. Six of these points are approximately equally distributed along the wall and are termed the labelled points. The labelled points have indices 1, 4, 7, 10, 13, 16 in the set of 18 points.

3.2.6.1. Volume

To calculate the volume of the left ventricle, the below steps are to be followed using contours from the AP2 and AP4 views. The volumes are calculated separately using these steps for end-systole and end-diastole at both rest and peak stress.

1. Align AP2 and AP4 contours such that their apical points are the same.
2. Extend both ends of both sets of contours to the lowest y coordinate over both sets of contours by appending a point at both ends of each contour with that y-coordinate and the same x-coordinate as the corresponding end points.
3. Interpolate contours with a quadratic spline to 1000 evenly spaced points. The implementation should match that in the interp1d class of scipy.interpolate in Python v3.6.5.
4. Divide the y axis into n_d elliptical discs by taking the distance between the highest and lowest y-coordinates and dividing by n_d, using $n_d = 400$. This gives the height of each disc, δ. An example disc generated from AP2 and AP4 contours can be seen in Figure 19.
5. The volume is calculated using the following formula:

$$V = \pi \times \delta \sum_{i=1}^{n_d} a_i b_i,$$

where a_i and b_i refer to the radii along the major and minor axis of the elliptical disc i using points on the AP2 and AP4 contour. The contours are assumed to be directly perpendicular to one another and to meet at the apex as per the standard Simpson's Biplane method.

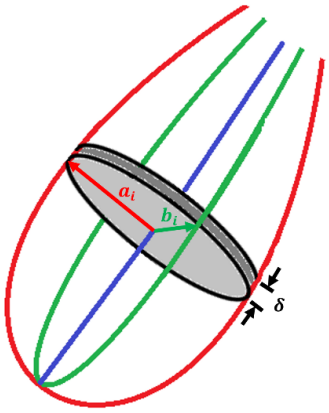


Figure 19: An example disc used when calculating volume following Simpson's biplane method – the red contour indicates the AP2 contour, and the green contour indicates the AP4 contour.

If there exist any points in one of the contours above the apical crossing point, the volume will extend upwards to the top of that contour by assuming only that contour's radius for the volume thus generating circular discs for the volume and not elliptical ones.

3.2.6.2 Volume difference, ejection fraction and strain

The ejection fraction is calculated using the formula:

$$EF_P = \frac{(V_P^D - V_P^S)}{V_P^D} \times 100$$

$$EF_R = \frac{(V_R^D - V_R^S)}{V_R^D} \times 100$$

The stroke volume is calculated using:

$$SV_R = V_R^D - V_R^S$$

$$SV_P = V_P^D - V_P^S$$

The difference in volumes at rest and stress for ED and ES is calculated using:

$$\delta V^S = V_R^S - V_P^S$$

$$\delta V^D = V_R^D - V_P^D$$

The cardiac output is calculated using:

$$CO_R = (V_R^D - V_R^S) \times HR_R$$

$$CO_P = (V_P^D - V_P^S) \times HR_P$$

Where HR_R and HR_P are the average heart rate at rest and peak respectively over all views.

Longitudinal strain is then calculated using:

$$\frac{p^S - p^D}{p^D} \times 100\%$$

Whereby p^S= Perimeter of systole and p^D = perimeter of the diastole contour calculated by the following equations (Figure 20):

- $p^S := \sum_{i=1}^{n-1} d^S_{i,i+1}$
- $p^D := \sum_{i=1}^{n-1} d^D_{i,i+1}$

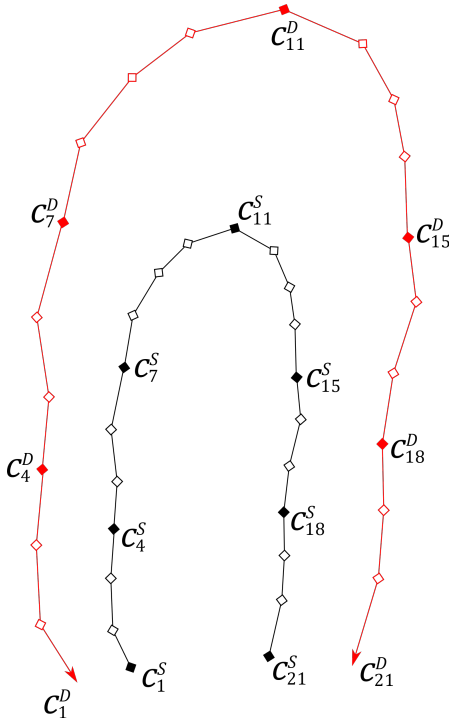


Figure 20: Example shown is p^D (red) and p^S (black)

3.2.7 Statistical Analysis

A receiver operator characteristic curve (ROC) will be used to assess the efficacy of the binary classification of contrast/non-contrast images. The efficacy of the network's autocontouring performance was assessed using the Sørensen-Dice Coefficient (DC) which is computed from the ratio of the intersection of the output masks (*Y*) and the known ground truth masks (*X*) as shown below:

$$DC = \frac{2|X \cap Y|}{|X| + |Y|}$$

The results for Ejection Fraction and Global Longitudinal Strain are validated in a separate validation study in the following chapter so will not be considered in the results below.

3.3 Results

3.3.1 Contrast/Non-contrast classifier

The nominal ROC produced by the model was 0.998 with training and validation losses in the same order of magnitude, the classifier correctly classified 100% of cases.

3.3.2 Autocontouring

The framework outlined created a generally effective and robust auto-contouring pipeline, with high degrees of accuracies measured by dice coefficient displayed in Table 6. Figure 21 shows how the mask developed by the UNet corresponds to the contours and Figure 22 illustrates contours produced in each view in both contrast and non-contrast data.

Model	Dice coefficient (%)
Contrast apical 2 and 4 chamber views	93.25
Non-contrast apical 2 and 4 chamber views	92.47
Parasternal short-axis view	92.10

Table 6: DC results for contrast 2 and 4 chamber, non-contrast 2 and 4 chamber and SAX.

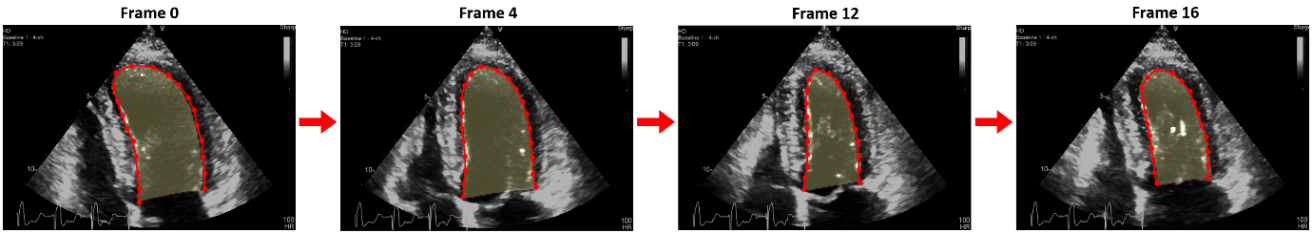


Figure 21: Series of image frames for a non-contrast dataset with their corresponding image mask from the U-net CNN overlaid in yellow and the final processed contour plotted in red.

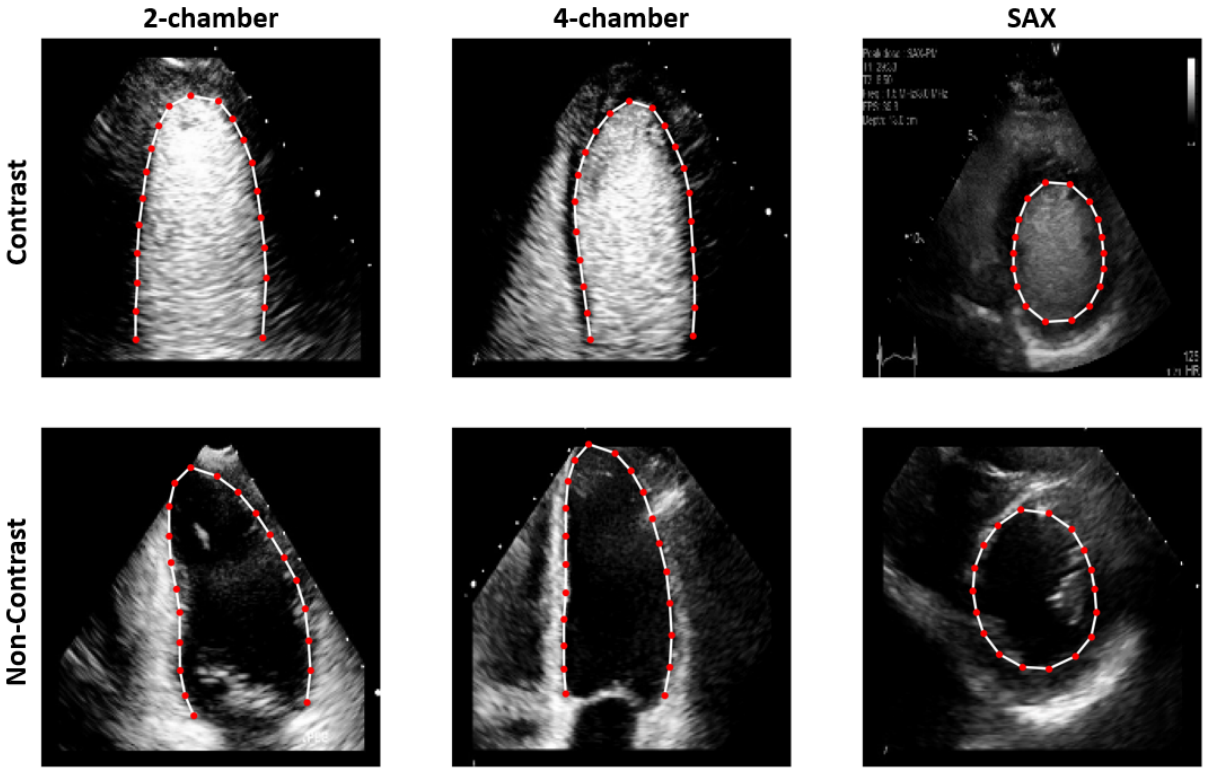


Figure 22: Results of the contours on contrast and non-contrast images in each view – AP4, AP2 and SAX.

3.4 Discussion

The binary classifier to detect whether a model is contrast or non-contrast has yet to show any incorrect classifications although has yet to be tested on further external datasets.

To our knowledge this is the first time that a classifier has been created to differentiate between contrast and non-contrast images. This is an important step because most AI classifiers have been used in general resting echocardiograms which are usually non-contrast compared with stress echocardiograms where contrast is more prevalent. Therefore it is critical to be able to process both types of images.

There have been studies, however, that have investigated a similar classification problem but trying to classify type of view instead. Madani et al. investigated view classification of 15 standard transthoracic echocardiogram views using a CNN with an accuracy of 97.8% (48). Similarly, Khamis et al. classified apical views using a CNN with an accuracy of 95% (49).

Although both studies demonstrate high degrees of accuracy, automated view identification for the majority of echocardiograms is not a problem due to embedded DICOM tags identifying the view from the echocardiogram acquisition protocol, which is how I was able to identify the views for this study. Furthermore, the image

classification results we achieved demonstrated a higher accuracy than both of these previous studies as we did not incorrectly classify a single image.

The Autocontouring framework also appears to have created a very accurate and robust pipeline for delineation of the LV endocardium in a wide-ranging echo dataset comprising of rest, stress, contrast and non-contrast echos. From the results, and visual inspection, the contouring appears to be able to track the endocardial border effectively in greater than 90% of cases compared to manually drawn contours. There have been many studies that have attempted to do this before with varying degrees of success (56,57,150–152). Domingos et al. attempted to use machine learning to delineate mid SAX echo views. This work was based on a small sample size of 25 datasets and is limited in its generalisability due to only focussing specifically on SAX views, in optimal, resting echocardiograms (56). Tsang et al. studied an automated 3D contouring software (HeartModel, Philips Healthcare) in 159 patients and compared the output measurements to CMR. This software performed relatively well, however, would not be applicable to the majority of stress echocardiogram scans and it remains to be seen how this would generalise to sub-optimal image qualities (151). Stebbing et al. attempted automated segmentation of the RV using ML with an accuracy of 90% but again was limited to optimal image quality and only validated on 35 patients (57).

These studies all showed promising applications but none would be applicable in the stress echo setting due to the different types of images (contrast/non-contrast) and the variable image quality, which is why we believe our framework represents an important improvement in the field. The Autocontouring results were not perfect,

however, and there is room for improvement to increase the DC scores. One of the greatest challenges was accurately identifying where the basal region starts. The base does not represent a hard edge, which caused issues with the U-net identifying it as a true boundary and then struggling to find image features that map to it. Consequently down-weighting this edge by diminishing the value of the mask to zero at the edge, as shown in Figure 23, may improve the performance of the network (145) as it will contribute less to the dice coefficient loss if the segmentation masks do not include this area.

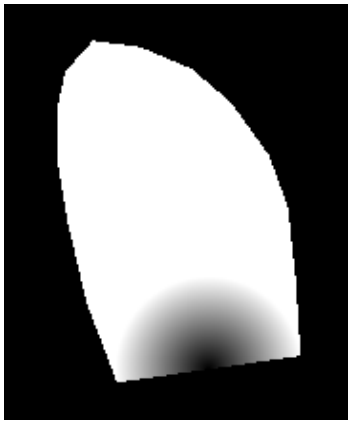


Figure 23: A mask used for training in which the lower edge of the mask has been gradually down-weighted to zero in all directions about the centre of the line. This will likely ensure that edge is not considered as a true boundary during training of the U-net.

Furthermore, if the initial contrast/non-contrast classifier is incorrect then that would produce an errant contour. Figure 24 shows an example of the incorrect contouring model (i.e. the contrast model) being deliberately applied to the same data as in Figure 21. It is clear that the contrast auto-contouring model has incorrectly detected

the bright region in the image as that containing the LV. This could become an issue if the framework is applied to further data.

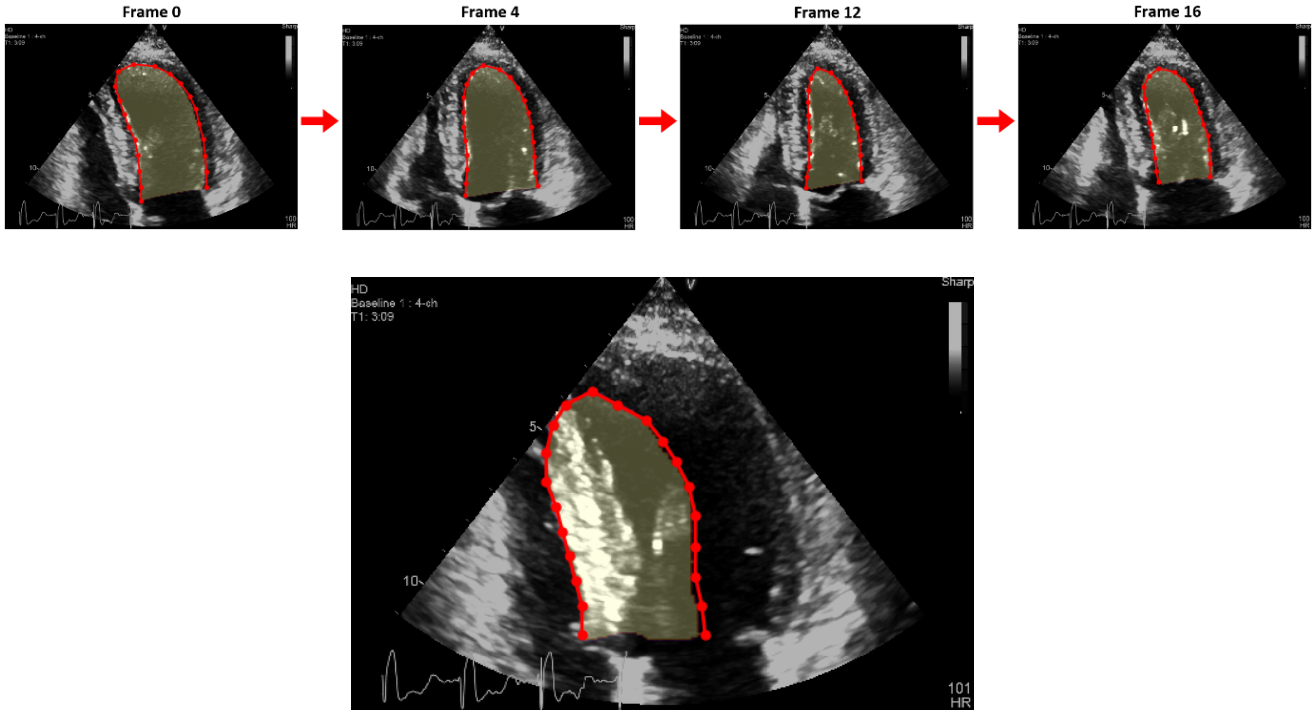


Figure 24: Taking the same dataset as in Figure 21 (above) and applying the contrast model yields the incorrect result shown here (below) in which the CNN has selected the bright region in the image as the region containing the LV.

The technology is also limited by image quality. Figure 25 shows an example of two datasets – one where the structure is present and one where the wall has disappeared. We can see that in the second image, the contour has incorrectly included an area greater than that of the LV as it has insufficient information to stay within the LV endocardium.

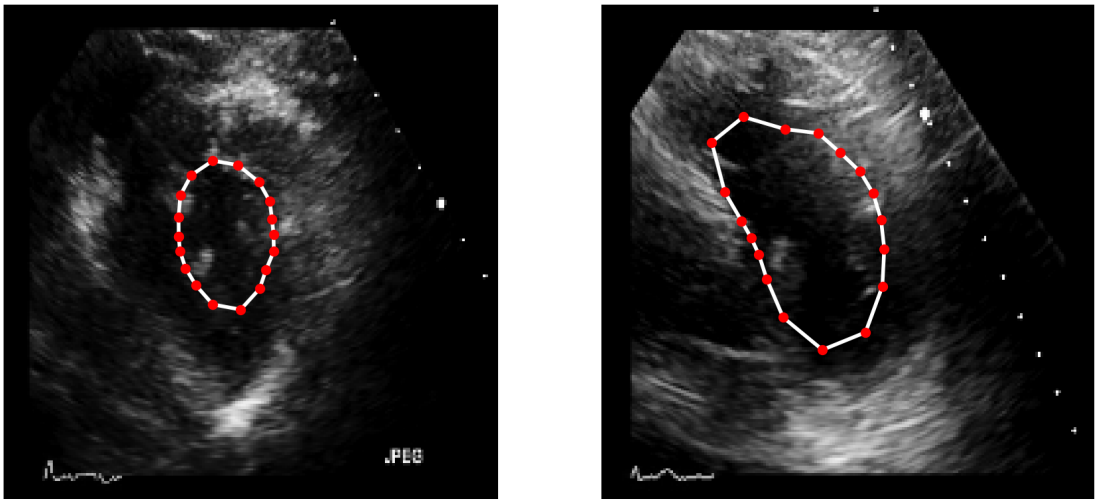


Figure 25: If the short-axis chamber walls are not fully present in most frames for non-contrast images (such as in the image on the left), the contouring struggles and ends up losing the structure of the LV walls, as seen in the image on the right.

Finally, the methodology of training and validating on manual contours in itself is flawed as manual contours often do not produce a perfect outline of the endocardium and can be highly variable (153).

In conclusion, image processing frameworks outlined in this work have proven to be very effective in terms of their DC performance and ability to generalise to validation data. More data would likely improve the contouring accuracies further and expanding on the technologies that I used to create the autocontours to incorporate further dimensionality, such as a 3D Unet may help improve it further. The results reached a high enough accuracy to apply this framework to the EVAREST dataset to extract features of left ventricular function which will provide further validation of this framework.

4.0 Comparison of Auto Ejection Fraction and Global Longitudinal Strain Against the Clinical Standard: A Validation Study

4.1 Introduction

In order to evaluate the Autocontouring framework, a formal validation study was performed to demonstrate that global longitudinal strain (GLS) and ejection fraction (EF) values calculated by the framework are accurate and substantially equivalent to the clinical standard (TomTec-Arena TTA2). The Autocontouring code was integrated into a proprietary software platform (EchoGo) by software development team at Ultromics Ltd. and will be hence referred to as EchoGo throughout this chapter.

There remains no established gold-standard method of assessing GLS on clinical data (140) and therefore no accepted ground truth. Thus, in order to assess the performance of EchoGo, GLS measurements acquired by a ratified (FDA cleared) medical software device, TomTec-Arena TTA2 (K150122) was used as the ground truth measurement.

The currently recommended 2D method of assessing left ventricular EF uses the biplane method of disks (modified Simpson's rule) to derive estimates of end diastolic and end systolic volume (EDV and ESV, respectively), from which LVEF can be calculated using the formula LVEF = (EDV-ESV)/EDV x 100% and presented as a percentage (5). This is the method employed by both EchoGo and TomTec-Arena TTA2. Therefore, to assess the performance, LVEF measures acquired by TomTec-Arena TTA2 software were used as the ground truth measurement.

Device repeatability testing aimed to assess measurement variability under the same conditions (i.e. replicate measurements taken on the same subject, same ultrasound

manufacturer, and same echocardiographer). EchoGo does not form part of the image acquisition phase and claiming standardisation of measurements across manufacturers was not be part of this validation.

This study aimed to compare GLS and LVEF calculated by EchoGo to those obtained by TomTec-Arena TTA2; and to assess the inter and intra observer variability on the automatically detected endocardial border of the LV.

4.2 Methods

4.2.1 Study Population

A retrospective dataset of 146 previously-acquired transthoracic echocardiograms (TTEs) were utilized for this performance testing. The TTEs were collected as part of the EVAREST clinical research study.

4.2.2 Data Selection

The performance testing data set consisted of TTEs selected from the EVAREST clinical research study. These data were independent and strictly segregated from algorithm training and development datasets. TTE DICOM studies were selected from the total dataset for this performance testing in order to include a range of physiological values of GLS and LVEF. Included in the dataset were data acquired across a representative sample of ultrasound manufacturers (GE and Philips) to verify that EchoGo performs acceptably across echocardiographic data obtained from multiple scanner platforms.

The dataset for the performance testing met the following criteria:

a. The datasets were obtained in accordance with standard guidelines from the European Association of Echocardiography and the American Society of Echocardiography using commercially available ultrasound machines;
b. All patients were examined in the left lateral decubitus position using the greyscale second-harmonic 2D imaging technique, with the adjustment of image contrast, frequency, depth, and sector width size for adequate frame rate and optimal LV border visualization;
c. Image acquisition was optimised to maximise LV areas while avoiding foreshortening of the LV, as recommended when acquiring images for LV volume measurement (5);
d. Views and stages were correctly labelled;
e. All required views were present;
f. Sufficient number of cardiac cycles were recorded;
g. Gain and contrast were appropriately adjusted for visualization;
h. LV were clearly visualized and in centre of image;
i. Image quality was sufficient for automated LV tracing to be performed.

4.2.3 Data Handling and Transfer

All TTE DICOM studies were deidentified and relabelled with a study specific ID. The study identification key was stored within Castor EDC 2019.1; an access restricted electronic database. Operators did not have access to this database. Test data was strictly segregated from algorithm training datasets and access was restricted.

4.2.4 Quality Control Process

Quality control was performed on all incoming echocardiographic studies by the Operators. This ensured that all information was present, the datasets had been obtained under the correct protocols, and all images captured the LV to a suitable degree of quality for assessment. Incoming datasets were only processed if they were of sufficient quality. The Autocontours required quality to be assessed by the trained Operators before any parameters were calculated.

Quality checks before the contouring confirmed the following:

- views and stages had been correctly labelled;
- all required views and stages were present;
- images contained appropriate end-diastolic or end-systolic frame;
- gain and contrast were appropriately adjusted for visualization;
- the LV was clearly visualized and centred;

Quality checks after contouring confirm the following:

- Auto-contouring provided valid traces of the LV.

4.2.5 Calculation of Global Longitudinal Strain and Ejection Fraction

Once quality control had been performed, the automated contour detection of the endocardium of the LV was reviewed and approved by the trained Operators. Identifying and outlining the LV is standard practice in echocardiography and present in other FDA cleared devices such as TomTec-Arena TTA2. In EchoGo, an auto-

contouring algorithm, developed using neural networks, places points around the LV that sufficiently capture the LV shape as described in the previous chapter.

For a patient, this outlining is detected for each set of acquired images at end-systole (ES) and end-diastole (ED) in the apical two-chamber (AP2) and four-chamber (AP4) views of the LV. These four contours were used in calculations for geometric parameters.

The following parameters were calculated for inclusion in the report:

1) Left ventricular ejection fraction (EF): standard calculation as included in TomTec-Arena TTA2 (K150122);
2) Global longitudinal strain (GLS): standard calculations as included in TomTec-Arena TTA2 (K150122).

4.2.6 Accuracy Assessment

DICOM TTEs were analysed by trained Operators using both EchoGo and TomTec-Arena TTA2 following the workflow described here:

1) The first Operator (Operator A) who had completed the training on the TomTec-Arena TTA2, was provided with the identified TTE DICOM dataset for validation (Dataset 01).

2) Operator A progressed through the worklist, selecting the designated DICOM studies sequentially to open with TomTec-Arena TTA2. Operator A applied automatic contours to the endocardium of the LV using TomTec-Arena TTA2; adjusted them manually, if necessary, and then calculated the GSL and LVEF

values according to the user manual. The results were recorded within Castor EDC 2019.1 an access restricted electronic database.

3) The same TTE DICOM studies (Dataset 01) were duplicated, relabelled so that the Operators were fully blinded, and the order of the studies randomised (Dataset 02). The second Operator, (Operator B) who had completed the training on EchoGo, was provided with this group of randomised DICOM datasets.

4) Operator B progresses through the Dataset 02, selecting the designated DICOM studies sequentially to open with EchoGo. The Operator reviewed and approved the automatically detected endocardial border of the LV on the images and from these regions of interest, GLS and LVEF measurements were calculated. The results were recorded within Castor EDC 2019.1 an access restricted electronic database.

4.2.7 Repeatability and Inter/Intra Observer Assessment

EchoGo provides fully automated GLS and LVEF measurements through contouring the LV, and as such does not require Operators to manually adjust the image contour. Following the automated detection of the LV contours, the Operators reviewed and approved the automatically detected endocardial border of the LV on the images. Therefore, inter and intra-observer variability was performed to assess the variability of the values derived by the automated contours.

1) A subset of TTE DICOM studies were randomly selected, duplicated, relabelled, and presented in a randomised order. This subset Dataset (03-B)

was given to Operator B to repeat analysis. As the studies were relabelled and randomised the Operator was fully blinded.

2) The Dataset 03-B was renamed and the order of studies randomised again, Dataset 03-C, and given to a third Operator (Operator C) who had completed the training on EchoGo.

3) Operators B & C then progressed through the Dataset 03-B & 03-C, selecting the designated DICOM studies sequentially to open with EchoGo. The Operators reviewed and approved the automatically detected endocardial border of the LV on the images and from these regions of interest, GLS and LVEF measurements were calculated. The results were recorded within Castor EDC 2019.1 an access restricted electronic database.

For repeatability to be assessed, the following conditions were met: the analysis of scans from the same patient and the same ultrasound machine acquired on the same day were repeated by the same Operator, using EchoGo under the same conditions.

A new dataset (Dataset 04) that consisted of images obtained from a subset of patients included in Dataset 02 (matched 1:1), and had been acquired by the same clinician using the same ultrasound machine, acquired on the same date, were given to Operator B and analysed under the same conditions as Dataset 02. A diagrammatic view of the performance testing data is shown below in Figure 26.

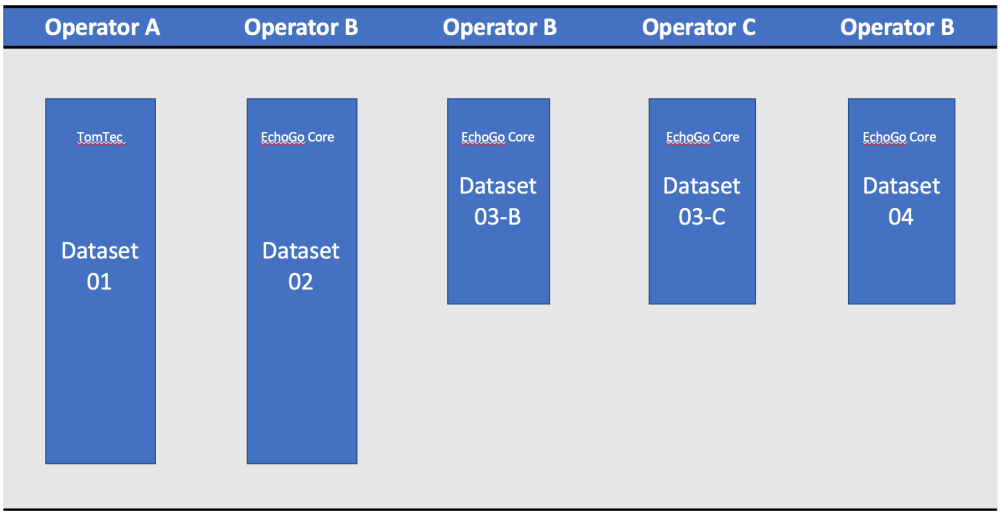


Figure 26: Diagrammatic overview of operators and datasets

4.3 Results

4.3.1 Demographics

The demographics of the patient cohort are summarized in Table 7.

Measure	Minimum	25% Quantile	50% Quantile	75% Quantile	Maximum
N=146					
Height (m)	1.47	1.62	1.70	1.76	1.95
Weight (kg)	44.0	69.0	76.5	87.0	130
BMI (kg/m^2)	19.1	24.0	26.7	29.4	50.8
Age	21.4	58.1	66.0	73.6	88.3

Table 7: Demographics of patient population

4.3.2 Quality assurance exclusions

Of the 146 recruited patients, 10 were excluded due to software loading failure and missing baseline views. The final dataset contained 136 patients for processing in both TomTec-Arena TTA2 and EchoGo. Table 8 summarises the scans rejected

following quality assurance review. A total of 113 contours from EchoGo passed quality assurance and are described in Table 9.

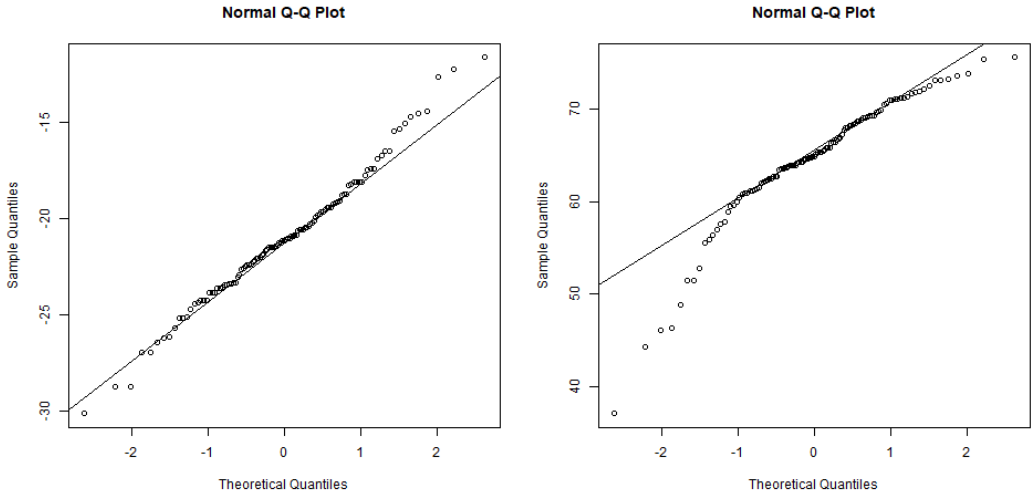
QA rejections	QA performed by	Number of scans analysed	Missing images	Contour incorrect	Apex cut off	Number after exclusions
Accuracy – TomTec-Arena TTA2	Operator A	136	0	0*	0	136
Accuracy – EchoGo	Operator B	136	2 (1.5%)	16 (11.8%)	5 (3.7%)	113 (83.1%)
Repeatability – Baseline 1	Operator B	53	0	4 (7.5%)	3 (5.7%)	46 (86.8%)
Repeatability – Baseline 2	Operator B	53	0	11 (20.8%)	3 (5.7%)	39 (73.6%)
Inter- and intra-variability	Operator B	10	0	2 (20%)	0	8 (80%)
Inter-variability	Operator C	10	0	3 (30%)	0	7 (70%)
Intra-variability	Operator B	10	0	2 (20%)	0	8 (80%)

Table 8: Summary of the rejected scans.

Measure	Range (%)	Mean (%)	Standard Deviation (%)	N
Ejection Fraction	29.0-79.0	62.0	9.5	113
Global Longitudinal Strain	-6.3- -28.2	-17.3	4.0	113

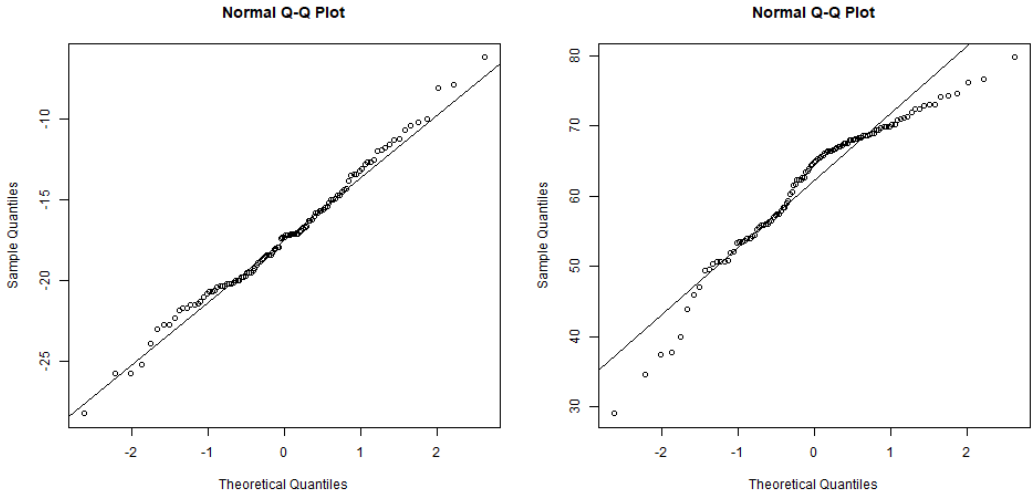
Table 9: Summary of the EchoGo scans that passed quality assurance of the contours

Normality of data was assessed using Q-Q plots (Figures 27 and 28). GLS demonstrated a normal distribution while EF appeared to be non-normal. However, Bland-Altman and RMSE analysis was performed in keeping with previously published reports for GLS and EF measures to allow for comparison and compliance with the performance testing plan.



Global Longitudinal Strain on EchoGo *Ejection Fraction on EchoGo*

Figure 27: Q-Q plots of GLS and EF measurements calculated on EchoGo.



Global Longitudinal Strain on TomTec *Ejection Fraction on TomTec*

Figure 28: Q-Q plots of GLS and EF measurements calculated on TomTec Arena TTA2.

4.3.3 Accuracy Results

The Accuracy for EF and GLS are summarised in Tables 10, 11 and Figures 29, 30. An RMSE of 8.05% was obtained for EF accuracy and GLS produced an RMSE of

5.10%. The RMSE and bias for EF was in acceptable limits (<10%, <5% respectively) but for GLS we did not meet our acceptance criteria (<5%, >-3% respectively) according to previous literature and FDA submissions (154,155). We also assessed results by Ultrasound manufacturer which yielded similar results in terms of RMSE as summarised by Tables 12 and 13 and Figures 31 and 32.

Measure	Bias (%)	Lower LoA (%)	Upper LoA (%)	RMSE	N
Ejection Fraction	2.6	-12.4	17.6	8.07	113
Global Longitudinal Strain	-3.7	-10.6	3.1	5.11 (3.84)*	113

**Bias adjusted*

Table 10: GLS and EF measurements comparison between TomTec Arena TTA2 vs. EchoGo

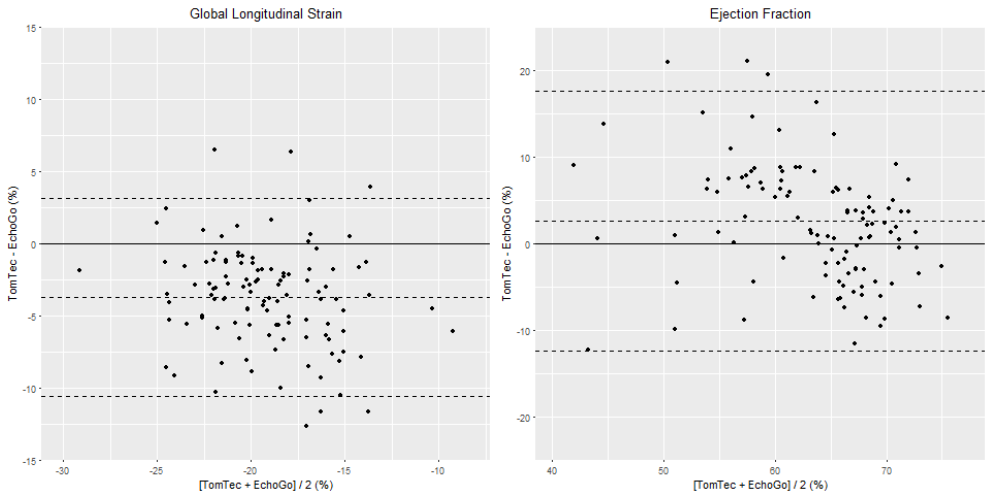


Figure 29: Bland-Altman analysis of EchoGo Accuracy

Complete Dataset n=113			
Ejection Fraction			
Measure	**Estimate (%)**	**Lower 95% CI (%)**	**Upper 95% CI (%)**
Bias	2.6	1.2	4.0
Upper LoA	17.6	15.2	20.0
Lower LoA	-12.4	-14.8	-10.0
RMSE	8.1	--	--
Global Longitudinal Strain			
Measure	**Estimate (%)**	**Lower 95% CI (%)**	**Upper 95% CI (%)**
Bias	-3.7	-4.4	-3.1
Upper LoA	3.1	2.0	4.2
Lower LoA	-10.6	-11.7	-9.5
RMSE	5.1	--	--

Table 11: Summary of Accuracy Results

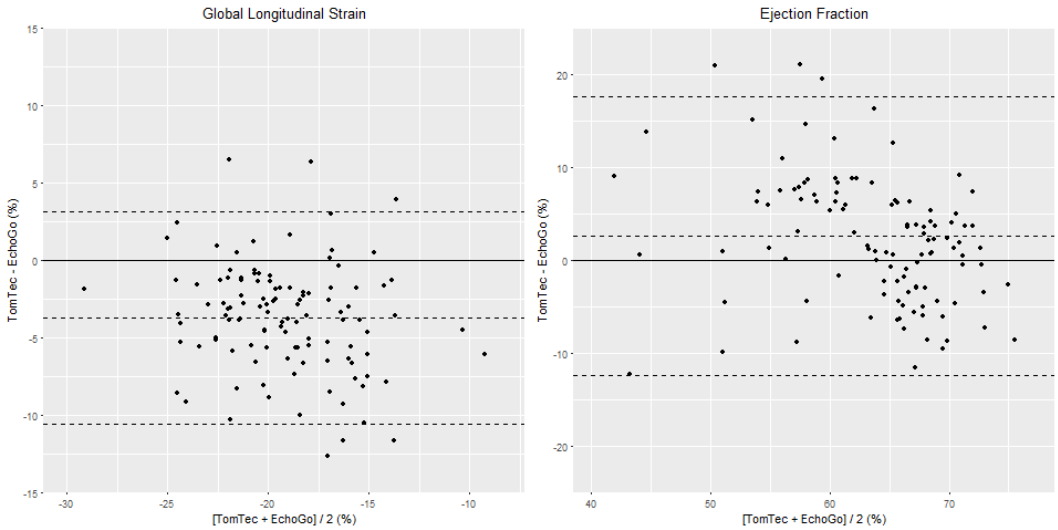


Figure 30: Bland-Altman analysis of all data

GE n=53			
Ejection Fraction			
Measure	**Estimate (%)**	**Lower 95% CI (%)**	**Upper 95% CI (%)**
Bias	2.5	0.3	4.8
Upper LoA	18.4	14.6	22.3
Lower LoA	-13.4	-17.2	-9.5
RMSE	8.2	--	--
Global Longitudinal Strain			
Measure	**Estimate (%)**	**Lower 95% CI (%)**	**Upper 95% CI (%)**
Bias	-4.0	-5.0	-3.1
Upper LoA	2.7	1.1	4.3
Lower LoA	-10.7	-12.4	-9.1
RMSE	5.3	--	--

Table 12: Accuracy results of data obtained by GE Ultrasound Device Manufacturer

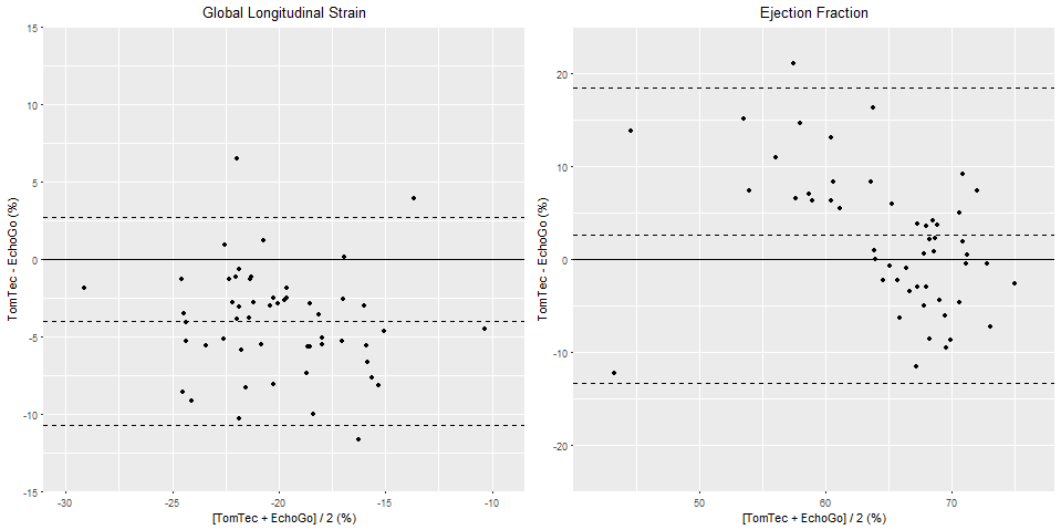


Figure 31: Bland-Altman analysis of data obtained by GE Ultrasound Device Manufacturer.

Philips n=60			
Ejection Fraction			
Measure	**Estimate (%)**	**Lower 95% CI (%)**	**Upper 95% CI (%)**
Bias	2.9	0.9	5.0
Upper LoA	18.5	14.9	22.0
Lower LoA	-12.6	-16.1	-9.0
RMSE	8.4	--	--
Global Longitudinal Strain			
Measure	**Estimate (%)**	**Lower 95% CI (%)**	**Upper 95% CI (%)**
Bias	-3.5	-4.5	-2.5
Upper LoA	3.9	2.2	5.6
Lower LoA	-10.9	-12.5	-9.2
RMSE	5.1	--	--

Table 13: Accuracy results of data obtained by Philips Ultrasound Device Manufacturer

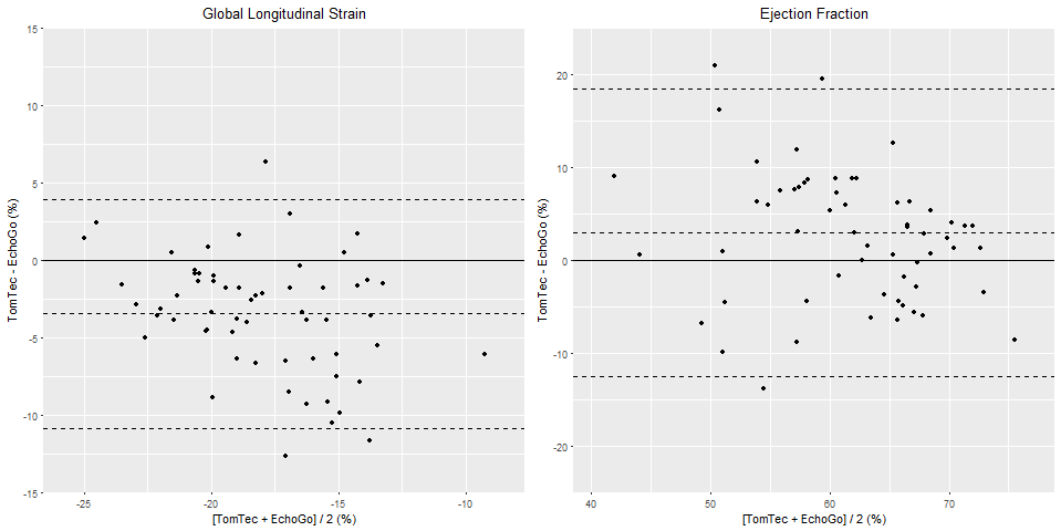


Figure 32: Bland-Altman analysis of data obtained by Philips Ultrasound Device Manufacturer.

4.3.4 Precision Results

The EF precision results demonstrated a mean bias of -1.4% and 95% LoA of ±14.3% and an RMSE of 7.36 was obtained for repeatability. GLS results demonstrated a mean bias of 0.4% and 95% LoA of ±7.9% and an RMSE of 2.76 was obtained for repeatability. These results are summarised in Table 14 and Figure 33.

Measure	Bias (%)	Lower LoA (%)	Upper LoA (%)	RMSE	N
Ejection Fraction	-1.4	-15.6	12.9	7.36	39
Global Longitudinal Strain	0.4	-4.9	5.9	2.76	39

Table 14: EchoGo repeatability results for Operator 1 vs. Operator 2

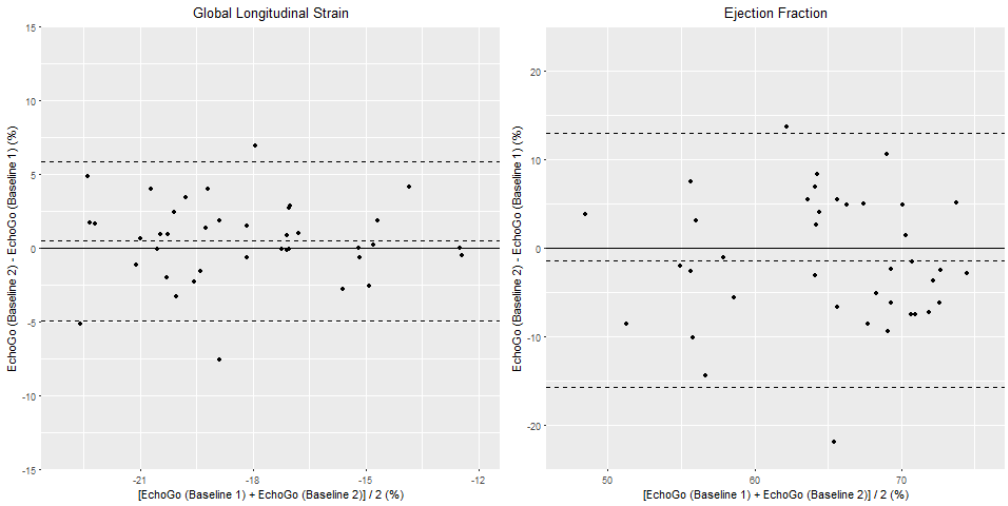


Figure 33: Bland-Altman analysis of EchoGo Repeatability

4.3.4.3 Inter- and intra-observer variability

There were 10 patients recruited for the Intra observer variability study. Out of these 10 patients, the Operator excluded the same 2 patients on both occasions. This left 8 sets of contours available for comparison.

Cohens Kappa Analysis of operator agreement:

- Number of observed agreements: 10 (100.00% of the observations)
- Number of agreements expected by chance: 6.8 (68.00% of the observations)
- Kappa= 1.000
- SE of kappa = 0.000
- 95% confidence interval: From 1.000 to 1.000

An RMSE of 0% was obtained for EF, which was in keeping with the prespecified acceptance criteria of 0%. An RMSE of 0% was obtained for GLS, which was in keeping with the prespecified acceptance criteria of 0%.

For inter-observer variability 10 patients were analysed, the first reader excluded 2 patients, and the second reader excluded 3 patients. The patients excluded by the first reader overlapped with the second reader, so this left 7 sets of contours available for comparison.

Cohens Kappa Analysis of reader agreement:

- Number of observed agreements: 9 (90.00% of the observations)
- Number of agreements expected by chance: 6.2 (62.00% of the observations)
- Kappa= 0.737
- SE of kappa = 0.241
- 95% confidence interval: From 0.265 to 1.000

An RMSE of 0% was obtained for EF, which was within the prespecified acceptance criteria of 0%. An RMSE of 0% was obtained for GLS, achieving the prespecified acceptance criteria of 0%.

4.4 Discussion

Building on the Autocontouring framework we have demonstrated we can extract clinically relevant measurements equivalent to the current clinical standard software TomTec. EF RMSE was within our acceptance criteria, GLS RMSE, however, did not meet our acceptance criteria.

The reason for not achieving the acceptance criteria for GLS is due to a known systematic bias in the TomTec measurements (140). Our validation testing

demonstrated a mean difference between EchoGo GLS and TomTec Arena TTA2 GLS of -3.7%. This finding is consistent with a large inter-vendor agreement study (140) that has previously demonstrated a significant bias of -2.45±2.10% (n=1302 echocardiographic examinations) between 8 vendor GLS measurements (Hitachi-Aloka, Esaote, GE, Philips, Samsung, Siemens, Toshiba, Epsilon) and TomTec GLS measurement. Post hoc analysis taking into account this published difference indicated the difference between EchoGo GLS measurements and TomTec GLS measurements should fall between -4.6% to -0.4%. Therefore the -3.7% difference was considered acceptable and in keeping with clinical expectations for inter-vendor GLS measurement agreement. GLS comparison analysis produced an RMSE of 5.10%, which when accounting for the expected bias described in Farsalinos et al. (2015), equates to a RMSE of 3.84% (140).

With regard to EF a previous inter-vendor comparison study of 63 subjects using two different post-processing software (156) demonstrated a negligible bias between image processing vendors (-0.16%). Thus, an expected bias was not accounted for in our validation testing and Root Mean Square Error (RMSE) calculation was not adjusted for published bias.

Our repeatability results for both EF and GLS met our acceptance criteria. Previous reports of EF suggested repeatability should produce a mean bias of approximately -1.2% with a 95% distribution of ±15.1% (125,157). This is in keeping with our validation results demonstrating a mean bias of -1.4% and 95% LoA of ±14.3%. An RMSE of 7.36 was obtained for EF repeatability. Previous reports of GLS suggested a repeatability with mean bias of -0.87% with a 95% distribution of ±8.5% was

expected (158). This was in keeping with our validation results demonstrating a mean bias of 0.4% and 95% LoA of ±7.9%. An RMSE of 2.76 was obtained for GLS. Furthermore, due to the nature of the contouring being fully automated intra and inter observer agreement was very high, therefore there was no possibility of variability. The only possible source of variability is in which images passed QA which was also very high with operators disagreeing on only one image out of 10.

Many previous studies have used AI to derive functional measurements (53,54,155,159). In one study, Cannesson et al. compared an AI derived auto EF software with Simpsons biplane measurements and CMR. This was validated in 200 patients. However, in 25% of these the software required manual editing of the contours. The EF results showed good correlation with both manual tracings and CMR (r=>0.9). However this study was limited in that it is not clear how the training data and validation data was segregated and the same clinicians that were manually editing the contours from the auto ejection fractions were also producing the manual ejection fractions (53).

Perhaps the most comparable study, Asch et al. used a deep learning base Automated EF software (EchoMD, Baylabs) and compared to Simpsons biplane measurements by 3 expert clinicians. In this study they achieved an RMSE of 8.3% which is higher than the RMSE we achieved in this study. Interestingly, this study also found that RMSE between clinicians measuring EF was greater than 9% (155). Furthermore, the EchoMD platform is not fully automated, it is only applicable to resting echo images and is yet to be tested on contrast images and stress images. Therefore, our platform demonstrates more accurate measurements and is

applicable to a wider array of echocardiograms, while providing a fully automated approach that means no variability is introduced by operators.

Although we have demonstrated good results, and significantly improved on the current technologies, there are limitations in this study due to there being no gold standard to measure against (160). Furthermore, we are comparing a fully automated technology (EchoGo) versus a semi-automated technology (TomTec). In addition, we know from a comparison study of strain software that TomTec has a known bias, which we did not account for prior to starting the study or when designating our acceptance criteria.

In conclusion, the results provide confidence that we can generate significant improvements on current technologies that are robust enough to take forward for use in further prediction work. The demonstration of consistent measures of EF and GLS from the autocontours provides confidence that they are correctly identifying and segmenting the left ventricle. This means additional novel features that we extract from the contours to create a classifier are likely to be representative of left ventricular geometry to support the diagnosis of coronary artery disease.

5.0 Training a Machine Learning Classifier to Predict Coronary Artery Disease

5.1 Introduction

CAD is a major cause of morbidity and mortality worldwide, affecting 2.3 million people in the UK (92). Whilst CAD can be diagnosed through a variety of imaging modalities, stress echocardiography is the most commonly used technique, used in over 50% of hospitals in the UK (161), and has the benefit of being a relatively safe and inexpensive test, free from ionising radiation (162).

Stress echocardiography allows the visualisation of myocardial wall motion across a range of heart rates, achieved either via an exercise protocol or drug infusion (e.g. dobutamine) (163). As heart rate increases, so should myocardial contractility. Should flow-limiting CAD be present (cardiac ischaemia), the region(s) of myocardium supplied by the obstructed vessel(s) exhibit abnormal contraction, known as regional wall motion abnormalities (RWMAs). Each myocardial segment is assigned a score by the interpreting cardiologist, depending on the level of contractile impairment (1= normal, 2= hypokinetic, 3= akinetic, 4= dyskinetic and 5= aneurysmal). This is used in combination with the number of myocardial segments exhibiting a RWMA, to determine the severity of cardiac ischaemia and whether any intervention is indicated (163).

Currently, RWMA scores are assigned based on the reviewing cardiologist's opinion of this semi-qualitative data. Whilst operator experience plays an important role in the correct interpretation of stress echocardiography (164), studies have demonstrated the inherently subjective nature of assigning wall motion scores (WMS) leading to high inter-observer variation (7). This is especially common in

patients with poor echo windows or those with poor endocardial definition. Whilst the limitation of poor visualisation can be overcome with the use of contrast, the subjectivity of assigning WMS remains (163).

As with all clinical tests stress echocardiography is not 100% sensitive or specific in the detection of CAD. Within leading academic centres, all studies, irrespective of size, have produced similar values with a sensitivity and specificity of 81 and 82%, respectively (64). This is similar to the highest values obtained from other functional imaging tests, such as myocardial perfusion scintigraphy (165) and stress CMR (166). This means that a significant proportion of patients receive either false positive or false negative results. These incorrect results could lead to unnecessary procedures and treatments or even the withholding of necessary, and potentially lifesaving interventions.

As a result of the limitations described above, there is great interest in the development of robust imaging biomarkers which can reliably and reproducibly provide a predictive index for prognostically significant CAD. This removal of the subjectivity of wall motion scoring, combined with powerful machine learning techniques, would allow for a more accurate risk stratification and hence guide the physician towards the most appropriate therapies (167–169).

With these limitations in mind, a diagnostic support tool was developed that quantifies myocardial features and uses machine learning to combine the features against three-month outcomes to generate a prediction for prognostically significant CAD.

5.2 Methodology

5.2.1 Study Population

Patients were recruited into the EVAREST study as outlined in chapter 2 and the datasets obtained were used to generate the training data. All patients over the age of 18 years were approached unless they had had a myocardial infarction in the preceding seven days or unstable cardiopulmonary disease. All aspects of data and sample collection were approved by the Oxfordshire Research Ethics Committee (14/SC/1437) and all participants gave written informed consent.

5.2.2 Outcomes Definition

Following the collection of one-year outcome data, it was considered by the EVAREST Adjudication Committee at the Oxford Research Echocardiography Core Laboratory, University of Oxford. This committee comprised of a panel of cardiologists who reviewed the follow-up data and assigned outcomes for each participant based on their clinical status (see Figure 34). The Adjudication Committee was blinded to the result of the stress echocardiogram. A positive outcome indicated the presence of clinically significant coronary artery disease (i.e. a stenosis > 70% luminal diameter in one of the main epicardial vessels or > 50% in the Left Main Stem, an FFR < 0.85, decision to revascularise, acute coronary event or cardiac-related death).

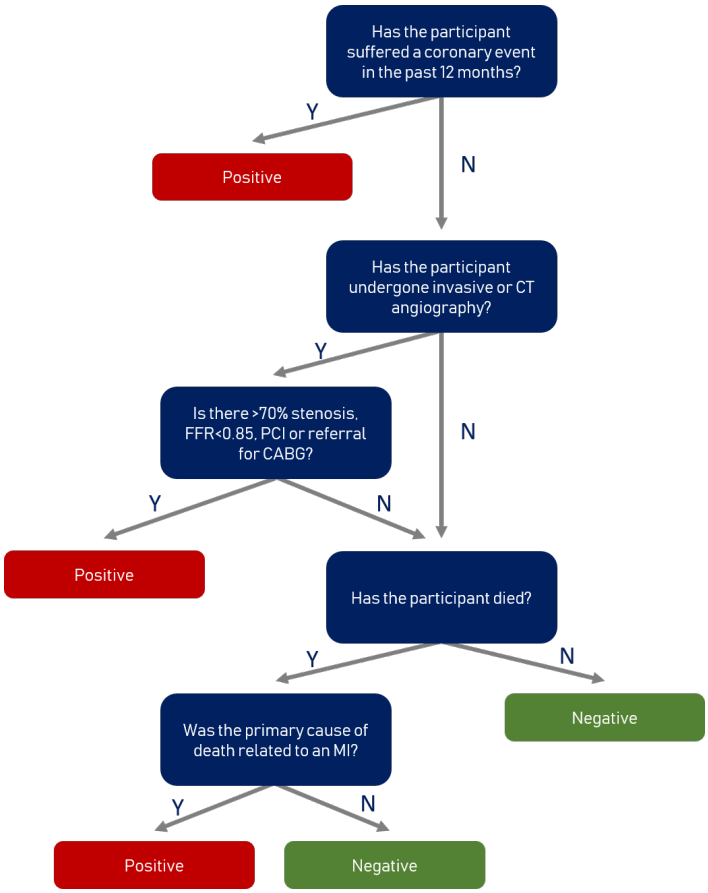


Figure 34: Decision tree demonstrating the outcome adjudication process.

5.2.3 Auto-contouring of left ventricular and cycle extraction

The auto-contouring framework from chapter 3 was used to define the LV border. All three auto-contouring models were utilised: (1) contrast apical 2 and 4 chamber views; (2) non-contrast apical 2 and 4 chamber views and; (3) parasternal short-axis view.

5.2.4 Biomarkers and feature engineering

Biomarkers such as ejection fraction, and global longitudinal strain are established measurements for the assessment of left ventricular function and will support the

diagnosis of CAD. While these measurements have been calculated and validated in the previous chapter, novel biomarkers that describe the shape of the contours were also extracted to help with the CAD prediction. Prior to this work, these novel biomarkers had not traditionally been considered for use in echocardiography and some were entirely novel conceptions. Their design and development therefore were captured as part of the IP generation within this research and some form part of the patent families filed during the DPhil.

They were designed taking into account clinical expertise in cardiac motion during stress studies and through discussion with other members of the research team about geometric parameters it was feasible to extract from LV contours. Both global and regional biomarkers were calculated. The feature dataset was further extended through feature transformation which allowed additional utilisation of information, calculating difference between the rest and equivalent stress biomarker, and the ratio between end-systolic and the equivalent end-diastolic biomarker. In addition, changes due to time were considered, calculating the rate of global and regional biomarkers based on the frame rate. This computation of extensive geometric, mechanical, and shape biomarkers, allowed for comprehensive extraction of information from the echocardiogram images utilising the same autocontouring framework from chapter 3. I have detailed the computation of some of these additional key features below.

5.2.4.1 Area

The area, A, for a polygon with number of sides n comprised of a set of ordered points $\{(x_i, y_i): i \in \{1, 2, \dots, n\}\}$ can be found using the Shoelace Formula, expressed as:

$$A = \frac{1}{2}\left|\sum_{i=1}^{n-1} x_i y_{i+1} + x_n y_1 - \sum_{i=1}^{n-1} x_{i+1} y_i - x_1 y_n\right|. \quad (2)$$

Calculate:

The area of the end-systole (A^S) or end-diastole (A^D) contours using equation (2) for the sets of points c^S or c^D respectively (Figure 35).

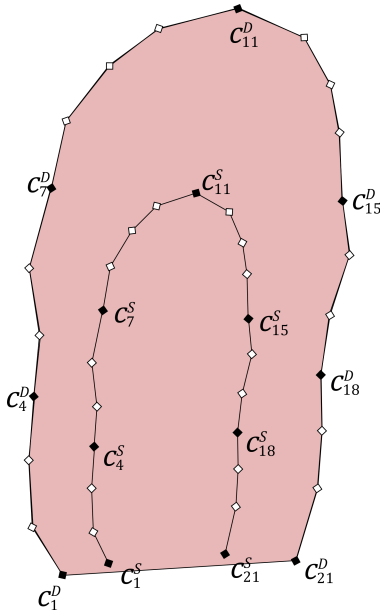


Figure 35: Example shown is A^D

The area of the segments between the end-systole and end-diastole contours as shown in Figure 36, using equation (2) for the sets of points:

- $\{c_1^D, c_2^D, c_3^D, c_4^D, c_4^S, c_3^S, c_2^S, c_1^S\}$
- $\{c_4^D, c_5^D, c_6^D, c_7^D, c_7^S, c_6^S, c_5^S, c_4^S\}$

- $\{c_7^D, c_8^D, c_9^D, c_{10}^D, c_{11}^D, c_{11}^S, c_{10}^S, c_9^S, c_8^S, c_7^S\}$
- $\{c_{11}^D, c_{12}^D, c_{13}^D, c_{14}^D, c_{15}^D, c_{15}^S, c_{14}^S, c_{13}^S, c_{12}^S, c_{11}^S\}$
- $\{c_{15}^D, c_{16}^D, c_{17}^D, c_{18}^D, c_{18}^S, c_{17}^S, c_{16}^S, c_{15}^S\}$
- $\{c_{18}^D, c_{19}^D, c_{20}^D, c_{21}^D, c_{21}^S, c_{20}^S, c_{19}^S, c_{18}^S\}$

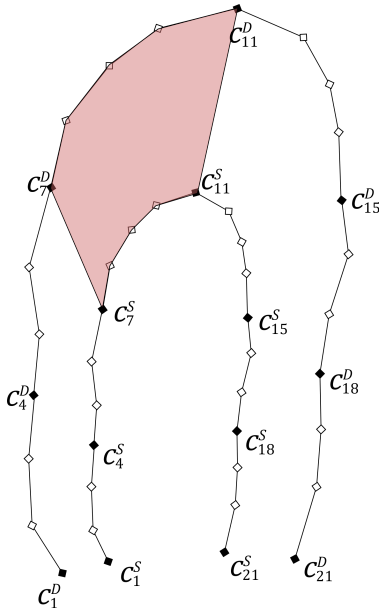


Figure 36: Example shown is $\{c_7^D, c_8^D, c_9^D, c_{10}^D, c_{11}^D, c_{11}^S, c_{10}^S, c_9^S, c_8^S, c_7^S\}$.

5.2.4.2 Rectangularity

The rectangularity, R, is the ratio of the enclosing area of a polygon A_{poly} (computed using the Shoelace Formula in equation (2) and the area of its minimum enclosing rectangle.

The area of the minimum enclosing rectangle A_{rect} for a polygon with number of sides n comprised of a set of ordered points $\{(x_i, y_i), i = 1, 2, \dots, n\}$ is found by rotating the set of points such that a rotation by angle α forms a set

$[(x'_{\alpha,i}, y'_{\alpha,i}), i = 1, 2, \ldots, n]$. The minimum enclosing rectangle is found by finding the such that:

$$A_{rect} = \text{argmin}_{\alpha}\{(x'_{\alpha,i,max} - x'_{\alpha,i,min}) \times (y'_{\alpha,i,max} - y'_{\alpha,i,min})\}.$$

This is implemented by rotating the contour through π/2 radians in increments of π/2000 radians, computing the bounding area aligned with x and y axis at each increment, and then taking the smallest area of all these (Figure 37).

The rectangularity R is then expressed as:

$$R = \frac{A_{poly}}{A_{rect}}.$$

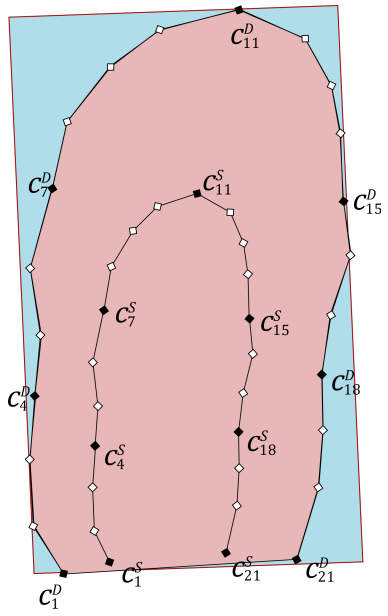


Figure 37: Example shown is for diastole (red area / red and blue area).

5.2.4.3 Solidity

The solidity, S, is the ratio of the enclosing area of a polygon A_{poly} (computed using the Shoelace Formula in equation (2)) and the area of its convex hull A_{ch} (Figure 38). The convex hull $\boldsymbol{C}$ is found by assigning a weight w_i to all the points p_i in set $\boldsymbol{P}$ such that all the weights are positive and sum to a value of one:

$$\boldsymbol{C} = \left\{ \sum_{i=1}^{|\boldsymbol{P}|} w_i p_i \,|(w_i \geq 0, \forall i) \wedge \sum_{i=1}^{|\boldsymbol{P}|} w_i = 1 \right\}.$$

The solidity S is then:

$$S = \frac{A_{poly}}{A_{ch}}.$$

For a convex shape, this will always give a value of 1.

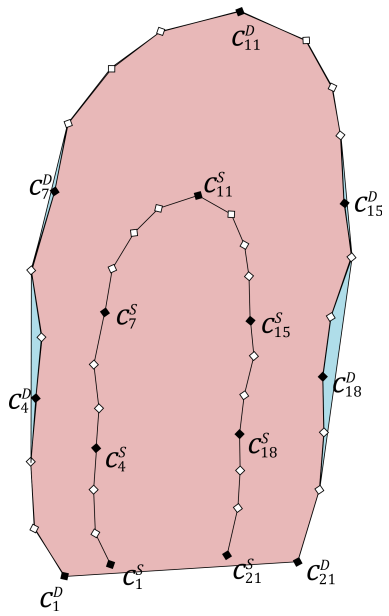


Figure 38: Example shown is for diastole (red area / red and blue area).

5.2.4.4 Principal Transformation

The principal transformation is proportional to the average Manhattan displacement between end-systole and end-diastole contour points, and is formulated as:

$$T = \frac{1-\sqrt{2}}{2n}\left(\sum_{i=1}^{n}|x_i^S - x_i^D| + \sum_{i=1}^{n}|y_i^S - y_i^D|\right).$$

5.2.4.5 Curvature

For a set of three points, we can calculate the curvature of the contour by using the Menger curvature formula, whereby,

$$c(x,y,z) = \frac{1}{R} = \frac{4A}{|x-y||y-z||z-x|}$$

Where x, y and z are the three points, A is the area enclosed by the triangle xyz, and R is the radius of the circle that fits through those three points (Figure 39).

The curvature is calculated at ED and ES for every view and for every triple of consecutive labelled points:

- $\{c_1^D, c_4^D, c_7^D\}$, $\{c_4^D, c_7^D, c_{11}^D\}$, $\{c_7^D, c_{11}^D, c_{15}^D\}$, $\{c_{11}^D, c_{15}^D, c_{18}^D\}$, $\{c_{15}^D, c_{18}^D, c_{21}^D\}$
- $\{c_1^S, c_4^S, c_7^S\}$, $\{c_4^S, c_7^S, c_{11}^S\}$, $\{c_7^S, c_{11}^S, c_{15}^S\}$, $\{c_{11}^S, c_{15}^S, c_{18}^S\}$, $\{c_{15}^S, c_{18}^S, c_{21}^S\}$

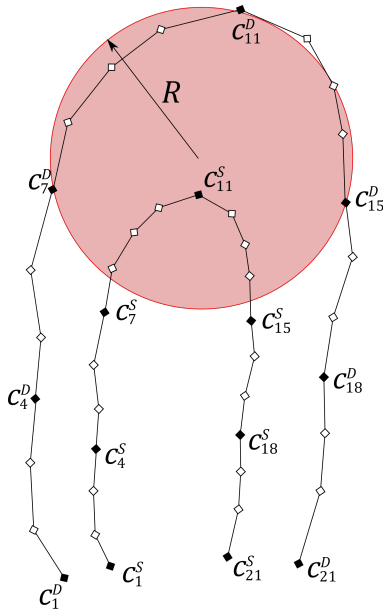


Figure 39: Example shown is for the set of points $\{c_7^D, c_{11}^D, c_{15}^D\}$.

5.2.4.6 Tortuosity

The tortuosity of a path is the length of the path divided by the straight-line length of that path (Figure 40).

$$\boxed{tortuosity = \tau = \frac{L}{C}}$$

We calculate the tortuosity for different sets of points:

- For segments: $\boxed{\{c_1^D, c_2^D, c_3^D, c_4^D\}}$, $\boxed{\{c_4^D, c_5^D, c_6^D, c_7^D\}}$, $\boxed{\{c_7^D, c_8^D, c_9^D, c_{10}^D, c_{11}^D\}}$, $\boxed{\{c_{11}^D, c_{12}^D, c_{13}^D, c_{14}^D, c_{15}^D\}}$, $\boxed{\{c_{15}^D, c_{16}^D, c_{17}^D, c_{18}^D\}}$, $\boxed{\{c_{18}^D, c_{19}^D, c_{20}^D, c_{21}^D\}}$ (also for end-systole)
- For walls: $\boxed{\{c_1^D, c_2^D, c_3^D, c_4^D, c_5^D, c_6^D, c_7^D, c_8^D, c_9^D, c_{10}^D, c_{11}^D\}}$, $\boxed{\{c_{11}^D, c_{12}^D, c_{13}^D, c_{14}^D, c_{15}^D, c_{16}^D, c_{17}^D, c_{18}^D, c_{19}^D, c_{20}^D, c_{21}^D\}}$ (also for end-systole)
- For the entire contour or for labelled points only

These features are calculated for each view, at both ED and ES.

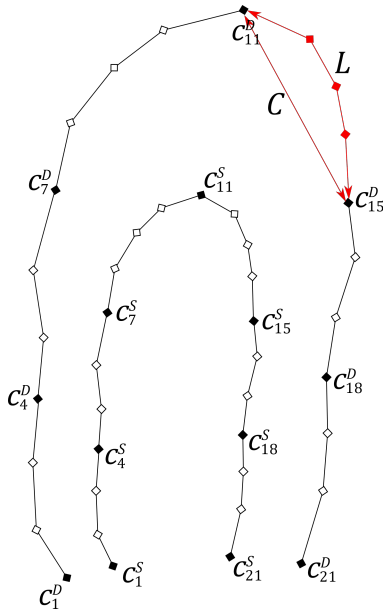


Figure 40: Example shown is for the set of points $\{c_{11}^D, c_{12}^D, c_{13}^D, c_{14}^D, c_{15}^D\}$.

5.2.4.7 Circularity ratio, circle variance, ellipse variance and eccentricity

The circularity ratio, circle variance, ellipse variance and eccentricity are calculated according to the methods previously described in literature (170).

The circularity ratio is the ratio of area to perimeter squared:

$$circ_{ratio} = \frac{A}{P^2}$$

Note that for a circle, the circularity ratio is the constant $1/4\pi$.

The circle variance is a measure of the roundedness of an object, calculated as the coefficient of variation of the distance between the centre of gravity of the contour mask, and the contour points (Figure 41).

First, calculate the centre of gravity as $g = (g_x, g_y)$:

$$g_x = \frac{1}{6A_s} \sum_{i=1}^{n} (x_i + x_{i+1})(x_i\, y_{i+1} - x_{i+1} y_i)$$

$$g_y = \frac{1}{6A_s} \sum_{i=1}^{n} (y_i + y_{i+1})(x_i\, y_{i+1} - x_{i+1} y_i)$$

Where A_s is the polygon's signed area,

$$A_s = \frac{1}{2}\sum_{i=1}^{n}(x_i\, y_{i+1} - x_{i+1} y_i)$$

The circle variance is then calculated,

$$V_i = \begin{pmatrix} x_i - g_x \\ y_i - g_y \end{pmatrix}$$

$$cv_i = \sqrt{V_i^T . V_i}$$

$$circle_{var} = \frac{\text{stddev}(cv)}{\text{mean}(cv)} = \frac{1}{\overline{cv}}\sqrt{\frac{\sum_{i=1}^{n}(cv_i - \overline{cv})^2}{n}}$$

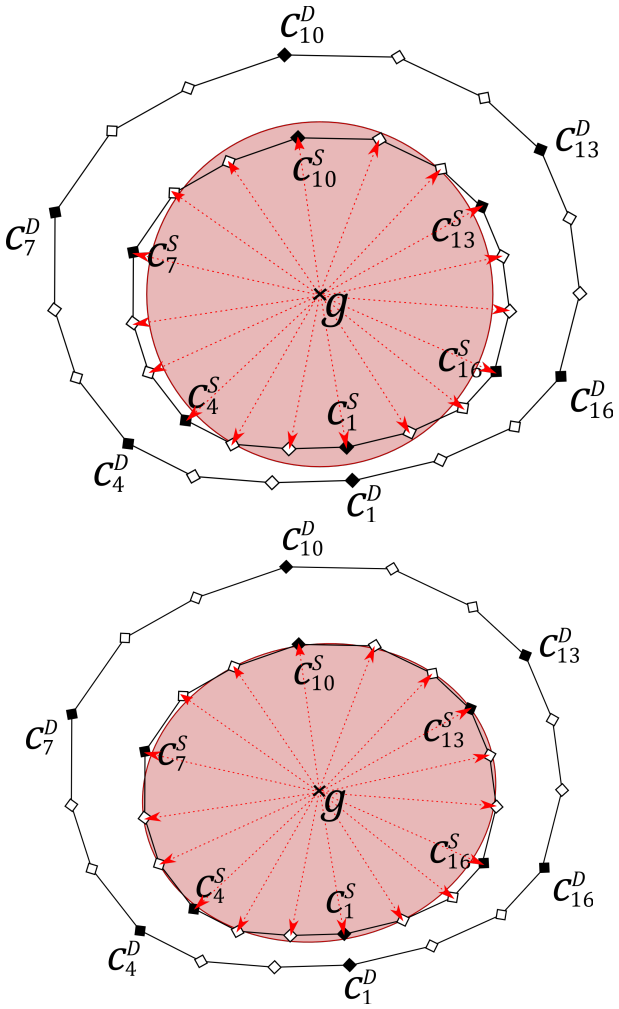


Figure 41: Example for end-systole

The ellipse variance is a measure of the roundedness of an object, where instead of comparison to a circle, the comparison is to an ellipse with the same covariance matrix as the shape. The covariance matrix is calculated as:

$$C_{ellipse} = \frac{1}{n}\sum_{i=1}^{n} V_i . V_i^T$$

Then the ellipse variance is calculated

$$ev_i = \sqrt{V_i^T . C_{ellipse}^{-1} . V_i}$$

$$ellipse_{var} = \frac{\text{stddev}(ev)}{\text{mean}(ev)} = \frac{1}{\overline{ev}}\sqrt{\frac{\sum_{i=1}^{n}(ev_i - \overline{ev})^2}{n}}$$

The eccentricity is calculated as the ratio of the smallest eigenvalue of the covariance matrix to the largest eigenvalue of the covariance matrix (Figure 42).

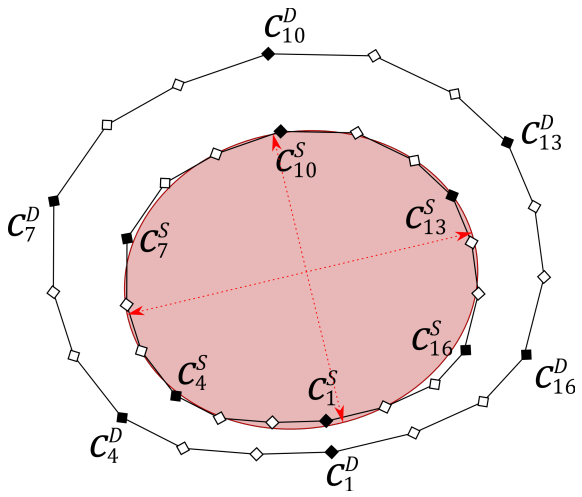


Figure 42: Example shown is eccentricity

5.2.4.8 Rotation

The polar coordinate of every labelled point is calculated using the centre of gravity of the contour as the origin. Let $\theta^S = \{\theta_1^S, \theta_2^S, \dots, \theta_n^S\}$ be the set of end-systole polar angles for labelled points (where $n = 7$ for 2-chamber and 4-chamber views, and $n = 6$ for short-axis views). Similarly let $\theta^D = \{\theta_1^D, \theta_2^D, \dots, \theta_n^D\}$ be the set of end-diastole polar angles.

For example, $\theta_1^S = \arctan\left(\frac{y_1^S - g_y}{x_1^S - g_x}\right)$, where (x_1^S, y_1^S) are the coordinates of the first labelled

end-systole point, and (g_x, g_y) are the coordinates of the end-systole centre of gravity.

For each triplet of labelled points, and for all labelled points, calculate the mean difference in angle from end-diastole to end-systole to approximate the rotation of the contour globally or regionally. We therefore calculate:

- $\frac{(\theta_1^D - \theta_1^S) + (\theta_2^D - \theta_2^S) + (\theta_3^D - \theta_3^S)}{3}$
- $\frac{(\theta_2^D - \theta_2^S) + (\theta_3^D - \theta_3^S) + (\theta_4^D - \theta_4^S)}{3}$
- $\frac{(\theta_3^D - \theta_3^S) + (\theta_4^D - \theta_4^S) + (\theta_5^D - \theta_5^S)}{3}$
- $\frac{(\theta_4^D - \theta_4^S) + (\theta_5^D - \theta_5^S) + (\theta_6^D - \theta_6^S)}{3}$
- $\frac{(\theta_5^D - \theta_5^S) + (\theta_6^D - \theta_6^S) + (\theta_7^D - \theta_7^S)}{3}$
- $\frac{(\theta_1^D - \theta_1^S) + (\theta_2^D - \theta_2^S) + (\theta_3^D - \theta_3^S) + (\theta_4^D - \theta_4^S) + (\theta_5^D - \theta_5^S) + (\theta_6^D - \theta_6^S) + (\theta_7^D - \theta_7^S)}{7}$

5.2.4.9 Leesonarity

The Leesonarity is a measure of how well the end-systole and end-diastole contours map on to one another. We demonstrate three ways of measuring Leesonarity, following different transformation methods (Figure 43).

Both contours are first translated by subtracting their centre of gravity, so both the end-systole and end-diastole are centred on the origin.

The first mapping, $lees_1$ measures the deviation of systole and diastole after rescaling.

Find λ such that the cost function, $cost = \sum_{i=1}^{n} d(c_i^D, \lambda c_i^S)$ is minimized. The value of the cost function at this minimum is $lees_1$.

The second mapping, $lees_2$ measures the deviation of systole and diastole after a rotation and a rescaling.

Find λ, θ such that the cost function, $cost = \sum_{i=1}^{n} d(c_i^D, \lambda \boldsymbol{R}(\theta).c_i^S)$ is minimized, where $\boldsymbol{R}(\theta)$ is the standard rotation matrix with parameter θ. The value of the cost function at this minimum is $lees_2$.

The third mapping, $lees_4$ measures the deviation of systole and diastole after a rotation and a rescaling and a translation.

Find λ, θ, b such that the cost function, $cost = \sum_{i=1}^{n} d(c_i^D, \lambda \boldsymbol{R}(\theta).c_i^S + b)$ is minimized, where $\boldsymbol{R}$ is the rotation matrix with parameter θ, and $b = (b_x, b_y)$ is a translation vector. The value of the cost function at this minimum is $lees_4$.

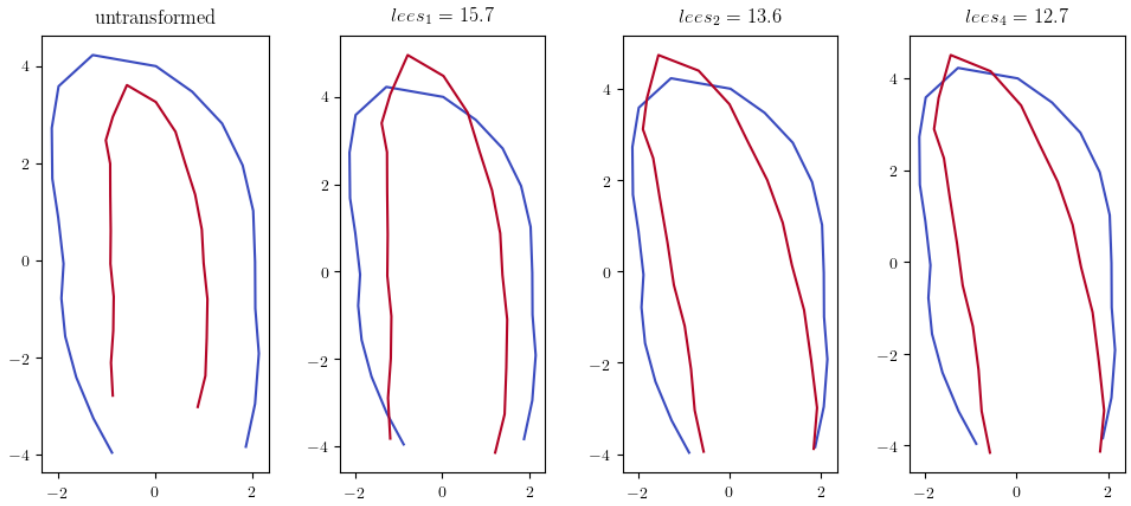


Figure 43: Leesonarity examples. Mapping of end-systole to end-diastole with varying precision as number of mapping parameters changes.

5.2.4.10 Calculation of rate features

Features calculated per view at end-diastole and end-systole can be calculated at every single frame between end-systole and end-diastole.

The time between frames dicom tag, “FrameTime” is the time between frames, and is used in the rate feature calculations:

1. calculate the feature value at every frame
2. calculate the average change from frame to frame of that feature using only a subset of the frames. The subset used is all frames between the third frame and the 3rd to last frame.
3. Divide the average change by the time between frames.

5.2.5 Building the Machine Learning Classifier

The code was written in the programming language Python, and the machine learning library scikit-learn is employed. The following 4 classifiers were used to build

predictive models on the training data: Random Forest, Support Vector Machine, Logistic Regression and Multi-Layer Perceptron. For each classifier a grid search is carried out to tune the hyperparameters, in which every combination of the hyper-parameter values is evaluated for each classifier, and the best performing hyper-parameters are then used to train the classifier. The hyper-parameters that are considered for optimisation are described below.

5.2.5.1 Random forest

rf_p_grid = {'randomforestclassifier__n_estimators': [50, 100, 200], # Number of trees

'randomforestclassifier__min_samples_split': [5, 10, 2]} # Min. No. of samples required to split a node

n_estimators: Number of trees in forest. The larger the number of trees, the better it generalises. But this comes with greater computational cost, and eventually diminishing returns.

min_samples_split: Minimum number of samples required to split a node, which controls overfitting.

5.2.5.2 Support Vector Machine

svm_p_grid = {'svc__C': [1, 10, 100], # Soft margin constant

'svc__gamma': [.0001, .001, .01]}

C: Soft margin constant, which controls the balance between keeping the street (i.e. gap between classes) as large and limiting the margin violations (i.e. instances that

end up in the middle of the street or even on the wrong side). A smaller C value leads to a wider street but more margin violations.

Gamma: The gamma parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’, and therefore determines the decision boundary. It is used as a regularisation hyperparameter: if the model is overfitting, the gamma value should be reduced, and if it underfitting, you should increase it.

5.2.5.3 Logistic regression

lr_p_grid = {'logisticregression__C': [0.01,0.1,1,10]} # regularisation parameter that controls complexity of model

C: regularisation parameter that constrains the model to make it simpler and reduce the risk of overfitting. The higher the C value, the less the model is regularised.

5.2.5.4 Multi-layer perceptron

```
mlp_p_grid = {'mlpclassifier__max_iter': [100, 200, 500],
                        'mlpclassifier__hidden_layer_sizes': [(100), (100, 100)]}
```

max_iter: maximum iterations, adjusting model parameters to minimise the loss function using an optimisation technique such as stochastic gradient descent.

hidden_layer_sizes: number of layers and nodes in each layer.

5.2.6 Validating the Models

To train and evaluate the predictive model, repeated k-fold nested cross-validation (CV) was adopted, as using just CV to tune a classifier that is itself evaluated using

CV produces a significantly biased estimate of the performance of the model (171). In nested CV the inner loop tunes the hyper-parameters of the classifiers and sampling is performed. The outer loop tests the models with the optimal hyper-parameters. Feature selection for each fold and repeat also occurs in the training fold of the outer loop, rather than the whole dataset, to ensure that overfitting does not occur (Figure 44). Both feature selection and hyper-parameter optimisation is carried it in total 20 times over the 2 folds and 10 repeats. As different hyper-parameters and features are selected each time, those that are selected the most are used for the final model. In particular, with regard to the features, it is believed that these will generalise the best to unseen data (172). For an overall prediction probability, they are averaged over all classifiers. This is known as soft voting, which takes into account the confidence level of the prediction of each classifier, and therefore tends to perform better than other ensemble techniques, such as hard voting (84). The Receiver Operator Characteristic Area Under Curve (ROC AUC) value were used to assess the performance of models.

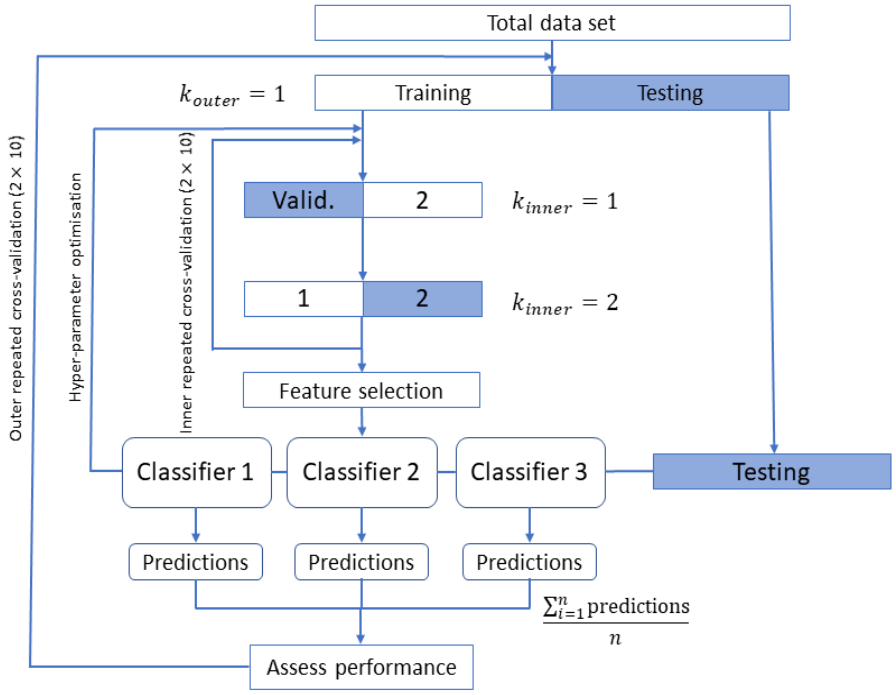


Figure 44: Nested cross fold validation used for evaluating model performance

5.3 Results

5.3.1 Patient Demographics

633 patients were available from the EVAREST dataset at the time of training the model. Tables 15 and 16 below show key demographical and patient population information.

Quantile	Age	Height	BMI	Weight
0.00	27.00	1.31	17.54	44.90
0.25	57.00	1.63	25.21	73.00
0.50	66.00	1.70	28.73	82.80
0.75	73.00	1.75	32.23	93.03
1.00	92.00	1.96	64.52	191.70
Mean	64.52	1.69	29.48	84.08
Std	11.46	0.10	5.73	17.60
Missing	2	92	114	45

Table 15: Patient Demographics

	Sex is Male	%	Coronary Event	%	Exercise Stress Performed	%	Previous CAD	%	Outcome +	%
Yes	356	56.24	14	2.21	91	14.38	252	39.81	195	31
No	277	43.76	480	75.83	541	85.47	333	52.61	438	69
Missing	0	0.00	139	21.96	1	0.16	48	7.58	0	0

Table 16: Sex, Previous History, Outcome, SE type

5.3.1 Optimised Classifier hyper-parameters

The hyper-parameters of the four classifiers are fixed to the values that are selected most frequently for the total number of times hyper-parameter optimisation is carried out during cross validation. These are as follows:

Random forest: n_estimators = 200, min_samples_split = 2

Support vector machine: C = 100, gamma = 0.0001

Logistic regression: C=1

Multi-layer Perceptron: hidden_layer = 100, max_iter = 200

5.3.2 Features selected

Using the Boruta feature selection method, 746 image features are selected at least once out of a possible 20 times, from a total of 6476. Since the training and testing dataset is different for each fold and repeat, a different feature set is selected each time. As can be seen from Table 17, about 70% of features were selected $\leq$ 5 times out of 20.

Number of times selected (out of 20)	Number of features (out of 746)
≤ 5	517
≤ 10	659
≤ 15	715

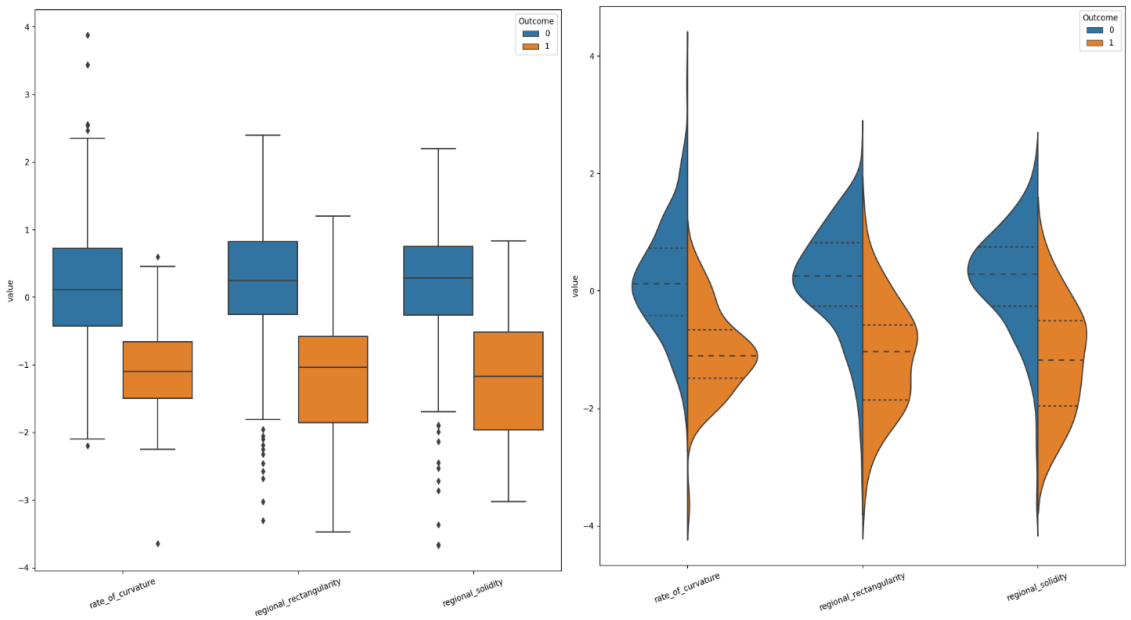
Table 17: Number of times features from the entire feature set were selected.

31 features were selected >75% of the time during cross validation of the model and these were the features chosen to be used in the final model. These are a combination of regional, and global features, describing the shape and distance, as well as the rate of change of these features between end-diastole and end-systole (Table 18).

Feature	No. of times selected	Feature	No. of times selected
rate_80_curv_section_24_4P	20	tort_section_16_ED_SP	18
sold_section_25_4P	20	rate_100_dist_section_34_SP	18
tort_section_15_ED_SP	20	rate_100_curv_triplet_026_4P	17
tort_section_15_ED_SR	20	rate_100_curv_triplet_125_2P	17
tort_section_15_ES_SP	20	rate_100_curv_triplet_245_4P	17
tort_section_15_ES_SR	20	rate_100_curv_triplet_246_4P	17
elip_var_ES_SR	19	rate_60_curv_section_24_4P	17
rate_100_curv_section_24_4P	19	rate_80_curv_triplet_125_4P	17
rate_80_curv_section_25_4P	19	sold_section_15_4P	17
rect_section_25_4P	19	rate_100_dist_pair_34_SP	17
sold_section_16_4P	19	rate_100_curv_section_25_4P	16
rate_80_curv_triplet_245_4P	18	rate_100_curv_section_26_4P	16
rect_section_16_4P	18	rate_100_curv_triplet_025_4P	16
rect_section_26_4P	18	rate_100_curv_triplet_245_2P	16
sold_section_24_4P	18	tort_section_05_ES_SR	16
sold_section_26_4P	18		

Table 18: Table shows how many times each feature was selected

Figure 45 shows three examples of key features (rate of curvature, regional rectangularity, and regional solidity) used in the final model, which were selected every single time (20/20) during the model training process. Figure 46 is a scatter plot and 3D visualisation of the aforementioned features, again visibly distinguishing between the classes.



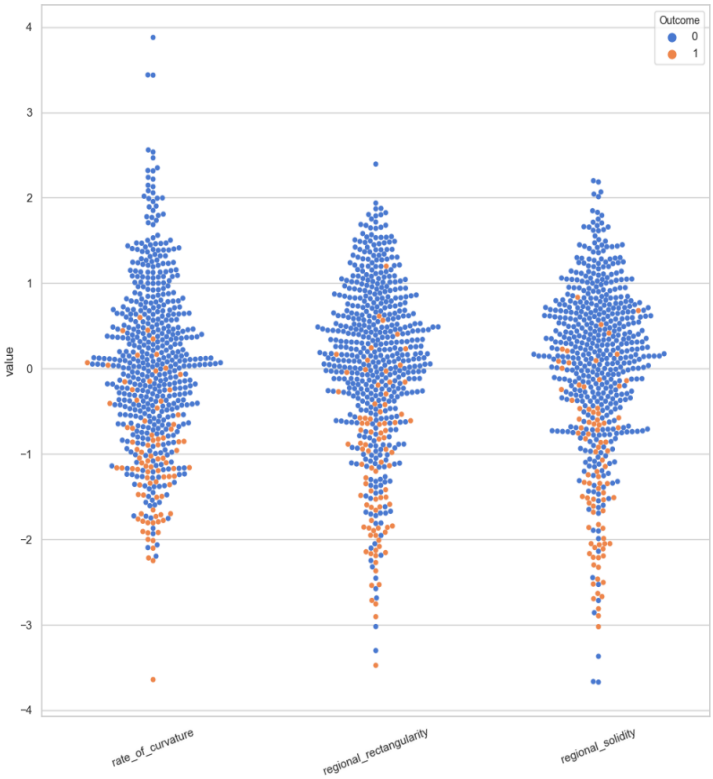


Figure 45 The feature values are normalised using Z-score. On the top left are box plots, on the top right are violin plots, and on the bottom are swarm plots, presenting examples of three significant features for the prediction of CAD. The dotted lines on the violin plots are the quartile ranges. The label 1 is positive and 0 is negative for CAD.

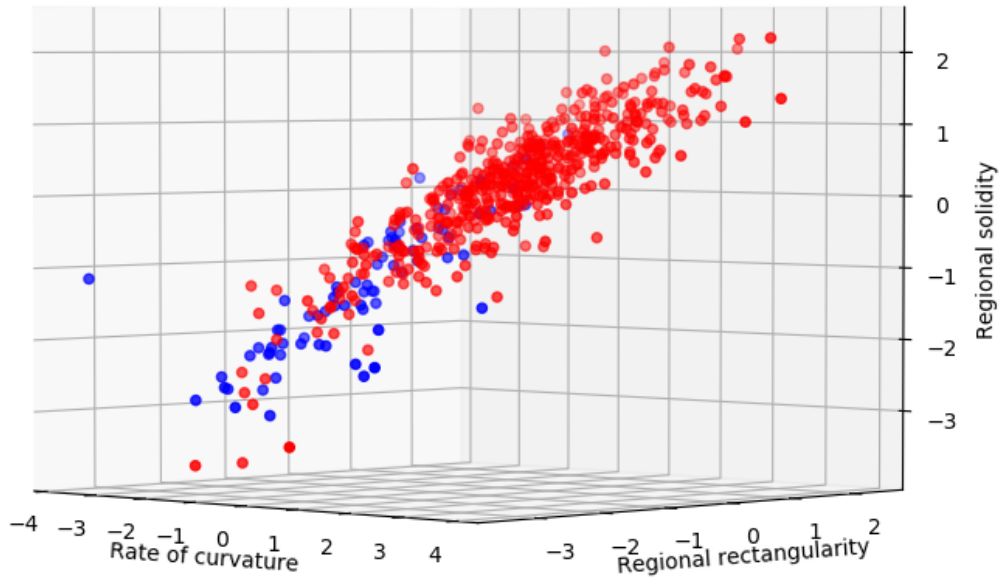


Figure 46: Scatter plot, where the red dots represent those that are labelled negative and the blue dots are those that are labelled positive for CAD.

Figure 47 shows box and violin plots demonstrating how features from just the resting echocardiogram images can effectively differentiate between the two classes.

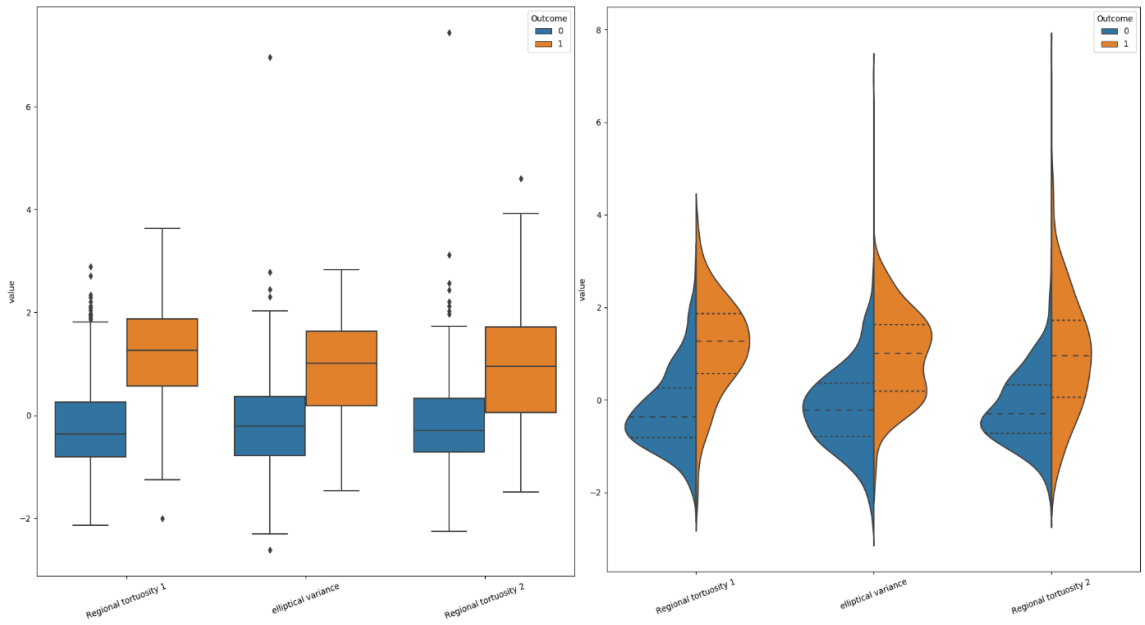


Figure 47: Box and Violin plot of resting features that are in final model

5.3.2 Model performance on training data

Once the final list of features was selected, and optimum hyper-parameters fixed, the average performance of all the classifiers is shown in Table 19 and ROC curves show in Figure 48. The ensemble with a 10 repeat 2-fold cross-validation on the training dataset yielded a ROC AUC of 0.934 and with a threshold of 0.2 the model can achieve a specificity of 0.856 and a sensitivity of 0.858. The ROC curve of the optimal ensemble is shown in Figure 49.

Threshold=0.2			
Classifiers	**ROC**	**Specificity**	**Sensitivity**
Ensemble SVM_RF_LR_MLP (soft voting)	0.934	0.856	0.858
SVM	0.929	0.853	0.851
RF	0.934	0.856	0.850
LR	0.911	0.848	0.801
MLP	0.920	0.859	0.831

Table 19: Model results from nested CV on training data

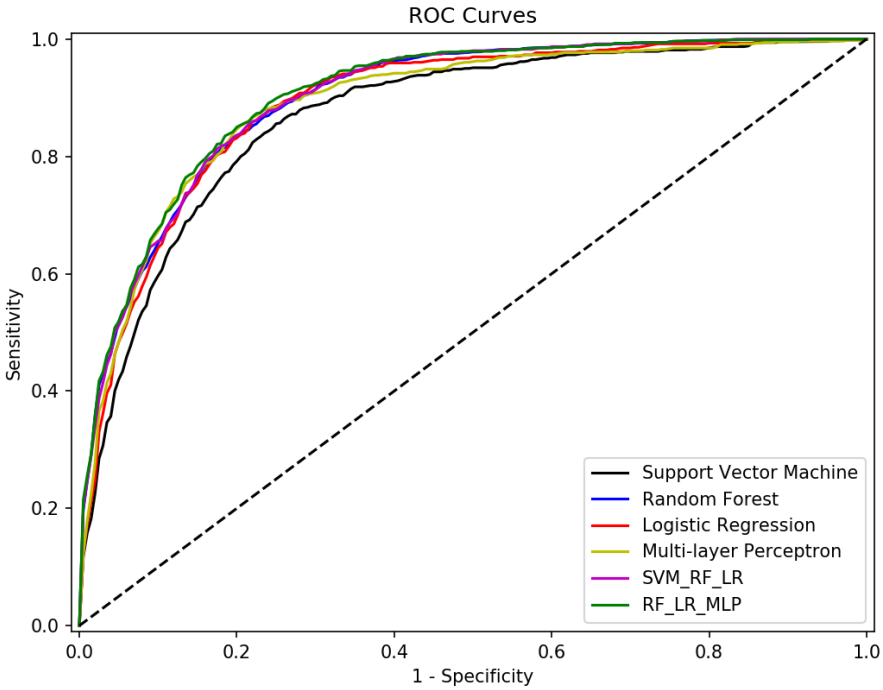


Figure 48: ROC curves of different classifiers on training data

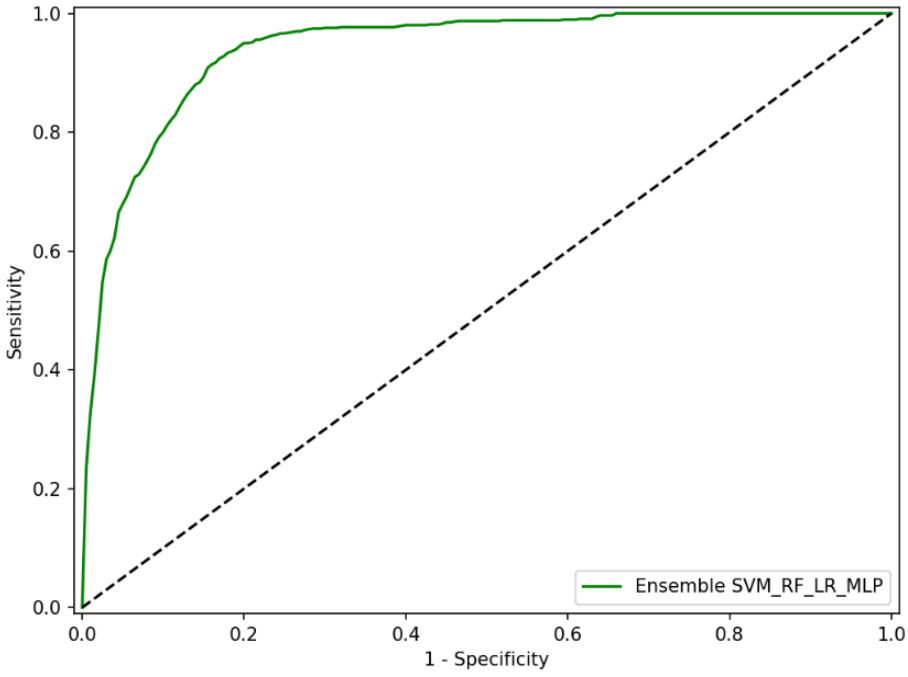


Figure 49: ROC curve of the ensemble model, which out performed the other classifiers.

The best ROC achieved by any individual feature is a ROC AUC of 0.850, as shown in Figure 50.

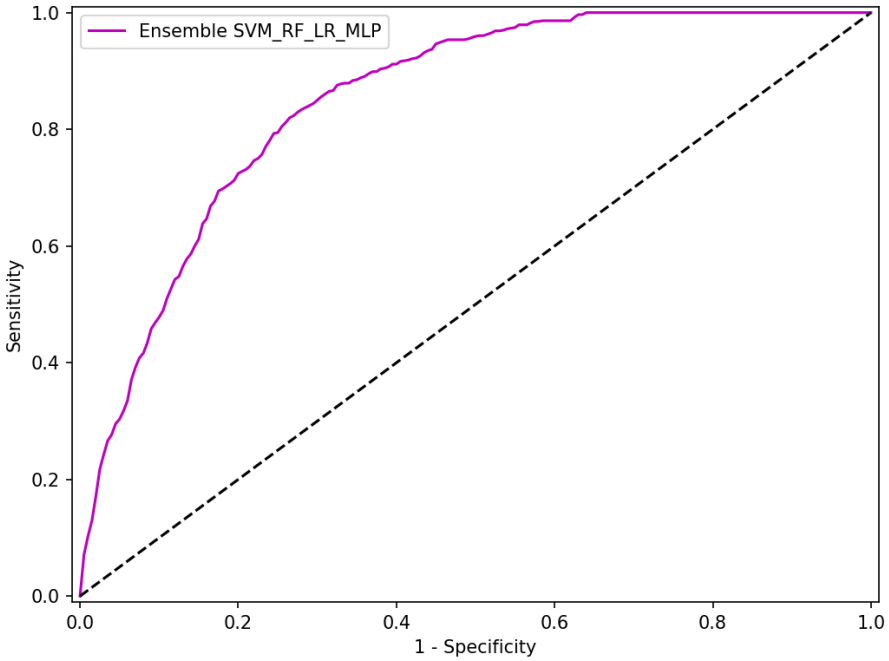


Figure 50: ROC curve with model trained on just regional solidity

5.4 Discussion

Combining the autocontouring framework, novel features and advanced supervised learning techniques enables accurate prediction of coronary artery disease. The optimal classifier constructed from an ensemble of all 4 classifiers used was able to achieve a ROC AUC of 0.934, sensitivity of 0.856 and specificity of 0.858. When this is compared to a review of the literature reporting clinical accuracies it represents a substantial improvement. Gelejinse et al. reported average sensitivities and specificities of 81% and 82 % with a range of 33-96% and 38-97% respectively. Therefore, this technology not only provides increased accuracy but also the potential to reduce clinical variability. Furthermore, creating a predictive model just

from one feature yielded very promising results with a ROC AUC of 0.850, which in itself may provide useful to support the clinical diagnosis of CAD.

31 features were chosen to be included in the final model because during the 2 x 10 nested cross validation they were the features that were selected most often. Further features were not considered as they are likely to be features that describe patterns in the training dataset, for a specific fold, and are unlikely to generalise well to unseen data (173). On the other hand, those that are selected the most, regardless of the split in the training/testing dataset, are likely to generalise the best. Therefore, the 31 features that were selected $\geq$ 16 times were used in the final fixed model. Beyond this, it was found that including further features resulted in overfitting of the training dataset and did not generalise better to unseen data. This is because the more features in a model, the more likely they will describe individual instances, directly learning the training dataset. Also, with each feature selected, the dimensionality of the classification problem increases. With each dimension, sufficient data points are required to mitigate the possibility of erroneous conclusions regarding the dividing line(s) of the classes (173).

The features that were selected most often were: regional solidity, regional rectangularity, rate of curvature and tortuosity. Figures 45-47 show distribution of feature values and probability density. These plots show a clear separation in the distribution of feature values with studies labelled with positive and negative outcomes, clearly displaying the power and ability of these features to predict CAD. Interestingly no known clinically derived features such as strain or EF appeared in the top 31 features. Further analysis will need to be conducted to understand why this was the case.

While most of the selected features are derived from peak images, as one would expect, what is compelling is that there are also several features that are calculated from just resting images. However, while the results of the resting features are very promising, it is currently not comparable to combined ability of the resting and peak features in predicting CAD.

There have been 2 earlier published studies attempting to quantify stress echocardiograms in order to improve the accuracy. Mansor introduced a technique based on a machine-learning model in combination with cavity area change. The group were able to combine 3 features to produce a local cavity area and a Hidden Markov Model was used to determine abnormal from normal. The group yielded a respectable accuracy at 84% against Cardiologist defined abnormal and normal scans (68). Chykeyuk improved upon this accuracy in 2011 when they used Support Vector Machines over the Hidden Markov Model. However, using cavity area change as a feature is operator dependent on identifying the centre of the ventricular cavity. It is also important to note with these studies is that the level of accuracy was defined against Cardiologist report rather than the patient outcome for the diagnosis of CAD (65). Therefore, unlike our study this work cannot claim to help improve the diagnosis of coronary artery disease, but rather just attempts to simulate WMS performed by a consultant cardiologist.

Despite the promising results there were some limitations associated with the study. Firstly, the sample size that we trained on was relatively small compared to other machine learning studies, particularly in other fields (174). The sample was also

slightly imbalanced which could lead to overfitting, ideally there would be a 50% split between positive and negative outcomes. Finally, the results were internally validated by cross validation and therefore remains to be seen how it performs on unseen, external data.

In conclusion, the performance of the final model clearly shows the great capabilities of machine learning, when leveraging such techniques in a medical setting. It can discern informative patterns in the data, in this complex classification problem. This can be further shown when comparing the performance of the model, to that of individual features. The next step is to validate this model in an external dataset.

6. Predicting Coronary Artery Disease: A Validation Study

6.1 Introduction

The trained model from the previous chapter was added as a module to the EchoGo software that already incorporated the autocontouring framework, this software version is hereby referred to as EchoGo Pro (Ultromics). In this chapter we describe the performance testing study conducted to evaluate diagnostic performance of EchoGo Pro.

Previous reports of standard SE interpretation accuracy vary, however Gelejinse et al. reports an average sensitivity of 81% and specificity of 82%. Consideration should be given to the fact that the comparator in most prior studies, including the aforementioned meta-analyses, encompass several confounders including purely anatomical measures and coronary angiography demonstrating stenosis > 50% or >70% being used as the cut-off for obstructive CAD (64). In this study, the comparators were more specific (including functional parameters and clinical outcomes), which represents the current standard of care. We elected to assess outcomes over a follow-up period of 1 year, as the prognostic accuracy of stress echocardiography has been well validated over this duration across multiple patient populations and clinical scenarios.

EchoGo Pro v1.0 software provides analysis of DICOM 3.0 compliant stress echo exams for patients undergoing assessment for coronary disease and provides a recommendation indicating whether there is a level of coronary disease which is prognostically significant, to support physician diagnosis (not for primary diagnosis

CAD, diagnosis of mild or moderate myocardial ischaemia, or localisation of CAD or myocardial ischaemia).

6.2 Methods

6.2.1 Study Database

This was a single-center retrospective study conducted on SE datasets received from Oregon Health and Science University (OHSU), Portland, Oregon. Data included for analysis in this study were retrospective SE images from male and female patients aged 18 years or above who received a stress-echocardiography test for suspected coronary artery disease a minimum of one year prior to data collection date to allow for collection of 12-month follow-up data (e.g. invasive angiogram results/hospital records). Given the retrospective nature of the study, high sample size, and negligible risk to involved subjects, waiver of the requirement of individual subjects' consent was requested from involved Institutional Review Boards (IRBs).

The inclusion criteria for the study were : (1) male or female, aged 18 years or above; (2) patient had undergone stress echocardiography for suspected coronary artery disease 1 year or more prior to data collection; (3) results of coronary angiogram (if performed) and medical records are obtainable for the 1-year following the SE test. Data were excluded if: (1) age < 18 years; (2) non-diagnostic SE (failure to reach target heart rate/double product, poor echocardiographic windows and other technical considerations). Relevant medical history, including history of current disease, other pertinent respiratory history, and information regarding underlying diseases were also recorded.

The SE datasets from all included cases were sent from the Syngo Dynamics PACS at OHSU to a Laurel Bridge Compass router. The router was configured to filter DICOM tags associated with Manufacturer, Make, Model and Stress Echo study definitions to determine the location of the burned in Patient Information. The router was also configured for each specific study and image type to crop the appropriate x,y quadrants from the image or not to crop the image if there is no patient information present. Additionally, DICOM tags that contained Patient Name, Medical Record Number, Patient ID and Patient Age was setup for removal by the router, prior to study transfer. The studies were sent from the Laurel Bridge router to the Ambra Cloud Suite whereby they were assigned a pseudo-anonymized patient identifier in place of the real name, in a numerical, incremental order. The data was accessed from AMBRA to perform the quality assessment and EchoGo Pro processing of the data.

Direct access to the collected data was granted to authorized representatives from the trial team for monitoring and audit of the study to ensure compliance with regulations. The Investigator prepared and maintained adequate and accurate source documents designed to record all observations. The study personnel entered data from source documents corresponding to a sample case into the protocol-specific electronic Case Report Form (eCRF). Subjects could not be identified by name in the study database or on any study documents to be collected. If a correction was required for an eCRF, the time and date stamps tracked the person entering or updating eCRF data and created an electronic audit trail. The investigator was responsible for all information collected on samples enrolled in this study.

All data collected during the course of this study was reviewed and verified for completeness and accuracy by the Investigator. The data was entered into a validated database at OHSU. A study log was kept on a secure database with the medical record review. The site data management group was responsible for data processing, in accordance with institutional procedural documentation.

The SE DICOM datasets were stored within Ambra Cloud Suite and processed via EchoGo Pro. Password protected electronic database was used for the clinical data and to record the physician's interpretation during the reader variability study. The database was safeguarded against unauthorized access by established security procedures; appropriate backup copies of the database and related software files will be maintained. Databases are backed up by the database administrator in conjunction with any updates or changes to the database. The Investigator ensured that study data remained accessible to the monitor, other authorized representatives of the Sponsor (or designee), IRB/IEC, and Regulatory Agency (e.g., FDA) inspectors upon request. A file for each subject was maintained that includes the copies of all source documentation related to that subject.

The data collection was conducted according to the Declaration of Helsinki, Protection of Human Volunteers (21 CFR 50), Institutional Review Boards (21 CFR 56), Obligations of Clinical Investigators (21 CFR 812) and ISO 14155:2016. To maintain confidentiality, evaluation forms, reports and other records were identified by a coded number and initials only. All study records were kept in a secure database. The Investigator complied with all applicable privacy regulations (e.g.,

Health Insurance Portability and Accountability Act of 1996, EU Data Protection Directive 95/46/EC).

6.2.2 Medical Record Review

Medical record review was performed by the study site to identify whether the patient had been admitted into the hospital with acute coronary syndrome. Evidence was recorded for incidences of cardiovascular events including admissions with chest pain, shortness of breath, heart failure or cardiovascular procedures, including coronary angiography, vascular investigations or surgery, or a record of death with myocardial infarction listed as primary cause. Coronary angiography was used as the gold standard to identify CAD. Patients who had a negative SE and did not undergo invasive angiography or other investigation during the one year following SE were considered to have a negative outcome if there was an absence of any cardiac events such as acute coronary syndrome or death from a cardiovascular cause recorded in their medical records. Patient records for this group were reviewed in the OHSU electronic medical record (EMR) system (Epic Systems Corporation, Verona, WI) and, to mitigate against lack of cardiovascular events being a result of loss to follow up, these patient records were also be reviewed in the Care Everywhere network of the Epic EMR platform. The Care Everywhere network integrates health care records, including outpatient visits, inpatient/emergency department visits, and all diagnostic testing, for hospitals and health care systems using the Epic EMR platform. In the absence of any medical encounter records at all, the social security death index was queried to check for any death during follow-up.

6.2.3 Assessment of Patient Outcomes

The patient outcome was obtained following the assessment of angiograms and patient records in the year following the data collection by the site. A positive outcome was defined as in the previous chapter, evidence of obstructive coronary artery disease (> 70% epicardial coronary artery stenosis with or without consequent percutaneous coronary intervention [PCI], or ≤ 70% epicardial coronary artery stenosis but with low Fractional Flow Reserve or low Instantaneous Wave-Free Ratio prompting PCI) during invasive coronary angiography done either electively or following admission for acute coronary syndrome; percutaneous or surgical coronary revascularization; or death with myocardial infarction listed as primary cause.

Where a participant had an event, or underwent angiography, the outcome was determined by clinical opinion of an individual with appropriate expertise e.g. board-certified clinician or cardiologist who undertook the angiography or reported the mortality. Where this was not available, or there was uncertainty, outcome was adjudicated by an adjudication committee at OHSU comprised of at least two board certified cardiologists, of which one was not involved in project oversight, data collection or analysis. The adjudication committee could exclude patients if, after inclusion, the committee subsequently decided there was uncertainty about diagnosis due to insufficient clinical information to confirm outcome, such as lack of adequate confirmatory invasive angiography data following other non-invasive tests.

6.2.4 Quality Assessment

Quality assessment (QA) was carried out on all SE datasets by echocardiographers accredited to American or British Society of Echocardiography standards who had been trained to use the software.

Data was checked to ensure the stress echo had been performed consistent with American Society of Echocardiography guidelines (175) including collection of required views, that all stages had been labelled appropriately, had been correctly triggered, had a heart rate that was appropriate to the stage, rather than errors caused by arrhythmias and had adequate endocardial border definition for visual assessment. Data which failed this part of the QA process was recorded as ineligible for the study and excluded.

6.2.5 Data Processing by EchoGo Pro

Echocardiograms were processed using EchoGo Pro v1.0. The images first underwent a quality review before the LV endocardial contour was applied automatically to the images and approved by an accredited echocardiographer that had undergone training. A report was generated including a recommendation for a higher or lower risk of prognostically significant CAD. The accuracy of EchoGo Pro v1.0 in predicting 12-month clinical outcomes was assessed.

6.3 Results

6.3.1 Patient Demographics

191 patients were included in the final dataset for validation the demographics of which are listed below in tables 20 and 21. 20 patients were rejected due to insufficient image quality.

Quantile	Ages	Height	BMI	Weight
0.00	26.56	1.40	4.23	12.25
0.25	55.71	1.63	24.58	70.76
0.50	63.30	1.68	28.03	81.19
0.75	68.89	1.75	32.58	97.30
1.00	87.23	2.03	59.98	167.83
Mean	62.58	1.69	29.52	84.67
Std	10.46	0.10	7.42	21.20
Missing	0	0	1	0

Table 20: Patient Demographics

	Sex is Male	%	Coronary Event	%	Exercise Stress Performed	%	Previous CAD	%	Outcome +	%
Yes	100	52.36	8	4.19	131	68.59	39	20.42	79	41
No	90	47.12	183	95.81	60	31.41	152	79.58	112	59
Missing	1	0.52	0	0.00	0	0.00	0	0.00	0	0

Table 21: Sex, Previous History, Outcome, SE type

6.3.2 Accuracy of Classifiers

The results of EchoGo Pro shown in table 22 alongside the other classifiers that we trained in the previous chapter to show as comparators. The ROC curves for these other classifiers are displayed in Figure 51 and EchoGo Pro which achieved a ROC AUC of 0.929, is shown in Figure 52.

Threshold 0.2	Accuracy	Sensitivity	Specificity
Ensemble SVM_RF_LR_MLP (EchoGo Pro)	0.929	0.864	0.911
SVM	0.894	0.886	0.892
RF	0.903	0.727	0.854
LR	0.885	0.841	0.873
MLP	0.867	0.864	0.866

Table 22: Results of each classifier on validation set

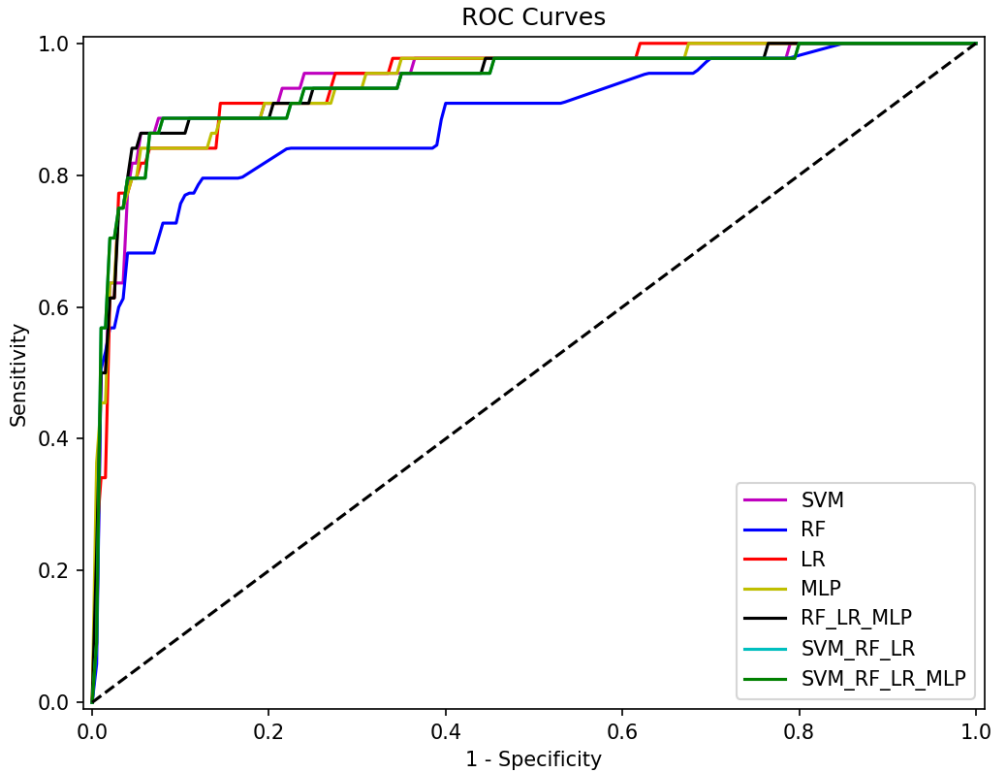


Figure 51: ROC curves for each classifier on validation dataset.

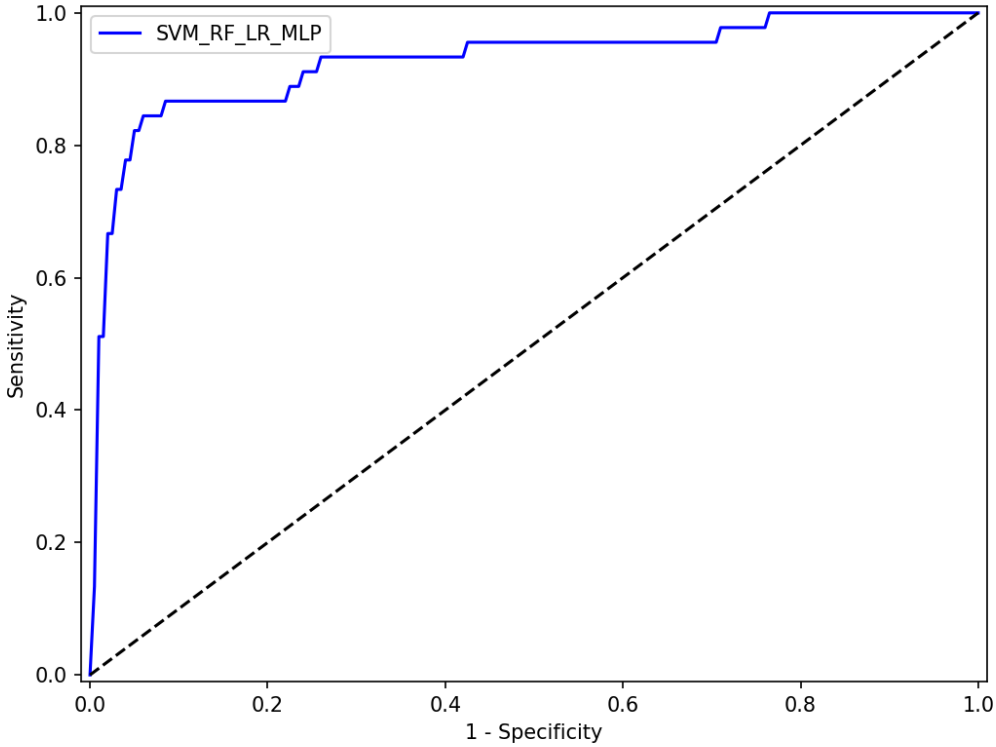


Figure 52: ROC curve for EchoGo Pro (SVM_RF_LR_MLP Ensemble)

6.3.3 Training versus Validation Data

When considering the potential clinical validity of an AI classifier in practice, an estimation of the heterogeneity of data between training and validation is informative. The results describe the difference between Rainier (validation) and Evarest (training) datasets. Tables 23-31 show differences in demographics, stress echo type and previous disease. The dataset means are matched well for age (EVAREST 65.5 vs Rainier 61.9), sex (EVAREST 56.3% Male vs Rainier 56.5% Male), Height (EVAREST 1.69 vs 1.71), weight (EVAREST 84.9 vs Rainier 87.9) and BMI (30.1 vs 30.1). However there was a significant mean difference in previous CAD (EVAREST 39.6% vs Rainer 19.1%).

Training data	Ages	Validation data	Ages
0	39.49	0	26.56
0.25	58.35	0.25	54.97
0.5	66.93	0.5	62.55
0.75	74.01	0.75	68.45
1	89.84	1	87.23
Mean	65.48	Mean	61.94
Std	12.08	Std	10.40
Missing	179	Missing	0

Table 23: Age difference

Training data	%	Validation data	%
Male	56.31	Male	56.55
Female	43.69	Female	43.27
Missing	0	Missing	0.18

Table 24: Sex Difference

Training data	Height	Validation data	HGT	HGT Converted
0	1.31	0	55	1.4
0.25	1.62	0.25	64	1.63
0.5	1.69	0.5	67	1.70
0.75	1.76	0.75	70	1.78
1	1.96	1	80	2.03
Mean	1.69	Mean	67.19	1.71
Std	0.10	Std	3.96	0.10
Missing	230	Missing	0	0

Table 25: Height difference

Training data	Weight	Validation data	WGT	WGT Converted
0	44.9	0	27	12.25
0.25	73	0.25	157	71.21
0.5	83	0.5	184	83.46
0.75	95	0.75	220	99.79
1	191.7	1	446	202.30
Mean	84.94	Mean	193.7218	87.87
Std	18.97	Std	53.26	24.16
Missing	183	Missing	0	0

Table 26: Weight difference

Training data	BMI	Validation data	BMI2
0	17.54	0	4.23
0.25	25.54	0.25	24.69
0.5	29.23	0.5	28.29
0.75	33.06	0.75	33.85
1	64.52	1	74.22
Mean	30.07	Mean	30.12
Std	6.25	Std	7.95
Missing	255	Missing	2

Table 27: BMI difference

Training data	%	Validation data	%
No	48.11	No	80.91
Yes	39.59	Yes	19.09
Missing	12.30	Missing	0

Table 28: Previous CAD

Training data	%	Validation data	%
Yes	15.93	Yes	8.55
No	7.10	No	91.45
Missing	76.97	Missing	0

Table 29: Previous MI

Training data	%	Validation data	%
No	36.59	No	86
Yes	9.46	Yes	14
Missing	53.94	Missing	0

Table 30: Previous CABG

Training data	%	Validation	%
Yes	85.17	Yes	30.9
No	14.35	No	69.1
Missing	0.47	Missing	0

Table 31: Dobutamine Stress Echo

6.4 Discussion

The results of EchoGo Pro on the validation set achieved a ROC AUC of 0.929, sensitivity of 86% and specificity of 91%. These results are very similar to those achieved in the training data from the previous chapter and would represent a

substantial improvement in clinical practice if EchoGo Pro maintained this level of accuracy.

These results represent the first time that a machine learning classifier has been validated on external data for the prediction of CAD in stress echocardiography. The other classifiers (RF, SVM, LR, MLP) that were trained in the previous chapter all performed well although, the performance dropped more than the ensemble performance, which provides further validation that use of an Ensemble classifier generalises better than individual classifiers.

When analysing the differences between the training and validation data we can see firstly that there are a many missing data in the training data which makes comparisons in certain categories difficult. However, it is important to note that the cohorts were well matched for age, gender, weight and BMI especially considering how these factors can affect ultrasound image quality.

One of the more striking differences between the datasets was type of stress echo where the training data favoured dobutamine over exercise stress echo and the validation data favoured exercise. This is most likely due to exercise echo being favoured in the US over dobutamine SE as it has been found to be more sensitive (93). Since the final results were very similar between the cohorts it does not seem to have made much of an impact, however it will be an interesting sub-analysis to perform in larger cohort studies.

There have been previous studies that have used machine learning techniques to predict other cardiovascular disease from echocardiograms with varying degrees of success. For example, Narula et al. attempted to determine the difference between left ventricular hypertrophy in athletes and left ventricular hypertrophy in hypertrophic cardiomyopathy. Given the risk of sudden cardiac death in hypertrophic cardiomyopathy accurate differentiation is important. In this study, similar to our validation study they developed an ensemble machine learning model consisting of multiple classifiers including support vector machines, random forests, and convolutional neural networks were able to differentiate between hypertrophic cardiomyopathy and athletes, with a sensitivity of 96% (11).

Furthermore, another study used machine learning classifiers to differentiate between constrictive pericarditis and restrictive cardiomyopathy which also have similar phenotypes. However, a machine learning model was able differentiate between these diseases, with an accuracy of 90%, using multiple echocardiography measurements (37,40).

Despite the accuracies of these studies, neither attempted to predict patient outcomes which is a big limitation. In the case of differentiating between athletes and hypertrophic cardiomyopathy, it is fine to differentiate, but what is arguably more important is to determine which of those patients that have hypertrophic cardiomyopathy go on to have a cardiac event or are at risk of sudden cardiac death. This illustrates the importance of using patient outcomes as the dependent variable and is one of the factors that differentiates our study.

Although the results achieved in this validation study were very accurate and replicated the training data results there were limitations. Firstly, the study was a retrospective sample of stress echocardiograms in a single centre database. While this provides good validation for the classifier due to it being unseen data from a different country, sample biases could have been introduced by the methods and procedures of data selection. A more effective study would be a multi-centre prospective trial.

Furthermore, although the technology performed well, it remains to be seen how the technology will perform in clinical practice, for example, the cardiologist performing the test may ignore the recommendation given by EchoGo Pro. Finally, the sample size of the study was relatively small, so ideally the technology needs to be validated on further datasets.

In conclusion, the technology we developed (EchoGo Pro) performed very accurately in the validation study and would represent a substantial improvement compared with current clinical practice according to the literature. Further testing is required to determine if this accuracy can be maintained in prospective datasets.

7. Future Developments

7.1 Autocontouring

The first area of work to be completed is refining of the auto contouring methodology. The accuracy of this image processing technique can have significant consequences on the ability to predict coronary artery disease as we are trying to detect subtle wall motion abnormalities. Therefore, it is imperative that autocontouring is as precise as possible. The first stage of refinements of the auto-contouring framework is already in process with the addition of further echo imaging datasets.

In addition, incorporating time domain behaviour in the CNN itself is likely to lead to the greatest improvements in performance as described in other publications (176). The next steps in this regard will likely be to construct a 3D U-net based approach (177,178) which incorporates time domain behaviour where the three dimensions are two spatial dimensions plus time. The sparse labelling of the image frames (i.e. just end-diastole and systole labels are available) makes training such a network challenging but previous studies have shown that training with sparse labelling is possible and leads to reasonable results (178).

As there is general uncertainty about which regions of an image frame contain the LV and which do not, even amongst experienced operators, a less deterministic contouring approach may work well with this problem. Recently such a network was proposed in the form of a probabilistic U-net (179). This outputs a likelihood of each pixel in an image belonging to a certain segmentation class and this may be suitable for contouring noisy and unclear echocardiogram images.

Finally, automation of the image quality assurance process is currently under development for implementation into the technology platform as this is currently being performed manually by experienced operators. Once completed, the Autocontouring platform will be fully automated and will not be subject to operator variability.

7.2 Feature Generation

Identifying new features and biomarkers has been critical in trying to detect coronary artery disease. Some of the biomarkers I have generated for this DPhil seem to be better predictors than currently known clinical markers such as EF and GLS. Further validation for these biomarkers such as solidity, rectangularity and tortuosity are required. These biomarkers may also be relevant to other disease areas associated with LV function such as Heart Failure, Cardiomyopathies, Valve Disease and Cardiotoxicity. It will be interesting to perform studies in the aforementioned diseases to see if these novel biomarkers translate. We already have studies being designed to investigate heart failure in collaboration with the Mayo Clinic and Cardiotoxicity with University of Alberta. If these features seem applicable to these other disease areas, we can follow the same pipeline outlined in this thesis and build predictive models for each disease.

Furthermore, I plan to continue to generate novel features that describe LV function and geometry. This can be done both by using our current approach of clinically-driven design of biomarkers that are relevant to specific diseases but also by using a novel machine learning approach. One such approach would be to use Deep Learning to identify and extract image based features that are predictive of disease

(180). By using this approach, we are not limited to our clinical knowledge and therefore we can develop biomarkers that are determined to be important by machine learning classifiers.

By combining the machine-driven biomarkers with our clinically derived biomarkers this should help enhance our predictive capabilities. Moreover, the addition of clinical metadata into our predictive classifiers may also enhance predictive capabilities therefore we have set-up a research project to investigate this further.

7.3 Future Validation Studies

Further validation of EchoGo Pro is underway in clinical trials using external datasets. We are commencing and have ongoing trials in multiple hospitals including: Oregon Health and Science University, University College San Diego, Cleveland Clinic, Mayo Clinic, Stanford University and University of Alberta. These studies will also evaluate the clinical benefit of EchoGo Pro by investigating whether the addition of EchoGo Pro reports during interpretation increases the physician interpretation of SEs compared to standard reading practice. For example, within OHSU, 8 cardiologists are performing reads with and without EchoGo Pro in a double-blind cross-over study with a 4 week washout period. This been shown to be an effective way to demonstrate how a technology improves a readers or in this case a cardiologist's performance. The results of this study are expected in December 2019.

Finally plans are in process to extend the EVAREST study and perform a randomised control multi-centre trial with EchoGo Pro. This should further help prove the validity of the technology and will build on our retrospective trials. This trial is being planned to commence in 2020 and will be performed across 18 of the EVAREST sites.

8. References

1. Maleki M, Esmaeilzadeh M. The evolutionary development of echocardiography. Iran J Med Sci [Internet]. 2012 Dec;37(4):222–32. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23390327

2. Al-Taweel A, Almahmoud MF, Khairandish Y, Ahmad M. Degenerative mitral valve stenosis: Diagnosis and management. Echocardiography. 2019 Oct;

3. Douglas PS, Garcia MJ, Haines DE, Lai WW, Manning WJ, Patel AR, et al. ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 Appropriate Use Criteria for Echocardiography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Associat. J Am Soc Echocardiogr. 2011/02/23. 2011;24(3):229–67.

4. Pysz P, Kozlowski M, Malczewska M, Adamczyk-Filipek E, Wojakowski W, Smolka G. Prospective registry validating the reproducibility of mitral paravalvular leak measurements in a standardized real-time three-dimensional transesophageal echocardiography algorithm for optimal choice of the closure device. Postep w Kardiol interwencyjnej = Adv Interv Cardiol. 2019;15(2):203–10.

5. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr [Internet]. 2015/01/07. 2015;28(1):1-39 e14. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25559473

6. Gaibazzi N, Porter TR, Agricola E, Cioffi G, Mazzone C, Lorenzoni V, et al. Prognostic value of echocardiographic calcium score in patients with a clinical indication for stress echocardiography. JACC Cardiovasc Imaging. 2015 Apr;8(4):389–96.

7. Hoffmann R, Lethen H, Marwick T, Arnese M, Fioretti P, Pingitore A, et al. Analysis of interinstitutional observer agreement in interpretation of dobutamine stress echocardiograms. J Am Coll Cardiol. 1996/02/01. 1996;27(2):330–6.

8. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol. 2019 Apr;28(2):73–81.

9. Leeson P, Fletcher AJ. Combining Artificial Intelligence With Human Insight to Automate Echocardiography. Vol. 12, Circulation. Cardiovascular imaging. United States; 2019. p. e009727.

10. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, et al. Deep Learning in Medical Imaging: General Overview. Korean J Radiol. 2017/07/04. 2017;18(4):570–84.

11. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography. J Am Coll Cardiol. 2016/11/26. 2016;68(21):2287–95.

12. Schank RC. What is AI, anyway? AI Mag. 1987;8(4):59.

13. Brynjolfsson E, Mitchell T. What can machine learning do? Workforce implications. Science (80-). 2017;358(6370):1530–4.

14. Mayr A, Binder H, Gefeller O, Schmid M. The evolution of boosting algorithms. From machine learning to statistical modelling. Methods Inf Med. 2014/08/13. 2014;53(6):419–27.

15. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart [Internet]. 2018/01/21. 2018;104(14):1156–64. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29352006

16. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. Emerg Artif Intell Appl Comput Eng. 2007;160:3–24.

17. Meherwar Fatima MP. Survey of Machine Learning Algorithms for Disease Diagnostic. J Intell Learn Syst Appl. 2017;9(1):1–16.

18. Talabis MRM, McPherson R, Miyamoto I, Martin JL, Kaye D. Chapter 1 - Analytics Defined. In: Talabis MRM, McPherson R, Miyamoto I, Martin JL, Kaye DBT-ISA, editors. Boston: Syngress; 2015. p. 1–12. Available from: http://www.sciencedirect.com/science/article/pii/B9780128002070000010

19. Krittanawong C, Tunhasiriwet A, Zhang H, Wang Z, Aydar M, Kitai T. Deep Learning With Unsupervised Feature in Echocardiographic Imaging. J Am Coll Cardiol. 2017/04/22. 2017;69(16):2100–1.

20. Sahba F, Tizhoosh HR, Salama MMA. A Reinforcement Learning Framework for Medical Image Segmentation. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings. 2006. p. 511–7.

21. Barto RSS and AG. Reinforcement Learning: An Introduction. Second edi. Cambridge, Massachusetts; London, England: The MIT Press;

22. Kubat M. An Introduction to Machine Learning. Cham: Springer International Publishing; 2017.

23. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature [Internet]. 2015/02/27. 2015;518(7540):529–33. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25719670

24. Gao J. Machine learning applications for data center optimization. 2014;

25. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 2010;11(Jul):2079–107.

26. Yamada A, Luis SA, Sathianathan D, Khandheria BK, Cafaro J, Hamilton-Craig CR, et al. Reproducibility of Regional and Global Longitudinal Strains Derived from Two-Dimensional Speckle-Tracking and Doppler Tissue Imaging between Expert and Novice Readers during Quantitative Dobutamine Stress Echocardiography. J Am Soc Echocardiogr [Internet]. 2014 Aug 1;27(8):880–7.

Available from: https://doi.org/10.1016/j.echo.2014.04.016

27. Awan SE, Sohel F, Sanfilippo FM, Bennamoun M, Dwivedi G. Machine learning in heart failure: ready for prime time. Curr Opin Cardiol [Internet]. 2017/12/02. 2018;33(2):190–5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29194052

28. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature [Internet]. 2017/01/25. 2017;542(7639):115–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28117445

29. Wolterink JM, Leiner T, Takx RAP, Viergever MA, Išgum I. Automatic coronary calcium scoring in non-contrast-enhanced ECG-triggered cardiac CT with ambiguity detection. IEEE Trans Med Imaging. 2015;34(9):1867–78.

30. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. Jama [Internet]. 2016/11/30. 2016;316(22):2402–10. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27898976

31. Cobey FC. Intelligent Algorithms in Perioperative Echocardiography: A New Era. J Am Soc Echocardiogr. 2017;30(10):A26–7.

32. Nishimura RA, Otto CM, Bonow RO, Carabello BA, Erwin 3rd JP, Guyton RA, et al. 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation [Internet]. 2014/03/05. 2014;129(23):2440–92. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24589852

33. Goodacre S, Locker T, Morris F, Campbell S. How useful are clinical features in the diagnosis of acute, undifferentiated chest pain? Acad Emerg Med. 2002;9(3):203–8.

34. Knackstedt C, Bekkers SC, Schummers G, Schreckenberg M, Muraru D, Badano LP, et al. Fully Automated Versus Standard Tracking of Left Ventricular Ejection Fraction and Longitudinal Strain: The FAST-EFs Multicenter Study. J Am Coll Cardiol. 2015/09/26. 2015;66(13):1456–66.

35. Thavendiranathan P, Popovic ZB, Flamm SD, Dahiya A, Grimm RA, Marwick TH. Improved interobserver variability and accuracy of echocardiographic visual left ventricular ejection fraction assessment through a self-directed learning program using cardiac magnetic resonance images. J Am Soc Echocardiogr [Internet]. 2013/09/03. 2013;26(11):1267–73. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23993695

36. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial Intelligence in Cardiology. J Am Coll Cardiol [Internet]. 2018/06/09. 2018;71(23):2668–79. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29880128

37. Sengupta PP, Huang YM, Bansal M, Ashrafi A, Fisher M, Shameer K, et al. Cognitive Machine-Learning Algorithm for Cardiac Imaging: A Pilot Study for

Differentiating Constrictive Pericarditis From Restrictive Cardiomyopathy. Circ Cardiovasc Imaging. 2016/06/09. 2016;9(6).

38. Levy F, Dan Schouver E, Iacuzio L, Civaia F, Rusek S, Dommerc C, et al. Performance of new automated transthoracic three-dimensional echocardiographic software for left ventricular volumes and function assessment in routine clinical practice: Comparison with 3 Tesla cardiac magnetic resonance. Arch Cardiovasc Dis. 2017/06/02. 2017;110(11):580–9.

39. Chao PK, Wang CL, Chan HL. An intelligent classifier for prognosis of cardiac resynchronization therapy based on speckle-tracking echocardiograms. Artif Intell Med. 2011/10/05. 2012;54(3):181–8.

40. Mahmoud A, Bansal M, Sengupta PP. New Cardiac Imaging Algorithms to Diagnose Constrictive Pericarditis Versus Restrictive Cardiomyopathy. Curr Cardiol Rep. 2017/04/14. 2017;19(5):43.

41. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med. 2001;23(1):89–109.

42. Arsanjani R, Dey D, Khachatryan T, Shalev A, Hayes SW, Fish M, et al. Prediction of revascularization after myocardial perfusion SPECT by machine learning in a large population. J Nucl Cardiol. 2014/12/07. 2015;22(5):877–84.

43. Betancur J, Otaki Y, Motwani M, Fish MB, Lemley M, Dey D, et al. Prognostic Value of Combined Clinical and Myocardial Perfusion Imaging Data Using Machine Learning. JACC Cardiovasc Imaging. 2017/10/23. 2018;11(7):1000–9.

44. Haro Alonso D, Wernick MN, Yang Y, Germano G, Berman DS, Slomka P. Prediction of cardiac death after adenosine myocardial perfusion SPECT based on machine learning. J Nucl Cardiol. 2018/03/16. 2018;

45. Dey D, Slomka PJ, Leeson P, Comaniciu D, Shrestha S, Sengupta PP, et al. Artificial Intelligence in Cardiovascular Imaging: JACC State-of-the-Art Review. J Am Coll Cardiol [Internet]. 2019;73(11):1317–35. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30898208

46. Wharton G, Steeds R, Allen J, Phillips H, Jones R, Kanagala P, et al. A minimum dataset for a standard adult transthoracic echocardiogram: a guideline protocol from the British Society of Echocardiography. Echo Res Pract. 2015;2(1):G9–24.

47. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2014. p. 806–13.

48. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning.

49. Khamis H, Zurakhov G, Azar V, Raz A, Friedman Z, Adam D. Automatic apical view classification of echocardiograms using a discriminative learning dictionary. Med Image Anal. 2016/11/07. 2017;36:15–21.

50. Maron BJ. Hypertrophic cardiomyopathy: a systematic review. Jama.

2002;287(10):1308–20.

51. Hatle LK, Appleton CP, Popp RL. Differentiation of constrictive pericarditis and restrictive cardiomyopathy by Doppler echocardiography. Circulation. 1989;79(2):357–70.

52. Gregg Belous David Rowlands AB. Segmentation of the Left Ventricle from Ultrasound Using Random Forest with Active Shape Model. Artificial Intelligence, Modelling and Simulation (AIMS). Kota Kinabalu, Malaysia: IEEE; 2013.

53. Cannesson M, Tanabe M, Suffoletto MS, McNamara DM, Madan S, Lacomis JM, et al. A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction. J Am Coll Cardiol. 2007/01/16. 2007;49(2):217–26.

54. Rahmouni HW, Ky B, Plappert T, Duffy K, Wiegers SE, Ferrari VA, et al. Clinical utility of automated assessment of left ventricular ejection fraction using artificial intelligence-assisted border detection. Am Hear J. 2008/02/26. 2008;155(3):562–70.

55. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. Echo Res Pract. 2018 Dec;5(4):R115–25.

56. Domingos JS, Stebbing R V, Leeson P, Noble JA, Wu G, Zhang D, et al. Structured Random Forests for Myocardium Delineation in 3D Echocardiography. In Cham: Springer International Publishing; 2014. p. 215–22.

57. Stebbing R V, Namburete AIL, Upton R, Leeson P, Noble JA. Data-driven shape parameterization for segmentation of the right ventricle from 3D+t echocardiography. Med Image Anal. 2015 Apr;21(1):29–39.

58. Cobey FC, Patel V, Gosling A, Ursprung E. The Emperor Has No Clothes: Recognizing the Limits of Current Echocardiographic Technology in Perioperative Quantification of Mitral Regurgitation. J Cardiothorac Vasc Anesth. 2017/06/27. 2017;31(5):1692–4.

59. Jeganathan J, Knio Z, Amador Y, Hai T, Khamooshian A, Matyal R, et al. Artificial intelligence in mitral valve analysis. Ann Card Anaesth. 2017/04/11. 2017;20(2):129–34.

60. Moghaddasi H, Nourian S. Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos. Comput Biol Med. 2016/04/16. 2016;73:47–55.

61. Grothues F, Smith GC, Moon JCC, Bellenger NG, Collins P, Klein HU, et al. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. Am J Cardiol. 2002;90(1):29–34.

62. Nagel E, Lehmkuhl HB, Bocksch W, Klein C, Vogel U, Frantz E, et al. Noninvasive diagnosis of ischemia-induced wall motion abnormalities with the use of high-dose dobutamine stress MRI: comparison with dobutamine stress

echocardiography. Circulation. 1999;99(6):763–70.

63. Ladapo JA, Blecker S, Douglas PS. Physician decision making and trends in the use of cardiac stress testing in the United States: an analysis of repeated cross-sectional data. Ann Intern Med. 2014/10/07. 2014;161(7):482–90.

64. Geleijnse ML, Krenning BJ, van Dalen BM, Nemes A, Soliman OI, Bosch JG, et al. Factors affecting sensitivity and specificity of diagnostic testing: dobutamine stress echocardiography. J Am Soc Echocardiogr. 2009/09/22. 2009;22(11):1199–208.

65. Chykeyuk K, Clifton DA, Noble JA. Feature extraction and wall motion classification of 2D stress echocardiography with relevance vector machines. In: Proceedings - International Symposium on Biomedical Imaging [Internet]. 2011. p. 677–80. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-80055040222&doi=10.1109%2FISBI.2011.5872497&partnerID=40&md5=88d27ff056064bdee748ee145c285c2d

66. Raghavendra U, Fujita H, Gudigar A, Shetty R, Nayak K, Pai U, et al. Automated technique for coronary artery disease characterization and classification using DD-DTDWT in ultrasound images. Biomed Signal Process Control. 2018;40:324–34.

67. Omar HA, Domingos JS, Patra A, Upton R, Leeson P, Noble JA. Quantification of cardiac bull's-eye map based on principal strain analysis for myocardial wall motion assessment in stress echocardiography. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). 2018. p. 1195–8.

68. Mansor S, Hughes NP, Noble JA. Wall motion classification of stress echocardiography based on combined rest-and-stress data. Med Image Comput Comput Assist Interv. 2008/11/06. 2008;11(Pt 2):139–46.

69. Liu S, Wang X, Liu M, Zhu J. Towards better analysis of machine learning models: A visual analytics perspective. Vis Informatics [Internet]. 2017;1(1):48–56. Available from: http://www.sciencedirect.com/science/article/pii/S2468502X17300086

70. Lakshminarayan K, Harp SA, Goldman RP, Samad T. Imputation of Missing Data Using Machine Learning Techniques. In: KDD. 1996. p. 140–5.

71. Denoeux T. A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Trans Syst Man Cybern. 1995;25(5):804–13.

72. Jayalakshmi T, Santhakumaran A. Statistical normalization and back propagation for classification. Int J Comput Theory Eng. 2011;3(1):1793–8201.

73. Al Shalabi L, Shaaban Z. Normalization as a preprocessing engine for data mining and the approach of preference matrix. In: 2006 International Conference on Dependability of Computer Systems. IEEE; 2006. p. 207–14.

74. Akbani R, Kwek S, Japkowicz N. Applying support vector machines to imbalanced datasets. In: European conference on machine learning. Springer; 2004. p. 39–50.

75. Rahman MM, Davis DN. Addressing the class imbalance problem in medical

datasets. Int J Mach Learn Comput. 2013;3(2):224.

76. Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.

77. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explor Newsl. 2004;6(1):20.

78. Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. In: Icml. Nashville, USA; 1997. p. 179–86.

79. Kira K, Rendell LA. A practical approach to feature selection. In: Machine Learning Proceedings 1992. Elsevier; 1992. p. 249–56.

80. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. J Stat Softw. 2010;36(11):1–13.

81. Kursa MB, Rudnicki WR, Kursa MMB. Package 'Boruta.' 2018.

82. Breiman L. Random Forests. 2001;1–33.

83. Huynh T, Gao Y, Kang J, Wang L, Zhang P, Lian J, et al. Estimating CT image from MRI data using structured random forest and auto-context model. IEEE Trans Med Imaging. 2015;35(1):174–83.

84. Géron A. Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems. 2017. 549 p.

85. Cortes C, Vapnik V. Support-Vector Networks. Mach Learn. 1995;20(3):273–97.

86. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics. 2000;16(10):906–14.

87. Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput. 2004;14(3):199–222.

88. Cox DR. the Regression Analysis of Binary Sequences. J R Stat Soc Ser B J R Stat Soc Ser B J R Stat Soc Ser B. 1958;20(2):215–42.

89. Ng AY, Jordan MI. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In: Advances in neural information processing systems. 2002. p. 841–8.

90. Gardner JW, Craven M, Dow C, Hines EL. The prediction of bacteria type and culture growth phase by an electronic nose with a multi-layer perceptron network. Meas Sci Technol. 1998;9(1):120.

91. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomed Eng Online. 2014;13(1):94.

92. Foundation BH. BHF Heart Statistics Factsheet – UK [Internet]. Vol. 2019. British Heart Foundation Website; 2018. Available from: https://www.bhf.org.uk/what-we-do/our-research/heart-statistics

93. Sicari R, Nihoyannopoulos P, Evangelista A, Kasprzak J, Lancellotti P, Poldermans D, et al. Stress Echocardiography Expert Consensus Statement--Executive Summary: European Association of Echocardiography (EAE) (a registered branch of the ESC). Eur Hear J [Internet]. 2008/11/11. 2009;30(3):278–89. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19001473

94. Sicari R, Cortigiani L. The clinical use of stress echocardiography in ischemic heart disease. Cardiovasc Ultrasound [Internet]. 2017/03/21. 2017;15(1):7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28327159

95. N. NA, Mauro G, P. HJ, P. FD, Rasha A-L. Mortality From Ischemic Heart Disease. Circ Cardiovasc Qual Outcomes [Internet]. 2019 Jun 1;12(6):e005375. Available from: https://doi.org/10.1161/CIRCOUTCOMES.118.005375

96. Grech ED. Pathophysiology and investigation of coronary artery disease. BMJ [Internet]. 2003 May 10;326(7397):1027 LP – 1030. Available from: http://www.bmj.com/content/326/7397/1027.abstract

97. Kolodgie FD, Gold HK, Burke AP, Fowler DR, Kruth HS, Weber DK, et al. Intraplaque hemorrhage and progression of coronary atheroma. N Engl J Med. 2003;349(24):2316–25.

98. R. GJ, F. KK. Development of the Lipid-Rich Core in Human Atherosclerosis . Arterioscler Thromb Vasc Biol [Internet]. 1996 Jan 1;16(1):4–11. Available from: https://doi.org/10.1161/01.ATV.16.1.4

99. Lee RT, Libby P. The unstable atheroma. Arterioscler Thromb Vasc Biol. 1997;17(10):1859–67.

100. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. Lancet. 2003;361(9351):13–20.

101. Erbel R, Heusch G. Coronary microembolization. Journal of the American College of Cardiology; 2000.

102. Picano E. Dipyridamole-echocardiography test: historical background and physiologic basis. Eur Heart J. 1989 Apr;10(4):365–76.

103. Gallagher KP, Matsuzaki M, Koziol JA, Kemper WS, Ross JJ. Regional myocardial perfusion and wall thickening during ischemia in conscious dogs. Am J Physiol. 1984 Nov;247(5 Pt 2):H727-38.

104. Ross JJ. Mechanisms of regional ischemia and antianginal drug action during exercise. Prog Cardiovasc Dis. 1989;31(6):455–66.

105. Picano E, Palinkas A, Amyot R. Diagnosis of myocardial ischemia in hypertensive patients. J Hypertens. 2001 Jul;19(7):1177–83.

106. Nihoyannopoulos P, Kaski JC, Crake T, Maseri A. Absence of myocardial dysfunction during stress in patients with syndrome X. J Am Coll Cardiol. 1991 Nov;18(6):1463–70.

107. Camici PG, Gistri R, Lorenzoni R, Sorace O, Michelassi C, Bongiorni MG, et

al. Coronary reserve and exercise ECG in patients with chest pain and normal coronary angiograms. Circulation. 1992 Jul;86(1):179–86.

108. Picano E, Lattanzi F, Masini M, Distante A, L'Abbate A. Usefulness of a high-dose dipyridamole-echocardiography test for diagnosis of syndrome X. Am J Cardiol. 1987 Sep;60(7):508–12.

109. Palinkas A, Toth E, Amyot R, Rigo F, Venneri L, Picano E. The value of ECG and echocardiography during stress testing for identifying systemic endothelial dysfunction and epicardial artery stenosis. Eur Heart J. 2002 Oct;23(20):1587–95.

110. Renker M, Baumann S, Rier J, Ebersberger U, Fuller SR, Batalis NI, et al. Imaging coronary artery disease and the myocardial ischemic cascade: clinical principles and scope. Radiol Clin North Am. 2015 Mar;53(2):261–9.

111. Ikonomidis I, Athanassopoulos G, Stamatelopoulos K, Lekakis J, Revela I, Venetsanou K, et al. Additive prognostic value of interleukin-6 at peak phase of dobutamine stress echocardiography in patients with coronary artery disease. A 6-year follow-up study. Am Hear J [Internet]. 2008;156(2):269–76. Available from: http://www.sciencedirect.com/science/article/B6W9H-4T2KV08-8/2/41982fb4c4cca422064dc766038e9bbc

112. Nicola DCC, G. MF, A. CS, Philip C, Fabian B, Joseph SU. Coronary Artery Computed Tomography Scanning. Circulation [Internet]. 2014 Mar 25;129(12):1341–5. Available from: https://doi.org/10.1161/CIRCULATIONAHA.113.002835

113. Jacobs JE, Boxt LM, Desjardins B, Fishman EK, Larson PA, Schoepf J, et al. ACR practice guideline for the performance and interpretation of cardiac computed tomography (CT). J Am Coll Radiol [Internet]. 2006;3(9):677–85. Available from: http://europepmc.org/abstract/MED/17412148

114. Caruso D, Eid M, Schoepf UJ, Jin KN, Varga-Szemes A, Tesche C, et al. Dynamic CT myocardial perfusion imaging. Eur J Radiol [Internet]. 2016;85(10):1893–9. Available from: http://www.sciencedirect.com/science/article/pii/S0720048X16302340

115. Pijls NHJ, Fearon WF, Tonino PAL, Siebert U, Ikeno F, Bornschein B, et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention in patients with multivessel coronary artery disease: 2-year follow-up of the FAME (Fractional Flow Reserve Versus Angiography for Multivessel Evaluation) study. J Am Coll Cardiol. 2010 Jul;56(3):177–84.

116. Shaw LJ, Berman DS, Maron DJ, Mancini GBJ, Hayes SW, Hartigan PM, et al. Optimal medical therapy with or without percutaneous coronary intervention to reduce ischemic burden: results from the Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation (COURAGE) trial nuclear substudy. Circulation. 2008 Mar;117(10):1283–91.

117. Meinel FG, De Cecco CN, Schoepf UJ, Nance JWJ, Silverman JR, Flowers BA, et al. First-arterial-pass dual-energy CT for assessment of myocardial blood supply: do we need rest, stress, and delayed acquisition? Comparison with SPECT. Radiology. 2014 Mar;270(3):708–16.

118. Mordi I, Stanton T, Carrick D, McClure J, Oldroyd K, Berry C, et al. Comprehensive dobutamine stress CMR versus echocardiography in LBBB and suspected coronary artery disease. JACC Cardiovasc Imaging. 2014 May;7(5):490–8.

119. Greenwood JP, Motwani M, Maredia N, Brown JM, Everett CC, Nixon J, et al. Comparison of cardiovascular magnetic resonance and single-photon emission computed tomography in women with suspected coronary artery disease from the Clinical Evaluation of Magnetic Resonance Imaging in Coronary Heart Disease (CE-MARC) Trial. Circulation. 2014 Mar;129(10):1129–38.

120. Schwitter J, Wacker CM, Wilke N, Al-Saadi N, Sauer E, Huettle K, et al. MR-IMPACT II: Magnetic Resonance Imaging for Myocardial Perfusion Assessment in Coronary artery disease Trial: perfusion-cardiac magnetic resonance vs. single-photon emission computed tomography for the detection of coronary artery disease: a comparative . Eur Heart J. 2013 Mar;34(10):775–81.

121. Parker MW, Iskandar A, Limone B, Perugini A, Kim H, Jones C, et al. Diagnostic accuracy of cardiac positron emission tomography versus single photon emission computed tomography for coronary artery disease: a bivariate meta-analysis. Circ Cardiovasc Imaging. 2012 Nov;5(6):700–7.

122. Underwood SR, Shaw LJ, Anagnostopoulos C, Cerqueira M, Ell PJ, Flint J, et al. Myocardial perfusion scintigraphy and cost effectiveness of diagnosis and management of coronary heart disease. Heart [Internet]. 2004;90 Suppl 5:v34-6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15254007

123. Hamon M, Fau G, Née G, Ehtisham J, Morello R. Meta-analysis of the diagnostic performance of stress perfusion cardiovascular magnetic resonance for detection of coronary artery disease. J Cardiovasc Magn Reson [Internet]. 2010/05/19. 2010;12:29. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20482819

124. Picano E, Lattanzi F, Orlandini A, Marini C, L'Abbate A. Stress echocardiography and the human factor: the importance of being expert. J Am Coll Cardiol [Internet]. 1991;17(3):666–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/1993786

125. Becher H, Chambers J, Fox K, Jones R, Leech GJ, Masani N, et al. BSE procedure guidelines for the clinical application of stress echocardiography, recommendations for performance and interpretation of stress echocardiography: a report of the British Society of Echocardiography Policy Committee. Heart [Internet]. 2004;90 Suppl 6:vi23-30. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15564422

126. Senior R, Monaghan M, Becher H, Mayet J, Nihoyannopoulos P. Stress echocardiography for the diagnosis and risk stratification of patients with suspected or known coronary artery disease: a critical appraisal. Supported by the British Society of Echocardiography. Heart [Internet]. 2005 Apr 1;91(4):427 LP – 436. Available from: http://heart.bmj.com/content/91/4/427.abstract

127. NICE. Chest Pain of Recent Onset: Assessment and Diagnosis [Internet]. Vol.

2019. 2016. Available from: https://www.nice.org.uk/guidance/cg95/chapter/recommendations

128. van der Sijde JN, Boiten HJ, van Domburg RT, Schinkel AF. Long-Term (>10 Years) Prognostic Value of Dobutamine Stress Echocardiography in a High-Risk Cohort. Am J Cardiol [Internet]. 2016/01/14. 2016;117(7):1078–83. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26839054

129. Steeds RP, Wheeler R, Battacharyya S, Reiken J, Nihoyannopulos P, Senior R, et al. Stress echocardiography in coronary artery disease: a practical guideline from the British Society of Echocardiography. Echo Res Pr [Internet]. 2019/03/01. 2019; Available from: https://www.ncbi.nlm.nih.gov/pubmed/30921767

130. Makani H, Bangalore S, Halpern D, Makwana HG, Chaudhry FA. Cardiac outcomes with submaximal normal stress echocardiography: a meta-analysis. J Am Coll Cardiol [Internet]. 2012/09/12. 2012;60(15):1393–401. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22981557

131. Beleslin BD, Ostojic M, Stepanovic J, Djordjevic-Dikic A, Stojkovic S, Nedeljkovic M, et al. Stress echocardiography in the detection of myocardial ischemia. Head-to-head comparison of exercise, dobutamine, and dipyridamole tests. Circulation [Internet]. 1994;90(3):1168–76. Available from: https://www.ncbi.nlm.nih.gov/pubmed/7916274

132. Noble J a., Boukerroui D. Ultrasound image segmentation: a survey. IEEE Trans Med Imaging. 2006;25(8):987–1010.

133. Chan TF, Vese LA. Active contours without edges. IEEE Trans Image Process. 2001;10(2):266–77.

134. Akgul YS, Kambhamettu C. A new multi-level framework for deformable contour optimization. Proceedings 1999 IEEE Comput Soc Conf Comput Vis Pattern Recognit (Cat No PR00149). 1999;26(I):465–70.

135. Milletari F, Ahmadi S-A, Kroll C, Plate A, Rozanski V, Maiostre J, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. Comput Vis Image Underst. 2017;164:92–102.

136. Xu K, Yang Y, Stone M, Jaumard-Hakoun A, Leboullenger C, Dreyfus G, et al. Robust contour tracking in ultrasound tongue image sequences. Clin Linguist Phon. 2016;30(3–5):313–27.

137. Nagata Y, Kado Y, Onoue T, Otani K, Nakazono A, Otsuji Y, et al. Impact of image quality on reliability of the measurements of left ventricular systolic function and global longitudinal strain in 2D echocardiography. Echo Res Pract [Internet]. 2018/02/05. 2018 Mar;5(1):27–39. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29432198

138. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, et al. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018;138(16):1623–35.

139. Medvedofsky D, Lang RM, Kruse E, Guile B, Weinert L, Ciszek B, et al. Feasibility of Left Ventricular Global Longitudinal Strain Measurements from

Contrast-Enhanced Echocardiographic Images. J Am Soc Echocardiogr. 2018 Mar;31(3):297–303.

140. Farsalinos KE, Daraban AM, Ünlü S, Thomas JD, Badano LP, Voigt JU. Head-to-Head Comparison of Global Longitudinal Strain Measurements among Nine Different Vendors: The EACVI/ASE Inter-Vendor Comparison Study. J Am Soc Echocardiogr. 2015;28(10):1171-1181.e2.

141. Aloysius N, Geetha M. A review on deep convolutional neural networks. Proc 2017 IEEE Int Conf Commun Signal Process ICCSP 2017. 2018;2018-Janua:588–92.

142. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst. 2012;1–9.

143. Nie D, Wang L, Gao Y, Sken D. Fully convolutional networks for multi-modality isointense infant brain image segmentation. Proc - Int Symp Biomed Imaging. 2016;2016-June:1342–5.

144. Ciresan D, Giusti A, Gambardella L, Schmidhuber J. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. Neural Inf Process Syst. 2012;1–9.

145. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2015;9351:234–41.

146. Ibtehaz N, Rahman MS. MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw. 2019 Sep;121:74–87.

147. Lieman-Sifry J, Le M, Lau F, Sall S, Golden D. Fastventricle: Cardiac segmentation with ENet. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2017. p. 127–38.

148. Chollet F. Keras. https://keras.io; 2015.

149. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016;

150. Bernard O, Bosch JG, Heyde B, Alessandrini M, Barbosa D, Camarasu-Pop S, et al. Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography. IEEE Trans Med Imaging. 2016;35:967–77.

151. Tsang W, Salgo IS, Medvedofsky D, Takeuchi M, Prater D, Weinert L, et al. Transthoracic 3D echocardiographic left heart chamber quantification using an automated adaptive analytics algorithm. JACC Cardiovasc Imaging. 2016;9:769–82.

152. Otani K, Nakazono A, Salgo IS, Lang RM, Takeuchi M. Three-dimensional echocardiographic assessment of left heart chamber size and function with fully automated quantification software in patients with atrial fibrillation. J Am Soc Echocardiogr. 2016;29:955–65.

153. Thavendiranathan P, Grant AD, Negishi T, Plana JC, Popovic ZB, Marwick TH. Reproducibility of echocardiographic techniques for sequential assessment of left ventricular ejection fraction and volumes: application to patients undergoing cancer chemotherapy. J Am Coll Cardiol. 2013 Jan;61(1):77–84.

154. FDA U. K173780 EchoMD Automated Ejection Fraction Software. Vol. 49. 2015.

155. Asch FM, Abraham T, Jankowski M, Cleve J, Adams M, Romano N, et al. ACCURACY AND REPRODUCIBILITY OF A NOVEL ARTIFICIAL INTELLIGENCE DEEP LEARNING-BASED ALGORITHM FOR AUTOMATED CALCULATION OF EJECTION FRACTION IN ECHOCARDIOGRAPHY. J Am Coll Cardiol [Internet]. 2019 Mar 12;73(9 Supplement 1):1447. Available from: http://www.onlinejacc.org/content/73/9_Supplement_1/1447.abstract

156. Patrianakos AP, Zacharaki AA, Kalogerakis A, Solidakis G, Parthenakis FI, Vardas PE. Two-dimensional global and segmental longitudinal strain: are the results from software in different high-end ultrasound systems comparable? Echo Res Pract. 2015 Mar;2(1):29–39.

157. Wood PW, Choy JB, Nanda NC, Becher H. Left ventricular ejection fraction and volumes: it depends on the imaging method. Echocardiography. 2014;31(1):87–100.

158. Ross Agner BF, Katz MG, Williams ZR, Dixen U, Jensen GB, Schwarz KQ. Left Ventricular Systolic Function Assessed by Global Longitudinal Strain is Impaired in Atrial Fibrillation Compared to Sinus Rhythm. J Atr Fibrillation. 2017 Dec;10(4):1437.

159. Belous ABG, Rowlands D. Artificial Intelligence, Modelling and Simulation (AIMS). Segmentation left Vent from ultrasound using random For with Act shape Model. 2013;

160. Januzzi JLJ, Chandrashekhar Y. Strain Echocardiography: The New Gold Standard for Imaging Ventricular Function? Vol. 70, Journal of the American College of Cardiology. United States; 2017. p. 955–7.

161. NICE. Chest pain of recent onset: Costing report. NICE, editor. Clinical Guideine 95. London; 2011.

162. Ntusi NA, Samuels P, Moosa S, Mocumbi AO. Diagnosing cardiac disease during pregnancy: imaging modalities. Cardiovasc J Afr [Internet]. 2016;27(2):95–103. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27213857

163. Marwick TH. Stress echocardiography. Heart [Internet]. 2003 Jan;89(1):113–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12482809

164. Picano E, Lattanzi F, Orlandini A, Marini C, L'Abbate A. Stress echocardiography and the human factor: The importance of being expert. J Am Coll Cardiol. 1991;17(3):666–9.

165. Underwood SR, Anagnostopoulos C, Cerqueira M, Ell PJ, Flint EJ, Harbinson M, et al. Myocardial perfusion scintigraphy: The evidence - A consensus conference organised by the British Cardiac Society, the British Nuclear

Cardiology Society and the British Nuclear Medicine Society, endorsed by the Royal College of Physicians of London and t. Eur J Nucl Med Mol Imaging. 2004;31(2):261–91.

166. Hamon M, Fau G, Née G, Ehtisham J, Morello R, Hamon M. Meta-analysis of the diagnostic performance of stress perfusion cardiovascular magnetic resonance for detection of coronary artery disease. J Cardiovasc Magn Reson [Internet]. 2010;12(1):29. Available from: http://dx.doi.org/10.1186/1532-429X-12-29

167. Henglin M, Stein G, Hushcha PV, Snoek J, Wiltschko AB, Cheng S. Machine Learning Approaches in Cardiovascular Imaging. Circ Cardiovasc Imaging. 2017;10(10).

168. Massalha S, Clarkin O, Thornhill R, Wells G, Chow BJW. Decision Support Tools, Systems, and Artificial Intelligence in Cardiac Imaging. Can J Cardiol. 2018 Jul;34(7):827–38.

169. Dreyer KJ, Geis JR. When Machines Think: Radiology's Next Frontier. Radiology. 2017 Dec;285(3):713–8.

170. Mingqiang Y, Kidiyo K, Joseph R. A Survey of Shape Feature Extraction Techniques. In: Yin P-Y, editor. Pattern Recognition. Rijeka: IntechOpen; 2008.

171. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics. 2006;7:1–8.

172. Wainer J, Cawley G. Nested cross-validation when selecting classifiers is overzealous for most practical applications. 2018.

173. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: A new perspective. Neurocomputing [Internet]. 2018;300:70–9. Available from: http://www.sciencedirect.com/science/article/pii/S0925231218302911

174. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep Learning in Radiology: Does One Size Fit All? J Am Coll Radiol [Internet]. 2018;15(3, Part B):521–6. Available from: http://www.sciencedirect.com/science/article/pii/S1546144017316721

175. Pellikka PA, Nagueh SF, Elhendy AA, Kuehl CA, Sawada SG. American Society of Echocardiography recommendations for performance, interpretation, and application of stress echocardiography. J Am Soc Echocardiogr. 2007 Sep;20(9):1021–41.

176. Jafari MH, Girgis H, Liao Z, Behnami D, Abdi A, Vaseli H, et al. A Unified Framework Integrating Recurrent Fully-Convolutional Networks and Optical Flow for Segmentation of the Left Ventricle in Echocardiography Data. Dlmia. 2018;10553:29–37.

177. Romera-Paredes B, Karthikesalingam A, Visentin D, Hughes J, Chopra R, Meyer C, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med. 2018;24(9):1342–50.

178. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. Lect Notes

Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2016;9901 LNCS:424–32.

179. Kohl SAA, Romera-Paredes B, Meyer C, De Fauw J, Ledsam JR, Maier-Hein KH, et al. A Probabilistic U-Net for Segmentation of Ambiguous Images. 2018;(NeurIPS):1–28.

180. Dara S, Tumma P. Feature Extraction By Using Deep Learning: A Survey. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). 2018. p. 1795–801.

181. Yach D, Hawkes C, Gould CL, Hofman KJ. The Global Burden of Chronic Diseases: Overcoming Impediments to Prevention and Control. Jama [Internet]. 2004;291(21):2616–22. Available from: http://jama.ama-assn.org/cgi/content/abstract/291/21/2616

182. Ikonomidis I, Athanassopoulos G, Lekakis J, Venetsanou K, Marinou M, Stamatelopoulos K, et al. Myocardial Ischemia Induces Interleukin-6 and Tissue Factor Production in Patients With Coronary Artery Disease: A Dobutamine Stress Echocardiography Study. Circulation [Internet]. 2005;112(21):3272–9. Available from: http://circ.ahajournals.org/cgi/content/abstract/112/21/3272

183. Feringa HHH, Elhendy A, Bax JJ, Boersma E, de Jonge R, Schouten O, et al. Baseline plasma N-terminal pro-B-type natriuretic peptide is associated with the extent of stress-induced myocardial ischemia during dobutamine stress echocardiography. Coron Artery Dis. 2006;17(3):255–9.

184. Christenson RH, deFilippi CP, Kreutzer D. Biomarkers of Ischemia in Patients With Known Coronary Artery Disease: Do Interleukin-6 and Tissue Factor Measurements During Dobutamine Stress Echocardiography Give Additional Insight? Circulation [Internet]. 2005;112(21):3215–7. Available from: http://circ.ahajournals.org

185. Baron M, Boulanger CM, Staels B, Tailleux A. Cell-derived microparticles in atherosclerosis: biomarkers and targets for pharmacological modulation? J Cell Mol Med [Internet]. 2011/11/05. 2011;16(7):1365–76. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22050954

186. Andrews RK, Berndt MC. Microparticles facilitate neutrophil/platelet crosstalk. Blood [Internet]. 2008/09/10. 2008;112(6):2174–5. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18779396

187. Bulut D, Tuns H, Mugge A. CD31+/Annexin V+ microparticles in healthy offsprings of patients with coronary artery disease. Eur J Clin Invest [Internet]. 2008/12/18. 2009;39(1):17–22. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19087127

188. Ueba T, Nomura S, Inami N, Nishikawa T, Kajiwara M, Iwata R, et al. Plasma level of platelet-derived microparticles is associated with coronary heart disease risk score in healthy men. J Atheroscler Thromb [Internet].

2010/04/10. 2010;17(4):342–9. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20379056

189. Nozaki T, Sugiyama S, Koga H, Sugamura K, Ohba K, Matsuzawa Y, et al. Significance of a multiple biomarkers strategy including endothelial dysfunction to improve risk stratification for cardiovascular events in patients at high risk for coronary heart disease. J Am Coll Cardiol [Internet]. 2009/08/08. 2009;54(7):601–8. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19660689

190. Bernal-Mizrachi L, Jy W, Fierro C, Macdonough R, Velazques HA, Purow J, et al. Endothelial microparticles correlate with high-risk angiographic lesions in acute coronary syndromes. Int J Cardiol [Internet]. 2004/11/25. 2004;97(3):439–46. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15561331

191. Augustine D, Ayers L V, Lima E, Newton L, Lewandowski AJ, Davis EF, et al. Dynamic release and clearance of circulating microparticles during cardiac stress. Circ Res [Internet]. 2013/10/18. 2014;114(1):109–13. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24141170

192. Bland JM ADG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1:307–10.

193. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med. 1996 Sep;3(9):895–900.

194. Carley S, Dosman S, Jones SR, Harrison M. Simple nomograms to calculate sample size in diagnostic studies. Emerg Med J. 2005 Mar;22(3):180–1.

195. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. Emerg Med J [Internet]. 2003 Sep;20(5):453–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12954688

196. Underwood SR, Anagnostopoulos C, Cerqueira M, Ell PJ, Flint EJ, Harbinson M, et al. Myocardial perfusion scintigraphy: the evidence. Eur J Nucl Med Mol Imaging [Internet]. 2004;31(2):261–91. Available from: http://dx.doi.org/10.1007/s00259-003-1344-5

197. Armstrong WF, Pellikka PA, Ryan T, Crouse L, Zoghbi WA. Stress echocardiography: recommendations for performance and interpretation of stress echocardiography. Stress Echocardiography Task Force of the Nomenclature and Standards Committee of the American Society of Echocardiography. J Am Soc Echocardiogr. 1998 Jan;11(1):97–104.

198. Galiuto L, Badano L, Fox K, Sicari R, Zamorano JL. The EAE Textbook of Echocardiography [Internet]. Oxford, UK: Oxford University Press; Available from: https://oxfordmedicine.com/view/10.1093/med/9780199599639.001.0001/med-9780199599639

9. Appendices

9.1 Appendix A – EVAREST Clinical Trial Protocol

SYNOPSIS

Study Title	The Use of Blood Biomarkers, Including Extracellular Vesicles, to Improve the Diagnostic Accuracy of Cardiac Assessment by Stress Echocardiogram	
Internal ref. no. / short title	Extracellular Vesicles At REst and STress – EVAREST Study	
Study Design	Multi-centre observational study	
Study Participants	Patients under investigation for cardiovascular disease who have been referred for a stress echocardiogram will be recruited in 2 phases. For the initial phase (group 1) results of the blood biomarker and the stress echocardiogram together will be compared with clinical outcome in the 12 months after the test to investigate whether the use of biomarkers in combination with echocardiogram can improve the diagnostic accuracy of this procedure. Recruitment will then continue into a second phase (Group 2) to allow assessment of accuracy based on analysis of ultrasound images alone.	
Planned Sample Size	8000	
Planned Study Period	7.5 years	
	Objectives	Endpoints
Primary	To establish whether the measurement of specific blood biomarkers, including extracellular vesicles, during a stress echocardiogram, can give additional prognostic value to the procedure. (Group 1 only)	Analysis of stress echocardiogram and any subsequent angiogram testing, in conjunction with biomarker levels

Secondary	To determine the impact of cardiac stress on the circulating levels of blood biomarkers, including extracellular vesicles which will provide information of the rate of release and clearance from the circulation.	Results of biomarker level analysis before, during and after cardiac stress
	To analyse any correlations between the extent of inducible ischemia generated during stress echocardiogram with changes in levels of blood biomarkers.	Results from correlation analysis, between the extent of inducible ischemia and changes in levels of blood biomarkers
	To evaluate the accuracy of image-based analysis of stress echo (Group 1 and 2 combined).	Analysis of stress echocardiogram and clinical outcome over the next 12 months (Group 1 and 2 combined).

ABBREVIATIONS

CAD	Coronary Artery Disease
CI	Chief Investigator
CRF	Case Report Form
CTRG	Clinical Trials & Research Governance, University of Oxford
DSE	Dobutamine Stress Echocardiogram
ELISA	Enzyme-linked immunosorbent assay
EMVs	Endothelial Cell Derived Microvesicles
EVs	Extracellular Vesicles
GCP	Good Clinical Practice
GP	General Practitioner
ICF	Informed Consent Form
IQR	Inter Quartile Range
MVs	Microvesicles
NHS	National Health Service
NRES	National Research Ethics Service
NTA	Nanoparticle Tracking Analysis
PI	Principal Investigator
PIL	Participant/ Patient Information Leaflet
PMVs	Platelet Derived Microvesicles
R&D	NHS Trust R&D Department
REC	Research Ethics Committee

SD	Standard Deviation
SOP	Standard Operating Procedure
TRPS	Tunable Resistive Pulse Sensing

BACKGROUND AND RATIONALE

Atherosclerosis and Stress Echocardiography

Atherosclerosis is a chronic disease which involves the development of lipid-rich lesions, called plaques, in the major arteries, such as the coronary and carotid arteries (181). Plaques in the coronary arteries can result in impairment of cardiac function as they can rupture and block the artery or limit the amount of blood that can flow through to the heart. This is referred to as ischemic heart disease which includes conditions such as angina or myocardial infarction.

In patients suspected of having ischemic heart disease, stress echocardiography is used to detect possible underlying coronary artery atherosclerosis and determine its severity. During this test ultrasound, non-invasive imaging which poses no risk to the patient, is used to monitor cardiac function while a drug is infused to speed up the heart rate or the patient is exercised to a standardised regime. If the heart muscle cannot receive enough blood during the stress due to flow limitation in a coronary artery the myocardium starts to become ischemic and is seen on echocardiography as an impairment of the heart wall pumping during the fast heart rates (when the heart should be pumping more forcefully). As with all clinical tests stress echocardiography does not have 100% specificity or sensitivity to detect atherosclerosis and there is interest in whether a combination of changes in certain blood markers pre and post-stress with the echocardiography findings could enhance the predictive capability of the stress test. Increases in inflammatory markers in the blood, such as Tissue Necrosis Factor, N-Terminal pro B type natriuretic peptide and Interleukin 6, have been associated with the presence and extent of myocardial ischemia during stress echocardiogram, but have not added significantly to diagnostic accuracy (111,182–184).

Extracellular Vesicles and Atherosclerosis

Cell-derived microvesicles (MVs) are sub-cellular membrane vesicles, ranging in size from 0.1μm to 1.0μm, released during cell activation, such as ischaemia, or cell death (185). They, along with exosomes, make up the biomarkers known as extracellular vesicles (EVs). MVs 'bud off' from the cell membrane of the parent cell and, therefore, express the surface markers of the cells they originate from (186). Whereas exosomes are released when multivesicular bodies fuse with the cell memebrane. MVs have been shown to play a role in inflammation, coagulation and endothelial dysfunction, all of which are important processes in cardiovascular disease. Previous studies have suggested that elevated circulating levels of MVs derived from endothelial cells and platelets predict future cardiovascular risk in healthy individuals (187–189). However, it has also been reported that endothelial cell MVs are actually decreased in patients with the most severe heart disease (190).

We have previously investigated the impact of cardiac stress, induced by dobutamine stress echocardiogram (DSE), on MV levels in 119 individuals referred for coronary artery disease investigation (191). We found that MVs derived from platelets, endothelial cells and erythrocytes were significantly elevated immediately following cardiac stress. However, this increase was only seen in individuals without vascular disease. In patients with vascular disease no such increase in MVs during cardiac stress was noted. In this initial study, 15 of the 119 participants went on to have angiogram. Those with a normal angiogram had no increase in MVs, and those with a positive angiogram showed a trend for increased MVs. The findings from the data in this study are novel and suggest MV release is a physiological response that protects cells from stress. During early disease development increased circulating levels may be a marker of greater stress exposure and once disease is established the ability to release additional MVs on stress appears to be lost (191). In order to test this theory, it would be necessary to further investigate levels of EVs during cardiac stress.

Potential Relevance of Extracellular Vesicles to Stress Echocardiography

In our previous study we noted one patient, who had a normal stress echocardiogram but subsequently had an out of hospital arrest and was found to have coronary disease. Retrospective review of the blood biomarkers demonstrated their MV levels had not risen in response to cardiac stress. Another patient was classified as having coronary disease based on their stress echocardiogram, but was subsequently found to have normal coronary arteries. Retrospective review of their MV levels showed an elevation during cardiac stress (191). These observations need to be interpreted with caution as they are case histories, but are compelling, and this pilot data raises the possibility that MV levels may provide additional diagnostic value to a traditional DSE for coronary artery disease. Therefore, we now wish to collect samples to determine biomarker levels, including circulating EV levels, pre and post stress echocardiography to further explore their use in prognosis, diagnosis and monitoring.

Other potential biomarkers to improve Stress Echocardiography

Although the primary aim of the study is to evaluate blood-based biomarkers, there have also been recent innovations in image-based biomarkers. These extract information from the ultrasound images that have been gathered in different ways to try and improve the diagnostic accuracy of the test. These include markers of transformation and strain. However, in contrast to blood-based markers these may vary depending on the imaging protocols and equipment used to collect the stress echocardiograms.

Purpose of the Study

To characterise the change in extracellular MVs and associated blood biomarkers during standardised stress echocardiogram protocols and understand the relevance of this change as a predictive marker of coronary disease.

Currently, stress echocardiogram does not have 100% sensitivity or specificity for detecting atherosclerosis. Patients with false positive results may go on to have further unnecessary more-invasive procedures and patients with false negative results may be missed. Therefore, improving the sensitivity and specificity of these tests would be beneficial.

We will be inviting 8000 adults with suspected coronary artery disease, referred for stress echocardiogram as part of their standard clinical care, to take part. The first phase of these participants (group 1) will be asked to give blood samples during the stress echocardiogram through a cannula (inserted for the standard clinical procedure). We will measure biomarkers, including extracellular vesicles, in blood samples collected before and after the stress echocardiogram. The sensitivity and specificity of these blood biomarkers will be compared to standard clinical review of stress echocardiograms in the full cohort (group 1 and 2)

It is anticipated that biomarker levels in samples collected before, during and after stress echocardiogram may provide the best diagnostic accuracy for this procedure and therefore this study may provide an opportunity to integrate biomarker measurements into routine clinical practice. It is not felt that there are any significant risks to patients taking part in the study.

OBJECTIVES AND OUTCOME MEASURES/ENDPOINTS

Objectives	Outcome Measures/Endpoints
Primary Objective To establish whether the measurement of specific blood biomarkers, including EVs, during a stress echocardiogram, can give additional prognostic value to the procedure (Group 1 only).	Analysis of stress echocardiogram and any subsequent angiogram testing, in conjunction with biomarker levels
Secondary Objectives To determine the impact of cardiac stress on the circulating levels of blood biomarkers, including EVs, which will provide information of the rate of release and clearance from the circulation. To analyse any correlations between the extent of inducible ischemia generated during stress echocardiogram with changes in levels of blood biomarkers. To evaluate the accuracy of image-based analysis of stress echo (Group 1 and 2 combined)	Results of biomarker level analysis before, during and after cardiac stress Results from correlation analysis, between the extent of inducible ischemia and changes in levels of blood biomarkers Analysis of stress echocardiogram and clinical outcome over the next 12 months (Group 1 and 2 combined).

STUDY DESIGN

This is a multi-centre observational study designed to investigate the impact of cardiac stress on levels of circulating biomarkers, particularly extracellular vesicles, and to examine whether the additional measurement of blood biomarkers (Group 1 only) and/or image biomarkers during stress echocardiography can give additional prognostic value to the procedure (Group 1 and 2). Participants are characterized as patients referred for stress echocardiogram testing to identify suspected coronary artery disease.

The study visit will be carried out during the planned routine clinical visit of the participant to the Cardiology Department at the (1) John Radcliffe Hospital OUH NHS Foundation Trust; (2) the Milton Keynes University Hospital NHS Foundation Trust; (3) the Royal United Hospital Bath NHS Foundation Trust; (4) the Great Western Hospital NHS Foundation Trust; (5) the Royal Berkshire Hospital NHS Foundation Trust; (6) the Wycombe Hospital, Buckinghamshire Healthcare NHS Foundation Trust; (7) the St George's University Hospital NHS Foundation Trust; (8) the Northwick Park Hospital, London North West University Healthcare NHS Trust; (9) the Bristol Heart Institute, University Hospitals Bristol NHS Foundation Trust; (10) the West Middlesex University Hospital, Chelsea and Westminster Hospital NHS Foundation Trust; (11) the Royal Blackburn Hospital, East Lancashire Hospitals NHS Trust; (12) the Broomfield Hospital, Mid Essex NHS Hospital Services NHS Trust; (13) the Royal Albert Edward Infirmary Hospital, Wrightington, Wigan and Leigh NHS Foundation Trust; (14) the Tameside Hospital, Tameside and Glossop Integrated Care NHS Foundation Trust; (15) the Bradford Royal Infirmary Hospital, Bradford Teaching Hospitals NHS Foundation Trust; (16) the Calderdale Royal Hospital and Huddersfield Royal Infirmary, Calderdale and Huddersfield NHS Foundation Trust; (17) the North Middlesex University Hospital NHS Trust; (18) the Northumbria Healthcare NHS Foundation Trust for their stress echocardiogram; (19) St Thomas' Hospital London, Guy's and St Thomas' NHS Foundation Trust; (20) King's College Hospital, King's College Hospital NHS Foundation Trust; (21) St Bartholomews Hospital London, Barts Health NHS Trust; (22) Peterborough City Hospital and Hinchingbrooke Hospital, North West Anglia NHS Foundation Trust; (23)

Ipswich Hospital NHS Trust ; (24) Lancashire Cardiac Centre, Blackpool Victoria Hospital, Blackpool Teaching Hospitals NHS Foundation Trust; (25) Northampton General Hospital, Northampton General Hospital NHS Trust; (26) Nottingham City Hospital, Nottingham University Hospitals NHS Trust; (27) Pinderfields Hospital, The Mid Yorkshire NHS Trust; (28) Poole Hospital, Poole Hospital NHS Foundation Trust; (29) Yeovil District Hospital, Yeovil District Hospital NHS Foundation Trust; (30) Basingstoke and North Hampshire Hospital, Hampshire Hospitals NHS Foundation Trust. The investigation will be carried out by the Consultant to which the patient has been referred and the research participation will not alter the normal clinical progress of the participant through their investigation. The additional components to the stress echocardiogram, including asking potential participants, with their consent, to take blood samples before and after the DSE (Group 1 only) and to record the details of the clinical procedure, will be performed by trained study investigators, identified on the delegation log. With the informed consent of the participant, the researchers will also collect medical information relevant to the research from the referral letter and medical records regarding the medical history of the participant, such as history of high blood pressure, current medication, height and weight, and evidence of coronary artery disease from previous investigations. Participants will also be asked to consent to further follow up by review of their medical notes and telephone call around after 1 year, in order to collect evidence of cardiovascular events or relevant investigations. Refer to section 7.5 for further detail.

PARTICIPANT IDENTIFICATION

Study Participants

The overall study group will consist of 8000 participants referred for stress echocardiogram by their clinical consultant in order to establish likelihood of CAD.

Inclusion Criteria

Participant is willing and able to give informed consent for participation in the study.

Male or Female, aged 18 years or above.

Being investigated with stress echocardiography for suspected coronary disease

Exclusion Criteria

The participant may not enter the study if ANY of the following apply:

Unable or unwilling to give valid consent for participation in the study.

STUDY PROCEDURES

This section describes procedures which will be carried out.

Recruitment

Participants will be recruited who are currently under investigation or follow up for cardiovascular disease. The routine clinical care team will identify potential participants that fit the inclusion/exclusion criteria from the waiting lists for routine outpatient stress echocardiograms or, if stress echos are performed as part of a patient's visit to the rapid access chest pain clinics, from these clinics.

Potential participants for routine outpatient stress echos will receive a letter of invitation, signed by their consultant, about the study posted to the address provided on the request form for the stress echocardiogram, at least one week prior to their appointment. A detailed participant information sheet will be provided at the same time. The participant will have an opportunity to contact the study team prior to the study visit to ask further questions about the study and the study team will also meet with the participant when they arrive for the procedure to discuss whether they are interested in taking part in the study. This is when the informed consent process will take place. Participants may also be recruited directly from the rapid access chest pain clinics should their clinician feel that a stress echo is indicated to further investigate their symptoms. They will be provided with the invitation letter and PIL during their appointment.

Informed Consent

The participant must personally sign and date the latest approved version of the Informed Consent form before any study specific procedures are performed.

Written and verbal versions of the Participant Information and Informed Consent will be presented to the participants detailing no less than: the exact nature of the study; what it will involve for the participant; the implications and constraints of the protocol; the known side effects and any risks involved in taking part. It will be clearly stated that the participant is free to withdraw from the study at any time for any reason without prejudice to future care, and with no obligation to give the reason for withdrawal.

In most cases, when the patient is referred for a stress echo, the participant will be allowed as much time as wished, and not less than 24 hours, to consider the information, and the opportunity to question the Investigator, their GP or other independent parties to decide whether they will participate in the study. There may be occasions, due to the nature of certain clinics, such as rapid access chest pain clinics, where it is not possible to allow 24 hours to read the PIL before their stress echo is performed. In these cases, the PIL will be given to the patient as soon as possible after referral for stress echo, and the participant will be given as much time as possible to consider the study information, before giving informed consent. The minimum time period between referral from the rapid access clinic to stress echo procedure will be 1 hour. It will be emphasised to the participant that they can decide to withdraw from the study at any time. The participant does not have to state a reason for wishing to withdraw from the study.

Written informed Consent will then be obtained by means of participant dated signature and dated signature of the person who presented and obtained the Informed Consent. The person who obtained the consent will be suitably qualified and experienced, and have been authorised to do so by the Chief Investigator/Principle Investigator. A copy of the signed Informed Consent will be given to the participant. The original signed form will be retained at the study site.

Study Visit for stress echocardiogram

The research study visit will be carried out during the planned clinical visit of the participant to the Cardiology Department at the clinical sites for their stress echocardiogram. The procedure will be carried out by the Consultant to which the patient has been referred and the research participation will not alter the normal clinical progress of the participant through their investigation. The additional components to the stress echocardiogram will be performed by trained study investigators and will be as follows:

Consent (15mins) - The participant will be provided with information about the study prior to their attendance for their stress study. On arrival they will have the opportunity to ask further questions and if they agree to take part will be asked to complete a consent form.

Blood tests (3-4 minutes at start and end of stress protocol and prior to discharge of patient – 10 minutes total) (Group 1 only)- As part of the DSE patients have a small plastic tube or 'cannula' inserted into a vein, usually in the back of the hand to allow administration of drugs. For the research study we will use this cannula to collect about 40mls of blood (8 teaspoons) before the procedure starts and another 40mls immediately at the end of the stress protocol. The cannula is normally left in the vein for an hour after the procedure while the patient recovers and a further blood 40ml sample will be taken prior to removal of the cannula.

Collecting images and results from clinical stress echocardiogram – Clinical details of the outcome of the stress echocardiogram will also be recorded by the researcher. This is all carried out as part of routine clinical care. With the consent of the patient, the ultrasound images will be anonymised and marked with a patient specific study identifier for storage by the research team. This will allow subsequent analysis of the ultrasound images to review accuracy of clinical interpretation and determine whether other features of myocardial function relate to biomarker variation

Medical record examination - This will be undertaken during the study visit by the researcher who will collect information from the referral letter and medical records regarding the medical history of the participant, such as history of high blood pressure, current medication, height and weight, evidence of coronary disease from previous investigations such as coronary angiography.

Baseline Assessments

Biomarker measurements (Group 1 only)

Blood Sample Collection & Storage

For EV measurement, citrate blood samples will be obtained peripherally prior to, immediately following and 1 hour post stress echocardiography in all patients. Citrate blood samples will be centrifuged at 3000g for 15 minutes within 10 minutes of the blood collection. Plasma will then be centrifuged again at 13000g for 2 minutes to produce platelet-poor-plasma. Aliquots of platelet-poor-plasma will be frozen immediately and stored at -80°C. For other biomarker measurement, EDTA, heparin and serum blood samples will be obtained. Samples will be centrifuged at 3000g for 15 minutes within 10 minutes of the blood collection. Plasma and serum will be frozen immediately and stored at -80°C. The samples of plasma, serum and DNA will be stored for at least 10 years and may be used in future ethically approved studies as our understanding of blood vessel function grows.

Measurement of Microvesicles (Group 1 only)

Aliquots of platelet-poor-plasma will be thawed at room temperature, and washed twice at 18,000g for 30 minutes. Levels of MVs derived from platelets, red blood cells, white blood cells and endothelial cells (cells which line the blood vessels) will be assessed by staining with monoclonal antibodies and acquiring data by flow cytometry. Platelet-poor-plasma sample will be incubated with the appropriate monoclonal antibody for 30 minutes at room temperature, protected from light, followed by addition of 900µl PBS-Calcium. Data will be acquired using a BD FACSCalibur® (Becton Dickinson, Oxford UK).

Levels of circulating EVs will also be measured by commercial coagulation assays to define the functional capability of these vesicles, and by novel technologies including Nanoparticle Tracking Analysis (NTA) and Tunable Resistive Pulse Sensing (TRPS) to accurately assess the size of the vesicles being measured.

Other Biomarker Measurement (Group 1 and 2)

Other biomarkers relevant to cardiovascular disease or ischemia will also be measured before and after stress echocardiogram. The biomarkers measured may change as technology develops and new biomarkers are discovered. All biomarker analyses will be performed blind to the patient's stress echocardiography and angiogram results.

Subsequent Visits

Follow up (Group 1 and 2)

All participants will be asked to consent to further follow up for one year following participation. This follow up will consist of the following.

Medical record review – medical records relating to the participant will be reviewed annually to identify if the patient has been admitted. Evidence of cardiovascular events including admissions with chest pain, shortness of breath, heart failure or cardiovascular procedures, including coronary angiography, vascular investigations or surgery. Coronary angiography will be used as the gold standard to identify coronary artery disease, and therefore this data will be essential to determine if the measurement of biomarkers improves the diagnostic accuracy of stress echocardiogram. In order to reduce the burden to these participants, where their records are kept in the Hospital where the stress echocardiogram took place, researchers will look these up without contacting participants

Participant contact (via 10 minute phone call)– participants will be contacted after one year by telephone call to determine whether they have been admitted to hospitals outside the clinical sites region with related events. Information on their hospital admission will be obtained from the participant. Where possible, attempts will be made to obtain precise information from the admitting hospital. As it cannot be predicted where these admissions will occur, specific ethical or regulatory requirements that need to be obtained will be considered on a case-by-case basis during the study progress.

With the consent of the participants, the records held with NHS digital and the Office for National Statistics (ONS) will be also reviewed for the duration of the study.

Discontinuation/Withdrawal of Participants from Study

Each participant has the right to withdraw from the study at any time. In addition, the Investigator may discontinue a participant from the study at any time if the Investigator considers it necessary for any reason including:

Ineligibility (either arising during the study or retrospectively if not known at time of screening)

Withdrawal of Consent

Loss to follow up

The reason for withdrawal will be recorded in the CRF.

Definition of End of Study

The end of the trial will be when the last analysis of the last blood sample has been carried out or when the final follow up phone call has been made, whichever is latest.

STATISTICS AND ANALYSIS

Description of Statistical Methods

Statistical analysis will be performed using GraphPad Prism 5 and SPSS version 15.0 (SPSS Inc., Chicago, IL, USA). Measurements will be expressed as the mean + standard deviation (SD) or median and interquartile range (IQR) if not normally distributed. Distribution of the variables will be assessed using the Shapiro-Wilk test. Variables with normal distribution will be compared using the paired Student's t-test. Variables with not normal distribution will be compared using Wilcoxon test. Limits of agreement between techniques will be presented and assessed visually using the Bland-Altman plot (192). Inter-observer and intra-observer variability will be calculated. Correlation between normally distributed variables will be compared using Pearson product moment correlation coefficient. In the case of non-normally distributed data Spearman's rank correlation will be used. Differences in demographics and baseline biomarker levels between patient groups will be compared by unpaired t-tests or Mann Whitney tests (if non-parametrically distributed). Changes in biomarker levels following stress echocardiography will be assessed by paired non-parametric T-tests (Wilcoxon matched-pairs signed rank tests). A p-value of <0.05 will be considered statistically significant. Statistical support has been provided for the calculation of number of participants and will continue to be provided by Dr Reza Oskrochi, Senior Lecturer in medical statistics and director of medical statistics at Oxford Brookes.

The Number of Participants

Following interim analysis of clinical outcome data in our patient group, we now have a more accurate assessment of prevalence of disease within the population. We have used this to determine sample size required for both the test of our blood-based markers as diagnostic tools and also of our comparator group of imaging-based measures. Follow up data from our current cohort indicate the likely prevalence of functionally significant coronary artery disease in EVAREST will be ~14% (27/189). Based on Buderer et al (193–195), to demonstrate a greater than 90% sensitivity with 95% confidence in a condition with 14% prevalence we will need 521 participants and for 90% specificity, with 95% confidence, 85 participants. Therefore, we will aim to collect up to 521 participants for blood-based biomarker assessment (Group 1).

In addition, we have estimated sample size for confidence with regard the accuracy of interpretation based on the images alone, including more novel image analysis techniques. To have 95% confidence that current practice is achieving at least 85% sensitivity and specificity we will require a sample size of 987 participants for sensitivity and 160 for specificity. However, as this is a multi-site study, we also want to allow exploration of the influence of other factors on stress echo accuracy performed based on image interpretation. To achieve this aim, we will continue recruitment up to a total of 8000 participants within the study sites. Therefore, we will continue recruitment for Group 2 only. This will ensure that, for example, we can separately study the 20% of patients (1600 patients) we expect to have stress echo performed without use of contrast and still maintain the same power to determine whether current practice achieves 85% sensitivity. Similarly, we will be able to take into account in analysis other factors such as study site, geographical location, image quality, patient demographics, stress echo protocol and echocardiography machine type.

Analysis of Outcome Measures/Endpoints

Coronary events are defined as Primary (Myocardial Infarction or Cardiac Death) or Secondary (hospital admission for angina confirmed as CAD by angiography with plaque >50% in left main stem or >70% in other epicardial vessels or positive non-invasive imaging). The biomarker levels, including EVs and imaging biomarkers, from positive and negative stress echocardiograms, which resulted in coronary angiography (gold standard for CAD diagnosis), will be analysed in order to determine whether the measurement of biomarkers gives additional prognostic value to the procedure. Results of biomarker level analysis before, during and after cardiac stress will determine the impact of cardiac stress on biomarkers and results from correlation analysis will determine if biomarker levels are dependent on the extent of disease.

DATA MANAGEMENT

Access to Data

Direct access will be granted to authorised representatives from the Sponsor or OUH NHS Trust for monitoring and/or audit of the study to ensure compliance with regulations.

Data Recording and Record Keeping

The study staff will ensure that the participants' anonymity is maintained. The participants will be identified only by initials and a participant ID number on the CRF and any electronic database. A study log will be kept on a secure drive with the patient's name and hospital number, to allow for the medical record review. Pseudo-anonymised data will be stored within a firewall and password-protected computer system within a swipe-card secured building. Any paper copies of identifying information collected will be kept separately in a lockable filing cabinet in a swipe-card secured building. Images that are processed by other parties for analysis purposes will only contain the study ID of the participant and will therefore be fully anonymised to the receipient.

Sample Storage

All samples will be retained in a secure environment for future analysis and will be stored in the University of Oxford Clinical Cardiovascular Research Facility at the John Radcliffe Hospital or the Clinical Immunology Laboratory, OUH NHS Trust under the custodianship of the Chief Investigator. The samples will be stored for at least 10 years and may be used in future research, as our understanding of how various blood markers vary during stress echocardiography evolves. Samples will be retained while an appropriate research ethical approval is in place. If samples will be analysed by collaborators outside the University of Oxford an appropriate material transfer agreement will be put in place and no identifying information will leave the site.

Data Sharing

With the consent of the participant, personal data (namely participant's NHS number, Date of Birth and post code) will be shared securely between a local site, the main study site at the University of Oxford and NHS digital/Office for National statistics (ONS) in order to request health outcome data from participants. Data will be transferred from each of the local sites to the main study site at the University of Oxford which will act as data controller and request such information from NHS digital/ONS. NHS digital/ONS will return the requested information to the data controller at the main site in pseudo-anonlymised form which is only identifyable to the data controller.

QUALITY ASSURANCE PROCEDURES

The study may be monitored, or audited in accordance with the current approved protocol, GCP, relevant regulations and standard operating procedures. Data will be evaluated for compliance with the protocol and accuracy in relation to source documents. The study will be conducted in accordance with procedures identified in the protocol and instrumentation for study measures monitored for accuracy. Study Investigators are trained in Good Clinical Practice and the study will be internally monitored for compliance with Good Clinical Practice by the Division of Cardiovascular Medicine Clinical Research Co-ordinator. The study may be audited by the authorised members of the University of Oxford as sponsors of the study.

ETHICAL AND REGULATORY CONSIDERATIONS

Declaration of Helsinki

The Investigator will ensure that this study is conducted in accordance with the principles of the Declaration of Helsinki.

Guidelines for Good Clinical Practice

The Investigator will ensure that this study is conducted in full conformity with relevant regulations.

Approvals

The protocol, informed consent form, participant information sheet and any proposed advertising material will be submitted to an appropriate Research Ethics Committee (REC), and host institution(s) for written approval.

The Investigator will submit and, where necessary, obtain approval from the above parties for all substantial amendments to the original approved documents.

Reporting

The CI shall submit once a year throughout the study, or on request, an Annual Progress report to the REC Committee, host organisation and Sponsor. In addition, an End of Study notification and final report will be submitted to the same parties.

Participant Confidentiality

The study staff will ensure that the participants' anonymity is maintained. The participants will be identified only by initials and a participant's ID number on any electronic database. Documents will be stored securely and only accessible by study staff and authorised personnel. The study will comply with General Data Protection Regulation (GDPR) and the associated Data Protection Act, 2018, which requires data to be anonymised as soon as it is practical to do so.

Expenses and Benefits

No additional visits to the hospital, outside of the patient's standard clinical care, will be required, therefore expenses will not be covered by the study team.

Language provisions

Some investigators speak additional languages and can translate where appropriate. It will not be feasible to provide other translation services, and so where a potential participant cannot adequately understand the study information they will not be included in the study.

Storage of Blood Samples and Images

It will be made clear to participants as part of the informed consent process how images and blood samples they provide will be used for this study and stored..

Genetic Testing

It is possible that some samples obtained from Group 1 participants will be used for genetic research. This research may be conducted by the study research team or collaborating research teams. The genetic tests would involve looking at common variations in genes that affect how blood vessels work. We do not propose to test for inherited genetic diseases. There is no evidence to suggest that the results of these genetic studies would have significant implications for the participant.

It will be explained to participants during the consent process that blood samples may be used for genetic testing. It will be explained that the genetic tests would involve looking at common variations in genes that affect how blood vessels work and that tests for inherited genetic diseases will not be performed. Genetic information is kept strictly confidential and results of genetic analysis will not be disclosed to participants or their treating physicians.

Other Ethical Considerations

If a participant loses their capacity to consent to the study as determined during the follow-up phone call, they will be withdrawn from the study

The investigator will ensure the participants understand their involvement in the study, including review of medical notes, review and use of images from the stress echocardiogram and follow-up contact about relevant investigations and admissions.

As we will be contacting participants one year after the study, there is the possibility that the participant may have passed away. In order to minimise the risk of causing any distress to family members, we will consult their medical records and the Personal Demographic Services in the SPINE database to identify such participants, prior to making the follow up call.

FINANCE AND INSURANCE

Funding

Main study and blood biomarkers analysis is funded by an NIHR HEE Healthcare Science Research Fellowship (NIHR-HCS-P13-04-001) with additional funding support for image processing and analysis through the Oxford Cardiovascular Clinical Research Facility (Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford) and a Sponsored Research Agreement with Ultromics Ltd (R51323/CN001, a University spinout company) and Lantheus Medical Imaging inc.

Insurance

The University has a specialist insurance policy in place which would operate in the event of any participant suffering harm as a result of their involvement in the research (Newline Underwriting Management Ltd, at Lloyd's of London, policy numbered: WD1200463). NHS indemnity operates in respect of the clinical treatment that is provided.

PUBLICATION POLICY

The Investigators will be involved in reviewing drafts of the manuscripts, abstracts, press releases and any other publications arising from the study. The Investigators will seek appropriate reviews of all manuscripts, abstracts, oral/slide presentations, book chapters based on this study before release. Authors will acknowledge that the study was funded by the NIHR. Authorship will be determined in accordance with the ICMJE guidelines and other contributors will be acknowledged.

9.2 Appendix B- Rainier Clinical Trial Protocol

PROTOCOL SYNOPSIS

TITLE	OHSU RAINIER Trial
SPONSOR	Ultromics Ltd
FUNDING ORGANIZATION	Ultromics Ltd
STUDY SITES	Oregon Health Science University Hospital, Portland, OR
RATIONALE	The aim of the trial is to evaluate the diagnostic performance of the EchoGo Pro v1.0 in the prediction of clinical outcomes and/or the assessment of coronary artery disease by coronary angiography or CT coronary angiography in the 1-year following stress testing. Furthermore, to investigate the clinical benefit of EchoGo Pro v1.0 by determining whether the use of results from EchoGo Pro v1.0 by an observer during analysis of stress echocardiography can reduce inconsistencies in clinical SE interpretation through a reduction in variance and/or a segmental increase in accuracy. A positive clinical outcome is defined as evidence of obstructive coronary artery disease (> 70% epicardial coronary artery stenosis with or without consequent percutaneous coronary intervention [PCI], or ≤ 70% epicardial coronary artery stenosis but with low Fractional

	Flow Reserve or low Instantaneous Wave-Free Ratio prompting PCI) during invasive coronary angiography done either electively or following acute coronary syndrome; percutaneous or surgical coronary revascularization; or death with myocardial infarction listed as primary cause. We will be accessing the investigational database from OHSU in the United States for the review of longitudinal outcome data at 1-year post SE testing on 370 cases. Applicable data will be collected retrospectively and in a consecutive manner over the time frame between 2011 and 2017 inclusive.
STUDY DESIGN	A single center retrospective study
PRIMARY OBJECTIVES	To assess the diagnostic performance of EchoGo Pro v1.0 in predicting 12-month clinical outcomes in patients who have been referred for SE testing for suspected CAD (EchoGo Pro Performance Study). To assess the clinical benefit of using EchoGo Pro v1.0 to increase clinical SE interpretation (Clinical Benefit Study).
PRIMARY ENDPOINTS	The accuracy of EchoGo Pro v1.0 in predicting 12-month clinical outcomes will be assessed using accuracy, sensitivity and specificity against criteria as defined in the literature, using one-sample portions test. A positive outcome is defined as evidence of obstructive coronary artery disease (> 70% epicardial coronary artery stenosis

	with or without consequent percutaneous coronary intervention [PCI], or ≤ 70% epicardial coronary artery stenosis but with low Fractional Flow Reserve or low Instantaneous Wave-Free Ratio prompting PCI) during invasive coronary angiography done either electively or following acute coronary syndrome; percutaneous or surgical coronary revascularization; or death with myocardial infarction listed as primary cause (EchoGo Pro Performance Study). The clinical benefit of using EchoGo Pro will be measured as a segmental increase in SE interpretation accuracy, sensitivity and specificity across reader groups using one-sample portions test (Clinical Benefit Study).
NUMBER OF SUBJECTS	370
INCLUSION CRITERIA	Male or female aged 18 years or above. Patient has undergone stress echocardiography for suspected coronary disease 1 year prior to the study start date. Results of coronary angiogram (if performed) and medical records are obtainable for the 1-year period following the SE test.
EXCLUSION CRITERIA	Age < 18 years Non-diagnostic SE (failure to reach target heart rate /double product, poor echocardiographic windows and other

	technical considerations)
TEST PRODUCT	EchoGo Pro v1.0 software EchoGo Pro v1.0 is a stand-alone software device. Input to the software will be anonymized DICOM 3.0 compliant SE datasets and results transferred under the supervision of a cardiologist. Images will be processed using EchoGo Pro v1.0. Images undergo a quality review by Ultromics trained, and accredited echocardiographer/ sonographer. A report will be generated including the recommendation for a higher or lower risk of prognostically significant CAD.
ESTIMATED DURATION OF STUDY	6 Months
CONCOMITANT MEDICATIONS	Allowed - No restrictions

BACKGROUND

Product Background

Ultromics is a University of Oxford spin-out combining advanced computational techniques, rare longitudinal datasets, and clinical insights to provide quantitative diagnostics for cardiovascular disease (CVD), the single leading cause of death worldwide. Coronary artery disease (CAD) is responsible for the deaths of over eight million people a year globally; it accounts for 25% of all deaths in the US and, if present trends continue, it is forecast that over half of all men and one in three women over forty will develop CAD over the course of their lives.

The most widely used diagnostic imaging technique for CVD is echocardiography, whereby the heart is imaged using ultrasound. In the diagnosis of CAD, a common procedure is stress echocardiography (SE). During this test, echocardiography is used to assess the function of the left ventricle (LV) whilst the heart rate is increased following a standardised regime, either by dobutamine infusion or exercise. If the heart muscle is not sufficiently perfused under stress, due to flow limitation in a coronary artery, the myocardium starts to become ischemic and is seen on echocardiography as an impairment of the LV wall pumping motion at higher heart rates, known as wall motion abnormalities.

The current clinical standard for diagnosis of CAD in stress echocardiography is wall motion scoring (WMS). This consists of the consultant cardiologist assessing LV wall motion in different segments of the LV during SE and assigning a discrete score between one and four to each segment. These scores are combined to produce an overall index which is used to determine the outcome of the test. WMS is the current clinical standard for diagnosing CAD in SE clinics and is a qualitative and highly subjective process, relying heavily on the physician's expertise. Incorrect diagnoses can have serious consequences. False negatives, whereby presence of disease is missed, may lead to a coronary event that may have been prevented by intervention taking place. This could result in death or costly hospitalization and treatment. False positives, whereby a patient is incorrectly diagnosed as needing intervention, may lead to unnecessary medication or invasive procedures such as angiography. Thus, it is critical that this diagnostic test is both accurate and reliable.

Results of applied meta-analysis to 62 studies and 6881 patients to assess accuracy of dobutamine stress echocardiography (DSE) estimate a sensitivity of 81% and a specificity of 82% (64). This is similar to the highest values obtained from other functional imaging tests, such as myocardial perfusion imaging (196) and stress cardiac magnetic resonance imaging perfusion (166). The interpretation of SE images is highly subjective, and therefore, although accuracy of exercise and pharmacologic SE has been demonstrated to be independent of age and gender (197), the interpretation can vary greatly depending on a variety of factors, such as image quality and operator training (163). For the detection of patients with CAD, sensitivity has been reported to range from 72% to 97%, and specificity ranges from 64% to 100% (197). This subjectivity can also result in high levels of inter- and intra-observer variability, leading to different results even when the same physician interprets a scan. One study looked at inter-institutional agreement across five centers and 150 patients, using experienced stress echocardiographers for DSE. It was found that, for normal or abnormal results, there was 73% agreement for at least four out of the five centers, with a mean kappa value of 0.37 (198). The percentage agreement was higher for CAD free patients (82%) or three-vessel CAD (100%) but was lower for one- or two-vessel disease (61% and 68%, respectively). Finally, agreement for the highest quality images was 100%, but only 43% for those with the lowest quality images. These findings demonstrate that numerous factors can contribute to the high level of variation. Furthermore, the Medical Advisory Secretariat (Canada; MAS) released an evidence-based analysis of stress echocardiography for the diagnosis of coronary artery disease (Ontario Health Technology Assessment Series 2010; Vol. 10, No. 9;) suggesting that publication bias and selection bias in older studies with strict inclusion and exclusion criteria

have resulted in optimistic reports of SE diagnostic performance, and report that true SE accuracy (in studies the include both dobutamine and exercise SE) could be between 76-80% and 75-90% sensitivity and specificity, respectively.

Ultromics aims to improve diagnosis of CVD using echocardiography by increasing the accuracy and reducing the variability of the information provided to a physician when assessing a patient, with the first application being the improvement of CAD detection using SE. The proposed workflow, described below, will allow the physician, a trained cardiologist, to receive additional information useful in forming their diagnosis. An example workflow is described below:

A physician performs a SE test following standard protocols which adhere to the American Society of Echocardiography (ASE) guidelines for the performance, interpretation, and application of Stress Echocardiography (175).

DICOM SE datasets and results are transferred to Ultromics using an approved third-party transfer system which also anonymizes the data.

Quality control is performed on the received anonymized dataset to check that it meets the ASE guidelines (175), all of the required views and stages are present and correctly labelled, and that image quality is sufficient for LV tracing to be performed.

Contours are automatically detected.

Auto-contours are presented on the screen.

A team of trained operators review and approve of the automatically detected endocardial border of the LV on the images.

From these regions of interest, geometric parameters are calculated that are known to relate to LV functionality and fed into a classification model that has been previously trained on datasets with known outcomes.

The output of the pre-trained model gives a binary recommendation as to whether the data are indicative of functional CAD or not. A report with the binary recommendation for a higher or lower risk of prognostically significant CAD is generated.
The report is returned using the same approved third-party transfer system and is used by the physician as an aid to diagnose CAD.

Ultromics is seeking regulatory clearance for the stand-alone software application that will be used for steps 3-8 to process anonymized data, namely EchoGo Pro v1.0. Ultromics aims to provide a thorough evaluation of the diagnostic performance of EchoGo Pro v1.0 and to determine whether the use of the EchoGo Pro report provides clinical benefit by reducing the inconsistencies in the interpretation of SEs by physicians.

Study Rationale

The trial data will be processed using EchoGo Pro v1.0 under identical conditions to those that are proposed for the device. There will be two primary objectives of the study – to evaluate diagnostic performance (EchoGo Performance Testing) and the clinical benefit of the device (Reader Variability).

EchoGo Pro Performance Study: The EchoGo Pro recommendation will be compared to confirmed clinical outcomes and the performance of the device will be assessed. Previous reports of standard SE interpretation accuracy vary, however the MAS evidence-base, calculated a realistic sensitivity of 76% and specificity of 75%. Consideration should be given to the fact that the comparator in most prior studies, including the aforementioned meta-analyses, encompass several confounder including purely anatomical measures and coronary angiography demonstrating stenosis > 50% or >70% being used as the cut-off for obstructive CAD. In our study, the comparators are more specific (including functional parameters and clinical outcomes), which represent the current standard of care. We elected to assess outcomes over a follow-up period of 1 year, as the prognostic accuracy of stress echocardiography has been well validated over this duration across multiple patient populations and clinical scenarios.

Clinical Benefit Study: The study will also evaluate the clinical benefit of EchoGo Pro v1.0 by investigating whether the addition of EchoGo Pro reports during interpretation increases the physician interpretation of SEs compared to standard reading practice.

USE/INDICATIONS FOR USE

Intended use

Based on an echocardiogram, indicate whether there is evidence of disease as an aid to diagnosis.

Indication for use

EchoGo Pro v1.0 software provides analysis of DICOM 3.0 compliant stress echo exams for patients undergoing assessment for coronary disease and provides a recommendation indicating whether there is a level of coronary disease which is prognostically significant, to support physician diagnosis (not for primary diagnosis CAD, diagnosis of mild or moderate myocardial ischemia, or localisation of CAD or myocardial ischemia).

To be used with stress-echo protocols include exercise, dobutamine, contrast, and non-contrast that contain A2, A4 and mid-ventricular short-axis views at rest and at peak stress.

Not for the assessment of myocardial perfusion, myocardial viability or valve disease within stress echo protocols.

Limitations

Minimum image quality requirements include endocardial visualization in at least 14 of 16 segments.

Contraindications

Not intended for use without physician interpretation.

RISK/BENEFIT ASSESSMENT

Risk

There is no risk to patient safety during the trial. All data used will be retrospectively gathered and will have been analysed by the patients' direct care team at the time of acquisition. Therefore, there will be no alteration to patient care. The only identified risk to patients would be any accidental disclosure of their medical details, however this has been extensively mitigated by strict procedures requiring all data, including images, to be deidentified by the patients' direct care team before leaving the clinical site.

Benefit

There is no direct benefit to the patients' whose data will be used in this trial. However, the RAINIER trial is expected to demonstrate that the CAD recommendation provided by the proposed device can reduce the variability seen in SE interpretation, particularly across readers at different levels of experience, and potentially improve the accuracy of SE interpretation. This could lead to fewer deaths and cardiac events taking place in at least the 1-year following the test, thereby improving patient well-being and reducing costs in the future.

STUDY OBJECTIVES

Primary Objectives

The coprimary objectives of this study are to:

Evaluate the diagnostic performance of EchoGo Pro v1.0 in predicting 12-month clinical outcomes in patients who have had a SE test for suspected CAD. A positive outcome is defined as evidence of obstructive CAD (> 70% epicardial coronary artery stenosis with or without consequent percutaneous coronary intervention [PCI], or ≤ 70% epicardial coronary artery stenosis but with low Fractional Flow Reserve or low Instantaneous Wave-Free Ratio prompting PCI) during invasive coronary angiography done either electively or following acute coronary syndrome; percutaneous or surgical coronary revascularization; or death with myocardial infarction listed as primary cause (EchoGo Performance Study).

To assess the clinical benefit of using EchoGo Pro v1.0 to increase clinical SE interpretation (Clinical Benefit Study).

STUDY DESIGN

Study Overview

This is a single-center retrospective study conducted on SE datasets received from a US site. Results from the proposed device will be compared to clinical outcomes within a 12 month period after SE for evaluation of CAD, to determine the accuracy of EchoGo Pro v1.0 at predicating clinical outcome, defined as evidence of obstructive coronary artery disease (> 70% epicardial coronary artery stenosis with or without consequent percutaneous coronary intervention [PCI], or ≤ 70% epicardial coronary artery stenosis but with low Fractional Flow Reserve or low Instantaneous Wave-Free Ratio prompting PCI) during invasive coronary angiography done either electively or following acute coronary syndrome; percutaneous or surgical coronary revascularization; or death with myocardial infarction listed as primary cause." (EchoGo Performance Study).

In addition, to evaluate the use of EchoGo Pro v1.0 in a clinical context, an assessment of physicians interpretation of SE images with and without EchoGo Pro v1.0 will be conducted (Clinical Benefit Study).

Definition of End of Study

The end of the trial will be when the final analysis of the final dataset is complete.

SUBJECT SELECTION

Study Population

Data included for analysis in this study will be retrospective data from male and female patients aged 18 years or above who received a stress-echocardiography test for suspected coronary artery disease a minimum of one year prior to data collection date to allow for collection of 12-month follow-up data (e.g. invasive angiogram results/hospital records).

Given the retrospective nature of the study, high sample size, and negligible risk to involved subjects, waiver of the requirement of individual subjects' consent will be requested from involved Institutional Review Boards (IRBs). No data collection will take place until this waiver is approved.

Inclusion Criteria

Male or female, aged 18 years or above

Patient has undergone stress echocardiography for suspected coronary artery disease 1 year or more prior to data collection

Results of coronary angiogram (if performed) and medical records are obtainable for the 1-year following the SE test

Exclusion Criteria

Age < 18 years

Non-diagnostic SE (failure to reach target heart rate/double product, poor echocardiographic windows and other technical considerations)

Medical History

Relevant medical history, including history of current disease, other pertinent respiratory history, and information regarding underlying diseases will be recorded.

STUDY PROCEDURES

This section describes procedures which will be carried out by the site during the trial.

Study Database

The data for this study will be obtained from the Oregon Health and Science University (Portland, Oregon) database of SEs (both pharmacologic and exercise), which were conducted for the assessment of suspected CAD and include definitive SE reports. Retrospective collection of data will be consecutive within the inclusion and outside the exclusion criteria as detailed in sections 6.2 and 6.3.

This study will be designed as a case-control study, with enrichment with outcome positive cases and case controls will be the outcome negative cases. Given the SE database structure and logistics employed at the Oregon Health and Science University, the final study database will be consecutively complied from 2011 and 2017, inclusive, using the following search terms: “positive” and/or “high risk”, and “negative” and/or “low risk”. This will ensure that all cases are collated and remain representative of the full clinical cohort.

Once the study database has been defined, the SE and the associated clinical outcome data will be collected up to a year after the SE for that individual and be reviewed consecutively and chronologically, working backwards from June 30 2017 until the pre-specified number of patients negative for CAD and positive for CAD have been identified according to the assessment of patient outcomes. Cases will be excluded if they are ineligible according to the inclusion exclusion criteria (section 6). Records will be kept of all exclusions.

Medical record review

Medical record review will be performed by the site to identify if the patient has been admitted into the hospital with acute coronary syndrome within one year following the SE procedure. Evidence will be recorded for incidences of cardiovascular events including admissions with chest pain, shortness of breath, heart failure or cardiovascular procedures, including coronary angiography, vascular investigations or surgery, or a record of death with myocardial infarction listed as primary cause. Coronary angiography will be used as the gold standard to identify CAD, and therefore this data will be essential to determine the diagnostic accuracy of SE and the recommendation from the report. Patients who have a negative SE and do not undergo invasive angiography or other investigation during the one year following SE will be considered to have a negative outcome if there is an absence of any cardiac events such as acute coronary syndrome or death from a cardiovascular cause recorded in their medical records Patient records for this group will be reviewed in the OHSU electronic medical record (EMR) system Epic (Epic Systems Corporation, Verona, WI) and, to mitigate against lack of cardiovascular events being a result of loss to follow up to other care centres, these patient records will also be reviewed in the Care Everywhere network of the Epic EMR platform. The Care Everywhere network integrates health care records, including outpatient visits, inpatient/emergency department visits, and all diagnostic testing, for hospitals and health care systems using the Epic EMR platform. In the absence of any medical encounter records at all, the social security death index will be queried to check for any death during follow-up.

Assessment of patient outcomes

The patient outcome is obtained by the assessment of angiograms and patient records in the year following the the patient's SE procedure. A positive outcome is defined as evidence of obstructive coronary artery disease (> 70% epicardial coronary artery stenosis with or without consequent percutaneous coronary intervention [PCI], or ≤ 70% epicardial coronary artery stenosis but with low Fractional Flow Reserve or low Instantaneous Wave-Free Ratio prompting PCI) during invasive coronary angiography done either electively or following admission for acute coronary syndrome; percutaneous or surgical coronary revascularization; or death with myocardial infraction listed as primary cause. Where a participant has had an event, or undergone angiography, outcome will be determined by clinical opinion of an individual with appropriate expertise e.g. board-certified clinician or cardiologist who undertook the angiography or reported the mortality. Where this is not available, or there is uncertainty, outcome will be adjudicated by an adjudication committee at OHSU comprised of at least two board certified cardiologists, of which one shall not be involved in project oversight, data collection or analysis. The adjudication committee can exclude patients if, after inclusion, the committee subsequently decides there is uncertainty about diagnosis due to insufficient clinical information to confirm outcome, such as lack of adequate confirmatory invasive angiography data following other non-invasive tests. Records of all exclusions will be kept.

Anonymisation of SE datasets

The SE datasets from all included cases will be sent from the Syngo Dynamics PACS at OHSU to a Laurel Bridge Compass router. The router will be configured to filter on DICOM tags associated with Manufacturer, Make, Model and Stress Echo study definitions to determine the location of the burned in Patient Information. The router will then be configured for each specific study and image type to crop the appropriate x,y quadrants from the image or not to crop the image if there is no patient information present. Additionally, DICOM tags that contain Patient Name, Medical Record Number, Patient ID and Patient Age will be setup for removal by the router, prior to study transfer. The studies will be sent from the Laurel Bridge router to the Ambra Cloud Suite whereby they will be assigned a pseudo-anonymized patient identifier in place of the real name, in a numerical, incremental order. The Ultromics team will access the data from AMBRA and proceed to the quality assessment and EchoGo Pro processing of the data.

Quality Assessment

QA will be carried out at Ultromics on all SE datasets by echocardiographers accredited to American or British Society of Echocardiography standards who have been trained in the Ultromics QA process. Data is checked to ensure the stress echo has been performed consistent with American Society of Echocardiography guidelines (175) including collection of required views , that all stages have been labelled appropriately, have been correctly triggered, have a heart rate that is appropriate to the stage, rather than errors caused by arrhythmias and have adequate endocardial border definition for visual assessment. Data which fail this part of the QA process will be recorded as ineligible for the study (exclusion criteria 2) and be excluded.

Processing Datasets for the EchoGo Performance Study

Echocardiograms will be processed using EchoGo Pro v1.0. The images will first undergo a quality review before the LV endocardial contour is applied automatically to the images and approved by an accredited echocardiographer that has undergone the Ultromics training program. A report will be generated including a recommendation for a higher or lower risk of prognostically significant CAD. The accuracy of EchoGo Pro v1.0 in predicting 12-month clinical outcomes will be assessed by accuracy, sensitivity and specificity calculations using criteria as defined in the literature.

Clinical Benefit Study

The Clinical Benefit Study will be performed with certified readers sourced from the US and/or UK. Reading physicians will be selected by study management team according to their clinical experience in interpreting SEs, defined based on accepted number of reads to achieve and maintain competence (124,175). Readers will be defined as: expert reader (greater than 10 years' experience of advanced echo and >1000 stress echo reads); standard reader (between 5 and 10 years' experience of advanced echo and 500-1000 stress echo reads); newly independent reader (1 to 5 years' experience and 100-500 stress echo reads); newly trained reader (having reached competence with ~100 clinical stress echo reads). Blinded SE datasets will be provided to the physicians, all physicians will read the SE datasets using a standardised SE optimised DICOM and their initial interpretation will be recorded into the eCRF. Then immediately after their first read of the SE datasets the reader will be presented with the EchoGo Pro v1.0 output report and their final interpretation will be recorded (sequential reader study). Physicians' decisions will be compared with the gold-standard of the 12-month clinical outcome data, to determine reader accuracy, sensitivity and specificity under "unaided" and "aided" conditions.

Blinding

All outcomes data will have been collected and recorded in the electronic CRF before any processing of images using EchoGo Pro v1.0 takes place. The eCRF will be user-restricted so that trained operators (echocardiographers) of the EchoGo Pro v1.0 software cannot access the outcome data.

During the reader study the reading physicians will be blinded to all other clinical data, including the invasive angiogram reports and CT angiogram reports, for the patients with the SE datasets. Physicians will also be unable to access each other's results in the eCRF. The reading physicians will be blinded to the enriched prevalence in the study to reduce bias.

DISCONTINUATION AND REPLACEMENT OF SUBJECTS

All data will be screened before inclusion in the study to ensure that it meets the eligibility criteria, therefore there will be no need to exclude cases for missing longitudinal data in medical records or for non-diagnostic SEs. If data are rejected for failing the quality assessment procedures, rejection will be logged, and data will be replaced with further case studies.

STATISTICAL ANALYSIS

Data analysis

The following endpoints will be calculated using standard methods:

Sensitivity;

Specificity;

Accuracy;

The data analysis will target two objectives; (1) demonstrating overall superiority to literature sensitivity and specificity; and (2) demonstrating a segmental increase in “aided” accuracy for each reader vs. “unaided”. Superiority of the device to literature accuracy, sensitivity and specificity and increased performance for each reader will be formally assessed using appropriate statistical methodologies, such as one-sample portions test and non-inferiority/superiority test. Significance will be defined as $p<0.05$.

Sample size justification

Preliminary bench testing provided estimates of sensitivity and specificity for EchoGo Pro of 87% and 81%, respectively. Using one-sample proportion tests, a sample size of 99 CAD positives and 103 CAD negatives would be required to detect a (one-sided) 10% point increase in sensitivity and specificity with a 5% type I error rate and 80% power. Similarly, a sample size of 99 (1:1 ratio CAD positives and CAD negatives) would be needed to detect a (one-sided) 10% point increase in accuracy with a 5% type I error rate and 80% power.

Bench testing suggests that readers have an accuracy, sensitivity and specificity increase of 17±5%, 24±14% and 10±4%, respectively, when using EchoGo Pro v1.0 as a second read. Using one-sample proportion tests, a sample size of 153 CAD positives and 106 CAD negatives would be required to detect a (one-sided) 10% point increase in sensitivity and specificity with a 5% type I error rate and 80% power. Similarly, a sample size of 139 (1:1 ratio CAD positives and CAD negatives) would be needed to detect a (one-sided) 10% point increase in accuracy with a 5% type I error rate and 80% power.

Therefore, our proposed sample size should be sufficient to detect the likely change, assuming similar performance as the bench testing data and a reasonable rate of sample drop-out (i.e. 20%). We are proposing 4 readers at least to take part at the Clinical Benefit study.

PROTOCOL VIOLATION

A protocol violation occurs when the subject, investigator, or Sponsor fails to adhere to significant protocol requirements affecting the inclusion, exclusion, subject safety and primary endpoint criteria. Protocol violations for this study include, but are not limited to, failure to meet inclusion/exclusion criteria and failure to comply with Good Clinical Practice (GCP) guidelines.

The Sponsor will determine if a protocol violation will result in withdrawal of a subject.

When a protocol violation occurs, it will be discussed with the investigator and a Protocol Violation Form detailing the violation will be generated. This form will be signed by a Sponsor representative and the Investigator. A copy of the form will be filed in the site's regulatory binder and in the Sponsor's files.

DATA COLLECTION, RETENTION AND MONITORING

Access to Data

Direct access will be granted to authorized representatives from the Sponsor or OSHU for monitoring and/or audit of the study to ensure compliance with regulations. The Investigator will prepare and maintain adequate and accurate source documents designed to record all observations.

Study personnel at each site will enter data from source documents corresponding to a sample case into the protocol-specific electronic Case Report Form (eCRF) OR paper CRF when the information corresponding to that sample is available.

Subjects will not be identified by name in the study database or on any study documents to be collected by the Sponsor (or designee), but will be identified by site abbreviation, date of study, unique study ID.

For eCRFs: If a correction is required for an eCRF, the time and date stamps track the person entering or updating eCRF data and creates an electronic audit trail. *For paper CRFs:* If a correction is made on a CRF, the study staff member will line through the incorrect data, write in the correct data and initial and date the change.

The Investigator is responsible for all information collected on samples enrolled in this study. All data collected during the course of this study must be reviewed and verified for completeness and accuracy by the Investigator. A copy of the CRF will remain at the Investigator's site at the completion of the study.

Site Data Management Procedures

The data will be entered into a validated database at OHSU. A study log will be kept on a secure database with the medical record review. The site data management group will be responsible for data processing, in accordance with institutional procedural documentation.

The Sponsor will ensure that the participants' anonymity is maintained. The participants will be identified only by site initials, a participant ID number and study UID on the CRF and any electronic database. A study log will be retained on a secure database with the medical record review. Pseudo-anonymised data will be stored within a firewall and password-protected computer system. Any paper copies of identifying information collected will be transferred into an electronic file and uploaded to secured database for retention.

All procedures for the handling and analysis of data will be conducted using good computing practices meeting FDA guidelines for the handling and analysis of data for clinical trials.

Ultromics will follow the company's standard operating procedures for data management. The SE DICOM datasets will be stored within Ambra Cloud Suite and will be processed via EchoGo Pro. Password protected electronic database will be used for the clinical data and to record the physician's interpretation during the reader variability study.

Quality Assurance Procedures

The database is safeguarded against unauthorized access by established security procedures; appropriate backup copies of the database and related software files will be maintained. Databases are backed up by the database administrator in conjunction with any updates or changes to the database.

Availability and Retention of Investigational Records

The Investigator must make study data accessible to the monitor, other authorized representatives of the Sponsor (or designee), IRB/IEC, and Regulatory Agency (e.g., FDA) inspectors upon request. A file for each subject must be maintained that includes the copies of all source documentation related to that subject. The Investigator must ensure the reliability and availability of source documents from which the information on the CRF was derived.

All study documents (patient files, copies of CRFs, Study File Notebook, etc.) must be kept secured for a period of two years following marketing of the investigational product or for two years after centers have been notified that the IDE has been discontinued. There may be other circumstances for which the Sponsor is required to maintain study records and, therefore, the Sponsor should be contacted prior to removing study records for any reason.

Monitoring

The study may be monitored, or audited in accordance with the current approved protocol, GCP, relevant regulations and standard operating procedures. Data will be evaluated for compliance with the protocol and accuracy in relation to source documents. The study will be conducted in accordance with procedures identified in the protocol and instrumentation for study measures monitored for accuracy. Study Investigators are trained in Good Clinical Practice. The study may be audited by the Sponsor or the authorized members of the Oregon Health Science University Hospital, Portland, Oregon.

By signing this protocol, the Investigator grants permission to the Sponsor (or designee), and appropriate regulatory authorities to conduct on-site monitoring and/or auditing of all appropriate study documentation.

Subject Confidentiality

In order to maintain subject confidentiality, only a site number, study UID will identify all study subjects on CRFs and other documentation submitted to the Sponsor. Additional subject confidentiality issues (if applicable) are covered in the Clinical Study Agreement.

ADMINISTRATIVE, ETHICAL, REGULATORY CONSIDERATIONS

The study will be conducted according to the Declaration of Helsinki, Protection of Human Volunteers (21 CFR 50), Institutional Review Boards (21 CFR 56), Obligations of Clinical Investigators (21 CFR 812) and ISO 14155:2016.

To maintain confidentiality, evaluation forms, reports and other records will be identified by a coded number and initials only. All study records will be kept in a secure database. Clinical information will not be released without prior anonymisation. The Investigator must also comply with all applicable privacy regulations (e.g., Health Insurance Portability and Accountability Act of 1996, EU Data Protection Directive 95/46/EC).

Protocol Amendments

Any amendment to the protocol will be written by the Sponsor. Protocol amendments cannot be implemented without prior written IRB/IEC approval.

The protocol will be submitted to an appropriate Institutional Review Board (IRB) / Independent Ethics Committee (IEC), and host institution(s) for written approval.

The Investigator will submit and, where necessary, obtain approval from the above parties for all substantial amendments to the original approved documents.

Institutional Review Boards and Independent Ethics Committees

The protocol will be reviewed and approved by the IRB/IEC of each participating center prior to study initiation. and will keep the IRB/IEC informed as to the progress of the study. The Investigator will obtain assurance of IRB/IEC compliance with regulations.

Any documents that the IRB/IEC may need to fulfil its responsibilities (such as protocol, protocol amendments, or other pertinent information) will be submitted to the IRB/IEC. The IRB/IECs written unconditional approval of the study protocol must refer to the study by exact protocol title and number and should identify the documents reviewed and the date of review.

Protocol modifications or changes may not be initiated without prior written IRB/IEC approval except when the change(s) involves only logistical or administrative aspects of the study. Such modifications will be submitted to the IRB/IEC and written verification that the modification was submitted and subsequently approved should be obtained. The IRB/IEC must be informed of revisions to other documents originally submitted for review; an annual update and/or request for re-approval; and when the study has been completed.

Informed Consent Form

Given the retrospective nature of the study, high sample size, and negligible risk to involved subjects, waiver of the requirement of individual subjects' consent will be confirmed from involved Institutional Review Boards (IRBs).

Publications

The preparation and submittal for publication of manuscripts containing the study results shall be in accordance with a process determined by mutual written agreement among the study Sponsor and participating institutions. The publication or presentation of any study results shall comply with all applicable privacy laws, including, but not limited to, the Health Insurance Portability and Accountability Act of 1996.

Funding

Funded by:

Ultromics Limited.

Registered Office:

James Cowper Kreston 2 Chawley Park, Cumnor Hill, Oxford, England, OX2 9GG

Investigator Responsibilities

By signing the Agreement of Investigator form, the Investigator agrees to conduct the study in accordance with Investigator Responsibilities as defined in §21 CFR 812.100 and 812.110, including but not limited to:

Conduct the study in accordance with the protocol and only make changes after notifying the Sponsor (or designee), except when to protect the safety, rights or welfare of subjects.

Personally conduct or supervise the study (or investigation).

Ensure that the requirements relating to obtaining informed consent and IRB review and approval meet federal guidelines, as stated in § 21 CFR, parts 50 and 56.

Ensure that all associates, colleagues and employees assisting in the conduct of the study are informed about their obligations in meeting the above commitments.

Maintain adequate and accurate records in accordance with §21 CFR 812.140 and to make those records available for inspection with the Sponsor (or designee).

Ensure that an IRB that complies with the requirements of §21 CFR part 56 will be responsible for initial and continuing review and approval of the clinical study.

Promptly prepare and submit reports to the IRB and the Sponsor (or designee) as defined in §21 CFR 812.140.

Seek IRB approval before any changes are made in the research study, except when necessary to eliminate hazards to the patients/subjects.

Sponsor’s responsibilities

To ensure an appropriate process of scientific critique has demonstrated that this research proposal is worthwhile and of high scientific quality.

To ensure any necessary indemnity or insurance arrangements, will be in place before this research starts. Insurance or indemnity policies will be renewed for the duration of the study where necessary.

To ensure that arrangements will be in place before the study starts for the research team to access resources and support to deliver the research as proposed.

To ensure that arrangements to allocate responsibilities for the management, monitoring and reporting of the research will be in place before the research starts.

To ensure that, in line with current GCP, the requirements of the primary objectives may be assessed.

9.3 Appendix C – Contributions to Work

Chapter 2 - Study Population

I set-up the EVAREST study with Dr Lisa Ayers including co-authoring the protocol and submission for ethical approval. I then set-up the first 5 sites and recruited the first 500 patients. This involved consenting patients, collecting, processing and analyzing images, collecting and processing blood samples, completing CRFs and collecting the patient outcomes. William Woodward produced the initial report of patient demographics in the EVAREST population that is included in this chapter.

Chatper 3 - Method Development Autocontouring

This work was done in collaboration with Arian Beqiri. I collected and provided all of the data for this work. I performed the manual QA for all of the images and manually contoured the majority of images. I also provided the quality assessment for the autocontours as they were processed and provided continuous assessment of the images during the iterative improvements of the autocontouring. Arian created the CNN pipeline for the autocontouring and adjusted the CNN model to match my QA feedback until satisfactory performance was achieved.

Chapter 4 - Validation of Autocontouring

This was a clinical trial that I conceived, designed and led. The key personnel in the work alongside myself was Andrew Parker who helped calculate the majority of the

features from the autocontour and Gary Woodward who helped construct the performance testing plan.

Chapter 5 - Prediction of Coronary Artery Disease

This was my main body of work and all of the developmental steps were used to complete these final chapters. I designed the outcome-based study including defining optimal outcome measures for prediction. I generated and prepared the image datasets including contour generation and QA of images as well as reviewed and audited participant outcomes. I designed the initial model for prediction of CAD then collaborated with Angela Mumith who was able to use the python machine learning packages to improve machine learning model and make it more robust. Angela and myself worked together on both selecting the features and building the classifier to predict coronary artery disease.