

Durham E-Theses

*Characterising Shape Variation in the Human Right
Ventricle Using Statistical Shape Analysis:
Preliminary Outcomes and Potential for Predicting
Hypertension in a Clinical Setting*

ELEANOR LAUREN DESMOND

How to cite:

DESMOND, ELEANOR LAUREN (2023) Characterising Shape Variation in the Human Right Ventricle Using Statistical Shape Analysis: Preliminary Outcomes and Potential for Predicting Hypertension in a Clinical Setting. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a <https://etheses.durham.ac.uk/id/eprint/14963/> is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

Characterising Shape Variation in the Human Right Ventricle Using Statistical Shape Analysis: Preliminary Outcomes and Potential for Predicting Hypertension in a Clinical Setting

Eleanor Lauren Desmond

A thesis presented for the degree of
Doctor of Philosophy



Department of Engineering
Durham University
United Kingdom
October 2022

"A person who can write a long letter with ease, cannot write ill."

— Jane Austen, *Pride and Prejudice*

Characterising Shape Variation in the Human Right Ventricle Using Statistical Shape Analysis: Preliminary Outcomes and Potential for Predicting Hypertension in a Clinical Setting

Eleanor Lauren Desmond

Submitted for the degree of Doctor of Philosophy
October 2022

Abstract

Variations in the shape of the human right ventricle (RV) have previously been shown to be predictive of heart function and long term prognosis in Pulmonary Hypertension (PH), a deadly disease characterised by high blood pressure in the pulmonary arteries. The extent to which ventricular shape is also affected by non-pathological features such as sex, body mass index (BMI) and age is explored in this thesis. If fundamental differences in the shape of a structurally normal RV exist, these might also impact the success of a predictive model. This thesis evaluates the extent to which non-pathological features affect the shape of the RV and determines the best ways, in terms of procedure and analysis, to adapt the model to consistently predict PH. It also identifies areas where the statistical shape analysis procedure is robust, and considers the extent to which specific, non-pathological, characteristics impact the diagnostic potential of the statistical shape model. Finally, recommendations are made on next steps in the development of a classification procedure for PH. The dataset was composed of clinically-obtained, cardiovascular magnetic resonance images (CMR) from two independent sources; The University of Pittsburgh Medical Center and Newcastle University. Shape change is assessed using a 3D statistical shape analysis technique, which topologically maps heart meshes through an harmonic mapping approach to create a unique shape function for each shape. Proper Orthogonal Decomposition (POD) was applied to the complete set of shape functions in order to determine and rank a set of shape features (i.e. modes and corresponding coefficients from the decomposition). MRI scanning protocol produced the most significant difference in shape; a shape mode associated with detail at the RV apex and ventricular length from apex to base strongly correlated with the MRI sequence used to record each subject. Qualitatively, a protocol which skipped slices produced a shorter RV with less detail at the apex. Decomposition of sex, age and BMI also derives unique RV shape descriptors which correspond to anatomically meaningful features. The shape features are shown to be able to predict presence of PH. The predictive model can be improved by including BMI as a factor, but these improvements are mainly concentrated in identification of healthy subjects.

Declaration

The work in this thesis is based on research carried out at the Department of Engineering, University of Durham, United Kingdom. No part of this thesis has been submitted elsewhere for any other degree or qualification. The content is all work produced by the author unless referenced to the contrary in the text.

Copyright © 2022 by Eleanor Lauren Desmond.

“The copyright of this thesis rests with the author. No quotations from it should be published without the author’s prior written consent and any information derived from it should be acknowledged”.

Publications

Conference Publications

Desmond, Eleanor L (2021), *Statistical Shape Analysis of the Human Right Ventricle Across Populations*, ASCE Engineering Mechanics Institute International Conference 2021.

Desmond, Eleanor L, Xu, Jing, Wong, Timothy C, Simon, Marc S, Giani, Stefano & Brigham, John C (2019), *Classification of Pulmonary Hypertension by Shape Analysis of the Human Right Ventricle: Robustness and Preliminary Association with Clinical Outcomes*, Proceedings of the 6th International Conference on Computational & Mathematical Biomedical Engineering CMBE19. Sendai City, Japan.

Journal Article

Xu, Jing, Desmond, Eleanor L., Wong, Timothy C., Neill, Colin G., Simon, Marc A. & Brigham, John C. (2022). *Right Ventricular Shape Feature Quantification for Evaluation of Pulmonary Hypertension: Feasibility and Preliminary Associations With Clinical Outcome*. Journal of Biomechanical Engineering 144(4):044502.

Acknowledgements

First, to my supervisor, Prof. John Brigham, I really can't thank you enough for your enthusiasm, encouragement, patience, diligence and good humour. You believed in me when I didn't believe in myself and I doubt I'd ever have finished this thesis were it not for the reminders to pull myself together on a semi-regular basis. I've always enjoyed our academic discourse and lively debates on sentence structure and I've never once regretted the day I walked into your office in November 2017. To Prof. William Coombs and Dr Stefano Giani: Will, thank you for stepping in as supervisor and for guiding me through the submission process; Stefano: thank you for helping me work through many practical aspects of implementing the methodology. To Prof. Andrew Gallant on my review team for always lending a sympathetic ear, and a very handy box of tissues.

I'm extremely grateful to Drs Tim Wong and Marc Simon at UPMC for providing access to the Pittsburgh data, and an invaluable medical perspective. I was also very appreciative of your constructive comments on my work and assistance with statistics. Thanks too are due to Dr Kieren Hollingsworth at the Newcastle University NMR Centre for facilitating access to the Newcastle data and for formative discussions on segmentation strategy. To Erin Sarosi, though we've never met, thank you for your Pittsburgh dataset segmentations and for having already asked all the questions I wanted to on the ITK-SNAP forum!

To all the CDIM group members; Jing Xu, Rob Zupan, Rodrigo Arauz-Sosa, Carlos Garcia-Verdugo, Soumaya Ouhssousou, Nathan Gavin, Amin Pourasghar, Dongge Jia and Adam Leach. Our group meetings have been enriching, enlightening and often incredibly amusing! Jing, thanks for all of your help throughout my PhD; for consistently digging out files from years gone by and for your patient (and repeated) explanations of the intricacies of running the code. Rob, thank you for willingly troubleshooting my remote access to your work computer, often at unsociable hours for one or both of us; for having really great taste in breakfast venues and for taking me to see a real-life West Virginian country road! Adam, thank you so much for building ADAMCUS, for answering my incessant and usually stupid questions and, of course, for teaching me to code in the first place. Carlos and Rodrigo, thanks for letting me stay with you in Pittsburgh and for the general entertainment! Nathan,

your unflappable patience with my fluctuating list of segmentation requirements was greatly appreciated, and I've always enjoyed our coffee bar (and actual bar) chats! Soumaya, thanks so much for your help with network troubleshooting and for letting me remote connect to your computer during the US night time.

To Dr Richard Williams and Aidan Duffy for being wonderful office mates (sorry Rich, it's not like you had a choice in the matter!). Thank you for keeping me grounded, for dealing with my emotional breakdowns, and particularly for not ripping down my door sign! Go team E009! To Liz and Andrea in the coffee bar who continue to enable my caffeine addiction. I have thoroughly appreciated the company and distraction, if not the extent of my tab!

To Durham University Solar Car for being a constant source of joy, and an unashamed time-sink. Without you lot, I'd never have embarked upon a Master's, let alone a PhD and *especially* not one in Engineering (although the hypothetical degree might have been finished on time!). I can't possibly give you all the credit you're due, but I extend most special thanks to Andrew, Sam, Will, Rob Battrum, Ollie, Usmaan, Hugh, Adam, Mike, Andrey, Roy, James, Toby, Kate, Alex Stemmler, Isaac, Liam, Jack, Frances, Rob Bird, Alex Skinner, Joe, Emily, Rachel, Owen, Morgan, Deacs, Ella, Flanders, Anith, Nick, Brian, Ved, Ioan, Matt, Dylan, and of course David! You have all helped to make me the person I am today - thank you for everything.

To Prof. David Sims-Williams for acting both as a mentor and a partner in many intricate and hare-brained logistics operations spanning several continents. Thanks for everything you've done for me. I also would like to mention all of the Engineering academic and administrative staff who've supported me over the years, as well as the Mechanical, Civil and Electrical workshop technicians who've helped me both with lab demonstrating and solar car-related antics.

Yet again I extend my sincerest thanks to Bethany Lines and Jack Towers, as well as to Ruth and Andrew Dean, without whom I would have been homeless for at least part of my degree. To my co-conspirators on John's MCR exec; especially Jenny, Jennie, Andrew, Ruth and Thodge. James; Joe, thank you for putting up with me! Finally, to Mum, Dad and Ciara and any other family members who might be politely pretending to read this thesis. Look, I've finally finished my PhD, so you can stop pestering me about it now! Seriously though, I could never have done it without your support, gentle[ish] cajoling and generous helpings of wine. Thank you all for believing in me.

Contents

Abstract	iii
Declaration	iv
Publications	v
Acknowledgements	vi
List of Figures	xxix
List of Tables	xxxvi
List of Abbreviations	xxxviii
1 Introduction	1
1.1 Pulmonary Hypertension	5
1.2 RV Structure and Function	9
1.3 Imaging Techniques	17
1.3.1 Echocardiography	18
1.3.2 Computed Tomography	19
1.3.3 Magnetic Resonance Imaging	20
1.4 Statistical Shape Analysis	23
1.5 Scope of the Thesis	29

1.5.1	Structure of the Thesis	30
1.6	Novel Contribution	31
1.6.1	Evaluates the Impact of Non-Pathological Characteristics on RV Shape	31
1.6.2	Determines Procedural Constraints Required to Achieve Consistent Shape Features	33
1.6.3	Proposes and Evaluates Approaches to Use the Shape Features to Pre- dict PH	34
2	Methodology	36
2.1	Demographic Information	36
2.2	Data Acquisition	37
2.3	Segmentation	41
2.4	Mesh Generation	42
2.5	Anatomical Boundary Selection	44
2.6	Image Registration	46
2.7	Topological Mapping	48
2.8	Statistical Decomposition	52
2.9	Output and Visualisation	55
3	Population-Based Differences	57
3.1	Introduction	57
3.2	Aims and Hypotheses	63
3.3	Methodology	65
3.3.1	Population Characteristics	66
3.3.2	Statistical Analysis	68
3.3.3	Qualitative Reporting	73
3.4	Results	75
3.4.1	Modal Decomposition	75

3.4.2	Modal Coefficient Correlation with Traits	76
3.4.3	Linear Discriminant Analysis	82
3.5	Conclusions	99
4	Robustness	102
4.1	Introduction	102
4.2	Aims and Hypotheses	110
4.2.1	Aims	110
4.2.2	Hypotheses	110
4.3	Methodology	111
4.3.1	Data Acquisition	112
4.3.2	Segmentation	112
4.3.3	Dateline selection	113
4.3.4	Smoothing criteria	113
4.3.5	Statistical Analysis	115
4.4	Results	124
4.4.1	Robustness to Changes in Dataset Size	124
4.4.2	Robustness to Changes in Dateline Selection	137
4.4.3	Robustness to Inter and Intra-observer Variability	146
4.4.4	Robustness to Changes in Smoothing	166
4.5	Conclusions	179
5	Diagnostic Metric	185
5.1	Introduction	185
5.2	Aims and Hypotheses	190
5.2.1	Aims	190
5.2.2	Hypotheses	191
5.3	Methodology	191

5.3.1	Training and evaluation of the classifiers	191
5.3.2	Support Vector Machines	192
5.4	Results	197
5.4.1	Assessing the Diagnostic Potential of Mode Shape Features	197
5.4.2	Application of Subject Characteristics to Improve the PH Model	201
5.4.3	Subdivision of Datasets to Control for Population Characteristics	210
5.4.4	Implementation of Support Vector Machines	222
5.5	Conclusions	225
6	Conclusion and Evaluation	228
6.1	Conclusions	228
6.2	Future Work	236
	References	240

List of Figures

1.1	a) Cross-sectional schematic of the healthy heart. The RV is shown in the blue region on the left side of the heart, and the LV is depicted in red on the right side. b) The same heart cross-section, but showing the effects of PH-associated flattening and hypertrophy of the base of the RV. Both Figures are depicted at end-Diastole. c) An expanded view of the RV, labelling the three main regions of the ventricle: the inflow tract, the apical body and the outflow tract.	10
1.2	Annotated MRI scan of the RV of a healthy subject from the dataset used in this thesis. The slice is taken in the short axis orientation, and represents the RV at end-Diastole.	12
1.3	Annotated MRI scan of the RV of a healthy subject form the dataset used in this thesis. The slice is taken in the long axis orientation, and represents the RV at end-Diastole. The heart depicted is the same as in Figure 1.2. RV and LV mark the right and left ventricles, and RA and LA mark the right and left atria. The subject is lying supine on the scanning bed, with their head pointing into the page.	12
1.4	Schematic of the RV at end-Diastole, the relaxed state, and end-Systole, the contracted state. This schematic shows the longitudinal orientation of the deeper fibres, and the circumferential layer of the outer, or more superficial fibres.	15

1.5	Schematics describing three common cardiac imaging techniques: a) Echocardiography, b) Cardiac CT (taken from Wright and Grey (2016)) and c) Cardiac MRI (reproduced with permission from Haynes and Holmes (2013))	18
2.1	Flowchart depicting the overall shape analysis workflow which was first established in Xu et al. (2021) and has been improved upon as a part of this thesis.	37
2.2	Top-down, cross-section illustration of the shape of the ventricles in a healthy patient which shows the contractile pattern at a) end-Systole and b) end-Diastole. The location of the dateline on this slice is marked by an arrow. This slice is located at the widest point. c) Provides an alternative view, which shows the location of these cross-sections in the heart.	39
2.3	Screenshot of segmentation slice at end-Diastole for a subject from the University of Pittsburgh dataset. Selections are made in the short-axis orientation using the polygon tool in ITK-SNAP.	43
2.4	Description of the segmentation process. a) Screenshot of segmentation slice at end-Diastole. Selections are made in the short-axis orientation using the polygon tool in ITK-SNAP. The point used for dateline selection is marked in white. b) A 3D graphic of the selections made for each slice, taken from ITK-SNAP but coloured green to differentiate the 2D slice selections from the 3D volume. The RV is oriented with the septal wall facing into the page to emphasise how the RV slices contours selected as part of the relate to the 3D representation. Depth is defined by the MRI slice thickness. The imaging planes are labelled x, y and z.	43

2.5 **a)** an example input RV at end-Diastole, from a non-Hypertensive RV from the Newcastle Dataset. This segmentation shape is exported from ITK-SNAP, where the initial segmentation as a .mha file, which is then imported into the commercial software package Simpleware. This figure is a direct screenshot from ITK-SNAP, and is equivalent to the graphic shown in Figure 2.4b). **b)** output of the same RV from Simpleware. At this stage, the RV has been smoothed using a discrete Gaussian filter and converted to a triangular FE mesh in .inp format, ready for import into ABAQUS FE Software. 44

2.6 A graphical representation of the unit sphere is displayed on the left in addition to an example FE mesh which represents the input into the harmonic mapping step. The sphere is annotated to indicate the localisation of the north pole, which corresponds to the pulmonary valve during the harmonic mapping and the south pole, which corresponds to the RV apex. The line joining them represents the dateline. The dateline, north and south poles are also labelled on the FE mesh to confirm the location of these anatomical regions. 46

2.7 User-interface of ADAMCUS, the specialised software designed to select anatomical reference points for mapping. The red sphere is used to select the North pole, the blue sphere to select the South pole. The yellow sphere is used to select any number of midpoints to guide the dateline. 47

2.8 Point clouds depicting example outputs in success and failure of the ICP rigid registration algorithm. **a)** Example failure of the iterative closest point algorithm to match correct anatomical points before manual correction. **b)** Example of a successful iterative closest point matching. Manual correction was not required in this case. 48

2.9	Exemplar results from the latitude and longitude shape mappings using Laplace's equations. a) Latitude shape mappings for Patient 11 at end-Diastole. Note the North and South Poles are included in this analysis. b) Longitude shape mappings for Patient 11 at end-Diastole. Note that the split occurs at the date-line. North and South poles are removed from this analysis.	50
2.10	Example 2D mapped x coordinate mesh before interpolation. There is a higher density of points around the equator than the poles. The x axis represents longitude coordinates, ranging from 0 to 2π radians and the y axis represents latitude coordinates, ranging from 0 to π radians.	51
2.11	Example 2D-mapped x coordinate mesh after interpolation. The interpolation step places these points on a regular 400x200 grid. The x axis represents longitude coordinates, ranging from 0 to 2π radians and the y axis represents latitude coordinates, ranging from 0 to π radians.	52
2.12	Example cumulative modal energy curve corresponding to each of the shape modes included in the analysis. The eigenvalue corresponding to each mode is a measure of the relative contribution of that particular shape feature to the overall RV shape when reconstructing all shapes in the dataset. The two dotted red lines indicate the number of modes which cover 90% and 99% of the total modal energy respectively.	54
2.13	Example qualitative reconstructions of the mean RV $\pm 2 \times SD$, used to visualise each shape feature.	56
3.1	Flowchart describing the shape analysis workflow	64
3.2	A comparison between PCA and LDA projections. PCA maximises the variance of all values, whereas LDA maximises the distance between individual groups, thereby enabling separation of values into pre-defined categories. Figure adapted from Muraleedharan (2021)	69

3.3 Labelled diagram of the RV, depicting the septal wall, free wall, apex and base regions and the position of the dateline. 74

3.4 Cumulative modal energy corresponding to each of the 192 shape modes and form modes decomposed from their respective POD decompositions. Shape modes are depicted by white circles and form modes by black squares. Dotted lines mark the number of modes required to cover 90% and 99% modal energy, respectively. 7 shape modes and 7 form modes cover 90% of the cumulative modal energy, whereas 30 shape modes and 28 form modes cover 99% of the cumulative modal energy. 75

3.5 **a)** The absolute values of the magnitude of the Pearson correlation coefficient between each of the first 30 modal coefficients corresponding to the end-Diastole shape and subject characteristics. **b)** The absolute value of the magnitude of the Pearson correlation coefficient between each of the first 30 modal coefficients corresponding to the end-Systole shape and subject characteristics. Exact Pearson values are reported for correlations that exceed 0.2. Only rows containing statistically significant values have been included in the analysis. 77

3.6 **A.** shows the magnitude of Mode 1 values superimposed on the mean RV (in both free-wall and septal view). Red areas indicate areas with the greatest motion and blue areas indicate areas with no motion. **B.** shows the mean RV shape $\pm 2 \times$ SD of the modal coefficient of Mode 1 ($x\mu_{ki}$) added and subtracted from the mean. This mean is a composite of end-Systolic and end-Diastolic RVs, and therefore contains phase information. 78

3.7 A box plot showing the distribution of Mode 1 coefficients for the end-Diastolic RVES shape between subjects scanned in Newcastle and those scanned in Pittsburgh. RVES from the Newcastle cohort exhibited larger, positive Mode 1 coefficient values, whereas Pittsburgh RVES exhibited lower, negative coefficients. The horizontal bars represent the mean. 80

3.8 The first two shape features produced as output from Form Analysis, where the RVES shapes were not scaled. The top (blue) shows Mode 1 results from Form Analysis, where the left hand is the mean RVES shape + 2 × SD of modal coefficients and the right hand side shows the mean - 2 × SD of the modal coefficients. The bottom (green) depicts Mean Mode 2 result from Form Analysis. The left hand side shows the mean + 2 × SD of Mode 2 coefficients and the RHS shows mean - 2 × SD of Mode 2 coefficients. Both mode effects are shown in the septal view. Mode 2 is clearly a volume related mode, and Mode 1 is a scanner related to shape feature. 81

3.9 **A.** the qualitative impact of Modes 6 and 23; the best modes for separating non-hypertensive RVES shapes by sex at end-Diastole, as determined by the LDA analysis. **B.** the impact of Mode 19; the best mode at end-Systole, as determined by LDA. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a female RV plotted on the mean (which includes healthy and diseased subjects) and the right side shows the same but for a male individual. The RVs are orientated with the septal wall facing out of the page. 85

- 3.10 **A.** the qualitative impact of Modes 17, 24, 13 and 8; the best modes for separating the 56 hypertensive RVES shapes by sex at end-Diastole. The female RV at end-Diastole exhibits greater rounding of the septal wall, indicated by the double-headed arrow, whereas a flattening is seen in the same location in male RVs, which is illustrated by the two crossed arrows. Basal flattening in the male RV is also indicated by an asterisk. **B.** the impact of Modes 6, 17 and 3; the best modes at end-Systole. At end-Systole, the female RV exhibits basal bulging, marked by '+', and the male RV displays basal flattening marked by an asterisk, and emphasised by the arrow. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a female RV plotted on the mean (which includes healthy and diseased subjects) and the right side shows the same but for a male individual. The RVs are orientated with the septal wall facing out of the page. 86
- 3.11 The mean RVES shape of all female RVs (in yellow), the overall mean RVES shape (in blue), and the mean male RVES shape (in red). These qualitative depictions contain both end-Systolic and end-Diastolic shape information. The mean female RVES displays greater septal wall curvature, whereas the mean male shape is flatter in the same region. 86
- 3.12 Density plots showing the separation of LDA function values between male and female subjects. Female RVES shapes are depicted in pink, and male RVES in blue. **A.** shows the function value density at end-Diastole (using Modes 6 & 23) and **B.** shows the function value density at end-Systole (Mode 19). At both cardiac phases, there is a clear split peak between the sexes. . . . 87

- 3.13 3D plots of the mean RV and the effect of obesity-related modes on RV shape in non-hypertensive RVs. **A.** the qualitative impact of Modes 20, 11 and 2; the best modes for separating RVES shapes by obesity at end-Diastole in non-hypertensive subject. The RHS shows that three combinations of Modes, **A. B. C.** separated by obesity at end-Systole with equal success when all combinations of the top five modes were tested. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a subject with BMI>28 plotted on the mean (which includes healthy and diseased subjects) and the right side shows the same but for subjects with BMI<28. At end-Diastole, the obese RVES shapes exhibit significant hypertrophy. At end-Systole, the non-obese RVs show greater septal curvature. The RVs are orientated with the septal wall facing out of the page. 88
- 3.14 **A.** the qualitative impact of Modes 6, 30, 12, 11; the best modes for separating the 56 hypertensive RVES shapes into obesity groups at end-Diastole. **B.** the impact of Modes 22, 28 and 4; the best separating modes at end-Systole. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a non-obese subject, plotted on the mean (which includes healthy and diseased subjects), and the right side shows the same but for an obese individual. The RVs are orientated with the septal wall facing out of the page. 91
- 3.15 The mean RVES shape of all subjects with a BMI less than 28 (in green), the overall mean RVES shape (in blue) and the mean RVES of subjects with a BMI greater than 28 (in purple). These qualitative depictions contain both end-Systolic and end-Diastolic shape information. The mean non-obese RVES exhibits greater septal curvature, whereas the mean obese RVES shows evidence of hypertrophy. 91

- 3.19 **A.** the qualitative impact of Modes 4, 13 and 9; the best modes for separating the 56 hypertensive RVES shapes by age at end-Diastole, **B.** the impact of Modes 8, 16 and 22; the best modes at end-Systole. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a younger subject plotted on the mean (which includes healthy and diseased subjects) and the right side shows the same but for an older individual. At end-Systole, the older individuals exhibit lengthening of the RV basal region, which is marked by the double headed arrow and also show generalised flattening in that area. The RVs are orientated with the septal wall facing out of the page. 96
- 3.20 The mean RVES shape of all subjects below the mean age of 60.5 years (in mint green), the overall mean RVES shape (in blue), and the mean RVES of subjects older than the mean age of 60.5 years (in pink). These qualitative depictions contain both end-Systolic and end-Diastolic shape information. The mean younger RVES exhibits greater septal curvature, marked by the longer double-headed arrow and greater detail at the apex, marked by the shorter arrow, whereas the mean older RVES shows basal flattening, marked by an asterisk. 97
- 3.21 Density Plot showing the LDA function value distribution of classified subjects at end-Diastole based on whether their age fell above or below the mean. Subjects below the mean age are depicted in pink, and subjects above the mean age are depicted in green. **A.** shows Modes 6, 3 & 13, **B.** shows Modes 6, 13 & 17 and **C.** shows Modes 6, 3 & 2. The older subjects show a concentrated peak around an LDA function value of 1 when separated by Modes 6, 3 & 2, and the younger subjects appear dispersed across the range. 98

3.22 Density Plot showing the LDA function value distribution of classified subjects at end-Systole, based on whether their age fell above or below the mean. Subjects below the mean age are depicted in pink, and subjects above the mean age are depicted in green. **A.** shows Modes 9 & 5. and **B.** shows Modes 9& 18. There is a concentrated peak of younger individuals when separated by Modes 9 & 5. 98

4.1 Screenshots of the short axis MRI scans of five RVs at end-Diastole. **a)** and **b)** were selected on the basis of being clear MRI representations of the RV. **c)** and **d)** were selected on the basis of being unclear representations of the RV, and **e)** was selected randomly. 114

4.2 Comparison between the two dateline selections plotted on the mesh of an example RV mesh (Patient 2). The mesh is presented in a different orientation compared to other RV mesh Figures to best exhibit the distance between the datelines, with the apex and south pole pointing out of the page. 115

4.3 Visualisation of the different levels of smoothing applied when different pixel filters are applied. **a)** depicts a less stringent recursive Gaussian filter (σ_2), **b)** depicts the standard smoothing filter described in the Methodology and utilised in Chapter 3 (σ_3). **c)** shows a more stringent recursive Gaussian filter (σ_4). 117

4.4 Reproduction of Figure 4.2 with datelines coloured by selection attempt. The original dateline nodes are depicted in red, the second attempt nodes in dark blue, and those nodes which overlap between the two selection attempts are shown in purple. The Hausdorff distance calculates the longest minimum distance between nodes selected in the first group and the second group. Again, the mesh is presented in a different orientation to other depictions of the RV, in order to better visualise the dateline placement, since it is located on a boundary between the free wall and the septal wall. 119

- 4.5 Difference in Modal energy distribution between the datasets. **a)** Modal energy profiles for shape analyses conducted when an additional 25 subjects were added and when the dataset size was reduced to 80, 60, 40 and 20 subjects respectively. The vertical dashed lines indicate Mode 7, which covers 90% of the total modal energy, and Mode 30, which describes 99% of the total modal energy. **b)** Difference between the original modal energy curve and those for shape analyses conducted an additional 25 subjects were added and when the dataset size was reduced to 80, 60, 40 and 20 subjects respectively. 125
- 4.6 Qualitative depiction of the RVES shapes when Modes 1, 6, 9 and 19 in the 121, 96 (original), 80, 60, 40 and 20 subject datasets are plotted on the Mean $\pm 2 \times$ SD. Mode 1 appears to encode the same qualitative shape for all dataset sizes. Mode 6 encodes a similar shape feature when the dataset for all except the 20 subject test. Mode 9 diverges when the dataset is increased to 121 subjects, and when the dataset is reduced to 20 subjects. Finally, Mode 19 differs in most comparisons, although the 60 subject test shares many similarities with the original. As the modes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. Hence, a mode can be observed to encode the same shape whether that mode is added or subtracted from the mean. 131
- 4.7 Difference in modal energy distribution between dateline sets. **a)** Modal energy profiles for shape analyses conducted when a new dateline selection was made. The vertical dashed lines indicate Mode 7, which covers 90% of the total modal energy, and Mode 30, which describes 99% of the total modal energy. **b)** Difference between the original modal energy curve and those for shape analyses conducted when the new dateline selection was made. 139

4.8 Box plots showing the distribution of the percentage difference values at end-Systole and end-Diastole for the four example modes. Plots are on a Log_{10} scale to enable the display of outliers without distorting the data. 143

4.9 Qualitative depiction of the RVES shapes when four key modes: 1, 6, 9 and 19 in original and new dateline datasets are plotted on the $\text{Mean} \pm 2 \times \text{SD}$. The original RVES shapes are coloured blue, whereas the dateline replacement sets are coloured red. Modes 1, 6 and 9 appear to encode similar shape features when the dateline replacement sets. Mode 19 does not appear to encode the same shape in the dateline replacement set. As the modes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. Hence, a mode can be observed to encode the same shape whether that mode is added or subtracted from the mean. 144

4.10 Heatmaps displaying the calculated Jaccard Indices between Observer 1 and Observer 2 for inter and intra-observer. A white region indicates no pixel overlap between slices, dark blue indicates a low overlap (20-30%), teal indicates a 50% overlap and yellow indicates a percentage pixel overlap in excess of 90%. Columns represent individual subject RVs, and rows represent slices of those RVs. The vertical black lines are trace lines. The distance of each line from the centre of the cell is proportional to the size of the measurement. 150

4.11 Example of a segmentation on a poor contrast MRI slice. The LHS shows an unmarked slice, the central image shows an attempt by Observer 1 and the RHS shows a segmentation attempt Observer 2. 150

- 4.12 Example 3D segmentation volumes produced from ITK-SNAP. These volumes are produced prior to smoothing. Both surfaces depict segmentation attempts of the same RVES. **a)** is the segmentation attempt by Observer 1 and **b)** is the attempt by Observer 2. The sharp twist half way down is observable in both attempts and is marked by the curved arrow on the 3D surface in **b)**. Notable differences between the two segmentations also exist, and are also marked by arrows. For example, there is a greater transverse difference between the dateline region and the apex in the segmentation attempt by Observer 2, and there is a broader RV outflow tract vehicle in the same segmentation. If no region is selected on a given slice, then 8mm of volume in the z-direction is not included. 152
- 4.13 Distribution of modal energy when comparing segmentations from two observers. **a)** Modal Energy Curve Displaying the Cumulative Modal Energy required to explain the shape in the initial shape analysis and when including five subjects re-segmented by each observer. **b)** Plot showing the difference between the original modal energy curve and the repeats with each observer. Observer 1 differences are marked with an X and Observer 2 differences are marked by a triangle. 157
- 4.14 Box plots showing the distribution of the percentage difference values at end-Systole (top) and end-Diastole (bottom) for the four example modes. Plots are on a Log_{10} scale to enable the display of outliers without distorting the data. 163

4.15 Qualitative depiction of the RVES shapes when four key modes: 1, 6, 9 and 19 in original and the Observer 1 and Observer 2 repeats are plotted on the Mean $\pm 2 \times$ SD. The original RVES shapes are coloured blue, whereas the dateline replacement sets are coloured red, the mode shapes from Observer 1 are coloured green and the mode shapes from Observer 2 are coloured yellow. Modes 1, 6, 9 and 19 appear to encode similar shape features across all three trials. As the modes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. Hence, a mode can be observed to encode the same shape whether that mode is added or subtracted from the mean. 164

4.16 Visualisation of the different levels of smoothing applied when the three different pixel filters are included. **a)** $\sigma 2$ smoothing contains 6162 elements, **b)** $\sigma 3$ smoothing contains 4378 elements and **c)** $\sigma 4$ smoothing contains 3682 elements. This Figure was reproduced in the methodology as Figure 4.3, to demonstrate the effect of smoothing filters. 168

4.17 Distribution of modal energy when different smoothing filters are applied. **a)** Modal energy profiles for shape analyses conducted when the recursive smoothing filter was altered to $\sigma 2$ pixels and $\sigma 4$ pixels. **b)** Plot showing the difference between the original modal energy curve and those for shape analyses conducted when the recursive smoothing filter was altered to $\sigma 2$ pixels and $\sigma 4$ pixels. 169

4.18 Qualitative depiction of the RVES shapes when four key modes: 1, 6, 9 and 19 in original and the σ_2 and σ_4 shape analysis repeats are plotted on the Mean $\pm 2 \times SD$. The original RVES shapes are coloured blue, the more stringent recursive Gaussian smoothing filter (σ_4) replacement sets are coloured mint green, and the less stringent smoothing filter (σ_2) shapes are coloured pink. Modes 1, 6 and 9 appear to encode the same shape features in all smoothing trials. The shape features encoded by Mode 19 share some similarities across the three smoothing filters (e.g. the flattened base, leading to a ‘triangular’ shape), however, there are also clear differences in the effect of the Mode. As the modes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. Hence, a mode can be observed to encode the same shape whether that mode is added or subtracted from the mean. 172

4.19 Box plots showing the distribution of the percentage difference values at end-Systole (top) and end-Diastole (bottom) for the four example modes. Plots are on a Log_{10} scale to enable the display of outliers without distorting the data. 175

5.1 Cross-section of the heart, showing the RV in blue and the LV in red. The standard RV free wall border, in a non-PH heart is marked by the red dotted line. The hypertrophic region is exterior to this region. This Figure is reproduced from Figure 1.1b. 186

5.2 The combined qualitative shape impact of the modal coefficients for Modes 15, 10, 20 and 21 of the most extreme non-PH subject, at end-Diastole, based on LDA function value, and the combined impact of the coefficients for the most extreme PH subject. Figures are the mean RV $\pm 2 \times SD$ 199

5.3 The combined qualitative impact of Modes 15, 10, 20 and 21 in subjects with the 25th and 75th percentile LDA scores at end-Diastole. The left side depicts the 25th percentile score, indicating a typical non-PH RVES, and the right side depicts the 75th percentile score which indicates a typical PH RVES shape. Asterisks in the basal region mark the rounding effect in the non-PH RV, as compared to the mean, and the flattening effect in the PH sufferer. Figures are the mean RV $\pm 2 \times$ SD. 200

5.4 The combined qualitative impact of Modes 2, 15, 14, 8 and 5, the five best modes for determining PH status at end-Systole. The left side shows the most extreme non-PH function value, and the right side shows the most extreme PH subject in terms of LDA function value. The extreme subject exhibits reduced septal curvature as compared to its non-PH counterpart, this difference in curvature is marked by arrows on both RVs. 201

5.5 Confusion matrices depicting the results of the decomposition at end-Diastole. Figure 5.5a shows the four shape modes only, Figure 5.5b adds sex as a factor, Figure 5.5c adds BMI as a factor and Figure 5.5d adds age as a factor. In all cases, ‘True Neg’ refers to subjects in the non-PH group, who are correctly identified as non-PH; ‘False Neg’ refers to PH sufferers who are incorrectly assigned to the non-PH group; ‘True Pos’ refers to PH sufferers who are correctly identified as such, and ‘False Pos’ to non-PH individuals who are incorrectly classified as PH patients. 204

- 5.6 Confusion matrices depicting the results of the decomposition at end-Systole. Figure 5.6a shows the five shape modes only, Figure 5.6b adds sex as a factor, Figure 5.6c adds BMI as a factor and Figure 5.6d adds age as a factor. In all cases, ‘True Neg’ refers to subjects in the non-PH group, who are correctly identified as non-PH; ‘False Neg’ refers to PH sufferers who are incorrectly assigned to the non-PH group; ‘True Pos’ refers to PH sufferers who are correctly identified as such, and ‘False Pos’ to non-PH individuals who are incorrectly classified as PH patients. 205
- 5.7 **a)** Histogram showing BMI value distribution in correctly identified cases using the LDA algorithm at end-Diastole. **b)** Histogram showing BMI value distribution in incorrectly identified cases. 207
- 5.8 Sex-based confusion matrices. The left hand column (Figures 5.8a, 5.8d, 5.8g, 5.8j) depicts the results of the decomposition using shape modes only. The central column shows the results when BMI is included as a factor, and the left hand column shows the results when age is included as a factor. . . . 214
- 5.9 BMI-based confusion matrices. The left hand column (Figures 5.9a, 5.9d, 5.9g, 5.9j) depicts the results of the decomposition using shape modes only. The central column shows the results when sex is included as a factor, and the left hand column shows the results when age is included as a factor. 217
- 5.10 Age-based confusion matrices. The left hand column (Figures 5.10a, 5.10d, 5.10g, 5.10j) depicts the results of the decomposition using shape modes only. The central column shows the results when BMI is included as a factor, and the left hand column shows the results when sex is included as a factor. . . . 221

List of Tables

1.1	The five PH classification groups defined by the WHO	7
2.1	Demographic breakdown of the subjects considered as part of this thesis. . .	37
2.2	A comparison of the scanning protocols employed at the Newcastle and Pittsburgh MRI centres.	38
3.1	Descriptive statistics based on subject BMI. BMI statistics are reported for a series of dichotomous categories: the two sexes (female and male), those below and above average age (based on an age threshold of 60.5 years), between the two centres (Newcastle and Pittsburgh) and between healthy and diseased subjects (Hypertensive and Non-Hypertensive). These descriptive statistics include all 96 subjects, regardless of hypertensive or non-hypertensive status.	67
3.2	Descriptive statistics based on subject age. Age statistics are reported for a series of dichotomous categories: the two sexes (female and male), those below and above average BMI (based on a BMI threshold of $28\text{kg}/\text{m}^3$), between the two centres (Newcastle and Pittsburgh) and between healthy and diseased subjects (Hypertensive and Non-Hypertensive). These descriptive statistics include all 96 subjects, regardless of hypertensive or non-hypertensive status.	67

3.3	Classification scores (% accuracy) of the best overall mode combinations for separating BMI, age and sex categories. The best combination of modes, identified using LDA, is reported for each characteristic and cardiac phase. The percentage accuracy after cross-validation is reported in the ‘Score’ column. The ‘Hyp’ column notes the percentage accuracy when the same LDA discriminant function, used to generate the accuracy score, was applied to hypertensive subjects only, and the ‘Applied Modes’ column reports the percentage accuracy of a new LDA, conducted on the hypertensive individuals.	82
4.1	Statistical Tests performed to evaluate the robustness of each of the four targeted interventions. The same output tests were performed for each intervention, but input tests varied based on the nature of the input data.	116
4.2	Absolute Pearson correlation coefficients comparing the first 20 modes in the 96 subject dataset to cases where dataset size from the original 96 subjects is changed by adding in or removing individuals at end-Systole.	127
4.3	Absolute Pearson correlation coefficients comparing the first 20 modes in the 96 subject dataset to cases where dataset size is changed by adding in or removing individuals at end-Diastole.	128
4.4	Sum of squared difference between the Mode 1 eigenvector when the dataset size is reduced from 96 subjects to 80, 40, 60 and 20 subjects, and increased to 121 subjects. This method accounts for a difference in sign of the mode.	132
4.5	Mean, minimum, maximum and standard deviation of the percentage difference values which compare results from selected modes across the different dataset sizes against the original 96 subject dataset from Chapter 3	133
4.6	Representative raw mode coefficient values for Mode 1 comparing results for different dataset sizes. As the dataset size reduction was performed using a random number generator, the NA values refer to cases where that subject did not form part of the smaller dataset. Six consecutive subjects are shown.	134

4.7 Exemplar LDA Analyses comparing a selection of classification results between the original analysis (non-hypertensive only) and the dataset alterations. Test Combination reports the combination of modes used as inputs in the LDA. Where more than one combination was found equally successful in Chapter 3, both were tested. ‘Classification %’ reports the best accuracy percentage after cross validation. ‘Top Five’ lists the best five modes identified using a new LDA, created based on the new coefficients from the robustness test. 136

4.8 Hausdorff Distance between the two datelines at both end-Systole and end-Diastole. 138

4.9 Absolute Pearson correlation values for the first 20 modes when comparing the 96 original subject mode coefficient values with the results obtained during a test where five of the 96 subjects had their datelines replaced. 140

4.10 Sum of squared difference between the top 5 modes and Modes 6, 9 and 19 when comparing the original mode eigenvectors to those created when 5 new datelines were included. 141

4.11 Mode coefficient values at Mode 1 when comparing the original 96 subject mode coefficient values at end-Systole to the five subjects with redrawn datelines. 142

4.12 Mean, minimum, maximum and standard deviation of the percentage difference values which compare results from selected modes from the new dateline dataset against the original 96 subject dataset from Chapter 3. 143

- 4.13 Exemplar LDA Analyses comparing classification results between the original analysis and the tests which include a new dateline. Test Combination reports the combination of modes used as inputs in the LDA. Where more than one combination was found equally successful in Chapter 3, both were tested. ‘Classification %’ reports the best accuracy percentage after cross validation. ‘Top Five’ lists the best five modes identified using a new LDA, created based on the new coefficients from the robustness test. 145
- 4.14 Overall Jaccard indices comparing pixel overlap across 5 subjects when comparing Observer 1 and Observer 2’s initial and repeat segmentations. This value is calculated by averaging Jaccard Indices across all slices for each subject; the denominator is the number of slices. 147
- 4.15 Raw inter- and intra-observer segmentation volume measurements (cm³) reported for Observer 1 and Observer 2 and their associated percentage differences, reported for selected subjects. 153
- 4.16 Means of coefficient of variation (CV%) for inter- and intra-observer variability. The reported values constitute the mean, median and range across the five subjects assessed by each observer. 156
- 4.17 Sum of squared difference between Modes 1, 6, 9, 19 and the various observers to the original shape analysis result. 158
- 4.18 Absolute Pearson correlation values for the first 20 modes when comparing the 96 original mode coefficients with those obtained during tests where 5 of the RVs were replaced by new segmentations. 159
- 4.19 Raw coefficient values for Mode 1 obtained in the statistical shape analysis procedure for when comparing the original 96 subject mode coefficients with those obtained when 5 repeat subjects were included by each observer. 161

4.20 Mean, minimum, maximum and standard deviation of the percentage difference values comparing the original 96 subject dataset with the shape analysis run between the two observers for Modes 1, 6, 9 and 19. 162

4.21 Exemplar LDA Analyses comparing classification results between the original analysis and the repeats testing inter and intra-observer variability. For the sake of brevity, this analysis depicts just one of the shape analysis results between observers. Test Combination reports the combination of modes used as inputs in the LDA. Where more than one combination was found equally successful in Chapter 3, both were tested. ‘Classification %’ reports the best accuracy percentage after cross validation. ‘Top Five’ lists the best five modes identified using a new LDA, created based on the new coefficients from the robustness test. 165

4.22 Absolute Pearson correlation values for the first 20 modes when comparing the 96 original subject mode coefficient values with those obtained during a test where the smoothing coefficients of five of the 96 subjects were altered to σ_2 and to σ_4 . Pearson Correlations were calculated to compare the σ_2 case to the original 96 subjects, the σ_4 case to the original 96 subjects, and finally σ_2 to σ_4 171

4.23 Correlation coefficient values at Mode 1 when comparing the original 96 subject mode coefficient values at end-Systole to those where the smoothing values of five subjects were altered to 2 and 4 pixels 174

4.24 Mean, minimum, maximum and standard deviation of the percentage difference values which compare results across the different smoothing filters for Modes 1, 6, 9 and 19. 176

4.25 Percentage Difference associated with the correlation coefficient values at Mode 1 when comparing the original 96 subject mode coefficient values at end-Systole. 177

4.26	Sum of squared difference between Modes 1, 6, 9 and 19 when comparing the two smoothing levels to the original shape analysis result, which used a smoothing coefficient of σ_3	177
4.27	Exemplar LDA Analyses comparing classification results between the original analysis and the σ_2 and σ_4 smoothing cases. Test Combination reports the combination of modes used as inputs in the LDA. Where more than one combination was found equally successful in Chapter 3, both were tested. ‘Classification %’ reports the best accuracy percentage after cross validation. ‘Top Five’ lists the best five modes identified using a new LDA, created based on the new coefficients from the robustness test.	178
5.1	Best Mode combinations and cross-validated percentage accuracy scores for characterising PH using Linear Discriminant Analysis	197
5.2	Structure Matrix showing the results of an LDA model built using Modes 15, 10, 20 and 21 at both end-Systole and end-Diastole	198
5.3	LDA cross-validated percentage accuracy scores when the maximum mode combinations for diagnosing PH (Table 5.1) are included and BMI, sex and age are included as factors in the analysis.	202
5.4	Confusion Matrices showing the breakdown of classification scores based on dataset and disease status grouping when classified by a combination of Modes 15, 10, 20 and 21 at end-Diastole	206
5.5	Best Combinations for characterising PH using Linear Discriminant Analysis when Modes 2-30 and BMI, sex and age are included as factors. Cross-validated percentage accuracy scores are also reported.	208

5.6 Composition of each subject category, total values and division into PH and non-PH groups. The cut-off for the age category was mean age across all subjects (60.5 years), and the cut-off for Obesity was a BMI of $28\text{kg}/\text{m}^3$. The Non-PH category is comprised of 63 subjects; 23 healthy ageing, 23 RVs with generalised hypertension and 17 non-PH subjects from the Pittsburgh cohort. 211

5.7 Top Five identified Modes, ranked in order of importance, when separate LDA analyses are performed on Male and Female subjects. Also reported are the associated classification percentages using these modes only, and when BMI and age are included as inputs to the model. 212

5.8 Top Five identified Modes, ranked in order of importance, when separate LDA analyses are performed on the obese and non-obese groups. Also reported are the associated classification percentages using these modes only, and when Sex and age are included as inputs to the model. 215

5.9 Top Five identified Modes, ranked in order of importance, when separate LDA analyses are performed on the below and above mean age groups. Also reported are the associated classification percentages using these modes only, and when BMI and Sex are included as inputs in the model. 218

5.10 Optimally tuned SVM results using Modes 2-30. 223

5.11 Optimally tuned SVM results using Modes 2-30 and BMI. BMI value is included as an input factor to the SVM, in addition to the shape mode coefficients. 224

5.12 Optimally tuned SVM results using Modes 2-30 and Sex. Sex is included as an input factor to the SVM, in addition to the shape mode coefficients. 224

List of Abbreviations

BMI Body Mass Index

CMR Cardiovascular Magnetic Resonance

CT Computed Tomography

CV% Mean of Coefficient of Variation

ECG Electrocardiogram

Echo Echocardiography

FEA Finite Element Analysis

FE Finite Element

FOV Field of View

ICP Iterative Closest Point

LDA Linear Discriminant Analysis

LHS Left Hand Side

LVEF Left Ventricular Ejection Fraction

LV Left Ventricle

MRI Magnetic Resonance Imaging

PAH Pulmonary Arterial Hypertension

PAP Pulmonary Arterial Pressure

PCA Principal Component Analysis

PCWP Pulmonary Capillary Wedge Pressure

PH Pulmonary Hypertension

POD Proper Orthogonal Decomposition

RHC Right Heart Catheterisation

RHS Right Hand Side

RVEF Right Ventricular Ejection Fraction

RVES Right Ventricular Endocardial Surface

RV Right Ventricle

SSM Statistical Shape Modelling

SVM Support Vector Machines

UPMC University of Pittsburgh Medical Center

WHO World Health Organisation

Chapter 1

Introduction

Over a lifetime, the heart constantly changes its shape and function to maintain level cardiac output; the total amount of blood pumped out of each ventricle of the heart, in a process known as remodelling (Cohn et al., 2000). Remodelling can be adaptive, for example, during physiological growth, or when undergoing athletic training, (e.g. Muhl et al., 2008; Weiner & Baggish, 2012), or adverse, which can indicate disease progression. The World Health Organisation (WHO) reports that cardiovascular diseases are the leading global cause of death, taking an estimated 17.9 million lives each year which representing ~32% of total lives lost (WHO 2021). It is therefore unsurprising that research has focused on identifying pathological shape trends which hasten diagnosis and thereby improve the prognosis of deadly diseases. The left ventricle (LV), for example, undergoes dilatation at the early stages of myocardial infarction to maintain blood supply to the circulatory system (Suinesiaputra et al., 2018). This thesis focuses specifically on prediction of Pulmonary Hypertension (PH), a type of high blood pressure which primarily affects the arteries of the lungs and the right side of the heart. PH is a progressive cardiovascular disease that significantly impacts the mechanical properties of the right ventricle (RV) as compensatory morphological changes take place to overcome the elevated blood pressure. PH is associated with high morbidity and mortality, in part because non-specific early disease symptoms often lead to delayed diagnosis (Vonk-Noordegraaf et

al., 2013; Xu et al., 2021).

The most common compensatory shape changes are an increased wall thickness and flattening of the ventricular base in response to the increased afterload caused by high blood pressure in the pulmonary arteries (Haddad et al., 2008; Vonk-Noordegraaf et al., 2013) (afterload is the pressure that the heart must work against to eject blood during systole and is described in greater detail in Section 1.2). However, such an approach still leaves the unknown component of natural variation. Separating out our understanding of shape features related to patient-specific factors, such as demography, age, obesity, or sex, has the potential to enhance our understanding of pathological shape trends by establishing the bounds of normal variation and determining whether any natural shape characteristics enhance or protect against the development of heart disease. Furthermore, the acute response of cardiac remodelling to a pathological event is often protective and beneficial (Suinesiaputra et al., 2018).

Biological shape change is especially challenging to measure given its multivariate, heterogeneous nature (Badrick, 2021). Therefore it is important to consider what a clinician or anatomist might define as ‘normal’ for an organ or structure. To recognise variation in anatomy, first, it is necessary to identify commonalities across individuals in size, form, relative position, appearance and function (Ambellan et al., 2019). Such knowledge is key in the context of the heart, where slight anatomical deviations can be either indicative of or the root cause of systemic heart failure and, ultimately, death. It is also essential to consider whether a fundamental shape difference is expected between individuals. If we take our definition of shape from Dryden and Mardia (2016), ‘shape’ is what remains after removing the effects of rotating, scaling and translating in space. For example, it might be expected for the female heart to be a smaller-scaled version of the male heart in accordance with observations in many other anatomical structures, for example in the lungs (LoMauro & Aliverti, 2018). Equally, we might expect slight differences in heart orientation (translation/rotation within the chest) between individuals of different sizes and fitness levels. However, a crucial unanswered question relates to whether there is a more fundamental difference inherent to population categories,

and if such variations exist, whether they contribute to impaired organ function in a way that should be accounted for in treatment design. Tight control of heart morphology during fetal development is necessary to ensure correct cardiac function (Tan & Lewandowski, 2019), as abnormalities can affect the overall viability of an organism. So, at first consideration, it is unlikely that significant deviations from a standard structure occur.

Developing our understanding of characteristic or integral shape differences could be a crucial ingredient in creating tailored medical tools. On the other hand, if no pattern between shape features and any biological or patient-specific factors is identifiable, this could point to a unique biometric signature. Such non-specificity could have exciting implications in other fields such as security (Rinaldi, 2015), in the same manner as the work by Duta (2009) on biometric technology based on hand-shape.

Non-surgical intervention is preferable in the vast majority of cases, both as preventative tests are more likely to be recommended, and clinically fragile individuals recover more swiftly. In addition to the already routine use of echocardiography in initial testing (NHS, 2020; Pamboucas & Nihoyannopoulos, 2006), over several decades, considerable evidence has accrued in support of Cardiac Magnetic Resonance (CMR) as a standard for the evaluation of many common medical conditions, and this has been reflected in the most recent cardiovascular care guidelines. The precision of measurement is superior to TTE, leading to improved inter- and intra-observer agreement. Similarly, the cost saving of using such a precise and reproducible test is significant, and this fact has been appreciated in clinical practice (Swift et al., 2014). However MRI is still not the most recommended protocol. The lagging recommendations are largely due to two factors (in the United Kingdom, but these experiences are common internationally): first that the previous guidelines were published in 2010, when CMR was not widely used in mainstream clinical practice, and secondly because there is wide geographic inequality in provision and infrastructure (Treibel et al., 2022).

Increased uptake of medical imaging methods such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), combined with associated developments in *post hoc* image

processing techniques, have been instrumental in enabling observation of the nature of shape changes and other indicative markers of pathology in a research environment. Yet, in clinical practice, shape and structural analysis is carried out using much simpler morphometric analyses on 2D images. Traditional anatomical descriptors such as lengths, diameters and angles are taken from a single image slice, as described in the Standardised Image Interpretation and Post-processing Guidelines of Schulz-Menger et al. (2020). Such methods are fast and provide the means to make largely accurate clinical decisions. However, they do not fully exploit the shape information in 3D techniques at the state of the art. Rodero et al. (2021) investigate the extent to which large changes in function are caused by large changes in shape and conclude that using only the modes of shape which capture the most variance (the first 9 modes explain 90% of variance) does not produce accurate synthetic models of heart function. Their research created a 3D model of the healthy heart integrating CT images of all four chambers with a series of rules to define the motion of muscular fibres. These methods might be considered to describe ‘local’ shape information, contrasting with the global approach of statistical shape analysis. It is important to strike a balance between practicality and superior information retrieval. For example, bringing computational tools from engineering and statistical analysis fields is often hindered by the complexity of the approaches, and the reliance on mathematical terms which are difficult to interpret by non-experts (Bruse, 2017).

This chapter will introduce the study by discussing essential background and context under four main subheadings. First, the symptoms of Pulmonary Hypertension will be described in depth; secondly, an overview of the role and anatomy of the RV will be given, thirdly, a brief description of non-invasive imaging techniques will be provided, and finally several statistical shape analysis techniques will be introduced and discussed. The remainder of the introduction is devoted to outlining the research aims and objectives and describing the overall structure of the thesis. The final section will discuss the scope of the work and highlight novel contributions to the literature. Each chapter is also accompanied by relevant literature review and therefore the effect of non-pathological factors on RV shape, and robustness methods will

be reviewed in greater detail later in the thesis.

1.1 Pulmonary Hypertension

Pulmonary Hypertension (PH) is a severe and progressive cardiopulmonary disease that significantly affects the shape, mechanical properties and overall mechanical function of the heart. Left untreated, PH can be deadly, as it can lead to right sided heart failure and ultimately, death. PH describes high blood pressure in the arteries of the lungs, and was initially defined as an increase in mean Pulmonary Arterial Pressure (PAP) beyond 25mmHg at rest (Galiè, Manes, et al., 2009), as assessed by Right Heart Catheterisation (RHC). For reference, resting PAP in healthy individuals is $14 \pm 3 \text{ mmHg}$ (Kovacs et al., 2009). The definition has been extended in recent years to include subjects with a mean PAP of above 20mmHg at rest, in conjunction with pulmonary vascular resistance greater than or equal to 3 Wood Units, as assessed by RHC (Simonneau et al., 2019). Wood Units are a measure of pulmonary vascular resistance, and average resistance amongst healthy controls is 2 (Kovacs et al., 2009).

The blood pressure in the pulmonary arteries is much lower than in systemic circulation, and therefore cannot be accurately or reliably measured by a cuff. RHC also allows for a more accurate, direct measurement of PAP. There has been some debate in the clinical literature regarding the efficacy of RHC. A randomised, controlled trial of the use of pulmonary artery catheters in high risk surgical patients found no benefit to therapy over standard care in high risk, elderly patients (Sandham et al., 2003). The ability of the RV to adapt to increased pulmonary pressures is the main determinant of clinical outcomes in patients with PH. Some RV responses to elevated pulmonary pressures are clearly adaptive, whereas others are thought to be maladaptive. Early recognition of the components of maladaptive change will hopefully enable timely interventions to prevent deterioration of RV function (Lahm et al., 2018).

The modern classification for PH was established in 1998 (Simonneau et al., 2004), and is updated every five years in line with the latest research and clinical practice (Simonneau et al.,

2009, 2013, 2019). A series of different pathological features characterise five diverse clinical PH groups identified by the WHO (2008). Despite all sitting under the broader PH umbrella, each group has distinct pathological and haemodynamic characteristics (Rich & Rich, 2014). Therefore, it is of critical importance to determine the underlying cause of PH, as prognosis and management strategies are often WHO group-dependent (Simonneau et al., 2013; Grünig & Peacock, 2015). Table 1.1 briefly describes each PH category. Despite heterogeneous underlying aetiologies, all forms of PH share similar clinical and haemodynamic findings. Affected individuals often present with shortness of breath, chest pain, fatigue and fainting, especially after intense exercise (Hambly et al., 2016); symptoms which can be associated with a wide array of other cardiac and lower-respiratory problems. This presents challenges to follow-up of suspected cases, as an absence of clear biomarkers for disease progression reflects the underlying complexity of PH pathophysiology (Vachieri & Gaine, 2012).

The non-specificity of early PH symptoms coupled with the disease's relative rarity leads to a tendency for PH to be confused with other, less serious, conditions such as anaemia (Hyduk et al., 2005). The earliest PAH (Group 1) symptoms, for example, often manifest as shortness of breath during exercise (van Wolferen et al., 2008; Rich & Rich, 2014). Consequently, patients tend to be diagnosed only after significant disease progression. This delay in the commencement of effective therapy, in conjunction with the debilitating and progressive nature of PH is extremely detrimental to long-term patient outcomes; the median survival time after diagnosis for untreated PAH was reported as 2.8 years in 1991 (D'Alonzo et al., 1991). Prognosis has not improved substantially despite significant research funding since the early 1990s. Chang et al. (2022) report that contemporary mortality rates for patients with PAH (Group 1), are 8% after one year, 16% after 2 years and 21% after 3 years.

After initial consultation, where symptoms, family history and interactions with other medication are considered, the accepted diagnostic process for PH consists of combination of non-invasive and invasive tests. Through the NHS, this is most commonly an echocardiogram, followed by invasive RHC to directly measure PAP (NHS, 2020). Other recommended

Table 1.1: The five PH classification groups defined by the WHO

Group	Type	Description
1	Pulmonary Arterial Hypertension (PAH)	PH caused by the distal pulmonary arteries becoming narrowed, thickened and stiff; in particular those less than $500\mu\text{m}$ in diameter. PAH can be idiopathic, inherited, induced by drugs or toxins, or caused by connective tissue disease.
2	PH caused by left heart disease	PH caused by disease in the left heart. Pathological changes in this group include enlarged and thickened pulmonary veins, dilatation of the pulmonary capillaries and interstitial oedema. The most common and lethal type (68.5% of a recent population-based study (Wijeratne et al., 2018)).
3	PH caused by lung disease and or hypoxaemia	PH hypoxic vasoconstriction, mechanical stress of hyperinflated lungs, inflammation and the toxic effect of cigarette smoke, The second most common type, around 47.0% of PH patients. Patients experience Group 2 & 3 PH simultaneously.
4	PH caused by chronic thromboembolism (CTEPH)	Occurs if the body is unable to dissolve blood clots in the lungs, which later undergo fibrosis producing scar tissue which can mechanically obstruct the pulmonary arteries. (9.0% of cases)
5	PH caused by unclear or multifactorial mechanisms	PH is secondary to other diseases in ways that are not well understood.

tests include an electrocardiogram (ECG) , often accompanied by exercise tests including a six-minute walk, and lung function tests, ventilation-perfusion scans and chest x-rays (Galiè, Hoeper, et al., 2009; NHS, 2020), all of which serve to help identify PH type.

Increasingly, CMR-based protocols have been introduced for screening, and can be used in conjunction with RHC for diagnostic purposes. RHC is still required to directly measure PAP (Pamboucas & Nihoyannopoulos, 2006). The hypertensive state can also be proxied by several metrics, most notably PAP and pulmonary capillary wedge pressure (PWCP). Clini-

cal guidelines notwithstanding, MRI has several inherent advantages in the diagnosis of PH, especially given recent advances in imaging quality and acquisition time (Peng et al., 2016). CMR is an accurate, reproducible tool for the quantification of ventricular mass and volumes in healthy subjects and those with PAH (WHO Group 1). It has been used successfully to establish normal ranges in healthy subjects (Pamboucas & Nihoyannopoulos, 2006). Since 2007, CMR-based research has led to significant developments in our understanding of PH and right ventricular remodelling, see Torbicki (2015). Specific CMR findings can be suggestive of, and consequently instrumental in, diagnosing the underlying cause of a particular PH case (Bradlow et al., 2012). For example, the presence of shunts and aberrant pulmonary veins may indicate PAH (Group 1). Alternatively, LV hypertrophy, or other wall motion abnormalities may indicate PH due to left heart disease (Group 2). Despite significant advances in the development of non-invasive imaging techniques and the identification of potential non-invasive indicators (Vonk-Noordegraaf & Galiè, 2011), RHC remains the clinically recommended procedure and is essential for confirmatory testing and diagnosis (Galiè et al., 2015).

Although there has been substantial progress in the understanding of both PH pathogenesis and treatment, right ventricular (RV) failure remains the most common fatal pathway in PH. Sitbon et al. (2015) undertook an important inter-study collaboration between the French Pulmonary Hypertension Network (FPHN) and the Registry to Evaluate Early and Long-term Pulmonary Arterial Hypertension Disease Management (REVEAL). This study took PAH survival algorithms developed in both studies and cross-applied them. Both the REVEAL risk score calculator (Benza et al., 2010) and the FPHN predictive algorithm (Humbert et al., 2010) were both proven accurate and well-calibrated when applied to the opposing dataset. Therefore, it is likely that prognostic generalisation across cohorts and continents is possible.

In recent years, predictive models for when fatality is most likely to occur have been proposed (Mauger et al., 2019; Xu et al., 2021). Many, if not all of these models have made use of advances in image processing, and a more holistic view of the modes of RV failure are likely to require advanced imaging and machine learning. Thus far, machine learning has

been limited in diagnostic models. However, Dawes et al. (2017) have published a survival-based model, noted in a 2019 review of best practice in PH diagnostics (Frost et al., 2019) as a helpful starting point.

Several PH-specific shape changes have been recorded. Broadly, shape change can be considered as *adaptive* or *mal-adaptive*, with the most common shape changes being a rounding out, or flattening of the base of the ventricle in response to high PAP and hypertrophy of the RV wall. An additional, widely observed example is that in RV failure secondary to PH, the septal wall position is shifted towards the left side of the heart with intact LV function (Friedberg & Redington, 2014; Dwivedi & Axel, 2017).

1.2 Structure and Function of the Right Ventricle

Understanding typical RV structure and function is significant to appreciating the impact and severity of observed variations in ventricular shape. The RV structure has several features which define its function. The RV pumps oxygen-poor, carbon dioxide-rich blood to the lungs via the pulmonary arteries, where it then becomes oxygenated and can be pumped around the body. Under normal conditions, the right and left ventricles (LV) are connected in series and therefore pump, on average, the same effective stroke volume (Haddad et al., 2008). The RV is the most anteriorly situated cardiac chamber, sitting directly behind the sternum at the front of the chest. The basic structure of the RV is displayed in diagrammatic form in Figure 1.1a and also in the two annotated screenshots of MRI imaging planes, Figures 1.2 and 1.3. The main structure of the organ consists of the apical body, inflow tract and outflow tract (Peng et al., 2016), and these regions are shown graphically in Figure 1.1c. The muscular wall which is shared with the LV is referred to as the septum, whereas the opposite wall is denoted as the free wall. These are marked on the diagram in Figures 1.1a and 1.1b.

Until recently, the RV was considered relatively unimportant as a component in the circulatory system, both in terms of its contribution to overall heart function and as a contributor to

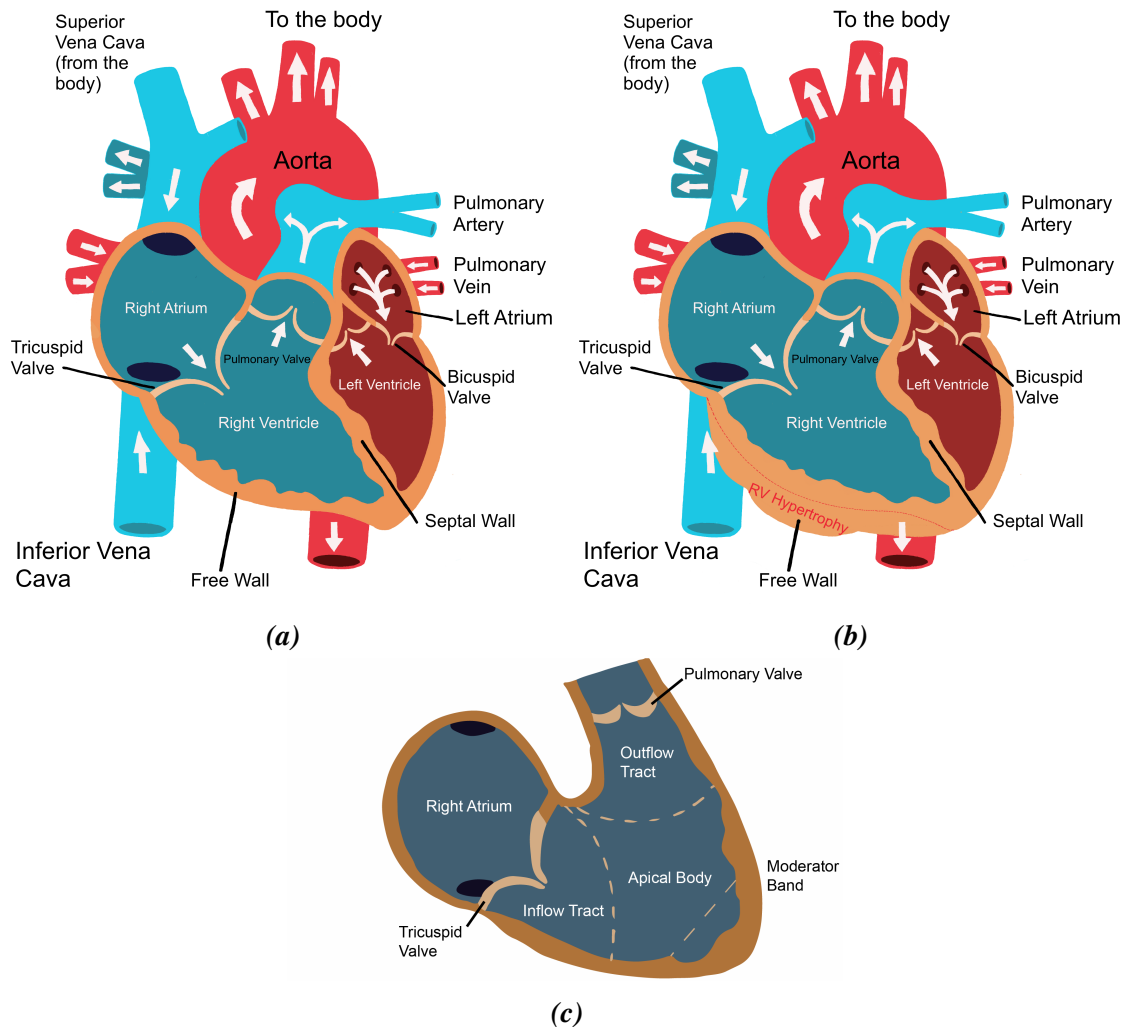
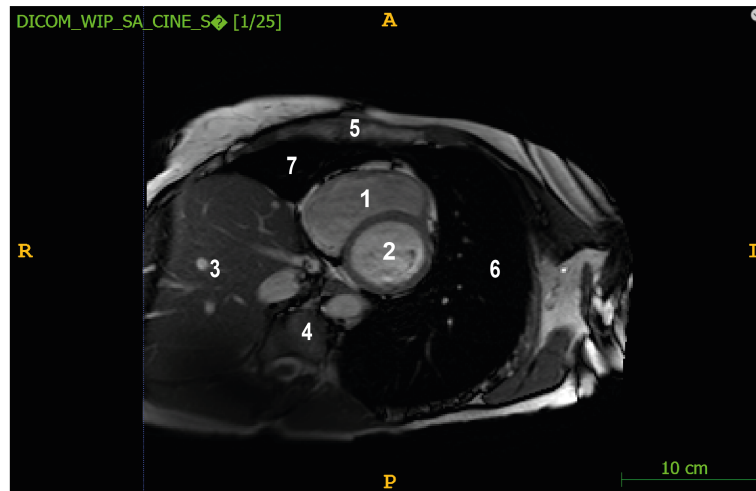


Figure 1.1: *a) Cross-sectional schematic of the healthy heart. The RV is shown in the blue region on the left side of the heart, and the LV is depicted in red on the right side. b) The same heart cross-section, but showing the effects of PH-associated flattening and hypertrophy of the base of the RV. Both Figures are depicted at end-Diastole. c) An expanded view of the RV, labelling the three main regions of the ventricle: the inflow tract, the apical body and the outflow tract.*

hypertensive heart failure (Sheehan & Redington, 2008). This general apathy originated from early studies (Starr et al., 1943; Kagan, 1952) which showed effectively no increase in venous pressure after sustaining significant damage to the RV free wall, leading to an underestimation of the importance of RV function in overall heart performance and in heart failure. Additional experiments in animals (Rose et al., 1955) showed that replacement of the RV-free wall was well tolerated and did not cause a reduction in cardiac output. These initial studies led to a general belief that the laws defining cardiac mechanics, and the therapeutic techniques derived from them, were as equally applicable to the RV as the LV (Walker & Buttrick, 2013) and, since overall pressures were lower (mean peak systolic RV pressure is 25mmHg vs 130mmHg in the LV (Fowler, 1980)), the RV was not considered worthy of significant research attention.

In contrast to the LV, which is near conical in shape, the RV is triangular when viewed from the front and, since it curves over the LV, crescent-shaped in a cross-sectional view (Ho, 2006). The RV's complex, irregular shape exists because its septal contour is indented by the more dominant LV (Sheehan & Redington, 2008). The irregular shape is difficult to see in the cross-sectional view in Figure 1.1a, but can be clearly observed in the top down (short-axis) view in Figure 1.2. In this MRI scan the RV is marked 1 and the LV is marked 2. The RV displays increased variability between short-axis MRI segments as compared to the LV, owing to its twisted shape (Haddad et al., 2008), and this increased variability has an impact on non-invasive imaging, which is discussed further in Section 1.3.

The large difference in RV volume and shape between each short axis slice makes *in silico* analysis more difficult than for the LV, since the intervening geometry between slices is more complex to infer. The ventricle's position in the chest also produces a poor acoustic window for 3D echo-cardiography and can cause low-quality images (Bernardino et al., 2021); the acoustic window is the location from which an ultrasound probe scans. Several factors, including obesity and the presence of foreign objects, can limit this window and the resulting scan quality. There are also challenges associated with evaluating right heart function using a standard short-axis MRI image stack, as was required in this thesis (Chaturvedi et al.,



- | | |
|---------------------|----------------|
| 1 - Right Ventricle | 5 - Chest Wall |
| 2 - Left Ventricle | 6 - Left Lung |
| 3 - Liver | 7 - Right Lung |
| 4 - Stomach | |

Figure 1.2: Annotated MRI scan of the RV of a healthy subject from the dataset used in this thesis. The slice is taken in the short axis orientation, and represents the RV at end-Diastole.

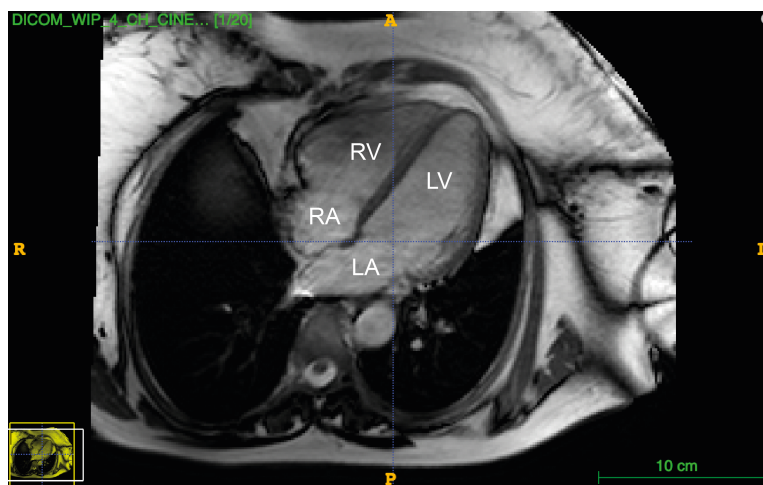


Figure 1.3: Annotated MRI scan of the RV of a healthy subject from the dataset used in this thesis. The slice is taken in the long axis orientation, and represents the RV at end-Diastole. The heart depicted is the same as in Figure 1.2. RV and LV mark the right and left ventricles, and RA and LA mark the right and left atria. The subject is lying supine on the scanning bed, with their head pointing into the page.

2016). This is because substantial through-plane, or horizontal motion occurs during systolic contraction, and it is difficult to distinguish the tricuspid valve and the thin wall of the RV from the right atrium in short-axis orientation.

In order to understand likely adaptive or maladaptive RV shape changes, it is first important to consider the function requirements of the RV. Afterload represents all the factors that contribute to total myocardial wall stress (or tension) during systolic ejection (Norton, 2001). It can be thought of as the amount of force or load against which the heart has to contract. It therefore represents pressure that the heart must generate in order to eject the blood during the contraction of ventricles (LaCombe et al., 2022). When, for example, there is constriction of the pulmonary arteries which leads to high PAP, the afterload increases. There is an inverse relationship between afterload and systolic performance which means that as cardiac output increases, afterload decreases and vice versa (i.e. vasodilation, or widening of the arteries decreases vascular resistance and therefore reduces afterload). Recall that cardiac output is the amount of blood ejected by the heart each minute, which also represents the total blood flow through the pulmonary circuit (Vincent, 2008). The exact equation for afterload depends on the shape of the cardiac chamber (LaCombe et al., 2022), and is therefore complex and subject to noticeable variation between individuals. The shape of the ventricle is significant because the pressure imparted to the blood is not the same as the tangential stress in the RV wall, and the mathematical relationship would be vastly simplified if the RV was of a regular geometric shape (Milnor, 1975).

As the RV is “coupled” to its low-impedance vascular bed, acute changes in afterload lead to major changes in overall RV pressure-volume relationships, and can cause a noticeable decline in ventricular performance; resulting in reduced cardiac output and oxygen delivery as the ability of the RV to pump blood into the pulmonary artery is impaired (Meyer et al., 2019). Conversely, a progressive rise in pulmonary arterial impedance, leads to a more progressive change towards an “LV” pressure-volume loop pattern (Redington et al., 1990). The RV to pulmonary artery coupling describes the relationship between RV contractility and RV

afterload. This may be affected in heart failure patients due to increased LV filling pressures that lead to PH and increased RV afterload.

Although the RV pumps the same stroke volume as the LV, it uses just 25% of the stroke work owing to the lower resistance of the pulmonary vasculature in comparison to that of the central vasculature system in the rest of the body (Voelkel et al., 2006). As a result, the RV wall can be thinner and more compliant. The RV free wall and septum contribute equally to RV performance (Badano et al., 2009). As might be expected from the RV's thinner wall structure, the RV is designed to function by displacing volume rather than generating pressure. Thus, abnormalities that cause pressure overload result in remodelling of the ventricle in response to this load (Vonk-Noordegraaf et al., 2013). As the RV changes shape, the interventricular septum becomes distorted (Méndez et al., 2011). RV stroke volume and wall motion depend on the interaction of the RV free wall, septal wall and outflow tract, all of which markedly differ in geometric and contractile properties (Innelli et al., 2008). The septal wall is shared with the LV, and contributes to the contractile function of both ventricles (Nguyen-Truong et al., 2021). Although evidence states that the septal wall is structurally and mechanically different on its left and right sides (Holland et al., 2010), the overall layered structure on the right side remains thicker and more muscular than the surrounding regions.

The RV typically has a circumferential moderator band (marked on the schematic in Figure 1.1c, however, unlike the LV, it lacks a middle layer and therefore, must rely more heavily on longitudinal shortening than the LV (Sheehan & Redington, 2008). The RV undergoes torsion as it contracts owing to its shape, curved around the LV. The lack of a middle layer in the RV means that there is a lack of continuity of fibres between the inlet and outflow valves. MRI tagging studies have documented the heterogeneity of normal RV motion and confirmed greater longitudinal shortening than that in the short-axis (Childs et al., 2011).

Dissection studies have demonstrated the differences in fibre architecture between the two ventricles (Sanchez-Quintana et al., 1996). Apart from the aforementioned septal wall differences, the muscle fibres in the RV generally form two layers; the outer - or superficial- ,

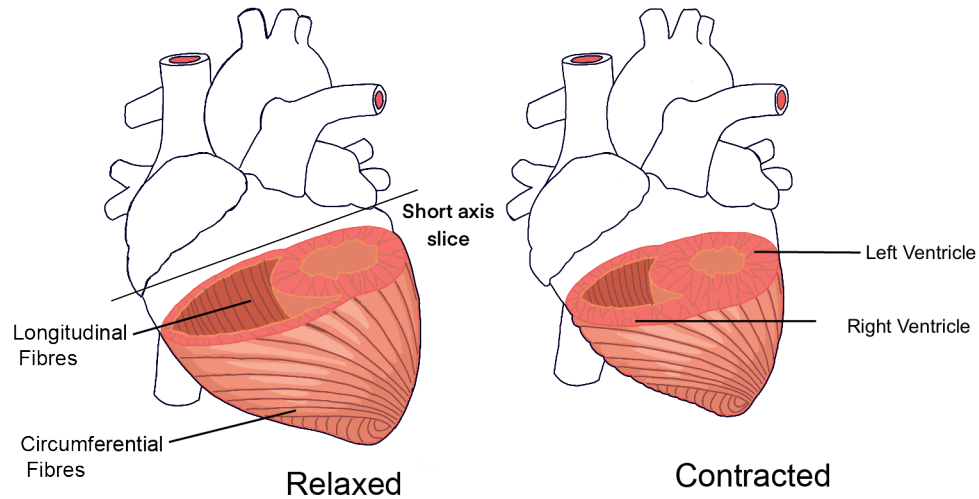


Figure 1.4: Schematic of the RV at end-Diastole, the relaxed state, and end-Systole, the contracted state. This schematic shows the longitudinal orientation of the deeper fibres, and the circumferential layer of the outer, or more superficial fibres.

circumferential layer of fibres and a deeper longitudinal layer of fibres running from the apex to the base (indicated in Figure 1.4). The outer, superficial layer of fibres continue on to the LV after reaching the RV apex. This continuity of fibres contributes to ventricular interdependence (Ho, 2006; Walker & Buttrick, 2013).

There are several other unique RV features which contribute to its contractile function and the ability to demarcate the ventricle on non-invasive images. For example, the tricuspid valve (marked in Figure 1.1a), which separates the RV from the right atrium, has the largest diametered annulus of all the cardiac valves, and is held in place by three papillary muscles. This ‘top heavy’ design feature renders the RV more vulnerable to structural deformation than the other cardiac chambers, especially under sustained pressure or volume load (Walker & Buttrick, 2013). The RV is characterised by coarser trabeculae (a complex mesh of muscular strands that protrude into the centre of both ventricular chambers) in comparison to the LV. The role of the trabeculae in adults is not yet fully understood, but they are hypothesised to influence vortex formation and promote mechanical efficiency (Dawes et al., 2018). These trabeculae often make identifying the RV endocardial border difficult when marking out anatomical regions from clinical images.

The initial (mal)adaptive response to pressure overload, such as that caused by PH, is myocardial hypertrophy, or the thickening of the muscular wall of the RV, and this is followed by gradual but progressive contractile impairment (Dias et al., 2002). Several mechanisms have been proposed to be involved in the progression of this dysfunction, including ischaemia; a decrease of blood supply to tissues, leading to a decrease in oxygen and nutrients to the affected area, changes in the gene expression of sarcomere proteins; the sarcomere is the basic contractile unit of muscle fibre, and is composed amongst others of actin and myosin proteins, which form cross bridges, the breaking of which is the basis of the muscular power stroke, and activation of the myocardial renin-angiotensin system which is involved in the regulation of tissue function.

RV septal curvature is the most commonly studied aspect of RV shape (e.g. Beyar et al., 1993; Sciancalepore et al., 2011). Abnormal septal motion is a commonly attributed sign of raised pulmonary arterial pressure. However, it is not routinely used to quantify the severity of PH (Sciancalepore et al., 2011). In the past, proxy measurements for ventricular shape and function based on both curvature (e.g. Reisner et al., 1994) and sphericity (Tischler & Niggel, 1993) have been considered, although all of these have been based on the LV. Reisner et al. (1994) made the assumption that the wall stress of the septal wall was proportional to that of the right ventricular free wall, thus creating a practical framework.

Having described some standard physiological characteristics, it is worth considering some of the classic responses to various RV diseases, and how these might impact the shape of the ventricle. The RV can respond in diverse ways to a pathological increase in pressure load. The relative success of its compensation depends on the stage of life, as well as the severity of the problem. Problems that occur earlier in life tend to be better tolerated than those which onset in adulthood.

Changes in RV shape associated with sex and obesity are discussed in greater detail in the Introduction to Chapter 3. The effect of oestrogen is purported to be responsible for the higher incidence of PH in this demographic; referred to as the oestrogen paradox, oestrogen

has also been shown to exert a protective effect in acute and hypoxia-induced PH in female patients (Earley & Resta, 2002). Statistical links could be inferred between patients of certain ages or risk factors, and their RV shape.

1.3 Imaging Techniques

Imaging techniques form a common part of all medical assessment, and are often employed early on as screening tests to determine the necessity of undergoing a Gold Standard diagnostic test (e.g. mammography, prior to biopsy in breast cancer (Maxim et al., 2014)). Although imaging is used in all aspects of medicine, the literature review in this section will focus on cardiac-specific imaging. Up until recently, invasive tests have been the Gold Standard and requirement for diagnosis of cardiac problems; specifically PH. Nevertheless, there is significant demand (e.g. Treibel et al., 2022) within routine clinical settings for the development of a Gold Standard for non-invasive diagnostic imaging in PH. Opinion pieces over the past decade have argued persuasively for the introduction of such a standard, yet clinical practice lags behind (e.g. Wood, 2009; Simonneau et al., 2019; Abman et al., 2015; Truong et al., 2020); so much so that the American College of Cardiology Foundation and American Heart Association (McLaughlin et al., 2009) together with the European Paediatric Pulmonary Disease Network (Latus et al., 2016) now recommend both MRI and CT for deriving prognostic information about the RV. However, the recently published United Kingdom standards for non-invasive cardiac imaging (Treibel et al., 2022) note that there have been alterations in mainstream clinical practice and a consequent increase in demand for routine diagnostic imaging in medical practice.

Grünig and Peacock (2015) reviewed non-invasive imaging and concluded that the development of more advanced imaging tools would increase pathophysiological factors which culminate in RV failure in PH, and that integrating such tools into diagnostic algorithms would optimise PH detection and management. There have been substantial advances in the avail-

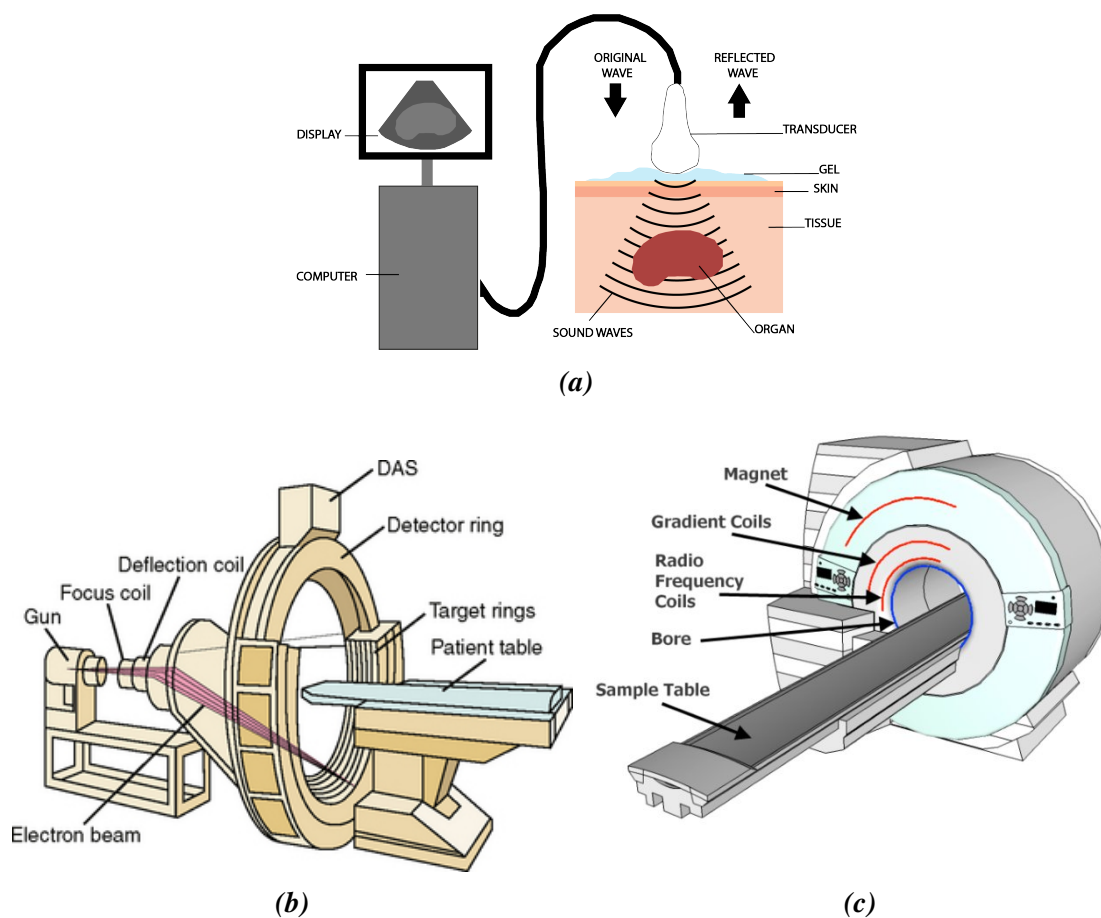


Figure 1.5: Schematics describing three common cardiac imaging techniques: **a)** Echocardiography, **b)** Cardiac CT (taken from Wright and Grey (2016)) and **c)** Cardiac MRI (reproduced with permission from Haynes and Holmes (2013))

ability and quality of imaging techniques, even since 2015.

A recent market simulation study revealed that breathing manoeuvre-enhanced CMR (a form of breath-hold CMR that intersperses periods of hyperventilation with a long breath-hold to increase oxygenation (Guensch et al., 2013)) had the highest shares of preference amongst both patients (59.6%) and physicians (32.7%) (Bertrand et al., 2022). The study concluded that a patient’s preference for a particular cardiovascular test was most determined by the risks and side-effects, and that they prescribed highest value to tests which minimised side effects, while doctors prioritised less costly tests for their patients.

1.3.1 Echocardiography

Echocardiography (Echo) uses ultrasound waves to create real-time 2D or 3D images. During an echocardiogram, a transducer (a handheld device, capable of emitting and receiving ultrasound waves) is placed on the patient's chest (Figure 1.5a). The transducer emits high-frequency sound waves that penetrate the chest and bounce off cardiac structures. The echoes produced by these sounds are detected by the transducer and used to create an image of the heart. Although used in a wide variety of cardiac applications, including the detection of congenital heart disease, heart murmur and abnormal heart valves (Otto, 2013), Echo is of specific benefit to screening for PH, and forms a routine first step towards diagnosis (Galiè, Hoyer, et al., 2009).

Images of cardiac structures are most commonly obtained non-invasively, in the context of transthoracic echo (TTE). However, higher definition images of the heart can be obtained by inserting an echo probe directly into the oesophagus in transoesophageal echocardiography (TOE) (Treibel et al., 2022). Echocardiography has an advantage over other cardiac imaging techniques of possessing both high temporal and spatial resolution, making it most applicable for visualising small and fast-moving structures, such as valves (Treibel et al., 2022). Machines are also highly portable and allowing tests to be performed in a wide variety of healthcare settings, including emergency scanning for rapid diagnosis. As with CMR, it does not use ionising radiation.

Limitations of echocardiography include the potential for low-quality images owing to poor acoustic windows in some patients. In echo scans there can be a greater range of image quality as compared to CT or MRI. Bernardino et al. (2021) have developed a technique for regionally assessing the volume of three relevant RV regions (apical, inlet and outflow) using a parcellation method based on geodesic distances to anatomical landmarks.

1.3.2 Computed Tomography

Computed Tomography (CT scanning) is commonly used in clinical practice. Cardiac CT scanning uses x-rays to image the heart. It works by using a specialised x-ray machine that rotates around the subject's body taking multiple images from different angles (Iranmanesh & Rubin, 2019). The x-ray machine emits a series of x-rays, and detectors within the machine measure the proportion which have passed through the patient's body (Figure 1.5b). These images are then combined by a computer to produce 3D images. Each snapshot is timed with heartbeat to avoid motion blurring from the heart's movement. CT can be performed with contrast-enhancement or in a non-contrast environment. Non-contrast cardiac CT uses only x-rays to create images of the heart while contrast-enhanced CT injects contrast dye to highlight the blood vessels and heart structures.

CT is advantageous due to the high quality images produced in a short study time (15-20 minutes) (Kwong & Yucel, 2003). CT has the disadvantage of exposing the patients to ionising radiation. However, the method can replace MRI for volumetric measurements in those patients for whom MRI scanning is more difficult, for example when the patient has a foreign body or metallic implant, or in cases of severe claustrophobia. CT is commonly used in oncological scanning (e.g. Zhao et al., 2016; van Timmeren et al., 2016). The requirement for contrast dye makes CT scanning unsuitable for individuals with poor kidney function (Kwong & Yucel, 2003).

1.3.3 Magnetic Resonance Imaging

Cardiac Magnetic Resonance (CMR) is the reference imaging modality for assessment of cardiac structure and function and accordingly, its use in clinical practice has become increasingly widespread (Raisi-Estabragh, Izquierdo, et al., 2020). CMR uses a strong magnetic field and radio waves to create detailed images of the heart and blood vessels. The technique works by aligning protons in the water molecules within the body's tissues with a magnetic field, then applying a brief pulse of radio waves to alter their alignment. As the protons return to their

original alignment, they emit signals that are picked up by the MRI machine, and which can be used to create an image of the heart (Hundley et al., 2010). A schematic of an MRI scanner is shown in Figure 1.5c.

MRI is capable of imaging the heart from any angle, does not use radiation and in general offers a more complete, and therefore accurate picture of the anatomical feature of interest (Kwong & Yucel, 2003). Further advantages of MRI include that it is not constrained by body size, which can be advantageous in older patients with limited ranges of motion, those with poor acoustic windows or surgical scars. (Truong et al., 2020). There are several crucial disadvantages of MRI, including the comparatively long study time (30-90 minutes (NHS, 2022b), compared to 10 minutes for CT (NHS, 2021) and 20-40 minutes for an echocardiogram (NHS, 2022a)) and the confined space. Approximately 12.5% of the population is claustrophobic (Vadakkan & Siddiqui, 2022) and the nature of the MRI scanner means that this modality is likely to be unsuitable for such patients. Similarly, a number of conditions may yet preclude subjects from undergoing MRI. Pacemakers and other intra-cardiac devices have varying levels of MRI compatibility and other metal objects, including reconstructive pins in the chest can obscure imaging. More specifically to PH, an MRI-based approach requires a longer scanning time and this can exclude subjects whose cardiopulmonary illness prevents them from long periods of breath-holding. The trade off to reduce scanning time is to take thicker short axis slices. Gilbert et al. (2018) and Hudsmith et al. (2005) both note that substantial slice thickness (5-10mm) leads to partial voluming near the apex and base of the RV, increasing error in boundary placement. There are no specific MRI sequences, preferred for the diagnosis of PH, but several cardiac MRI sequences are commonly used. These include SSFP Cine Imaging for function and volumetric assessments. Other common types include phase contrast velocity encoded imaging, late gadolinium enhancement and T1/T2 mapping. The latter two techniques can be used to assess prognosis (Sharma et al., 2021).

Kocaoglu et al. (2020) argue that retrospectively cardiac gated 2D segmented k-space cine balanced steady state free precession is the preferred CMR sequence for quantitatively assess-

ing cardiac function. The high intrinsic contrast between blood and myocardial muscle, in addition to a high signal to noise ratio makes this scan sequence the preferred option. However, short axis scans using this sequence are routinely acquired using breath-holds because the steady state free precession sequence is susceptible to artefacts from respiratory motion and disruption of magnetisation (Scheffler & Lehnhardt, 2003).

Compared to traditional echocardiography CMR does not suffer from speckle artefacts and produces good contrast between soft tissues (Peng et al., 2016). Images can be obtained in a wide variety of orientations which allows researchers to specify the image planes to optimise their view of required anatomical structures. All structures in the heart can be imaged without having to alter the acoustic window (Gilbert et al., 2018). MRI can offer high temporal and spatial resolution without exposure to ionising radiation (Farrell et al., 2022). MRI temporal resolution is a flexible parameter, which can be adjusted as a trade-off to lower imaging time (Lin & Alessio, 2009). Multiple studies have noted the high reproducibility of MRI-derived ventricular measures including volumes which have led to the modality's acceptance as the reference standard for non-invasive volumetric and functional measures (Geva, 2014; Truong et al., 2020).

A 2017 review by Ptaszyńska-Kopczyńska et al. (2017) found that commonplace non-invasive procedures allowed only approximate estimation of parameters derived from RHC and the current medical consensus is that CMR can be helpful in the diagnosis of PH only when used in conjunction with other techniques, including RHC and echocardiography (Grünig & Peacock, 2015; Galiè et al., 2015). It is, however, accepted that the development of more advanced imaging tools will increase our understanding of the pathophysiological features that contribute to RV failure in PH (Grünig & Peacock, 2015). Developing a reliable, clinically-accepted, non-invasive gold standard for confirmation of PH would be a landmark in automated disease diagnostics, and would pave the way for totally non-invasive diagnostic procedures.

As alluded to in Section 1.2, the RV has a complex crescent shape, which spans both slices

and phases in CMR scans (Luo et al., 2016) and also sees high variability in shape between subjects and between different phases of the cardiac cycle (Moolan-Feroze et al., 2016), which can be heightened in the event of cardiac pathology (Simon et al., 2009). Consequently, segmenting the RV has proved challenging compared to the LV and has seen comparatively little research. Poor contrast on cavity borders further hinders attempts to segment the RV computationally. However, increasing evidence of the importance of RV function to cardiac health (e.g. Caudron et al., 2012) has led to several attempts to automate the segmentation process. The most notable example was the MICCAI 2012 RV Segmentation Challenge (Grosgeorge et al., 2013; Petitjean et al., 2015). Whilst the outcome of the Segmentation Challenge failed to produce any clinically acceptable solutions, substantiating claims that it is difficult to accurately and automatically produce a 3D segmentation model from a series of 2D scan slices, the segmentation challenge provided cardiac MRI datasets which have been instrumental in more recent segmentation studies (e.g. Zuluaga et al., 2020).

As segmentation defines the shape of the initial mesh representing the ventricle, accuracy is paramount. Consequently, interactive drawing of the desired segmentation by trained medical practitioners, rather than the use of automated techniques, has historically been considered the only acceptable method for partitioning regions of anatomical interest (Warfield et al., 2004). However, this manual approach suffers from the typical drawbacks associated with human error, including inter and intra-expert variability as well as a significantly increased time and cost (Bonnemains et al., 2011); it is estimated that it takes an experienced clinician 15 minutes per patient to perform an RV segmentation (Petitjean et al., 2015). Automating the segmentation process would assist with designing a more complete shape analysis framework. Until 2016, no computational attempt had equalled manual performance (Peng et al., 2016). In the past five years, there have been several proposed. Avendi et al. (2017), for example uses convolutional neural networks and stacked autoencoders for automatic detection and initial segmentation of the RV. They trained the algorithm on 16 cardiac MRI datasets from the MICCAI 2012 RV segmentation challenge database.

1.4 Statistical Shape Analysis

Statistical shape analysis, as the the name perhaps implies, is the use of statistical methods to model variations in shape which may not be observable by eye. Shape is defined as the remaining geometric information, once translation, rotation and scaling has been removed (Dryden & Mardia, 2016). Statistical shape methods can quantify shape differences, generate shapes that are representative of cohorts or extract and correlate shape to other relevant measurements. Statistical shape analysis is widely applicable to many disciplines where there may be a quantifiable difference in profile, whether prognostic or demographic (Dryden & Mardia, 1998).

While much of the shape analysis research conducted to date has focussed on the often-imaged brain and heart; Thompson et al. (2004) studied changes of shape the hippocampus in Alzheimer’s patients, and Meda et al. (2012) observed hippocampal changes associated with schizophrenia, techniques have been applied to a vast array of areas, both medical and non-medical fields. For example, Verwaerde et al. (2021) use statistical shape analysis to monitor the shape of pregnant uteri throughout the duration of the pregnancy. Statistical shape models have also been used to diagnose liver disease, (e.g. Kohara et al., 2010) and to create a tuneable model predicting foot shape deviations based on subject-specific shape and demographic parameters (Stanković et al., 2020). Other applications are in archaeology (Woods et al., 2017) and computer-aided design (Kozic et al., 2010; Joskowicz, 2018).

The heart’s capacity for remodelling; both adaptive and maladaptive, makes it a compelling candidate for shape analysis study. There has been investigation into adverse cardiac remodelling in response to Repaired Tetralogy of Fallot (e.g. Mansi et al., 2011; Sheehan & Redington, 2008; McLeod et al., 2013) and this has provided opportunity for the implementation of statistical shape analysis.

Statistical shape analysis frameworks for classification purposes typically involve building a correspondence between the shapes in the given set, statistically decomposing a group of models into fundamental shape components, and then building features from the shape com-

ponents that can cluster the shape set into various groupings (da Fontoura Costa & Cesar Jr., 2010). The statistical shape analysis method may also build a classifier associated with the application of interest. A wide variety of shape have been deployed within statistical shape analysis frameworks, tailored to the specific features, restrictions, and objectives necessitated by specific applications. Furthermore, the statistical shape analysis work to-date has primarily focused on developing the mathematical representation of the shapes in a given set along with the preprocessing methods necessary to build a correspondence between these shapes, including aspects of topological mapping, shape alignment/registration, and parameterisation, while much less consideration has been given to the method of statistically decomposing the shape set once this correspondence is set.

Historically, statistical studies of morphology were based on pre-defined measurements of size, area and volume (Thompson, 1917). These studies can indicate anomalies between individuals, however they do not fully capture the complexity of a shape and can not necessarily distinguish normal anatomical variation from pathological alterations. Thus, with the improvement of medical imaging techniques based on both ultrasound and MR, statistical methods have come to the fore. Whereas previously pre-defined morphological features were examined, it is now possible to use statistical methods to tell us what the features of interest are from the data. Statistical Shape Analysis has gained traction as a method for computationally aided disease diagnosis through its potential to precisely locate the morphological changes between healthy and pathological structures (Styner et al., 2006). Shape analysis at its core can be thought of in the pattern recognition context, and the process can be divided into two phases: first, to identify and extract features from measured data, in this case medical images and secondly, to measure the correlation between these features and variables of interest. This second step could also be extended to create a classification tool capable of predicting variables of interest, such as blood pressure, for diagnostic purposes from those features. The fundamental features could also assist with improving future image processing endeavours (Heimann & Meinzer, 2009). Shape Analysis generally operates by obtaining a

representation of a biological structure's shape from medical image data, then performing statistical analysis on such images to study the geometrical variations and correlate these with symptoms.

Before shape features can be derived using statistical shape analysis, an applicable and robust method must be selected. This is crucial since differences in analysis methodology are a key cause of variation in obtained results. Several shape analysis frameworks have been proposed, both discrete and continuous (Wu et al., 2014) and such procedures can be categorised based on a series of criteria, including whether the method is applied globally or locally. A global method searches for features that describe the entire shape of the object, whereas local methods attempt to extract a feature from some part of the total shape. Global methods are the most suitable for analysing RV shape, as examining complete 3D representations best describes ventricular shape and hence provides better data to approximate pathological shape changes. It is possible to subcategorise shape analysis methods based on whether the method uses pre-defined features or features derived from the analysis. Global examples of pre-defined features include volume and size, whereas derived features could be Proper Orthogonal Decomposition (POD) modes. Many methods create an atlas which could later be used for semi-automated model generation. Three main approaches for capturing shape information are: using skeletonisation and the medial axis; spherical harmonics, and the direct mapping POD method. Direct mapping POD was first developed in earlier work by this research group (Wu et al., 2012, 2013).

The first step of a statistical shape analysis approach is the extraction of shape features from a set of input images. Several shape descriptors have been proposed for application in medical image analysis, and the precise choice of representation depends on the application (Golland et al., 2005). Popular approaches are discussed in Zhang and Golland (2016) and include using anatomical landmarks to represent the geometrical features of a region (Stegmann & Gomez, 2002). Landmarks capture the shape of the object's outside contour or boundary, however can be limited by a lack of available features. Frangi et al. (2002) used a landmark-based

representation of both the LV and RV at end-diastole and was the first to introduce the concept of statistical shape analysis for cardiac structures. A landmark approach to representing 3D surfaces always introduces loss of detail (Zhang et al., 2010). If such errors can be minimised or the landmarking loss is shown not to impact important shape regions, this approach can substantially decrease shape selection time.

Skeletonisation to create 2D and 3D medial shape descriptions, (e.g. Jang et al., 2019; Golland et al., 1999) has also been used to analyse shape. In these methods, the medial axis of a model is created from the set of all points which have more than one closest point to the object's boundary. A medial description can decouple extraneous features such as thickness and position from shape. However, medial shape analysis struggles to reliably compute a stable model in the presence of variability in biological shape (Styner et al., 2003) and a medial axis often fails to convey the complexity of biological structures. This method has fallen out of favour in recent years owing to the availability of more powerful modelling computers.

SPHARM, a series of packages utilising spherical harmonics initially showed some promise and received much early attention (Gerig et al., 2001; Shen et al., 2003). SPHARM is approximated by breaking the object into a string of linear harmonics. All of these linear harmonics store degrees of information; the lower order harmonics store information about the basic shape, whereas the higher order harmonics store finer details of the shape. Nitzken (2015) notes that using a traditional SPHARM expansion, a standard least-squares fitting model does not accurately approximate the 3D shape of the brain cortex, which may result in some loss of shape. The use of an iterative fitting model on brain cortical surfaces (Chung et al., 2008) allowed for a much more accurate shape approximation.

An alternative approach to modelling shape boundaries is to consider the space in which the shape sits. These models, where the space surrounding the shape is deformed to match the target image are considered 'atlases' (Davies, 2002). A simple shape atlas shows the average structural information as a *mean shape*, and considers variations around that mean (Bookstein, 1989).

Young and Frangi (2009) take the definition of an atlas to be “an alignment of data maps from different domains, either population or individualized, which enables querying of relations ... to construct the big picture’. An atlas, therefore, can be taken more broadly than in Dawes’ definition, as a statistical method of integrating information from different sources.

Recent attention is focused on using point clouds and the creation of continuous meshes to obtain shape descriptions. Wu et al. (2013), for example, used mesh-based continuous descriptions of the right ventricular endocardial surface (RVES) throughout 9 phases of the cardiac cycle and then applied a harmonic mapping method to convert the surfaces into relative shape change functions to build correspondences. The success of any given shape analysis method largely depends on the number of degrees of freedom used, as the intent of process is to discretise the shape and consequently it is desirable to retain as much information as possible. The advantage of such continuous, mesh-based techniques is that in excess of 3000 individual points are mapped, giving a much more nuanced description of the shape. On the other hand, Styner et al. (2003) used just 60 spherical harmonics and measurements of volume and consider only one degree of freedom. Wu et al. (2013) used an approach for quantitative analysis and statistical decomposition that was both straightforward to implement and well suited to medical image analysis. This was particularly true of images where at least one clearly identifiable anatomical reference line could be determined.

The basic idea of harmonic shape image mapping is to map a 3D surface patch to a 3D domain and encode the shape information of this 3D shape onto the 2D image. This simplifies the surface matching problem to a 2D image matching problem. Harmonic shape images are beneficial because they preserve the shape and continuity of the underlying surfaces (Zhang & Hebert, 1999). Shape analysis, particularly of the RV, has yet to transition into standard medical practice, and is only now being considered as part of larger-scale clinical research (Zhang & Golland, 2016). This is in large part a result of the complex shape of the RV and the observed variations both between and within subjects. The recent attention to developing statistical shape analysis methods both in the heart and in other organs is very promising

for the classification of pathologies and for the development of non-invasive diagnostic tools. However, these research papers have approached the analysis in a variety of ways, using different mathematical methods at all stages of the analysis process, from image registration to statistical decomposition techniques without much justification for the selection of their methods, or for why one method supersedes another.

Open-source SSM tools rely on different modelling approaches and make different assumptions to establish surface correspondence. It is therefore difficult to evaluate the success of shape models. Davies (2002) suggests that shape models can be intrinsically evaluated using quantitative metrics representing correspondence quality. However, this method has been criticised since relevant shape information can be lost whilst retaining strong evaluation measures (Ericsson & Karlsson, 2007).

1.5 Scope of the Thesis

The scope of this thesis is to improve a clinically applicable statistical shape analysis methodology, first developed in Wu et al. (2013), and extended to MRI in Xu et al. (2021), which would enable rapid, accurate and non-invasive assessment of the degree of PH pathology based on CMR scanned images. The methodology applied and extended a previously developed statistical shape analysis method to a new dataset, containing 96 subjects from two hospitals; University of Pittsburgh Medical Center and Newcastle University.

The particular focus of this work is to determine whether there are specific RV shape features that differ based on the biological, or population-based characteristics; sex, age and obesity status, independent of disease-induced changes. This thesis will also assess the robustness of the statistical shape analysis pipeline to four common sources of input discrepancy; variations to the dataset size, alteration of the anatomical reference points used to orient the ventricles before the initial parameterisation, assessment of inter and intra-observer variability and variation of the degree of recursive gaussian smoothing filter, and determine a new,

diagnostic procedure for PH, accounting for any population-based features which improve identification of PH sufferers, based on RV shape features. The thesis will make recommendations for optimal categorisation of PH, based on the results obtained in Chapters 3 and 4.

The scope of the thesis does not extend to the proposition of a definitive diagnostic procedure for PH. Although recommendations will be made for improving diagnostic success in a clinical environment, the sample size is not yet sufficient to apply this method directly to a hospital setting. With the exception of quality of life improvements, adaptations of the code to combine analyses of new and existing segmentations and several minor fixes, no developments to the fundamental mathematics or implementation of the statistical shape method have been made. The scope of this thesis will not include any development of non-linear, machine learning-based models to assess diagnosis of PH. The available haemodynamic data differed between the two centres and therefore direct comparisons of certain metrics; specifically traditional measures of the state of PH in the Newcastle dataset such as PAP and PCWP were impossible between the cohorts. The RV models in this thesis were segmented in the short-axis only, as long-axis images were unavailable for the 50 RVs derived from Pittsburgh.

1.5.1 Structure of the Thesis

This thesis is comprised of six chapters.

- * **Chapter 1** introduces the background to the problem, discusses the thesis scope and outlines the novel contribution of this PhD.
- * **Chapter 2** sets out the methodology adopted to produce the results presented in this thesis. It outlines initial data preparation and describes the statistical shape analysis and visualisation methods utilised in all data chapters. Methods specific to a given chapter are described within that chapter.
- * **Chapter 3** applies shape analysis techniques to a dataset of 96 subjects to determine whether non-pathological shape features cause measurable differences in right ventric-

ular shape between dichotomous groups. This chapter also characterises the differences between groups and plots the mean RVs for each category.

- * **Chapter 4** determines the robustness of the method when the input data is altered, including by increasing and decreasing the size of the input dataset, changing the placement of the anatomical poles and dateline which form the boundary conditions, altering the smoothing parameters and assesses inter and observer variability.
- * **Chapter 5** combines the results from the population-based analysis in Chapter 3 and the robustness analysis in Chapter 4 to establish a diagnostic metric for Pulmonary Hypertension, which takes into account inherent population-based differences. This chapter also performs an SVM analysis to determine the efficacy of a non-linear model.
- * **Chapter 6** concludes the thesis and presents recommendations for the application of statistical shape analysis in a clinical setting.

1.6 Novel Contribution

In this section, novel contributions to the literature derived from the thesis are individually summarised. Overall, the significant contribution of this work was extension of a proof of concept shape analysis methodology from Wu et al. (2012) and Xu et al. (2021) to determine that shape characteristics unique to the natural features; sex, age and BMI exist. The approach focuses on non-hypertensive individuals, which enabled these unique signatures to be identified, free from any pathological interactions. The initial analysis incorporated 46 never before segmented RVs, along with 50 RVs from Xu et al. (2021) to create a new twin-centre dataset; some healthy, some with generalised hypertension and some with PH. The statistical shape analysis method was found to be sensitive to methodological attributes; the most significant shape feature (Mode 1) correlated strongly with the MRI scanning protocol used, and no other feature. Therefore, there is potential to calibrate the shape analysis procedure to remove arte-

facts of the technique. Methods which do not change the underlying finite element mesh, including altering the size of the dataset and the location of the anatomical poles and dateline, demonstrate the most significant robustness to changes in input. The shape analysis method was able to separate individuals with PH from those without (including 33 individuals with generalised hypertension). Accounting for BMI in the model increased the overall classification percentage of PH from non-PH, however this result often sacrificed correct diagnosis of PH, and the percentage gains were achieved from improved identifications of healthy ageing subjects.

1.6.1 Evaluates the Impact of Non-Pathological Characteristics on RV Shape

Chapter 3 evaluates the effects of three biological characteristics on RV shape using an anatomically consistent statistical shape model (first developed in Wu et al. (2012) and improved in Xu et al. (2021)). The analysis outputs a series of ranked features (modes), and a linear combination of these features describes each individual RV, with the amount of each shape mode exhibited by the modal coefficient.

This chapter examines correlations of mode coefficients to biological features and also assesses the possibility of separating non-hypertensive RVs into dichotomous groups using a linear discriminant function; thereby showing, through their mathematical separation, that there is a difference in overall shape between the groups.

The most important finding is that several non-pathological features affect the shape of the RV, independent of hypertension. Notably, the MRI scanning protocol used; specifically the choice of slice thickness, used to create the initial diagnostic images accounted for a significant shape component. The MRI scanning protocols used are described in Section 2.2. The impact of scanner was concentrated into a single shape feature, which means that this impact can be easily isolated and excluded, leaving the biological or population-based shape features. One benefit related to the potential of extraction of method-based features in this manner is

that several datasets with disparate origins can be combined and analysed together, and that there is potential to identify and ultimately calibrate for other shape features introduced by the procedure.

This chapter identifies that all three biological feature categories tested; male and female individuals, younger and older individuals (split by the mean age of 60.5 years) and obese and non-obese individuals (separated by a BMI of $28\text{kg}/\text{m}^3$), demonstrated quantifiable difference in RV shape. BMI was not subject to age-based correction, as this information was available only for the Newcastle dataset. The categories were assessed on non-hypertensive subjects, to ensure that the unique features that were derived exist independently of hypertension. Unique shape features were identified at both cardiac phases; end-Diastole and end-Systole, as in all cases, the linear discriminant model was capable of separating the dichotomous feature groups with greater than 72.5% accuracy. Sex, which was best categorised by coefficient values for Mode 19 in end-Systolic RVs, and obesity, best described by Modes 20, 11 and 2 at end-Diastole resulted in the most significant impact on RV shape. In each case, the best combination of shape features could correctly separate RVs 77.5% of the time. The shape features most prominent in the obese RV showed rounding and hypertrophy. Non-obese RVs did not show the same hypertrophy, and exhibited enhanced RV curvature. Male RVs at end-Systole displayed basal flattening, which is similar to that exhibited in hypertensive hearts.

1.6.2 Determines Procedural Constraints Required to Achieve Consistent Shape Features

This thesis assesses robustness of the statistical shape analysis procedure (outlined in Chapter 2). Four targeted perturbations were made to the input; altering the dataset size, altering the placement of user-identified anatomical location points, including MRI segmentations by a new observer to compare inter- and intra-observer variability and finally, altering the stringency of the recursive gaussian smoothing filter. Robustness is examined by assessing how sensitive the extracted shape features (modes) are to external modification, and also by

evaluating how sensitive the mode coefficients (or how much of each individual shape feature is required to reconstruct any given RV).

All four tested interventions derive the same overall mode shapes from the input dataset, suggesting that the procedure produces an accurate deconstruction. However, the mode coefficients are more sensitive to alterations in input value. Changing the dataset size, was the least sensitive to changes in modal coefficient, followed by changing the image segmentation. While still robust over the first 13 modes, altering the anatomical reference points for the harmonic mapping (poles and dateline) was more sensitive. Adjusting the smoothing filter caused substantial differences in recorded modal coefficient, demonstrating only two statistically significant correlations.

Overall, the statistical shape analysis method is largely insensitive to changes which do not alter the underlying RV models. This includes changing the dataset size, and selecting new anatomical data points. Conversely, the greatest differences occurred in situations where shape of the models were altered, for example when altering the smoothing filter, or from scanners using different scanning protocols; particularly those which alter MRI slice thickness and/or continuity.

Several practical recommendations can be made based on the robustness analysis. Scans from varying origins can be combined without issue; the specific shape features are contained within a single mode and can be isolated, as is demonstrated in Chapter 3. Additionally, dataset size can be altered without substantially changing the modal coefficients, the remaining methodological parameters should remain untouched to ensure the consistency of derived shape features.

1.6.3 Proposes and Evaluates Approaches to Use the Shape Features to Predict PH

This thesis explored classification strategies with the aim of providing guidance on how to use RV shapes in conjunction with biological characteristics to predict the state of PH. Clinically

accepted shape-based descriptors of PH were identified when the statistical shape analysis was applied to all 96 subjects. It was possible to separate individuals with PH from the remainder of the cohort using a combination of four shape features for end-Diastolic RVs and five shape features for end-Systolic RVs, thus creating a baseline descriptor. When reconstructions of these shape features are plotted they qualitatively correspond to medical descriptions of the effect of PH on the RV described in the literature, including a rounding of the ventricle. This remains the case despite the inclusion of 46 new subjects of varying ventricular health, combining two different datasets with varied scanning parameters.

One specific mode in this dataset; Mode 15, consistently correlated with the presence of PH, and is also associated with BMI. Accounting for BMI in the model increased the overall classification percentage of PH from non-PH, however this result often sacrificed correct diagnosis of PH, and the percentage gains were achieved from improved identifications of healthy ageing subjects. Age and sex had a substantially smaller impact on predictive models for PH.

Chapter 5 also briefly explores the possibility of using a non-linear method to predict PH using RV shape features. Given the complexity of heart shape it was anticipated that a more complex classification boundary might improve diagnosis of PH. Applying non-linear, machine learning models (Support Vector Machines) do not improve the classification potential of the model, and also are more computationally expensive, leading to the conclusion that linear models are sufficient to predict PH in this dataset.

Chapter 2

Methodology

The shape analysis framework used in this thesis was developed by Wu et al. (2012) but has been altered and extended, first by Xu et al. (2021) and then as part of this work. The flow chart in Figure 2.1 shows the overall workflow. Each step of the process, from image segmentation, registration and meshing to the statistical decomposition and production of results will be discussed in depth in Sections 2.3-2.9. Given a dataset of medical images, first, a three-dimensional closed-surface representation of the shape of interest (the human right ventricle) is segmented or extracted from an MRI image stack. Subsequently, the set of shapes is aligned using an image registration algorithm, and correspondence is established between hearts by topologically mapping each shape to a common reference domain. Lastly, the set of mapped shapes is statistically decomposed to identify a set of common shape features and the corresponding significance of their representation in each shape in the dataset.

2.1 Demographic Information

A total 121 subjects were considered in this thesis; 50 subjects from the University of Pittsburgh and 71 subjects from Newcastle University. Of those, 96 were used in the initial statistical shape analysis (Chapter 4), and in the appraisal of a diagnostic metric for PH (Chapter 5). The remaining 25 subjects were considered for robustness tests (Chapter 4). Table 2.1

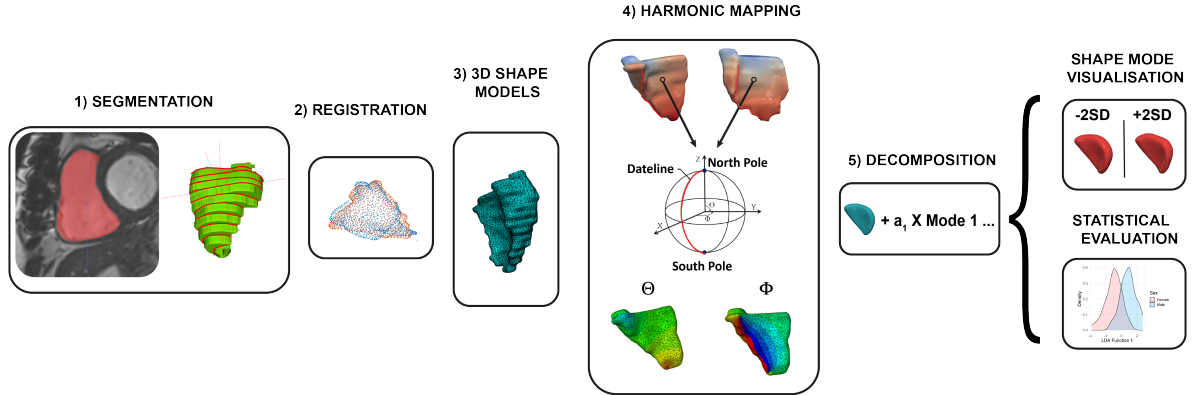


Figure 2.1: Flowchart depicting the overall shape analysis workflow which was first established in Xu et al. (2021) and has been improved upon as a part of this thesis.

Table 2.1: Demographic breakdown of the subjects considered as part of this thesis.

	Original 96		121 Total	
Sex	Female	Male	Female	Male
	44	52	54	67
BMI	<28	>28	<28	>28
	47	48	68	52
Age	<60.5	>60.5	<60.5	>60.5
	41	55	46	75
Centre	Newcastle	Pittsburgh	Newcastle	Pittsburgh
	46	50	71	50
Diagnosis	PH	Non-PH	PH	Non-PH
	33	63	33	88

provides a breakdown of the subjects on the basis of demographics. More detailed subject information is provided in Sections 3.3, 4.3 and 5.3, the relevant Chapter Methodologies).

2.2 Data Acquisition

Double-blind anonymised haemodynamic patient data was obtained from two institutions; Newcastle University and the University of Pittsburgh Medical Center (UPMC). Double-blind refers to the divorce of meta-data from the images. The UPMC dataset is composed of clinically obtained cardiovascular magnetic resonance (CMR) images from patients who under-

Table 2.2: A comparison of the scanning protocols employed at the Newcastle and Pittsburgh MRI centres.

Feature	UPMC	Newcastle
Coil Channels	32	6
Slice Thickness	6mm	8mm
Slice Skip	4mm	0mm
Phases per R-R	25	30
Flip Angle	51°	40°
Matrix	256×144	256×256
FOV	270mm	350mm
TE	1.11 ms	3.7/1.9ms

went both CMR and right heart catheterisation. Images were obtained using a 1.5T Siemens Magnetom Espree machine equipped with a 32-channel cardiac coil. Standard breath-held cine imaging was acquired with steady-state free precession in the short-axis orientation, spanning base to apex. Image stacks were taken in 6mm thick slices, skipping 4mm between each slice. Typical imaging parameters included 30 phases per R-R interval, matrix 256 by ~ 144 , flip angle 51°, TE= 1.11 ms, acceleration factor 3. MRI images were extracted at two distinct cardiac phases for each patient; end-Diastole and end-Systole. Figure 2.2 shows examples of an end-Systolic RV and an end-Diastolic RV in the short axis orientation. At end-Systole, the RV is at its most contracted, and the muscular ventricle wall is thicker, whereas at end-Diastole, the RV is relaxed, the cavity volume is greater and the muscular wall is thinner.

The Newcastle University dataset formed part of an ageing study. Images were acquired using a 3T Philips Intera Achieva scanner. A dedicated 6-channel cardiac coil (Phillips) was used to obtain a stack of balanced, breath-held, steady-state free precession images in the short axis view. 14 slices were taken over 25 phases with a slice thickness of 8mm and a 0mm inter-slice gap. FOV = 350mm, TE= 3.7/1.9ms, turbo factor 17, flip angle 40°, resolution 1.37, temporal duration ~ 40 ms per phase, dependent on heart rate.

Steady-state free-precession refers to an MRI pulse sequence that is commonly used to

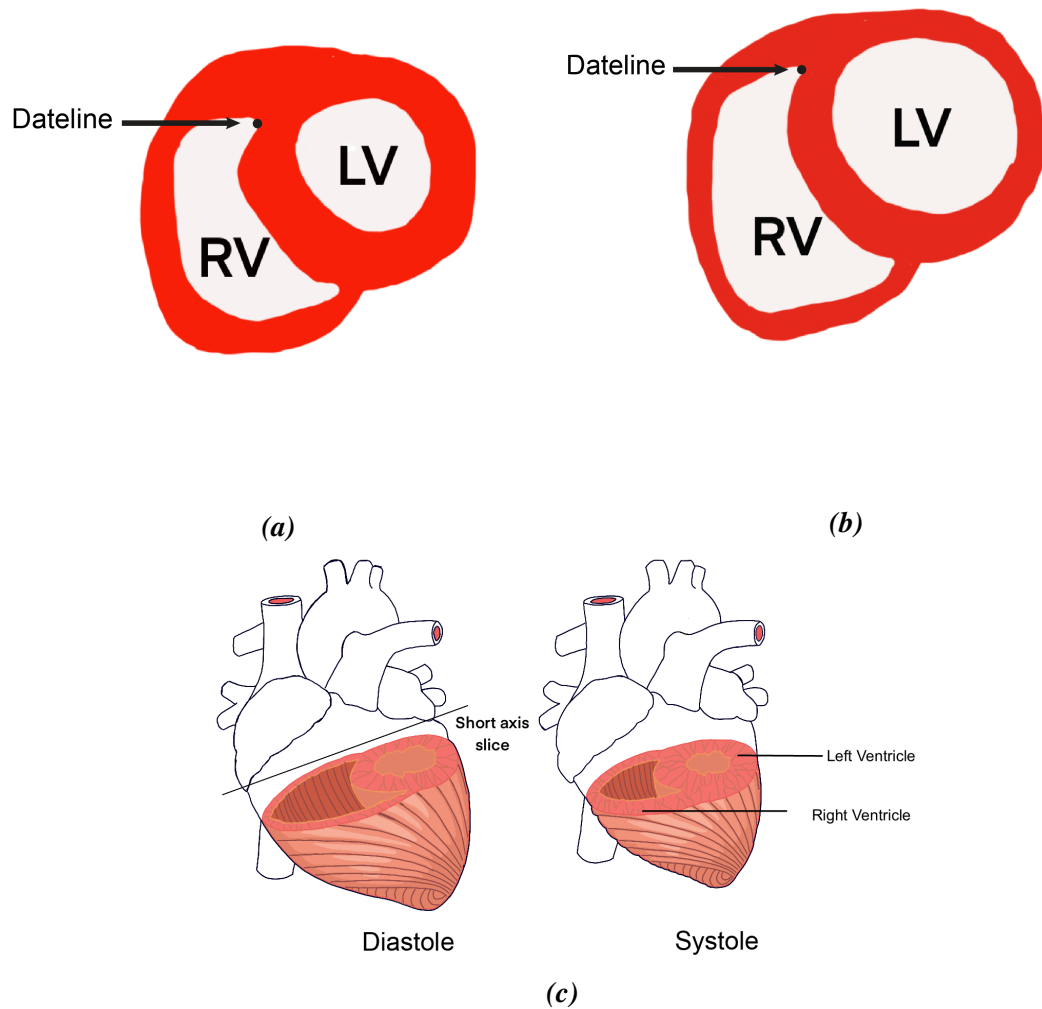


Figure 2.2: Top-down, cross-section illustration of the shape of the ventricles in a healthy patient which shows the contractile pattern at **a)** end-Systole and **b)** end-Diastole. The location of the dateline on this slice is marked by an arrow. This slice is located at the widest point. **c)** Provides an alternative view, which shows the location of these cross-sections in the heart.

image tissues. ‘Steady-state’ refers to the fact that the magnetic signal produced by the tissues remains constant over time, and free-precession to the fact that the signal precesses ‘freely’ (like a spinning top) in response to the magnetic field applied by the scanner. There are a series of alternative pulse sequences available, and the choice of pulse sequence is related to the specific imaging goals, the type of tissue being imaged and the presence of certain medical conditions.

The matrix size is the number of frequency encoding steps (in one direction) and the number of phase encoding steps in the other direction of the image plane. When all other factors remain constant, increasing the number of frequency encodings or the number of phase steps improves the image resolution. The frequency encoding depends on how rapidly the scanner samples the free-induction-decay signal. Increasing the sampling rate does not cause a time penalty, but increasing the number of phase steps results in a proportional increase in image acquisition time. Hence, in breath-held MRIs for symptomatic patients, the number of phase steps is reduced to ensure patient comfort. The magnetic resonance echo containing the frequency- and phase-encoded spatial information which is necessary to construct a cardiac image is decomposed using a Fourier transform.

K-space is a mathematical representation of the raw data collected during an MRI scan. It is a 2D or 3D grid in which each point represents a unique combination of mathematical gradients applied during a scan. During MRI, a strong magnetic field is applied to the patient’s body which causes protons in the tissues to align with the magnetic field. Subsequently, a sequence of radio-frequency pulses is used to manipulate the alignment of those protons, which generates a complex signal containing information about the tissues being imaged.

The information encoded within this complex signal is then encoded into k-space using magnetic field gradients. Each point in k-space represents a different combination of the magnetic field gradients applied during the scan. The amplitude and phase of the signal at each k-space point represents the amount of signal coming from different spatial locations. After the data is collected in k-space, it is transformed into an image using a Fourier transform. The

complex signal at each point in k-space is therefore transformed into a corresponding pixel in the final image. The shape and size of k-space determines the resolution and contrast of the final image. The Field of View (FOV) is the size of the area covered by the matrix. Dividing the FOV by the matrix size defines the in-plane voxel size: decreasing the FOV improves image resolution, whereas increasing the FOV in either direction increases the size of the voxels and decreases image resolution.

Slice thickness determines voxel depth. The slice thickness is almost always the largest dimension of the voxel in 2D imaging. The resolution perpendicular to the image plane is the poorest. This factor relates to the maximum strength of the z-gradient coils as well as time constraints, and limits the number of slices available. 3D imaging using z-direction phase encoding is capable of smaller slice thickness than 3D imaging but again carries a time penalty proportional to the number of slices taken.

2.3 Segmentation

Segmentation is the process of creating a mask to partition particular regions of an image. RV surfaces were obtained by manually segmenting short-axis images of the RVES. Boundaries were decided in consultation with a combination of short-axis and long-axis image stacks in the case of the UPMC images, however long-axis images were not available for the Newcastle University image data.

Surfaces were segmented by hand to ensure that the most anatomically accurate representations of RVES shape were segmented within the constraints of the medical image quality.

Image segmentation was performed in ITK-SNAP (Yushkevich et al., 2016), using the polygon tool to accurately select points along the endocardial surface, tracing the shape of the RV. This procedure is depicted in Figure 2.3 and an example completed trace of a single slice is depicted in Figure 2.4a. ITK-SNAP outputs a metaimage (.mha) file which retains the selected region of interest in addition to a fixed slice-depth related to the MRI slice thickness.

Figure 2.4b demonstrates how a 2D selection on each slice produces the overall 3D model, by incorporating slice thickness. The 2D selection made in ITK-SNAP is shown in red, and the green segments illustrate how the 2D slice is extended downwards to produce a basic 3D representation of the RV. Newcastle segmentations were undertaken by the author, whereas Pittsburgh segmentations were part of a pre-existing dataset and were performed by summer student Erin Sarosi. All segmentation was overseen by the same trained cardiologist to control for quality and the segmentation process was blind to population and haemodynamic data.

The endocardial surface was chosen over the epicardial surface for segmentation owing to the relatively higher contrast of the blood-tissue boundary layer, which makes the inner wall easier to identify on MRI images, and therefore leads to a more accurate surface extraction. This was particularly important at end-Diastole when the ventricle was fully relaxed and the features of interest were most difficult to distinguish. Manual segmentation was used to give the most accurate surface representations of ventricular shape, within the constraints of MRI image quality. Other similar shape analysis studies use semi-automatic segmentation (Mauger et al., 2019; Leary et al., 2012) to save processing time, however, at the time of writing the semi-automatic methods available required such a significant amount of manual intervention to accurately segment the human RV, that an entirely manual approach appeared superior. Despite careful attention to accuracy during the segmentation process, the reliance on short-axis images, especially in the University of Pittsburgh dataset which was generated using a 4mm skip MRI scanning protocol, leaves a small possibility that the scan was curtailed before the apical slice resulting in the tip of the apex being missed. The Newcastle University MRI scans were filtered to ensure total RV coverage, so this issue was less prevalent.

2.4 Mesh Generation

Closed 3D surfaces of each heart surface were obtained by creating a mask to isolate the segmented surface, resampling the slices to even the pixel spacing to 1x1x1mm using Nearest

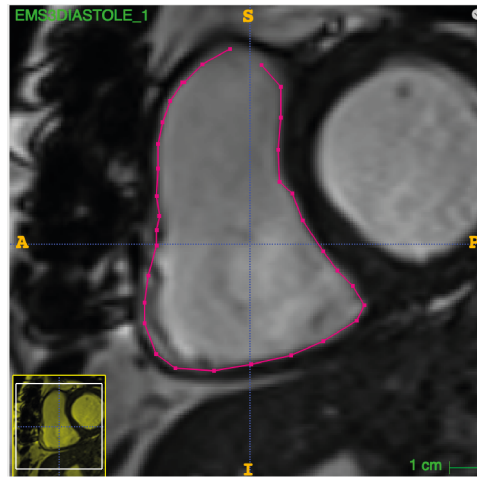


Figure 2.3: Screenshot of segmentation slice at end-Diastole for a subject from the University of Pittsburgh dataset. Selections are made in the short-axis orientation using the polygon tool in ITK-SNAP.

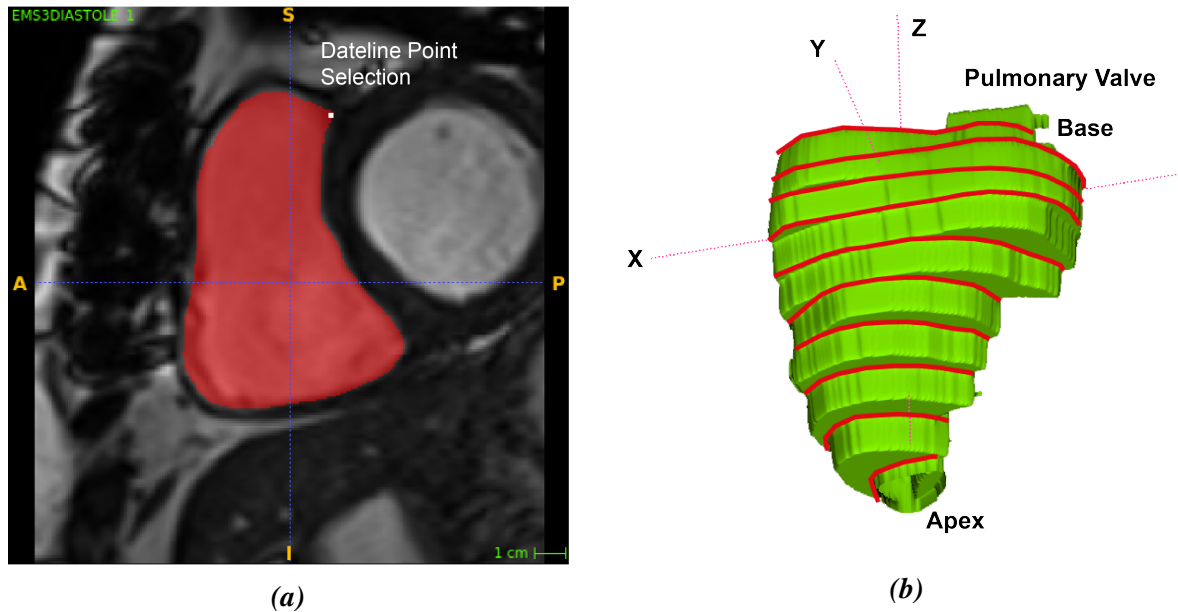


Figure 2.4: Description of the segmentation process. **a)** Screenshot of segmentation slice at end-Diastole. Selections are made in the short-axis orientation using the polygon tool in ITK-SNAP. The point used for dateline selection is marked in white. **b)** A 3D graphic of the selections made for each slice, taken from ITK-SNAP but coloured green to differentiate the 2D slice selections from the 3D volume. The RV is oriented with the septal wall facing into the page to emphasise how the RV slices contours selected as part of the relate to the 3D representation. Depth is defined by the MRI slice thickness. The imaging planes are labelled x , y and z .

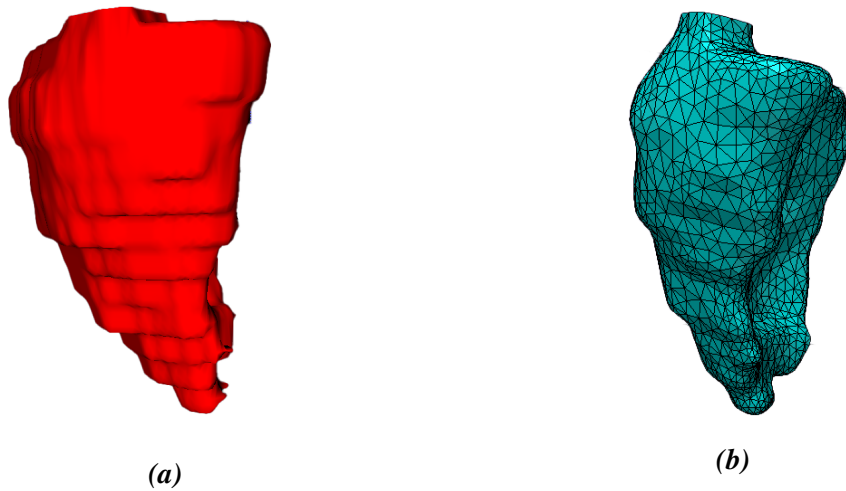


Figure 2.5: *a)* an example input RV at end-Diastole, from a non-Hypertensive RV from the Newcastle Dataset. This segmentation shape is exported from ITK-SNAP, where the initial segmentation as a .mha file, which is then imported into the commercial software package Simpleware. This figure is a direct screenshot from ITK-SNAP, and is equivalent to the graphic shown in Figure 2.4b). *b)* output of the same RV from Simpleware. At this stage, the RV has been smoothed using a discrete Gaussian filter and converted to a triangular FE mesh in .inp format, ready for import into ABAQUS FE Software.

Neighbour interpolation and smoothing the interpolated surfaces using a standard recursive Gaussian filter within the commercial imaging software package Simpleware ScanIP (Synopsys Ltd., Exeter, UK). The nearest neighbour interpolation algorithm is a relatively simple method of resampling an image by assigning the value of the nearest pixel in the original image to the corresponding pixel in the new image.

A $\sigma 3$ smoothing filter was selected to strike the balance between introducing unwanted artefacts of the 3D model and removing too much material, leading to an inaccurate model. Internal cavity volume measurements were also taken from Simpleware at this stage. The 3D finite element models of each ventricular surface are saved as Abaqus input (.inp) files.

2.5 Anatomical Boundary Selection

Four reference points on the RV mesh were chosen as boundary conditions; the north and south poles, the dateline and the west elements, with the aim of making the analysis anatomically consistent across the dataset.

During the harmonic mapping step (described in Section 2.7), each mesh is mapped to the unit sphere such that the location of the pulmonary valve (the valve that separates the top of the RV from the pulmonary artery) aligns to the north pole of the sphere and the apex (located at the bottom of the chamber and which describes the region where all the walls converge) is positioned at the south pole of the sphere. The north and south poles form the endpoints of the third selected region; the dateline. As shown by the red line on the example mesh in Figure 2.6, the dateline traces the anterior border between the free wall and the septum of the RV. The dateline is the split-line during the calculation of spherical coordinates and represents the boundary between 0 and 2π radians. Finally, the west elements are selected (those on the left hand side of the border), which denote the direction of the split.

The location of the poles and datelines were selected manually using ADAMCUS (Leach, 2019) to ensure that the most anatomically accurate representation was obtained. Each region was selected with reference to the MRI stacks in ITK-SNAP, and the location of the anterior border was identified in each slice, however exact placement was limited by the location of the nearest node on the mesh. The apex is defined with respect to the left ventricle, and, as the RV sits at an oblique angle relative to the LV, it should be noted that the most basal part of the ventricle is not always at the very tip (which can be observed in Figure 2.6).

ADAMCUS is custom-made software written in the Godot game engine (version 3.2.1) which was developed in consultation with the author. The author has also made minor quality of life improvements to the original code, such as the addition of a reset button.

ADAMCUS imports the Abaqus .inp files which were saved in Section 2.4. These are data files in ASCII format which define the mesh geometry by describing elements and nodes. ADAMCUS provides a user interface which enables the user to select and save the poles, dateline and west elements, ready for downstream analysis.

ADAMCUS saves separate text (.txt) files containing the nodes and elements which make up the 3D surface, as well as the selected poles, dateline and west-elements. ADAMCUS determines the dateline by calculating the distance between the selected north and south pole,

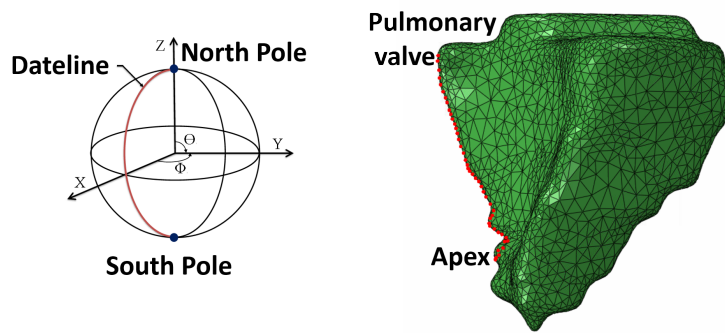


Figure 2.6: A graphical representation of the unit sphere is displayed on the left in addition to an example FE mesh which represents the input into the harmonic mapping step. The sphere is annotated to indicate the localisation of the north pole, which corresponds to the pulmonary valve during the harmonic mapping and the south pole, which corresponds to the RV apex. The line joining them represents the dateline. The dateline, north and south poles are also labelled on the FE mesh to confirm the location of these anatomical regions.

and plotting a hamiltonian line with the shortest path between nodes. The user is able to define midpoints through which the dateline must pass, giving finer control over final path selection. Figure 2.7 demonstrates the user interface for the selection of the two poles, dateline and west elements.

2.6 Image Registration

An Iterative Closest Point (ICP) rigid registration algorithm (Bergstroem & Edlund, 2014) was applied in MATLAB to align the set of RV shapes and remove the effects of organ-level size, orientation and position from the shape analysis. This was achieved by rotating, translating and scaling the point clouds against a template heart. The template RV for image registration was taken as the first Systolic heart (Newcastle, age 50, non-hypertensive), and all RVs were aligned to this surface.

The template heart was centred prior to registration by finding the mean x, y and z coordinates and then subtracting those from each point. Before the ICP algorithm was applied, point clouds were first scaled using RV mesh volumes taken from Simpleware ScanIP. Each point cloud was scaled relative to the volume of the template RV, after the resampling and

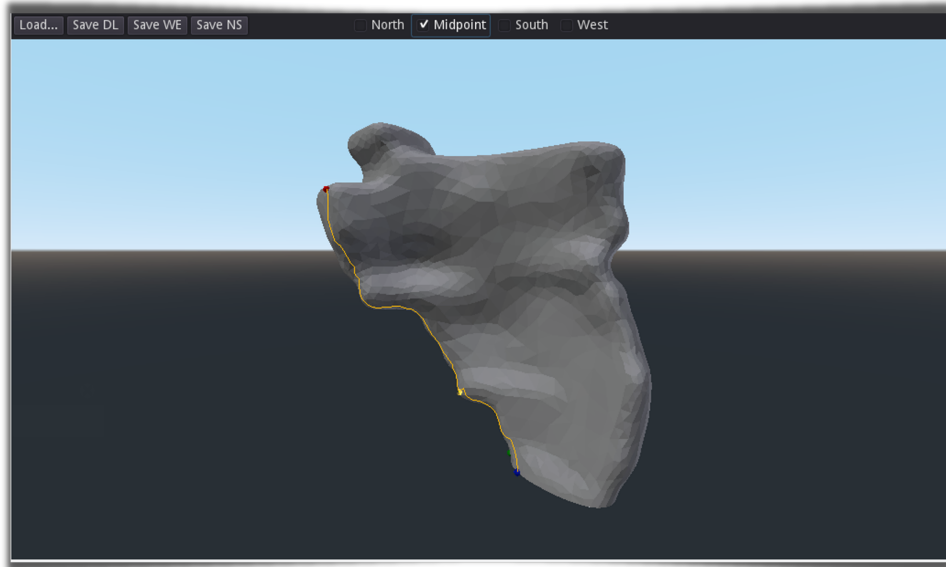


Figure 2.7: User-interface of ADAMCUS, the specialised software designed to select anatomical reference points for mapping. The red sphere is used to select the North pole, the blue sphere to select the South pole. The yellow sphere is used to select any number of midpoints to guide the dateline.

smoothing stages had taken place. ICP iteratively fits the data to the template by minimising the sum of square errors between the closest template and data points. Figure 2.8b demonstrates a successful alignment using the ICP algorithm. The ICP algorithm determines a $[3 \times 3]$ rotation matrix (R) and a $[3 \times 1]$ translation vector (T) which aligns each RV surface to the template in the following way: the registered data = $R \times$ point cloud + T . A manual correction was applied by the author to minimise the difference between the two pulmonary valves and apexes in cases where this was not achieved using the ICP algorithm. In the most disparate RVs, the closest mathematical match between the point clouds did not orient the ventricles anatomically (e.g. Figure 2.8a). In Figures 2.8a and 2.8b the template RV is depicted in blue and the registration candidate is depicted in orange.

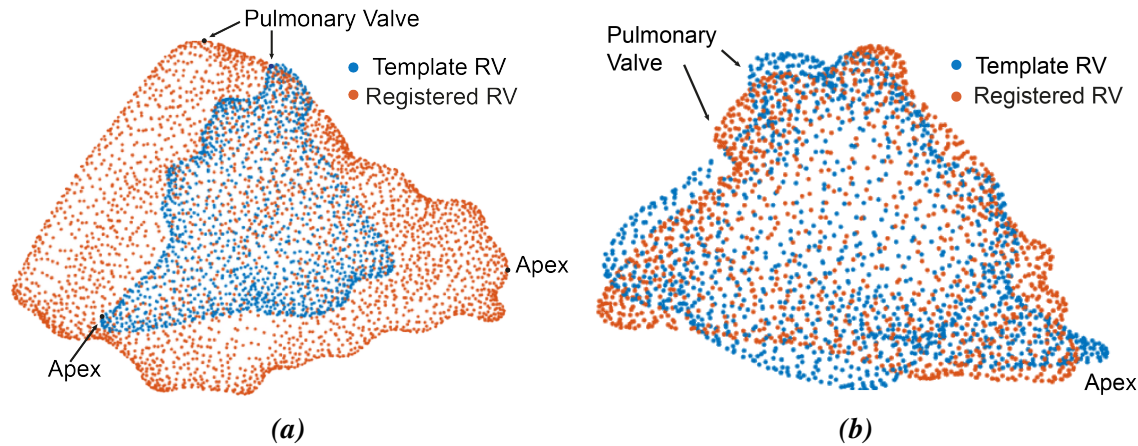


Figure 2.8: Point clouds depicting example outputs in success and failure of the ICP rigid registration algorithm. **a)** Example failure of the iterative closest point algorithm to match correct anatomical points before manual correction. **b)** Example of a successful iterative closest point matching. Manual correction was not required in this case.

$$R = \begin{matrix} R_z(\alpha) & R_y(\beta) & R_x(\gamma) \\ \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} \cos(\beta) & -\sin(\beta) & 0 \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\gamma) & -\sin(\gamma) \\ 0 & \sin(\gamma) & \cos(\gamma) \end{bmatrix} \end{matrix} \quad (2.1)$$

Equation 2.1 shows the rotation matrix used to manually rotate heart surfaces in 3D space to find the closest correspondence between apex and pulmonary valves.

2.7 Topological Mapping

A harmonic mapping method (as described in Wu et al. (2012)) was applied to create a correspondence between all RVES shapes in the dataset. The harmonic mapping approach takes advantage of the fact that the closed RVES surface is topologically equivalent to a sphere (genus-0) to enable each RVES to be uniquely mapped to the surface of a unit sphere. The term genus refers to the number of “holes” an orientable surface has. The correspondence

between any 3D location on the surface of a given heart mesh and the 2D location on the unit sphere is determined from the solution to the following two Laplace's equations (shown here in the theoretical form), which give the spherical coordinates, θ (latitude) and ϕ (longitude):

$$\nabla^2\theta(\vec{x}) = 0, \forall \vec{x} \in \Omega_k \quad \nabla^2\phi(\vec{x}) = 0, \forall \vec{x} \in \Omega_k \quad (2.2)$$

where $\Omega_k \subset \mathbb{R}^3$ is the k^{th} continuous, 3D, non-overlapping closed RVES surface domain and \vec{x} are the standard Cartesian coordinates with respect to the origin, for the given heart surface. This method was chosen because of its ability to maintain high anatomical consistency. The problem is discretised using the finite element method to approximate solutions defining the shape using the nodal values and associated element interpolation. The harmonic mapping requires at least two reference points, the north pole (Γ_n) and south pole (Γ_s) for the latitude mapping, and an additional reference line, referred to as the dateline (Γ_d), for the longitude mapping. These define the boundary conditions for the Laplace's equation.

The dateline represents the line where the longitude passes $360^\circ(2\pi$ radians) and therefore has a non-unique mapped value on the sphere. In other words, the nodes with a longitude of 0 and a longitude of 2π have the same value. To solve this, the surface is cut along the dateline and 'unwrapped' to create two independent, non-overlapping boundaries (and disjoint values of 0 and 2π): an east dateline (Γ_d) and a west dateline (Γ_d^o). The values of the spherical coordinates along these reference lines are assigned as:

$$\theta = 0, \forall \vec{x} \in \Gamma_n, \quad \theta = \pi, \forall \vec{x} \in \Gamma_s, \quad (2.3)$$

$$\phi = 0, \forall \vec{x} \in \Gamma_d, \quad \phi = 2\pi, \forall \vec{x} \in \Gamma_d^o. \quad (2.4)$$

Standard finite element analysis (FEA) through the commercial FEA software ABAQUS was used to solve the Laplace's equations for each RVES. To do this, each segmented RVES

was converted into a continuous non-overlapping mesh of linear triangular membrane elements using Simpleware (described in Section 2.4), and an in-house script was used to identify the values of ϕ or θ corresponding to the given boundary conditions (Equations 2.3 and 2.4 respectively) for each of the two respective finite element analyses (to solve the two Laplace's equations in Equation 2.7). Specifically, static heat transfer analyses with unit conductivity were used within ABAQUS, with the temperature field providing the distribution of ϕ and θ from the analyses, with their corresponding boundary conditions as applied temperature. Note that element size was determined to be sufficient through mesh convergence with representative examples in prior work (e.g. Wu et al., 2013). The nodal values of ϕ and θ throughout the meshes were then exported for further processing.

After mapping, the one-to-one correspondence can be inverted, which involves swapping the geometrical and field parameters, replacing the nodal coordinates with the ϕ and θ values and replacing the vector field with the original spatial coordinates (x , y and z) for each node. Therefore, any given RVES ($\vec{x} \in \Omega_k$) can be described continuously as a shape function, with respect to a common domain for all surfaces:

$$\vec{x} = \vec{x}(\theta, \phi), \quad \forall \theta \in [0, \pi] \text{ and } \phi \in [0, 2\pi]. \quad (2.5)$$

The position vector on the RVES becomes a three-dimensional vector of real-valued functions, representing points in cartesian space, over the spherical coordinates, ϕ and θ . As such, every RVES in the dataset can be compared in a quantitatively consistent manner. Figure 2.9 shows example output of the harmonic mapping process, plotted on the RV shape. Figure 2.9a displays the latitude mapping (i.e. solving for θ) and 2.9b displays the longitude mapping (i.e. solving for ϕ).

Figure 2.10 shows the spherical coordinates plotted on a 2D grid, after the harmonic mapping process has taken place. The longitude (ϕ) is plotted on the x axis and the latitude (θ) on the y axis. Each point on the RV mesh is now represented by a (θ, ϕ) coordinate. It is apparent that a higher density of points exists around the equator than at the poles (along the horizontal

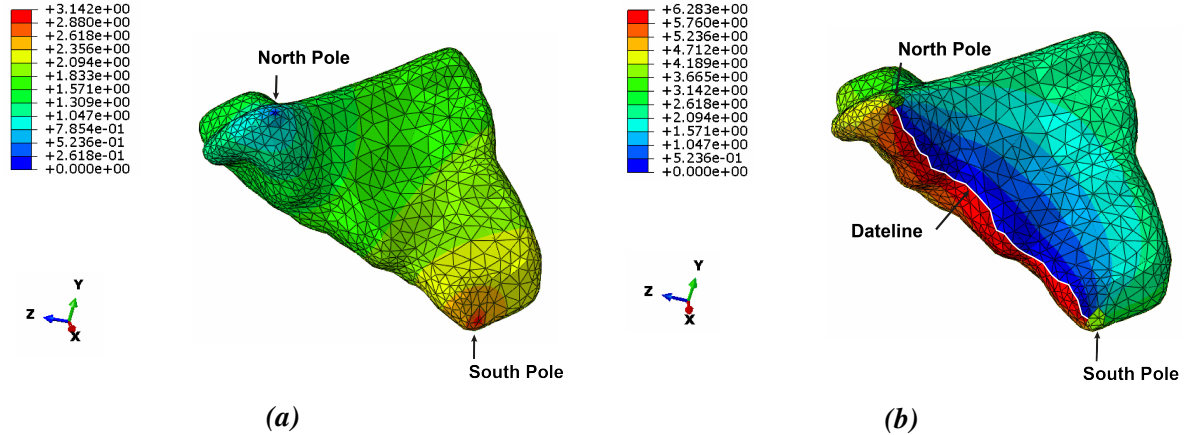


Figure 2.9: Exemplar results from the latitude and longitude shape mappings using Laplace's equations. **a)** Latitude shape mappings for Patient 11 at end-Diastole. Note the North and South Poles are included in this analysis. **b)** Longitude shape mappings for Patient 11 at end-Diastole. Note that the split occurs at the dateline. North and South poles are removed from this analysis.

edges). To correct this and to develop a finer resolution for the purposes of statistical decomposition, a linear interpolation was performed to plot points on a 400×200 grid. The input grid refers to the mesh intersection points. This grid size was selected as it was sufficiently small to capture the variations between mappings while also giving the elements a fixed square size in the (ϕ, θ) domain. Each interpolated RVES is comprised of 80,601 points in total. Figure 2.11 shows an example coordinate mesh after interpolation

2.8 Statistical Decomposition

The Proper Orthogonal Decomposition (POD) method was applied to statistically decompose shape features. POD is a well-established technique that allows a higher-dimensional system to be approximated by a low-dimensional one. It involves finding the dominant patterns, or modes, that describe the variability in the data, and then re-representing the data using a smaller number of these modes. The modes are a linear combination of orthonormal basis functions which capture the most important features of the data. The POD modes are ranked in order of their contribution to variability in the data; the first few modes capture the largest variability, while the lower order modes capture progressively smaller amounts of variability

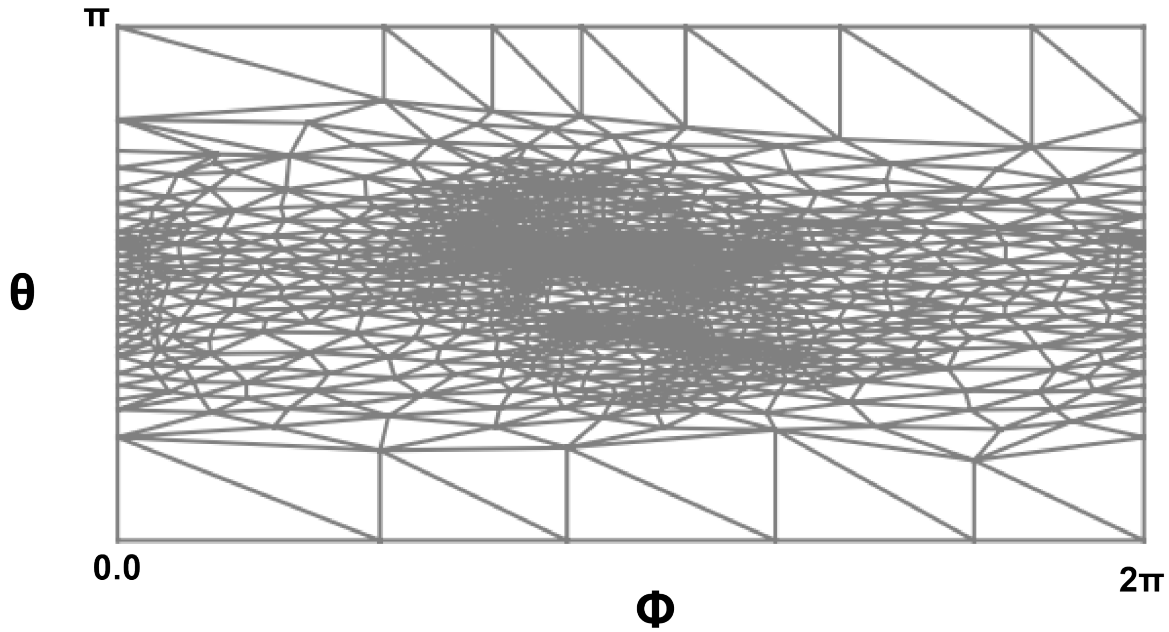


Figure 2.10: Example 2D mapped x coordinate mesh before interpolation. There is a higher density of points around the equator than the poles. The x axis represents longitude coordinates, ranging from 0 to 2π radians and the y axis represents latitude coordinates, ranging from 0 to π radians.

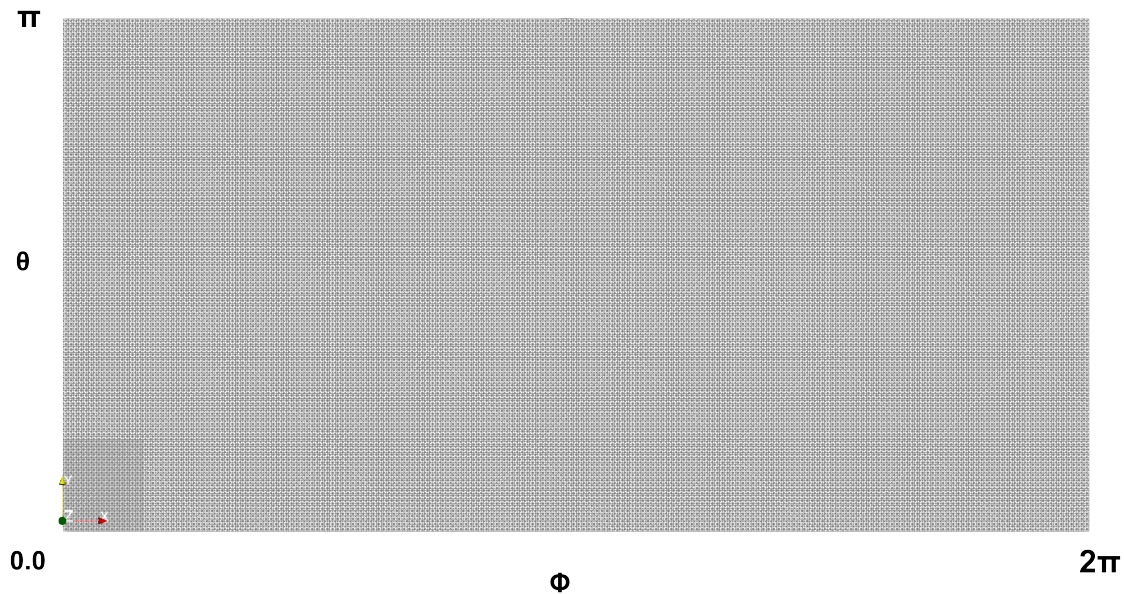


Figure 2.11: Example 2D-mapped x coordinate mesh after interpolation. The interpolation step places these points on a regular 400×200 grid. The x axis represents longitude coordinates, ranging from 0 to 2π radians and the y axis represents latitude coordinates, ranging from 0 to π radians.

(Chatterjee, 2000).

POD can also, interchangeably, be referred to as Principal Component Analysis (PCA). However the term POD is used in this application to the continuous form of the mapped shape function (as defined here by the finite element nodal values and associated element interpolation), in comparison to the finite optimal axes associated with PCA output. The form of decomposition applied assumes that any shape function from the dataset could be decomposed as follows:

$$\vec{x}_k(\theta, \phi) = \vec{x}_\mu(\theta, \phi) + \sum_{i=1}^m a_{ki} \vec{v}_i(\theta, \phi), \quad (2.6)$$

where the translation function, \vec{x}_μ is the mean of all RVES shapes, and a_{ki} is the real-valued coefficient that approximates the k^{th} shape function with the i^{th} mode, \vec{v}_i . POD identifies the basis (i.e modes) that minimises the average of the L_2 -norm of the difference between each dataset and the best approximation of the dataset using that basis. This leads to the following optimisation problem:

$$\min_{\vec{v}_i(\theta, \phi)_{i=1}^m} \langle \|\vec{x}_k(\theta, \phi) - \vec{x}_k^*(\theta, \phi)\|_{L_2}^2 \rangle, \quad (2.7)$$

where $\vec{x}_k^*(\theta, \phi)$ is the best approximation to $\vec{x}_k(\theta, \phi)$ through Equation 2.6, which can be obtained using the projection operator.

The method of snapshots was applied to form the following m -dimensional eigenvalue problem, where m is the number of RVES shapes in the dataset, which can be solved to identify the set of POD modes.

$$\frac{1}{n} \sum_{k=1}^n A_{jk} C_k^{(i)} = \lambda^{(i)} C_j^{(i)} \quad (2.8)$$

where

$$A_{jk} = \int_0^{2\pi} \int_0^\pi (\vec{x}_j - \vec{\bar{x}}) \cdot (\vec{x}_k - \vec{\bar{x}}) \sin(\theta) d\theta d\phi \quad (2.9)$$

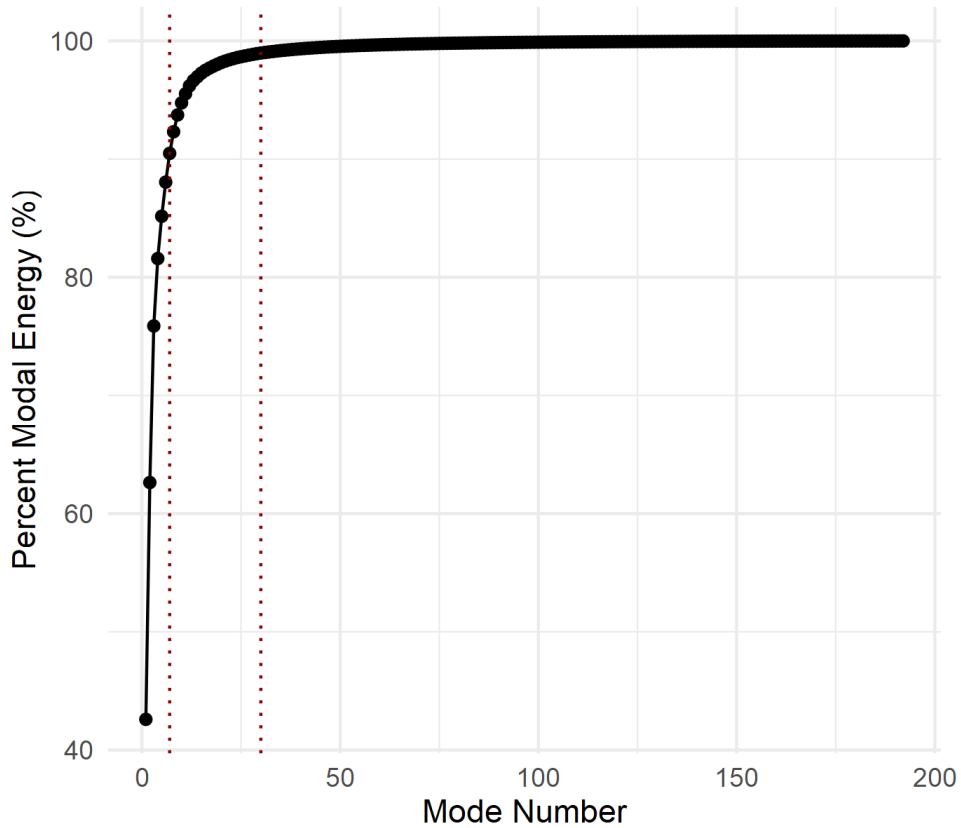


Figure 2.12: Example cumulative modal energy curve corresponding to each of the shape modes included in the analysis. The eigenvalue corresponding to each mode is a measure of the relative contribution of that particular shape feature to the overall RV shape when reconstructing all shapes in the dataset. The two dotted red lines indicate the number of modes which cover 90% and 99% of the total modal energy respectively.

and

$$C_k^{(i)} = \int_0^{2\pi} \int_0^\pi \vec{p}_k(\theta, \phi). \quad (2.10)$$

The eigenvalue corresponding to each mode is a measure of the relative contribution of that particular shape feature to the overall RV shape when reconstructing all shapes in the dataset. Therefore, a typical rule is to consider the modes with the largest corresponding eigenvalues, such that the summation of the eigenvalues of the created modes is in the range of 99% of the total sum of all eigenvalues in the set. The statistical decomposition was performed using the FEniCS suite of packages (Logg et al., 2012; Logg & Wells, 2010).

2.9 Output and Visualisation

The process to visualise the 3D shapes of the POD modes was the reverse of the process to produce patients the 2D shape functions (Equation 2.5) from the 3D RVES meshes and mapping results. The geometric and field parameters were simply swapped again so that the nodal coordinates of the mode were replaced by the x, y and z field coordinates.

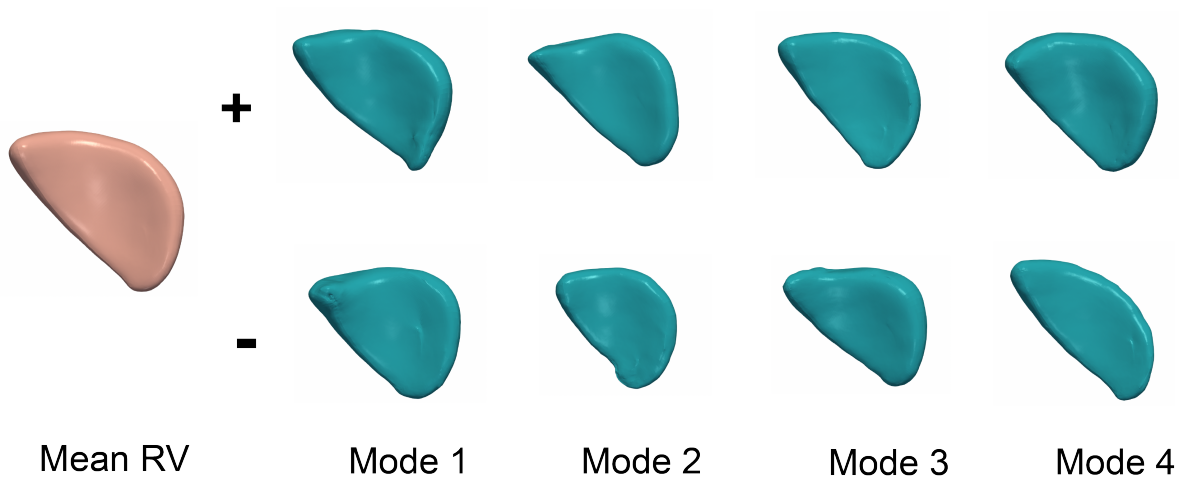
As described in Section 2.8, the POD shape analysis produces a mean RVES shape. Each RVES can be reconstructed from this mean RV plus each shape mode (feature) multiplied by some coefficient which represents the ‘amount’ of each shape contained in any given individual’s RV. The output can be interpreted qualitatively, by plotting the mean shape \pm any given shape mode or quantitatively by using statistical metrics to determine correlations between those shape features and biological metrics. Equation 2.11 describes how the reconstructions are formed from the mean RV \pm each coordinate multiplied by $2 \times$ the standard deviation of the range of modal coefficients for each mode.

$$\vec{x}_\mu(\theta, \phi) \pm \vec{x}_i(\theta, \phi) * 2 \sqrt{\frac{\sum_k (a_{ki} - \bar{a}_{ki})^2}{N}} \quad (2.11)$$

Figure 2.13 gives a graphical representation of the qualitative mode reconstruction, using the mean and the first four modes as examples.

This method is used to derive mode shapes and coefficients in the three subsequent data chapters (Chapters 3, 4, 5). Any chapter-specific variations to the methodology are discussed in their respective Methodology sections (Sections 3.3, 4.3 and 5.3).

Figure 2.13: Example qualitative reconstructions of the mean $RV \pm 2 \times SD$, used to visualise each shape feature.



Chapter 3

Evaluation of Population-Based Differences in Ventricular Shape

3.1 Introduction

The RV of the human heart exhibits quantifiable shape variation between individuals, both genetic and in response to environmental factors (Mauger et al., 2019). Cases where shape changes are indicative of pathology have been well documented, especially when related to cardiopulmonary disease. It has long been recognised that RV shape alters in response to variations in pressure and volume loading.

However, natural variation in heart shape which is not indicative of pathology or related to an impairment in function is poorly understood. It is hypothesised that variations in RV phenotype are also affected by benign genetic and environmental factors such as sex, geographic region and age, and that novel correlations also exist between such demographic features and RV shape. Shape variation attributable to a broad spectrum of demographic characteristics and risk factors is likely to play an important role in modulating the cardiovascular response to disease and therefore can be crucially integrated into a predictive metric for detecting cardiopulmonary disease. The impact of three main phenotypes will be examined in this chapter:

sex, obesity (as defined by a Body Mass Index (BMI) of $> 28\text{kg}/\text{m}^2$) and age (as defined by an age greater than the group mean of 60.53 years).

RV remodelling most often involves changes in size, wall thickness and regional curvature (Addetia et al., 2018) but can also be represented by changes in sphericity and eccentricity. Understanding of natural variation in right ventricular shape has been hampered by technical difficulties with RV imaging and RV geometry estimates (Walker & Buttrick, 2013) owing to the RV's anatomical intricacies. As described in detail in Section 1.2, the RV shape is very complex in contrast to the conical LV. It is triangular when viewed from the front and curves over the LV, whereas when viewed in cross-section, as in short-axis MRI scans, its appearance is crescent shaped (Ho, 2006).

There are inherent distinctions in RV shape between the sexes (as noted by Tandri et al., 2006; Kawut et al., 2011). However, there is very limited literature comparing RV shape in healthy individuals. With the exception of Addetia et al. (2018), the literature is focused on shape changes as part of a disease state. Lakatos et al. (2020) examined the effects of sex amongst other demographic factors, on RV motion in healthy volunteers. Past studies have described superior RV systolic function in healthy females (Martínez-Sellés et al., 2015; Keen et al., 2021) which persisted even after adjustment for body size and left ventricular function (Tandri et al., 2006; Kawut et al., 2019). In the Multi-Ethnic Study of Atherosclerosis (MESA), a prospective study of men and women aged 45-84 years old who were screened to be clear of cardiovascular disease, it was found that men have 8% higher RV mass ($p < 0.001$) and 7% higher ventricular volumes ($p < 0.001$) (Kawut et al., 2011), where relationships were characterised by bivariate and multivariate linear regression. Men also have lower RV ejection fractions (absolute decrease of 4%, $p < 0.001$) as determined by CMR. This trend was consistent across all ethnic groups (Caucasians, African Americans, Hispanics and Chinese Americans) and mirrors the trend observed in the left ventricle in both volume (Salton et al., 2002) and ejection fraction (Chung et al., 2006). A recent collaboration between Queen Mary University of London, the University of Barcelona and the University of Southamp-

ton (Raisi-Estabragh, 2020; Cetin et al., 2020) used CMR to assess the LV structure of 667 healthy individuals (309 men and 358 women) from the UK Biobank Imaging study. Male heart muscle was found to be dominated by coarser textures, whereas the female ventricle was predominated by finer muscle grains. The same study also identified significant differences in overall shape, including larger heart surface area in male hearts even after controlling for body size. Sex-based differences in heart shape became less prominent with increasing age, while texture differences persisted across all age groups (45-54, 55-64, 65-74) and were most pronounced in the oldest individuals. Addetia et al. did not identify any inter-gender differences in regional curvature (Addetia et al., 2018).

RV shape differences are magnified in several RV-centric diseases which can have important prognostic implications. Studies of heart failure patients have consistently found a survival benefit in females, which is more pronounced in ischaemic patients (those with reduced blood flow to the heart as a result of partial or complete blockage of the coronary arteries) and that this effect is explained by the RV (Jacobs et al., 2014). One of the best-researched sex based divergences in the RV is related to survival in idiopathic Pulmonary Arterial Hypertension (PAH). Although the incidence of hypertension is greater in female patients, the survival rates far exceed that of their male counterparts. A pair-matched controlled study to characterise right ventricular shape differences in patients with idiopathic PAH found that when correcting for both age and sex, RV mass tended to be higher in female individuals (Swift et al., 2015). These findings indicate that the female RV with PAH undergoes the same, if not more hypertrophy than male ventricles. Moreover, some interesting recent analysis (Tello et al., 2020) documents that women with PAH have improved RV contractility and RV-pulmonary artery coupling; women have an effective median arterial elastance [Ea] of 0.85 compared with 0.52 in male subjects. Pulmonary effective arterial elastance is the ratio of RV end-Systolic pressure and stroke volume, better described in Brener, Burkhoff, and Sunagawa (2020), and represents a concise overview of both steady and pulsatile components of the arterial load. Females also exhibit a better RV response to medical therapy which

further contributes to their improved survival.

Female LVs adapt to pressure overload differently, exhibiting more prominent concentric (rather than eccentric) remodelling (Kuch et al., 1998). They also suffer a more profound response to hypertension and obesity than men. There is reasonable evidence to suggest that sex hormones directly modulate RV function in PAH, and that this is a contributing factor to the differing responses between the sexes. Keen et al. (2021) report links between oestrogen (the main female-predominant sex hormone), testosterone (the main male-predominant sex hormone), and dehydroepiandrosterone (a precursor for multiple sex hormones in both males and females) with RV dysfunction. Rider et al. (2015) found that obese subjects show sex-specific differences in right ventricular hypertrophy. However, in this study (n=739), there was no correlation between end-Diastolic volume and BMI in men. Some interaction effects between the shape and function-related impacts of various characteristics are evident.

Obesity is a known risk factor for cardiovascular diseases as higher body mass increases cardiac workload by increasing total blood volume and cardiac output (Sokmen et al., 2013). Sokmen et al. (2013) examined RV shape change in obese, but otherwise healthy, young adults (<40 years) using echocardiography and tissue Doppler imaging. This study found that obesity was associated with RV and right atrial dilatation, indicated by increased diameter and increased RV free wall thickness. Isolated obesity is associated with subclinical impairment in RV function at both end-Systole and end-Diastole, and with dilatation of right cardiac chambers in young, otherwise healthy subjects. These unfavourable changes were more pronounced in severely obese subjects and appear linked to the degree of obesity. Chahal et al. (2012) also investigated the relationship between obesity and RV structure and function using patients from the MESA RV study (mean age 61.5 years). RV mass and volumes were greater in overweight and obese participants compared with lean ones (6% greater in overweight and 14% greater in obese) after adjusting for other demographic factors and LV mass.

Right ventricular ejection fraction (RVEF) showed an inverse relationship to both waist circumference and waist to hip ratio both before and after adjusting for left ventricular ejection

fraction (LVEF) (i.e. larger waist circumference was associated with poorer RVEF). Ejection Fraction (EF) is a measure of cardiac efficiency and represents the percentage of blood pumped out of a filled ventricle with each heart beat. A normal LVEF is around 50-70%.

They report results that were independent of body size indicators, such as height and the corresponding LV parameter, and therefore conclude that the relationship between obesity and RV morphology cannot solely be attributed to the presence of morbidities such as hypertension or diabetes (Chahal et al., 2012). Danias et al. (2003) used CMR to identify a similar relationship in a cohort of 25 obese men and 25 slim men (mean age approximately 30 years) and argued that a relationship between elevated RV mass, end systolic volume and end diastolic volume indicated that obesity was independently associated with RV remodelling. The results from Danias et al. (2003) on the other hand did not persist after accounting for height, although this effect could be related to the relatively small sample size. Various studies have therefore revealed conflicting data regarding the effects of obesity on the RV and much still remains unknown about RV structure and function variations in obese subjects (Chahal et al., 2012; Wong et al., 2006; Otto et al., 2004; Yildirimturk et al., 2011). Sphericity (associated with a width to height ratio) is known to be associated with adverse remodelling processes in both asymptomatic populations and patients with cardiovascular disease. Mauger et al. (2019) identified that increases in the sphericity shape mode were found in patients with obesity and angina.

There is a substantial overlap of size based RV parameters with those detected in pathological conditions which makes evaluation of the normal effects of increased size challenging. D'Andrea et al. (2014) noted this while assessing RV adaptive remodelling in highly trained athletes. The main changes associated with exercise are cardiac enlargement with preserved compliance and contractility and an increase in free wall thickness. The same authors had previously described a disproportionate load on the RV during exercise.

The heart has a tendency to undergo structural changes as individuals age. The majority of analysis of age-related shape change is conducted on the left ventricle (e.g. Kuch et al.

(1998)). For example, Strait & Lakatta (2012) describe the significant structural change in the heart and vasculature associated with ageing. Such changes include arterial stiffening, leading to increased systolic and pulse pressure; an increase in myocardial thickness as a result of increased cardiomyocyte size and an increase in the size of the various changes. The heart's overall shape transitions from elliptical to spheroid (Strait & Lakatta, 2012) because of an asymmetrical change where the interventricular septum is affected more than the free-wall (Hees et al., 2002). However, there has been some recent studies which focus specifically on the RV. A community-based cohort study found that subclinical RV dysfunction was present in nearly one fifth of elderly individuals (mean = 76 years, STDEV = 5) (Nochioka et al., 2018). After adjustment for a series of factors including age, sex and ethnicity, it was found that each 5% decrease in RVEF was associated with a 20% increase in hazard for death or hospitalisation for heart failure (Nochioka et al., 2018).

Addetia et al. (2018) used 3D echocardiography to derive curvature indices for shape analysis in healthy individuals with the aim of facilitating future 3D echocardiographic study of adverse remodelling in the right ventricle. In previous work (Addetia et al., 2016), this group have described the action of RV contraction as “bellows-like” . They found no statistically significant difference in curvature between genders, but that in older subjects (here classified as ≥ 55 years of age) the RV free wall and RV outflow tract were flatter. They also found that the inflow tract became less flattened and the apex less pointed during the transition from diastole to systole and concluded that these changes are related to increased stiffness of the RV in older subjects. One study observed a slight but significant decrease in end-systolic and end-diastolic volume, associated with a significant increase in LVEF and RVEF as patient age increased (Fiechter, 2013). Changes in ventricular structure and function attributable to the ageing process may play an important role in modulating the cardiovascular response to disease (Innelli et al., 2008). Innelli et al. (2008) also evaluated the influence of age on RV tricuspid annulus measurements in a population of healthy subjects and proposed reference values across various age bands, split by decade. Exclusion criteria for this study were: smok-

ing and history of any acute or chronic cardiopulmonary disease.

The central hypothesis of this work is that multiple physical traits of an individual are encoded within the shape of their RV, and that the effects of these traits on RV shape can be measured and isolated. It follows, therefore, that natural (or non-pathological) variations can be discriminated from pathological variations.

Taking the framework established by Wu et al. (2013) for statistical shape analysis of the RV, the aim of this study is to evaluate a dataset of clinically obtained CMR images from two centres and assess correlation of shape features extracted from the RV to a range of demographic measures, including subject sex and age as well as the centre they were assessed in. It was hoped that shape information associated with the origin of the subjects could be isolated.

The flowchart in Figure 3.1 demonstrates the analytical steps conducted in this Chapter in order to identify key shape features which differ between populations. Section 3.3.1 describes the clinically obtained CMR dataset, Section 3.3.2 briefly details the methodology of the statistical shape analysis, while Section 3.4 presents and discusses the shape analysis results, along with a linear discriminant investigation of the uniqueness of specific RV features, in relation to their separation potential. This is followed by concluding remarks.

3.2 Aims and Hypotheses

This chapter will investigate the following hypotheses:

1. A measurable difference will exist in RV shape, as measured by modal coefficients between the two datasets (Newcastle and Pittsburgh).
2. Differences between groups will be most pronounced at end-Diastole, owing to greater coverage by the MRI scanner.
3. There will be no significant difference in shape between male and female RVs, as measured by the ability to categorise individuals using Linear Discriminant Analysis, since

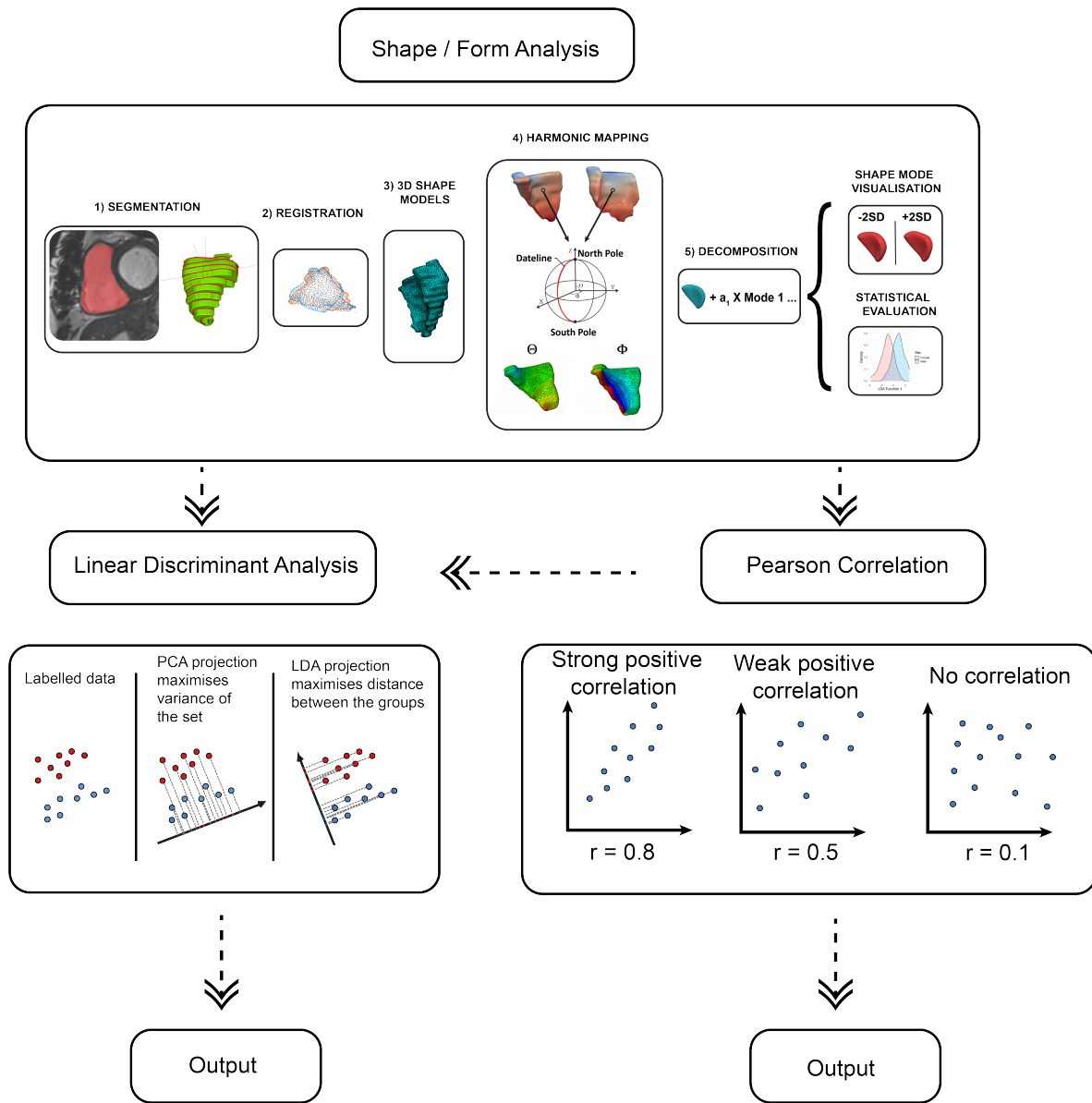


Figure 3.1: Flowchart describing the shape analysis workflow

the POD method removes the effect of size and volume. There will be no fundamental shape difference between the sexes once size is controlled for.

4. A measurable difference will exist between RVs in the young and old groups, based on evidence that there is a less pronounced sex-based difference in older patients. This is suggestive of shape features that are specific to the ageing hearts.
5. There will be a significant shape difference observed across all groups when a form analysis is undertaken. However, this will relate to the inclusion of RV volume ('size'). Notably, because the end-Systolic and end-Diastolic phases are decomposed together, leading to two different morphs per individual.

3.3 Methodology

The overall statistical shape analysis process used to determine key shape features is detailed in Chapter 2, and summarised in Figure 3.1. To briefly recapitulate, first short-axis MRI scans are manually segmented, and then smoothed and meshed to produce 3D finite element models of the RV shape. Each model is scaled and then registered to a common reference using an ICP algorithm. The RV shape models are topologically mapped to the unit sphere using a harmonic mapping process, thus creating a unique shape function for each RV. The complete set of shape functions are then statistically decomposed using POD to determine and rank a set of shape features (i.e. modes and corresponding coefficients from decomposition). Having derived these modes and coefficients, the modal coefficients can be utilised to determine which shape features are significant to the patient population.

At the same time, an exploratory form analysis (Dryden & Mardia, 1998; Mitteroecker et al., 2013) was also undertaken (considering both shape and size) to assess whether the inclusion of volume-related information, in addition to shape, made an impactful difference to the ability to discriminate between traits. The form analysis included an additional step at the ICP registration, which removed the mesh scaling effects after centring had taken place.

3.3.1 Population Characteristics

Ninety-six individuals at two phases were used in this study ($n=50$ from UPMC and $n=46$ from Newcastle University) giving a total of 192 heart shapes. There were 44 females and 52 males in the cohort. The mean BMI was 28.7 (SD 7.42) and the mean age was 60.53. Healthy and remodelled RVs were decomposed together as part of this dataset, and similarly both end-Systole and end-Diastole were included in the statistical decomposition. The benefit of including both phases in the statistical decomposition is to allow the potential isolation of phase-independent shape features. Similarly, to understand the nature of the dataset, and therefore identify disease-specific shape features, the POD modes must capture a range of diseased and healthy features, and decomposing both phases together increases the number of shape features included in the total analysis, and produces a neutral mean. Analysis of the mode coefficients is undertaken separately for each phase, so that specific Systolic and Diastolic features can be identified, however these could be combined to give a better overall picture of ventricular shape throughout the motion of the heart cycle. Despite different loadings at the two cardiac phases, the POD method should identify and capture these features as part of its dimension reduction.

Tables 3.1 and 3.2 give some descriptive statistics, including mean, median and standard deviation of the 96 subjects included in the statistical shape analysis. From Table 3.1 it is evident that the BMI was higher in male individuals than female (29.7, vs 27.7), and was higher in individuals below mean age compared to those above mean age (30.3 in subjects younger than 60.5 years compared to 27.6 in subjects greater than 60.5 years). Mean BMI was also greater in the Pittsburgh cohort as compared to Newcastle (30.3 vs 27.1). The most pronounced difference in mean BMI was between Hypertensive subjects and non-Hypertensive subjects (31.2 vs 25.4).

All factors were characterised as dichotomous variables. Age was defined as those above or below the mean age (60.53 years). Obesity was defined as those who are clinically overweight (BMI of $> 28\text{kg}/\text{m}^3$). Groups were selected to align with medical literature, while en-

Table 3.1: Descriptive statistics based on subject BMI. BMI statistics are reported for a series of dichotomous categories: the two sexes (female and male), those below and above average age (based on an age threshold of 60.5 years), between the two centres (Newcastle and Pittsburgh) and between healthy and diseased subjects (Hypertensive and Non-Hypertensive). These descriptive statistics include all 96 subjects, regardless of hypertensive or non-hypertensive status.

BMI	F	M	Age < 60.5	Age > 60.5	Newc	Pitt	Non-Hyp	Hyp
Mean	27.7	29.7	30.3	27.6	27.1	30.3	25.4	31.2
Median	25.7	28.3	28.6	26.8	26.7	28.9	24.8	30.7
St Dev	7.88	6.96	9.19	5.64	4.88	8.90	4.27	8.27
Min.	15.0	19.9	18.8	15.0	18.9	15.0	18.81	15.0
Max.	58.3	61.3	61.3	39.1	38.0	61.3	36.3	61.3

Table 3.2: Descriptive statistics based on subject age. Age statistics are reported for a series of dichotomous categories: the two sexes (female and male), those below and above average BMI (based on a BMI threshold of $28\text{kg}/\text{m}^3$), between the two centres (Newcastle and Pittsburgh) and between healthy and diseased subjects (Hypertensive and Non-Hypertensive). These descriptive statistics include all 96 subjects, regardless of hypertensive or non-hypertensive status.

Age	F	M	BMI < 28	BMI > 28	Newc	Pitt	Non-Hyp	Hyp
Mean	59.0	61.0	63.1	57.9	63.6	57.7	59.0	61.7
Median	59.9	63.0	61.7	60.0	61.4	60.1	60.2	63.4
St Dev	10.8	14.6	10.3	15.0	8.47	15.5	11.3	14.0
Min.	34.3	18.4	27.7	18.4	50.4	18.37	27.7	18.4
Max.	79.5	80.0	79.5	80.0	80.0	79.2	79.2	80.0

sureing that the groups remained roughly equal in size. Correlation was assessed using Pearson Product-Moment correlation coefficients for continuous variables and Point Biserial correlations for categorical variables. The Pearson coefficient is given by:

$$r_{df} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \quad (3.1)$$

and assigns a value between -1 and 1, where 0 indicates no correlation, 1 is total positive correlation and -1 is total negative correlation. x_i represents the initial mode coefficient values in a test (the results obtained in Chapter 3), y_i represents the comparison mode coefficient values, \bar{x} represents the mean of those groups and df represents the number of degrees of freedom, where $df = n-2$.

All exploratory analysis was performed using IBM SPSS 27 (IBM, Armonk NY), while Figures were produced using *ggplot2* (Wickham, 2016); R version 4.1.2 (2021-11-01).

3.3.2 Statistical Analysis

Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) (Fisher, 1936) was applied to the mode coefficient values produced using POD (procedure described in Section 2.8) to build classifiers which attempt to separate RVs based on their population characteristics.

LDA is a dimensionality reduction technique; it helps to reduce high dimensional data onto a lower dimensional plane. Although their aims are similar, LDA approaches dimensionality reduction in a different way to Principal Component Analysis (PCA)/POD (the technique used to derive the shape features in this thesis). Figure 3.2 shows a simplified diagram to emphasise how LDA works in comparison to PCA in its achievement of dimensionality reduction. LDA performs well as a classification tool, since it takes labelled groups, and maximises variance between those groups. In other words, LDA searches for those vectors in the underlying space

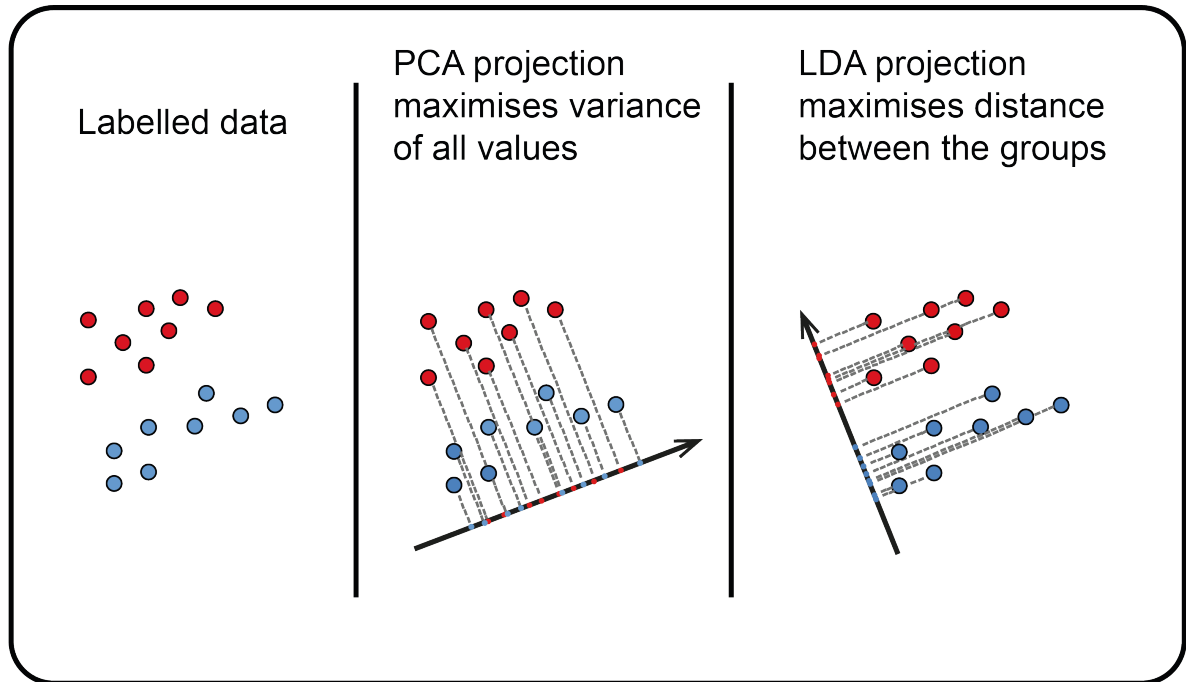


Figure 3.2: A comparison between PCA and LDA projections. PCA maximises the variance of all values, whereas LDA maximises the distance between individual groups, thereby enabling separation of values into pre-defined categories. Figure adapted from Muraleedharan (2021)

that best discriminate among classes, rather than those vectors which best describe the data. On the other hand, PCA maximises the variance of all values in a dataset, by choosing new axes for the dimensions. The new axes are an orthogonal projection into lower dimensional space and most clearly separate the groups, i.e. the data is rotated in such a way as to align the directions in which the data is most spread out with the principal axes of the graph. This is a useful tool for dimensionality reduction in most cases, and is used to derive the shape modes earlier in the methodology, but as those separations do not necessarily fall across desired group boundaries, PCA is not an efficient tool for classifying labelled data (Martinez & Kak, 2001).

LDA can be described mathematically in the following way. Given the set of k samples (in this case mode shapes), m_1, m_2, \dots, m_k where each sample is represented by a features (containing an individual modal coefficients for each subject) $a_{ki_1}, a_{ki_2} \dots a_{ki_n}$, where each sample contains a_{ki_i} belongs to one of c classes, y_1, y_2, \dots, y_c . In this Chapter, each class is a dichotomous feature, for example above and below mean age. The goal of LDA is to find an a' -dimensional feature space, where $a' < a$, and where the classes are well separated. This is

achieved using a three step process. The first step is to calculate the separability between different classes (i.e. the difference between the means of the different classes); this is known as the between-class variance. The second step is to calculate the distance between the mean and the samples of each class, which is called the within-class variance, and the third step is to construct the lower dimensional space which *maximises* the between-class variance and *minimises* the within-class variance.

As a first step, the mean vectors are computed for each class:

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^{n_j} a_{ki} \quad (3.2)$$

for $j = 1, 2, \dots, c$, where n_j is the number of samples in class j . The second step is to compute the within-class scatter matrix:

$$S_W = \sum_{j=1}^{n_j} (a_{ki} - \mu_j)^2 \quad (3.3)$$

and then the between-class scatter matrix, which represents the distance between the mean of the i^{th} class (μ_i) and the total mean (μ):

$$S_B = \sum_{j=1}^c n_j * (\mu_j - \mu)(\mu_j - \mu)^T \quad (3.4)$$

where

$$\mu = \frac{1}{n} \sum_{i=1}^n a_{ki} \quad (3.5)$$

and

$$n = \sum_{j=1}^c n_j \quad (3.6)$$

Compute the eigenvectors of the scatter matrices:

$$S_w^{(-1)} * S_b * W = \lambda W \quad (3.7)$$

where W is a matrix of eigenvectors, and λ is a diagonal matrix of eigenvalues.

Select k eigenvectors, where k is $c-1$, and form a new feature space:

$$Z = W^T * X, \quad (3.8)$$

where X is a matrix of feature vectors in the original feature space. Then it is possible to use the new feature space for classification. For example, given a new sample x , compute its feature vector in the new feature space, $Z = W^T * x$, and assign it to the class whose mean vector is closest to it in the new feature space.

Implementation Strategy

Cumulative modal energy was plotted to determine the number of shape and/or form modes which accounted for 99% of the total shape variation. These modes were taken forward for further analysis. Preliminary tests applied LDA to the modal coefficients (derived in Equation 2.6) obtained for all non-Hypertensive individuals using the modes which accounted for the first 99% of the cumulative total modal energy, and which met the cutoff threshold.

Each preliminary test identified the five shape modes/features with the largest standardised structure matrix coefficients, i.e. those factors with the highest correlations to the discriminant function. Each element of the structure matrix is the correlation between a predictor variable (or mode number in this analysis) and a discriminant function, after adjusting for the correlations between the predictor variables. The larger the absolute value of an element, the stronger the correlation between that variable and the corresponding discriminant function. The structure matrix therefore provides a valuable method of identifying the most important modes, based on their ability to separate the RVs (modal coefficient values) into the demographic groups specified in the input. The discriminant function in LDA is a linear combination of

the predictor variables which defines a hyperplane separating the groups in p -dimensional feature space, where p is the number of predictor variables.

Five modes were selected in particular, as up until the fifth mode, adding in a factor to the LDA model improved discrimination between groups, however, after adding in a sixth mode as a factor there was no improvement in the ability of the LDA model to discriminate between groups, indicating that five modes was sufficient to capture the salient modes in separating the majority of dichotomous variables. A second round of LDA was then applied using subsets of these five modes to find the combination which gave the best classification result.

In many cases creating the initial LDA model caused over-fitting (near perfect class separation on the training data, with chance performance on the test data). Overfitting is a commonly reported drawback of the LDA process (e.g. Luo et al., 2011). In high dimensions LDA has a tendency towards overfitting as the computations rely on the inversion of the within-class covariance matrix. For high dimensional matrices, inversion is a sensitive operation that can only be reliably performed if the estimate of within-class covariance is very accurate. In high dimensions, it is much more difficult to obtain a precise estimate of within-class covariance, and in practice, far more datapoints than dimensions are required, otherwise the within-class covariance matrix will be almost singular (the eigenvalues will be very low), which will cause overfitting.

To reduce the impact of overfitting, the initial LDA results were used to identify five modes with the largest standardised structure matrix coefficients, i.e. the modes with the strongest correlations to the discriminant function. A second round of LDA was then applied testing every combination, and all subsets, of the top five modes to find the combination which gave the best classification result. The linear classifier was then applied to the remaining 56 hypertensive subjects to observe how the separation changes when subjects with a disease known to change RV shape are introduced. Separate classifiers were produced for hearts at end-systole and end-diastole. The success of this classifier was compared to a separate LDA, generated using the 56 hypertensive subjects.

Results are displayed by reporting the highest successful separation percentage after leave one out cross-validation, and by plotting the density of LDA function value against dichotomous variable group. Leave one out cross-validation operates by splitting the dataset into a training set and a testing set, using all but one observation to form part of the training set. The model is then used to predict the response value of the single observation that has been left out, and calculate the mean squared error:

$$\text{MSE} = \frac{1}{n} \sum (y_i - f(x_i))^2 \quad (3.9)$$

where n is the total number of observations, y_i is the response value of the i^{th} observation and $f(x_i)$ the predicted response value of the i^{th} observation. The closer the prediction is to the observation, the smaller the mean squared error. This process is then repeated for each subject in the dataset, leaving out a different observation from the training set each time. The test mean squared error is taken as the average of all the leave one out tests:

$$\text{Test MSE} = \frac{1}{n} \sum \text{MSE}_i \quad (3.10)$$

where n is the total number of observations in the dataset and MSE_i is the test MSE for the i^{th} iteration of fitting the model. Leave one out cross-validation is implemented in SPSS (IBM, Armonk NY).

Qualitative images of the effect of key modes were also produced by plotting a 3D mesh of the mean heart and adding each mode eigenvector multiplied by the coefficient of the individual with the most extreme positive and negative LDA function values for the population characteristic in question.

3.3.3 Qualitative Reporting

The combined qualitative impact of shape modes on subject RVs was visualised by plotting the overall mean RV shape (encompassing both healthy and diseased subjects) plus the modal

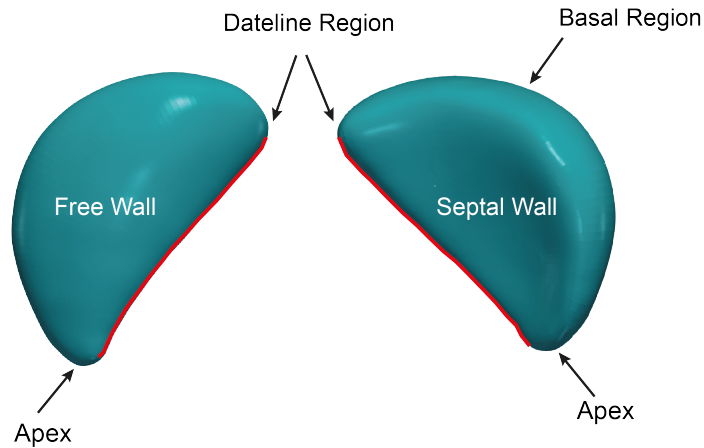


Figure 3.3: Labelled diagram of the RV, depicting the septal wall, free wall, apex and base regions and the position of the dateline.

coefficients of the RV identified as having the most ‘extreme’ discriminant function value (i.e. the most positive and the most negative).

Discriminant function values are derived from the linear combination of factors (the selected Mode coefficients) and their canonical weightings, as determined by the LDA. As such, the greatest negative value will be held by the RV that most closely resembled the characteristics. The discriminant function values were ranked, and the largest and smallest numbers were selected. 3D models were created by multiplying the modal coefficient values held by that RV for the salient modes.

General trends in RV shape across categories were determined by plotting group means. 3D visualisations were created by plotting the non-hypertensive mean, created from 40 non-hypertensive RVs in addition to individual subgroup means, which were displayed alongside the overall mean. Mean values are calculated by averaging x, y and z coordinates separately over each point (80601 locations) for each individual included in the required category (e.g. 58 non-obese RV shapes, as end-Systole and end-Diastole are plotted together). The x, y and z co-ordinates were then plotted in 3D space to produce mean RV shapes.

Figure 3.3 shows a labelled diagram of the RV, which depicts various terms used to describe shape change throughout this Chapter. The septal and free wall are marked, in addition to the apex and basal regions.

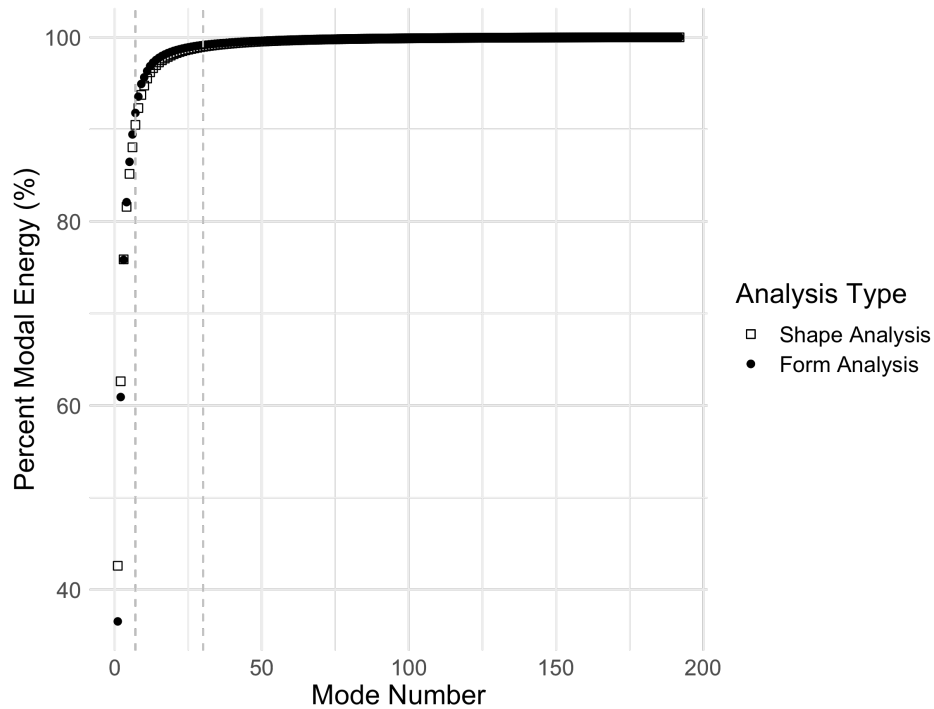


Figure 3.4: Cumulative modal energy corresponding to each of the 192 shape modes and form modes decomposed from their respective POD decompositions. Shape modes are depicted by white circles and form modes by black squares. Dotted lines mark the number of modes required to cover 90% and 99% modal energy, respectively. 7 shape modes and 7 form modes cover 90% of the cumulative modal energy, whereas 30 shape modes and 28 form modes cover 99% of the cumulative modal energy.

3.4 Results and Discussion

3.4.1 Modal Decomposition

Figure 3.4 shows the cumulative modal energy of the 192 modes (snapshots) obtained from the statistical shape analysis and form analysis procedures. Recall that Form Analysis includes volume information, as well as shape information. The modal energy represents the summation of the eigenvalues up to the specified number, divided by the sum of all 192 eigenvalues produced from the modal decomposition in both tests. The eigenvectors are a ranked representation of the variance in the dataset, and the corresponding eigenvalues are ordered from highest to lowest.

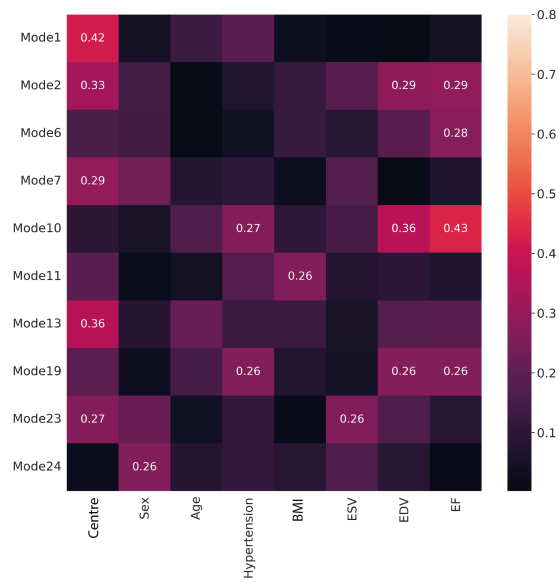
As is typical, based on the cumulative modal energy, a relatively small number of modes

are required to capture the shape variation present in the dataset of the RVES of 96 subjects at end-systole and end-diastole. Only the first 7 modes are required to capture over 90% of the cumulative modal energy and the first 30 modes capture 99% of the total modal energy. The form analysis results share a similar energy profile with the decomposition into shape modes; again only the first 7 modes are required to capture over 90% of the cumulative modal energy, and the first 28 modes capture 99% of the cumulative modal energy (Figure 3.4). The first mode in the form analysis covers a smaller percentage modal energy than the first mode in the shape analysis (36.5% vs. 42.5%).

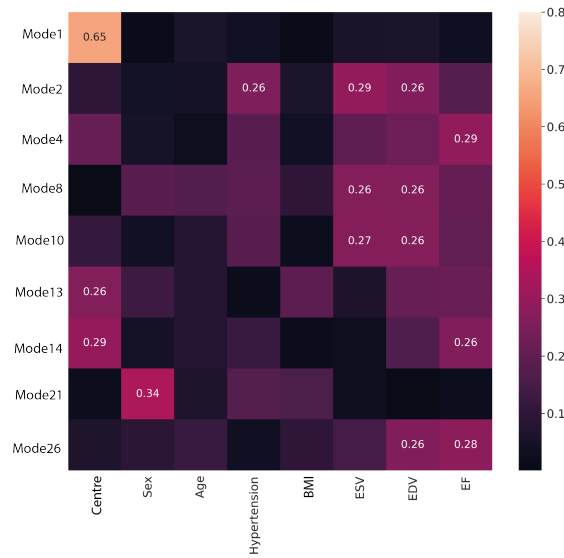
Heimann and Meinzer (2009) comment that a good model should be able to capture 90% of the total variance in the training set with less than 12 modes. The objective of the shape analysis is to search for shape features that uniquely describe specific subject characteristics, such as age and sex, regardless of their importance to the overall reconstruction of the shape. Therefore, it was necessary to consider more modes than strictly needed to accurately reconstruct the ventricle shape, even if such modes represent a relatively small component of that overall shape. In this study, the first 30 modes were selected for further evaluation. Compared with previous statistical shape analysis procedures performed by this group (Xu et al., 2021) which required just 15 modes from 100 snapshots to capture 99% of the shape information, important shape features are more widely distributed across the earlier modes. This delocalisation or dilution of features is perhaps related to the two different scanners and centres which make up the dataset. Variance in scanner settings, could result in a separate set of shape features being identified for each half of the dataset.

3.4.2 Modal Coefficient Correlation with Traits

The heatmaps in Figure 3.5 present key Pearson correlation results when shape modal coefficients are correlated against subject characteristics. Modal coefficients are derived from Equation 2.6, and represent the ‘amount’ of a given shape mode (eigenvalue) contained within any given individual’s RV. The absolute correlation is given, and only rows containing statis-



(a)



(b)

Figure 3.5: **a)** The absolute values of the magnitude of the Pearson correlation coefficient between each of the first 30 modal coefficients corresponding to the end-Diastole shape and subject characteristics. **b)** The absolute value of the magnitude of the Pearson correlation coefficient between each of the first 30 modal coefficients corresponding to the end-Systole shape and subject characteristics. Exact Pearson values are reported for correlations that exceed 0.2. Only rows containing statistically significant values have been included in the analysis.

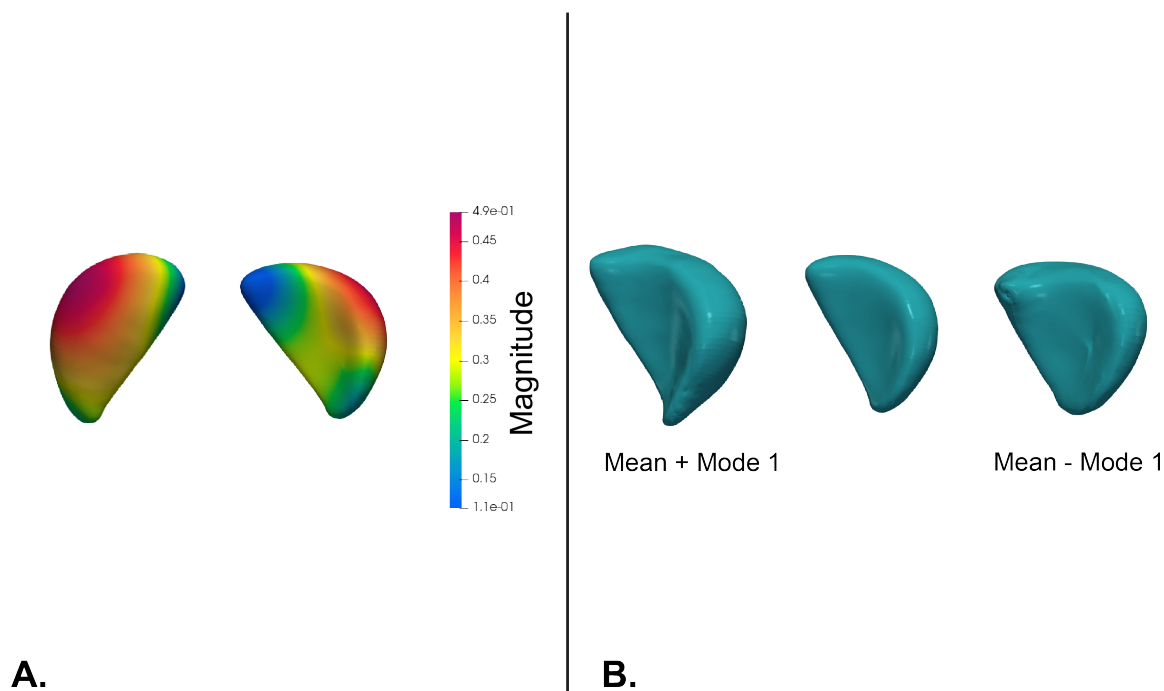


Figure 3.6: **A.** shows the magnitude of Mode 1 values superimposed on the mean RV (in both free-wall and septal view). Red areas indicate areas with the greatest motion and blue areas indicate areas with no motion. **B.** shows the mean RV shape $\pm 2 \times SD$ of the modal coefficient of Mode 1 ($x\mu_{ki}$) added and subtracted from the mean. This mean is a composite of end-Systolic and end-Diastolic RVs, and therefore contains phase information.

tically significant results are reported for ease of interpretation. Exact coefficient scores are reported for coefficients which exceed 0.2. Several Modal Coefficients correlated with subject characteristics: for example, Mode 10 correlates with ejection fraction and end-diastolic volume at end-diastole ($r_{94} = 0.36$ and $r_{94} = 0.43$, respectively) and Mode 21 correlates with Sex at end-Systole ($r_{94} = 0.34$). Mode 1 coefficient correlates with subject dataset only. This contrasts with the other modes, which all associate significantly with a much wider array of factors as might be expected, given the heterogeneous nature of ventricle shape.

Mode 1 is more strongly correlated to dataset than any other factor tested at both end-Systole ($r_{94} = 0.419$, $p = 0.000$) and end-Diastole: ($r_{94} = 0.654$, $p = 0.000$) (Figure 3.5). This finding, in conjunction with the fact that the only other significant association with Mode 1 is a weak positive correlation with height ($r_{94} = 0.232$, $p = 0.024$) (a factor which itself does not significantly correlate with dataset), suggests that the relationship is largely explained by

factors related to the analysis procedure. Figure 3.6 shows the mean right ventricle produced from this dataset and the mean RV \pm the modal coefficient of Mode 1 multiplied by two times the standard deviation of coefficients. This method of reporting shape change is beneficial, because the modal coefficient is a measure of the ‘amount’ of Mode 1 experienced by each RV in the dataset. Therefore, using this value $\pm 2 \times SD$ gives an accurate representation of the shape features encoded by that mode, exaggerated for visualisation purposes in a way that aligns with the variance in the dataset (see Section 2.9 for a more detailed description). It can be observed qualitatively from this figure that increasing the Mode 1 coefficient is congruent with an elongation of the right ventricle, and with an increase in captured detail at the apex of the ventricle.

The scanning protocol in Newcastle used 8mm slice thickness, skipping 0mm between slices, whereas the Pittsburgh scanner protocol was 6mm skipping 4mm. Figure 3.7 indicates that RVs from the Newcastle dataset have a higher Mode 1 coefficient (mean = 20.2, SD = 4.86) than subjects in the Pittsburgh dataset (mean = -8.87, SD = 4.38). This information, together with the fact that adding Mode 1 to the mean RV leads to greater detail at the apex, and a longer, thinner body overall, whereas removing Mode 1 results in a rounder RV with less detail at the apex, suggests involvement of MRI scanner and/or scanning protocol. The continuous scans performed on the individuals in the Newcastle dataset (i.e. with no skipped geometry between slices), would ensure that the apex is captured in the scanning window in a greater number of subjects. This is also suggested by the longer shape, as more of the ventricle is covered. The longer RV when Mode 1 is added to the mean (depicted in Figure 3.6) also explains the weakly significant correlation with height, as a taller subject possibly suggests a longer RV.

As the relationship is explained by the experimental procedure, no further conclusions can be drawn about the impact of nationality or related anthropological measures on the shape of the RV. Moreover, as shown in Figure 3.8, the strength of the relationship is such that it is also the highest ranked shape mode in a form analysis, where heart volume was not controlled in

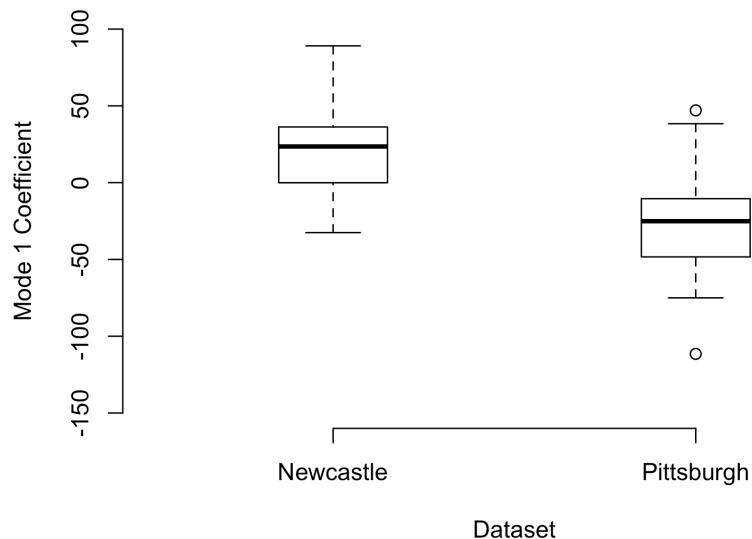


Figure 3.7: A box plot showing the distribution of Mode 1 coefficients for the end-Diastolic RVES shape between subjects scanned in Newcastle and those scanned in Pittsburgh. RVES from the Newcastle cohort exhibited larger, positive Mode 1 coefficient values, whereas Pittsburgh RVES exhibited lower, negative coefficients. The horizontal bars represent the mean.

the analysis.

Having established one strongly significant subject-related factor from the Pearson Correlation coefficient analysis, there were a series of other notable correlations: for example Mode 10 with volume-related measures at end-Diastole, and Mode 21 with sex at end-Systole, although none of these were as singly nor as strongly associated with one subject-related factor. Clearly, interactions between several shape modes are of importance. To further explore these promising but complex, shape-based correlations, the next step was to use LDA to produce a function which maximises the separation between dichotomous groups using combinations of the modes. It is important to note that the higher modes contain very little energy (often less than 1% of the cumulative energy), and although they may contain a clear shape signal, their variance may be less than other uncertainties, for example the accuracy of the initial segmentation.

Exploratory Pearson correlations were performed using RVs from all 96 subjects included

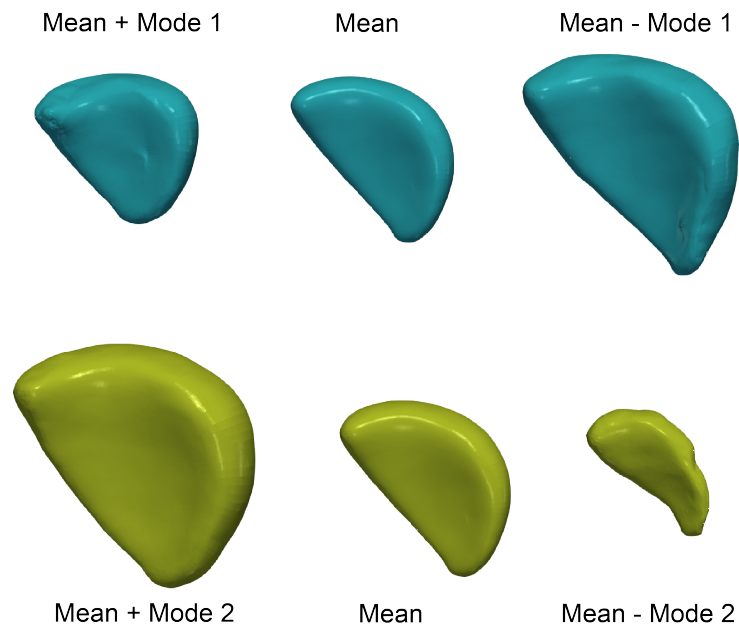


Figure 3.8: The first two shape features produced as output from Form Analysis, where the RVES shapes were not scaled. The top (blue) shows Mode 1 results from Form Analysis, where the left hand is the mean RVES shape + $2 \times SD$ of modal coefficients and the right hand side shows the mean - $2 \times SD$ of the modal coefficients. The bottom (green) depicts Mean Mode 2 result from Form Analysis. The left hand side shows the mean + $2 \times SD$ of Mode 2 coefficients and the RHS shows mean - $2 \times SD$ of Mode 2 coefficients. Both mode effects are shown in the septal view. Mode 2 is clearly a volume related mode, and Mode 1 is a scanner related to shape feature.

Table 3.3: Classification scores (% accuracy) of the best overall mode combinations for separating BMI, age and sex categories. The best combination of modes, identified using LDA, is reported for each characteristic and cardiac phase. The percentage accuracy after cross-validation is reported in the ‘Score’ column. The ‘Hyp’ column notes the percentage accuracy when the same LDA discriminant function, used to generate the accuracy score, was applied to hypertensive subjects only, and the ‘Applied Modes’ column reports the percentage accuracy of a new LDA, conducted on the hypertensive individuals.

Characteristic	Phase	Best Combination	Score	Hyp	Applied Modes
Sex	Diastole	6, 23	72.5	53.6	60.7
Sex	Systole	19	77.5	53.6	42.8
Obesity	Diastole	20, 11, 2	77.5	54.5	61.8
Obesity	Systole	15, 16, 6; 15, 29, 16 & 16, 13, 20	72.5	47.2	36.4
Age	Diastole	6, 3, 13; 6, 3, 2 & 6, 13, 17	75.0	44.7	64.3
Age	Systole	9, 5 & 9, 18	72.5	48.2	57.1

in the statistical decomposition. However, to clearly identify whether a shape signal exists without any co-morbidity associated with hypertension, LDA was applied to non-Hypertensive subjects in the first instance. Having examined whether population-based modes impact RV shape, the discriminant function was then applied to hypertensive individuals, in order to determine whether any population based shape differences were also observable in diseased subjects.

3.4.3 Linear Discriminant Analysis

Relationship of Sex to Ventricular Shape

Table 3.3 reports the best overall combinations, in terms of percentage accuracy, of modes (LDA features) to classify non-hypertensive subjects based on the three main characteristic groups. As referenced in the Methodology (Section 3.3), all subjects were decomposed together in the initial POD, but the resultant mode coefficients were assessed separately both for end-Systolic and end-Diastolic RVs, and for non-hypertensive and hypertensive groups. The best overall mode combination for separating 40 non-hypertensive subjects by their categorised sex at end-Diastole used Modes 6 and 23, as determined by the ability of an LDA model constructed from these features to separate the groups. This combination was able

to separate the sexes with 72.5% accuracy after cross-validation. Table 3.3 also shows that applying the same linear discriminant function (also referred to as model) generated using these modes to the remaining 56 hypertensive individuals achieved 53.57% accuracy in classifying hypertensive subjects at end-Diastole. To further isolate the impact of specific shape features, independent of adverse remodelling associated with hypertension, the model was subsequently applied to the two hypertensive sub-categories (subjects with PH and those with generalised hypertension). In this case, the model was 52.17% accurate when separating 23 hypertensive subjects and 54.54% accurate when separating the 33 individuals with PH. Thus, there is no difference in the model's aptitude for assessing pulmonary hypertensive subjects as opposed to generalised hypertensive individuals.

At end-Systole, Mode 19 alone was best classifier of sex and able to separate 77.5% subjects into the correct categories. Applying the cut-off to hypertensive individuals achieved 53.6% separability. Creating a new model directly on the 56 hypertensive subjects was less successful still; separating just 42.8% of individuals. When the original model was assessed on the two sub-categories of hypertensive RV, it was apparent that Mode 19 was most effective at separating the 23 individuals from Newcastle: 65.2% of these subjects were successfully separated using the model, whereas only 45.5% of individuals with PH were successfully separated. These facts taken together point to the fact that Mode 19 is especially prevalent and able to separate individuals of different sexes in Newcastle individuals, but that this effect is not observed to the same extent in subjects from Pittsburgh. As the modes are ranked in order of their contribution to the total shape energy, Mode 19 alone contributes just 0.16% of that total modal energy. Nevertheless, its over-emphasised importance to RV separation by sex, suggests that this mode in particular is a strong contributor to that shape.

Figure 3.9 a) shows the combined qualitative impact of Modes 6 and 23 on the shape of the RV. This visualisation is produced from the decomposition of the entire 96 subject dataset, including diseased subjects. Of these, the left-most RV is composed of the mean heart plus the combined coefficients of a subject that is at an extreme LDA function value (in other words,

‘the most female’) and the right RV exhibits the extreme LDA function value for male subjects. The combined effect of Modes 6 and 23 indicates a rounding out of the ventricular free wall in women, and a flattening of the same region in men. In particular, end-Diastolic male hearts show a flattened basal surface, and then a dipped region in the free-wall (see Figure 3.3 for a depiction of these regions). There is also a notch at the apex in females. The rounding effect observed in male individuals is also seen in diseases such as PH (Wu et al., 2012), whereas this effect is not as noticeable in female RVs. The shared effect may mask sex-based differences in hypertensive subjects and explains why separation is less successful when applying the model to hypertensive individuals. Analysis in Chapter 5 of this thesis further explores the relationship between hypertension and sex. The RHS of Figure 3.9 shows the impact of Mode 19 on RVs in the two categories. The left of these images indicates the qualitative effect of adding the Mode 19 coefficient of the female RV with the greatest absolute LDA discriminant function score to the mean RV, whereas the right image shows the impact of adding the mode 19 coefficient of the male RV with the greatest absolute discriminant score to the mean. It is notable that, at end-Systole, it is the female RV that experiences notching, and that this occurs further down the free-wall, away from the base, whereas the male RV exhibits the triangular bulge shape. The seemingly ‘swapped’ shapes could be indicative of an earlier contraction phase in female RVs.

Table 3.3 demonstrates that, when the same modes were used to create a new LDA model specifically trained on hypertensive subjects, the separation was somewhat more successful. This model was able to separate subjects with 60.7% accuracy. These findings suggest that Modes 6 and 23 are able to separate subjects when hypertension is introduced as a factor, although not as successfully as when trained only on non-hypertensive individuals. This result is unsurprising given the clear shape features associated with hypertension (observed in Xu et al., 2021), which are likely to overpower any sex-specific impacts. Finally, the same algorithm was used to identify Modes and create a new LDA model, specifically trained on all 56 hypertensive subjects. The overall best separator of sex at end-Diastole used Modes

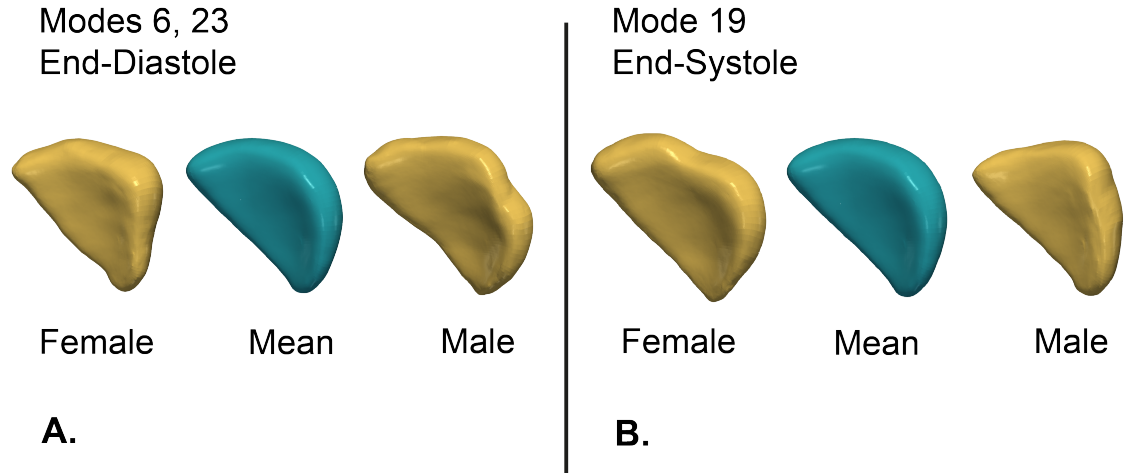


Figure 3.9: *A. the qualitative impact of Modes 6 and 23; the best modes for separating non-hypertensive RVES shapes by sex at end-Diastole, as determined by the LDA analysis. B. the impact of Mode 19; the best mode at end-Systole, as determined by LDA. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a female RV plotted on the mean (which includes healthy and diseased subjects) and the right side shows the same but for a male individual. The RVs are orientated with the septal wall facing out of the page.*

17, 24, 13 and 8. which was 76.8% accurate after cross-validation. Figure 3.10 b) shows that the best combination for splitting hypertensive subjects results in increased septal curvature in female subjects, and that male individuals exhibit slight septal flattening and also flattening in the mid to basal region of the free wall.

Comparing the modes that best separate sex in non-hypertensive individuals with those that best separate in those who are hypertensive indicates that the shape of the female ventricle is broadly similar in both cases, but that the male heart exhibits a flatter septal wall in hypertensive right ventricles.

The overall best mode combination for separating the 56 hypertensive individuals at end-Systole used Modes 6, 17 and 3 and was 75.5% effective after cross-validation. Figure 3.10 shows that in hypertensive RVs, female individuals have a noticeably rounder free wall region compared to male individuals. This region is central between base and apex.

Figure 3.11 displays the overall mean RV, the mean male RV and the mean female RV. This figure is produced with the intention of comparing the statistical shape analysis reconstructions to real life counterparts. Here, the mean male RV (encompassing both systolic and

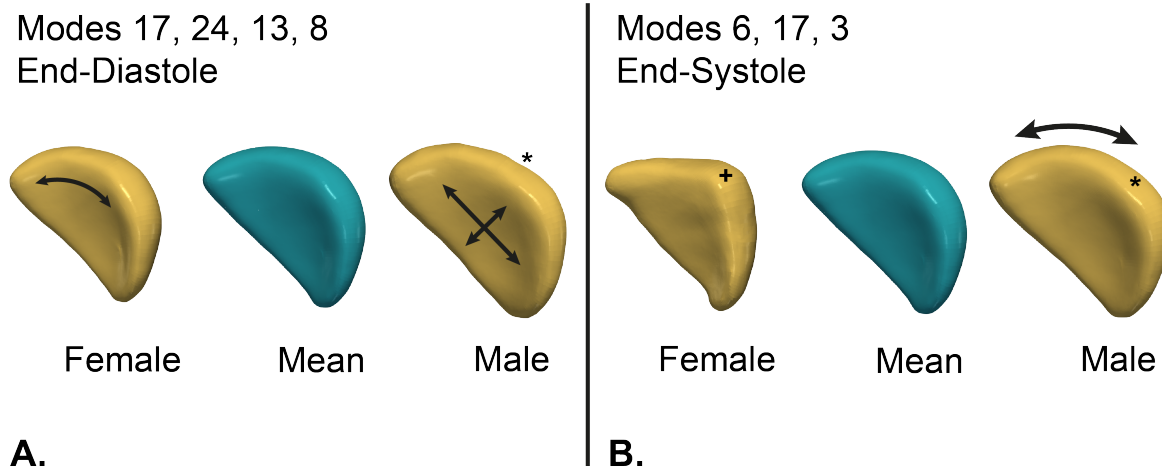


Figure 3.10: **A.** the qualitative impact of Modes 17, 24, 13 and 8; the best modes for separating the 56 hypertensive RVES shapes by sex at end-Diastole. The female RV at end-Diastole exhibits greater rounding of the septal wall, indicated by the double-headed arrow, whereas a flattening is seen in the same location in male RVs, which is illustrated by the two crossed arrows. Basal flattening in the male RV is also indicated by an asterisk. **B.** the impact of Modes 6, 17 and 3; the best modes at end-Systole. At end-Systole, the female RV exhibits basal bulging, marked by '+', and the male RV displays basal flattening marked by an asterisk, and emphasised by the arrow. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a female RV plotted on the mean (which includes healthy and diseased subjects) and the right side shows the same but for a male individual. The RVs are orientated with the septal wall facing out of the page.

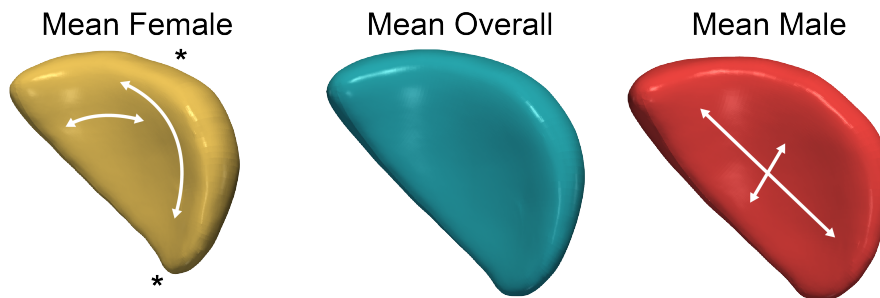


Figure 3.11: The mean RVES shape of all female RVs (in yellow), the overall mean RVES shape (in blue), and the mean male RVES shape (in red). These qualitative depictions contain both end-Systolic and end-Diastolic shape information. The mean female RVES displays greater septal wall curvature, whereas the mean male shape is flatter in the same region.

diastolic shapes) and the mean female heart were separately plotted to examine notable variations in shape features across the groups as a whole. The annotated plots show the same shape patterns as those illustrated in the statistical shape analysis, thus adding credence to the statistical shape analysis methodology in determining key features associated with demographic descriptions. Figure 3.11 clearly illustrates the flatter septal wall shape in male RVs

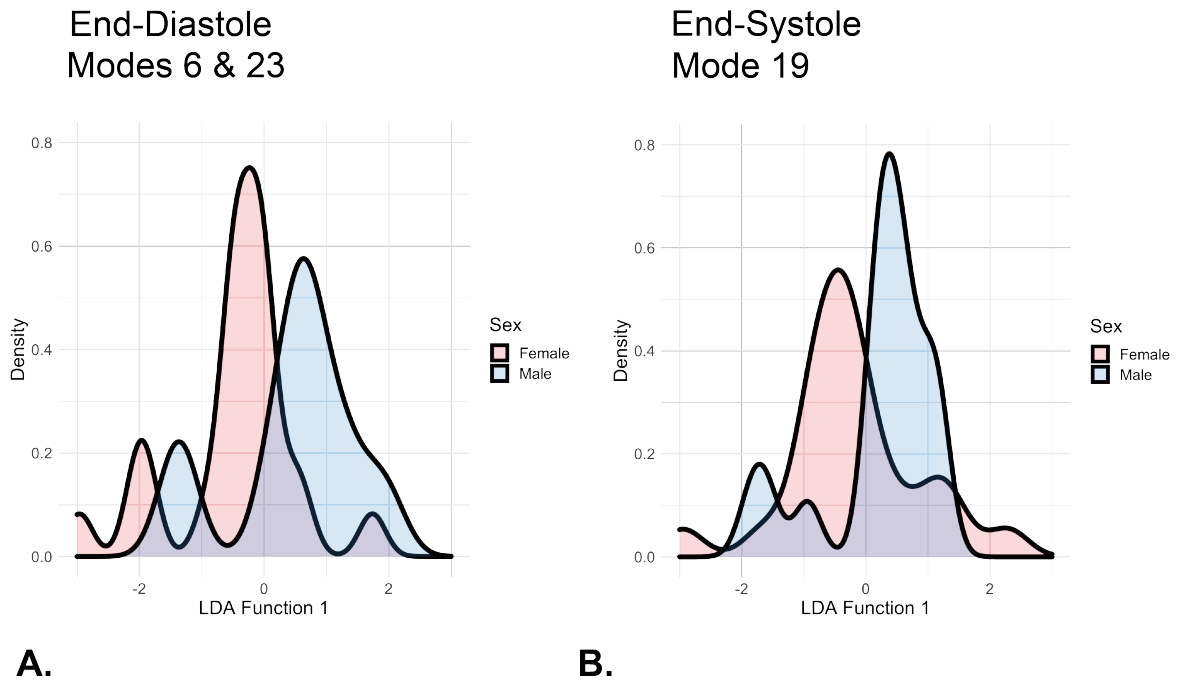


Figure 3.12: Density plots showing the separation of LDA function values between male and female subjects. Female RVES shapes are depicted in pink, and male RVES in blue. **A.** shows the function value density at end-Diastole (using Modes 6 & 23) and **B.** shows the function value density at end-Systole (Mode 19). At both cardiac phases, there is a clear split peak between the sexes.

compared to female RVs, and the mean female right heart displayed a smaller overall mean shape (despite the initial registration algorithm controlling for volume), which indicates more that the male ventricle bulged out in these areas, and also exhibited a slight flattening, or “drop off” in the basal region. Conversely, male hearts have a slight bulge in this same region. The effect can also be seen in the density plots at end-Systole and end-Diastole (Figure 3.12). In both cases, there is clear separation between the peak female LDA function score and the peak male function score, with the male score being largely concentrated above 0 and the female score largely concentrated below 0. At end-Diastole (Modes 6 & 23), the female peak density is greater indicating a smaller distribution of female LDA function values, and at end-Systole (Mode 19), the male peak density is greater.

All non-hypertensive subjects can be separated with over 70% accuracy, which indicates that there is a clear difference in RV shape between males and females. This evidence is

further corroborated in Figure 3.11 where the effect of subject sex on ventricular shape can be observed qualitatively.

Relationship of Obesity to Ventricular Shape

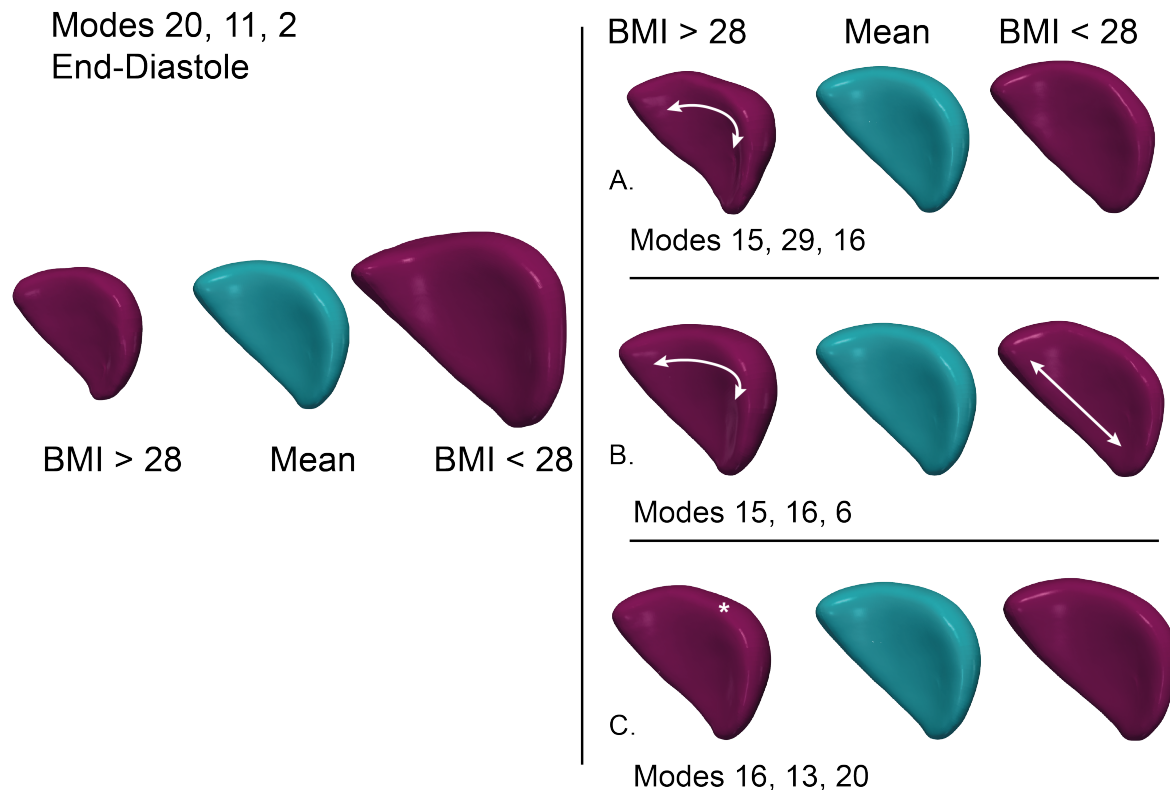


Figure 3.13: 3D plots of the mean RV and the effect of obesity-related modes on RV shape in non-hypertensive RVs. **A.** the qualitative impact of Modes 20, 11 and 2; the best modes for separating RVES shapes by obesity at end-Diastole in non-hypertensive subject. The RHS shows that three combinations of Modes, **A. B. C.** separated by obesity at end-Systole with equal success when all combinations of the top five modes were tested. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a subject with BMI>28 plotted on the mean (which includes healthy and diseased subjects) and the right side shows the same but for subjects with BMI<28. At end-Diastole, the obese RVES shapes exhibit significant hypertrophy. At end-Systole, the non-obese RVs show greater septal curvature. The RVs are orientated with the septal wall facing out of the page.

Individuals were separated based on a BMI cut-off of $28\text{kg}/\text{m}^3$. This value falls at the centre of the “clinically overweight” category (Eveleth, 1996). The best combination for separa-

tion at end-Diastole was Modes 20, 11 and 2, which was able to separate subjects with 77.5% accuracy after cross-validation (Table 3.3). Successful separation fell to 54.5% when applying the same model to 56 subjects exhibiting hypertension. Sub-divisions based on obesity show that the 23 generalised hypertensive individuals were almost as successfully separated (72.7%) and subjects with PH were separated with 42.4% accuracy. Therefore, the combination of modes trained on BMI in non-hypertensive individuals were much more successful at separating individuals in the Newcastle cohort. Figure 3.13 shows that the most extreme LDA function based on the combination of Modes 20, 11 and 2 caused a significant free-wall expansion. Using these modes to create a new model, trained directly on hypertensive individuals was more successful, and correctly identified 61.8% of subjects (Table 3.3), which suggests that the modes may interact in different ways when applied specifically to hypertensive subjects.

The best combination identified when applying the algorithm directly to the hypertensive individuals was Modes 6, 30, 12, 11, which accurately classified 78.2% of individuals after cross-validation (Table 3.3). Mode 11 is common amongst both groups. The most extreme obese subject by this marker is part of the Newcastle group, which exhibited a lower mean BMI overall (Table 3.1).

At end-Systole, three mode combinations (Modes 15, 29 and 16; Modes 15, 16 and 6 and Modes 16, 13 and 20) separated individuals by their BMI category with equal success. When applying the model to 56 hypertensive subjects, the combination of Modes 16, 13 and 20 performed the best, albeit with only 47.2% accuracy. When applying this combination to more refined sub-categories (i.e. to the generalised hypertensive subjects and those with PH), the model correctly classified 45.4% of generalised subjects and 48.5% of PH subjects. Creating a new discriminant function using Modes 16, 13 and 20 also had a poor success rate, classifying just 36.4% of individuals after cross-validation. Clearly, the obesity-related shape signatures at end-Systole are masked in hypertensive individuals, whereas the unique features exhibited at end-Diastole are maintained even in subjects showing symptoms of hypertension. There

is no direct physiological reason why RV shape at end-Systole should correlate with BMI, excepting obesity-related structural changes. It could also therefore be argued that the lack of classification power reflects an overall lack of difference in shape in this cardiac phase. However, the presence of shape-related differences in the non-hypertensive group instead points towards complex interactions between pathological shape change and that associated with obesity, which could be explored further with a larger sample size.

The qualitative impact of higher or lower BMI on individuals is shown in Figure 3.13. Again, the modes multiplied by coefficients of most characteristic individuals in each category are plotted on the mean heart, giving a graphical indication of the combined effect of the identified shape features. At end-Diastole, Modes 20, 11 and 2 were the key separating modes. In the overweight category, the most characteristic overweight individual exhibits a clear rounding out, or hypertrophic effect. This roundness is evident throughout the ventricular free-wall, and makes the shape appear ‘larger’, although this is an illusion, as all ventricular meshes were registered against a template shape to remove any scaling effect. Chahal et al. (2012) reported a similar effect, with increased RV mass and volume in overweight individuals, even after controlling for heart size. At end-Systole, the representative individuals with a BMI > 28 clearly showed increased septal curvature, as compared to the mean. This was most noticeable in the combination of Modes 15, 29 and 16, and Modes 15, 16 and 6. This effect matches that observed in (Mauger et al., 2019), where the sphericity shape mode was found to increase in subjects with obesity and angina.

Table 3.3 indicates that the best overall combination of modes for separating hypertensive individuals at end-Systole was: 22, 28 and 4. This combination was able to separate 80% of the hypertensive subjects after cross-validation. Interestingly, this case led to a rounder ventricle in the subject with a BMI below 28 and a thinner RV in the overweight subject. As this is an end-Systolic effect, the relationship could possibly be attributed to relatively improved systolic function. The qualitative shape change produced by these modes is displayed in Figure 3.14.

Figure 3.15 gives a qualitative representation of the mean overall RV, as well as the mean

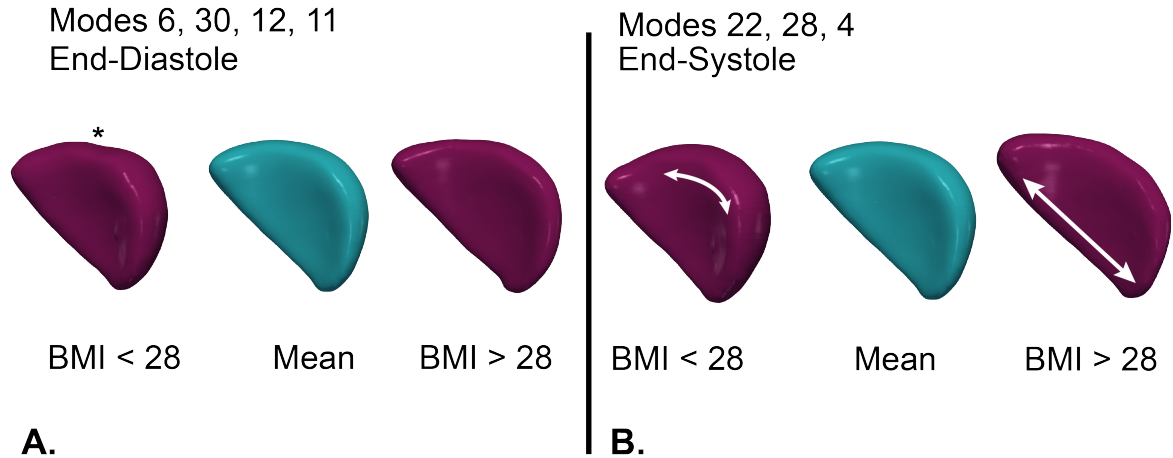


Figure 3.14: **A.** the qualitative impact of Modes 6, 30, 12, 11; the best modes for separating the 56 hypertensive RVES shapes into obesity groups at end-Diastole. **B.** the impact of Modes 22, 28 and 4; the best separating modes at end-Systole. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a non-obese subject, plotted on the mean (which includes healthy and diseased subjects), and the right side shows the same but for an obese individual. The RVs are orientated with the septal wall facing out of the page.

non-obese and the mean obese right ventricles. The mean obese heart clearly shows a similar effect to the combined Modes and coefficients 20, 11 and 2. Again, the shape exhibits significant rounding around the base and free wall. Whereas the mean non-obese heart more closely resembles the greater septal curvature effect observed in the end-Systolic hearts.

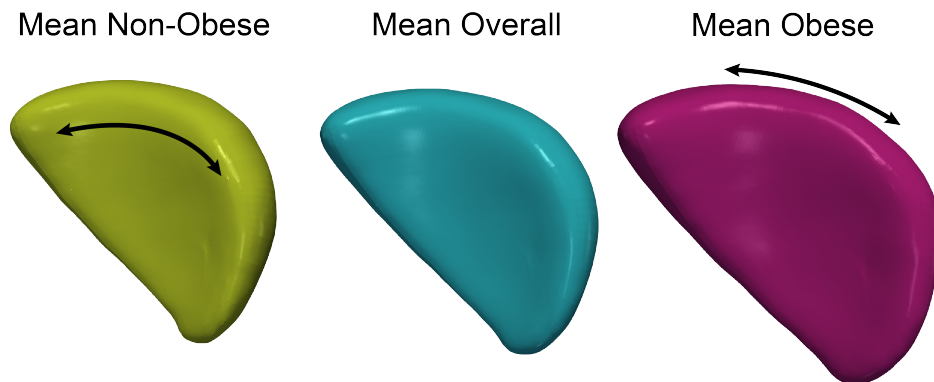


Figure 3.15: The mean RVES shape of all subjects with a BMI less than 28 (in green), the overall mean RVES shape (in blue) and the mean RVES of subjects with a BMI greater than 28 (in purple). These qualitative depictions contain both end-Systolic and end-Diastolic shape information. The mean non-obese RVES exhibits greater septal curvature, whereas the mean obese RVES shows evidence of hypertrophy.

Figures 3.16 and 3.17 depict the density of LDA function scores achieved by the classified

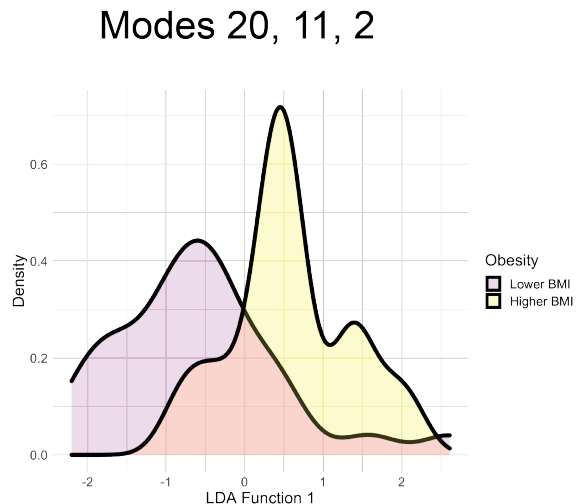


Figure 3.16: Density distributions for the separation of LDA function values between obese and non-obese individuals at end-Diastole. BMI values below 28 are shown in pink, and BMI values greater than or equal to 28 are shown in yellow.

individuals. It should be noted that these reflect the subjects as classified by the algorithm, and not just the correctly classified RVs. At end-Diastole (Figure 3.16), there is a greater density of similar LDA function values in the higher BMI category, whereas there is a greater spread of function values in the lower BMI group. At end-Systole, Modes 15, 29 and 16, the most concentrated peak is in the lower BMI group.

The greater number of different modes that are able to separate obesity amongst sub-categories suggests that there are many shape features impacted by weight and that it may be impossible to segregate specific features using a linear method. This may be resolved by creating a non-linear classifier, however, it also seems intuitive that there are many other pathologies, inter-related with BMI that could be obscuring any given linear signal. From Figure 3.13 it can be observed that there is a common curvature (increased concavity) of the septal wall in subjects with a BMI below 28, and a flatter (“D-shape”) in overweight subjects. The same effect is noticeable when plotting the effect of the three top Mode combinations.

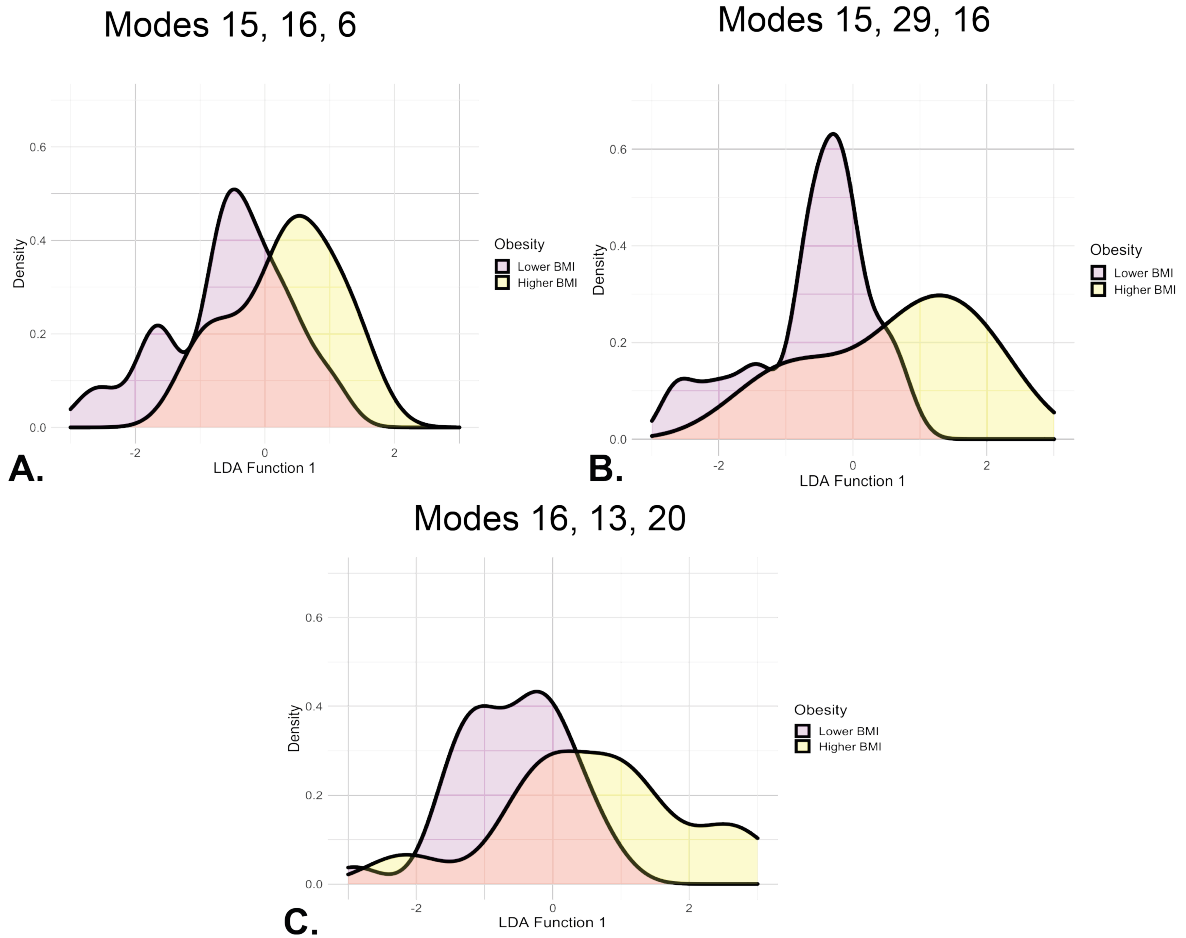


Figure 3.17: Density distributions for the separation of LDA function values between obese and non-obese individuals at end-Systole. BMI values below 28 are shown in pink, and BMI values greater than or equal to 28 are shown in yellow. The top left image shows Modes 15, 16 and 6, the top right image shows Modes 15, 29 and 16 and the bottom image shows Modes 16, 13 and 20. Modes 15, 29 and 16 shows a concentrated peak of LDA values between -1 and 0 for non-obese individuals, but a disperse distribution of LDA values in obese individuals.

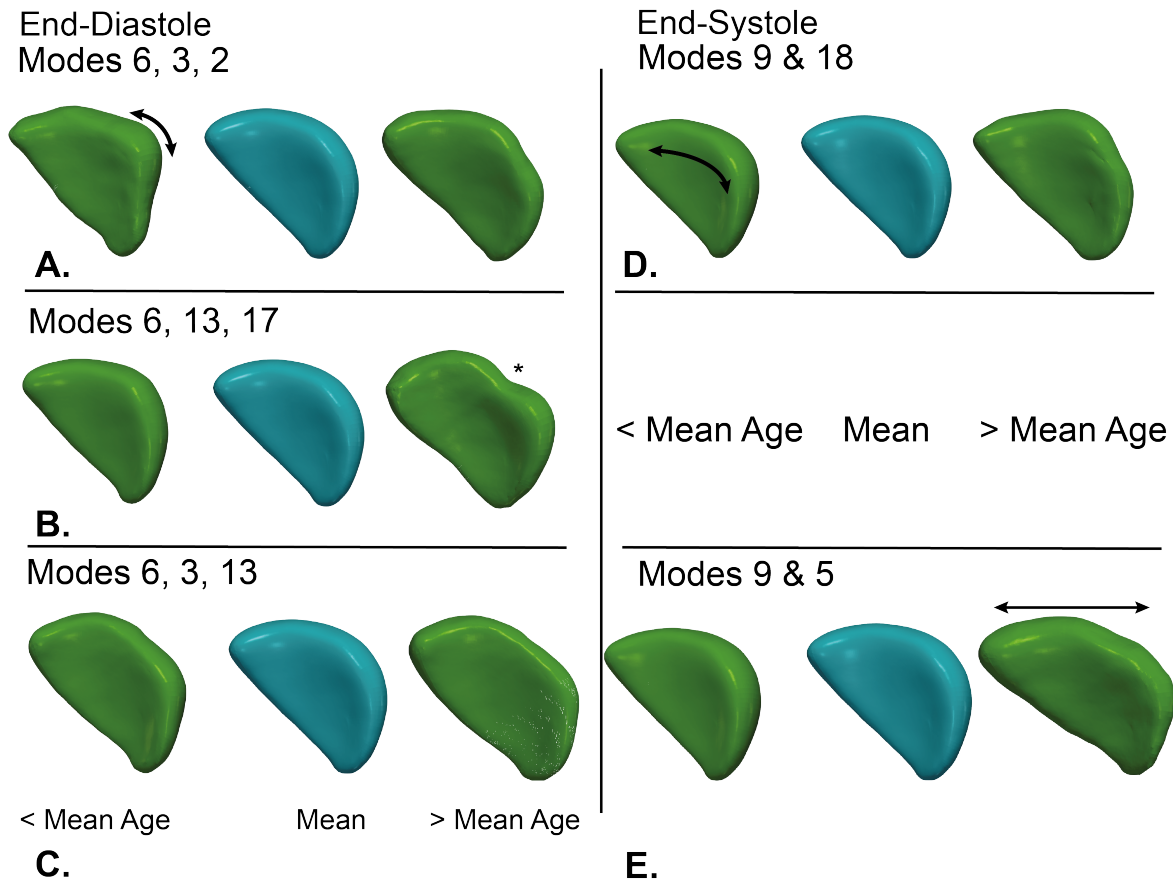


Figure 3.18: *A.* the qualitative impact of Modes 9 and 5 and Modes 9 and 18; the joint two best mode combinations for separating non-hypertensive RVES shapes into groups based on age at end-Systole. *B.* three combinations of Modes are equally good at separating RVES by age at end-Diastole. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a subject younger than the mean age of 60.5, plotted on the mean RV (which includes healthy and diseased subjects), and the right side shows the same but for subjects older than the mean age. The RVs are orientated with the septal wall facing out of the page.

Relationship of Age to Ventricular Shape

At end-Diastole, three Mode combinations were equally successful at separating individuals into groups based on age. These were: Modes 6, 3, 2; 6, 13, 17 and 6, 3, 13, which were able to classify young and old subjects with 75% accuracy (Table 3.3). The equal weighting of three separate mode combinations suggests that there is a complex relationship between ageing and shape at end-Diastole and that it cannot be easily isolated using a linear classifier. Nevertheless, the high initial separation accuracy demonstrates that population-based shape differences do exist. The same subject was identified by all models as being the exemplar

older RV, i.e. that this specific RV had the highest linear discriminant score in all three cases, an indication that it was the ‘most typical old’ RV.

The same inputs and discriminant cut-off, created using the non-hypertensive individuals was then applied to the hypertensive subjects. All modal coefficients were produced during the initial POD, but in this case, only coefficients from the 56 diseased individuals were included in the analysis. The best of the combinations was Modes 6, 13 and 17, although it correctly classified only 44.7% of hypertensive subjects (Table 3.3). The success was fairly evenly split across the two centres; 47.8% Newcastle subjects successfully identified, and 42.4% Pittsburgh subjects. Figure 3.18 shows that the three different mode combinations have varying effects on the shape of the RV, including increased septal curvature in younger subjects, a widening of the RV in older subjects, and basal flattening in older subjects. Creating a new model using Modes 6, 3 and 17 produces a classifier that can separate 64.3% of subjects by age after cross-validation. The best combination at separating hypertensive subjects at end-Diastole was Modes 4, 13, 9, which correctly separated 73.2% of individuals after cross-validation. Figure 3.19 suggests that there is a slight, but insubstantial increase in septal curvature in non-hypertensive younger subjects.

At end-Systole, two mode combinations were equally good at separating the 40 non-hypertensive hearts into two average age groups. These were Modes 9 and 5, identified through building up the best combinations from the Newcastle/Pittsburgh subgroups and Modes 9 and 18, identified from applying the preliminary LDA algorithm directly to the 40 subjects. Both combinations successfully separated 72.5% of individuals. Neither combination was successful when the model was applied to the 56 hearts at end-Systole. Modes 9 and 5 successfully categorised just 37.5% of individuals. Modes 9 and 18 were more successful, linearly separating 48.2% individuals into age categories, but both remained worse than chance. Modes 9 and 5 were able to separate 53.6% of individuals into the correct age categories when the model was specifically trained on that group. Modes 9 and 18 similarly were able to separate 57.1% of individuals into the correct age categories. Modes 9 and 18 are 45.5%

successful at separating the PH subgroup and 52.1% effective at separating the generalised hypertensive group. On the other hand, Modes 9 and 5 only separated 39.1% of individuals with generalised hypertension and 36.4% of individuals with PH. It is clear from this analysis that Mode 9 plays a large part in the separation of sub-groups at end-Systole in non-hypertensive individuals. The two pairs of modes separate equally successfully, but produce such different shapes at the extreme - with Mode 5 seemingly contributing to a stretch and Mode 18 to a hump.

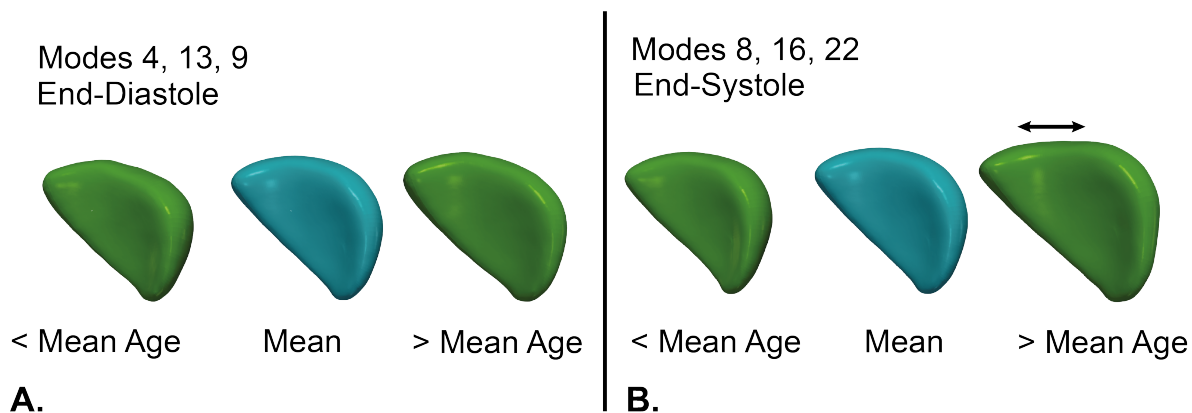


Figure 3.19: *A. the qualitative impact of Modes 4, 13 and 9; the best modes for separating the 56 hypertensive RVES shapes by age at end-Diastole, B. the impact of Modes 8, 16 and 22; the best modes at end-Systole. The left-most RV shape on both sides depicts the modal coefficients of the most extreme LDA function value for a younger subject plotted on the mean (which includes healthy and diseased subjects) and the right side shows the same but for an older individual. At end-Systole, the older individuals exhibit lengthening of the RV basal region, which is marked by the double headed arrow and also show generalised flattening in that area. The RVs are orientated with the septal wall facing out of the page.*

The best mode combination, assessed by building up from the hypertensive sub-groups, was Modes 8, 16 and 22, which successfully separated a cross-validated 71.4% of individuals by age. The older subject with the most extreme LDA function value was incorrectly classified, which indicates an imperfect classification function for assessing subject age. Figure 3.19 shows an increase in length of the RV basal region, in addition to flattening in the same area.

To confirm the results of the shape analysis, Figure 3.20 shows the overall mean RV, the mean young RV (individuals below the mean age for the cohort, of 60.53 years), and the mean

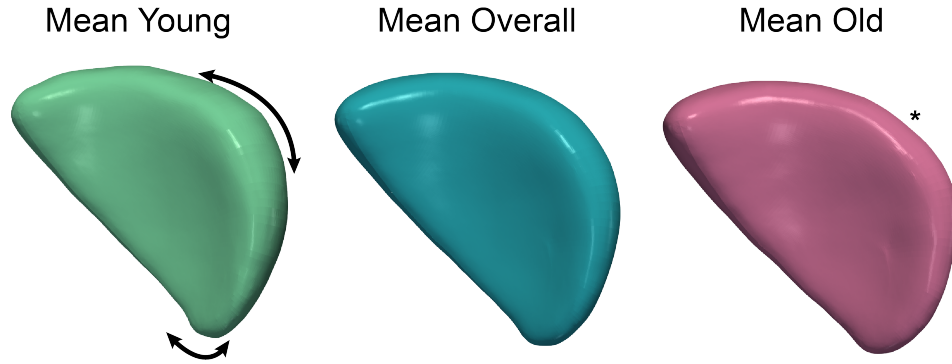


Figure 3.20: The mean RVES shape of all subjects below the mean age of 60.5 years (in mint green), the overall mean RVES shape (in blue), and the mean RVES of subjects older than the mean age of 60.5 years (in pink). These qualitative depictions contain both end-Systolic and end-Diastolic shape information. The mean younger RVES exhibits greater septal curvature, marked by the longer double-headed arrow and greater detail at the apex, marked by the shorter arrow, whereas the mean older RVES shows basal flattening, marked by an asterisk.

older RV. The mean young heart exhibits a notable increase in basal rounding, as compared to the mean old heart, and perhaps more strikingly shows apex curvature. In both Figure 3.18 and Figure 3.20, younger individuals exhibit greater septal curvature in non-hypertensive subjects than their older counterparts. Therefore, similar shape features are observable by examining sub-group means as were picked out by the LDA classifier.

The specific shape features displayed suggest that the mean young heart more often contains apex geometry that can be detected by the MRI scanner. This could indicate that the younger heart is longer. There is some evidence in the literature (Strait & Lakatta, 2012) that suggests that older hearts often display flattening of the RV free-wall, which can be observed in these qualitative images (Figure 3.18, Figure 3.20).

LDA density plots are shown in Figures 3.21 and 3.22. At end-Diastole, the clearest division of peaks was displayed by the model created by Modes 6, 3, and 2, and shows a peak density of older RVs. At end-Systole, there is an observable density peak in the below-mean-age individuals when Modes 9 & 5 are used to create the model, whereas the older individuals have function values extending across the entire range. The implication is perhaps that this particular model is able to identify younger hearts, but that detecting older individuals is more difficult.

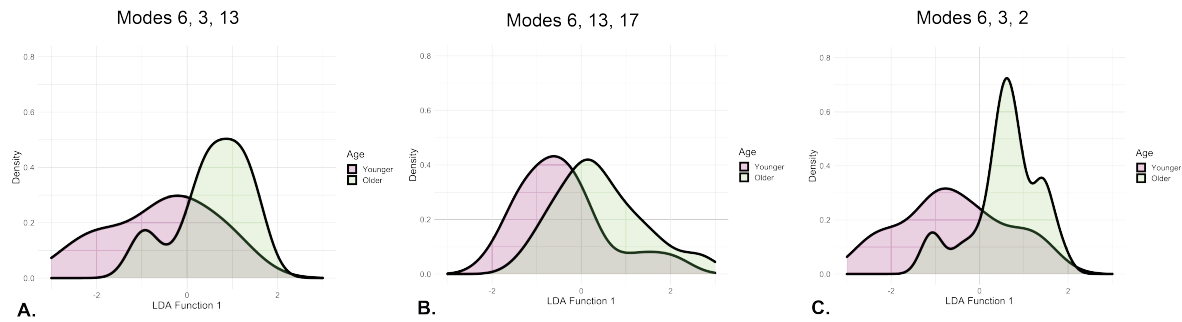


Figure 3.21: Density Plot showing the LDA function value distribution of classified subjects at end-Diastole based on whether their age fell above or below the mean. Subjects below the mean age are depicted in pink, and subjects above the mean age are depicted in green. **A.** shows Modes 6, 3 & 13, **B.** shows Modes 6, 13 & 17 and **C.** shows Modes 6, 3 & 2. The older subjects show a concentrated peak around an LDA function value of 1 when separated by Modes 6, 3 & 2, and the younger subjects appear dispersed across the range.

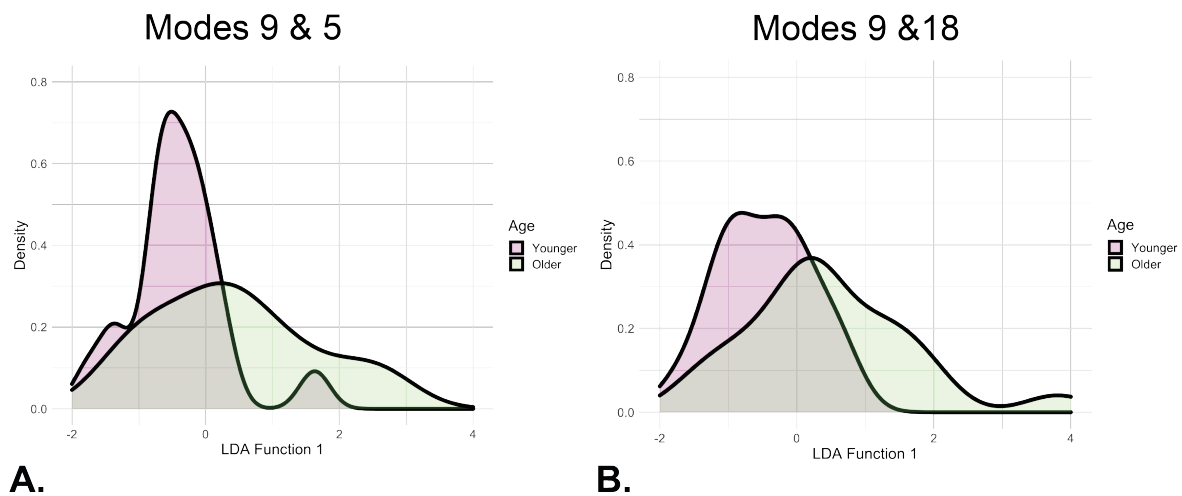


Figure 3.22: Density Plot showing the LDA function value distribution of classified subjects at end-Systole, based on whether their age fell above or below the mean. Subjects below the mean age are depicted in pink, and subjects above the mean age are depicted in green. **A.** shows Modes 9 & 5. and **B.** shows Modes 9 & 18. There is a concentrated peak of younger individuals when separated by Modes 9 & 5.

The effect observed in Figure 3.18, where the optimal end-Systolic modes (Modes 9 & 18 and Modes 9 & 5) share a common ventricular shape in the younger age group (below ‘mean’), but exhibit considerable variation in older subjects is mirrored in the LDA density plots in Figure 3.22. This implies that there is no uniform shape-trait associated with age related change and/or dysfunction, but that a series of different Modes affect individuals at different stages of the healthy and dysfunctional ageing process. The relatively small sample

size of 96 subjects means that it is difficult to isolate specific features.

3.5 Conclusions

This chapter sought to investigate whether normal, non-pathological characteristics, such as sex, age, BMI and centre affect RV shape. The first key conclusion is that it is possible to separate subjects into groups based on statistically-decomposed shape features, suggesting that unique shape features encoded in individuals exist.

The initial hypotheses were that it would be possible to differentiate between the ventricular shape of obese and non-obese individuals (BMI above and below 28) and age (above and below the cohort mean of (60.5 years), but that it would not be possible to distinguish male and female hearts, nor would it be possible to distinguish between the two demographic regions on account of there being too many confounding factors. The results presented in this chapter show that it is possible to separate hearts into two groups on the basis of sex, BMI and age using a tailored combination of shape modes and therefore indicate that these shape modes are associated with specific RV features. All of the above cases achieved a cross-validated percentage score of between 72.5% and 77.5% in both cardiac phases, which might be considered a ‘good’ score within the context of this data.

The coefficient of one mode for the end-Systole RV shape (Mode 19) was found to be associated with sex. This mode coefficient classified individuals into groups based on their sex with 77.5% accuracy, which was an improvement over any other combination of the top five shape features in the LDA, indicating that, for this dataset, the shape feature encoded by this mode is the best predictor of sex. This mode also has a relevant qualitative impact on the free wall of the right ventricle, which accords with observations in the literature. The same impact that sex has on the RV can be viewed from the mean female and mean male surfaces. Obesity can be classified with the same success rate as sex (77.5%), but using a much wider array of Modes. Three separate triplets of Modes are all equally accurate at linearly separating the

obese from the non-obese subjects. The fact that more groups were equally successful means that different combinations of individuals were captured using the linear separator, and that there are a series of different features that trained the split. It would be interesting to compare the results obtained in obese subjects to those of an athletic build. Obesity is considered a pathological condition and therefore cannot optimally assess the impact of increased RV size on ventricular shape and function in the absence of a known source of dysfunction. Whereas, those of athletic build will also have a corresponding increase in right ventricular volume, but may have other compensatory shape changes that exclude them from a blanked risk category such as might be drawn up by including shape information only.

Age is separable using two different pairs of Modes at end-Systole and three combinations of three Modes at end-Diastole. Similar to obesity, this indicates that there are specific differences in shape in ageing subjects but they are complex and difficult to isolate with a dataset of just 96 hearts. The mode coefficients at end-Diastole which best separate age when applied to a group of hypertensive subjects only were able to separate 64.3% of the dataset accurately. This implies that the RV undergoes age-specific shape changes, independent of hypertension.

One substantial benefit of statistical shape analysis that has been demonstrated in this chapter is that it is possible to identify and extract shape information associated with the method. The technique identified statistical differences associated with scanning protocol over and above any other differing factors. The most obvious shape difference between any tested group was associated with the scanning protocol (8mm skip 0mm vs 6mm skip 4mm), this demonstrates a tangible effect of scanning centre on model shape. However, the effect is constrained almost exclusively within a single shape mode (Mode 1) which means that the relevant shape information can be easily removed from the model when applied in a clinical setting. This effect is of such importance that the form analysis results, which did not control for RV volume, also produced this shape feature as Mode 1. In addition to the clinical value associated with the knowledge that although scanning information is responsible for the largest concentration of shape information, it can also be removed from the analysis. This result

demonstrates that two different data sources, with different MRI scanner protocol settings can be combined and compared with the same parameters.

Overall, applying the LDA classifier trained on non-hypertensive individuals to those with hypertension was unsuccessful at separating hypertensive subjects. The best case classifiers (Sex at both heart cycle stages and obesity at end-Diastole) were just over 50% accurate. However, when generating a new LDA model using the same key Modes for separating non-hypertensive subjects, the classification percentages differed more substantially. This indicates that the shape modes themselves can be important classifiers, but that the LDA function must be generated with the specific dataset in order to provide useful results. Here, there was a noticeable difference between the most successful classifier of hypertensive subjects; age at end-Diastole and the least successful classifier; obesity at end-Systole (64.3% vs 36.4%).

One potential problem might lie with the selected values used for classification cutoffs. These are influenced by the very high BMI values ($> 50\text{kg}/\text{m}^3$) and the young individuals (< 20 years) in the PH group. The non-hypertensive individuals may not be adequately split by this biased classification metric. Future steps will aim to link these specific shape features to outcome data so that specific pathological shape changes can be isolated. With such a small sample size, there is always concern related to overfitting the model, but the cross-validation mechanism is intended to counter this.

It is clear from Pearson Correlation coefficients and from the LDA separation that subject characteristics can have varying effects at the two main heart cycle stages. It is therefore of clinical value to assess end-Systolic and end-Diastolic shape change separately (i.e. both phases can be statistically decomposed at the same time, but statistical analysis conducted on modal coefficients from each group). Analysing the modal coefficients separately ensures that phase specific shape features are more easily isolated. Notably, the effects of obesity at end-Diastole are more apparent and are shared between hypertensive and non-hypertensive subjects, whereas there is no clear signal at end-Systole.

Chapter 4

Robustness Testing of Shape Results

Using Targeted Interventions

4.1 Introduction

Biological data can be highly variable and difficult to reproduce, even in controlled environments. Reviews into the importance of biological variation (Simundic et al., 2015; Badrick, 2021; Voelkl et al., 2020) indicate that this variability, once controlled for, gives rise to the central sources of variation which impact function, rather than peripheral noise. The shape analysis method utilised in this thesis involves a series of different steps in which additional variation can be introduced; for example, through the selection of the anatomical reference poles and datelines and through altering the size and composition of the input datasets. It is therefore, beneficial to understand the processes which contribute the greatest variability in the system and also to consider the relative benefits of reducing this variability in critical areas.

The increased availability of high-resolution *in-vivo* anatomical images in a clinical setting has led to a new generation of open-source computational tools designed to utilise these images to model structure shapes and their variability within populations. Despite the pub-

lication of many new pipelines both in a cardiac context (e.g. Mauger et al., 2019; Rodero et al., 2021; McLeod et al., 2012), and otherwise; for example Crawford et al. (2019) in glioblastoma research and Lu et al. (2013) in modelling liver damage in car crash victims, there has been remarkably little research into the replicability of such procedures. This has been noted by Gao et al. (2014) in particular, who have argued that shape analysis is a field in need of careful validation. Upon undertaking a substantive literature review, they found no reports on the evaluation and affirmation of shape analysis frameworks; although their work was largely focused on mapping human brain structures. Gao et al. (2014) proposed their own validation method, where defined deformations are applied to a series of shapes, and then the various shape analysis algorithms are applied to compute the mean shape and capture deformations. To record results, they calculated P-value maps and plotted them onto the respective mean shape. Goparaju et al. (2018) also systematically assesses the outcome of widely-used off-the-shelf SSM tools. Although recent research efforts have turned to the evaluation of methods, there is more that could be achieved, especially with regard to mapping variation in the input pipelines to the success and consistency of the output. There are potential areas for discrepancy in the MRI scans, segmentation process, the selection of appropriate boundaries for the harmonic mapping process (i.e. determining which anatomical features correspond to the top of the unit sphere). Rodero et al. (2021) comment on the impact of imaging modality on shape analysis output. They suggest that the first modes are the most likely to be consistent between different input values and even imaging modalities. This is because modes are ranked in order of the degree of shape information they explain. The first few modes are likely to be shared even in divergent inputs. Heimann and Meinzer (2009) have also previously reviewed statistical shape models for 3D medical image analysis. As part of the discussion, they identified several challenges associated with transitioning from a research to a commercial or clinical environment. Training data must be sufficiently general so that it can be applicable across a series of sites. As has been noted earlier in this thesis, many research environments train and test on the same dataset, whereas practical applications involve the combination of

a variety of sources, all of which introduce their own idiosyncrasies.

A series of different components of the analysis procedure have the potential to impact shape analysis results. Such aspects include variations in scanner protocol, individual decisions on RVES boundaries in the manual segmentation process, the selection of fiducial points on the mesh, the stringency of the smoothing filter applied to the RV mesh and the potential for randomness introduced in the numerics. Different methods may be more or less susceptible to variations due to the implementation of the various approaches. Literature relevant to each of these components will now be discussed.

Imaging Modality

First, the largest potential for disparity of results relates to the imaging modality, as well as the wide array of available image scanners and scanning protocols. It has already been demonstrated in Chapter 3 that the statistical shape analysis is sensitive to slice thickness, with more detail at the apex being identified in the 8mm skip 0mm scanning protocol, as compared to 6mm skip 4mm. Unfortunately, owing to the retrospective nature of the data collection in this thesis, more in depth robustness tests, such as repeats of the same subject in different labs or scanners could not be performed in this investigation. Some literature review is conducted, however, for the sake of completeness.

MRI reproducibility, in particular, is an active research area. As MRI has become a routine diagnostic modality, several studies aimed towards quantifying the impact of variability in scanner protocol on shape feature extraction (e.g. Raisi-Estabragh, Gkontra, et al., 2020). Raisi-Estabragh et al. (2020) assessed the repeatability of CMR images across multiple centres and scanning vendors in a test-retest study using a varied mix of cases. Test-retest studies, where the same test is performed on the same sample at a different time point, are the recommended form of reproducibility studies. Images in the study were segmented according to a specific, predetermined protocol which selected three regions of interest; the RV, the LV and the left ventricular myocardium at both cardiac phases. They extracted 280 radiomics

(voxel-level) features in total, and ranked them according to intra-class correlation coefficient. Repeatability of RV blood pool shape features ranged from moderate to excellent, with mean intra-class correlation ranging from 0.556 to 0.941. The most robust RV features were minor axis length and volume features. Jang et al. (2020) also considered repeatability of cardiac texture features, albeit focusing on the LV. They conclude that only a small subset of myocardial radiomic features are reproducible. They also comment that MRI protocols can be set up to provide reasonable contrast for differentiating between soft tissues, but imaging homogeneity still exists, which can present a challenge for automatic segmentation, in addition to that posed by the geometric complexity of the heart.

Aside from cardiac studies, a series of recent efforts have been devoted to the robustness of oncological imaging using standard test-retest studies (e.g. Zwanenburg et al., 2019). Most are based on a CT modality, which presents different challenges to MRI scanning. Image features need to be robust to differences in patient positioning within the scanner, acquisition method and segmentation to ensure reproducibility. Although test-retest studies are the recommended form of reproducibility studies, it is often impossible to test specific phenotypes of interest using this method, and they also preclude the use of retrospective images from clinical use, which cuts off a valuable data source. This is especially important since clinical testing does not benefit from the same set up time and resources available in research scans, and can introduce artefacts into the data. For this reason and also with the goal of increasing sample size, Zwanenburg et al. (2019) investigated combinations of image perturbations to determine feature robustness. They performed five separate types of image perturbation: noise addition, translation, rotation volume/growth shrinkage and super voxel-based contour randomisation on an oncological dataset.

Segmentation Variability

The next most important source of input variability is image segmentation. After accounting for variation associated with imaging guidelines and protocols, the extraction of 3D surfaces

provides further opportunity for method-based variations to occur. There is a possibility for the introduction of uncertainty of interpretation when creating the geometric models using segmentation. This is particularly the case in cardiac models, and more specifically the right ventricle, owing to the requirement for user judgment and/or complete manual input. The complex 3D shape of the RV; crescent-shaped in top-down view, but more triangular in side profile as it curves around the LV, means that automatic segmentation methods do not always adequately capture the shape, especially at the base and apex. Resultantly, manual segmentation protocols remain common in the literature (Tate et al., 2020), despite automatic segmentation becoming the standard for other tissues. Manual segmentation methods do increase uncertainty regarding both inter- and intra-observer variability, and this must be quantified. Sardanelli et al. (2008) compared manual segmentation between a beginner reader (25 cardiac segmentations performed in preparation for the study) and an experienced radiologist and found that both readers obtained a bias ranging between 8.0% and 9.9% for the RV when using both semi-automated and manual contouring methods. A small but significant difference in segmentation time was found between readers using the semi-automated method.

Tate et al. (2020) used shape analysis to assess segmentation variability. They note that despite the emergence of simulation pipelines into clinical use, variability in clinical interpretation remains largely unquantified. Tate et al. (2020) found high areas of variability at the RV apex and outflow tract. Although the area of largest standard deviation can occur at different locations along the annulus of the pulmonary valve. This study generated a parameterised shape model that can be used in cardiac modelling pipelines to compute uncertainty. Additionally, Peng et al. (2016) review segmentation techniques with a view to better understanding the role of cardiac segmentation in the final clinical assessment. Peng et al. (2016) also examine the use of long-axis images in cardiac segmentation, as these are important in the clinical use of CMR. The manual delineation of heart structures from MRI is very time consuming and subject to previously established inter-and intra-observer variability. With the exception of Tate et al. (2020), there has been very limited assessment of how altering the

shape meshes specifically impacts the output. If statistical shape analysis is to be employed in mainstream settings, whether clinical or commercial, an understanding of the degree of acceptable variation in input is crucial.

Lamata et al. (2014) tested the segmentation fitting accuracy of their automatic ventricular mesh personalisation algorithm using a Euclidean distance, expressed in millimetres from the nodes of the isosurface of the image to the closest point of the external surface of their mesh. They then integrated the error through the surface of the mesh using Gaussian quadrature. They found the regions of greatest fitting error to be at the open-faced basal site and also at the apex. This agrees with similar findings in Mauger et al. (2019).

Zhang et al. (2010) randomly divided 4D datasets in healthy and repaired Tetralogy of Fallot populations into 5 groups, each containing 5 datasets. They compared millilitre volumes of LV blood pool, myocardium mass and RV, in addition to calculating the overlap ratio and percentage volume difference and determined the mean and standard deviation of average surface positioning errors, overlap ratios and percentage volume differences.

Several studies have used two-way random-effects ANOVA and Bland-Altman plots to quantify inter and intra-observer differences (e.g. Kiefer et al., 2018). In addition to these statistical measures addition to mean absolute differences were calculated and presented as mean plus or minus standard deviation. Bland-Altman plots provide an approach of assessing agreement between two quantitative measures without assuming that either is the gold standard (Morey et al., 2009). This technique is performed as standard in many clinical robustness analyses, for example (Raisi-Estabragh, Gkontra, et al., 2020) which is described above. However, in this Chapter only a small subset of inputs are altered which means that alternative methods of testing agreement must be assessed. Taha et al. (2014) describe a formal method for selecting the most suitable metric for evaluating segmentation depending on use-case and then expand upon it in Taha and Hanbury (2015) to provide a set of guidelines based on interim results. The recommendation is to use distance-based metrics rather than overlap-based methods when the aim is to test the general alignment of the segments, espe-

cially when the difference between the two shapes is likely to be small. However, in this case, this thesis will use a Jaccard Indices, which are an overlap metric to compare pixel selections on a per-slice basis. Although it is highly likely that the central part of the RV will demonstrate a strong overlap, the unusual shape of the RV in the apex and outflow tract regions implies that there will be some slices where only a very small overlap is observed. Inter- and intra-observer comparisons will also examine overall volume.

Anatomical Reference Selection

The next significant source of variability associated with the shape analysis method employed in this thesis is the choice of poles and datelines; the anatomical boundary conditions which ensure that the correct anatomical regions of each heart (the apex and base) are situated on the north and south pole of the unit sphere, and that the same ‘edge’, which traces the anterior boundary between the septal wall and the free wall is assigned as the ‘dateline’, which represents the $[0, \pi]$ coordinate boundary after harmonic mapping (see Section 2.5). Anatomical consistency at the point of harmonic mapping is one of the unique features associated with the statistical shape analysis in this work, and therefore, there is not as much relevant literature to draw upon.

Wu (2013), who performed the underpinning work for this thesis, examined a combination of different dateline possibilities on a CT-derived right ventricular mesh. The different options included prescribing three different anatomical options for dateline selection as well as performing a series of systematic perturbations to the reference points themselves, including moving the north and south poles by 6 and 9mm respectively, shifting the entire dateline to the east or to the west and shifting one half of the dateline, to create an s-shaped perturbation. The conclusion from Wu’s thesis was that the selection of the anatomical region was significant to the modal results, but that errors in dateline selection changed the L_2 -norm (the chosen error metric) of the original mapping results by less than 0.5%.

Shi et al. (2008) mapped 3D surfaces using a similar harmonic mapping approach. They

have developed an automatic stable landmark detection algorithm, which finds features that are intrinsic to the surface geometry of the test subject; in this case, structures in the sub-cortex of the brain, and then uses these to guide harmonic maps to the sphere in a similar fashion to the poles and dateline in this analysis. To test the robustness of their method Shi et al. (2008) created an atlas by averaging shapes from the control group and then computed the displacement from each vertex on the surface to the corresponding vertex on the atlas. In order to correct for multiple comparisons, they applied 1 million permutations. They projected a checkerboard pattern on the original unit sphere and on the initial and final maps to confirm that the correct mapping regions occurred, which represents an innovative visualisation solution, and report a P value of 0.000076.

In other statistical shape analysis procedures, the boundary conditions are not as prescribed. Bowman et al. (2015) provides an in-depth analysis using two case studies (breast tissue identification and sexual dimorphism in human lips) of anatomical curve creation. This is an extension of a landmark-based method which better enables 3D shape description. The modified harmonic mapping method employed in this thesis combines the benefits of landmark shape analysis guiding the shape with the added degrees of freedom for shape associated with using the full shape information available with a point cloud. There is also very little literature associated with alteration of boundary condition placement, which renders the topic worthy of investigation.

Smoothing Filters

One final source of input variation tested involves the degree of Gaussian smoothing applied to the finite element meshes before the mapping and statistical decomposition steps. To the author's knowledge, there have been no direct assessments of the impact of altering the level of smoothing in a shape analysis procedure.

4.2 Aims and Hypotheses

4.2.1 Aims

The experimental focus of the work presented is to characterise the robustness of the statistical shape analysis based method when various modifications are made to the input. These modifications will be: changing the size of the input dataset (i.e. the number of input meshes) up to a maximum of 121 subjects and down to an extreme minimum of 20 subjects; introducing inter- and intra- observer differences in RV manual segmentation; altering the position of the poles and datelines, which locate the meshes on the unit sphere according to their anatomy and increasing and decreasing the stringency of the recursive Gaussian smoothing filter.

4.2.2 Hypotheses

1. In general, although inputs are likely to differ, as these are being altered as part of the robustness test, the output modes will remain unchanged, since recognised anatomical features have been identified in Chapter 3.
2. The same shape features will be extracted using the statistical shape analysis procedure, but these will not necessarily be in the same order of precedence as was determined in the initial 96 subject set.
3. The process will be robust to addition of subjects, since the same methods of data processing are used. Reducing the size of the dataset will decrease robustness at the lowest extremes. There may be fluctuations in robustness depending on which subjects are included in the random subject dataset.
4. Small changes to the dateline selections induced by different observers will not substantially change the overall mapping results.
5. The statistical shape analysis results will be robust to segmentation by a different observer. The same images are used, and each individual performing the segmentation

received advice from a cardiologist, so it would be expected that the inputs will not be substantially different, and therefore should not impact the coefficient results. Testing the difference in segmentation involves making alterations to both the mesh and the dateline. A larger difference in input should correspond to a larger difference in output.

6. Alterations to the strictness of the smoothing filter will substantially change statistical shape analysis results. Too stringent a smoothing filter will eradicate important shape features, while too relaxed a filter will maintain the ‘steps’ in the mesh, which are created as an artefact of the segmentation.
7. The coefficients experienced by different surfaces are likely to exhibit greater variation but a larger percentage difference in coefficient will not affect the order of the mode shapes.

4.3 Methodology

To assess the aims of this chapter (described in Section 4.2), robustness analysis included identifying the difference in input data as well as the resultant effect on shape features at the point of output. Keeping one parameter constant while perturbing other properties enables us to observe how sensitive the statistical shape analysis is to the implemented changes. While the shape analysis methodology to assess output remains as described in Chapter 2, the methods related to data acquisition and statistical analysis are described in this section.

The differences in initial input will be assessed mathematically, using a series of methods; Jaccard Indices and volume differences for comparing segmentations and Hausdorff distances for comparing dateline selection. In addition, the resultant difference in output will be tested by correlating the mode coefficients using a Pearson correlation to determine whether they still code for the same modes; by qualitatively comparing the modes that have been isolated and by comparing the mean squared difference between the modes themselves. The following sections will describe the specific perturbations made in more detail.

4.3.1 Data Acquisition

The set of 96 subjects described in Subsection 3.3.1 was combined with an additional set of 25 non-hypertensive hearts was acquired from Newcastle University. These further RVES were segmented, smoothed and meshed in accordance with the methodology outlined in Chapter 2. These hearts were added to the 96 subjects assessed in Chapter 3 to form a total dataset of 121 subjects and 242 hearts. All of the additional subjects included to increase the dataset size are non-hypertensive subjects from the Newcastle dataset. They were initially excluded as the scan images were not quite as clear as the initial 46 subjects. Nonetheless, they are a valuable test of the non-hypertensive results.

4.3.2 Segmentation

MRI images were segmented twice each by two observers in two separate sessions. There was no formal blinding of the dataset to reduce recall bias, however, each set of segmentations was performed after a minimum one month gap, and subjects were not able to refer to their previous segmentations. This was considered sufficient to prevent alterations in the segmented shapes based on memory. Further detail related to the statistical assessment of variability between observers is given in Subsection 4.3.5 (Statistical Analysis).

For the inter- and intra-observer variability tests, two independent observers each segmented a series of ten hearts, five from the Pittsburgh dataset and five from the Newcastle dataset. Although, ultimately only subjects from the Newcastle dataset were analysed in this robustness test. All selected RVs were from non-Hypertensive individuals. Two hearts were chosen on the basis of being clear visualisations of the RV, two on the basis of being ‘unclear’ visualisations of the RV and one was randomly selected. Cross-sections of the five selected RVs are displayed in Figure 4.1. Although neither observer received formal medical training, all segmentation was overseen by a trained cardiologist. Segmentations were performed using ITK-SNAP software (Yushkevich et al., 2016), on images in the short axis only to ensure consistency across the centres, since no long axis images were available for the images from

the Pittsburgh dataset. Meshing, smoothing and anatomical reference point selection on the new meshes were all performed by the author. Additional detail on assessing the robustness of pole and dateline selection is included in Subsection 4.3.3. The images were all registered to the same template, as described in the Methodology section (Chapter 2) in an effort to maintain consistency.

4.3.3 Dateline selection

Robustness to changes in dateline was assessed using blind sampling. The same observer blindly marked datelines on five heart meshes, selected according to the description in Subsection 4.3.2, i.e. by identifying the location of the anterior border of the RV in each slice, and using the custom software, ADAMCUS, to select the closest corresponding node on each mesh. Dateline selection also included selecting new end-points, which correspond to the pulmonary valve and apex on the RV segmentation, which then become the north and south poles of the unit sphere after harmonic mapping. The statistical shape analysis procedure was then applied using these alternate datelines. Figure 4.2 gives an example of the difference between two separate dateline selections on the same RV mesh.

4.3.4 Smoothing criteria

In the population-based analysis, discussed in Chapter 3, a recursive Gaussian filter with a sigma smoothing factor of $\sigma 3$ was applied to the RV mesh. This filter is a standard option as part of the Simpleware ScanIP software, but the stringency can be adjusted manually. The Gaussian smoothing operator performs a weighted average of surrounding pixels based on the Gaussian distribution, where central pixels are given more weight than neighbouring pixels; the standard deviation σ determines the amount of smoothing. The number 3, therefore refers to the standard deviation of the recursive smoothing algorithm. To assess the impact of varying the level of smoothing which was applied, that σ value was adjusted to 2 and 4, where 2 reduced the amount of smoothing and 4 increased it. Figure 4.3 demonstrates the

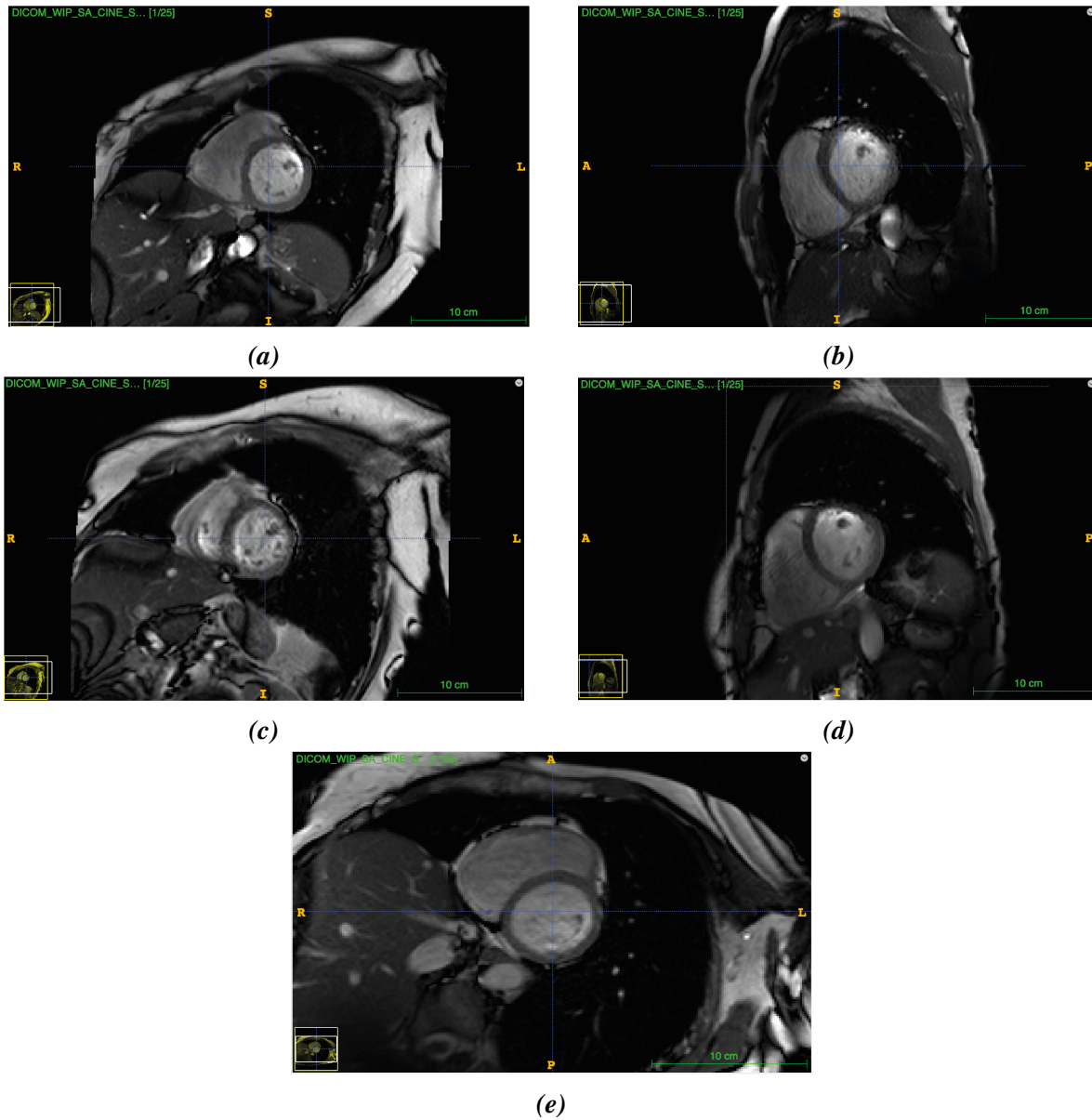


Figure 4.1: Screenshots of the short axis MRI scans of five RVs at end-Diastole. **a)** and **b)** were selected on the basis of being clear MRI representations of the RV. **c)** and **d)** were selected on the basis of being unclear representations of the RV, and **e)** was selected randomly.

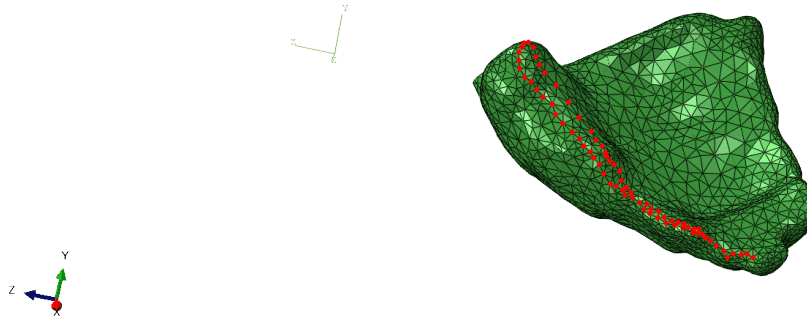


Figure 4.2: Comparison between the two dateline selections plotted on the mesh of an example RV mesh (Patient 2). The mesh is presented in a different orientation compared to other RV mesh Figures to best exhibit the distance between the datelines, with the apex and south pole pointing out of the page.

effect of the three smoothing filters on the finite element mesh. Too little smoothing leaves an artefact of the artificial slice thickness included in the initial segmentation and too much risks eliminating vital shape features. In the Newcastle subjects, as an example, the MRI scan includes approximately 14 slices of 8mm thickness. As the short-axis MRI scans produce a series of 2D images, the segmentation image a 2D polygon traced on each slice and then extrudes it in the z-direction by 8mm. Thus, a series of ‘steps’ are created in the pre-smoothed model which do not reflect reality. There is also a risk that the smoothing does not identify the twisted shape, as it curves around the LV. The degree of smoothness, and therefore this trade-off was judged by eye, but was informed by previous publications using this method (e.g. Xu et al., 2021).

4.3.5 Statistical Analysis

A series of statistical metrics were applied to compare the differences between the modified and unmodified groups at the point of input into the model, and their subsequent effect on the shape analysis output. In all cases, with the exception of perturbations to dataset size,

Intervention	Input Tests	Output Tests
Dataset Size	NA	<ul style="list-style-type: none"> ● Inspect Modal energy plots ● Inspect 3D reconstructions ● Pearson Correlations ● Percentage Difference ● Sum of Squared Difference ● Application of LDA
Dateline Selection	<ul style="list-style-type: none"> ● Hausdorff Distance 	<ul style="list-style-type: none"> ● Inspect Modal energy plots ● Inspect of 3D reconstructions ● Pearson Correlations ● Percentage Difference ● Sum of Squared Difference ● Application of LDA
Segmentation Differences	<ul style="list-style-type: none"> ● Jaccard Indices ● Volume Percentage Difference ● CV% 	<ul style="list-style-type: none"> ● Inspect of Modal energy plots ● Inspect 3D reconstructions ● Pearson Correlations ● Percentage Difference ● Sum of Squared Difference ● Application of LDA
Smoothing Differences	<ul style="list-style-type: none"> ● Comparison of mesh elements 	<ul style="list-style-type: none"> ● Inspect Modal energy plots ● Inspect 3D reconstructions ● Pearson Correlations ● Percentage Difference ● Sum of Squared Difference ● Application of LDA

Table 4.1: Statistical Tests performed to evaluate the robustness of each of the four targeted interventions. The same output tests were performed for each intervention, but input tests varied based on the nature of the input data.

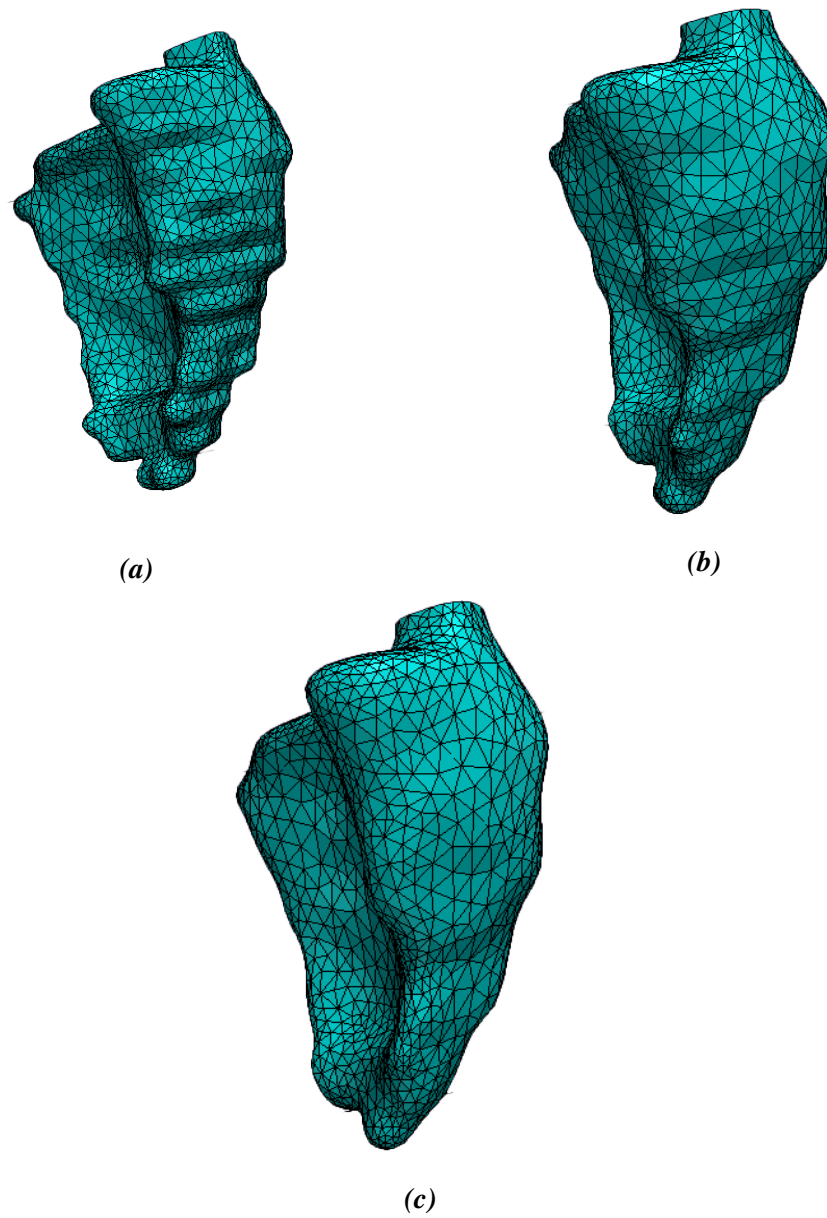


Figure 4.3: Visualisation of the different levels of smoothing applied when different pixel filters are applied. **a)** depicts a less stringent recursive Gaussian filter ($\sigma 2$), **b)** depicts the standard smoothing filter described in the Methodology and utilised in Chapter 3 ($\sigma 3$). **c)** shows a more stringent recursive Gaussian filter ($\sigma 4$).

five subjects (ten ventricle shapes since both cardiac phases are included) were subjected to altered conditions, as described in this methodology section, as part of the robustness test. A small number of meshes were changed, to ensure that the overall shape analysis is capable of withstanding alteration of a small number of hearts first before extending to a wider range,

which is outside the scope of this thesis. Retaining the majority of meshes also provides a point of comparison, where the impact can be measured against the unchanged meshes. Changing just five subjects precludes the use of common medical robustness measures, such as Bland-Altman plots as it gives too few input data points to produce a comparison.

While measurement error should ideally be independent of the actual sample, in biology this is almost never the case. Popović and Thomas (2017) describe an extreme case where the error in systolic strain rate measurement widens dramatically with decreasing animal size. This is particularly notable in this dataset, where although all ventricle meshes were registered to the same template, there are substantial differences in the size of the mesh between end-Systolic and end-Diastolic shapes. Off the back of this, the recommendation is that error reporting should be expressed as a percentage.

The conclusions drawn from Chapter 3 regarding the impact on population are only relevant if they are repeatable in different conditions, and with different MRI scanners (although test-retest repeats are outside the scope of this thesis) and therefore, this chapter aims to quantify changes in robustness of the data when it is perturbed in different ways. This is with a view to assessing whether the results can be interpreted accurately in a clinical setting and to determine specific mitigations that could improve the diagnostic potential of a statistical shape analysis based diagnostic. The specific changes that will be assessed are the size of the dataset; both reducing and increasing the number of subjects, altering meshes created by changing the individual performing the segmentation, altering the dateline reference points for the harmonic mapping, and changing the level of smoothing applied.

Recommendations can then be made on the requirements for reproducibility and the effectiveness of the current method without further intervention.

Quantifying Input Differences

Hausdorff Distances The variation in the nodes selected as the north pole, south pole and dateline inputs was quantified using a Hausdorff distance. The Hausdorff distance is a point to

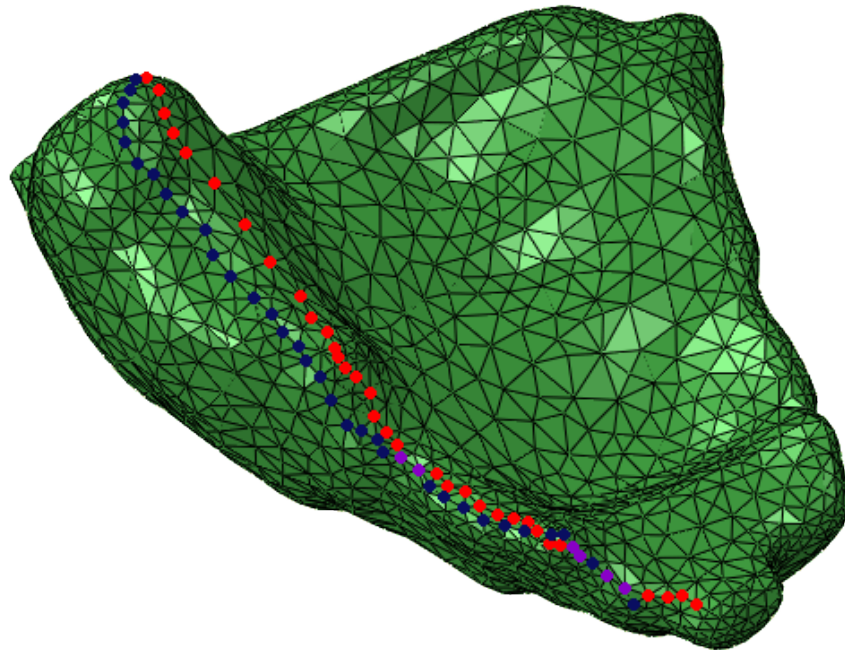


Figure 4.4: *Reproduction of Figure 4.2 with datelines coloured by selection attempt. The original dateline nodes are depicted in red, the second attempt nodes in dark blue, and those nodes which overlap between the two selection attempts are shown in purple. The Hausdorff distance calculates the longest minimum distance between nodes selected in the first group and the second group. Again, the mesh is presented in a different orientation to other depictions of the RV, in order to better visualise the dateline placement, since it is located on a boundary between the free wall and the septal wall.*

point matching method that is commonly used in shape comparison strategies. The distance defines the maximum deviation between two models, and this method finds the maximum of the closest pairings between all points on the dateline (see, for example, Figure 4.4). Assuming that D_i is the set of points on the original dateline $D_i = d_{i1} \dots d_{im}$ and D_j is the set of points on the new dateline $D_j = d_{j1} \dots d_{jn}$ where m and n are the number of dateline nodes in each set respectively, d represents individual coordinates of points on the dateline in cartesian space and $\|\cdot\|$ denotes the the Euclidean norm. The Hausdorff distance is given by:

$$H(D_i, D_j) = \max(h(D_i, D_j), h(D_j, D_i)) \quad (4.1)$$

with

$$h(D_i, D_j) = \max_{d_j \in D_j} \min_{d_i \in D_i} \|\vec{d}_i - \vec{d}_j\|, \quad (4.2)$$

where both h and H are in mm.

Segmentation volumes were recorded from Simpleware (Synopsys, Exeter, UK) after the Gaussian smoothing step, but before meshing in all cases where the overall mesh was altered as part of the robustness assessment (i.e. when testing inter- & intra-observer variability and smoothing).

Calculating the percentage difference in segmentation volume is a simple way to assess the overall difference between two RV segmentations since volumes are evaluated in Simpleware before the application of Gaussian smoothing filter. However, this calculation can often give an inaccurate picture of the extent of the disparity. For example, if one observer chooses not to segment an apical slice, the resultant change in the volume is proportionally greater than the impact on the shape. Therefore, two additional measures; Jaccard indices and means of coefficient of variation (CV%) were calculated in an attempt to more fairly quantify the differences in input RVES.

Jaccard Indices were calculated as a method of describing the similarity of the starting sample sets, before the shape analysis procedure was applied. The Jaccard index is defined as:

$$J(\text{Slice 1, Slice 2}) = \frac{|\text{Slice 1} \cap \text{Slice 2}|}{|\text{Slice 1} \cup \text{Slice 2}|}, \quad (4.3)$$

where Slice 1 is an MRI slice from the first segmentation and Slice 2 is the equivalent slice from the second segmentation. This method was used to quantify pixel overlap between the MRI image slices and thereby determine areas on each sample slice where there was agreement between both segmentations. Separate Jaccard Indices were produced to compare the difference in pixels selected between observers, and to compare between the initial and repeat measurements taken by each assessor on a slice by slice basis. Per slice overlap is calculated in MATLAB by assigning a value of 0 to pixels not selected in either MRI slice, a value of 1 to those selected in one segmentation, but not the other, and a value of 2 to those selected in both slices. The intersection over union calculation is then performed dividing the number of pixels assigned exactly 2 by the number assigned either 1 or 0. Both percentage overlap, and overlap on a slice by slice basis are reported. Total percentage overlap is reported as an average of all slices for each RV.

Means of coefficient of variation (CV%) of mesh volume were calculated as a further method of analysing variability in the initial sample data when comparing two observers. The CV% is the standard deviation of paired measurements divided by the mean value computed for each subject, calculated as follows:

$$CV\% = \frac{\sigma}{\mu} \times 100 \quad (4.4)$$

where σ is the standard deviation of a pair of measurements (e.g. the volume recorded by Observer 1's segmentation and that recorded by Observer 2 for the same RVES), and μ is the mean of those volumes which is then averaged over the five test subjects and reported as a per-

centage. For each case, four CV% were calculated and compared: two repeated measurements of volume by Observer 1; two repeated measurements by Observer 2; the first measurements of Observer 1 and 2 and the second measurements of the observers. This method is commonly used in literature for assessing semi-automated segmentation methods (Bae et al., 2009), and reports a single, informative percentage value. Bland-Altman plots, which depict the difference between two measures over their average, are commonly used to assess segmentation difference. It was decided not to produce such plots in this work, since a subset of just 5 subjects were altered in each test, which would give an insufficient sample size to provide useful data.

Quantifying Output Differences

Statistical decomposition (described in Section 2.8) produces eigenvectors representing the strongest shape descriptors within the dataset (‘the modes’) which have associated modal coefficients; values which describe how much of any given mode produces the best reconstruction of a particular RVES (the strength of the shapes). Each subject is reconstructed from the summation of the mean heart and each of the 192 shape modes multiplied by a representative modal coefficient, as was described in Equation 2.6.

To assess the impact of altering the input parameters on the overall shape features derived from the shape analysis (the output); a series of tests were conducted which fall into two broad categories. The first group directly compares the mode shapes produced through the statistical decomposition. This was achieved by comparing the modal energy plots from each subsequent statistical shape analysis; through calculating the sum of squared difference between the shape modes, and through inspection of 3D reconstructions of the mode shapes, when these are added to the mean ($\pm 2 \times SD$), as described in Equation 2.11, and Figure 2.13.

The second category of tests compares the mode *coefficients*, i.e. whether the representation of each shape feature within individual subjects has altered. Pearson correlations were computed to determine whether the mode coefficients are statistically correlated with one an-

other, or with cognate modes. A second test analysed percentage differences to directly compare mode coefficient values. Finally, LDAs were performed using a selection of the best identified test combinations in Chapter 3 to contrast key Linear Discriminant scores and determine the clinical applicability of the Population Analysis. The LDA procedure followed that described in Subsection 3.3.2. Specifically, Mode 19, which correlated with subject sex, Modes 9 & 5, and Modes 9 & 18, which correlated with Age at end-Systole, Modes 20, 11, 2 which correlated with Obesity at end-Diastole.

Pearson product-moment correlation coefficient values were calculated to measure the strength of the linear relationship between the output mode coefficients. The Pearson coefficient is defined in Equation 3.3.1 and assigns a value between -1 and 1, where 0 indicates no correlation, 1 is total positive correlation and -1 is total negative correlation. The subscript represents the number of degrees of freedom, where $df = n-2$.

The sum of squared differences was calculated to compare the mode eigenvectors directly. This method calculates the squared difference between every point on the interpolated 400x200 grid as follows:

$$D = \sqrt{(\vec{\mu}_{ki} - \vec{\mu}_{kj})^2} \quad \text{or} \quad D = \sqrt{(\vec{\mu}_{ki} + \vec{\mu}_{kj})^2} \quad (4.5)$$

where μ_{kj} is a point on the original grid, and μ_{kj} is a point on the new grid.

As briefly outlined in the methodology (Chapter 2), since the mode shapes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. This means that the same mode might have the opposite sign without changing the significance or effect. Therefore, there is a possibility that two modes which are in actuality the same appear to have a larger difference. This would not be accounted for by just taking the absolute value. To combat this, an addition is performed instead of a subtraction in cases where the signs differ.

The second step is then to calculate a second total distance between them, which was

implemented as follows:

$$\text{Total Distance} = \frac{1}{n} \sqrt{\sum D^2}, \quad (4.6)$$

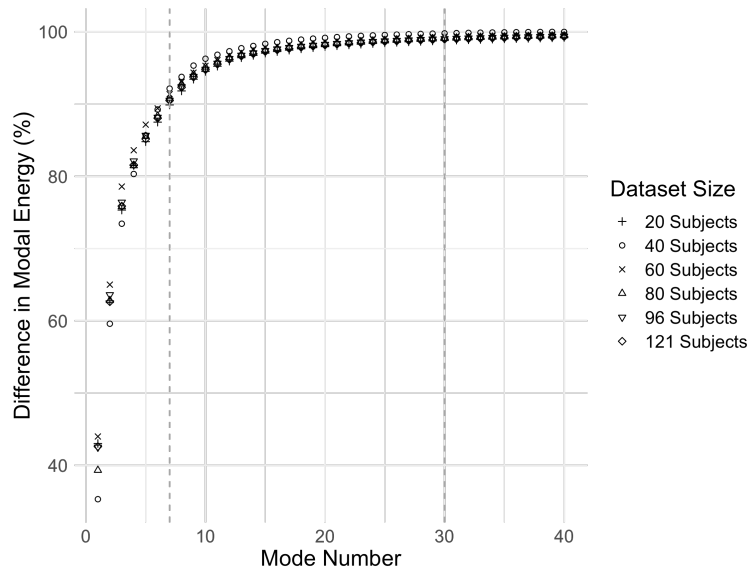
where n is the total length of the mode vector.

4.4 Results and Discussion

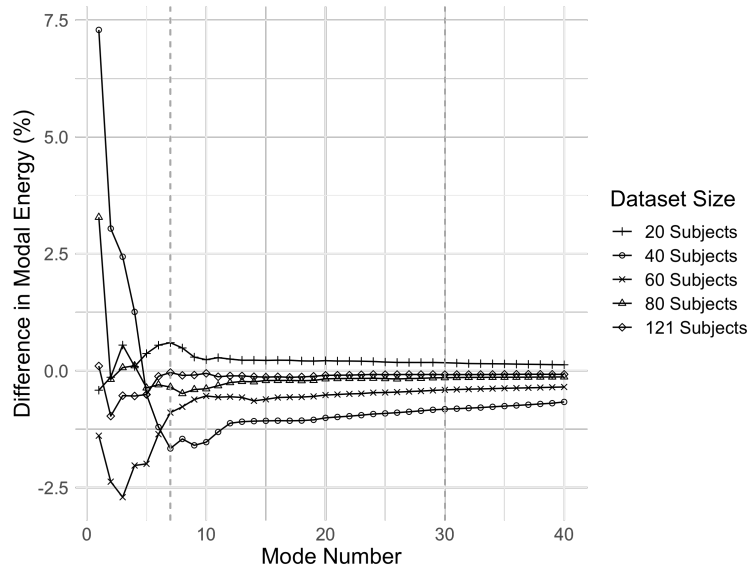
In this section, the results of the four specific interventions described in the Introduction (Section 4.1) will be described. Each section will first include analysis of the impact that the methodological change has introduced to the input. This is assessed in different ways depending on the specific assessment. Then, the impact of changes to the overall shape analysis output will be reported. The shape analysis procedure will be evaluated using a series of different metrics; examining the modal energy curves, assessing Pearson Product-Moment correlation coefficients, through examination of the qualitative mode shapes produced, through inspection of the raw modal coefficients and through exemplar LDA analyses, which compare a selection of classification scores.

4.4.1 Robustness to Changes in Dataset Size

To assess the robustness of the mode generation system in response to alterations to the dataset size, the original 96 subject dataset was also reduced to 80 subjects, 60 subjects, 40 subjects and 20 subjects. The composition of each of these reduced test sets was selected by a random number generator. The selections were overlapping, but each was taken as a subset of the original 96 subjects, so the choices are not subsets of one another, and the 40 subject set was not necessarily a subset of the 60 subject set.



(a)



(b)

Figure 4.5: Difference in Modal energy distribution between the datasets. **a)** Modal energy profiles for shape analyses conducted when an additional 25 subjects were added and when the dataset size was reduced to 80, 60, 40 and 20 subjects respectively. The vertical dashed lines indicate Mode 7, which covers 90% of the total modal energy, and Mode 30, which describes 99% of the total modal energy. **b)** Difference between the original modal energy curve and those for shape analyses conducted an additional 25 subjects were added and when the dataset size was reduced to 80, 60, 40 and 20 subjects respectively.

Impact of Changes to Dataset Size Output

Figure 4.5 shows the modal energy curve profiles for the decomposition when dataset size is altered. As described in Section 2.8, the end-Systolic and end-Diastolic shapes are decomposed together. Figure 4.5a demonstrates that the modal energy curve profiles are very similar when the dataset size is both increased to 121 subjects and systematically reduced. Seven modes are needed to capture 90% of the modal energy in all but the most extreme scenario (dataset size reduced to 20 subjects), which only requires 6 modes. The 80 subject dataset uses 29 modes to capture 99% of the modal energy; the 60 subject set requires 28 modes; the 40 subject set requires 24 modes and the 20 subject dataset needs 19 modes to capture 99% of the modal energy. In most scenarios, the distribution of modal energy would not be expected to change, but the fact that such change is minimal even after the addition and removal of large numbers of individuals from the analysis is a promising indicator of robustness. Figure 4.5b indicates that the maximum percentage difference between the tests was just 7.5%, even at 20 subjects.

Tables 4.2 and 4.3 report Pearson correlations of the modal coefficients between the 96 subject dataset used in Chapter 3 and the altered set of shape analyses. At end-Diastole, there is a relatively small change in the information conveyed by the shape modes. The modal coefficients are strongly correlated for the first 17 Modes when the dataset size is reduced from 96 subjects to 80 subjects. The mode shapes are also qualitatively similar (Figure 4.6).

Tables 4.2 and 4.3 also provide evidence that the shapes encoded by Modes 6 and 7 have interchanged in the 80 subject analysis. The comparatively low (although still statistically significant) absolute Pearson correlations of 0.597 and 0.451 respectively are corroborated when considering the qualitative mode shapes in Figure 4.6. When correlating the Mode 6 coefficients from the 96 subject analysis to the Mode 7 coefficients from the 80 subject analysis the absolute correlation is an improved $r_{78} = 0.839$. Likewise, correlating 96 subject Mode 7 coefficients to 80 subject Mode 6 gives a correlation of $r_{78} = 0.810$.

At end-Diastole, when the dataset size is increased to 121 individuals, the correlation is very strong and does not fall below $r_{96} = 0.878$ for the first 16 modes. The first 4 modes are

Table 4.2: Absolute Pearson correlation coefficients comparing the first 20 modes in the 96 subject dataset to cases where dataset size from the original 96 subjects is changed by adding in or removing individuals at end-Systole.

Mode No.	121 Subjects	80 Subjects	60 Subjects	40 Subjects	20 Subjects
1	0.999	0.999	0.998	0.999	0.995
2	0.998	0.995	0.983	0.985	0.942
3	0.986	0.996	0.990	0.948	0.937
4	0.988	0.997	0.980	0.981	0.968
5	0.989	0.994	0.992	0.132	0.232
6	0.959	0.467	0.801	0.424	0.386
7	0.951	0.576	0.886	0.417	0.011
8	0.986	0.975	0.875	0.479	0.414
9	0.964	0.962	0.882	0.875	0.445
10	0.878	0.973	0.887	0.355	0.245
11	0.972	0.966	0.329	0.230	0.755
12	0.929	0.941	0.488	0.887	0.174
13	0.948	0.963	0.900	0.610	0.274
14	0.969	0.937	0.915	0.667	0.295
15	0.958	0.926	0.902	0.803	0.312
16	0.930	0.914	0.702	0.458	0.438
17	0.651	0.931	0.701	0.143	0.205
18	0.435	0.523	0.782	0.588	0.155
19	0.175	0.002	0.697	0.153	0.443
20	0.322	0.579	0.312	0.019	0.110

Table 4.3: Absolute Pearson correlation coefficients comparing the first 20 modes in the 96 subject dataset to cases where dataset size is changed by adding in or removing individuals at end-Diastole.

Mode No.	121 Subjects	80 Subjects	60 Subjects	40 Subjects	20 Subjects
1	0.999	0.999	0.996	0.999	0.976
2	0.998	0.996	0.991	0.993	0.971
3	0.982	0.996	0.991	0.962	0.940
4	0.982	0.995	0.976	0.983	0.880
5	0.985	0.991	0.989	0.866	0.658
6	0.977	0.597	0.866	0.748	0.372
7	0.955	0.451	0.846	0.528	0.636
8	0.976	0.955	0.814	0.602	0.198
9	0.950	0.948	0.922	0.781	0.533
10	0.940	0.990	0.888	0.008	0.336
11	0.950	0.954	0.380	0.184	0.435
12	0.958	0.949	0.319	0.850	0.159
13	0.959	0.963	0.953	0.585	0.570
14	0.956	0.905	0.903	0.539	0.259
15	0.953	0.928	0.859	0.547	0.091
16	0.913	0.890	0.685	0.432	0.699
17	0.527	0.926	0.605	0.159	0.339
18	0.510	0.552	0.733	0.560	0.229
19	0.274	0.080	0.735	0.183	0.298
20	0.319	0.455	0.065	0.529	0.140

robust in all dataset size perturbations tested (with the weakest correlation being $r_{96} = 0.968$ at end-Systole, and $r_{96} = 0.880$ at end-Diastole), including reducing the dataset size to just 20 subjects. This highlights that the key shape descriptors are captured even in a much reduced dataset. After Mode 4 at end-Systole, there is a dramatic reduction in Pearson correlation when the dataset is reduced to 40 subjects and 20 subjects, indicating that the mode shapes are not shared after this point. The same mode robustness patterns are shared at both end-Systole and end-Diastole. This is a helpful indication for future analysis, because the dataset was decomposed together, but systolic and diastolic components are assessed separately. Therefore, the similar robustness patterns demonstrate that both heart phase shapes are captured equally.

Mode 19, which was found to be a mode associated with sex-based differences at end-Systole in Chapter 3, does not correlate significantly with the 96 subject dataset in any test except for the 60-subject dataset. This cautions the user to be careful with the lower-energy modes, unless a significantly larger dataset is used as such modes may be encoding different shape features between analyses. Although, conclusions regarding separation into sex-based groups do not directly translate to different dataset sizes, the similarity of the decomposed shape feature and the relative correlations insinuates that a correction could easily be applied. This would enable direct comparisons of meaningful shape features. On the other hand, Modes 9 & 6 both correlate strongly with their equivalent coefficient values in all dataset sizes, indicating that these age-based predictor modes remain the same across all perturbations to dataset. Perhaps the main outcome from these findings is that, despite evidence from the modal energy curve (Figure 4.7a), the higher order modes (above 16/17) do not convey useful shape information beyond a specific dataset and are prone to noise. A key component of the statistical shape analysis method is that the features are ranked; therefore, this outcome is not unexpected.

Figure 4.6 depicts visualisations of the mean $RV \pm 2 \times$ standard deviation for a series of modes, selected because of their ability to classify non-medical subject characteristics in the previous chapter, at 121 subjects, 80 subjects, 60 subjects, 40 subjects and 20 subjects. This

image demonstrates that there is very little qualitative difference between the mode shapes when increasing and decreasing the size of the dataset, and provides convincing evidence that perturbations related to dataset size do not substantially change the derived mode features, at least over the first 9 modes. It is imperative to highlight that Mode 1 retains the distinctive shape features seen in the original population-based analysis, affecting ventricle length and apex detail. The 20 subject dataset (depicted in olive green) clearly differed in all cases but Mode 1. Mode 9 differed when the dataset size increased to 121 subjects - in the 121-subject sample, it appears to represent greater curvature at the base, whereas in the 96-subject sample the bulging is closer to the apex. Mode 19 only correlated with the 60-subject dataset. It is clear from Figure 4.6 that this statistical relationship is based on the distinctive ‘hump’ shape at the base of the RV, near the pulmonary valve. The Mode shapes for the 20 subject dataset clearly diverge after Mode 1.

A further way of validating the robustness of the method to fluctuations in Dataset size was to look directly at the mode eigenvectors shapes themselves. Table 4.4 compares the sum of squared difference at Mode 1 between 96 subjects and the various 80-40-60-20 extremes. This method is described by Equation s 4.3.5 and 4.3.5. As expected, there is a small overall mean squared difference between 96 and 95 subjects (1.31×10^{-5}), and the difference increases in the more extreme datasets (e.g. 96 vs 20). Subjects 96 vs 40 show a relatively small sum of squared differences (9.32×10^{-5}), despite a substantial change in input, which is surprising. The addition of 25 subjects to form the 121-subject group causes an approximately similar change in the sum of squared difference value to the 80-subject comparison. This suggests that increasing the difference in input value causes a similar change to the output modes. Nevertheless, the surprisingly small difference between 96 and 40 subjects indicates that the sum of squared differences may not always be the best metric for comparing statistical shape analysis measures. These values are for Mode 1 only, which as has been shown in previous Tables (4.2, 4.3 & 4.7) to have remained largely static.

Table 4.5 lists a comparison of percentage differences across the dataset sizes. The mean,

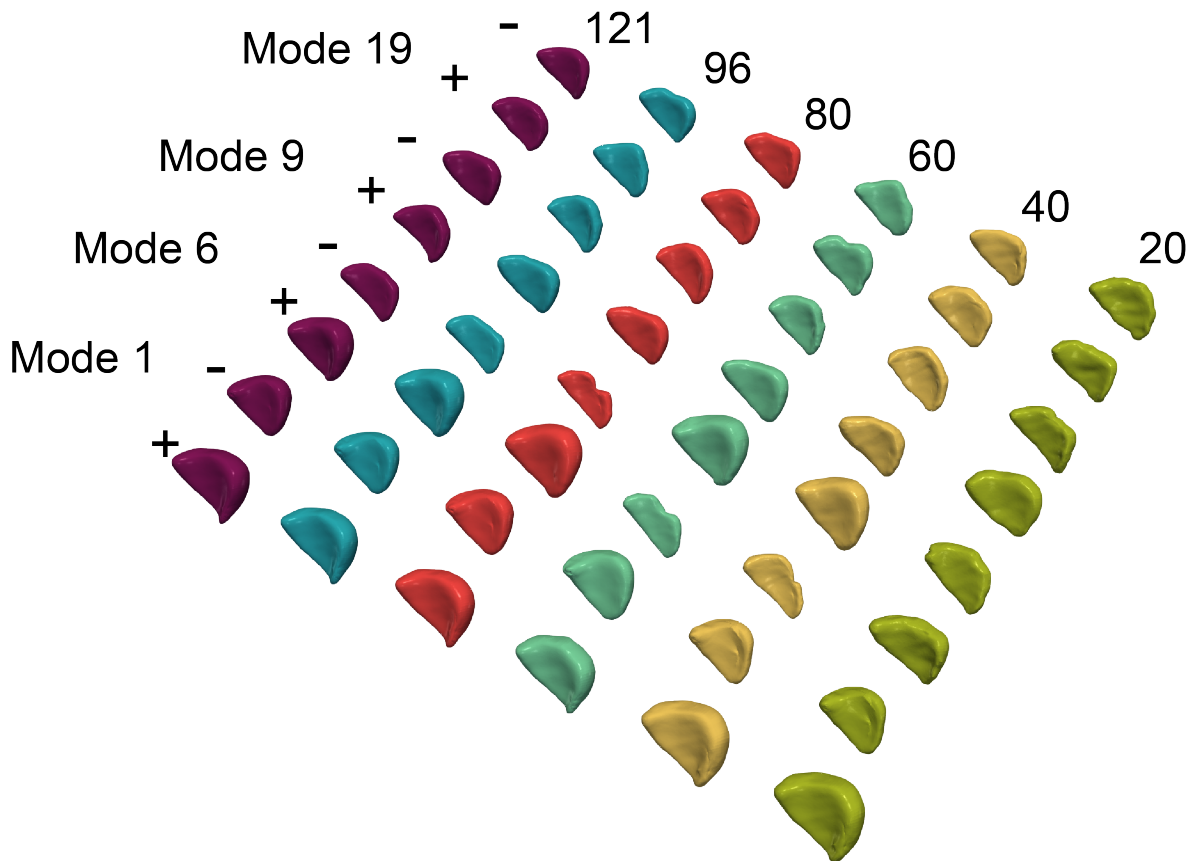


Figure 4.6: *Qualitative depiction of the RVES shapes when Modes 1, 6, 9 and 19 in the 121, 96 (original), 80, 60, 40 and 20 subject datasets are plotted on the $\text{Mean} \pm 2 \times \text{SD}$. Mode 1 appears to encode the same qualitative shape for all dataset sizes. Mode 6 encodes a similar shape feature when the dataset for all except the 20 subject test. Mode 9 diverges when the dataset is increased to 121 subjects, and when the dataset is reduced to 20 subjects. Finally, Mode 19 differs in most comparisons, although the 60 subject test shares many similarities with the original. As the modes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. Hence, a mode can be observed to encode the same shape whether that mode is added or subtracted from the mean.*

Table 4.4: Sum of squared difference between the Mode 1 eigenvector when the dataset size is reduced from 96 subjects to 80, 40, 60 and 20 subjects, and increased to 121 subjects. This method accounts for a difference in sign of the mode.

Dataset	Difference	Dataset	Difference	Dataset	Difference
96 vs 95	1.31×10^{-5}	121 vs 96	6.18×10^{-5}	80 vs 60	6.63×10^{-5}
96 vs 94	1.66×10^{-5}	121 vs 80	7.06×10^{-5}	80 vs 40	7.66×10^{-5}
96 vs 80	5.52×10^{-5}	121 vs 60	1.07×10^{-4}	80 vs 20	1.28×10^{-4}
96 vs 60	1.00×10^{-4}	121 vs 40	1.01×10^{-4}	60 vs 40	1.11×10^{-4}
96 vs 40	9.32×10^{-5}	121 vs 20	1.40×10^{-4}	60 vs 20	1.01×10^{-4}
96 vs 20	1.50×10^{-4}			40 vs 20	1.76×10^{-4}

minimum, maximum and standard deviation of the percentage differences are reported. The percentage difference calculation uses absolute mode coefficient values to account for the fact that the sign is not relevant, and takes into account the size of the dataset.

The results in Table 4.5 in general accord with the Pearson correlations shown in Tables 4.2 and 4.3. The smallest mean percentage difference is between the two datasets which are closest in size. The two shape analyses where the size was only very slightly altered: from 96 to 95 subjects and from 96 to 94 subjects, show the smallest mean percentage difference. The percentage difference between the original values and the altered dataset sizes is the smallest at Mode 1. This again can be explained through the fact that the modes are ranked, meaning that Mode 1 explains the largest proportion of modal energy, and therefore would be expected to be most similar. The standard deviation of percentage difference is always smaller at end-Diastole than at end-Systole. Again, this is unsurprising since the end-Diastolic hearts, on average are larger and are also much easier to capture in the segmentation process.

Table 4.6 shows a subset of raw mode coefficient values to give an example of coefficient sizes and the absolute impact that altering input dataset size has on these values. There is a greater difference in raw values in the more extreme datasets, but it is clear that there is an overall trend of mode features in all cases.

Table 4.5: Mean, minimum, maximum and standard deviation of the percentage difference values which compare results from selected modes across the different dataset sizes against the original 96 subject dataset from Chapter 3

		Mode 1		Mode 6		Mode 9		Mode 19	
		Dia	Sys	Dia	Sys	Dia	Sys	Dia	Sys
121	Mean	44.7	159.7	39.4	73.1	65.0	92.3	282.1	267.6
	Min	0.2	0.2	0.1	0.6	0.0	0.0	4.2	3.2
	Max	991.2	13318.4	2355.2	5134.8	2503.8	5967.9	5718.9	6096.1
	SD	122.6	1357.6	252.0	523.9	291.3	611.5	913.4	945.9
95	Mean	-0.3	32.8	-0.2	4.6	9.5	119.6	13.6	16.9
	Min	0.1	0.0	0.0	0.0	0.0	0.8	0.2	0.0
	Max	44.4	3141.6	95.5	622.8	428.7	7998.1	547.5	573.3
	SD	7.3	320.3	14.3	64.2	59.6	820.1	76.7	80.2
94	Mean	2.8	37.3	8.1	17.6	22.6	113.3	110.9	128.1
	Min	0.0	0.0	0.0	0.0	0.1	0.8	0.1	0.9
	Max	154.7	3452.0	455.6	1500.7	806.1	6973.2	2955.0	3826.2
	SD	20.3	353.9	55.9	155.3	122.1	720.8	416.2	543.4
80	Mean	9.6	12.8	185.4	444.2	105.8	245.6	271.9	223.6
	Min	0.3	0.1	1.7	1.7	0.3	0.7	6.7	1.7
	Max	560.0	712.9	6329.7	26016.1	2629.9	16402.3	4953.1	4052.9
	SD	86.4	80.0	786.7	2909.3	480.0	1821.7	787.2	716.4
60	Mean	42.4	241.3	33.9	103.7	56.1	72.5	134.7	362.8
	Min	0.1	0.2	2.2	1.9	0.1	1.1	0.8	0.5
	Max	1999.3	14477.9	1857.8	2059.2	1295.3	697.6	2410.6	10986.1
	SD	267.0	1851.7	237.7	353.6	185.3	153.0	380.1	1512.0
40	Mean	4.5	1.2	51.0	635.1	278.8	95.9	139.8	44.1
	Min	0.8	0.6	5.0	5.6	0.3	4.3	8.1	7.0
	Max	143.1	72.9	1041.5	15927.8	5248.0	1269.2	1882.9	809.4
	SD	28.3	13.0	189.5	2521.2	977.2	262.7	425.5	148.1
20	Mean	14.3	5.8	-30.1	126.2	269.2	1318.4	352.8	87.4
	Min	1.2	0.1	5.0	9.5	19.3	13.1	1.6	1.4
	Max	499.9	116.5	107.9	580.5	4299.6	24137.7	3138.5	1131.6
	SD	103.2	26.0	33.0	161.9	921.0	5234.1	807.4	294.1

Table 4.6: Representative raw mode coefficient values for Mode 1 comparing results for different dataset sizes. As the dataset size reduction was performed using a random number generator, the NA values refer to cases where that subject did not form part of the smaller dataset. Six consecutive subjects are shown.

Subject No.	121	96	95	94	80	60	40	20
6	-20.8	-16.4	16.2	15.2	-21.5	-20.8	-17.7	-25.8
7	66.5	67.8	-68.1	-68.2	66.7	70.1	NA	67.0
8	-63.3	-58.0	57.9	56.8	-64.4	-64.9	-62.1	NA
9	2.4	5.2	-5.3	-6.1	1.7	2.3	NA	-2.6
10	82.8	84.1	-84.0	-84.6	NA	NA	NA	NA
11	-12.0	-7.5	7.0	NA	-11.1	-10.2	-6.2	NA
12	70.2	72.0	NA	NA	69.3	71.5	NA	NA

Application of LDA

Next, the conclusions obtained in Chapter 3 were assessed by applying them to these new robustness analyses. LDA analyses were performed on selected groups to determine the impact of dataset size on the classification results obtained in Chapter 3. The intention of this test was to determine whether altering the dataset size and the accompanying differences to the derived modal coefficients impacted the ability of a linear discriminant model to classify based on features, as was performed in Chapter 4. The strong Pearson correlations and visible qualitative agreement outside of the extreme changes to dataset size suggest that the best mode combinations for classification in the original 96 subject analysis should be similarly successful here. The selected combinations were 121 subjects; 95 subjects, to determine the impact of making only a very small adjustment; 80 subjects and 40 subjects. The 20-subject dataset analysis was not included since testing more covariates (i.e. the top 30 Modes) than input subjects is not recommended. Table 4.7 displays this selection of LDA classification results with the reduced and increased dataset sizes. Of the 40 subject group, 17 subjects were non-hypertensive, and of the 80 subject group, 37 subjects were non-hypertensive.

Applying Mode 19 alone, which, by itself was the most distinctive shape feature for separating sex in non-hypertensive subjects, did not separate any dataset well (range 54.1%-64.7%

successful classification). The highest classification accuracy (64.7%) was in the 40-subject test, and the lowest classification accuracy was in the 80-subject dataset (54.1%). The classification rate was also poor (55.4%) when the input size was increased to 121 individuals. However, in this category, Mode 18 was a stronger predictor of the dataset, able to correctly identify 60% of the 121 subjects after cross-validation. Therefore, it could be argued that although adding more subjects has slightly altered the specific mode number, the identified shape features are not changing. There is also a possibility that adding in 25 additional non-hypertensive subjects has diluted the specific features and that Mode 19 does capture shape features related to hypertensive individuals.

The best overall alignment with the original data was in categorising obesity. In the original 96 subject dataset, Modes 20, 11 and 2 were the optimal separators based on obesity. The classification range was 58.8%-74.4%, with the best classification percentage in the 95-subject category, where just one subject was removed. Modes 20, 11 and 2 were able to correctly separate 62.5% of the 121 subject dataset, and 67.6% of the 80 subject dataset and the discriminant analysis is statistically significant in these cases. In general, it can be discerned that the percentage of correctly separated individuals is greater when the best separator modes are included in the list of ‘Top Five’ by structure matrix (as described in Section 3.3.2).

Removing a single subject from the analysis in the 95-subject dataset produced an unexpectedly different result at end-Systole. One specific test stood out as producing significantly poorer results than the others. Modes 9, 5 and 9, 18, which were the best classifiers by age, could only separate 35.9% and 30.8% of the individuals correctly, whereas the same modes correctly classified 72.5% of the original 96 individuals. However, looking at the most accurate classifier column indicates that Modes 10 and 19 are the two most important classifier modes in terms of their structure matrix value; the structure matrix value reports correlation to the discriminant function, and can be interpreted as factor loadings to identify the largest absolute correlations to the discriminant function. It is therefore plausible that Modes 9 and 10 and Modes 18 and 19 have been switched in this case. The ‘Top Five’ modes indicate a

Table 4.7: Exemplar LDA Analyses comparing a selection of classification results between the original analysis (non-hypertensive only) and the dataset alterations. Test Combination reports the combination of modes used as inputs in the LDA. Where more than one combination was found equally successful in Chapter 3, both were tested. ‘Classification %’ reports the best accuracy percentage after cross validation. ‘Top Five’ lists the best five modes identified using a new LDA, created based on the new coefficients from the robustness test.

Dataset	Trait	Phase	Test Combination	Classification %	Top Five
40 Subject	Sex	Systole	19	64.7	3, 9, 15, 8, 7
40 Subject	Age	Systole	9, 5	58.8	29, 9, 3, 4, 2
40 Subject	Obesity	Systole	15, 16, 6 & 15, 29, 16	52.9	9, 10, 3, 8, 4
40 Subject	Sex	Diastole	6, 23	64.7	15, 5, 7, 6, 8
40 Subject	Age	Diastole	6, 13, 17 & 6,3,13	64.7	6, 7, 13, 8, 11
40 Subject	Obesity	Diastole	20, 11, 2	58.8	2, 15, 11, 23, 12
80 Subject	Sex	Systole	19	54.1	23, 20, 28, 22, 27
80 Subject	Age	Systole	9, 5	59.5	9, 20, 22, 6, 29
80 Subject	Obesity	Systole	15, 16, 6	67.6	15, 20, 30, 16, 13
80 Subject	Sex	Diastole	6, 23	56.8	6, 5, 25, 29,20
80 Subject	Age	Diastole	6, 3, 2	73.0	6, 3, 5, 28, 17
80 Subject	Obesity	Diastole	20, 11, 2	67.6	2, 11, 3, 13, 26
95 Subject	Sex	Systole	19	64.1	29, 27, 24, 19, 22
95 Subject	Age	Systole	9, 5	35.9	10, 19, 24, 6, 7
95 Subject	Obesity	Systole	15, 29, 16	64.1	16, 17, 30, 14, 19
95 Subject	Sex	Diastole	6, 23	69.2	6, 22, 7, 23, 30
95 Subject	Age	Diastole	6, 13, 17	76.9	6, 3, 29, 15, 7
95 Subject	Obesity	Diastole	20, 11, 2	74.4	2, 20, 11, 13, 24
121 Subject	Sex	Systole	19	55.4	18, 29, 28, 21, 26
121 Subject	Age	Systole	9, 5 & 9,18	41.5	20, 4, 7, 23, 10
121 Subject	Obesity	Systole	15, 29, 16	63.1	20, 15, 18, 13, 16
121 Subject	Sex	Diastole	6, 23	50.8	15, 7, 6, 11, 14
121 Subject	Age	Diastole	6, 3, 13	64.6	27, 7, 14, 6, 8
121 Subject	Obesity	Diastole	20, 11, 2	61.5	2, 26, 13, 24, 27

close relationship between the shape features of importance, even if the classification percentages do not reflect such a close agreement. Nevertheless, it is surprising that the omission of a single subject caused such a noticeable impact on the age classification results. A likely explanation relates to the relatively small amount of modal energy captured by these lower order modes. Although they clearly contribute important age-related shape information, describing such a small amount of the total modal energy (less than 1%) increases the possibility that small changes result in specific modes encoding different shape features between tests.

4.4.2 Robustness to Changes in Dateline Selection

The next test was to compare robustness to the selection of new poles and a new dateline. As with the analysis in Subsection 4.4.1, firstly the distance between the input values were assessed, before considering that impact on the output. Figure 4.2 gives a qualitative example of the degree of variability of point selection.

Impact of Changes to Dateline Input

Table 4.8 contains the Hausdorff Distances in mm between the original dateline selection and the second dateline selection at both end-Systole and end-Diastole. Hausdorff distance represents the largest distance between any point on the two datelines, so is indicative of a worst case scenario. The greatest distance between dateline selections was 6.56mm, and the smallest distance observed was 2.09mm. Overall, the distance between the initial and repeat datelines was greater at end-Diastole (range 3.81-6.56) than end-Systole (range 2.09-3.27). This variation is likely to be associated with the greater size of the diastolic ventricle, and is also influenced by the differences in the lengths of the selected datelines, which is dependent on each observer's determination of apex positioning. In the case of Subject 2 at end-Diastole, the second attempt to select the south pole led to the dateline containing 2 fewer points. Although the short-axis images provide a frame of reference, user input was still required regarding mapping the segmentation edge in ITK-SNAP to the smoothed finite

Table 4.8: Hausdorff Distance between the two datelines at both end-Systole and end-Diastole.

Subject	Diastole	Systole
2	6.56	2.88
3	4.97	2.09
4	3.81	3.27
10	3.95	2.32
22	5.13	2.62

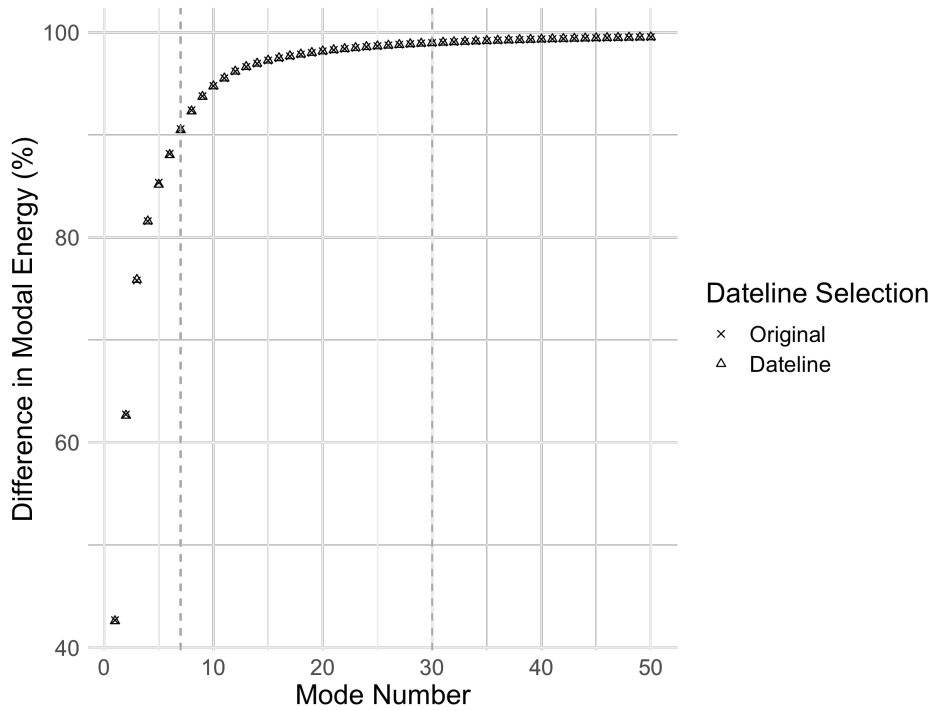
element mesh. Selection is also constrained by the location of nodes on the mesh; by necessity, it is limited to the nearest available node.

Impact of Changes to Dateline Output

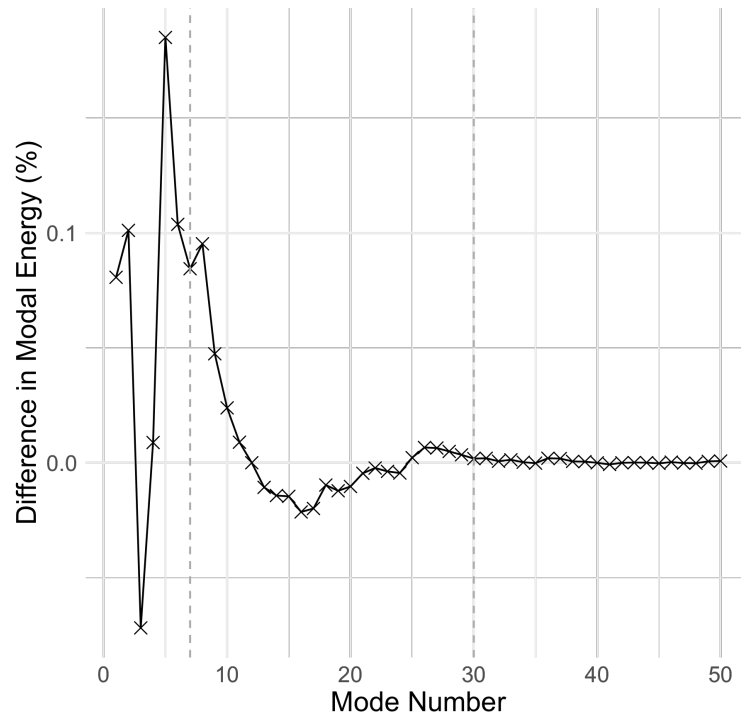
Having established that the differences between dateline selection can be up to 6.5mm, Tables 4.9, 4.10 and 4.11, now examine the impact of that dateline selection on the final statistical shape analysis results, since altering the dateline alters the position of the anatomical boundary when mapping to the unit sphere. Figures 4.7a and 4.7b show the modal energy curve for the shape analysis conducted after the new dateline selection was made. There is extremely close alignment in percentage modal energy between the modes derived in the original dataset and after the new dateline selection. Changing the selected dateline impacts the shape mapping. This accords with expectation, since the meshes themselves were not altered.

Table 4.9 compares the Pearson correlation coefficients between the original and replaced dateline shape analyses at end-Systole and end-Diastole. The first 13 modes are strongly correlated (the weakest correlation is $r_{94} = 0.817$, and all others are above $r_{94} = 0.9$). It is possible that Modes 14 and 15 encode the same features and have interchanged, since they both show uncharacteristically weak correlation coefficients in comparison to adjacent modes ($r_{94} = 0.384$ for Mode 14 and $r_{94} = 0.355$ for Mode 15). Therefore, as with the robustness to dataset size analyses, if this swap is taken into account, the first 16 modes are matched to cognates.

The sum of the squared difference between modes derived from the original and the new



(a)



(b)

Figure 4.7: Difference in modal energy distribution between dateline sets. **a)** Modal energy profiles for shape analyses conducted when a new dateline selection was made. The vertical dashed lines indicate Mode 7, which covers 90% of the total modal energy, and Mode 30, which describes 99% of the total modal energy. **b)** Difference between the original modal energy curve and those for shape analyses conducted when the new dateline selection was made.

Table 4.9: Absolute Pearson correlation values for the first 20 modes when comparing the 96 original subject mode coefficient values with the results obtained during a test where five of the 96 subjects had their datelines replaced.

Mode No.	Replaced Dateline	
	Systole	Diastole
1	0.988	0.993
2	0.943	0.936
3	0.961	0.974
4	0.817	0.991
5	0.946	0.974
6	0.942	0.984
7	0.914	0.992
8	0.944	0.953
9	0.967	0.981
10	0.953	0.986
11	0.969	0.965
12	0.938	0.975
13	0.958	0.995
14	0.384	0.347
15	0.355	0.316
16	0.971	0.960
17	0.133	0.026
18	0.027	0.133
19	0.453	0.564
20	0.509	0.457

Table 4.10: Sum of squared difference between the top 5 modes and Modes 6, 9 and 19 when comparing the original mode eigenvectors to those created when 5 new datelines were included.

Comparison	Sum Squared Diff.
Mode 1	3.34×10^{-5}
Mode 2	8.46×10^{-5}
Mode 3	8.46×10^{-5}
Mode 4	1.04×10^{-4}
Mode 5	1.09×10^{-4}
Mode 6	7.14×10^{-5}
Mode 9	4.71×10^{-5}
Mode 19	5.89×10^{-4}

dateline selections is shown in Table 4.10. These values were calculated for the first 5 modes in addition to Modes 6, 9 and 19; assessed as a result of their importance in defining sex and age-based differences, respectively. This analysis diverges slightly from the comparison of dataset sizes in Subsection 4.4.1, largely owing to the number of dataset analyses conducted in that section. Interestingly, the difference in relative value for Modes 6 and 9, which are key shape modes for determining sex and age-based differences, was much smaller than Modes 4 and 5, suggesting that these noteworthy shape features are preserved. Although, as posited in Subsection 4.4.1, it appears that the sum of squared difference is a less useful metric for determining the robustness of results; such differences are not upheld in the Pearson correlation analysis in Table 4.9.

Table 4.11 shows the Mode 1 coefficient values for the first 15 subjects and their associated percentage differences. The raw coefficient values are more prone to change than measures of mode similarity, which accounts for some very large percentage difference values. It is noteworthy that, while overall, the two shape analysis runs are similar, the subjects with altered datelines (Subjects 2, 3, 4 and 10 in this sample) exhibit substantially larger percentage differences. Subject 3, in particular, differs by 132%. The average percentage difference for the five replaced datelines is 69.7%, whereas the average for the remaining subjects is just 4.47%. It is surprising to note that the greatest percentage difference was not observed in Subject 4;

Table 4.11: Mode coefficient values at Mode 1 when comparing the original 96 subject mode coefficient values at end-Systole to the five subjects with redrawn datelines.

Subject Number	Original Result	Replaced Dateline	Percent Difference
1	56.6	-55.2	2.5
2	-32.5	24.4	24.8
3	24.3	-56.3	132
4	36.3	-49.6	36.7
5	89.1	-89.9	0.9
6	50.8	-49.7	2.3
7	54.1	-54.4	0.6
8	-16.4	18.4	12.5
9	17.9	-16.5	7.5
10	33.3	-61.9	86.0
11	30.8	-31.9	3.5
12	22.9	-22.5	1.8
13	52.6	-52.7	0.2
14	12.7	-10.9	14.3
15	-18.1	18.7	3.1

the individual with the largest Hausdorff distance (and thereby the largest absolute difference in the dateline selections).

Table 4.12 shows the mean, maximum, minimum and standard deviation values of the percentage differences calculated for Modes 1, 6, 9 and 19. This information is also displayed as a box plot in Figure 4.8. The box plots are transformed onto a Log_{10} scale to ensure that all points are recorded, since there are a significant number of outliers and removing them would preclude a true representation of the data. Table 4.12 suggests that the mean percentage differences of Modes 1, 6 and 9 are comparable at end-Diastole (range: 16.0%-19.6%), and end-Systole (range: 44.2%-74.2%). The mean of Mode 19 is substantially greater at both phases (248.5% at end-Diastole and 330.0% at end-Systole). Figure 4.8 also demonstrates that all tests show a substantial spread across the range of 0-10,000%. Overall the percentage difference values for the dateline analyses are more closely clustered than those for the dataset tests, with the exception of the very minor 95 and 94 subject changes.

Table 4.12: Mean, minimum, maximum and standard deviation of the percentage difference values which compare results from selected modes from the new dateline dataset against the original 96 subject dataset from Chapter 3.

	Mode 1		Mode 6		Mode 9		Mode 19	
	Dia	Sys	Dia	Sys	Dia	Sys	Dia	Sys
Mean	19.6	57.4	16.0	44.2	18.2	74.2	248.5	330.0
Min	0.1	0.0	0.3	0.1	0.1	1.0	1.3	0.4
Max	908.7	4140.7	928.2	2144.4	810.8	3944.8	6892.4	9549.8
SD	120.0	424.0	100.4	250.6	112.6	435.5	951.6	1364.2

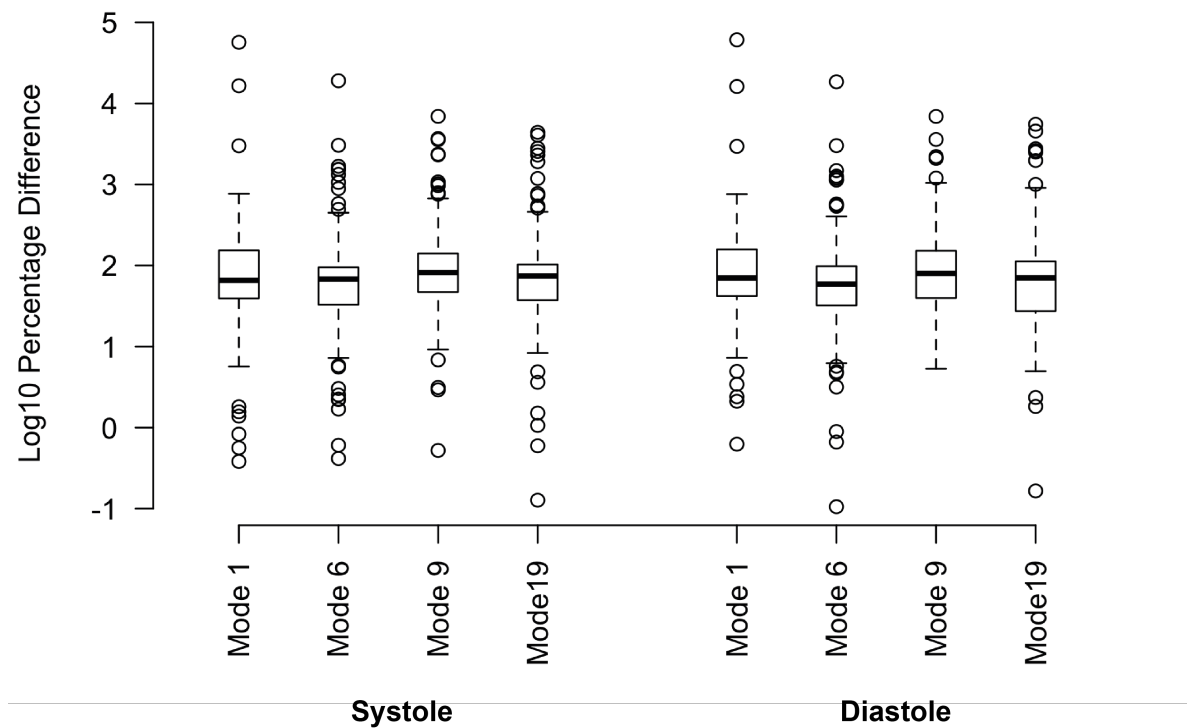


Figure 4.8: Box plots showing the distribution of the percentage difference values at end-Systole and end-Diastole for the four example modes. Plots are on a Log_{10} scale to enable the display of outliers without distorting the data.

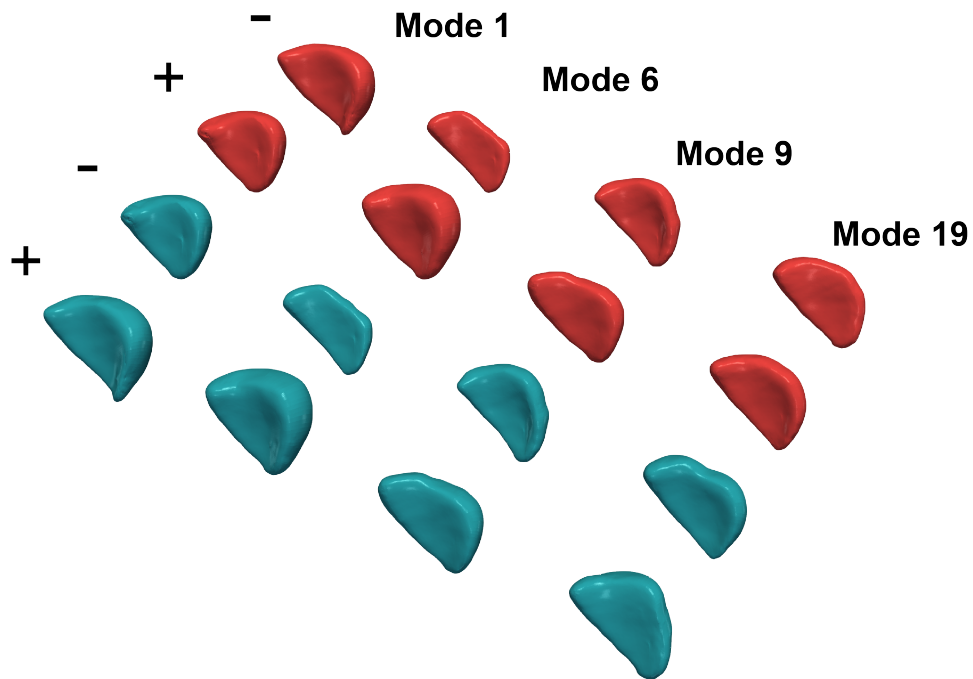


Figure 4.9: Qualitative depiction of the RVES shapes when four key modes: 1, 6, 9 and 19 in original and new dateline datasets are plotted on the $\text{Mean} \pm 2 \times \text{SD}$. The original RVES shapes are coloured blue, whereas the dateline replacement sets are coloured red. Modes 1, 6 and 9 appear to encode similar shape features when the dateline replacement sets. Mode 19 does not appear to encode the same shape in the dateline replacement set. As the modes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. Hence, a mode can be observed to encode the same shape whether that mode is added or subtracted from the mean.

Figure 4.9 depicts the mode shapes for Modes 1, 6, 9 and 19 in both the original and new datasets plotted on the $\text{mean} \pm 2 \times$ standard deviation of the modal coefficients experienced by every subject. Mode 1 again exhibits characteristic increased length and detail in the apex region, which is associated with the MRI scanning protocol. Modes 1, 6 and 9 all show the same shape in the original (red) and dateline-replaced (blue) sets. However, Mode 19 does not appear to encode the same mode features. This qualitative result agrees with the Pearson correlation coefficients which are shown in Table 4.9. The shape modes remained robust for the first 16 ranked features in terms of their Pearson correlation coefficients, assuming that Modes 14 and 15 are switched.

Table 4.13: Exemplar LDA Analyses comparing classification results between the original analysis and the tests which include a new dateline. Test Combination reports the combination of modes used as inputs in the LDA. Where more than one combination was found equally successful in Chapter 3, both were tested. ‘Classification %’ reports the best accuracy percentage after cross validation. ‘Top Five’ lists the best five modes identified using a new LDA, created based on the new coefficients from the robustness test.

Characteristic	Phase	Test Combination	Classification %	Top Five
Sex	Systole	19	25.0	27, 8, 17, 7, 18
Age	Systole	9, 5	70.0	5, 19, 23, 9, 14
Obesity	Systole	15, 16, 6	65.0	19, 29, 13, 18, 6
Sex	Diastole	6, 23	60.0	6, 23, 28, 7, 24
Age	Diastole	6, 13, 17 & 6, 3, 13	65.0	6, 22, 18, 7, 29
Obesity	Diastole	20, 11, 2	72.5	11, 24, 13, 2, 26

Application of LDA

LDA analyses comparing classification results between the initial 96 subjects and the dateline replacement analysis are displayed in Table 4.13. Obesity at end-Diastole (Modes 20, 11 & 2) proves the most discriminatory characteristic. The dateline robustness test shows a 72.5% classification rate, compared to 77.5% in the initial decomposition (c.f. Table 3.3). In the dateline comparison, Modes 11 and 2 feature in the top five modes based on structure matrix value. Overall, the classification percentages recorded in Table 4.13 are closer to the original test results, with the notable exception of sex at end-Systole (of which the best initial discriminator was Mode 19 only). In this test, only 25% of hearts were successfully classified. The conspicuously poor performance of the LDA classifier could be attributed to several causes. The conclusions of Chapter 3 and the qualitative images shown in Figure 4.9 display a prominent ‘hump’ feature encoded by Mode 19. This feature is not replicated in the dateline robustness test. Although, Mode 18’s presence in the ‘Top Five’ selection indicates that again, this noticeable basal feature was represented, even if not described by exactly the same mode. Such switches become increasingly likely in the higher order modes, and therefore the identification of matching features, still shows the usefulness of this decomposition.

Wu et al. (2013) determined that the harmonic mapping method was sensitive to dateline

selection. This also appears to be true in the following analysis too, as a relatively small shift in dateline positioning (of the magnitude observed in Figure 4.2) has a large effect on modal coefficient results in those subjects which were changed. However, there is no large difference in the overall shape analysis result, especially not on the scale of the inter- and intra-observer variability analyses.

4.4.3 Robustness to Inter and Intra-observer Variability

Neither of the interventions implemented up until this point have made specific alterations to the finite element mesh. This Subsection will introduce a much more substantial change to the hearts included in the analysis. Differences in the input when the initial MRI image segmentations are performed by the same observer after a long hiatus or by an entirely new observer are assessed using three measures. The first, a Jaccard Index, is designed to look at the overlap of the selections on a per-pixel basis; the second is a volume calculation, examining the overall impact of the segmentation choice on the size of the RV mesh and finally a CV% compares the relative variation between and within observers. Jaccard Indices and CV% are described in detail in Section 4.3.5.

Impact of Changes to Observer Input

First, Jaccard Indices for each slice are displayed in the heatmaps in Figure 4.10. These figures draw attention to the regions with the greatest and smallest differences in pixel overlap between the two MRI slices. Specifically, it is evident from Figures 4.10a, 4.10b and 4.10c that pixel selection is much more consistent in diastolic hearts than systolic hearts. Darker green and blue shades are indicative of an overlap 50%-70% of pixels, whereas areas in yellow are in the region of 90% overlap across segmentation slices. The consistency at end-Diastole is almost certainly based on the larger coverage area at end-Diastole, in combination with the more obscure and difficult to identify apical slices at end-Systole, especially in the subjects from UPMC. An overall Jaccard index is also reported in Table 4.14 for each heart surface

that was tested.

Table 4.14 reports overall Jaccard Indices, which are a measure of pixel overlap comparing slice by slice segmentations for each RVES tested. The table displays overall percentage overlap, averaged across all slices. The range of recorded Jaccard Index values was 48.0-81.3%. The lowest overall Jaccard overlap was 45.6%, indicating a total match of less than half of the selection. This value was observed at end-Systole, comparing the two observers' initial segmentation attempts. Table 4.14 demonstrates that, in general, pixel overlap is greater in end-Diastolic segmentations than end-Systolic segmentations. The combined data from Table 4.14 and Figure 4.10) provides strong evidence that manual segmentation can be extremely subjective.

Table 4.14: Overall Jaccard indices comparing pixel overlap across 5 subjects when comparing Observer 1 and Observer 2's initial and repeat segmentations. This value is calculated by averaging Jaccard Indices across all slices for each subject; the denominator is the number of slices.

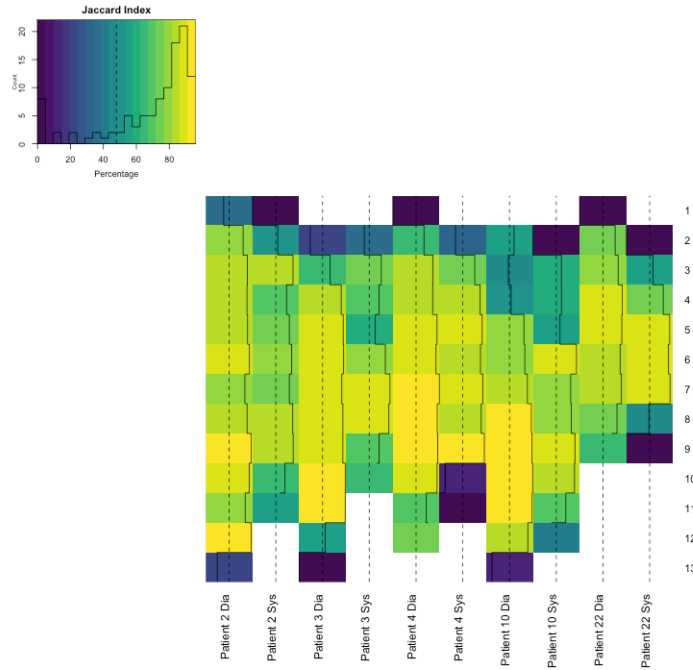
Observer	Subject 2		Subject 3		Subject 4		Subject 10		Subject 22	
	Dia	Sys	Dia	Sys	Dia	Sys	Dia	Sys	Dia	Sys
1 Initial vs 1 Repeat	77.3	66.0	71.9	69.8	76.7	63.9	72.5	63.3	70.7	55.2
2 Initial vs 2 Repeat	74.3	64.1	78.4	52.5	76.8	70.3	79.8	65.6	65.5	48.0
1 Initial vs 2 Initial	75.3	62.9	78.2	49.6	78.0	68.4	73.2	71.9	68.9	54.5
2 Initial vs 1 Repeat	77.5	64.9	79.9	45.6	72.0	58.1	81.3	61.2	77.5	59.6

Subject 10, overall was the most consistently matched RV across all repeats at end-Systole, with a Jaccard Index range of 63.3%-71.9%; the highest overall index for an end-Systolic subject, and Subject 22 was the least consistently matched across all repeats, with a maximum overlap of less than 60% (range 48.0%-59.6%) (Table 4.14), although Subject 3 depicted the overall lowest degree of similarity (45.9%). The noticeably lower Jaccard Index scores exhibited by Subject 22 can be explained by the relatively poor quality MRI scan, which made the segmentation choice difficult. Figure 4.11 depicts a single slice of this segmentation, and the initial segmentations by Observer 1 and Observer 2. The differences in the segmentations, even in a central slice demonstrate how impact of scanning conditions can affect an observer's decision on segmentation boundaries. In the central image, the observer has interpreted the

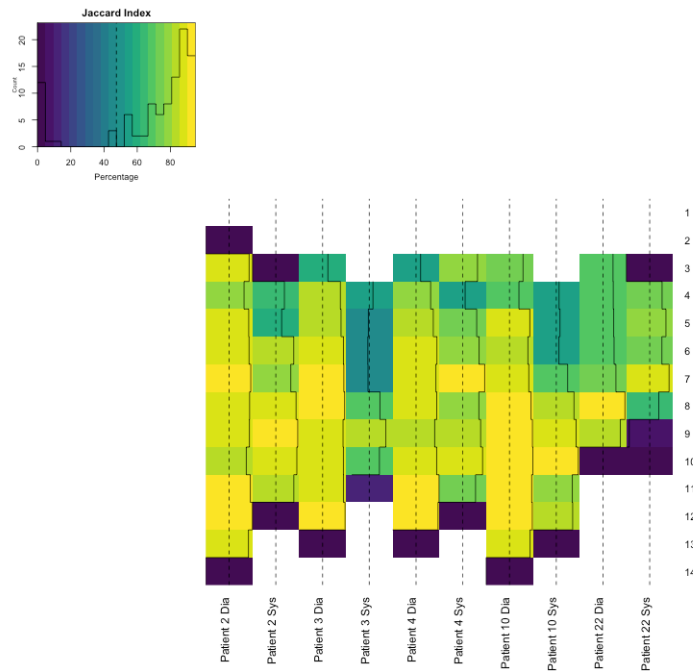
dark, trabeculated region around the free-wall border as being within the endocardial surface, and has thus selected that region in the segmentation, whereas the observer in the right hand image has interpreted that same region as being part of the RV wall, and has therefore drawn a more conservative boundary. The segmentation standard is to include RV trabeculations in the ventricular cavity, i.e. within the segmentation (Petitjean et al., 2015). Subject 3 at end-Systole demonstrated the smallest degree of similarity between repeat segmentations. The first segmenter appears to have produce more internally consistent results (overlap range of 55.2%-77.3% compared to 48.0%-79.8%). It is notable that the experience level of the two segmenters in this analysis was more comparable to the beginner reader in Sardanelli et al. (2008), and therefore consistency could be improved by the employment of more experienced professionals if manual segmentation is to be used as the standard input. Further, a set of ‘ground rules’ could be established to avoid inconsistencies such as that described above and observed in Figure 4.11. For example, any trabeculation should be considered to be within the endocardial surface.

Figure 4.12 shows two attempts to segment Subject 4 at end-Systole; the surface with the largest pre-shape analysis variation between individuals. Although many of the same anatomical attributes are identified by both observers, for example the noticeable twist half way down the septal wall, there is some disagreement (and therefore lack of pixel overlap) in the base region. This is marked by an arrow on Figure 4.12. The central (6th and 7th) slices are most similar in the repeat analyses performed by the first observer.

Table 4.15 compares measurements of mesh volume obtained from Simpleware. The table reports raw values and percentage differences. Inter-observer variability was greater at end-Systole than at end-Diastole, and this was especially evident in the first set of comparisons. For example, there was a 91.6% difference in end-Systolic segmentation volume for Subject 22, as compared to a 24.9% difference at end-Diastole. Overall, there was a more substantial inter-observer difference in the first set of segmentations by each observer (range 1.83-91.6%), in comparison to the second set, which exhibited a range of 0.38-22.61% difference. In terms

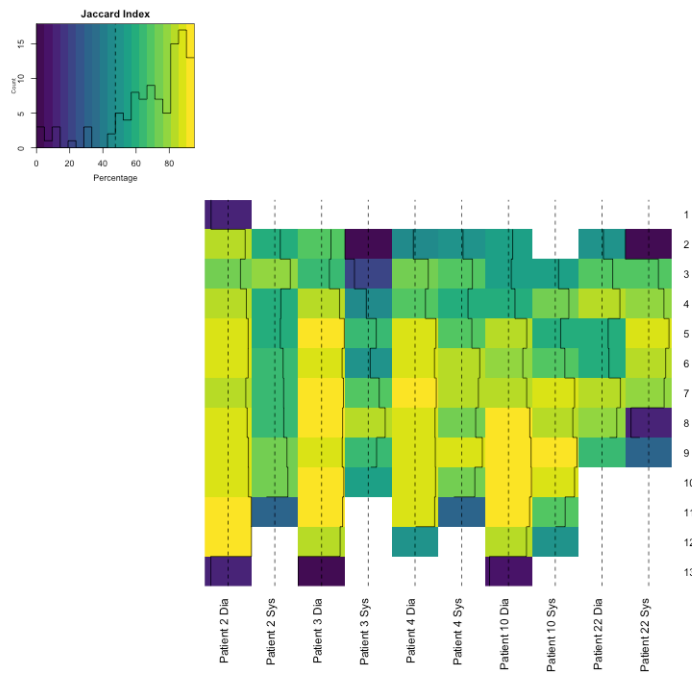


(a) Observer 1 Initial vs Repeat



(b) Observer 2 Initial vs Repeat

of intra-observer variability, Observer 2 demonstrated greater intra-observer variability than Observer 1. For example, the percentage difference in volume for Subject 22 was 56.8%



(c) Observer 1 vs Observer 2

Figure 4.10: Heatmaps displaying the calculated Jaccard Indices between Observer 1 and Observer 2 for inter and intra-observer. A white region indicates no pixel overlap between slices, dark blue indicates a low overlap (20-30%), teal indicates a 50% overlap and yellow indicates a percentage pixel overlap in excess of 90%. Columns represent individual subject RVs, and rows represent slices of those RVs. The vertical black lines are trace lines. The distance of each line from the centre of the cell is proportional to the size of the measurement.

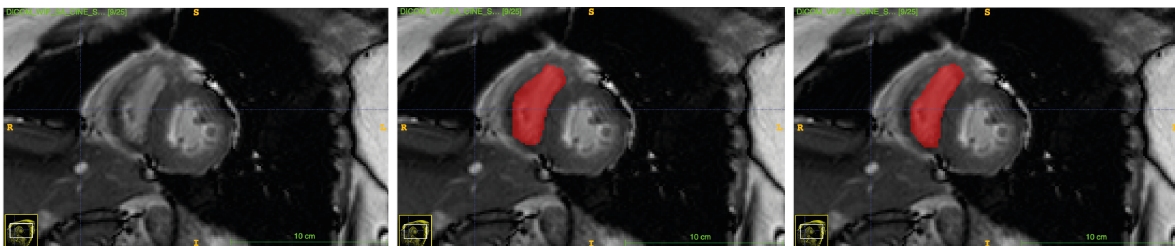


Figure 4.11: Example of a segmentation on a poor contrast MRI slice. The LHS shows an unmarked slice, the central image shows an attempt by Observer 1 and the RHS shows a segmentation attempt Observer 2.

between the two segmentations performed by the second observer, as compared to 5.34% when segmented by the first observer. All of these volume differences accord with regions of poor overlap identified using Jaccard Indices (Table 4.14), although the Jaccard Indices, in general show higher agreement than would be suggested through consideration of the volumes alone.

The difference in volume is likely exacerbated by the way in which ITK-SNAP (Yushkevich et al., 2016) deals with slice thickness. If no pixels are selected on a particular slice, then 8mm depth in the z axis is automatically lost in that region. This is depicted in Figure 4.12, which shows 3D models selected by both observers where one segmentation unintentionally missed the apex or bottom most slice, thereby reducing overall volume. Such a loss could amount to a volume change of approximately 2% by itself. Removing the bottom slice segmentation of Subject 4 in ITK-SNAP resulted in a volume change, as recorded in that software, of 1.15cm^3 or 1.39% of the total volume. As can be observed by the Jaccard Index heatmaps, the low overlap was concentrated in these tricky top and bottom slices, whereas the agreement was much greater in the central regions, where the yellow colour indicates a pixel overlap in excess of 90%. Personal correspondence with one of the cardiologists involved in the initial data collection at UPMC, confirmed that the pulmonary valve region is an area of low frequency of pathological shape change. Therefore, the segmentation of this region could be excluded to prevent unnecessary differences.

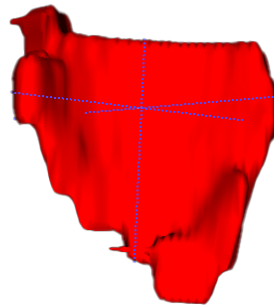
It is important to consider that the volume measurements in Table 4.15 were recorded after smoothing had taken place. Calculating robustness metrics using the post-smoothing volumes gives a closer representation of the size of ventricular meshes in the statistical shape analysis procedure, but also means that the smoothing process has the potential to remove more material in some areas than others, especially near the apex and base of the ventricle. The recursive Gaussian filter applied in this particular analysis operates by applying the smoothing filter to three adjacent pixels. Assessment of the impact of smoothing filters is discussed in greater detail in Subsection 4.4.4, as altering the number of impacted pixels can remove

Table 4.15: Raw inter- and intra-observer segmentation volume measurements (cm^3) reported for Observer 1 and Observer 2 and their associated percentage differences, reported for selected subjects.

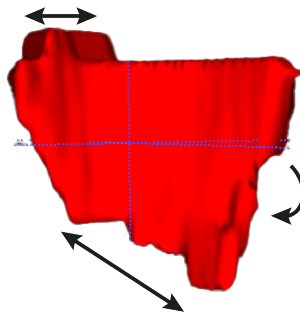
	Observer 1		Observer 2		Inter-Obs Sys	Inter-Obs Dia
Set 1	Systole	Diastole	Systole	Diastole	Set 1 % Diff	% Diff
2	113.9	242.7	134.5	220.8	18.09	-9.02
3	78.1	171.8	93.3	166.0	19.46	-8.78
4	84.6	138.9	93.0	126.7	9.93	-3.38
10	62.8	147.7	83.2	145.0	32.48	-1.83
22	26.2	64.3	50.2	80.3	91.60	24.88
Set 2	Systole	Diastole	Systole	Diastole	Set 2 % Diff	% Diff
2	101.1	217.5	83.7	236.4	-17.21	8.69
3	95.1	178.8	73.6	165.4	-22.61	-7.49
4	63.8	130.7	77.5	131.2	21.47	0.38
10	72.5	138.7	69.0	142.2	-4.83	2.52
22	27.6	72.6	21.7	67.3	-21.38	-7.30
	% Diff		% Diff			
	Systole	Diastole	Systole	Diastole		
	Intra-Obs 1		Intra-Obs 2			
	-11.24	-10.38	-37.77	7.07		
	21.77	4.07	-21.11	-0.36		
	-24.59	-5.90	-16.67	3.55		
	15.45	-6.09	-17.07	-1.93		
	5.34	12.91	-56.77	-16.19		

required geometry.

The volume comparisons and CV% displayed in Tables 4.15 and 4.16 indicate a relatively large difference in volume measurement, both between the two observers and within their two separate measurements. Both observers were less consistent at end-Systole than at end-Diastole. Inter-observer variability was greater than the intra-observer variability: a mean difference of 34.3% at end-Systole, and 28.0% at end-Diastole was achieved in the first set, compared to a mean intra-observer variability of 15.7% at end-Systole and 7.87% at end-Diastole for Observer 1, and 29.9% at end-Systole and 5.82% at end-Diastole for Observer



(a)



(b)

Figure 4.12: Example 3D segmentation volumes produced from ITK-SNAP. These volumes are produced prior to smoothing. Both surfaces depict segmentation attempts of the same RVES. **a)** is the segmentation attempt by Observer 1 and **b)** is the attempt by Observer 2. The sharp twist half way down is observable in both attempts and is marked by the curved arrow on the 3D surface in **b)**. Notable differences between the two segmentations also exist, and are also marked by arrows. For example, there is a greater transverse difference between the dateline region and the apex in the segmentation attempt by Observer 2, and there is a broader RV outflow tract vehicle in the same segmentation. If no region is selected on a given slice, then 8mm of volume in the z-direction is not included.

2. It is noteworthy that the data was skewed with particularly large values in individual segmentations, e.g that of Subject 22 which suffered from unclear MRI images. The smallest overall percentage difference was the intra-observer difference between Observer 2's sets at end-Diastole. There was a noticeable improvement in inter-observer variability in the second set of segmentations, which perhaps alludes to an improvement in skill. The difference in input data for these segmentations was larger than seen elsewhere in the literature. Most notably in the intra/inter-observer volume values reported in Couto et al. (2020) where the mean difference in end-Diastolic volume did not exceed 1.1% and the mean difference in ESV did not exceed 1.9%. This paper did have the advantage of simultaneously comparing the long-axis, 4 chamber view which increased difficulty of observing the boundaries at the apex and base. This problem has been debated in the literature (e.g. Clarke et al., 2012; Alfakih et al., 2003). Although, Couto et al. (2020) did not find a significant difference between a four-chamber segmentation method and short-axis only. Their imaging protocol involved a 6mm skip 0mm slice selection, which may have improved similarity at the apex slices. Sardanelli et al. (2008) compared manual contouring between two readers with vastly differing levels of experience and concluded that RV systolic volume agreement was lower for both the beginner segmenter (an engineer who had performed 25 cardiac segmentations in preparation) and the radiologist. They report a bias of -3.9% for the less experienced reader and 2.3% for the interobserver value, although the experienced radiologist reported an intra-observer bias value of 0.0%, where bias is the mean of the differences between the two measurements. The scanning protocol was very similar to the Newcastle University setup, with an 8mm slice thickness and 0mm inter-slice gap. One clear outcome, illustrated from all of the pre-shape analysis measures was that the segmentation process employed in this thesis yielded far less consistent segmentation results than observed in the literature.

Table 4.16 reports the mean, median and range of means of coefficient of variation (CV%) for each individual segmentation attempt. CV% is defined in Section 4.3.5, and reports the standard deviation of a pair of measurements, divided by the mean value for each subject. The

Table 4.16: Means of coefficient of variation (CV%) for inter- and intra-observer variability. The reported values constitute the mean, median and range across the five subjects assessed by each observer.

	Observer 1	Observer 2	Obs 1 vs Obs 2
Diastole			
Mean CV %	9.20	7.88	6.05
Median CV %	9.2	6.3	3.49
Range CV %	20.3	8.13	17.3
Systole			
Mean CV %	11.2	26.3	19.0
Median CV %	10.1	16.7	20.9
Range CV %	16.1	43.2	23.6

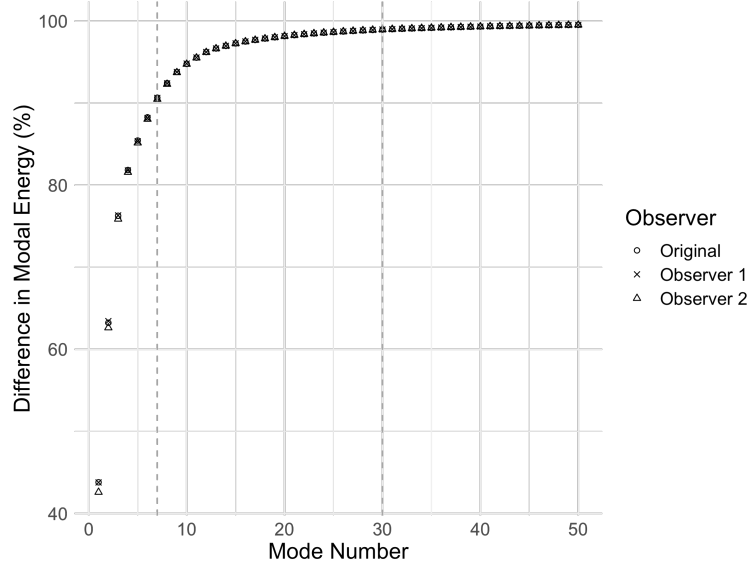
values reported in this table show that both observers have a lower median CV% than mean CV% (10.1% vs 11.2% for Observer 1, and 16.7% vs 26.2% for Observer 2), which indicates that the results are influenced by a few large outliers. The ranges of CV% again demonstrates that Observer 2 is less consistent than Observer 1 in their input segmentations. The CV% values for inter- and intra- observer comparisons, suggest that these are slightly more consistent than the repeats performed by Observer 2. End-Diastolic measurements are more accurate than end-Systolic measurements in all cases, which again is likely related to the larger blood pool volume, especially in the central segmentation slices. These values are somewhat worse than the literature, although in most cases this was largely a result of outliers, particularly in those RV scans with boundaries that were difficult to interpret. Caudron et al. (2012), for example, report an end-Diastolic range of CV% between 9.3-13.1% at end-Diastole and 12.0-16.7% at end-Systole for intra-observer variability using a similar methodology. Their inter-observer variability ranges were 11.9-14.5% at end-Diastole and 12.4-20.9% at end-Systole. This study enrolled 60 consecutive RVs, and assessed the inter and intra-observer variability of three observers of differing experience. One significant cause of lower accuracy in the results obtained in this thesis is likely to relate to segmentation using short-axis images only, whereas Caudron et al. (2012) had access to long axis and 4-chamber views.

Impact of Changes to Observer Output

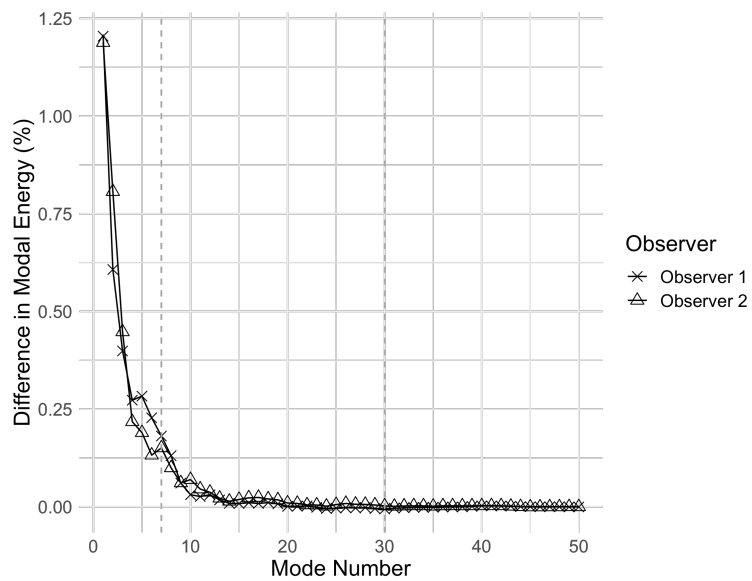
Having established that there is a larger than anticipated variation in ventricular input mesh both between and within observer, the following tables (Tables 4.17, 4.18, 4.19, 4.20 and 4.21) document the impact of this variability on the output. The results demonstrate the effect of this variation both qualitatively, through examining similarities between the results obtained in the population analysis conducted in Chapter 3, and quantitatively through examining the modal energy profile. The percentage difference in output coefficients and the statistical correlation between the modal values obtained when the different segmentations are also included.

Figure 4.13 reports the modal energy curves achieved when the segmentations are performed by different observers. As with the robustness to changes in dataset size, the curves indicate that there is only a small difference in modal energy curve profile when segmentations are performed by different observers or as repeats by the same observer. The first 7 modes still describe 90% of the modal energy and the first 30 modes describe 99% of the modal energy. The first three modes of the Observer 1 and Observer 2 repeats have a slightly higher cumulative eigenvalue total than the original, but this does not affect the results. Figure 4.13b plots the percentage difference in cumulative modal energy between the original and the two Observer repeats. This plot seeks to clarify the differences between shape analysis results. In both cases, the maximum percentage difference from the original values does not exceed 1.2%. The modal energy difference for Observer 1 is very slightly greater between Modes 4 and 8 as compared to Observer 2, although these variations are on the order of 0.1%.

Table 4.17 reports the sum of squared difference values between the mode eigenvectors obtained by each observer for a series of modes, which were found to be relevant to specific population features in Chapter 3, a form of direct comparison between the mode shapes. The difference was smallest between the Mode 1 eigenvectors across all Observer trials. This result is consistent with expectation, given that the Modes are ranked in order of their contribution to the overall shape, and also since that Mode 1 clearly represents the impact of MRI scanner protocol, which was found to be the most dominant effect on ventricle shape in this thesis.



(a)



(b)

Figure 4.13: Distribution of modal energy when comparing segmentations from two observers. **a)** Modal Energy Curve Displaying the Cumulative Modal Energy required to explain the shape in the initial shape analysis and when including five subjects re-segmented by each observer. **b)** Plot showing the difference between the original modal energy curve and the repeats with each observer. Observer 1 differences are marked with an X and Observer 2 differences are marked by a triangle.

Table 4.17: Sum of squared difference between Modes 1, 6, 9, 19 and the various observers to the original shape analysis result.

Comparison	Sum Squared Difference			
	Mode 1	Mode 6	Mode 9	Mode 19
Observer 1 - Observer 1	2.13×10^{-5}	9.54×10^{-5}	1.03×10^{-4}	2.62×10^{-4}
Observer 2 - Observer 2	2.76×10^{-5}	8.23×10^{-5}	8.99×10^{-5}	1.68×10^{-4}
Observer 1 - Observer 2	2.16×10^{-5}	6.16×10^{-5}	8.31×10^{-5}	2.82×10^{-4}

Similarly, the general trend is that the sum of squared difference increases as the mode number increases. However, evidence from the analysis where the dataset size was reduced by 1, from 96-95 individuals (Table 4.4) indicates that the sum of squared difference values plateaus at the value, even when an almost imperceptible change is made to the dataset.

Table 4.18 compares Pearson correlation values for the first 20 modes between the two observers and also when each Observer's segmentation values are correlated with the original data.

The Pearson Correlation values delineated in Table 4.18 provide strong evidence that, despite a large difference in the individual modal coefficient values in the replaced RVs, this has had no impact on the shape modes produced. The Pearson correlation values remain greater than ± 0.9 with the exception of a few individual cases, and no value is less than $r_{94} = 0.765$. There is a strong correlation between those values identified as Mode 19 $r_{94} = 0.873$ in Observer 2's segmentation attempt. The Pearson correlations remain both consistent and very strong throughout the 20 modes included in this assessment. This is contrary to expectation, since it was anticipated that large changes such as introducing a new individual to perform the segmentation would result in larger changes in modal coefficient, and therefore weaker correlations. Equally, as shown in the Jaccard Heatmaps (Figure 4.10), there appeared to be a relatively large difference in the input to this shape analysis result.

Table 4.19 reports the raw mode coefficient values for Mode 1 obtained as part of the statistical shape analysis procedure. The table shows the modal coefficient values when Observer 1 is compared against the original, Observer 2 is compared against the original and

Table 4.18: Absolute Pearson correlation values for the first 20 modes when comparing the 96 original mode coefficients with those obtained during tests where 5 of the RVs were replaced by new segmentations.

Mode No.	Observer 1	Observer 2	Obs 1 vs 2
1	0.926	0.988	0.968
2	0.933	0.977	0.955
3	0.986	0.981	0.977
4	0.972	0.994	0.988
5	0.951	0.987	0.966
6	0.971	0.978	0.959
7	0.958	0.959	0.980
8	0.971	0.975	0.940
9	0.946	0.973	0.968
10	0.972	0.987	0.987
11	0.831	0.993	0.858
12	0.936	0.996	0.943
13	0.981	0.977	0.996
14	0.898	0.871	0.854
15	0.952	0.861	0.932
16	0.964	0.949	0.956
17	0.765	0.918	0.887
18	0.779	0.892	0.887
19	0.907	0.905	0.930
20	0.962	0.928	0.929

when the two repeats are compared against one another. As the Newcastle subjects included in the original analysis were segmented by the first observer, the Observer 1 repeat analysis can also be taken as a comparison to the original 96 subject dataset used in Chapter 3. While the overall correspondence between shape analyses is good over the 96 subject comparison, the coefficients themselves are significantly different in the replaced hearts (Subjects 2, 3, 4, 10 and 22). The smallest percentage difference in one of the replaced segmentations was -37.9% (range 37.9% - 133.3%), and this is substantially greater than the largest percentage difference in a case where the mesh was not altered (13.9%).

It is obvious from Table 4.19 that the segmentation models Subjects 2, 3, 4 and 10 were

Table 4.19: Raw coefficient values for Mode 1 obtained in the statistical shape analysis procedure for when comparing the original 96 subject mode coefficients with those obtained when 5 repeat subjects were included by each observer.

Subject Number	Original Result	Observer 1	Observer 2	% Diff. 1	% Diff. 2
1	56.6	-55.2	-57.2	-2.4	1.2
2	-32.5	-0.5	-14.6	-98.4	-55.0
3	24.3	-15.1	-51.5	-37.9	112.2
4	36.3	-11.2	-74.2	-69.1	104.5
5	89.1	-87.8	-91.1	-1.4	2.2
6	50.8	-49.4	-51.5	-2.8	1.3
7	54.1	-52.4	-55.9	-3.1	3.3
8	-16.4	17.7	16.1	7.7	-1.7
9	17.9	-16.3	-18.7	-8.7	4.6
10	33.3	-77.6	-56.2	133.3	69.0
11	30.8	-30	-32.8	-2.6	6.7
12	22.9	-21.7	-24.1	-5.2	5.4
13	52.6	-50.9	-54.3	-3.4	3.1
14	12.7	-10.9	-13.2	-13.9	3.5
15	-18.1	20.3	17	12.2	-5.9

included by the new observer, and these show a much larger percentage difference in correlation than the other hearts. This suggests that the process of mode generation can be replicated, but that the coefficients themselves are quite sensitive to change based on segmentation. It is difficult to isolate whether this change is also related to dateline selection, since the generation of an entirely new mesh necessitates the selection of new dateline points. However, the impact of dateline selection assessed in 4.3.3 suggests that this effect is only minor. Nevertheless, Table 4.19 demonstrates that often there are examples where there is close agreement between the raw coefficient values (e.g. the three absolute Mode 1 coefficient values: 56.6 vs 55.2 vs 57.2).

Table 4.20 describes mean, minimum, maximum values of percentage difference between raw modal coefficients as well as the standard deviation for the series of four ‘marker modes’ described earlier in the results section as contributing towards a particular feature in Chapter

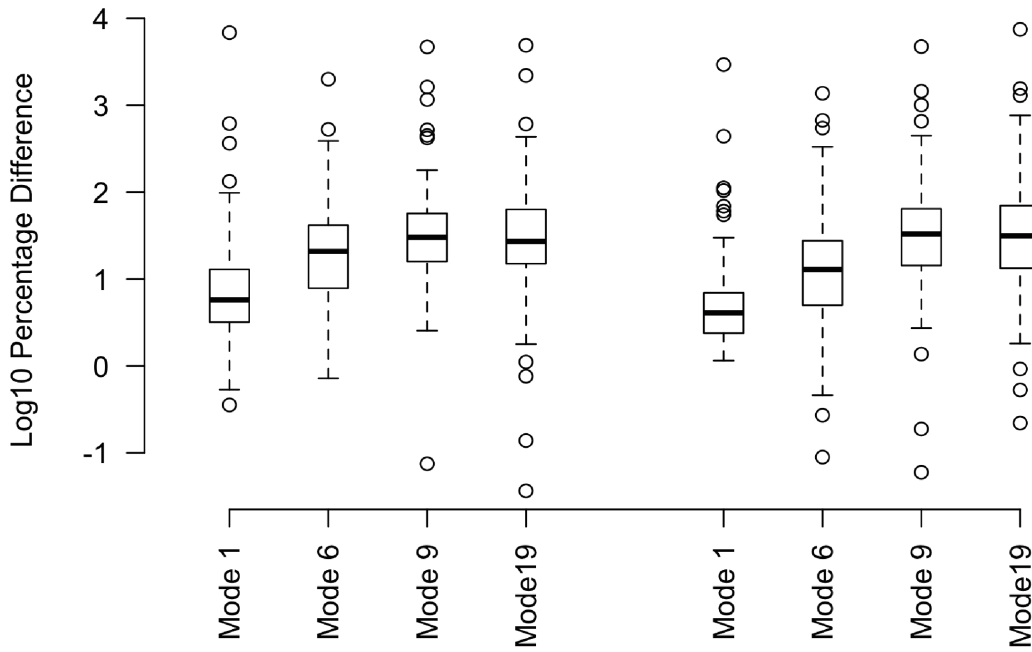
Table 4.20: Mean, minimum, maximum and standard deviation of the percentage difference values comparing the original 96 subject dataset with the shape analysis run between the two observers for Modes 1, 6, 9 and 19.

		Mode 1		Mode 6		Mode 9		Mode 19	
		Dia	Sys	Dia	Sys	Dia	Sys	Dia	Sys
Observer 1	Mean	61.4	83.0	14.7	38.7	14.9	89.1	0.1	96.4
	Min	0.4	0.4	0.0	0.7	0.0	0.1	0.5	0.0
	Max	1649.3	6859.1	971.8	1997.7	576.6	4697.4	8583.9	4903.6
	SD	263.0	698.1	105.2	213.8	89.2	514.1	935.1	543.9
Observer 2	Mean	24.0	37.5	20.2	30.0	14.5	85.7	0.1	137.2
	Min	0.1	1.2	0.2	0.1	0.6	0.1	0.9	0.2
	Max	1197.6	2942.0	868.2	1375.0	771.7	4740.6	9478.1	7485.7
	SD	133.0	301.1	108.4	165.9	95.9	510.6	1039.5	785.8

3. These results report the difference between the first segmentation attempt by each observer and the original. From these output values, it appears that the output coefficient results for Observer 1 show greater difference than those reported for Observer 2 (e.g. a mean of 61.4% vs 24.0% for Mode 1 at end-Diastole). The same values are plotted on a Log_{10} scale in Figure 4.14 to enable observation of the general trends and outliers. In all cases, Mode 1 demonstrates the highest degree of consistency, and the lowest percentage difference. The most substantial takeaway from this specific form of analysis, however, is that there are often very large deviations in modal coefficient value (9478% for Mode 19 at end-Diastole), and that resultantly raw modal coefficient values should not be taken by themselves as a record of the success of a statistical decomposition. Therefore, the percentage difference in coefficients cannot alone be used to assess the reproducibility of the shape analysis method under different input conditions.

Figure 4.15 shows the same important mode shapes (1, 6, 9 and 19) in the original shape analysis, for the Observer 1 repeat segmentation and for both Observer 2 analyses. In all four examples, the mode shapes are almost identical. The recurrence of such shapes indicates

Systole



Diastole

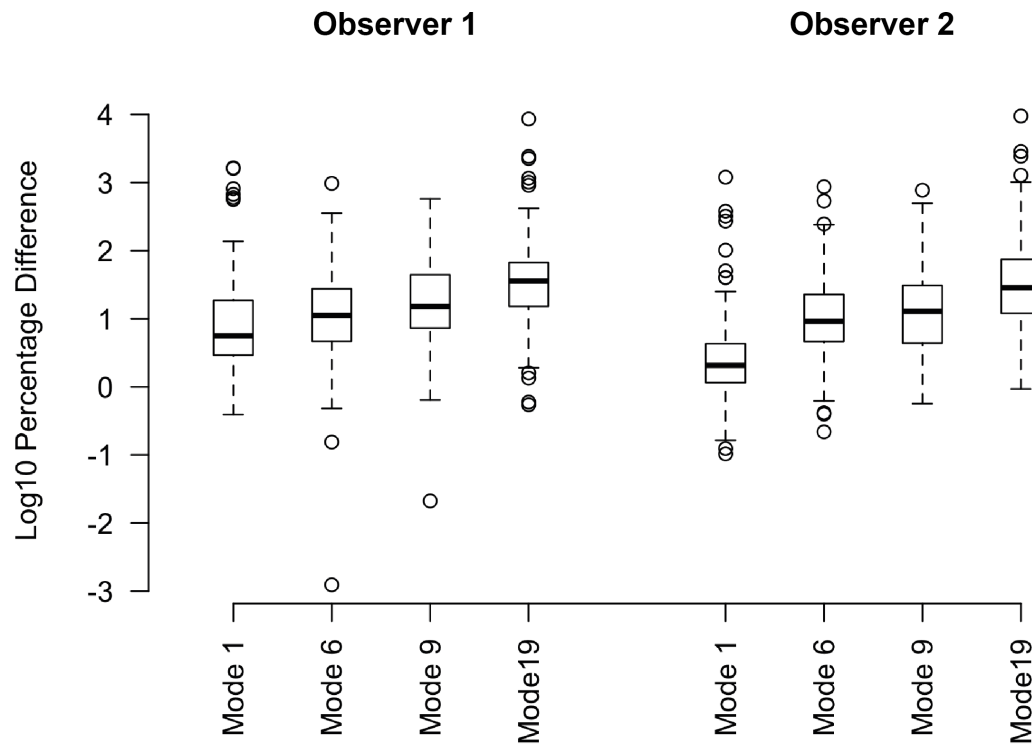


Figure 4.14: Box plots showing the distribution of the percentage difference values at end-Systole (top) and end-Diastole (bottom) for the four example modes. Plots are on a Log_{10} scale to enable the display of outliers without distorting the data.

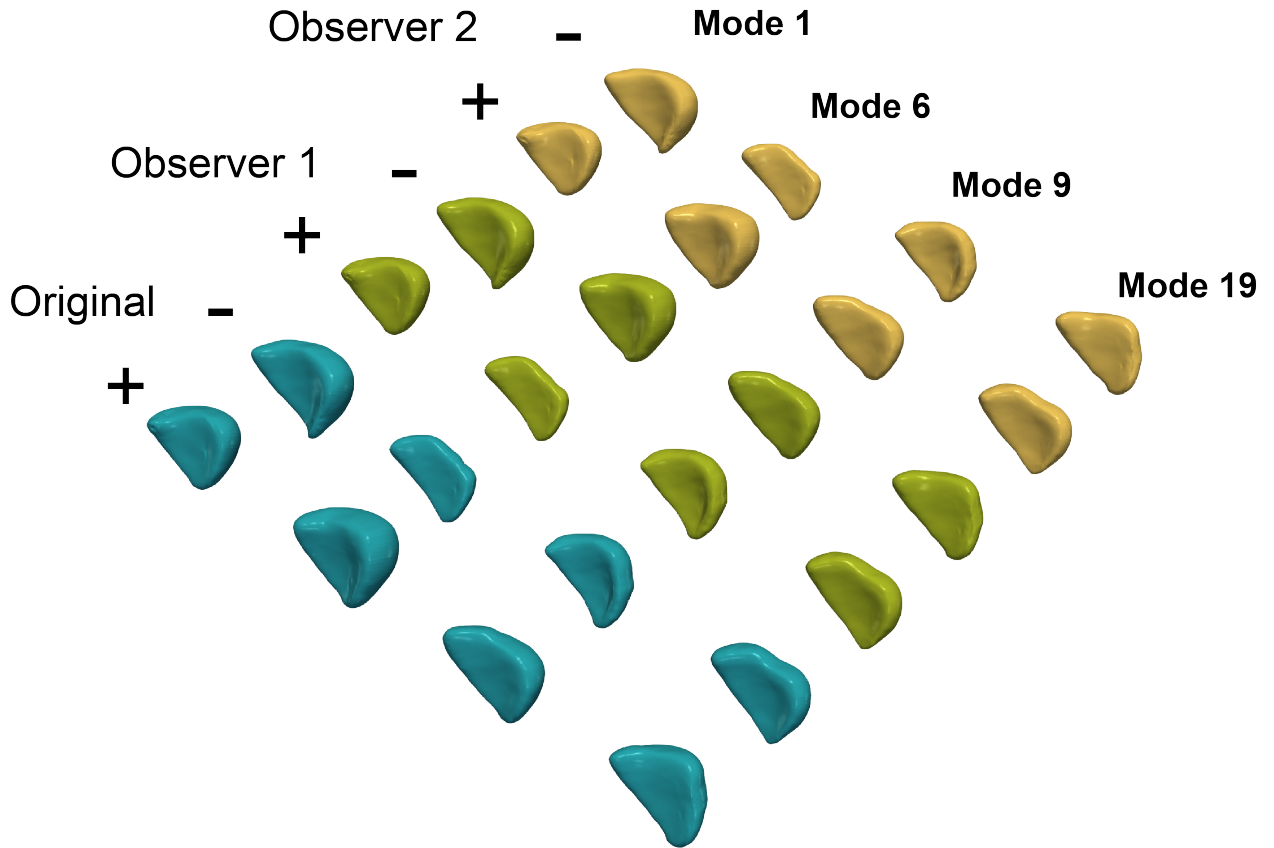


Figure 4.15: *Qualitative depiction of the RVES shapes when four key modes: 1, 6, 9 and 19 in original and the Observer 1 and Observer 2 repeats are plotted on the $\text{Mean} \pm 2 \times \text{SD}$. The original RVES shapes are coloured blue, whereas the dateline replacement sets are coloured red, the mode shapes from Observer 1 are coloured green and the mode shapes from Observer 2 are coloured yellow. Modes 1, 6, 9 and 19 appear to encode similar shape features across all three trials. As the modes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. Hence, a mode can be observed to encode the same shape whether that mode is added or subtracted from the mean.*

that, qualitatively at least, the shape analysis results are robust to alterations in the segmented volume, and resultant mesh. This result accords with the Pearson Correlation coefficients shown in Table 4.18, where the absolute Pearson Correlation values exceed 0.9; including at Mode 19. The close Pearson correlation values raise an interesting point regarding relative differences, as the maximum percentage differences from Table 4.20 (specifically for Mode 19 at end-Diastole) are in excess of 8000%, yet remain correlated to one another.

Table 4.21: Exemplar LDA Analyses comparing classification results between the original analysis and the repeats testing inter and intra-observer variability. For the sake of brevity, this analysis depicts just one of the shape analysis results between observers. Test Combination reports the combination of modes used as inputs in the LDA. Where more than one combination was found equally successful in Chapter 3, both were tested. ‘Classification %’ reports the best accuracy percentage after cross validation. ‘Top Five’ lists the best five modes identified using a new LDA, created based on the new coefficients from the robustness test.

	Characteristic	Phase	Test Combination	Classification %	Top Five
1	Sex	Systole	19	60.0	23, 24, 20, 21, 28
1	Age	Systole	9, 5 & 9, 18	52.5	26, 8, 17, 21, 4
1	Obesity	Systole	15, 29, 16	57.5	27, 2, 26, 17, 19
1	Sex	Diastole	6, 23	27.5	17, 26, 14, 16, 18
1	Age	Diastole	6, 13, 17	67.5	19, 7, 6, 11, 12
1	Obesity	Diastole	20, 11, 2	40.0	12, 14, 9, 3, 21
2	Sex	Systole	19	57.5	8, 6, 28, 7, 21
2	Age	Systole	9, 5	70.0	29, 9, 17, 23, 27
2	Obesity	Systole	15, 16, 6 & 15, 29, 16	65.0	15, 17, 13, 29, 20
2	Sex	Diastole	6, 23	67.5	23, 28, 6, 7, 22
2	Age	Diastole	6, 3, 2	67.5	6, 3, 15, 29, 25
2	Obesity	Diastole	20, 11, 2	72.5	20, 11, 24, 13, 2

Application of LDA

Finally, an LDA analysis was performed to identify the ability of this new statistical model to linearly classify the important modes which were identified in the Population Analysis Chapter (Chapter 3), the results of which are shown in Table 4.21.

There are several examples of close agreement, where the top separating modes identified in Chapter 3 achieved similar degrees of success in the Observer trials. This is particularly evident for Observer 2 at end-Systole, where Modes 9 and 5 were able to correctly separate 70.0% of individuals into age-based groups, which indicates that the subjects in the dataset are constructed from similar amounts of these shape modes. Modes 20, 11, 2 at end-Diastole were able to identify 72.5% of cross-validated cases correctly by obesity status. The LDA using Modes 9 and 5 as inputs successfully separated 70.0% of ventricles by age, but Modes 9 and 18 separated 60% of individuals by age. Mode 19 was able to successfully classify just

57.5% of non-hypertensive hearts into groups based on sex after cross-validation.

Comparatively, when Observer 1's attempts were analysed using LDA, Mode 19 at end-Systole successfully separated 60% of subjects by sex. The two best classifiers based on age (Modes 9 & 5 and Modes 9 & 18) together separated just 52.5% of individuals. Mode 19 at end Systole was just 27.5% successful at separating individuals by sex. Modes 20, 11 and 2 at obesity could separate 40% of individuals and Modes 6, 3 and 13 for obesity could separate 60% of right ventricles. Mode 19 did not correlate significantly with any cognate modes in the new model by Observer 1. There is a very low degree of successful classification of sex at end-Diastole in the Observer 1 repeats. Therefore, the internal consistency of Observer 1, demonstrated by the Jaccard Indices, is not reproduced in the output. In general, the effect of sex on RV shape is better elucidated in these inter- and intra-observer variability trials as compared to in dataset size or dateline selection.

It is important to note that this test has only replaced 5 subjects, rather than a whole sample. Therefore, it is now pertinent to consider the impact that adding or reducing the number of subjects included in the analysis will have on the outputs using the same measurement criteria. Given the non-trivial inter and intra- observer variability scores, it is surprising how robust the qualitative and quantitative shape analysis results are in comparison.

4.4.4 Robustness to Changes in Smoothing

The final perturbation assessed in this chapter was the degree of smoothing filter applied. The number of pixels impacted by the recursive Gaussian smoothing filter implemented in Simpleware ScanIP (Synopsis, Exeter, UK) was reduced to 2 pixels at each step, and increased to 4 pixels.

Impact of Changes to Smoothing Input

The recursive Gaussian smoothing filter applied during the meshing process has a substantial qualitative impact on the mesh structure. Figure 4.16 displays the qualitative mode shapes,

when the smoothing filters are altered. It is apparent that the least stringent smoothing filter σ_2 contains significantly more detail at the apex, there appears to be a small cleft at the bottom, whereas the apex detail is lost entirely at the equivalent region when the strictest smoothing filter, σ_4 is applied. This is reflected in the degree of mesh complexity, where the σ_2 contains 6162 elements; significantly more than the 3682 elements required to mesh the runs with the largest amount of smoothing. While reducing the smoothing coefficient minimises data-loss, particularly around the apex slices, it also maintains the ‘step-like structure’ which is an artefact of the slice-based imaging protocol and is not reflective of the true-life shape of the right ventricle. Conversely, a more stringent smoothing filter removes features critical to the shape analysis.

It should be noted that the visualisations rendered in Simpleware (Figure 4.16) are often reflections of the true RV segmentations. The software is unable to determine the correct orientation of the slices produced in ITK-SNAP (whether to infer the 2D image as top down or bottom up), which can result in the RV segmentations appearing to be inverted. This has no bearing on the production of the final mesh or to downstream application to shape analysis.

Impact of Changes to Smoothing Output

Having established that there is a significant qualitative impact associated with altering the rigour of the smoothing filter, the impact of these alterations to the shape analysis output were then assessed. Figures 4.17a and 4.17b compare the modal energy curves of the original 96 subjects with the σ_3 smoothing filter to the σ_2 and σ_4 smoothing tests. Yet again, it is clear that the three tests share a very similar modal energy profile. Although, there is a small amount of variation in the first three modes (more so than with interventions that did not impact the shape of the mesh), 7 modes are still required to cover 90% modal energy. As with the changes to dataset size, the smoothing is causing a substantive difference to the shape, as the filter is necessarily adding or removing shape features as it resolves the pixel slices. Therefore, it is yet again an endorsement of the robustness of the procedure that such little

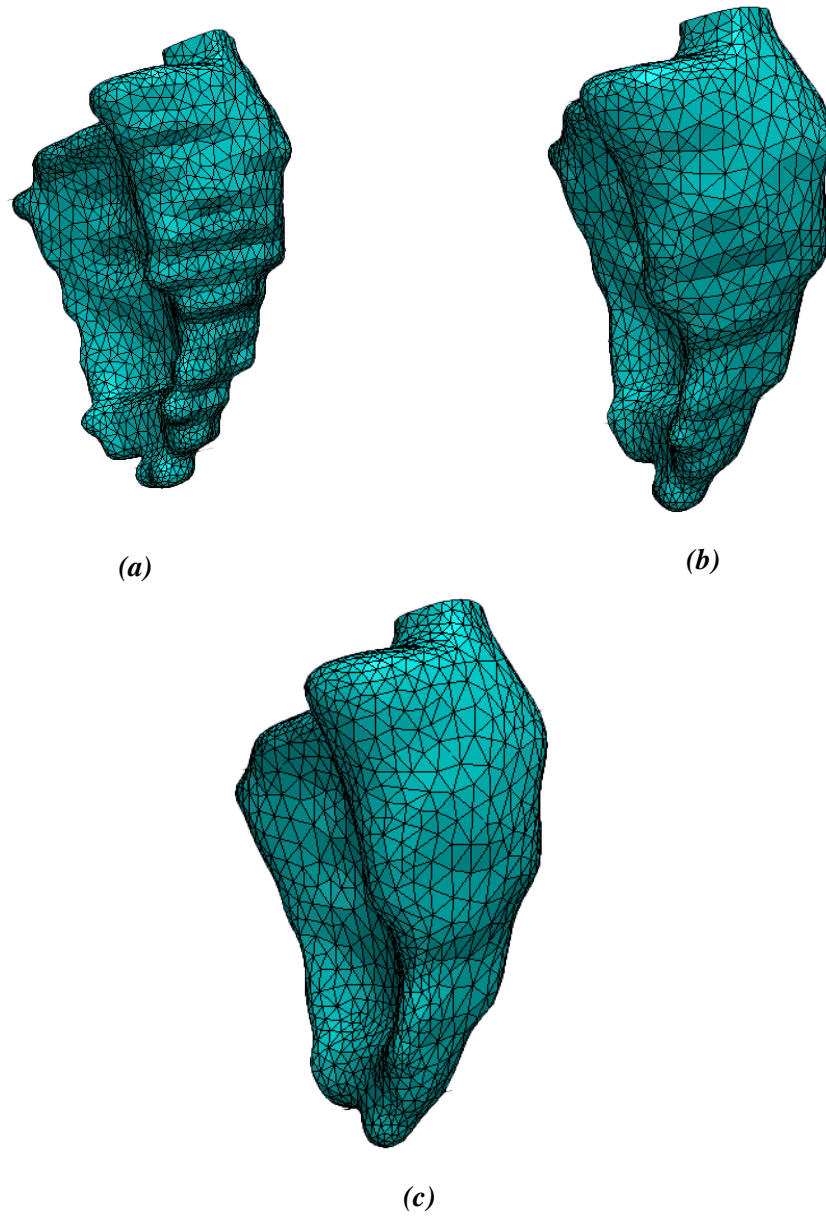
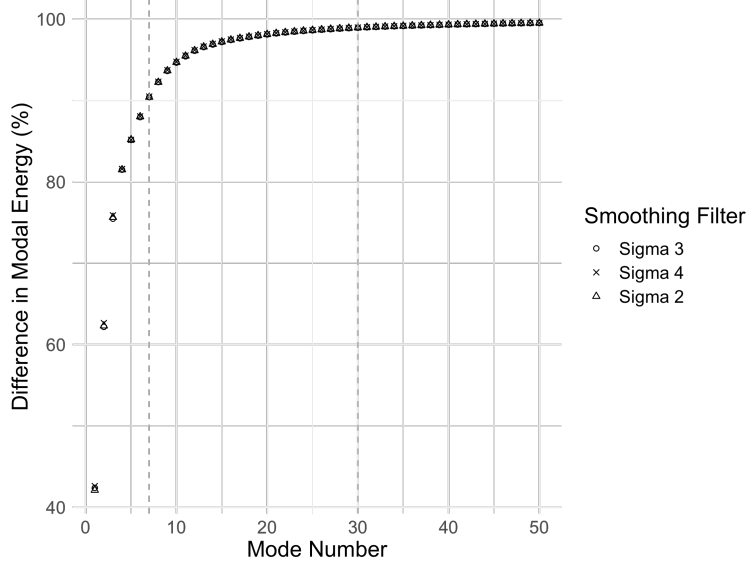


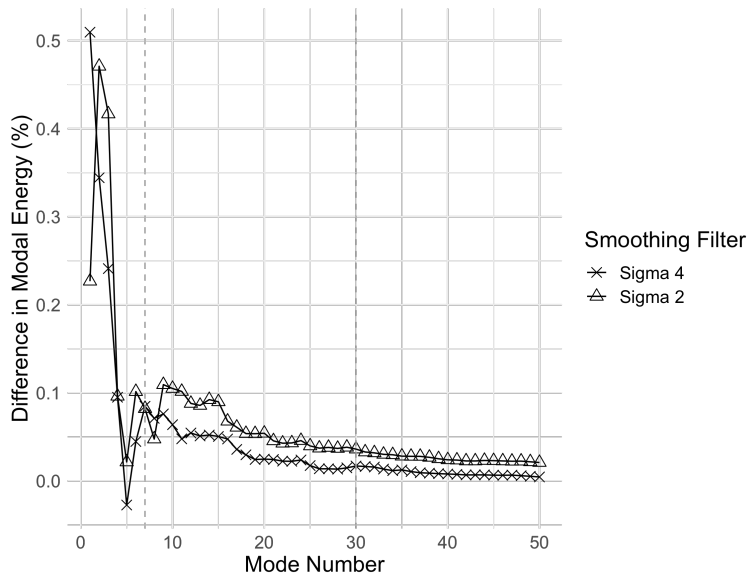
Figure 4.16: Visualisation of the different levels of smoothing applied when the three different pixel filters are included. **a)** $\sigma 2$ smoothing contains 6162 elements, **b)** $\sigma 3$ smoothing contains 4378 elements and **c)** $\sigma 4$ smoothing contains 3682 elements. This Figure was reproduced in the methodology as Figure 4.3, to demonstrate the effect of smoothing filters.

difference in the modal energy profile exists. The higher variation in the first three modes is perhaps indicative of the addition/removal of shape information that comes from impacting the Gaussian smoothing filter.

To assess impact on the modal coefficients (the strength of the shape feature), Table 4.22



(a)



(b)

Figure 4.17: Distribution of modal energy when different smoothing filters are applied. **a)** Modal energy profiles for shape analyses conducted when the recursive smoothing filter was altered to $\sigma 2$ pixels and $\sigma 4$ pixels. **b)** Plot showing the difference between the original modal energy curve and those for shape analyses conducted when the recursive smoothing filter was altered to $\sigma 2$ pixels and $\sigma 4$ pixels.

shows Pearson correlation values for the first 20 modes produced from an analysis where the recursive Gaussian smoothing filter stringency is increased to σ_4 and when the stringency is reduced to σ_2 . The output shape analysis differs much more substantially in this test than in all previous cases. There is no statistically significant case at end-Systole (with the best correlation coefficient value being 0.255). At end-Diastole, only Modes 6 and 7 are statistically significant, and even then only weakly so (Mode 6 for σ_2 is 0.363, and Mode 7 for σ_2 is 0.344). Even Mode 1, which would be expected to correlate the most strongly since the modes are ranked does not show a statistically significant Pearson correlation between either smoothing variant and the original. This suggests that the choice of smoothing filter is very important to the ability to draw conclusions on the role of shape features across various datasets, especially since, what appear to be cognate shapes are not being conclusively matched when the smoothing filter is altered. Although it can be argued that smoothing filter is too important a parameter to be arbitrarily chosen when intending to use this method to reflect intrinsic characteristics, the fact that results remain stable when different filters are applied instead suggests that the comparative shape analysis remains robust, so long as the same smoothing filter is applied to all meshes.

Figure 4.18 further bolsters this position by showing the mode shape features for key modes. There is clear, qualitative evidence that Mode 1 still affects RV length and apex detail in this robustness test. In fact, qualitatively, Modes 1, 6 and 9 encode the same shape features as their counterparts in the original shape decomposition. It is therefore surprising, given this evidence of similarity, that there is such a substantial difference in robustness as recorded by the Pearson correlations in Table 4.22).

One explanation for this unusual result is that the shape of Mode 1 (which encodes the largest proportion of variation) is substantially altered upon the inclusion of σ_2 and σ_4 subjects, which would then have a substantial knock on effect to all subsequent modes. However, Figure 4.18 demonstrates, at least qualitatively, that there is no change in the features captured by Mode 1. The most plausible explanation, therefore, is that higher order Modes, such as

Table 4.22: Absolute Pearson correlation values for the first 20 modes when comparing the 96 original subject mode coefficient values with those obtained during a test where the smoothing coefficients of five of the 96 subjects were altered to σ_2 and to σ_4 . Pearson Correlations were calculated to compare the σ_2 case to the original 96 subjects, the σ_4 case to the original 96 subjects, and finally σ_2 to σ_4 .

Mode No.	Systole			Diastole		
	σ_2	σ_4	σ_2 vs σ_4	σ_2	σ_4	σ_2 vs σ_4
1	0.209	0.255	0.986	0.149	0.093	0.986
2	0.015	0.003	0.985	0.044	0.035	0.985
3	0.139	0.093	0.986	0.053	0.027	0.986
4	0.219	0.224	0.957	0.016	0.018	0.957
5	0.036	0.031	0.970	0.135	0.125	0.970
6	0.237	0.225	0.967	0.363	0.347	0.967
7	0.224	0.192	0.960	0.344	0.321	0.960
8	0.029	0.040	0.981	0.037	0.007	0.981
9	0.102	0.114	0.975	0.027	0.007	0.975
10	0.024	0.002	0.956	0.144	0.160	0.956
11	0.026	0.030	0.938	0.153	0.141	0.938
12	0.039	0.063	0.930	0.153	0.112	0.930
13	0.084	0.012	0.937	0.276	0.269	0.937
14	0.016	0.089	0.114	0.164	0.032	0.114
15	0.113	0.101	0.179	0.007	0.005	0.179
16	0.129	0.077	0.930	0.118	0.089	0.930
17	0.269	0.260	0.925	0.045	0.123	0.925
18	0.023	0.041	0.903	0.125	0.082	0.903
19	0.018	0.047	0.928	0.140	0.118	0.928
20	0.032	0.019	0.933	0.170	0.218	0.933

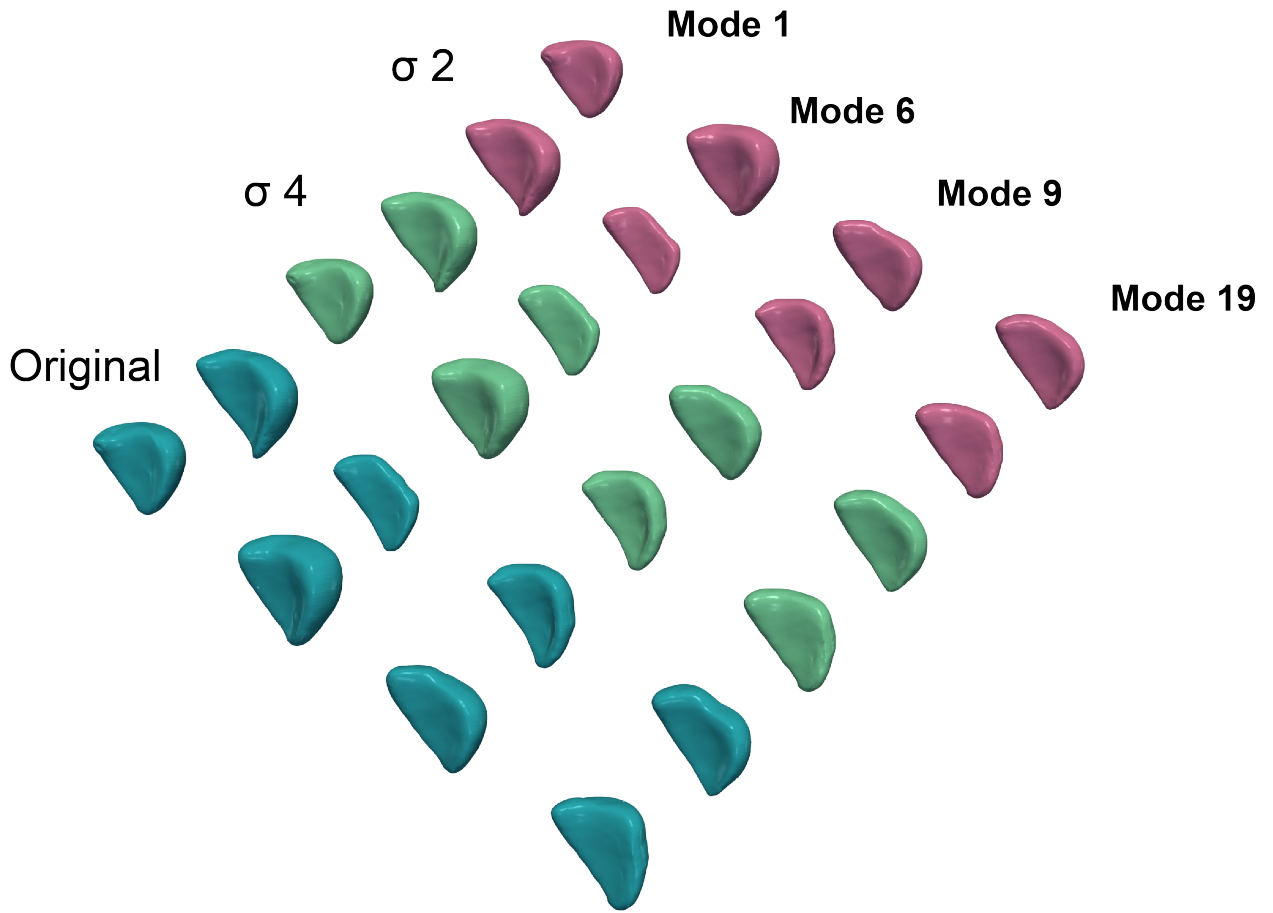


Figure 4.18: Qualitative depiction of the RVES shapes when four key modes: 1, 6, 9 and 19 in original and the $\sigma 2$ and $\sigma 4$ shape analysis repeats are plotted on the $\text{Mean} \pm 2 \times \text{SD}$. The original RVES shapes are coloured blue, the more stringent recursive Gaussian smoothing filter ($\sigma 4$) replacement sets are coloured mint green, and the less stringent smoothing filter ($\sigma 2$) shapes are coloured pink. Modes 1, 6 and 9 appear to encode the same shape features in all smoothing trials. The shape features encoded by Mode 19 share some similarities across the three smoothing filters (e.g. the flattened base, leading to a ‘triangular’ shape), however, there are also clear differences in the effect of the Mode. As the modes are eigenvectors, they provide only direction and their magnitude is arbitrary, including whether positive or negative. Hence, a mode can be observed to encode the same shape whether that mode is added or subtracted from the mean.

Modes 17 and 18 encode shape features related to the level of smoothing, and this then affects the modal coefficients experienced by individual subjects. Such an explanation addresses why the qualitative, and direct mathematical comparisons of mode shape are very similar, but that the comparisons of mode coefficient are different. Although σ_2 and σ_4 have opposing effects (increasing or decreasing smoothing), they have similar global impacts on RV shape, and are therefore likely to impact other modal coefficients in the same way, leading to much higher observed correlations between σ_2 and σ_4 , than between the altered smoothing tests and σ_3 .

Mode 1 still affects RV length and apex detail in this robustness test. There is greater detail at the apex in σ_2 (the heart with less smoothing), which is in agreement with expectation, since less material was removed when the filter was applied. Mode 1 encodes the same shape even when a greater or lesser level of smoothing is applied to the five subject sample. Furthermore, the lower energy modes clearly display the sharp shelves in the σ_2 analysis (with Modes 17 and 18 being representative examples). The alteration of certain lower energy modes to reflect the unsmoothed or over-smoothed shapes may, as a consequence, impact the relative amount of each mode contained within a given subject's RV. This would provide a plausible explanation as to why the Pearson correlations in Table 4.22 are very weak, but those between the σ_2 and σ_4 trials are similar to each other, i.e. that the same regions are affected by the smoothing level. It is very interesting that the end-Systolic and end-Diastolic Pearson correlations are identical in the σ_2 - σ_4 comparisons.

Correlation coefficient values for Mode 1 across the three smoothing levels were directly compared in Tables 4.23 and 4.25. The large percentage differences between Mode 1 and the original 96 subject case (with the exception of Subjects 1 and 15) corroborates the lack of significant correlation observed in Table 4.22. However, interestingly, the raw Mode 1 coefficients from the σ_2 and σ_4 repeats are very similar to one another. This does not align with any expectation related to the input files, since visually the original smoothing filter is closer to either σ_2 or σ_4 than the two extremes of smoothing are to each other. A maximum absolute percentage difference of 11.3% when comparing the two altered trials, e.g 240.3%

Table 4.23: Correlation coefficient values at Mode 1 when comparing the original 96 subject mode coefficient values at end-Systole to those where the smoothing values of five subjects were altered to 2 and 4 pixels

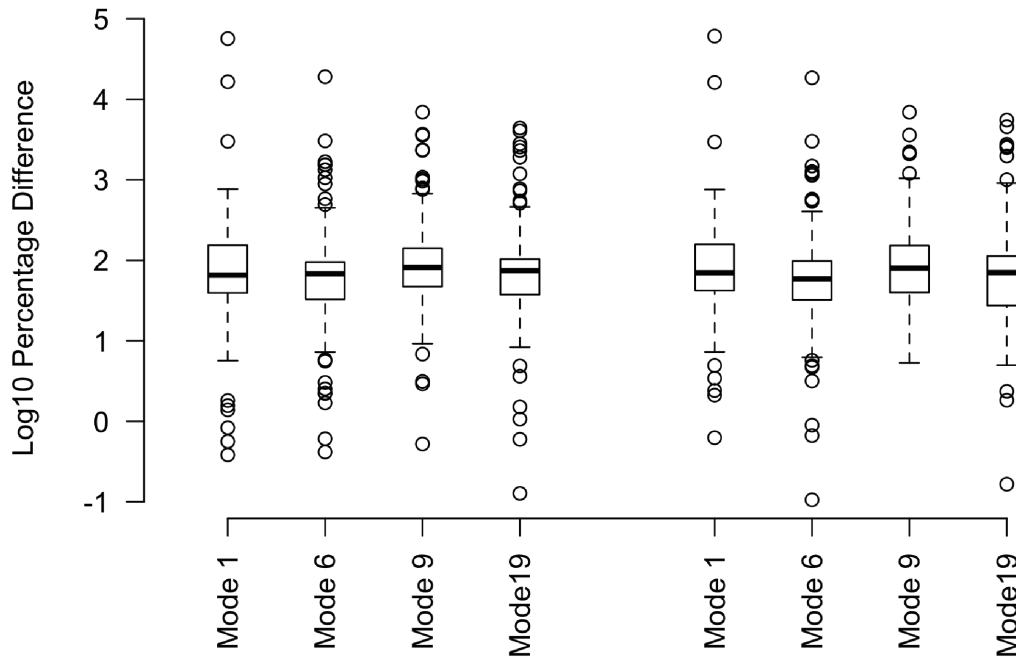
Subject Number	Original Result	$\sigma 2$	$\sigma 4$
1	-55.2	55.6	-56.2
2	-32.5	89.3	-90.0
3	24.3	50.0	-50.8
4	36.3	53.7	-55.2
5	89.1	-17.9	17.2
6	50.8	16.8	-17.7
7	54.1	31.5	-32.1
8	-16.4	22.7	-23.1
9	17.9	52.4	-53.5
10	3.3	11.2	-12.5
11	30.8	-18.8	17.5
12	22.9	34.5	-35.7
13	52.6	12.4	-13.5
14	12.7	-30.6	29.3
15	-18.1	-18.0	16.8

when $\sigma 2$ is compared with the original, and 278.8% when $\sigma 4$ is compared to the original.

Figure 4.19 displays box plots which show the distribution of the percentage difference values at both end-Systole and end-Diastole. They are reported on a Log_{10} scale to enable the display of outliers without distorting the data. This scaling factor is particularly important when assessing the impact of smoothing filter, since there are substantial differences in modal coefficient between the altered smoothing trials and the original result. The general trend shown from these box plots is that at end-Diastole, the percentage difference between shape analyses increases with increasing mode number, with Mode 1 exhibiting the lowest percentage difference, and Mode 19 the greatest. Although all direct measurements of raw modal coefficient reported in this thesis have demonstrated large changes between analyses (see Tables 4.6, 4.11, 4.20 by way of example.)

Table 4.25 reports the percentage differences between the raw mode coefficients described

Systole



Diastole

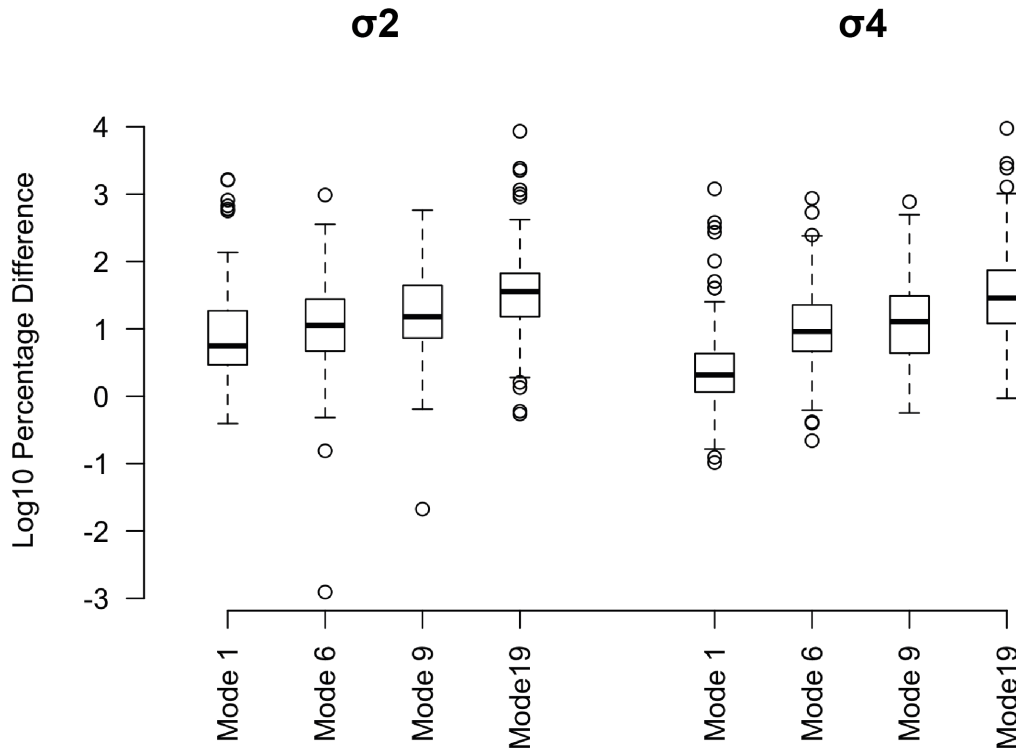


Figure 4.19: Box plots showing the distribution of the percentage difference values at end-Systole (top) and end-Diastole (bottom) for the four example modes. Plots are on a Log_{10} scale to enable the display of outliers without distorting the data.

Table 4.24: Mean, minimum, maximum and standard deviation of the percentage difference values which compare results across the different smoothing filters for Modes 1, 6, 9 and 19.

		Mode 1		Mode 6		Mode 9		Mode 19	
		Dia	Sys	Dia	Sys	Dia	Sys	Dia	Sys
$\sigma 2$	Mean	234.4	850.6	318.8	342.0	704.7	281.8	431.2	239.8
	Min	0.5	0.4	1.9	0.4	3.9	5.3	0.4	0.0
	Max	6073.9	56904.0	10886.4	19092.0	21653.7	6061.8	18775.7	4410.1
	SD	726.1	5990.9	1365.5	1968.1	3003.1	842.9	2071.9	756.6
$\sigma 4$	Mean	230.1	889.2	304.0	317.6	684.9	305.3	416.7	258.3
	Min	1.2	0.6	0.1	0.1	0.6	2.1	2.9	0.0
	Max	6096.1	61067.1	9798.0	18520.7	19219.5	7127.3	16838.6	5537.1
	SD	729.3	6388.6	1230.6	1905.8	2824.6	952.9	1913.0	857.3

in Table 4.24. The combined evidence from these tables demonstrates that Mode 9 exhibits a greater percentage difference at end-Diastole than end-Systole at both extremes of smoothing.

Table 4.26 shows the sum of squared difference values for 4 example modes, across three smoothing filters, selected on the basis of their capability to classify age (Modes 6 and 9) and sex (Mode 19). The sum of the squared difference between Mode 1 values does not differ across robustness assessment of the three smoothing filters (Table 4.26). The difference between smoothing levels is the smallest at Mode 1, which accords with the sum of squared difference results presented in Table 4.10, whereas the distance between the mode eigenvectors in Mode 19 is much larger (e.g. 4.60×10^{-4} when comparing $\sigma 2$ and $\sigma 3$). Again, this result appears reasonable since the mode features are ranked in order of their contribution to the total modal energy.

The $\sigma 2$ smoothing and $\sigma 4$ smoothing values retain a similar profile when compared to the original 96 subjects ($\sigma 3$), but are quite different to each other. Considering Mode 1 alone, the sum of squared difference between the different smoothing levels is smaller than all cases of changing the dataset size (with the exception of the 96-95 subject case, which we would expect to be almost identical (see Table 4.4)).

Table 4.25: Percentage Difference associated with the correlation coefficient values at Mode 1 when comparing the original 96 subject mode coefficient values at end-Systole.

Subject No.	% Diff Original - σ_2	% Diff Original - σ_4	% Diff σ_2 - σ_4
1	-0.7	-1.8	-1.1
2	-174.8	-176.9	-0.8
3	-105.8	-109.1	-1.6
4	-47.9	-52.1	-2.8
5	80.0	80.7	3.6
6	67.0	65.2	-5.7
7	41.8	40.7	-1.9
8	-38.4	-40.9	-1.8
9	-192.6	-198.9	-2.1
10	-240.3	-278.8	-11.3
11	39.0	43.2	6.9
12	-50.6	-55.9	-3.5
13	76.5	74.3	-9.1
14	-140.6	-130.7	4.1
15	0.6	7.2	6.7

Table 4.26: Sum of squared difference between Modes 1, 6, 9 and 19 when comparing the two smoothing levels to the original shape analysis result, which used a smoothing coefficient of σ_3

Comparison	Sum Squared Diff.			
	Mode 1	Mode 6	Mode 9	Mode 19
$\sigma_2 - \sigma_3$	2.91×10^{-5}	9.31×10^{-5}	8.01×10^{-5}	4.60×10^{-4}
$\sigma_4 - \sigma_3$	2.05×10^{-5}	9.49×10^{-5}	6.34×10^{-5}	3.55×10^{-4}
$\sigma_2 - \sigma_4$	2.27×10^{-5}	1.38×10^{-4}	1.10×10^{-4}	2.37×10^{-4}

Table 4.27: Exemplar LDA Analyses comparing classification results between the original analysis and the σ_2 and σ_4 smoothing cases. Test Combination reports the combination of modes used as inputs in the LDA. Where more than one combination was found equally successful in Chapter 3, both were tested. ‘Classification %’ reports the best accuracy percentage after cross validation. ‘Top Five’ lists the best five modes identified using a new LDA, created based on the new coefficients from the robustness test.

	Characteristic	Phase	Test Combination	Classification %	Top Five
σ_2	Sex	Systole	19	47.5	23, 20, 25, 10, 16
σ_2	Age	Systole	9, 18	57.5	17, 8, 30, 4, 26
σ_2	Obesity	Systole	15, 29, 16	60.0	27, 3, 26, 29, 17
σ_2	Sex	Diastole	6, 23	32.5	17, 26, 15, 16, 12
σ_2	Age	Diastole	6, 13, 17	65.0	7, 19, 8, 6, 11
σ_2	Obesity	Diastole	20, 11, 2	45.0	12, 15, 9, 3, 22
σ_4	Sex	Systole	19	52.5	23, 10, 20, 27, 16
σ_4	Age	Systole	9, 18	52.5	17, 26, 8, 21, 4
σ_4	Obesity	Systole	15, 16, 6	62.5	27, 3, 17, 16, 29
σ_4	Sex	Diastole	6, 23	50.0	17, 16, 26, 18, 14
σ_4	Age	Diastole	6, 3, 13	67.5	7, 19, 12, 8, 5
σ_4	Obesity	Diastole	20, 11, 2	52.5	14, 12, 9, 11, 3

Application of LDA

Table 4.27 shows the application of a linear discriminant model to the two alternate smoothing cases. As expected based on the Pearson Correlation Results reported in Table 4.22, none of the exemplar LDA analyses resulted in successful classifications of any of the population characteristics. The best achieved correlation was 67.5% (separation of age at end-Diastole when using a σ_4 filter). However, none of the top five identified modes from Chapter 3 appear in this separation. Modes 17 and 18 appear in many of the top five identified selections for both of the smoothing filter tests (e.g. σ_2 age and obesity at end-Systole and σ_4 sex at end-Diastole, amongst others). The least successful percentage accuracy based on shape modes was in the attempt to separate sex at end-Systole in the σ_2 trial.

4.5 Conclusions

This chapter presents and evaluates the impact of a series of interventions on the input and output data presented in Chapter 4. The intention was to determine the constraints on the statistical shape analysis process necessary to produce consistent shape features and, in doing so, to infer useful details from their characterisation. The results reported in this chapter first assess the difference in input data when the dataset is perturbed in four specific ways, and then consider how that difference translates both to the shape modes produced at output, and to the composition of modes within individuals.

Robustness is examined by assessing how sensitive the extracted shape features (modes) are to external modification, and also by evaluating how sensitive the mode coefficients (or how much of each individual shape feature is required to reconstruct any given RV). For the sake of interpretation, Pearson correlations are used to demonstrate *relative* agreement between the various analyses, and 3D reconstructions to consider robustness of the mode shapes.

All four tested interventions derive the same overall mode *shapes* from the input dataset. This suggests that the direct-mapping statistical shape analysis procedure produces an accurate deconstruction, whereas, the mode *coefficients* are more sensitive to alterations in input value. Changing the dataset size, was the least sensitive to changes in modal coefficient, followed by changing the segmentation. While still robust over the first 13 modes, altering the anatomical boundaries was more sensitive. Adjusting the smoothing filter caused substantial differences in recorded modal coefficient, demonstrating only two statistically significant correlations.

The results of the statistical shape analysis are robust to perturbations in the size of the dataset, both when it is increased to 121 subjects and when it is reduced systematically. The qualitative mode shapes in particular encode features of similar appearance for Modes 1, 6 and 9. This suggests that once a ‘correct’ segmentation has been identified, accurate diagnoses can be made on the basis of these shapes. The fact that Mode 19 does not produce the same shape, qualitatively, nor does it correlate strongly with any intervention based on dataset size (with

the exception of the 60 subject dataset at both phases ($r_{58} = 0.697$ at end-Systole, $r_{58} = 0.735$), is indicative that caution should be used when interpreting the results associated with the lower order modes if intending cross-comparability of specific mode numbers between datasets. The modal energy profiles in all cases have shown that at least 30 modes are required to cover 99% of the shape information contained in this dataset. Therefore, these modes are still crucial to the identification of interesting shape features, but that there is likely to be a substantial difference in the specific features that they encode.

The second intervention, controlling the placement of the poles and dateline, also showed agreement in relative mode coefficient throughout the first 13 modes directly, and up to the first 16, when several direct substitutions are made ($r_{94} > 0.914$). It appears that the roles of Modes 14 and 15 may have been swapped. Therefore, the statistical shape analysis method can also be considered insensitive to alteration of anatomical boundary placement. Dateline alteration resulted in a more sizeable impact both on mode shapes and on modal coefficients than was observed in any dataset alteration above 40 subjects. At the outset, it was anticipated that a significant flaw in this shape analysis framework would be related to the level of personal judgement involved in selecting the points along the septal-free wall boundary, beneath the pulmonary valve, as there is a Hausdorff Distance difference of up to 6.5mm, even in selections made by the same observer.

The third intervention was to assess differences in segmentation. Alterations to the segmentation, despite showing more substantial difference at the point of input than had been observed in previous literature (Sardanelli et al., 2008; Caudron et al., 2012), showed agreement in the output coefficients as far as the first 16 modes. The number of correlated modes is not quite as large as in the agreement between analyses where the dataset size was varied, but represents a small improvement over the dateline results. These similarities suggest that when different individuals segment the initial MRI scans, the variations in the first 15 or so modes produce shape features that do not impact the most common shape features. The fundamentally different shape features, those which may impact a categorisation or a diagnostic

result, are likely to be concentrated in the higher order modes, those between Modes 16-30. This could limit the applicability of shape analysis based selection, as the relatively large differences in observer variability could render manual segmentation insufficiently accurate. Therefore, the recommendation would be to continue to develop highly accurate automatic segmentation protocols and to focus on producing optimally contrasted MRI images. In the short-medium term this may translate to semi-automatic methods with manual checking.

The final intervention involved assessing the impact of alterations to the smoothing filter. The shape filter was anticipated to have the largest impact on the shape analysis output, owing to the introduction of vertical steps from the sharp sides of crude 3D models produced in ITK-SNAP, or from the removal of important shape features owing to an overly rigorous filter. Therefore, it is unsurprising that these input differences have changed the results of the decomposition. While the sum of squared difference between the individual modes is very low (especially at Mode 1), there is no significant linear correlation between the original mode coefficients and those derived from the smoothing tests. In contrast to the Pearson coefficients, the qualitative evidence from the plotted mode shapes indicates that the fundamental components still look similar. The lower energy modes (particularly Modes 17 and 18) encode the aforementioned steps. It is possible that these features are also concentrated and could be removed from overall analysis. Nonetheless, they appear to have impacted the modal coefficients for all of the other modes, leading to poor statistical correlations (i.e. one subject, especially where significant twisting has occurred may have a more jagged edge, which is usually removed using the smoothing filter, which may influence all other modal coefficients). Therefore, it is necessary to conclude that direct comparison of mode coefficients between individuals cannot be controlled sufficiently for clinical use. When considering shape analysis output, the correlation between coefficients for each mode suggests that the output shape analysis differs much more substantially in this test than in all previous cases (see Table 4.22). The smoothing parameter must be kept constant, as the alteration of smoothing factor caused the most significant fall in Pearson coefficient.

Several other notable observations can be derived from examination of this dataset. The first is that Mode 1 encoded the same shape feature across all four analyses, as evidenced through directly comparing the mode shapes using the sum of squared differences, and also through direct comparison of the mode coefficient values. This continuity confirms scanning protocol as the most substantial factor affecting RV shape in this dataset. Additionally, as a consideration of the differences in dateline, dateline is set at a location of the geometry where there is more natural variation, whereas the opposite septal/free-wall boundary is more consistent between slices. Therefore, there could be an argument for establishing the dateline on the opposite side for the sake of improved consistency. The fact that dateline selection caused more substantial variation points to the fact that controlling for anatomy (in the form of segmentation) is an important finding, and is a benefit of the particular (direct-mapping) POD approach to statistical shape analysis used in this thesis.

As has been demonstrated by the statistical shape analysis results in preceding work by Wu (2013), the location of the anatomical boundary has a significant impact on the overall results. However, in this section, the dramatic impact on the modes themselves is not replicated, despite resulting in more substantial variation than was observed in many of the dataset size tests. The Hausdorff distance values may be exaggerating the overall difference in these dateline selections, particularly since the selections are all made with prior anatomical knowledge. Nevertheless, the implementation of an automated dateline selection algorithm could be an important intervention if this shape analysis were to be implemented in a clinical setting, where the individuals selecting the dateline may be less familiar with right ventricular geometry, or are required to make the selections when under time pressure. Such an intervention could be as simple as demarcating a pixel on the short-axis scan during segmentation, and then using the 3D locations of these points to inform mesh refinement. However, it remains a sensible choice to control the boundary manually in areas which are more prone to variation, since it is here that the greatest benefit of human intervention is gained. Therefore, although the development of an automated boundary selection method could improve the repeatability

of these results, this may be detrimental to anatomical matching between the RV and the mathematical representation. A future analysis could determine whether the difference in shape analysis output is reduced if the poles are maintained but the dateline is altered (i.e. ensuring that the anatomical top and bottom of the heart are mapped to the same place, regardless of the split selection). However, since dateline alterations, in general, affect the shape analysis results less strongly than other pipeline changes such as altering smoothing factor, or scanner selection, this seems unnecessary.

Overall, the statistical shape analysis method is largely insensitive to changes which do not alter the underlying RV models. This includes changing the dataset size, and selecting new anatomical data points. Conversely, the greatest differences occurred in situations where shape of the models were altered, for example in smoothing, or from scanners with different imaging parameters. Therefore, several practical recommendations can be made based on the robustness analysis. Scans from varying origins can be combined without issue; the specific shape features are contained within a single mode and can be isolated, as is demonstrated in Chapter 3, and dataset size can be altered without substantially changing the modal coefficients, although the user should be wary of reducing the dataset size much beyond 40 subjects. However, the remaining methodological parameters should remain untouched to ensure the consistency of derived shape features. In those interventions where the mode coefficients strongly correlate, none do so for more than the first 17 or 18 modes. This suggests, although the first 30 shape features should continue to be assessed, as anything up to 99% of the modal energy can be assumed to be geometrically important, it cannot automatically be assumed that those features encode the same specific shapes, and therefore the mode number should not automatically be assumed to encode the same feature. In other words, if visual analysis of the mode shapes is seen in conjunction with very strong ($r_{94} > 0.9$) Pearson correlation coefficients, similarity of interpretation is highly probable.

It is likely that including large numbers of hearts selected by a variety of individuals will change the specific features of importance, but in general it appears that the main difference

in the dataset remains the choice of scanner, and that this difference supersedes any other difference. This is encouraging, as it suggests that once key quantitative parameters are fixed, such as the smoothing parameter and the scan slice thickness, the protocol is robust to many of the more subjective criteria. Inter and intra-observer variability remains an inevitable source of variation, but as will be discussed in more detail in the application of the method to PH (Chapter 5), key areas of anatomical variability are identified, even if there are individual problem cases where there is a substantial difference in variability. Depending on the particular use case for this diagnostic shape analysis pipeline, there are a few different approaches that could be taken to mitigate the impact of human error or difference in judgement. For instance, each segmentation could be reviewed by two different individuals, and flagged if the two choices differ by a certain threshold percentage. Or, as suggested in Subsection 4.4.3, a series of defined segmentation rules could be established to attempt to mitigate the impact of human choice, as compared to human error. Alternatively, an automated algorithm could be used as a baseline, and then manually reviewed. Inevitably, if any kind of statistical shape analysis is to be used in routine medical environment, the speed of processing must be prioritised, which represents an additional argument in favour of semi-automated segmentation. There is promising indication from the results obtained in this chapter that the shape features are robust to targeted variations in pipeline inputs.

Regardless of other qualitative interventions, one crucial take away from this chapter is that the parameters cannot be altered from the original without causing noticeable changes to the statistical shape analysis output, even if that difference is obvious and categorisable. This is the case even in this small test example, where only five subjects were altered.

Chapter 5

Appraisal of a Diagnostic Metric that can be applied to Pulmonary Hypertension

5.1 Introduction

PH is a progressive cardiovascular disease that significantly impacts the mechanical properties of the heart. Remodelling of the RV occurs in response to high blood pressure in the arteries of the lungs and causes a defined, progressive change in shape and mechanical properties, which may be indicative of increased potential for heart failure. PH is associated with high morbidity and mortality, in part because non-specific early disease symptoms often lead to delayed diagnosis (Vonk-Noordegraaf et al., 2013; Xu et al., 2021). The heart is able to compensate for RV dysfunction up to a point, where initial symptoms are indistinguishable from a wide array of cardio-pulmonary ailments, ranging from breathlessness during exercise to angina and heartburn. The relatively sudden onset of RV failure has cemented the disease's reputation as a silent killer: the mean survival rate is 2.8 years after diagnosis (D'Alonzo et al., 1991). Despite significant research funding since the early 1990s, this figure has not improved substantially. Chang et al. (2022) report that contemporary mortality rates for patients with PAH are 8% after one year, 16% after 2 years and 21% after 3 years.

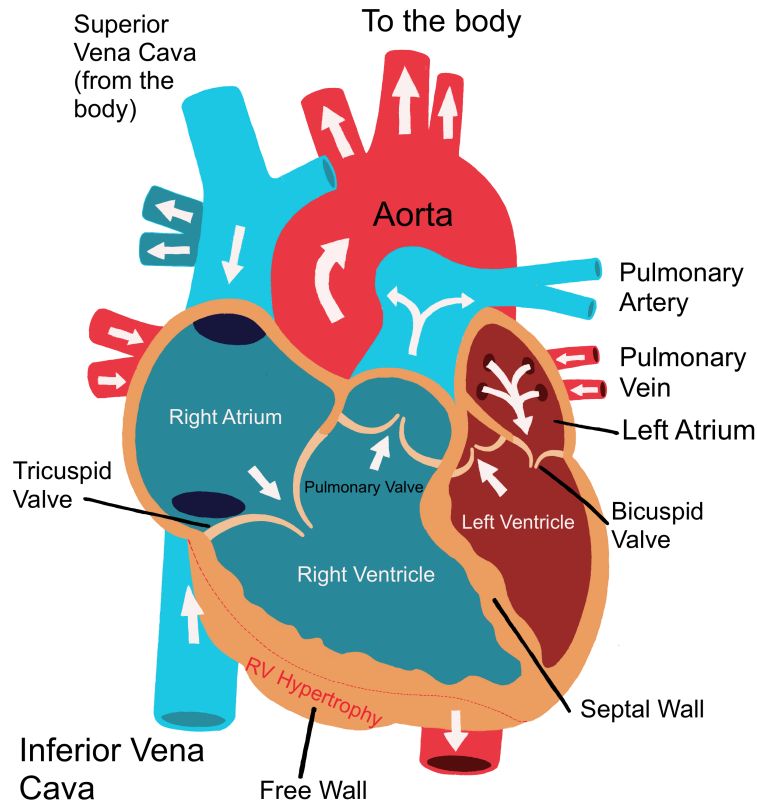


Figure 5.1: Cross-section of the heart, showing the RV in blue and the LV in red. The standard RV free wall border, in a non-PH heart is marked by the red dotted line. The hypertrophic region is exterior to this region. This Figure is reproduced from Figure 1.1b.

PH noticeably alters the shape and function of the RV. The RV experiences hypertrophy (an increase in volume related to increase in cell size) in response to raised PAP, and this causes flattening of the base of the ventricle (visible in Figure 5.1, which is reproduced from Chapter 1). Hypertensive RVs have also been shown to be more spherical, and experience increased cross-sectional area in the middle and base-segments. Other aspects of RV shape, including geometric shortening, have been associated with prognosis of PH sufferers, and right heart failure is associated with functional adaptation (Vonk-Noordegraaf et al., 2013).

The earlier chapters of this thesis have demonstrated that specific biological characteristics are uniquely associated with identifiable shape changes in the RV. Examples include the flattening of the ventricular base in male RVs, and an overall rounding or hypertrophy in overweight subjects. What has yet to be determined is whether such discrepancies impact PH

diagnosis. As the shape of the RV determines PH, it is critical to evaluate the ability to use the shape metrics identified as part of this thesis, and in the literature, alone or in combination with other key individual biological characteristics to predict the presence and/or state of PH. Improving and automating methods of recording shape change will allow for better understanding and prediction of the effects of PH on the mechanical behaviour of the human right ventricle. The ultimate intention of which is to bring shape analysis strategies based on non-invasive, clinically attainable imaging techniques and measurable haemodynamics into mainstream diagnostic medicine. Such research forms part of a continuing collaboration with cardiologists and imaging experts. Statistical shape analysis procedures have shown promise in characterising the shape change in the RV and in matching this change to a disease state (Wu et al., 2012, 2014; Xu et al., 2021; Leary et al., 2012). Reported differences in response to PH related to sex, age and obesity will be briefly reviewed in the following paragraphs.

Differences between the sexes have been reported in the prevalence, severity, treatment response and survival rates of PH (Rodríguez-Arias & García-Álvarez, 2021). One of the most notable differences is that PAH (WHO Group 1 disease) is more frequent in women, however, once affected, female patients present better prognosis; a phenomenon known as the oestrogen paradox. Swift et al. (2015) performed a pair-matched study, with subjects matched by age and sex and concluded that male patients with idiopathic PAH have proportionally worse RV function despite similar afterload. Swift et al. (2015) hypothesise that this difference is caused by more effective adaptive remodelling in female hearts. Chapter 3 determined that non-hypertensive female RVs exhibited a rounder septal wall, and that male RVs were overall much flatter in the same region. This difference in shape may be associated with the male RV's difficulty in compensating.

Obesity is also associated with PH, and has been shown to modify patient outcomes (Frank et al., 2020). In this study, multivariable regression models tested associations of BMI and measures of pulmonary vascular haemodynamics and determined that higher BMI was independently associated with greater odds of PH disease, and that patients with PH had greater

risk of mortality. This cohort comprised 8940 subjects who underwent RHC in a hospital setting between 2005-2016. The statistical shape analysis method employed earlier in this thesis identified that individuals in the higher BMI group exhibit a significant rounding out of the free wall at the base. Increased BMI is also associated with a lateral lengthening of the RV.

In the first registry-based studies, PH was thought to primarily affect younger female patients (Rich et al., 1987) with a mean age of 40. However, as more cohorts have been examined, mean and median ages at diagnosis have been steadily increasing. The main issue with diagnosis of PH (and most notably PAH) in elderly patients is that it is difficult to discriminate potential pulmonary vascular disease from the normal ageing process (Berra, Noble, Soccia, Beghetti, & Lador, 2016). Age-related physiological changes of the cardiovascular and respiratory systems should be considered if PH is suspected in elderly patients, as the normal ageing process could either lead to over-diagnosis of PH or alternatively an underestimation of PAH. Truong et al. (2020) produced a comprehensive review of non-invasive imaging methods and common biomarkers used to detect right ventricle dysfunction.

Clearly, therefore, biological features impact the incidence and severity of PH symptoms. However, it remains to be determined whether biological features also impact RV shape in a context that either helps or hinders the diagnostic process. This Chapter will identify the impact of including the biological features, sex, BMI and age on the diagnostic capacity of the model.

Precise measurement of RV function is fundamental to patient management. The RV's contractile pattern is distinctive and consists of three main phases. The first is a shortening along the longitudinal axis and the traction of the tricuspid annulus towards the apex. The second is inward (radial) movement of the RV free wall, in a process often known as the bellows effect and finally the interventricular septum bulges during LV contraction and the free wall stretches over the septum. This leads to shortening in the anterior-posterior axis (Tokodi et al., 2021). Through capturing both end-Systolic and end-Diastolic shape features, it may be possible to identify such unique contractile shapes. Measurement of RV to LV-

diameter ratio at end-Systole in evaluating outcomes in children with PH (Jone et al., 2014), showed that the RV/LV ratios were lower in controls as compared to PH patients and that this ratio correlated with mean PAP and systolic PAP as a percentage of systemic pressure, amongst other metrics. This non-invasive metric can be easily obtained in a clinical setting and could be used in conjunction with a statistical shape analysis model. Bidviene et al. (2021) assessed regional RV shape changes in pressure (PH) and volume (repaired Tetralogy of Fallot) overload conditions in terms of their relations with RV function and mechanics. They found that repaired Tetralogy of Fallot and PH patients had flatter free walls at both the body and apex regions and a flatter RV inflow tract. Moriyama et al. (2020) used RV-specific speckle tracking echocardiography of patients with chronic thromboembolic PH (Group 4) to reveal that each of three regions of the RV (inlet, outlet and apex) presents a unique response to haemodynamic changes during PH treatment, and as such, these are useful indicators for assessing RV function. Particularly, impaired function in the outlet region was reflective of myocardial damage indicators in non-PH conditions.

This chapter will also determine whether a linear classifier, as has been used to assess relationships between the shape features thus far, provides the best available discriminating properties, or whether a non-linear classifier would improve diagnostic success. The availability and easy of obtaining vast amounts of clinical data makes healthcare a strong setting for the use of advanced data techniques. Non-linear behaviour often occurs within human systems owing to their complexity and dynamic response to a range of factors. It is highly likely that these factors cannot be adequately described using linear models (Higgins, 2002). This chapter examines whether non-linear classification using Support Vector Machines (SVM) substantially improves classification of healthy from diseased subjects, and therefore is worth pursuing.

In the past, SVM-based classifiers have been used to diagnose PH, however, to the author's knowledge none have used shape or texture-based components. Shang et al. (2021) created classifiers which focused on genomic characteristics, transcriptomics and gene expression

profiles related to PH. This study found that an optimal gene-based classifier was able to accurately distinguish between healthy individuals and those with PAH as determined by ROC (Receiver Operating Characteristic) curve analysis, which is used commonly used in machine learning to assess overall diagnostic performance of a test. However, Shang et al. (2021) ultimately conclude that their SVM results are unconvincing in light of the small sample size. Leha et al. (2019) applied five different machine learning algorithms including SVMs to a database of 90 subjects with invasively determined PAP. They performed a 10 times repeated 3-fold cross-validation, leading to 3 folds of 28 subjects per repetition. The makeup of the dataset was 68/90 subjects exhibiting PH, as determined by PAP. They determined that SVMs have an AUC (area under ROC curve; the higher the AUC, the better the performance) of 0.83 and 95% CI ranging from 0.73-0.93 when based on echocardiography-derived measures. This machine learning algorithm determined that estimated right atrial pressure was only of minimal importance. All machine learning methods in this case recorded AUC accuracies exceeding 78%. Narula et al. (2016) note that machine learning is not principally more advantageous in the diagnosis of PH than manual assessment by a skilled practitioner. However, it offers a high-throughput, contemporary solution to standardise diagnostic criteria, and to assist in assigning overall likelihood of PH.

Having established that there are important shape differences in right ventricles based on sex, age and obesity that had not previously been considered in diagnostic pipelines, a new diagnostic metric for identifying PH and generalised hypertension can be created. A linear metric, using LDA will be assessed in the first instance, and then a non-linear classifier will be implemented in an attempt to improve upon the classification that already exists.

5.2 Aims and Hypotheses

5.2.1 Aims

This chapter will evaluate the capability of the shape analysis tools to be used to predict PH. This chapter will identify specific features that can be applied to the diagnosis of PH. Three methods will be used to incorporate shape and haemodynamic information from the datasets. First, by *brute force*, the optimal combination of Modes for diagnosing PH will be identified. Secondly, known demographic features will be included in such a linear model. Finally, a non-linear model will be considered.

This chapter aims to confirm whether the specific shape changes related to PH which have been seen in the literature are also observable in this dataset; to evaluate the impact of determine whether it is possible to improve upon the shape analysis procedure's ability to categorise subjects with PH by including sex and obesity in the model.

5.2.2 Hypotheses

1. In line with findings from the literature (e.g. Mauger et al., 2019; Xu et al., 2021), there will be a specific shape mode combination that will define anatomical shape features that are specific to PH.
2. The shape of the RV is a determinate of PH, both in diagnosis and prognosis, and so the aim is to evaluate the ability to use this shape metric alone, or with other important individual features to predict the presence and/or state of PH.
3. Including obesity, sex and age parameters in a linear classifier will improve that classifier's ability to identify individuals with PH as these were found to uniquely impact RV shape.
4. A non-linear model will be able to more accurately distinguish PH hearts from those with generalised hypertension and healthy individuals than the more crude linear model.

5.3 Methodology

5.3.1 Training and evaluation of the classifiers

The first step to evaluate the ability to use RV shape to predict PH is to establish how easily PH can be predicted using LDA, using the same process as was implemented in Chapter 3. First, a linear discriminant model was applied to the POD modal coefficients generated in Chapter 3 for all 96 RVs. End-Systole and end-Diastole were assessed separately. LDA was used to build classifiers to separate RVs based on their disease status, i.e. whether they were PH sufferers. The dataset is comprised of 33 subjects with confirmed PH and 63 non-PH. The non-PH group consists of 23 healthy ageing, 23 RVs with generalised hypertension (defined as having a diagnosis of hypertension at a local general practice, but no other cardiovascular diagnosis, diabetes or dialysis-dependent renal failure (Parikh et al., 2017)) and 17 non-PH subjects from Pittsburgh.

In the same manner as described in Chapter 4, the LDA identified the five shape modes with the largest standardised structure matrix coefficients, i.e. the modes with the highest correlations to the discriminant function (see Subsection 3.4.3 for a more in-depth explanation). A second round of LDA was then applied using subsets of these top five modes to find the combination which gave the best classification result and separating function. The percentage of RVs successfully categorised was recorded.

Subsequently, the same three biological factors that were included in Chapter 3; sex, age and BMI were introduced one by one as cofactors in the analysis to determine the impact of including this additional shape information on the ability to detect PH. Sex was a binary parameter, and age and BMI were continuous. The LDA analysis was performed using SPSS (Armonk, NY). Finally, the hearts were split into their sex, age and BMI categories before applying LDA, to determine whether the same, pre-identified modes impacted the classifier.

5.3.2 Support Vector Machines

Support Vector Machines (SVMs) are a subclass of supervised classifiers that attempt to partition a feature space into two or more groups. The machines achieve this by finding an optimal means of separating such groups based on known class labels. In complex cases SVMs are able to partition in a non-linear fashion, which they achieve by means of a kernel function. SVMs are a sophisticated class of separator, but as with the linear discriminant functions are also prone to overfitting; although arguably less so than other machine learning algorithms such as artificial neural networks (Golland et al., 2001). Standard SVM models construct a hyperplane or *decision boundary*; a line in multidimensional space that best divides the input space into two disjoint regions. The main goal of SVM is to divide the dataset in such a way that it finds a maximum margin hyperplane, i.e., it finds a plane that maximises the distance between the closest pair of data points which belong to opposing classes. These points are the “support vectors”, since they support, or determine the decision boundary.

SVMs are advantageous because they are the only model which behaves in a linear fashion which is also capable of classifying data which is not linearly separable (using the ‘kernel trick’, which is described in this section). Linear models have the advantage over non-linear models such as neural networks, in that the coefficient weightings directly correspond to the feature’s importance in the model, which makes them much easier to understand and interpret. This is crucial when analysing already parameterised data like the POD modes, since the aim is to correlate specific shape features to anatomical outputs. The kernel SVM decision regions for non-linear kernels operate by creating non-linear combinations of the features of interest to uplift these samples onto a higher-dimensional feature space, where a linear boundary can be used to separate classes.

A hyperplane in d dimensions is given as the set of all points $x_i \in \mathbb{R}^d$ that satisfy the equation $h(x) = 0$, where $h(x)$ is the hyperplane function defined as follows:

$$\begin{aligned}
 h(\vec{x}) &= \vec{w}^T \vec{x} + b \\
 &= w_1 x_1 + w_2 x_2 + \dots + w_d x_d + b
 \end{aligned}
 \tag{5.1}$$

\vec{w} is a d dimensional weight vector and b is a scalar, known as the bias. Therefore points that lie on the hyperplane satisfy the following equation:

$$h(\vec{x}) = \vec{w}^T \vec{x} + b = 0. \tag{5.2}$$

The hyperplane is thus defined as the set of all points such that $\vec{w}^T \vec{x} = -b$. The fundamental idea behind SVMs is to choose the canonical hyperplane, specified by the weight vector \vec{w} and the bias b that yields the maximum margin among all possible separating hyperplanes. If δ_h^* represents the margin for hyperplane $h(x) = 0$, then the maximum margin hyperplane is found by solving the following optimisation problem:

$$h^* = \operatorname{argmax} \{ \delta_h^* \} = \operatorname{argmax} \left\{ \frac{1}{\|w\|} \right\} \tag{5.3}$$

The kernel takes the low dimensional input space and transforms it into higher-dimensional space to enable the successful determination of a non-linear separation line. In other words, the kernel converts non-separable problems into separable problems by increasing the number of dimensions. A specific advantage of the SVM method is that it is able to employ a technique known as the kernel trick. This trick allows the data to be represented only through a set of similarity comparisons between the original data observations x , which retain their original coordinates in the lower dimensional space, instead of explicitly applying the transformations necessary to represent the features in higher dimensional space. Thus, the kernel trick is able to avoid the potentially extreme computational costs associated with applying the transformation.

Four of the most popular kernels will be tested in this chapter; a Linear kernel, which predicts a new input using the dot product of the input and the support vectors:

$$k(X, Y) = X \cdot Y + C \quad (5.4)$$

where X and Y are vectors/observations in the input space and C is a regularisation parameter; a Radial Basis Function kernel (considered very similar to a Gaussian kernel):

$$k(X, Y) = \exp(-\gamma \cdot \|X - Y\|^2), \quad \gamma > 0 \quad (5.5)$$

where $\gamma = \frac{1}{2\sigma^2}$ and $\|X - Y\|^2$ is the squared Euclidean norm.

a Sigmoid kernel;

$$k(X, Y) = \tanh(\gamma \cdot X^T Y + C), \quad (5.6)$$

and a Polynomial kernel;

$$k(X, Y) = (\gamma \cdot X^T Y + C)^d, \quad \gamma > 0 \quad (5.7)$$

A Polynomial kernel is a more generalised form of the linear kernel. Degree, d , is a parameter required only for polynomial kernels, and controls the flexibility of the curve.

All cases, except for the linear kernel include the variable parameters are C and gamma (γ). C is the penalty parameter and represents the misclassification or error term. This determines how much error is acceptable in the SVM optimisation and allows control of the trade-off between the decision boundary and the misclassification term. When C is high it will classify all data points correctly but at the risk of overfitting.

Gamma (γ): defines the distance at which points can influence calculation of the line of separation. When the gamma value is high, only nearby points will have a large influence whereas a low gamma value means that further away points will also be considered in the determination of the decision boundary. As a result of this, high gamma values typically produce highly flexible decision boundaries, and low gamma boundaries produce a more linear

decision boundary.

Implementation

The 96 subject dataset was randomly split into a training set and a testing set in a 75:25 ratio. The input was Mode coefficient values generated in Chapter 3, for those modes selected as the ‘Top Five’ based on LDA. Each SVM kernel was trained on a training set (72 subjects) and applied to a test set (24 subjects). The SVM output was a cross-validated accuracy score, expressed as a percentage, and this result was averaged over 100 attempts, each involving a different 75:25 split. This produced an accurate representation of the entire dataset and ensured that statistical power was not compromised due to the need to subdivide a relatively small sample size. The 100 attempts were run sequentially, in a loop, and mean accuracy and standard deviation values were reported over all the attempts.

All mode coefficient values were scaled prior to application of the SVM model, to remove any effects associated with the decreasing standard deviation in the higher order modes. As the modes are ranked in order of their contribution to the total shape decomposition, Mode 2 will necessarily have larger coefficient values, and a greater range than Mode 17, for instance. Testing was performed using the `caret` and `e1071` packages in R version 4.1.2 (Bird Hippie).

Parameter tuning was undertaken to improve the success of the SVM. A cross-validated grid search was performed to identify combinations of SVM input parameters (C , γ , degree, polynomial and `coef0`) which will separate PH from non-PH groups with the greatest accuracy. Grid search is one method of hyperparameter tuning, and involves defining a grid of possible values for each hyperparameter, and then exhaustively evaluating the model’s performance on every possible combination of values within that grid. In the first instance, exponentially growing sequences of C and γ were tested to narrow down on a useful range of values (as demonstrated in Hsu et al., 2016). After initial exploratory analysis using the `caret` package, the useful range was identified to be $C=0.1:10$, `coef0` = 0.1:10, $\gamma=0.1:10$ (this was performed by inspection), and the parameters were tuned according to these ranges. The `svm.tune()`

Table 5.1: Best Mode combinations and cross-validated percentage accuracy scores for characterising PH using Linear Discriminant Analysis

Diagnosis	Phase	Mode Combination	Accuracy
PH	Diastole	15, 10, 20, 21	72.9%
PH	Systole	2, 15, 14, 8, 5	68.8%

command in R's `e1071` package was used to tune the narrowed range of parameters and determine the best combination for accurately separating PH from non-PH. The entire process, from the sample split, training and test set division, and the parameter tuning was performed for each iteration in the loop, so every random 75:25% split of RVs underwent a separate tuning process. In practice, the kernels and parameters were applied, and the analysis performed by specifying options in the R function `svm.tune()`. Owing to the long computational time associated with some of the more complex kernels tested (e.g. sigmoid and polynomial), the SVM analysis was performed using the Hamilton HPC service at Durham University.

5.4 Results and Discussion

5.4.1 Assessing the Diagnostic Potential of Mode Shape Features

Initial characterisations applied LDA to all 96 RVs, using the coefficients of Modes 2-30 as inputs. The initial characterisation aimed to produce the best classifier for separating individuals with PH from those without. Mode 1 coefficients are excluded from the analysis, as they encode features related to the scanning protocol (as determined from the population-based analysis in Chapter 3). In each case, after initial identification of the top five discriminating shape modes based on their structure matrix scores, sub-combinations of modes were tested to find the overall best separating combination to diagnose PH. A more in-depth description of this process, and the mode selection criteria, is covered in the Chapter 4 Methodology (Section 4.3).

Table 5.1 reports the successful classification percentages, after leave-one-out cross-val-

idation, for the maximum separating mode combinations achieved at both cardiac phases. PH can be predicted with 72.9% accuracy using a combination of 4 modes at end-Diastole; Modes 15, 10, 20, 21 (identified as the best discriminating modes using LDA). This result confirms that a specific shape associated with PH is observable in PH-sufferers, as compared to the rest of the group. PH sufferers were classified on the basis of PAP exceeding 25mmHg, and were not stratified based on the severity of their illness. The categorisation of PH sufferers was provided with the data from UPMC. At end-Systole, all five of the top modes, based on structure matrix were required to capture 68.8% of subjects correctly (shown in Table 5.1). This classification score is not quite as strong as at end-Diastole, but still indicates the presence of a fundamental shape difference at end-Systole in individuals with PH. There are noticeable markers which are unique to PH at both end-Systole and end-Diastole. The maximum diagnostic separation at End-Systole uses all five of the top identified modes to achieve the maximum diagnostic separation, which is perhaps indicative of the fact that more than 5 modes are involved in optimal shape characterisation.

Predictive accuracy was very slightly better at end-Diastole than at end-Systole (72.9% vs 68.8% (Table 5.1)). Notably, Mode 15 features in both phases, which suggests that this shape feature is a specific predictor of PH. Table 5.2 describes the top modes in the Linear Discriminant model based on their structure matrix values, and establishes Mode 15 as the strongest individual discriminating variable related to PH at end-Diastole (-0.680), and the second strongest discriminating variable at end-Systole (0.441), further substantiating the claim. As described in the overall Methodology in Chapter 2, the end-Systolic and end-Diastolic shapes were decomposed together, although the resultant mode coefficients were analysed separately. Therefore, Mode 15 is encoding the same feature, which is a strong discriminator at both phases of the cardiac cycle. Chapter 3 identified Mode 15 as an important factor in the classification of obesity in non-Hypertensive individuals at end-Systole. This points to a potential inter-connection between PH and the obesity and its impact on RV shape. It is likely, therefore, that BMI will be a confounding factor in identifying RV shape features

Table 5.2: Structure Matrix showing the results of an LDA model built using Modes 15, 10, 20 and 21 at both end-Systole and end-Diastole

Diastole		Systole	
Discriminating Variable	LDA Function	Discriminating Variable	LDA Function
Mode 15	-.680	Mode 2	-0.462
Mode 10	.566	Mode 15	0.441
Mode 20	.434	Mode 14	0.418
Mode 21	.424	Mode 8	-0.342
		Mode 5	0.295

which best predict PH.

Using the shape features encoded by these four modes in combination produces an overall shape that aligns with evidence from the literature. The qualitative shape displayed in Figure 5.2 demonstrates a rounding out and flattening of the base of the right ventricle, which aligns with descriptions of PH-related shape change in previous shape based literature (e.g. Mauger et al., 2019; Xu et al., 2021) and in medical descriptors of the condition (Vonk-Noordegraaf et al., 2013). The unusual shape of the non-PH extreme RV, with a noticeable bulge at the base, which is marked by an arrowhead in Figure 5.2, prompted the suggestion that this subject was not reflective of the cohort as a whole. To confirm whether this result was distorted from plotting the coefficients of one anomalous subject, the 75th and 25th percentile subjects were also analysed in Figure 5.3).

Figure 5.3 depicts 3D reconstructions of the RV produced when plotting the mode coefficients from the 75% and 25% percentile function scores when the RV is constructed using Modes 15, 10, 20 and 21. Ventricular shapes confirm that PH patients are likely to exhibit a ‘rounding out’ of the base of the ventricle, which can be explained by the increase in pulmonary pressure associated with PH (as described in Simonneau et al., 2019). The area of greatest difference between the 75% subject and the 25% subject is indicated by an asterisk. The non-PH subject in Figure 5.3 showed bulging in the same region of the RV base as observed in the extreme individual, although this effect was far less pronounced than in Figure

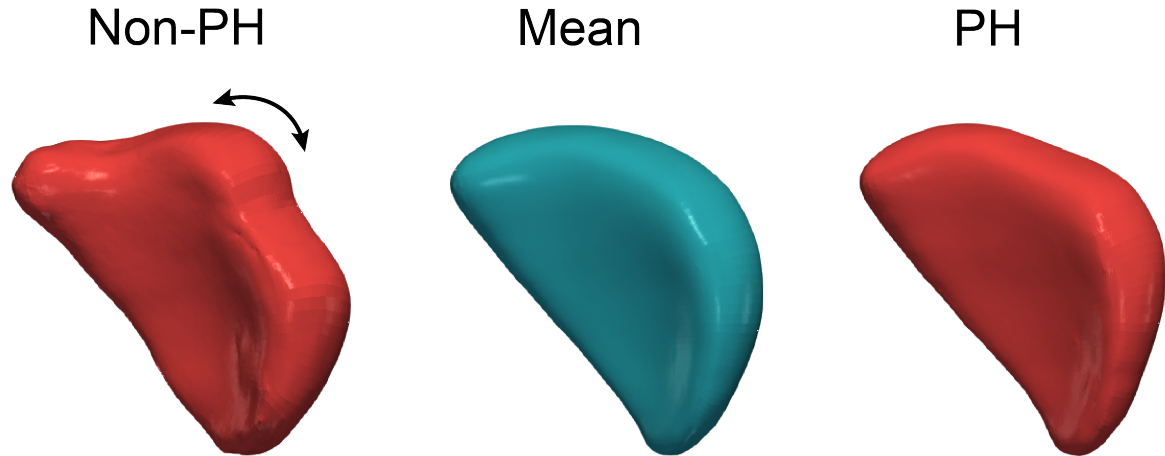


Figure 5.2: The combined qualitative shape impact of the modal coefficients for Modes 15, 10, 20 and 21 of the most extreme non-PH subject, at end-Diastole, based on LDA function value, and the combined impact of the coefficients for the most extreme PH subject. Figures are the mean RV $\pm 2 \times$ SD.

5.2. Therefore, it can be discerned that non-hypertensive subjects have a tendency to a slight protrusion higher up on the free wall, nearer to the base. On the other hand, the PH patient, as reconstructed by Modes 15, 10, 20 and 21, remains largely unchanged between the most extreme subject's LDA score and the 75th percentile subject's LDA score, indicating that the flattened shape is likely to be a prominent identifying feature of individuals with PH. Again, these results are consistent with previous shape analysis papers (Xu et al., 2021; Mauger et al., 2019). Thus, the best identifying mode combination shows that there is a flat region at the base of the heart in PH sufferers.

Figure 5.4 shows the ventricular models produced when plotting the mode coefficients from most extreme PH and non-PH RVES, as determined by the model created using the top five modes at end-Systole. The 3D models are constructed from the mean RV shape plus the combined Modes 2, 15, 14, 8 and 5 multiplied by the modal coefficient values for the extreme non-PH and PH RVs. The extreme PH subject exhibits reduced septal curvature (marked by a black arrow) as compared to its non-PH counterpart, which accord with previous literature (Xu et al., 2021). Previous shape analyses have also found that the presence of PH has caused a flattening of this wall. The RVES of the exemplar non-PH subject displays

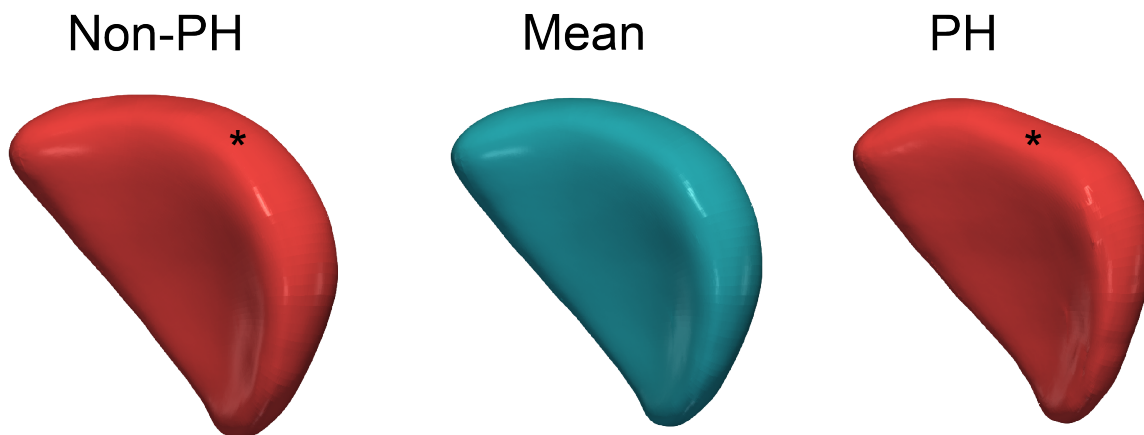


Figure 5.3: The combined qualitative impact of Modes 15, 10, 20 and 21 in subjects with the 25th and 75th percentile LDA scores at end-Diastole. The left side depicts the 25th percentile score, indicating a typical non-PH RVES, and the right side depicts the 75th percentile score which indicates a typical PH RVES shape. Asterisks in the basal region mark the rounding effect in the non-PH RV, as compared to the mean, and the flattening effect in the PH sufferer. Figures are the mean RV $\pm 2 \times$ SD.

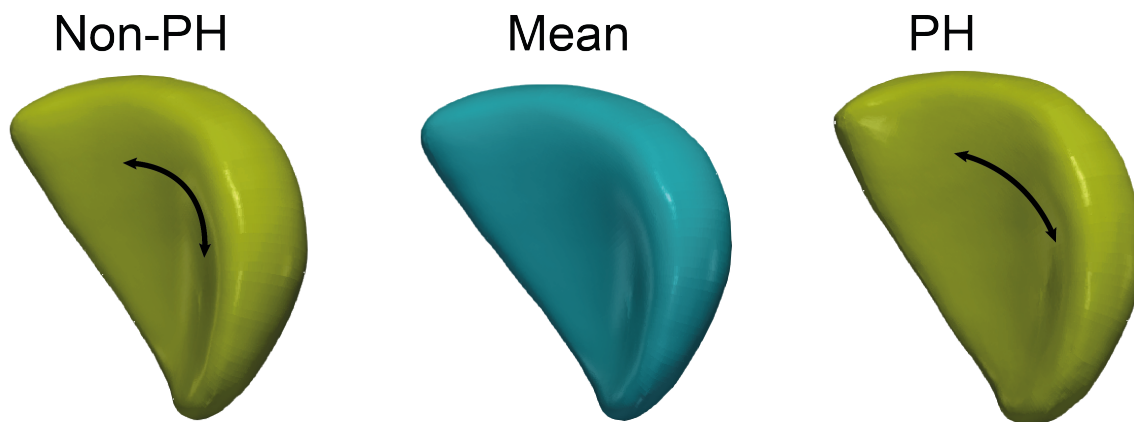


Figure 5.4: The combined qualitative impact of Modes 2, 15, 14, 8 and 5, the five best modes for determining PH status at end-Systole. The left side shows the most extreme non-PH function value, and the right side shows the most extreme PH subject in terms of LDA function value. The extreme subject exhibits reduced septal curvature as compared to its non-PH counterpart, this difference in curvature is marked by arrows on both RVs.

slightly increased curvature of the septal wall, especially lower down in the RV, closer to the apex. However, these shape features were associated with end-Systole, rather than end-Diastole. There is also evidence of characteristic flattening at the base of the ventricle in the extreme RV with PH. Having established that this shape analysis method has identified characteristic and distinguishing features of PH, the next step was to attempt to improve this model by including the biological characteristics that were assessed in Chapter 3.

5.4.2 Application of Subject Characteristics to Improve the PH Model

The results from Subsection 5.4.1 establish that PH patients share shape characteristics with obese individuals. Therefore, it seems plausible that accounting for an individual's BMI would improve a diagnostic classification model for PH. As the results reported in Chapter 3 (Table 3.3) demonstrate, that combinations of shape modes can be found which separate the biological features for the three sets of biological combinations tested.

This section added BMI, sex and age as factors into the predictive models to characterise their respective effects on diagnosis of PH. Sex and age were considered for the sake of completeness despite less obvious qualitative evidence for their immediate impact. Each population-based metric was assessed in a separate LDA, where the input was the discriminatory set of shape features identified in Subsection 5.4.1: Modes 15, 10, 20 and 21 for the end-Diastolic RVs, and Modes 2, 15, 14, 8 and 5 for the end-Systolic RVs, as well as the feature of interest, added as an additional factor. There are 5 total inputs in each end-Diastolic test and 6 total inputs in each end-Systolic test. Cross-validated (using leave-one-out cross-validation) accuracy scores are reported, in addition to the best combinations, in the same manner as the initial LDA analyses reported earlier in the chapter.

Table 5.3 reports the cross-validated percentage accuracy scores when LDA analyses are conducted to identify the presence of PH, using the best combinations from Table 5.1 and sex, BMI and age, in turn. In both end-Diastolic and end-Systolic RVs, the inclusion of BMI slightly improves the LDA function's ability to discriminate PH (from 72.9% to 74.7% and

Table 5.3: LDA cross-validated percentage accuracy scores when the maximum mode combinations for diagnosing PH (Table 5.1) are included and BMI, sex and age are included as factors in the analysis.

Phase	Mode Combination	Accuracy (%)	+Sex	+BMI	+Age
Diastole	15, 10, 20, 21	72.9	71.9	74.7	69.8
Systole	2, 15, 14, 8, 5	68.8	67.7	70.5	65.6

from 68.8% to 70.5% respectively), whereas, adding in subject sex very slightly reduces the ability to diagnose PH (down to 71.9% at end-Diastole, and to 67.7% at end-Systole). The addition of age also reduces the successful classification of PH, from 72.9% to 69.8% at end-Diastole and from 68.8% to 65.6% at end-Systole.

As a further means of evaluation, confusion matrices are reported in Figure 5.5 for end-Diastole and Figure 5.6 for end-Systole. Comparing Figures 5.5a and 5.6a demonstrates that the greater diagnostic potential at end-Diastole is concentrated in the correct identification of non-hypertensive RVs, whereas the end-Systolic shape can better diagnose PH (23/33 vs 25/33 correctly classified). The improvement exhibited when BMI was included as an input was concentrated in the greater number of ‘true negative’ values observed. This was the case at end-Diastole (comparing 5.5a and 5.5c), where one more individual was correctly classified as non-hypertensive, and at end-Systole (through comparing 5.6a, and 5.6c) where 5 more subjects are correctly classified as non-hypertensive, but at the cost of a reduction in correct identification of true positive individuals, i.e. those who suffer from hypertension.

At end-Diastole, including sex in the LDA slightly improved the overall ability of the diagnostic model to identify the RVES of PH sufferers; correctly identified RVs rose by 1 from 23/33 to 24/33 (Figure 5.5b), whereas accounting for age slightly reduced diagnostic success in terms of true positives and true negatives (Figure 5.5d). At end-Systole, on the other hand, including sex in the maximal separating model slightly reduced the LDA function’s ability to discriminate RVES without PH. The number of correctly classified individuals fell from 41/63 to 40/63 (Figure 5.6b), whereas, the model had no impact on the ability to predict PH. Figure 5.6d demonstrates that including age in the model marginally improved identification

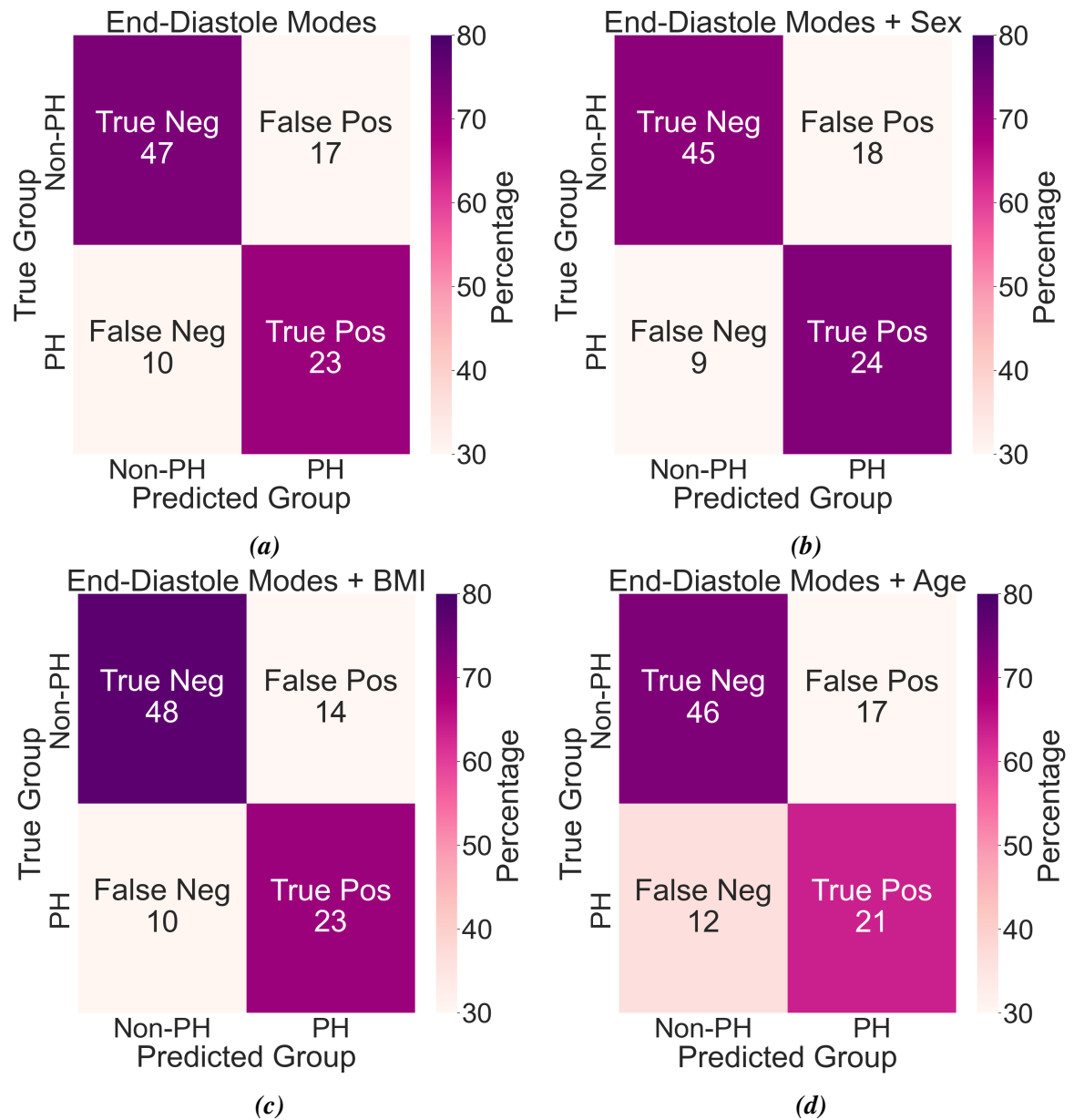


Figure 5.5: Confusion matrices depicting the results of the decomposition at end-Diastole. Figure 5.5a shows the four shape modes only, Figure 5.5b adds sex as a factor, Figure 5.5c adds BMI as a factor and Figure 5.5d adds age as a factor. In all cases, ‘True Neg’ refers to subjects in the non-PH group, who are correctly identified as non-PH; ‘False Neg’ refers to PH sufferers who are incorrectly assigned to the non-PH group; ‘True Pos’ refers to PH sufferers who are correctly identified as such, and ‘False Pos’ to non-PH individuals who are incorrectly classified as PH patients.

Table 5.4: Confusion Matrices showing the breakdown of classification scores based on dataset and disease status grouping when classified by a combination of Modes 15, 10, 20 and 21 at end-Diastole

		Newc. Healthy	Newc. Hypertensive	Pitt. PH	Pitt. Non-PH
Modes	Correct	15	18	24	14
	Incorrect	8	5	9	3
Modes + BMI	Correct	17	16	24	16
	Incorrect	6	6	9	1
Modes + Age	Correct	15	20	23	13
	Incorrect	8	3	10	4
Modes + Sex	Correct	16	18	25	13
	Incorrect	7	5	8	4

of true negative values (those without PH), but also caused a noticeable drop in performance when attempting to diagnose PH (25/33 vs 21/33).

Table 5.4 breaks down the specific starting cohort to which the various correctly and incorrectly classified subjects belonged, whether Newcastle confirmed healthy, Newcastle generalised hypertensive, Pittsburgh non-hypertensive, or Pittsburgh PH. The inclusion of BMI in the model improved classification of the non-hypertensive subjects from Pittsburgh (from 14 to 16 correct), and also in the healthy ageing group from Newcastle (from 15 to 17). There was also a slight reduction in the ability of modes 15, 10, 20 and 21 to classify the generalised hypertensive subjects from the Newcastle dataset (18 vs 16). Therefore accounting for BMI led to increased classification of the non-hypertensive individuals, likely discounting those of lower BMI.

Table 5.4 shows a breakdown of the successful and unsuccessful categorisations using a combination of Modes 15, 10, 20 and 21; the best combination for identifying PH on the basis of structure matrix value. In this analysis, 33 of the 96 subjects (Pitt. PH) were sufferers. From this Table, it can be seen that including BMI in the predictive model improves the identification of the healthy ageing RVs from the Newcastle dataset and the non-PH group from Pittsburgh (altering the split from 15 correct, 8 incorrect to 17 correct, 6 incorrect). There is no change in the breakdown of PH patients correctly identified, and there is a slight reduction in the number

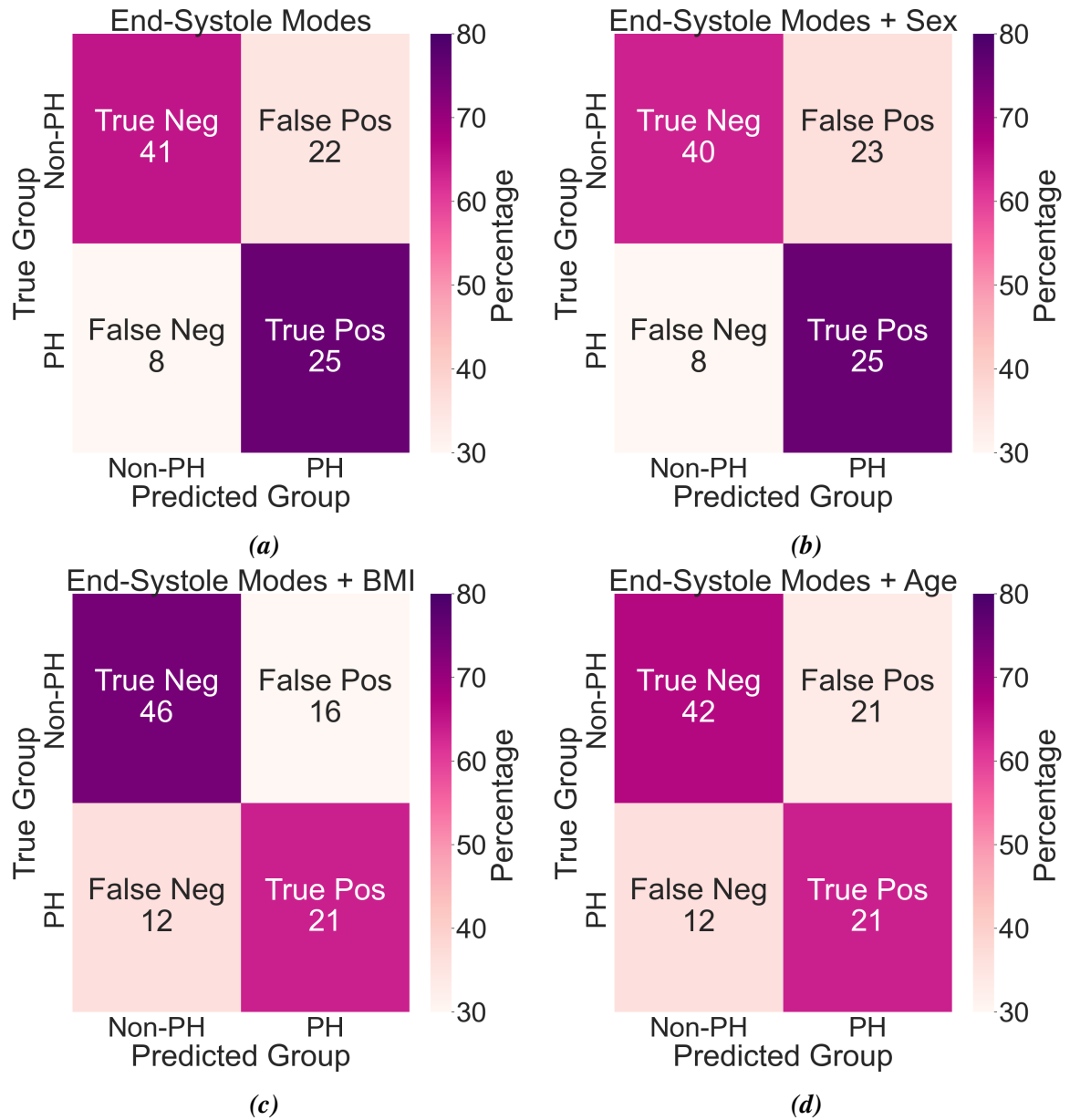


Figure 5.6: Confusion matrices depicting the results of the decomposition at end-Systole. Figure 5.6a shows the five shape modes only, Figure 5.6b adds sex as a factor, Figure 5.6c adds BMI as a factor and Figure 5.6d adds age as a factor. In all cases, ‘True Neg’ refers to subjects in the non-PH group, who are correctly identified as non-PH; ‘False Neg’ refers to PH sufferers who are incorrectly assigned to the non-PH group; ‘True Pos’ refers to PH sufferers who are correctly identified as such, and ‘False Pos’ to non-PH individuals who are incorrectly classified as PH patients.

of generalised hypertensive subjects from Newcastle that have been correctly identified (from 18/5 to 16/6). The slight difference in total subject number when BMI is included relates to the fact that height and weight data is missing for one individual, preventing calculation of BMI. However, it is important to note that these sample sizes are small and the misclassification of two individuals, although indicative, is insufficient to confirm the existence of a trend. The confusion matrices indicate that whilst they are unable to improve overall separability (Table 5.5), accounting for sex very slightly improves the overall ability to diagnose PH to 25 correct, 8 incorrect, and accounting for age decreases the ability of the model to detect PH. Accounting for age does improve the correct identification of the 23 generalised hypertensive subjects.

The results reported in this section suggest that BMI can be an important distinguisher of borderline PH cases, but that the main benefits are in the designation of confirmed healthy individuals. This result must be further explored, since the result is often achieved at the expense of correct diagnosis. For this reason, it is critical to understand, with more granularity which RVs are being classified or misclassified. Therefore, Figure 5.7 shows histograms which describe the distribution of BMI in correctly and incorrectly identified subjects at end-Diastole (using Modes 15, 10, 20 and 21). Figure 5.7b in particular shows a concentration of lower BMI scores in correctly categorised PH individuals, and that histogram peak of the incorrectly diagnosed PH individuals (depicted in blue in these histograms) suggest that they often had higher BMI scores. The implication of this result is that the misdiagnosis of individuals is occurring because of obesity-related changes to the RV.

The addition of the non-pathological factors to the maximum separating combinations was also complemented by a second test, the results of which are reported in Table 5.5. The inclusion of each population-based metric was assessed in a separate LDA, where the input was shape modes 2-30 as well as the feature of interest, i.e. 30 total inputs. Each analysis was performed at both end-Systole and end-Diastole. Combinations of the top five modes were then assessed to find the best diagnostic function. Cross-validated (using leave-one-out cross-validation) accuracy scores are reported, in addition to the best combinations, in the same

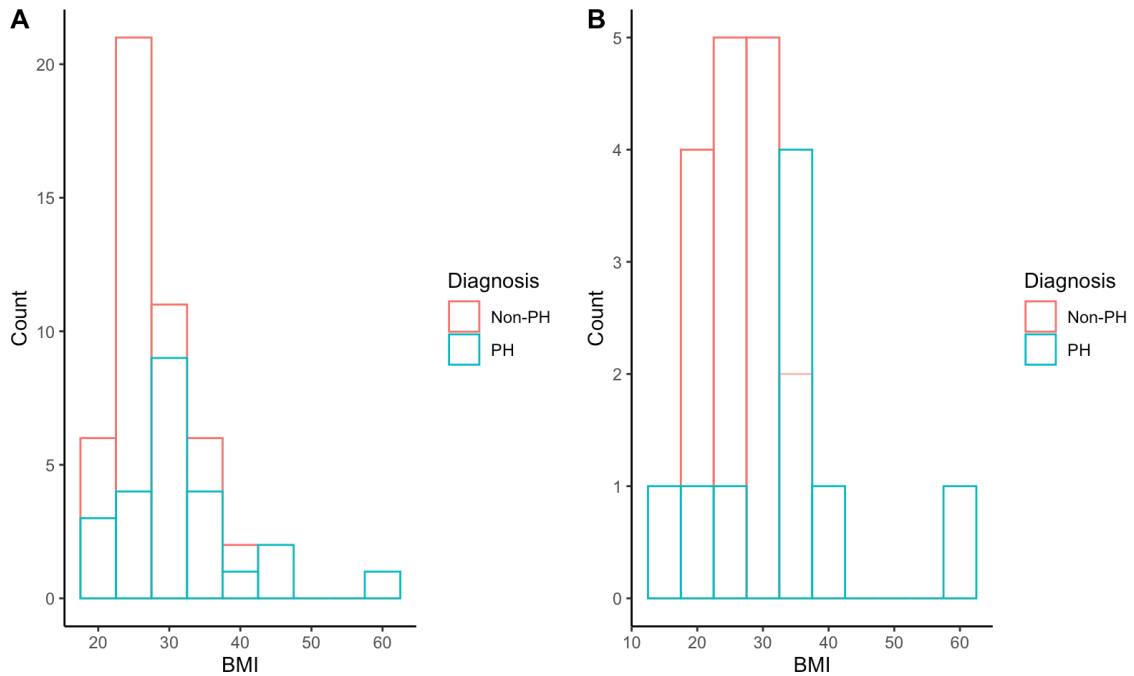


Figure 5.7: *a)* Histogram showing BMI value distribution in correctly identified cases using the LDA algorithm at end-Diastole. *b)* Histogram showing BMI value distribution in incorrectly identified cases.

manner as the initial LDA analyses reported earlier in the chapter. No confusion values are reported for the top five results, since otherwise the effects of age and sex would be unchanged from the initial analysis, based on the shape features alone.

Table 5.5 lists the results when BMI was included as a factor in the LDA analysis. By also accounting for BMI in the model, the successful separation percentage rose to 74.7% after cross-validation using the top five input features: BMI, and Modes 15, 2, 10, 20. The best combination of modes at end-Systole achieved 75.8% accuracy in separation. This was achieved using a combination of BMI, Modes 15 and 13. Mode 15 has a role in identifying obese individuals at end-Systole, as well as being a key in diagnosing PH. Mode 13 was also identified in Chapter 3 as an important mode for separating older from younger non-hypertensive individuals. It is interesting that, once BMI is included in the discriminating LDA function, the impact of Mode 13 becomes visible. The appearance of such interactions warrants consideration of smaller subdivisions of the dataset, to control for interactions of the

Table 5.5: Best Combinations for characterising PH using Linear Discriminant Analysis when Modes 2-30 and BMI, sex and age are included as factors. Cross-validated percentage accuracy scores are also reported.

Diagnosis	Phase	Combination	Accuracy (%)
BMI	Diastole	BMI, 15, 10, 2, 20	74.7
BMI	Systole	BMI, 15, 13	75.8
Sex	Diastole	15, 10, 20, 21	72.9
Sex	Systole	2, 15, 14, 13; 2, 15, 5, 13; 15, 14, 5, 13	70.8
Age	Diastole	15, 10, 20, 21	72.9
Age	Systole	2, 15, 14, 5	67.7

population characteristics.

Adding sex as an input to the diagnostic model, in conjunction with shape modes 2-30, correctly classifies 70.8% of the input subjects, based on their diagnosis. This represents a slight improvement over the baseline, even though sex did not feature as one of the top five modes. Although each subset of the input factors creates an entirely new LDA function, the inclusion of sex interacted with Mode 13, and increased its importance to the model. Sex also made no significant appearance in the ranked features, based on structure matrix values in the end-Diastolic group (structure matrix coefficient of -0.124, compared to -0.482 for Mode 15). Furthermore, the same top five modes were identified, and the optimal separating combination for diagnosing PH remained the same as the initial results, reported in Table 5.1. Therefore, accounting for sex as part of this LDA model worsened the diagnostic potential, as compared to using fundamental shape features, derived from the RV.

Table 5.5 also confirms that accounting for age, the third of the three predictors, in the discriminant model has no substantial effect. Age does not feature in the top five discriminating modes at either end-Systole or end-Diastole, nor does it alter the top five discriminating modes. The best discriminating combinations remained: Modes 15, 10, 20 and 21 (72.9% separation) and Modes 2, 15, 14, 8, 5 (68.8% separation) at end-Systole. Therefore, accounting for age in the linear discriminant model did not improve the ability to diagnose PH in any meaningful way.

The most substantive finding of the initial characterisation was that the shape modes which best predicted PH were left unchanged from the initial analysis, which used shape features only, whereas BMI emerged as the strongest discriminating component in the model at end-Diastole. This effect was mirrored at end-Systole, although in this case a smaller subset of features (BMI, Mode 15 and Mode 13) emerged as the optimal combination.

From the initial application of BMI, age and sex to a discriminating shape model containing all 96 individuals, it appears that BMI has the most significant impact on diagnostic potential. Notably, BMI is always a strong discriminating factor based on its correlation to the discriminant function (or structure matrix value). There also seems to be a slight improvement in the ability of a model to diagnose PH in diastolic subjects when sex is included. Including age in a model alongside shape modes appears to make little difference to the diagnostic potential.

5.4.3 Subdivision of Datasets to Control for Population Characteristics

Earlier work in this thesis has established the presence of distinctions in right ventricular shape based on sex, age and obesity status when removed from a disease context and has established that BMI and sex can improve shape-based separation. Such substantive shape differences may warrant the need for separate diagnostic models for subjects, and therefore, to test this, the final set of linear experiments performed separate LDAs for each of the dichotomous groups, enabling any interactions between characteristics to be extracted and observed. To evaluate the effect of non-pathological characteristics on the diagnostic potential of RV shape features, it is important to consider two angles. First, whether including those population factors positively or negatively impacts the ability to diagnose PH using the shape feature baseline combinations, which were described at the opening of this chapter (Subsection 5.4.1, Table 5.1), and secondly, what effect introducing such factors into the model has on the most prominent separating shape features; i.e. whether the introduction of sex or age as a feature changes which modes separate the data with the highest accuracy. To answer the first of these

Table 5.6: Composition of each subject category, total values and division into PH and non-PH groups. The cut-off for the age category was mean age across all subjects (60.5 years), and the cut-off for Obesity was a BMI of $28\text{kg}/\text{m}^3$. The Non-PH category is comprised of 63 subjects; 23 healthy ageing, 23 RVs with generalised hypertension and 17 non-PH subjects from the Pittsburgh cohort.

Category	Total	Non-PH	PH
Male	52	32	20
Female	44	31	13
Young (< 60.5 years)	41	26	15
Old (≥ 60.5 years)	55	37	18
Non-Obese (< $28\text{kg}/\text{m}^3$)	48	38	10
Obese ($\geq 28\text{kg}/\text{m}^3$)	47	24	23

questions, LDA was performed using the maximal separating combinations for end-Systole and end-Diastole as inputs (see Subsection 3.3.2 for an overview, and Section 5.3 for specific description of the training and evaluation of classifiers in this chapter), separately split into the two population subgroups. Then, the two remaining factors (i.e. sex and age) were also added into that model, in addition to the key shape modes, creating a total of five or six input factors into each analysis. Cross-validated accuracy percentages were reported. To answer the second question, a separate LDA test was performed, including Modes 2-30 plus each factor in turn, to derive the top five ranked modes for each subdivided group.

Table 5.6 breaks down the PH vs non-PH individuals in each category and also describes the number of subjects featured in each category. The numbers in the Obese and non-Obese are almost equal (47 vs 48), as this was one decision criterion for the split at $28\text{kg}/\text{m}^2$. It is therefore striking how many more PH sufferers there are in the higher BMI ('Obese') group compared to the non-Obese group. A smaller proportion of PH individuals are in the old category than the young category. This is initially surprising, but can be explained by the composition of the Pittsburgh PH dataset, where many of the referrals were for young patients. There are more male subjects in this dataset than female (52 vs 44), but noticeably fewer women with a diagnosis of PH.

Table 5.7 shows the best classification and modes when performing LDA on male and

female subjects separately. This dataset consisted of 44 female subjects, of whom 13 suffered from PH and 31 did not (Table 5.6). The results described in this table show that the maximum separating mode combination is significantly better at identifying PH in male subjects as compared to female subjects in all cases. At end-Diastole, the successful classification percentage is only slightly above chance in female RVs (54.5%). In female individuals, accounting for BMI improves that classification from 54.5% to 67.4%, whereas age slightly worsens it (from 54.5% to 52.5%). In male RVs, accounting for BMI improves classification of BMI to 80.8% and adding age into the model very slightly improves the result to 75%. It is therefore evident that the PH related shape features, identified across the whole cohort are much more strongly represented in male hearts than female hearts. This is slightly surprising, given the overlap of features between male RVs identified in Chapter 3 and those associated with PH. Although, when BMI is included as a factor, that negative effect is somewhat mitigated.

Figure 5.8 displays confusion matrices, which illustrate the success of the PH prediction by dividing out the results into true and false negatives and positives. Examining the confusion matrices shows that the gains in female diagnosis of PH when BMI is included in the model all occur as a result of improved categorisation of healthy female individuals: 22 correct, 8 incorrect compared to 16/15 (Figures 5.8a & 5.8b). There is a slight reduction in the successful classification of female PH sufferers. In male individuals, the effect is similar, with the majority of gains observed in the true negative (non-PH) category. Although the male model successfully identifies one additional RV with PH. The inclusion of age in the model also worsens the discriminating capacity in determining PH status. The number of correctly assigned PH patients falls by one to 13/20 (Figure 5.8i).

At end-Systole, only 63.6% of the RVs were correctly separated based on PH. As can be seen from Table 5.7 and the confusion matrices in Figure 5.8, including BMI in this analysis slightly improved the classification from 63.6% to 65.1%. Whereas, accounting for age reduced the successful classification percentage slightly to 61.4%. The number of correctly identified PH sufferers was highest in the baseline classification, and adapting the model to

Table 5.7: Top Five identified Modes, ranked in order of importance, when separate LDA analyses are performed on Male and Female subjects. Also reported are the associated classification percentages using these modes only, and when BMI and age are included as inputs to the model.

Phase	Analysis	Classification %	+BMI	+Age	Ranked Modes
Systole	Female	63.6	65.1	64.1	8, 2, 7, 26, 19
Systole	Male	71.2	67.3	73.1	14, 15, 20, 2, 13
Diastole	Female	54.5	67.4	52.5	5, 19, 23, 15, 18
Diastole	Male	73.1	80.8	75.0	10, 15, 20, 11, 17

include BMI and age both resulted in the successful identification of PH in female subjects falling from 8/13 to 7/13 (Figures 5.8e & 5.8f). PH status was correctly deduced in 71.2% of the male-only group using the five top shape features. This was an improvement over the baseline rate of 68.8% (calculated across the 96 subject group, and shown in Table 5.1). Accounting for BMI reduced the accuracy of the LDA model in male subjects, whereas accounting for age slightly improved the classification to 73.1%.

The top five modes reported for male RVs are also reported in Table 5.7 and much more closely match the overall best classifier of PH than the female features. The ranked modes for male RVs at end-Diastole included Modes 10, 15 and 20 and the ranked features at end-Systole also included Modes 14, 15. On the other hand, the top features for discriminating PH in the female RVs differed far more substantially from the baseline. BMI was important in the classification of PH in both sexes, but was a more significant factor in females than in males. Mode 15 also appears as an important identifier of PH in both sexes. Mode 19 and Mode 23, which are both Modes that appear in the identification of sex are important classifiers of female subjects, as well as Mode 5 which appears in the classification of age in Chapter 3.

In all but one of the tests for either sex, including BMI in the separating model resulted in an overall improvement in the classification result. The exception was males at end-Systole, where a slight reduction in classification percentage was observed (from 71.2% to 67.3%). This improvement tended to manifest itself in the correct identification of subjects without PH and in actuality represented a worsening ability to differentiate the individuals with PH.

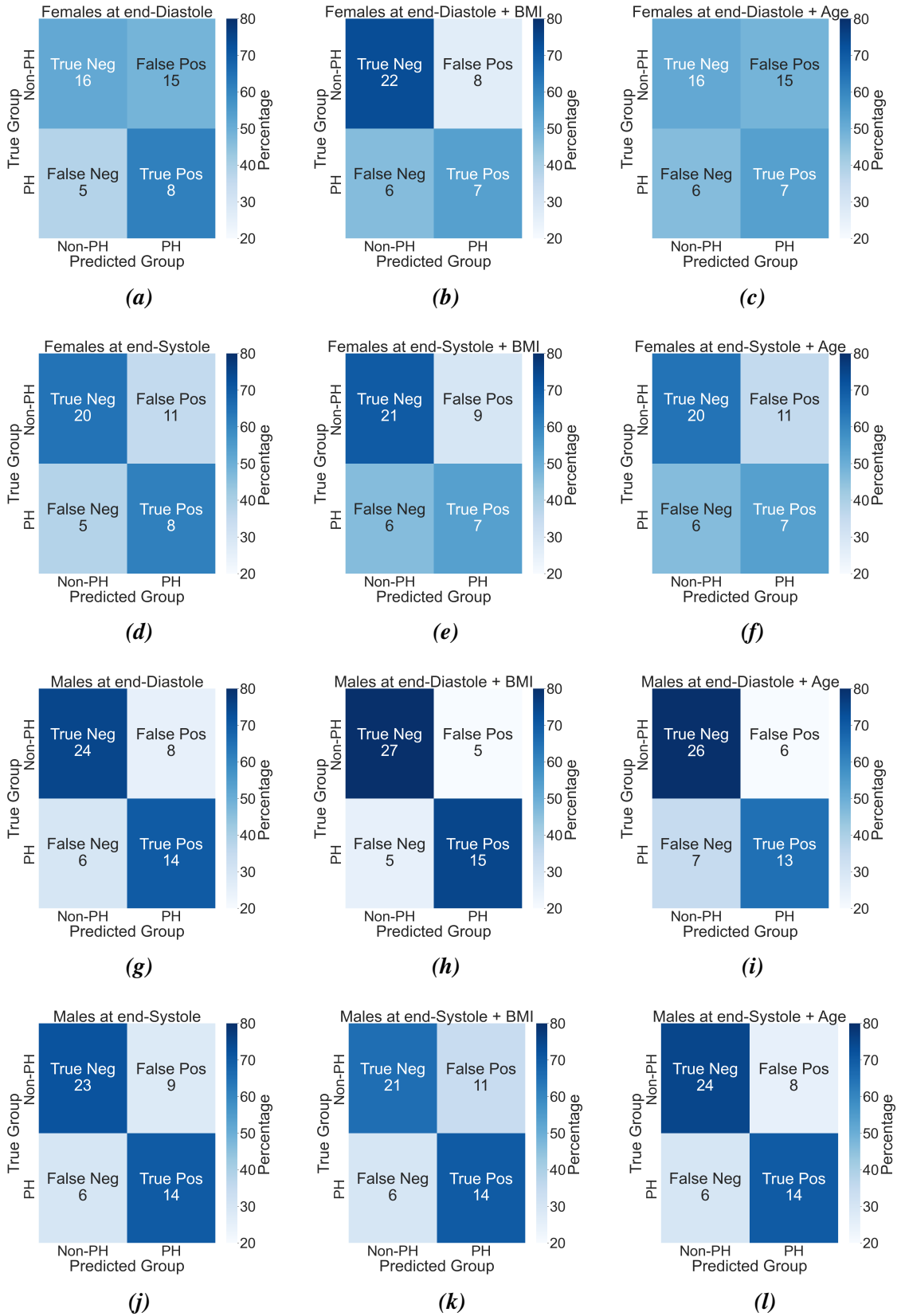


Figure 5.8: Sex-based confusion matrices. The left hand column (Figures 5.8a, 5.8d, 5.8g, 5.8j) depicts the results of the decomposition using shape modes only. The central column shows the results when BMI is included as a factor, and the right hand column shows the results when age is included as a factor.

Table 5.8: Top Five identified Modes, ranked in order of importance, when separate LDA analyses are performed on the obese and non-obese groups. Also reported are the associated classification percentages using these modes only, and when Sex and age are included as inputs to the model.

Phase	Analysis	Classification %	+Sex	+Age	Ranked Modes
Systole	Non-Obese	66.7	66.7	64.6	15, 2, 29, 26, 21
Systole	Obese	63.8	66.0	66.0	6, 29, 11, 22, 26
Diastole	Non-Obese	72.9	70.8	68.8	20, 11, 8, 15, 10
Diastole	Obese	66.0	66.0	68.1	2, 21, 17, 27, 15

This is evident from the comparing the confusion matrices from the classification using just shape features to those where BMI is also included in the model (e.g. the end-Diastolic female values depicted in Figures 5.8a and 5.8b). Tests involving a larger sample size would further clarify these relationships.

It is notable that BMI was the greatest discriminating factor in both male and female individuals, but did not improve the ability to detect PH. Having isolated BMI's contribution in the different sex groups, the next test considered the impact of different modes when subjects were subdivided based on obesity. Table 5.8 shows the top five modes when the RV groups were assessed separately based on their obesity status. The lower BMI group (below $28\text{kg}/\text{m}^3$) comprised 38 individuals without a PH diagnosis and 10 individuals with a PH diagnosis. The higher BMI group consisted of 47 individuals, of whom 24 were non PH and 23 suffered from PH.

Table 5.8 shows the top five modes when the RV groups were assessed separately based on their obesity status. The lower BMI group (below $28\text{kg}/\text{m}^3$) comprised 38 individuals without a PH diagnosis and 10 individuals with a PH diagnosis. The higher BMI group consisted of 47 individuals, of whom 24 were non PH and 23 suffered from PH. Figure 5.9 describes BMI-based confusion matrices. These describe the breakdown of the classification percentage into true and false negative and positive results. The colour scheme is indicative of the percentage concentrated in each group, with darker blue values in the region of 80%.

At end-Diastole, Modes 15, 10, 20 and 21 were able to separate *non-obese* RVs with PH

with a success rate matching that achieved across the whole cohort (72.9%). The inclusion of sex and age as inputs to the model reduced the efficacy of this classification to 70.8% and 68.8% respectively. Comparison of Figures 5.9g, 5.9h & 5.9i indicates that the poorer performance of the classifier was related to a slight reduction in correctly classified PH patients (from 7/10 to 6/10). Although this reduction in successful classification is observed when sex or age is included in the model, it is worthy of remark that the non-obese dataset contains notably fewer RVs than the obese dataset. In *obese* RVs at end-Diastole, the overall success of the diagnostic classifier is worse: it is able to correctly categorise just 66% of individuals based on their PH status; 16/24 non-PH and 15/23 with PH (Figure 5.9a). Including sex in a predictive model comprised only of obese individuals has no impact on the percentage accuracy or the categorisation splits (Figure 5.9b), and including subject age very slightly improves the accuracy score to 68.1%, where the improvement comes from the correct classification of an additional PH sufferer (Figure 5.9c).

The maximum separating shape modes for identifying PH at end-Systole (Modes 2, 15, 14, 8 and 5) were a better match for the non-obese RVs than obese RVs (66.7% compared to 63.8%). Through inspecting Figures 5.9a & 5.9g, the performance gain in the non-obese RVs is concentrated in strong identification of non-hypertensive individuals, whereas the stronger separation performance of the LDA in obese individuals was able to correctly separate PH. Accounting for sex in the model had no impact on non-obese individuals (66.7% accuracy), whereas accounting for age slightly reduced the successful classification rate for PH to 64.6% of individuals. The confusion matrices in Figures 5.9j, 5.9k & 5.9l demonstrate that a diagnostic model can only accurately separate 50% of hearts with PH in non-obese RVs, regardless of intervention by including subject age or sex in the diagnostic model. The baseline shape modes are not a clear diagnostic indicator of the presence of PH in non-obese individuals. In obese individuals, both of the included biological factors constituted an improvement over separating groups using the five shape modes alone; including sex or age as an input to the LDA increased diagnostic success to 66%. However, this still remains less successful than

Table 5.9: Top Five identified Modes, ranked in order of importance, when separate LDA analyses are performed on the below and above mean age groups. Also reported are the associated classification percentages using these modes only, and when BMI and Sex are included as inputs in the model.

Phase	Analysis	Classification %	+BMI	+Sex	Ranked Modes
Systole	Younger	68.3	75.0	73.2	28, 2, 25, 14, 10
Systole	Older	63.6	60.0	63.6	15, 28, 2, 14, 8
Diastole	Younger	65.9	72.5	63.4	15, 6, 7, 2, 10
Diastole	Older	72.7	72.7	67.3	15, 10, 21, 2, 5

the baseline separation. Therefore, the modes identified as the best discriminators of PH at end-Systole are not as effective in subjects that are obese, as compared to the population as a whole. Separating out the RVs based on BMI hindered the accuracy of the baseline shape model in most cases, again strengthening the argument of considering BMI when evaluating a diagnostic model.

In terms of considering the important ranked modes in each of these subcategories, several prominent Modes from the Population analysis (Chapter 3) reappear when using shape features to predict PH. Modes 20, 11 and 2 were identified as key discriminators of obesity at end-Diastole, and these return here. In Table 5.8, Modes 20 and 11 are the top two modes in the obese group, and Mode 2 is the top discriminating mode for the non-obese subjects.

Finally, two separate LDAs were performed on the younger and older age groups. Table 5.9 shows the ranked mode and classification scores for the above and below-mean age groups. The younger age group contained 41 subjects; of these, 26 did not have PH and 15 were PH sufferers. There were 55 older individuals in the dataset, and of these, 37 did not have PH and 18 were PH sufferers. It is important to combine these results with those from the confusion matrices in Figure 5.10. An improved separator does not always mean that PH itself is distinguished. The improvements might be in clarifying the boundaries of other categories.

At end-Diastole, when the four best modes used to diagnose PH across the entire cohort are applied to the younger (below mean age) individuals, 65.9% of RVs are successfully classified. Including BMI in the model improves the LDA model's ability to correctly identify PH, with

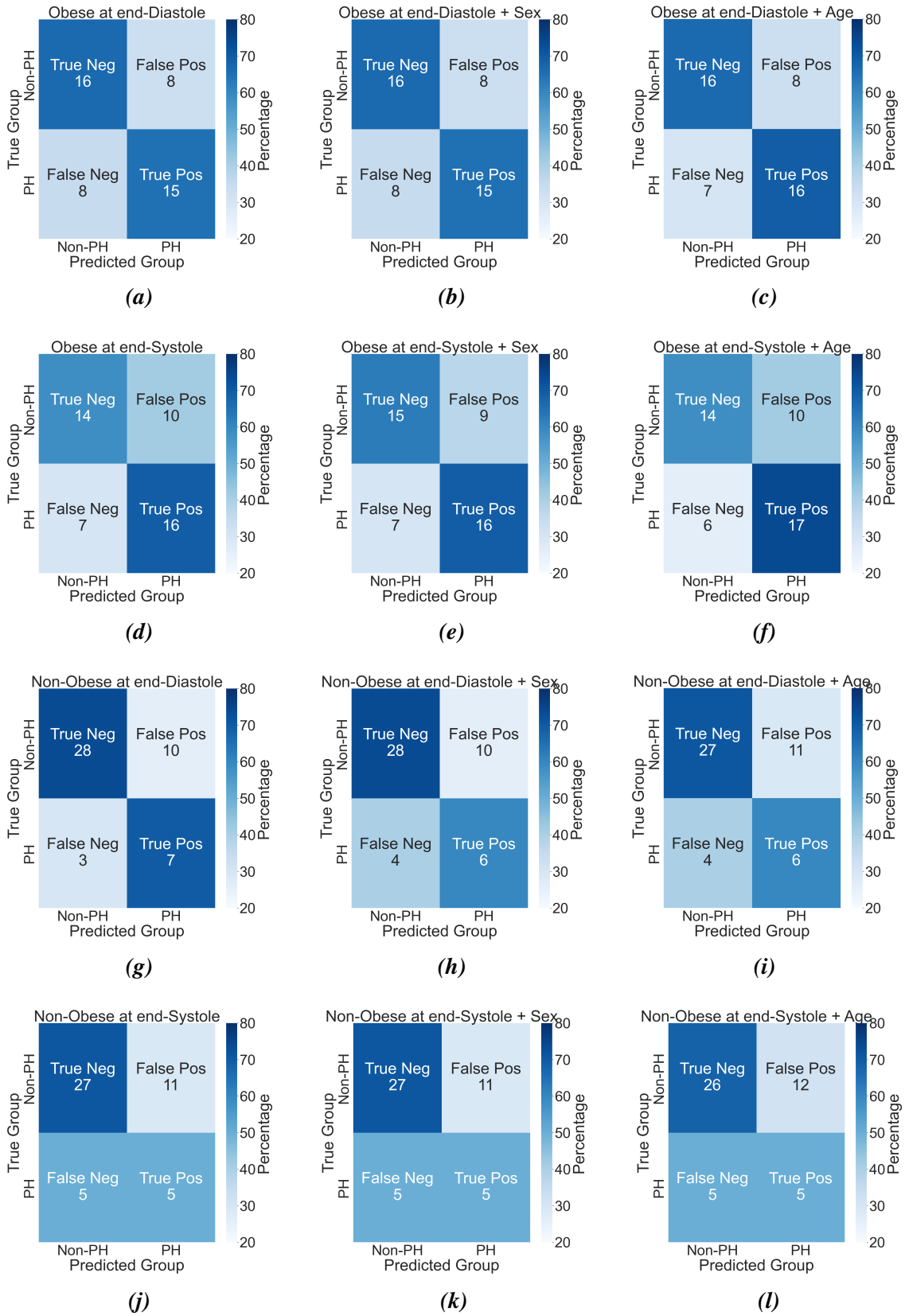


Figure 5.9: BMI-based confusion matrices. The left hand column (Figures 5.9a, 5.9d, 5.9g, 5.9j) depicts the results of the decomposition using shape modes only. The central column shows the results when sex is included as a factor, and the right hand column shows the results when age is included as a factor.

a percentage accuracy of 72.5%. Figure 5.10h suggests that this improvement is concentrated in the non-PH group: the split in the confusion matrices showed a rise from 16/26 correctly identified when the classifier was built using shape modes only to 19/25 upon the inclusion of BMI. The number of correctly identified PH sufferers ('true positives') fell from 11/15 to 10/15. It can be stated that including BMI in a model enables younger subjects without PH to be more accurately separated from those with PH, but has no effect, or a slightly detrimental effect on the ability to detect those with PH. It is important to consider that accounting for subjects with a higher BMI also makes it harder to distinguish between those individuals who are hypertensive but who do not have PH. Finally, including sex in the model slightly reduces the model's ability to discriminate PH from 65.6% to 63.4% accuracy. Figure 5.10i indicates that the additional misclassification occurs in the PH group. Therefore, it appears that BMI is an important predictor of PH in younger subjects, whereas sex is not. Modes 15, 10, 20 and 21 have relatively better success separating the older subjects. The classification percentage for older individuals is in line with that achieved when analysing the whole cohort (72.7%), and this result remains unchanged when BMI is added into the model. Although comparing Figures 5.10a & 5.10b indicates that the addition of BMI slightly improves diagnosis of PH (from 11/18 to 12/18). The efficacy of the separation is reduced to 67.3% when sex is included as a factor. Comparing the confusion matrix in Figure 5.10c, accounting for subject sex in the model slightly reduced the successful classification of both the true negative and true positive groups. It is apparent that the characteristic shape changes which are associated with PH diagnosis in the original 96 subject group are better aligned with the shape of the older RVs than the younger ones, but that this alignment largely manifests itself in the ability to separate the shape of RVES without PH.

Evidence from Table 5.9 also confirms that end-Systolic modes perform better on the younger subjects than the older subjects. The maximum separating shape features correctly identify PH status in 68.3% of RVs. The confusion matrix in Figure 5.10j indicates that greater percentage success was achieved in predicting individuals with PH (11/15) than predicting

those without. Accounting for BMI in the younger group improved the LDA's ability to separate individuals up to 75%, with no change in the number of PH identified, but an increase from 16/26 to 19/25 non-PH (Figure 5.10k). On the other hand, accounting for BMI in the older RVs lowered that efficacy from 63.6% to 60.0%. When sex is added to the model in addition to the five maximum separating features, there is a slight improvement from 68.3% to 73.2% separation in the younger subjects, albeit less so than when BMI is included. In the older individuals, no change is observed. Therefore, it appears that including BMI in the model is important to help distinguish PH in younger individuals (Figure 5.10k), but is often a hindrance in older subjects. This is the same trend as was observed in evaluating the effect of separating factors in end-Diastolic RVs. Similarly, including age in the LDA predictive model affords a very slight benefit to detecting subjects with PH in the younger group. The older group are much better accounted for by the end-Diastolic shape modes, compared to the best end-Systolic shape modes when BMI is added to the model.

When evaluating the impact of the inclusion of age as a factor on the top five modes identified based on their structure matrix values, several common modes again feature prominently in this analysis. Mode 15 is present as an important discriminator in both age categories at end-Diastole, and in the older ventricles at end-Systole. Although crucially, not in the +BMI test at end-Systole. As confirmed by the confusion matrices, this could still point to the importance of Mode 15 in diagnosing PH. BMI also features in older end-Systolic hearts as well as at end-Diastole.

These results suggest that BMI improves detection of non-hypertensive subjects, and their ability to discount generalised hypertensive individuals. However, the cost is a reduced detection of PH alone. In the older group at end-Diastole, accounting for subject sex in the linear discriminant model affords a slight benefit to detecting subjects with PH. From the wealth of analyses conducted in this section. Two interactions in particular stand out, and both involve the inclusion of BMI alongside shape-based features. The first of these is the separation of male RVs using the end-Diastolic modes (Figure 5.8h) and the second is separation of the

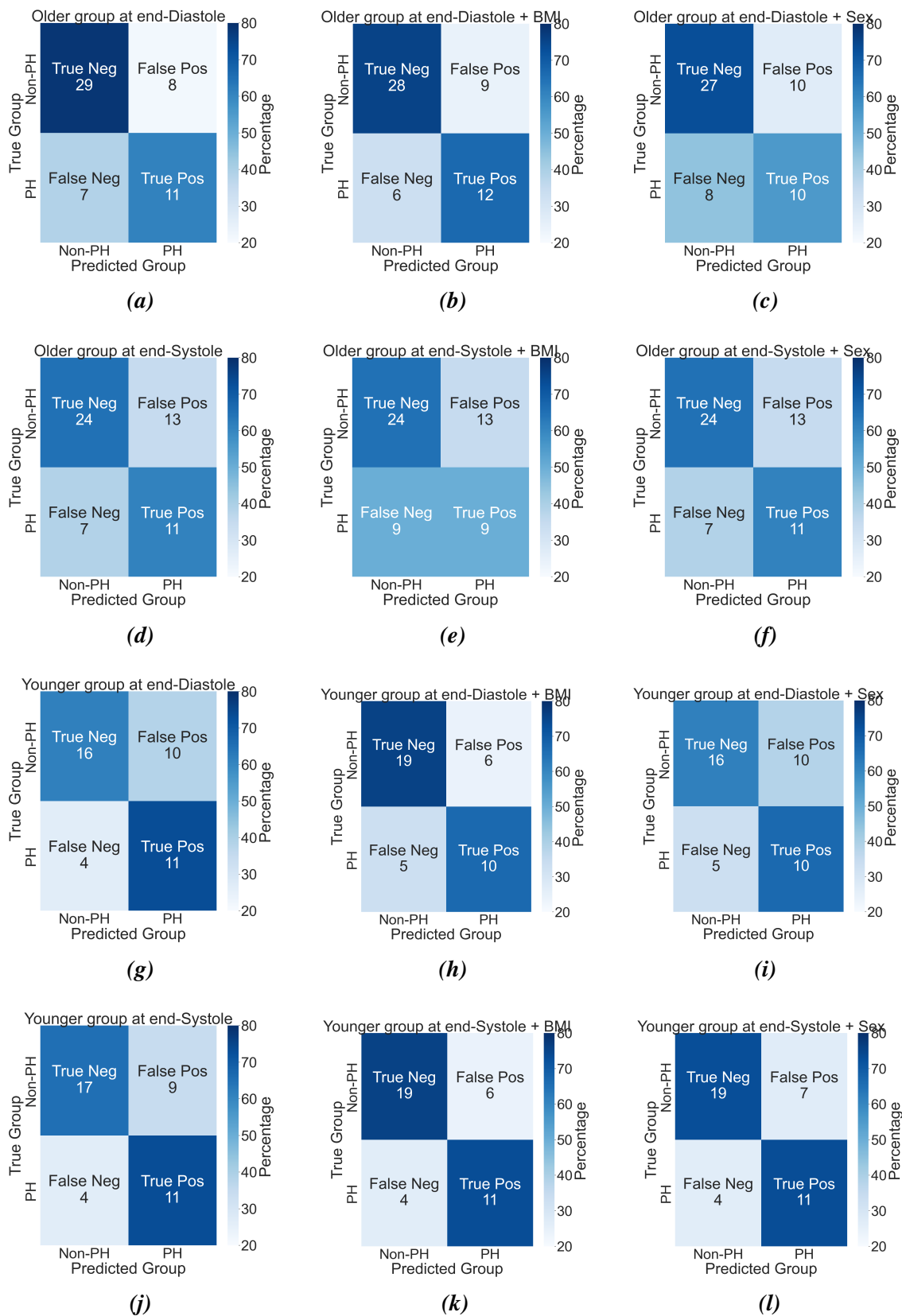


Figure 5.10: Age-based confusion matrices. The left hand column (Figures 5.10a, 5.10d, 5.10g, 5.10j) depicts the results of the decomposition using shape modes only. The central column shows the results when BMI is included as a factor, and the right hand column shows the results when sex is included as a factor.

younger group of end-Systolic shapes (Figure 5.10k). The linking factor between these two successful analyses is the inclusion of BMI in the model.

The shared shape characteristics between male individuals and PH sufferers may be masking the ability to detect hypertension. In the same manner, the stronger potential of shape modes to identify PH in obese RVs, despite an overall drop in the model's ability to successfully classify individuals without PH. Therefore, a more nuanced, non-linear classifier should be employed as this might improve differentiation. Several non-linear classifiers were considered for this analysis, including a kernel support vector machine (SVM). One advantage of using an SVM is the ability to 'map higher dimensions' without dramatically increasing computing resource.

5.4.4 Implementation of Support Vector Machines

SVMs were implemented in an attempt to improve upon the classification success of the linear discriminant model applied in Subsection 5.4.1. Although 72.9% of the of RVs could be correctly categorised based on their PH status at end-Diastole, there is a possibility of non-linear interactions between the shape features which would result in significantly greater diagnostic success with a non-linear classification method. The default parameters were tested and then a grid search was used to tune the optimal cost and γ hyper-parameters as described in Subsection 5.3.2 of the methodology. The results were cross-validated and averaged over 100 attempts to ensure that the 75:25 sampling ratio did not impact the results from any given analysis. Taking into account the impact of including the various non-pathological characteristics in the linear discriminant models (Subsection 5.4.3), BMI has by far the greatest effect on the ability of shape modes to divide the cohort into various categories. This was followed by sex. Age had largely no impact, or an adverse impact on linear categorisation, so was not included at this stage. Modes 2-30 were included as inputs to the initial SVM. Then 2 subsequent analyses were performed, in turn adding BMI and sex.

Table 5.10 shows the mean accuracy and standard deviation of the four tested SVM kernels

at both end-Systole and end-Diastole. Table 5.11 provides the results when BMI is added, and Table 5.12 when sex is added to the SVM. At end-Diastole using just Modes 2-30, the best averaged result of 71.3% was achieved with a Polynomial kernel, while a straightforward linear kernel achieved 70.0% separation accuracy. A radial kernel accomplished a slightly lower mean accuracy of 67.6%, but with lower standard deviation than the other samples (2.4%). Including BMI at end-Diastole (Table 5.11) resulted in a slight improvement in linear accuracy from 71.3 to 72.6%, and in sigmoid accuracy from 67.0% to 68.0% but no substantive difference in the performance of the polynomial or radial kernels. As with the SVMs including the shape modes only, the radial kernel demonstrated a narrower standard deviation than the other kernels (2.6 vs > 9). Finally, Table 5.12 shows the implementation of the four SVM kernels when Modes 2-30 and Sex were included in the model. The best performing kernel was linear, and the mean classification was 71.6%, which represented an improvement over the classification using only shape modes, although not over the linear kernel's mean accuracy when BMI was included (Table 5.11). The sigmoid and radial kernels predicted PH with poorer accuracy than using the 29 shape features alone. In general, the models performed better at end-Diastole than end-Systole, mirroring a trend seen throughout this thesis. The best average accuracy achieved using any SVM (72.6%) still did not improve upon the separation achieved in the linear discriminant analysis.

At end-Systole, no mean accuracy score exceeded 70%. The mean accuracy and standard deviation achieved at end-Systole when averaged across 100 training splits is contained in Table 5.10. Using shape Modes 2-30, the best obtained mean value was achieved by implementing a polynomial kernel (69.3%), the radial kernel was the second best predictor of PH (68.5%) and this result also had the smallest standard deviation (4.2%). Tuning a polynomial kernel took approximately 48 hours in comparison to several minutes for the linear kernel, owing to the larger number of input parameters. Consequently, there seems to be no additional benefit to using a more complex kernel in diagnosing PH using this dataset. The linear and sigmoid kernels were able to predict PH using RV shape with 64.2% and 62.8% accuracy,

Table 5.10: Optimally tuned SVM results using Modes 2-30.

Kernel	Accuracy (%)	Std Dev
Diastole		
Linear	70.0	8.1
Radial	67.6	2.4
Sigmoid	67.0	8.5
Polynomial	71.3	8.4
Systole		
Linear	64.2	8.4
Radial	68.5	4.2
Sigmoid	62.8	8.0
Polynomial	69.3	8.4

Table 5.11: Optimally tuned SVM results using Modes 2-30 and BMI. BMI value is included as an input factor to the SVM, in addition to the shape mode coefficients.

Kernel	Accuracy (%)	Std Dev
Diastole		
Linear	72.6	9.6
Radial	68.0	2.6
Sigmoid	68.0	9.4
Polynomial	71.4	9.9
Systole		
Linear	67.1	9.6
Radial	68.4	3.7
Sigmoid	62.0	8.67
Polynomial	67.8	7.8

Table 5.12: Optimally tuned SVM results using Modes 2-30 and Sex. Sex is included as an input factor to the SVM, in addition to the shape mode coefficients.

Kernel	Accuracy (%)	Std Dev
Diastole		
Linear	71.6	8.1
Radial	67.0	2.1
Sigmoid	64.8	9.5
Polynomial	69.3	8.3
Systole		
Linear	64.6	8.3
Radial	69.0	3.6
Sigmoid	62.7	7.12
Polynomial	67.2	8.0

respectively. When BMI was included in the SVM model, in addition to the shape modes at end-Systole (Table 5.11), the radial kernel was the most successful at identifying individuals with PH (68.4%), this result did not improve upon that achieved using shape modes only. The results obtained by the linear and polynomial kernels were almost identical (67.1% and 67.8% respectively). Including BMI resulted in an improvement in accuracy of the linear kernel, but a slight reduction in accuracy of the polynomial kernel (albeit with a lower standard deviation value). The sigmoid kernel remained the poorest kernel for dividing RV shape based on PH. Table 5.12 reports the mean accuracy and standard deviation for the shape modes and Sex at end-Systole. The best mean accuracy achieved upon the addition of sex was a mean cross-validated accuracy score of 69.0%. Again, this used a radial kernel. The result represented a 0.6% improvement over the analysis using shape modes only (Table 5.10), which is not a significant bonus.

In all cases, the radial kernel exhibited the smallest standard deviation, suggesting much closer agreement between trials using this kernel when different training and test splits were used. Since the performance of the radial kernel was often at the higher end of the cross-

validated, tuned scores, a radial kernel appears the best available choice. At end-Diastole, the best overall accuracy was achieved using a linear kernel and including BMI in the model. At end-Systole, the best possible solution was achieved using a polynomial kernel and by separating using the shape modes only. However, overall, the results obtained in this section do not suggest that there is a substantial benefit to using a non-linear kernel over the linear discriminant scores that have been considered throughout the remainder of this work.

5.5 Conclusions

The hypotheses outlined at the beginning of this chapter were that 1) a specific shape mode combination will define anatomical shape features that are specific to PH; 2) including population-based information such as BMI or sex in the model will improve that model's ability to distinguish individuals with PH given the unique shape features that could be predicted by groups of modes in Chapter 3, and 3) utilising a non-linear model will improve overall accuracy. A baseline predictor for RV using shape modes was identified within this dataset, and then this was used to evaluate the impact of the three non-hypertensive characteristics.

First, the results obtained in this chapter suggest that Mode 15, which exhibits a flattening of the ventricular base congruent with PH, is involved in separating diseased individuals from non-diseased individuals. This aligns with previous findings from the literature which identify flattening of the RV base in addition to hypertrophic rounding, as key shape-based features associated with PH. A four mode combination: Modes 15, 10, 20 and 21 is able to correctly identify 72.5% of the 96 individuals with PH at end-Diastole, and a five mode combination: Modes 2, 15, 14, 8, 5 is able to correctly predict PH in 68.8% of individuals. The impacts of PH on the RV are easier to identify in end-Diastolic hearts than end-Systolic hearts. However, Mode 15 is also an important determiner of PH in systolic hearts. Since the end-Systolic and end-Diastolic RVES were all statistically decomposed together, Mode 15 is encoding the same shape feature in both. Therefore, in accordance with the first hypothesis,

a baseline differentiator of PH using the available shape modes was identified, which could then be applied to evaluate the effect of accounting for the non-hypertensive characteristics.

Regarding the second hypothesis, it is possible to improve upon the detection of PH in a statistical shape model by including a component of BMI at both end-Diastole and end-Systole. Examination of structure matrix components show that BMI is the strongest discriminator of PH. As Mode 15 was involved in classifying non-hypertensive individuals based on their obesity (into categories of BMI > 28 and BMI < 28), BMI was included in the final model to improve classification of PH. However, although the models which included BMI often improved the classifier, this was often in non-PH individuals and in several analyses reduced the successful classification of PH. The ability of the diagnostic model to correctly identify subjects with PH was reduced in several cases; notably in the LDA assessment of all 96 individuals at end-Systole, and in some of the subcategory breakdowns, for example, the older subject group at end-Systole. The findings associated with dividing the analyses up into subcategories could be crucial for curating better weighted diagnostic models.

Finally, the third hypothesis was disproven in this chapter. Although in certain isolated cases, slightly improved results were achieved when applying a polynomial SVM kernel, this code took over 24 hours to run on Durham University's high performance computer. The improvement in classification results does not warrant the vastly increased processing over use of a linear discriminant separator or a linear kernel in an SVM. Further, it appears that the modes identified using linear methods are sufficient to identify the strongest disease-related features, and there is no significant benefit to the additional computing power required to undertake a full cross-validated SVM sweep. There seems to be no specific added benefit, using a SVM-based classifier at least, of non-linear classification of the statistical shape analysis method for the purposes of disease diagnosis. Although a certain degree of caution should be taken when recommending that one method should not be pursued in the future, since the findings may only apply to this modestly sized dataset. Although this may be the case only for this dataset.

There is a possibility that in this dataset, using more shape modes would enable a more sensitive diagnostic function to be created. The top five were taken in Chapter 3 as the smaller sample sizes, created when only the non-hypertensive subjects were analysed, precluded the inclusion of a large number of input parameters, owing to the probability of overfitting, and this method was repeated to produce the baseline classifier. Nonetheless, utilising a smaller number of modes in the baseline classifier makes it easier to observe the impact of including the population-based characteristics.

Chapter 6

Conclusion and Evaluation

6.1 Conclusions

This thesis presents a detailed characterisation of shape variation in the right ventricle of the human heart using a non-invasive, anatomically consistent, statistical shape analysis procedure. Shape information was derived from manually segmented, short axis MRI scans. The aim was to determine any unique shape features associated with non-prognostic demographics, such as sex, age, obesity and nationality and to test new diagnostic models for PH which incorporate such biological characteristics.

The overall goal was to improve non-invasive diagnosis of PH by making use of the shape of the right ventricle along with additional biological characteristics. To achieve this, the introduction to this thesis set out three primary research objectives. The first was to evaluate the relationship between the shape of the RV and non-pathological features in order to determine its capability to discern the effect of disease on RV shape, separately from the effect of non-pathological features. The second was to assess the robustness of the shape analysis approach to determine the constraints on the analysis process necessary to produce consistent shape features and related characterisation. Lastly, classification strategies were explored to provide guidance on how to use RV shapes, along with biological characteristics to predict

the state of PH. In this thesis Chapter 1 presented the introduction and initial literature review and established the scope of the thesis. Chapter 2 detailed the overall statistical shape analysis procedure used to produce the shape features analysed and compared throughout the remainder of the thesis. Chapter 3 isolated and characterised specific shape features associated with the three biological characteristics noted above, Chapter 4 assessed the robustness of the statistical shape method. Finally, Chapter 5 combined the results from the natural characteristic analysis to determine the optimum diagnostic protocol achievable using the data available to diagnose PH. Chapters 3, 4 and 5 all contain novel material directly or indirectly associated with determining a metric that can optimally diagnose PH in this dataset.

Apart from the three primary research objectives, several other methodological targets have been achieved. The work described in this thesis combines two cardiac datasets, whereas previous applications of the same method used either a small dataset to demonstrate proof of concept (Wu et al., 2012), or only took patients from the same study centre (Xu et al., 2021). The shape analysis procedure has again shown capacity to analyse the RV in an anatomically consistent way, irrespective of the image acquisition methodology. It has successfully decomposed hearts across a spectrum of variations in ventricle size, gender amongst other factors. The applicability of the PH diagnostics across diverse ventricular shapes confirms that the method is equally applicable across a broad spectrum of hearts from disparate origins.

The first research objective was assessed in Chapter 3, and evaluated whether RV shape features based on characteristics, such as sex, age, BMI and demography, impact the shape of the RV and typified specific shape changes associated with variation. Initial designation of the relation of shape features to population characteristics used Pearson coefficients to determine potential correspondences.

Inspection of the Pearson coefficients before LDA led to some interesting observations. Shape information associated with the MRI scanner and scanning protocol proved to have the strongest signal isolated using the statistical shape analysis procedure. The impact of scanner was observed through quantifying the correlation between modal coefficients (i.e. the magni-

tude of the shape feature) and membership of either the Newcastle or the Pittsburgh dataset. Crucially, this shape information was constrained within a single shape mode (Mode 1, or the best overall descriptor of the dataset), which describes elongation/shortening of the ventricle and an increase or decrease in detail at the apex. Therefore, one of the significant advantages of this statistical shape analysis procedure is that method-related features can be isolated and removed, leaving only shape information directly related to the comparison feature of interest.

After ascertaining that the relationship was often more complex and multi-faceted than could be described using a simple linear descriptor (an LDA), the relationship between non-pathological characteristics was evaluated using an inverse problem context; ie. taking the premise that two population characteristics can be linearly separated using shape feature combinations as evidence that a material difference in overall shape between those individuals exists. The LDA analyses found fundamental differences in RV shape between all dichotomous populations tested: male and female RVs, younger and older hearts (above/below mean age 60.53) and between those with lower and higher BMIs (above/below 28 kg/m^3). Although all 96 subjects were included in the initial decomposition, only the modal coefficients of the non-hypertensive RVs were included in the LDA analyses to ensure that only shape features inherent to each of the biological groups were extracted, without concern about the interaction of pathology. These features were all separated successfully by an LDA function when defined by combinations of the top five shape modes. The maximum separation accuracy scores ranged between 72.5% and 77.5% in all cases.

Of the biological features considered, sex had the greatest effect on RV shape at end-Systole, and obesity had the greatest effect on RV shape at end-Diastole. In both cases, 77.5% of non-hypertensive hearts could be categorised correctly using the shape modes of the RV. The value of Mode 19 for the end-Systole RV shape was the best separator of sex; the qualitative appearance of this mode shows female individuals exhibiting a substantial cut-out or notch at the base of the ventricle, whereas male individuals display a bulge or point in the same region, leading to an overall triangular appearance in the RV. The RVs of obese patients

are rounded and hypertrophic at end-Diastole, and exhibit free-wall expansion. This shape was best described by a combination of Modes 20, 11 and 2. Conversely, the non-obese individuals were characterised by increased septal wall curvature. These shape features were confirmed by plotting the mean obese RV and the mean non-obese RVs, which account for both end-Systolic and end-Diastolic shapes simultaneously.

Finally, when the non-hypertensive RVs were divided into groups based on age, three different combinations of shape modes at end-Diastole; Modes 6, 3, & 13; 6, 3 & 2 and 6, 13 & 17 had an equal effect on the shape of the RV. These modes each separated the older patients from the younger patients with 75.0% accuracy. There were several distinguishing shape features between the two age groups. Older patients showed a flattening of the septal wall, which was especially evident at end-Systole. Older patients showed a much more substantial lateral lengthening; where greater RV hypertrophy was observed nearer the base of the RV, towards the outflow tract and pulmonary artery. Younger patients exhibited greater detail at the RV apex at both end-Systole and end-Diastole in addition to greater septal-wall curvature.

Having determined several distinguishing characteristics in each of the patient groups, the same separating function was then applied to a group of hypertensive individuals to help determine whether these features could still be observed in a diseased cohort. When applying the same discriminant cut-off to hypertensive individuals, none of the analyses could separate the groups with an accuracy greater than 54.5%. The poor separation potential implies that the overarching shape changes associated with hypertension mask the specific shape features identified in non-hypertensive individuals. Again, obesity at end-Diastole, as characterised by Modes 20, 11 and 2, stands out as the best set of separating features. This is the case both when the same model, trained on the non-hypertensive individuals is applied to the remaining hypertensive RVs, and when a new model, specific to the hypertensive RVs is created using the same modes (this case separated 61.6% of RVs).

Chapter 4 varied the input data in four controlled ways; altering the size of the dataset, moving the anatomical reference points, substituting the individual performing the initial MRI

segmentation and altering the smoothing filter. It then examined how those input variations affected the decomposed shape information. This chapter examined how sensitive the modes (shape features) were to external modifications, and also examined the sensitivity of the mode coefficients (the ‘amount’) of each shape feature in each RVES. While there was some significant variation in juxtaposing direct coefficient values, Pearson correlations were proven to be highly consistent in cases where close qualitative alignment was observed and therefore proved the clearest metric to determine robustness; in comparisons of tests with high Pearson correlations, there is confidence that the shape analyses produce the same results and are insensitive to variation. In cases where the absolute Pearson correlation exceeded 0.9, examination of the qualitative models suggests that the two modes can be seen to encode the same shape features. The 3D shape models indicate that the modes are encoding the same shape (of the examples plotted Modes 1, 6 and 9 encoded the same features in almost all cases), and the Pearson correlations give confidence that the patients are experiencing *relatively* similar amounts of that feature. A combination of Pearson correlations and qualitative analysis was therefore used to quantify robustness.

All four tested interventions produced the same overall mode shapes, as evidenced through qualitative reconstructions and through the small sum of square difference between key eigenvectors. The dataset size reduction to just 20 patients demonstrated that modes can be coerced to encode different features using a limited number of RV shapes from this dataset, thus lending credibility to the analysis.

On the other hand, the mode coefficients were more sensitive to alterations in input. For example, changing the size of the dataset by either including patients (to increase the total dataset size to 121 patients) or excluding them, reducing the dataset size to 80, and then 60 patients did not substantially impact the shape modes over the first 16 Modes. This intervention was the most consistent of all attempts applied. Tests of both inter- and intra-observer variability were insensitive in terms of change in modal coefficient throughout the first 16 modes in all cases with Pearson correlations exceeding $r_{94}=0.9$, and all correlations

exceeded $r=0.765$. Tests which replaced the dateline, north & south poles; thereby altering their placement on the harmonic maps was more sensitive than the previous two described results. Alterations to the dateline did not impact the mode shapes over the first 14 modes. Previous work by Wu (2013) altered anatomical dateline replacement in a more systematic and substantial manner and also determined an impact on the POD results. Although inter- and intra-observer variation was more sizeable than expected and greater than other similar studies in the literature (e.g. Hudsmith et al., 2005; Caudron et al., 2012), the resultant impact of the mode shapes, and their associated coefficients was smaller. This was surprising given both the changes to the underlying segmentation, and the fact that such a change also necessitated selecting dateline points. The modal coefficient values were very sensitive to alterations in smoothing filter; only the correlation coefficients between Modes 6 and 7 were statistically significant, and the correlation values did not exceed $r_{94} = 0.4$. A possible explanation for this relates to the decomposition picking up steps or ridges in the RVs, related to the initial segmentation in ITK-SNAP, and that they are experiencing relatively higher amounts (coefficients) of these modes, compared to the original analysis.

It was assumed that interventions which involve making changes to the underlying segmentation and the FE mesh on which the harmonic mapping is performed, would have a larger impact on replicability than those which make smaller changes, such as moving the dateline or mutating the size of the overall input set. In practice, this was largely true, except that the mode coefficients were less sensitive than anticipated to changes in image segmentation. It is likely that some of the differing mode features identified during the robustness tests are related to aspects of the method. As was determined by the ability to detect the MRI scanner in Chapter 3, known procedural differences can be isolated in an anatomically meaningful way. Therefore the sensitivity of the shape analysis to smoothing features and inter/intra-observer variations could be sequestered and removed as an artefact of the process.

Based on the results from Chapter 4, several practical recommendations can be made. While various different scanning protocols can be combined and accounted for successfully,

the consistency of the remaining methodological parameters such as smoothing and segmentation must be maintained to prevent the risk of dissonant results. In the future, once this method is applied to greater sample sizes and anatomical shape features relevant to both pathological and non-pathological RV variations are confirmed, for example, further methodology-based shape modes may be identified enabling relaxation of the requirement for the same segmenter and boundary conditions. Furthermore, to address the challenges of wider clinical implementation, if the methodological parameters must be tightly controlled, it would be possible to incorporate a degree of automation to the pipeline. Automation of processes sensitive to run-by-run variation would undoubtedly achieve higher consistency. However, risks also exist in relinquishing manual control; especially in the case of segmentation, losing crucial shape information related to the significant cross-plane motion observed at end-Systole. Nevertheless, as automation of RV segmentation continues to improve at the current rate, as evidenced in recent papers by Zuluaga et al. (2020) and Tran et al. (2020), it seems that a computational selection might come to produce a more accurate and consistent picture of the RV, which would focus the derived shape features on anatomical differences rather than methodological ones.

Finally, Chapter 5 proposed and evaluated approaches to use shape features to predict PH. Four shape modes, 15, 10, 20 and 21 produced the best separation of PH. The same rounding out and flattening of the base RV was observed in this dataset, matching observations from Xu et al. (2021) and Mauger et al. (2019). This result was obtained even with the differences in the dataset. This dataset is composed of RVs from two different centres which employed different MRI scanning parameters. The inclusion of healthy ageing and generalised hypertensive patients dilutes the number of PH patients in the cohort. Despite these sources of statistical diversion in the dataset, it was still possible to separate PH.

Mode 15, in particular, stood out, both in its ability to classify patients with PH, and due to its physical appearance, which corresponded with qualitative observations from the literature (Mauger et al., 2019; Xu et al., 2021). This shape feature was identified in Chapter 3 as being one of the strongest classifiers of BMI in non-hypertensive RVs. Therefore, it is likely that

this shape feature has a pivotal role both in encoding shape differences associated with patient obesity and in those shape features which separate the RVs of PH patients. Recall that the Modes are ranked in order of importance to the overall shape of the RV across the 96 subjects at 2 phases. A mode can be important or indicative of a PH-related shape feature, even if it only captures the 15th largest amount of modal energy.

Sex, BMI and age-based information were then applied to a disease diagnostic model to determine whether the diagnosis of PH could be improved by incorporating the differences. Including an individual's BMI as a predictive variable in addition to the modal coefficients was the only modification which produced significant improvement in PH detection. However, although BMI improved the prediction scores, this improvement was concentrated on non-hypertensive individuals, and often the PH patient prediction was worsened when BMI was included. This is consistent with standard expectations since the inclusion of obesity resulted in the softening of the boundary between generalised hypertensive patients and PH patients; in general being overweight leads to an increased risk of developing cardiopulmonary illness and also an increased likelihood of having a larger, flatter heart.

Both LDA and non-linear analysis in the form of SVMs were used to isolate PH features. The non-linear analysis did not provide any benefit over the linear discriminant models; the best-achieved separation in any SVM test was 72.6%, which was very slightly worse than the best combination of the 96 RVs at end-Diastole using an LDA. The implication of this finding is that it is not worth the additional set-up and computational time required to run a non-linear analysis, at least on a dataset of this size. However, a linear analysis separates meaningful, anatomically consistent PH descriptors without requirement for non-linear analysis. Perhaps, in a substantially larger cohort, application of machine learning techniques might identify previously unseen diagnostic relationships.

This thesis has considered both end-Systolic and end-Diastolic shape information, to ensure that no phase specific shape features were missed. Especially since evidence of impaired contractile function can present differently when the RV is fully contracted, compared to when

if is fully relaxed, and this could manifest itself as a specific shape feature in one or other of the phases. Nevertheless, with the exception of Mode 19 being the strongest discriminator of sex at end-Systole, there seems to be no uniquely derived shape information that is attributable solely to end-Systole, and therefore, for the sake of simplicity of data analysis this phase could be omitted in future analyses.

6.2 Recommendations for Future Work

The research performed in the development of this thesis could be extended in the following directions:

- Extending the sample size to include a much larger set of scanners and scanning protocols. A larger dataset more accurately approximates the mean shape of the heart; anatomical consistency and closeness to the ‘normal’ RV shape are likely to give both more stable results and produce better descriptive features. The features associated with scanner type were very easily separated in the population-based study in Chapter 3, but this may not always be the case when considering a more comprehensive array of scanning parameters. As an example, the power of the UK Biobank (Sudlow et al., 2015), or the Icelandic Heart Association AGES study (Harris et al., 2007) databases could be used as a source of readily available cardiac MRI data. The UK Biobank, in particular is running a CMR study, due to be completed in 2023, which will make 20,000 CMR scans readily available for research purposes (for a fee). Manual segmentation of all four cardiac chambers has been completed for the first 5000 scans (Raisi-Estabragh, Harvey, et al., 2020). Furthermore, the Newcastle University study comprises 150 RVs in total, of which only 71 have been segmented for use in this thesis (46 in the initial decomposition, and a further 25 in Chapter 4). These could also be accessed to further increase the size of the dataset. It is important to note that the Newcastle MRIs were not focused on the RV, and that the initial 46 RVs were selected on the basis of clarity

and completeness of the RV in the scan.

- It was noted in Chapter 4 that owing to the nature of the MRI scan collection in this thesis, it was impossible to truly compare the robustness to scanner and scanning protocol. Performing studies using the same patient in different MRI scanners would provide valuable insight into the repeatability of the shape analysis procedure in different environments. Findings from this study may mitigate the need for a uniform scanning protocol recommendation if it is found that the precise scanning environment makes less of a difference than found in this thesis.
- The development of a specific scanning protocol that sets a specific number of slices, which is sensitive both to the requirement of maximising the shape information collected, would be beneficial, especially in the apex region, while remaining sensitive to the difficulties with maintaining a breath hold, experienced by patients with debilitating cardiopulmonary symptoms.
- The analysis is currently performed using a series of software packages, both commercial and open-source, including proprietary software from the MRI scanner, Simpleware, ABAQUS and FEniCS. A significant amount of processing time and effort is devoted to data reformatting to allow for transition between programs and operating systems, and furthermore, much of this software is niche and expensive. While some headway has been made towards reducing the number of separate stages involved in the analysis, such as the creation of ADAMCUS and the inclusion of various helper scripts. More work is still needed to create a seamless and reproducible pipeline that could subsequently be incorporated into a clinical setting. Performing the harmonic mapping and statistical decomposition in MATLAB or an open-source alternative would also alleviate the number of separate steps.
- The binary categories selected for evaluation shape analysis features form artificial ‘hard’ boundaries (such as the mean age of 60.5 years), which may prevent correct

classification in borderline individuals, and therefore obscure the identification of shape features possessed by some of the RVs near the cut off. The impact of altering these cut offs could be explored in more detail in an additional test. Alternatively, if given access to a larger dataset, a larger number of categories could be tested, which would improve the resolution of the study. For instance, a particular mode might feature more prominently the over 70s.

- As was determined by the ability to detect MRI scanner in Chapter 3, known procedural differences can be isolated using the statistical shape analysis method in an anatomically meaningful way. Therefore in future tests, the sensitivity of the shape analysis to smoothing features and inter/intra-observer variations could be sequestered and removed as an artefact of the analysis. The same correction could be applied to as yet unknown method-related signatures. Alternatively, such differences could be measured and calibrated, providing opportunities to train both humans and equipment.
- Recently published advancements in automated shape segmentation (e.g Zuluaga et al., 2020; Tran et al., 2020) could dramatically improve the throughput and speed of the process, and this could facilitate the development of a set of standard reference segmentations.
- Performing image perturbations (or ‘data augmentation’), for example, by artificially inducing shape features would allow a relatively easy way of increasing the sample size and honing in on shape features associated with disease. Potential algorithms could be as described in Shorten and Khoshgoftaar (2019). Data augmentation could be performed on the sample as a whole, or solely on subjects with PH, if there was a desire to increase the number of samples in a diseased state. This would be most useful in the PH and robustness-related chapters, as they are more confirmatory in nature than exploratory. This could be employed in the same way as Amodeo et al. (2021), who used data augmentation techniques in SVM analysis to cope with a small sample size.

The techniques employed in their analysis included adding modified copies of existing data to the dataset to strategically increase the number of hearts available, which will be advantageous to the training of predictive models. Incidentally, it is important to mention that Amodeo et al. (2021) employed a manual segmentation technique.

- There is scope for further robustness analysis checks now that a comparison framework has been identified. For example, the impact of patient positioning in the scanner on RV shape could be assessed. This represents yet another methodology-based source of variation that this statistical shape analysis method has proven adept at identifying.
- Long-axis images of the segmentations were not available for many of the scans in the cohort. The quality, accuracy and consistency of the manual image segmentations would be improved by performing manual image segmentations with access to both orientations simultaneously.
- Incorporating information from local shape calculations, such as sphericity, eccentricity or curl, could amplify useful shape information on a per-slice basis. This information, in conjunction with the overarching global information provided by the shape analysis presented in this thesis, could crystallise our understanding of specific shape features.
- With slight modification, the shape analysis procedure could be applied to many other organs, especially those where it is anticipated that there would be a pathological or commonly observed shape change. There are several examples in the literature (e.g. Lu et al., 2013) who use statistical shape analysis to characterise shape change in the livers of car crash victims.

References

- Abman, S. H., Hansmann, G., Archer, S. L., Ivy, D. D., Adatia, I., Chung, W. K., Hanna, B. D., Rosenzweig, E. B., Raj, J. U., Cornfield, D., Stenmark, K. R., Steinhorn, R., Thébaud, B., Fineman, J. R., Kuehne, T., Feinstein, J. A., Friedberg, M. K., Earing, M., Barst, R. J., Keller, R. L., Kinsella, J. P., Mullen, M., Deterding, R., Kulik, T., Mallory, G., Humpl, T., & Wessel, D. L. (2015, nov). Pediatric pulmonary hypertension. *Circulation*, *132*(21), 2037–2099. doi: 10.1161/cir.0000000000000329
- Addetia, K., Maffessanti, F., Muraru, D., Singh, A., Surkova, E., Mor-Avi, V., Badano, L. P., & Lang, R. M. (2018, may). Morphologic analysis of the normal right ventricle using three-dimensional echocardiography–derived curvature indices. *Journal of the American Society of Echocardiography*, *31*(5), 614–623. doi: 10.1016/j.echo.2017.12.009
- Addetia, K., Maffessanti, F., Yamat, M., Weinert, L., Narang, A., Freed, B. H., Mor-Avi, V., & Lang, R. M. (2016, jul). Three-dimensional echocardiography-based analysis of right ventricular shape in pulmonary arterial hypertension. *European Heart Journal – Cardiovascular Imaging*, *17*(5), 564–575. doi: 10.1093/ehjci/jev171
- Alfakih, K., Plein, S., Bloomer, T., Jones, T., Ridgway, J., & Sivananthan, M. (2003, jun). Comparison of right ventricular volume measurements between axial and short axis orientation using steady-state free precession magnetic resonance imaging. *Journal of Magnetic Resonance Imaging*, *18*(1), 25–32. doi: 10.1002/jmri.10329
- Ambellan, F., Lamecker, H., von Tycowicz, C., & Zachow, S. (2019). Statistical shape models: Understanding and mastering variation in anatomy. In *Advances in exper-*

- imental medicine and biology* (pp. 67–84). Springer International Publishing. doi: 10.1007/978-3-030-19385-0-5
- Amodeo, I., Nunzio, G. D., Raffaelli, G., Borzani, I., Griggio, A., Conte, L., Macchini, F., Condò, V., Persico, N., Fabietti, I., Ghirardello, S., Pierro, M., Tafuri, B., Como, G., Cascio, D., Colnaghi, M., Mosca, F., & Cavallaro, G. (2021, nov). A machine and deep learning approach to predict pulmonary hypertension in newborns with congenital diaphragmatic hernia (CLANNISH): Protocol for a retrospective study. *PLOS ONE*, *16*(11), e0259724. doi: 10.1371/journal.pone.0259724
- Avendi, M. R., Kheradvar, A., & Jafarkhani, H. (2017, feb). Automatic segmentation of the right ventricle from cardiac MRI using a learning-based approach. *Magnetic Resonance in Medicine*, *78*(6), 2439–2448. doi: 10.1002/mrm.26631
- Badano, L. P., Gingham, C., Easaw, J., Muraru, D., Grillo, M. T., Lancellotti, P., Pinamonti, B., Coghlan, G., Marra, M. P., Popescu, B. A., & Vita, S. D. (2009, oct). Right ventricle in pulmonary arterial hypertension: haemodynamics, structural changes, imaging, and proposal of a study protocol aimed to assess remodelling and treatment effects. *European Journal of Echocardiography*, *11*(1), 27–37. doi: 10.1093/ejehocardiography/jep152
- Badrick, T. (2021, jan). Biological variation: Understanding why it is so important? *Practical Laboratory Medicine*, *23*, e00199. doi: 10.1016/j.plabm.2020.e00199
- Bae, K., Shim, H., Tao, C., Chang, S., Wang, J., Boudreau, R., & Kwok, C. (2009, dec). Intra- and inter-observer reproducibility of volume measurement of knee cartilage segmented from the OAI MR image set using a novel semi-automated segmentation method. *Osteoarthritis and Cartilage*, *17*(12), 1589–1597. doi: 10.1016/j.joca.2009.06.003
- Benza, R. L., Miller, D. P., Gomberg-Maitland, M., Frantz, R. P., Foreman, A. J., Coffey, C. S., Frost, A., Barst, R. J., Badesch, D. B., Elliott, C. G., Liou, T. G., & McGoon, M. D. (2010, jul). Predicting survival in pulmonary arterial hypertension. *Circulation*, *122*(2), 164–172. doi: 10.1161/circulationaha.109.898122
- Bergstrom, P., & Edlund, O. (2014, feb). Robust registration of point sets using iteratively

- reweighted least squares. *Computational Optimization and Applications*, 58(3), 543–561. doi: 10.1007/s10589-014-9643-2
- Bernardino, G., Hodzic, A., Langet, H., Legallois, D., Craene, M. D., Ballester, M. Á. G., Saloux, É., & Bijmens, B. (2021, jul). Volumetric parcellation of the cardiac right ventricle for regional geometric and functional assessment. *Medical Image Analysis*, 71, 102044. doi: 10.1016/j.media.2021.102044
- Berra, G., Noble, S., Soccac, P. M., Beghetti, M., & Lador, F. (2016, mar). Pulmonary hypertension in the elderly: a different disease? *Breathe*, 12(1), 43–49. doi: 10.1183/20734735.003416
- Bertrand, T., Bartlett-Esquilant, G., Fischer, K., & Friedrich, M. G. (2022, feb). Patient and physician preferences for non-invasive diagnostic cardiovascular imaging technologies: a discrete choice experiment. *Journal of Patient-Reported Outcomes*, 6(1). doi: 10.1186/s41687-022-00419-0
- Beyar, R., Dong, S. J., Smith, E. R., Belenkie, I., & Tyberg, J. V. (1993, dec). Ventricular interaction and septal deformation: a model compared with experimental data. *American Journal of Physiology-Heart and Circulatory Physiology*, 265(6), H2044–H2056. doi: 10.1152/ajpheart.1993.265.6.h2044
- Bidviene, J., Muraru, D., Maffessanti, F., Ereminiene, E., Kovács, A., Lakatos, B., Vaskelyte, J.-J., Zaliunas, R., Surkova, E., Parati, G., & Badano, L. P. (2021, jan). Regional shape, global function and mechanics in right ventricular volume and pressure overload conditions: a three-dimensional echocardiography study. *The International Journal of Cardiovascular Imaging*, 37(4), 1289–1299. doi: 10.1007/s10554-020-02117-8
- Bonnemains, L., Mandry, D., Marie, P.-Y., Micard, E., Chen, B., & Vuissoz, P.-A. (2011, aug). Assessment of right ventricle volumes and function by cardiac MRI: Quantification of the regional and global interobserver variability. *Magnetic Resonance in Medicine*, 67(6), 1740–1746. doi: 10.1002/mrm.23143
- Bookstein, F. (1989, jun). Principal warps: thin-plate splines and the decomposition of

- deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(6), 567–585. doi: 10.1109/34.24792
- Bowman, A. W., Katina, S., Smith, J., & Brown, D. (2015, jun). Anatomical curve identification. *Computational Statistics & Data Analysis*, *86*, 52–64. doi: 10.1016/j.csda.2014.12.007
- Bradlow, W. M., Gibbs, J. S. R., & Mohiaddin, R. H. (2012). Cardiovascular magnetic resonance in pulmonary hypertension. *Journal of Cardiovascular Magnetic Resonance*, *14*(1), 6. doi: 10.1186/1532-429x-14-6
- Brener, M. I., Burkhoff, D., & Sunagawa, K. (2020, mar). Effective arterial elastance in the pulmonary arterial circulation. *Circulation: Heart Failure*, *13*(3). doi: 10.1161/circheartfailure.119.006591
- Bruse, J. L. (2017). *Shape and function in congenital heart disease: a translational study using image, statistical and computational analyses* (Unpublished doctoral dissertation). University College London.
- Caudron, J., Fares, J., Lefebvre, V., Vivier, P.-H., Petitjean, C., & Dacher, J.-N. (2012, aug). Cardiac MRI assessment of right ventricular function in acquired heart disease. *Academic Radiology*, *19*(8), 991–1002. doi: 10.1016/j.acra.2012.03.022
- Cetin, I., Raisi-Estabragh, Z., Petersen, S. E., Napel, S., Piechnik, S. K., Neubauer, S., Ballester, M. A. G., Camara, O., & Lekadir, K. (2020, nov). Radiomics signatures of cardiovascular risk factors in cardiac MRI: Results from the UK biobank. *Frontiers in Cardiovascular Medicine*, *7*. doi: 10.3389/fcvm.2020.591368
- Chahal, H., McClelland, R. L., Tandri, H., Jain, A., Turkbey, E. B., Hundley, W. G., Barr, R. G., Kizer, J., Lima, J. A., Bluemke, D. A., & Kawut, S. M. (2012, feb). Obesity and right ventricular structure and function. *Chest*, *141*(2), 388–395. doi: 10.1378/chest.11-0172
- Chang, K. Y., Duval, S., Badesch, D. B., Bull, T. M., Chakinala, M. M., Marco, T. D., Frantz, R. P., Hemnes, A., Mathai, S. C., Rosenzweig, E. B., Ryan, J. J., Thenappan, T., Allen,

- R., Bartolome, S., Benza, R., Cadaret, L., Eggert, M., Elwing, J., Fineman, J., Foley, R., Ford, H. J., Hirsch, R., Grinnan, J., Ivy, D. D., Kawut, S., Kennedy, J., Klinger, J., Leary, P., Mazimba, S., Ramani, G., Raina, A., Runo, J., Swisher, J., Varghese, N., White, R. J., Williamson, T., Yung, D., Zamanian, R., & Zwicke, D. (2022, may). Mortality in pulmonary arterial hypertension in the modern era: Early insights from the pulmonary hypertension association registry. *Journal of the American Heart Association*, *11*(9). doi: 10.1161/jaha.121.024969
- Chatterjee, A. (2000). An introduction to the proper orthogonal decomposition. *Current Science*, *78*(7), 808-817.
- Chaturvedi, A., Whitnah, J., Maki, J. H., Baran, T., & Mitsumori, L. M. (2016, dec). Horizontal long axis imaging plane for evaluation of right ventricular function on cardiac magnetic resonance imaging. *Journal of Clinical Imaging Science*, *6*, 52. doi: 10.4103/2156-7514.197076
- Childs, H., Ma, L., Ma, M., Clarke, J., Cocker, M., Green, J., Strohm, O., & Friedrich, M. G. (2011, aug). Comparison of long and short axis quantification of left ventricular volume parameters by cardiovascular magnetic resonance, with ex-vivo validation. *Journal of Cardiovascular Magnetic Resonance*, *13*(1). doi: 10.1186/1532-429x-13-40
- Chung, A. K., Das, S. R., Leonard, D., Peshock, R. M., Kazi, F., Abdullah, S. M., Canham, R. M., Levine, B. D., & Drazner, M. H. (2006, mar). Women have higher left ventricular ejection fractions than men independent of differences in left ventricular volume. *Circulation*, *113*(12), 1597–1604. doi: 10.1161/circulationaha.105.574400
- Chung, M. K., Hartley, R., Dalton, K. M., & Davidson, R. J. (2008). Encoding cortical surface by spherical harmonics. *Statistica Sinica*, *18*, 1269-1291.
- Clarke, C. J., Gurka, M. J., Norton, P. T., Kramer, C. M., & Hoyer, A. W. (2012, jan). Assessment of the accuracy and reproducibility of RV volume measurements by CMR in congenital heart disease. *JACC: Cardiovascular Imaging*, *5*(1), 28–37. doi: 10.1016/j.jcmg.2011.05.007

- Cohn, J. N., Ferrari, R., & Sharpe, N. (2000, mar). Cardiac remodeling—concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling. *Journal of the American College of Cardiology*, *35*(3), 569–582. doi: 10.1016/s0735-1097(99)00630-0
- Couto, M., Souto, M., Martínez, A., Maceira, A., Vieira, C., Pumar, J. M., & Croisille, P. (2020, jan). Accuracy of right ventricular volume and function assessed with cardiovascular magnetic resonance: comparison with echocardiographic parameters. *Clinical Imaging*, *59*(1), 61–67. doi: 10.1016/j.clinimag.2019.10.002
- Crawford, L., Monod, A., Chen, A. X., Mukherjee, S., & Rabadán, R. (2019, oct). Predicting clinical outcomes in glioblastoma: An application of topological and functional data analysis. *Journal of the American Statistical Association*, *115*(531), 1139–1150. doi: 10.1080/01621459.2019.1671198
- da Fontoura Costa, L., & Cesar Jr., R. M. (2010). *Shape analysis and classification*. CRC Press. doi: 10.1201/9781420037555
- D'Alonzo, G. E., Barst, R. J., Ayres, S. M., Bergofsky, E. H., Brundage, B. H., Detre, K. M., Fishman, A. P., Goldring, R. M., Groves, B. M., & Kernis, J. T. (1991, September). Survival in patients with primary pulmonary hypertension. results from a national prospective registry. *Annals of internal medicine*, *115*, 343–349.
- D'Andrea, A., Gerche, A. L., Golia, E., Teske, A. J., Bossone, E., Russo, M. G., Calabrò, R., & Baggish, A. L. (2014, sep). Right heart structural and functional remodeling in athletes. *Echocardiography*, *32*, S11–S22. doi: 10.1111/echo.12226
- Danias, P. G., Tritos, N. A., Stuber, M., Kissinger, K. V., Salton, C. J., & Manning, W. J. (2003). Cardiac structure and function in the obese: A cardiovascular magnetic resonance imaging study. *Journal of Cardiovascular Magnetic Resonance*, *5*(3), 431–438. doi: 10.1081/jcmr-120022259
- Davies, R. (2002). *Learning shape: optimal models for analysing shape variability* (Unpublished doctoral dissertation). University of Manchester.

- Dawes, T. J. W., Cai, J., Quinlan, M., de Marvao, A., Ostrowski, P. J., Tokarczuk, P. F., Watson, G. M. J., Wharton, J., Howard, L. S. G. E., Gibbs, J. S. R., Cook, S. A., Wilkins, M. R., & O'Regan, D. P. (2018, aug). Fractal analysis of right ventricular trabeculae in pulmonary hypertension. *Radiology*, *288*(2), 386–395. doi: 10.1148/radiol.2018172821
- Dawes, T. J. W., de Marvao, A., Shi, W., Fletcher, T., Watson, G. M. J., Wharton, J., Rhodes, C. J., Howard, L. S. G. E., Gibbs, J. S. R., Rueckert, D., Cook, S. A., Wilkins, M. R., & O'Regan, D. P. (2017, may). Machine learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension: A cardiac MR imaging study. *Radiology*, *283*(2), 381–390. doi: 10.1148/radiol.2016161315
- Dias, C. A., Assad, R. S., Caneo, L. F., Abduch, M. C. D., Aiello, V. D., Dias, A. R., Marcial, M. B., & Oliveira, S. A. (2002, nov). Reversible pulmonary trunk banding. II. an experimental model for rapid pulmonary ventricular hypertrophy. *The Journal of Thoracic and Cardiovascular Surgery*, *124*(5), 999–1006. doi: 10.1067/mtc.2002.124234
- Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. Wiley.
- Dryden, I. L., & Mardia, K. V. (2016). *Statistical shape analysis: with applications in R* (Vol. 995). John Wiley & Sons.
- Duta, N. (2009, nov). A survey of biometric technology based on hand shape. *Pattern Recognition*, *42*(11), 2797–2806. doi: 10.1016/j.patcog.2009.02.007
- Dwivedi, A., & Axel, L. (2017, oct). Abnormal motion patterns of the interventricular septum. *JACC: Cardiovascular Imaging*, *10*(10), 1281–1284. doi: 10.1016/j.jcmg.2017.06.014
- Earley, S., & Resta, T. C. (2002, jul). Estradiol attenuates hypoxia-induced pulmonary endothelin-1 gene expression. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *283*(1), L86–L93. doi: 10.1152/ajplung.00476.2001
- Ericsson, A., & Karlsson, J. (2007, aug). Measures for benchmarking of automatic correspondence algorithms. *Journal of Mathematical Imaging and Vision*, *28*(3), 225–241.

doi: 10.1007/s10851-007-0018-5

- Eveleth, P. B. (1996). Physical status: The use and interpretation of anthropometry. report of a WHO expert committee. *American Journal of Human Biology*, 8(6), 786–787. doi: 10.1002/(sici)1520-6300(1996)8:6<786::aid-ajhb11>3.0.co;2-i
- Farrell, C., Balasubramanian, A., Hays, A. G., Hsu, S., Rowe, S., Zimmerman, S. L., Hassoun, P. M., Mathai, S. C., & Mukherjee, M. (2022, jan). A clinical approach to multimodality imaging in pulmonary hypertension. *Frontiers in Cardiovascular Medicine*, 8. doi: 10.3389/fcvm.2021.794706
- Fiechter. (2013, feb). Age-related normal structural and functional ventricular values in cardiac function assessed by magnetic resonance. *BMC Medical Imaging*, 13(1). doi: 10.1186/1471-2342-13-6
- Fisher, E. (1936). Linear discriminant analysis. *Statistics & Discrete Methods of Data Sciences*, 392, 1-5.
- Fowler, N. O. (1980). *Cardiac diagnosis and treatment*. Harper Row.
- Frangi, A., Rueckert, D., Schnabel, J., & Niessen, W. (2002, sep). Automatic construction of multiple-object three-dimensional statistical shape models: application to cardiac modeling. *IEEE Transactions on Medical Imaging*, 21(9), 1151–1166. doi: 10.1109/tmi.2002.804426
- Frank, R. C., Min, J., Abdelghany, M., Paniagua, S., Bhattacharya, R., Bhambhani, V., Pomerantsev, E., & Ho, J. E. (2020, mar). Obesity is associated with pulmonary hypertension and modifies outcomes. *Journal of the American Heart Association*, 9(5). doi: 10.1161/jaha.119.014195
- Friedberg, M. K., & Redington, A. N. (2014, mar). Right versus left ventricular failure. *Circulation*, 129(9), 1033–1044. doi: 10.1161/circulationaha.113.001375
- Frost, A., Badesch, D., Gibbs, J. S. R., Gopalan, D., Khanna, D., Manes, A., Oudiz, R., Satoh, T., Torres, F., & Torbicki, A. (2019, jan). Diagnosis of Pulmonary Hypertension. *European Respiratory Journal*, 53(1), 1801904. doi: 10.1183/13993003.01904-2018

- Galiè, N., Hoepfer, M., Humbert, M., Torbicki, A., Vachiery, J.-L., Barbera, J. A., Beghetti, M., Corris, P., Gaine, S., Gibbs, J. S., Gomez-Sanchez, M. A., Jondeau, G., Klepetko, W., Opitz, C., Peacock, A., Rubin, L., Zellweger, M., Simonneau, G., Vahanian, A., Auricchio, A., Bax, J., Ceconi, C., Dean, V., Filippatos, G., Funck-Brentano, C., Hobbs, R., Kearney, P., McDonagh, T., McGregor, K., Popescu, B. A., Reiner, Z., Sechtem, U., Sirnes, P. A., Tendra, M., Vardas, P., Widimsky, P., Sechtem, U., Attar, N. A., Andreotti, F., Aschermann, M., Asteggiano, R., Benza, R., Berger, R., Bonnet, D., Delcroix, M., Howard, L., Kitsiou, A. N., Lang, I., Maggioni, A., Nielsen-Kudsk, J. E., Park, M., Perrone-Filardi, P., Price, S., Domenech, M. T. S., Vonk-Noordegraaf, A., Zamorano, J. L., & and. (2009, aug). Guidelines for the diagnosis and treatment of pulmonary hypertension: The task force for the diagnosis and treatment of pulmonary hypertension of the european society of cardiology (ESC) and the european respiratory society (ERS), endorsed by the international society of heart and lung transplantation (ISHLT). *European Heart Journal*, *30*(20), 2493–2537. doi: 10.1093/eurheartj/ehp297
- Galiè, N., Humbert, M., Vachiery, J.-L., Gibbs, S., Lang, I., Torbicki, A., Simonneau, G., Peacock, A., Noordegraaf, A. V., Beghetti, M., Ghofrani, A., Sanchez, M. A. G., Hansmann, G., Klepetko, W., Lancellotti, P., Matucci, M., McDonagh, T., Pierard, L. A., Trindade, P. T., Zompatori, M., & Hoepfer, M. (2015, aug). 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension. *European Respiratory Journal*, *46*(4), 903–975. doi: 10.1183/13993003.01032-2015
- Galiè, N., Manes, A., Negro, L., Palazzini, M., Bacchi-Reggiani, M. L., & Branzi, A. (2009, sep). A meta-analysis of randomized controlled trials in pulmonary arterial hypertension. *European Heart Journal*, *30*(4), 394–403. doi: 10.1093/eurheartj/ehp022
- Gao, Y., Riklin-Raviv, T., & Bouix, S. (2014, apr). Shape analysis, a field in need of careful validation. *Human Brain Mapping*, *35*(10), 4965–4978. doi: 10.1002/hbm.22525
- Gerig, G., Styner, M., Shenton, M. E., & Lieberman, J. A. (2001). Shape versus size: Im-

- proved understanding of the morphology of brain structures. In *Medical image computing and computer-assisted intervention – MICCAI 2001* (pp. 24–32). Springer Berlin Heidelberg. doi: 10.1007/3-540-45468-3-4
- Geva, T. (2014, jan). Is MRI the preferred method for evaluating right ventricular size and function in patients with congenital heart disease? *Circulation: Cardiovascular Imaging*, 7(1), 190–197. doi: 10.1161/circimaging.113.000553
- Gilbert, K., Forsch, N., Hegde, S., Mauger, C., Omens, J. H., Perry, J. C., Pontré, B., Suinesiaputra, A., Young, A. A., & McCulloch, A. D. (2018, jan). Atlas-based computational analysis of heart shape and function in congenital heart disease. *Journal of Cardiovascular Translational Research*, 11(2), 123–132. doi: 10.1007/s12265-017-9778-5
- Golland, P., Grimson, W., Shenton, M., & Kikinis, R. (2005, feb). Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis*, 9(1), 69–86. doi: 10.1016/j.media.2004.07.003
- Golland, P., Grimson, W. E. L., & Kikinis, R. (1999). Statistical shape analysis using fixed topology skeletons: Corpus callosum study. In *Lecture notes in computer science* (pp. 382–387). Springer Berlin Heidelberg. doi: 10.1007/3-540-48714-x-33
- Golland, P., Grimson, W. E. L., Shenton, M. E., & Kikinis, R. (2001). Deformation analysis for shape based classification. In *Lecture notes in computer science* (pp. 517–530). Springer Berlin Heidelberg. doi: 10.1007/3-540-45729-1_54
- Goparaju, A., Csecs, I., Morris, A., Kholmovski, E., Marrouche, N., Whitaker, R., & Elhabian, S. (2018). On the evaluation and validation of off-the-shelf statistical shape modeling tools: A clinical application. In *Shape in medical imaging* (pp. 14–27). Springer International Publishing. doi: 10.1007/978-3-030-04747-42
- Grosgeorge, D., Petitjean, C., Dacher, J.-N., & Ruan, S. (2013, September). Graph cut segmentation with a statistical shape model in cardiac mri. *Computer Vision and Image Understanding*, 117, 1027–1035. doi: 10.1016/j.cviu.2013.01.014
- Grünig, E., & Peacock, A. J. (2015, nov). Imaging the heart in pulmonary hypertension:

- an update. *European Respiratory Review*, 24(138), 653–664. doi: 10.1183/16000617.0058-2015
- Guensch, D. P., Fischer, K., Flewitt, J. A., Yu, J., Lukic, R., Friedrich, J. A., & Friedrich, M. G. (2013, sep). Breathing manoeuvre-dependent changes in myocardial oxygenation in healthy humans. *European Heart Journal - Cardiovascular Imaging*, 15(4), 409–414. doi: 10.1093/ehjci/jet171
- Haddad, F., Hunt, S. A., Rosenthal, D. N., & Murphy, D. J. (2008, mar). Right ventricular function in cardiovascular disease, part i: Anatomy, physiology, aging, and functional assessment of the right ventricle. *Circulation*, 117(11), 1436–1448. doi: 10.1161/circulationaha.107.653576
- Hambly, N., Alawfi, F., & Mehta, S. (2016, may). Pulmonary hypertension: diagnostic approach and optimal management. *Canadian Medical Association Journal*, 188(11), 804–812. doi: 10.1503/cmaj.151075
- Harris, T. B., Launer, L. J., Eiriksdottir, G., Kjartansson, O., Jonsson, P. V., Sigurdsson, G., Thorgeirsson, G., Aspelund, T., Garcia, M. E., Cotch, M. F., Hoffman, H. J., & and, V. G. (2007, jan). Age, gene/environment susceptibility-reykjavik study: Multidisciplinary applied phenomics. *American Journal of Epidemiology*, 165(9), 1076–1087. doi: 10.1093/aje/kwk115
- Haynes, H., & Holmes, W. (2013). Geomorphological methods. In B. G. Society (Ed.), (chap. The Emergence of Magnetic Resonance Imaging (MRI) for 3D Analysis of Sediment Beds).
- Hees, P. S., Fleg, J. L., Lakatta, E. G., & Shapiro, E. P. (2002, dec). Left ventricular remodeling with age in normal men versus women. *The American Journal of Cardiology*, 90(11), 1231–1236. doi: 10.1016/s0002-9149(02)02840-0
- Heimann, T., & Meinzer, H.-P. (2009, aug). Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13(4), 543–563. doi: 10.1016/j.media.2009.05.004

- Higgins, J. P. (2002). Nonlinear systems in medicine. *The Yale journal of biology and medicine*, 75, 247–260.
- Ho, S. Y. (2006, apr). Anatomy, echocardiography, and normal right ventricular dimensions. *Heart*, 92(suppl1), i2–i13. doi: 10.1136/hrt.2005.077875
- Holland, M. R., Gibson, A. A., Bauer, A. Q., Peterson, L. R., Schaffer, J. E., Bach, R. G., Cresci, S., & Miller, J. G. (2010, oct). Echocardiographic tissue characterization demonstrates differences in the left and right sides of the ventricular septum. *Ultrasound in Medicine & Biology*, 36(10), 1653–1661. doi: 10.1016/j.ultrasmedbio.2010.07.007
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2016). A practical guide to support vector classification. <http://www.csie.ntu.edu.tw/~cjlin>.
- Hudsmith, L., Petersen, S., Francis, J., Robson, M., & Neubauer, S. (2005, oct). Normal human left and right ventricular and left atrial dimensions using steady state free precession magnetic resonance imaging. *Journal of Cardiovascular Magnetic Resonance*, 7(5), 775–782. doi: 10.1080/10976640500295516
- Humbert, M., Sitbon, O., Chaouat, A., Bertocchi, M., Habib, G., Gressin, V., YaïlŁci, A., Weitzenblum, E., Cordier, J.-F., Chabot, F., Dromer, C., Pison, C., Reynaud-Gaubert, M., Haloun, A., Laurent, M., Hachulla, E., Cottin, V., Degano, B., JaiŁs, X., Montani, D., Souza, R., & Simonneau, G. (2010, jul). Survival in patients with idiopathic, familial, and anorexigen-associated pulmonary arterial hypertension in the modern management era. *Circulation*, 122(2), 156–163. doi: 10.1161/circulationaha.109.911818
- Hundley, W. G., Bluemke, D. A., Finn, J. P., Flamm, S. D., Fogel, M. A., Friedrich, M. G., Ho, V. B., Jerosch-Herold, M., Kramer, C. M., Manning, W. J., Patel, M., Pohost, G. M., Stillman, A. E., White, R. D., & Woodard, P. K. (2010, jun). ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance. *Circulation*, 121(22), 2462–2508. doi: 10.1161/cir.0b013e3181d44a8f
- Hyduk, A., Croft, J. B., Ayala, C., Zheng, K., Zheng, Z.-J., & Mensah, G. A. (2005, Novem-

- ber). Pulmonary hypertension surveillance—united states, 1980-2002. *Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002)*, 54, 1–28.
- Innelli, P., Esposito, R., Olibet, M., Nistri, S., & Galderisi, M. (2008, dec). The impact of ageing on right ventricular longitudinal function in healthy subjects: a pulsed tissue doppler study. *European Journal of Echocardiography*, 10(4), 491–498. doi: 10.1093/ejechocard/jen313
- Iranmanesh, A., & Rubin, G. D. (2019, nov). CT in cardiac applications. In *Computed tomography* (pp. 427–458). Springer International Publishing. doi: 10.1007/978-3-030-26957-9_23
- Jacobs, W., van de Veerdonk, M. C., Trip, P., de Man, F., Heymans, M. W., Marcus, J. T., Kawut, S. M., Bogaard, H.-J., Boonstra, A., & Noordegraaf, A. V. (2014, jun). The right ventricle explains sex differences in survival in idiopathic pulmonary arterial hypertension. *Chest*, 145(6), 1230–1236. doi: 10.1378/chest.13-1291
- Jang, J., Hwang, H.-J., Tschabrunn, C. M., Whitaker, J., Menze, B., Anter, E., & Nezafat, R. (2019, jun). Cardiovascular magnetic resonance-based three-dimensional structural modeling and heterogeneous tissue channel detection in ventricular arrhythmia. *Scientific Reports*, 9(1). doi: 10.1038/s41598-019-45586-1
- Jang, J., Ngo, L. H., Mancio, J., Kucukseymen, S., Rodriguez, J., Pierce, P., Goddu, B., & Nezafat, R. (2020, jun). Reproducibility of Segmentation-based Myocardial Radiomic Features with Cardiac MRI10.1038/s41598-018-36938-4. *Radiology: Cardiothoracic Imaging*, 2(3), e190216. doi: 10.1148/ryct.2020190216
- Jone, P.-N., Hinzman, J., Wagner, B. D., Ivy, D. D., & Younoszai, A. (2014, feb). Right ventricular to left ventricular diameter ratio at end-systole in evaluating outcomes in children with pulmonary hypertension. *Journal of the American Society of Echocardiography*, 27(2), 172–178. doi: 10.1016/j.echo.2013.10.014
- Joskowicz, L. (2018). Future perspectives on statistical shape models in computer-aided orthopedic surgery: Beyond statistical shape models and on to big data. In *Computer*

- assisted orthopaedic surgery for hip and knee* (pp. 199–206). Springer Singapore. doi: 10.1007/978-981-10-5245-317
- Kagan, A. (1952, jun). Dynamic responses of the right ventricle following extensive damage by cauterization. *Circulation*, 5(6), 816–823. doi: 10.1161/01.cir.5.6.816
- Kawut, S. M., Lima, J. A., Barr, R. G., Chahal, H., Jain, A., Tandri, H., Praestgaard, A., Bagiella, E., Kizer, J. R., Johnson, W. C., Kronmal, R. A., & Bluemke, D. A. (2011, jun). Sex and race differences in right ventricular structure and function. *Circulation*, 123(22), 2542–2551. doi: 10.1161/circulationaha.110.985515
- Kawut, S. M., Pinder, D., Al-Naamani, N., McCormick, A., Palevsky, H. I., Fritz, J., Smith, K. A., Mazurek, J. A., Doyle, M. F., MacLean, M. R., DeMichele, A., & Mankoff, D. A. (2019, nov). Fulvestrant for the treatment of pulmonary arterial hypertension. *Annals of the American Thoracic Society*, 16(11), 1456–1459. doi: 10.1513/annalsats.201904-328rl
- Keen, J., Prisco, S. Z., & Prins, K. W. (2021, jan). Sex differences in right ventricular dysfunction: Insights from the bench to bedside. *Frontiers in Physiology*, 11. doi: 10.3389/fphys.2020.623129
- Kiefer, L. S., Fabian, J., Lorbeer, R., Machann, J., Storz, C., Kraus, M. S., Wintermeyer, E., Schlett, C., Roemer, F., Nikolaou, K., Peters, A., & Bamberg, F. (2018, may). Inter- and intra-observer variability of an anatomical landmark-based, manual segmentation method by MRI for the assessment of skeletal muscle fat content and area in subjects from the general population. *The British Journal of Radiology*, 20180019. doi: 10.1259/bjr.20180019
- Kocaoglu, M., Pednekar, A. S., Wang, H., Alsaied, T., Taylor, M. D., & Rattan, M. S. (2020, jul). Breath-hold and free-breathing quantitative assessment of biventricular volume and function using compressed SENSE: a clinical validation in children and young adults. *Journal of Cardiovascular Magnetic Resonance*, 22(1). doi: 10.1186/s12968-020-00642-y

- Kohara, S., Tateyama, T., Foruzan, A., & Chen, Y.-W. (2010). Application of statistical shape model to diagnosis of liver disease. *2nd International Conference on Software Engineering and Data Mining (SEDM)*.
- Kovacs, G., Berghold, A., Scheidl, S., & Olschewski, H. (2009, mar). Pulmonary arterial pressure during rest and exercise in healthy subjects: a systematic review. *European Respiratory Journal*, *34*(4), 888–894. doi: 10.1183/09031936.00145608
- Kozic, N., Weber, S., Bãijchler, P., Lutz, C., Reimers, N., Ballester, M. Á. G., & Reyes, M. (2010, jun). Optimisation of orthopaedic implant design using statistical shape space analysis based on level sets. *Medical Image Analysis*, *14*(3), 265–275. doi: 10.1016/j.media.2010.02.008
- Kuch, B., Muscholl, M., Luchner, A., Dãűring, A., Riegger, G., Schunkert, H., & Hense, H. (1998, oct). Gender specific differences in left ventricular adaptation to obesity and hypertension. *Journal of Human Hypertension*, *12*(10), 685–691. doi: 10.1038/sj.jhh.1000689
- Kwong, R. Y., & Yucel, E. K. (2003, oct). Computed tomography scan and magnetic resonance imaging. *Circulation*, *108*(15). doi: 10.1161/01.cir.0000086899.32832.ec
- LaCombe, P., MATariq, & Lappin., S. (2022). *Physiology, afterload reduction*. [Internet]. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK493174/>
- Lahm, T., Douglas, I. S., Archer, S. L., Bogaard, H. J., Chesler, N. C., Haddad, F., Hennes, A. R., Kawut, S. M., Kline, J. A., Kolb, T. M., Mathai, S. C., Mercier, O., Michelakis, E. D., Naeije, R., Tuder, R. M., Ventetuolo, C. E., Vieillard-Baron, A., Voelkel, N. F., Vonk-Noordegraaf, A., & Hassoun, P. M. (2018, aug). Assessment of right ventricular function in the research setting: Knowledge gaps and pathways forward. an official american thoracic society research statement. *American Journal of Respiratory and Critical Care Medicine*, *198*(4), e15–e43. doi: 10.1164/rccm.201806-1160st
- Lakatos, B. K., Nabeshima, Y., Tokodi, M., Nagata, Y., Tõsér, Z., Otani, K., Kitano, T., Fãbián, A., Ujvári, A., Boros, A. M., Merkely, B., Kovács, A., & Takeuchi, M.

- (2020, aug). Importance of nonlongitudinal motion components in right ventricular function: Three-dimensional echocardiographic study in healthy volunteers. *Journal of the American Society of Echocardiography*, 33(8), 995–1005.e1. doi: 10.1016/j.echo.2020.04.002
- Lamata, P., Sinclair, M., Kerfoot, E., Lee, A., Crozier, A., Blazevic, B., Land, S., Lewandowski, A. J., Barber, D., Niederer, S., & Smith, N. (2014, feb). An automatic service for the personalization of ventricular cardiac meshes. *Journal of The Royal Society Interface*, 11(91), 20131023. doi: 10.1098/rsif.2013.1023
- Latus, H., Kuehne, T., Beerbaum, P., Aplitz, C., Hansmann, G., Muthurangu, V., & Moledina, S. (2016, apr). Cardiac MR and CT imaging in children with suspected or confirmed pulmonary hypertension/pulmonary hypertensive vascular disease. expert consensus statement on the diagnosis and treatment of paediatric pulmonary hypertension. the european paediatric pulmonary vascular disease network, endorsed by ISHLT and DGPK. *Heart*, 102(Suppl 2), ii30–ii35. doi: 10.1136/heartjnl-2015-308246
- Leach, A. (2019). Adamcus.
- Leary, P. J., Kurtz, C. E., Hough, C. L., Waiss, M.-P., Ralph, D. D., & Sheehan, F. H. (2012, jan). Three-dimensional analysis of right ventricular shape and function in pulmonary hypertension. *Pulmonary Circulation*, 2(1), 34–40. doi: 10.4103/2045-8932.94828
- Leha, A., Hellenkamp, K., Unsöld, B., Mushemi-Blake, S., Shah, A. M., Hasenfuß, G., & Seidler, T. (2019, oct). A machine learning approach for the prediction of pulmonary hypertension. *PLOS ONE*, 14(10), e0224453. doi: 10.1371/journal.pone.0224453
- Lin, E., & Alessio, A. (2009, nov). What are the basic concepts of temporal, contrast, and spatial resolution in cardiac CT? *Journal of Cardiovascular Computed Tomography*, 3(6), 403–408. doi: 10.1016/j.jcct.2009.07.003
- Logg, A., Mardal, K.-A., & Wells, G. (Eds.). (2012). *Automated solution of differential equations by the finite element method*. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-23099-8

- Logg, A., & Wells, G. N. (2010, apr). DOLFIN. *ACM Transactions on Mathematical Software*, 37(2), 1–28. doi: 10.1145/1731022.1731030
- LoMauro, A., & Aliverti, A. (2018, may). Sex differences in respiratory function. *Breathe*, 14(2), 131–140. doi: 10.1183/20734735.000318
- Lu, Y.-C., Kemper, A. R., Gayzik, S., Untaroiu, C. D., & Beillas, P. (2013, nov). Statistical modeling of human liver incorporating the variations in shape, size, and material properties. In *SAE technical paper series*. SAE International. doi: 10.4271/2013-22-0012
- Luo, D., Ding, C., & Huang, H. (2011, aug). Linear discriminant analysis: New formulations and overfit analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 25(1), 417–422. doi: 10.1609/aaai.v25i1.7926
- Luo, G., An, R., Wang, K., Dong, S., & Zhang, H. (2016, sep). A deep learning network for right ventricle segmentation in short-axis MRI. In *2016 computing in cardiology conference (CinC)*. Computing in Cardiology. doi: 10.22489/cinc.2016.139-406
- Mansi, T., Voigt, I., Leonardi, B., Pennec, X., Durrleman, S., Sermesant, M., Delingette, H., Taylor, A. M., Boudjemline, Y., Pongiglione, G., & Ayache, N. (2011, sep). A statistical model for quantification and prediction of cardiac remodelling: Application to tetralogy of fallot. *IEEE Transactions on Medical Imaging*, 30(9), 1605–1616. doi: 10.1109/tmi.2011.2135375
- Martinez, A., & Kak, A. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228–233. doi: 10.1109/34.908974
- Martínez-Sellés, M., Pérez-David, E., Yotti, R., Jiménez-Borreguero, J., Loughlin, G., Gallego, L., Ayesta, A., Olivera, M., Bermejo, J., & Fernández-Avilés, F. (2015, oct). Gender differences in right ventricular function in patients with non-ischaemic cardiomyopathy. *Netherlands Heart Journal*, 23(12), 578–584. doi: 10.1007/s12471-015-0753-y
- Mauger, C., Gilbert, K., Lee, A. M., Sanghvi, M. M., Aung, N., Fung, K., Carapella, V., Piechnik, S. K., Neubauer, S., Petersen, S. E., Suinesiaputra, A., & Young, A. A. (2019,

- jul). Right ventricular shape and function: cardiovascular magnetic resonance reference morphology and biventricular risk factor morphometrics in UK biobank. *Journal of Cardiovascular Magnetic Resonance*, 21(1). doi: 10.1186/s12968-019-0551-6
- Maxim, L. D., Niebo, R., & Utell, M. J. (2014, sep). Screening tests: a review with examples. *Inhalation Toxicology*, 26(13), 811–828. doi: 10.3109/08958378.2014.955932
- McLaughlin, V. V., Archer, S. L., Badesch, D. B., Barst, R. J., Farber, H. W., Lindner, J. R., Mathier, M. A., McGoon, M. D., Park, M. H., Rosenson, R. S., Rubin, L. J., Tapson, V. F., & Varga, J. (2009, apr). ACCF/AHA 2009 expert consensus document on pulmonary hypertension. *Circulation*, 119(16), 2250–2294. doi: 10.1161/circulationaha.109.192230
- McLeod, K., Mansi, T., Sermesant, M., Pongiglione, G., & Pennec, X. (2012, aug). Statistical shape analysis of surfaces in medical images applied to the tetralogy of fallot heart. In *Modeling in computational biology and biomedicine* (pp. 165–191). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-31208-3-5
- McLeod, K., T., M., Sermesant, M., Pongiglione, G., & Pennec, X. (2013). Statistical shape analysis of surfaces in medical images applied to the tetralogy of fallot heart. In (p. 165-191). Springer Berlin Heidelberg.
- Meda, S. A., Narayanan, B., Liu, J., Perrone-Bizzozero, N. I., Stevens, M. C., Calhoun, V. D., Glahn, D. C., Shen, L., Risacher, S. L., Saykin, A. J., & Pearlson, G. D. (2012, apr). A large scale multivariate parallel ICA method reveals novel imaging–genetic relationships for alzheimer's disease in the ADNI cohort. *NeuroImage*, 60(3), 1608–1621. doi: 10.1016/j.neuroimage.2011.12.076
- Méndez, C., Soler, R., Rodriguez, E., López, M., Álvarez, L., Fernández, N., & Montserrat, L. (2011, apr). Magnetic resonance imaging of abnormal ventricular septal motion in heart diseases: a pictorial review. *Insights into Imaging*, 2(4), 483–492. doi: 10.1007/s13244-011-0093-4
- Meyer, T. E., Shih, J. A., & Harrington, C. (2019). Acute heart failure and pulmonary edema.

- In *Cardiac intensive care* (pp. 180–198.e4). Elsevier. doi: 10.1016/b978-0-323-52993-8.00018-7
- Mihl, C., Dassen, W. R. M., & Kuipers, H. (2008, apr). Cardiac remodelling: concentric versus eccentric hypertrophy in strength and endurance athletes. *Netherlands Heart Journal*, 16(4), 129–133. doi: 10.1007/bf03086131
- Milnor, W. R. (1975, may). Arterial impedance as ventricular afterload. *Circulation Research*, 36(5), 565–570. doi: 10.1161/01.res.36.5.565
- Mitteroecker, P., Gunz, P., Windhager, S., & Schaefer, K. (2013). A brief review of shape, form, and allometry in geometric morphometrics, with applications to human facial morphology. *Hystrix, the Italian Journal of Mammalogy*, 24. doi: 10.4404/hystrix-24.1-6369
- Moolan-Feroze, O., Mirmehdi, M., & Hamilton, M. (2016, apr). Right ventricle segmentation using a 3d cylindrical shape model. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*. IEEE. doi: 10.1109/isbi.2016.7493207
- Morey, R. A., Petty, C. M., Xu, Y., Hayes, J. P., Wagner, H. R., Lewis, D. V., LaBar, K. S., Styner, M., & McCarthy, G. (2009, apr). A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *NeuroImage*, 45(3), 855–866. doi: 10.1016/j.neuroimage.2008.12.033
- Moriyama, H., Kawakami, T., Kataoka, M., Hiraide, T., Kimura, M., Endo, J., Kohno, T., Itabashi, Y., Seo, Y., Fukuda, K., & Murata, M. (2020, nov). Regional right ventricular abnormalities implicate distinct pathophysiological conditions in patients with chronic thromboembolic pulmonary hypertension. *Journal of the American Heart Association*, 9(21). doi: 10.1161/jaha.120.018096
- Muraleedharan, V. (2021). *What is linear discriminant analysis (lda)*. Online. Retrieved from <https://vivekmuraleedharan73.medium.com/what-is-linear-discriminant-analysis-lda-7e33ff59020a>
- Narula, S., Shameer, K., Salem Omar, A. M., Dudley, J. T., & Sengupta, P. P. (2016, Novem-

- ber). Machine-learning algorithms to automate morphological and functional assessments in 2d echocardiography. *Journal of the American College of Cardiology*, 68, 2287–2295. doi: 10.1016/j.jacc.2016.08.062
- Nguyen-Truong, M., Liu, W., Doherty, C., LeBar, K., Labus, K., Puttlitz, C., Easley, J., Monnet, E., Chicco, A., & Wang, Z. (2021, dec). The interventricular septum is biomechanically distinct from the ventricular free walls. *Bioengineering*, 8(12), 216. doi: 10.3390/bioengineering8120216
- NHS. (2020, January). *Pulmonary hypertension diagnosis*. Retrieved from <https://www.nhs.uk/conditions/pulmonary-hypertension/diagnosis/>
- NHS. (2021). *Ct scan*. Retrieved from <https://www.nhs.uk/conditions/ct-scan/>
- NHS. (2022a). *Echocardiogram*. Retrieved from <https://www.nhs.uk/conditions/echocardiogram/>
- NHS. (2022b). *Mri scan*. Retrieved from <https://www.nhs.uk/conditions/mri-scan/what-happens/>
- Nitzken, M. (2015). *Shape analysis of the human brain*. (Doctoral dissertation). doi: 10.18297/etd/2067
- Nochioka, K., Roca, G. Q., Claggett, B., Biering-Sørensen, T., Matsushita, K., Hung, C.-L., Solomon, S. D., Kitzman, D., & Shah, A. M. (2018, oct). Right ventricular function, right ventricular–pulmonary artery coupling, and heart failure risk in 4 US communities. *JAMA Cardiology*, 3(10), 939. doi: 10.1001/jamacardio.2018.2454
- Norton, J. M. (2001, mar). TOWARD CONSISTENT DEFINITIONS FOR PRELOAD AND AFTERLOAD. *Advances in Physiology Education*, 25(1), 53–61. doi: 10.1152/advances.2001.25.1.53
- Otto, C. M. (2013). *Textbook of clinical echocardiography expert consult - online and print*. Elsevier - Health Sciences Division.
- Otto, M. E., Belohlavek, M., Khandheria, B., Gilman, G., Svatikova, A., & Somers, V. (2004, jun). Comparison of right and left ventricular function in obese and nonobese men. *The*

- American Journal of Cardiology*, 93(12), 1569–1572. doi: 10.1016/j.amjcard.2004.02.073
- Pamboucas, C., & Nihoyannopoulos, P. (2006, jan). Cardiovascular magnetic resonance: Its role in the diagnosis and evaluation of pulmonary arterial hypertension. *Revista Española de Cardiología (English Edition)*, 59(8), 755–760. doi: 10.1016/s1885-5857(07)60039-5
- Parikh, J. D., Hollingsworth, K. G., Wallace, D., Blamire, A. M., & MacGowan, G. A. (2017, may). Left ventricular functional, structural and energetic effects of normal aging: Comparison with hypertension. *PLOS ONE*, 12(5), e0177404. doi: 10.1371/journal.pone.0177404
- Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S. E., & Frangi, A. F. (2016, jan). A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 29(2), 155–195. doi: {10.1007/s10334-015-0521-4}
- Petitjean, C., Zuluaga, M. A., Bai, W., Dacher, J.-N., Grosgeorge, D., Caudron, J., Ruan, S., Ayed, I. B., Cardoso, M. J., Chen, H.-C., Jimenez-Carretero, D., Ledesma-Carbayo, M. J., Davatzikos, C., Doshi, J., Erus, G., Maier, O. M., Nambakhsh, C. M., Ou, Y., Ourselin, S., Peng, C.-W., Peters, N. S., Peters, T. M., Rajchl, M., Rueckert, D., Santos, A., Shi, W., Wang, C.-W., Wang, H., & Yuan, J. (2015, jan). Right ventricle segmentation from cardiac MRI: A collation study. *Medical Image Analysis*, 19(1), 187–202. doi: 10.1016/j.media.2014.10.004
- Popović, Z. B., & Thomas, J. D. (2017, jun). Assessing observer variability: a user’s guide. *Cardiovascular Diagnosis and Therapy*, 7(3), 317–324. doi: 10.21037/cdt.2017.03.12
- Ptaszyńska-Kopczyńska, K., Krentowska, A., Sawicka, E., Skoneczny, A., Jasiewicz, M., Knapp, M., Musiał, W. J., Sobkowicz, B., & Kamiński, K. A. (2017, mar). The strengths and weaknesses of non-invasive parameters obtained by echocardiography and cardiopulmonary exercise testing in comparison with the hemodynamic assessment

- by the right heart catheterization in patients with pulmonary hypertension. *Advances in Medical Sciences*, 62(1), 39–44. doi: 10.1016/j.advms.2016.06.001
- Raisi-Estabragh, Z. (2020). *Architecture of the heart different between women and men and with age*. Online. Retrieved from <https://www.bhf.org.uk/what-we-do/news-from-the-bhf/news-archive/2020/august/esc-heart-shape-structure-men-women-qmul>
- Raisi-Estabragh, Z., Gkontra, P., Jaggi, A., Cooper, J., Augusto, J., Bhuva, A. N., Davies, R. H., Manisty, C. H., Moon, J. C., Munroe, P. B., Harvey, N. C., Lekadir, K., & Petersen, S. E. (2020, dec). Repeatability of cardiac magnetic resonance radiomics: A multi-centre multi-vendor test-retest study. *Frontiers in Cardiovascular Medicine*, 7. doi: 10.3389/fcvm.2020.586236
- Raisi-Estabragh, Z., Harvey, N. C., Neubauer, S., & Petersen, S. E. (2020, nov). Cardiovascular magnetic resonance imaging in the UK biobank: a major international health research resource. *European Heart Journal - Cardiovascular Imaging*, 22(3), 251–258. doi: 10.1093/ehjci/jeaa297
- Raisi-Estabragh, Z., Izquierdo, C., Campello, V. M., Martin-Isla, C., Jaggi, A., Harvey, N. C., Lekadir, K., & Petersen, S. E. (2020, mar). Cardiac magnetic resonance radiomics: basic principles and clinical perspectives. *European Heart Journal - Cardiovascular Imaging*, 21(4), 349–356. doi: 10.1093/ehjci/jeaa028
- Redington, A. N., Rigby, M. L., Shinebourne, E. A., & Oldershaw, P. J. (1990, jan). Changes in the pressure-volume relation of the right ventricle when its loading conditions are modified. *Heart*, 63(1), 45–49. doi: 10.1136/hrt.63.1.45
- Reisner, S. A., Azzam, Z., Halmann, M., Rinkevich, D., Sideman, S., Markiewicz, W., & Beyar, R. (1994, jan). Septal/free wall curvature ratio: A noninvasive index of pulmonary arterial pressure. *Journal of the American Society of Echocardiography*, 7(1), 27–35. doi: 10.1016/s0894-7317(14)80415-x
- Rich, J. D., & Rich, S. (2014, nov). Clinical diagnosis of pulmonary hypertension. *Circula-*

- tion, 130(20), 1820–1830. doi: 10.1161/circulationaha.114.006971
- Rich, S., Dantzker, D. R., Ayres, S. M., Bergofsky, E. H., Brundage, B. H., Detre, K. M., Fishman, A. P., Goldring, R. M., Groves, B. M., & Koerner, S. K. (1987, August). Primary pulmonary hypertension. a national prospective study. *Annals of internal medicine*, 107, 216–223.
- Rider, O. J., Lewis, A. J., Lewandowski, A. J., Ntusi, N., Nethononda, R., Petersen, S. E., Francis, J. M., Pitcher, A., Banerjee, R., Leeson, P., & Neubauer, S. (2015, jan). Obese subjects show sex-specific differences in right ventricular hypertrophy. *Circulation: Cardiovascular Imaging*, 8(1). doi: 10.1161/circimaging.114.002454
- Rinaldi, A. (2015, dec). Biometrics' new identity—measuring more physical and biological traits. *EMBO reports*, 17(1), 22–26. doi: 10.15252/embr.201541677
- Rodero, C., Stocchi, M., Marciniak, M., Longobardi, S., Whitaker, J., O'Neill, M. D., Gillette, K., Augustin, C., Plank, G., Vigmond, E. J., Lamata, P., & Niederer, S. A. (2021, apr). Linking statistical shape models and simulated function in the healthy adult human heart. *PLOS Computational Biology*, 17(4), e1008851. doi: 10.1371/journal.pcbi.1008851
- Rodriguez-Arias, J. J., & García-Álvarez, A. (2021, oct). Sex differences in pulmonary hypertension. *Frontiers in Aging*, 2. doi: 10.3389/fragi.2021.727558
- Rose, J. C., Cosimano, S. J., Hufnagel, C. A., & Massullo, E. A. (1955, nov). THE EFFECTS OF EXCLUSION OF THE RIGHT VENTRICLE FROM THE CIRCULATION IN DOGS 1. *Journal of Clinical Investigation*, 34(11), 1625–1631. doi: 10.1172/jci103215
- Salton, C. J., Chuang, M. L., O'Donnell, C. J., Kupka, M. J., Larson, M. G., Kissinger, K. V., Edelman, R. R., Levy, D., & Manning, W. J. (2002, mar). Gender differences and normal left ventricular anatomy in an adult population free of hypertension. *Journal of the American College of Cardiology*, 39(6), 1055–1060. doi: 10.1016/s0735-1097(02)01712-6

- Sanchez-Quintana, D., Anderson, R. H., & Ho, S. Y. (1996, sep). Ventricular myoarchitecture in tetralogy of fallot. *Heart*, *76*(3), 280–286. doi: 10.1136/hrt.76.3.280
- Sandham, J. D., Hull, R. D., Brant, R. F., Knox, L., Pineo, G. F., Doig, C. J., Laporta, D. P., Viner, S., Passerini, L., Devitt, H., Kirby, A., & Jacka, M. (2003, jan). A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. *New England Journal of Medicine*, *348*(1), 5–14. doi: 10.1056/nejmoa021108
- Sardanelli, F., Quarenghi, M., Leo, G. D., Boccaccini, L., & Schiavi, A. (2008). Segmentation of cardiac cine MR images of left and right ventricles: Interactive semiautomated methods and manual contouring by two readers with different education and experience. *Journal of Magnetic Resonance Imaging*, *27*(4), 785–792. doi: 10.1002/jmri.21292
- Scheffler, K., & Lehnhardt, S. (2003, aug). Principles and applications of balanced SSFP techniques. *European Radiology*, *13*(11), 2409–2418. doi: 10.1007/s00330-003-1957-x
- Schulz-Menger, J., Bluemke, D. A., Bremerich, J., Flamm, S. D., Fogel, M. A., Friedrich, M. G., Kim, R. J., von Knobelsdorff-Brenkenhoff, F., Kramer, C. M., Pennell, D. J., Plein, S., & Nagel, E. (2020, mar). Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update. *Journal of Cardiovascular Magnetic Resonance*, *22*(1). doi: 10.1186/s12968-020-00610-6
- Sciancalepore, M. A., Maffessanti, F., Patel, A. R., Gomberg-Maitland, M., Chandra, S., Freed, B. H., Caiani, E. G., Lang, R. M., & Mor-Avi, V. (2011, jun). Three-dimensional analysis of interventricular septal curvature from cardiac magnetic resonance images for the evaluation of patients with pulmonary hypertension. *The International Journal of Cardiovascular Imaging*, *28*(5), 1073–1085. doi: 10.1007/s10554-011-9913-3
- Shang, Z., Sun, J., Hui, J., Yu, Y., Bian, X., Yang, B., Deng, K., & Lin, L. (2021, nov). Construction of a support vector machine–based classifier for pulmonary arterial hypertension patients. *Frontiers in Genetics*, *12*. doi: 10.3389/fgene.2021.781011
- Sharma, M., Burns, A. T., Yap, K., & Prior, D. L. (2021, jun). The role of imaging in

- pulmonary hypertension. *Cardiovascular Diagnosis and Therapy*, *11*(3), 859–880. doi: 10.21037/cdt-20-295
- Sheehan, F., & Redington, A. (2008, jul). The right ventricle: anatomy, physiology and clinical imaging. *Heart*, *94*(11), 1510–1515. doi: 10.1136/hrt.2007.132779
- Shen, L., Ford, J., Makedon, F., & Saykin, A. (2003). A surface-based approach for classification of 3d neuroanatomic structures. *Dartmouth Computer Science Technical Report TR2003464*.
- Shi, Y., Lai, R., Kern, K., Sicotte, N., Dinov, I., & Toga, A. W. (2008). Harmonic surface mapping with laplace-beltrami eigenmaps. In *Medical image computing and computer-assisted intervention – MICCAI 2008* (pp. 147–154). Springer Berlin Heidelberg. doi: 10.1007/978-3-540-85990-118
- Shorten, C., & Khoshgoftaar, T. M. (2019, jul). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1). doi: 10.1186/s40537-019-0197-0
- Simon, M. A., Deible, C., Mathier, M. A., Lacomis, J., Goitein, O., Shroff, S. G., & Pinsky, M. R. (2009, aug). Phenotyping the right ventricle in patients with pulmonary hypertension. *Clinical and Translational Science*, *2*(4), 294–299. doi: 10.1111/j.1752-8062.2009.00134.x
- Simonneau, G., Galiè, N., Rubin, L. J., Langleben, D., Seeger, W., Domenighetti, G., Gibbs, S., Lebrec, D., Speich, R., Beghetti, M., Rich, S., & Fishman, A. (2004, jun). Clinical classification of pulmonary hypertension. *Journal of the American College of Cardiology*, *43*(12), S5–S12. doi: 10.1016/j.jacc.2004.02.037
- Simonneau, G., Gatzoulis, M. A., Adatia, I., Celermajer, D., Denton, C., Ghofrani, A., Sanchez, M. A. G., Kumar, R. K., Landzberg, M., Machado, R. F., Olschewski, H., Robbins, I. M., & Souza, R. (2013, dec). Updated clinical classification of pulmonary hypertension. *Journal of the American College of Cardiology*, *62*(25), D34–D41. doi: 10.1016/j.jacc.2013.10.029
- Simonneau, G., Montani, D., Celermajer, D. S., Denton, C. P., Gatzoulis, M. A., Krowka, M.,

- Williams, P. G., & Souza, R. (2019, jan). Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *European Respiratory Journal*, *53*(1), 1801913. doi: 10.1183/13993003.01913-2018
- Simonneau, G., Robbins, I. M., Beghetti, M., Channick, R. N., Delcroix, M., Denton, C. P., Elliott, C. G., Gaine, S. P., Gladwin, M. T., Jing, Z.-C., Krowka, M. J., Langleben, D., Nakanishi, N., & Souza, R. (2009, jun). Updated clinical classification of pulmonary hypertension. *Journal of the American College of Cardiology*, *54*(1), S43–S54. doi: 10.1016/j.jacc.2009.04.012
- Simundic, A.-M., Bartlett, W. A., & Fraser, C. G. (2015, jan). Biological variation: a still evolving facet of laboratory medicine. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, *52*(2), 189–190. doi: 10.1177/0004563214567478
- Sitbon, O., Benza, R. L., Badesch, D. B., Barst, R. J., Elliott, C. G., Gressin, V., Lemarié, J.-C., Miller, D. P., Rouzic, E. M.-L., Simonneau, G., Frost, A. E., Farber, H. W., Humbert, M., & McGoon, M. D. (2015, apr). Validation of two predictive models for survival in pulmonary arterial hypertension. *European Respiratory Journal*, *46*(1), 152–164. doi: 10.1183/09031936.00004414
- Sokmen, A., Sokmen, G., Acar, G., Akcay, A., Koroglu, S., Koleoglu, M., Yalcintas, S., & Aydin, M. N. (2013). The impact of isolated obesity on right ventricular function in young adults. *Arquivos Brasileiros de Cardiologia*. doi: 10.5935/abc.20130142
- Stanković, K., Huysmans, T., Danckaers, F., Sijbers, J., & Booth, B. G. (2020, aug). Subject-specific identification of three dimensional foot shape deviations using statistical shape analysis. *Expert Systems with Applications*, *151*, 113372. doi: 10.1016/j.eswa.2020.113372
- Starr, I., Jeffers, W. A., & Meade, R. H. (1943, sep). The absence of conspicuous increments of venous pressure after severe damage to the right ventricle of the dog, with a discussion of the relation between clinical congestive failure and heart disease. *American Heart Journal*, *26*(3), 291–301. doi: 10.1016/s0002-8703(43)90325-4

- Stegmann, M. B., & Gomez, D. D. (2002). A brief introduction to statistical shape analysis.
- Strait, J. B., & Lakatta, E. G. (2012, jan). Aging-associated cardiovascular changes and their relationship to heart failure. *Heart Failure Clinics*, 8(1), 143–164. doi: 10.1016/j.hfc.2011.08.011
- Styner, M., Gerig, G., Lieberman, J., Jones, D., & Weinberger, D. (2003, September). Statistical shape analysis of neuroanatomical structures based on medial models. *Medical image analysis*, 7, 207–220.
- Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Levitt, J. J., Shenton, M. E., & Gerig, G. (2006). Framework for the statistical shape analysis of brain structures using spharm-pdm. *The insight journal*, 242–250.
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015, mar). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine*, 12(3), e1001779. doi: 10.1371/journal.pmed.1001779
- Suinesiaputra, A., Dhooge, J., Duchateau, N., Ehrhardt, J., Frangi, A. F., Gooya, A., Grau, V., Lekadir, K., Lu, A., Mukhopadhyay, A., Oksuz, I., Ablin, P., Parajuli, N., Pennek, X., Pereanez, M., Pinto, C., Piras, P., Rohe, M.-M., Rueckert, D., Saring, D., Sermesant, M., Siddiqi, K., Alba, X., Tabassian, M., Teresi, L., Tsaftaris, S. A., Wilms, M., Young, A. A., Zhang, X., Medrano-Gracia, P., Alessandrini, M., Allen, J., Bai, W., Cimen, S., Claes, P., & Cowan, B. R. (2018, mar). Statistical shape modeling of the left ventricle: Myocardial infarct classification challenge. *IEEE Journal of Biomedical and Health Informatics*, 22(2), 503–515. doi: 10.1109/jbhi.2017.2652449
- Swift, A. J., Capener, D., Hammerton, C., Thomas, S. M., Elliot, C., Condliffe, R., Wild, J. M., & Kiely, D. G. (2015, may). Right ventricular sex differences in patients with idiopathic pulmonary arterial hypertension characterised by magnetic resonance imaging: Pair-matched case controlled study. *PLOS ONE*, 10(5), e0127415. doi: 10.1371/journal

.pone.0127415

- Swift, A. J., Wild, J. M., Nagle, S. K., Roldán-Alzate, A., François, C. J., Fain, S., Johnson, K., Capener, D., van Beek, E. J., Kiely, D. G., Wang, K., & Schiebler, M. L. (2014, mar). Quantitative magnetic resonance imaging of pulmonary hypertension. *Journal of Thoracic Imaging, 29*(2), 68–79. doi: 10.1097/rti.0000000000000079
- Taha, A. A., & Hanbury, A. (2015, aug). Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging, 15*(1). doi: 10.1186/s12880-015-0068-x
- Taha, A. A., Hanbury, A., & del Toro, O. A. J. (2014, oct). A formal method for selecting evaluation metrics for image segmentation. In *2014 IEEE international conference on image processing (ICIP)*. IEEE. doi: 10.1109/icip.2014.7025187
- Tan, C. M. J., & Lewandowski, A. J. (2019, sep). The transitional heart: From early embryonic and fetal development to neonatal life. *Fetal Diagnosis and Therapy, 47*(5), 373–386. doi: 10.1159/000501906
- Tandri, H., Daya, S. K., Nasir, K., Bomma, C., Lima, J. A., Calkins, H., & Bluemke, D. A. (2006, dec). Normal reference values for the adult right ventricle by magnetic resonance imaging. *The American Journal of Cardiology, 98*(12), 1660–1664. doi: 10.1016/j.amjcard.2006.07.049
- Tate, J., Zemezmi, N., Good, W., van Dam, P., Brooks, D., & MacLeod, R. (2020, dec). Shape analysis of segmentation variability. In *2020 computing in cardiology conference (CinC)*. Computing in Cardiology. doi: 10.22489/cinc.2020.466
- Tello, K., Richter, M. J., Yogeswaran, A., Ghofrani, H. A., Naeije, R., Vanderpool, R., Gall, H., Tedford, R. J., Seeger, W., & Lahm, T. (2020, oct). Sex differences in right ventricular–pulmonary arterial coupling in pulmonary arterial hypertension. *American Journal of Respiratory and Critical Care Medicine, 202*(7), 1042–1046. doi: 10.1164/rccm.202003-0807le
- Thompson, D. W. (1917). *On growth and form*. Cambridge University Press.

- Thompson, P. M., Hayashi, K. M., de Zubicaray, G. I., Janke, A. L., Rose, S. E., Semple, J., Hong, M. S., Herman, D. H., Gravano, D., Doddrell, D. M., & Toga, A. W. (2004, aug). Mapping hippocampal and ventricular change in alzheimer disease. *NeuroImage*, 22(4), 1754–1766. doi: 10.1016/j.neuroimage.2004.03.040
- Tischler, M. D., & Niggel, J. (1993, sep). Exercise echocardiography in combined mild mitral valve stenosis and regurgitation. *Echocardiography*, 10(5), 453–457. doi: 10.1111/j.1540-8175.1993.tb00058.x
- Tokodi, M., Staub, L., Budai, Á., Lakatos, B. K., Csákvári, M., Suhai, F. I., Szabó, L., Fábíán, A., Vágó, H., Tőser, Z., Merkely, B., & Kovács, A. (2021, mar). Partitioning the right ventricle into 15 segments and decomposing its motion using 3d echocardiography-based models: The updated ReVISION method. *Frontiers in Cardiovascular Medicine*, 8. doi: 10.3389/fcvm.2021.622118
- Torbicki, A. (2015, jun). Right ventricle in pulmonary hypertension: Echocardiography strikes back? *Circulation: Cardiovascular Imaging*, 8(6), e003518–e003518. doi: 10.1161/circimaging.115.003518
- Tran, C. T., Halicek, M., Dormer, J. D., Tandon, A., Hussain, T., & Fei, B. (2020, feb). Fully automated segmentation of the right ventricle in patients with repaired tetralogy of fallot using u-net. In B. S. Gimi & A. Krol (Eds.), *Medical imaging 2020: Biomedical applications in molecular, structural, and functional imaging*. SPIE. doi: 10.1117/12.2549052
- Treibel, T. A., Kelion, A., Ingram, T. E., Archbold, R. A., Myerson, S. G., Menezes, L. J., Morgan-Hughes, G. J., Schofield, R., Keenan, N. G., Clarke, S. C., Keys, A., Keogh, B., Masani, N., Ray, S., Westwood, M., Pearce, K., Colebourn, C. L., Bull, R. K., Greenwood, J. P., Roditi, G. H., & Lloyd, G. (2022). United kingdom standards for non-invasive cardiac imaging: recommendations from the imaging council of the british cardiovascular society. *Heart*. Retrieved from <https://heart.bmj.com/content/early/2022/05/24/heartjnl-2022-320799> doi: 10.1136/heartjnl-2022-320799

- Truong, U., Meinel, K., Haddad, F., Koestenberger, M., Carlsen, J., Ivy, D., & Jone, P.-N. (2020, oct). Update on noninvasive imaging of right ventricle dysfunction in pulmonary hypertension. *Cardiovascular Diagnosis and Therapy*, *10*(5), 1604–1624. doi: 10.21037/cdt-20-272
- Vachieri, J.-L., & Gaine, S. (2012, nov). Challenges in the diagnosis and treatment of pulmonary arterial hypertension. *European Respiratory Review*, *21*(126), 313–320. doi: 10.1183/09059180.00005412
- Vadakkan, C., & Siddiqui, W. (2022, January). Claustrophobia. *StatPearls*.
- van Timmeren, J. E., Leijenaar, R. T., van Elmpt, W., Wang, J., Zhang, Z., Dekker, A., & Lambin, P. (2016, dec). Test–retest data for radiomics feature stability analysis: Generalizable or study-specific? *Tomography*, *2*(4), 361–365. doi: 10.18383/j.tom.2016.00208
- van Wolferen, S. A., Marcus, J. T., Westerhof, N., Spreeuwenberg, M. D., Marques, K. M., Bronzwaer, J. G., Henkens, I. R., Gan, C. T.-J., Boonstra, A., Postmus, P. E., & Vonk-Noordegraaf, A. (2008, dec). Right coronary artery flow impairment in patients with pulmonary hypertension. *European Heart Journal*, *29*(1), 120–127. doi: 10.1093/eurheartj/ehm567
- Verwaerde, J., Laforet, J., Marque, C., & Rassineux, A. (2021, sep). Statistical shape analysis of gravid uteri throughout pregnancy by a ray description technique. *Medical & Biological Engineering & Computing*, *59*(10), 2165–2183. doi: 10.1007/s11517-021-02402-1
- Vincent, J.-L. (2008). Understanding cardiac output. *Critical Care*, *12*(4), 174. doi: 10.1186/cc6975
- Voelkel, N. F., Quaife, R. A., Leinwand, L. A., Barst, R. J., McGoon, M. D., Meldrum, D. R., Dupuis, J., Long, C. S., Rubin, L. J., Smart, F. W., Suzuki, Y. J., Gladwin, M., Denholm, E. M., & Gail, D. B. (2006, oct). Right ventricular function and failure: Report of a national heart, lung, and blood institute working group on cellular and molecular mechanisms of right heart failure. *Circulation*, *114*(17), 1883–1891. doi: 10.1161/circulationaha.106.632208

- Voelkl, B., Altman, N. S., Forsman, A., Forstmeier, W., Gurevitch, J., Jaric, I., Karp, N. A., Kas, M. J., Schielzeth, H., de Castele, T. V., & Wajrbel, H. (2020, jun). Reproducibility of animal research in light of biological variation. *Nature Reviews Neuroscience*, 21(7), 384–393. doi: 10.1038/s41583-020-0313-3
- Vonk-Noordegraaf, A., & Galie, N. (2011, nov). The role of the right ventricle in pulmonary arterial hypertension. *European Respiratory Review*, 20(122), 243–253. doi: 10.1183/09059180.00006511
- Vonk-Noordegraaf, A., Haddad, F., Chin, K. M., Forfia, P. R., Kawut, S. M., Lumens, J., Naeije, R., Newman, J., Oudiz, R. J., Provencher, S., Torbicki, A., Voelkel, N. F., & Hassoun, P. M. (2013, dec). Right heart adaptation to pulmonary arterial hypertension. *Journal of the American College of Cardiology*, 62(25), D22–D33. doi: 10.1016/j.jacc.2013.10.027
- Walker, L. A., & Buttrick, P. M. (2013, feb). The right ventricle: Biologic insights and response to disease: Updated. *Current Cardiology Reviews*, 9(1), 73–81. doi: 10.2174/157340313805076296
- Warfield, S., Zou, K., & Wells, W. (2004, jul). Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7), 903–921. doi: 10.1109/tmi.2004.828354
- Weiner, R. B., & Baggish, A. L. (2012, jun). Exercise-induced cardiac remodelling: the need for assessment of regional myocardial function. *The Journal of Physiology*, 590(12), 2829–2830. doi: 10.1113/jphysiol.2012.233585
- WHO. (2021). *Cardiovascular diseases (cvds)* (Tech. Rep.). World Health Organization. Retrieved from [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- WHO Commission on Social Determinants of Health. (2008). Closing the gap in a generation: Health equity through action on the social determinants of health. *Final Report of the Commission on Social Determinants of Health*.

- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York. publisher = Springer-Verlag New York,. Retrieved from <https://ggplot2.tidyverse.org>
- Wijeratne, D. T., Lajkosz, K., Brogly, S. B., Lougheed, M. D., Jiang, L., Housin, A., Barber, D., Johnson, A., Doliszny, K. M., & Archer, S. L. (2018, feb). Increasing incidence and prevalence of world health organization groups 1 to 4 pulmonary hypertension. *Circulation: Cardiovascular Quality and Outcomes*, *11*(2). doi: 10.1161/circoutcomes.117.003973
- Wong, C. Y., O'Moore-Sullivan, T., Leano, R., Hukins, C., Jenkins, C., & Marwick, T. H. (2006, feb). Association of subclinical right ventricular dysfunction with obesity. *Journal of the American College of Cardiology*, *47*(3), 611–616. doi: 10.1016/j.jacc.2005.11.015
- Wood, W. C. (2009, nov). Non-invasive anatomy and physiology: the new 'gold standard'? *Heart*, *96*(6), 411–412. doi: 10.1136/hrt.2009.183731
- Woods, C., Fernee, C., Browne, M., Zakrzewski, S., & Dickinson, A. (2017, dec). The potential of statistical shape modelling for geometric morphometric analysis of human teeth in archaeological research. *PLOS ONE*, *12*(12), e0186754. doi: 10.1371/journal.pone.0186754
- Wright, G. K., & Grey, M. W. (2016, March). *Computed tomography*. <https://radiologykey.com/computed-tomography-8/>. Retrieved from <https://radiologykey.com/computed-tomography-8/>
- Wu, J. (2013). *In vivo human right ventricle shape and kinematic analysis with and without pulmonary hypertension* (Unpublished doctoral dissertation). University of Pittsburgh.
- Wu, J., Simon, M. A., & Brigham, J. C. (2014, jul). A comparative analysis of global shape analysis methods for the assessment of the human right ventricle. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *4*(6), 327–343. doi: 10.1080/21681163.2014.941442

- Wu, J., Wang, Y., Simon, M. A., & Brigham, J. C. (2012, nov). A new approach to kinematic feature extraction from the human right ventricle for classification of hypertension: a feasibility study. *Physics in Medicine and Biology*, *57*(23), 7905–7922. doi: 10.1088/0031-9155/57/23/7905
- Wu, J., Wang, Y., Simon, M. A., Sacks, M. S., & Brigham, J. C. (2013, mar). A new computational framework for anatomically consistent 3d statistical shape analysis with clinical imaging applications. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *1*(1), 13–27. doi: 10.1080/21681163.2013.764610
- Xu, J., Desmond, E. L., Wong, T. C., Neill, C. G., Simon, M. A., & Brigham, J. C. (2021, nov). Right ventricular shape feature quantification for evaluation of pulmonary hypertension: Feasibility and preliminary associations with clinical outcome. *Journal of Biomechanical Engineering*, *144*(4). doi: 10.1115/1.4052495
- Yildirimturk, O., Tayyareci, Y., & Aytekin, S. (2011, may). The impact of body mass index on right ventricular systolic functions in normal and mildly obese healthy patients: A velocity vector imaging study. *Echocardiography*, *28*(7), 746–752. doi: 10.1111/j.1540-8175.2011.01422.x
- Young, A. A., & Frangi, A. F. (2009, apr). Computational cardiac atlases: from patient to population and back. *Experimental Physiology*, *94*(5), 578–596. doi: 10.1113/expphysiol.2008.044081
- Yushkevich, P. A., Gao, Y., & Gerig, G. (2016, aug). ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE. doi: 10.1109/embc.2016.7591443
- Zhang, D., & Hebert, M. (1999). Harmonic maps and their applications in surface matching. In *Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (cat. no PR00149)*. IEEE Comput. Soc. doi: 10.1109/cvpr.1999.784731
- Zhang, H., Wahle, A., Johnson, R., Scholz, T., & Sonka, M. (2010, feb). 4-d cardiac MR

- image analysis: Left and right ventricular morphology and function. *IEEE Transactions on Medical Imaging*, 29(2), 350–364. doi: 10.1109/tmi.2009.2030799
- Zhang, M., & Golland, P. (2016, oct). Statistical shape analysis: From landmarks to diffeomorphisms. *Medical Image Analysis*, 33, 155–158. doi: 10.1016/j.media.2016.06.025
- Zhao, B., Tan, Y., Tsai, W.-Y., Qi, J., Xie, C., Lu, L., & Schwartz, L. H. (2016, mar). Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Scientific Reports*, 6(1). doi: 10.1038/srep23428
- Zuluaga, M. A., Cardoso, M. J., & Ourselin, S. (2020). Automatic right ventricle segmentation using multi-label fusion in cardiac mri. doi: 10.48550/ARXIV.2004.02317
- Zwanenburg, A., Leger, S., Agolli, L., Pilz, K., Troost, E. G. C., Richter, C., & Löck, S. (2019, jan). Assessing robustness of radiomic features by image perturbation. *Scientific Reports*, 9(1). doi: 10.1038/s41598-018-36938-4