Computational Methods for Augmenting the Echocardiographic Assessment of Low-Flow, Low-Gradient Aortic Valve Stenosis

---

Marcell Illyés

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF ENGINEERING

COLLEGE OF SCIENCE AND ENGINEERING

November 2025

Abstract

Aortic stenosis (AS) is the most common valvular disease requiring intervention in adults in developed countries. One of its most diagnostically challenging subtypes is low-flow low-gradient (LFLG) AS, characterised by discordant echocardiographic grading criteria. Current guidelines recommend dobutamine stress echocardiography (DSE) to resolve diagnostic uncertainty; however, DSE is often inconclusive and carries procedural risks. This research investigates whether resting echocardiographic recordings alone can distinguish true-severe from pseudo-severe LFLG AS.

First, statistical analyses and classical machine learning (ML) techniques were applied to standard and novel echocardiographic predictors to assess their ability to separate severity grades and flow states. The results show that the decision boundaries derived from concordant AS cases do not generalise well to LFLG AS. Although ML models achieved 79.3% accuracy in synthetic data generated from concordant AS, performance decreased when evaluated in the LFLG AS cohort in both rest and stress states.

Second, a novel strain imaging pipeline was developed to estimate three-dimensional myocardial deformation from 2D triplanar echocardiographic data. This enabled calculation of 3D strain parameters over time without full 3D acquisitions. The global and segmental longi tudinal strain values did not show significant differences between the 3D and triplanar methods, confirming the validity of the pipeline. A qualitative analysis also suggested that regional strain values were lower in patients with severe LFLG AS.

Finally, deep learning (DL) models that combine transformer-based feature extractors with multiple-instance learning were trained on a public dataset for AS severity classification. The test accuracy showed marginal improvements in the public data set with an average improvement of 2.16% over the predefined test sets, and generalisation to clinical data from patients with concordant AS achieved a balanced accuracy of 67.5%, compared to 62.5% from previous models. Additionally, the model performance in LFLG AS was assessed under resting or both resting and stress states.

Table of Contents

Abstract **i**

List of Tables **v**

List of Figures **vii**

Acknowledgment **xvi**

List of Abbreviations **xvii**

1 Introduction **1**

1.1 Motivation 1

1.2 Contribution 2

1.3 Outline 3

2 Clinical Background **4**

2.1 Form and Function 4

2.2 Echocardiography 6

2.3 Aortic Stenosis 10

2.3.1 Pathophisiology 10

2.3.2 Clinical Assessment 10

2.3.3 Low-Flow, Low-Gradient Phenotypes 13

3 Literature Review **17**

3.1 Limitations of Current Practice 17

3.1.1 Technical Limitations 17

3.1.2 Physiological Limitations 18

3.2 Emerging Research Directions in AS Diagnosis 21

3.2.1 Novel Clinical Predictors 21

3.2.2 Myocardial Function by Speckle-tracking Echocardiography 23

3.2.3 Machine Learning in Aortic Stenosis 27

3.3 Summary 31

4 Materials **32**

4.1 Low-Flow, Low-Gradient Aortic Stenosis Data Set 32

4.2 Concordant Aortic Stenosis Data Set 34

4.3 3D Echocardiography Data Set 35

4.4 Tufts Medical Echocardiogram Data Set 35

5 Analysis of Clinical Predictors Using Statistical and Machine Learning Methods 37
5.1 Introduction . . . 37
5.2 Methodology . . . 38
5.2.1 Univariate Analysis . . . 38
5.2.2 Multivariate Analysis . . . 41
5.2.3 Classification Algorithms . . . 43
5.2.4 Model Selection . . . 46
5.2.5 Applicability to Low-Flow, Low-Gradient Aortic Stenosis and Synthetic Data . . . 48
5.3 Results and Discussion . . . 50
5.3.1 Univariate Analysis . . . 50
5.3.2 Multivariate Analysis . . . 53
5.3.3 Classifier Performance . . . 56
5.4 Conclusion . . . 58
6 Triplane Speckle-Tracking Echocardiography for 3D Left Ventricular Strain Analysis 61
6.1 Introduction . . . 61
6.2 Related Work . . . 63
6.2.1 Reconstruction Techniques . . . 63
6.3 Solid Mechanics . . . 65
6.3.1 Position and Displacement . . . 65
6.3.2 Deformation Gradient and Strain Tensors . . . 65
6.3.3 Strains for Triangular Surface Elements . . . 67
6.3.4 Directional Strain Measures . . . 68
6.4 Image Processing . . . 69
6.4.1 Image Registration . . . 69
6.4.2 Transforms and Orientations . . . 72
6.5 Methodology . . . 73
6.5.1 Verification of Strain Measures . . . 73
6.5.2 Strains Measures from 3D Echocardiography . . . 74
6.5.3 Strain Measures from Synthetic Triplane Data . . . 78
6.5.4 Strain Measures from Clinical Triplane Data . . . 80
6.6 Results and Discussion . . . 83
6.6.1 Verification of Strain Measures . . . 83
6.6.2 3D and Triplane Measurements . . . 85
6.6.3 Regional Strain Assessment in Low-Flow, Low-Gradient Aortic Stenosis 91
6.7 Conclusion . . . 94

7 Deep Learning Approaches to Aortic Stenosis Prediction **95**
7.1 Introduction . . . 95
7.2 Deep Learning Foundations . . . 96
7.2.1 Learning Algorithm . . . 96
7.2.2 Maximum Likelihood Estimate . . . 97
7.2.3 Gradient-Based Optimisation . . . 99
7.2.4 Feedforward Neural Network . . . 100
7.2.5 The Convolutional Neural Network . . . 101
7.2.6 Additional Components . . . 103
7.3 Related Work . . . 105
7.3.1 Multiple-Instance Learning . . . 106
7.3.2 Transformers . . . 108
7.4 Methodology . . . 111
7.4.1 Training Strategy . . . 111
7.4.2 Generalisation Performance . . . 112
7.4.3 Model Explainability . . . 113
7.5 Results on TMED2 . . . 115
7.5.1 Training and Validation Performance . . . 115
7.5.2 Attention Weight Analysis . . . 121
7.5.3 Saliency Maps . . . 122
7.6 Generalisation Performance on Clinical Data . . . 124
7.6.1 Inference on Concordant Aortic Stenosis . . . 124
7.6.2 Inference on Low-Flow, Low-Gradient Aortic Stenosis . . . 126
7.6.3 Domain Shift Analysis . . . 127
7.7 Discussion and Conclusion . . . 129
8 Conclusion **130**
8.1 Key Findings . . . 130
8.2 Suggestions and Future Directions . . . 132
8.2.1 Methodology Improvements . . . 132
8.2.2 Broader Research Opportunities . . . 133
8.3 Final Remarks . . . 135
A Appendix A **136**
B Appendix B **145**
C Appendix C **146**

List of Tables

2.1 Reference values for grading aortic stenosis. Although classification employs discrete cutoff values, it is essential to consider the measures of disease severity as a continuous spectrum. 14

4.1 Distribution of severity labels and average number of labeled images per study across the training, validation, and test splits for each fold in the Tufts Medical Echocardiogram Dataset. 36

5.1 Overview of the hypothesis tests performed in this chapter. Each test is defined by its objective, the comparison groups involved, and the corresponding sample sizes used for statistical analysis. AS60 refers to the concordant aortic stenosis (AS) data set, while DSE12 denotes the low-flow, low-gradient (LFLG) AS cohort, who underwent dobutamine stress echocardiography with both resting and stress recordings. 40

5.2 Statistical tests on the clinical predictors using the concordant aortic stenosis (AS) cohort and the low-flow, low-gradient AS data set. The table columns display the p-values for each corresponding test outlined in Table 5.1, with statistically significant values highlighted. 52

5.3 Statistical tests on the principal component representation of the clinical data using the concordant aortic stenosis (AS) cohort and the low-flow, low-gradient AS data set. The table columns display the p-values for each corresponding test outlined in Table 5.1, with statistically significant values highlighted. 55

6.1 The comparison of average regional displacements at peak systole between the 3D and triplane methods, using all patients from the myocardial wall motion analysis data set (n = 9). For each of the 17 segments, the table reports the mean and standard deviation for both 3D and triplane methods, the mean difference (3D-Triplane), the root mean square error (RMSE), and the p-value from the Wilcoxon signed-rank test assessing the statistical significance between the two methods. Units for distance metrics are in millimetres. . . 88

6.2 The comparison of average regional longitudinal strain at peak systole between the 3D and triplane methods, using all patients from the myocardial wall motion analysis data set (n = 9). For each of the 17 segments, the table reports the mean and standard deviation for both 3D and triplane methods, the mean difference (3D–Triplane), the root mean square error (RMSE), and the p-value from the Wilcoxon signed-rank test assessing the statistical significance between the two methods. Units for strain metrics are in percentages. 89

6.3 The comparison of average regional circumferentail strain at peak systole between the 3D and triplane methods, using all patients from the myocardial wall motion analysis data set (n = 9). For each of the 17 segments, the table reports the mean and standard deviation for both 3D and triplane methods, the mean difference (3D-Triplane), the root mean square error (RMSE), and the p-value from the Wilcoxon signed-rank test assessing the statistical significance between the two methods. Units for strain metrics are in percentages. 90

7.1 Per-patient predictions under rest and stress for all models. Hard voting columns report predicted class labels only. Soft voting columns report the class-probability distributions across the three classes, with the predicted class (maximum probability) shown in bold. Bottom rows show mean log-loss (MLL) and number of correct predictions per state. 126

List of Figures

2.1 The anatomical structure of the heart, showing the four main chambers as well as the semilunar and atrioventicular valves. The arrows indicate the direction of the blood flow, while the structures colored in blue and red represent the pulmonary and systemic circulatory systems, respectively. 4

2.2 Wiggers diagram showing the varying pressure and volume curves at each cardiac phase, along with the corresponding electrocardiogram. The different phases are determined by the opening and closing times of the valves. . . . 6

2.3 The acoustic windows during transthoracic echocardiography as illustrated on the left and the cardiac imaging planes of the heart as shown on the right. The three major axes are indicated for the superior-inferior, medial-lateral and anterior-posterior directions. 7

2.4 Examples of common echocardiographic imaging modes. Morphological assessment is performed using B-mode and M-mode imaging as shown in (a) and (b), while Doppler techniques provide quantitative evaluation of blood flow and tissue velocities as illustrated in (c) and (d). 9

2.5 An illustration of hemodynamic measurements obtained by continuous-wave (CW) Doppler ultrasound (US) at the left ventricular outflow tract (LVOT) and by pulsed-wave (PW) Doppler US at the level of the aortic valve (AV). The arrows indicate the direction of blood flow, while the areas shaded in red represent the velocity-time integral (VTI) measurements used by the continuity equation to calculate the aortic valve area. Note the difference in scales for peak velocities between the two measurements. 13

2.6 The clinical grading algorithm of aortic stenosis (AS). In the high gradient track all cases are considered severe, while the low gradient track is further refined. The flow state is split based on the indexed stroke volume (SV_i) in cases of discordance between the aortic valve area (AVA) and mean transvalvular gradient to differentiate between low flow and normal flow conditions. In low flow states, the ejection fraction (EF) is used to measure global left ventricular (LV) function, categorising patients with paradoxical and classic low-flow, low-gradient (LFLG) AS. Calcium scoring by computed tomography (CT) is recommended for normal flow and paradoxical LFLG AS patients, while dobutamine stress echocardiography (DSE) is recommended in classic LFLG AS cases. To differentiate between pseudo-severe and true severe AS, the AVA and mean gradients are examined during the increased flow conditions and the severity is determined if the parameters adhere to the grading criteria presented in Table 2.1. Inconclusive scans may arise as result of lack of flow reserve or persisting AVA-gradient discordance, where experimental predictors may be used to guide treatment. 16

3.1 The different directions used to quantify strains in the assessment of left ventricular (LV) function. The longitudinal direction follows the myocardium from the base to the apex, the circumferential direction is defined by the perimeter of the LV as seen on the short axis view, while the radial direction spans the thickness of the myocardial wall. 24

3.2 Myocardial segments at different left ventricular levels as depicted from an apical view. The long-axis, defined through the apex and mid-base points, divides the myocardium into two parts, which are further split based on the length of the myocardium on each side. 26

3.3 The distinction between explicit programming of rule-based systems, Machine Learning (ML) and Deep Learning (DL) algorithms. The shaded areas represent the functional components learnt from data, indicating that DL extends ML by an added layer of abstraction. 29

4.1 Clinical examples of routine echocardiographic measurements. The top row depicts hemodynamic parameters obtained at the level of the aortic valve (AV) and at the level of the left ventricular outflow tract (LVOT). The middle row shows the anatomical measurements of the left ventricle (LV) and LVOT diameter, while the bottom row illustrates the volumetric measurements of the LV and the assessment of LV contractile function by global longitudinal strain (GLS). 33

4.2 Example recordings of 3D ultrasound volumes from the myocardial wall motion analysis data set as displayed at the end-diastolic [left] and end-systolic frames within the cardiac cycle. The arrows represent the three major axes and standard orientation labels are provided for right (R), left (L), superior (S), inferior (I), and anterior (A) directions. 35

5.1 Flow diagram for choosing the appropriate statistical test. The left column presents the parametric tests, whereas the right column lists the non-parametric tests for continuous dependent variables and categorical independent variables. 40

5.2 The decision boundary of an SVM as depicted in the linearly separable case (left) and as shown in the linearly non-separable case (right). The orange and blue points belong to the positive and negative classes respectively. The parameters of the model define a hyperplane (a line in two dimensions),H_0, such that $H_0 = \{\mathbf{x} \in \mathbb{R}^f : \mathbf{x}\mathbf{w}^T + b = 0\}$. 45

5.3 The distribution of clinical predictors by aortic stenosis (AS) severity in the concordant AS cohort, as depicted via a box and whisker plot. The corresponding features from the low-flow, low-gradient AS cohort is shown on top. Square markers represent the baseline, while the circular markers the stress states. The clinical cutoff values are marked by a dashed line for maximum velocity through the aortic valve (AV Vmax), the mean transvalvular gradient (AV mean PG), aortic valve area (AVA), and indexed AVA (AVA_i). 51

5.4 The distribution of the first eight principal components by aortic stenosis (AS) severity in the concordant AS cohort, as illustrated on a box and whisker plot. The corresponding principal components from the low-flow, low-gradient AS cohort is overlaid on top. Square markers on the left side of each box show the baseline, while the circular markers on the right the stress states. The percentage values represent the explained variance ratio by the given component. 54

5.5 The absolute loadings of the clinical predictors for the first three principle components representing the contribution of each feature to the given component. 55

5.6 Decision boundaries of different classification algorithms as shown on the subset of the principal component (PC) space where the separation of severity grades is most prominent. Patients with concordant aortic stenosis (AS) and low-flow, low-gradient (LFLG) AS are shown in the left column, while the synthetically generated patients are displayed in the right column. AS60 represents the concordant AS data set, while DSE12 refers to the LFLG AS cohort. Arrows indicate the intrapatient stress response from the baseline state. Accuracies are displayed for the AS60 and synthetic data sets using two and eight PCs. 57

5.7 Decision boundaries of different classification algorithms as plotted on the feature space covered by the two discriminant vectors from linear discriminant analysis. Patients with concordant aortic stenosis (AS) and low-flow, low-gradient (LFLG) AS are shown in the left column, while the synthetically generated patients are displayed in the right column. AS60 represents the concordant AS data set, while DSE12 refers to the LFLG AS cohort. Arrows indicate the intrapatient stress response from the baseline state. Accuracies are displayed for the AS60 and synthetic data sets using one and two discriminant vectors. 59

6.1 Triplane acquisition mode capturing three planes at 60° angles, resulting in three separate 2D + time B-mode recordings of the left ventricle with a fixed geometric relationship between the planes over time. 62

6.2 Deformation of an object from the reference configuration (Ω_0) to the current configuration (Ω_t). An infinitesimal line element $d\mathbf{X}$ connecting points P and Q in Ω_0 is mapped to $d\mathbf{x}$ between points P' and Q' in Ω_t. The displacement vectors are denoted by $\mathbf{u}(\mathbf{X}, t)$ at P and $\mathbf{u}(\mathbf{X} + d\mathbf{X}, t)$ at Q. 66

6.3 Representation of the deformation gradient $\mathbf{F}$ through a sequence of transformations acting on a triangular surface element. The element is first mapped from the reference configuration to the isoparametric configuration via $\mathbf{F}_X^{-1}$, and then from the isoparametric configuration to the deformed configuration via $\mathbf{F}_x$. In the reference configuration, the edges $\mathbf{A}_1$ and $\mathbf{A}_2$ are mapped to $\mathbf{a}_1$ and $\mathbf{a}_2$ in the deformed configuration. 67

6.4 Schematic representation of the deformable image registration framework. The moving image I_t is interpolated using the current spatial transformation s_i and compared to the fixed image I_0 via a similarity measure $\mathcal{D}(I_0, I_t \circ s_i)$. The optimiser computes an update to the transformation based on the gradient of the objective. Smoothness of the deformation field is enforced implicitly through Gaussian filtering of both the similarity gradient and the updated transformation field or via an explicit regularisation term in the objective function. 70

6.5 The displacement field from a deformable image registration tool as displayed on the three major anatomical planes for components x, y and z. 71

6.6 The 17 segment model of the left ventricle as visualised in 3D (left), and as depicted on a bullseye plot with the region labels listed below (right). 75

6.7 Short axis interpolation using Catmull-Rom splines through the selected control points, generating N_θ number of equally spaced points [left] and the basis functions used to construct the spline [right]. The annotations highlight how the coordinates of the interpolated point, $\mathbf{p}(\tau)$ on the segment between $\mathbf{c}_0$ and $\mathbf{c}_1$ are calculated using the local parametric coordinate τ and the basis functions. 77

6.8 Long-axis (LAX) interpolation using Catmull-Rom splines through the corresponding control points on each short-axis (SAX) curve. The left figure shows the equally spaced interpolated points along the LAX direction (N_z) generated between SAX contours (apical to basal), with the apex marked in red. The right figure displays the resulting 3D surface mesh created from these interpolated points, visualised in ParaView. 78

6.9 The developed pipeline showing the steps needed to acquire a triangular surface mesh series following the myocardial midline of the left ventricle from a triplane DICOM cine series. $I_{\text{seq}}[i, j]$ uses the DICOM image data and tags to split the recording into three separate cine series. $\mathbf{A}_{\text{DICOM}}(\alpha)$ denotes the angle-dependant spatial transformations required to assemble the different mesh series in the same physical space. 82

6.10 The triangular surface mesh of cylindrical geometry as shown in the reference configuration [left] and as depicted in the deformed configuration [right]. The pointwise correspondance between the two meshes allowed the verification of the strain calculations. 83

6.11 Visualisation of the local strain directions in the deformed configuration, showing the longitudinal , circumferential, and radial directions across the mesh surface in blue, red and green, respectively. 84

6.12 Stretch values in the circumferential [left], longitudinal [centre], and radial directions [right] of the reference geometry as depicted on a polar plot. Each plot shows the spatial variations in directional stretch across the mesh surface, where the angular position represents the circumferential location, while the radial distances correspond to the height measured from the top of the cylinder. 84

6.13 The comparison of global mesh distance metrics [left] and left ventricular (LV) volumes [right] over the normalised cardiac cycle between 3D and triplane methods. The left figure displays the symmetric Hausdorff distance, th pointwise root mean square error (RMSE), and the symmetric mean boundary distance (SMBD), indicating the temporal variation in mesh distance. The right figure displays mean LV volume curves from 3D and triplane reconstructions. The shaded regions represent the standard deviations across all patients. 85

6.14 The comparison of global longitudinal [left] and circumferential [right] strain values across the normalised cardiac cycle for all patients, derived using 3D and triplane methods. Solid lines represent the mean global strain across all patients, while shaded areas indicate one standard deviation. 86

6.15 Boxplot comparison of average global displacement [left], longitudinal strain (middle), and circumferential strain [right] at peak systole between the 3D and triplane methodologies across all patients in the myocardial wall motion analysis data set. Displacement is reported in millimetres, while strain values are expressed as percentages. 87

6.16 Regional displacement comparison between 3D and triplane methodologies at peak systole across all patients in the myocardial wall motion analysis data set, displayed using a 17-segment bull's eye plot. The left figure shows the mean difference, the centre figure displays the root mean square error (RMSE), and the right figure illustrates the corresponding statistical significance with the associated segmental p-values. The inner rings represent apical regions, and the outer rings represent basal segments. 88

6.17 Regional longitudinal strain comparison between 3D and triplane methodologies at peak systole across all patients in the myocardial wall motion analysis data set, displayed using a 17-segment bull's eye plot. The left figure shows the mean difference, the centre figure displays the root mean square error (RMSE), and the right figure illustrates the corresponding statistical significance with the associated segmental p-values. 89

6.18 Regional circumferential strain comparison between 3D and triplane methodologies at peak systole across all patients in the myocardial wall motion analysis data set, displayed using a 17-segment bull's eye plot. The left figure shows the mean difference, the centre figure displays the root mean square error (RMSE), and the right figure illustrates the corresponding statistical significance with the associated segmental p-values. 90

6.19 Segmental analysis of left ventricular deformation parameters at peak systole in patients with moderate (blue circles) and severe (red squares) aortic stenosis (AS), across all 17 standard left ventricular segments. The top figure shows myocardial displacement [mm], the middle figure shows longitudinal strain [%], and the bottom figure shows circumferential strain [%]. The dashed lines represent group-wise means for each parameter. Error bars denote one standard deviation. 92

6.20 Segmental assessment of left ventricular (LV) contractile reserve in three patients with low-flow, low-gradient aortic stenosis during dobutamine stress echocardiography, compared to baseline (resting) measurements. Differences in peak systolic regional displacement [top], longitudinal strain [middle], and circumferential strain [bottom] are plotted per LV segment. Negative values for longitudinal and circumferential strain indicate increased contractility in response to stress. Dashed lines represent the mean change across all segments for each patient. 93

7.1 A fully connected feedforward neural network with m input units, 5 hidden units, and 3 output units. The input layer $(x_1, \ldots, x_m)$ is connected to the hidden layer $(h_1, \ldots, h_5)$ via the weight matrix $\mathbf{W}^{[1]}$ and bias vector $\mathbf{b}^{[1]}$, followed by the output layer (z_1, z_2, z_3), computed using the weight matrix $\mathbf{W}^{[2]}$ and bias vector $\mathbf{b}^{[2]}$. Information flows from left to right. 100

7.2 The convolution operation as performed on a three dimensional input using a set of filters. The set of filters are shifted across the input from the top-left to the bottom-right corner in a sliding-window manner, and the output is calculated by summing the element-wise products of the corresponding regions. 102

7.3 Common activation functions used in deep learning. The sigmoid and hyperbolic tangent (tanh) functions introduce nonlinearity by squashing input values into bounded intervals $[0, 1]$ and $[-1, 1]$, respectively. The rectified linear unit (ReLU) allows unbounded positive outputs while setting all negative inputs to zero. 103

7.4 Residual connections propagating the input to a layer to its output unaltered, stabilizing gradient flow during backpropagation. 105

7.5 Architecture of the supervised attention multiple-instance learning (SAMIL) model. The network comprises three main components: a feature extractor $f(\cdot)$, which converts each image from a transthoracic echocardiogram study into a feature representation; a pooling layer $\sigma(\cdot)$, which aggregates these instance-level features into a bag-level representation through learnt attention weights while preserving permutation invariance; and an output layer $g(\cdot)$, which maps the bag-level representation to the predicted aortic stenosis (AS) severity class. A supervised attention (SA) module further guides the attention weights toward clinically relevant views (parasternal long-axis and parasternal short-axis), through an additional view-relevance loss L_{SA}. . . . 108

7.6 The Fully Convolutional Transformer architecture for medical image segmentation. 110

7.7 Visual overview of the domain pairs used for shift analysis. Six domains are considered: the training, validation and test splits from the Tufts Medical Echocardiogram Dataset (TMED) denoted as TMED-Train, TMED-Val, TMED-Test, the concordant aortic stenosis (AS) cohort indicated by AS60, and recordings from the low-flow low-gradient AS patients at rest and at the combined rest and stress state denoted as DSE12-rest, and DSE12-all. Solid lines represent comparisons between domains from the same hospital (e.g., training to validation), while dashed lines indicate cross-hospital comparisons. 115

7.8 Training and validation performance of the supervised attention multiple instance learning model across the pre-defined cross-validation splits of the Tufts Medical Echocardiogram Dataset. Cross-entropy loss, view regularisation loss, and balanced accuracy are shown from top to bottom. The left column presents the results for the training set, while the left column shows the results for the validation set. 117

7.9 Training and validation performance of the fully convolutional transformer feature extractor multiple instance learning model with varying view regularisation strengths on split 3 of the Tufts Medical Echocardiogram Dataset. Cross-entropy loss, view regularisation loss, and balanced accuracy are shown from top to bottom. The left column presents the results for the training set, while the right column shows the results for the validation set. Smoothing was applied to the raw output values (seen in matching transparent colors) to help visualisation. 118

7.10 Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 3 of the Tufts Medical Echocardiogram Dataset. The columns contain the confusion matrices for the training, validation and test sets from left to right for the models selected at the best validation accuracy. Precision (P) and recall (R) values are provided for each class in the given set and balanced accuracy (B.acc.) is calculated for each set to assess overall generalisation performance. 119

7.11 Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 3 of the Tufts Medical Echocardiogram Dataset. The columns contain the ROC curves for multiple binary classification tasks for the training validation and test sets for models selected at the best validation accuracy. 120

7.12 Average view relevance scores for the top 10 attended images across the test set in split 3 of the Tufts Medical Echocardiogram Dataset. Values close to 1 indicate that clinically relevant views (parasternal short axis and parasternal long axis) were used for the AS predictions. 121

7.13 Gradient-weighted class activation maps (Grad-CAM) for the Fully Convolutional Transformer (FCT) feature extractor and the original supervised attention multiple instance learning (SAMIL) architectures. Backpropagation was performed for the the correctly predicted significant aortic stenosis (AS) severity class on the top attended input image. 123

7.14 Zero-shot inference on the concordant aortic stenosis cohort using the models from all three data splits of the Tufts Medical Echocardiogram Dataset. Confusion matrices are shown for models with the best balanced accuracy (b.acc.) on the validation set from split 1, split 2 and split 3, as seen from left to right. The top row presents the results for the fully convolutional transformer (FCT) feature extractor architecture, while the bottom row shows the results for the supervised attention multiple instance learning (SAMIL) architecture. Class-specific precision (P) and recall (R) metrics are overlayed on the corresponding diagonal entries. 124

7.15 Model ensembling using majority voting [top] and probability averaging [bottom]. Stochastic inference via MCD was also conducted for Fully Convolutional Transformer (FCT) using five random seeds. Confusion matrices are shown for the FCT, FCT with MCD and SAMIL models from left to right. 125

7.16 The class-wise distribution of the low rank approximations of bag-level representations in different data sets using the first two principal components as depicted for the FCT architecture in split 3. 127

7.17 Domain shift analysis showing the classwise Fréchet Distance (FD) across different domains. The horizontal line indicates the value of the centroid shift, while the shaded areas represent the average FD distance across the classes. 128

A.1 Clinical predictors measured in the DSE12 data set 137

A.2 Clinical predictors measured in the DSE12 data set continued 138

A.3 Clinical predictors measured in the AS60 data set 139

A.4 Clinical predictors measured in the AS60 data set continued. 140

A.5 Clinical predictors measured in the AS60 data set continued. 141

A.6 Clinical predictors measured in the AS60 data set continued. 142

A.7 Clinical predictors measured in the AS60 data set continued. 143

A.8 Clinical predictors measured in the AS60 data set continued. 144

C.1 Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 1 of the Tufts Medical Echocardiogram Dataset. The columns contain the confusion matrices for the training validation and test sets for models selected at the best validation accuracy. 147

C.2 Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 1 of the Tufts Medical Echocardiogram Dataset. The columns contain the ROC curves for multiple binary classification tasks for the training validation and test sets for models selected at the best validation accuracy. 147

C.3 Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 2 of the Tufts Medical Echocardiogram Dataset. The columns contain the confusion matrices for the training validation and test sets for models selected at the best validation accuracy. 148

C.4 Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 2 of the Tufts Medical Echocardiogram Dataset. The columns contain the ROC curves for multiple binary classification tasks for the training validation and test sets for models selected at the best validation accuracy. 148

C.5 Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 3 of the Tufts Medical Echocardiogram Dataset. The columns contain the confusion matrices for the training validation and test sets for models selected at the best validation accuracy. 149

C.6 Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 3 of the Tufts Medical Echocardiogram Dataset. The columns contain the ROC curves for multiple binary classification tasks for the training validation and test sets for models selected at the best validation accuracy. 149

C.7 Top 10 ranked images by attention weights for the FCT and SAMIL architectures in the test set of split 3 in the Tufts Medical Echocardiogram Dataset. Both models succeed in attending to relevant views (PLAX,PSAX) as shown by average view relevance scores close to 1. 150

Acknowledgment

This project would not have been possible without the consistently precise guidance and commitment of my supervisors, Ankush Aggarwal and Andrew McBride. Their expertise, patience, and encouragement have shaped not only this work, but also my growth as a researcher.

Special thanks is due to Chaitanya Kahul for providing encouragement and mentorship throughout my exploration into the realm of deep learning, as well as Athanasios Tragakis for his friendly and patient collaborative efforts. I have learnt a tremendous amount through our work together.

I would like to express my gratitude to Richard Good and Piotr Sonecki for their invaluable insights into clinical practice. Their work, guidance, and support have been a source of both knowledge and motivation throughout my research.

Thanks must be extended to all my colleagues and friends in the Glasgow Computational Engineering Centre, whose kindness has made a real difference throughout this journey. In particular, I am grateful to Claire and Jude, whose positive spirit has brightened even the most challenging days.

I am immeasurably grateful to my partner, Aidan, for always being a constant source of inspiration and joy. Her love and encouragement have made even the most difficult moments bearable, and I could not have completed this journey without her.

Lastly, I would like to thank my family and especially my mom for their unwavering love and support throughout my PhD. Although striving for greatness is important, it is equally vital to appreciate the gift of life.

List of Abbreviations

Medical Terms

A2C	Apical two-chamber
A4C	Apical four-chamber
AHA	American Heart Association
AS	Aortic stenosis
AT	Acceleration time
AV	Aortic valve
AVA	AV area
AVA_i	indexed AVA
BLS	Basal longitudinal strain
BPM	Beats per minute
BSA	Body surface area
CMR	Cardiac magnetic resonance
CO	Cardiac output
CSA	Cross-sectional area
CT	Computed tomography
CW	Continuous-wave
DSE	Dobutamine stress echochardiography
ECG	Electrocardiogram
EDV	End-diastolic volume
EF	Ejection fraction
EOA	Effective orifice area
ESV	End-systolic volume
ET	Ejection/envelope time
GCS	Global circumferential strain
GLS	Global longitudinal strain
HF	Heart failure
HR	Heart rate
IOV	Interobserver variability
IVS	Interventricular septum
LA	Left atrium
LFLG	Low-flow, low-gradient
LV	Left ventricle
LVID	LV internal diameter

LVOT	LV outflow tract
LVPW	LV posterior wall
MRI	Magnetic resonance imaging
MWMA	Myocardial wall motion analysis
MV	Mitral valve
PG	Pressure gradient
PLAX	Parasternal long-axis
PRF	Pulse repetition frequency
PSAX	Parasternal short-axis
PW	Pulsed-wave
RA	Right atrium
RV	Right ventricle
SAVR	Surgical aortic valve replacement
STE	Speckle-tracking echocardiography
SV	Stroke volume
SV_i	indexed SV
TAVR	Transcatheter aortic valve replacement
TEE	Transesophageal echocardiography
TTE	Transthoracic echocardiography
US	Ultrasound
VTI	Velocity-time integral

Other

AI	Artificial intelligence
AUC	Area under curve
BACC	Balanced accuracy
BN	Batch normalisation
CE	Cross-entropy
CNN	Convolutional neural network
DICOM	Digital Imaging and Communications in Medicine
DL	Deep learning
FCT	Fully convolution transformer
FN	False negative
FP	False positive
FPR	False positive rate
Grad-CAM	Gradient-weighted class activation map
KL	Kullback–Leibler
KNN	K-nearest neighbour
MHSA	Multi-head self-attention
MIL	Multiple instance learning
ML	Machine learning

MLE	Maximum likelihood estimate
MLP	Multilayer perceptron
NIfTI	Neuroimaging Informatics Technology Initiative
NLL	Negative log-likelihood
NLP	Natural language processing
ReLU	Rectified linear unit
RNN	Recurrent neural network
ROC	Receiver operating characteristics
SA	Supervised attention
SVM	Support vector machine
TN	True negative
TMED	Tufts Medical Echocardiogram Dataset
TP	True positive
TPR	True positive rate
ViT	Vision transformer
VTK	Visualization toolkit

1. Introduction

Aortic stenosis (AS) is the most common valvular disease among adults in developed countries that requires intervention (Vahanian et al., 2022). It is characterised by progressive narrowing of the effective aortic valve area due to restricted leaflet mobility, often caused by valve calcification or congenital malformations such as bicuspid aortic valves (Rajamannan et al., 2011). The resulting obstruction of systolic blood flow from the left ventricle (LV) to the aorta causes increased resistance that the LV must overcome to eject blood, which often leads to adaptive remodelling of the LV myocardium. Over time, these adaptations can appear as a variety of structural and functional abnormalities, including altered hemodynamic conditions and impaired myocardial contractility, which contribute to the heterogeneous presentation of AS (Lindman et al., 2016).

1.1 Motivation

One of the most diagnostically challenging subtypes is the low-flow, low-gradient (LFLG) phenotype (Baumgartner et al., 2017). Patients with classical LFLG AS exhibit deceptively mild hemodynamic characteristics with reduced LV function, despite showing signs of advanced leaflet calcification and structural degeneration. The apparent discordance between hemodynamic and anatomical severity complicates the routine echocardiographic evaluation and limits the utility of conventional grading algorithms (Clavel et al., 2016). As a result, patients are frequently referred for additional diagnostic procedures such as dobutamine stress echocardiography (DSE) or computed tomography (CT) calcium scoring to resolve this uncertainty. Although these tests can provide incremental information, they are resource intensive and often yield inconclusive results, prolonging the diagnostic process without clear benefit in all cases (Baumgartner et al., 2017).

This thesis investigates whether standard echocardiographic recordings alone are sufficient to accurately determine the severity of LFLG AS, with the goal of reducing the dependence on additional diagnostic procedures such as DSE. It is hypothesised that novel computational approaches applied to routine echocardiographic data can improve classification performance and enhance clinical interpretability in cases of classical LFLG AS.

1.2 Contribution

To explore this hypothesis, three distinct methodologies were employed.

- **Statistical and classical machine learning-based analysis of clinical predictors**. A comprehensive statistical analysis framework was developed to test various hypotheses regarding the separation of different hemodynamic states and severity grades using standard and novel clinical markers. Next, dimensionality reduction and classical machine learning algorithms were employed to evaluate whether the diagnostic criteria established for concordant AS can be adapted to the LFLG phenotype. This approach also examined the stress response patterns captured during DSE within a multivariate feature space, offering novel insights into the physiological differences between severe and non-severe LFLG AS. This constitutes the first known study to investigate the response to DSE using reduced-dimensional representations of echocardiographic parameters alone in the LFLG AS cohort.
- **Triplane speckle-tracking echocardiography for 3D left ventricular strain analysis**. A novel technique was designed to estimate regional LV myocardial strain by reconstructing three-dimensional deformation from triplane echocardiographic acquisitions. This approach aimed to capture spatially resolved myocardial dysfunction in patients with LFLG AS, offering a more detailed assessment than global functional measures such as ejection fraction or global longitudinal strain.
- **A novel deep learning based severity classification algorithm**. Finally, deep learning models were trained to directly predict AS severity from a set of echocardiographic images. The generalisability of these models was evaluated in concordant and LFLG AS populations, evaluating their ability to identify severe disease from routine imaging data alone.

Together, these methods investigate whether routine echocardiographic data is sufficient to improve the accuracy of AS severity assessment, with a focus on the LFLG subtype. By applying statistical analysis, machine learning, biomechanical modelling, and deep learning, this thesis evaluates the extent to which standard imaging can replace or reduce the need for additional tests such as DSE.

1.3 Outline

This thesis is structured into 7 chapters as follows.

- **Chapter 2** introduces the cardiovascular system and its main components. It describes the principles of transthoracic echocardiography (TTE), including standard views and imaging modes used to evaluate anatomical and hemodynamic parameters. The chapter also outlines the clinical workflow for assessing AS severity, with particular emphasis on the classification of LFLG phenotypes and the role of stress echocardiography and CT calcium scoring in discordant cases, constituting the clinical foundation for the research explored in this thesis.
- **Chapter 3** provides a comprehensive review of the literature on current limitations in the echocardiographic evaluation of AS, with a focus on the LFLG phenotype. It highlights key technical and physiological constraints of standard parameters and diagnostic shortcomings of DSE. Emerging research directions are detailed, including novel hemodynamic markers, myocardial strain imaging, and machine learning-based approaches.
- **Chapter 4** provides a summary of the various medical imaging data sets used in this thesis, serving as a reference point for the techniques introduced in Chapters 5,6 and 7.
- **Chapter 5** first presents the methodology for the statistical analysis framework and ML techniques. This includes various hypothesis tests to investigate the separation of AS grades and hemodynamic states, as well as dimensionality reduction and ML classification models using echocardiographic parameters to predict AS. The objective was to assess whether models trained on concordant AS generalise to LFLG AS under varying flow conditions and remain consistent with clinical guidelines.
- **Chapter 6** introduces a novel methodology to obtain 3D regional LV strain parameters using triplane echocardiography, with the aim of improving myocardial assessment in patients with LFLG AS. The methodology is applied to clinical data from patients undergoing DSE to explore baseline and stress-induced differences in regional myocardial deformation between different grades of AS.
- **Chapter 7** presents a novel DL approach to predict the severity of AS from a set of echocardiographic images. A baseline multiple instance learning (MIL) model and an enhanced variant using a fully convolutional transformer (FCT) were trained and evaluated on public and clinical data sets. Their performance was assessed using standard metrics and interpretability techniques.
- **Chapter 8** outlines the key contributions of the thesis and acknowledges the limitations of the proposed approaches. It also discusses their potential clinical implications and highlights possible directions for future research, concluding with a summary of the overall findings.

2. Clinical Background

This chapter introduces the cardiovascular system and the core physiological concepts in Section 2.1, followed by an overview of TTE, including standard imaging views and modes used to assess cardiac anatomy and function in Section 2.2. It also outlines the clinical workflow for evaluating AS, details key echocardiographic predictors, and discusses the classification of LFLG phenotypes, as well as the role of DSE and CT calcium scoring in discordant cases in Section 2.3.

2.1 Form and Function

The cardiovascular system comprises the heart and a network of interconnected blood vessels, which include arteries, veins, and capillaries, as well as approximately five litres of blood. Its primary function is to maintain a stable state within the body (i.e. homeostasis) by supplying it with oxygen and nutrients while removing carbon dioxide and metabolic waste.

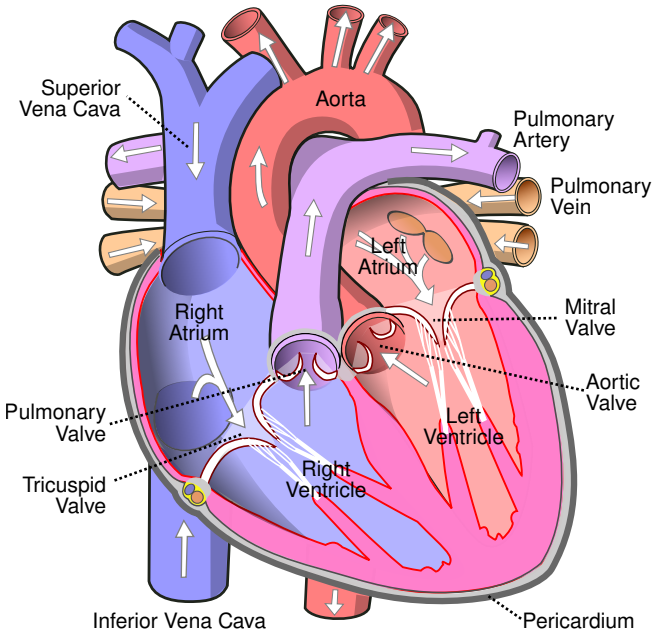


Figure 2.1: The anatomical structure of the heart, showing the four main chambers as well as the semilunar (aortic and pulmonary) and atrioventicular valves (mitral and tricuspid). The arrows indicate the direction of the blood flow, while the structures colored in blue and red represent the pulmonary and systemic circulatory systems, respectively. (Figure adapted from Wikipedia (2024)).

At the core of the cardiovascular system lies the heart, a complex muscular organ responsible for pumping blood through the vessels of the circulatory system. As illustrated in Figure 2.1,

the heart consists of four distinct chambers: the right atrium (RA), right ventricle (RV), left atrium (LA), and left ventricle (LV). These chambers regulate blood flow through the heart, which occurs via two main circulatory circuits: the pulmonary and systemic circulation. The pulmonary circulation, responsible for gas exchange, involves the path of blood flow from the right side of the heart to the lungs. It begins as deoxygenated blood returns from the body to the right atrium via the superior and inferior vena cava. The blood then flows through the tricuspid valve into the right ventricle and is subsequently pumped through the pulmonary valve into the pulmonary artery, which carries it to the lungs for gas exchange. Oxygenated blood then returns to the left atrium via the pulmonary veins, passes through the mitral valve to the LV, and is finally ejected into the aorta through the aortic valve (AV) to supply the systemic circulation.

However, blood flow does not remain constant throughout the circulatory system; it is governed by the cyclic contraction of the heart chambers initiated by a sequence of electrical impulses. This contraction leads to increased pressure within the heart chambers, generating a pressure gradient between the neighbouring chambers and/or vessels, ultimately causing the valves to open, allowing blood to exit the chamber. In order to gain a deeper understanding of the heart's physiology, the events occurring in the cardiac cycle are typically separated into various phases. These phases are determined by the timing of electrical activity (assessed using electrocardiography (ECG)), along with the opening and closing stages of the heart valves. A Wiggers diagram, as shown in Figure 2.2, is commonly used to represent these concurrent events, illustrating changes in relevant cardiac parameters over time.

The R wave of the ECG signal initiates an isovolumic contraction, defined by an increase in ventricular pressure without a change in ventricular volume. When the ventricular pressure exceeds the pressure in the ventricular outflow tract, the semilunar valves open and the ejection phase begins. Here, ventricular volume decreases until the semilunar valves close, by which systole is concluded. The diastolic phase begins with an isovolumic relaxation, characterised by a decrease in ventricular pressure, while the ventricular volume remains constant. This is followed by a rapid inflow of blood after the atrioventricular valves open, and consequently the ventricular volume increases substantially by passive filling. This is followed by the diastasis phase, marked by a slower ventricular filling that lasts until the onset of atrial contraction, triggered by the P-wave in the ECG signal. Atrial systole further increases ventricular volume and causes a slight increase in atrial and ventricular pressures. Diastole is concluded by ventricular depolarisation, indicated by the R wave, along with the closure of the atrioventricular valves, resulting in all parameters reverting to their initial states.

Understanding the cardiac cycle facilitates the identification of pathologies such as heart failure and stenotic valves, which are characterised by modified hemodynamic states. Ultrasound imaging continues to be the primary method for evaluating cardiac function, allowing

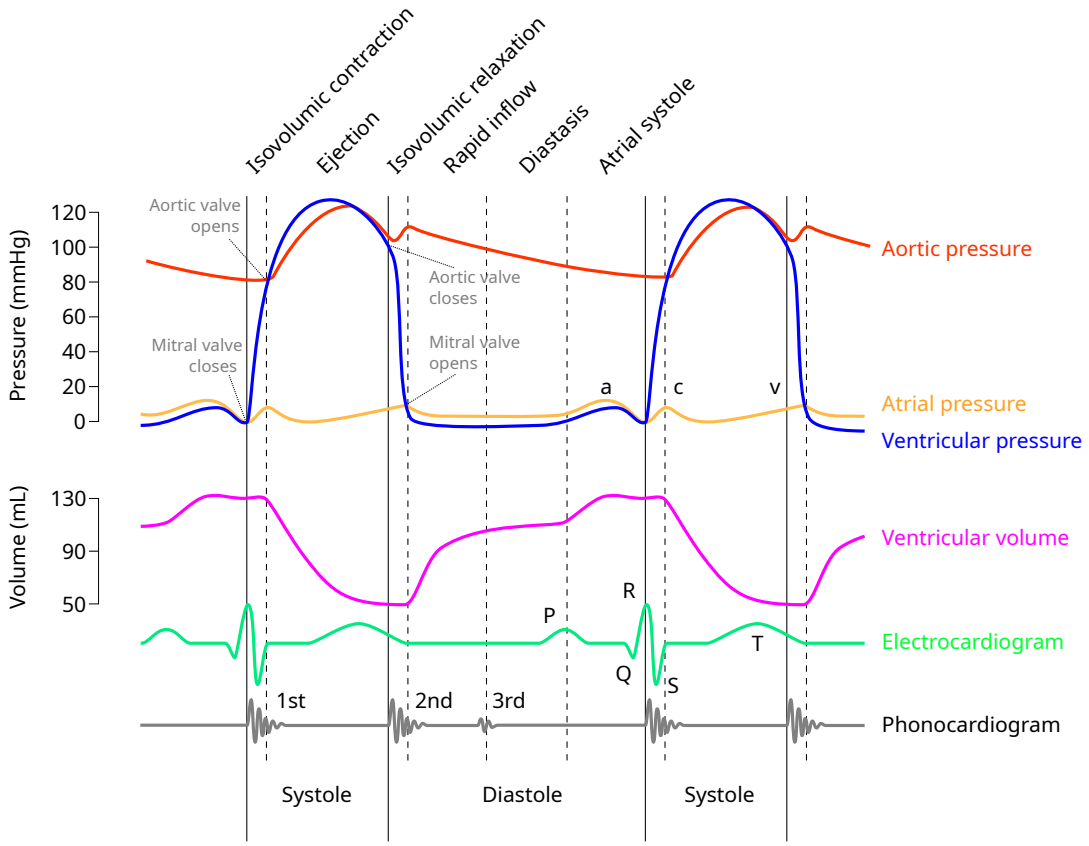


Figure 2.2: Wiggers diagram showing the varying pressure and volume curves at each cardiac phase, along with the corresponding electrocardiogram. The different phases are determined by the opening and closing times of the valves. (Figure adapted from (Wikipedia, 2016)).

measurement of morphological and hemodynamic parameters throughout the cardiac cycle, and it remains an essential instrument in clinical diagnostics.

2.2 Echocardiography

In the evaluation of cardiovascular anatomy and function, transthoracic echocardiography (TTE) is the primary imaging technique due to its cost-effective, noninvasive, and real-time data acquisition capabilities. This modality employs high-frequency sound waves produced by a piezoelectric transducer to penetrate the body, with the reflected echoes being utilised to construct an image of the heart.

Ultrasound (US) is significantly reflected by bones, resulting in acoustic shadows cast on structures located behind them, which complicates image acquisition. To mitigate these challenges, the TTE probe must be positioned to avoid the bones (e.g. between the ribs), limiting the number of possible access points. These access points, known as acoustic windows, allow the heart to be observed only from specific view perspectives, which contributes the first component in describing an echocardiographic view. These windows include the parasternal, apical, subcostal, and suprasternal positions, as depicted in Figure 2.3.

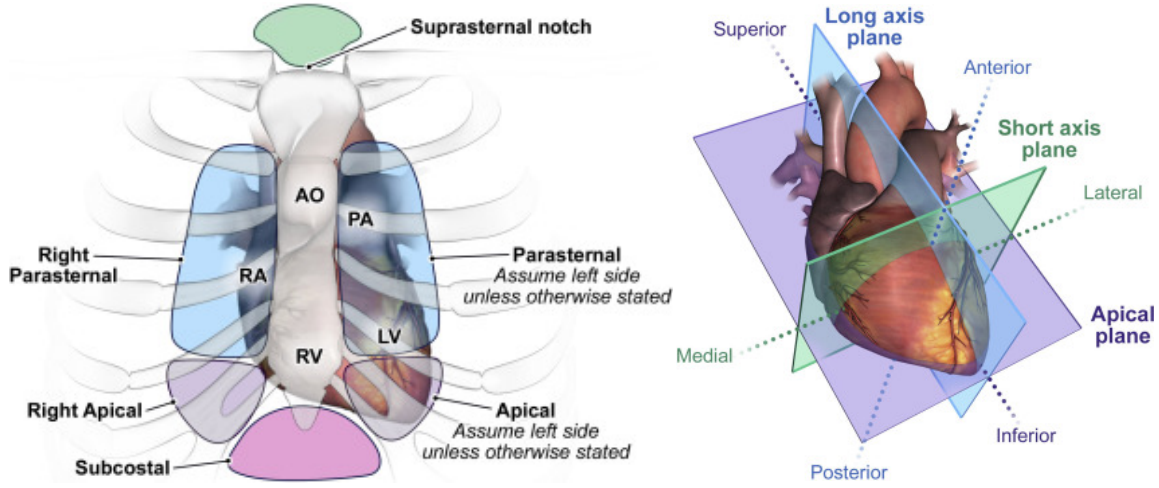


Figure 2.3: The acoustic windows during transthoracic echocardiography as illustrated on the left and the cardiac imaging planes of the heart as shown on the right. The three major axes are indicated for the superior-inferior, medial-lateral and anterior-posterior directions. (Figure adapted from Mitchell et al. (2019)).

When performing 2D echocardiography through these acoustic windows, it is essential to align the obtained recordings with the standard tomographic planes of the heart (Mitchell et al., 2019). These cardiac imaging planes function as the secondary descriptor of an echocardiographic view. The three major planes are given by the long-axis plane, the short-axis plane, and the apical plane. Note that the acoustic windows allow for image acquisition along more than one cardiac plane, since the rotation and tilt of the probe can be adjusted. For example, the parasternal window allows for capturing along both the short- and long-axis planes, which are perpendicular.

The long-axis plane is defined through the long axis of the LV, characterised by a line that extends from the apex to the centre of the mitral valve (MV) (Cerqueira et al., 2002). An additional landmark point is selected to fully determine the plane, typically the centre of the left ventricular outflow tract (LVOT) or the centre of the AV. The corresponding echocardiographic views include the parasternal long-axis (PLAX) view and the apical two-chamber view (A2C), which show the main structures of the heart from the anterior-posterior and superior-inferior directions.

The short axis plane is defined as a plane perpendicular to the long axis and provides cross-sectional views of the main structures at different levels, e.g. at the MV or the AV. The corresponding echocardiographic view is the parasternal short-axis (PSAX) view, encompassing the medial-lateral and anterior-posterior axes.

The apical plane is perpendicular to both the short- and long-axis planes, showing all 4 chambers simultaneously in the medial-lateral and superior-inferior directions. The corresponding echocardiographic views include the apical four-chamber (A4C) and apical five-chamber (A5C) views.

In addition to the diverse set of echocardiographic views, various imaging modes are available, allowing for different types of measurements. To assess the morphology of the heart, B-mode US is employed, producing a greysacle image of cardiac structures. In this mode, "B" stands for brightness, indicating that the intensity of each pixel is associated with the received signal amplitude. The vertical axis within the image displays depth information, whereas the horizontal axis corresponds to the lateral resolution of the scan.

The M mode captures the motion of structures along a single US scan line, displaying how their depth changes over time. This mode provides high temporal resolution, making it useful for evaluating highly dynamic cardiac structures, for example, the rapid movement of the MV or AV (Feigenbaum, 2010), as well as for determining the timing of valve openings and closures (Lang et al., 2015).

Doppler methods are frequently used to gain insight into the velocity profile of blood or tissue. These techniques exploit the Doppler effect, in which the received frequency is modified relative to the emitted frequency as a result of a moving source. By measuring the difference between the emitted and received signals, the velocity magnitude of the moving structure can be calculated as

$$v = \frac{c\Delta f}{2f_0 \cos(\theta)}, \tag{2.1}$$

where c is the speed of sound propagation in the tissue, Δf is the measured frequency shift, f_0 is the frequency of the emitted US signal and θ is the angle of insonation (the angle between the moving object and the US beam) (Kremkau, 1992). Note that this only measures the velocity component parallel to the beam; therefore, an accurate alignment is required for Doppler measurements, rendering θ as close to 0 as possible.

In continuous-wave (CW) Doppler US, the probe continuously sends and receives the signal, sampling all velocities along the beam path. As a result, this method lacks spatial resolution, i.e. it is not able to localise where the velocity was measured. However, this technique yields the best velocity range, making it the ideal candidate for measuring high velocities, which are often present in AS.

Pulsed wave (PW) Doppler imaging, on the other hand, utilises intermittent pulses of US signals to localise the velocity components along its beam. By adjusting the timing between the consequent pulses, also referred to as the time delay, the depth at which the velocity is measured can be adjusted. To allow enough time for deeper echoes to return before the next pulse is sent, the time delay is increased. While allowing for more precise spatial resolution, this method is limited by aliasing. This occurs because of a trade-off between the depth of the sample volume and the maximum observable frequency

$$\Delta f_{\text{max}} = \frac{\text{PRF}}{2}, \tag{2.2}$$

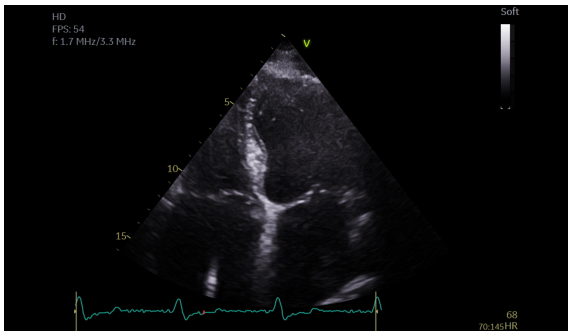
where PRF denotes the pulse repetition frequency, which is equal to the inverse of the time

delay. Thus, by increasing the time delay, the velocity range is decreased. This limitation of maximum observable frequency is due to aliasing and is also known as the Nyquist limit.

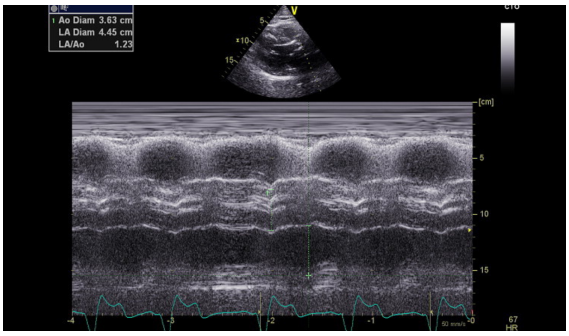
The collection of CW and PW Doppler US techniques is referred to as spectral Doppler imaging, since they both generate velocity-time waveforms by decomposing the Doppler signal into its constituent frequency shifts via the Fourier transform. This means that each frequency shift corresponds to a specific velocity within the sampled region and these waveforms represent the spectral distribution of blood flow velocities over time.

Another Doppler technique overlays the mean blood velocities obtained within an extended sample region on the B-mode echocardiograms to visualise the direction and speed of the blood at each point. This method can be used to examine turbulent regions and plays a role in assessing valvular regurgitation (Anavekar and Oh, 2009).

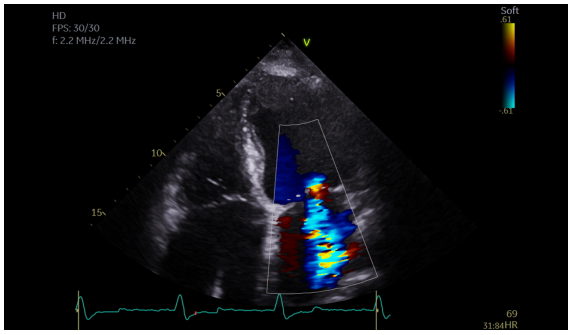
Figure 2.4 displays examples from the echocardiographic modes described earlier. Although colour Doppler and M-mode US both provide valuable information on cardiac function, the most well-established measurements use spectral Doppler to quantify hemodynamic parameters and B-mode US to assess cardiac morphology. The echocardiographic parameters measured during a standard examination to evaluate AS will be discussed in the next section.



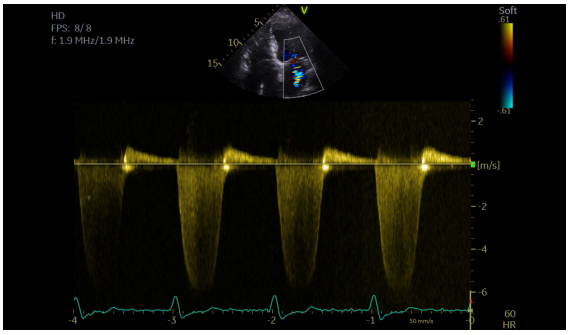
(a) B-mode



(b) M-mode



(c) Color Doppler



(d) Spectral Doppler

Figure 2.4: Examples of common echocardiographic imaging modes. Morphological assessment is performed using B-mode and M-mode imaging as shown in (a) and (b), while Doppler techniques provide quantitative evaluation of blood flow and tissue velocities as illustrated in (c) and (d).

2.3 Aortic Stenosis

2.3.1 Pathophisiology

AS is a type of valvular heart disease characterised by a narrowing of the effective aortic valve area (AVA) due to limited leaflet mobility, frequently resulting from calcific degeneration or a congenital malformation. This narrowing obstructs systolic outflow from the LV into the aorta. Consequently, the afterload — the resistance against which the LV must contract to eject blood — increases. To overcome this elevated systemic vascular resistance and maintain adequate cardiac output, the LV develops concentric hypertrophy, which is indicated by the thickening of the ventricular walls without dilating the chambers.

Depending on how the LV adapts to pressure overload, disease progression can take two main courses (Carabello, 2013). In diastolic heart failure (HF), the initial compensatory mechanism of hypertrophy leads to reduced ventricular compliance, impairing the heart's ability to fill during diastole. The resulting diastolic dysfunction causes an increase in the end-diastolic pressure of the LV, which can be transmitted back to the LA and pulmonary circulation, leading to symptoms of HF. This is also referred to as HF with preserved ejection fraction (EF), as the systolic function of the LV remains intact. In the case of systolic HF, over time the adaptive mechanism fails, resulting in increased wall stress and a dilated LV chamber, which reduces systolic contractility. This is known as HF with reduced EF because the systolic function of LV is impaired.

Understanding the distinct mechanisms of disease progression in AS allows for accurately categorising its severity. While symptoms such as chest pain and breathlessness may suggest the presence of AS, the diagnosis is confirmed through echocardiographic evaluation.

2.3.2 Clinical Assessment

The clinical indicators observed during a standard echocardiographic examination to assess AS can be divided into two main categories: the first includes parameters associated with the dimensions and movements of anatomical structures, while the second comprises hemodynamic parameters that measure the characteristics of blood flow. Although the two categories are intrinsically linked, their measurement requires distinct US modes and views to be used.

Anatomic Parameters

To assess the size of the LV cavity and thickness of the myocardial walls, the LV internal dimension (LVID), the interventricular septum (IVS) thickness, and the LV posterior wall (LVPW) thickness are recorded. LVID denotes the width of the mid-cavity, measured between its inner walls, IVS thickness is determined from the right ventricular side to the left ventricular border of the septum, whereas LVPW thickness is measured on the opposite side, indicating the thickness of the posterior wall. These linear measurements are acquired in the PLAX view with the measurement line perpendicular to the long-axis of the LV.

Volumetric measurements are derived from 2D or 3D echocardiography by tracing the boundary between the myocardium and the LV cavity. The former employs the summation of disks method (also known as the Simpson biplane method), in which the LV volume is calculated by dividing the cavity into a series of elliptical disks, whose dimensions are derived from the endocardial traces in the A4C and A2C views. Since these views are approximately perpendicular to each other, they define the minor and major axes of each elliptical disk, and summing the individual disk volumes along the long-axis dimension provides the approximation for the LV volume.

Acquiring the end-diastolic volume (EDV) and end-systolic volume (ESV) allows the calculation of the stroke volume (SV) through

$$\mathrm{SV} = \mathrm{EDV} - \mathrm{ESV}, \tag{2.3}$$

reflecting the volume of blood ejected by the LV with each contraction. The ejection fraction (EF) is then defined as a percentage by

$$\mathrm{EF} = 100\frac{\mathrm{SV}}{\mathrm{EDV}}, \tag{2.4}$$

serving as a normalised measure of global systolic function that adjusts for variations in heart size. Another way of addressing patient variability is to standardise measurements according to body surface area (BSA). This way of standardisation in clinical practice is referred to as indexing a parameter to BSA, so the adjusted SV is called indexed SV, expressed as

$$\mathrm{SV_i} = \frac{\mathrm{SV}}{\mathrm{BSA}}. \tag{2.5}$$

Cardiac output (CO) is defined as the volume of blood ejected by the LV per minute, which is the product of SV and heart rate (HR), so that

$$\mathrm{CO} = \mathrm{SV} \cdot \mathrm{HR}. \tag{2.6}$$

Although an approximation of the LV volume from linear measurements (e.g. LVID) exists, such as the Teicholz or cubic method, their use is discouraged, as they are based on crude geometric assumptions that do not hold under many pathologies (Lang et al., 2015).

The cross-sectional area (CSA) of the LVOT is estimated by measuring the diameter below the AV in the zoomed-in PLAX view and assuming a circular geometry by

$$\mathrm{CSA_{LVOT}} = \pi\left(\frac{d_{\mathrm{LVOT}}}{2}\right)^2. \tag{2.7}$$

It is not possible to perform a similar approximation of the AVA due to the irregular shape of the opening, hence other techniques are used to obtain this parameter. Although 2D planimetry can be used in the PSAX view to trace the effective orifice area (EOA) in mid-systole, when the valve opening is maximal, this method is highly dependent on image quality and accurate alignment. Current guidelines suggest estimating AVA using hemodynamic parameters (Otto et al., 2021) and recent studies also confirm a better correlation of this method with the values obtained by catheterisation (Rong et al., 2021).

Hemodynamic Parameters

To assess the flow conditions present in patients with AS, spectral Doppler recordings are used. By aligning the US beam with the direction of blood flow through the AV and setting the acquisition mode to CW Doppler, the velocity range is maximised and hemodynamic parameters are obtained at the level of the AV. The maximum velocity, $\mathrm{V_{max}}$, is marked by the highest point on the outer edge of the velocity envelope. The envelope time (ET) is determined by noting t_0 and t_1 as the start and end of the ejection phase, then finding the difference. By tracing the entire envelope during the forward flow and calculating its area, the velocity-time integral (VTI) is obtained, expressed as

$$\mathrm{VTI} = \int_{t_0}^{t_1} v(t)dt. \tag{2.8}$$

The mean velocity is then calculated by

$$\mathrm{V_{mean}} = \frac{\mathrm{VTI}}{\mathrm{ET}} \tag{2.9}$$

The peak gradient across the AV is obtained through the simplified Bernoulli equation (Baumgartner et al., 2009) using the maximum velocity, so that

$$\Delta\mathrm{P_{max}} = 4\mathrm{V_{max}^2}, \tag{2.10}$$

while the mean gradient is calculated by averaging the instantaneous mean gradients (Baumgartner et al., 2009), such that

$$\Delta P_{\mathrm{mean}} = \frac{1}{t_1 - t_0}\int_{t_0}^{t_1} 4v(t)^2 dt \approx \frac{1}{n}\sum_{i=1}^{n} 4v(t_n)^2, \tag{2.11}$$

where the integral is approximated with a finite set of n equally spaced samples between t_0 and t_1.

To obtain the same parameters at the level of the LVOT, PW Doppler US is used, placing the sample volume just below the AV, where a drop in maximum velocity occurs. The AVA is then estimated through the continuity equation as

$$\mathrm{CSA_{AV}VTI_{AV}} = \mathrm{CSA_{LVOT}VTI_{LVOT}}, \tag{2.12}$$

where the cross-sectional area (CSA) of the LVOT is obtained by equation 2.7 and the VTI through equation 2.8. Figure 2.5 demonstrates hemodynamic measurements at the AV and LVOT levels and presents the continuity equation.

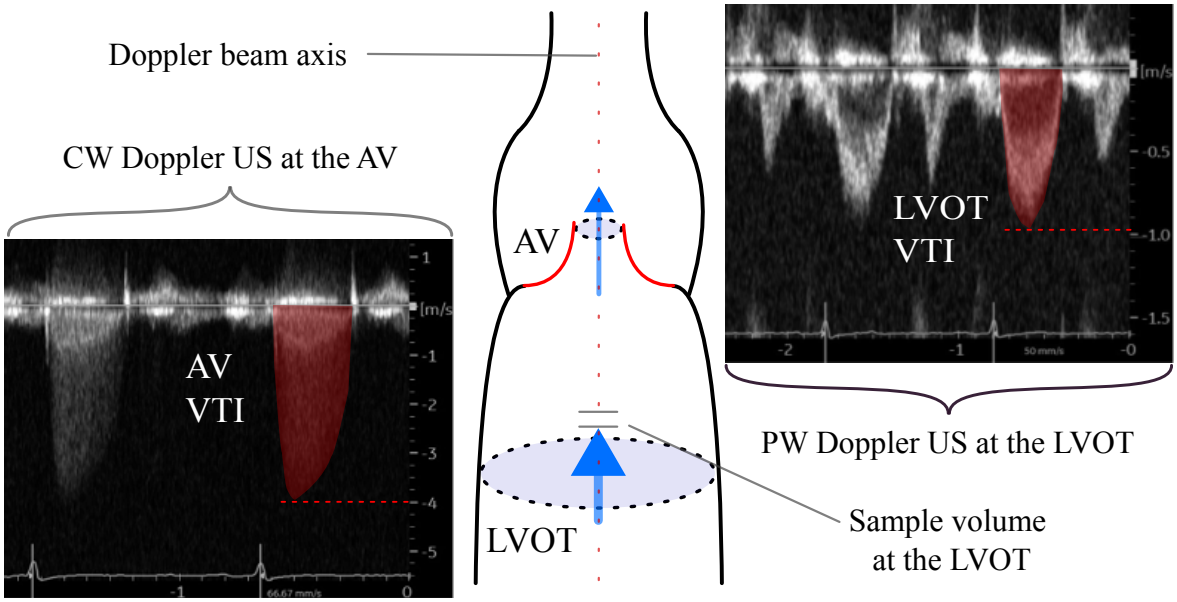


Figure 2.5: An illustration of hemodynamic measurements obtained by continuous-wave (CW) Doppler ultrasound (US) at the left ventricular outflow tract (LVOT) and by pulsed-wave (PW) Doppler US at the level of the aortic valve (AV). The arrows indicate the direction of blood flow, while the areas shaded in red represent the velocity-time integral (VTI) measurements used by the continuity equation to calculate the aortic valve area. Note the difference in scales for peak velocities between the two measurements.

2.3.3 Low-Flow, Low-Gradient Phenotypes

Given the continuous range of disease progression from mild aortic sclerosis to severe AV calcification, the definition of different AS severity grades can be considered somewhat arbitrary. Nevertheless, the timing and nature of the intervention are determined based on specific AS grades (Otto et al., 2021). Current clinical guidelines suggest classifying AS using hemodynamic parameters such as maximum velocity and mean transvalvular gradient, along the AVA and/or indexed AVA (Baumgartner et al., 2017) according to the values presented in Table 2.1. Any parameter falling into the severe column suggests the presence of severe AS.

In most cases, prognostic markers are in agreement; however, discordance may arise in which specific parameters classify the patient into a different severity category than others. This occurs in approximately one in three cases of AS (Silva et al., 2022), when the valve area is below 1.0 cm^2, indicating severe AS, and the mean gradient remains below 40 mmHg with

	Aortic sclerosis	Mild	Moderate	Severe
Peak velocity [m/s]	≤2.5	2.6-2.9	3.0-4.0	≥4.0
Mean gradient [mmHg]	-	<20	20-40	≥40
AVA [cm2]	-	>1.5	1.0-1.5	<1.0
Indexed AVA [cm2/m2]	-	>0.85	0.6-0.85	<0.6

Table 2.1: Reference values for grading aortic stenosis (Baumgartner et al., 2017). Although classification employs discrete cutoff values, it is essential to consider the measures of disease severity as a continuous spectrum.

normal transvalvular flow, suggesting mild or moderate AS.

In these instances, the flow state is further investigated using the indexed SV, differentiating between patients with normal flow and low flow conditions. Normal flow is defined by $SV_i > 35ml/m^2$, while the low-flow low-gradient (LFLG) phenotype is considered with SV_i below this value. Global LV systolic function is assessed among cases of LFLG type AS to further categorise disease pathology. A LVEF ≥ 50% is defined as a paradoxical LFLG AS, as it demonstrates a low flow state despite the preserved LV function. In contrast, a LVEF less than 50% signifies the classic LFLG scenarios.

To verify the severity of AS in patients with normal flow and paradoxical LFLG type AS, calcium scoring by computed tomography (CT) is used to measure valve calcification (Cueff et al., 2011; Clavel et al., 2013). A calcium score exceeding 2000 for men and 1200 for women indicates severe AS, whereas scores falling below these thresholds suggest non-severe AS.

In classic LFLG cases, further examination is performed by stress testing, which involves the administration of a drug called dobutamine, in an attempt to improve LV contractility. The objective of dobutamine stress echocardiography (DSE) is to assess changes in transvalvular hemodynamic parameters and AVA at increased flow rates, as well as to determine the presence of flow reserve, quantified by an increase in SV.

The diagnostic rationale is as follows: In true severe AS, advanced disease produces a significantly restricted EOA and secondary LV systolic dysfunction, resulting in a misleadingly low baseline transaortic velocity. During DSE, additional flow reserve manifests itself as an increase in peak velocity and mean gradient rather than an additional enlargement of the valve area, since the maximum AV opening had already been reached at resting heart rate.

However, in moderate (pseudo-severe) AS, LV dysfunction is suspected to be a symptom of another pathology (e.g. cardiomyopathy, myocardial infarction), and the reduced EOA at baseline reflects insufficient LV contractile energy to fully open the valve. The dobutamine-induced improvement in LV contractility in this context increases AVA with only a slight increase in the transvalvular gradient, thus distinguishing moderate from true severe stenosis (Baumgartner et al., 2017). Differentiating between true-severe and pseudo-severe AS

is especially important considering the different interventions available for each case. Although true severe patients are prescribed transcatheter or surgical aortic valve replacement (TAVR/SAVR) to relieve LV afterload, thus potentially allowing the LV to regain some of its contractile function, pseudo-severe patients are not expected to benefit from this procedure, and heart failure therapy is the recommended course of treatment (Fougeres et al., 2012).

The DSE protocol requires incremental administration of low-dose dobutamine and evaluation of the main hemodynamic and anatomical predictors at each stage using TTE, until one of the following criteria is met:

- The maximum dobutamine dose is reached.
- The test is considered positive by observing one of the following: (i) an increase in AVA above the 1 cm^2 threshold, consistent with pseudo-severe AS (Carabello, 2002); (ii) an increase in maximal transvalvular velocity and mean gradient beyond the severe limits (4 m/s and 40 mmHg, respectively) while AVA remains < 1cm^2, indicating true severe AS; or (iii) absence of adequate flow reserve, which — regardless of AS severity — identifies a subgroup at elevated risk for postoperative mortality (Monin et al., 2003).
- The maximum HR is reached, specified by exceeding 100 beats per minute (BPM), or 20 BPM above the resting HR.
- An adverse event occurred, such as arrhythmia, blood pressure drop, or other symptoms.

DSE is considered inconclusive when sufficient flow reserve is not generated (i.e. the increase in SV is less than 20%), or the AVA-gradient discordance persists under elevated flow conditions. In these cases, the projected valve area can be used among other experimental markers; however, a CT calcium scoring examination is required to accurately determine the true severity of AS. Figure 2.6 shows a flow chart of the entire diagnostic procedure.

Generally, diagnosing LFLG AS is quite complex, susceptible to measurement errors, and carries potential risks to patients due to the DSE procedure. This consideration serves as the main motivation for the thesis, which investigates the potential of computational methods to advance clinical practice.

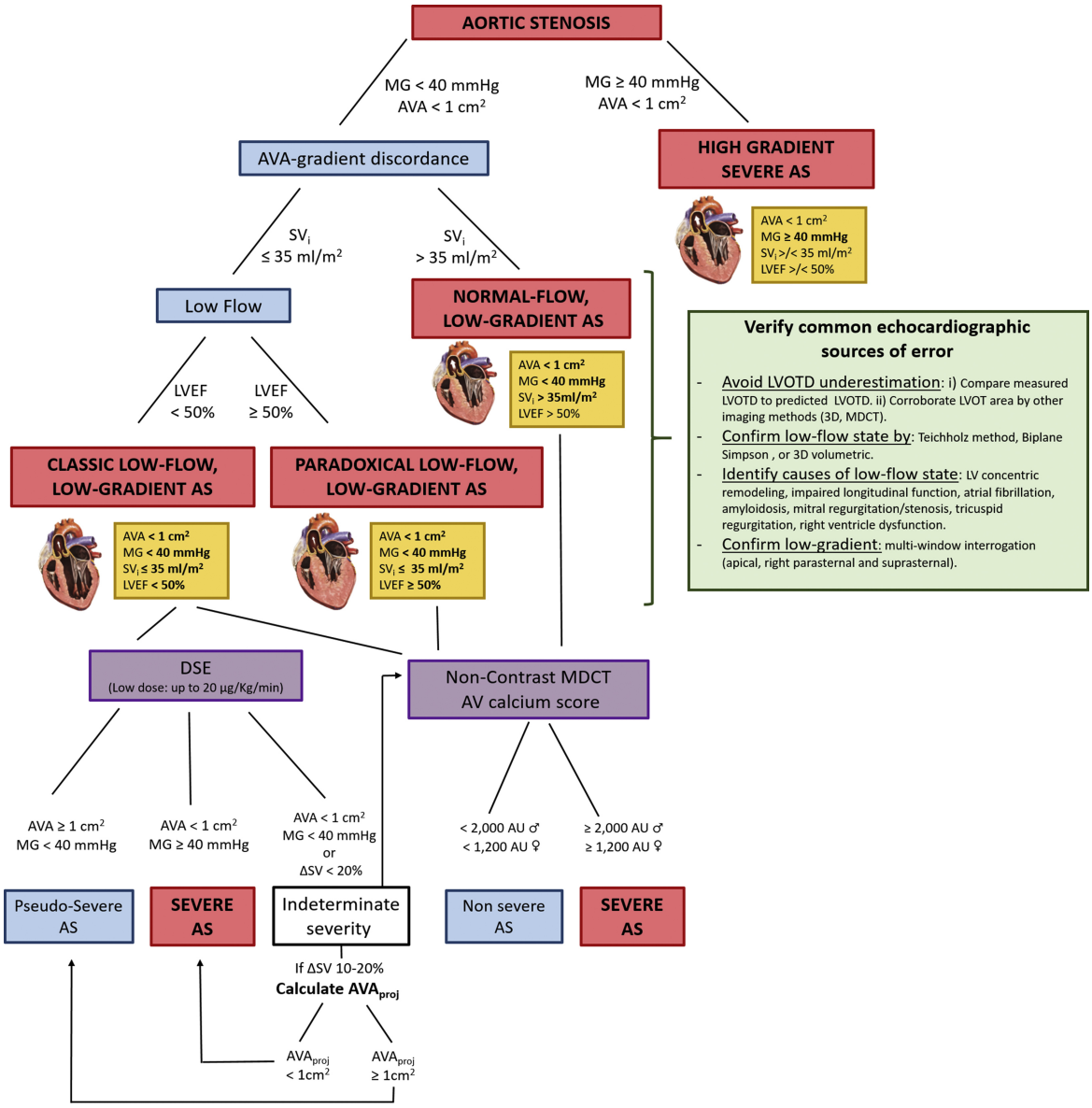


Figure 2.6: The clinical grading algorithm of aortic stenosis (AS). In the high gradient track all cases are considered severe, while the low gradient track is further refined. The flow state is split based on the indexed stroke volume (SV_i) in cases of discordance between the aortic valve area (AVA) and mean transvalvular gradient to differentiate between low flow and normal flow conditions. In low flow states, the ejection fraction (EF) is used to measure global left ventricular (LV) function, categorising patients with paradoxical and classic low-flow, low-gradient (LFLG) AS. Calcium scoring by computed tomography (CT) is recommended for normal flow and paradoxical LFLG AS patients, while dobutamine stress echocardiography (DSE) is recommended in classic LFLG AS cases. To differentiate between pseudo-severe and true severe AS, the AVA and mean gradients are examined during the increased flow conditions and the severity is determined if the parameters adhere to the grading criteria presented in Table 2.1. Inconclusive scans may arise as result of lack of flow reserve or persisting AVA-gradient discordance, where experimental predictors may be used to guide treatment. (Figure adapted from Silva et al. (2022)).

3. Literature Review

This chapter presents a comprehensive overview of emerging research directions in echocardiographic evaluation of AS. Section 3.1 reviews the limitations of current clinical practice, Section 3.2 examines the key literature that addresses these limitations, while Section 3.3 summarises the main findings and identifies the research gap that motivates the work presented in this thesis.

3.1 Limitations of Current Practice

Although TTE remains the primary noninvasive modality for assessing AS due to its accessibility and ability to provide hemodynamic data, it is inherently constrained by several limitations. These challenges become more pronounced in patients with LFLG AS, where even minor inaccuracies in acquisition or interpretation can significantly affect diagnostic reliability and disease classification.

3.1.1 Technical Limitations

One of the most prevalent technical limitations of TTE in AS evaluation is significant interobserver and intraobserver variability in routine measurements. This is particularly evident in the measurement of the diameter of the LVOT, which, when squared, has a substantial influence on the AVA derived from the continuity equation, as expressed in Equation 2.12. The current consensus states 5-8% variability in LVOT diameter measurements, making it the largest contributor to the AVA error (Baumgartner et al., 2017). A recent investigation on the reproducibility of standard echocardiographic predictors by Manna et al. (2024) demonstrated that the same day re-examination could lead to a change in AS classification in 27% of the patients, underscoring the clinical impact of observer-dependent variability. In this study, they also found that two-thirds of the disagreement in AS grades could be attributed to differences in AVA, while a discrepancy in $V_{\max}$ contributed only one-fourth of the differences in classifications. This finding also reflects the differences in the magnitudes of interobserver variability (IOV) for AVA and $V_{\max}$, which are 12.7% and 3.5%, respectively, further highlighting the effect of measurement errors in clinical practice.

Although the variability described above is primarily attributable to random measurement errors, these arise from deviations around the true value and tend to average out over repeated assessments. For example, variability in Doppler beam alignment can lead to an underestima-

tion of transvalvular velocities and gradients, and inconsistent placement of the PW Doppler sample volume may result in sampling flow acceleration instead of true LVOT flow - both introducing random errors in AVA estimation via the continuity equation (Ring et al., 2021).

In addition, echocardiographic evaluation is also subject to systematic errors, leading to a consistent bias in clinical predictors due to methodological assumptions or imaging limitations. Differentiating between these two types of error is essential, not only for understanding the sources of diagnostic variability but also for identifying which limitations may be mitigated through an improved acquisition and grading protocol and which require alternative measurement methodology.

An example of systematic bias in 2D TTE is the assumption that the LVOT has a circular cross-sectional geometry. In reality, LVOT is often elliptical, as observed by alternative imaging modalities such as CT, 3D transesophageal echocardiography (TEE), and cardiac magnetic resonance (CMR) imaging. This geometric simplification leads to a median underestimation of the LVOT CSA by 20% to 28% compared to 3D-TEE, CT, and CMR planimetry, which in turn results in an underestimation of AVA when using the continuity equation (Velders et al., 2023).

Another methodological limitation arises from utilising the chamber quantification of the LV through Simpson's biplane rule, which serves as the primary technique in 2D TTE. This method relies on geometric assumptions that i) the A2C and A4C planes are perpendicular and centrally aligned along the long axis of the LV and ii) the basal end of the LV is perpendicular to the long axis. Kim et al. (2022) examined the dependence of LV volume on viewing and basal orientation angles when calculated by different methods found that Simpson's rule suffers from a consistent underestimation of LV volume, which is further magnified by increasing deviation from the idealised geometric assumptions when compared to the ground truth values. Cowie et al. (2023) found that this method underestimates SV by an average of 18.3% compared to the gold standard CMR technique.

3.1.2 Physiological Limitations

Beyond technical limitations, the echocardiographic assessment of AS is also restricted by fundamental physiological assumptions. Standard clinical predictors such as AVA, SV and LVEF are influenced by hemodynamic states and loading conditions that are often not taken into account to accurately reflect true myocardial function or valvular severity. The following subsections explore how such physiological dependencies introduce errors and ambiguities in current practice.

Flow Dependence

Although the AS grading algorithm, as presented earlier in Figure 2.6, uses SV_i to differentiate between normal and low flow states, volume is fundamentally different from flow. The transvalvular flow rate is defined as

$$Q = \frac{SV}{ET}, \tag{3.1}$$

which also accounts for the time component of the ejection phase. This relationship illustrates that low flow states can result not only from reduced SV but also from an increase in ET. The contribution of ejection dynamics to Q becomes particularly relevant in patients with AS, as ET is often prolonged and only reduces to a seemingly normal range due to LV dysfunction associated with an advanced stage of the disease or concomitant HF (Lewis et al., 1977). In contrast, a reduced SV may be compensated for by a shorter ET to yield sufficient transvalvular flow.

Inadequate representation of flow in AS grading can result in an inaccurate assessment of true AVA, since maximum valve opening may not be induced under low flow conditions. This is supported by a recent study by Namasivayam et al. (2020) investigating the effect of AVA on overall mortality in patients with AS at varying flow rates. They have shown that below the cutoff point of 242ml/s, AVA $\leq 1cm^2$ was a poor prognostic marker of severe AS, while above this threshold AVA $\leq 1cm^2$ was highly associated with mortality.

Ejection Fraction

Another physiological limitation concerns the use of EF as a surrogate for global LV function. While the classification of LFLG AS into classical and paradoxical phenotypes is based on LVEF, with the classical phenotype defined as LVEF $< 50\%$ and paradoxical as LVEF $\geq 50\%$, this distinction is based on an arbitrary cut-off point and does not fully capture the complexity of LV dysfunction. As expressed earlier in equation 2.4, LVEF is the ratio of SV to EDV. Although in acute conditions such as heart attack, a reduction in SV accurately reflects systolic dysfunction, in chronic and stable conditions, a preserved SV with elevated EDV can still produce reduced LVEF, indicating diastolic dysfunction rather than true systolic impairment (Weir et al., 2025). Thus, the assumption that low flow in classical LFLG AS is solely due to reduced contractility may be misleading, since diastolic dysfunction frequently coexists and contributes to the low flow state. Depending solely on LVEF to differentiate between the subtypes of LFLG AS could obscure the inherent variability in the LV function.

Another significant limitation of employing EF to assess LV function in patients with AS is its ineffectiveness to detect early myocardial dysfunction. This limitation is due to the potential for EF to remain above the recommended threshold of 50% in the initial stages of the disease, aided by compensatory hypertrophy that maintains cardiac output despite elevated valvular resistance. Consequently, a reduction in EF below 50% typically indicates an advanced disease associated with the onset of symptoms and permanent myocardial damage (Henkel et al., 2012; Ito et al., 2018). This is supported by the fact that patients with AS and reduced EF

face increased surgical risks and significantly worse long-term survival compared to patients with preserved EF, while many do not fully recover LV function after surgery, implying that reduced EF often represents a late-stage and irreversible dysfunction (Dahl et al., 2015; Toya et al., 2021; Spilias et al., 2022).

Dobutamine Stress Echocardiography

DSE is used to differentiate true severe from pseudo-severe AS in patients with classical LFLG AS, however, several limitations undermine its diagnostic utility. A key issue is the reliance on SV increase by 20% to define contractile reserve, yet approximately one third of patients do not meet this criterion, resulting in diagnostic uncertainty regarding the severity of AS (deFilippi et al., 1995; Weir et al., 2025). Furthermore, while SV may not increase, transvalvular flow rate can still normalise due to heart rate–induced shortening of ET, thus allowing for a better approximation of true AVA (Clavel et al., 2010; Pibarot and Dumesnil, 2012). Vamvakidou et al. (2017) found that a more conclusive assessment of AS severity could be obtained using Q as opposed to SV as a measure of contractile reserve, highlighting the inadequacy of SV alone as a marker of flow response.

The diagnostic utility of DSE is further limited by heterogeneity in flow response, which can vary due to concomitant conditions such as coronary artery disease and cardiac amyloidosis, affecting up to 52% and 30% of the population with LFLG AS, respectively (Tribouilloy et al., 2009; Pibarot and Dumesnil, 2012; Ternacle et al., 2019). This is particularly relevant in cardiac amyloidosis, which not only shows a similar LV pathophisiology seen in LFLG AS - such as increased wall thickness and impaired systolic and diastolic function - but may also contribute to disease progression (Jaiswal et al., 2023; Weir et al., 2025). Although current data indicate that patients with both AS and cardiac amyloidosis have poor surgical outcomes and a worse overall prognosis compared to those with lone AS, the presence of concomitant disease introduces considerable diagnostic uncertainty and further complicates management decisions in this already complex AS subgroup (Nitsche et al., 2021).

Finally, a recent investigation found that the accuracy of DSE in identifying true severe AS when using standard criteria (mean transvalvular gradient $\geq$ 40mmHg and AVA$\leq$ 1cm^2) was 47% (Annabi et al., 2018). Another research by Adrichem et al. (2024) has strongly advised against using CT-derived AV calcium scoring as a substitute for DSE in patients with classical LFLG AS. Their study found that the recommended calcium score thresholds resulted in a sensitivity of 44.3% and a specificity of 56.5% when differentiating between severe and moderate AS. Together, these limitations underscore the need for alternative methods or better disease interpretations to improve the accuracy and reliability of the assessment of AS severity in the low-flow state.

3.2 Emerging Research Directions in AS Diagnosis

3.2.1 Novel Clinical Predictors

In recent years, there has been growing interest in identifying novel markers that better reflect the underlying hemodynamics and myocardial function. These predictors aim to overcome flow dependence, serve as an early marker of the disease to provide refined insight into the pathophysiology of AS and to enhance diagnostic precision beyond traditional metrics such as AVA, SV_i and LVEF. This section explores the current evidence supporting their utility and prognostic value in the context of LFLG AS.

Transvalvular Flow Rate

To address the misrepresentation of flow by SV_i, the transvalvular flow rate, Q (Eq. 3.1), is used to provide a more physiologically accurate measure of the aortic valve opening stimulus. A recent study by Bar Gil et al. (2025) has shown that reduced Q is an independent predictor of all-cause mortality in patients with AS, serving as the main evidence for incorporating Q into standard diagnostic and risk stratification algorithms for AS. Furthermore, Vamvakidou et al. (2019) found that in low-gradient AS, Q was also associated with medium-term mortality in patients undergoing TAVR, while SV_i was unable to discriminate between survival groups. Finally, current data from Hungerford et al. (2025) suggest that Q measured at peak stress is able to differentiate true-severe from pseudo-severe AS in patients with LFLG AS undergoing DSE, while no differences in Q were found between the severity groups at rest.

In addition to its diagnostic utility, the transvalvular flow rate also allows inferring the degree to which the AVA calculated by the continuity equation reflects the true valvular opening. Chahal et al. (2015) examined patients with LFLG AS who underwent DSE and concluded that significant changes in AVA were only observed when the resting flow rate was below 200 ml/s, while AVA in patients with normal resting Q values were representative of the severity of AS. This is further reinforced by Namasivayam et al. (2020), who found that the recommended cut-off point for AVA$\leq$ 1cm^2 was only prognostic for mortality when Q was greater than 242 ml/s, suggesting that flow dependence is more accurately assessed by Q and not by SV_i.

Projected Valve Area

The projected aortic valve area (AVA_{proj}) is used to mitigate the underestimation of the true AVA that occurs under low flow conditions. Unlike standard DSE criteria, which depend on an increase in SV_i, AVA_{proj} estimates the valve area at a normalised transvalvular flow rate of 250 ml/s, thus addressing flow dependence (Blais et al., 2006; Clavel et al., 2010). It is defined as

$$AVA_{proj} = AVA_{rest} + \frac{AVA_{peak} - AVA_{rest}}{Q_{peak} - Q_{rest}}(250 - Q_{rest}), \tag{3.2}$$

where both Q and AVA are measured in the resting and maximum stress states during DSE with units ml/s and cm^2, respectively. AVA_{proj} is then extrapolated from these to estimate the valve area the patient would have at a normal flow rate. Although this approach still requires an increase in Q by a minimum of 15%, it offers a more reliable distinction between true severe and pseudosevere AS in patients with limited flow reserve (Blais et al., 2006).

A prospective study analysing the utility of DSE in patients with classical LFLG AS by Annabi et al. (2018) revealed that AVA_{proj} had better performance in predicting the severity of AS compared to current clinical guidelines (70% vs. 47%). They have also found that a low projected valve area is associated with an increased risk of mortality, while high transvalvular gradients did not predict mortality. Another study by Sato et al. (2019) examined an AS population with a low gradient and a reduced LVEF. Their findings indicated that AVA_{proj} is more effective than standard DSE criteria in identifying patients who would benefit from AVR.

Dimensionless Index

The Dimensionless Index (DI) is a Doppler-based parameter defined as the ratio of the VTI measured at the level of the LVOT to the VTI measured at the level of the AV, that is

$$DI = \frac{\mathrm{VTI}_{\mathrm{LVOT}}}{\mathrm{VTI}_{\mathrm{AV}}}. \tag{3.3}$$

In contrast to the AVA derived from the continuity equation, DI does not require measurement of the LVOT diameter, bypassing geometric assumptions and eliminating a major source of random measurement error. In the absence of AS, the DI is expected to be close to 1, while a DI<0.25 has been proposed as a marker of severe AS (Oh et al., 1988; Baumgartner et al., 2017).

Multiple studies have evaluated the prognostic relevance of DI in AS. Rusinaru et al. (2015) demonstrated that in patients with AS and preserved EF, progressively lower DI values were associated with a gradual decline in event-free survival. Altes et al. (2020) examined a cohort of AS patients with preserved EF and low transvalvular gradients and found that a DI below 0.25 effectively identified a subset of patients who were at increased risk of mortality and could benefit from aortic valve replacement. Another study by Namasivayam et al. (2022a) has shown that DI<0.25 was associated with a higher risk of mortality in low gradient severe AS.

Acceleration Time to Ejection Time

Although AS has long been documented to affect ejection dynamics, only in recent years has there been evidence to establish reference values for their diagnostic utility in native AS (Gamaza-Chulián et al., 2017; Maréchaux and Tribouilloy, 2021). The use of ET alone is limited by the variability in HR and LV function. The time interval from the beginning of the systolic flow to the maximum transvalvular velocity defines the acceleration time (AT), and the ratio of AT to ET serves as a more accurate marker of the ejection dynamics. Moderate

AS has been observed to have a rapid flow acceleration, while severe AS is characterised by a slower, more rounded flow contour, as observed in the spectral Doppler recordings, reflecting the delayed and reduced valve opening (Bermejo et al., 2000).

Contemporary evidence from Altes et al. (2019) suggests a threshold of >0.34 to identify severe AS. They have also shown that AT/ET correlated well with traditional markers of severity of AS, such as peak velocity, mean transvalvular gradient, AVA and CT-derived aortic valve calcium scores, demonstrating its strong prognostic value. Another study by Einarsen et al. (2021) found that AT/ET >0.32 was independently associated with a significantly higher risk of cardiovascular death and hospitalisation for HF in a large cohort of asymptomatic patients with mild to moderate or severe AS with low gradient. Chong et al. (2024) further validated AT/ET in a population with paradoxical LFLG AS, identifying a cutoff value of ≥0.35 as a strong predictor of adverse cardiac outcomes. In this cohort, patients with AT/ET ≥0.35 who underwent AVR had markedly improved survival, highlighting the value of this marker in guiding treatment decisions. Finally, a recent study by Sancar et al. (2022) demonstrated that baseline AT<100ms successfully distinguished between true severe and pseudosevere cases in patients with classical LFLG AS undergoing DSE, achieving a sensitivity of 77% and a specificity of 87%. While their study was limited in size (n=67), this finding strongly motivates the use of ejection dynamics prior to DSE in patients with reduced EF and low flow conditions.

3.2.2 Myocardial Function by Speckle-tracking Echocardiography

All of the markers described previously aim to address the limitations of standard echocardiographic techniques with regard to flow; however, the pathophysiology of AS also concerns LV remodeling. In response to the constraints of using EF fraction to reflect LV function, speckle tracking echocardiography (STE) has become a valuable tool to provide a more accurate and detailed representation of myocardial deformation (Karlsen et al., 2019; Mihos et al., 2025). This technique relies on analysing the unique acoustic patterns present in echocardiographic recordings and tracking the displacement of these speckles from one frame to another. By capturing the movement of the LV myocardium throughout the cardiac cycle, several deformation parameters can be obtained to quantify its contractraction. The most well established parameters include the strain measures, which for a one-dimensional linear deformation expresses the relative change in length of a line segment as

$$\epsilon(t) = \frac{L(t) - L_0}{L_0}, \tag{3.4}$$

where $L(t)$ represents the length of the line segment at time t, while L_0 represents the reference length, commonly measured at the end of diastole.

Strains are typically expressed in the longitudinal, circumferential, and radial directions, as shown in Figure 3.1, reflecting the different types of contractions that the LV exhibits. For

example, longitudinal strain captures the deformation of the myocardium from the base to the apex, indicating relative shortening of the LV along the long axis during systole. In addition, circumferential strain captures the deformation along the circumference of the myocardial wall, whereas radial strain measures the extent of the wall thickening across the width of the myocardium. Although deformation parameters can be obtained in all parts of the LV, the most common metrics include strain measures that evaluate the global LV systolic function by averaging the strain values in each subregion (Anastasiou et al., 2023). In this context, global longitudinal strain (GLS) acts as an alternative metric to assess LV contractility with improved prognostic utility compared to LVEF (Stanton et al., 2009).

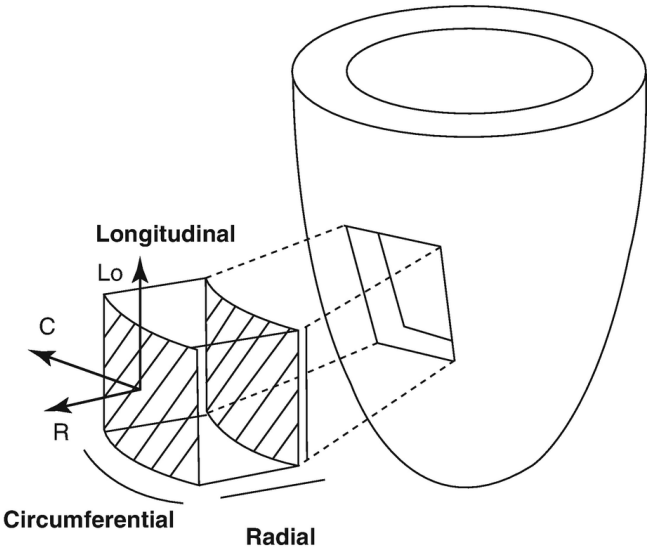


Figure 3.1: The different directions used to quantify strains in the assessment of left ventricular (LV) function. The longitudinal direction follows the myocardium from the base to the apex, the circumferential direction is defined by the perimeter of the LV as seen on the short axis view, while the radial direction spans the thickness of the myocardial wall. (Figure adapted from Karvandi and Ranjbar (2023)).

The potential advantages of incorporating GLS into clinical practice include its ability to detect early myocardial dysfunction before LV hypertrophy develops and LVEF declines (Imbalzano et al., 2011; Kusunose et al., 2014; Vollema et al., 2018). This may be due to the sensitivity of longitudinally aligned subendocardial fibres to elevated wall stress (making them the first to be affected by increased afterload), as fibrotic changes in AS typically begin subendocardially and predominantly impair longitudinal function, which is not adequately captured by EF (Weidemann et al., 2009). To support this, a recent study by Slimani et al. (2020) investigated stress-strain indices based on GLS and global circumferential strain (GCS) in patients with AS undergoing TAVR. They have found that patients with lower preoperative indices demonstrated a significantly greater extent of biopsy-confirmed myocardial fibrosis than patients without. GLS could additionally assist in distinguishing different types of LV hypertrophy and help identify whether LV dysfunction is associated with AS-induced remodeling or concomitant amyloidosis-related myocardial impairment (Afonso et al., 2012; Castaño et al., 2017).

Global Longitudinal Strain in Aortic Stenosis

Multiple studies have examined the prognostic importance of STE-derived GLS in the AS population with reduced and preserved LVEF. Ng et al. (2018) demonstrated that patients with impaired GLS defined as < 14% are subject to a markedly higher risk of long-term mortality in a subgroup of severe AS and preserved LVEF (> 55%). The research carried out by Wang et al. (2022) also confirmed that reduced GLS was strongly associated with worse cardiovascular outcomes in patients with asymptomatic AS, regardless of EF, severity of AS, or mean transvalvular gradient, highlighting the importance of incorporating GLS into clinical diagnostics. A study by Dahou et al. (2015) examined the utility of GLS in patients with LG AS and reduced EF. Although resting GLS did not differentiate between long-term survival groups after adjustment for early mortality, stress |GLS| < 10% was independently associated with worse outcomes in patients treated conservatively. Furthermore, stress LVEF was not predictive of mortality in this subgroup, indicating the additional prognostic value of GLS over conventional echocardiographic measures and confirming the benefit of using DSE in this AS phenotype.

GLS has also been shown to have a prognostic value in the LFLG AS phenotype. Sato et al. (2014) examined a cohort of severe paradoxical LFLG AS and found that a preserved |GLS| > 17% was associated with a significantly higher rate of event-free survival at a two-year follow-up compared to patients with reduced GLS (97.2% vs. 57.5%). They have also concluded that this prognostic variability from poor to benign could be attributed to the heterogeneity of distinct AS phenotypes present in the paradoxical LFLG AS group defined by standard echocardiographic metrics that could be differentiated by GLS. D'Andrea et al. (2019) investigated the predictive value of resting GLS on flow reserve during DSE and reverse remodeling of the LV six months after TAVR in patients with severe classical LFLG AS. They have shown that using a cutoff value of |GLS| < 12% was able to predict the lack of flow reserve with a sensitivity of 84% and a specificity of 93%. They have also found that in patients without flow reserve, GLS improved significantly after TAVR while LVEF did not, indicating the improved representation of LV function by GLS. Kim et al. (2020) observed that an improvement in GLS during DSE by more than 2% in patients with severe classical LFLG AS resulted in better survival compared to those without. They have also shown that GLS has an additive predictive value for mortality, in addition to maximum transvalvular velocity.

Regional Deformation Parameters

Although the assessment of global LV function by STE is increasingly adopted in routine echocardiographic examination, current evidence on the impact of AS on regional deformation parameters of the LV remains limited, especially in the various subtypes of AS. This is particularly relevant, since different types of geometric adaptation of the LV (for example, concentric remodeling and hypertrophy) have been observed to cause distinct alterations in longitudinal strain values at different myocardial levels (Cramariuc et al., 2010). These levels

are defined with respect to the long axis of the LV, which extends from the apex to the mid-base point, dividing the left and right regions of interest when visualised in a 2D apical view, as shown in Figure 3.2. The myocardium along each side of the long axis is typically divided into four segments: the apical cap, the apical segment, the mid-cavity (or middle) segment, and the basal segment. The apical cap comprises approximately 1/7 of the total myocardial length on either side, whereas the apical, middle and basal segments each consist of 2/7 of the myocardium.

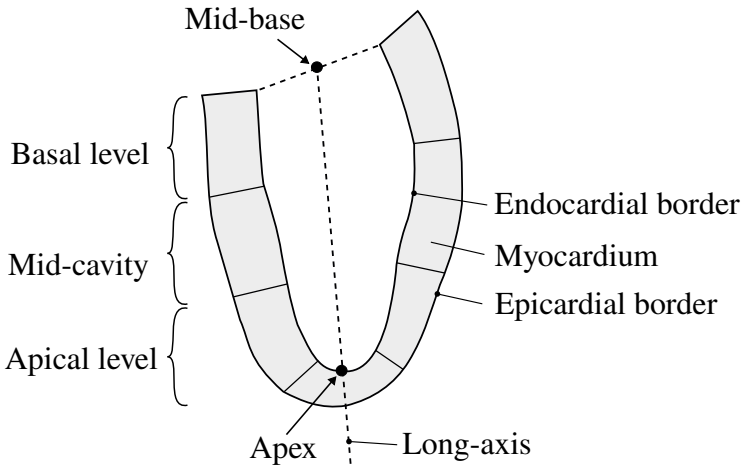


Figure 3.2: Myocardial segments at different left ventricular levels as depicted from an apical view. The long-axis, defined through the apex and mid-base points, divides the myocardium into two parts, which are further split based on the length of the myocardium on each side. (Figure adapted from Voigt et al. (2015)).

Some studies have explored the importance of regional deformation parameters in patients with AS. Attias et al. (2013) found that both GLS and basal longitudinal strain (BLS) were significantly different between AS groups when stratified by symptomatic status and LVEF, however, only BLS was independently associated with symptomatic status. A study by Carstensen et al. (2016) examined regional longitudinal strains in asymptomatic moderate and severe AS. They showed that the longitudinal strains measured in the basal and mid-cavity segments differed between the survival groups when divided by AVR referral. Furthermore, $|\text{BLS}| < 13\%$ was able to identify patients with significantly lower event-free survival rates, while GLS thresholds were not associated with future AVR, highlighting the importance of evaluating regional parameters alongside global strain metrics. A study by Dahl Pedersen et al. (2020), investigating the ratio of apical to basal strains in patients with AS who underwent TAVR, found that a significantly higher risk of mortality was associated with a ratio greater than 4. Recent research by Thellier et al. (2024) found that increased LV dyssynchrony as measured by myocardial dispersion (the deviation in time to peak strain of each myocardial segment) is associated with a higher risk of mortality in severe AS.

In addition to various strain measures, regional rotation parameters obtained by STE were also explored in several smaller studies. Tumenbayar et al. (2018) examined apical rotation in patients with severe AS and preserved LVEF and compared it with a cohort exhibiting

LV hypertrophy. They found that patients with severe AS had significantly increased apical rotation compared to the hypertrophy group. Increased apical rotation was also shown to be associated with reduced survival in a separate study by Holmes et al. (2015) that examined patients with severe AS. More recently, Hjertaas et al. (2023) provided evidence that basal rotation, as well as apical-basal twist, differ significantly between AS severity grades, indicating that rotation parameters may also contribute to the clinical diagnosis of AS.

3.2.3 Machine Learning in Aortic Stenosis

A different approach to improving the echocardiographic assessment of AS from the investigation of novel markers involves the optimisation of decision rules that separate severity grades or distinct patient phenotypes. The current clinical grading algorithm can be considered a rule-based system, in which echocardiographic recordings are interpreted based on hand-designed anatomical and hemodynamic characteristics, and the relationship between severity or phenotype classes and clinical predictors is determined manually. This determination is often made through studies that perform survival analysis and risk stratification using single-feature cutoff values. However, this approach poses significant limitations, since it overlooks the inherent relationship between echocardiographic variables.

Machine learning (ML) algorithms aim to address this by incorporating multiple features simultaneously, facilitating the discovery of more complex relationships between clinical markers and AS pathology. ML algorithms can be divided into supervised and unsupervised categories according to their learning paradigms. In supervised learning, the algorithm is presented with both input features and ground-truth output labels, and the model learns the relationship between input and output variables. In contrast, unsupervised learning aims to uncover patterns present in the input data without explicit labelling of the output variables. The former can be used to refine the classification of the severity or prognostic outcome of AS, while the latter can be used to reveal distinct patient phenotypes.

Classical Machine Learning Algorithms

Research on the use of supervised ML techniques that incorporate only echocardiographic data remains sparse, with only limited evidence on their utility in patients with AS. Namasivayam et al. (2022b) demonstrated that their model using nine selected echocardiographic and clinical variables from patients with moderate and severe AS (n=1130) was able to predict a composite outcome of AVR and mortality with an AUC of 0.74 and 0.78 in internal and external validation sets (n=540), respectively. In addition, their model was also able to identify patients with high-risk mortality in the low-gradient severe AS cohort (n=383). Similarly, Itelman et al. (2025) developed an ML algorithm using standard echocardiographic parameters to predict whether the progression of patients with mild and moderate AS (n=3443) to a severe stage occurred within 5 years. Their model demonstrated a sensitivity of 81%

and specificity 85% with an AUC of 0.91, highlighting its strong predictive performance in identifying high-risk patients. Recently, Wrzosek et al. (2025) evaluated the performance of a supervised ML algorithm in differentiating between moderate AS and low-gradient severe AS, reaching an AUC of 0.853 when using only six routine echocardiographic parameters from 135 patients with AS.

Studies exploring unsupervised ML algorithms in patients with AS are more common. Sengupta et al. (2021) used topological data analysis to identify high-risk and low-risk phenogroups using only echocardiographic predictors within the clinical classification algorithm (AVA_i, SV_i, LVEF, MG and Vmax) in 1052 patients with AS. Consequently, they examined the integrity of these groups in two external validation cohorts (n=752 and n=160). They found that high-risk patients had significantly higher calcium scores derived from CT examination and increased myocardial fibrosis measured by CMR. The prognostic significance of these groups was also investigated by time-to-event analysis, as indicated by AVR or death. They showed that high-risk patients face significantly poorer clinical outcomes, which relationship remained when the groups were further stratified by clinical guideline-based cutoff values for LVEF, AVA and mean transvalvular gradient. Several other studies have also explored this method using laboratory and clinical patient data along with echocardiographic recordings (Kwak et al., 2020; Bohbot et al., 2022; Kusunose et al., 2023).

Recently, Sen et al. (2025) explored the use of different clustering algorithms in patients with moderate AS (n=2469), identifying four distinct groups associated with different prognostic profiles. Using an external validation cohort of 1358 patients with moderate AS, they found that groups with low flow and cardiovascular comorbidities were associated with the worst clinical outcomes, with an increased risk of AVR, hospitalisation for HF, and mortality. Furthermore, AVR mainly benefited the calcific AV group, while it did not significantly affect the remaining groups, highlighting the additive value of subphenotyping AS in determining the appropriate time and nature of treatment strategies. Similarly, Lachmann et al. (2021) investigated the heterogeneity of severe AS using echocardiographic and hemodynamic data in 366 patients undergoing TAVR, revealing four additional groups presenting with different prognostic implications.

Deep Learning

Deep learning (DL) refers to a specific subset of ML in which models are represented by several successive layers of neural networks. The potential advantages of employing DL models over classic ML algorithms lie in their ability to process echocardiographic recordings directly by learning not only the mapping between input features and output labels, but also by discovering latent feature representations. The distinction between the different approaches of rule-based systems, classical ML, and DL, is shown in Figure 3.3.

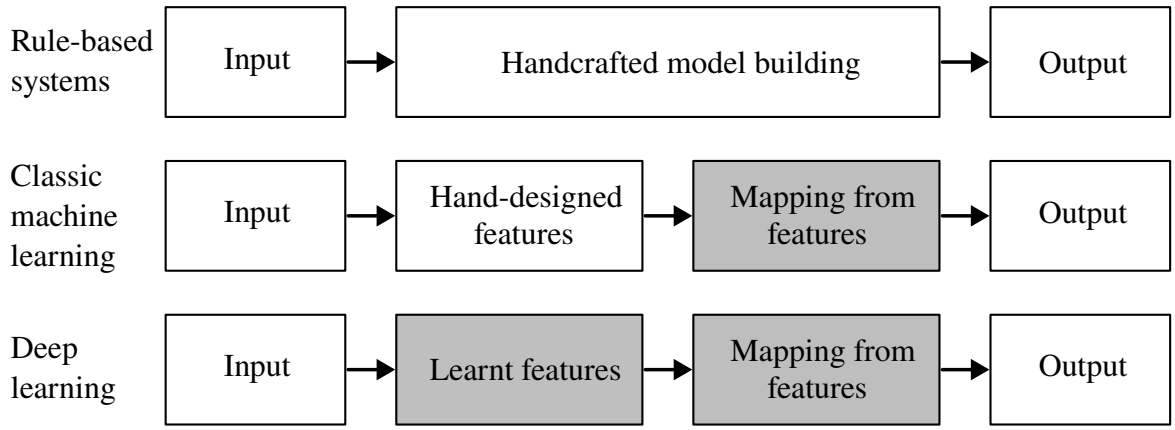


Figure 3.3: The distinction between explicit programming of rule-based systems, Machine Learning (ML) and Deep Learning (DL) algorithms. The shaded areas represent the functional components learnt from data, indicating that DL extends ML by an added layer of abstraction. (Figure adapted from Janiesch et al. (2021)).

The most straightforward application of DL algorithms in the echocardiographic assessment of AS involves the automatic prediction of standard parameters to reduce IOV and improve reproducibility. Since clinicians perform several tasks when interpreting a single echocardiographic recording, initial DL research focused on reproducing simple tasks such as recognising a specific view or predicting a single anatomical landmark or parameter. The view classifier algorithm developed by Madani et al. (2018) using 267 echocardiograms demonstrated excellent performance in recognising 12 different types of routine echocardiographic views with a test accuracy of 97.8%. Gungor et al. (2023) expanded the view classification task by automating the detection of the aortic and mitral valves using 11150 B-mode cine loops. More complex tasks have involved segmentation of the LV. Leclerc et al. (2019) applied deep learning to the end-systolic and end-diastolic frames of 2D TTE recordings from 500 patients to predict the endocardial and epicardial borders. Their best models demonstrated good alignment with the reference values for ESV and EDV, with Pearson's correlation coefficients of 0.964 and 0.954, respectively. Ouyang et al. (2020) assessed the LV function over time using more than 10000 annotated 2D B-mode cine series. Their model showed strong agreement with clinicians in the calculation of EF, and segmentation masks for the LV cavity for end-systolic and end-diastolic frames showed a high overlap with expert annotations, measured by the dice similarity coefficient.

In recent years, automatic measurement of several additional anatomical and hemodynamic parameters has also been achieved. Strain measures were examined by Salte et al. (2023) and Nyberg et al. (2024) using A4C and A2C echocardiographic recording from 72 and 80 patients, respectively, demonstrating that deep learning models enabled markedly improved reproducibility compared to human experts in both global and regional longitudinal strains. Excellent correlation was also demonstrated for Vmax and VTI measured by CW and PW spectral Doppler echocardiography at various levels by Jeon et al. (2024) employing 25854 Doppler recordings, with an average correlation coefficient of 0.95 and 0.965 respectively. Most recently, a comprehensive DL model developed by Sahashi et al. (2025) using more

than 150000 studies showed high accuracy and reproducibility compared to sonographer annotations in predicting nine anatomical and nine hemodynamic parameters. Although the methods used in the studies described above are valuable for their ability to reduce the workload of clinicians required to examine a patient and improve overall reproducibility, they do not inherently contribute to a better understanding of the disease pathology alone.

The great success of DL in the automatic interpretation of echocardiograms has also motivated the development of models specifically aimed at the detection of AS. Initial efforts focused on performing a simplified task of distinguishing between severe and non-severe AS using single-view PLAX videos from 2D TTE recordings (n=17570), as explored by Holste et al. (2023). Their model demonstrated excellent predictive performance in three separate external validation cohorts, with AUC scores ranging from 0.942 to 0.978. Ginsberg et al. (2021) jointly used PLAX and PSAX views by majority voting to identify severe AS in a cohort of 2247 patients, achieving an F1 accuracy score of 96.5%.

Successive studies have aimed to improve the classification of multiple severity levels using integrated approaches that include additional downstream tasks, along with simultaneous prediction of AS, to facilitate concordance with the clinical decision-making process and thus enhance model explainability. A model introduced by Huang et al. (2024) was trained to prioritise clinically relevant views given a set of 2D B-mode images per study using labelled and unlabelled data from 577 and 5287 patients, respectively. Their model achieved a balanced accuracy of 76% in the three-way severity classification task (no AS, early AS, significant AS). Ahmadi et al. (2024) improved on these results by employing DL models trained jointly with 2D PSAX and PLAX cine series, reporting accuracies of 95.2% for AS detection and 78.1% for four-class severity classification.

The most recent advances in diagnosing AS from 2D TTE have been a major leap in both scope width and accuracy. Holste et al. (2025) developed a comprehensive DL system capable of performing 18 diagnostic tasks and 21 routine parameter estimation tasks simultaneously using over 1 million echocardiographic recordings. Their system demonstrated a median AUC of 0.91 for classification tasks and a normalised mean average error of 0.13 for regression tasks. Park et al. (2025) proposed a new approach by predicting a continuous score rather than a distinct AS class alongside the prediction of routine echocardiographic parameters ($V_{\max}$, AVA and mean transvalvular gradient) using 8427 patients. Their model obtained an accuracy of as high as 96.8% for the three-way AS classification in an external test set (n=772). Their AS score was also associated with prognostic implications for the composite outcome of AVR, HF, and mortality, indicating an improved representation of the disease pathology compared to previous methods. Furthermore, while all studies excluded discordant AS from the training procedure, this study was the first to explicitly evaluate LFLG AS phenotypes in the test set, thus investigating the possibility of obtaining markers under normal flow conditions that differentiate between severe and moderate LFLG AS. Their model predicted somewhat

higher overall scores for severe than for moderate LFLG AS; however, no distinction was made between the classical and paradoxical cases, and the overlap between the distributions remained quite large, making it difficult to accurately distinguish between severe and moderate cases.

3.3 Summary

Although TTE remains the main method for evaluating AS, it is limited by several factors, including high variability in the measurement of standard parameters and systematic biases introduced by geometric and physiological assumptions. The current clinical grading algorithm also presents with important constraints by incorporating clinical markers that exhibit flow dependence, such as AVA, or do not capture the complexity of disease pathology, such as LVEF. Furthermore, the recommended protocol for DSE remains inconclusive in terms of AS severity in approximately one in three patients, while it poses potential risks to patients.

Several novel markers have emerged in recent years to advance the diagnostic pipeline for AS, yet their utility in differentiating true-severe and pseudo-severe classical LFLG AS remains mostly unexplored. These markers include both an improved hemodynamic representation of AS through transvalvular flow rate, projected valve area, or acceleration time, as well as detailed LV myocardial assessment through regional and global STE-derived deformation parameters. In addition, current data on the impact of DSE on these markers are sparse, with little to no evidence on class-specific responses of different disease phenotypes. Furthermore, contemporary ML literature on the use of echocardiographic data alone to derive composite markers to improve severity and subgroup classification of AS is also limited.

DL algorithms have shown excellent agreement with expert annotations in predicting standard echocardiographic parameters, and their use in the diagnosis of AS has improved significantly in recent years through the refinement of training objectives. All existing DL studies excluded discordant AS from their training procedure, and their performance on the classical and paradoxical LFLG AS phenotype remains unknown. Finally, despite DSE being the recommended test in classical LFLG AS to confirm disease severity, to the best of our knowledge, no research has investigated the additive value of stress testing in differentiating between true-severe and pseudo-severe AS using DL algorithms.

4. Materials

This chapter outlines the various sources and types of data used throughout this thesis, providing a reference point for the techniques introduced in later chapters. Since the overall objective of this thesis was to benefit the assessment of LFLG AS, each method was evaluated on clinical data from a cohort of 12 patients, termed as DSE12. This data set comprises echocardiographic recordings of 12 patients who underwent DSE at Golden Jubilee National Hospital, and is described in Section 4.1. For statistical analysis and machine learning techniques, an additional data set of 60 patients with concordant AS was used. This is denoted as the AS60 data set and is detailed in Section 4.2. The strain imaging methodology, which focuses on obtaining 3D LV deformation parameters from triplane echocardiographic data, leverages an open-source data set called Myocardial Wall Motion Analysis (MWMA) (Puyol-Anton et al., 2019), described in Section 4.3. Finally, DL models were developed using the Tufts Medical Echocardiographic Dataset (TMED) (Huang et al., 2022), as presented in Section 4.4.

4.1 Low-Flow, Low-Gradient Aortic Stenosis Data Set

The central data set of this thesis comprises various echocardiographic recordings and clinical measurements from 12 patients diagnosed with LFLG AS who underwent DSE at the Golden Jubilee National Hospital, along with their respective diagnostic labels. This is denoted as the DSE12 data set. The following markers were recorded by clinicians or measured manually according to the guidelines recommendations.

- Hemodynamic parameters from spectral Doppler recordings performed at the level of the AV and LVOT. Routine clinical predictors included peak velocity (V_{max}), mean velocity (V_{mean}), peak pressure gradient (maxPG), mean pressure gradient (meanPG), velocity time integral (VTI) and envelope time (Env.Ti). The corresponding heart rate for each recording was also observed. An additional measurement of the acceleration time was conducted at the AV level.
- Standard one-dimensional anatomical measurements of the LV and LVOT from the PLAX view included the LVID, LVPW, IVS thickness, and LVOT diameter.
- LV volumetric and functional measurements were taken from the A4C and A2C views. EDV and ESV were calculated using the Simpsons biplane or Teicholz method, while GLS was obtained from GE EchoPac software.

From these primary measurements additional clinical predictors were derived. The AVA was calculated from the continuity equation assuming a circular LVOT geometry as introduced in Equation 2.12. SV, EF, CO, Q, and DI were obtained using the equations described in Chapter 2 and 3. BSA was used to index SV and AVA, and additional novel markers that could be evaluated from the above measurements such as AT/ET and meanPG/AVA were also calculated. An example of the various types of echocardiographic recordings used is shown in Figure 4.1, while the complete patient data from the DSE12 data set is included in Appendix A.

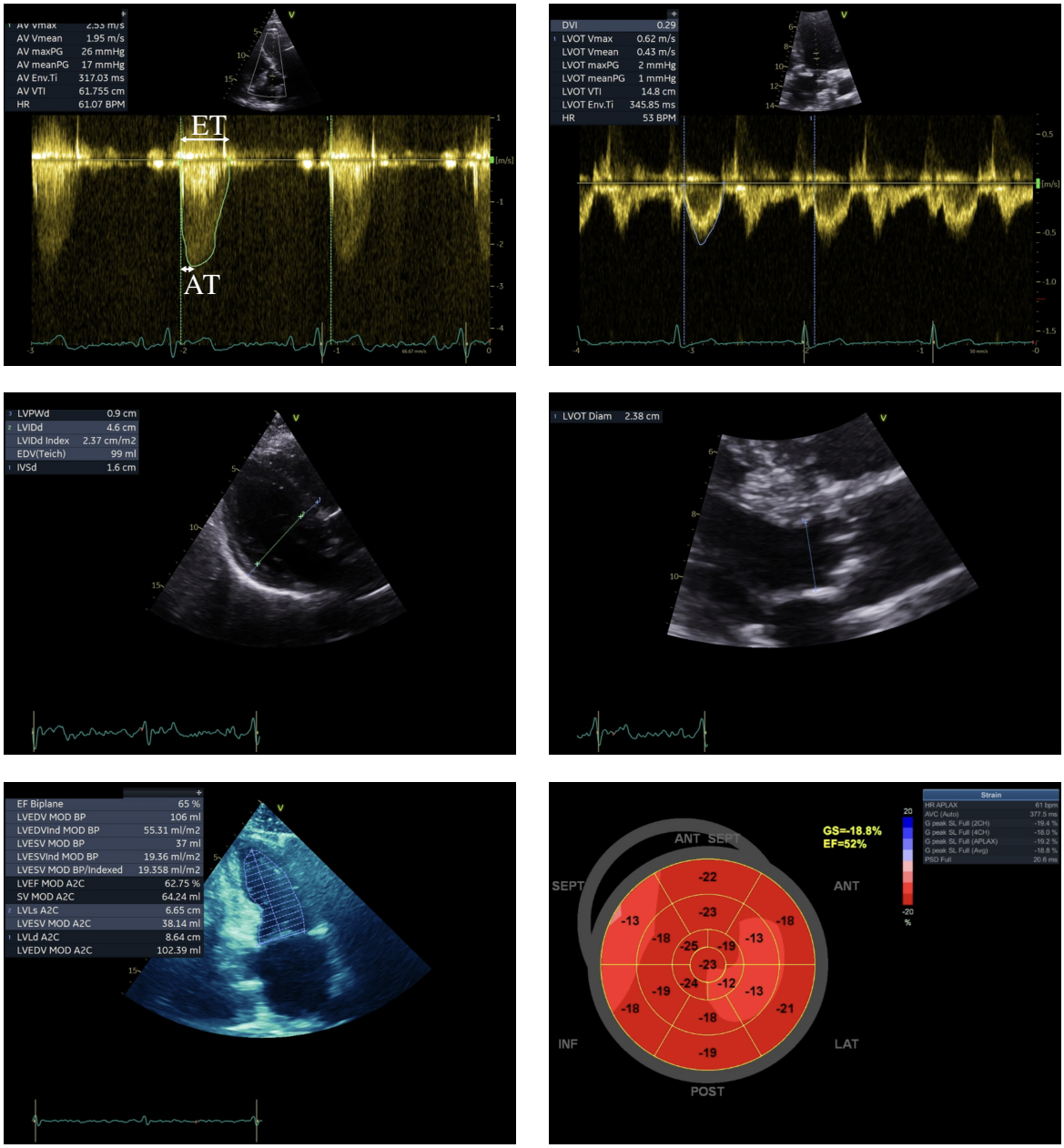


Figure 4.1: Clinical examples of routine echocardiographic measurements. The top row depicts hemodynamic parameters obtained at the level of the aortic valve (AV) and at the level of the left ventricular outflow tract (LVOT). The middle row shows the anatomical measurements of the left ventricle (LV) and LVOT diameter, while the bottom row illustrates the volumetric measurements of the LV and the assessment of LV contractile function by global longitudinal strain (GLS).

Some predictors were excluded from the data set. For example, the calculation of SV by Doppler echocardiograhy resulted in inconsistencies with those measured by volumetric quantification of the LV, while the calculation of projected valve area also proved to be inaccessible in most cases due to inadequate flow reserve. GLS values were only recorded for 3 patients; therefore, these were also excluded from the statistical analysis. For other missing values, such as BSA, the average of the AS60 data set was used, as this represented a larger cohort, thus providing a more reliable estimate of the population mean.

The main characteristics of the DSE12 dataset with respect to AS severity and flow conditions are as follows. Of the 12 patients, 8 had paired recordings collected at both baseline (resting) and stress-induced conditions, allowing intrapatient comparisons of hemodynamic and morphological changes. Among these eight patients, one scan was marked as inconclusive, while three presented severe AS and four moderate AS. Of the remaining four patients, who had only baseline data, three had moderate AS and one had severe AS.

4.2 Concordant Aortic Stenosis Data Set

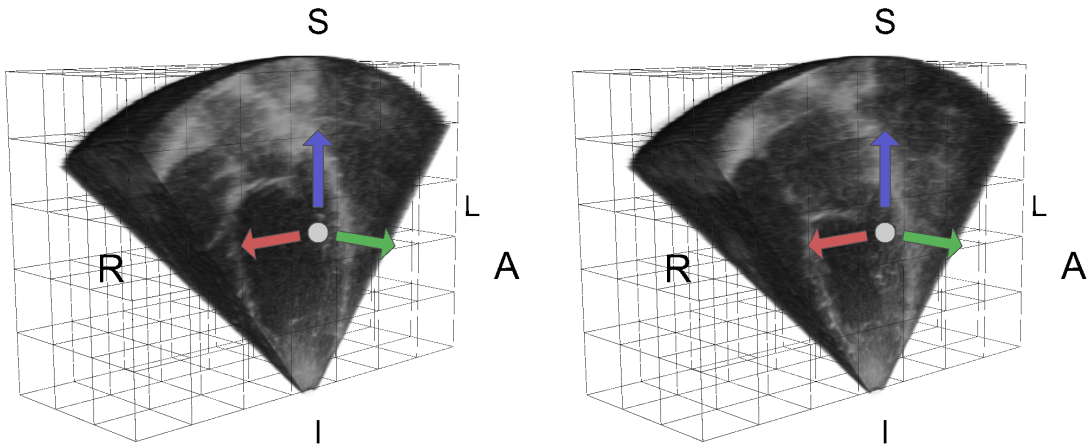
The data set called AS60 comprises echocardiographic recordings of 60 patients with concordant AS, indicating that their maximum transvalvular velocity and/or mean pressure gradient were in agreement with the area of the aortic valve. All recordings were acquired under resting heart rate conditions. Each patient was assigned an AS severity label of mild, moderate, or severe, according to standard clinical guidelines presented in Table 2.1. The measurements used in this data set include an identical set of hemodynamic and anatomical predictors as described in the DSE12 data set, with an additional GLS recorded for the majority of patients (41 out of 60) using EchoPac software. The complete measurement data from the AS60 data set are provided in Appendix A.

The distribution of AS grades in the AS60 data set was balanced, with 20 patients in each category. One patient (Patient 44) was excluded from the analysis due to inconsistencies between their measurements and the diagnosis label. Triplane recordings were available for seven patients in total, five with mild AS and one in each of the moderate and severe categories. However, regional myocardial function was not evaluated in this cohort and GLS was measured using commercial software.

This data set served multiple purposes. First, it allowed for various hypothesis tests using a statistical framework to assess significant differences between AS severity grades and different flow conditions, as presented in Chapter 5. Second, it was used to train ML models aimed at predicting the severity of AS using only echocardiographic predictors. Finally, it provided an external validation cohort to evaluate the generalisation performance of trained DL algorithms on previously unseen clinical data, as explored in Chapter 7.

4.3 3D Echocardiography Data Set

Given the inaccessibility of retrospective 3D US data from LFLG and concordant AS cohorts, the triplane methodology was developed using an external data set, called Myocardial Wall Motion Analysis (MWMA). This incorporated apical full-volume 3D US sequences of the LV of 9 healthy participants as described by Puyol-Anton et al. (2019). These recordings were stored in the Neuroimaging Informatics Technology Initiative (NIfTI) format (Cox et al., 2004). The voxel spacing ranged from 0.7 mm to 1.0 mm, with image sizes typically spanning from 208 × 240 × 176 to 272 × 256 × 224 voxels along the X, Y, and Z axes, respectively. Each image sequence contained between 19 and 31 frames, starting with the end-diastolic frame as the first frame in the series. An example is shown in Figure 4.2.



EDV (t = 0) ESV (t = 12)

Figure 4.2: Example recordings of 3D ultrasound volumes from the myocardial wall motion analysis data set as displayed at the end-diastolic [left] and end-systolic frames within the cardiac cycle. The arrows represent the three major axes and standard orientation labels are provided for right (R), left (L), superior (S), inferior (I), and anterior (A) directions.

This data set was used to simulate a synthetic triplane acquisition mode, compare myocardial deformation parameters between 3D and triplane imaging approaches, and evaluate the reconstruction error in the LV geometry.

4.4 Tufts Medical Echocardiogram Data Set

Currently, TMED2 is the only publicly available data set that facilitates the development of AS severity classification algorithms from TTE recordings, and its use in DL research has been widely adopted as a benchmark to evaluate new architectures. It comprises of:

- The view and diagnosis labelled set with 599 studies from 577 unique patients, where all patients have an AS diagnostic label, and some images from each study have view label annotations. This is also referred to as DEV479 in the literature as the number of

studies in the development set (training and validation partitions) adds up to 479.

- The view labelled set with 705 studies from 703 unique patients associated with the view labels only, and can be used for the development of view classifier algorithms.
- The unlabelled set with 5486 studies of 5287 patients without any associated view labels or AS severity grades, and is commonly used for pre-training purposes.

Patients are classified according to diagnostic grades as having no, mild, mild to moderate, moderate or severe cases of AS as assigned by clinical guidelines, while the view labels correspond to PLAX, PSAX, A4C, A2C, or A4CA2C/other anatomical view classes. All studies comprise a varying number of greyscale images with a size of 112x112 pixels stored in a PNG format. Although diagnostic labels denote five distinct grades of AS severity, the usual learning objective in this data set is to predict three separate categories: no AS, early AS, or significant AS. In this classification regime, mild and mild to moderate cases constitute the early AS group, while moderate and severe correspond to the significant AS category.

Recent studies have also used three predefined cross-validation splits to assess the performance of DL models. These splits are structured to ensure equal distributions of studies with regard to severity labels and the number of images per study across each fold. A summary of each fold is presented in Table 4.1, including the distribution of severity labels, as well as the number of images per study in the training, validation, and test splits. Given that DEV479 is considered a common benchmark for the development of learning algorithms that aim to predict the severity of AS, the models developed in Chapter 7 will use this data set for training and validation purposes, and will also evaluate their generalisation performance using private clinical data from the AS60 and DSE12 data sets.

Split	**Number of Studies**				**Avg. Images per Study**		
	no_AS	**early_AS**	**significant_AS**	**All**	**no_AS**	**early_AS**	**significant_AS**
Fold 1							
Train	76	103	181	360	13.0	21.0	95.5
Val	25	34	60	119	14.0	28.0	102.0
Test	25	34	60	119	14.0	35.0	114.5
Fold 2							
Train	76	103	181	360	13.0	22.0	100.5
Val	25	34	60	119	14.0	23.0	97.0
Test	25	34	60	119	10.0	28.0	110.0
Fold 3							
Train	76	103	181	360	13.5	22.0	99.5
Val	24	34	60	118	13.0	22.0	95.5
Test	26	34	60	120	13.5	31.0	98.0
Total							
All	126	171	301	598	—	—	—

Table 4.1: Distribution of severity labels and average number of labeled images per study across the training, validation, and test splits for each fold in TMED2.

5. Analysis of Clinical Predictors Using Statistical and Machine Learning Methods

This chapter presents the statistical analysis and application of classical machine learning techniques using standard and novel echocardiographic predictors. Section 5.1 reiterates the research gaps addressed in this study, followed by a detailed description of the methodological framework in Section 5.2. The results and key findings are presented in Section 5.3. Section 5.4 includes the conclusions of this chapter.

5.1 Introduction

With emerging evidence supporting the clinical utility of novel echocardiographic markers (such as the ratio of acceleration time to ejection time and transvalvular flow rate) in the evaluation of both concordant and discordant AS, the refinement of the current AS grading algorithm may be achieved by analysing these predictors between different AS phenotypes and severity grades. Although DSE remains the recommended examination for classical LFLG AS, the impact of this procedure on individual predictors remains sparsely explored, with no severity-specific cut-off values for most markers. DSE can also remain inconclusive, in which case it provides little or no benefit in the assessment of the severity of AS. The efficacy of DSE in differentiating severe and pseudo-severe classical LFLG AS is not yet established when using novel clinical markers. In addition, the cut-off values for AVA and mean transvalvular gradient as observed at maximum stress state in classical LFLG AS is identical to the concordant AS population, which assumes that stress testing causes predictors to conform to normal flow conditions. This assumption has also not been validated yet.

In this chapter, the examination of both standard and novel echocardiographic predictors is covered with respect to the different hemodynamic states and severity grades using a comprehensive statistical analysis framework. The efficacy of DSE is also inferred by examining intrapatient stress responses in a wide range of echocardiographic predictors. Classical ML techniques are employed in the concordant AS cohort to predict the severity of AS and their generalisation performance was assessed in both resting and stress states to facilitate learning a representation of the severity of the disease that is applicable in the resting state.

5.2 Methodology

This section outlines the methods used to assess the efficacy of DSE in stratifying the diagnosis of LFLG AS, utilising various clinical predictors documented within the AS60 and DSE12 data sets. First, a univariate analysis was performed to explore the distribution of each predictor and to perform several hypotheses tests. Subsequently, the study progressed to using multivariate analysis methods to identify the main combination of features that affect the variability and class separation present in the AS60 data set. Next, various ML models were implemented to classify the severity of AS using the AS60 data set. Then, these models were evaluated for their ability to generalise to the DSE12 data set and a set of synthetically generated patients. This assessment aimed to determine the efficacy of DSE in classifying the severity of LFLG AS, learnt by decision rules under normal flow conditions, and to examine the degree to which the decision rules complied with clinical guidelines.

5.2.1 Univariate Analysis

As the simplest approach to analysing the recorded data is through univariate analysis, the first technique used in this section examined each feature in the clinical data set on its own, without considering how the features might interact with one another. To explore the main trends, the mean and standard deviation of each severity group (mild, moderate, and severe AS) were calculated.

After examining the feature distributions individually, the analysis progressed to performing statistical tests aimed at evaluating various hypotheses:

- Test **I.** compared the distribution of clinical predictors between the different severity groups of AS within the AS60 data set to explore which markers provide the best distinction between the AS groups under normal flow conditions. The null hypothesis, denoted as H_0, stated that, under normal flow conditions, there are no significant differences between the group means/medians. The test included a total of 59 patients, as the measurements of one patient did not match clinical guidelines.
- Test **II.A/1** was conducted to determine if the clinical predictors evaluated previously in test **I** still differentiate the groups in the DSE12 data set recorded in the baseline state, where LFLG hemodynamic conditions applied. H_0 here stated that there are no significant differences between the baseline means/medians of severity groups under LFLG hemodynamic conditions. The test incorporated a total of 11 patients from the DSE12 data set, as one patient's diagnosis was inconclusive.
- Test **II.A/2** assessed the similarity between the baseline recordings of DSE12 and the AS60 patients across each severity groups to evaluate the resemblance in the distribution of clinical predictors between the two data sets. H_0 proposed that no

significant difference exists between the DSE12 baseline and AS60 measurements within the respective severity groups. Given that the DSE12 data set lacked patients diagnosed with mild AS, comparisons were made with the moderate and severe groups, consisting of 7 and 4 patients, respectively.

- Test **II.A/3** conducted a similar analysis as **II.A/2**, however, it utilised the stress recordings from the DSE12 data set instead of the baseline. H_0 stated that there is no statistically significant difference between the DSE12 stress and AS60 recordings within the corresponding severity groups. Similarly, the test was performed for groups with moderate and severe AS, comprising 4 and 3 patients, respectively.
- Test **II.B/1** examined whether changes in the stress response could substantially influence predictors within individuals, regardless of the severity of AS. H_0 stated that there is no significant difference between the means/medians of the baseline and stress recordings. As the diagnostic label did not pertain to this test, the patient with inconclusive scans was readmitted, resulting in a total of 8 patients.
- Test **II.B/2** investigated whether the change of each feature from baseline to stress differed between the severity groups. H_0 stated that the means/medians of the change in predictors from baseline to stress do not significantly differ among the severity groups. The DSE12 subjects involved in this test are identical to those in test **II.A/3**.

Table 5.1 summarises the various hypothesis tests conducted in this chapter, including the objectives and comparison groups of each test, as well as the associated sample sizes.

To perform the appropriate statistical test for each hypothesis, continuous clinical predictors, acting as dependent variables, were categorised. The Shapiro-Wilk test was used to check whether the feature distribution adhered to a normal distribution. A p-value greater than 0.05 indicated that the outcome variable conformed to a normal distribution, allowing for parametric statistical tests. In contrast, p-values below 0.05 suggested that the distribution of the feature was skewed, necessitating the use of non-parametric statistical tests. Subsequently, the number of categorical variables for comparison was classified into two types. When comparing more than two groups (e.g., in test **I.** and **II.A/1**, where the independent variables comprised three severity groups), the Analysis of Variance (ANOVA) was selected for the parametric test, whereas the Kruskal-Wallis H-test was utilised as the non-parametric counterpart. These tested for statistically significant differences in means/medians across the independent groups. To compare two independent groups (as seen in test **II.A/2** and **II.A/3**) with a normally distributed outcome variable, the Student's t-test was employed. Conversely, the Mann-Whitney U-test was utilised for comparing two independent groups when the predictors were skewed. To compare paired groups, such as in test **II.B/1** where baseline and stress recordings were interrelated, the Wilcoxon Signed-Rank test was applied as the non-parametric test, while the paired t-test was employed for parametric analysis. Figure 5.1 shows a flow chart illustrating the selection process for the suitable test.

Test	Objective	Comparison	Sample Size
I	Assess which echocardiographic predictors distinguish AS severity under normal flow conditions.	Mild vs Moderate vs Severe in AS60	59
II.A/1	Determine which predictors differentiate AS severity under LFLG baseline conditions.	Moderate vs Severe in DSE12 (baseline)	11
II.A/2	Compare baseline LFLG recordings with normal flow data.	AS60 vs DSE12 (baseline), Moderate (7) and Severe (4)	11
II.A/3	Compare stress LFLG recordings with normal flow data.	AS60 vs DSE12 (stress), Moderate (4) and Severe (3)	7
II.B/1	Assess overall change in predictors due to stress testing.	Baseline vs Stress within DSE12 (paired)	8
II.B/2	Evaluate whether the stress response differs by AS severity.	Δ(Baseline – Stress) across Moderate and Severe groups	7

Table 5.1: Overview of the hypothesis tests performed in this chapter. Each test is defined by its objective, the comparison groups involved, and the corresponding sample sizes used for statistical analysis. AS60 refers to the concordant aortic stenosis (AS) data set, while DSE12 denotes the low-flow, low-gradient (LFLG) AS cohort, who underwent dobutamine stress echocardiography with both resting and stress recordings.

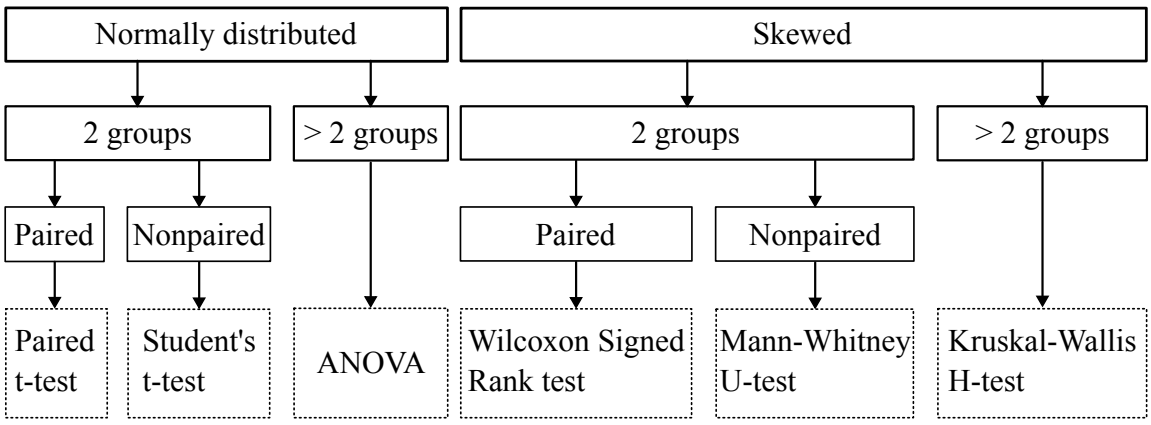


Figure 5.1: Flow diagram for choosing the appropriate statistical test. The left column presents the parametric tests, whereas the right column lists the non-parametric tests for continuous dependent variables and categorical independent variables. (Figure adapted from An (2019)).

In cases of small sample sizes ($n < 10$), the Shapiro-Wilk test was considered to have limited statistical power, therefore non-parametric analysis was preferred even when p-values exceeded 0.05. When statistical tests indicated significant differences among more than two groups, post hoc tests were performed to examine pairwise differences between the groups. Tukey's HSD was selected as the parametric post-hoc test, while Dunn's test served as the non-parametric counterpart. Since performing multiple hypotheses tests increases the chance of encountering unlikely outcomes even when the null hypothesis is true, thereby raising the risk of incorrectly rejecting the null hypothesis, a correction was applied when running post hoc tests to reduce the false-positive error rate. After performing the appropriate test for each hypotheses with the corresponding post hoc tests and corrections, a p-value less than 0.05 resulted in the rejection of the null hypothesis.

5.2.2 Multivariate Analysis

After conducting the initial univariate analysis utilising the clinical predictors, the study expanded to incorporating multivariate data analysis techniques.

Principal Component Analysis

To uncover the dominant combinations of features that describe as much of the data as possible (i.e. maximise the variance), an unsupervised learning technique, called principal component analysis (PCA), was used. The initial step involved representing the clinical data using a design matrix $\mathbf{X} \in \mathbb{R}^{p \times f}$, where p denotes the total number of patients and f indicates the number of features used in the AS60 data set. The features were then standardised to have 0 mean and unit variance

$$\hat{X}_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j}, \tag{5.1}$$

by subtracting the mean of each column, μ_j, from the data points and dividing by the standard deviation of the corresponding feature, σ_j. The covariance matrix was then calculated from the standardised data matrix as

$$\mathbf{C} = \frac{1}{p-1}\hat{\mathbf{X}}^T\hat{\mathbf{X}}, \tag{5.2}$$

capturing the relationship between each feature. The next step in PCA involved the eigen decomposition on the covariance matrix as

$$\mathbf{CV} = \mathbf{VD}, \tag{5.3}$$

where $\mathbf{V} \in \mathbb{R}^{f \times f}$ comprises the eigenvectors, $\mathbf{v}$, in its columns, and $\mathbf{D} \in \mathbb{R}^{f \times f}$ is a diagonal matrix containing the corresponding eigenvalues, λ. The principal components are then defined by projecting the standardised data matrix to the eigenvectors as

$$\mathbf{T} = \hat{\mathbf{X}}\mathbf{V}. \tag{5.4}$$

To compute the fraction of variance, ϵ_r, captured by the first r components, the eigenvalues were used as

$$\epsilon_r = \frac{\sum_{k=1}^{r} \lambda_k}{\sum_{k=1}^{f} \lambda_k}. \tag{5.5}$$

The optimal number for r was chosen so that ϵ_r exceeded 95%. Truncating $\mathbf{V}$, by keeping only the first r columns, yielding $\mathbf{V}_r$, allowed for a lower-dimensional representation of the data matrix as

$$\mathbf{T}_r = \hat{\mathbf{X}}\mathbf{V}_r. \tag{5.6}$$

To examine the contribution of the clinical features to the principal components, the coefficients of the eigenvectors, also referred to as the loadings, were examined. Having obtained $\mathbf{T}_r$, the analysis was extended to employ the same descriptive techniques and statistical tests as previously described in the univariate analysis section, but this time using the reduced-dimensional representation of the data matrix in PCA-space, rather than the clinical features themselves.

Linear Discriminant Analysis

PCA was used to identify directions of maximum variance, thereby preserving as much information as possible while reducing the dimensionality of the data. In contrast, linear discriminant analysis (LDA) was employed to maximise the separation between known AS severity grades. By using the existing clinical labels for mild, moderate, and severe AS, LDA identifies projection directions that optimise class separation within a wider multivariate feature space. This may result in decision boundaries that translate more effectively to LFLG AS than those based on individual guideline-defined predictors alone. Therefore, LDA not only further reduces the dimensionality of the input but may also provide additional insight beyond standard guideline criteria in diagnostically challenging phenotypes.

The aim of LDA - to find an optimal projection matrix, that both maximises the separation between classes and reduces the variance within each class - is expressed as a maximisation objective via

$$\mathbf{W} = \underset{\mathbf{W}}{\operatorname{argmax}} \frac{|\mathbf{W}^T\mathbf{S}_B\mathbf{W}|}{|\mathbf{W}^T\mathbf{S}_W\mathbf{W}|} \quad . \tag{5.7}$$

The columns of $\mathbf{W} \in \mathbb{R}^{r\times k}$ are the projection directions, called discriminant vectors, $\mathbf{S}_B \in \mathbb{R}^{r\times r}$ represents the between-class scatter matrix measuring the separation between class means, and $\mathbf{S}_W \in \mathbb{R}^{r\times r}$ is the within-class scatter matrix encompassing the spread of each class around their means. To implement this supervised learning algorithm, first, the patient diagnosis labels (mild, moderate, and severe AS) were assigned to the class labels 0, 1 and 2, respectively. Next, the number of discriminant vectors was chosen through $k = min(n_c - 1, r)$, where n_c is the number of distinct classes, while the value of r was identical to that previously found in the PCA. It is important to note that to maximise the expression found on the right-hand side of equation 5.7, also referred to as the Rayleigh quotient, one can increase

the numerator, which assesses class separation, and/or decrease the denominator, which represents the within-class spread.

To calculate the scatter matrices, the classwise mean vectors, $\boldsymbol{\mu}_c$, and the global mean vector, $\boldsymbol{\mu}$ of the rank r PCA representation of the data matrix are calculated first, as

$$\boldsymbol{\mu}_c = \frac{1}{p_c} \sum_{\mathbf{x} \in \mathbf{T}_{r,c}} \mathbf{x} \quad \text{, for} \quad c \in \{0, 1, 2\}, \qquad \text{and} \qquad \boldsymbol{\mu} = \frac{1}{p} \sum_{i=1}^{p} \mathbf{x}_i. \tag{5.8}$$

Here, $\mathbf{T}_{r,c}$ represents the subset of the reduced PCA representation of the data matrix of patients with class label c, and p_c is the number of patients in class c. From the above equations, the between-class and within-class scatter matrices are formulated as follows

$$\mathbf{S}_B = \sum_{c=0}^{2} p_c (\boldsymbol{\mu}_c - \boldsymbol{\mu})^T (\boldsymbol{\mu}_c - \boldsymbol{\mu}) \qquad \text{and} \qquad \mathbf{S}_W = \sum_{c=0}^{2} \sum_{\mathbf{x} \in \mathbf{T}_{r,c}} (\mathbf{x} - \boldsymbol{\mu}_c)^T (\mathbf{x} - \boldsymbol{\mu}_c). \tag{5.9}$$

Since $\mathbf{S}_B$ and $\mathbf{S}_W$ are computed directly from the input data matrix and the corresponding class labels, they remain constant for a given data set. Therefore, maximising the Rayleigh quotient depends only on the discriminant vectors, $\mathbf{w}$, which are found through solving the eigenvalue problem

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}. \tag{5.10}$$

After finding $\mathbf{W}$, the rank r PCA decomposition was projected to the discriminant vectors through

$$\mathbf{X}_{LDA} = \mathbf{T}_r \mathbf{W}. \tag{5.11}$$

5.2.3 Classification Algorithms

To complement the univariate statistical analysis, performed to test the validity of various null hypotheses, classification algorithms were developed to learn the specific decision rules that best separate AS severity groups in the multivariate case. Several algorithms were tested on both PCA and LDA reduced feature spaces of the AS60 data set.

K-Nearest Neighbour

The simplest non-parametric model for predicting the class labels of unseen data points, $\mathbf{x}_u$,is the k-nearest neighbour algorithm. This algorithm calculates the Euclidean distances between the new sample and the input examples as

$$d_i = \|\mathbf{x}_u - \mathbf{x}_i\|_2 \tag{5.12}$$

and selects the class labels with the K lowest distances, represented by the set $Y_K = \{y_{i_1}, .., y_{i_K}\}$. The class label is then assigned to the new sample by a majority vote be-

tween the K closest examples as

$$\hat{y} = \underset{c\in\{0,1,2\}}{\operatorname{argmax}} \sum_{y\in Y_K} \delta_{yc} \quad \text{, where} \quad \delta_{yc} = \begin{cases} 1 & \text{if } y = c \\ 0 & \text{if } y \neq c \end{cases} \tag{5.13}$$

Selecting the hyperparameter K, has a great influence on how the decision boundaries are assigned, generally adhering to the rule that low values (e.g. $K \leq 3$) tend to overfit the input data, while setting the number of neighbours too high ($K \geq 10$) can result in poor resemblance of local structures.

Logistic Regression

Multinomial logistic regression is a linear model that estimates the probability that an observation belongs to each of the possible classes. This is achieved by first computing the class scores, $\mathbf{z} \in \mathbb{R}^{n_c}$, for an input row vector $\mathbf{x} \in \mathbb{R}^f$ by

$$\mathbf{z} = \mathbf{x}\mathbf{W} + \mathbf{b}, \tag{5.14}$$

where $\mathbf{W} \in \mathbb{R}^{f\times n_c}$ and $\mathbf{b} \in \mathbb{R}^{n_c}$ represent the weight and bias parameters of the model, while f and n_c are the number of input features and the number of classes, respectively. These scores are then normalised through the softmax function, $\sigma : \mathbb{R}^{n_c} \to (0,1)^{n_c}$, to represent the probability scores for each class by

$$P(y = c|\mathbf{x}) = \sigma(z_c) \quad \text{with} \quad \sigma(z_c) = \frac{e^{z_c}}{\sum_{j=0}^{2} e^{z_j}}. \tag{5.15}$$

Class labels are assigned to the new sample by taking the class with the highest probability, such that

$$\hat{y} = \underset{c\in\{0,1,2\}}{\operatorname{argmax}}\, \sigma(z_c). \tag{5.16}$$

The parameters of the model are learnt via gradient-based optimisation, which is detailed in Chapter 7.

Support Vector Machine

While logistic regression employs a probabilistic approach aimed at maximising the likelihood of the correct class, support vector machines (SVM) are utilised to identify the optimal hyperplane (a linear decision boundary in a high-dimensional space) that separates data points into their corresponding categories. In cases where the input data are linearly separable, this is achieved by maximising the margin, which is defined as the distance between the nearest points of each class and the hyperplane.

For a binary classification, where the input labels are $y_i \in \{-1, 1\}$, the SVM computes the class score of an input row vector $\mathbf{x}_i \in \mathbb{R}^f$ by

$$z_i = \mathbf{x}_i\mathbf{w}^T + b, \tag{5.17}$$

where the parameters of the model $\mathbf{w} \in \mathbb{R}^f$ and b, represent the normal vector to the

hyperplane and the intercept term, respectively. Since any point lying on the hyperplane is assigned to 0 by definition, as shown in Figure 5.2, the SVM classifies the inputs as belonging to the positive class if the score exceeds 0, and predicts it belonging to the negative class if the score is negative as

$$\hat{y}_i = \text{sgn}(z_i), \tag{5.18}$$

where sgn() refers to the sign function. It is important to note that in the linearly separable case there exist two other hyperplanes, H_1 and H_2 parallel to H_0, that not only ensure that the predicted score has the correct sign but also that the absolute value of the class scores are greater than 1, i.e. $|z_i| \geq 1$.

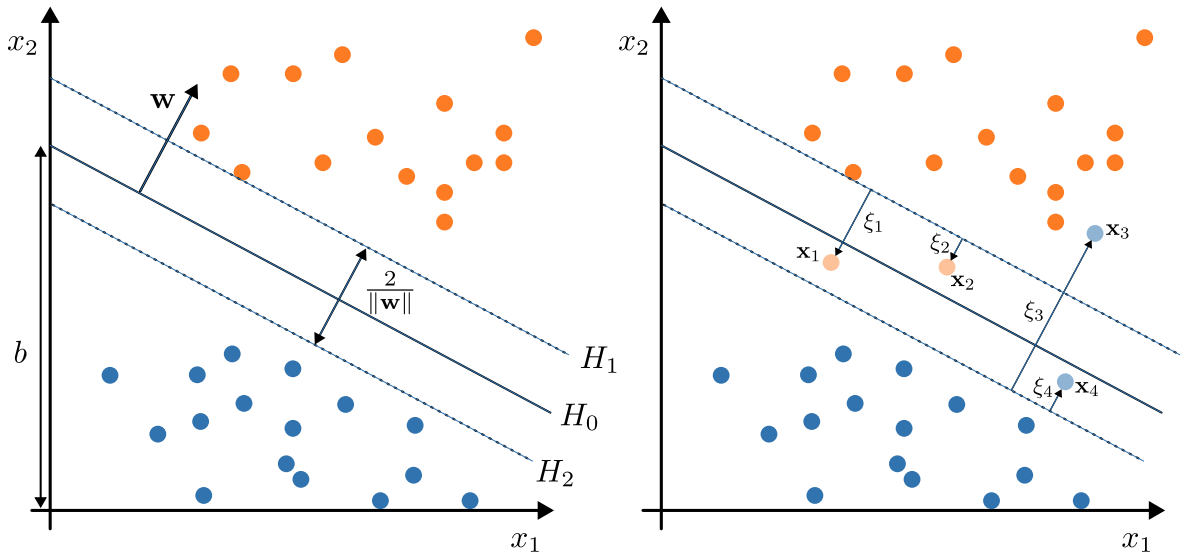


Figure 5.2: The decision boundary of a support vector machine classifier as depicted in the linearly separable case [left] and as shown in the linearly non-separable case [right]. The orange and blue points belong to the positive and negative classes respectively. The parameters of the model define a hyperplane (a line in two dimensions), H_0, such that $H_0 = \{\mathbf{x} \in \mathbb{R}^f : \mathbf{x}\mathbf{w}^T + b = 0\}$. (Figure adapted from García-Gonzalo et al. (2016)).

To find a linear decision boundary with the largest separation between classes, the margin, measured by $\frac{2}{\|\mathbf{w}\|}$, is maximised. Since this objective is equivalent to minimising the inverse of the margin, SVM finds the optimal hyperplane that achieves correct classification for all inputs via

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\| \quad \text{subject to} \quad y_i z_i \geq 1 \quad \forall i \in \{1, .., n\}, \tag{5.19}$$

where n represents the number of input examples. This classifier is called the hard-margin SVM.

For the linearly non-separable case, SVM can be altered in two ways:

- Modifying the optimisation objective and relaxing its constraint to allow some misclassification. This algorithm still produces linear decision boundaries, but is preferred when the data is almost linearly separable with some amount of overlap between classes.
- Projecting the input data to a higher-dimensional feature space where the data is linearly

separable. This algorithm is suited for finding nonlinear decision boundaries.

The former is known as soft-margin SVM, and its optimisation objective is expressed by introducing a new variable into equation 5.19 as

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \frac{1}{2}\|\mathbf{w}\| + C\sum_{i=1}^{n}\xi_i \quad \text{subject to} \quad y_i z_i \geq 1 - \xi_i \quad \forall i \in \{1,..,n\}, \tag{5.20}$$

where ξ_i measures the extent of misclassification by calculating the distance from the hard margins through

$$\xi_i = \max(0, 1 - y_i z_i). \tag{5.21}$$

Misclassified points are characterised by $\xi > 1$, while points that are correctly classified and lie within the margin satisfy $0 \leq \xi \leq 1$. For correct predictions that are beyond the margin, $\xi = 0$. Due to the significant impact of large ξ values on the optimisation objective, this method makes the SVM susceptible to outliers; however, modifying the hyperparameter C allows for the adjustment of this effect.

The latter is called the kernel SVM, referring to the kernel function used to obtain the high-dimensional feature space in which a linear decision boundary may exist. Common kernel functions include the polynomial and Gaussian kernels.

To perform multiclass classification with SVM algorithms, n_c number of independent binary classifiers are developed $f_c(\mathbf{x}) = \mathbf{x}\mathbf{w}_c^T + b_c$ on the one-vs-rest grouping of class labels and aggregating their prediction by taking the maximum class scores as

$$\hat{y} = \underset{c\in\{0,1,2\}}{\operatorname{argmax}} f_c(\mathbf{x}). \tag{5.22}$$

5.2.4 Model Selection

After defining the models, the next step was to evaluate their performance using a set of input examples paired with their corresponding diagnosis labels $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, to identify the most appropriate algorithm for the classification task. Assessing several models with various hyperparameter configurations across different feature spaces - a process known as model selection- facilitates the development of models capable of generalising to new, unseen data.

Performance Metrics

The model selection process begins by establishing the appropriate performance metric through which the models are evaluated. In the context of a classification task, the typically selected metrics are as follows:

- Accuracy is simply defined as the ratio of correct predictions to the number of examples

as

$$\mathrm{A} = \frac{1}{n}\sum_{i=1}^{n}\delta_{y_i\hat{y}_i}, \tag{5.23}$$

where y_i is the ground truth label, for example i, $\hat{y}_i$ is the label predicted by the model, and δ is the Krockner delta function introduced in equation 5.13. This metric is well-suited for data sets with equal proportions of class labels.

- Recall measures the ratio of correctly predicted class c examples (also known as true positives) to the total number of class c examples, such that

$$\mathrm{R}_c = \frac{\sum_{i=1}^{n}\delta_{y_i c}\delta_{\hat{y}_i c}}{p_c}, \tag{5.24}$$

where p_c represent the number of samples belonging to class c. This metric is also referred to as the sensitivity or true positive rate (TPR), reflecting the model's ability to detect positive classes correctly.

- The precision is calculated as the ratio of true positives to the total number of class c predictions, so that

$$\mathrm{P}_c = \frac{\sum_{i=1}^{n}\delta_{y_i c}\delta_{\hat{y}_i c}}{\sum_{i=1}^{n}\delta_{\hat{y}_i c}} \tag{5.25}$$

measuring how well the model is able to avoid false positives.

- Balanced accuracy is often preferred over standard accuracy in cases where class labels are unevenly distributed within a data set (i.e., the data set is imbalanced) to avoid the majority class from disproportionately influencing the performance. It is calculated by taking the average of the class recall values as

$$\mathrm{B} = \frac{1}{n_c}\sum_{c=1}^{n_c}\mathrm{R}_c, \tag{5.26}$$

where n_c denotes the number of classes.

- The F1 score is used to combine the precision and recall metric to measure both performances simultaneously as

$$F_1 = \sum_{c=1}^{n_c}\frac{2R_cP_c}{R_c + P_c}. \tag{5.27}$$

Cross-Validation

After choosing the performance metrics, the various models were evaluated against each other using the AS60 data set. To estimate how well the models would perform on unseen data, cross-validation was used. This technique splits all available data into different partitions, each serving specific functions. The training set is used to determine the parameters of the classification algorithm, while the validation set is used to evaluate the generalisation capability of the trained model. It should be noted that often there is separate held out test set to evalute the final performance of the model. However, this approach was not feasible in this study due to the size of the AS60 cohort, as further partitioning would have substantially reduced the amount of data available for training and validation. Therefore, reporting on the

final test performance was omitted from the analysis and a more qualitative method was used, in which the learnt decision boundaries were examined with respect to the feature distribution of the DSE12 patients and synthetically generated patients classified according to clinical guidelines. This enabled an exploratory analysis of the utility of the models in clinically relevant settings, rather than providing an estimate of predictive performance.

Given the constraints of the available clinical data, stratified k-fold cross-validation was considered the optimal method for the validation strategy. In this approach, the AS60 data set was randomly partitioned into k approximately equal size folds using stratified sampling, thus preserving the overall AS severity class distribution within each fold. Subsequently, the learning algorithm underwent training on the combined $k - 1$ folds and was evaluated using the held-out partition. This process was repeated k times, using each fold precisely once as the validation set. The model performance is then reported as the average validation score across the folds given the chosen metric.

Although the model parameters are acquired from the training data, certain parameters are set manually prior to the training process. These can include the number of neighbours for the KNN algorithm, the regularisation strength used in the loss function of the logistic regression, or the type of kernel function used in the SVM. These predetermined parameters are referred to as hyperparameters, and they significantly impact the decision boundaries the model learns, although they remain fixed once training begins. To address the challenge of manually examining all hyperparameter combinations, grid search is employed. This method conducts a systematic search through a predefined parameter grid containing all potential values for each hyperparameter, and it keeps track of the optimal hyperparameter settings.

The hyperparameter grid was defined by the number of PCA components, the number of LDA components, the regularisation term for objective functions used by the logistic regression and SVM classifiers, the type of kernel used by the SVM classifier, as well as the number of neighbours used for the KNN algorithm. The algorithm performing stratified k-fold cross-validation with grid search hyperparameter tuning is shown in Algorithm 1.

5.2.5 Applicability to Low-Flow, Low-Gradient Aortic Stenosis and Synthetic Data

The applicability of each tuned classification algorithm to the LFLG state was then assessed using the DSE12 data set. This analysis examined whether the decision rules learnt from concordant AS remained clinically meaningful when applied to patients with LFLG hemodynamic conditions. In addition, a synthetically generated data set was used to examine how well clinical guidelines adhered to the learnt decision boundaries.

To investigate how the model applies to patients with LFLG AS, the DSE12 data set was transformed into the low-dimensional feature space used by the classification algorithm.

Algorithm 1 The stratified k-fold cross-validation algorithm with grid search for hyperparameter tuning of each classifier.

Require: data set (X, y) with N samples, Parameter grid $\mathcal{G}$, Number of folds K
Ensure: Best parameters θ^*, Best validation score s^*

Initialize $s^* \leftarrow 0$, $\theta^* \leftarrow \emptyset$
for each parameter combination $\theta \in \mathcal{G}$ **do**
    Initialize list of validation scores $\mathcal{S} \leftarrow [\,]$
    Generate K stratified folds $\{(X_k^{\text{train}}, y_k^{\text{train}}), (X_k^{\text{val}}, y_k^{\text{val}})\}_{k=1}^{K}$
    for each fold $k = 1$ to K **do**
        Fit scaler on X_k^{train} and transform both X_k^{train}, X_k^{val}
        Fit PCA with n_{PCA} on X_k^{train} and transform both X_k^{train}, X_k^{val}
        Fit LDA with n_{LDA} on X_k^{train} and transform both sets
        Fit classifier with hyperparameters from θ on transformed X_k^{train}
        Predict on transformed X_k^{val} and compute score s_k
        Append s_k to $\mathcal{S}$
    end for
    Compute mean score $\bar{s} = \frac{1}{K} \sum_{k=1}^{K} s_k$
    if $\bar{s} > s^*$ **then**
        $s^* \leftarrow \bar{s}$, $\theta^* \leftarrow \theta$
    end if
end for
return θ^*, s^*

Subsequently, the model's decision surface was visualised, as it was learnt under normal flow conditions, and the baseline and stress recordings of DSE12 were overlaid. This allowed for a visual assessment of the stress response, addressing multiple questions:

- Did stress testing alter the predicted severity?
- Did stress testing align patients more closely with their true class labels?
- Did stress testing cause patients with different grades of AS to become more similar or more distinct?

A similar approach was considered to assess the concordance of the learnt decision functions with clinical guidelines. The decision boundaries of each model were visualised in the reduced feature spaces and the overlap between a synthetically generated set of patients classified by the clinical guidelines was assessed. In order to create synthetic data, the feature distribution observed in the AS60 data set was employed. Assuming that the recorded features followed a normal distribution within the extended population, new patient instances, denoted as $\tilde{\mathbf{x}}$, were sampled from a multivariate Gaussian distribution by

$$\tilde{\mathbf{x}}_j \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \quad \text{for} \quad j = 1, .., m \quad , \tag{5.28}$$

where $\boldsymbol{\mu}$ and Σ are the mean feature vector and covariance matrix of the AS60 data set, respectively, while m denotes the number of synthetic patients. The generated data was then filtered to exclude physiologically impossible patients, and diagnosis labels were assigned following clinical guidelines.

5.3 Results and Discussion

This section provides the results obtained from applying the methodology described previously for performing a statistical analysis on the clinical predictors observed in the AS60 and DSE12 data sets. First, univariate analysis and hypotheses testing are covered, followed by the evaluation of multivariate techniques. Finally, the classification algorithms training procedure and generalisation performance is detailed.

5.3.1 Univariate Analysis

To investigate the primary trends present in the AS60 data set, a box and whisker plot was used, illustrating the distribution of clinical predictors in each severity class as shown in Figure 5.3. The DSE12 measurements from the baseline and stress states are overlayed on the corresponding features to further aid in the visual comparison of the two distinct flow conditions.

From the plot it can be deduced that a clear separation of classes occurs in patients with normal flow conditions for predictors included in the clinical recommendations (AV maximum velocity, AV mean pressure gradient, AVA), which confirms that the cutoff values can distinguish the severity of AS. It can also be inferred that for the same predictors, DSE12 patients recorded at resting heart rate exhibit little to no separation in terms of the severity of AS. However, stress testing appeared to slightly improve the distinction between classes for AV Vmax. It is apparent that attributes such as AV mean velocity and AV maximum pressure gradient, which are closely aligned with those employed in the guidelines, show a similar trend, both in normal flow states and under LFLG conditions.

Although hemodynamic measurements obtained at the level of the LVOT and features related to the LV volume (such as EDV, ESV, and EF) do not seem to assist in class differentiation, AV envelope time and AV VTI show a promising stress response by diverging patients with different class labels. A possible explanation might be that DSE led to a more significant increase in heart rate in patients with moderate AS, thus reducing envelope time by a greater amount.

It is noted that the greatest discrepancy in the DSE12 data set, when compared with the AS60 patient data, is related to predictors associated with LV volumetric measurements. This is likely caused by the measurement software's dependency on estimating these predictors from two-dimensional measurements, such as those derived from the Simpson biplane or the Teicholz method.

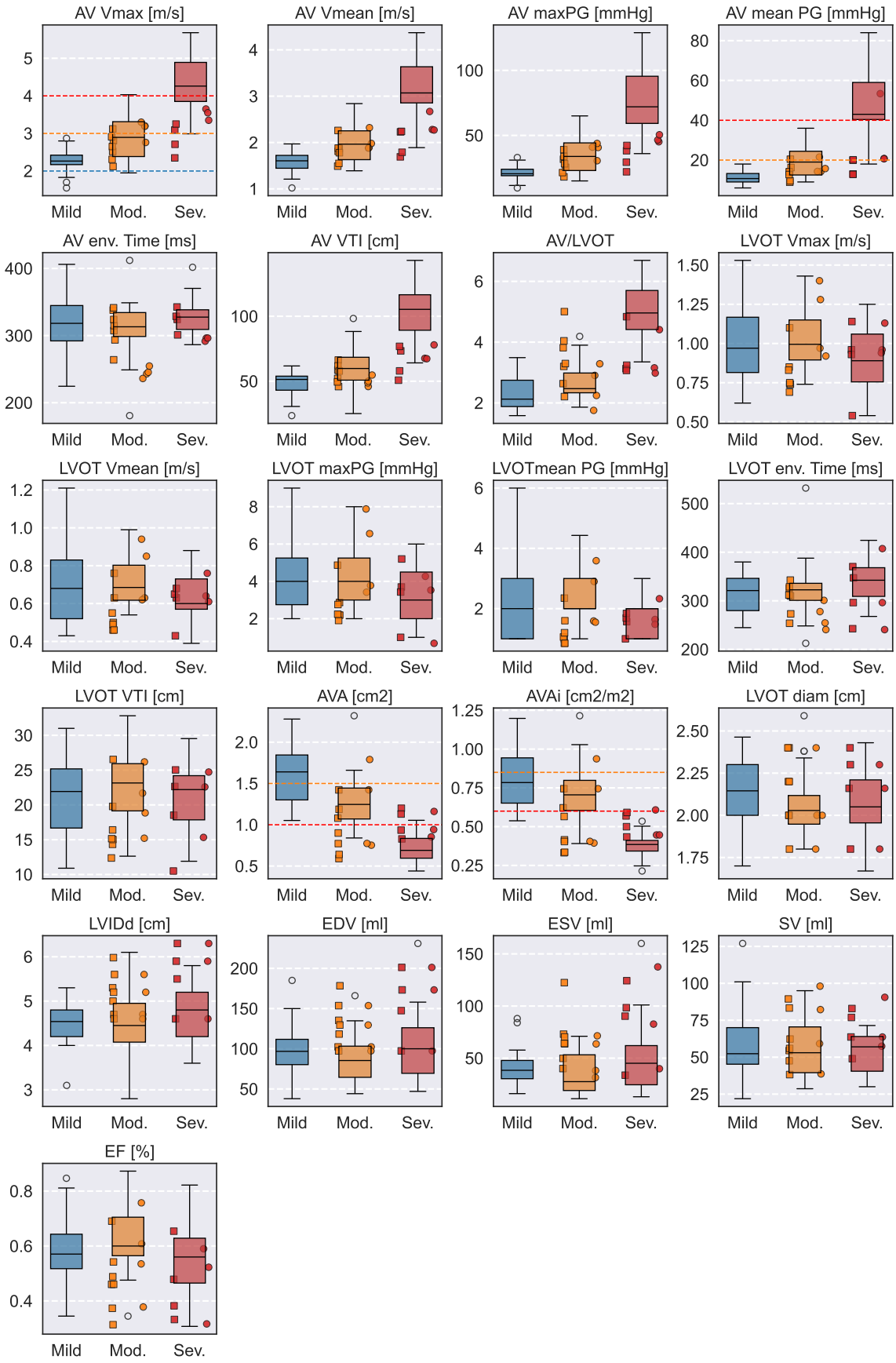


Figure 5.3: The distribution of clinical predictors by aortic stenosis (AS) severity in the concordant AS cohort, as depicted via a box and whisker plot. The corresponding features from the low-flow, low-gradient AS cohort is shown on top. Square markers represent the baseline, while the circular markers the stress states. The clinical cutoff values are marked by a dashed line for maximum velocity through the aortic valve (AV Vmax), the mean transvalvular gradient (AV mean PG), aortic valve area (AVA), and indexed AVA (AVA_i).

Following the preliminary qualitative assessment of the main trends within the AS60 and DSE12 data sets, a series of statistical tests were performed to quantitatively evaluate the different research questions. In Table 5.2, the p-values for each hypothesis tested are compiled, with rows indicating the predictor under examination and columns denoting the specific test conducted. P-values less than 0.05 are highlighted, indicating statistically significant differences.

Predictor / Test	I.	II.A/1	II.A/2	II.A/3	II.B/1	II.B/2
AV Vmax	**1.87e-08**	0.660	0.472/**0.019**	0.813/0.201	**0.041**	0.857
AV Vmean	**1.87e-08**	0.611	0.472/**0.019**	0.816/0.145	0.066	0.629
AV maxPG	**1.87e-08**	0.660	0.472/**0.028**	0.813/0.201	**0.041**	0.857
AV mean PG	**1.87e-08**	0.666	0.185/**0.021**	0.816/0.344	**0.041**	0.400
AV env. Time	5.67e-01	0.193	0.956/0.935	0.158/0.145	**0.041**	0.400
AV VTI	**1.87e-08**	0.269	0.472/**0.019**	0.401/0.145	1.000	0.400
AV/LVOT	**2.18e-08**	0.881	0.110/**0.048**	0.813/0.145	0.342	0.229
LVOT Vmax	3.15e-01	0.475	**0.032**/0.935	0.813/0.508	0.123	0.057
LVOT Vmean	3.15e-01	0.709	**0.032**/0.935	0.816/0.573	0.070	0.074
LVOT maxPG	3.35e-01	0.475	**0.047**/0.935	0.813/0.899	0.505	0.057
LVOT mean PG	2.81e-01	0.284	**0.032**/0.242	0.813/1.000	0.070	0.114
LVOT env. Time	2.81e-01	0.825	0.551/0.663	0.347/0.573	0.209	1.000
LVOT VTI	6.93e-01	0.611	0.055/0.810	0.813/0.975	0.917	0.400
AVA	**1.87e-08**	0.674	**0.049**/0.054	0.813/0.186	0.209	0.229
AVAi	**2.33e-08**	0.697	**0.034**/**0.037**	0.813/0.195	0.209	0.229
LVOT diam	4.44e-01	0.590	0.472/0.663	0.813/0.975	1.000	1.000
LVIDd	4.44e-01	0.607	**0.049**/0.131	0.491/0.289	1.000	1.000
EDV	4.30e-01	0.607	**0.032**/0.170	0.401/0.289	1.000	1.000
ESV	3.15e-01	0.476	**0.032**/0.170	0.401/0.334	0.605	0.229
SV	8.73e-01	0.794	0.503/0.170	0.689/0.289	0.605	0.229
EF	2.81e-01	0.878	**0.032**/0.549	0.813/0.573	0.574	0.229

Table 5.2: Statistical tests on the clinical predictors using the concordant aortic stenosis (AS) cohort and the low-flow, low-gradient AS data set. The table columns display the p-values for each corresponding test outlined in Table 5.1, with statistically significant values highlighted.

Column **I.** -comparing AS classes in normal flow states - confirms that statistically significant differences exist for all predictors used in the clinical guidelines. In addition, it also states that closely related measurements, such as AV mean velocity, AV maximum pressure gradient, AV VTI, have a similar predictive value in differentiating AS severity. However, for column **II.A/1**, when evaluating the same hypotheses under LFLG hemodynamic conditions, neither predictor demonstrated the ability to discriminate between classes.

In columns **II.A/2** and **II.A/3**, the differences in distributions between the two flow conditions (i.e., comparing AS60 to DSE12 recordings) were evaluated for each class, under baseline and stress states, respectively. The left portion of each column represents the moderate class, while the right portion corresponds to the severe class, as the DSE12 data set lacked patients with mild AS.

From column **II.A/2** it can be seen that the hemodynamic parameters measured at the level of the AV exhibit a clear difference between the AS classes. This suggests that patients with severe LFLG AS conform better to the feature distribution present in moderate patients under normal flow conditions, since no differences were previously found between AS classes at baseline in DSE12 according to the test results presented in column **II.A/1**. This confirms the diagnostic challenge of using the recommended predictors in grading LFLG AS. In contrast, the opposite phenomenon is observed when assessing hemodynamic measurements at the level of the LVOT and predictors related to LV function (EDV, ESV, EF), indicating that patients with moderate LFLG AS more closely resemble those in the severe category under normal flow conditions. Examining the results of the test **II.A/3** suggests that DSE caused patients to be closer to their respective groups under normal flow conditions, however, the sample size here was very limited and the statistical power remained low.

Test **II.B/1** assessed the impact of DSE on altering predictors without considering class labels, thereby enabling the identification of features that most significantly contributed to the stress state representation of LFLG AS. The results showed that the flow measurements obtained at the AV were the most affected, which confirmed the presence of increased flow states, while a significant impact could not be observed for any other features.

To examine the class-specific response to stress testing, differences in predictors from baseline to stress were analysed across various severity categories. The p-values presented in column **II.B/2** illustrate these findings, showing no significant differences for either predictor. It is important to note that the statistical power of this hypothesis testing is limited due to the small sample size.

5.3.2 Multivariate Analysis

After examining the results of the univariate analysis and hypothesis testing on clinical predictors, the study expanded to multivariate techniques. PCA was performed to capture the combination of features that explain most of the variance within the A60 data set, and the DSE12 recordings were transformed by the learnt transformation matrix. Figure 5.4 illustrates the distribution of principal components (PC) by class for AS60 measurements, with an overlay of the PCA representation of the DSE12 patient data for both baseline and stress states. The fractional explained variance suggests that $\sim 95\%$ of the data distribution can be explained by only 6 components, successfully achieving a reduction in dimensions. Furthermore, PC1 demonstrates a noticeable separation of classes within AS60, and PC3 also appears to possess class-discriminative capabilities.

By examining the loadings of each component, corresponding to the coefficients of $\mathbf{V}_r$, as illustrated in Figure 5.5, the contribution of each clinical predictor can be inferred. Since PCA is an unsupervised dimensionality-reduction method, the PCs themselves do not have inherent

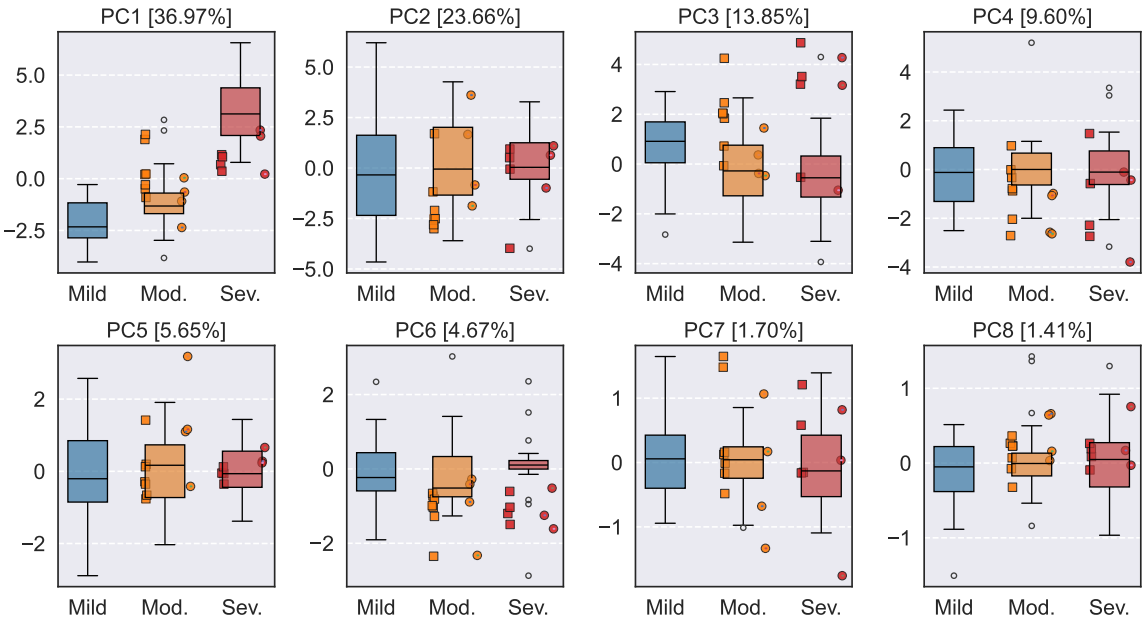


Figure 5.4: The distribution of the first eight principal components by aortic stenosis (AS) severity in the concordant AS cohort, as illustrated on a box and whisker plot. The corresponding principal components from the low-flow, low-gradient AS cohort is overlaid on top. Square markers on the left side of each box show the baseline, while the circular markers on the right the stress states. The percentage values represent the explained variance ratio by the given component.

physiological meaning; however, they can be examined post hoc based on the predictors with the largest absolute weights. The assignment of loading-based summary labels to PCs may further aid clinical interpretability and help identify the most relevant markers. Using this approach, the first three PCs showed functional separation in their dominant contributing loadings. PC1 was primarily influenced by flow-related measurements recorded at the AV, PC2 by flow-related measurements obtained at the LVOT, and PC3 by predictors related to LV size and function, including EDV, ESV, SV and LVIDd. It can also be seen that EF, despite being a clinically used marker of LV function, did not contribute substantially to the first three PCs.

Subsequently, Table 5.3 displays the results produced by applying the same statistical tests used in the univariate analysis to the first 8 PCs. This examined whether any of the dominant combinations of features observed under normal flow conditions possesses any class-discriminative ability that applies in the LFLG state.

Test **I.** indicates that PC1 has statistically significant differences among the AS severity groups within the AS60 population. This suggests that the most significant indicators of AS arise from the flow-state measurements taken at the AV. PC3, which evaluates the LV parameters, has the second-smallest p-value under normal flow conditions. A similar response is observed for test **II.A/1** relative to the univariate analysis of clinical predictors, as neither PC can distinguish between AS severity groups within LFLG hemodynamic conditions.

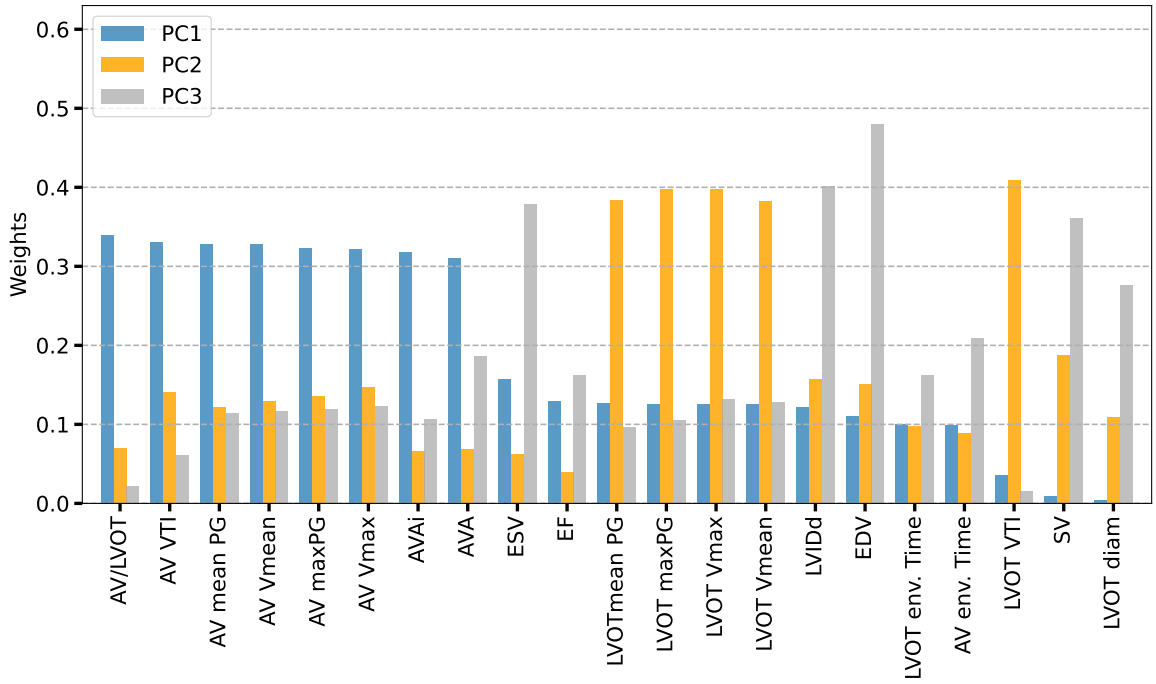


Figure 5.5: The absolute loadings of the clinical predictors for the first three principle components representing the contribution of each feature to the given component.

Predictor / Test	I.	II.A/1	II.A/2	II.A/3	II.B/1	II.B/2
PC1 [36.97%]	**5.07e-08**	0.476	**0.020**/**0.022**	0.763/0.431	0.641	0.136
PC2 [23.66%]	8.59e-01	0.572	0.096/0.897	0.852/1.000	0.104	0.136
PC3 [13.85%]	2.11e-01	0.555	**0.020**/0.073	0.836/0.603	**0.031**	0.315
PC4 [9.60%]	8.59e-01	0.831	0.186/0.501	**0.027**/0.652	**0.042**	0.639
PC5 [5.65%]	8.59e-01	0.281	0.808/0.907	0.360/0.652	**0.031**	0.315
PC6 [4.67%]	8.59e-01	0.425	**0.020**/**0.024**	0.618/0.166	0.104	0.462
PC7 [1.70%]	8.59e-01	0.769	0.651/0.585	0.842/1.000	0.286	1.000
PC8 [1.41%]	8.59e-01	0.411	0.493/0.890	0.360/0.698	0.104	0.827

Table 5.3: Statistical tests on the principal component representation of the clinical data using the concordant aortic stenosis (AS) cohort and the low-flow, low-gradient AS data set. The table columns display the p-values for each corresponding test outlined in Table 5.1, with statistically significant values highlighted.

For test **II.A/2** that evaluates the similarity between the AS60 and DSE12 patients at rest for each class, significant differences were found for PC1 and PC6 for both classes. This could imply that an intermediate representation is found for LFLG AS that neither resembles the moderate nor severe groups under normal flow states, but does not differentiate between severity according to the p-values from **II.A/1**

Test **II.A/3** demonstrated that the stress DSE12 recordings aligned with those of the AS60 patients, with an exception for PC4 in the moderate group. However, since the sample size for this group decreased substantially from 7 to 4, this outcome was considered to have low statistical power. No statistically significant differences were found in test **II.B/2** that examined the class-specific differences in stress response, while the test **II.B/1** suggests that PCs 3-5 undergo significant changes during DSE.

5.3.3 Classifier Performance

After performing the statistical test on individual PCs, various classifiers were developed using a multidimensional feature space of AS60. Initially, hyperparameter tuning for a logistic regression model, a KNN classifier, a kernel SVM, and a Gaussian process classifier was conducted using stratified k-fold cross-validation with 10 folds, in conjunction with a hyperparameter grid search. After determining the optimal hyperparameters, the algorithms were implemented on the AS60 data set, and training accuracies were documented.

The initial feature space was chosen to correspond to the top 2 PCs that exhibit the most significant class separation in the AS60 data set, as determined by the p values listed in column **I.** of Table 5.3. Figure 5.6 illustrates the decision boundaries learnt by each classifier from the AS60 data set, utilising the reduced representation of these two principal components. To evaluate the conformity of clinical guidelines to the learnt boundaries, the synthetic data are superimposed in the figure's right columns. The DSE12 patients are displayed in the left columns, with arrows indicating stress responses to evaluate how applicable the decision functions were in the LFLG condition. As can be seen from the right columns of the plot, logistic regression and Gaussian process classifier align the most closely with the clinical guidelines. This is similarly indicated by the prediction accuracy of the synthetic data using two principal components (PCs). KNN tends to overfit on AS60, whereas kernel SVM leads to compacted clusters. Using 8 PCs improves the training accuracies for all parametric models, while slightly reducing their generalisation performance to synthetic data. Most of the DSE12 baseline recordings get misclassified, however, the stress states result in a somewhat improved classification performance.

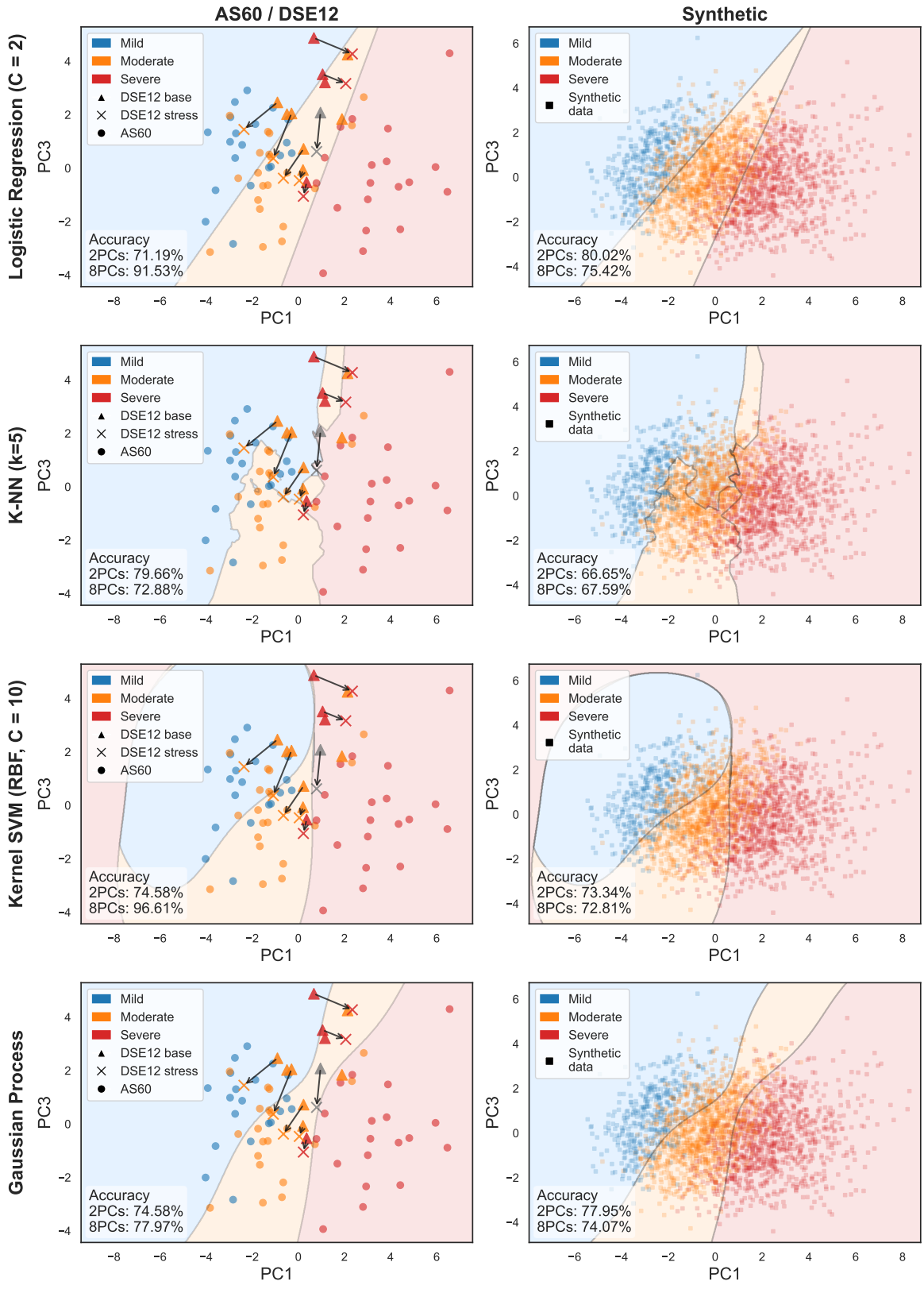


Figure 5.6: Decision boundaries of different classification algorithms as shown on the subset of the principal component (PC) space where the separation of severity grades is most prominent. Patients with concordant aortic stenosis (AS) and low-flow, low-gradient (LFLG) AS are shown in the left column, while the synthetically generated patients are displayed in the right column. AS60 represents the concordant AS data set, while DSE12 refers to the LFLG AS cohort. Arrows indicate the intrapatient stress response from the baseline state. Accuracies are displayed for the AS60 and synthetic data sets using two and eight PCs.

After training the classifiers on the PCA representations, LDA was performed to further reduce the dimension of the data while promoting class separation. A similar approach was taken when tuning the hyperparameters for each classifier in the LDA reduced feature space of AS60, i.e. grid search with stratified 10-fold cross-validation was performed.

Figure 5.7 presents the learnt decision boundaries from AS60 using two discrimant vectors. In the right column of the figure, the synthetically generated patients are superimposed to assist with visualising the clinical guidelines, while the left column presents the DSE12 patients with arrows representing the stress response. It is observed again that all parameteric models result in higher training accuracies for two components as opposed to one, but at the cost of generalisation performance to the synthetic data. Although the generalisation performance of the kernel SVM and Gaussian process classifier suggests moderate concordance to clinical guidelines using two discriminant vectors, the triple-point phenomena are not easily interpreted in a clinical setting. Nevertheless, the kernel SVM appears to benefit from the tightly clustered moderate group by correctly predicting almost all stress responses in the DSE12 cohort (6 out of 7). It is important to note that employing solely the first discriminant vector for classification leads to a decision boundary that is more clinically reasonable, as it eliminates the presence of triple points. However, in this case, LFLG patients cannot be distinguished at baseline and only demonstrate a mild divergence during DSE.

5.4 Conclusion

This chapter involved a comprehensive statistical analysis of standard and novel echocardiographic predictors in relation to both the severity of AS and hemodynamic states. This analysis confirmed that predictors under resting conditions do not differentiate true-severe from pseudo-severe AS. It also highlighted that, while the resting echocardiographic predictors in the LFLG population differ from those seen in concordant AS, DSE shifts the distribution of these predictors to be more aligned with those in concordant AS, thus demonstrating the clinical relevance of DSE in resolving diagnostic uncertainty. In contrast, no class-specific stress response has been observed in this data set; however, this may be due to the limited statistical power.

To complement this analysis, ML algorithms were developed to classify the severity of AS using a reduced feature space, derived via PCA and LDA, applied to both routine and novel predictors. This study represents the first attempt to characterise the stress response in LFLG AS within such a low-dimensional space and to evaluate how the decision boundaries learnt from a concordant AS cohort generalise to this more complex phenotype.

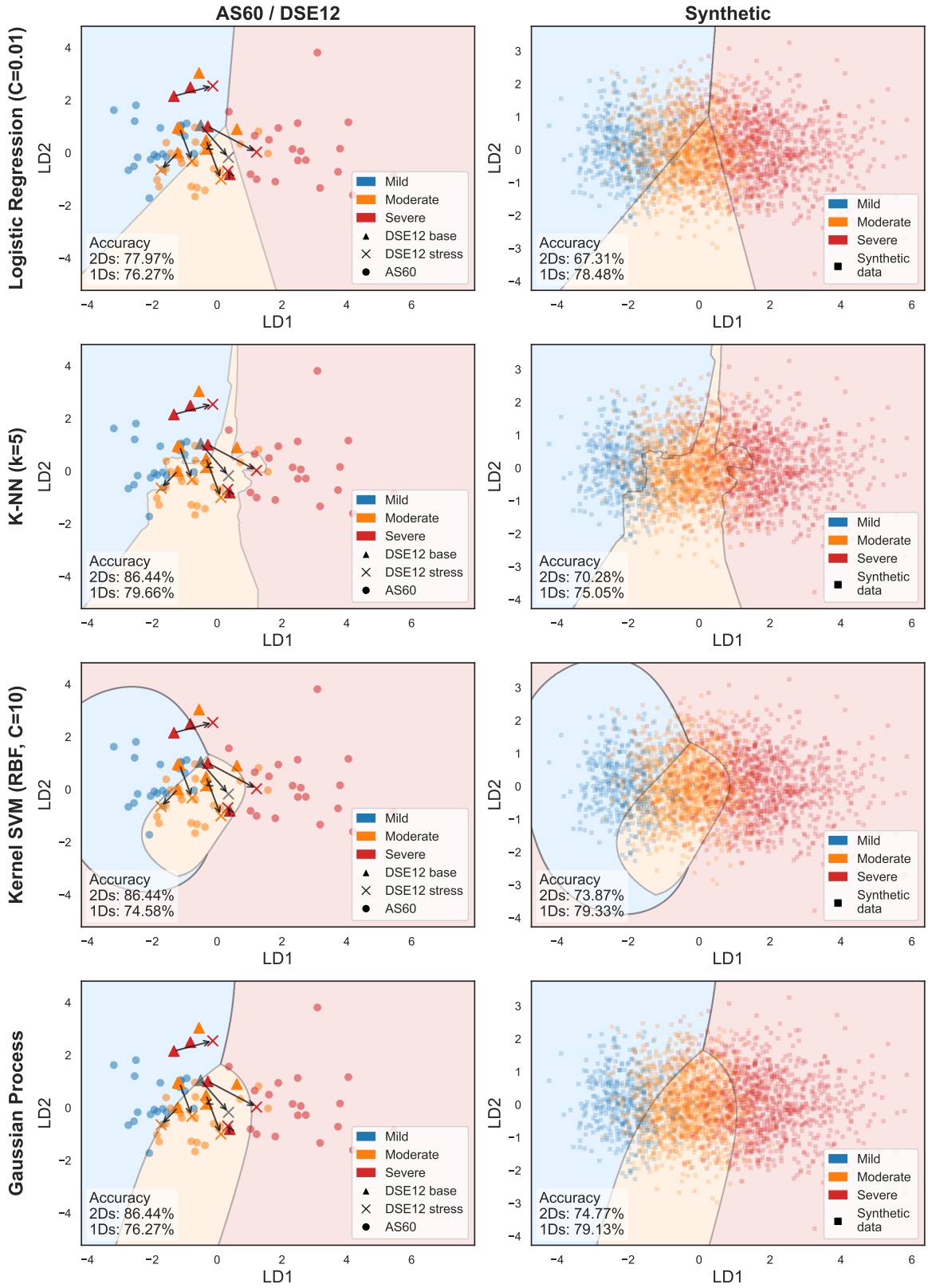


Figure 5.7: Decision boundaries of different classification algorithms as plotted on the feature space covered by the two discriminant vectors from linear discriminant analysis. Patients with concordant aortic stenosis (AS) and low-flow, low-gradient (LFLG) AS are shown in the left column, while the synthetically generated patients are displayed in the right column. AS60 represents the concordant AS data set, while DSE12 refers to the LFLG AS cohort. Arrows indicate the intrapatient stress response from the baseline state. Accuracies are displayed for the AS60 and synthetic data sets using one and two discriminant vectors.

The ML models performed well in the synthetic validation data, achieving an accuracy of up to 79.3% when using only the first discriminant vector, which demonstrates the utility of the learnt low-dimensional representation. In particular, this first discriminant vector can itself serve as a novel class-discriminative feature, which can be automatically calculated from routine clinical measurements, and potentially aid clinician decision-making.

Performance decreased when applied to the LFLG AS cohort; however, due to the small sample size, definitive conclusions regarding generalisation performance could not be drawn. Both PCA and LDA feature spaces showed improved class separation under stress states, further supporting the diagnostic value of DSE in this subgroup.

Together, these findings suggest that the optimal classification criteria for LFLG AS cannot be reliably obtained from this set of predictors, as observed in the concordant AS cohort. Larger clinical studies are needed, however, to validate these results, with predictors recorded consistently at every stage of the DSE protocol. Additionally, the framework developed can be extended to include other markers, which can further improve the classification accuracy. To ensure discriminative model performance in patients with LFLG AS, discordant cases must be included during training, and AS severity in these cases should ideally be assigned based on CT-derived calcium scoring or prospective clinical endpoints, allowing the identification of optimal thresholds for routine echocardiographic predictors.

6. Triplane Speckle-Tracking Echocardiography for 3D Left Ventricular Strain Analysis

This chapter concerns the application of strain imaging to assess the movement and deformation of the LV myocardium in patients with LFLG AS. First, the comparison of different techniques to derive global and regional strain measures is described in Section 6.1, followed by a brief summary of the related work in Section 6.2. Next, the deformation parameters are introduced with regard to a solid mechanics framework in Section 6.3, and in Section 6.4 the image registration framework is described to support the methodology outlined in Section 6.5. Finally, the results to compare the different strain parameters in the triplane and 3D modalities, as well as within the DSE12 data set, are detailed in Section 6.6. Section 6.7 outlines the main conclusions of this chapter.

6.1 Introduction

As described in Chapter 2, current clinical guidelines suggest the use of EF to evaluate global LV function and to define the classical and paradoxical phenotypes of LFLG AS. Despite its well-established utility, LVEF has several limitations. As an indicator of fractional volumetric change in the LV cavity, it is a global marker of LV function and is unable to provide information on regional signs of myocardial decline. In addition, it does not distinguish between systolic and diastolic impairment and is also inadequate to capture subclinical myocardial dysfunction. To address these constraints, STE-derived strain measures have emerged as valuable clinical markers in evaluating LV contractility to provide a more detailed and accurate representation of myocardial dysfunction, with robust prognostic utility in patients with AS.

Various modalities are available to derive strain parameters from echocardiograms, including 2D and 3D techniques. Although 3D echocardiography and STE are accessible in more than 95% of European laboratories, their adoption into routine practice remains limited due to image quality and operator dependence (Nabeshima et al., 2020). A recent survey showed that one in three centres used 3D US in routine examinations, and only 2 out of 96 laboratories reported employing 3D techniques during stress testing (Ajmone Marsan et al., 2020). In addition, the survey also highlighted significant heterogeneity in practice, with more than 70% of the centres relying on equipment from multiple vendors. This is especially relevant in the context of a recent meta-analysis comparing 2D and 3D strain values between major

vendors, revealing substantial vendor-specific differences in both GLS and GCS measurements (Nabeshima et al., 2020). This difference may arise from the alternate definitions used to define the region of interest and the various ways in which commercial software calculates strains for 2D and 3D geometries. For example, some software performs subendocardial tracking, while other tracks the entire myocardium. Furthermore, the calculation of GLS measured in 2D and in 3D is fundamentally different, given the geometric constraints, hence the direct comparison of these values remains physiologically limited and establishing reference values for GLS may require the consideration of both of these approaches.

Given the apparent under-utilisation of 3D STE in stress testing and the limitations of commercial software to represent 3D strains from multiple separate 2D recordings, this study investigated whether triplane acquisition mode could improve the quantification of 3D LV deformation parameters. This mode relies on US transducers simultaneously acquiring three planes at 60° angles, rather than obtaining the full volume, as shown in Figure 6.1. This has the potential to improve the spatial and temporal resolution of the scans, to reduce stitching artefacts resulting from the multi-beat reconstruction of 3D volumes, and to provide a more standardised apical view of the LV (Trache et al., 2014). The feasibility of triplane acquisition is comparable to 2D echocardiography and has also been shown to improve concordance with the gold standard CMR method in quantifying LV volumes (Malm et al., 2006). In contrast, GLS derived from 2D STE is generally more feasible than that obtained from 3D STE due to the technical expertise required to capture the entire LV through a single acoustic window (Nabeshima et al., 2020). The advantages of 3D STE over standard 2D and triplane techniques include its ability to assess the true three-dimensional geometry of the LV myocardium and to capture all components of systolic deformation simultaneously, thus allowing a more accurate quantification of rotation and twist.

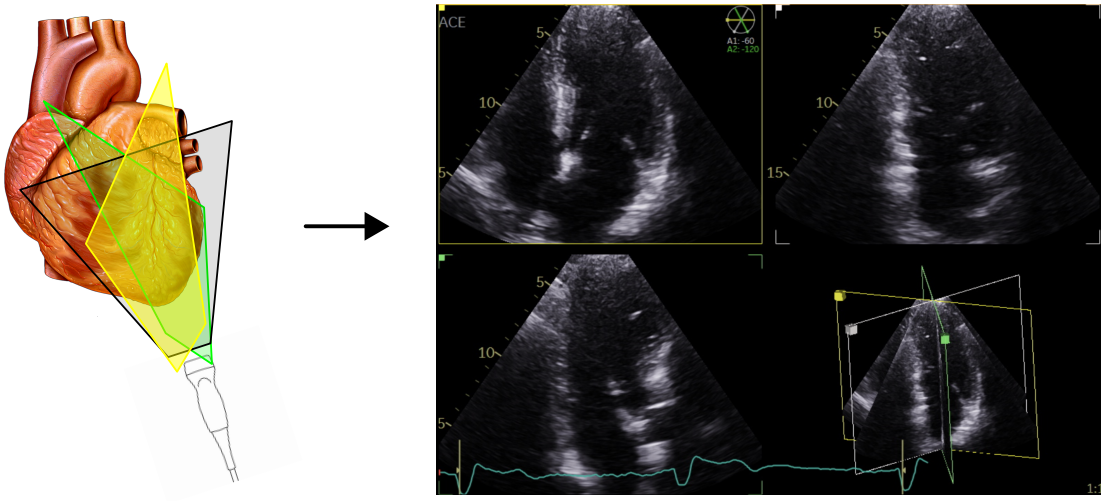


Figure 6.1: Triplane acquisition mode capturing three planes at 60° angles, resulting in three separate 2D + time B-mode recordings of the left ventricle with a fixed geometric relationship between the planes over time.

In this chapter, a novel method is introduced to obtain 3D LV strain parameters from triplane

echocardiographic recordings. This method is then validated using the MWMA data set. The method is then used in the DSE12 data set to investigate baseline characteristics of moderate and severe LFLG AS and to examine class-specific responses to stress testing in the LFLG AS cohort.

6.2 Related Work

Given the importance of accurate LV evaluation for understanding the pathology of AS, various 2D and 3D echocardiographic modalities allow quantification of the LV chamber and assessment of both systolic and diastolic volumes. The agreement between these techniques, while well established for volumetric and global strain measures, remains less explored for regional strain parameters.

Multiple studies have been conducted to compare global strains between 3D and traditional 2D echocardiography; however, the benefit of employing triplane echocardiography to obtain LV deformation parameters has been less researched. Trache et al. (2014) compared the GLS values obtained by standard 2D, triplane and 3D STE methods using commercial software and found that the 2D and triplane methods showed excellent agreement, with a Pearson's correlation coefficient of 0.95. In contrast, only a good overall correlation was found between the 2D and 3D methods, with correlation coefficients ranging from 0.81 to 0.83. They also observed that the 3D method produced consistently lower GLS values, and the volume rate had a significant impact on these results, causing the correlation to decrease when the volume rate exceeded 30 volumes per second, indicating that the 2D and 3D measurement methods are not interchangeable.

More recently, Plášek et al. (2022) showed that the commercial software used in echocardiography machines, while showing agreement between multiplanar and 3D methods for GLS, almost all regional longitudinal strains differed significantly, prohibiting accurate quantification of regional myocardial deformation parameters. This study demonstrates the need for an improved approach to obtain regional strain parameters from multiplanar echocardiography that better reflect the true 3D deformation.

6.2.1 Reconstruction Techniques

The goal of reconstructing the entire 3D geometry of the LV from multiple 2D echocardiographic recordings has been studied since the early 1990s, with the introduction of commercial 3D echocardiography machines (Gill and Klas, 2007). Martin et al. (1993) used TEE and elaborate scanning techniques to incrementally span the entire myocardium, providing an ordered dense set of 2D planes to obtain EDV and ESV. Gustavsson et al. (1993) employed TTE recordings proposing conventional spatial interpolation techniques using three long-axis and one short-axis plane to examine 3D myocardial movement over time with manually annotated

contours in each time frame. However, these initial approaches used idealised geometric assumptions on the position and orientation of the acquisition planes or time-consuming scanning protocols that could not be performed in routine practice.

More recently, Rajan et al. (2016) used cubic spline interpolation techniques to reconstruct the LV using three long-axis and additional short-axis TTE recordings from porcine data to obtain volume and flow rate over time. Their methodology improved upon standard previous approaches by incorporating optimisation objectives to estimate the position and orientation of the long-axis slices compared to the short-axis planes. Later, they expanded their work with improved assumptions on the relation between the long- and short-axis planes over time (taking into account that the SAX plane samples the myocardium at a different LV level at each time frame) and used it to develop computational modelling ready meshes (Hedayat et al., 2020).

A distinct approach to creating 3D LV surface meshes from medical imaging data includes the use of template-fitting methods. This involves the gradual deformation of a reference shape (often a reduced parametric model) that captures the main characteristics of the myocardium until an optimal alignment is achieved with the scan data. This has the advantage of being less dependent on high-resolution recordings than conventional interpolation techniques and is often preferred when data is sparse (Willems et al., 2024). Banerjee et al. (2021) proposed a fully automated pipeline for multiplanar reconstruction of the LV and RV from MRI recordings. They employed a statistical shape model to fit the different segmentation masks produced by a DL algorithm, reducing the total misalignment error from an average of 1.03 mm to 0.56 mm for LV endocardial contours. Joyce et al. (2022) implemented a similar pipeline for acquiring volumetric LV meshes from multiplanar MRI data by proposing a novel differentiable mesh voxelisation approach to the fitting regime, obtaining volume and global strain measures through the entire cardiac cycle. More recent research by Willems et al. (2024) used echocardiographic recordings from the A4C, A2C and PSAX views and used non-uniform rational B-splines to fit the endocardial and epicardial surface of the LV from a truncated ellipsoid template. The disadvantage of these fitting techniques over conventional interpolation methods includes the potential bias from over-simplified assumptions regarding the shape of the template which can lead to the inability to capture pathological geometries.

The contemporary literature is more prominent on DL methods for end-to-end fusion of multiple 2D echocardiographic views to produce 3D LV geometries (Stojanovski et al., 2022; Chen et al., 2025; Laumer et al., 2025); however, these methods do not readily facilitate the derivation of accurate strain measures due to the lack of point-to-point correspondance between the generated meshes over time. Although most of the above research focused on conventional 2D techniques for multiplanar reconstruction, some studies have explored the use of triplane acquisition modes. Dangi et al. (2015) used triplane TEE combined with conventional interpolation techniques to estimate EDV and ESV, while Vixège et al. (2022)

used colour Doppler to reconstruct flow patterns. However, no studies were found using triplane echocardiography to reconstruct the LV to derive regional deformation parameters in healthy individuals or patients with AS.

6.3 Solid Mechanics

To quantify the deformation of the LV myocardium throughout the cardiac cycle, a solid mechanics framework is employed. In this formulation, the LV is modelled as a continuous body embedded in three-dimensional space that undergoes deformation from an undeformed reference state to a deformed configuration. This section introduces fundamental kinematic quantities such as displacement, deformation gradient, and strain tensor, which describe the relationship between the undeformed and deformed configurations. A numerical approximation of the deformation gradient is outlined using triangular surface elements for regional strain analysis.

6.3.1 Position and Displacement

To represent the LV as a continuum body in three-dimensional space, $\mathbb{R}^3$, it is modelled as a collection of material particles. The configuration of this body before deformation, at an initial time $t = 0$, is referred to as the reference configuration, denoted by Ω_0, and serves as a baseline for the analysis of deformation. In this configuration, each material point occupies a unique position vector $\mathbf{X}$, referred to as the material or reference coordinate. Given a Cartesian coordinate system with an orthonormal basis $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$, the coordinates of the material points can be expressed as a linear combination of these basis vectors, such that

$$\boldsymbol{X} = X_1\boldsymbol{e_1} + X_2\boldsymbol{e_2} + X_3\boldsymbol{e_3}\,. \tag{6.1}$$

When the body undergoes deformation, each material point moves to a new position vector $\mathbf{x}$ in the current or deformed configuration at time t, denoted by Ω_t. The position $\mathbf{x}$ is related to the reference configuration through a continuous deformation function $\boldsymbol{\chi}()$, expressed as $\mathbf{x} = \boldsymbol{\chi}(\mathbf{X}, t)$. The displacement of a material point is then defined as the difference between its current position and its original position, so that

$$\boldsymbol{u}(\boldsymbol{X}, t) = \boldsymbol{\chi}(\mathbf{X}, t) - \boldsymbol{X}. \tag{6.2}$$

6.3.2 Deformation Gradient and Strain Tensors

The deformation of a continuum body can be described by examining the motion of an infinitesimal line element $d\mathbf{X}$ in the reference configuration. Under deformation, this line element moves and transforms to a new vector $d\mathbf{x}$ in the deformed configuration, as shown in Figure 6.2.

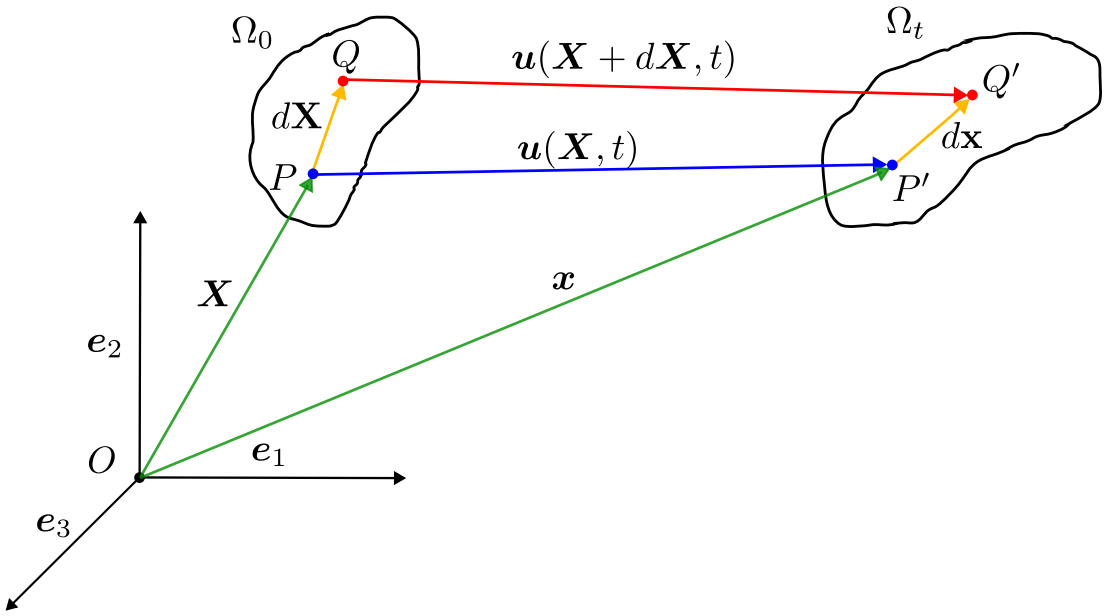


Figure 6.2: Deformation of an object from the reference configuration (Ω_0) to the current configuration (Ω_t). An infinitesimal line element $d\mathbf{X}$ connecting points P and Q in Ω_0 is mapped to $d\mathbf{x}$ between points P' and Q' in Ω_t. The displacement vectors are denoted by $\mathbf{u}(\mathbf{X},t)$ at P and $\mathbf{u}(\mathbf{X}+d\mathbf{X},t)$ at Q.

The deformation from the reference to the deformed configuration can be locally characterised by the deformation gradient tensor, $\mathbf{F}$, defined as

$$d\boldsymbol{x} = \boldsymbol{F}d\boldsymbol{X}, \qquad \text{where} \qquad \mathbf{F} = \frac{\partial \boldsymbol{\chi}(\mathbf{X},t)}{\partial \mathbf{X}} \tag{6.3}$$

The gradient of this displacement, called the displacement gradient tensor, denoted by $\nabla\mathbf{u}$, measures how the displacement varies with position in the reference configuration. From Equations 6.3 and 6.2 it can be expressed as

$$\nabla \boldsymbol{u}(\mathbf{X},t) = \frac{\partial \mathbf{u}(\mathbf{X},t)}{\partial \mathbf{X}} = \mathbf{F} - \mathbf{I} \tag{6.4}$$

From these measures, the Green–Lagrange strain tensor, $\mathbf{E}$, can be expressed as

$$\boldsymbol{E}(\mathbf{X},t) = \frac{1}{2}(\boldsymbol{F}^T\boldsymbol{F} - \mathbf{I}), \tag{6.5}$$

capturing the change in length and orientation of the infinitesimal line element.

While the above definitions describe deformation in a continuous domain, numerical implementation requires their discretisation. The implementation in this thesis represents the LV myocardium as a triangulated surface mesh rather than as a complete volumetric discretisation, as the multiplanar interpolation technique used to reconstruct the LV geometry was considered insufficient in accurately representing the small changes in the radial direc-

tion. Therefore, the continuous deformation measures introduced above are discretised over triangular surface elements.

6.3.3 Strains for Triangular Surface Elements

To discretise the continuous LV surface, the reference geometry is divided into non-overlapping triangular elements. Each triangle locally approximates the surface geometry, allowing the entire domain to be represented by a finite set of triangular patches. This enables regional deformation parameters to be estimated by examining the deformation of individual triangular elements defined by their nodal coordinates.

To obtain the deformation gradient for a triangular surface element from Ω_0 to Ω_t, the deformation gradients to Ω_0 and to Ω_t from a unit triangle with $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$ as its edge vectors were expressed separately. This unit triangle is referred to as the isoparametric configuration of a triangular element and can be used to derive the deformation gradient from Ω_0 to Ω_t.

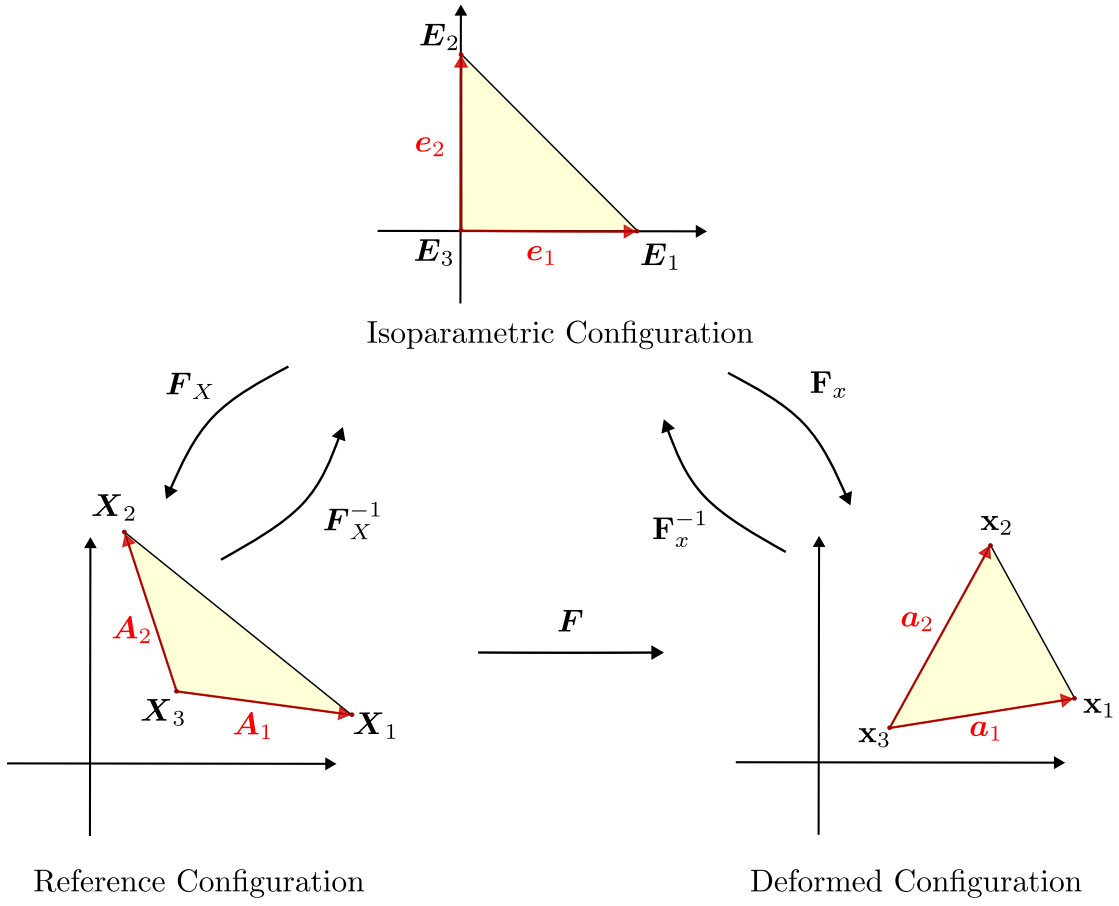


Figure 6.3: Representation of the deformation gradient $\mathbf{F}$ through a sequence of transformations acting on a triangular surface element. The element is first mapped from the reference configuration to the isoparametric configuration via $\mathbf{F}_X^{-1}$, and then from the isoparametric configuration to the deformed configuration via $\mathbf{F}_x$. In the reference configuration, the edges $\mathbf{A}_1$ and $\mathbf{A}_2$ are mapped to $\mathbf{a}_1$ and $\mathbf{a}_2$ in the deformed configuration.

Assuming the strain is constant over the triangular element, the in-plane components of the covariant basis vectors in the reference and deformed configurations can be calculated as

$$\boldsymbol{A_1} = \boldsymbol{X_1} - \boldsymbol{X_3}, \quad \boldsymbol{A_2} = \boldsymbol{X_2} - \boldsymbol{X_3}, \quad \boldsymbol{a_1} = \boldsymbol{x_1} - \boldsymbol{x_3}, \quad \boldsymbol{a_2} = \boldsymbol{x_2} - \boldsymbol{x_3}\,, \tag{6.6}$$

where $\boldsymbol{A}$ and $\boldsymbol{a}$ correspond to the edge vectors of the element in Ω_0 and Ω_t, respectively.

Expressing the two deformation gradients that map the isoparametric configuration to the reference and deformed configurations yields

$$\boldsymbol{F}_x = \frac{\partial \mathbf{x}}{\partial \mathbf{E}} = \begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{n} \end{bmatrix}, \quad \boldsymbol{F}_X = \frac{\partial \mathbf{X}}{\partial \mathbf{E}} = \begin{bmatrix} \boldsymbol{A_1} \\ \boldsymbol{A_2} \\ \boldsymbol{N} \end{bmatrix}, \tag{6.7}$$

which describes how the basis vectors are transformed from the unit triangle to the given configuration. $\mathbf{N}$ denotes the normal vector in the reference configuration, while $\mathbf{n}$ is the normal vector in the deformed configuration.

From Equation 6.7, $\boldsymbol{F}$ can be calculated as

$$\boldsymbol{F} = \boldsymbol{F}_x \boldsymbol{F}_X^{-1}, \tag{6.8}$$

representing a series of transformations, first from the reference configuration to the isoparametric configuration by $\boldsymbol{F}_X^{-1}$, then from the isoparametric configuration to the deformed configuration by $\boldsymbol{F}_x$, as shown in Figure 6.3. It should be noted that, under the assumption of constant strain within each linear triangular element, the resulting strain field is piecewise constant and may therefore be discontinuous across element boundaries. For the purposes of this study, the strain values were interpreted as regional deformation estimates rather than as a fully continuous myocardial strain field. This was considered appropriate because the subsequent analysis focused on global and segmental strain summaries, where local element-wise discontinuities are averaged within anatomically defined regions.

6.3.4 Directional Strain Measures

LV myocardial strains are routinely measured in the longitudinal, circumferential, and radial directions, demonstrating the different types of contractions the LV undergoes during systole. To calculate the stretch, λ, of a line element in the direction of an arbitrary vector, $\boldsymbol{v}$, in the reference configuration, the Cauchy-Green deformation tensor, $\mathbf{C} = \mathbf{F}^T\mathbf{F}$, is used as

$$\lambda_{\mathbf{v}}(\mathbf{X}, t) = \sqrt{\boldsymbol{v}^T \boldsymbol{C} \boldsymbol{v}}\,, \tag{6.9}$$

which defines the ratio of the length of a deformed line element to the length of the corresponding undeformed line element. To obtain directional strains in the deformed configuration, $\epsilon_{\mathbf{v}}$, the Green-Lagrange strain measure is used as

$$\epsilon_{\mathbf{v}}(\mathbf{X}, t) = \mathbf{v}^T \mathbf{E} \mathbf{v}. \tag{6.10}$$

6.4 Image Processing

Expressing the motion and deformation of the LV myocardium within a solid mechanics framework enabled the computation of regional strain measures and displacement values, based on the material point correspondence between the reference configuration and the deformed configuration. Manually annotating these points across all time frames within the cardiac cycle is labour intensive and prone to interobserver and intraobserver variability. To address this, an image registration approach can be implemented that facilitates the automatic propagation of a reference configuration, requiring it to be defined only once per patient. However, the resulting deformation field remains dependent on image quality, spatial and temporal resolution, and the assumptions imposed by the registration algorithm. Non-uniform myocardial motion, through-plane motion, and regions with weak image texture may not be fully captured, which can affect the accuracy of the derived displacement and strain measures. In this section, the general image registration framework is described with additional details on the physical interpretation of image data through common orientation and transformation matrices.

6.4.1 Image Registration

The goal of image registration is to estimate a spatial transformation that maps each point in the moving image, $I_t : \Omega_t \rightarrow \mathbb{R}$ (representing the deformed configuration) to its corresponding location in the fixed image, $I_0 : \Omega_0 \rightarrow \mathbb{R}$ (representing the reference configuration). This is achieved by iteratively updating the current transformation, $s_i : \Omega_t \rightarrow \Omega_0$ as follows. First, the moving image is interpolated to produce a warped image in the fixed image space, denoted by $I_t \circ s_i$. Next, a similarity measure is computed between the warped image and the fixed image. A commonly used similarity measure is the normalised cross-correlation, expressed as

$$\mathcal{D}_{NCC}(I_0, I_t \circ s) = \frac{\sum_{\mathbf{x}\in\Omega_x}(I_0(\mathbf{x}) - \bar{I}_0)(I_t \circ s(\mathbf{x}) - \bar{I}_t)}{\sqrt{\sum_{\mathbf{x}\in\Omega_x}(I_0(\mathbf{x}) - \bar{I}_0)^2 \sum_{\mathbf{x}\in\Omega_x}(I_t \circ s(\mathbf{x}) - \bar{I}_t)^2}}, \tag{6.11}$$

where Ω_x refers to the region of overlap between the two images in which the metric is calculated and $\bar{I}_0$ and $\bar{I}_t$ are the mean intensities of the fixed and the moving images within this region. Another similarity metric employs the sum of the squared differences over the entire image, such that

$$\mathcal{D}_{SSD}(I_0, I_t \circ s) = \sum_{\mathbf{x}\in\Omega_0}(I_0(\mathbf{x}) - I_t \circ s(\mathbf{x}))^2. \tag{6.12}$$

After computing the similarity between the fixed image and the warped moving image, the current transformation is updated via gradient-based optimisation. This involves calculating the gradient of the objective function, which may consist solely of the similarity metric or include an additional explicit regularisation term that acts on the transformation to pe-

nalise irregularities in the displacement field and prevent large deformations. The general optimisation objective can be expressed as

$$s^* = \arg\min_s \ \mathcal{E}(s; I_0, I_t) = \arg\min_s \Big[\mathcal{D}(I_0,\ I_t \circ s) + \lambda\, \mathcal{R}(s)\Big], \tag{6.13}$$

where λ controls the strength of the explicit regularisation term, $\mathcal{R}(s)$. Alternatively, smoothing can be used to implicitly regularise the displacement field. Smoothing is often achieved by applying a Gaussian kernel to the calculated gradient, thus causing neighbouring voxels to perform a similar update to the displacement field. In addition, the updated displacement field can also be smoothed by a Gaussian kernel to further reduce noise or local folding. This implicit regularisation technique helps maintain the diffeomorphic properties of the transformation ($s : \Omega_t \rightarrow \Omega_0$ is differentiable and has a differentiable inverse $s^{-1} : \Omega_0 \rightarrow \Omega_t$), which ensures that the estimated motion remains anatomically plausible. The general image registration framework is shown in Figure 6.4.

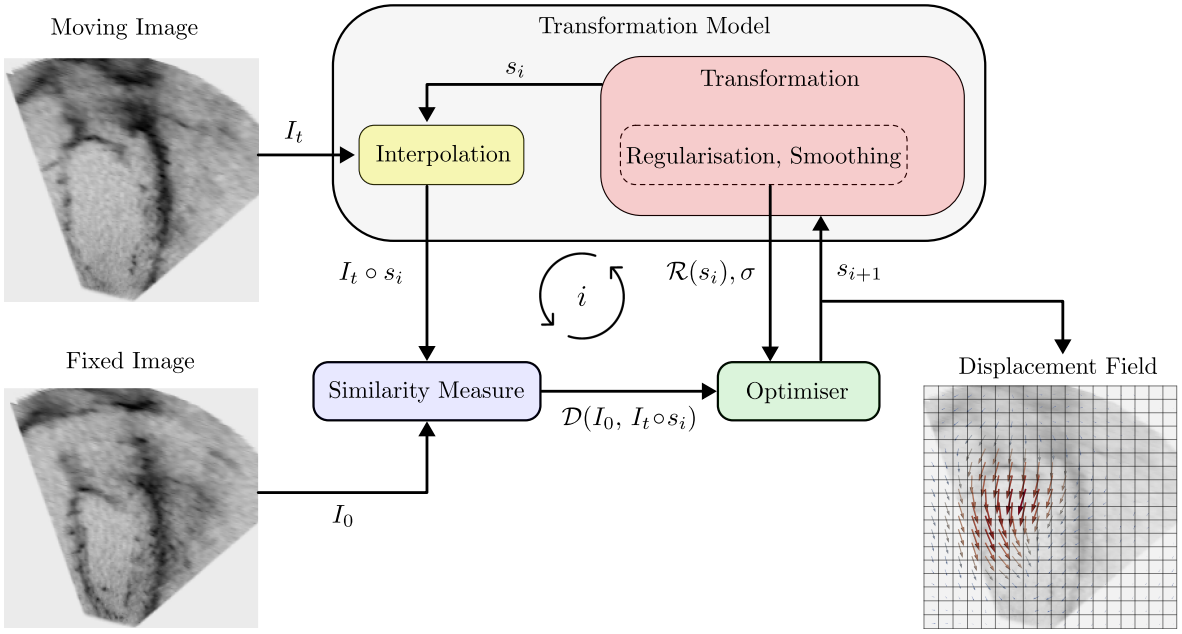


Figure 6.4: Schematic representation of the deformable image registration framework. The moving image I_t is interpolated using the current spatial transformation s_i and compared to the fixed image I_0 via a similarity measure $\mathcal{D}(I_0, I_t \circ s_i)$. The optimiser computes an update to the transformation based on the gradient of the objective function. Smoothness of the deformation field is enforced implicitly through Gaussian filtering of both the similarity gradient and the updated transformation field or via an explicit regularisation term in the objective function. Figure adapted from Curiale et al. (2017).

The output of the image registration process is a displacement field that maps the deformed configuration to the reference configuration. Since s_i is invertible, $\mathbf{u}(\mathbf{X}, t)$ can also be obtained in the reference configuration and Equation 6.2 can be used to obtain the deformed configuration. An example of a displacement field, visualised in the three main anatomical planes, is shown in Figure 6.5.

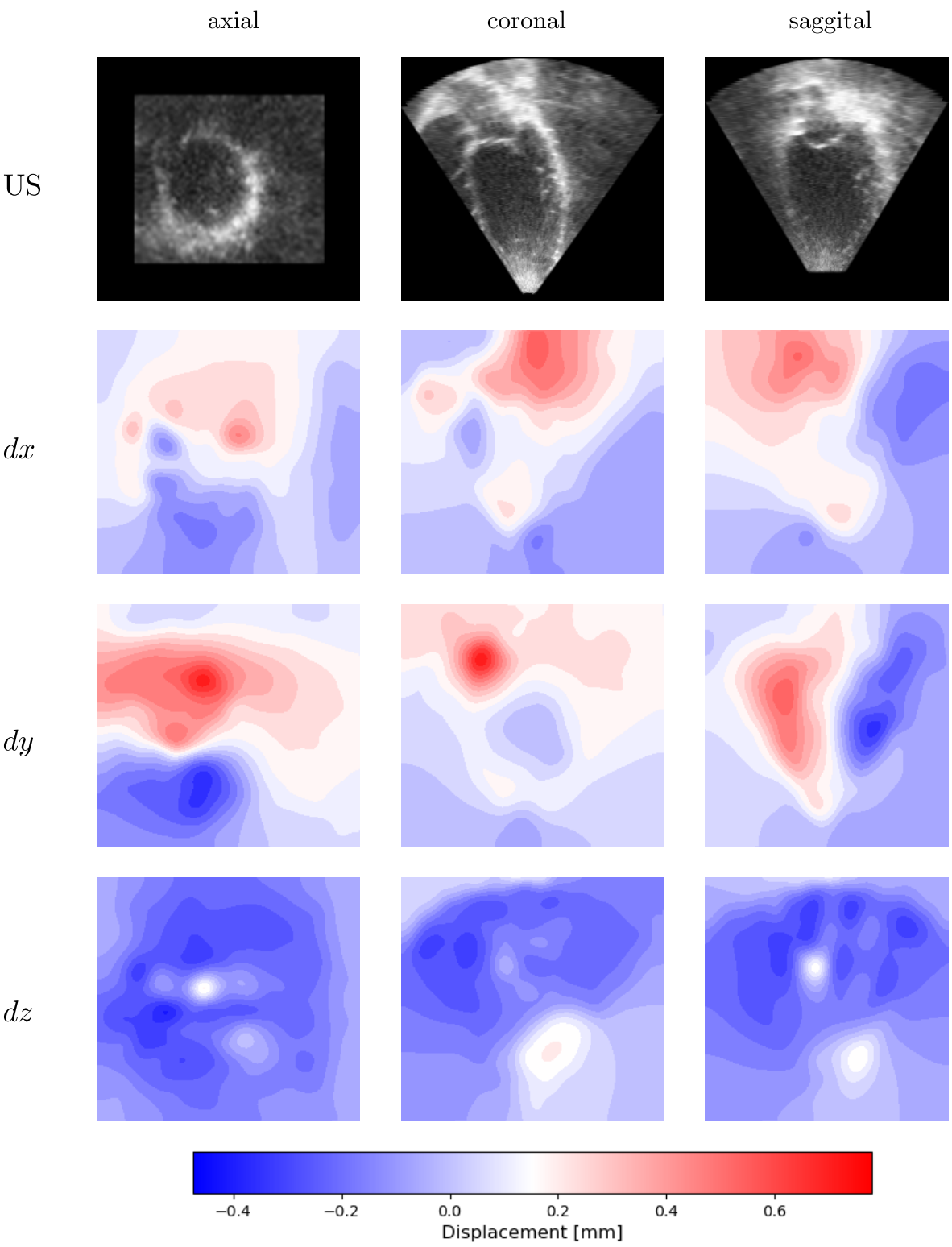


Figure 6.5: The displacement field from a deformable image registration tool as displayed on the three major anatomical planes for components x, y and z.

6.4.2 Transforms and Orientations

As image data are stored in computer memory as an array of pixel/voxel intensities, it lacks physical interpretability on its own. A transformation matrix needs to be defined alongside these values to create a mapping between pixel space and physical coordinates. This is generally described as a 4x4 affine matrix to represent translation by a vector $\mathbf{t}$ as $\boldsymbol{A}_{T(\mathbf{t})}$, scaling by vector $\mathbf{s}$ as $\boldsymbol{A}_{S(\mathbf{s})}$ and rotation about vector $\mathbf{v}$ by angle θ as $\boldsymbol{A}_{R_{\mathbf{v}}(\theta)}$. The affine matrices are constructed as

$$\boldsymbol{A}_{T(\mathbf{t})} = \begin{bmatrix} \boldsymbol{I} & \mathbf{t} \\ 0 & 1 \end{bmatrix}, \boldsymbol{A}_{S(\mathbf{s})} = \begin{bmatrix} \text{diag}(\mathbf{s}) & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{A}_{R_{\mathbf{v}}(\theta)} = \begin{bmatrix} R_{\mathbf{v}}(\theta) & 0 \\ 0 & 1 \end{bmatrix}, \tag{6.14}$$

where $\boldsymbol{I}$ is a 3x3 identity matrix, and $\text{diag}(\mathbf{s})$ represents a 3x3 diagonal matrix. Rotation matrices about the standard orthonormal basis vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are given by

$$R_{\mathbf{x}}(\gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\gamma & -sin\gamma \\ 0 & sin\gamma & cos\gamma \end{bmatrix}, R_{\mathbf{y}}(\beta) = \begin{bmatrix} cos\beta & 0 & sin\beta \\ 0 & 1 & 0 \\ -sin\beta & 0 & cos\beta \end{bmatrix}, R_{\mathbf{z}}(\alpha) = \begin{bmatrix} cos\alpha & -sin\alpha & 0 \\ sin\alpha & cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.15}$$

A counterclockwise rotation of α about $\boldsymbol{z}$ is also called a yaw angle, while the counterclockwise rotation of β about $\boldsymbol{y}$ is called the pitch angle and the counterclockwise rotation of γ about $\boldsymbol{x}$ is called the roll angle. The rotation matrix about an arbitrary unit vector $\mathbf{v}$ by angle θ in the counter-clockwise direction is expressed by Rodrigues' formula

$$R_{\mathbf{v}}(\theta) = \boldsymbol{I} + (sin\theta)\boldsymbol{K} + (1 - cos\theta)\boldsymbol{K}^2, \quad \text{with} \quad \boldsymbol{K} = \begin{bmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{bmatrix}. \tag{6.16}$$

Transform values are set to reposition an object in space without deforming it,hence given the above translation, scaling, and rotation values, an affine transform sequence can be composed as

$$\boldsymbol{A} = \boldsymbol{A}_{T(\mathbf{t})}\boldsymbol{A}_{R_{\mathbf{y}}(\beta)}\boldsymbol{A}_{R_{\mathbf{z}}(\alpha)}\boldsymbol{A}_{R_{\mathbf{x}}(\gamma)}\boldsymbol{A}_{S(\mathbf{s})}, \tag{6.17}$$

which first scales the object, then rotates it, and finally translates it.

While the above transformations were described within the same coordinate system to construct the overall affine matrix, medical imaging datasets often originate from various acquisition systems and are stored in different file formats that rely on distinct coordinate system conventions, also referred to as orientations. Properly understanding and managing these orientations ensures consistent anatomical interpretation across modalities and software tools. Two of the most widely used coordinate systems in medical imaging are the RAS (Right-Anterior-Superior) and LPS (Left-Posterior-Superior) conventions. In the RAS system (used in NIfTI files and in 3D Slicer), the coordinate axes are defined such that:

- The X-axis increases towards the right side of the patient.

- The Y-axis increases towards the anterior direction (front).
- The Z-axis increases towards the superior direction (up).

In contrast, the LPS system (used in DICOM data and in ITKSNAP) defines:

- The X-axis increases towards the left side of the patient.
- The Y-axis increases towards the posterior direction (back).
- The Z-axis also increases towards the superior direction, consistent with the RAS system.

When transferring data between these different coordinate systems, orientation conversion is required. This is achieved through the construction of an orientation matrix, $\mathbf{A}_O$ that reflects the conversion from one coordinate system to another. For example, the conversion from LPS to RAS is given by

$$\mathbf{A}_{O_{LPS\to RAS}} = \text{diag}(-1, -1, 1, 1) \tag{6.18}$$

and the transformation of an affine matrix from LPS to RAS is computed as

$$\mathbf{A}_{RAS} = \mathbf{A}_{O_{LPS\to RAS}} \mathbf{A}_{LPS} \mathbf{A}^{-1}_{O_{LPS\to RAS}}. \tag{6.19}$$

6.5 Methodology

This section provides an explanation of the methods used to evaluate LV regional deformation parameters from B-mode US in patients with LFLG AS. While the statistical analysis performed in Chapter 5 concerned the investigation of global markers of disease severity, this chapter focuses on more complex measures; and uses exclusively the B-mode triplane recordings from the DSE12 dataset. This approach was considered due to the unavailability of retrospective 3D data. First, the verification of the strain calculations is described. Next, the pipeline is introduced to compare regional strain measures derived from 3D US with regional strain measures calculated from triplane recordings. Finally, the specifics of the clinical study that uses the triplane methodology are outlined.

6.5.1 Verification of Strain Measures

To verify the strain calculations described in Equations 6.7 and 6.8, a cylinder was used as a simple reference geometry. The cylinder was parameterised by by its radius, r, and its height, h. The material points were defined by covering the curved surface of the cylinder via N_z number of equally spaced circles along its axis. Each circle consisted of N_θ number of points placed equidistantly along the circumference. To ensure consistent ordering for the creation of the triangular surface elements, the point IDs were assigned incrementally, first iterating over the points within a circle and then along the axis, starting from the base of the cylinder.

Triangles were constructed using adjacent points from a lower circle and their corresponding points in the upper circle, forming two triangles per quadrilateral strip segment. For further details, refer to Algorithm 2 found in Appendix B.

Calculating the strain directions in the reference configuration for a triangular element with material points $\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3}$ was performed as follows. First, the normal vector was calculated by taking the cross product of two edge vectors as

$$\boldsymbol{N_r} = (\boldsymbol{X_1} - \boldsymbol{X_3}) \times (\boldsymbol{X_2} - \boldsymbol{X_3}) \tag{6.20}$$

to define the radial direction, and then this was normalised by dividing it by its magnitude, so that

$$\boldsymbol{v_r} = \frac{\boldsymbol{N_r}}{\|\boldsymbol{N_r}\|} \ . \tag{6.21}$$

The circumferential direction was simply defined by approximating the tangent via a normalised edge vector as

$$\boldsymbol{v_c} = \frac{\boldsymbol{X_1} - \boldsymbol{X_3}}{\|\boldsymbol{X_1} - \boldsymbol{X_3}\|} \ , \tag{6.22}$$

which was made possible by the ordering constraints detailed above. Finally, the longitudinal direction was calculated by taking the cross product of the radial and circumferential vectors as

$$\boldsymbol{v_l} = \boldsymbol{v_r} \times \boldsymbol{v_c} \ . \tag{6.23}$$

To create the deformed configuration, the geometry described above was modified so that its radius would increase by a constant scaling factor, s, along its height. Ensuring that both meshes were created with an equal number of points with their ordering kept the same allowed for the calculation of the regional strain measures analytically, serving as the ground truth value for the verification.

6.5.2 Strains Measures from 3D Echocardiography

Once the reliability of the strain calculations was assessed, the clinically used strain imaging pipeline was implemented using the MWMA dataset introduced in Chapter 4. This was carried out to ensure that the values obtained here can serve as a reference for evaluating the accuracy of the triplane measurements, which will be detailed in the next section.

Modelling the LV myocardium as a continuous deformable solid began with defining the appropriate geometry of the LV. The 17 segment model defined by the AHA is the most widely used clinical model, which divides the LV into distinct regions based on their anatomical positions. These regions are often illustrated on a bull's eye diagram, as shown in Figure 6.6, to facilitate the graphical depiction of various parameters for clinical assessment. The bull's eye diagram resembles a polar plot, comprised of 17 segments with the apex situated at the centre. Continuing from the centre towards the boundary of the plot, the apical, middle, and basal levels of the LV are included. The apical level is divided into four parts, whereas the

middle and basal levels are subdivided into six equal segments each. The interior lining of the LV is called the endocardial border, the exterior surface is called the epicardial border, and the average of these two delineates the myocardial midline. The long axis (LAX) of the LV is defined by the line passing through the apex and the mid-base point, located in the centre of the basal endpoints of the myocardium. A short axis (SAX) plane is then defined as a plane that is perpendicular to the LAX.

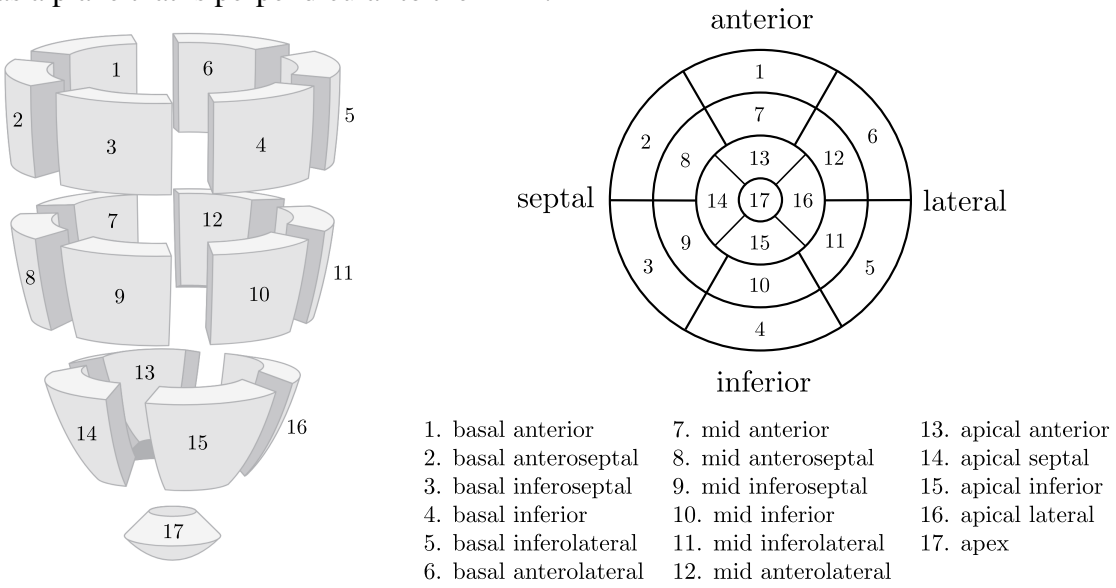


Figure 6.6: The 17 segment model of the left ventricle as visualised in 3D [left], and as depicted on a bullseye plot with the region labels listed below [right]. (Figure adapted from Hjertaas et al. (2023)).

Using 4D US NifTI files from each patient, LV surface meshes were generated at the peak diastolic frames. This started by creating markup files in 3D Slicer to define the anatomical landmarks of the LV myocardium. First, landmark points were placed at the apex, **a**, and the mid-base point, **b**, to define the LAX. Then, the rotation matrix needed to align the LAX with the z axis was calculated. This involved the construction of a unit length vector, **r** about which the counterclockwise rotation angle is defined, through the cross product of the z-axis and a normalised vector in the direction of the long axis, $\mathbf{v}_l = \frac{\mathbf{b}-\mathbf{a}}{\|\mathbf{b}-\mathbf{a}\|}$, as

$$\mathbf{r} = \mathbf{v}_l \times \mathbf{e}_3 \tag{6.24}$$

and calculating the rotation angle, θ, by

$$\theta = \arccos(\mathbf{v}_l \cdot \mathbf{e}_3). \tag{6.25}$$

The rotation matrix was then computed using Rodrigues' rotation formula as described in Equation 6.16. The overall affine transformation was given by

$$\mathbf{A}_{LAX} = \begin{bmatrix} R_{\mathbf{r}}(\theta) & 0 \\ 0 & 1 \end{bmatrix}. \tag{6.26}$$

After applying this transformation to the image at the diastolic frame, the SAX planes were

realigned with the axial plane. This allowed for the consistent placement of anatomical landmarks across the slices. Next, in each selected SAX plane, eight control points were placed manually to outline the midwall of the LV myocardium. 3D Slicer's default spline interpolation was then applied through these control points to generate a smooth representation of the myocardial contour, and the control points were then adjusted until the closest match was found between the interpolation curve and the midwall. Four SAX planes were annotated in this way, selected to correspond to the boundaries of the basal, mid-cavity, and apical segments, as defined by the AHA myocardial segmentation model as shown in Figure 6.6.

Once the landmark points were defined in the LAX aligned configuration of the LV, they were transformed back to the original image space by applying the inverse transform, $\mathbf{A}_{LAX}^{-1}$, and saved using 3D Slicer's markup format.

To generate the myocardial contour from the saved landmark points in each SAX plane, a closed Catmull–Rom spline was used. In this interpolation technique, each control point, $\mathbf{c}_i$, is associated with a single tangent vector, $\mathbf{m}_i$, which serves as both the outgoing tangent from the point into the next segment and the incoming tangent into the point from the previous segment. This ensures C1 continuity, meaning that the first derivative is continuous across segment boundaries, resulting in a smooth, anatomically more plausible curve.

The tangent vectors at each control point are estimated via the spatial coordinates of three adjacent control points and their associated parametric coordinate t_i, as

$$\mathbf{m}_i = \frac{(t_{i+1} - t_i)(\mathbf{c}_i - \mathbf{c}_{i-1})}{(t_i - t_{i-1})(t_{i+1} - t_{i-1})} + \frac{(t_i - t_{i-1})(\mathbf{c}_{i+1} - \mathbf{c}_i)}{(t_{i+1} - t_i)(t_{i+1} - t_{i-1})} \quad . \tag{6.27}$$

To improve interpolation consistency, a non-uniform chordal parameterisation was implemented. This involved calculating the cumulative Euclidean distances between the control points to construct a grid based on the chord length. These distances were then normalised and used as the parametric coordinate of the spline, as calculated by

$$t_i = \frac{\sum_{j=0}^{i} \|\mathbf{c}_{j+1} - \mathbf{c}_j\|}{\sum_{j=0}^{7} \|\mathbf{c}_{j+1} - \mathbf{c}_j\|} \tag{6.28}$$

Since the Catmull–Rom method uses cubic splines to construct the curve, the coordinates of each interpolated point, $\mathbf{p}(\tau)$, for a segment between $\mathbf{c}_i$ and $\mathbf{c}_{i+1}$ can be expressed as

$$\mathbf{p}(\tau) = h_{00}(\tau)\mathbf{c}_i + h_{10}(\tau)(t_{i+1} - t_i)\mathbf{m}_i + h_{01}(\tau)\mathbf{c}_{i+1} + h_{11}(\tau)(t_{i+1} - t_i)\mathbf{m}_{i+i} \quad , \tag{6.29}$$

where, $\tau = \frac{t - t_i}{t_{i+1} - t_i}$, represents a local parameter, remapped from the global parametric coordinates, and the basis functions, $h(\tau)$, are given via

$$\begin{aligned} h_{00}(\tau) &= 2\tau^3 - 3\tau^2 + 1 \quad , h_{10}(\tau) = \tau^3 - 2\tau^2 + \tau \\ h_{01}(\tau) &= -2\tau^3 + 3\tau^2 \quad\quad , h_{11}(\tau) = \tau^3 - \tau^2 \end{aligned} \quad . \tag{6.30}$$

The spline was created with a periodic boundary condition to grant continuity around the endpoints. For each SAX plane, a Catmull–Rom spline was first evaluated at regular intervals to produce N_θ uniformly spaced points along the closed myocardial contour. To ensure consistent ordering across all slices, the resulting set of interpolated points was then reordered so that the point with the smallest angle relative to the x axis served as the starting point. This reordering also enforced a counterclockwise direction, facilitating the alignment of interpolation points across slices. An example of the SAX interpolation regime is shown in Figure 6.7.

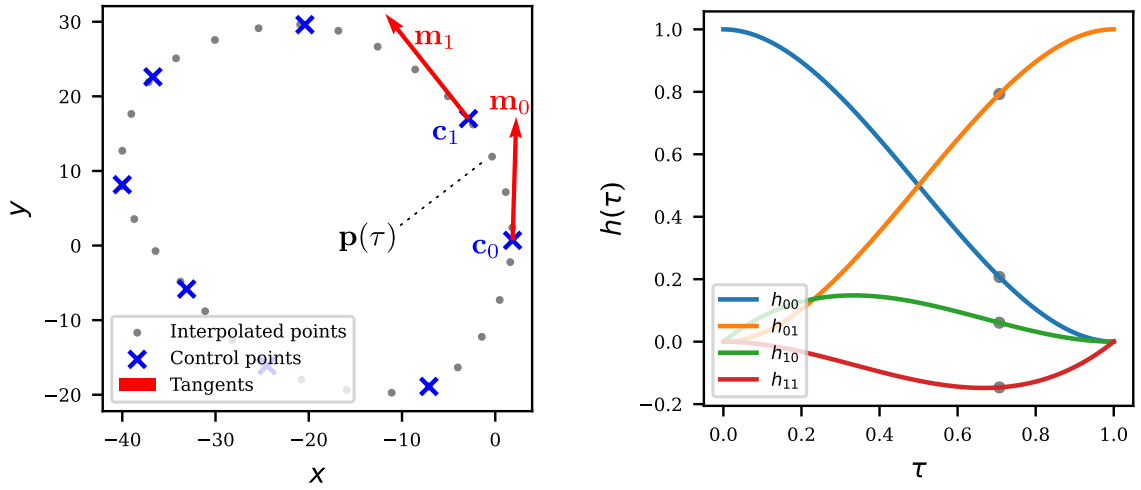


Figure 6.7: Short axis interpolation using Catmull-Rom splines through the selected control points, generating N_θ number of equally spaced points [left] and the basis functions used to construct the spline [right]. The annotations highlight how the coordinates of the interpolated point, $\mathbf{p}(\tau)$ on the segment between $\mathbf{c}_0$ and $\mathbf{c}_1$ are calculated using the local parametric coordinate τ and the basis functions.

Once the contours were generated at each SAX plane, a similar LAX interpolation was performed to create the remaining material points. This was done by selecting a single point from each SAX curve that corresponded to the same index from the reordered sets and using these and the apex as control points. Setting the number of points interpolated along the LAX to be N_z resulted in a geometry comparable to the cylindrical one described in the previous section; therefore, the same techniques could be used for the generation of triangular elements and the calculation of the strain directions.

Long-axis (LAX) interpolation using Catmull-Rom splines through the corresponding control points on each short-axis (SAX) curve. The left figure shows the equally spaced interpolated points along the LAX direction (N_z) generated between SAX contours (apical to basal), with the apex marked in red. The right figure displays the resulting 3D surface mesh created from these interpolated points, visualised in ParaView.

To track the myocardium through the cardiac cycle, the meshes created in the end dias-

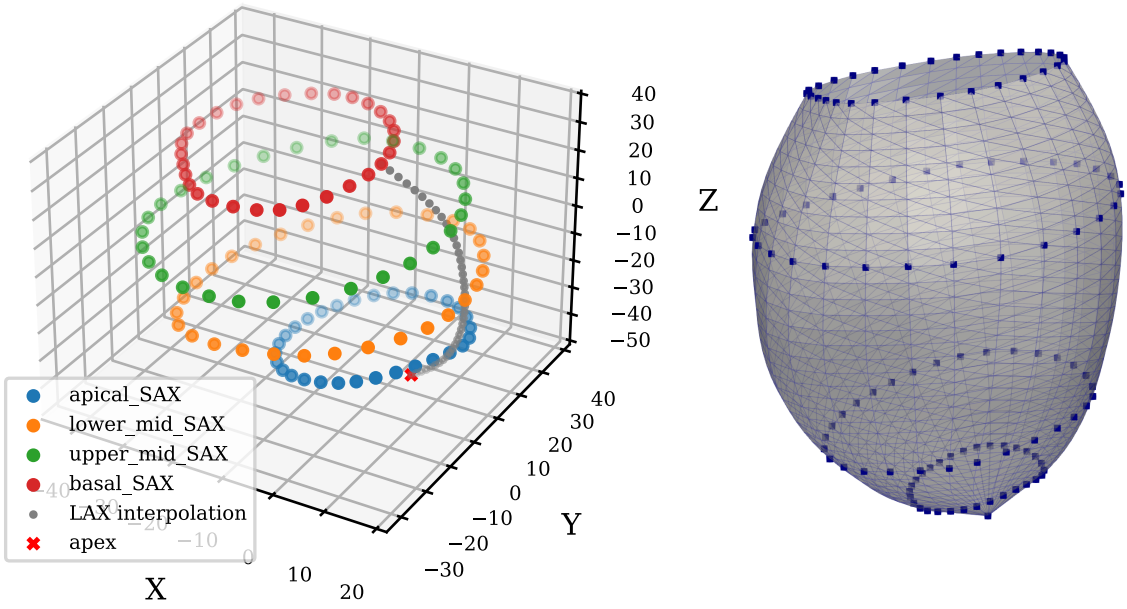


Figure 6.8: Long-axis (LAX) interpolation using Catmull-Rom splines through the corresponding control points on each short-axis (SAX) curve. The left figure shows the equally spaced interpolated points along the LAX direction (N_z) generated between SAX contours (apical to basal), with the apex marked in red. The right figure displays the resulting 3D surface mesh created from these interpolated points, visualised in ParaView.

tolic frames were propagated by applying the affine transformations and displacement fields obtained from the deformable image registration.

6.5.3 Strain Measures from Synthetic Triplane Data

To assess the accuracy of the triplane pipeline compared to the 3D methodology, synthetic triplane data were created from the previously used 3D recordings by slicing them at consecutive 60° angles along the LAX, mimicking triplane acquisition. This was achieved by computing an affine matrix that performed the following:

- Transformed the image to identity space by applying the inverse affine matrix, $\mathbf{A}_{H^{-1}}$ stored in the header of each NIfTI file. This was done to avoid misalignment in the slicing operation used to produce the planes by ensuring that the transforms and images were in the same space and had the same orientation. Since these affine matrices only stored scale and translation values they could be characterized by scaling and translation vectors as, $\mathbf{s}_{H^{-1}}$ and $\mathbf{t}_{H^{-1}}$.
- Translated the apex to the origin via $\mathbf{T}_O = \mathbf{T}(-(\mathbf{a}\mathbf{s}_{H^{-1}} + \mathbf{t}_{H^{-1}}))$, where $\mathbf{a}$ is the landmark point coordinates of the apex annotated in the previous section.
- Aligned the LAX with the z-axis, via $\mathbf{A}_{H^{-1}}\mathbf{A}_{LAX}\mathbf{A}_H$, where $\mathbf{A}_{LAX}$ is identical to the one in the previous section.

- Rotated the image around the z-axis by $\mathbf{R}_z(\alpha)$, where $\alpha = 0°, 60°, 120°$
- Rotated the image around the y-axis by $\mathbf{R}_y(90°)$ to place the acquisition plane to align with the XY-plane
- Translated the image by $\mathbf{T}_{XY}$ to align the acquisition plane with the voxel space.

The overall affine matrix used to create the planes was the composition of the above transformations, expressed as:

$$\mathbf{A}_{2D}(\alpha) = \mathbf{T}_{XY}\mathbf{R}_y(90°)\mathbf{R}_z(\alpha)\mathbf{A}_{LAX}\mathbf{T}_O\mathbf{A}_{H^{-1}}. \tag{6.31}$$

Applying this transformation to the 3D + time US recordings and slicing them via taking the first XY-plane in voxel coordinates resulted in 2D + time image sequences at different acquisition angles. The 3D mesh, generated at the end-diastolic frame in the previous section, was also transformed via this affine matrix, and landmark points were placed at the intersection of the mesh and the 2D image, using 3D Slicer. This allowed for the accurate reconstruction of the LV geometry from triplane images, enabling an in-depth analysis of the image registration parameters on the error metrics for displacement and regional strain values.

After defining the landmark points on each plane, a Catmull-Rom interpolation was used to create a smooth representation of the myocardial midwall along the LAX by generating N_z number of points.

Performing image registration in all three planes between the consecutive frames, yielded displacement fields that could be applied to each LAX contour separately. After propagating these temporary meshes throughout the cardiac cycle, the inverse transforms, $\mathbf{A}_{2D}^{-1}(\alpha)$ were applied at each timestep to assemble the contours in the original space. Since each LAX curve had an equal number of points, SAX interpolation could be implemented by choosing the corresponding point IDs from each contour. To facilitate the comparison of the final LV meshes between the triplane and 3D pipelines, N_θ points were created along the SAX. The triangular elements and the strain directions were calculated as described in the previous sections.

Error Metrics

Denoting the mesh created via the 3D methodology with matrix $\mathbf{M}$, and the mesh generated by the triplane pipeline with $\hat{\mathbf{M}}$, where both matrices representing the mesh coordinates are in $\mathbb{R}^{N_\theta N_z \times 3}$, allows the calculation of pointwise displacement metrics. The root mean square error for corresponding points is calculated as

$$RMSE_{corr} = \sqrt{\frac{1}{N_\theta N_z}\sum_{i=1}^{N_\theta N_z} \|\mathbf{M}_i - \hat{\mathbf{M}}_i\|^2}, \tag{6.32}$$

while the maximum distance between corresponding points is reported, such that

$$MAX_{corr} = \max_i \|\mathbf{M}_i - \hat{\mathbf{M}}_i\| \tag{6.33}$$

Closest point displacement metrics included the directed Hausdorff distances defined as

$$H(\mathbf{M} \to \hat{\mathbf{M}}) = \max_i \min_j \|\mathbf{M}_i - \hat{\mathbf{M}}_j\| \quad , H(\hat{\mathbf{M}} \to \mathbf{M}) = \max_i \min_j \|\hat{\mathbf{M}}_i - \mathbf{M}_j\|, \tag{6.34}$$

the symmetric Hausdorff distance expressed as the maximum of the two directed distances as

$$H(\mathbf{M}, \hat{\mathbf{M}}) = \max(H(\mathbf{M} \to \hat{\mathbf{M}}), H(\hat{\mathbf{M}} \to \mathbf{M})) \tag{6.35}$$

and the symmetric root mean square error for the closest points computed as:

$$RMSE_{closest} = \sqrt{\frac{1}{2N_\theta N_z} \sum_{i=1}^{N_\theta N_z} (\min_j \|\mathbf{M}_i - \hat{\mathbf{M}}_j\|^2) + \frac{1}{2N_\theta N_z} \sum_{i=1}^{N_\theta N_z} (\min_j \|\hat{\mathbf{M}}_i - \mathbf{M}_j\|^2)}. \tag{6.36}$$

6.5.4 Strain Measures from Clinical Triplane Data

After establishing the range of error associated with measuring regional LV deformation parameters from the synthetic triplane data set, the triplane recordings from the DSE12 data set were examined using this method to investigate the prognostic value of the derived features in patients with LFLG AS. The following section details the data processing pipeline that is applied to the clinical data to calculate LV deformation measures.

First, the triplane DICOM files of the LV were exported to three separate NIfTI image series based on the different acquisition angles. Since in the DICOM format Bidgood et al. (1997), the files contain both pixel data and metadata in the form of tags, these tags were considered to determine the transformation matrix needed to interpret the DICOM images in 3D space. The following tag data were used for the reconstruction of the LV in the physical space:

- The Region Location Min X0, Region Location Min Y0, Region Location Max X1 and Region Location Max Y1 tags indicating the offsets in pixel indices with respect to the upper left corner of the entire image. This was used to extract the appropriate sequence regions from DICOM file as

$$I_{seq}[i, j] = I_{DICOM}[X1 - X0 + i, Y1 - Y0 + j]. \tag{6.37}$$

- PhysicalDeltaX and PhysicalDeltaY defining the width and height of each pixel with p_x and p_y. This defines the parameters for the scaling vector as $\mathbf{s}_{DICOM} = (p_x, p_y, p_y)$. Since no spacing information is given for the z-axis for a 2D image, and $p_x \approx p_y$ for all planes, s_z was approximated by p_y.
- The units of the above tags, specified by PhysicalUnitsXDirection and PhysicalUnitsXDirection respectively. This allowed for quantifying the units of the displacements.

The affine matrix for the extracted image sequences at angles $\alpha = 0°, 60°, 120°$, was calculated as

$$\mathbf{A}_{DICOM}(\alpha) = \mathbf{R}_z(\alpha)\mathbf{S}(\mathbf{s}_{DICOM})\mathbf{T}(\mathbf{t}_{DICOM}), \tag{6.38}$$

where the translation vector $\mathbf{t}_{DICOM}$ is calculated from the Region Location tags as $\mathbf{t}_{DICOM} = [-\frac{X1-X0}{2}, 0, 0]$, since the symmetric imaging cone was centred along the width of the sequence.

Subsequently, the myocardial contours at end-diastole for each angle were annotated using 3D Slicer. This was achieved by placing landmark points in each LAX plane at the apex and at the boundaries of the basal, mid-cavity, and apical segments defined by the AHA segmentation model. Using a Catmull-Rom interpolation as previously described through these landmark points to generate the LAX curves resulted in intermediate meshes with N_z points.

Next, a deformable image registration was performed in each plane between the consecutive time frames, resulting in a displacement field at each time step. This is denoted by $\mathbf{u}_\alpha(x)_t$. The intermediate LAX meshes were propagated via these deformation fields through one cardiac cycle.

Finally, the warped mesh series at each plane was placed in the same physical space using the affine matrix $\mathbf{A}_{DICOM}(\alpha)$, calculated from the DICOM tag data. Utilising a single apex and performing a Catmull-Rom interpolation in the SAX direction via the respective LAX control points within each plane yielded 3D meshes at every time interval. The generation of triangular elements and the calculation of strain directions was performed similarly to the method described in the previous section. To avoid unrealistic interpolation curves around the apex resulting from a set of closely clustered control points, an additional mesh offset parameter was defined that truncated the LAX interpolation by N_{offset} points. The overall pipeline is shown in Figure 6.9.

Comparing regional displacements and strains between different AS severity groups and stress states allows for the assessment of the clinical utility of these markers and the efficacy of DSE.

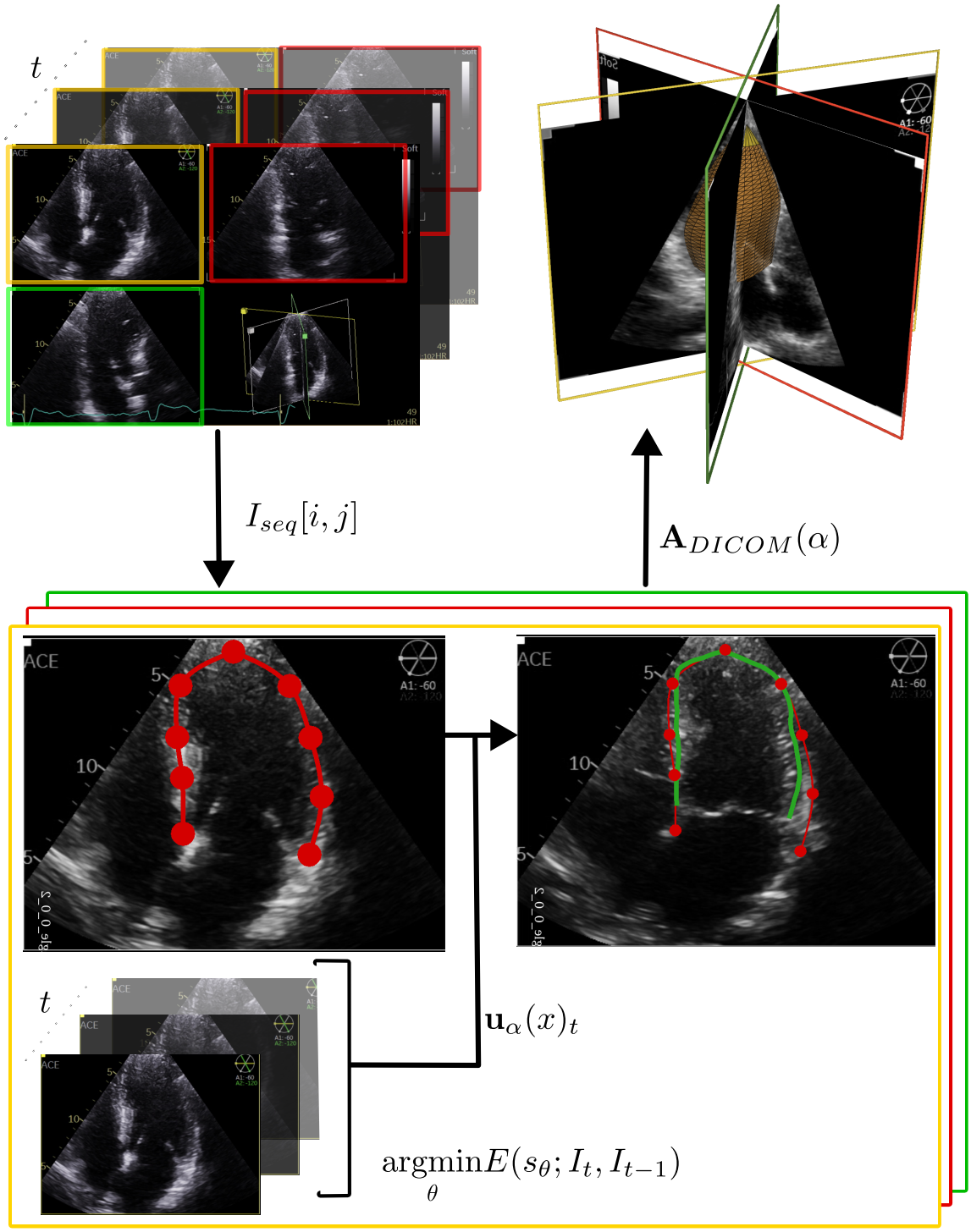


Figure 6.9: The developed pipeline showing the steps needed to acquire a triangular surface mesh series following the myocardial midline of the left ventricle from a triplane DICOM cine series. $I_{\text{seq}}[i,j]$ uses the DICOM image data and tags to split the recording into three separate cine series. $\mathbf{A}_{\text{DICOM}}(\alpha)$ denotes the angle-dependant spatial transformations required to assemble the different mesh series in the same physical space.

6.6 Results and Discussion

In this section, the main results for the strain imaging methodology are detailed. First, strain verification is covered, followed by the validation of the triplane methodology in deriving regional deformation parameters compared to the standard 3D method, using the MWMA dataset. Finally, an exploratory clinical study is presented that examines the regional strain parameters of patients with LFLG AS for different severity groups of AS using the triplane methodology in a subset of the DSE12 data set.

6.6.1 Verification of Strain Measures

To estimate the accuracy of the strain calculations, first a known deformation was applied to a cylindrical reference geometry and the element-wise stretch values in the longitudinal, circumferential and radial directions were compared to their theoretical values. A cylinder with radius $r = 12.5$ and height $h = 102$ was chosen as the undeformed body in the reference configuration. A function was applied to these coordinates so that the radius would increase linearly along the height of the cylinder, increasing in the z direction, leaving the top undeformed and dilating the base with a scaling factor $s = 11$. The two configurations are illustrated in Figure 6.10.

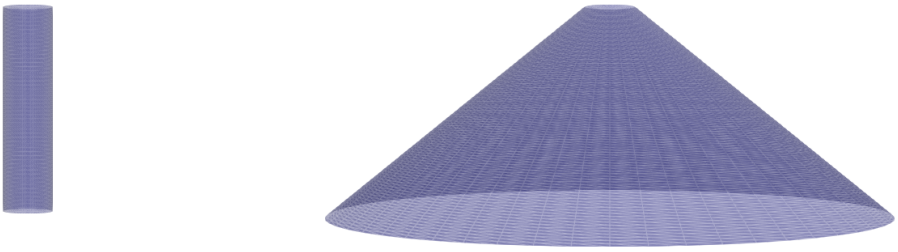


Figure 6.10: The triangular surface mesh of cylindrical geometry as shown in the reference configuration [left] and as depicted in the deformed configuration [right]. The pointwise correspondance between the two meshes allowed the verification of the strain calculations.

After applying the deformation to the reference geometry, the stretch directions were calculated for each triangular surface element. The vectors in the longitudinal, circumferential and radial directions for each element are shown in Figure 6.11 illustrating the top part of the deformed cylinder.

To obtain the theoretical values for the stretches along these directions, the transformation of an individual triangle from its reference configuration to its deformed configuration was examined. Since the element was a surface element, there was no change in length in the radial direction, and the stretch remained 1 for all elements. The circumferential stretch corresponded to the change in length of the edges of the triangular elements lying on the XY plane; hence, this was equal to the change of radius and was a function of the z coordinate increasing towards the base of the cylinder.

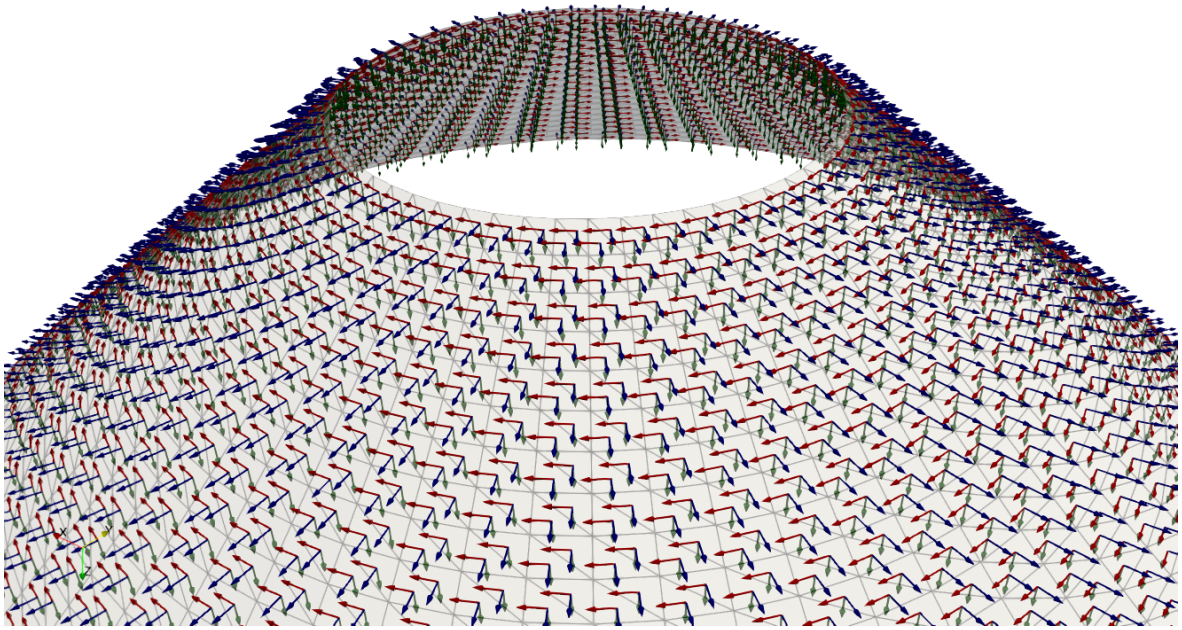


Figure 6.11: Visualisation of the local strain directions in the deformed configuration, showing the longitudinal , circumferential, and radial directions across the mesh surface in blue, red and green, respectively.

The longitudinal stretch was given by

$$\sqrt{1 + (\frac{(s-1)r}{h})^2} = 1.5817\,,$$

which expresses the length of a line element along the surface of the cylinder and only depends on the height, the radius, and the scaling of the deformed geometry. The directional stretch values are shown in Figure 6.12 using a polar plot representation.

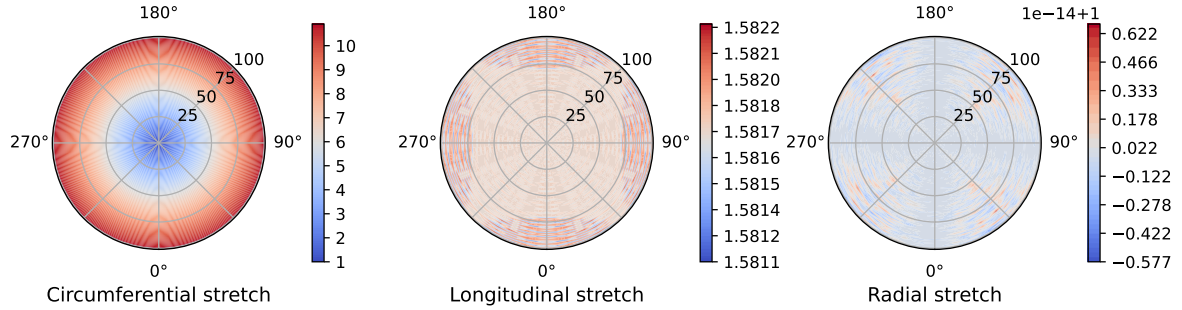


Figure 6.12: Stretch values in the circumferential [left], longitudinal [centre], and radial directions [right] of the reference geometry as depicted on a polar plot. Each plot shows the spatial variations in directional stretch across the mesh surface, where the angular position represents the circumferential location, while the radial distances correspond to the height measured from the top of the cylinder.

The radial stretch only produces numerical errors in the range of 10^{-14} compared to its theoretical value of 1. The circumferential stretch accurately aligns with the variation in radius as one progresses from the base to the top of the cylinder, and the longitudinal stretch is closely aligned with the theoretical value, with an error approximately equal to 5×10^{-4}. This is caused by the slightly incorrect vector directions when calculating the circumferential

direction as edge vectors instead of their analytical values; however, this was preferred due to it being used later in the LV geometry calculations as well, which did not have an analytical solution.

6.6.2 3D and Triplane Measurements

After establishing the error associated with the strain calculations, the tracking accuracy of the image registration process was investigated. This was done by calculating the symmetric mean boundary distance and Hausdorff distance metrics as well as the corresponding points RMSE for each patient at every time frame. Since each patient had a different number of frames for a full cardiac cycle, the time was normalised across patients using linear interpolation. Reporting the mean and standard deviation at each normalised time step for the distance and volume metrics, as shown in Figure 6.13, allowed for the in-depth analysis of the tracking accuracy over time.

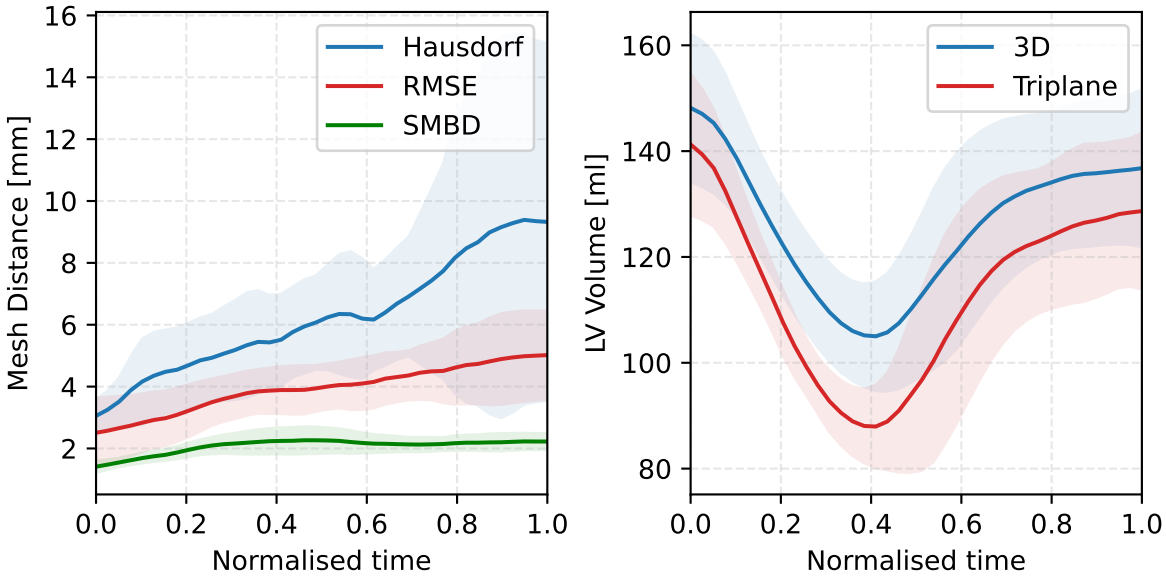


Figure 6.13: The comparison of global mesh distance metrics [left] and left ventricular (LV) volumes [right] over the normalised cardiac cycle between 3D and triplane methods. The left figure displays the symmetric Hausdorff distance, th pointwise root mean square error (RMSE), and the symmetric mean boundary distance (SMBD), indicating the temporal variation in mesh distance. The right figure displays mean LV volume curves from 3D and triplane reconstructions. The shaded regions represent the standard deviations across all patients.

Although mesh distance metrics at the beginning of the cardiac cycle are small, they become more prominent towards the end of the cardiac cycle due to the triplane image registration process providing a much tighter tracking of the LV wall, when compared to the 3D method. Note that SMBD stays mostly constant throughout the cardiac cycle, suggesting that the overall mesh distance is constant. However, the Hausdorff metric increases significantly towards the end of the cardiac cycle, which indicates that an outlier region causes the meshes to be distant from each other. This outlier is most likely caused by tracking of the mitral valve at closure, as it detaches from the LV wall and returns to a closed state rather than keeping

the region to follow the LV myocardium. This tracking error is likely due to the fact that the regularisation applied during the triplane registration process was too low, causing the displacement field to be affected by sharp and localised changes.

This difference in tracking performance is also reflected in the volume curves. Initially, the two methods produce similar volumes, but diverge near peak systole. The triplane image registration method provides tighter tracking due to its reduced spatial dimensionality, leading to lower estimated volumes at peak systole compared to the 3D method. Toward the end of the cardiac cycle, the volume difference decreases again, indicating that the meshes are generally well aligned aside from a few local outliers detected by the Hausdorff metric and the effects of differing regularisation levels.

Figure 6.14 illustrates the average global longitudinal and circumferential strains over normalised time intervals across all patients in the MWMA data set. A similar pattern is seen for the global circumferential strain as observed for the volume curves, in which initial strain values agree, while they begin to diverge towards peak systole and return to closer alignment by the end of the cardiac cycle. For global longitudinal strain, the two methods show great alignment until peak systole and diverge significantly at the end of the cardiac cycle. This phenomenon may also be explained by the erroneous tracking of the mitral valve, which causes a small region to undergo a large deformation, thus increasing the average strain, which consequently results in a greater discrepancy between the two methods.

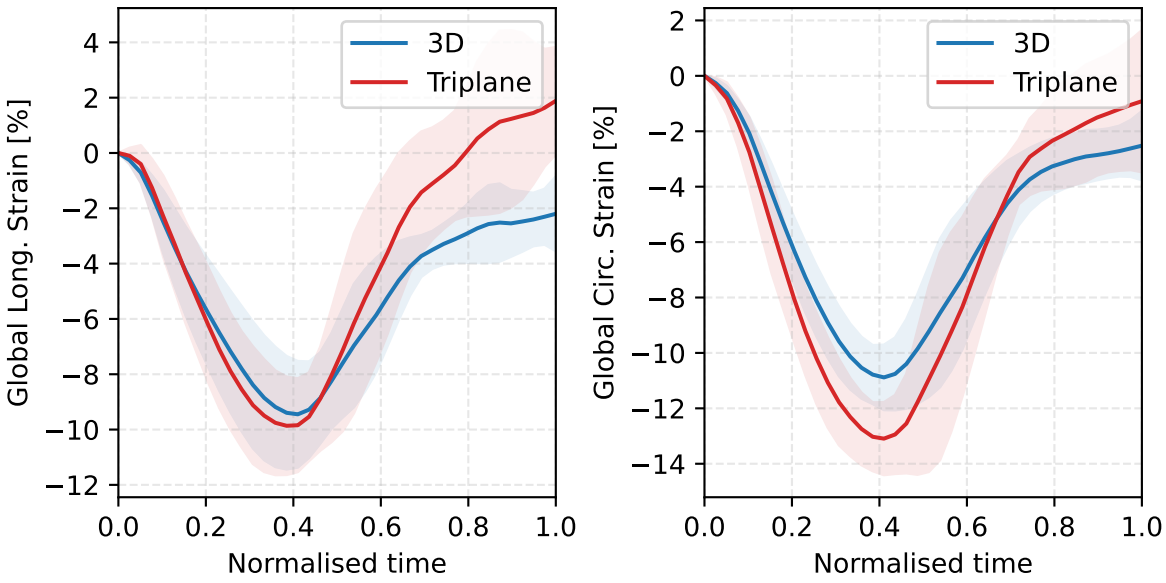


Figure 6.14: The comparison of global longitudinal [left] and circumferential [right] strain values across the normalised cardiac cycle for all patients, derived using 3D and triplane methods. Solid lines represent the mean global strain across all patients, while shaded areas indicate one standard deviation.

To further evaluate the differences between the two methods, a comparison of global deformation parameters measured at peak systole was conducted. In Figure 6.15, boxplots of the average regional displacement and strain values derived from the 3D and triplane pipelines are depicted for the peak systolic frame. The average global displacement and longitudinal deformation were observed to show good agreement between the two methodologies, while the mean regional circumferential strain values were slightly lower for triplane measurements.

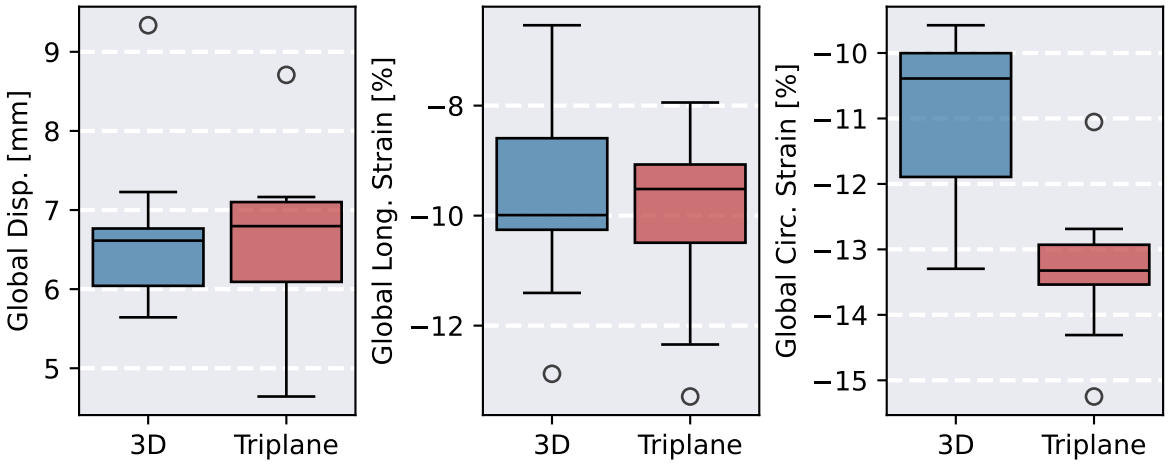


Figure 6.15: Boxplot comparison of average global displacement [left], longitudinal strain (middle), and circumferential strain [right] at peak systole between the 3D and triplane methodologies across all patients in the myocardial wall motion analysis data set. Displacement is reported in millimetres, while strain values are expressed as percentages.

Finally, statistical analysis of regional deformation parameters of the LV myocardium between 3D and triplane derived values at peak systole was performed using a Wilcoxon signed-rank test using Bonferroni correction. The metrics reported for regional displacement, longitudinal and circumferential strains are shown in Tables 6.1-6.3, while the segmental mean difference, RMSE and statistical significance are also shown using a bull's eye plot as illustrated in Figures 6.16-6.18.

From Table 6.1 it is observed that the regional displacement values differed by a maximum of 2.2 mm RMSE. Statistically significant differences were found between the two methods for two regions: one basal (segment 1) and one apical (segment 14). The basal segments are also observed to exhibit larger displacements compared to the apical segments, supporting the findings of Leitman and Tyomkin (2025). Segmental longitudinal strains showed good agreement for all regions, but large deviation with a maximum mean difference of -5.36%. No statistically significant differences were found for this metric. The basal segments are also seen to exhibit lower absolute longitudinal deformations compared to the apical segments. For regional circumferential strains, the two methods showed good agreement with no statistically significant differences for any region. The maximum difference between the two methods for circumferential strain was found to be less significant than longitudinal strain with 6.2% RMSE.

Segment	3D (Mean ± SD)	Triplane (Mean ± SD)	Mean Diff.	RMSE	p-value
1	10.081 ± 1.875	11.893 ± 2.44	-1.812	2.170	0.066
2	12.251 ± 2.966	13.755 ± 2.071	-1.503	2.079	0.465
3	11.407 ± 1.664	10.92 ± 2.783	0.487	2.073	1.000
4	10.261 ± 1.946	10.648 ± 2.186	-0.387	2.051	1.000
5	10.143 ± 2.341	11.188 ± 2.1	-1.045	1.990	1.000
6	9.65 ± 1.845	10.055 ± 2.473	-0.406	1.467	1.000
7	8.943 ± 1.991	9.886 ± 2.106	-0.943	1.450	0.930
8	10.837 ± 2.049	11.619 ± 1.385	-0.782	1.420	1.000
9	9.215 ± 1.219	9.292 ± 2.347	-0.076	1.621	1.000
10	7.436 ± 1.445	7.774 ± 1.601	-0.339	1.254	1.000
11	6.957 ± 1.613	7.6 ± 2.201	-0.643	1.441	1.000
12	7.413 ± 1.854	7.789 ± 1.893	-0.376	1.082	1.000
13	7.644 ± 2.056	8.328 ± 1.832	-0.684	0.876	0.199
14	7.198 ± 0.814	8.548 ± 0.951	-1.349	1.512	0.066
15	4.616 ± 1.415	5.174 ± 1.32	-0.558	0.729	0.332
16	4.636 ± 1.012	4.831 ± 1.191	-0.195	0.740	1.000
17	4.178 ± 0.548	5.01 ± 0.481	-0.832	0.989	0.133

Table 6.1: The comparison of average regional displacements at peak systole between the 3D and triplane methods, using all patients from the myocardial wall motion analysis data set (n = 9). For each of the 17 segments, the table reports the mean and standard deviation for both 3D and triplane methods, the mean difference (3D-Triplane), the root mean square error (RMSE), and the p-value from the Wilcoxon signed-rank test assessing the statistical significance between the two methods. Units for distance metrics are in millimetres.

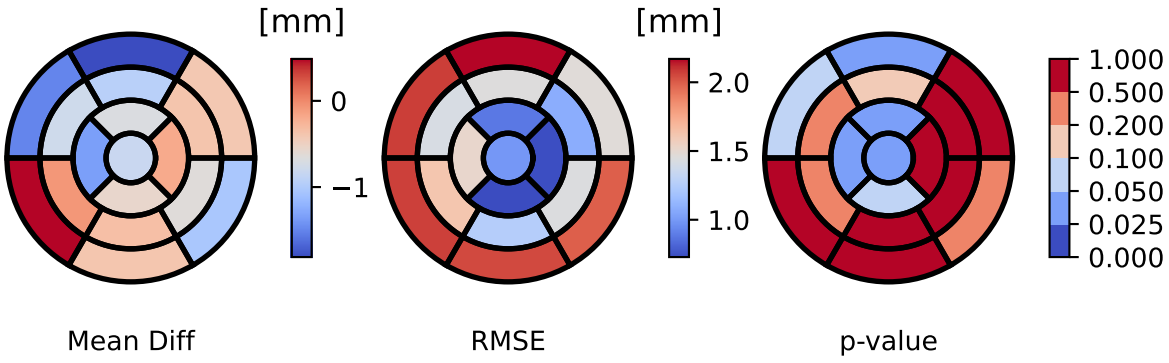


Figure 6.16: Regional displacement comparison between 3D and triplane methodologies at peak systole across all patients in the myocardial wall motion analysis data set, displayed using a 17-segment bull's eye plot. The left figure shows the mean difference, the centre figure displays the root mean square error (RMSE), and the right figure illustrates the corresponding statistical significance with the associated segmental p-values. The inner rings represent apical regions, and the outer rings represent basal segments.

Segment	3D (Mean ± SD)	Triplane (Mean ± SD)	Mean Diff.	RMSE	p-value
1	−3.37 ± 5.49	−4.14 ± 7.74	0.76	5.57	1.00
2	0.81 ± 4.32	1.96 ± 7.56	-1.15	6.60	1.00
3	−2.96 ± 6.12	2.39 ± 14.70	-5.36	13.42	1.00
4	−5.82 ± 3.43	−5.24 ± 9.23	-0.59	8.51	1.00
5	−9.50 ± 4.69	−12.25 ± 8.56	2.75	6.61	1.00
6	−9.82 ± 4.42	−12.33 ± 7.61	2.51	5.45	1.00
7	−6.84 ± 4.45	−7.03 ± 3.19	0.18	4.51	1.00
8	−13.17 ± 3.40	−12.56 ± 2.45	-0.60	2.76	1.00
9	−13.74 ± 2.90	−12.02 ± 3.38	-1.72	3.77	1.00
10	−12.42 ± 2.51	−13.25 ± 5.47	0.83	5.86	1.00
11	−11.79 ± 4.12	−13.88 ± 4.67	2.09	5.10	1.00
12	−10.03 ± 3.47	−7.44 ± 6.41	-2.59	6.99	1.00
13	−11.46 ± 2.52	−14.30 ± 3.39	2.85	4.74	0.93
14	−14.69 ± 2.45	−17.38 ± 3.96	2.69	5.06	1.00
15	−12.38 ± 2.70	−11.18 ± 5.08	-1.20	4.74	1.00
16	−11.46 ± 4.14	−13.25 ± 4.69	1.79	5.85	1.00
17	−10.42 ± 1.46	−11.27 ± 2.85	0.85	4.93	1.00

Table 6.2: The comparison of average regional longitudinal strain at peak systole between the 3D and triplane methods, using all patients from the myocardial wall motion analysis data set (n = 9). For each of the 17 segments, the table reports the mean and standard deviation for both 3D and triplane methods, the mean difference (3D–Triplane), the root mean square error (RMSE), and the p-value from the Wilcoxon signed-rank test assessing the statistical significance between the two methods. Units for strain metrics are in percentages.

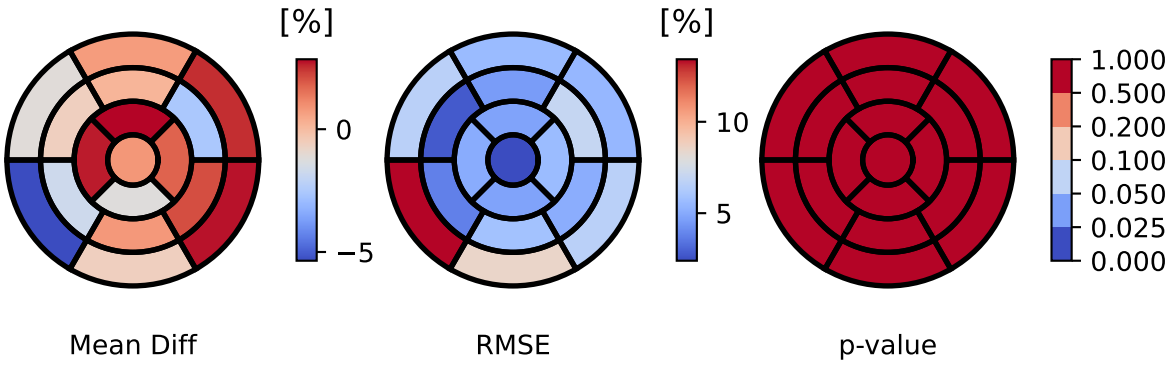


Figure 6.17: Regional longitudinal strain comparison between 3D and triplane methodologies at peak systole across all patients in the myocardial wall motion analysis data set, displayed using a 17-segment bull's eye plot. The left figure shows the mean difference, the centre figure displays the root mean square error (RMSE), and the right figure illustrates the corresponding statistical significance with the associated segmental p-values.

Segment	3D (Mean ± SD)	Triplane (Mean ± SD)	Mean Diff.	RMSE	p-value
1	−8.91 ± 2.32	−9.91 ± 1.89	1.00	2.68	1.00
2	−9.34 ± 2.47	−11.36 ± 1.84	2.02	4.13	1.00
3	−7.34 ± 2.48	−11.36 ± 1.56	4.02	4.94	0.07
4	−5.66 ± 3.87	−9.49 ± 3.03	3.83	6.17	1.00
5	−9.93 ± 2.66	−12.09 ± 1.38	2.17	3.59	0.93
6	−12.32 ± 3.24	−10.57 ± 1.79	-1.75	3.79	1.00
7	−11.50 ± 2.89	−13.59 ± 1.65	2.09	3.60	1.00
8	−12.87 ± 2.60	−14.41 ± 2.02	1.55	3.27	1.00
9	−11.27 ± 1.50	−14.04 ± 1.27	2.77	3.03	0.07
10	−9.89 ± 3.05	−13.35 ± 2.87	3.47	5.48	1.00
11	−10.44 ± 4.32	−14.31 ± 1.28	3.86	5.45	0.93
12	−12.26 ± 2.48	−13.24 ± 2.04	0.97	3.23	1.00
13	−13.94 ± 2.49	−14.61 ± 1.32	0.67	2.60	1.00
14	−12.28 ± 2.30	−14.55 ± 1.52	2.26	3.46	0.46
15	−9.99 ± 2.58	−12.81 ± 1.91	2.82	3.73	0.33
16	−11.56 ± 3.35	−14.43 ± 1.12	2.87	4.39	1.00
17	−12.33 ± 2.84	−15.15 ± 3.20	2.82	4.26	0.93

Table 6.3: The comparison of average regional circumferentail strain at peak systole between the 3D and triplane methods, using all patients from the myocardial wall motion analysis data set (n = 9). For each of the 17 segments, the table reports the mean and standard deviation for both 3D and triplane methods, the mean difference (3D-Triplane), the root mean square error (RMSE), and the p-value from the Wilcoxon signed-rank test assessing the statistical significance between the two methods. Units for strain metrics are in percentages.

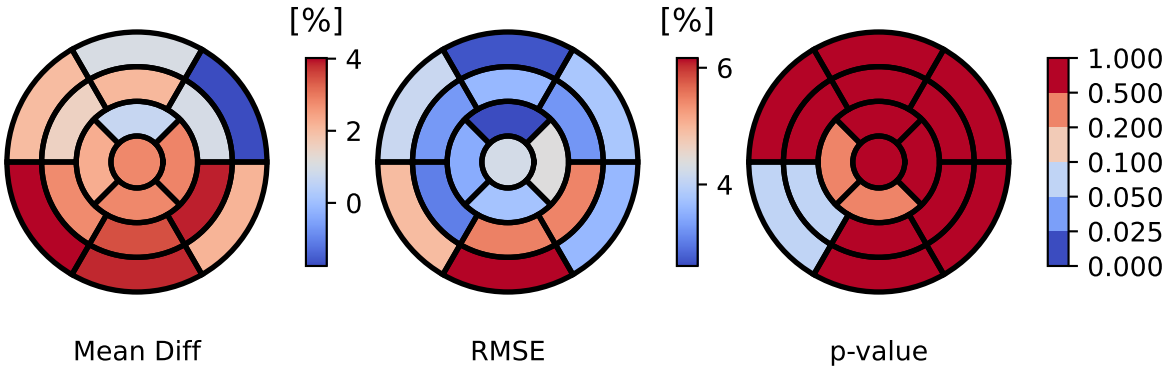


Figure 6.18: Regional circumferential strain comparison between 3D and triplane methodologies at peak systole across all patients in the myocardial wall motion analysis data set, displayed using a 17-segment bull's eye plot. The left figure shows the mean difference, the centre figure displays the root mean square error (RMSE), and the right figure illustrates the corresponding statistical significance with the associated segmental p-values.

6.6.3 Regional Strain Assessment in Low-Flow, Low-Gradient Aortic Stenosis

This section presents an analysis of regional deformation metrics derived from triplane echocardiograms of patients included in the DSE12 data set. First, the different severity groups were compared at resting heart rate, followed by an intrapatient examination of the differences between stress and resting states.

Regional Differences at Baseline in Moderate and Severe Aortic Stenosis

To investigate the clinical relevance of stress echocardiography concerning LV function, the differences among AS severity groups are initially assessed at the baseline resting heart rate. By analysing the mean and standard deviation for each group for every AHA segment, as depicted in Figure 6.19, the reference values for each group are established. Although longitudinal strain does not clearly show a difference between AS groups, displacement and circumferential strains indicate greater LV contractility within the moderate AS group. Displacement values also correspond to greater movements in the basal and middle segments for all severity groups compared to the apical segments, which is consistent with the expected motion of the LV (Leitman and Tyomkin, 2025). Circumferential strains for the basal segments in moderate AS are also observed to have lower absolute values compared to the middle and apical segments, a finding previously demonstrated by Hjertaas et al. (2023).

Intrapatient Stress Response Analysis

To conclude the clinical study, the deformation parameters were compared between the baseline and stress states to assess regional LV contractile reserve for each patient. The results, shown in Figure 6.6.3, report the differences per-region for displacement, longitudinal, and circumferential strain. Patient 2 exhibited a counter effect of an expected stress test, with lower magnitudes of longitudinal and circumferential strain during DSE than at rest, possibly due to inadequate image acquisition or poor image tracking. Patient 1, who had moderate AS, indicated greater longitudinal strain reserves compared to patient 3, who had severe AS. This result may reinforce the finding of Kim et al. (2020), in which a greater longitudinal strain reserve was associated with better survival for patients with LFLG AS.

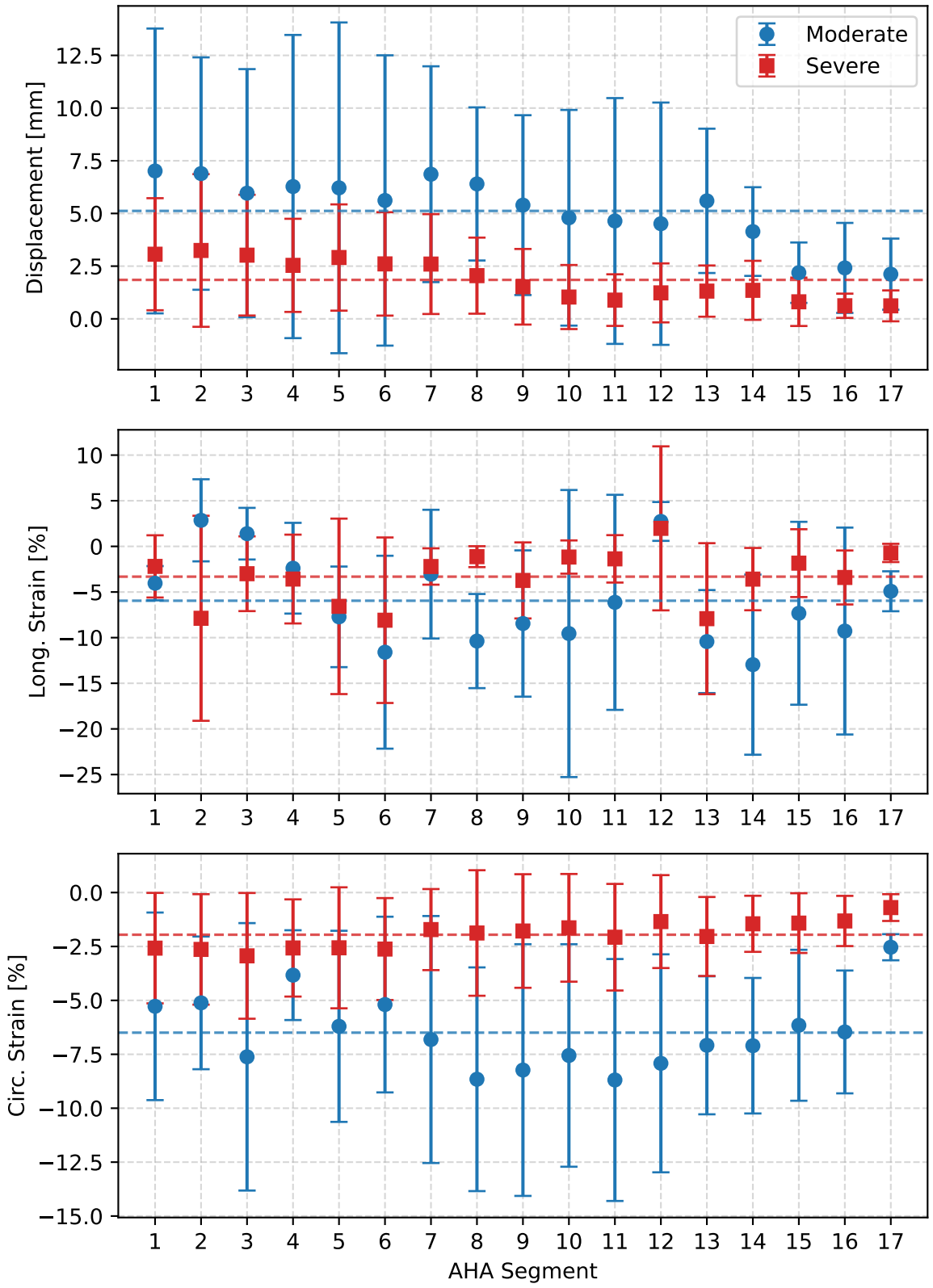


Figure 6.19: Segmental analysis of left ventricular deformation parameters at peak systole in patients with moderate (blue circles) and severe (red squares) aortic stenosis (AS), across all 17 standard left ventricular segments. The top figure shows myocardial displacement [mm], the middle figure shows longitudinal strain [%], and the bottom figure shows circumferential strain [%]. The dashed lines represent group-wise means for each parameter. Error bars denote one standard deviation.

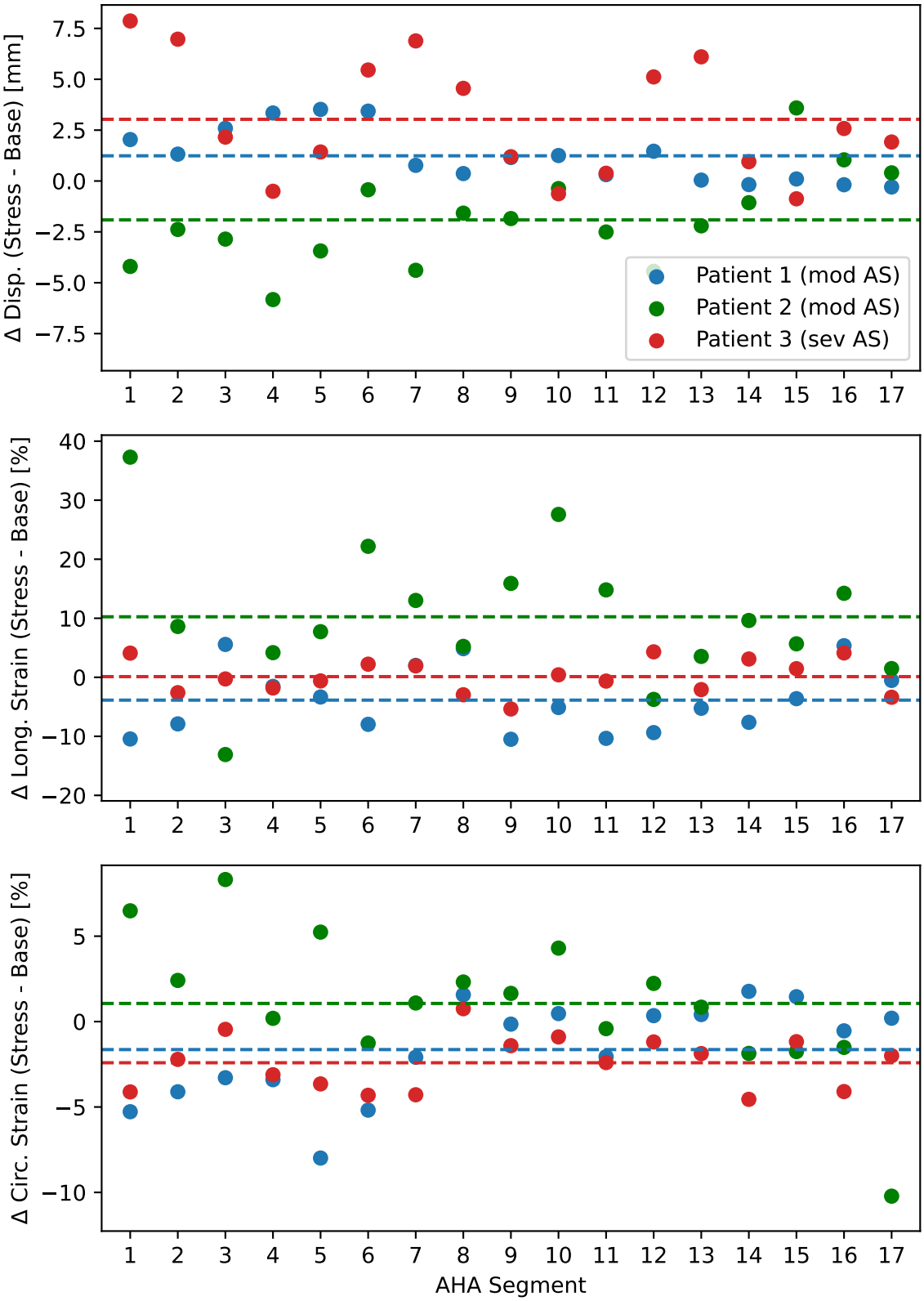


Figure 6.20: Segmental assessment of left ventricular (LV) contractile reserve in three patients with low-flow, low-gradient aortic stenosis during dobutamine stress echocardiography, compared to baseline (resting) measurements. Differences in peak systolic regional displacement [top], longitudinal strain [middle], and circumferential strain [bottom] are plotted per LV segment. Negative values for longitudinal and circumferential strain indicate increased contractility in response to stress. Dashed lines represent the mean change across all segments for each patient.

6.7 Conclusion

In this chapter, the development of a multiplanar reconstruction and image registration pipeline was presented to estimate the regional parameters of the LV myocardial deformation from triplane echocardiographic recordings. The validation results against full-volume 3D STE demonstrated good agreement for segmental longitudinal and circumferential strains as well as global longitudinal strain, confirming the accuracy and clinical viability of the proposed method. The proposed pipeline enables accurate quantification of 3D LV deformation parameters from triplane recordings, eliminating the need for full-volume 3D acquisitions and thus supporting clinical applicability in more challenging imaging protocols, such as DSE. In addition, the pipeline can also allow retrospective analysis of triplane data, which offers a valuable tool for investigating regional myocardial mechanics in various cardiac pathologies.

Although the sample size was limited, the initial results suggested that the segmental circumferential strain values measured at rest are consistently lower for patients with severe AS compared to those with moderate AS. Furthermore, patients with moderate AS exhibited greater myocardial displacement at rest across all LV segments than those with moderate AS, supporting the differentiation of severe and pseudo-severe cases of LFLG AS at rest. Finally, segmental longitudinal strain reserve may also serve as a clinically relevant marker for AS severity in classical LFLG AS.

7. Deep Learning Approaches to Aortic Stenosis Prediction

This chapter concerns the application of DL models for the prediction of the severity of AS from TTE recordings. After a brief introduction (Section 7.1) and a theoretical background on DL (Section 7.2) the state-of-the-art architectures related to AS prediction from TTE are covered in Section 7.3. Next, the methodology for training, evaluating, and conducting a comprehensive analysis of a novel architecture is covered in Section 7.4. Consequently, the results comparing the training and generalisation performance of the novel architecture with the state-of-the-art methods on the public TMED2 data set are described in Section 7.5. Finally, the generalisation performance of the models is investigated in an external clinical data set of concordant AS and a cohort of patients with LFLG AS recorded at rest and stress conditions as described in Section 7.6. The conclusions of this chapter are summarised in Section 7.7.

7.1 Introduction

Previous chapters have focused on investigating ML algorithms using manually extracted indicators of AS severity, such as LVEF and AVA, as well as employing image registration and multiplanar reconstruction techniques to quantify LV myocardial wall motion. Although both methods provide valuable physiological insight, they rely on clinical expertise to identify relevant predictors of AS within echocardiographic recordings, thus being limited by current clinical understanding. In contrast, DL employs a data-driven approach in which class-discriminative representations are learnt directly from image data.

Using recent advances in DL, such as attention mechanisms, multiple-instance learning (MIL), and multitask learning frameworks, DL models have shown improved performance compared to previous architectures in a variety of learning tasks. These tasks included medical image segmentation (Tragakis et al., 2023), AS classification using TTE recordings (Huang et al., 2024), or the automatic prediction of multiple anatomical and hemodynamic parameters from echocardiographic recordings (Holste et al., 2025). Although AS classification using TTE recordings has been demonstrated in several studies in recent years (Ahmadi et al., 2024; Huang et al., 2025; Park et al., 2025), all models were trained with the exclusion of the discordant AS population. Consequently, current data on the generalisation performance of these models in the LFLG AS cohort remain sparse, with no evidence on the utility of stress

testing or their discriminative capability between severe and pseudo-severe classical LFLG AS.

In this chapter, a novel DL architecture is introduced that combines transformer-based encoder blocks with the MIL framework to predict the severity of AS. The model is compared to existing approaches using the public TMED2 data set. Generalisation performance across hospitals is assessed using an external cohort of patients with concordant AS. A distinct LFLG AS group recorded under both resting and stress conditions is used to examine how well the learnt classifiers identify the severity of AS in this subgroup, as well as to infer the utility of stress testing in disease stratification.

7.2 Deep Learning Foundations

7.2.1 Learning Algorithm

The most comprehensive definition of machine learning, as described by Mitchell (1997), is the study of algorithms that can learn from data. This can be formally expressed as learning algorithms that improve their performance on some task through experience, fully characterising the learning problem.

In the context of computer vision, the task typically corresponds to problems such as image classification, segmentation, or image-based regression. The objective of image classification is to accurately specify a categorical output variable, $y \in \{1, .., k\}$, which corresponds to the provided input image $X \in \mathbb{R}^{H \times W}$. The task is then achieved by learning a mapping $f : \mathbb{R}^{H \times W} \rightarrow \{1, \ldots, k\}$. The task of image segmentation is to assign a categorical label to every pixel in the image so that the mapping learnt is represented as $f : \mathbb{R}^{H \times W} \rightarrow \{1, \ldots, k\}^{H \times W}$, while in regression the task involves predicting a continuous output variable from the input image such as the peak transvalvular velocity or ESV.

The experience of a learning algorithm is the data set it encounters during training. A common way to describe a data set is with a design matrix containing a different observation in each row and a different feature in each column of the matrix. In cases where the input examples do not have the same dimensions, rather than describing the data set as a matrix, it is described as a collection of input and output pairs, $S = \{(X^{(i)}, y^{(i)}); i = 1, .., n\}$. A single instance of an input and output pair is also called a training example, and each example is used to guide the algorithm to improve its predictions. In medical imaging data sets, an input example $X^{(i)}$ may correspond to a single image or video, or a set of images and videos, while the labels are defined through the learning objective.

The performance metric provides a quantitative measure of how well the learnt function

f performs on the given task. For a classification task, the most common metrics include accuracy, recall, precision, balanced accuracy, and the F1 score as defined earlier in Chapter 5 through Equations 5.23 - 5.27. Additionally, the confusion matrix can be used to summarise the predictions of the algorithm in a tabular format. By presenting the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), it provides a more detailed insight of the algorithms predictions, identifying areas of misclassification. Moreover, the Receiver Operating Characteristic (ROC) curve is often used to assess the efficacy of a binary classification algorithm by examining its performance via the true positive rate (TPR) and the false positive rate (FPR) defined as

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad ; \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \tag{7.1}$$

Every point on the ROC curve corresponds to a distinct classification threshold, which depicts the trade-off between sensitivity and specificity at each threshold level. The area under the curve (AUC) is then calculated from these values to provide a scalar measure of the algorithm's discriminative capability, where an AUC of 1.0 denotes a perfect classifier and 0.5 reflects performance equivalent to random guessing.

For image segmentation, common performance metrics include the dice similarity coefficient (DSC) and intersection over union (IoU). These metrics are measured between the ground truth annotation (A) and the predicted label map (B), as

$$\text{DSC} = \frac{2\,|A \cap B|}{|A| + |B|}, \quad \text{and} \quad \text{IoU} = \frac{|A \cap B|}{|A \cup B|}, \tag{7.2}$$

where the size of each set is determined by the number of pixels contained in the set. For regression tasks, performance metrics include the mean squared error (MSE) and the coefficient of determination (R^2) defined as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^{(i)})^2, \quad \text{and} \quad R^2 = 1 - \frac{\sum_i (\hat{y}^{(i)} - y^{(i)})^2}{\sum_i (y^{(i)} - \bar{y}^{(i)})^2}, \tag{7.3}$$

where $\hat{y}^{(i)}$ and $y^{(i)}$ denote the predicted and ground truth outputs for the given example, respectively, while $\bar{y}^{(i)}$ is the mean of the observed data.

7.2.2 Maximum Likelihood Estimate

One of the most common ways learning algorithms are derived involves the use of the maximum likelihood estimate (MLE) principle introduced by Fisher (1922). This principle assumes that given a set of training examples $\mathbb{X} = \{\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}\}$ drawn independently from an unknown but fixed data-generating distribution $p_{\text{data}}(\mathbf{X})$, the true data distribution can be approximated by a model, represented as a family of parametric probability distributions $p_{\text{model}}(\mathbf{X}; \boldsymbol{\theta})$, defined over the same space. The MLE formulation then aims to find the optimal solution for $\boldsymbol{\theta}$ that maximises the probability of the observed data under the model,

so that

$$\boldsymbol{\theta}_{\mathrm{ML}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p_{\mathrm{model}}(\mathbb{X}, \boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \prod_{i=1}^{n} p_{\mathrm{model}}(\mathbf{X}^{(i)}; \boldsymbol{\theta}) \quad , \tag{7.4}$$

which equals the product of the likelihood of each example, since the examples are assumed to be independent. For better numerical stability, the above equation is implemented in software as the sum over the logarithm of each likelihood, so that

$$\boldsymbol{\theta}_{\mathrm{ML}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{n} \log p_{\mathrm{model}}(\mathbf{X}^{(i)}; \boldsymbol{\theta}). \tag{7.5}$$

Since the argument does not change when the expression is rescaled, Equation 7.5, can be rewritten to be expressed as an expectation with respect to the observed distribution $\hat{p}_{\mathrm{data}}$, given the training data, such that

$$\boldsymbol{\theta}_{\mathrm{ML}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \log p_{\mathrm{model}}(\mathbf{X}^{(i)}; \boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathbb{E}_{\mathbf{X} \sim \hat{p}_{\mathrm{data}}} \log p_{\mathrm{model}}(\mathbf{X}; \boldsymbol{\theta}). \tag{7.6}$$

This formulation can also be viewed as a method that minimises the dissimilarity between the empirical data distribution $\hat{p}_{\mathrm{data}}$ and the model distribution p_{model}, with the degree of dissimilarity measured by the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) defined as

$$D_{\mathrm{KL}}(\hat{p}_{\mathrm{data}} || p_{\mathrm{model}}) = \mathbb{E}_{\mathbf{X} \sim \hat{p}_{\mathrm{data}}}(\log \hat{p}_{\mathrm{data}}(\mathbf{X}) - \log p_{\mathrm{model}}(\mathbf{X}; \boldsymbol{\theta})). \tag{7.7}$$

Therefore, minimising the KL divergence between the two distributions is equivalent to the MLE formulation since $\hat{p}_{\mathrm{data}}$ does not depend on the model parameters. Minimising the KL divergence is also equivalent to minimising the cross-entropy (CE) in the context of a supervised learning task, where $\hat{p}_{\mathrm{data}}$ is given, since

$$\mathrm{CE} = H(\hat{p}_{\mathrm{data}}, p_{\mathrm{model}}) = H(\hat{p}_{\mathrm{data}}) + D_{\mathrm{KL}}(\hat{p}_{\mathrm{data}} || p_{\mathrm{model}}), \tag{7.8}$$

where $H(\hat{p}_{\mathrm{data}})$, the entropy of the data distribution is fixed for a set of training examples.

For a supervised classification task with k number of class labels, the output variable is often represented as a one-hot encoded vector $y^{(i)} \in \{0, 1\}^k$. This vector takes the value of 1 for the true class label and is 0 everywhere else. In this setting the cross-entropy becomes equivalent to the negative log-likelihood expressed as

$$\mathcal{L}_{CE}(\mathbf{X}^{(i)}, y^{(i)}, \boldsymbol{\theta}) = \mathcal{L}_{NLL}(\mathbf{X}^{(i)}, y^{(i)}, \boldsymbol{\theta}) = -\sum_{j=1}^{k} y_j^{(i)} \log \hat{y}_j^{(i)}, \tag{7.9}$$

where $\hat{y}_j^{(i)} = p_{\mathrm{model}}(y = j | \mathbf{X}^{(i)}; \boldsymbol{\theta})$ is the model's predicted probability of class j for input $\mathbf{X}^{(i)}$, and $y_j^{(i)}$ is the true probability of class j. Note that this expression simplifies to the predicted negative log-probability of the true class label due to the one-hot encoded vector being 0 everywhere else.

7.2.3 Gradient-Based Optimisation

The above section outlined how finding the optimal model parameters via MLE can be obtained through minimising the cross-entropy between the observed data and the predicted model distribution. This section describes how gradient-based optimisation is used to guide the model parameters to achieve this minimisation objective. First the objective function or cost function to be minimised $J(\boldsymbol{\theta})$ is expressed as the average of per-example losses over the training data, such that

$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_{CE}(\mathbf{X}^{(i)}, y^{(i)}, \boldsymbol{\theta}) \quad , \tag{7.10}$$

where n is the number of training examples. Next, the derivative of the cost function is calculated with respect to the model parameters $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$ to provide information on the slope of the cost function. The derivative of the cost function is expressed as the average of the per-example derivatives of the individual losses, so that

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{CE}(\mathbf{X}^{(i)}, y^{(i)}, \boldsymbol{\theta}). \tag{7.11}$$

Note that a slope of 0 indicates that the parameters have settled in local or global minima or a saddle point, whereas non-zero gradients indicate that the cost function can be further decreased by changing the model parameters in the opposite direction of the gradient. This constitutes the next step of gradient-based optimisation, in which the model parameters are updated, such that

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon \nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \quad , \tag{7.12}$$

where ϵ denotes the learning rate, i.e. the magnitude of the step taken in the opposite direction of the gradient. This approach is referred to as steepest descent or gradient descent.

However, the steepest descent involves iterating over all training examples to perform a single update of the parameters. This is often not feasible due to the computational resource requirements of storing all gradients for all of the training data. To overcome this limitation, the gradient is approximated by a subset or batch of training examples drawn uniformly from the training data $\mathbb{B} = \{\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n')}\}$, so that

$$\mathbf{g} = \frac{1}{n'} \nabla_{\boldsymbol{\theta}} \sum_{i=1}^{n'} \mathcal{L}_{CE}(\mathbf{X}^{(i)}, y^{(i)}, \boldsymbol{\theta}), \tag{7.13}$$

and the parameters are updated via

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon \mathbf{g}. \tag{7.14}$$

This method is called Stochastic Gradient Descent (SGD) as the gradient is approximated from a randomly drawn subset of training examples. There are several extensions to SGD that improve its convergence stability. The most well established method is called Adaptive Moment Estimation (Adam) (Kingma and Ba, 2014), which combines adaptive learning rates with momentum-based techniques that use previous gradients to perform a parameter update.

7.2.4 Feedforward Neural Network

Feedforward neural networks, also called multilayer perceptrons (MLP), constitute the foundation of many deep learning models. MLPs approximate a mapping between input and output variables, by stacking many nodes, or neurons, in a layered structure, as shown in Figure 7.1. Each node within a layer receives input from all nodes in the previous layer, calculates a weighted sum of these inputs through a linear transformation, and then applies a nonlinear activation function ϕ to generate its output.

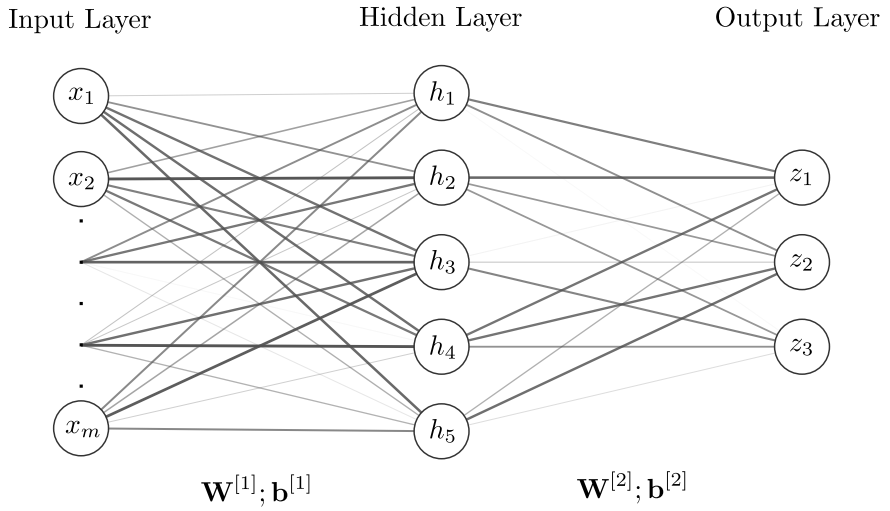


Figure 7.1: A fully connected feedforward neural network with m input units, 5 hidden units, and 3 output units. The input layer $(x_1, \ldots, x_m)$ is connected to the hidden layer $(h_1, \ldots, h_5)$ via the weight matrix $\mathbf{W}^{[1]}$ and bias vector $\mathbf{b}^{[1]}$, followed by the output layer (z_1, z_2, z_3), computed using the weight matrix $\mathbf{W}^{[2]}$ and bias vector $\mathbf{b}^{[2]}$. Information flows from left to right.

The nodes in the input layer represent the input features the model is supposed to process (i.e. individual pixels for an image), while each subsequent layer calculates its activations through a learnable weight matrix $\mathbf{W}^{[l]} \in \mathbb{R}^{d_l \times d_{l-1}}$ and bias vector $\mathbf{b}^{[l]} \in \mathbb{R}^{d_l}$. Here, l indexes the number of layers in the network and d_l represents the number of nodes in layer l.

Nonlinear activation functions are used on top of the affine transformations, allowing for learning nonlinear relationships between layers, giving the activation of the hidden and output layer for the given example as

$$\mathbf{h} = \phi(\mathbf{W}^{[1]}\mathbf{x} + \mathbf{b}^{[1]}), \quad \text{and} \quad \mathbf{z} = \phi(\mathbf{W}^{[2]}\mathbf{h} + \mathbf{b}^{[2]}). \tag{7.15}$$

Since the output layer is responsible for producing the prediction, the number of output neurons corresponds to the number of possible values the output variable can take, considering a classification task. To generate a valid probability distribution, the softmax function, σ, is used on top of the final activations, which ensures the values add up to one and are all

non-negative, such that

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}. \tag{7.16}$$

The output label is chosen by taking the class with the maximum predicted probability as

$$y = \underset{i \in \{1,\ldots,k\}}{\operatorname{argmax}} \sigma(\mathbf{z})_i. \tag{7.17}$$

7.2.5 The Convolutional Neural Network

Convolutional neural networks (CNNs) are a specific type of neural network designed to process data with a grid-like structure, such as imaging or time series data. CNNs have been widely used in computer vision tasks, such as the U-Net architecture for segmentation (Ronneberger et al., 2015) and the ResNet architecture for classification (He et al., 2015a). A standard layer in a CNN is composed of three distinct stages. In the first stage, several convolutional filters are applied to the input to generate a linear combination of the input features. Next, these combined values are passed through nonlinear activation functions to introduce nonlinearity to the model, which step is also referred to as the 'detector' stage. Finally, the third stage aggregates these thresholded values over a subregion via a pooling operation. A mathematical description of each stage is provided in the following subsections.

Convolution

In the context of NNs, convolution refers to the operation that applies a set of filters, also called kernels, to an input. Each filter is progressively shifted over the input in a sliding window manner, and the element-wise product is calculated between the kernel and the corresponding subregion of the input, as illustrated in Figure 7.2.

For a three-dimensional input image with c_{in} number of input channels, $\mathbf{X} \in \mathbb{R}^{c_{in} \times H \times W}$ and c_{out} number of kernels with learnable weights $\mathbf{W} \in \mathbb{R}^{c_{out} \times c_{in} \times K_H \times K_W}$, the convolution operation is expressed as

$$y_{c_{out},i,j} = b_{c_{out}} + \sum_{c_{in},u,v} x_{c_{in},i+u,j+v} w_{c_{out},c_{in},u,v} \quad , \tag{7.18}$$

where u and v iterate on the height and width of the kernel K_H and K_W, while i and j index the output along its height and weight, respectively. Of note, the kernel weights used in the convolutional layer are learnt from data; thus, the output is indifferent to whether these layers perform strict convolution or the cross-correlation operation. Often, zero-padding is applied before performing the convolution operation so that output's height and width remain unchanged.

There are many variants of the standard convolution operation described above that extend the capabilities of this technique, such as dilating the convolutional kernel to enable learning multiscale feature sets (Yu and Koltun, 2016), or applying depth-wise convolution to separate

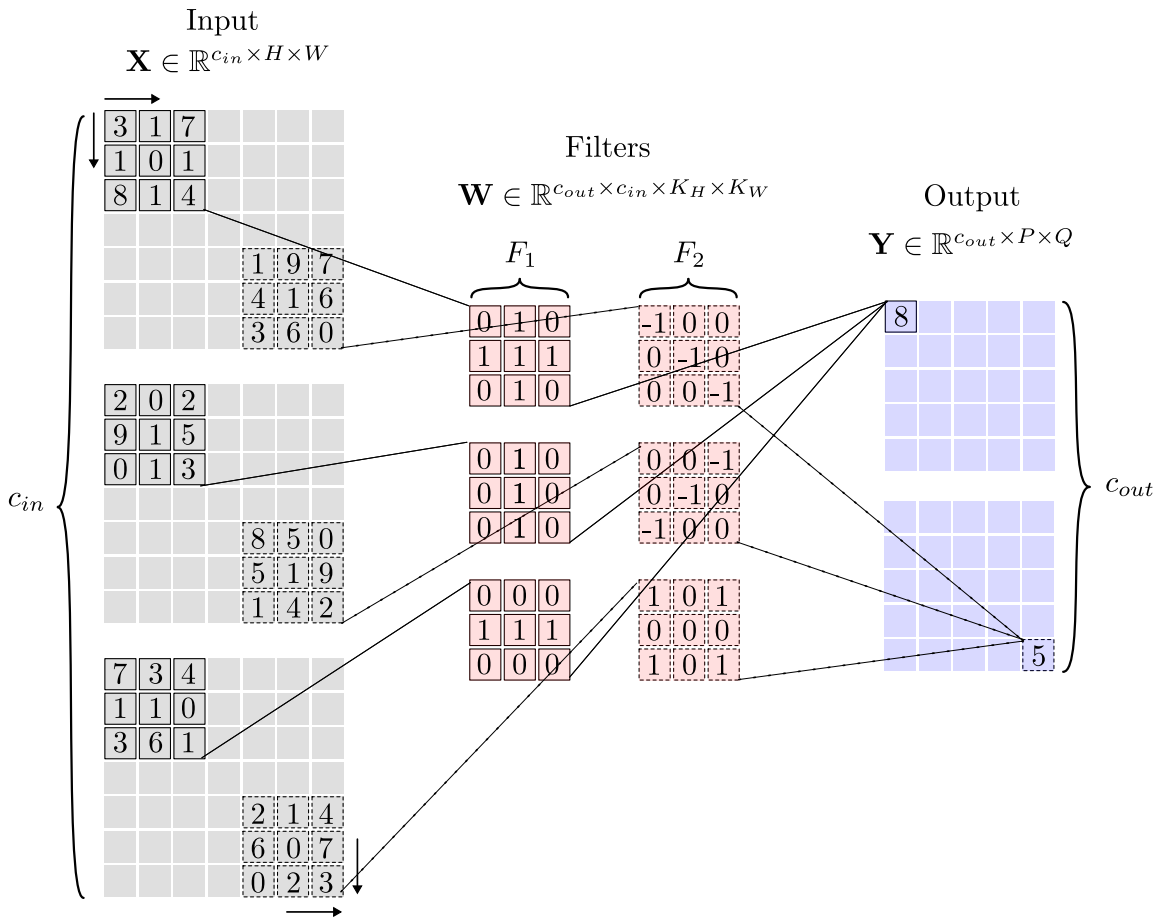


Figure 7.2: The convolution operation as performed on a three dimensional input using a set of filters. The set of filters are shifted across the input from the top-left to the bottom-right corner in a sliding-window manner, and the output is calculated by summing the element-wise products of the corresponding regions.

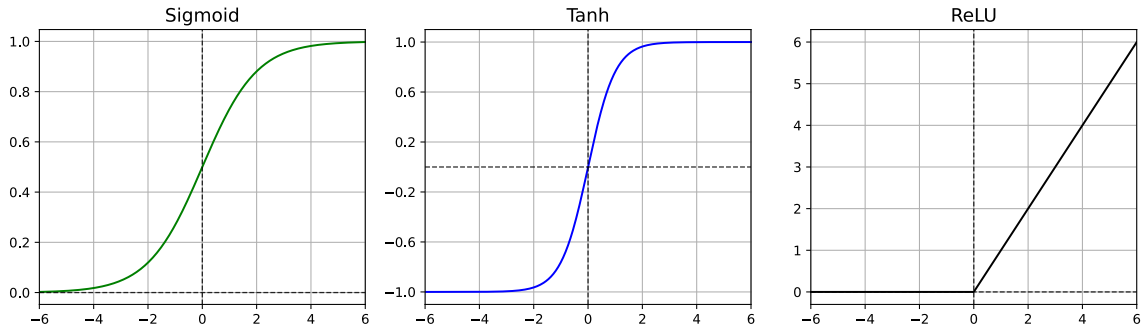
the spatial and cross-filter interactions (Chollet, 2016). In essence, however, they all leverage sparse interactions as opposed to a fully connected layer, thus reducing the computational requirements of the model and introducing parameter sharing.

Activation Function

The detector stage of a convolutional layer is responsible for introducing nonlinearity to the model by applying nonlinear functions to the output of the convolution operation. This, in turn, also has the effect of acting as a threshold for the range of the unit. The most common types of activation functions are illustrated in Figure 7.3. The sigmoid and hyperbolic tangent (tanh) functions introduce nonlinearity by squashing input values into bounded intervals $[0, 1]$ and $[-1, 1]$, respectively. The rectified linear unit (ReLU) allows for unbounded positive outputs by applying an identity function for positive inputs while setting all negative inputs to zero.

Pooling

The pooling operation summarises the responses of a layer in a neighbourhood of neurons. It first divides the input into non-overlapping rectangular regions based on the kernel shape and then computes a summary statistic for each region by aggregating the values of the neurons



$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad \tanh(x) = \frac{2}{1+e^{-2x}} - 1 \qquad \mathrm{ReLU}(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

Figure 7.3: Common activation functions used in deep learning. The sigmoid and hyperbolic tangent (tanh) functions introduce nonlinearity by squashing input values into bounded intervals $[0, 1]$ and $[-1, 1]$, respectively. The rectified linear unit (ReLU) allows unbounded positive outputs while setting all negative inputs to zero.

within it. The output of the pooling operation is then obtained by substituting each region of the input with the corresponding calculated scalar value. This reduces the spatial dimensions by a factor equal to the kernel stride, which helps the network remain invariant to small translations.

Average pooling is expressed as

$$y_{i,j} = \frac{1}{K_H K_W} \sum_{u=0}^{K_H-1} \sum_{v=0}^{K_W-1} x_{si+u,sj+v} \quad , \tag{7.19}$$

where s denotes the stride of the kernel, K_H and K_W denote the height and width of the kernel, respectively. Similarly, max pooling is also often applied to return the maximum value of each region, so that

$$y_{i,j} = \max_{0 \leq m < K_H} \max_{0 \leq n < K_W} x_{si+u,sj+v}. \tag{7.20}$$

7.2.6 Additional Components

Modern DL architectures often employ additional mechanisms beyond convolutional and fully connected layers to improve stability during training or to enhance generalisation performance. The most common components include normalisation, dropout, and residual connections.

Normalisation

Normalisation is a technique applied to the activations of a layer's neurons by subtracting the mean and dividing by the standard deviation of some input feature set of the layer. It is used to improve the training stability of a model by preventing vanishing or exploding gradients during backpropagation, while reducing the network's dependence on the scale and distribution of the input data.

There are different ways in which normalisation can be applied depending on how the feature set is chosen to calculate the mean and variance; however, the normalised activation of a neuron, $\hat{a}$ can generally be expressed as

$$\hat{a} = \frac{a - \mu}{\sqrt{\sigma^2 + \epsilon}}\gamma + \beta \quad , \tag{7.21}$$

where a represents the unnormalised activation of a single neuron, μ and σ are the mean and standard deviation of the feature set, γ and β are learnable gain and bias parameters, respectively. For numerical stability, a small number, ϵ, is introduced which usually has an order of magnitude of 10^{-5}.

For batched three-dimensional inputs $\mathbf{X} \in \mathbb{R}^{N \times C \times H \times W}$, where H and W denote the height and width of each feature, c indicates the number of channels or kernels within the layer, and n is the batch size, the batch norm operation is defined by calculating the layer statistics across the batch and spatial dimensions for each channel c as

$$\mu_c = \mathbb{E}_{n,i,j}[x_{n,c,i,j}], \quad \sigma_c^2 = \mathrm{Var}_{n,i,j}[x_{n,c,i,j}], \tag{7.22}$$

For the layer norm operation, the statistics are calculated across all channels and spatial dimensions for each sample n, so that

$$\mu_n = \mathbb{E}_{c,i,j}[x_{n,c,i,j}], \quad \sigma_n^2 = \mathrm{Var}_{c,i,j}[x_{n,c,i,j}]. \tag{7.23}$$

Dropout

Dropout is a regularisation technique used to improve the generalisation performance of ML algorithms by sampling a random subset of the network's neurons at each forward pass during training (Srivastava et al., 2014). The remaining non-output units are 'dropped out', i.e., their activations and gradients are set to zero. The binary selection mask applied to the network is determined by a specified probability called the dropout rate. By randomly dropping out neurons, the model is prevented from relying too much on individual units and feature interactions, thereby promoting robustness in the learnt representation.

Dropout can be interpreted as an efficient way to approximate training and evaluating an ensemble of neural networks that share parameters, which is referred to as bagging. This method also has a regularising effect as it prevents hidden units from learning the same representation; a phenomenon known as co-adaptation (Hinton et al., 2012).

Residual Connections

For very deep networks, where the number of successive layers is large, the repeated composition of nonlinear transformations can cause gradients to vanish or explode during training, which prevents an effective and stable optimisation. To address this limitation, residual connections can be used, which propagate both activations and gradients through skip connections that bypass certain parts of the model. Skip connections provide an identity path from the input of a layer to the output of a deeper layer as shown in Figure 7.4. This method

facilitates gradient flow, allowing deeper networks to be trained, which significantly increases model performance in several image classification tasks (He et al., 2015b).

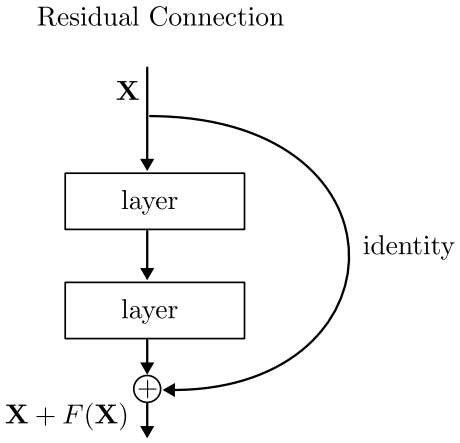


Figure 7.4: Residual connections propagating the input to a layer to its output unaltered, stabilizing gradient flow during backpropagation. (Figure adapted from Wikipedia (2023)).

7.3 Related Work

CNN architectures still constitute the basis for many image processing tasks, such as ResNet for classification and U-Net for segmentation tasks. Early research on DL methods applied to echocardiographic recordings by Madani et al. (2018) used standard CNNs to recognise 12 different anatomical views from 2D TTE images, achieving a test set accuracy of 97.8%. Leclerc et al. (2019) employed U-Nets to segment the LV at the end-diastolic and end-systolic frames outperforming previous models, while recently Holste et al. (2023) used ResNets to predict severe AS from a single PLAX cine loop, with an AUROC of 0.978.

Modern DL models, however, often incorporate additional training objectives or employ refined frameworks to improve model performance and tackle more complex problems. One of the most significant challenges in predicting the severity of AS arises from the fact that routine echocardiographic examinations produce multiple types of recordings from various anatomical views but are assigned only a single label for the overall severity grade. Ginsberg et al. (2021) addressed this by aggregating the video-level predictions from PLAX and PSAX recordings through majority voting (taking the most common label) to assign a single patient-level diagnosis. The initial work by Huang et al. (2021) proposed the aggregation of predictions using a weighted average of 'relevant' views, where the relevance score was assigned by a separately trained view classifier that assigned high values only to PLAX and PSAX views.

7.3.1 Multiple-Instance Learning

Although aggregating AS severity predictions from individual images or videos can produce a patient-level diagnosis, this strategy does not accurately mimic the clinical evaluation process, in which clinicians assess the study as a whole rather than label each recording separately. The prediction of a single output label given an unordered set of input features describes the goal of multiple-instance learning (MIL), an approach that has been explored by Huang et al. (2024) in their subsequent study.

The training objective of the MIL method for AS prediction using the TMED2 data set is formulated as a supervised learning task, where the model is trained on a set of N input–output pairs $S = \{(\mathbf{X}^{(i)}, y^{(i)}); i = 1, ...N\}$. Each input $\mathbf{X}^{(i)} = \{X_1, ..., X_{K_i}\}$ consists of a variable number of 2D TTE images from different anatomical views, and the corresponding output $y^{(i)} \in \{0, 1, 2\}$ denotes the severity grades of no AS, early AS or significant AS. The collection of images within an input example is also referred to as a 'bag' in the MIL terminology, representing the entire study as opposed to an instance that refers to a single image within a study.

The model introduced by Huang et al. (2024) following the MIL formulation consists of three main components: a feature extraction module $f(\cdot)$, which processes each individual instance within a study independently; a pooling layer $\sigma(\cdot)$, which combines the extracted instance-level features into a bag-level representation ensuring permutation invariance; and an output layer $g(\cdot)$, which maps a class label to the obtained bag-level representation. For the feature extraction module, a CNN architecture with a fully connected layer at the end was used to project each instance to a low-dimensional vector so that

$$h_k = f(X_k), \tag{7.24}$$

where $X_k \in \mathbb{R}^{H \times W}$ denotes a greyscale input image with a resolution of 112×112, and $h_k \in \mathbb{R}^M$ is the low-dimensional representation of the image.

The pooling layer then aggregates these features through a weighted average to produce a bag-level representation $z \in \mathbb{R}^M$, such that

$$z = \sum_{k=1}^{K_i} a_k h_k, \quad \text{with} \quad a_k = \frac{\exp(w^T \tanh(U h_k))}{\sum_{j=1}^{K} \exp(w^T \tanh(U h_j))}, \tag{7.25}$$

where $U \in \mathbb{R}^{L \times M}$ and $w \in \mathbb{R}^L$ are parameters of the model learnt during training. The attention weights a_k were passed through the softmax function to produce a valid probability distribution over the K instances.

A standard softmax function following a fully connected layer was chosen for the output module to produce a valid probability distribution over the three severity grades present in

the TMED2 data set, given a bag-level representation $z = \sigma(f(\mathbf{X})$, so that

$$P(y = c|\mathbf{X}) = g(z)_c \quad \text{with} \quad g(z) = \text{softmax}(Wz + b), \tag{7.26}$$

where $W \in \mathbb{R}^{3\times M}$ and $b \in \mathbb{R}^3$ denote the weight and bias terms of the fully connected layer, respectively.

The model was trained using gradient-based optimisation with a cross-entropy loss function calculated between the study-level diagnosis labels and the model's output predictions, such that

$$W^*, b^*, U^*, w^*, \theta^* = \underset{W,b,U,w,\theta}{\text{argmin}} \sum_i \mathcal{L}_{CE}\left(y^{(i)}, g_{W,b}(\sigma_{U,w}(f_\theta(\mathbf{X}^{(i)})))\right). \tag{7.27}$$

Huang et al. (2024) found that the architecture and training regime described above produced suboptimal classification performance and that the attention scores learnt by the model did not correspond to anatomically relevant views in which the aortic valve was visible (i.e., the PLAX and PSAX views). To address this limitation, they extended their MIL framework by introducing a supervised attention (SA) module. This mechanism steers the learnt attention weights a_k in Equation 7.25, to collate a bag-level representation from the anatomically relevant views only. This was achieved by incorporating an independently trained view classifier $v : \mathcal{X} \rightarrow [0, 1]$ that assigns to each image a relevance score corresponding to the likelihood that the given image is of a relevant view showing the aortic valve. The relevance scores produced by the view classifier $R = \{r_1, ...r_K\}$ were used to define an additional SA loss function that measured the distance between the learnt attentions, $A = \{a_1, ...a_K\}$, and the assigned view relevance scores via the KL-divergence, so that

$$\mathcal{L}_{\text{SA}} = D_{\text{KL}}(R||A) = \sum_{k=1}^{K} r_k \log \frac{r_k}{a_k}. \tag{7.28}$$

The view relevance scores were passed through a softmax function with a tunable temperature hyperparameter τ_v prior to the calculation of the loss function to ensure a valid probability distribution over R, such that

$$r_k = \frac{\exp(v(X_k)/\tau_v)}{\sum_{j=1}^{K} \exp(v(X_j)/\tau_v)}. \tag{7.29}$$

The model was then optimised using the composite objective function formulated by adding the view relevance dissimilarity measure $\mathcal{L}_{\text{SA}}$ and the study-level cross-entropy loss between the predicted and the true diagnosis labels $\mathcal{L}_{\text{CE}}$, as

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{SA}}, \tag{7.30}$$

where λ represents the tunable view regularisation hyperparameter.

By introducing the SA module the model, however, loses flexibility in attending to specific images more than others, as attention weights will greatly depend on the distribution of

the view relevance scores assigned by the view classifier. To mitigate this issue, the last component of their SAMIL model consists of additional learnable parameters that combine both the supervised attention scores and flexible attention scores to produce a bag-level representation as

$$z = \sum_{k=1}^{k} c_k h_k \quad \text{with} \quad c_k = \frac{a_k b_k}{\sum_{j=i}^{K} a_j b_j} \quad \text{and} \quad b_k = \frac{\exp(w_b^T \tanh(U_b h_k))}{\sum_{j=1}^{K} \exp(w_b^T \tanh(U_b h_j))}. \tag{7.31}$$

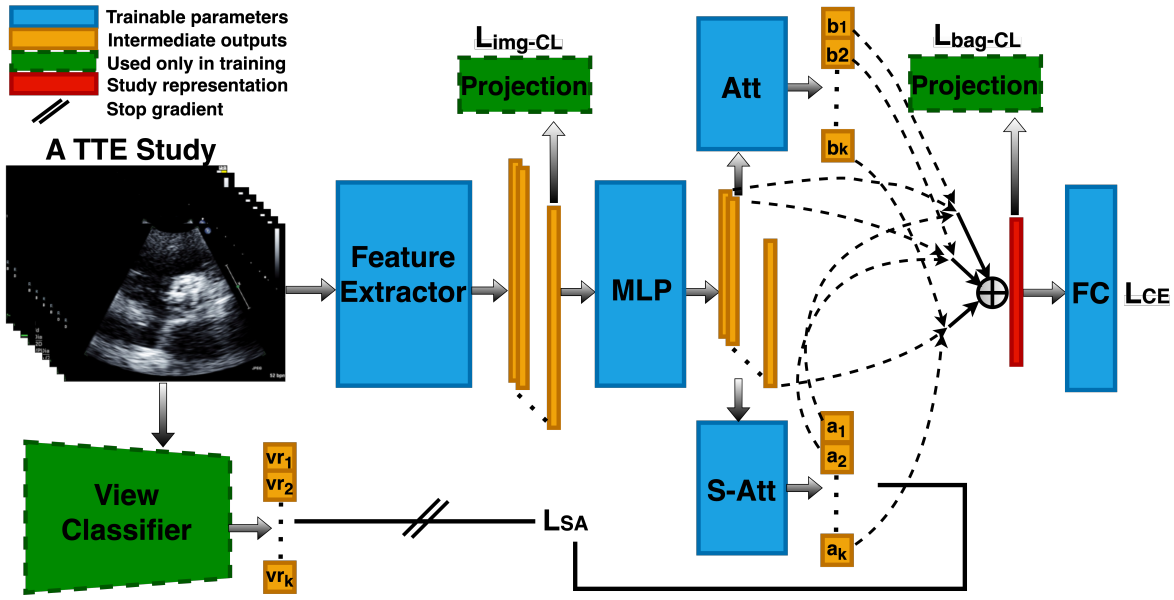


Figure 7.5: Architecture of the supervised attention multiple-instance learning (SAMIL) model. The network comprises three main components: a feature extractor $f(\cdot)$, which converts each image from a transthoracic echocardiogram study into a feature representation; a pooling layer $\sigma(\cdot)$, which aggregates these instance-level features into a bag-level representation through learnt attention weights while preserving permutation invariance; and an output layer $g(\cdot)$, which maps the bag-level representation to the predicted aortic stenosis (AS) severity class. A supervised attention (SA) module further guides the attention weights toward clinically relevant views (parasternal long-axis and parasternal short-axis), through an additional view-relevance loss L_{SA}. (Figure adapted from Huang et al. (2024)).

An overview of the MIL framework is illustrated in Figure 7.5. In a later study following their MIL framework, Huang et al. (2025) also incorporated Doppler images, outperforming all existing models for the AS classification task in the TMED2 data set.

7.3.2 Transformers

The MIL framework detailed above improved on conventional DL models by introducing a refined training regime; however, a distinct approach to enhancing computer vision architectures involves the use of self-attention mechanisms, which allow the discovery of long-range spatial dependencies more explicitly than standard convolutional layers. This technique gained widespread recognition by the introduction of the transformer model (Vaswani et al., 2017), which was first used in natural language processing (NLP) as a successor to recurrent neural networks (RNN). The adaptation of transformers to image data followed shortly;

the Vision Transformer (ViT)(Dosovitskiy et al., 2020) architecture outperformed CNNs on various image classification benchmarks and led to its extensive use for computer vision problems. Recently, Tragakis et al. (2023) have developed the Fully Convolutional Transformer (FCT) architecture, which combines the merits of CNN and transformer architectures, achieving state-of-the-art performance in several medical image segmentation datasets. Multiple efforts have also been made to incorporate transformer-based approaches to AS severity classification using 2D TTE recordings (Avola et al., 2024; Ahmadi et al., 2024), however, their adoption into the MIL framework has not yet been achieved.

Model Overview

The FCT model for medical image segmentation follows a standard U-Net shape, as shown in Figure 7.6, where the input image is first processed through a series of encoder blocks to obtain a low-dimensional representation and then upsampled by the decoder blocks to produce a label map at the original image resolution. The encoder blocks are composed of two distinct components after applying conventional layer normalisation, convolution, and pooling operations:

- First, the convolutional attention module performs a patch embedding function that splits input features into overlapping regions through the depth-wise convolution operation and flattens these 2D patches to form linear token embeddings for the query (Q), key (K) and value (V) representations. These embeddings then serve as inputs to the Multi-Head Self-Attention (MHSA) module. Lastly, the MHSA output is reshaped to match the resolution of the input features. This approach constitutes the first novelty of this model compared to designs like ViT, in which embeddings are created by first flattening non-overlapping patches of the input features and then obtaining the embeddings for Q, K and V through linear projections with additional positional encoding.
- Next, the wide-focus module is used to process the MHSA output. This module employs parallel convolutions, with one branch performing standard, while the other two performing dilated convolutions with varying receptive fields. The outputs from the multi-branch convolution are fused via summation and feature aggregation using an additonal convolutional layer.

Residual connections are incorporated into both components to facilitate the propagation of features through each layer. The original FCT model is composed of five encoder blocks that enable the processing of multi-scale inputs via consecutive average pooling of the input image to match the feature representation of the given layer. The last block is also called the bottleneck encoder, which is responsible for the latent space encoding of the input data. Decoder blocks employ standard upsampling and convolution operations, with multi-scale outputs for the deep supervision of the segmentation task, in which ground truth labels are downsampled to match the output dimension of the corresponding layer.

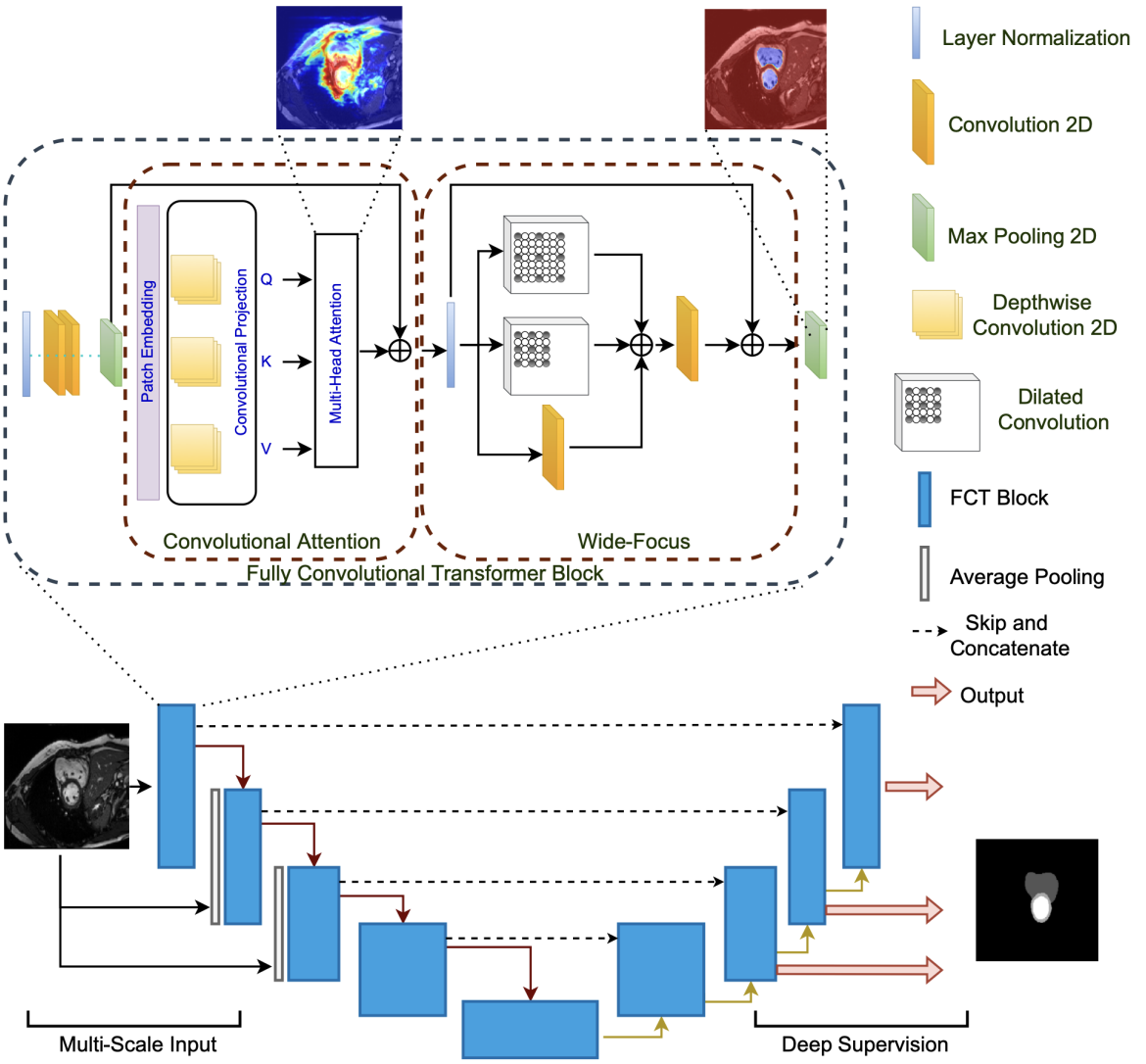


Figure 7.6: The Fully Convolutional Transformer (FCT) architecture for medical image segmentation. (Figure adapted from Tragakis et al. (2023)).

The MHSA module forms the foundation of most transformer architectures, enabling models to capture joint relationships across different representation subspaces. Each head within this module performs the scaled dot product attention defined as

$$\alpha(Q, K, V) = \sigma\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{7.32}$$

where $Q \in \mathbb{R}^{N \times d_k}$, $K \in \mathbb{R}^{N \times d_k}$ and $V \in \mathbb{R}^{N \times d_v}$ represent the query, key, and weight matrices, respectively, and σ is the softmax function. N denotes the sequence length, which for an input feature with height H and width W, is chosen to be HW, corresponding to the length of the flattened feature vector, and d_k and d_v are the embedding dimensions for queries/keys and values, respectively. For the MHSA module this operation is performed in parallel h times, where h indicates the number of attention heads, and their outputs are concatenated via a learnable weight matrix W_0, as

$$\begin{aligned} \text{MHSA}(Q, K, V) &= \text{Concat}(h_1, ..., h_h)W_o, \\ \text{where } h_i &= \alpha(Q_i, K_i, V_i). \end{aligned} \tag{7.33}$$

Linear projections are used to obtain patch embeddings for each head by multiplication of

the corresponding learnable weight matrices so that

$$Q_i = QW_i^Q \quad , K_i = KW_i^K \quad , \text{and} \quad V = VW_i^V . \tag{7.34}$$

7.4 Methodology

As described in the previous section, both the MIL framework and transformer-based approaches have the potential to improve the diagnostic capability of modern computer vision algorithms. Recent work suggests that the combination of these two techniques introduces further gains in multiple medical imaging data sets (Keshvarikhojasteh et al., 2024). However, this approach has not yet been tested for AS classification.

To integrate the multi-head attention–based transformer architecture proposed by Tragakis et al. (2023) into the MIL framework for AS diagnosis, the original FCT model was first modified. This entailed retaining only the first four encoder blocks, allowing a lower-level representation of the input images via the bottleneck layer. This representation was then passed through a fully connected layer to produce image-level features. The input dimensions were also modified to match the resolution of the images in the TMED2 data set. The resulting encoder-only FCT architecture was then used as a feature extractor layer $f(\cdot)$ within the MIL framework introduced by Huang et al. (2024), replacing the standard CNN used in the original SAMIL architecture. The created model using the modified FCT architecture as its feature extractor is referred to as FCT for simplicity.

7.4.1 Training Strategy

Data normalisation was applied to the input images prior to training using the mean and standard deviation statistics provided by the authors of the TMED2 data set. Both the SAMIL and FCT models were trained from scratch to ensure a fair comparison between the architectures. The hyperparameters for the SAMIL model, including temperature (Equation 7.29), view regularisation strength (Equation 7.30), learning rate and weight decay, were set according to published values.

Adam and SGD optimisers were employed with a 20- to 30-epoch learning rate warm-up followed by a cosine decay schedule. Data augmentation was applied by randomly selecting two operations from random cropping, rotation, mirroring, translation, and brightness shift. The batch size was set to one due to the data loading constraints of the MIL framework. The total loss combined cross-entropy and view relevance terms, with temperature and view-regularisation parameters controlling the relative contribution of the SA loss.

Both models were trained across all three predefined training splits using the pre-trained view classifiers corresponding to each split to produce view relevance scores for the super-

vised attention loss. As view classifiers were trained on raw input images, each study was unnormalised using the same mean and standard deviation before performing a forward pass through these networks. The models with the best validation performance were selected using the balanced accuracy performance metric.

Dropout regularisation was implemented in the FCT architecture, with its rate included among the tunable hyperparameters. The number of attention heads was also tuned, ranging from one to two per encoder block. Training was terminated early if validation performance did not improve for 20–30 consecutive epochs. Model training and hyperparameter selection were performed exclusively using the view-and-diagnosis-labelled subset of the TMED2 data set. The predefined TMED2 training and validation splits were used for parameter optimisation and model selection, respectively, while the predefined TMED2 test sets were reserved for held-out evaluation to assess generalisation performance. All experiments were performed on NVIDIA A100 GPUs.

7.4.2 Generalisation Performance

The trained models were then evaluated on the predefined TMED2 test sets and two external clinical cohorts to assess their performance across different settings. To compare model performance across all three data splits, the predefined test sets from TMED2 were used. To examine cross-cohort applicability, the concordant AS cohort from Golden Jubilee National Hospital, referred to as AS60, was employed. To further evaluate the models in a distinct clinical subgroup, the DSE12 data set representing the LFLG AS cohort was used. The recordings in the AS60 and DSE12 data sets were pre-processed to produce greyscale images of 112×112 pixels by extracting the first frame from each cine loop. Doppler recordings were excluded from all analyses. Furthermore, the DSE12 data set was divided into two groups: one comprising recordings acquired at rest and another combining recordings from both the rest and stress states. This allowed for examining the impact of stress testing on model performance within the LFLG AS cohort.

The performance metrics used included balanced accuracy and the ROCAUC of binary classifiers trained to distinguish between the following severity groupings: (i) no AS vs. any AS (early or significant), (ii) non-significant AS (no or early) vs. significant AS, and (iii) early vs. significant AS. Confusion matrices were also examined to further assess classification performance in different severity classes.

Model Ensembling

Since training on the predefined splits from TMED2 resulted in 3 separately train models for both SAMIL and FCT architectures, the test inference on the held-out clinical data was standardised. Multiple approaches were used to aggregate the prediction between the different models. For the SAMIL architecture, both majority voting and probability averaging were

applied. In the former, the final class label corresponds to the class most frequently predicted among the three models for a given input. In the latter, the predicted probability distributions from each model are averaged and the class with the highest mean probability is selected as the final prediction.

It should be noted that the reported class probabilities represent point estimates of the model's predictive distribution over AS severity classes, and they may reflect predictive ambiguity in the input data, such as overlapping imaging features between adjacent severity classes, which is related to aleatoric uncertainty. However, they do not explicitly separate aleatoric uncertainty from epistemic uncertainty. The uncertainty arising from limited training data or distribution shift between TMED2 and the clinical AS60 and DSE12 cohorts is not fully captured by these outputs, thus the predicted probabilities should be interpreted as relative model confidence scores rather than calibrated measures of diagnostic certainty.

Additionally, another model-ensembling technique was used for the FCT architecture, as this model included dropout regularisation during training. By keeping dropout active during inference, the model produces stochastic outputs that sample a different subset of the trained model at every forward pass. By applying probability averaging using the output distribution of subsequent forward passes of the same input, the output distribution becomes more robust when evaluated on slightly different test sets. This method is called Monte Carlo dropout (MCD) inference (Gal and Ghahramani, 2016).

7.4.3 Model Explainability

Attention Weight Analysis

To verify that the models based their predictions on clinically relevant views, the learnt attention scores were analysed. Specifically, the view relevance scores a_k of the ten most highly attended images from each study were averaged within the data splits, and the mean score for each rank position was observed, so that

$$a_k = \frac{1}{N} \sum_{i=1}^{N} r_{ik}, \tag{7.35}$$

where i indexes the studies within each data set and k denotes the image ranked k in terms of attention within each study. A score close to 1 indicates that the model focuses on clinically relevant PSAX or PLAX views, whereas lower values suggest that the model uses additional views for its output.

Saliency Maps

To gain insight into the specific regions that contribute the most to the models prediction for a given class, a gradient-weighted class activation map (Grad-CAM) is often used (Selvaraju et al., 2016). This method relies on calculating the gradient of the model's output probability for a given class, y^c with respect to a feature map within a convolutional layer $A^k \in \mathbb{R}^{H \times W}$,

such that

$$\alpha_k^c = \frac{1}{HW} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}, \tag{7.36}$$

where H and W denote the height and width of the feature map, respectively, and the scalar α_k^c is the average of all gradients within the given feature map. After obtaining the relative importance of a feature map, an input image is forward passed through the convolutional layer, and the activations of each feature map are weighted through α_k^c before applying the nonlinearity so that

$$L_{\text{Grad-CAM}}^c = \text{ReLU}(\sum_k \alpha_k^c A_k^c), \tag{7.37}$$

which produces a single activation map for the entire convolutional layer. Upsampling the activation map to match the resolution of the input image reveals the specific parts of the image that impact the models prediction the most for the given class.

Domain Shift Analysis

Domain shift refers to the phenomenon in which the underlying data distribution changes between different domains. This shift may occur in the input distribution P(X), the label distribution P(Y), or their joint distribution P(X,Y), which could result in a decreased model performance when applied to data that differ from those seen during training.

In the context of AS classification from echocardiographic recordings, the shift in input distribution may arise from differences in data acquisition, such as the use of different ultrasound probes or vendors, variations in postprocessing parameters (e.g., gain and contrast), which affect image brightness, or interobserver variability in acquisition angle. A shift in the label distribution could occur when the model is evaluated in different patient cohorts. For example, the TMED2 data set includes three severity labels (no AS, early AS, and significant AS), while the concordant AS patients in the AS60 data set contain only early and significant AS cases, and patients with LFLG AS in the DSE12 data set represent significant AS cases only. Finally, the joint distribution shift includes the most challenging case in which both the input and label distributions change simultaneously. This can occur due to differences in disease pathology, such as LFLG AS exhibiting distinct image characteristics compared to concordant AS.

To assess the extent to which domain shift occurs, a commonly used metric, the Fréchet Distance (FD) is used. As FD measures the distance between two multivariate Gaussian distributions, the assumption was made that the feature distribution of each class label could be modelled as

$$P(X \mid Y = c) \approx \mathcal{N}\left(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c\right), \tag{7.38}$$

where $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ denote the empirical mean and covariance of the feature distribution for class c.

PCA was used to obtain a low-dimensional feature vector for each patient using bag-level representations within the training set from the trained models. The number of PCs was chosen so that the explained variance exceeded 99.9%. Projecting each set to PCs obtained from the training set and observing the mean and covariance of each class allowed the calculation of the FD between the different domains.

To allow fair comparison across different model spaces, all features are normalised using the statistics of the PC representation of the reference training set $\mathbf{X}_{\text{ref}}$, so that:

$$\tilde{\mathbf{X}} = \frac{\mathbf{X} - \boldsymbol{\mu}_{\text{ref}}}{\boldsymbol{\sigma}_{\text{ref}}},$$

where $\boldsymbol{\mu}_{\text{ref}}$ and $\boldsymbol{\sigma}_{\text{ref}}$ are the mean and standard deviation of PCs in the reference domain.

The FD between domains $\mathcal{D}_A, \mathcal{D}_B$ is defined as

$$\text{FD}^2(\mathcal{D}_A, \mathcal{D}_B) = \|\boldsymbol{\mu}_A - \boldsymbol{\mu}_B\|_2^2 + \text{Tr}\left(\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_B - 2\left(\boldsymbol{\Sigma}_A\boldsymbol{\Sigma}_B\right)^{1/2}\right), \tag{7.39}$$

where $\boldsymbol{\mu}_A, \boldsymbol{\mu}_B$ are the mean vectors of the corresponding domains, $\boldsymbol{\Sigma}_A, \boldsymbol{\Sigma}_B$ are their covariance matrices, $\text{Tr}(\cdot)$ is the trace of a matrix and $\|\cdot\|_2$ denotes the Eucledian distance. Note that the first term corresponds to a shift in the centroid, while the second a change in shape of the distribution. The different domain comparisons are illustrated in Figure 7.7.

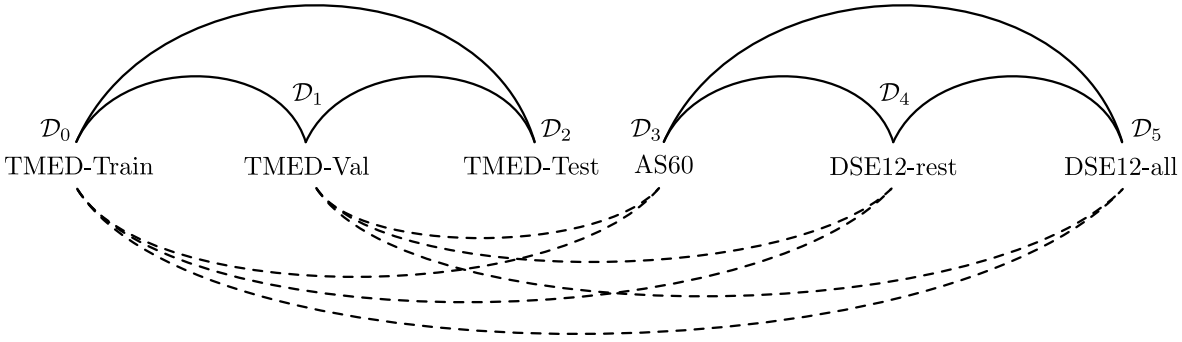


Figure 7.7: Visual overview of the domain pairs used for shift analysis. Six domains are considered: the training, validation and test splits from the Tufts Medical Echocardiogram Dataset (TMED) denoted as TMED-Train, TMED-Val, TMED-Test, the concordant aortic stenosis (AS) cohort indicated by AS60, and recordings from the low-flow low-gradient AS patients at rest and at the combined rest and stress state denoted as DSE12-rest, and DSE12-all. Solid lines represent comparisons between domains from the same hospital (e.g., training to validation), while dashed lines indicate cross-hospital comparisons.

7.5 Results on TMED2

7.5.1 Training and Validation Performance

To quantify the gains in introducing a novel feature extractor into the MIL framework, both FCT and SAMIL models were trained and evaluated using the same predefined TMED2 data splits and the same performance metrics. To ensure that the comparison between the two

architectures was as fair as possible, the number of encoder layers for the FCT model was chosen so that the resulting model capacity would match that of the SAMIL architecture, with 2.31 million trainable parameters for SAMIL and 2.32 million for FCT. The SAMIL architecture was treated as a published baseline and was trained from scratch using the hyperparameters reported in the original study, where these parameters had already been selected for this model. In contrast, the FCT architecture introduced a modified feature extractor and therefore required hyperparameter selection using random search. Thus, the comparison was controlled with respect to data partitioning, evaluation protocol, and parameter count, while the difference in hyperparameter selection reflects the use of an established baseline compared with a newly introduced architecture. Figure 7.8 shows the training process of the SAMIL architecture on the TMED2 view-and-diagnosis labelled set across all data splits. The cross-entropy loss, the view regularisation loss, and the balanced accuracy metrics are depicted in the training and validation sets for each epoch.

It is observed that both CE and view regularisation losses are minimised for the training splits, which results in an increase in model accuracy, confirming that the optimisation objective is well defined. In addition to the reduction in training losses, the models reached an overfitting regime after around 75 epochs, as indicated by the increase in CE loss in the validation splits. This increase is not observed for the regularisation losses, which is most likely caused by the high λ parameter, which results in the overall loss function being dominated by the view regularisation loss.

The difference in training accuracies between splits is small, reaching a balanced accuracy of around 95% for all models. However, there is a slight difference among the splits in validation accuracy, with split 3 being the highest with a balanced accuracy of around 73%. Note that the training and validation curves exhibited significant noise, possibly due to the use of a small batch size for stochastic gradient-based optimisation techniques.

The FCT architecture was then trained with varying hyperparameters using a random search for the number of heads, dropout rate, view regularisation strength, learning rate, weight decay, and temperature in each data split. The training process for split 3 is shown in Figure 7.9 for models with varying view regularisation strengths (λ) to assess the importance of this parameter. It is noted that increasing the view regularisation strength from 1 to 10 resulted in faster convergence with regards to both CE loss and balanced accuracy in addition to the view regularisation loss as seen in the training split.

It is also observed that higher view regularisation strengths result in greater overfit as indicated by the validation CE loss. The view regularisation loss in the validation split is also lower for higher values of λ. However, the difference in validation accuracy only shows marginal gains for $\lambda = 10$ compared to other models, with a highest balanced accuracy of around 78%. The highest validation split accuracy did not correspond to the highest training split

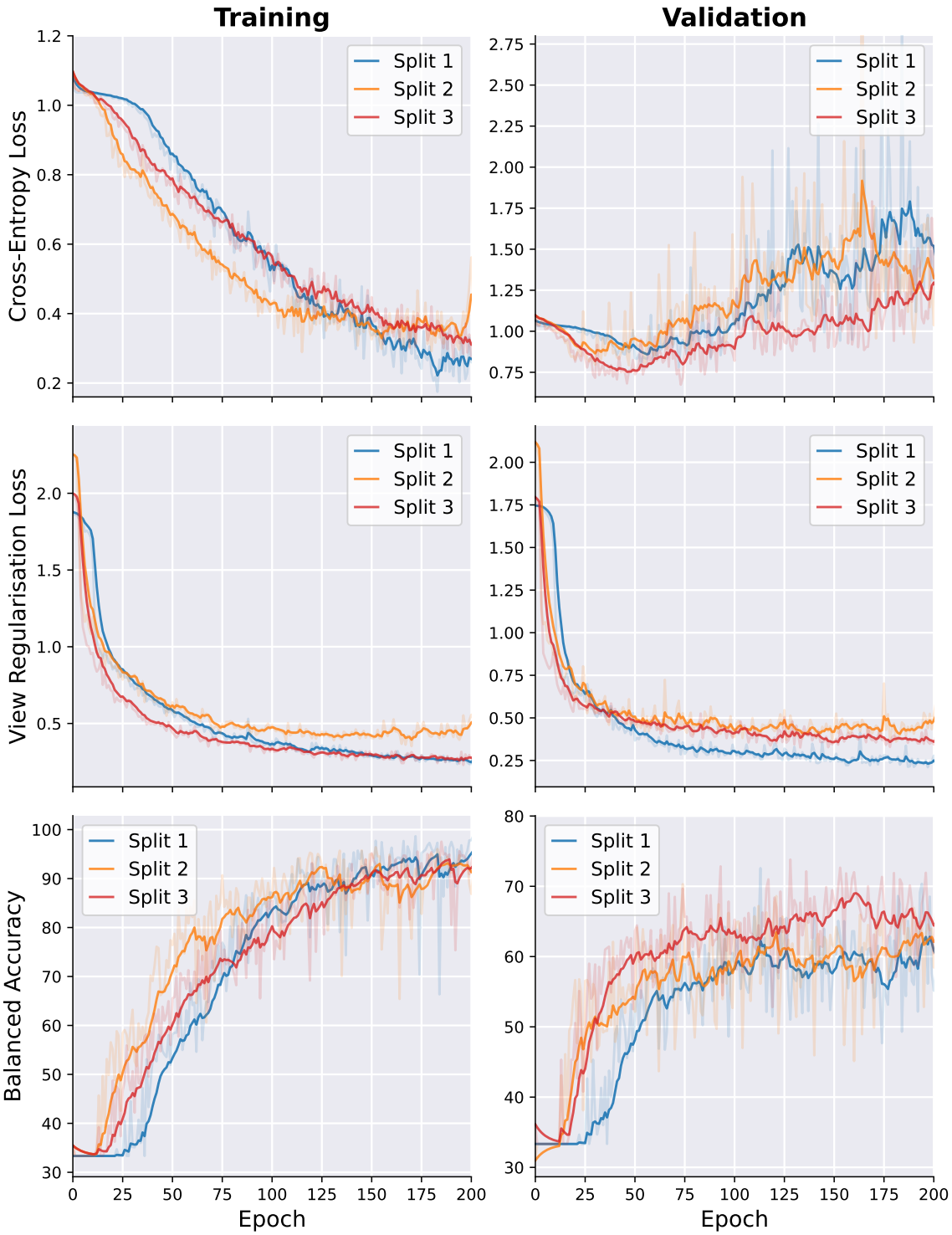


Figure 7.8: Training and validation performance of the supervised attention multiple instance learning model across the pre-defined cross-validation splits of the Tufts Medical Echocardiogram Dataset. Hyperparameters were set for each split according to the reported values in (Huang et al., 2024). Cross-entropy loss, view regularisation loss, and balanced accuracy are shown from top to bottom. The left column presents the results for the training set, while the right column shows the results for the validation set. Smoothing was applied to the raw output values (seen in matching transparent colors) to help visualisation.

accuracy. These results illustrate a trade-off between view relevance and generalisation, that is, increasing λ facilitates faster convergence and more consistent learning across views, but can limit flexibility, leading to poorer generalisation performance on unseen data.

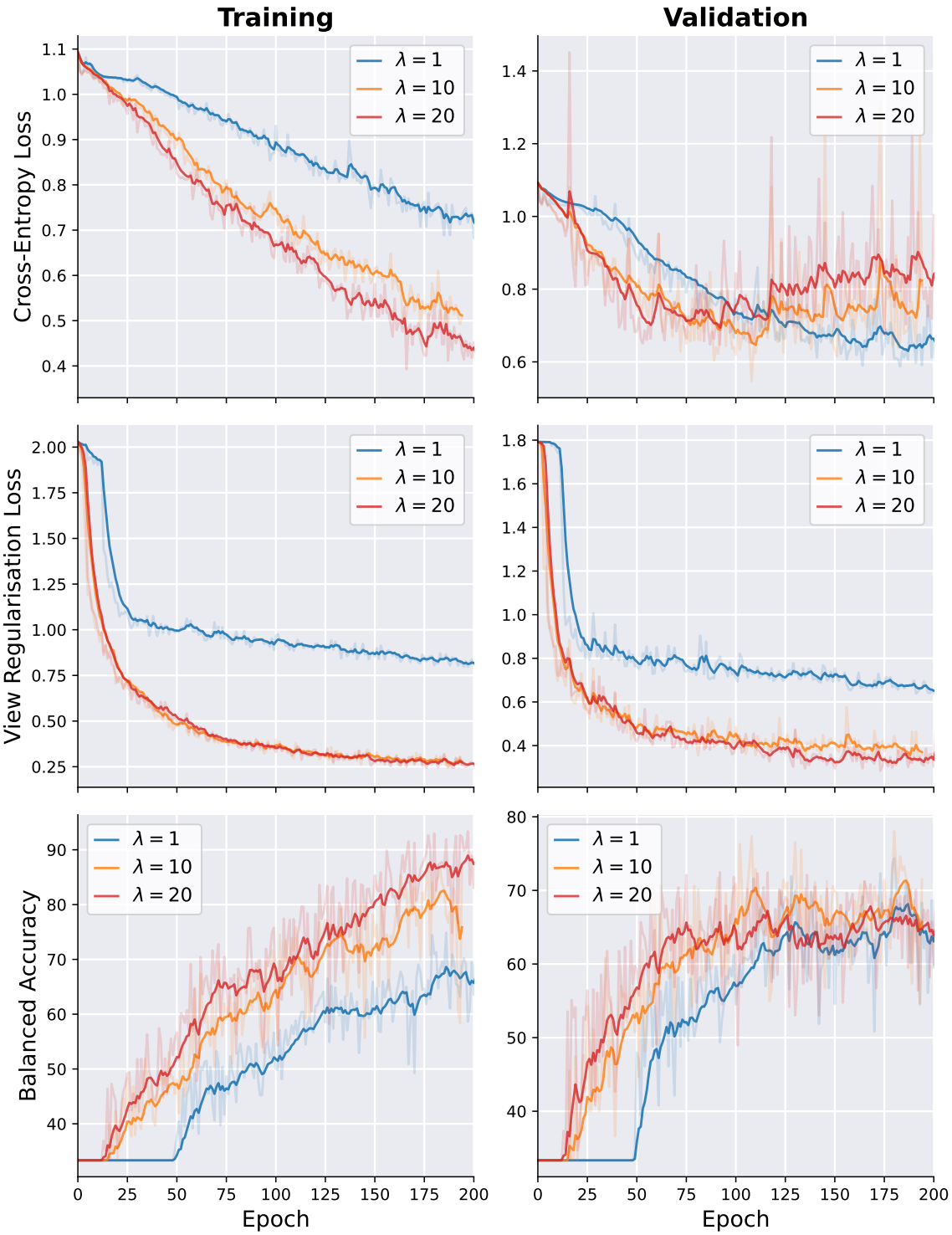


Figure 7.9: Training and validation performance of the fully convolutional transformer feature extractor multiple instance learning model with varying view regularisation strengths on split 3 of the Tufts Medical Echocardiogram Dataset. Cross-entropy loss, view regularisation loss, and balanced accuracy are shown from top to bottom. The left column presents the results for the training set, while the right column shows the results for the validation set. Smoothing was applied to the raw output values (seen in matching transparent colors) to help visualisation.

The inference results for the training, validation, and test sets of the models with the best validation accuracy in split 3 of the view and diagnosis labelled part of the TMED2 data set are summarised in Figure 7.10 using confusion matrices. Data for all splits are provided in

Appendix C. Class-specific recall and precision values are overlayed on diagonal entries to facilitate direct comparison of classwise performances. Additionally, balanced accuracy is reported for each set to assess the overall generalisation capability of both models.

The SAMIL model achieved high performance in the training set with a balanced accuracy of 95.25%, but this did not generalise well, with validation and test accuracies decreasing to 77.71% and 71.53%, respectively, indicating possible overfitting. In contrast, the FCT architecture demonstrated more uniform performance, with moderate training accuracy (80.25%) that was maintained across the validation (78.03%) and test (79.46%) sets. This stability suggests a better generalisation.

Class-specific metrics indicate that the FCT model achieved good accuracy for the early AS class in all splits and a high recall in the test set for the significant AS class (85%), which is clinically important for ensuring high sensitivity to advanced disease. The SAMIL architecture maintained good recall for the 'no AS' and 'significant AS' classes in both validation and test splits; however, a drop in recall for the early AS is observed in the test set.

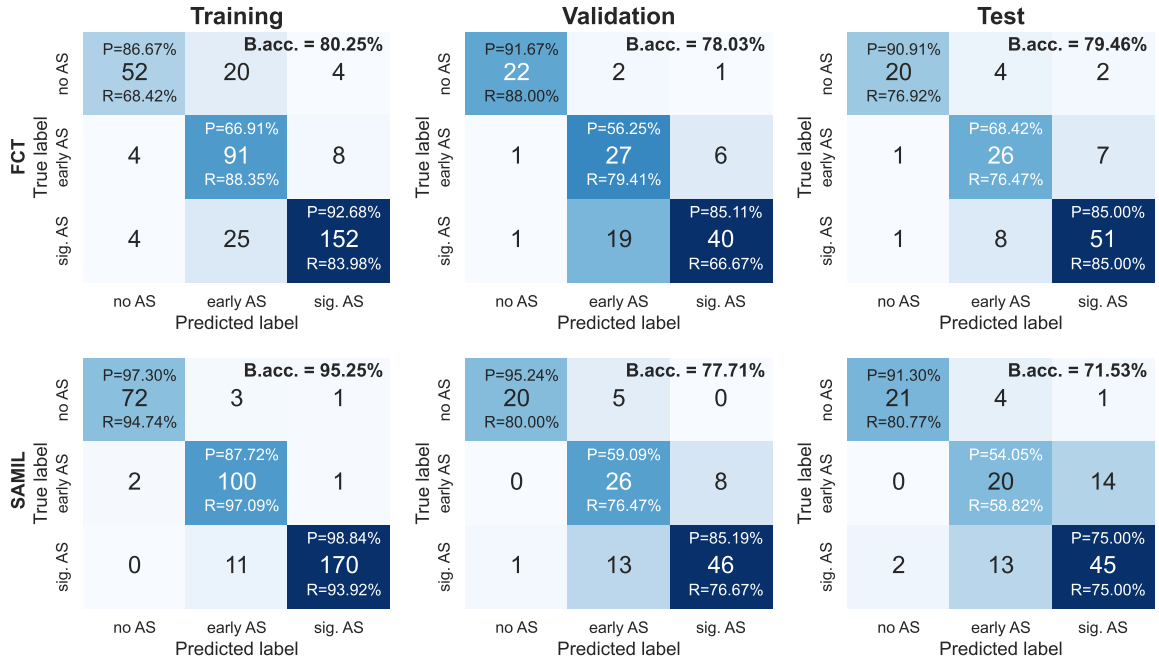


Figure 7.10: Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 3 of the Tufts Medical Echocardiogram Dataset. The columns contain the confusion matrices for the training, validation and test sets from left to right for the models selected at the best validation accuracy. Precision (P) and recall (R) values are provided for each class in the given set and balanced accuracy (B.acc.) is calculated for each set to assess overall generalisation performance.

ROC curves and corresponding AUC scores were used to evaluate the model's discriminative performance on binary classification tasks defined in the following severity groupings: i) no AS vs. some AS, ii) not significant AS vs. significant AS and iii) early AS vs. significant AS. Figure 7.11 shows these results for the models with the best validation accuracy in split 3 of the view and diagnosis labelled part of the TMED2 data set. Data for all splits are provided in Appendix C.

Across all data splits and models, the highest AUC values were consistently observed for the no AS vs. some AS classification task, indicating that differentiating between healthy and disease populations constitutes the most attainable learning problem. In contrast, the lowest AUCs were associated with the early AS vs. significant AS task across all splits and models, highlighting the difficulty in obtaining class-discriminative features in the advanced disease population. Among the models, the FCT architecture demonstrated robust performance across all task in the test set with AUC scores of 0.863 (early vs. significant AS), 0.891 (not significant vs. significant AS), and 0.953 (no vs. some AS). Although the SAMIL model achieved excellent test performance on the no AS vs. some AS task (AUC = 0.983), this resulted in reduced AUC scores on the other two tasks (AUC = 0.804 for early AS vs. significant AS and AUC=0.871 for not significant AS vs. significant AS). These results suggest that FCT preserved slightly better discriminative power with more robust performance across classification tasks.

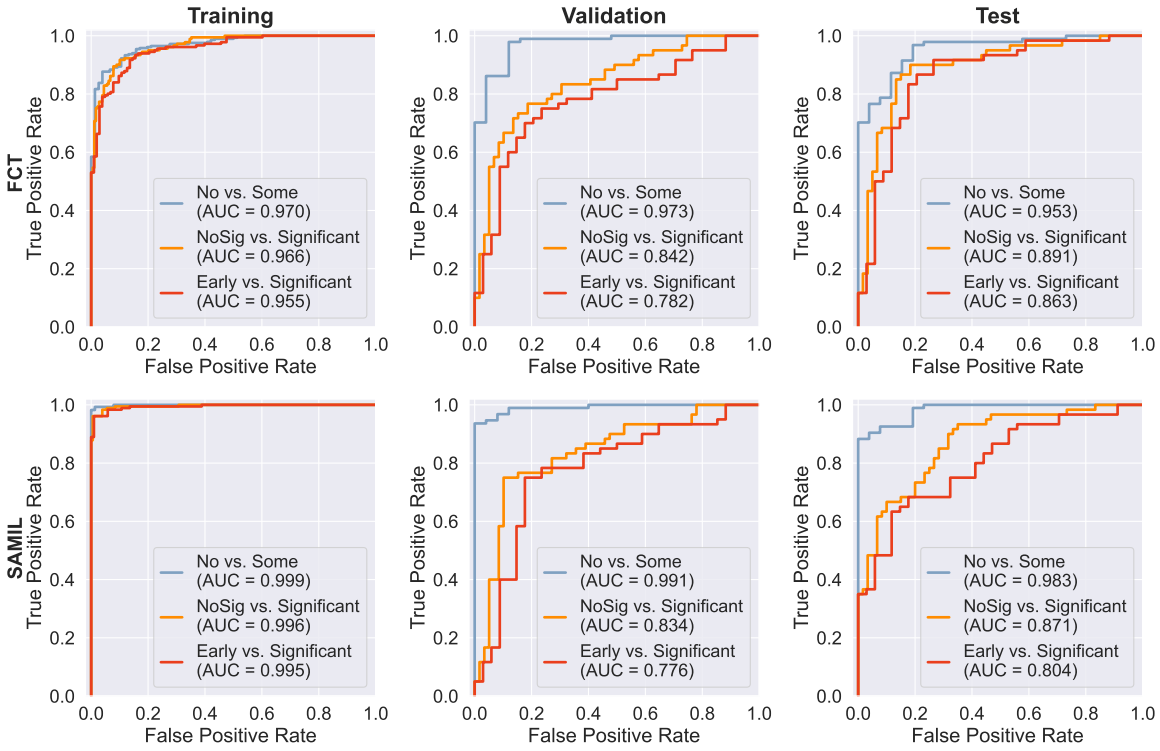


Figure 7.11: Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 3 of the Tufts Medical Echocardiogram Dataset. The columns contain the ROC curves for multiple binary classification tasks for the training validation and test sets for models selected at the best validation accuracy.

7.5.2 Attention Weight Analysis

To assess whether the models focused on clinically relevant views when making predictions, the view relevance scores r_k of the most attended images as ranked by the supervised attention scores a_k were analysed. Images with high attention scores (close to 1) are considered important by the model for predicting disease severity, while those with low scores (close to 0) have little influence on the final decision. By examining the view relevance scores of the most attended images ranked according to supervised attention scores, the model's ability to identify relevant views can be evaluated. Examples of the top ten images attended for the SAMIL and FCT architectures are shown in Figure C.7, included in Appendix C, which indicates that both models succeeded in prioritising the PSAX and PLAX views.

The above qualitative sanity check is complemented by measuring the relative view relevance scores of the top ten attended images over all the test examples and observing the average. Figure 7.12 shows the average attention scores for the top ten attended images in the test set of split 3 of the TMED2 data set, using the models with the best validation accuracy. The results indicate that the top attended images are assigned high view relevance scores (close to 1), which confirms that the models predict disease severity based on clinically relevant views.

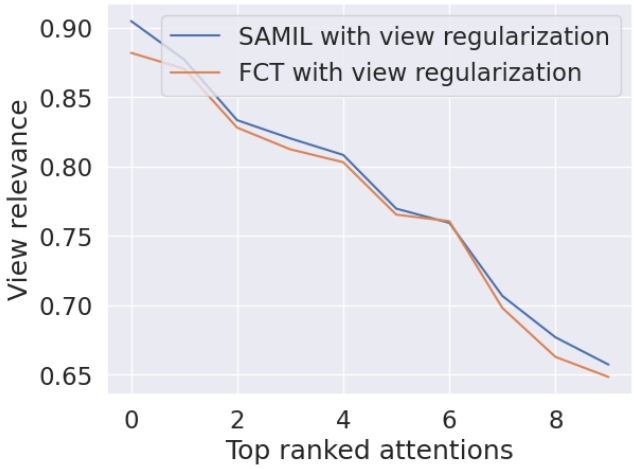


Figure 7.12: Average view relevance scores for the top 10 attended images across the test set in split 3 of the Tufts Medical Echocardiogram Dataset. Values close to 1 indicate that clinically relevant views (parasternal short axis and parasternal long axis) were used for the AS predictions.

It should be noted that in the MIL framework, the attention weights indicate which images contributed most strongly to the aggregation of instance-level features into the bag-level representation; however, they may be influenced by dataset-specific or architectural bias. The robustness of the observed attention patterns should be assessed using external data sets, while architecture-related bias could be further examined through complementary analyses such as occlusion-based perturbation tests, or masking experiments comparing the effect of removing highly attended and low-attention images on the predicted probabilities.

7.5.3 Saliency Maps

After demonstrating that the models prioritise PLAX and PSAX views to classify disease severity, the specific regions that contribute the most to the prediction within an image were examined using the Grad-CAM algorithm. The results for a single instance for each model in the PSAX and PLAX views are shown in Figure 7.13, backpropagating the significant AS class to the top attended images. The last convolutional layer was chosen for the target activation maps for both architectures and the generated CAMs were upscaled to match the dimensions of the input images.

For the images examined, the FCT model better localised the AV, focussing on the opening between the leaflets in the PLAX view, while the SAMIL architecture used part of the LVOT for its prediction. The localisation of the AV in the PSAX view was also more accurate for the FCT architecture in this example. Note that quantifying the accuracy of the AV localisation task would have required annotating the entire data set for the AV segmentation task, which was considered outside of the scope of this project.

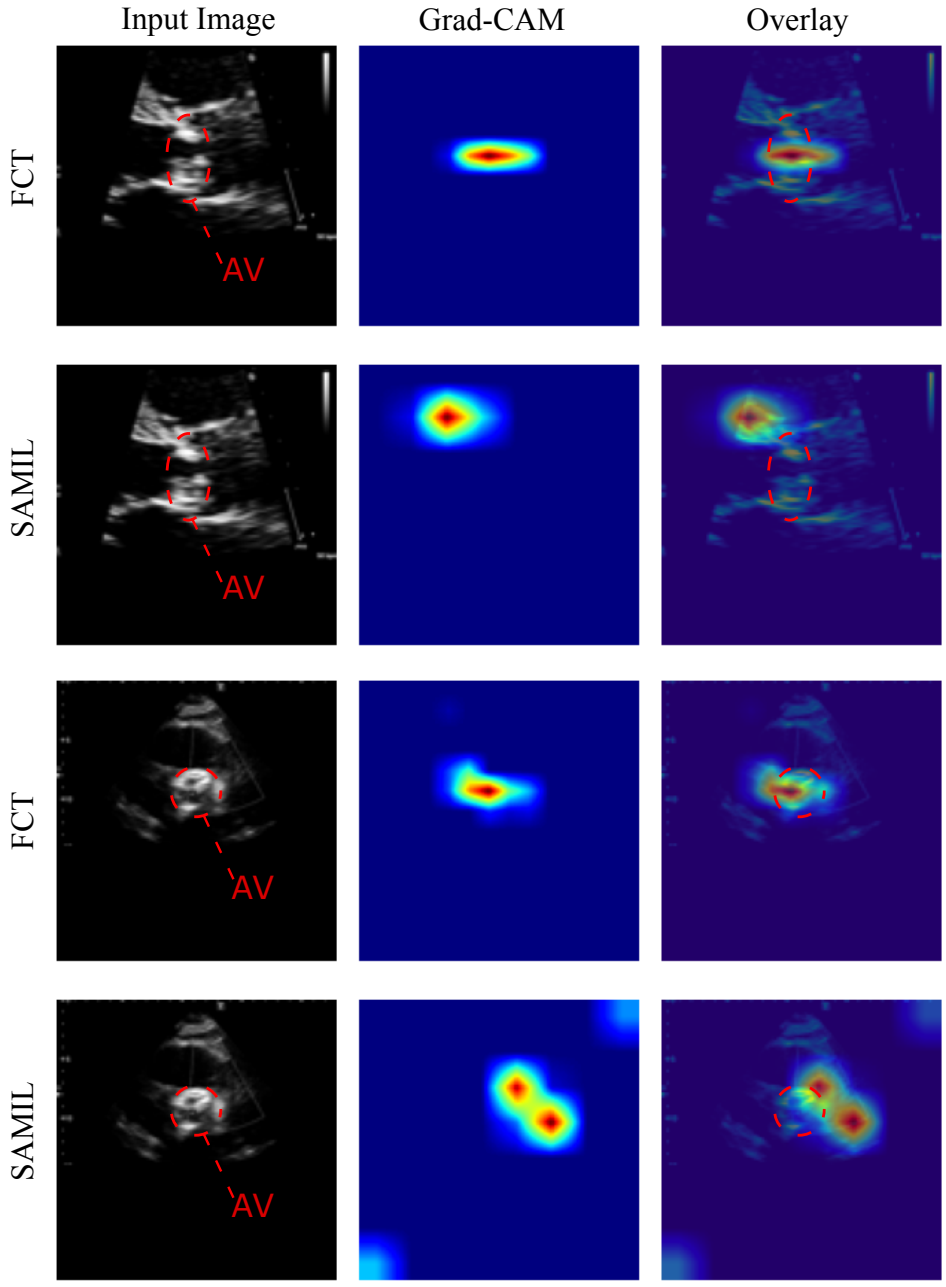


Figure 7.13: Gradient-weighted class activation maps (Grad-CAM) for the Fully Convolutional Transformer (FCT) feature extractor and the original supervised attention multiple instance learning (SAMIL) architectures. Backpropagation was performed for the the correctly predicted significant aortic stenosis (AS) severity class on the top attended input image.

7.6 Generalisation Performance on Clinical Data

7.6.1 Inference on Concordant Aortic Stenosis

Using the models with the best validation accuracy from each TMED2 data split, the concordant AS data set was used to assess their generalisation performance. The inference results for the zero-shot evaluation in the AS60 data set are illustrated in Figure 7.14. Class-specific precision and recall values are overlayed on the corresponding diagonal entries of the confusion matrices, and balanced accuracy is calculated as the average of the recall values for the early AS and significant AS classes in the absence of the no AS class.

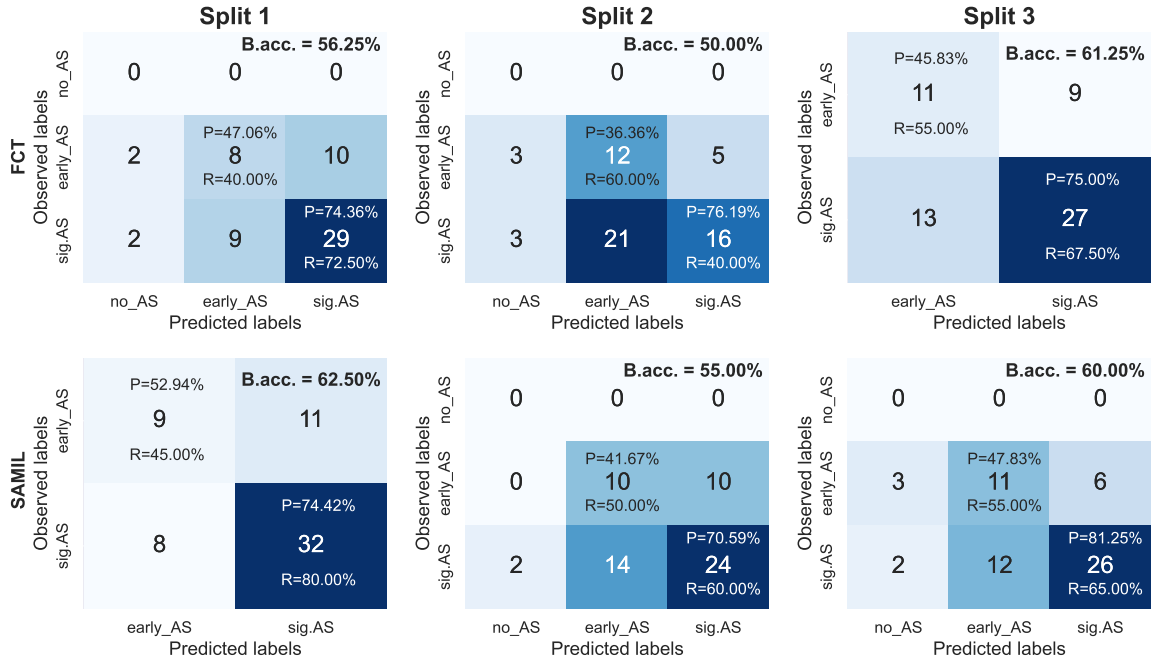


Figure 7.14: Zero-shot inference on the concordant aortic stenosis cohort using the models from all three data splits of the Tufts Medical Echocardiogram Dataset. Confusion matrices are shown for models with the best balanced accuracy (b.acc.) on the validation set from split 1, split 2 and split 3, as seen from left to right. The top row presents the results for the fully convolutional transformer (FCT) feature extractor architecture, while the bottom row shows the results for the supervised attention multiple instance learning (SAMIL) architecture. Class-specific precision (P) and recall (R) metrics are overlayed on the corresponding diagonal entries.

The results indicate that a significant drop in performance is present for both models as measured by the balanced accuracy metric. In split 3 of the TMED2 data set, the balanced accuracy decreased from 79.46% to 61.25% for the FCT, and from 71.53% to 60.0% for the SAMIL architectures. This reduction in accuracy is not only due to the absence of the easily identifiable no AS class, but also to a decrease in recall for both early AS (76.47% to 55% for split 3) and significant AS classes (85% to 68% for split 3), which could signify a shift in the input or joint distributions.

Because inference performance varied considerably across splits, particularly for the FCT model (with balanced accuracy ranging from 50% to 61.25%), model ensembling was applied to improve robustness. For the SAMIL model majority voting and probability averaging were used. For the FCT architecture additional MCD was used at inference using five random seeds. The inference results in the AS60 data set from aggregating the outputs of the corresponding models across TMED2 data splits and random seeds are presented in Figure 7.15.

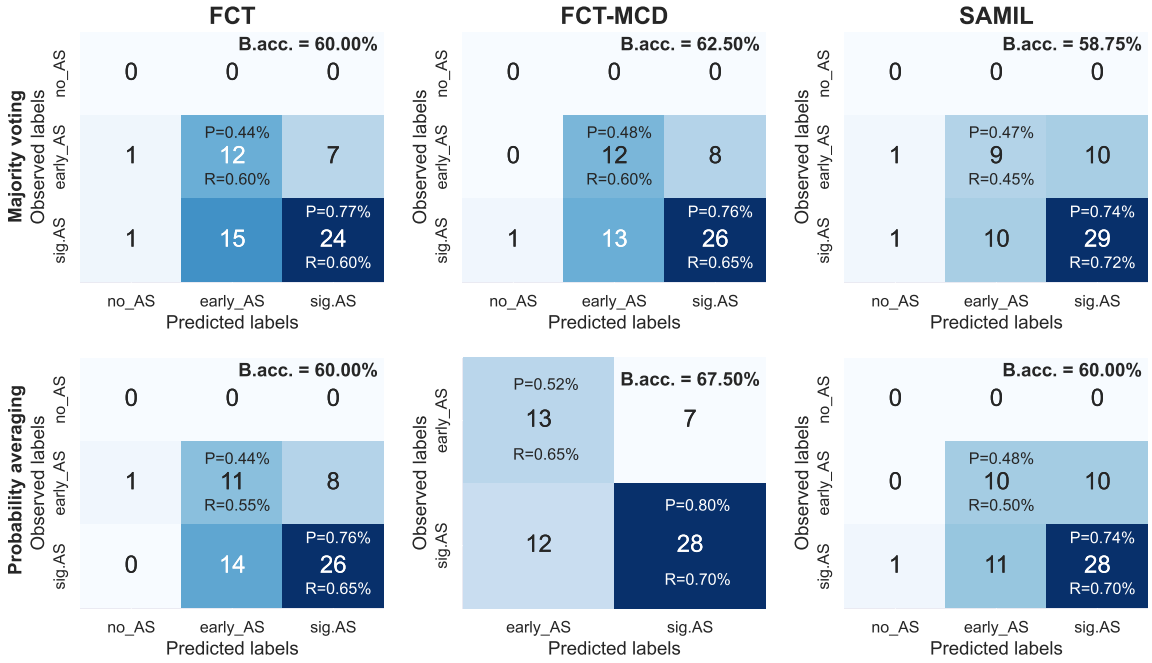


Figure 7.15: Model ensembling using majority voting [top] and probability averaging [bottom]. Stochastic inference via MCD was also conducted for Fully Convolutional Transformer (FCT) using five random seeds. Confusion matrices are shown for the FCT, FCT with MCD and SAMIL models from left to right.

The majority vote approach did not introduce a notable improvement for either model, as the highest balanced accuracy remained around 62.5%. However, the combination of probability averaging and MCD resulted in a significant increase in performance for the FCT architecture, reaching a balanced accuracy of 67.5%, which outperforms the best SAMIL test accuracy on the AS60 data set by 5%. Notably the higher recall value for the early AS class contributed the performance gains for the FCT architectures, indicating that the SAMIL model relied more on the no AS class to achieve a high balanced accuracy.

7.6.2 Inference on Low-Flow, Low-Gradient Aortic Stenosis

The trained models were then evaluated on the DSE12 data set, which consisted of echocardiographic recordings of 12 patients with LFLG AS who underwent DSE. This evaluation was performed to assess the generalisation capability of the models across disease phenotypes and to examine the utility of DSE in resolving diagnostic uncertainty. Most of the patients in the DSE12 data set had available resting and stress echocardiograms, while some patients (1, 4 and 5) only had resting recordings. The AS prediction and confidence scores for each model ensembling technique are summarised in Table 7.1. The number of correct predictions, as well as the mean log-likelihood (MLL) measured between the predicted probability distribution and the one-hot encoded class label, is presented in the bottom rows for both states.

Patient	State	Majority voting			Probability averaging		
		FCT	**FCTwD**	**SAMIL**	**FCT**	**FCT-MCD**	**SAMIL**
P01	Rest	1	1	1	[0.01, **0.79**, 0.19]	[0.01, **0.89**, 0.10]	[0.04, **0.66**, 0.30]
	All	1	1	1	[0.01, **0.79**, 0.19]	[0.01, **0.89**, 0.10]	[0.04, **0.66**, 0.30]
P02	Rest	1	1	1	[0.12, **0.76**, 0.13]	[0.05, **0.87**, 0.08]	[0.01, **0.69**, 0.30]
	All	1	1	1	[0.11, **0.74**, 0.15]	[0.06, **0.85**, 0.09]	[0.04, **0.70**, 0.27]
P03	Rest	1	1	1	[0.06, **0.56**, 0.38]	[0.02, **0.60**, 0.38]	[0.00, **0.51**, 0.49]
	All	1	1	1	[0.05, 0.46, **0.48**]	[0.02, **0.59**, 0.39]	[0.00, **0.52**, 0.48]
P04	Rest	1	1	1	[0.22, **0.73**, 0.06]	[0.25, **0.71**, 0.04]	[0.05, **0.78**, 0.17]
	All	1	1	1	[0.22, **0.73**, 0.06]	[0.25, **0.71**, 0.04]	[0.05, **0.78**, 0.17]
P05	Rest	0	2	2	[0.37, 0.19, **0.43**]	[0.33, 0.22, **0.45**]	[0.16, 0.26, **0.58**]
	All	0	2	2	[0.37, 0.19, **0.43**]	[0.33, 0.22, **0.45**]	[0.16, 0.26, **0.58**]
P06	Rest	1	2	2	[0.09, 0.38, **0.53**]	[0.05, 0.34, **0.62**]	[0.01, 0.42, **0.56**]
	All	1	1	2	[0.20, 0.36, **0.44**]	[0.12, 0.42, **0.46**]	[0.01, 0.41, **0.57**]
P07	Rest	2	2	2	[0.08, 0.16, **0.77**]	[0.02, 0.16, **0.82**]	[0.03, 0.13, **0.84**]
	All	2	2	2	[0.13, 0.12, **0.76**]	[0.03, 0.11, **0.87**]	[0.03, 0.10, **0.87**]
P08	Rest	2	2	2	[0.30, 0.06, **0.63**]	[0.22, 0.05, **0.74**]	[0.28, 0.23, **0.49**]
	All	0	2	0	[**0.67**, 0.03, 0.29]	[0.47, 0.03, **0.50**]	[**0.49**, 0.13, 0.37]
P09	Rest	1	2	2	[0.24, **0.38**, 0.38]	[0.13, 0.35, **0.52**]	[0.00, 0.39, **0.61**]
	All	2	2	2	[0.17, 0.31, **0.52**]	[0.12, 0.29, **0.59**]	[0.00, 0.37, **0.62**]
P10	Rest	2	2	2	[0.08, 0.26, **0.66**]	[0.06, 0.23, **0.71**]	[0.01, 0.40, **0.59**]
	All	2	2	2	[0.07, 0.21, **0.72**]	[0.04, 0.17, **0.80**]	[0.01, 0.29, **0.70**]
P11	Rest	2	2	2	[0.21, 0.12, **0.67**]	[0.10, 0.10, **0.80**]	[0.19, 0.24, **0.57**]
	All	2	2	2	[0.35, 0.07, **0.57**]	[0.28, 0.06, **0.67**]	[0.23, 0.23, **0.54**]
P12	Rest	2	2	1	[0.10, 0.14, **0.76**]	[0.02, 0.11, **0.86**]	[0.09, 0.34, **0.58**]
	All	2	2	2	[0.22, 0.16, **0.61**]	[0.11, 0.15, **0.74**]	[0.16, 0.27, **0.57**]
MLL	Rest	–	–	–	0.58	0.58	**0.50**
	All	–	–	–	0.63	0.62	**0.51**
Correct	Rest	5	**8**	7	7	**8**	**8**
	All	5	7	7	**8**	**8**	**8**

Table 7.1: Per-patient predictions under rest and stress for all models. Hard voting columns report predicted class labels only. Soft voting columns report the class-probability distributions across the three classes, with the predicted class (maximum probability) shown in bold. Bottom rows show mean log-loss (MLL) and number of correct predictions per state.

The results indicate that DSE did not positively affect the number of correct predictions or the MLL for either ensembling approach. There is also no difference between the number of correct predictions across FCT and SAMIL architectures when probability averaging and MCD is used.

7.6.3 Domain Shift Analysis

Lastly, the shift in distribution across different domains is examined visually by plotting the class-wise feature distribution of each domain and quantitatively by the class-wise FD distance. The low-rank approximation of the bag-level representations of each study in the different domains was obtained using PCA. The projections of the bag-level representations on the first two PCs are shown in Figure 7.16 for each domain.

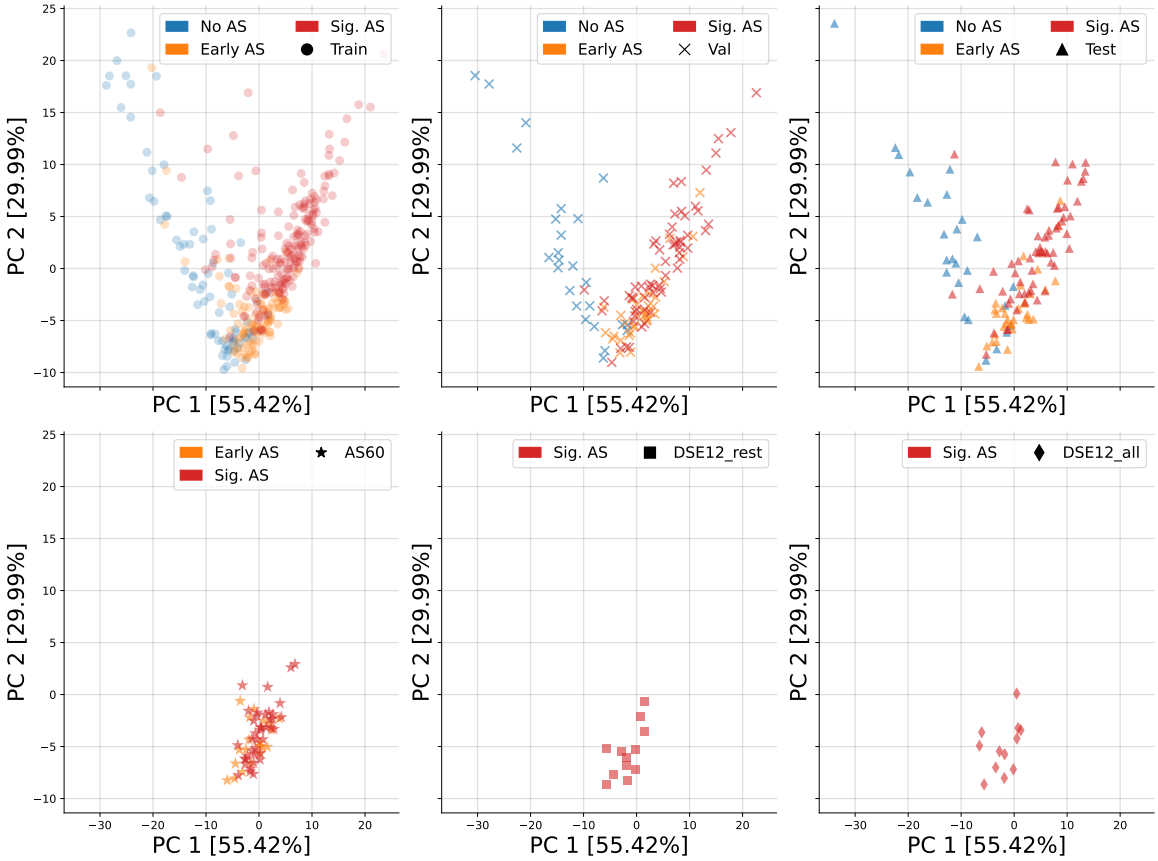


Figure 7.16: The class-wise distribution of the low rank approximations of bag-level representations in different data sets using the first two principal components as depicted for the FCT architecture in split 3.

It can be seen that the separation of class labels in the TMED2 data set is prominent, with visible clusters along both PCs. This is not observed for the AS60 data set, where both the early and significant AS classes occupy the same feature space as the early AS in the TMED2 data set, indicating that the decrease in the early AS and the significant AS recall values (as seen earlier in Figure 7.14) is caused by domain shift. The DSE12 representations follow a similar distribution to that seen in the AS60 data set for the significant AS class.

To quantify the degree to which domain shift occurs, the class-wise FD metrics are presented in Figure 7.17 comparing the different representations in the TMED2, AS60, and DSE12 data sets. The FCT and SAMIL architectures were chosen by the best validation accuracy in split 3 of the TMED2 data set.

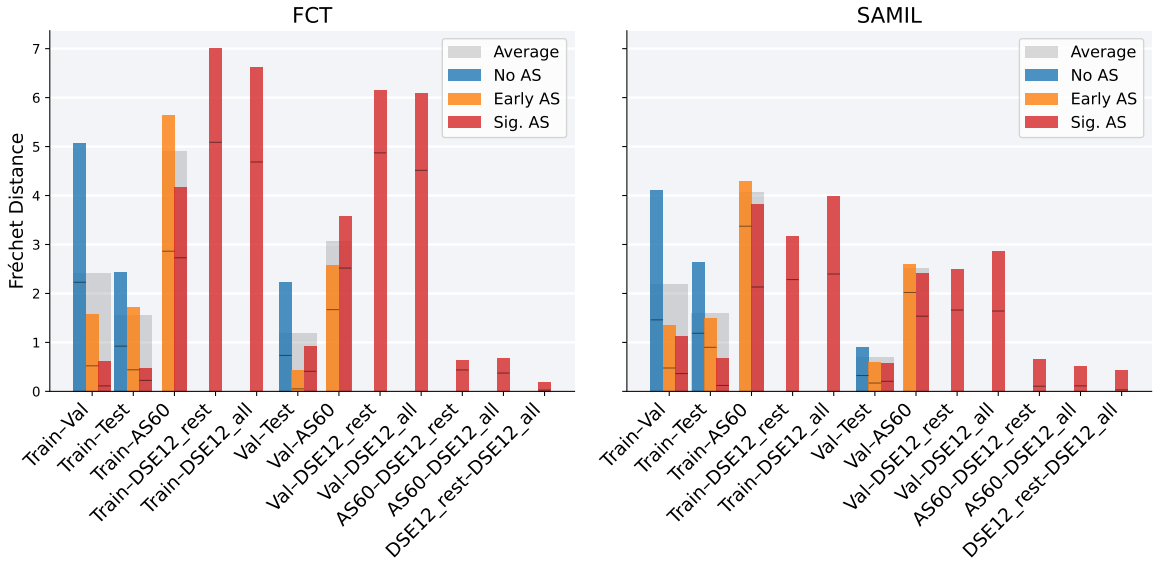


Figure 7.17: Domain shift analysis showing the classwise Fréchet Distance (FD) across different domains. The horizontal line indicates the value of the centroid shift, while the shaded areas represent the average FD distance across the classes.

The results show that there is a significant shift from the representations learnt in the training data of the TMED2 data set to the representations obtained in the DSE12 data set by the FCT architecture. This shift is also present when the DSE12 patients are compared with the validation set of TMED2. Less prominent is the shift from the training set to the AS60 data and from the validation set to the AS60 data, however, these values still indicate a high domain shift. The domain distance is small between the AS60 and DSE12 patients, confirming that the recordings came from the same acquisition setup.

7.7 Discussion and Conclusion

In this chapter, a novel DL architecture was developed to predict AS from TTE recordings by incorporating transformer-based feature extractors into the MIL framework. The developed model showed slight gains in the public TMED2 data set compared to the convolutional feature extractor and outperformed the SAMIL model (Huang et al., 2024) in the AS60 data set. The generalisation performance of the proposed architecture to concordant AS was good, suggesting that DL models trained with this AS detection objective can be applied in a clinical setting. However, further performance improvements are expected to support clinical application as a screening tool, possibly with the inclusion of an additional temporal dimension, as shown by Ahmadi et al. (2024).

The models also passed multiple explainability tests. The AS diagnosis labels were assigned based on clinically relevant PLAX and PSAX views. Using the Grad-CAM algorithm, the proposed model also showed gains in the detection of the AV, without explicitly training on this task. This finding suggests that the AS classification objective is linked to the pathology of the AV at the image level, and DL models are suited to obtain imaging markers related to AV pathology directly from echocardiographic recordings.

The evaluation of the models on LFLG AS was limited due to the sample size and the merged class labels for moderate and severe AS in the available training data set. However, domain shift analysis indicated that patients with LFLG AS resemble concordant AS in the learnt latent space representation, and training data should therefore include discordant cases of AS to allow the derivation of different classification thresholds for different phenotypes of AS.

8. Conclusion

This chapter includes the overall discussion and the conclusion of this work. First, the key findings are summarised in Section 8.1. Next, suggestions and future directions of the presented studies are covered in Section 8.2. Finally, closing remarks are presented in Section 8.3.

8.1 Key Findings

This research investigated whether resting echocardiographic predictors alone could reliably determine the severity of AS and further assessed the additive diagnostic value of DSE. A multifaceted approach was adopted to explore a wide range of computational methods that could be applied to routine echocardiographic recordings. These methods were presented in three dedicated chapters.

In Chapter 5 a comprehensive statistical analysis of standard and novel echocardiographic predictors was performed in relation to both the severity of AS and hemodynamic states. This analysis confirmed that the predictors examined under resting conditions do not differentiate true-severe from pseudo-severe AS. It also highlighted that, while the resting echocardiographic predictors in the LFLG population differ from those seen in concordant AS, DSE shifts the distribution of these predictors to be more aligned with those in concordant AS, thus demonstrating the clinical relevance of DSE in resolving diagnostic uncertainty. In contrast, no class-specific stress response has been observed in this data set; however, the statistical power of this specific test was considered low. Together, these results indicate that a more refined approach is required to differentiate true-severe from pseudo-severe AS at rest than only observing global anatomical and hemodynamic markers.

In addition to statistical analysis, ML algorithms were also developed to predict the severity of AS using a reduced feature space derived from PCA and LDA applied to routine and novel echocardiographic predictors. This work was the first to characterise the stress response of patients with LFLG AS in a reduced feature space and to evaluate the generalisation of learnt decision boundaries to this phenotype from a concordant AS cohort. Although the models performed well on synthetic validation data with an accuracy of 79.3%, they were unable to generalise to LFLG AS, neither at rest nor under stress states. Nevertheless, model performance under stress conditions was improved, further supporting the added diagnostic value of DSE in this subgroup. These results highlight the importance of developing phenotype-

specific classification thresholds for standard clinical markers to improve the utility of DSE in differentiating true-severe and pseudo-severe AS.

In Chapter 6 the development of a multiplanar reconstruction and image registration pipeline was carried out to estimate regional LV myocardial deformation parameters. The validation results against 3D STE data demonstrated strong agreement for segmental longitudinal and circumferential strains, confirming the accuracy and clinical potential of the proposed method. The proposed pipeline can support clinical applicability in more challenging imaging protocols, such as DSE, and also allow retrospective analysis of triplane data to investigate regional myocardial mechanics in various cardiac pathologies.

In Chapter 7 a novel DL architecture was developed to predict AS from TTE recordings by incorporating transformer-based feature extractors into the MIL framework. The created model showed slight gains on the public TMED2 data set compared to the convolutional feature extractor and outperformed the SAMIL model on the AS60 data set. The proposed model also passed multiple explainability tests, supporting the development of DL models capable of extracting imaging markers related to AV pathology directly from echocardiography.

In addition to the findings related to the methodological advancements, several insights were gained with regard to the optimal design choices for future clinical studies. Since all models relied on AS severity labels assigned according to current clinical guidelines, all improvements in learnt representations are tied to these classification rules and do not reflect more clinically meaningful outcomes such as hospitalisation, valve replacement, or mortality. To enable learning discriminative features in relation to these outcomes, additional prospective studies are needed that record adverse events along with standard echocardiographic measurements. Furthermore, guideline-based thresholds are calibrated under normotensive blood pressure and do not generalise to patients with elevated afterload due to hypertension, further restricting their applicability. To adopt a more robust marker of disease severity, CT calcium scoring could be used to define valve pathology and AS grade more reliably, thus reducing diagnostic ambiguity.

Finally, the classification models were trained exclusively in patients with concordant AS as seen in multiple previous studies related to AS prediction (Huang et al., 2025; Park et al., 2025; Ahmadi et al., 2024). This excludes AS cases with altered hemodynamic states, such as LFLG AS, thus preventing the models from learning phenotype-specific decision rules. Consequently, the models could only be evaluated on their ability to generalise the learnt decision boundaries to LFLG AS and whether DSE shifted representations to better align with concordant decision patterns. However, classification frameworks should account not only for AS severity grades, but also for distinct phenotypes, as demonstrated by Sengupta et al. (2021).

8.2 Suggestions and Future Directions

8.2.1 Methodology Improvements

Several improvements can be made to enhance model performance or improve robustness. This section outlines some of these approaches for each method that can be implemented while remaining within the constraints of the currently available data sets.

Analysis of Clinical Predictors

Although a comprehensive statistical analysis framework was developed to test various research hypotheses, the set of ML algorithms explored was relatively limited. Given the hierarchical nature of the current clinical grading algorithm, there is an opportunity to refine threshold-based classification using tree-based models, such as decision trees, XGBoost (Chen and Guestrin, 2016), or random forests. These algorithms offer strong performance on structured data and can also be used for feature selection. In addition, they may also provide more interpretable decision rules by optimising clinical thresholds for the selected features.

Left Ventricular Strain Analysis

The current mesh generation algorithm was developed using myocardial midline contours to create triangular surface meshes. However, this approach prohibits the measurement of radial strains. By incorporating the endocardial and epicardial borders of the LV, three-dimensional volumetric meshes could be produced to estimate the myocardial thickness, mass, and regional deformation parameters measured in the radial direction. Additionally, the myocardial dispersion coefficient could be included in the qualitative analysis to measure LV dyssynchrony based on the variability in time-to-peak segmental strain values.

The multiplanar reconstruction technique could also be extended to include additional SAX planes. This would not only enable observation of the true in-plane components of radial and circumferential strains, but would also allow a more constrained mesh-fitting approach. The mesh generation would then entail an optimisation problem in which a parametric LV model is fitted to the in-plane contours of the LV, and the distance between the estimated mesh and the annotations is minimised. Such approaches have been explored by Joyce et al. (2022), demonstrating volumetric mesh inference from MRI data to support three-dimensional LV deformation analysis. The image registration framework could also be further optimised to reduce the accumulation of additive noise introduced during frame-to-frame propagation of the initial mesh.

Deep Learning

The proposed model was trained using a supervised learning objective, which limited the amount of data that could be used for training to the fully labelled part of the TMED2 data set. Using a semi-supervised learning objective, the unlabelled part of the data set could also be used for study-level pre-training. This strategy has been demonstrated by Huang et al. (2025) to yield additional performance gains in the classification of AS severity. Furthermore, the use of a loss function tailored to ordinal classification could improve prediction consistency by enforcing a natural order of disease severity.

In addition to improving predictive performance, future work should consider how model uncertainty is communicated alongside the predicted AS severity label. Although the DL model developed in this thesis is not intended for immediate clinical deployment, future translation would require predictions to be accompanied by calibrated confidence scores or flags indicating low-confidence or borderline cases. In this way, uncertain predictions could be used to prompt additional investigations, such as referral for DSE or CT calcium scoring, where clinically appropriate, and future work would be required to calibrate the predicted probabilities, validate uncertainty thresholds, and determine how these thresholds should guide referral decisions.

8.2.2 Broader Research Opportunities

This section introduces broader research avenues that are outside the scope of this thesis but are closely related to the current research question. These directions would require additional data to implement or alternative methodologies to be considered.

Temporal data

DL models could benefit from expanding the data set to include an additional time dimension, as demonstrated by Ahmadi et al. (2024) for the AS classification task. Future directions for DL using temporal echocardiographic data could also involve reconstruction of 3D LV geometry from multiple 2D B-mode recordings, as demonstrated by Laumer et al. (2025).

In addition, patient-specific hemodynamic patterns could be examined by recording ventricular and atrial pressure and volume curves, flow rates over time, and the timing of valve opening and closure, and incorporating these data into a lumped parameter model Virag and Lulić (2008). These models can be used to estimate key parameters of the cardiovascular system, such as vascular resistance, arterial compliance, chamber elastance, and valvular impedance. By fitting such models to temporal data, it becomes possible to simulate the dynamic behaviour of the circulatory system under patient-specific conditions, providing deeper physiological insight in the progression of AS Laubscher et al. (2022).

Multimodal data

The use of additional imaging data from different modalities, such as CT or MRI can be used to improve the diagnostic utility of TTE. For example, predicting CT-derived calcium scores from TTE recordings could provide a non-invasive marker of valve pathology, particularly useful in patients with discordant AS phenotypes. Alternatively, composite markers could be obtained by multimodal data fusion, as demonstrated by Puyol-Antón et al. (2017) to derive LV myocardial strain parameters by combining ultrasound and MRI recordings.

Integrated approach

Finally, future work could also explore the integration of the statistical, biomechanical, and DL approaches developed in this thesis within a unified diagnostic framework. Rather than combining the methods only through a late-stage ensemble of predictions, a more interpretable strategy would be to use each method for a distinct task. For example, DL models could estimate echocardiographic view, probe position, and orientation, thereby supporting patient-specific geometric reconstruction from routine 2D or 3D recordings. These reconstructed geometries could then be used to develop biomechanical models of LV and valve function, from which clinically relevant parameters such as regional strain, myocardial work, pressure estimates, compliance, and flow-related measures could be derived. Statistical and ML methods could subsequently be applied to these biomechanical parameters or to latent DL representations for AS severity classification. This approach would allow each stage of the pipeline to be validated independently while still enabling information from imaging, biomechanics, and clinical predictors to be combined, which may provide a more interpretable and clinically tractable alternative to a purely end-to-end DL model.

8.3 Final Remarks

This research has investigated the potential of using resting echocardiographic recordings to accurately classify the severity of LFLG AS. Although no specific resting echocardiographic features were identified that could consistently distinguish between severe and non-severe LFLG AS, the developed AS classification algorithms demonstrated promising generalisation to both synthetic validation data and clinical data from patients with concordant AS. The proposed methods also advanced previous work in several key areas: i) the use of supervised ML algorithms applied to a reduced feature representation of both standard and novel clinical predictors, ii) the development of an image registration and multiplanar reconstruction pipeline to estimate regional LV myocardial deformation parameters from triplane echocardiography, and iii) the design of a novel transformer-based MIL model for AS classification.

This multifaceted approach established a solid foundation for future work aiming to provide a comprehensive solution to the echocardiographic assessment of AS. The development of such comprehensive AS screening systems is likely to incorporate a mixture of the examined methods, where global and spatio-temporal markers, as well as raw imaging data, are used to aid clinical decision making. This effort will require collaborative efforts from both clinicians and researchers not only to standardise echocardiographic acquisition protocols, but also to reduce measurement variability in routine predictors and to define common guidelines for data collection, model development, and clinical validation.

A. Appendix A

DSE12 patient	1 base	1 stress	2 base	2 stress	3 base	3 stress	4 base	4 stress	5 base	5 stress	6 base	6 stress
Severity	Moderate		Moderate		Severe		Moderate		Moderate		Moderate	
AV Vmax [m/s]	3.120	x	2.120	x	2.350	x	2.810	3.200	2.310	3.190	2.910	x
AV Vmean [m/s]	2.260	x	1.490	x	1.690	x	1.870	1.890	1.550	1.980	1.810	x
AV maxPG [mmHg]	39.040	x	18.120	x	22.000	x	31.670	40.960	21.390	40.710	33.790	x
AV mean PG [mmHg]	20.430	x	8.880	x	13.000	x	13.988	14.280	9.590	15.730	16.230	x
AV env. Time [ms]	293.580	x	307.340	x	301.040	x	298.500	243.050	337.960	265.480	341.410	x
AV VTI [cm]	66.300	x	45.870	x	50.790	x	58.560	45.920	52.120	48.690	61.910	x
AV HR [BPM]	66.000	x	69.000	x	65.850	x	72.000	138.000	46.000	74.000	60.000	x
AV/LVOT	4.043	x	3.201	x	4.837	x	2.207	1.754	2.635	2.245	5.005	x
LVOT Vmax [m/s]	0.740	x	0.740	x	0.540	x	1.100	1.400	0.850	1.280	0.690	x
LVOT Vmean [m/s]	0.500	x	0.460	x	0.430	x	0.760	0.940	0.630	0.850	0.490	x
LVOT maxPG [mmHg]	2.190	x	2.190	x	1.000	x	4.870	7.880	2.870	6.560	1.900	x
LVOTmean PG [mmHg]	1.000	x	0.850	x	1.000	x	2.350	3.590	1.600	2.900	1.060	x
LVOT env. Time [ms]	325.680	x	310.180	x	242.650	x	342.590	277.770	312.030	254.600	253.750	x
LVOT VTI [cm]	16.400	x	14.330	x	10.500	x	26.530	26.180	19.780	21.690	12.370	x
LVOT HR [BPM]	60.000	x	72.000	x	69.000	x	88.000	134.000	47.000	76.000	59.000	x
AVA [cm2]	0.590	x	1.187	x	0.935	x	1.423	1.790	0.640	0.753	0.900	x
AVAi [cm2/m2]	0.332	x	0.621	x	0.500	x	0.745	0.937	0.335	0.394	0.415	x
BSA [m2]	1.777	x	1.910	x	1.871	x	1.910	1.910	1.910	1.910	2.175	x
LVOT diam [cm]	2.200	x	2.200	x	2.400	x	2.000	2.000	2.400	2.400	2.400	x
LVIDd [cm]	5.300	x	5.000	x	5.500	x	5.600	5.600	5.200	5.200	5.980	x
IVSd [cm]	1.000	x	1.100	x	1.200	x	2.000	2.100	0.900	0.900	1.430	x
LVPWd [cm]	0.900	x	1.100	x	1.100	x	0.700	0.800	0.800	0.800	1.450	x
EDV [ml]	135.343	x	118.243	x	147.421	x	153.664	153.664	129.507	129.507	178.360	x
ESV [ml]	73.033	x	63.798	x	98.421	x	70.360	71.459	40.070	31.434	122.428	x
SV [ml]	62.310	x	54.445	x	49.000	x	83.304	82.205	89.437	98.074	55.932	x
EF [%]	0.460	x	0.460	x	0.332	x	0.542	0.535	0.691	0.757	0.314	x
CO [l/min]	3.926	x	3.838	x	3.304	x	6.664	11.180	4.159	7.356	3.328	x
Svi [ml/m2]	35.063	x	28.505	x	26.193	x	43.615	43.039	46.826	51.347	25.721	x
MG/EOA [mmHg/cm2]	34.627	x	7.481	x	13.907	x	9.833	7.977	14.984	20.891	18.033	x
Q [ml/s]	212.243	x	177.150	x	162.769	x	279.076	338.223	264.639	369.420	163.827	x
AT [ms]	116.780	x	103.700	x	133.330	x	141.660	110.000	136.900	123.890	104.250	x
AT/ET	0.398	x	0.337	x	0.443	x	0.475	0.453	0.405	0.467	0.305	x

Figure A.1: Clinical predictors measured in the DSE12 data set

7 base	7 stress	8 base	8 stress	9 base	9 stress	10 base	10 stress	11 base	11 stress	12 base	12 stress
Severe		Inconclusive		Severe		Severe		Moderate		Moderate	
2.710	3.350	2.780	3.600	3.090	3.650	3.250	3.550	2.520	2.760	2.650	3.300
1.790	2.280	1.930	2.600	2.230	2.670	2.240	2.270	1.880	1.880	1.780	2.320
29.410	45.120	31.100	52.000	38.210	46.500	42.250	50.500	25.450	30.550	28.180	43.810
12.800	20.790	14.920	30.000	19.950	53.380	20.120	20.650	14.150	14.280	12.700	21.530
324.070	296.290	361.110	279.600	328.700	315.780	342.600	296.300	263.890	254.630	324.000	236.110
57.950	67.610	69.770	73.400	73.380	77.960	76.820	67.320	49.490	50.000	57.790	54.750
51.000	90.000	58.000	76.000	72.000	78.000	63.000	81.000	99.000	111.000	73.000	98.000
3.129	4.410	3.338	3.191	3.241	3.154	3.069	2.981	3.295	3.287	3.817	2.905
0.960	0.960	0.900	1.100	0.930	0.940	1.140	1.130	0.830	0.920	0.750	0.970
0.630	0.640	0.600	0.800	0.650	0.610	0.680	0.760	0.550	0.630	0.460	0.620
3.700	0.680	3.000	5.000	3.430	3.530	5.200	4.270	2.760	3.420	2.240	3.780
1.570	1.640	2.000	3.000	1.690	1.480	1.840	2.330	1.200	1.590	0.850	1.550
296.290	240.700	327.900	279.600	347.220	407.400	370.370	296.290	273.140	240.740	328.700	300.920
18.520	15.330	20.900	23.000	22.640	24.720	25.030	22.580	15.020	15.210	15.140	18.850
50.000	68.000	63.000	70.000	72.000	74.000	64.000	84.000	95.000	120.000	75.000	92.000
1.200	0.942	0.849	0.970	1.130	1.161	0.829	0.853	0.772	0.774	1.080	1.422
0.569	0.446	0.437	0.499	0.592	0.608	0.434	0.447	0.404	0.405	0.565	0.745
2.109	2.109	1.944	1.944	1.910	1.910	1.910	1.910	1.910	1.910	1.910	1.910
2.300	2.300	1.900	1.900	2.160	2.160	1.800	1.800	1.800	1.800	2.000	2.000
6.300	6.300	5.400	5.400	5.900	5.900	4.600	4.600	4.700	4.700	4.600	4.600
1.400	1.300	0.900	0.900	0.900	0.900	0.900	0.900	1.200	1.200	1.400	1.400
1.100	1.000	0.800	0.800	1.300	1.300	0.700	0.700	0.700	0.700	1.200	1.200
201.187	201.187	141.314	141.314	173.211	173.211	97.336	97.336	102.361	102.361	97.336	97.336
124.280	137.527	82.086	65.314	90.292	82.674	33.675	39.906	64.159	63.676	49.796	38.147
76.907	63.660	59.227	76.000	82.919	90.537	63.661	57.430	38.202	38.685	47.540	59.189
0.382	0.316	0.419	0.538	0.479	0.523	0.654	0.590	0.373	0.378	0.488	0.608
3.884	5.029	3.583	5.548	5.970	6.881	4.042	4.738	3.706	4.468	3.518	5.623
36.460	30.180	30.460	39.086	43.413	47.402	33.331	30.068	20.001	20.254	24.890	30.989
10.667	22.080	17.576	30.928	17.655	45.965	24.279	24.206	18.331	18.457	11.759	15.136
237.316	214.857	164.015	271.817	252.263	286.709	185.818	193.824	144.764	151.927	146.727	250.684
127.650	118.200	132.380	85.580	134.660	115.780	112.150	110.320	112.420	98.650	119.400	100.940
0.394	0.399	0.367	0.306	0.410	0.367	0.327	0.372	0.426	0.387	0.369	0.428

Figure A.2: Clinical predictors measured in the DSE12 data set continued

AS60	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
Severity	Mild	Mild	Mild	Mild	Mild	Mild	Mild	Mild	Mild	Mild
AV Vmax [m/s]	2.400	1.700	2.300	1.830	2.260	2.250	2.090	2.530	2.390	1.550
AV Vmean [m/s]	1.270	1.210	1.670	1.230	1.610	1.600	1.500	1.950	1.690	1.020
AV maxPG [mmHg]	23.000	11.500	21.000	13.370	20.410	20.310	17.440	26.000	23.000	9.600
AV mean PG [mmHg]	9.000	6.000	13.000	6.000	10.420	10.300	9.000	17.000	13.000	6.000
AV env. Time [ms]	406.110	296.290	262.010	270.830	388.880	358.880	314.810	317.030	303.930	224.530
AV VTI [cm]	51.468	30.560	43.747	33.350	55.890	53.550	41.130	61.755	51.257	23.510
AV HR [BPM]	47.510	55.000	75.000	79.000	64.000	72.000	75.000	61.070	57.340	75.000
AV/LVOT	3.170	1.841	3.420	1.724	3.347	2.738	1.582	2.670	2.080	2.157
LVOT Vmax [m/s]	0.850	0.770	0.700	1.030	0.710	0.830	1.250	0.960	1.150	0.740
LVOT Vmean [m/s]	0.520	0.500	0.520	0.750	0.470	0.510	0.920	0.660	0.860	0.520
LVOT maxPG [mmHg]	3.000	2.000	2.000	4.000	2.000	3.000	6.000	4.000	5.000	2.000
LVOTmean PG [mmHg]	1.000	1.000	1.000	2.000	1.000	1.000	4.000	2.000	3.000	1.000
LVOT env. Time [ms]	310.920	314.820	248.030	284.060	356.330	314.810	284.130	348.470	333.330	244.800
LVOT VTI [cm]	16.200	16.600	12.800	19.340	16.700	19.560	26.000	23.100	24.660	10.900
LVOT HR [BPM]	51.000	63.000	76.000	79.000	65.000	73.000	89.000	57.000	59.000	67.000
AVA [cm2]	1.307	1.706	1.370	2.203	1.135	1.147	1.791	1.283	1.780	1.052
AVAi [cm2/m2]	0.684	0.893	0.650	1.030	0.538	0.600	0.871	0.650	0.890	0.652
BSA [m2]	1.910	1.910	2.108	2.140	2.111	1.910	2.057	1.941	2.000	1.613
LVOT diam [cm]	2.300	2.000	2.440	2.200	2.200	2.000	1.900	2.090	2.170	1.700
LVIDd [cm]	5.000	4.200	4.200	4.900	4.100	4.578	4.300	4.600	4.800	3.100
IVSd [cm]	1.100	1.100	1.400	1.300	1.000	1.287	2.500	1.600	1.200	1.000
LVPWd [cm]	1.000	1.600	1.200	0.900	1.100	1.190	1.200	1.000	0.700	0.900
EDV [ml]	118.243	78.578	77.000	111.000	74.223	96.265	80.850	97.336	107.520	37.916
ESV [ml]	22.321	32.212	34.000	24.606	36.369	57.759	34.670	46.721	47.436	16.015
SV [ml]	95.922	46.366	43.000	86.394	37.854	38.506	46.180	50.615	60.084	21.901
EF [%]	0.811	0.590	0.558	0.778	0.510	0.400	0.571	0.520	0.559	0.578
CO [l/min]	4.725	2.736	3.247	6.825	2.442	2.792	3.787	2.988	3.495	1.555
Svi [ml/m2]	50.211	24.271	20.401	40.376	17.931	20.156	22.446	26.078	30.042	13.579
MG/EOA [mmHg/cm2]	6.886	3.518	9.489	2.723	9.178	8.980	5.024	13.254	7.303	5.704
Q [ml/s]	236.197	156.490	164.116	318.996	97.340	107.295	146.692	159.653	197.691	97.542
AT [ms]	99.660	103.530	53.730	83.980	83.560	65.080	82.390	72.910	60.000	38.230
AT/ET	0.245	0.349	0.205	0.310	0.215	0.181	0.262	0.230	0.197	0.170
GLS [%]	x	x	x	x	-16.000	-13.100	-11.900	-16.100	x	x

Figure A.3: Clinical predictors measured in the AS60 data set

Patient 11	Patient 12	Patient 13	Patient 14	Patient 15	Patient 16	Patient 17	Patient 18	Patient 19	Patient 20
Mild	Mild	Mild	Mild	Mild	Mild	Mild	Mild	Mild	Mild
2.270	2.170	2.230	2.610	2.180	2.170	2.870	2.430	2.800	2.420
1.620	1.480	1.490	1.860	1.520	1.340	1.970	1.670	1.910	1.820
20.620	19.000	20.000	27.260	19.000	19.000	33.000	24.000	31.000	24.000
11.000	10.000	10.000	16.000	11.000	8.000	18.000	13.000	17.000	14.000
319.440	349.340	338.860	325.580	343.230	357.750	279.480	243.670	306.550	319.650
50.550	51.654	50.548	53.490	52.236	48.044	54.922	40.665	58.443	58.087
72.000	49.500	56.160	44.000	54.320	61.580	84.190	102.690	76.590	62.730
2.090	3.490	1.886	1.864	2.100	2.220	2.774	1.745	1.885	2.150
0.970	0.620	1.160	1.250	0.960	0.970	0.980	1.510	1.530	1.190
0.700	0.430	0.820	0.920	0.660	0.660	0.740	0.940	1.210	0.800
4.000	2.000	5.000	6.000	4.000	4.000	4.000	9.000	9.000	6.000
2.000	1.000	3.000	4.000	2.000	2.000	2.000	4.000	6.000	3.000
338.740	345.850	335.690	347.220	379.910	327.310	267.250	247.380	256.770	358.880
22.150	14.800	26.800	28.690	24.900	21.700	19.800	23.300	31.000	27.800
72.000	53.000	72.000	49.000	53.000	57.000	84.000	98.000	81.000	61.000
2.110	1.190	1.871	1.520	1.650	1.880	1.630	1.610	1.836	2.280
1.198	0.646	0.773	0.796	0.769	0.985	0.689	0.941	0.951	1.078
1.761	1.843	2.420	1.910	2.146	1.909	2.367	1.711	1.931	2.143
2.370	2.300	2.120	1.900	2.100	2.303	2.400	1.892	2.100	2.463
4.600	4.700	5.300	4.300	4.800	4.000	4.000	4.400	5.000	4.500
1.000	1.200	1.300	1.300	1.500	1.100	1.600	0.800	1.000	1.300
1.100	1.300	1.600	1.300	1.400	1.000	1.400	0.900	1.000	0.900
140.000	103.200	150.000	83.067	114.000	106.000	95.000	78.140	185.000	92.446
88.000	42.980	23.000	54.432	41.000	37.000	49.000	21.600	84.000	39.752
52.000	60.220	127.000	28.635	73.000	69.000	46.000	56.540	101.000	52.694
0.371	0.584	0.847	0.345	0.640	0.651	0.484	0.724	0.546	0.570
3.744	3.086	8.138	1.332	3.917	4.091	3.868	5.674	7.958	3.260
29.524	32.673	52.477	14.989	34.022	36.152	19.435	33.046	52.313	24.591
5.213	8.405	5.346	10.527	6.667	4.255	11.042	8.075	9.258	6.140
162.785	172.382	374.786	87.951	212.685	192.872	164.591	232.035	329.473	164.849
68.960	50.420	77.550	94.730	88.140	77.410	69.380	53.570	80.110	70.230
0.216	0.144	0.229	0.291	0.257	0.216	0.248	0.220	0.261	0.220
-10.900	-17.000	-18.500	x	x	-18.800	-11.900	-23.700	-18.700	-16.900

Figure A.4: Clinical predictors measured in the AS60 data set continued.

Patient 21	Patient 22	Patient 23	Patient 24	Patient 25	Patient 26	Patient 27	Patient 28	Patient 29	Patient 30
Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
3.670	3.070	2.760	3.350	2.150	3.140	2.440	2.040	1.950	2.720
2.830	1.890	2.120	2.290	1.540	2.240	1.670	1.490	1.390	1.900
54.000	38.000	30.000	45.000	19.000	39.460	24.000	17.000	15.000	29.570
35.000	18.000	20.000	24.000	11.000	20.090	13.000	10.000	9.000	19.000
311.790	301.310	314.410	303.930	332.750	332.200	338.860	322.270	180.790	314.810
88.290	57.032	66.705	69.706	51.273	63.800	56.607	47.925	25.100	50.900
112.810	55.850	62.740	61.890	65.620	68.000	61.200	62.230	120.530	60.000
4.190	1.860	2.650	3.900	2.430	2.358	2.388	2.510	1.987	2.658
0.800	1.430	1.100	0.770	0.900	1.340	1.010	0.880	0.980	0.920
0.610	0.930	0.750	0.570	0.650	0.950	0.710	0.600	0.680	0.650
3.000	8.000	5.000	2.000	3.000	7.000	4.000	3.000	4.000	3.000
2.000	4.000	3.000	1.000	2.000	4.000	2.000	2.000	2.000	2.000
343.230	330.130	337.990	314.410	322.270	322.910	335.370	317.030	212.540	325.580
21.100	30.700	25.200	17.900	21.100	27.060	23.700	19.100	12.630	19.150
67.000	70.000	65.000	65.000	63.000	71.000	59.000	62.000	118.000	57.000
0.840	2.320	1.390	1.040	1.350	1.202	1.237	1.290	1.426	1.181
0.391	1.215	0.757	0.501	0.721	0.852	0.609	0.689	0.869	0.618
2.148	1.909	1.823	2.076	1.885	1.411	2.031	1.857	1.642	1.910
2.380	2.340	2.170	2.380	2.044	1.900	1.940	2.030	1.900	2.000
5.200	5.800	4.300	4.000	4.900	4.600	4.700	5.200	3.300	4.300
1.600	1.200	1.300	1.500	1.100	0.900	1.400	1.300	1.700	1.000
0.900	1.200	1.500	1.500	1.000	0.800	1.400	0.800	1.800	1.800
134.780	128.000	76.000	66.000	111.000	99.000	100.000	103.000	44.133	83.067
59.150	53.000	33.000	29.000	54.000	39.000	41.000	54.000	11.166	54.432
75.630	75.000	43.000	37.000	57.000	60.000	59.000	49.000	32.967	28.635
0.561	0.586	0.566	0.561	0.514	0.610	0.590	0.476	0.747	0.345
6.800	4.719	2.746	2.347	3.666	4.170	3.546	3.044	3.932	1.675
35.204	39.278	23.592	17.824	30.245	42.522	29.046	26.385	20.080	14.989
41.667	7.759	14.388	23.077	8.148	16.715	10.510	7.752	6.312	16.083
242.567	248.913	136.764	121.739	171.300	180.614	174.113	152.046	182.352	90.960
98.420	73.940	81.780	96.230	82.670	115.250	125.740	93.280	59.700	104.260
0.316	0.245	0.260	0.317	0.248	0.347	0.371	0.289	0.330	0.331
-19.800	-21.700	-13.300	-14.800	-18.400	-20.000	-18.200	-15.800	x	x

Figure A.5: Clinical predictors measured in the AS60 data set continued.

Patient 31	Patient 32	Patient 33	Patient 34	Patient 35	Patient 36	Patient 37	Patient 38	Patient 39	Patient 40
Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
3.010	2.220	2.100	3.620	3.300	2.980	2.640	2.810	4.030	3.470
2.130	1.540	1.480	2.060	2.420	2.030	1.830	1.660	2.840	2.300
36.250	20.000	18.000	52.000	44.000	36.000	28.000	31.640	65.000	48.000
25.700	11.000	10.000	21.000	26.000	19.000	15.000	13.000	36.000	26.000
347.220	309.170	321.400	303.930	282.970	248.910	290.830	380.000	345.850	288.210
69.660	47.646	47.521	62.570	68.351	50.544	53.235	68.700	98.289	66.347
49.000	65.800	68.840	78.970	91.840	75.280	56.820	52.000	53.840	99.570
2.752	2.100	2.915	2.410	2.377	3.510	2.260	2.094	3.795	3.190
1.120	0.940	0.740	1.330	1.340	0.780	1.010	1.180	0.950	1.140
0.790	0.690	0.540	0.980	0.990	0.550	0.740	0.620	0.670	0.840
5.000	4.000	2.000	7.000	7.170	2.000	4.000	6.000	4.000	5.000
3.000	2.000	1.000	4.000	4.430	1.000	2.000	3.000	2.000	3.000
370.370	330.130	303.930	263.760	291.700	264.630	317.030	532.000	387.770	248.330
25.310	22.700	16.300	26.000	28.750	14.400	23.600	32.810	25.900	20.800
50.000	64.000	70.000	75.000	92.840	75.000	58.000	53.000	53.000	103.000
0.924	1.660	1.187	1.080	1.256	1.500	1.510	1.500	0.850	0.990
0.614	0.812	0.591	0.626	0.777	1.029	0.758	0.794	0.413	0.580
1.504	2.044	2.009	1.707	1.615	1.458	1.984	1.888	2.075	1.711
1.800	2.100	2.100	1.820	1.950	2.590	2.083	2.000	2.027	2.000
3.700	4.100	4.500	4.200	4.700	2.800	5.100	4.400	6.100	3.900
1.300	1.300	1.100	0.800	1.000	1.100	1.000	1.300	1.300	1.200
0.900	1.200	1.000	0.800	1.100	1.200	1.100	1.300	1.500	0.800
58.126	76.000	92.000	60.000	104.000	55.000	75.000	87.689	166.000	59.000
23.832	19.000	23.000	19.000	20.000	17.000	26.000	11.166	71.000	19.000
34.295	57.000	69.000	41.000	84.000	38.000	49.000	76.524	95.000	40.000
0.590	0.750	0.750	0.683	0.808	0.691	0.653	0.873	0.572	0.678
1.698	3.699	4.790	3.156	7.757	2.855	2.813	4.017	5.075	4.051
22.801	27.882	34.347	24.014	52.009	26.057	24.692	40.523	45.787	23.385
27.811	6.627	8.422	19.444	20.708	12.665	9.934	8.669	42.353	26.263
98.769	184.365	214.686	134.899	296.851	152.666	168.483	201.378	274.686	138.788
117.210	71.420	119.700	83.340	77.610	104.080	79.490	95.000	108.230	91.220
0.338	0.231	0.372	0.274	0.274	0.418	0.273	0.250	0.313	0.317
-19.700×		-17.700×		-17.100×		-19.500×		-15.900	-14.800

Figure A.6: Clinical predictors measured in the AS60 data set continued.

Patient 41	Patient 42	Patient 43	Patient 44	Patient 45	Patient 46	Patient 47	Patient 48	Patient 49	Patient 50
Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe
3.910	3.650	5.680	1.860	5.180	4.570	4.380	3.220	2.990	5.010
2.890	2.820	4.370	1.170	3.670	3.300	3.070	2.460	1.890	3.950
61.000	53.000	129.000	13.840	107.000	83.000	77.000	42.000	36.000	100.000
43.000	35.000	84.000	8.000	61.000	51.000	44.000	26.000	18.000	69.000
332.560	314.410	327.510	295.130	331.880	345.850	343.230	330.130	338.860	301.310
96.020	88.692	143.020	34.700	121.942	114.028	105.322	81.099	64.112	119.103
55.000	62.730	70.460	52.000	57.440	71.560	60.580	53.260	74.030	67.350
4.830	6.010	5.660	2.421	5.490	4.800	5.330	3.540	5.390	6.540
1.130	0.800	1.040	0.870	1.080	0.940	0.890	0.730	0.540	0.770
0.820	0.580	0.730	0.490	0.643	0.680	0.580	0.540	0.390	0.590
5.000	3.000	4.000	3.000	5.000	4.000	3.000	2.000	1.000	2.000
3.000	1.000	2.000	2.000	2.000	2.000	2.000	1.000	1.000	2.000
386.040	335.370	345.850	292.680	345.620	348.470	338.860	424.450	301.310	309.170
22.300	19.400	25.300	14.330	22.200	23.800	19.800	22.900	11.900	18.200
53.000	64.000	75.000	55.000	63.000	70.000	62.000	55.000	68.000	70.000
0.960	0.490	0.530	1.569	0.660	0.730	0.840	0.970	0.830	0.440
0.503	0.247	0.260	0.809	0.364	0.326	0.391	0.535	0.422	0.214
1.909	1.971	2.048	1.939	1.813	2.239	2.148	1.813	1.966	2.056
2.430	1.670	1.950	2.200	2.150	2.110	2.386	2.090	2.390	1.930
3.900	5.400	5.400	5.100	4.900	4.500	4.700	4.200	5.000	4.800
1.400	1.300	1.400	1.200	1.400	1.600	1.700	1.600	1.200	1.600
1.100	1.000	1.300	1.200	1.100	1.300	2.000	1.300	1.100	1.600
65.910	158.000	119.000	150.000	112.814	133.320	82.000	73.000	100.000	147.800
34.932	101.000	58.000	96.000	60.920	63.200	31.000	13.000	59.000	76.360
30.978	57.000	61.000	54.000	51.894	70.120	51.000	60.000	41.000	71.440
0.470	0.361	0.513	0.360	0.460	0.526	0.622	0.822	0.410	0.483
1.673	3.612	4.437	2.889	3.125	4.963	3.126	3.248	2.912	4.906
16.231	28.922	29.791	27.847	28.621	31.314	23.739	33.093	20.853	34.746
44.792	71.429	158.491	5.099	92.424	69.863	52.381	26.804	21.687	156.818
93.149	181.292	186.254	182.970	156.365	202.747	148.588	181.747	120.994	237.098
153.060	121.320	129.700	82.080	103.770	119.400	129.320	113.070	92.960	130.740
0.460	0.386	0.396	0.278	0.313	0.345	0.377	0.343	0.274	0.434
-13.000	-9.200	-15.100×		-16.000×		-12.000	-16.800×		-11.700

Figure A.7: Clinical predictors measured in the AS60 data set continued.

Patient 51	Patient 52	Patient 53	Patient 54	Patient 55	Patient 56	Patient 57	Patient 58	Patient 59	Patient 60
Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe	Severe
3.170	4.870	4.260	3.990	4.550	4.120	4.910	3.790	4.940	4.140
2.410	3.730	3.060	3.040	3.490	3.130	3.600	2.750	3.710	3.050
40.000	95.000	72.000	64.000	83.000	68.000	96.000	57.460	98.000	68.000
25.000	61.000	43.000	40.000	53.000	43.000	57.000	40.750	63.000	42.000
293.450	303.930	317.030	401.750	317.030	286.460	303.930	322.000	370.310	337.990
70.859	113.318	96.871	122.187	110.575	89.794	109.295	88.650	137.308	103.201
62.210	57.890	67.680	51.420	67.750	88.530	73.630	65.000	59.840	61.890
3.900	6.690	3.350	4.960	4.180	5.740	6.530	3.838	4.650	4.756
0.740	0.710	1.250	0.810	1.140	0.690	0.910	1.000	1.240	0.840
0.510	0.520	0.880	0.600	0.740	0.560	0.640	0.747	0.730	0.580
2.000	2.000	6.000	3.000	5.000	2.000	3.000	4.000	6.000	3.000
1.000	1.000	3.000	2.000	2.000	1.000	2.000	2.000	2.000	2.000
342.360	327.510	288.210	412.230	358.950	279.480	267.830	309.020	402.000	377.290
17.500	16.900	25.300	24.600	26.400	15.700	17.000	23.100	29.529	21.700
61.000	67.000	68.000	52.000	65.000	85.000	85.000	63.000	61.000	60.000
0.870	0.690	0.810	0.620	0.630	0.570	0.490	1.054	0.620	0.728
0.465	0.365	0.400	0.360	0.385	0.398	0.267	0.491	0.385	0.359
1.871	1.890	2.025	1.714	1.636	1.432	1.835	2.148	1.610	2.027
2.040	2.428	1.860	1.980	1.830	2.050	2.010	2.270	1.960	2.100
5.100	5.600	3.600	4.200	3.900	4.000	5.300	5.800	4.900	4.800
1.200	1.300	1.400	1.300	1.400	1.000	1.400	1.500	1.100	1.400
1.200	1.200	1.300	1.300	1.400	1.000	1.500	1.300	1.000	1.100
99.000	231.000	47.000	75.000	61.000	58.000	63.000	138.000	112.814	107.520
32.000	160.000	17.000	18.000	24.000	25.000	23.000	92.000	45.126	47.309
67.000	71.000	30.000	57.000	37.000	33.000	40.000	46.000	67.688	60.211
0.677	0.307	0.638	0.760	0.607	0.569	0.635	0.333	0.600	0.560
4.128	4.434	2.035	2.947	2.456	2.863	3.173	2.944	4.090	3.670
35.810	37.558	14.815	33.250	22.611	23.042	21.796	21.414	42.032	29.709
28.736	88.406	53.086	64.516	84.127	75.439	116.327	38.661	101.613	57.699
228.318	233.606	94.628	141.879	116.708	115.199	131.609	142.857	182.789	178.145
113.300	137.540	111.520	146.060	84.030	103.640	88.020	137.440	128.490	95.070
0.386	0.453	0.352	0.364	0.265	0.362	0.290	0.427	0.347	0.281
x	-9.700	x	-18.100	-17.500	-16.600	-11.200	-9.600	-19.000	-17.800

Figure A.8: Clinical predictors measured in the AS60 data set continued.

B. Appendix B

Algorithm 2 Generate Triangular Mesh on Curved Surface of a Cylinder

Require: Number of circles N_z, points per circle N_θ, height H, radius R
Ensure: Point list $\mathcal{P}$ and triangle list $\mathcal{T}$

Initialize empty point list $\mathcal{P} \leftarrow [\,]$
Initialize empty triangle list $\mathcal{T} \leftarrow [\,]$
$\Delta z \leftarrow H/(N_z - 1)$
▷ Generate points: clockwise first, then upward
for $j = 0$ to $N_\theta - 1$ **do** ▷ Loop over angle
    $\theta \leftarrow 2\pi \cdot j/N_\theta$
    for $i = 0$ to $N_z - 1$ **do** ▷ Loop over height
        $z \leftarrow i \cdot \Delta z$
        $x \leftarrow R \cdot \cos(\theta)$
        $y \leftarrow R \cdot \sin(\theta)$
        Append (x, y, z) to $\mathcal{P}$
    end for
end for
▷ Generate triangles
for $j = 0$ to $N_\theta - 1$ **do**
    $j_{\text{next}} \leftarrow (j + 1) \mod N_\theta$
    for $i = 0$ to $N_z - 2$ **do**
        $p_0 \leftarrow j \cdot N_z + i$
        $p_1 \leftarrow j_{\text{next}} \cdot N_z + i$
        $q_0 \leftarrow j \cdot N_z + (i + 1)$
        $q_1 \leftarrow j_{\text{next}} \cdot N_z + (i + 1)$
        Append triangle (p_0, p_1, q_0) to $\mathcal{T}$ ▷ Upward triangle
        Append triangle (p_1, q_1, q_0) to $\mathcal{T}$ ▷ Downward triangle
    end for
end for

C. Appendix C

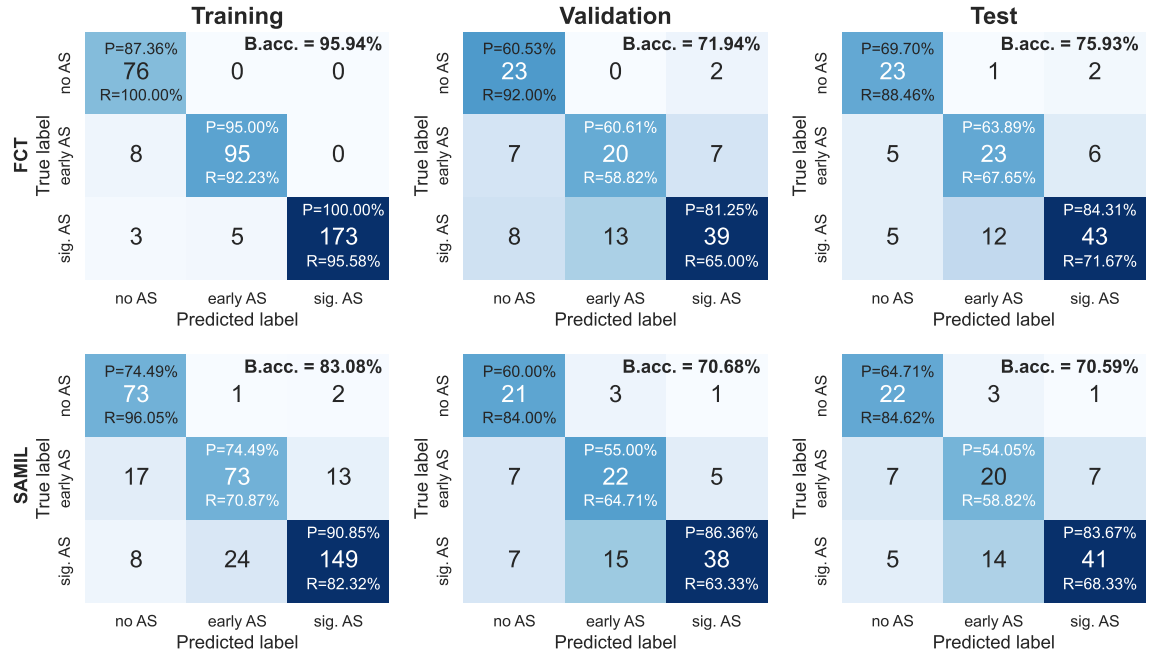


Figure C.1: Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 1 of the Tufts Medical Echocardiogram Dataset. The columns contain the confusion matrices for the training validation and test sets for models selected at the best validation accuracy.

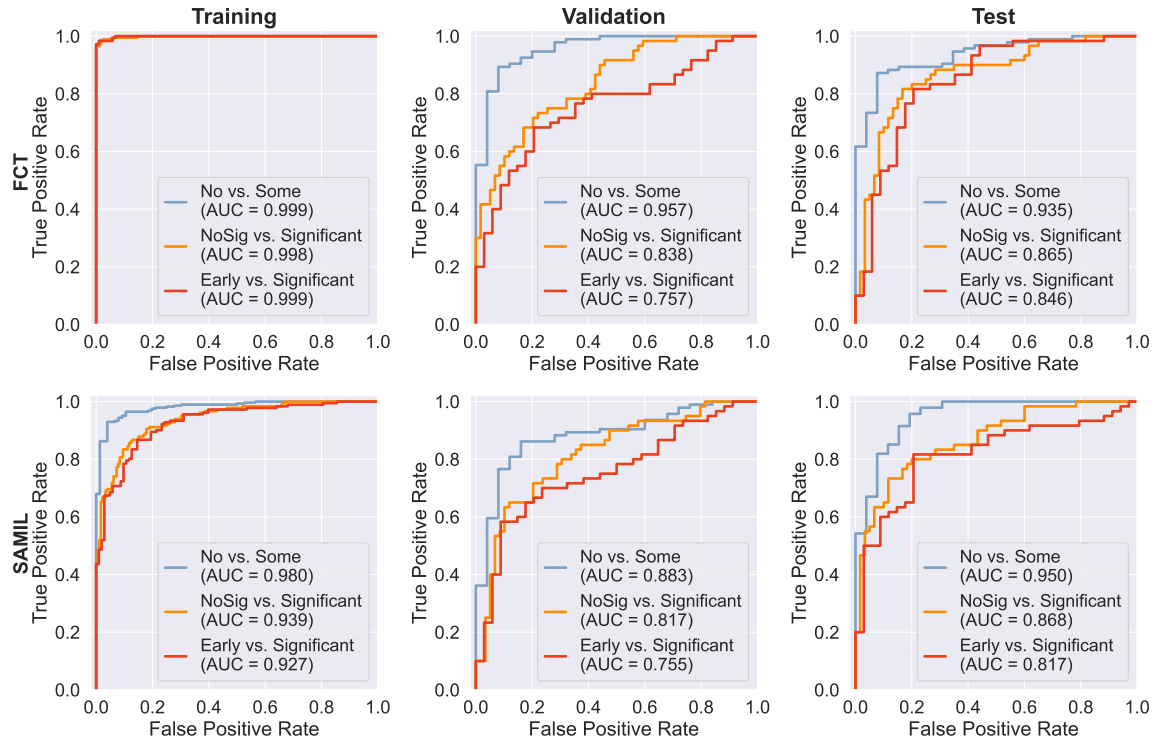


Figure C.2: Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 1 of the Tufts Medical Echocardiogram Dataset. The columns contain the ROC curves for multiple binary classification tasks for the training validation and test sets for models selected at the best validation accuracy.

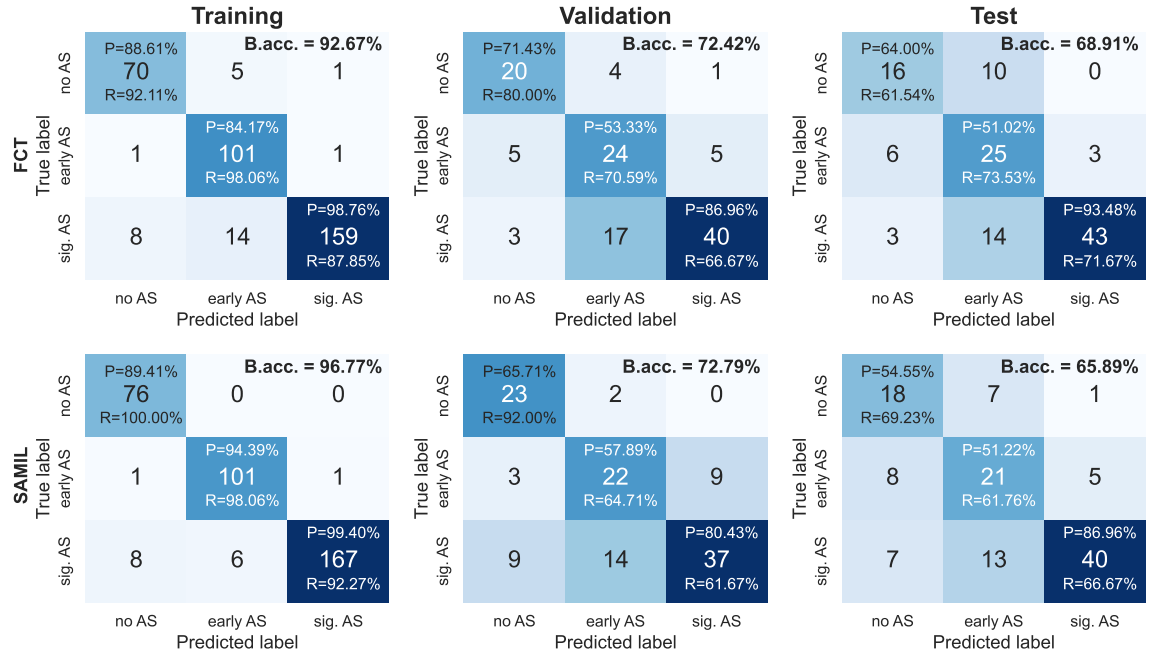


Figure C.3: Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 2 of the Tufts Medical Echocardiogram Dataset. The columns contain the confusion matrices for the training validation and test sets for models selected at the best validation accuracy.

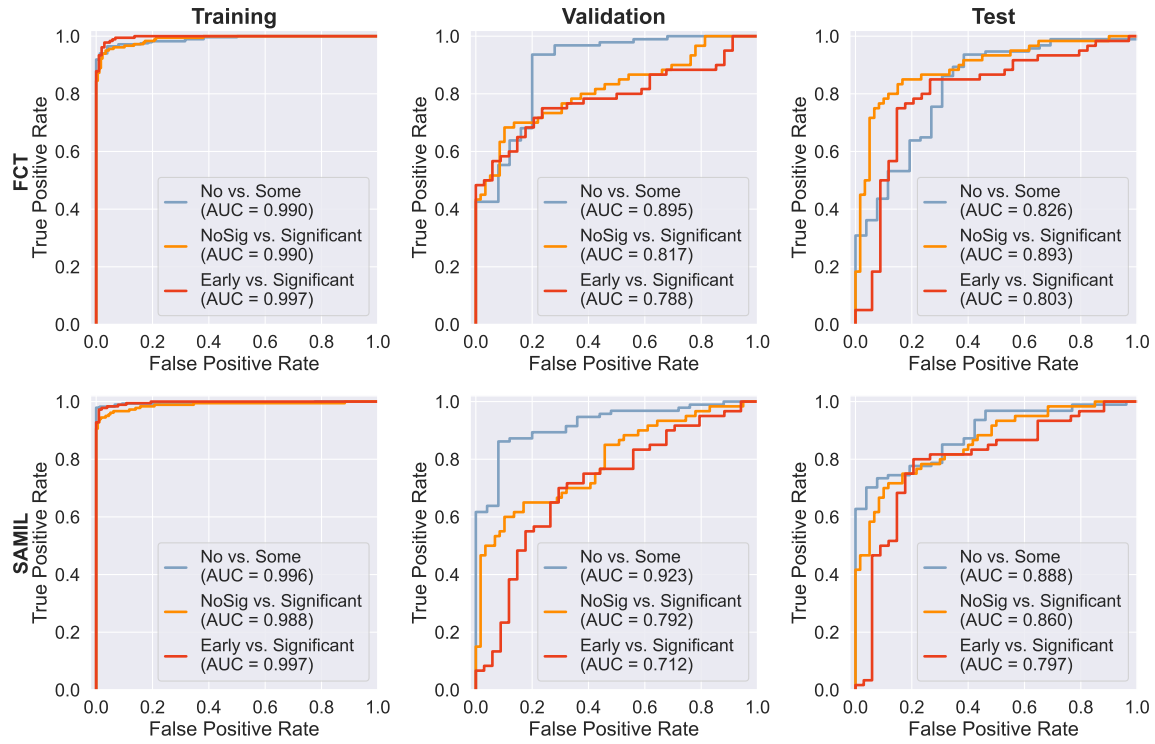


Figure C.4: Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 2 of the Tufts Medical Echocardiogram Dataset. The columns contain the ROC curves for multiple binary classification tasks for the training validation and test sets for models selected at the best validation accuracy.

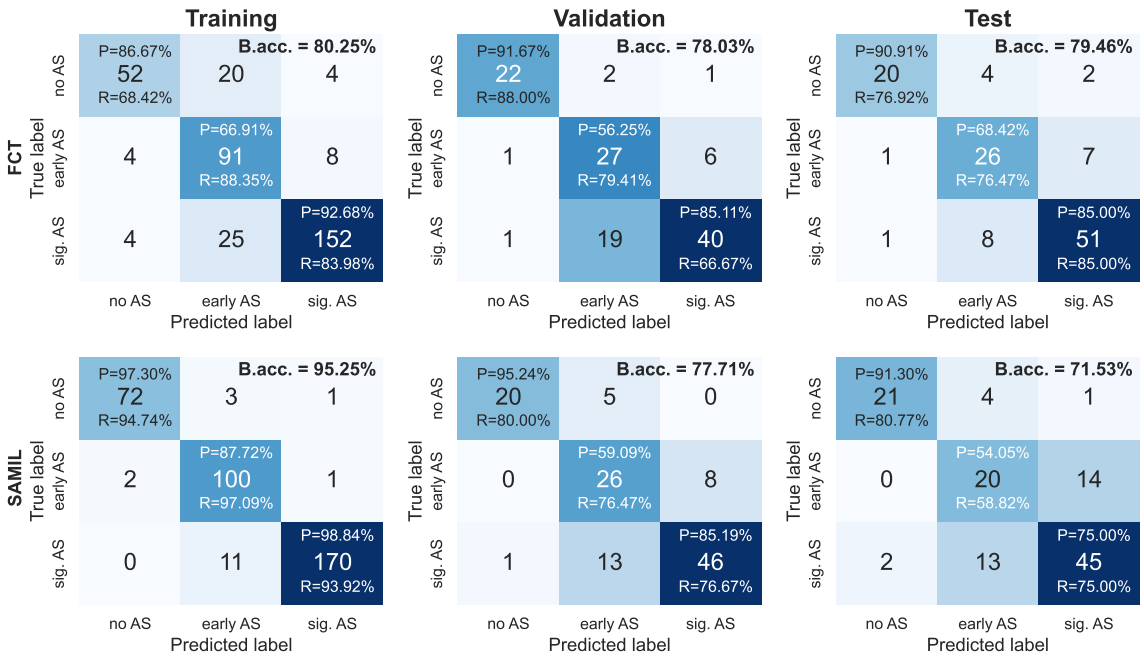


Figure C.5: Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 3 of the Tufts Medical Echocardiogram Dataset. The columns contain the confusion matrices for the training validation and test sets for models selected at the best validation accuracy.

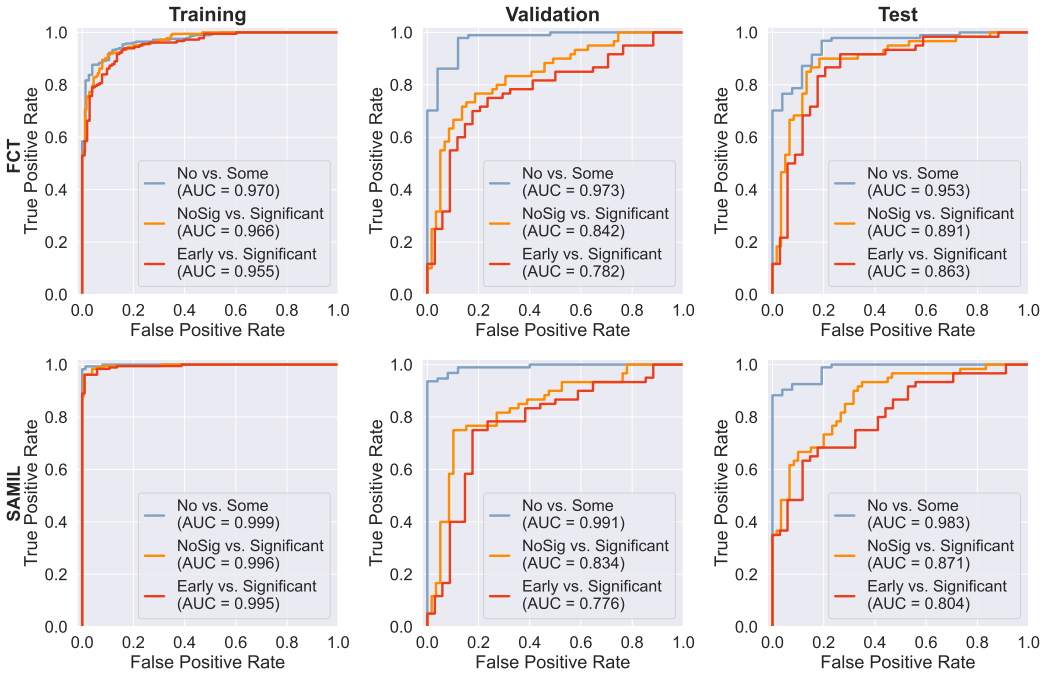


Figure C.6: Inference results for the FCT feature extractor [top] and SAMIL [bottom] model architectures on split 3 of the Tufts Medical Echocardiogram Dataset. The columns contain the ROC curves for multiple binary classification tasks for the training validation and test sets for models selected at the best validation accuracy.

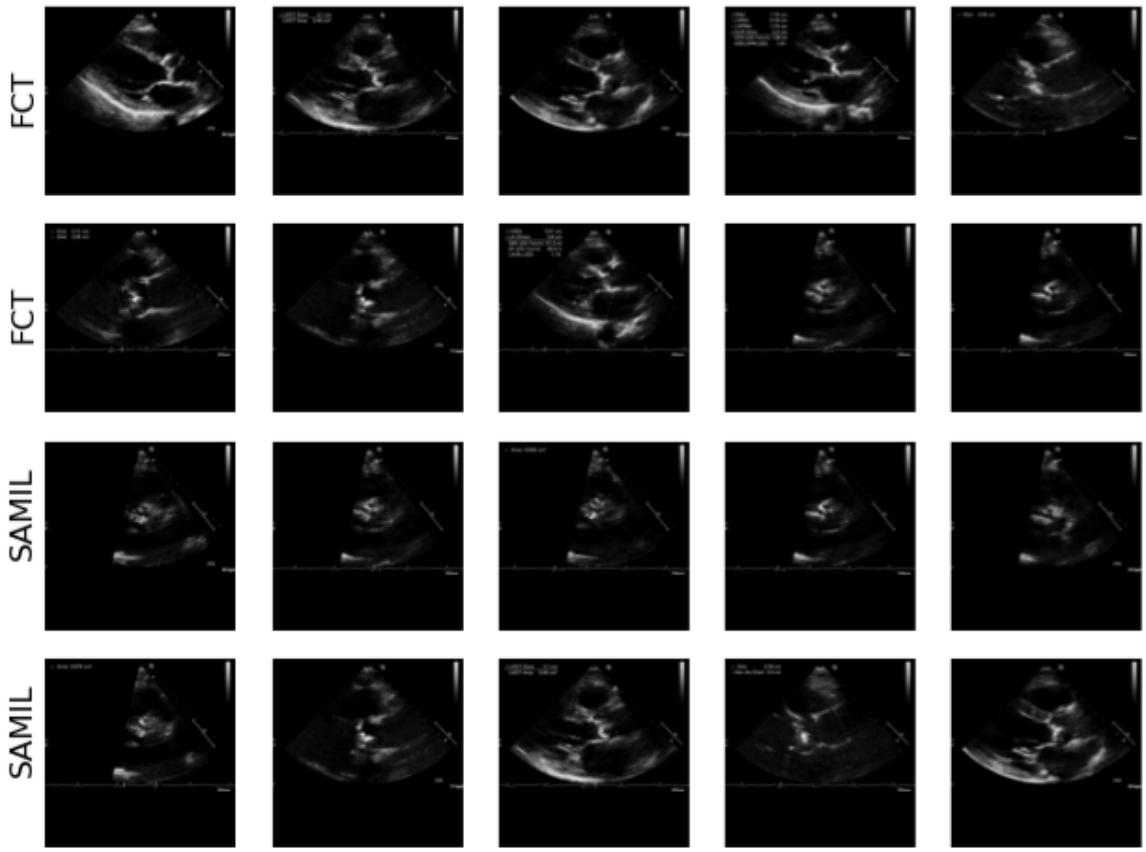


Figure C.7: Top 10 ranked images by attention weights for the FCT and SAMIL architectures in the test set of split 3 in the Tufts Medical Echocardiogram Dataset. Both models succeed in attending to relevant views (PLAX,PSAX) as shown by average view relevance scores close to 1.

Bibliography

Alec Vahanian, Friedhelm Beyersdorf, Fabien Praz, Milan Milojevic, Stephan Baldus, Johann Bauersachs, Davide Capodanno, Lenard Conradi, Michele De Bonis, Ruggero De Paulis, Victoria Delgado, Nick Freemantle, Martine Gilard, Kristina H Haugaa, Anders Jeppsson, Peter Jüni, Luc Pierard, Bernard D Prendergast, J Rafael Sádaba, Christophe Tribouilloy, Wojtek Wojakowski, ESC/EACTS Scientific Document Group, Franz-Josef Neumann, Patrick Myers, Magdy Abdelhamid, Stephan Achenbach, Riccardo Asteggiano, Fabio Barili, Michael A Borger, Thierry Carrel, Jean-Philippe Collet, Dan Foldager, Gilbert Habib, Christian Hassager, Alar Irs, Bernard Iung, Marjan Jahangiri, Hugo A Katus, Konstantinos C Koskinas, Steffen Massberg, Christian E Mueller, Jens Cosedis Nielsen, Philippe Pibarot, Amina Rakisheva, Marco Roffi, Andrea Rubboli, Evgeny Shlyakhto, Matthias Siepe, Marta Sitges, Lars Sondergaard, Miguel Sousa-Uva, Guiseppe Tarantini, Jose Luis Zamorano, Fabien Praz, Milan Milojevic, Stephan Baldus, Johann Bauersachs, Davide Capodanno, Lenard Conradi, Michele De Bonis, Ruggero De Paulis, Victoria Delgado, Nick Freemantle, Martine Gilard, Kristina H Haugaa, Anders Jeppsson, Peter Jüni, Luc Pierard, Bernard D Prendergast, J Rafael Sádaba, Christophe Tribouilloy, Wojtek Wojakowski, ESC National Cardiac Societies, Yasmina Benchabi, Aram Chilingaryan, Bernhard Metzler, Yasmin Rustamova, Vadim Shumavets, Patrizio Lancellotti, Elnur Smajic, Diana Trendafilova-Lazarova, Jure Samardzic, Maria Karakyriou, Tomas Palecek, Jordi Sanchez Dahl, Marwa Sayed Meshaal, Kairit Palm, Marko Virtanen, Claire Bouleti, Zviad Bakhutashvili, Stephan Achenbach, Maria Boutsikou, Attila Béla Kertész, Ragnar Danielsen, Yan Topilsky, Paolo Golino, Rustem Tuleutayev, Shpend Elezi, Alina Kerimkulov, Ainars Rudzitis, Sigita Glaveckaite, Rouguiatou Sow, Daniela Cassar Demarco, Nebojsa Bulatovic, Aicha Aouad, Renée Van Den Brink, Emilija Antova, Jan Otto Beitnes, Andrzej Ochala, Regina Ribeiras, Dragos Vinereanu, Olga Irtyuga, Branislava Ivanovic, Iveta Simkova, Ariana González Gómez, Giovanna Sarno, Giovanni B Pedrazzini, Walid Bsata, Lilia Zakhama, Levent Korkmaz, Sergii Cherniuk, Mohammed Y Khanji, and Islamjan Sharipov. 2021 esc/eacts guidelines for the management of valvular heart disease. *European Heart Journal*, 43(7):561–632, February 2022. ISSN 0195-668X, 1522-9645. doi: 10.1093/eurheartj/ehab395. URL https://academic.oup.com/eurheartj/article/43/7/561/6358470.

Nalini M. Rajamannan, Frank J. Evans, Elena Aikawa, K. Jane Grande-Allen, Linda L. Demer, Donald D. Heistad, Craig A. Simmons, Kristyn S. Masters, Patrick Mathieu, Kevin D. O'Brien, Frederick J. Schoen, Dwight A. Towler, Ajit P. Yoganathan, and Catherine M. Otto. Calcific aortic valve disease: Not simply a degenerative process:

A review and agenda for research from the national heart and lung and blood institute aortic stenosis working group executive summary: Calcific aortic valve disease – 2011 update. *Circulation*, 124(16):1783–1791, October 2011. ISSN 0009-7322, 1524-4539. doi: 10.1161/CIRCULATIONAHA.110.006767. URL https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.110.006767.

Brian R. Lindman, Marie-Annick Clavel, Patrick Mathieu, Bernard Iung, Patrizio Lancellotti, Catherine M. Otto, and Philippe Pibarot. Calcific aortic stenosis. *Nature Reviews Disease Primers*, 2(1):16006, March 2016. ISSN 2056-676X. doi: 10.1038/nrdp.2016.6. URL https://www.nature.com/articles/nrdp20166.

Helmut Baumgartner, Judy Hung, Javier Bermejo, John B. Chambers, Thor Edvardsen, Steven Goldstein, Patrizio Lancellotti, Melissa LeFevre, Fletcher Miller, Jr., and Catherine M. Otto. Recommendations on the Echocardiographic Assessment of Aortic Valve Stenosis: A Focused Update from the European Association of Cardiovascular Imaging and the American Society of Echocardiography. *J. Am. Soc. Echocardiogr.*, 30(4):372–392, April 2017. ISSN 1097-6795. doi: 10.1016/j.echo.2017.02.009.

Marie-Annick Clavel, Julien Magne, and Philippe Pibarot. Low-gradient aortic stenosis. *European Heart Journal*, 37(34):2645–2657, September 2016. ISSN 0195-668X, 1522-9645. doi: 10.1093/eurheartj/ehw096. URL https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehw096.

Wikipedia. Diagram of the human heart - wikipedia, 8 2024. URL https://en.wikipedia.org/wiki/File:Diagram_of_the_human_heart_(cropped).svg. https://en.wikipedia.org/wiki/File:Diagram_of_the_human_heart_(cropped).svg, Wikimedia Commons. Public domain. Accessed 2025-08-03.

Wikipedia. Wiggers Diagram - Wikipedia, 7 2016. URL https://en.wikipedia.org/wiki/File:Wiggers_Diagram_2.svg. https://en.wikipedia.org/wiki/File:Wiggers_Diagram_2.svg, Wikimedia Commons. Public domain. Accessed 2025-08-03.

Carol Mitchell, Peter S. Rahko, Lori A. Blauwet, Barry Canaday, Joshua A. Finstuen, Michael C. Foster, Kenneth Horton, Kofo O. Ogunyankin, Richard A. Palma, and Eric J. Velazquez. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: Recommendations from the american society of echocardiography. *Journal of the American Society of Echocardiography*, 32(1):1–64, January 2019. ISSN 08947317. doi: 10.1016/j.echo.2018.06.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731718303183.

Manuel D. Cerqueira, Neil J. Weissman, Vasken Dilsizian, Alice K. Jacobs, Sanjiv Kaul, Warren K. Laskey, Dudley J. Pennell, John A. Rumberger, Thomas Ryan, and Mario S. Verani. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. *Circulation*, 105(4):539–542, 2002. doi: 10.1161/hc0402.102975. URL https://www.ahajournals.org/doi/abs/10.1161/hc0402.102975.

Harvey Feigenbaum. Role of m-mode technique in today's echocardiography. *Journal of the American Society of Echocardiography*, 23(3):240–257, 2010. ISSN 0894-7317. doi: https://doi.org/10.1016/j.echo.2010.01.015. URL https://www.sciencedirect.com/science/article/pii/S0894731710000581.

Roberto M. Lang, Luigi P. Badano, Victor Mor-Avi, Jonathan Afilalo, Anderson Armstrong, Laura Ernande, Frank A. Flachskampf, Elyse Foster, Steven A. Goldstein, Tatiana Kuznetsova, Patrizio Lancellotti, Denisa Muraru, Michael H. Picard, Ernst R. Rietzschel, Lawrence Rudski, Kirk T. Spencer, Wendy Tsang, and Jens-Uwe Voigt. Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the american society of echocardiography and the european association of cardiovascular imaging. *Journal of the American Society of Echocardiography*, 28 (1):1–39.e14, January 2015. ISSN 08947317. doi: 10.1016/j.echo.2014.10.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731714007457.

Frederick W. Kremkau. Doppler principles. *Seminars in Roentgenology*, 27(1):6–16, 1992. ISSN 0037-198X. doi: https://doi.org/10.1016/0037-198X(92)90042-Z. URL https://www.sciencedirect.com/science/article/pii/0037198X9290042Z. Noninvasive Vascular Imaging.

Nandan S. Anavekar and Jae K. Oh. Doppler echocardiography: A contemporary review. *Journal of Cardiology*, 54(3):347–358, December 2009. ISSN 09145087. doi: 10.1016/j.jjcc.2009.10.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0914508709002731.

Blase A. Carabello. Introduction to aortic stenosis. *Circulation Research*, 113(2):179–185, July 2013. ISSN 0009-7330, 1524-4571. doi: 10.1161/CIRCRESAHA.113.300156. URL https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.113.300156.

Catherine M. Otto, Rick A. Nishimura, Robert O. Bonow, Blase A. Carabello, John P. Erwin, Federico Gentile, Hani Jneid, Eric V. Krieger, Michael Mack, Christopher McLeod, Patrick T. O'Gara, Vera H. Rigolin, Thoralf M. Sundt, Annemarie Thompson, and Christopher Toly. 2020 acc/aha guideline for the management of patients with valvular heart disease: A report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. *Circulation*, 143(5), February 2021. ISSN 0009-7322, 1524-4539. doi: 10.1161/CIR.0000000000000923. URL https://www.ahajournals.org/doi/10.1161/CIR.0000000000000923.

Lisa Q. Rong, Irbaz Hameed, Antonino Di Franco, Mohamed M. Rahouma, Faiza M. Khan, Michelle Demetres, Jonathan W. Weinsaft, Richard B. Devereux, and Mario Gaudino. A pairwise meta-analytic comparison of aortic valve area determined by planimetric versus hemodynamic methods in aortic stenosis. *International Journal of Cardiology*, 322: 77–85, January 2021. ISSN 0167-5273. doi: 10.1016/j.ijcard.2020.09.003. URL https://www.sciencedirect.com/science/article/pii/S0167527320337372.

Helmut Baumgartner, Judy Hung, Javier Bermejo, John B. Chambers, Arturo Evangelista, Brian P. Griffin, Bernard Iung, Catherine M. Otto, Patricia A. Pellikka, and Miguel Quiñones. Echocardiographic assessment of valve stenosis: Eae/ase recommendations for clinical practice. *Journal of the American Society of Echocardiography*, 22(1):1–23, January 2009. ISSN 08947317. doi: 10.1016/j.echo.2008.11.029. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731708007761.

Iria Silva, Erwan Salaun, Nancy Côté, and Philippe Pibarot. Confirmation of aortic stenosis severity in case of discordance between aortic valve area and gradient. *JACC: Case Reports*, 4(3):170–177, February 2022. ISSN 26660849. doi: 10.1016/j.jaccas.2021.11.009. URL https://linkinghub.elsevier.com/retrieve/pii/S2666084921010342.

C. Cueff, J.-M. Serfaty, C. Cimadevilla, J.-P. Laissy, D. Himbert, F. Tubach, X. Duval, B. Iung, M. Enriquez-Sarano, A. Vahanian, and D. Messika-Zeitoun. Measurement of aortic valve calcification using multislice computed tomography: correlation with haemodynamic severity of aortic stenosis and clinical implication for patients with low ejection fraction. *Heart*, 97(9):721–726, May 2011. ISSN 1355-6037. doi: 10.1136/hrt.2010.198853. URL https://heart.bmj.com/lookup/doi/10.1136/hrt.2010.198853.

Marie-Annick Clavel, David Messika-Zeitoun, Philippe Pibarot, Shivani R. Aggarwal, Joseph Malouf, Phillip A. Araoz, Hector I. Michelena, Caroline Cueff, Eric Larose, Romain Capoulade, Alec Vahanian, and Maurice Enriquez-Sarano. The complex nature of discordant severe calcified aortic valve disease grading. *Journal of the American College of Cardiology*, 62(24):2329–2338, December 2013. ISSN 07351097. doi: 10.1016/j.jacc.2013.08.1621. URL https://linkinghub.elsevier.com/retrieve/pii/S073510971305153X.

E. Fougeres, C. Tribouilloy, M. Monchi, H. Petit-Eisenmann, S. Baleynaud, A. Pasquet, C. Chauvel, D. Metz, C. Adams, D. Rusinaru, P. Gueret, and J.-L. Monin. Outcomes of pseudo-severe aortic stenosis under conservative treatment. *European Heart Journal*, 33 (19):2426–2433, October 2012. ISSN 0195-668X, 1522-9645. doi: 10.1093/eurheartj/ehs176. URL https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehs176.

Blase A. Carabello. Aortic stenosis. *New England Journal of Medicine*, 346(9):677–682, February 2002. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMcp010846. URL http://www.nejm.org/doi/abs/10.1056/NEJMcp010846.

Jean-Luc Monin, Jean-Paul Quéré, Mehran Monchi, Hélène Petit, Serge Baleynaud, Christophe Chauvel, Camélia Pop, Patrick Ohlmann, Claude Lelguen, Patrick Dehant, Christophe Tribouilloy, and Pascal Guéret. Low-gradient aortic stenosis: Operative risk stratification and predictors for long-term outcome: A multicenter study using dobutamine stress hemodynamics. *Circulation*, 108(3):319–324, July 2003. ISSN

0009-7322, 1524-4539. doi: 10.1161/01.CIR.0000079171.43055.46. URL https://www.ahajournals.org/doi/10.1161/01.CIR.0000079171.43055.46.

Daniel Manna, Maria Eliasson, Odd Bech-Hanssen, and Thomas Lindow. Reproducibility of echocardiographic measures of aortic stenosis severity and its impact on grading of severity. *Journal of the American Society of Echocardiography*, 37(3):370–372.e2, March 2024. ISSN 08947317. doi: 10.1016/j.echo.2023.11.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731723005886.

Liam Ring, Benoy N. Shah, Sanjeev Bhattacharyya, Allan Harkness, Mark Belham, David Oxborough, Keith Pearce, Bushra S. Rana, Daniel X. Augustine, Shaun Robinson, and Christophe Tribouilloy. Echocardiographic assessment of aortic stenosis: a practical guideline from the british society of echocardiography. *Echo Research and Practice*, 8 (1):G19–G59, March 2021. ISSN 2055-0464. doi: 10.1530/ERP-20-0035. URL https://echo.biomedcentral.com/articles/10.1530/ERP-20-0035.

Bart J. J. Velders, Rolf H. H. Groenwold, Nina Ajmone Marsan, Arie-Pieter Kappetein, Rob A. F. De Lind Van Wijngaarden, Jerry Braun, Robert J. M. Klautz, and Michiel D. Vriesendorp. Improving accuracy in diagnosing aortic stenosis severity: An in-depth analysis of echocardiographic measurement error through literature review and simulation study. *Echocardiography*, 40(9):892–902, September 2023. ISSN 0742-2822, 1540-8175. doi: 10.1111/echo.15664. URL https://onlinelibrary.wiley.com/doi/10.1111/echo.15664.

Woo-Jin Cho Kim, Arian Beqiri, Adam J. Lewandowski, Esther Puyol-Antón, Deborah C. Markham, Andrew P. King, Paul Leeson, and Pablo Lamata. Beyond simpson's rule: Accounting for orientation and ellipticity assumptions. *Ultrasound in Medicine and Biology*, 48(12):2476–2485, December 2022. ISSN 03015629. doi: 10.1016/j.ultrasmedbio.2022.07.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0301562922005051.

Brian Cowie, Leah Wright, Ben Costello, Kristel Janssens, Erin Howden, Darragh Flannery, Steve Foulkes, Roman Kluger, and Andre La Gerche. Measurement of stroke volume with echocardiography compared to gold standard cardiac magnetic resonance imaging: An observational study. *Journal of Cardiothoracic and Vascular Anesthesia*, 37(9):1639–1645, September 2023. ISSN 10530770. doi: 10.1053/j.jvca.2023.05.024. URL https://linkinghub.elsevier.com/retrieve/pii/S1053077023003300.

R P Lewis, S E Rittogers, W F Froester, and H Boudoulas. A critical review of the systolic time intervals. *Circulation*, 56(2):146–158, August 1977. ISSN 0009-7322, 1524-4539. doi: 10.1161/01.CIR.56.2.146. URL https://www.ahajournals.org/doi/10.1161/01.CIR.56.2.146.

Mayooran Namasivayam, Wei He, Timothy W. Churchill, Romain Capoulade, Shiying Liu, Hang Lee, Jacqueline S. Danik, Michael H. Picard, Philippe Pibarot, Robert A. Levine,

and Judy Hung. Transvalvular flow rate determines prognostic value of aortic valve area in aortic stenosis. *Journal of the American College of Cardiology*, 75(15):1758–1769, April 2020. ISSN 07351097. doi: 10.1016/j.jacc.2020.02.046. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109720344259.

Lachlan Weir, Thomas Meredith, Christopher Hayward, David Muller, Michael Feneley, and Mayooran Namasivayam. Low-flow low-gradient aortic stenosis: Foundations, challenges, and solutions. *Current Treatment Options in Cardiovascular Medicine*, 27(1):27, December 2025. ISSN 1092-8464, 1534-3189. doi: 10.1007/s11936-025-01082-6. URL https://link.springer.com/10.1007/s11936-025-01082-6.

Danielle M. Henkel, Joseph F. Malouf, Heidi M. Connolly, Hector I. Michelena, Maurice E. Sarano, Hartzell V. Schaff, Christopher G. Scott, and Patricia A. Pellikka. Asymptomatic left ventricular systolic dysfunction in patients with severe aortic stenosis. *Journal of the American College of Cardiology*, 60(22):2325–2329, December 2012. ISSN 07351097. doi: 10.1016/j.jacc.2012.08.988. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109712044877.

Saki Ito, William R. Miranda, Vuyisile T. Nkomo, Heidi M. Connolly, Sorin V. Pislaru, Kevin L. Greason, Patricia A. Pellikka, Bradley R. Lewis, and Jae K. Oh. Reduced left ventricular ejection fraction in patients with aortic stenosis. *Journal of the American College of Cardiology*, 71(12):1313–1321, March 2018. ISSN 07351097. doi: 10.1016/j.jacc.2018.01.045. URL https://linkinghub.elsevier.com/retrieve/pii/S073510971830322X.

Jordi S. Dahl, Mackram F. Eleid, Hector I. Michelena, Christopher G. Scott, Rakesh M. Suri, Hartzell V. Schaff, and Patricia A. Pellikka. Effect of left ventricular ejection fraction on postoperative outcome in patients with severe aortic stenosis undergoing aortic valve replacement. *Circulation: Cardiovascular Imaging*, 8(4):e002917, April 2015. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.114.002917. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.114.002917.

Teppei Toya, Satsuki Fukushima, Yusuke Shimahara, Shingo Kasahara, Junjiro Kobayashi, and Tomoyuki Fujita. Reverse left ventricular remodelling after aortic valve replacement for severe aortic insufficiency. *Interactive CardioVascular and Thoracic Surgery*, 32 (6):846–854, May 2021. ISSN 1569-9285. doi: 10.1093/icvts/ivab020. URL https://academic.oup.com/icvts/article/32/6/846/6134831.

Nikolaos Spilias, Trejeeve Martyn, Kara J. Denby, Serge C. Harb, Zoran B. Popovic, and Samir R. Kapadia. Left ventricular systolic dysfunction in aortic stenosis: Pathophysiology, diagnosis, management, and future directions. *Structural Heart*, 6(5):100089, October 2022. ISSN 24748706. doi: 10.1016/j.shj.2022.100089. URL https://linkinghub.elsevier.com/retrieve/pii/S2474870622018796.

Christopher R. deFilippi, DuWayne L. Willett, M. Elizabeth Brickner, Christopher P. Appleton, Clyde W. Yancy, Eric J. Eichhorn, and Paul A. Grayburn. Usefulness of dobutamine echocardiography in distinguishing severe from nonsevere valvular aortic stenosis in patients with depressed left ventricular function and low transvalvular gradients. *The American Journal of Cardiology*, 75(2):191–194, January 1995. ISSN 00029149. doi: 10.1016/S0002-9149(00)80078-8. URL https://linkinghub.elsevier.com/retrieve/pii/S0002914900800788.

Marie-Annick Clavel, Ian G. Burwash, Gerald Mundigler, Jean G. Dumesnil, Helmut Baumgartner, Jutta Bergler-Klein, Mario Sénéchal, Patrick Mathieu, Christian Couture, Rob Beanlands, and Philippe Pibarot. Validation of conventional and simplified methods to calculate projected valve area at normal flow rate in patients with low flow, low gradient aortic stenosis: The multicenter topas (true or pseudo severe aortic stenosis) study. *Journal of the American Society of Echocardiography*, 23 (4):380–386, April 2010. ISSN 08947317. doi: 10.1016/j.echo.2010.02.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731710000994.

Philippe Pibarot and Jean G. Dumesnil. Low-flow, low-gradient aortic stenosis with normal and depressed left ventricular ejection fraction. *Journal of the American College of Cardiology*, 60(19):1845–1853, November 2012. ISSN 07351097. doi: 10.1016/j.jacc.2012.06.051. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109712041058.

Anastasia Vamvakidou, Navtej Chahal, and Roxy Senior. Lack of stroke volume determined flow reserve does not always preclude assessment of severity of aortic stenosis in low-flow low-gradient state during dobutamine echocardiography. *JACC: Cardiovascular Imaging*, 10(4):491–493, April 2017. ISSN 1936878X. doi: 10.1016/j.jcmg.2016.04.001. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X1630242X.

Christophe Tribouilloy, Franck Lévy, Dan Rusinaru, Pascal Guéret, Hélène Petit-Eisenmann, Serge Baleynaud, Yannick Jobic, Catherine Adams, Bernard Lelong, Agnès Pasquet, Christophe Chauvel, Damien Metz, Jean-Paul Quéré, and Jean-Luc Monin. Outcome after aortic valve replacement for low-flow/low-gradient aortic stenosis without contractile reserve on dobutamine stress echocardiography. *Journal of the American College of Cardiology*, 53(20):1865–1873, May 2009. ISSN 07351097. doi: 10.1016/j.jacc.2009.02.026. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109709007049.

Julien Ternacle, Laura Krapf, Dania Mohty, Julien Magne, Annabelle Nguyen, Arnault Galat, Romain Gallet, Emmanuel Teiger, Nancy Côté, Marie-Annick Clavel, François Tournoux, Philippe Pibarot, and Thibaud Damy. Aortic stenosis and cardiac amyloidosis. *Journal of the American College of Cardiology*, 74(21):2638–2651, November 2019. ISSN 07351097. doi: 10.1016/j.jacc.2019.09.056. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109719379264.

Vikash Jaiswal, Vibhor Agrawal, Yashita Khulbe, Muhammad Hanif, Helen Huang, Maha Hameed, Abhigan Babu Shrestha, Francesco Perone, Charmy Parikh, Sabas Ivan Gomez, Kusum Paudel, Jerome Zacks, Kendra J Grubb, Salvatore De Rosa, and Alessia Gimelli. Cardiac amyloidosis and aortic stenosis: a state-of-the-art review. *European Heart Journal Open*, 3(6):oead106, November 2023. ISSN 2752-4191. doi: 10.1093/ehjopen/oead106. URL https://academic.oup.com/ehjopen/article/doi/10.1093/ehjopen/oead106/7310830.

Christian Nitsche, Paul R. Scully, Kush P. Patel, Andreas A. Kammerlander, Matthias Koschutnik, Carolina Dona, Tim Wollenweber, Nida Ahmed, George D. Thornton, Andrew D. Kelion, Nikant Sabharwal, James D. Newton, Muhiddin Ozkor, Simon Kennon, Michael Mullen, Guy Lloyd, Marianna Fontana, Philip N. Hawkins, Francesca Pugliese, Leon J. Menezes, James C. Moon, Julia Mascherbauer, and Thomas A. Treibel. Prevalence and outcomes of concomitant aortic stenosis and cardiac amyloidosis. *Journal of the American College of Cardiology*, 77(2):128–139, January 2021. ISSN 07351097. doi: 10.1016/j.jacc.2020.11.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109720377354.

Mohamed-Salah Annabi, Eden Touboul, Abdellaziz Dahou, Ian G. Burwash, Jutta Bergler-Klein, Maurice Enriquez-Sarano, Stefan Orwat, Helmut Baumgartner, Julia Mascherbauer, Gerald Mundigler, João L. Cavalcante, Éric Larose, Philippe Pibarot, and Marie-Annick Clavel. Dobutamine stress echocardiography for management of low-flow, low-gradient aortic stenosis. *Journal of the American College of Cardiology*, 71(5):475–485, February 2018. ISSN 07351097. doi: 10.1016/j.jacc.2017.11.052. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109717417785.

Rik Adrichem, Thijmen W. Hokken, Sjoerd Bouwmeester, Ola Abdelkarim, Birgit Vogel, David I. Blusztein, Verena Veulemans, Jurrien H. Kuneman, Marcel L. Geleijnse, Sarah Verhemel, Mark M.P. Van Den Dorpel, Isabella Kardys, Pim A.L. Tonino, Su Min Chang, Nadeen N. Faza, Stephanie Jou, Hiroki A. Ueyama, Joanna Bartkowiak, Tobias Zeus, Jeroen J. Bax, Philippe B. Bertrand, Rebecca T. Hahn, Susheel K. Kodali, Stamatios Lerakis, Roxana Mehran, Stephen H. Little, Patrick Houthuizen, and Nicolas M. Van Mieghem. Diagnostic value of aortic valve calcification levels in the assessment of low-gradient aortic stenosis. *JACC: Cardiovascular Imaging*, 17(8):847–860, August 2024. ISSN 1936878X. doi: 10.1016/j.jcmg.2024.03.014. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X24001293.

Elad M. Bar Gil, Ruslan Sergienko, Nir Roguin, Shoham Birman, and Sergio L. Kobal. Prognostic value of transvalvular flow rate in aortic stenosis: Implications for risk stratification. *Echocardiography*, 42(1):e70077, January 2025. ISSN 0742-2822, 1540-8175. doi: 10.1111/echo.70077. URL https://onlinelibrary.wiley.com/doi/10.1111/echo.70077.

Anastasia Vamvakidou, Wenying Jin, Oleksandr Danylenko, Navtej Chahal, Rajdeep Khattar, and Roxy Senior. Low transvalvular flow rate predicts mortality in patients with low-gradient aortic stenosis following aortic valve intervention. *JACC: Cardiovascular Imaging*, 12(9):1715–1724, September 2019. ISSN 1936878X. doi: 10.1016/j.jcmg.2018.01.011. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X18301165.

Sara Hungerford, Eleanor Rye, Hari Sritharan, Peter Hansen, Christopher Choong, and Ravinay Bhindi. Re-exploration of flow rate in low-flow, low-gradient aortic stenosis. *Catheterization and Cardiovascular Interventions*, page ccd.70029, August 2025. ISSN 1522-1946, 1522-726X. doi: 10.1002/ccd.70029. URL https://onlinelibrary.wiley.com/doi/10.1002/ccd.70029.

Navtej S. Chahal, Maria Drakopoulou, Ana M. Gonzalez-Gonzalez, Ramasamy Manivarmane, Rajdeep Khattar, and Roxy Senior. Resting aortic valve area at normal transaortic flow rate reflects true valve area in suspected low-gradient severe aortic stenosis. *JACC: Cardiovascular Imaging*, 8(10):1133–1139, October 2015. ISSN 1936878X. doi: 10.1016/j.jcmg.2015.04.021. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X15005380.

Claudia Blais, Ian G. Burwash, Gerald Mundigler, Jean G. Dumesnil, Nicole Loho, Florian Rader, Helmut Baumgartner, Rob S. Beanlands, Boris Chayer, Lyes Kadem, Damien Garcia, Louis-Gilles Durand, and Philippe Pibarot. Projected valve area at normal flow rate improves the assessment of stenosis severity in patients with low-flow, low-gradient aortic stenosis: The multicenter topas (truly or pseudo-severe aortic stenosis) study. *Circulation*, 113(5):711–721, February 2006. ISSN 0009-7322, 1524-4539. doi: 10.1161/CIRCULATIONAHA.105.557678. URL https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.105.557678.

Kimi Sato, Kesavan Sankaramangalam, Krishna Kandregula, Jennifer A. Bullen, Samir R. Kapadia, Amar Krishnaswamy, Stephanie Mick, L. Leonardo Rodriguez, Richard A. Grimm, Venu Menon, Milind Y. Desai, Lars G. Svensson, Brian P. Griffin, and Zoran B. Popović. Contemporary outcomes in low-gradient aortic stenosis patients who underwent dobutamine stress echocardiography. *Journal of the American Heart Association*, 8(6):e011168, March 2019. ISSN 2047-9980. doi: 10.1161/JAHA.118.011168. URL https://www.ahajournals.org/doi/10.1161/JAHA.118.011168.

Jae K. Oh, Charles P. Taliercio, David R. Holmes, Guy S. Reeder, Kent R. Bailey, James B. Seward, and A.Jamil Tajik. Prediction of the severity of aortic stenosis by doppler aortic valve area determination: Prospective doppler-catheterization correlation in 100 patients. *Journal of the American College of Cardiology*, 11(6):1227–1234, June 1988. ISSN 07351097. doi: 10.1016/0735-1097(88)90286-0. URL https://linkinghub.elsevier.com/retrieve/pii/0735109788902860.

Dan Rusinaru, Dorothée Malaquin, Sylvestre Maréchaux, Nicolas Debry, and Christophe Tribouilloy. Relation of dimensionless index to long-term outcome in aortic stenosis with preserved lvef. *JACC: Cardiovascular Imaging*, 8(7):766–775, July 2015. ISSN 1936878X. doi: 10.1016/j.jcmg.2015.01.023. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X15002673.

Alexandre Altes, Nicolas Thellier, Dan Rusinaru, Wassima Marsou, Yohann Bohbot, Gagandeep Chadha, Blandine Leman, Pierre Paquet, Pierre Vladimir Ennezat, Christophe Tribouilloy, and Sylvestre Maréchaux. Dimensionless index in patients with low-gradient severe aortic stenosis and preserved ejection fraction. *Circulation: Cardiovascular Imaging*, 13(10):e010925, October 2020. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.120.010925. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.120.010925.

Mayooran Namasivayam, Timothy Churchill, Romain Capoulade, Philippe Pibarot, Jacqueline Suk Danik, Michael H. Picard, Robert A. Levine, and Judy W. Hung. Dimensionless index complements transvalvular flow rate in solving discordant low gradient aortic stenosis. *Journal of the American College of Cardiology*, 79(9):1205, March 2022a. ISSN 07351097. doi: 10.1016/S0735-1097(22)02196-9. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109722021969.

Sergio Gamaza-Chulián, Enrique Díaz-Retamino, Santiago Camacho-Freire, Dolores Ruiz-Fernández, Alejandro Gutiérrez-Barrios, and Jesús Oneto-Otero. Acceleration time and ratio of acceleration time to ejection time in aortic stenosis: New echocardiographic diagnostic parameters. *Journal of the American Society of Echocardiography*, 30(10):947–955, October 2017. ISSN 08947317. doi: 10.1016/j.echo.2017.06.001. URL https://linkinghub.elsevier.com/retrieve/pii/S089473171730442X.

Sylvestre Maréchaux and Christophe Tribouilloy. Acceleration time in aortic stenosis: A new life for an old parameter. *Circulation: Cardiovascular Imaging*, 14(1):e012234, January 2021. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.120.012234. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.120.012234.

Javier Bermejo, J.Carlos Antoranz, Miguel A. Garcia-Fernandez, M.Mar Moreno, and Juan Luis Delcan. Flow dynamics of stenotic aortic valves assessed by signal processing of doppler spectrograms. *The American Journal of Cardiology*, 85(5):611–617, March 2000. ISSN 00029149. doi: 10.1016/S0002-9149(99)00820-6. URL https://linkinghub.elsevier.com/retrieve/pii/S0002914999008206.

Alexandre Altes, Maximilien Sochala, David Attias, Julien Dreyfus, Yohann Bohbot, Manuel Toledano, Laurent Macron, Cedric Renard, Gagandeep Chadha, Arianne Truffier, Raphaëlle A. Guerbaai, Pierre Vladimir Ennezat, Pierre Graux, Christophe Tribouilloy, and Sylvestre Maréchaux. Correlates of the ratio of acceleration time to ejection time in patients with aortic stenosis: An echocardiographic and computed tomography study. *Archives*

of Cardiovascular Diseases, 112(10):567–575, October 2019. ISSN 18752136. doi: 10.1016/j.acvd.2019.06.004. URL https://linkinghub.elsevier.com/retrieve/pii/S1875213619301299.

Eigir Einarsen, Dana Cramariuc, Edda Bahlmann, Helga Midtbo, John B. Chambers, and Eva Gerdts. Higher acceleration/ejection time ratio predicts impaired outcome in aortic valve stenosis. *Circulation: Cardiovascular Imaging*, 14(1):e011467, January 2021. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.120.011467. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.120.011467.

Adrian Chong, Jonathan Sen, Reza Reyaldeen, Sudhir Wahi, Quan Huynh, William Y.S. Wang, and Thomas H. Marwick. Prognostication and interventional guidance using acceleration-ejection time ratio in undifferentiated paradoxical low-flow low-gradient aortic stenosis. *JACC: Cardiovascular Imaging*, 17(11):1290–1301, November 2024. ISSN 1936878X. doi: 10.1016/j.jcmg.2024.05.015. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X24002274.

Kadriye Memic Sancar, Gamze Babur Guler, Halil Ibrahim Tanboga, Aysel Turkvatan Cansever, Ali Riza Demir, Arda Guler, Meltem Tekin, Begum Uygur, Ali Birant, Yalcin Avci, Ekrem Guler, and Mehmet Erturk. The diagnostic role of "acceleration time" measurement in patients with classical low flow low gradient aortic stenosis with reduced left ventricular ejection fraction. *The International Journal of Cardiovascular Imaging*, 39(3): 481–489, November 2022. ISSN 1875-8312. doi: 10.1007/s10554-022-02745-2. URL https://link.springer.com/10.1007/s10554-022-02745-2.

Sigve Karlsen, Thomas Dahlslett, Bjørnar Grenne, Benthe Sjøli, Otto Smiseth, Thor Edvardsen, and Harald Brunvand. Global longitudinal strain is a more reproducible measure of left ventricular function than ejection fraction regardless of echocardiographic training. *Cardiovascular Ultrasound*, 17(1):18, December 2019. ISSN 1476-7120. doi: 10.1186/s12947-019-0168-9. URL https://cardiovascularultrasound.biomedcentral.com/articles/10.1186/s12947-019-0168-9.

Christos G. Mihos, Jennifer E. Liu, Kelley M. Anderson, Maria Alexandra Pernetz, Jamie M. O'Driscoll, Gerard P. Aurigemma, Francisco Ujueta, Priscilla Wessly, on behalf of the American Heart Association Council on Peripheral Vascular Disease; Council on Cardiovascular, Stroke Nursing;, and Council on Clinical Cardiology. Speckle-tracking strain echocardiography for the assessment of left ventricular structure and function: A scientific statement from the american heart association. *Circulation*, page CIR.0000000000001354, August 2025. ISSN 0009-7322, 1524-4539. doi: 10.1161/CIR.0000000000001354. URL https://www.ahajournals.org/doi/10.1161/CIR.0000000000001354.

Vasileios Anastasiou, Stylianos Daios, Maria-Anna Bazmpani, Dimitrios Moysidis, Thomas Zegkos, Theodoros Karamitsos, Antonios Ziakas, and Vasileios Kamperidis. Shifting from left ventricular ejection fraction to strain imaging in aortic stenosis. *Diagnostics*,

13(10):1756, May 2023. ISSN 2075-4418. doi: 10.3390/diagnostics13101756. URL https://www.mdpi.com/2075-4418/13/10/1756.

Tony Stanton, Rodel Leano, and Thomas H. Marwick. Prediction of all-cause mortality from global longitudinal speckle strain: Comparison with ejection fraction and wall motion scoring. *Circulation: Cardiovascular Imaging*, 2(5):356–364, September 2009. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.109.862334. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.109.862334.

Mersedeh Karvandi and Saeed Ranjbar. *Generalized Strain Components of the Left Ventricular Myocardium*, page 53–59. Springer International Publishing, Cham, 2023. ISBN 9783031290459. doi: 10.1007/978-3-031-29046-6_5. URL https://link.springer.com/10.1007/978-3-031-29046-6_5.

Egidio Imbalzano, Concetta Zito, Scipione Carerj, Giuseppe Oreto, Giuseppe Mandraffino, Maurizio Cusmà-Piccione, Gianluca Di Bella, Carlo Saitta, and Antonino Saitta. Left ventricular function in hypertension: New insight by speckle tracking echocardiography: Preclinical systolic dysfunction in hypertension. *Echocardiography*, 28(6):649–657, July 2011. ISSN 07422822. doi: 10.1111/j.1540-8175.2011.01410.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-8175.2011.01410.x.

Kenya Kusunose, Andrew Goodman, Roosha Parikh, Tyler Barr, Shikhar Agarwal, Zoran B. Popovic, Richard A. Grimm, Brian P. Griffin, and Milind Y. Desai. Incremental prognostic value of left ventricular global longitudinal strain in patients with aortic stenosis and preserved ejection fraction. *Circulation: Cardiovascular Imaging*, 7(6):938–945, November 2014. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.114.002041. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.114.002041.

E. Mara Vollema, Tadafumi Sugimoto, Mylène Shen, Lionel Tastet, Arnold C. T. Ng, Rachid Abou, Nina Ajmone Marsan, Bart Mertens, Raluca Dulgheru, Patrizio Lancellotti, Marie-Annick Clavel, Philippe Pibarot, Philippe Genereux, Martin B. Leon, Victoria Delgado, and Jeroen J. Bax. Association of left ventricular global longitudinal strain with asymptomatic severe aortic stenosis: Natural course and prognostic value. *JAMA Cardiology*, 3(9):839, September 2018. ISSN 2380-6583. doi: 10.1001/jamacardio.2018.2288. URL http://cardiology.jamanetwork.com/article.aspx?doi=10.1001/jamacardio.2018.2288.

Frank Weidemann, Sebastian Herrmann, Stefan Störk, Markus Niemann, Stefan Frantz, Volkmar Lange, Meinrad Beer, Stefan Gattenlöhner, Wolfram Voelker, Georg Ertl, and Jörg M. Strotmann. Impact of myocardial fibrosis in patients with symptomatic severe aortic stenosis. *Circulation*, 120(7):577–584, August 2009. ISSN 0009-7322, 1524-4539. doi: 10.1161/CIRCULATIONAHA.108.847772. URL https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.108.847772.

Alisson Slimani, Julie Melchior, Christophe De Meester, Sophie Pierard, Clotilde Roy, Mihaela Amzulescu, Caroline Bouzin, Frédéric Maes, Agnès Pasquet, Anne-Catherine Pouleur, David Vancraeynest, Bernhard Gerber, Gebrine El Khoury, and Jean-Louis Vanoverschelde. Relative contribution of afterload and interstitial fibrosis to myocardial function in severe aortic stenosis. *JACC: Cardiovascular Imaging*, 13(2):589–600, February 2020. ISSN 1936878X. doi: 10.1016/j.jcmg.2019.05.020. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X19305546.

Luis Afonso, Ashok Kondur, Mengistu Simegn, Ashutosh Niraj, Pawan Hari, Ramanjit Kaur, Preeti Ramappa, Jyotiranjan Pradhan, Deepti Bhandare, Kim A Williams, Sandip Zalawadiya, Aurelio Pinheiro, and Theodore P Abraham. Two-dimensional strain profiles in patients with physiological and pathological hypertrophy and preserved left ventricular systolic function: a comparative analyses. *BMJ Open*, 2(4):e001390, 2012. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2012-001390. URL https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2012-001390.

Adam Castaño, David L Narotsky, Nadira Hamid, Omar K Khalique, Rachelle Morgenstern, Albert DeLuca, Jonah Rubin, Codruta Chiuzan, Tamim Nazif, Torsten Vahl, Isaac George, Susheel Kodali, Martin B Leon, Rebecca Hahn, Sabahat Bokhari, and Mathew S Maurer. Unveiling transthyretin cardiac amyloidosis and its predictors among elderly patients with severe aortic stenosis undergoing transcatheter aortic valve replacement. *European Heart Journal*, 38(38):2879–2887, October 2017. ISSN 0195-668X, 1522-9645. doi: 10.1093/eurheartj/ehx350. URL https://academic.oup.com/eurheartj/article/38/38/2879/4055642.

Arnold C T Ng, Edgard A Prihadi, M Louisa Antoni, Matteo Bertini, See Hooi Ewe, Nina Ajmone Marsan, Dominic Y Leung, Victoria Delgado, and Jeroen J Bax. Left ventricular global longitudinal strain is predictive of all-cause mortality independent of aortic stenosis severity and ejection fraction. *European Heart Journal - Cardiovascular Imaging*, 19(8):859–867, August 2018. ISSN 2047-2404, 2047-2412. doi: 10.1093/ehjci/jex189. URL https://academic.oup.com/ehjcimaging/article/19/8/859/4055928.

Yuan Wang, Minghui Zhang, Hui Chen, and Hongwei Li. Prognostic value of global longitudinal strain in asymptomatic aortic stenosis: A systematic review and meta-analysis. *Frontiers in Cardiovascular Medicine*, 9:778027, February 2022. ISSN 2297-055X. doi: 10.3389/fcvm.2022.778027. URL https://www.frontiersin.org/articles/10.3389/fcvm.2022.778027/full.

Abdellaziz Dahou, Philipp Emanuel Bartko, Romain Capoulade, Marie-Annick Clavel, Gerald Mundigler, Samuel Larue Grondin, Jutta Bergler-Klein, Ian Burwash, Jean G. Dumesnil, Mario Sénéchal, Kim O'Connor, Helmut Baumgartner, and Philippe Pibarot. Usefulness of global left ventricular longitudinal strain for risk stratification in low ejection fraction, low-gradient aortic stenosis: Results from the multicenter true or pseudo-severe aortic stenosis study. *Circulation: Cardiovascular Imaging*, 8(3):e002117, March

2015. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.114.002117. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.114.002117.

Kimi Sato, Yoshihiro Seo, Tomoko Ishizu, Masaaki Takeuchi, Masaki Izumo, Kengo Suzuki, Eiji Yamashita, Shigeru Oshima, Yoshihiro J. Akashi, Yutaka Otsuji, and Kazutaka Aonuma. Prognostic value of global longitudinal strain in paradoxical low-flow, low-gradient severe aortic stenosis with preserved ejection fraction. *Circulation Journal*, 78 (11):2750–2759, 2014. ISSN 1346-9843, 1347-4820. doi: 10.1253/circj.CJ-14-0726. URL https://www.jstage.jst.go.jp/article/circj/78/11/78_CJ-14-0726/_article.

Antonello D'Andrea, Andreina Carbone, Eustachio Agricola, Lucia Riegler, Simona Sperlongano, Giampaolo Tocci, Raffaella Scarafile, Tiziana Formisano, Cristina Capogrosso, Maurizio Cappelli Bigazzi, Eduardo Bossone, Maurizio Galderisi, and Paolo Golino. Predictive value of left ventricular myocardial deformation for left ventricular remodeling in patients with classical low-flow, low-gradient aortic stenosis undergoing transcatheter aortic valve replacement. *Journal of the American Society of Echocardiography*, 32(6):730–736, June 2019. ISSN 08947317. doi: 10.1016/j.echo.2019.03.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731719301294.

Darae Kim, Geu-Ru Hong, Jong-Won Ha, and Chi Young Shim. Association of left ventricular global longitudinal strain during dobutamine stress echocardiography with adverse events in patients with severe low-flow, low-gradient aortic stenosis. *Journal of the American Society of Echocardiography*, 33(11):1424–1425, November 2020. ISSN 08947317. doi: 10.1016/j.echo.2020.03.022. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731720302078.

D. Cramariuc, E. Gerdts, E. S. Davidsen, L. Segadal, and K. Matre. Myocardial deformation in aortic valve stenosis: relation to left ventricular geometry. *Heart*, 96(2):106–112, January 2010. ISSN 1355-6037. doi: 10.1136/hrt.2009.172569. URL https://heart.bmj.com/lookup/doi/10.1136/hrt.2009.172569.

Jens-Uwe Voigt, Gianni Pedrizzetti, Peter Lysyansky, Tom H. Marwick, Helen Houle, Rolf Baumann, Stefano Pedri, Yasuhiro Ito, Yasuhiko Abe, Stephen Metz, Joo Hyun Song, Jamie Hamilton, Partho P. Sengupta, Theodore J. Kolias, Jan D'Hooge, Gerard P. Aurigemma, James D. Thomas, and Luigi Paolo Badano. Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging. *Eur. Heart J. Cardiovasc. Imaging*, 16(1):1–11, January 2015. ISSN 2047-2404. doi: 10.1093/ehjci/jeu184.

David Attias, Laurent Macron, Julien Dreyfus, Jean-Luc Monin, Eric Brochet, Laurent Lepage, Guillaume Hekimian, Bernard Iung, Alec Vahanian, and David Messika-Zeitoun. Relationship between longitudinal strain and symptomatic status in aortic stenosis. *Journal of the American Society of Echocardiography*, 26(8):868–874, August 2013. ISSN

08947317. doi: 10.1016/j.echo.2013.05.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731713003106.

Helle Gervig Carstensen, Linnea Hornbech Larsen, Christian Hassager, Klaus Fuglsang Kofoed, Jan Skov Jensen, and Rasmus Mogelvang. Basal longitudinal strain predicts future aortic valve replacement in asymptomatic patients with aortic stenosis. *European Heart Journal – Cardiovascular Imaging*, 17(3):283–292, March 2016. ISSN 2047-2404, 2047-2412. doi: 10.1093/ehjci/jev143. URL https://academic.oup.com/ehjcimaging/article-lookup/doi/10.1093/ehjci/jev143.

Anders Lehmann Dahl Pedersen, Jonas Agerlund Povlsen, Anne Dybro, Tor Skibsted Clemmensen, Anders Hostrup Larsen, Bertil Ladefoged, and Steen Hvitfeldt Poulsen. Prevalence and prognostic implications of increased apical-to-basal strain ratio in patients with aortic stenosis undergoing transcatheter aortic valve replacement. *Journal of the American Society of Echocardiography*, 33(12):1465–1473, December 2020. ISSN 08947317. doi: 10.1016/j.echo.2020.07.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731720304739.

Nicolas Thellier, Alexandre Altes, Michael Rietz, Aymeric Menet, Jeremy Layec, François Outteryck, Ludovic Appert, Christophe Tribouilloy, and Sylvestre Maréchaux. Additive prognostic value of left ventricular dispersion and deformation in patients with severe aortic stenosis. *JACC: Cardiovascular Imaging*, 17(3):235–245, March 2024. ISSN 1936878X. doi: 10.1016/j.jcmg.2023.09.010. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X23004369.

Maidar Tumenbayar, Kazuto Yamaguchi, Hiroyuki Yoshitomi, Akihiro Endo, and Kazuaki Tanabe. Increased apical rotation in patients with severe aortic stenosis assessed by three-dimensional speckle tracking imaging. *Journal of Echocardiography*, 16(1):28–33, March 2018. ISSN 1349-0222, 1880-344X. doi: 10.1007/s12574-017-0347-3. URL http://link.springer.com/10.1007/s12574-017-0347-3.

Anthony A. Holmes, Cynthia C. Taub, Mario J. Garcia, Jian Shan, and David P. Slovut. Increased apical rotation in severe aortic stenosis is associated with reduced survival: A speckle-tracking study. *Journal of the American Society of Echocardiography*, 28(11):1294–1301, November 2015. ISSN 08947317. doi: 10.1016/j.echo.2015.07.029. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731715005544.

Johannes Just Hjertaas, Eigir Einarsen, Eva Gerdts, Marina Kokorina, Christian Arvei Moen, Stig Urheim, Sahrai Saeed, and Knut Matre. Impact of aortic valve stenosis on myocardial deformation in different left ventricular levels: A three-dimensional speckle tracking echocardiography study. *Echocardiography*, 40(10):1028–1039, October 2023. ISSN 0742-2822, 1540-8175. doi: 10.1111/echo.15668. URL https://onlinelibrary.wiley.com/doi/10.1111/echo.15668.

Mayooran Namasivayam, Paul D Myers, John V Guttag, Romain Capoulade, Philippe Pibarot, Michael H Picard, Judy Hung, and Collin M Stultz. Predicting outcomes in patients with aortic stenosis using machine learning: the aortic stenosis risk (asterisk) score. *Open Heart*, 9(1):e001990, May 2022b. ISSN 2053-3624. doi: 10.1136/openhrt-2022-001990. URL https://openheart.bmj.com/lookup/doi/10.1136/openhrt-2022-001990.

Edward Itelman, Yaron Shapira, Alon Shechter, Nadav Loebl, Yuval Altman, Leor Perl, and Ran Kornowski. Prediction of aortic stenosis progression using artificial intelligence. *JACC: Advances*, page 102121, August 2025. ISSN 2772963X. doi: 10.1016/j.jacadv.2025.102121. URL https://linkinghub.elsevier.com/retrieve/pii/S2772963X25005460.

Michal Wrzosek, Mikolaj Buchwald, Patryk Czernik, Szymon Kupinski, Karina Zatorska, Anna Jasinska, Dariusz Zakrzewski, Juliusz Pukacki, Cezary Mazurek, Robert Pekal, and Tomasz Hryniewiecki. Diagnosing severe low-gradient vs moderate aortic stenosis with artificial intelligence based on echocardiography images. *Journal of Imaging Informatics in Medicine*, April 2025. ISSN 2948-2933. doi: 10.1007/s10278-025-01497-4. URL https://link.springer.com/10.1007/s10278-025-01497-4.

Partho P. Sengupta, Sirish Shrestha, Nobuyuki Kagiyama, Yasmin Hamirani, Hemant Kulkarni, Naveena Yanamala, Rong Bing, Calvin W.L. Chin, Tania A. Pawade, David Messika-Zeitoun, Lionel Tastet, Mylène Shen, David E. Newby, Marie-Annick Clavel, Phillippe Pibarot, Marc R. Dweck, Éric Larose, Ezequiel Guzzetti, Mathieu Bernier, Jonathan Beaudoin, Marie Arsenault, Nancy Côté, Russell Everett, William S.A. Jenkins, Christophe Tribouilloy, Julien Dreyfus, Tiffany Mathieu, Cedric Renard, Mesut Gun, Laurent Macron, Jacob W. Sechrist, Joan M. Lacomis, Virginia Nguyen, Laura Galian Gay, Hug Cuéllar Calabria, Ioannis Ntalas, Bernard Prendergast, Ronak Rajani, Arturo Evangelista, and João L. Cavalcante. A machine-learning framework to identify distinct phenotypes of aortic stenosis severity. *JACC: Cardiovascular Imaging*, 14(9): 1707–1720, September 2021. ISSN 1936878X. doi: 10.1016/j.jcmg.2021.03.020. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X21002862.

Soongu Kwak, Yunhwan Lee, Taehoon Ko, Seokhun Yang, In-Chang Hwang, Jun-Bean Park, Yeonyee E. Yoon, Hack-Lyoung Kim, Hyung-Kwan Kim, Yong-Jin Kim, Goo-Yeong Cho, Dae-Won Sohn, Sungho Won, and Seung-Pyo Lee. Unsupervised cluster analysis of patients with aortic stenosis reveals distinct population with different phenotypes and outcomes. *Circulation: Cardiovascular Imaging*, 13(5):e009707, May 2020. ISSN 1941-9651, 1942-0080. doi: 10.1161/CIRCIMAGING.119.009707. URL https://www.ahajournals.org/doi/10.1161/CIRCIMAGING.119.009707.

Yohann Bohbot, Olivier Raitière, Pierre Guignant, Matthieu Ariza, Momar Diouf, Dan Rusinaru, Alexandre Altes, Mesut Gun, Chloé Di Lena, Laura Geneste, Nicolas Thellier, Sylvestre Maréchaux, Fabrice Bauer, and Christophe Tribouilloy. Unsupervised clustering of patients with severe aortic stenosis: A myocardial continuum. *Archives*

of Cardiovascular Diseases, 115(11):578–587, November 2022. ISSN 18752136. doi: 10.1016/j.acvd.2022.06.007. URL https://linkinghub.elsevier.com/retrieve/pii/S1875213622001759.

Kenya Kusunose, Takumasa Tsuji, Yukina Hirata, Tomonori Takahashi, Masataka Sata, Kimi Sato, Noor Albakaa, Tomoko Ishizu, Jun'ichi Kotoku, Yoshihiro Seo, JSE-TAVI investigators, Masaki Izumo, Atsushi Okada, Chisato Izumi, Shu Inami, Yasuharu Takeda, Toshinari Onishi, Yuki Izumi, Akiko Kumagai, Tomoko Fukuda, Naohiko Takahashi, Takeshi Kitai, Hiroyuki Iwano, Shigeo Sugawara, Kazumi Akasaka, Kenji Harada, Yoshiko Masaoka, Kazuaki Tanabe, Takahiro Sakamoto, and Takeshi Takamura. Unsupervised cluster analysis reveals different phenotypes in patients after transcatheter aortic valve replacement. *European Heart Journal Open*, 4(1):oead136, December 2023. ISSN 2752-4191. doi: 10.1093/ehjopen/oead136. URL https://academic.oup.com/ehjopen/article/doi/10.1093/ehjopen/oead136/7481830.

Jonathan Sen, Sudhir Wahi, William Vollbon, Marcus Prior, Alex G.C. De Sá, David B. Ascher, Quan Huynh, and Thomas H. Marwick. Definition and validation of prognostic phenotypes in moderate aortic stenosis. *JACC: Cardiovascular Imaging*, 18(2):133–149, February 2025. ISSN 1936878X. doi: 10.1016/j.jcmg.2024.06.013. URL https://linkinghub.elsevier.com/retrieve/pii/S1936878X24002511.

Mark Lachmann, Elena Rippen, Tibor Schuster, Erion Xhepa, Moritz Von Scheidt, Costanza Pellegrini, Teresa Trenkwalder, Tobias Rheude, Anja Stundl, Ruth Thalmann, Gerhard Harmsen, Shinsuke Yuasa, Heribert Schunkert, Adnan Kastrati, Karl-Ludwig Laugwitz, Christian Kupatt, and Michael Joner. Subphenotyping of patients with aortic stenosis by unsupervised agglomerative clustering of echocardiographic and hemodynamic data. *JACC: Cardiovascular Interventions*, 14(19):2127–2140, October 2021. ISSN 19368798. doi: 10.1016/j.jcin.2021.08.034. URL https://linkinghub.elsevier.com/retrieve/pii/S1936879821015715.

Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning. *Electronic Markets*, 31(3):685–695, September 2021. ISSN 1019-6781, 1422-8890. doi: 10.1007/s12525-021-00475-2. URL https://link.springer.com/10.1007/s12525-021-00475-2.

Ali Madani, Ramy Arnaout, Mohammad Mofrad, and Rima Arnaout. Fast and accurate view classification of echocardiograms using deep learning. *npj Digital Medicine*, 1(1):6, March 2018. ISSN 2398-6352. doi: 10.1038/s41746-017-0013-1. URL https://www.nature.com/articles/s41746-017-0013-1.

Derya Gol Gungor, Bimba Rao, Cynthia Wolverton, and Ismayil Guracar. View classification and object detection in cardiac ultrasound to localize valves via deep learning, 2023. URL https://arxiv.org/abs/2311.00068. arXiv Preprint, Available from: https://arxiv.org/abs/2311.00068.

Sarah Leclerc, Erik Smistad, Joao Pedrosa, Andreas Ostvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, Thomas Grenier, Carole Lartizien, Jan Dhooge, Lasse Lovstakken, and Olivier Bernard. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE Transactions on Medical Imaging*, 38(9):2198–2210, September 2019. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2019.2900516. URL https://ieeexplore.ieee.org/document/8649738/.

David Ouyang, Bryan He, Amirata Ghorbani, Neal Yuan, Joseph Ebinger, Curtis P. Langlotz, Paul A. Heidenreich, Robert A. Harrington, David H. Liang, Euan A. Ashley, and James Y. Zou. Video-based ai for beat-to-beat assessment of cardiac function. *Nature*, 580(7802): 252–256, April 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-2145-8. URL https://www.nature.com/articles/s41586-020-2145-8.

Ivar M. Salte, Andreas Østvik, Sindre H. Olaisen, Sigve Karlsen, Thomas Dahlslett, Erik Smistad, Torfinn K. Eriksen-Volnes, Harald Brunvand, Kristina H. Haugaa, Thor Edvardsen, Håvard Dalen, Lasse Lovstakken, and Bjørnar Grenne. Deep learning for improved precision and reproducibility of left ventricular strain in echocardiography: A test-retest study. *Journal of the American Society of Echocardiography*, 36 (7):788–799, July 2023. ISSN 08947317. doi: 10.1016/j.echo.2023.02.017. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731723001396.

John Nyberg, Andreas Østvik, Ivar M Salte, Sindre Olaisen, Sigve Karlsen, Thomas Dahlslett, Erik Smistad, Torfinn Eriksen-Volnes, Harald Brunvand, Thor Edvardsen, Kristina H Haugaa, Lasse Lovstakken, Havard Dalen, and Bjørnar Grenne. Deep learning improves test–retest reproducibility of regional strain in echocardiography. *European Heart Journal - Imaging Methods and Practice*, 2(4):qyae092, October 2024. ISSN 2755-9637. doi: 10.1093/ehjimp/qyae092. URL https://academic.oup.com/ehjimp/article/doi/10.1093/ehjimp/qyae092/7833373.

Jaeik Jeon, Jiyeon Kim, Yeonggul Jang, Yeonyee E. Yoon, Dawun Jeong, Youngtaek Hong, Seung-Ah Lee, and Hyuk-Jae Chang. A unified approach for comprehensive analysis of various spectral and tissue doppler echocardiography. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, page 1–5, Athens, Greece, May 2024. IEEE. ISBN 9798350313338. doi: 10.1109/ISBI56570.2024.10635387. URL https://ieeexplore.ieee.org/document/10635387/.

Yuki Sahashi, Hirotaka Ieki, Victoria Yuan, Matthew Christensen, Milos Vukadinovic, Christina Binder-Rodriguez, Justin Rhee, James Y. Zou, Bryan He, Paul Cheng, and David Ouyang. Artificial intelligence automation of echocardiographic measurements. *JACC*, page S0735109725073437, September 2025. ISSN 07351097. doi: 10.1016/j.jacc.2025.07.053. URL https://linkinghub.elsevier.com/retrieve/pii/S0735109725073437.

Gregory Holste, Evangelos K Oikonomou, Bobak J Mortazavi, Andreas Coppi, Kamil F Faridi, Edward J Miller, John K Forrest, Robert L McNamara, Lucila Ohno-Machado, Neal Yuan, Aakriti Gupta, David Ouyang, Harlan M Krumholz, Zhangyang Wang, and Rohan Khera. Severe aortic stenosis detection by deep learning applied to echocardiography. *European Heart Journal*, 44(43):4592–4604, November 2023. ISSN 0195-668X, 1522-9645. doi: 10.1093/eurheartj/ehad456. URL https://academic.oup.com/eurheartj/article/44/43/4592/7248551.

Tom Ginsberg, Ro-ee Tal, Michael Tsang, Calum Macdonald, Fatemeh Taheri Dezaki, John Van Der Kuur, Christina Luong, Purang Abolmaesumi, and Teresa Tsang. *Deep Video Networks for Automatic Assessment of Aortic Stenosis in Echocardiography*, volume 12967, page 202–210. Springer International Publishing, 2021. ISBN 9783030875824. doi: 10.1007/978-3-030-87583-1_20. URL https://link.springer.com/10.1007/978-3-030-87583-1_20.

Zhe Huang, Benjamin S. Wessler, and Michael C. Hughes. Detecting heart disease from multi-view ultrasound images via supervised attention multiple instance learning, April 2024. URL http://arxiv.org/abs/2306.00003. arXiv:2306.00003.

N. Ahmadi, M. Y. Tsang, A. N. Gu, T. S. M. Tsang, and P. Abolmaesumi. Transformer-based spatio-temporal analysis for classification of aortic stenosis severity from echocardiography cine series. *IEEE Transactions on Medical Imaging*, 43(1):366–376, January 2024. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2023.3305384. URL https://ieeexplore.ieee.org/document/10218329/.

Gregory Holste, Evangelos K. Oikonomou, Márton Tokodi, Attila Kovács, Zhangyang Wang, and Rohan Khera. Complete ai-enabled echocardiography interpretation with multitask deep learning. *JAMA*, 334(4):306, July 2025. ISSN 0098-7484. doi: 10.1001/jama.2025.8731. URL https://jamanetwork.com/journals/jama/fullarticle/2835630.

Jiesuck Park, Jiyeon Kim, Jaeik Jeon, Yeonyee E. Yoon, Yeonggul Jang, Hyunseok Jeong, Youngtaek Hong, Seung-Ah Lee, Hong-Mi Choi, In-Chang Hwang, Goo-Yeong Cho, and Hyuk-Jae Chang. Artificial intelligence-enhanced comprehensive assessment of the aortic valve stenosis continuum in echocardiography. *eBioMedicine*, 112:105560, February 2025. ISSN 23523964. doi: 10.1016/j.ebiom.2025.105560. URL https://linkinghub.elsevier.com/retrieve/pii/S2352396425000040.

Esther Puyol-Anton, Bram Ruijsink, Bernhard Gerber, Mihaela Silvia Amzulescu, Helene Langet, Mathieu De Craene, Julia A. Schnabel, Paolo Piro, and Andrew P. King. Regional multi-view learning for cardiac motion analysis: Application to identification of dilated cardiomyopathy patients. *IEEE Transactions on Biomedical Engineering*, 66(4):956–966, April 2019. ISSN 0018-9294, 1558-2531. doi: 10.1109/TBME.2018.2865669. URL https://ieeexplore.ieee.org/document/8437130/.

Zhe Huang, Gary Long, Benjamin Wessler, and Michael C. Hughes. Tmed 2: A dataset for semi-supervised classification of echocardiograms. In *DataPerf workshop at ICML*, 2022. URL https://tmed.cs.tufts.edu/papers/HuangEtAl_TMED2_DataPerf_2022.pdf.

Robert W Cox, John Ashburner, Hadley Breman, Kate Fissell, Christian Haselgrove, Christopher J Holmes, Jack L Lancaster, David E Rex, Stephen M Smith, James B Woodward, and Stephen C Strother. A (sort of) new image data format standard: Nifti-1. *NeuroImage*, 22:S137, 2004.

Jella An. Biostatistics 101: Perspective on Statistics in Ophthalmology. *ResearchGate*, 2019. doi: 10.19080/JOJO.2019.07.555724. URL https://www.researchgate.net/publication/337180076_Biostatistics_101_Perspective_on_Statistics_in_Ophthalmology.

Esperanza García-Gonzalo, Zulima Fernández-Muñiz, Paulino Jose Garcia Nieto, and Marta Menéndez. Hard-rock stability analysis for span design in entry-type excavations with learning classifiers. *Materials*, June 2016. doi: 10.3390/ma9070531.

Yosuke Nabeshima, Yoshihiro Seo, and Masaaki Takeuchi. A review of current trends in three-dimensional analysis of left ventricular myocardial strain. *Cardiovascular Ultrasound*, 18(1):23, December 2020. ISSN 1476-7120. doi: 10.1186/s12947-020-00204-3. URL https://cardiovascularultrasound.biomedcentral.com/articles/10.1186/s12947-020-00204-3.

Nina Ajmone Marsan, Blazej Michalski, Matteo Cameli, Tomaz Podlesnikar, Robert Manka, Marta Sitges, Marc R Dweck, and Kristina H Haugaa. Eacvi survey on standardization of cardiac chambers quantification by transthoracic echocardiography. *European Heart Journal - Cardiovascular Imaging*, 21(2):119–123, February 2020. ISSN 2047-2404, 2047-2412. doi: 10.1093/ehjci/jez297. URL https://academic.oup.com/ehjcimaging/article/21/2/119/5670593.

Tudor Trache, Stephan Stöbe, Adrienn Tarr, Dietrich Pfeiffer, and Andreas Hagendorff. The agreement between 3d, standard 2d and triplane 2d speckle tracking: effects of image quality and 3d volume rate. *Echo Research and Practice*, 1(2):71–83, November 2014. ISSN 2055-0456, 2055-0464. doi: 10.1530/ERP-14-0025. URL https://echo.biomedcentral.com/articles/10.1530/ERP-14-0025.

Siri Malm, Sigmund Frigstad, Einar Sagberg, Per Arvid Steen, and Terje Skjarpe. Real-time simultaneous triplane contrast echocardiography gives rapid, accurate, and reproducible assessment of left ventricular volumes and ejection fraction: A comparison with magnetic resonance imaging. *Journal of the American Society of Echocardiography*, 19(12):1494–1501, December 2006. ISSN 08947317. doi: 10.1016/j.echo.2006.06.021. URL https://linkinghub.elsevier.com/retrieve/pii/S0894731706006626.

Jiří Plášek, Tomáš Rychlý, Diana Drieniková, Ondřej Cisovský, Tomáš Grézl, Miroslav Homza, and Jan Václavík. The agreement of a two- and a three-dimensional speckle-tracking global longitudinal strain. *Journal of Clinical Medicine*, 11(9):2402, April 2022. ISSN 2077-0383. doi: 10.3390/jcm11092402. URL https://www.mdpi.com/2077-0383/11/9/2402.

Edward A. Gill and Berthold Klas. Three-dimensional echocardiography: An historical perspective. *Cardiology Clinics*, 25(2):221–229, May 2007. ISSN 07338651. doi: 10.1016/j.ccl.2007.05.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0733865107000343.

Roy W. Martin, G. Bashein, Michael L. Nessly, and Florence H. Sheehan. Methodology for three-dimensional reconstruction of the left ventricle from transesophageal echocardiograms. *Ultrasound in Medicine and Biology*, 19(1):27–38, January 1993. ISSN 03015629. doi: 10.1016/0301-5629(93)90015-G. URL https://linkinghub.elsevier.com/retrieve/pii/030156299390015G.

Tomas Gustavsson, Ragnar Pascher, and Kenneth Caidahl. Model based dynamic 3d reconstruction and display of the left ventricle from 2d cross-sectional echocardiograms. *Computerized Medical Imaging and Graphics*, 17(4–5):273–278, July 1993. ISSN 08956111. doi: 10.1016/0895-6111(93)90017-H. URL https://linkinghub.elsevier.com/retrieve/pii/089561119390017H.

Krishnan Navaneetha Rajan, Zeying Song, Kenneth R. Hoffmann, Marek Belohlavek, Eileen M. McMahon, and Iman Borazjani. Automated three-dimensional reconstruction of the left ventricle from multiple-axis echocardiography. *Journal of Biomechanical Engineering*, 138(1):011003, January 2016. ISSN 0148-0731, 1528-8951. doi: 10.1115/1.4031977. URL https://asmedigitalcollection.asme.org/biomechanical/article/doi/10.1115/1.4031977/371257/Automated-ThreeDimensional-Reconstruction-of-the.

Mohammadali Hedayat, Tatsat R. Patel, Taeouk Kim, Marek Belohlavek, Kenneth R. Hoffmann, and Iman Borazjani. A hybrid echocardiography-cfd framework for ventricular flow simulations. *International Journal for Numerical Methods in Biomedical Engineering*, 36(7):e3352, July 2020. ISSN 2040-7939, 2040-7947. doi: 10.1002/cnm.3352. URL https://onlinelibrary.wiley.com/doi/10.1002/cnm.3352.

Robin Willems, Lex Verberne, Olaf Van Der Sluis, and Clemens V. Verhoosel. Echocardiogram-based ventricular isogeometric cardiac analysis using multi-patch fitted nurbs. *Computer Methods in Applied Mechanics and Engineering*, 425:116958, May 2024. ISSN 00457825. doi: 10.1016/j.cma.2024.116958. URL https://linkinghub.elsevier.com/retrieve/pii/S0045782524002147.

Abhirup Banerjee, Julià Camps, Ernesto Zacur, Christopher M. Andrews, Yoram Rudy, Robin P. Choudhury, Blanca Rodriguez, and Vicente Grau. A completely automated

pipeline for 3d reconstruction of human heart from 2d cine magnetic resonance slices. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2212):20200257, December 2021. ISSN 1364-503X, 1471-2962. doi: 10.1098/rsta.2020.0257. URL https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0257.

Thomas Joyce, Stefano Buoso, Christian T. Stoeck, and Sebastian Kozerke. Rapid inference of personalised left-ventricular meshes by deformation-based differentiable mesh voxelization. *Medical Image Analysis*, 79:102445, July 2022. ISSN 13618415. doi: 10.1016/j.media.2022.102445. URL https://linkinghub.elsevier.com/retrieve/pii/S1361841522000901.

David Stojanovski, Uxio Hermida, Marica Muffoletto, Pablo Lamata, Arian Beqiri, and Alberto Gomez. *Efficient Pix2Vox++ for 3D Cardiac Reconstruction from 2D Echo Views*, volume 13565, page 86–95. Springer International Publishing, Cham, 2022. ISBN 9783031169014. doi: 10.1007/978-3-031-16902-1_9. URL https://link.springer.com/10.1007/978-3-031-16902-1_9.

Zhurong Chen, Jinhua Chen, Wei Zhuo, Wufeng Xue, and Dong Ni. 3d heart reconstruction from sparse pose-agnostic 2d echocardiographic slices. *arXiv Preprint*, 2025. doi: 10.48550/ARXIV.2507.02411. URL https://arxiv.org/abs/2507.02411. https://arxiv.org/abs/2507.02411.

Fabian Laumer, Lena Rubi, Michael A. Matter, Stefano Buoso, Gabriel Fringeli, François Mach, Frank Ruschitzka, Joachim M. Buhmann, and Christian M. Matter. 2d echocardiography video to 3d heart shape reconstruction for clinical application. *Medical Image Analysis*, 101:103434, April 2025. ISSN 13618415. doi: 10.1016/j.media.2024.103434. URL https://linkinghub.elsevier.com/retrieve/pii/S1361841524003591.

Shusil Dangi, Yehuda K. Ben-Zikri, Nathan Cahill, Karl Q. Schwarz, and Cristian A. Linte. Endocardial left ventricle feature tracking and reconstruction from tri-plane transesophageal echocardiography data. In Robert J. Webster and Ziv R. Yaniv, editors, *Proceedings of SPIE — Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*, page 941505, Orlando, Florida, United States, March 2015. doi: 10.1117/12.2082541. URL http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2082541.

Florian Vixège, Alain Berod, Pierre-Yves Courand, Simon Mendez, Franck Nicoud, Philippe Blanc-Benon, Didier Vray, and Damien Garcia. Full-volume three-component intraventricular vector flow mapping by triplane color doppler. *Physics in Medicine and Biology*, 67(9):095004, May 2022. ISSN 0031-9155, 1361-6560. doi: 10.1088/1361-6560/ac62fe. URL https://iopscience.iop.org/article/10.1088/1361-6560/ac62fe.

Ariel Hernán Curiale, Gonzalo Vegas Sánchez-Ferrero, and Santiago Aja-Fernández. Techniques for tracking: image registration. *ResearchGate*, June 2017. doi: 10.1049/PBHE013E_ch15.

W. Dean Bidgood, Jr., Steven C. Horii, Fred W. Prior, and Donald E. Van Syckle. Understanding and Using DICOM, the Data Interchange Standard for. *J. Am. Med. Inform. Assoc.*, 4 (3):199, May 1997. doi: 10.1136/jamia.1997.0040199.

Marina Leitman and Vladimir Tyomkin. Longitudinal displacement for left ventricular function assessment. *Journal of Cardiovascular Development and Disease*, 12(2):53, January 2025. ISSN 2308-3425. doi: 10.3390/jcdd12020053. URL https://www.mdpi.com/2308-3425/12/2/53.

Athanasios Tragakis, Chaitanya Kaul, Roderick Murray-Smith, and Dirk Husmeier. The fully convolutional transformer for medical image segmentation, 2023. URL https://arxiv.org/abs/2206.00566.

Zhe Huang, Xiaowei Yu, Benjamin S. Wessler, and Michael C. Hughes. Semi-supervised multimodal multi-instance learning for aortic stenosis diagnosis. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, page 1–5, Houston, TX, USA, April 2025. IEEE. ISBN 9798331520526. doi: 10.1109/ISBI60581.2025.10981205. URL https://ieeexplore.ieee.org/document/10981205/.

Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., USA, 1 edition, 1997. ISBN 0070428077.

Ronald A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368, 1922.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951. doi: 10.1214/aoms/1177729694.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv PrePrint*, 2014. doi: 10.48550/ARXIV.1412.6980. URL https://arxiv.org/abs/1412.6980.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. URL http://arxiv.org/abs/1505.04597.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv PrePrint*, 2015a. doi: 10.48550/ARXIV.1512.03385. URL https://arxiv.org/abs/1512.03385.

Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions, 2016. URL https://arxiv.org/abs/1511.07122.

François Chollet. Xception: Deep learning with depthwise separable convolutions. *CoRR*, abs/1610.02357, 2016. URL http://arxiv.org/abs/1610.02357.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. URL http://jmlr.org/papers/v15/srivastava14a.html.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012. URL http://arxiv.org/abs/1207.0580.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015b. URL https://arxiv.org/abs/1512.03385.

Wikipedia. Residual Block - Wikipedia, May 2023. URL https://upload.wikimedia.org/wikipedia/commons/b/ba/ResBlock.png. https://upload.wikimedia.org/wikipedia/commons/b/ba/ResBlock.png, Wikimedia Commons. Public domain. Accessed 2025-08-03.

Zhe Huang, Gary Long, Benjamin Wessler, and Michael C. Hughes. A new semi-supervised learning benchmark for classifying view and diagnosing aortic stenosis from echocardiograms, 2021. URL https://arxiv.org/abs/2108.00080.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017. URL http://arxiv.org/abs/1706.03762.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929, 2020. URL https://arxiv.org/abs/2010.11929.

Danilo Avola, Irene Cannistraci, Marco Cascio, Luigi Cinque, Alessio Fagioli, Gian Luca Foresti, Emanuele Rodolà, and Luciana Solito. Mv-ms-fete: Multi-view multi-scale feature extractor and transformer encoder for stenosis recognition in echocardiograms. *Computer Methods and Programs in Biomedicine*, 245:108037, March 2024. ISSN 01692607. doi: 10.1016/j.cmpb.2024.108037. URL https://linkinghub.elsevier.com/retrieve/pii/S0169260724000336.

Hassan Keshvarikhojasteh, Josien Pluim, and Mitko Veta. Multi-head attention-based deep multiple instance learning, 2024. URL https://arxiv.org/abs/2404.05362.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Maria Florina Balcan and Kilian Q. Weinberger,

editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1050–1059, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/gal16.html.

Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. *CoRR*, abs/1610.02391, 2016. URL http://arxiv.org/abs/1610.02391.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *arXiv prePrint*, 2016. doi: 10.48550/ARXIV.1603.02754. URL https://arxiv.org/abs/1603.02754.

Zdravko Virag and Fabijan Lulić. Modeling of aortic valve dynamics in a lumped parameter model of left ventricular-arterial coupling. *ANNALI DELL'UNIVERSITA' DI FERRARA*, 54(2):335–347, December 2008. ISSN 0430-3202, 1827-1510. doi: 10.1007/s11565-008-0051-3. URL http://link.springer.com/10.1007/s11565-008-0051-3.

Ryno Laubscher, Johan Van Der Merwe, Jacques Liebenberg, and Philip Herbst. Dynamic simulation of aortic valve stenosis using a lumped parameter cardiovascular system model with flow regime dependent valve pressure loss characteristics. *Medical Engineering and Physics*, 106:103838, August 2022. ISSN 13504533. doi: 10.1016/j.medengphy.2022.103838. URL https://linkinghub.elsevier.com/retrieve/pii/S135045332200087X.

Esther Puyol-Antón, Matthew Sinclair, Bernhard Gerber, Mihaela Silvia Amzulescu, Hélène Langet, Mathieu De Craene, Paul Aljabar, Paolo Piro, and Andrew P. King. A multimodal spatiotemporal cardiac motion atlas from mr and ultrasound data. *Medical Image Analysis*, 40:96–110, August 2017. ISSN 13618415. doi: 10.1016/j.media.2017.06.002. URL https://linkinghub.elsevier.com/retrieve/pii/S1361841517300890.