

Quantitative Analysis of Cardiac Magnetic Resonance in Population Imaging



Abdolrahman Attar Qazaani

University of Leeds

School of Computing

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

October, 2020

*To all the nurses, doctors and front-line health care
workers around the world ...*

Intellectual Property Statement

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. The contribution of the candidate and the other authors to this work has been explicitly indicated as follows.

I am the main author of all the following publications. I led the design of the study and wrote the manuscript, including the analysis and discussion of the results. I also wrote the analysis code and carried out the data processing, statistical analysis, and experiments. The other authors' contribution was in helping me designing the study, providing the data annotations, discussion of the results, and review of the manuscript.

- **Chapter 2:**

- 1) Attar, Rahman, et al. "Quantitative CMR population imaging on 20,000 subjects of the UK Biobank imaging study: LV/RV quantification pipeline and its evaluation." *Medical Image Analysis* 56 (2019): 26-42.

- 2) Attar, Rahman, et al. "High throughput computation of reference ranges of biventricular cardiac function on the UK Biobank population cohort." *MICCAI International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, Cham, 2018.

- **Chapter 3:**

- 1) Attar, Rahman, et al. "Automatic 3D+t Four-Chamber CMR Quantification of the UK Biobank: integrating anatomical and demographic priors at scale." under review.

2) Attar, Rahman, et al. "3D Cardiac Shape Prediction with Deep Neural Networks: Simultaneous Use of Images and Patient Metadata." MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.

- **Chapter 4:**

1) Attar, Rahman, et al. "Volumetric Epicardial Fat Quantification of UK Biobank Imaging Study." under review.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Abdolrahman Attar Qazaani to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

© 2020 The University of Leeds and Abdolrahman Attar Qazaani.

Acknowledgements

I would like to thank my supervisor, Professor Alejandro Frangi, for giving me the opportunity to pursue a PhD degree and work alongside him at CISTIB. I am thankful for his enthusiastic leadership, continuous constructive criticism and friendly advice, all of which helped me become a better scientist.

I am very thankful to all my fellow labmates over the last four years in Sheffield and Leeds, for all the good memories. Special thanks to Dr Marco Pereanez, Dr Christopher Bowles, Dr Andres Diaz-Pinto, and Dr Nishant Ravikumar for the numerous friendly and fruitful meetings and team discussions. I also thank our team of software engineers, Milton Hoz de Vila, Alejandro Daniel Lebrero, and Kattia Lisette Hoz de Vila for all the technical support throughout different parts of the project.

I appreciate the Doctoral Scholarship covering the full tuition fee and living cost stipends from the Department of Electronic and Electrical Engineering, University of Sheffield (the first two years of my PhD) and the School of Computing, University of Leeds (the second two years of my PhD). I also acknowledge AWS Research Grants that generously provided Cloud Credits throughout my PhD to run all the large-scale experiments presented in this thesis.

Last but not least, I want to express my deepest gratitude to Romina[♥] for her loving support and endless patience during my studies. I am deeply grateful to my parents for their unflagging love and continues support throughout my life.

Abstract

According to the World Health Organisation, cardiovascular diseases are the most prevalent cause of death worldwide and taking nearly 18 million lives each year. Identifying individuals at risk of cardiovascular diseases, and ensuring they receive appropriate treatment in time can prevent premature deaths. Early quantitative assessment of cardiac function, structure, and motion support preventive care and early cardiovascular treatment. Therefore, fully automated analysis and interpretation of large-scale population-based cardiovascular magnetic resonance imaging studies become of high importance. This analysis helps to identify patterns and trends across population groups, and accordingly, reveal insights into key risk factors before diseases fully develop.

To date, few large-scale population-level cardiac imaging studies have been conducted. UK Biobank (UKB) is currently the world's most extensive prospective population study, which in addition to various biological and physical measurements, contain cardiovascular magnetic resonance (CMR) images to establish cardiovascular imaging-derived phenotypes. CMR is an essential element of multi-organ multi-modality imaging visits for patients in multiple dedicated UK Biobank imaging centres that will acquire and store imaging data from 100,000 participants by 2023.

This thesis introduces CMR image analysis methods that appropriately scales up and can provide a fully automatic 3D analysis of the UKB CMR studies. Without manual user interactions, our pipeline performs end-to-end image analytics from multi-view cine CMR images all the way to anatomical and functional quantification. Besides, our pipelines provide 3D anatomical models of cardiac structures which enable the extraction of detailed information of the morphodynamics of the cardiac structures for subsequent associations to genetic, omics, lifestyle habits, exposure information, and other available information in population imaging studies. We present the quantification results from 40,000 subjects of the UK Biobank at 50 time-frames, i.e. two million image volumes.

CONTENTS

1	Introduction: Background, Motivation and Contribution	1
1.1	Heart Anatomy and Function	1
1.2	Cardiovascular Disease	3
1.3	Quantitative Cardiac Image Analysis	4
1.4	Large-scale Population-level Cardiac Image Analysis	5
1.5	Technical Background and Evaluation Criteria	7
1.5.1	Background Theory	7
1.5.2	Data and Evaluation Metrics	10
1.6	Thesis Contributions and Overview	12
2	Quantitative Population Analysis of Cardiac Ventricles	15
2.1	Introduction	16
2.2	Methodology	18
2.2.1	Workflow Integration and Execution	19
2.2.2	Data Pre-processing	20
2.2.3	Quality Analysis	20
2.2.4	Segmentation	21
2.2.5	Quantification	24
2.3	Experiments and Results	28
2.3.1	Segmentation Accuracy	29
2.3.2	Estimation of Cardiac Function Indexes	33
2.3.3	Hardware and Computational Cost	48
2.3.4	Sub-Cohort Analysis	50
2.4	Summary	53

CONTENTS

3	Quantitative Population Analysis of Cardiac Chambers	55
3.1	Introduction	56
3.2	Method	60
3.2.1	Reference 3D Cardiac Shapes	60
3.2.2	Point Distribution Model (PDM)	64
3.2.3	Network Architecture and Loss Function	65
3.3	Experiments and results	70
3.3.1	Data and Annotations	70
3.3.2	Implementation and Training	71
3.3.3	Accuracy of Predicted Shapes	71
3.3.4	Accuracy of Cardiac Function Indexes	77
3.3.5	Sub-Cohort Analysis	89
3.3.6	Impact of Metadata on Shape Accuracy	92
3.3.7	Contribution of Slices on Shape Accuracy	93
3.4	Summary	96
4	Quantitative Population Analysis of Epicardial Fat Tissue	101
4.1	Introduction	102
4.2	Method	106
4.2.1	Image Analysis Method	106
4.2.2	Data	109
4.2.3	Method Validation	109
4.3	Statistical Analysis and Interpretation	112
4.3.1	Association of EFT with HbA1c	112
4.3.2	Association of EFT with Demographics	113
4.3.3	Association of EFT with Cardiac Function	114
4.3.4	Association of EFT with Genetic Variations	118
4.4	Summary	120
5	Conclusions	123
5.1	Summary and Achievements	123
5.2	Limitations and Further Research Directions	124
5.3	Code and Results Availability	127

CONTENTS

List of Publications	135
Biosketch	137
References	139

CONTENTS

LIST OF FIGURES

1.1	Chambers of the heart. The heart has four chambers: two atria and two ventricles. (Modified from source: © Healthwise, Incorporated)	2
1.2	Illustration of the DSC, MCD and HD metrics used to report segmentation accuracy.	11
2.1	Schematic showing our fully automatic image parsing framework for large-scale analysis of cardiac ventricles. CMR images first go through the pre-processing phase, then flow into both the quality analysis and segmentation phases, which in turn communicate with one another, finally producing output that the last phase of quantification handles. . .	19
2.2	Results of our segmentation algorithm parameter optimisation. Here, the set of parameters that jointly yielded the best results for the DSC, MCD and HD metrics was test 4.	25
2.3	Segmentation accuracy expressed in terms of the DSC, MCD and HD measures.	35
2.4	Examples segmentation results at the ED and ES phases illustrating our three degrees of quality for automated segmentation contours versus manual contours. Red: automated segmentations. Green: ground-truth segmentations.	36

LIST OF FIGURES

2.5	Illustrating the repeatability of various cardiac functional indexes comparing the manual and automated analysis of 4,620 subjects from the UKB cohort. The top row shows Bland-Altman plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available. The black horizontal lines denote the mean difference (i.e. bias), whereas the two red dashed lines denote limits of agreement (LoA) i.e. ± 1.96 standard deviations from the mean. The second row shows correlation plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available.	41
2.6	Distributions of various cardiac functional indexes comparing results of manual and automatic analyses of 4,620 subjects. The top row shows probability distribution plots, whereas the bottom row shows Q-Q plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available.	43
2.7	Segmental LV parameters of 4,620 subjects presented as bulls-eye displays.	44
2.8	Regional analysis of LV shapes covering 40,000 subjects in terms of distribution of wall thickness at ED and ES phases. Here, red, blue and green lines indicate ground-truth values for 4,620 subjects, automated values for 4,620 subjects and automated values for 40,000 subjects, respectively. In all plots, the y-axis represents the relative frequency.	46
2.9	Regional analysis of LV shapes covering 40,000 subjects in terms of distribution of wall motion and thickening. Here, red, blue and green lines indicate ground-truth values for 4,620 subjects, automated values for 4,620 subjects and automated values for 40,000 subjects, respectively. In all plots, the y-axis represents the relative frequency.	47
2.10	Male (blue star marker) and Female (red circle marker) clinical indexes showing their mean value per age group.	52
3.1	An example 3D stack of 2D manual contours on SAX, two- and four-chamber LAX view slices.	61

3.2	The template is registered to each stack of 2D manual delineations to produce high resolution and smooth triangular mesh models for each subject. Then all the new generated reference shapes are used to create the point distribution model.	64
3.3	Representation of the mean \pm 3 standard deviation (SD) of the first five modes of variation in the 3D shape models of four-chambered cardiac. .	66
3.4	MCSI-Net extracts a high-level representation of the image from SAX and LAX views using four FCNs. It concatenates the image features together along with the output of an MLP network applied to the metadata. Through a multitask learning approach, fully connected layers with ReLU activation functions produce the k parameters in PCA space and their corresponding transformation parameters which provide the 3D shape of the cardiac.	68
3.5	Example segmentation results for short-axis and long-axis slices at the ED and ES phases illustrating the quality for automated segmentation contours versus manual contours. Blue: automated segmentations. Red: ground-truth segmentations. The automated method segments all the time frames. However, only ED and ES frames are shown, as manual analysis only annotates ED and ES frames.	73
3.6	Quantification results of 2 million CMR image volumes. Illustrating the mean (solid lines) and standard deviation (shaded area) of the ventricular and atrial volume (ml), LV mass (g) of 40k subjects of the UKB over the 50 time points.	81
3.7	Illustrating the repeatability of various cardiac functional indexes comparing the manual and automated analysis of 600 subjects from the UKB cohort; Bland-Altman plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available. The solid horizontal lines denote the bias or the mean difference (Automatic – Manual), whereas the two dashed lines denote limits of agreement (LoA) i.e. \pm 1.96 standard deviations from the mean.	87

LIST OF FIGURES

3.8	Illustrating the repeatability of various cardiac functional indexes comparing the manual and automated analysis of 600 subjects from the UKB cohort; Correlation plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available. The dashed and solid line denote the identity and linear regression lines, respectively.	88
3.9	Distributions of various cardiac functional indexes comparing results of manual and automatic analyses of 600 subjects. It shows probability distribution plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available.	90
3.10	Q-Q plots for various cardiac functional indexes computed both manually and automatically of 600 subjects.	91
3.11	Illustration of the quality of the shapes in different views. Cardiac shapes generated from a fraction of image data using only BMA, only SAX, or only LAX. Each colour corresponds to an individual subject.	97
4.1	Sample CMR image showing the adipose tissue (bright regions) surrounding the epicardial boundary (dashed line).	103
4.2	The proposed method using the CMR images volumes produces 3D mesh models of the cardiac ventricles in a full cardiac cycle. Then EFT is quantified within the search region around the epicardial wall.	108
4.3	Example segmentation results of eight random short-axis slices of different subjects at random phases of a cardiac cycle illustrating the quality for automated segmentation versus manual.	111
4.4	Distribution of EFT volume (ml) and HbA1c (mmol/mol) of 40,000 subjects of the UKB.	113
4.5	Distribution of EFT over three different groups of Non-Diabetic, Prediabetic, and Diabetic shows an extremely significant difference. Numbers in parenthesis shows the percentage of each category in the entire population.	114
4.6	Distribution of EFT for male and female subjects. Female subjects have higher EFT compared with male. Numbers in parenthesis shows the percentage of each category in the entire population.	115

4.7 Distribution of EFT for subjects with normal BMI and overweight shows overweight subjects have higher EFT. Numbers in parenthesis shows the percentage of each category in the entire population. 116

4.8 Distribution of EFT for two age groups. Older subjects have higher EFT compared with younger. Numbers in parenthesis shows the percentage of each category in the entire population. 117

4.9 Genomic loci associated with EFT. Manhattan plot depicting the genome-wide association study results of EFT traits. The red line indicates the Bonferroni significance threshold. A less stringent threshold ($p < 10^{-5}$) is highlighted in blue. 121

5.1 The fully automatic image parsing workflow for the analysis of cardiac ventricles in parallel. Left: The workflow includes the following modules: DO: Data Organisation, IQA: Image Quality Assessment, OD: Organ Detection, MI: Model Initialisation, S: Segmentation, SQA: Segmentation Quality Assessment, Q: Quantification, DS: Data Sink. Right: The quantitative functional analysis of a large database in parallel mode. DB: Database, DG: Data Grabber, n: number of subjects, and S_i : i^{th} subject of the dataset. 128

5.2 The schematic of the fully automatic workflow for the analysis of cardiac chambers. Green arrows indicate a chunk of data to be processed in every round. Chunk size, which is the number of subjects to be processed simultaneously, depends on the virtual machine’s computational resources. 130

5.3 Addition of a new module to the existing workflow for the fully automatic analysis of epicardial fat tissue in a large scale dataset. We use the privilege information given by our cardiac segmentation workflow to quantify fat tissue accurately. 132

LIST OF TABLES

2.1	The list of 54 distinct sets of segmentation parameters used in our segmentation algorithm parameter optimisation. As noted in the text, test 4 was the best choice.	26
2.2	Specific datasets used for training and testing the methods proposed and presented in this chapter.	30
2.3	Segmentation results based on the different test sets. The metrics used are DSC, MCD and HD. We compare manual with automatic methods, and error between human observers. M represents the manual ground-truth, LV_{endo} represents LV endocardium, LV_{myo} represents LV myocardium, and RV_{endo} represents RV endocardium. Table values are shown as mean \pm standard deviation.	31
2.4	Regional segmentation accuracy of LV myocardium based on the AHA 17-segment model covering 600 subjects. Values indicate mean \pm standard deviation. Bold indicates the cases in which our algorithm (i.e. A_S) outperformed algorithm B.	34
2.5	Categories of segmentation quality for the different approaches presented in this chapter.	37
2.6	A summary of subjects used in our analysis after quality control measures were applied.	38
2.7	Summarising the differences in clinical measures derived from our proposed method and manual segmentation. Here, GT represents the ground-truth values provided by manual segmentation. Values indicate mean \pm standard deviation.	39

LIST OF TABLES

2.8	The difference in clinical measures between the automatic and manual segmentations, as well between measurements by different human observers. M: ground truth provided by manual segmentation. VF: Automatic segmentation obtained from the automatic segmentation software inlineVF D13A. Values indicate mean \pm standard deviation.	40
2.9	Segmental LV parameters of 4,620 subjects obtained from manual and automatic approaches. Upper rows correspond to shapes generated from the manual segmentation and lower rows to those obtained with the automatic approach.	45
2.10	Comparing one of the current state-of-the-art CNN-based methods proposed by Bai et al. with our proposed framework.	49
2.11	Male (M) and Female (F) ventricular reference ranges detailing mean, lower reference limit and upper reference limit by age group. Reference limits are derived by the upper and lower bounds of the 95% prediction interval for each parameter at each age group.	51
3.1	Parameters used for region-wise and global registration in the CPD method.	63
3.2	Comparison of shape prediction accuracy using different FCNs in terms of DSC, MCD(mm) and HD(mm) for LV endo-/epicardium, RV, LA and RA.	69
3.3	Summary of the metadata (MD) of 40,000 subjects used in this study. According to the nature of the metadata, we have grouped them into four categories, i.e. MD ₁ : demographics, MD ₂ : blood pressure, MD ₃ : biological samples, and MD ₄ : lifestyle. Continuous values indicate mean \pm standard deviation.	72
3.4	Recent cardiac segmentation methods and reported accuracy compared with manual delineation in terms of the three common metrics: DSC, MCD(mm), and HD(mm), where available.	74
3.5	Specific UKB datasets used for training and testing the methods proposed and presented in this chapter.	76

3.6 Dice similarity coefficient. Segmentation results based on the different test sets. We compare manual with automatic methods, and error between human observers. LV_{endo} represents LV endocardium, LV_{myo} represents LV myocardium, RV_{endo} represents RV endocardium, $LA_{2\text{ch}}$ represents LA in two chamber LAX view, $LA_{4\text{ch}}$ represents RA in four chamber LAX view, and $RA_{4\text{ch}}$ represents represents RA in four chamber LAX view. Table values are shown as mean \pm standard deviation. 78

3.7 Mean contour distance. Segmentation results based on the different test sets. We compare manual with automatic methods, and error between human observers. LV_{endo} represents LV endocardium, LV_{myo} represents LV myocardium, RV_{endo} represents RV endocardium, $LA_{2\text{ch}}$ represents LA in two chamber LAX view, $LA_{4\text{ch}}$ represents RA in four chamber LAX view, and $RA_{4\text{ch}}$ represents represents RA in four chamber LAX view. Table values are shown as mean \pm standard deviation. 79

3.8 Hausdorff distance. Segmentation results based on the different test sets. We compare manual with automatic methods, and error between human observers. LV_{endo} represents LV endocardium, LV_{myo} represents LV myocardium, RV_{endo} represents RV endocardium, $LA_{2\text{ch}}$ represents LA in two chamber LAX view, $LA_{4\text{ch}}$ represents RA in four chamber LAX view, and $RA_{4\text{ch}}$ represents represents RA in four chamber LAX view. Table values are shown as mean \pm standard deviation. 80

3.9 Summarising the differences in ventricular measures derived from our proposed method and manual segmentation. Here, GT represents the ground-truth values provided by manual segmentation. A, B, and C represents the quantification results derived from automated segmentation methods. Values indicate mean \pm standard deviation. 83

3.10 Summarising the differences in atrial measures derived from our proposed method and manual segmentation. Here, GT represents the ground-truth values provided by manual segmentation. A, B, and C represents the quantification results derived from automated segmentation methods. Values indicate mean \pm standard deviation. 84

LIST OF TABLES

3.11	The absolute difference in clinical measures between the automatic and manual segmentations, as well between measurements by different human observers. M: ground truth provided by manual segmentation. A, B, and C represents the clinical indexes derived from the automated segmentation methods. Values indicate mean \pm standard deviation. . .	85
3.12	The relative difference in clinical measures between the automatic and manual segmentations, as well between measurements by different human observers. M: ground truth provided by manual segmentation. A, B, and C represents the clinical indexes derived from the automated segmentation methods. Values indicate mean \pm standard deviation. . .	86
3.13	Example comparison of cardiac function on large-scale datasets using automatically derived clinical indexes of 1600 subjects. Illustration of cardiac remodelling on healthy subjects vs subjects with myocardial infarction (MI).	92
3.14	Comparison of shape prediction accuracy using only images (IMG) or images with different type of metadata (IMG+MD) in terms of DSC, MCD(mm) and HD(mm) for LV endo-/epicardium and RV, LA, and RA on 600 UKB subjects.	94
3.15	Comparison of shape prediction accuracy using different set of images i.e. only SAX images, only LAX images, only three main SAX images (BMA: base, mid, and apex) or combination of SAX and LAX images in terms of DSC, MCD(mm) and HD(mm) for LV endo-/epicardium and RV, LA, and RA on 600 UKB subjects.	98
4.1	Summary of the most recent articles studied EFT and its association with other biomarkers.	104
4.2	Summary of the clinical characteristics of 40,000 subjects of the UKB analysed in this study. Continuous values indicate mean \pm standard deviation.	110
4.3	Segmentation accuracy. The comparison of the automated segmentation results with the reference manual contours in terms of DSC(%), MCD(mm) and HD(mm).	112

LIST OF TABLES

4.4 Association of the low and high volumes of EFT with cardiac function indexes of all the UKB subjects, Non-Diabetic, Prediabetic, and Diabetic subjects. 119

Abbreviations

2D	Two-Dimensional	LVEDV	Left Ventricle End-Diastolic Volume
3D	Three-Dimensional	LVEF	Left Ventricle Ejection Fraction
ASM	Active Shape Models	LVESV	Left Ventricle End-Systolic Volume
CMR	Cardiovascular Magnetic Resonance	LVM	Left Ventricle Myocardial Mass
CNN	Convolutional Neural Network	LVSV	Left Ventricle Stroke Volume
CPD	Coherent Point Drift	MI	Myocardial Infarction
CT	Computed Tomography	MLP	Multilayer Perceptron
CVD	Cardiovascular Disease	MRI	Magnetic Resonance Imaging
ED	End-Diastolic	PCA	Principal Component Analysis
EFT	Epicardial Fat Tissue	PDM	Point Distribution Model
EM	Expectation-Maximisation	RA	Right Atrium
ES	End-Systolic	RAEDV	Right Atrium End-Diastolic Volume
FCN	Fully Convolutional Networks	RAEF	Right Atrium Ejection Fraction
GMM	Gaussian Mixture Model	RAESV	Right Atrium End-Systolic Volume
GT	Ground-Truth	RASV	Right Atrium Stroke Volume
GWAS	Genome-Wide Association Study	RV	Right Ventricle
IAM	Intensity Appearance Model	RVEDV	Right Ventricle End-Diastolic Volume
IQA	Image Quality Assessment	RVEF	Right Ventricle Ejection Fraction
LA	Left Atrium	RVESV	Right Ventricle End-Systolic Volume
LAEDV	Left Atrium End-Diastolic Volume	RVSV	Right Ventricle Stroke Volume
LAEF	Left Atrium Ejection Fraction	SAX	Short Axis
LAESV	Left Atrium End-Systolic Volume	SD	Standard Deviation
LASV	Left Atrium Stroke Volume	SNP	Single-Nucleotide Polymorphism
LAX	Long Axis	SQA	Segmentation Quality Assessment
LoA	Limits of Agreement	UKB	UK Biobank
LV	Left Ventricle		

CHAPTER 1

Introduction: Background, Motivation and Contribution

1.1 Heart Anatomy and Function

Heart as a crucial organ to our survival pumps blood through the circulatory system to provide oxygen and nutrients to cells and carries away unwanted carbon dioxide and metabolic waste products [1]. The heart which is located between the lungs, in the middle compartment of the chest, has four chambers, as shown in Figure 1.1: two atria and two ventricles, and they function as follows:

- The right atrium (RA) receives deoxygenated blood from the body and pumps it to the right ventricle.
- The right ventricle (RV) pumps the deoxygenated blood to the lungs.
- The left atrium (LA) receives oxygenated blood from the lungs and pumps it to the left ventricle.
- The left ventricle (LV) pumps the oxygenated blood to the body.

1. INTRODUCTION: BACKGROUND, MOTIVATION AND CONTRIBUTION

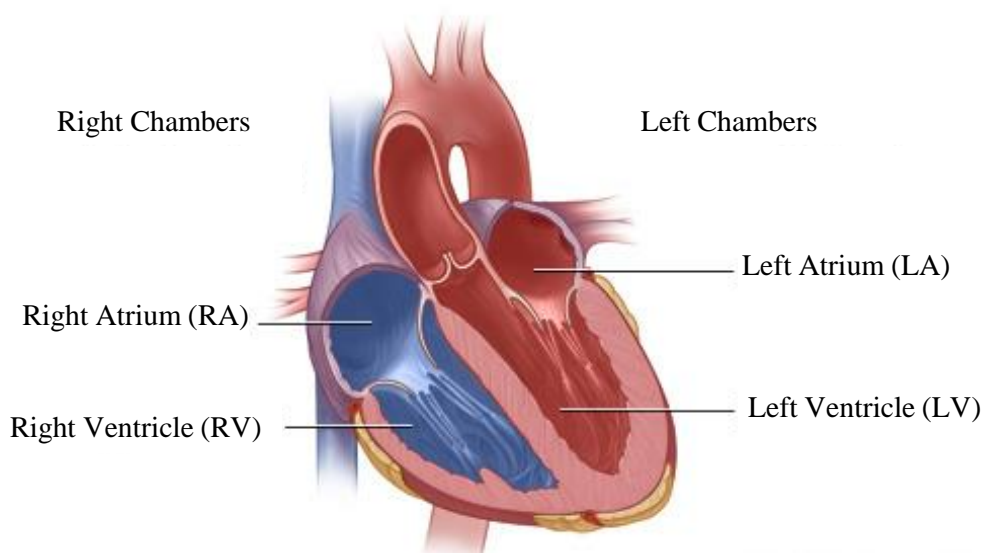


Figure 1.1: Chambers of the heart. The heart has four chambers: two atria and two ventricles. (Modified from source: © Healthwise, Incorporated)

The atria are smaller than the ventricles with thinner and less muscular walls and act as blood receiving chambers while ventricles that are connected to the arteries act as pumping chambers to send blood out of the heart. The ventricular wall has three layers: epicardium (the external layer), myocardium (the central layer), which contains the muscle for contraction, and endocardium (the internal layer). Considering adequate blood pressure is needed to pump the oxygenated blood to various parts of the body, the LV wall is thicker than the RV wall [2].

This cardiac function happens continuously through the full cardiac cycle, which is a sequence of electrical and mechanical movements that happens throughout the relaxation (diastole) and contraction (systole) phases of the heart. The ventricular diastolic stage involves the blood transfer from atria to ventricles, and the ventricular systole involves the blood transfer from ventricles to the pulmonary artery and aorta. The heart pumps blood with a rhythm managed by a group of pacemaking cells in the sinoatrial node generating a current to contract the heart and passes through the atrioventricular node and the conduction system of the heart [1].

1.2 Cardiovascular Disease

Cardiovascular disease (CVD) also called heart and circulatory disease is a general term for conditions that affect cardiac morphology and function. According to the World Health Organisation (WHO) [3], CVD is the most prevalent cause of death worldwide, accounting for nearly 18 million deaths each year.

CVD affects nearly seven million people in the UK and is a significant cause of disability and death [4]. The long term plan of the National Health Service (NHS) classifies CVD as a clinical priority and the most prominent condition where lives can be saved by the NHS over the next ten years. The plan sets the goal for the NHS to help prevent thousands CVD cases over the next ten years.

There are several risk factors for CVDs that each has moderate effects and interact with each other in complex ways, such as smoking, stress, alcohol, high blood pressure, high blood cholesterol, being physically inactive, being overweight, diabetes, family history of heart disease, ethnic background, sex, and age. Our current understanding of these risk factors of CVDs is mainly due to prospective population-based research studies such as the Framingham Heart Study [5], the Monitoring trends and determinants in Cardiovascular disease (MONICA) project [6], the INTERHEART study [7], the East German population-based Study of Health in Pomerania (SHIP) [8], the Jackson Heart Study [9], the Multi-Ethnic Study of Atherosclerosis (MESA) [10], the Dallas Heart Study [11], and the Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-Reykjavik) [12], all of which have demonstrated the advantage of population-based longitudinal studies for predicting and preventing CVDs.

In contrast to retrospective studies where the follow-up information is not available, prospective cohort studies provide the principal requirements and information for a better assessment of the risk factors and their individual and collective contribution in the progression of diseases including CVDs. However, prospective cohort studies require large-scale population-level studies because only a relatively small proportion with a particular background will develop a particular condition. When it is known which individuals have developed some particular condition, then further detailed investigations can be conducted [13].

Identifying individuals at risk of CVDs and ensuring they receive appropriate and timely treatment can help prevent premature deaths. However, diagnosis of CVD is

often made at late symptomatic stages, which leads to late interventions and decreased efficacy of medical care. Early quantitative assessment of the cardiac structure, motion, and function support preventive care and early cardiovascular treatment. Therefore, mechanisms for fully automated analysis and interpretation of large-scale population-based imaging studies are of high importance. This analysis helps to identify patterns and trends across population groups, and accordingly, provides insights into key risk factors before CVDs fully develop.

1.3 Quantitative Cardiac Image Analysis

Assessment of cardiac function can be achieved through different baseline tests such as blood test or electrocardiogram (ECG), or different imaging techniques that are available in clinical practice to generate images of the heart from different views, such as echocardiogram (echo), chest x-ray, computed tomography (CT), nuclear imaging, and magnetic resonance imaging (MRI) which enable non-invasive qualitative and quantitative assessment of cardiac function and structure to provide support for diagnosis, disease monitoring, treatment planning, and prognosis.

Among all, Cardiovascular Magnetic Resonance (CMR) imaging provides accurate morphological information and good quality soft-tissue contrast of a human heart. These images are an excellent source of visual information to monitor and analyse cardiac function for early diagnosis of heart abnormalities. CMR has fundamental advantages over other imaging techniques that encourage the community for its use in clinical practice and research, where it has established itself as the non-invasive gold standard for assessing cardiac chambers for a wide range of CVDs. CMR offers accurate and reproducible tomographic, static, or cine images of high spatial and temporal resolution in any desired plane without exposure to contrast agents or ionising radiation. As such, long-axis and short-axis views can be acquired to allow a visual, qualitative assessment of function [13, 14].

However, due to a large amount of visual information within the CMR images (different slices from different views at different time points), it is required to detect and segment the target anatomical regions for the further steps of quantification and clinical assessment. However, clinicians have been using manual or semi-automatic

approaches for CMR image analysis for years, which is time-consuming and prone to subjective errors. It is a major clinical challenge to derive quantitative and clinically relevant information from CMR images automatically. In this regard, fully automated cardiac image segmentation is an essential first step to partition the image into several anatomically meaningful regions, based on which quantitative measures can be obtained. Typically, the anatomical regions of interest for CMR image segmentation are the cardiac four chambers, i.e. LV, RV, LA, and RA.

There is an extensive literature of CMR segmentation methods (see reviews [15–21]). However, most published methods have been developed and tested using datasets of a few dozen images which in many cases are private databases that are not open access. During the last few years, the dimensionality of datasets used in the CMR image analysis has significantly increased. This presents an unprecedented challenge in the community since the existing algorithms do not necessarily perform equally accurate and efficient when dealing with these extraordinary high dimensions, i.e. a considerable number of subjects with spatial and temporal imaging data. New technologies, frameworks and algorithms are thus needed to efficiently process these large quantities of data to derive useful information.

To date, few really large-scale population-level cardiac imaging studies have been conducted. One such study stands out for its sheer size, careful implementation, and availability of top quality expert annotation; the UK Biobank (UKB). The resulting massive imaging datasets (targeting ca. 100,000 subjects) has put published approaches for cardiac image quantification to the test.

1.4 Large-scale Population-level Cardiac Image Analysis

Population imaging studies generate data for developing and implementing personalised health strategies to prevent or more effectively treat diseases. Large prospective epidemiological studies acquire imaging for pre-symptomatic populations. These studies enable the early discovery of alterations due to impending disease and enable early identification of individuals at risk.

Currently, patients with similar CVDs symptoms often receive the same treatments at late symptomatic stages, which leads to late interventions and decreased efficacy

1. INTRODUCTION: BACKGROUND, MOTIVATION AND CONTRIBUTION

of medical care. In this context, healthcare needs new biomarkers from large-scale databases of clinical and biological data [22, 23]. Biomarkers which correspond to different stages of disease progression have various functions. Biomarkers can support patients who seem that have no disease, those who have the risk of having a disease, those who are suspected of having disease, and those with clear pathology indicators [24].

Computation of accurate and reproducible predictive biomarkers derived from cardiac structural and functional measurements enable investigation of the disease progression and its association with those imaging-derived phenotypes. The combination of non-invasive CMR imaging with clinical data offers a rich source of big cardiac data, which opens up new issues of exploration to improve our understanding of the progression of CVDs across different population groups [13, 25]. This kind of study contributes toward the global move to predictive, preventive, personalised, and participatory medicine through big data analysis [26–28]. To study the possible associations within genetic, lifestyle factors, imaging-derived phenotypes, and subsequent risk of a wide range of diseases, it is essential to perform imaging in very large numbers of individuals as only a relatively small proportion of them will develop any particular condition during follow-up [29].

Large-scale population-based imaging studies of CVDs are becoming possible due to the advent of standardised, robust non-invasive imaging methods and infrastructure for big data analysis [30]. However, images constitute essentially of pixels and voxels, resulting in extremely high-dimensional feature spaces of low semantic value. Fundamentally, the goals are to process, aggregate, and reduce these raw intensity image signals to transform the images into higher-level representations (e.g., clusters, labels, shapes, biomarkers, etc.). Thus, such studies pose new challenges requiring automatic image analysis, an essential pre-requisite to automatically and robustly process the image data and extract information about the cardiovascular morphology and function using segmentation techniques [27, 31].

The UKB is a population-based prospective study, established to investigate the determinants of disease in middle and old age [32]. UKB is the largest and most detailed imaging study to date. The UKB imaging enhancement aims to perform brain, cardiac and abdominal MRI, full body dual-energy X-ray absorptiometry and a carotid

ultrasound scan on 100,000 of the existing 500,000 UKB participants before the end of 2023 [29]. In addition to the collection of extensive baseline questionnaire data, biological samples and physical measurements, CMR is utilised to provide cardiovascular imaging-derived phenotypes [13]. Over 45,000 participants have undergone an assessment, already making the UKB imaging enhancement by far the largest multi-modal imaging study in the world. At this time, CMR scans of 40,000 subjects have been released and are available for health-related research.

The amount of imaging data collected on such a large number of participants is truly unique. Yet it is the combination of these data with the wealth of other phenotypic, genetic and medical record information available in UKB that provides a powerful resource to address previously unanswerable research questions. Consequently, it is of high importance to have fully automated methods to quantify CMR-derived phenotypes and biomarkers from this large-scale study to identify early markers of pathology and their genetic and lifestyle determinants for preventive care and early cardiovascular treatment.

1.5 Technical Background and Evaluation Criteria

This section aims at giving a brief introduction to the fundamental theory behind the proposed segmentation approaches presented in the following chapters. We first explain the baseline technique of Principal Component Analysis (PCA). Subsequently, we show what evaluation metrics we use to assess the quality of our proposed methods presented in the next chapters.

1.5.1 Background Theory

Statistical Models for image processing were introduced by Cootes et al. [33]. These models can be used for different applications, including image segmentation in various domains. These models describe statistical variations as seen in a set of example images, in which corresponding landmark points are annotated. The shapes, which are spanned by the landmarks, are aligned using the Procrustes analysis to compensate for translation, rotation and scaling differences between the shape samples. The mean shape of this set of aligned shapes is calculated, and modes of shape variation are

1. INTRODUCTION: BACKGROUND, MOTIVATION AND CONTRIBUTION

computed using PCA. PCA reduces data by geometrically projecting them onto lower dimensions called principal components, intending to find the best summary of the data using a limited number of components.

To illustrate this, assume a training set of M shapes, each described by N points in \mathcal{R}^3 , i.e., $\mathbf{x}_j^i = (\mathbf{x}_j^i, \mathbf{y}_j^i, \mathbf{z}_j^i)$ with $i = 1, \dots, M$ and $j = 1, \dots, N$.

Further, let $\mathbf{s}_i = (\mathbf{x}_1^i, \mathbf{y}_1^i, \mathbf{z}_1^i, \dots, \mathbf{x}_N^i, \mathbf{y}_N^i, \mathbf{z}_N^i)^T$ be the i -th vector representing the shape of the i -th surfaces of cardiac chambers. Here, all nuisance pose parameters (e.g., translation, rotation and scaling) have been removed from using generalised Procrustes analysis. Considering this set covering a particular class of shapes, we will always observe some degree of inter-point correlation. Thus there could exist a shape representation accounting for correlation between points. If some point movements were to be correlated, this could be exploited to reduce dimensionality.

In our case, we will explore a linear transformation of the data:

$$\mathbf{s}'_i = \mathbf{L} \mathbf{s}_i \tag{1.1}$$

The shape class mean and covariance of \mathbf{s} and \mathbf{s}' are then as follows:

$$\bar{\mathbf{s}} = \frac{1}{M} \sum_{i=1}^M \mathbf{s}_i \tag{1.2}$$

$$\mathbf{C} = \frac{1}{M-1} \sum_{i=1}^M (\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})^T \tag{1.3}$$

$$\bar{\mathbf{s}}' = \frac{1}{M} \sum_{i=1}^M \mathbf{s}'_i = \frac{1}{M} \sum_{i=1}^M \mathbf{L} \mathbf{s}_i = \mathbf{L} \bar{\mathbf{s}} \tag{1.4}$$

$$\begin{aligned}
 \mathbf{C}' &= \frac{1}{M-1} \sum_{i=1}^M (\mathbf{s}'_i - \bar{\mathbf{s}}') (\mathbf{s}'_i - \bar{\mathbf{s}}')^T \\
 &= \frac{1}{M-1} \sum_{i=1}^M (\mathbf{L}\mathbf{s}_i - \mathbf{L}\bar{\mathbf{s}}) (\mathbf{L}\mathbf{s}_i - \mathbf{L}\bar{\mathbf{s}})^T \\
 &= \frac{1}{M-1} \sum_{i=1}^M \mathbf{L} (\mathbf{s}_i - \bar{\mathbf{s}}) (\mathbf{L} (\mathbf{s}_i - \bar{\mathbf{s}}))^T \\
 &= \frac{1}{M-1} \sum_{i=1}^M \mathbf{L} (\mathbf{s}_i - \bar{\mathbf{s}}) (\mathbf{s}_i - \bar{\mathbf{s}})^T \mathbf{L}^T \\
 &= \mathbf{L} \left(\frac{1}{M-1} \sum_{i=1}^M (\mathbf{s}_i - \bar{\mathbf{s}}) (\mathbf{s}_i - \bar{\mathbf{s}})^T \right) \mathbf{L}^T \\
 &= \mathbf{L}\mathbf{C}\mathbf{L}^T
 \end{aligned} \tag{1.5}$$

Then, if we limit ourselves to orthogonal transformations (i.e. $\mathbf{L}^{-1} = \mathbf{L}^T$) left-multiplication by \mathbf{L}^T in Equation 1.5 yields:

$$\mathbf{L}^T \mathbf{C}' = \mathbf{C}\mathbf{L}^T \tag{1.6}$$

Substitution of \mathbf{L}^T by Φ yields:

$$\mathbf{C}\Phi = \Phi\mathbf{C}' \tag{1.7}$$

From Equation 1.7 it is seen that if Φ is chosen as the (column) eigenvectors of the symmetric matrix \mathbf{C} , then the covariance of the transformed shapes, \mathbf{C}' , becomes a diagonal matrix of eigenvalues. In the case of correlated points the smallest eigenvalues will be (close to) zero and the corresponding eigenvectors could be omitted from Φ , thus reducing the length of \mathbf{s}' .

In conclusion, to establish a linear transform that de-correlate data vectors, the transformation matrix must be the eigenvectors of the covariance matrix of the original data. In order to back transform from the new set of variables, \mathbf{s}' , we invert Equation 1.1, remembering that \mathbf{L} is orthogonal:

$$\mathbf{s} = \mathbf{L}^{-1}\mathbf{s}' = \mathbf{L}^T\mathbf{s}' = \Phi\mathbf{s}' \tag{1.8}$$

Typically one would apply PCA on variables with zero mean (notice that the Φ is

unchanged):

$$\mathbf{s}' = \mathbf{L}(\mathbf{s} - \bar{\mathbf{s}}) \quad , \quad \mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi}\mathbf{s} \quad (1.9)$$

In summary, the shape covariance is represented in a low-dimensional space or PCA of the shape. That produces l eigenvectors $\mathbf{\Phi} = [\varphi_1\varphi_2\dots\varphi_l]$, and corresponding eigenvalues $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_l)$ of the covariance matrix computed via Singular Value Decomposition. Hence, assuming the shape class follows a multi-dimensional Gaussian probability distribution, any shape in the shape class can be approximated from the following linear generative model:

$$\mathbf{s} \approx \bar{\mathbf{s}} + \mathbf{\Phi}\mathbf{b} \quad (1.10)$$

where \mathbf{b} are shape parameters restricted to $|\mathbf{b}_i| \leq \beta\sqrt{\lambda_i}$; we typically set $\beta = 3$ to capture 99.7% of shape variability. The shape parameters of \mathbf{s} can then be estimated as follows:

$$\mathbf{b} = \mathbf{\Phi}_l^T(\mathbf{s} - \bar{\mathbf{s}}) \quad (1.11)$$

Here, the entries of \mathbf{b} are the projection coefficients of mean-centred shapes ($\mathbf{s} - \bar{\mathbf{s}}$) along the columns of $\mathbf{\Phi}$.

1.5.2 Data and Evaluation Metrics

We collected the UKB data under access applications number 2964 and 11350. This study complies with the Declaration of Helsinki. The study was covered by the ethical approval for UKB studies from the National Health Service National Research Ethics Service on 17th June 2011 (Ref 11/NW/0382) and extended on 10th May 2016 (Ref 16/NW/0274) with informed consent obtained from all participants. The full CMR protocol in the UKB has been described in detail elsewhere [34]. Researchers can apply to use the UKB resource for health-related research that is in the public interest¹.

For quantitative assessment of our proposed methods, we evaluate the performance of the automated method in two ways, respectively using commonly used metrics for

¹<https://www.ukbiobank.ac.uk/register-apply>

segmentation accuracy assessment and clinical measures derived from segmentations.

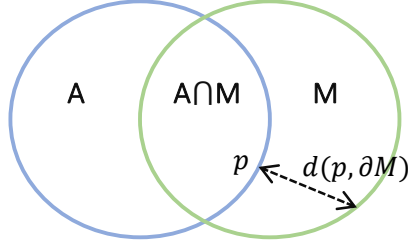


Figure 1.2: Illustration of the DSC, MCD and HD metrics. \mathbf{A} and \mathbf{M} are two sets representing automated segmentation and manual segmentation contours. The DSC metric is the ratio of the intersection over the average area of the two sets. The MCD first calculates, for each point on one contour, its distance to the other contour, then calculates the mean across all the points. The HD calculates the maximum distance between the two contours.

We compute the segmentation accuracy using three key metrics: Dice Similarity Coefficient (DSC), Mean Contour Distance (MCD) and Hausdorff Distance (HD). Figure 1.2 illustrates the definitions of the three aforementioned metrics, each of which is detailed below. First, the DSC evaluates the overlap between automated segmentation \mathbf{A} and manual segmentation \mathbf{M} ; we define DSC as follows:

$$DSC = \frac{2|\mathbf{A} \cap \mathbf{M}|}{|\mathbf{A}| + |\mathbf{M}|} \quad (1.12)$$

DSC is between 0 and 1, with a higher DSC indicating a better match between the two segmentations. The MCD and HD measures evaluate the mean and maximum distance, respectively, between segmentation contours $\partial\mathbf{A}$ and $\partial\mathbf{M}$. These measures are defined as follows:

$$MCD = \frac{1}{2|\partial\mathbf{A}|} \sum_{p \in \partial\mathbf{A}} d(p, \partial\mathbf{M}) + \frac{1}{2|\partial\mathbf{M}|} \sum_{q \in \partial\mathbf{M}} d(q, \partial\mathbf{A}) \quad (1.13)$$

$$HD = \max(\max_{p \in \partial\mathbf{A}} d(p, \partial\mathbf{M}), \max_{q \in \partial\mathbf{M}} d(q, \partial\mathbf{A})) \quad (1.14)$$

where $d(p, \partial)$ denotes the minimal distance from point p to contour ∂ . The lower the distance metric, the better the agreement.

We also evaluate the accuracy of clinical measures, which are derived from image segmentations. We report clinical cardiac functional indexes derived from manual and automated segmentation such as atrial and ventricular end-diastolic and end-systolic volume. To reproduce the reference ranges for cardiac structure and function, we first extract contours corresponding to the intersection between our 3D triangular meshes and CMR image slices. Then, for the ventricular indexes calculated on SAX slices, we use Simpson’s method of integration, whereby a cardiac 3D volume can be approximated by summing the areas within 2D segmentation contours and multiplying by the inter-slice spacing. Similarly, for the atrial indexes, we calculate the volume according to the area-length method on LAX slices. Having quantified cardiac chamber volumes across all the time points, we report the following sets of indexes, where applicable: the LV end-diastolic volume (LVEDV) and end-systolic volume (LVESV), LV stroke volume (LVSV), LV ejection fraction (LVEF), LV myocardial mass (LVM), RV end-diastolic volume (RVEDV) and end-systolic volume (RVESV), RV stroke volume (RVSV), RV ejection fraction (RVEF), LA end-diastolic volume (LAEDV) and end-systolic volume (LAESV), LA stroke volume (LASV), LA ejection fraction (LAEF), RA end-diastolic volume (RAEDV) and end-systolic volume (RAESV), RA stroke volume (RASV), and RA ejection fraction (RAEF).

1.6 Thesis Contributions and Overview

The main goal of this thesis is to provide fully automated workflows for large-scale image analysis of cardiac MRI sequences at all time points across the cardiac cycle and derive relevant clinical measures. We present and evaluate our proposed pipelines that properly scale up and can provide a fully automatic analysis of the UKB CMR study. Without any manual user interactions, our pipelines perform end-to-end image analytics from multi-view cine CMR images all the way to anatomical and functional quantification. The accuracy of clinical measures is comparable to human expert performance. These methods would assist clinicians in CMR image analysis and diagnosis with an automated and objective way for deriving clinical measures, therefore reducing

cost and improving work efficiency. They would also facilitate large-population imaging studies, such as the UKB, which aims to conduct CMR imaging scans of 100,000 subjects. Fully automated methods are crucial for analysing such a large amount of images and extracting clinically relevant information for subsequent clinical studies.

This thesis has addressed the quantitative analysis of population CMR imaging with the two following specific objectives:

- A fully automatic image parsing workflow with embedded quality control to perform end-to-end image analytics from multi-view cine CMR images all the way to anatomical and functional quantification, and evaluate its performance on a large-scale CMR imaging study.
- An accurate 3D modelling of cardiac chambers to enable the extraction of detailed information of the morphodynamics of the cardiac chambers for further study of its association to genetic, omics, lifestyle habits, exposure information, and other information provided in population imaging studies.

The proposed solutions to the objectives as mentioned above, are presented in the next chapters, which represent three specific contributions, with one chapter dedicated to each of them:

- **Chapter 2** presents a fully automatic, high throughput image parsing workflow for the analysis of CMR images, and test its performance on the UKB cardiac dataset. The proposed pipeline is capable of performing end-to-end image processing, including data organisation, image quality assessment, shape model initialisation, segmentation, segmentation quality assessment, and functional parameter computation; all without any user interaction. This study is the first published attempt tackling the fully automatic 3D analysis of the UKB population study, providing reference ranges for all key cardiovascular functional indexes, from both left and right ventricles of the heart.
- **Chapter 3** presents a new method for the accurate 3D modelling of cardiac four chambers. In this chapter, we propose a novel deep neural network using both CMR images and patient metadata to predict cardiac shape parameters

directly. The proposed method uses the promising ability of statistical shape models to simplify shape complexity and variability together with the advantages of convolutional neural networks for the extraction of solid visual features.

- **Chapter 4** presents a new method for quantification of epicardial fat tissue. Having developed fully automated methods for cardiac ventricles segmentation, we develop an extended quantification method for the extraction and volumetric quantification of fat tissue around the epicardial boundary. This study is the first attempt looking into the large-scale quantification of the epicardial fat tissue and its association with other information which reveal exciting findings.

Each of these three chapters is self-contained, and an adaptation of the articles that are under review, or already published in a peer-reviewed conference/journal papers. Finally, **Chapter 5** concludes the thesis and discusses the outlook and future work.

CHAPTER 2

Quantitative Population Analysis of Cardiac Ventricles

Population imaging studies generate data for developing and implementing personalised health strategies to prevent, or more effectively treat disease. Large prospective epidemiological studies acquire imaging for pre-symptomatic populations. These studies enable the early discovery of alterations due to impending disease, and enable early identification of individuals at risk. Such studies pose new challenges requiring automatic image analysis. To date, few large-scale population-level cardiac imaging studies have been conducted. One such study stands out for its sheer size, careful implementation, and availability of top quality expert annotation; the UK Biobank (UKB). The resulting massive imaging datasets (targeting ca. 100,000 subjects) has put published approaches for cardiac image quantification to the test. In this chapter, we present and evaluate a cardiac magnetic resonance (CMR) image analysis pipeline that properly scales up and can provide a fully automatic analysis of the UKB CMR study. Without manual user interactions, our pipeline performs end-to-end image analytics from multi-view cine CMR images all the way to anatomical and functional bi-ventricular quantification.

All this, while maintaining relevant quality controls of the CMR input images, and resulting image segmentations. To the best of our knowledge, this is the first published attempt to fully automate the extraction of global and regional reference ranges of all key functional cardiovascular indexes, from both left and right cardiac ventricles, for a population of 40,000 subjects imaged at 50 time frames per subject, for a total of two million CMR volumes. In addition, our pipeline provides 3D anatomical bi-ventricular models of the heart. These models enable the extraction of detailed information of the morphodynamics of the two ventricles for subsequent association to genetic, omics, lifestyle habits, exposure information, and other information provided in population imaging studies. We validated our proposed CMR analytics pipeline against manual expert readings on a reference cohort of 4,620 subjects with contour delineations and corresponding clinical indexes. Our results show broad significant agreement between the manually obtained reference indexes, and those automatically computed via our framework. 80.67% of subjects were processed with mean contour distance of less than 1 pixel, and 17.50% with mean contour distance between 1 and 2 pixels. Finally, we compare our pipeline with a recently published approach reporting on UKB data, and based on deep learning. Our comparison shows similar performance in terms of segmentation accuracy with respect to human experts.

2.1 Introduction

Cardiovascular disease (CVD) is the most prevalent cause of death worldwide [35]. Diagnosis of CVDs is often made at late symptomatic stages, leading to late interventions at high cost and with substantially decreased efficacy of treatment. Early quantitative assessment of cardiac function that allows for proper preventive care, and early cardiovascular treatment is therefore paramount. To support such an approach, large-scale population-based imaging studies of CVDs are increasingly possible given the advent of standardised robust non-invasive imaging methods, and the infrastructure for big data analysis [30]. These advancements open further opportunities for gaining new information about the development and progression of CVDs across various population groups [27, 31].

The analysis and interpretation of cardiac structural and functional indexes in large-

scale population imaging data can help identify patterns and trends across population groups, and accordingly, reveal insights into key risk factors before CVDs fully develop. Established to investigate the determinants of a disease, the UK Biobank (UKB) is one of the world’s largest prospective population studies [34]. The UKB data contain extensive baseline questionnaire data, biological samples, physical measurements, and cardiovascular magnetic resonance (CMR) images to establish cardiovascular imaging-derived phenotypes [13]. CMR is an important component of multi-organ multi-modality imaging visits for patients in multiple dedicated UKB imaging centres that will acquire and store imaging data from 100,000 participants by 2023.

In terms of population sample size, experimental setup, and quality control, the most reliable reference ranges for cardiovascular structure and function found in the literature are those reported by Petersen et al. [36], in which CMR scans were manually delineated and analysed by a team of eight expert observers using the commercially available cvi42 post-processing software (Version 5.1.1, Circle Cardiovascular Imaging Inc., Calgary, Canada). The expert team comprised of biomedical engineers, radiologists, image analysts and cardiologists, evaluated the quality of every image, and performed delineations. In cases where the image quality was doubtful, the team jointly decided upon exclusion. These reference values (delineations and volumes) comprise 4,620 subjects and are used in our present study to validate our proposed framework and workflow.

In this chapter, we present a novel fully automatic 3D image parsing workflow with embedded quality control, and evaluate its performance on the UKB. We validate our results by comparing with published manual analysis and one state-of-the-art method. Our proposed workflow is capable of segmenting the cardiac ventricles and generating global and regional clinical reference ranges comparable to those obtained by human raters and flagship methods.

In addition to comparing against manual measurements, we also compare our performance against one state-of-the-art method, i.e., the recent work by Bai et al. [37] in which the authors propose a 2D convolutional neural network (CNN)-based segmentation method for analysis of the UKB CMR images. Though in our study, we processed a much greater number of subjects (40,000), we performed experiments with smaller subsets of data to make direct comparisons with the existing literature. We

are interested in showing the advantages of true 3D shape analysis, over 2D CNN-based techniques, which, due to their per-slice disjoint nature, and absence of global constraints, lack the ability to infer or extrapolate noisy or missing data. We believe true 3D analysis is valuable, or even essential, for further structural analysis of regional myocardial function. Our 3D generative-based approach ensures global coherence of the cardiac anatomy and naturally lends itself to further analysis in which full 3D anatomy is necessary; for example, in mechanical and flow simulations.

Finally, since the power of population studies lies in the ability to provide normative reference values for sub-populations, enabling more patient-specific evaluation, we provide reference ranges for cardiac clinical indexes in sub-populations based on age-group and gender.

The main contributions of this chapter are, first, reproducing the cardiac functional index ranges derived from expert delineations reported in [34], and providing additional 3D-based ranges of local variation. Second, showcasing a fully scalable framework, capable of processing arbitrarily large population imaging studies, in a completely automatic manner. In this chapter we demonstrate this by processing 40,000 subjects from the UKB study, each comprised of 50 time frames for a total of two million image volumes, starting from raw input data, through data cleaning, quality assessment, 3D segmentation, volume computation, and statistical analysis.

The remainder of this chapter is organised as follows. In Section 2.2, we present our strategy for data processing scalability, and detail each of the modules comprising our image quantification pipeline. In Section 2.3, we present a thorough evaluation of our pipeline, both from technical, and clinical perspectives, including detailed statistics on global and local cardiovascular indexes. Finally, in Section 2.4, we present final remarks.

2.2 Methodology

Illustrated in Figure 2.1, our CMR image parsing pipeline consists of the following four phases: (1) pre-processing; (2) quality analysis; (3) segmentation; and (4) quantification. In the subsections that follow, we describe the methods used within each step and our design choices. In the next subsection, we highlight the framework used to

integrate this pipeline and streamline its execution both in terms of scalability and distributability.

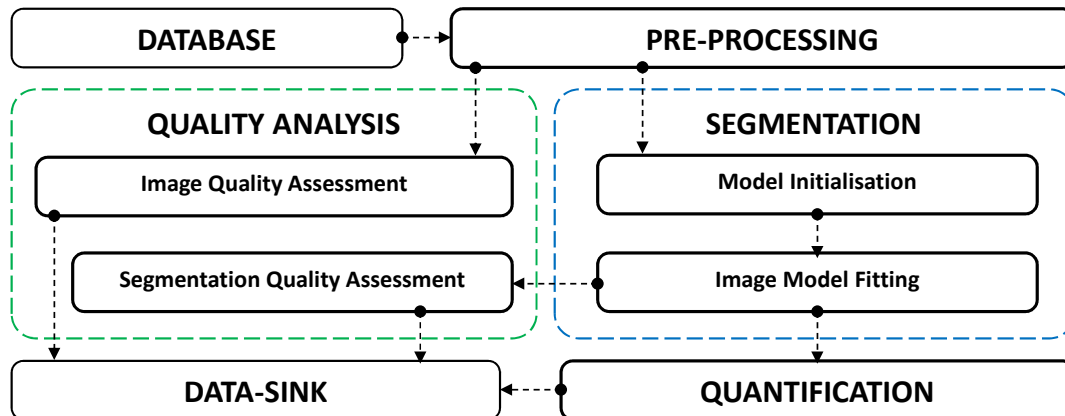


Figure 2.1: Schematic showing our fully automatic image parsing framework for large-scale analysis of cardiac ventricles. CMR images first go through the pre-processing phase, then flow into both the quality analysis and segmentation phases, which in turn communicate with one another, finally producing output that the last phase of quantification handles.

2.2.1 Workflow Integration and Execution

To scale both data access and computation, we propose a modular pipeline and developed an in-house cloud-based image analytics framework called MULTI-X ¹ [38]. MULTI-X enables both distributed access to data storage and distributed execution of image analysis pipelines on the cloud. Further, MULTI-X facilitates secure access and execution, component integration and interoperability (e.g., across different programming languages, frameworks, operating systems, and hardware), workflow execution, monitoring, and execution report generation. MULTI-X can also serve as middleware between storage and computing cloud providers (e.g., Amazon Web Services, Google-Cloud, and Microsoft Azure), workflow managers (e.g., Taverna and Nipype), data sources (e.g., UKB servers) and analytics tools providers. In our implementation, we selected Nipype [39] as the workflow manager. Further, we selected Amazon Web Ser-

¹<https://multi-x.org>

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

vices¹ to provide high-performance storage and computing in a cloud-based environment. More specifically, an Amazon Simple Storage Service (S3) provided unstructured data storage, Amazon Redshift provided data warehousing for petabyte-scale data analysis, and Amazon’s Elastic Cloud Computing (EC2) enabled on-demand adaptive cloud computing.

2.2.2 Data Pre-processing

Once data obtained from the UKB, they were transferred to a secure AWS S3 server accessible from an experimental deployment of MULTI-X, the aforementioned cloud-based infrastructure for our pipeline-oriented image analytics. When new CMR data are available, this fully automated pipeline provides detailed image analytics. The UKB Imaging Study undertakes detailed MRI scans of key vital organs of the human body using specialised imaging protocols that extend CMR. For each volunteer, relevant CMR subseries are extracted from the full imaging study, viz. short-axis (SAX) and long-axis (LAX) two-, three- and four-chamber CMR images and analysed through the next steps.

2.2.3 Quality Analysis

At least two quality analysis modules are required to ensure the reliability of the extracted cardiac indexes. The first module assesses the quality of the input images, whereas the second module assesses the quality of the quantification outputs, i.e., the generated 3D segmentations. Each of these is described in the subsections that follow.

Image Quality Assessment

Despite careful and strict imaging protocols, a significant portion of the data collected in population imaging studies, inevitably falls outside standard operating procedures. To ensure the quality and correctness of the collected data, thereby optimising the accuracy of the generated segmentation results, an image quality assessment (IQA) module detects suboptimal images whose inclusion in subsequent analysis would impair aggregated statistics over the entire cohort.

¹<https://aws.amazon.com>

More specifically, because the absence of basal and/or apical slices in SAX views forms the most frequently occurring problem affecting the accuracy of volumetric measurements and corresponding clinical indexes [40], our IQA module detects situations in which these slices are missing. In our design, SAX slices are processed independently through two CNN classifiers that determine the presence/absence of basal and apical slices, respectively. Details of the algorithm we used to achieve this effect can be found in the work published by Zhang et al. [41].

Segmentation Quality Assessment

Regarding segmentation quality assessment (SQA), large anatomical variations found across subject populations [42] and other forms of poor image quality beyond full ventricular coverage can cause image segmentation failures. We therefore propose an automated self-diagnosis mechanism for detecting unsatisfactory segmentation results. Flagged images can then be either re-processed with revised parameters or discarded from subsequent statistical analyses. We incorporate a segmentation quality assessment approach presented by Alba et al. [43]. The SQA module uses a random forest classifier trained to distinguish between successful and unsuccessful segmentations based on intensity features around the blood pool and myocardial boundaries.

2.2.4 Segmentation

For the segmentation phase of our workflow, we use SAX and LAX CMR images to estimate the approximate position and orientation of the cardiac ventricles. We then initialise the segmentation of the cardiac structure following a Sparse Active Shape Model (SPASM) approach [44]. More specifically, SPASM is used to segment the full cardiac cycle and retrospectively determine the end-diastolic (ED) and end-systolic (ES) phases of the cycle based on the frames showing maximum and minimum volumes, respectively. Before running our segmentation approach across all subjects, we applied grid search optimisation to a subset of 50 subjects to identify the parameters having the greatest impact on segmentation performance; we describe this further in Section 2.2.4.

Model Initialisation

To automatically initialise the model, we use the method proposed by Alba et al. [43] with a further step to improve bi-ventricular model initialisation. First, the location of the LV is determined via a rough estimate of the intersection of slices from the SAX and LAX views. Next, a random forest regressor trained with two complementary feature descriptors (i.e., the Histogram of Oriented Gradients and Gabor Filters) predicts the landmark positions for the LV. We extend this to take into account image features corresponding to the RV, thereby improving the initial estimate for the location of the bi-ventricular heart. We then use these landmarks to estimate pose parameters that place a mean shape model near the heart. Finally, we use these pose parameters to initialise the first image volume in the set of images for the cardiac cycle (i.e., 50 cardiac phases). Subsequent time frames are automatically initialised via the shape model fitted to the immediately preceding cardiac phase.

Image Model Fitting

In this subsection, we consider how we fit the image model. First, the cardiac LV and RV segmentations are obtained via the aforementioned SPASM segmentation method that improves on the Active Shape Models (ASM) approach [33] by addressing the sparsity found in imaging modalities such as CMR in which image information is sparsely distributed across the entirety of the image. The main components of the SPASM method are the Point Distribution Model (PDM), the Intensity Appearance Model (IAM) and a model matching algorithm.

The PDM encodes the mean and variance of the endocardial and epicardial shapes of the LV and the endocardial shape of the RV. The PDM is constructed during training using principal component analysis (PCA) on a set of generalised Procrustes-aligned shapes that preserve a 98% variance, as explained in details in Section 1.5.1. Next, we build an IAM based on intensity information across all corresponding landmarks in all training shapes \mathbf{s}_i . More specifically, IAMs capture the local intensity distribution along cardiac boundaries. We proceed by sampling one-dimensional intensity profiles normal to the myocardial boundaries. Each profile has a length of $m = 15$ pixels. For the i -th landmark, we estimate mean intensity profile $\bar{\mathbf{g}}_i$ and corresponding image intensity covariance \mathbf{S}_{g_i} .

During image segmentation, the intersections of the current shape model instance with all image planes collectively define a stack of two-dimensional contours in \mathcal{R}^3 . The algorithm proceeds by searching for the intensity profile location along the normal to the contours and over the imaging planes for each landmark. To derive the best-matching position or candidate point \mathbf{y}_i for each landmark, we minimise the Mahalanobis distance between a profile sampled at candidate position \mathbf{y}_i , $\mathbf{g}_i(\mathbf{y}_i)$ and corresponding model $\{\bar{\mathbf{g}}_i, \mathbf{S}_{g_i}\}$ as follows:

$$\mathbf{y}_i^o = \underset{\mathbf{y}_i}{\operatorname{arg\,min}}((\mathbf{g}(\mathbf{y}_i) - \bar{\mathbf{g}}_i)^T \mathbf{S}_{g_i}^{-1} (\mathbf{g}(\mathbf{y}_i) - \bar{\mathbf{g}}_i)) \quad (2.1)$$

Given the sparse nature of CMR images, it is not uncommon during fitting to have mesh triangles that do not intersect with any image slices in the stack. In this situation, the points that comprise these triangles would not be updated or displaced by the IAM, instead, these points would be passively updated by fitting of the PDM. A mechanism that propagates displacements from points that are image-driven to nearby points that are not, is therefore necessary. SPASM implements a displacement propagation strategy modelled as a Gaussian kernel centred at any given image-driven point q by propagating its effect to a neighbouring point p based on Gaussian kernel

$$\mathbf{w}(p, q) = \exp\left\{-\frac{\|p - q\|^2}{2\sigma^2}\right\} \quad (2.2)$$

where σ is the width of the kernel. Having a non-zero Gaussian kernel is not an indispensable feature of the algorithm as non-image driven points would be indirectly updated by the PDM, nevertheless, this feature adds smoothness to the evolution of the surface mesh, and speeds up convergence of the algorithm.

Parameter Optimisation

SPASM segmentation is affected by four key parameters. We ran an exhaustive grid optimisation scheme to determine the best combination of parameters. The individual parameters and corresponding ranges that we tested were as follows:

1. Freedom of the PDM measured in standard deviations from the mean i.e. $\beta =$

2, 2.5, 3.

2. Length of the image sampling profile \mathbf{g}_i used during image feature search, measured in pixels i.e. $l = 5, 10, 15$ pixels.
3. Standard deviation of the Gaussian kernel for the point displacement propagation feature i.e. $\sigma = 5, 7, 9$ mm.
4. Image orientations to use, i.e. using only SAX images or using both SAX and LAX during segmentation i.e. $v = SAX, ALL$.

Table 2.1 shows each of the 54 (i.e. $3 \times 3 \times 3 \times 2 = 54$) unique parameter combinations we used with our algorithm to segment 50 randomly selected subjects that had already been manually delineated by clinicians. Next, we computed the segmentation accuracy using three key metrics: Dice Similarity Coefficient (DSC), Mean Contour Distance (MCD) and Hausdorff Distance (HD). These metrics were defined on Equations 1.12, 1.13 and 1.14, respectively, in Section 1.5.2.

Figure 2.2 shows three boxplots summarising the results based on the three metrics, for each of the 54 test parameter sets. The x-axis on each of the boxplots (DSC, MCD, HD) shows the test number, and the tests are sorted from best to worst performance. On Figure 2.2 it can be seen that the best performing parameter test set is test number 4, which appears at the left-most end of each of the plots. Specifically, the best parameter values were: $\beta = 2$, $l = 5$, $\sigma = 7$ and $v = ALL$. We used this parameter set for segmentation thereon.

2.2.5 Quantification

For the final phase of our workflow, we computed a thorough set of functional parameters based on blood-pool and myocardial volumes. To reproduce the reference ranges reported by Petersen et al. [36], our quantification module performs volume computations using Simpson’s method of integration, whereby a cardiac 3D volume can be approximated by summing the areas within 2D segmentation contours, and multiplying by the inter-slice spacing. Because the output of our segmentation are 3D triangular meshes, before using Simpson’s rule, we had to extract contours corresponding to the intersection between our segmentation and CMR image slices. The 3D model we use

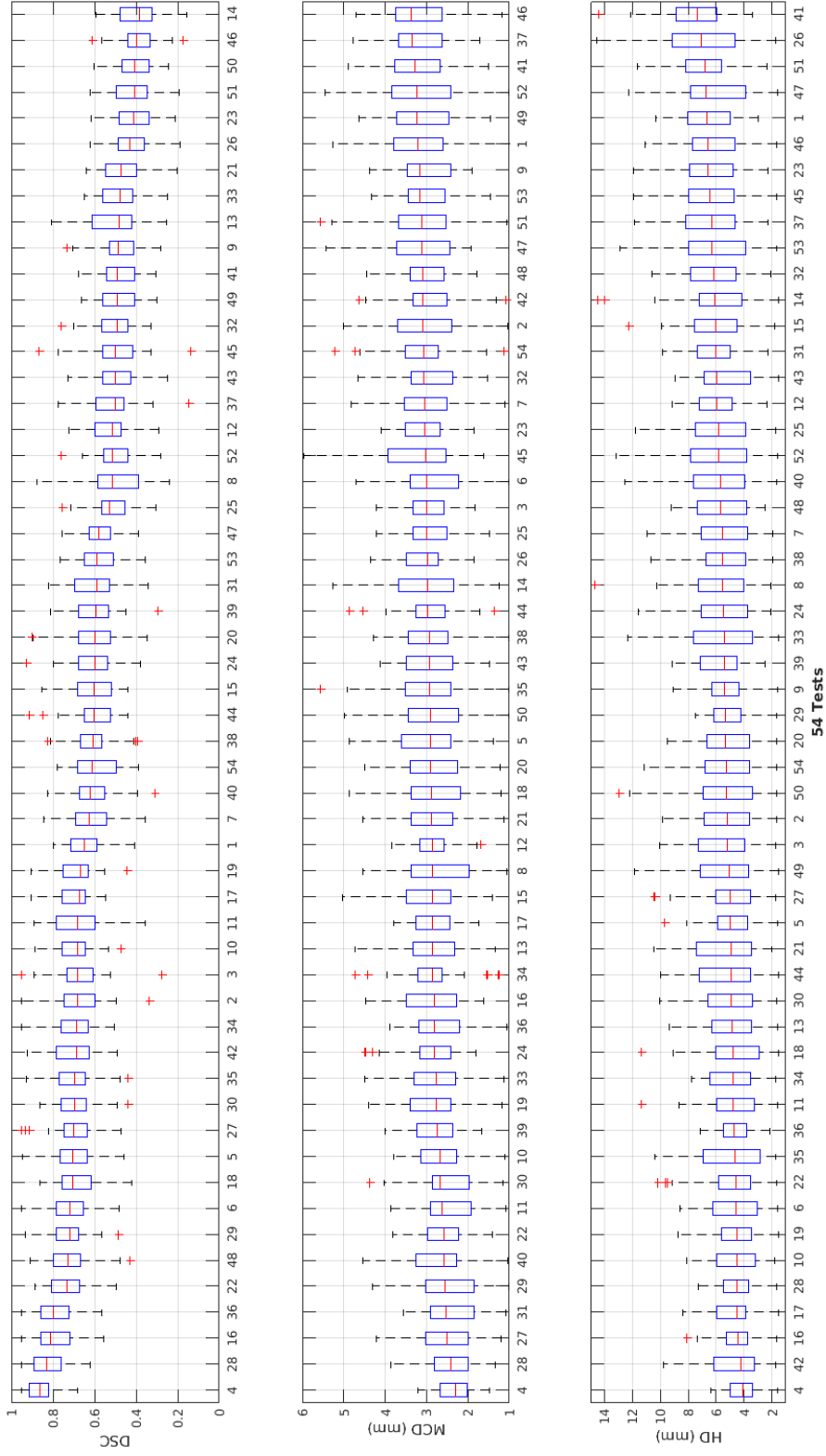


Figure 2.2: Results of our segmentation algorithm parameter optimisation. Here, the set of parameters that jointly yielded the best results for the DSC, MCD and HD metrics was test 4.

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

Table 2.1: The list of 54 distinct sets of segmentation parameters used in our segmentation algorithm parameter optimisation. As noted in the text, test 4 was the best choice.

Test	β	l	σ	v	Test	β	l	σ	v
01	2	5	5	SAX	28	2.5	10	7	ALL
02	2	5	5	ALL	29	2.5	10	9	SAX
03	2	5	7	SAX	30	2.5	10	9	ALL
04	2	5	7	ALL	31	2.5	15	5	SAX
05	2	5	9	SAX	32	2.5	15	5	ALL
06	2	5	9	ALL	33	2.5	15	7	SAX
07	2	10	5	SAX	34	2.5	15	7	ALL
08	2	10	5	ALL	35	2.5	15	9	SAX
09	2	10	7	SAX	36	2.5	15	9	ALL
10	2	10	7	ALL	37	3	5	5	SAX
11	2	10	9	SAX	38	3	5	5	ALL
12	2	10	9	ALL	39	3	5	7	SAX
13	2	15	5	SAX	40	3	5	7	ALL
14	2	15	5	ALL	41	3	5	9	SAX
15	2	15	7	SAX	42	3	5	9	ALL
16	2	15	7	ALL	43	3	10	5	SAX
17	2	15	9	SAX	44	3	10	5	ALL
18	2	15	9	ALL	45	3	10	7	SAX
19	2.5	5	5	SAX	46	3	10	7	ALL
20	2.5	5	5	ALL	47	3	10	9	SAX
21	2.5	5	7	SAX	48	3	10	9	ALL
22	2.5	5	7	ALL	49	3	15	5	SAX
23	2.5	5	9	SAX	50	3	15	5	ALL
24	2.5	5	9	ALL	51	3	15	7	SAX
25	2.5	10	5	SAX	52	3	15	7	ALL
26	2.5	10	5	ALL	53	3	15	9	SAX
27	2.5	10	7	SAX	54	3	15	9	ALL

for segmentation is comprised of two structures; the LV and the RV. The LV is a closed water-tight mesh comprising both endocardial and epicardial walls. The RV is an open mesh representing only the RV endocardium. The RV has two openings, the atrio-ventricular valve opening, and pulmonary valve opening. The LV and RV sit adjacent to each other but are not connected.

We computed both global and regional morphological and functional indexes. Global indices include chamber volumes, stroke volume, ejection fraction and myocardial mass.

Regional or local indices include myocardial wall thickness, wall motion and thickening computed for every segment in the AHA-17 cardiac subdivision scheme [45].

The global assessment of cardiac function is based on the following volumetric measurements [46]:

- End-Diastolic Volume (EDV) (ml): the volume of blood in the LV or RV before contraction. This is the highest ventricular volume of blood in the cardiac cycle.
- End-Systolic Volume (ESV) (ml): the volume of blood in the LV or RV at the end of contraction. This is the lowest ventricular volume of blood in the cardiac cycle.
- Stroke Volume (SV) (ml): the volume of blood pumped from the ventricle per beat obtained by subtracting the ESV from the EDV for a given ventricle. This term can be applied to either of the two ventricles.
- Ejection Fraction (EF) (%): the fraction of blood ejected from a ventricle of the heart with each heartbeat. This measure shows the pumping efficiency of the heart and is calculated by dividing the SV by the EDV. Note that the left ventricular EF (LVEF) is a measure of the efficiency of pumping blood into the body's systemic circulation, whereas the right ventricular EF (RVEF) is a measure of the efficiency of pumping blood into pulmonary circulation (i.e. the lungs).
- Left Ventricular Mass (LVM) (g): to compute LVM, we assume that the volume of the myocardium is equal to the total volume contained within the epicardial borders of the ventricle minus the chamber volume. Given these standard assumptions, LVM is calculated by multiplying the volume by the density of the muscle tissue (1.05 g/cm^3).

The regional assessment of cardiac function is based on the following indexes obtained from the LV myocardial shapes and computed locally based on the AHA 17-segment model. In contrast to the global indexes, where comparison with manual analysis was desired, and therefore 2D techniques were required (Simpson's rule), this segmental analysis was performed directly on 3D shapes, and using 3D techniques. Every measurement was computed on a per-point basis, and then averaged across all subjects, for every AHA-17 segment.

- LV Wall Thickness (mm): the distance between the endocardial and epicardial walls of the myocardium at ED and ES. Wall thickness may be used to quantify regional dysfunction, e.g. in myocardial ischaemia or after myocardial infarction. Myocardial thickness was measured as the average point-to-surface distance for every AHA-17 segment across the population.
- LV Wall Thickening (mm): the difference in the wall thickness measurement between ED and ES. Our models do not include papillary muscle or trabecular tissue, nor do the manual contours we compare our measurements with.
- LV Wall Motion (mm): the root-mean-squared distance between the location of mesh points at ED and ES averaged per AHA-17 region of the myocardium.

In the next section, we present and compare all the aforementioned global and regional clinical indexes obtained through manual and automatic segmentation.

2.3 Experiments and Results

We evaluated the performance of our proposed automated workflow by using common metrics for segmentation accuracy assessment (i.e. the aforementioned DSC, MCD and HD measures), comparing these measures against the ground-truth values obtained through manual delineation by clinicians and using clinical cardiac bi-ventricular functional indexes derived from manual and automated segmentations such as EDV, ESV and LVM.

We also compared our results with those reported by Bai et al. [37]. In Table 2.2, we present the data we used for training, testing and evaluating our workflow. Of the 4,870 available subjects in the UKB with manual segmentations, 250 random subjects were selected for PDM training, with 170 image volumes from a previous study by Tobon-Gomez et al. [47] used for IAM training. The remaining 4,620 subjects in the UKB with manual delineations were used as test datasets to evaluate the performance of our proposed automatic approach, labelled A_S in the table. To compare our results with those of Bai et al. [37], denoted B in the table, we used the same training and testing datasets, reporting the results as A_L in the table. As an additional assessment, we conducted a quantitative evaluation of human performance by measuring

the inter-observer variability among the segmentations performed manually by three different clinical experts. Here, we randomly selected 50 subjects; each subject was independently analysed by three expert observers labelled O1, O2 and O3. We compare segmentation results on the same set of subjects to show automated versus human performance, as well as the performance of our workflow on a larger dataset.

Input images and output segmentation contours were automatically quality controlled to ensure that input image volumes had full coverage of the heart i.e. included both basal and apical slices and to verify the quality of the output segmentations. Because our aim here is to properly evaluate segmentation accuracy, all segmentation results (including outliers) were included in the statistics in Section 2.3.1. In contrast, results presented in Section 2.3.2 are based only on good quality images and segmentations, i.e., excluding those deemed suboptimal by SQA and/or not providing full coverage by IQA.

2.3.1 Segmentation Accuracy

To quantify segmentation accuracy, we applied the three aforementioned metrics, each of which is detailed in Section 1.5.2.

Table 2.3 presents DSC, MCD and HD measures that compare automated and manual segmentation results; evaluations were performed on test sets consisting of 50, 600 and 4,620 subjects which have not been used to train the PDM or IAM. Here, the group of 50 subjects is the same set used to evaluate inter-observer variability, whereas the set of 600 subjects is the same set used as a test set by Bai et al. [37] in which a deep learning approach was used for segmentation. The large set of 4,620 subjects is all UKB cases with manual delineations that have not been used for shape and appearance model training.

In Table 2.3, the mean and standard deviations of DSC for the LV_{endo} , LV_{myo} and RV_{endo} with $n = 4,620$ were 0.93 ± 0.05 , 0.87 ± 0.05 , and 0.87 ± 0.07 , respectively, indicating excellent agreement between manual delineations and automated segmentations. We also observe that DSC measures for the LV_{myo} and RV_{endo} cases were less than that of the LV_{endo} case. One possible reason DSC values for the LV_{myo} are lower is that its annular shape has a larger perimeter (i.e. endo and epicardial edge) causing equal overlap shifts to produce greater error compared to the LV_{endo} and RV_{endo} .

Table 2.2: Specific datasets used for training and testing the methods proposed and presented in this chapter.

Label	Method	Training/Tuning Data	Test Data
B	Method by Bai et al. [37]	4,275 subjects from UKB	1) 600 subjects from UKB 2) 50 subjects from UKB
As	Our method (Small training dataset)	PDM: 250 subjects from UKB IAM: 170 subjects from [47]	1) 4,620 subjects from UKB 2) 600 subjects from UKB 3) 50 subjects from UKB
AL	Our method (Large training dataset)	PDM: 4,275 subjects from UKB IAM: 4,275 subjects from UKB	600 subjects from UKB
O1-O3	Three expert readers	Manual contours	50 subjects from UKB

Table 2.3: Segmentation results based on the different test sets (see Table 2.2) used in [37], and [36] (n=50, 600, and 4,620). The metrics used are DSC, MCD and HD. We compare manual with automatic methods, and error between human observers. M represents the manual ground-truth provided by [36]. LV_{endo} represents LV endocardium, LV_{myo} represents LV myocardium, and RV_{endo} represents RV endocardium. Table values are shown as mean \pm standard deviation.

(a) DSC									
Test-set	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	B vs M (n=50)	A _S vs M (n=50)	B vs M (n=600)	A _S vs M (n=600)	A _L vs M (n=600)	A _S vs M (n=4620)
LV _{endo}	0.94 \pm 0.04	0.92 \pm 0.04	0.93 \pm 0.04	0.94 \pm 0.04	0.93 \pm 0.03	0.94 \pm 0.04	0.93 \pm 0.05	0.94 \pm 0.04	0.93 \pm 0.05
LV _{myo}	0.88 \pm 0.02	0.87 \pm 0.03	0.88 \pm 0.02	0.87 \pm 0.03	0.88 \pm 0.03	0.88 \pm 0.03	0.87 \pm 0.04	0.87 \pm 0.03	0.87 \pm 0.05
RV _{endo}	0.87 \pm 0.06	0.88 \pm 0.05	0.89 \pm 0.05	0.86 \pm 0.07	0.87 \pm 0.06	0.90 \pm 0.05	0.88 \pm 0.06	0.89 \pm 0.05	0.87 \pm 0.07

(b) MCD (mm)									
Test-set	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	B vs M (n=50)	A _S vs M (n=50)	B vs M (n=600)	A _S vs M (n=600)	A _L vs M (n=600)	A _S vs M (n=4620)
LV _{endo}	1.00 \pm 0.25	1.30 \pm 0.37	1.21 \pm 0.48	1.08 \pm 0.30	1.28 \pm 0.39	1.04 \pm 0.35	1.21 \pm 0.36	1.06 \pm 0.35	1.18 \pm 0.41
LV _{myo}	1.16 \pm 0.34	1.19 \pm 0.25	1.21 \pm 0.36	1.18 \pm 0.31	1.20 \pm 0.34	1.14 \pm 0.40	1.23 \pm 0.48	1.13 \pm 0.35	1.23 \pm 0.50
RV _{endo}	2.00 \pm 0.79	1.78 \pm 0.45	1.87 \pm 0.74	2.20 \pm 0.92	1.79 \pm 0.80	1.78 \pm 0.70	1.80 \pm 0.80	1.74 \pm 0.61	1.80 \pm 0.69

(c) HD (mm)									
Test-set	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	B vs M (n=50)	A _S vs M (n=50)	B vs M (n=600)	A _S vs M (n=600)	A _L vs M (n=600)	A _S vs M (n=4620)
LV _{endo}	2.84 \pm 0.70	3.31 \pm 0.90	3.25 \pm 0.96	3.46 \pm 1.05	3.21 \pm 0.97	3.16 \pm 0.98	3.29 \pm 1.04	3.15 \pm 0.96	3.44 \pm 1.08
LV _{myo}	3.70 \pm 1.16	3.82 \pm 1.07	3.76 \pm 1.21	4.06 \pm 1.16	3.91 \pm 1.20	3.92 \pm 1.37	3.97 \pm 1.43	3.90 \pm 1.29	3.98 \pm 1.49
RV _{endo}	7.56 \pm 5.51	7.35 \pm 2.19	7.14 \pm 2.20	9.02 \pm 3.54	7.41 \pm 4.11	7.25 \pm 2.70	7.54 \pm 3.20	7.21 \pm 2.62	7.84 \pm 3.19

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

Further, the RV is a more challenging structure to segment compared to the LV. This is due to the sub-pixel thickness of the RV myocardium, the larger presence of trabeculations in the cavity with signal intensities similar to that of the myocardium, the more complex crescent shape of the RV, which, varies from base to apex, and considerable variability in shape and intensity of the chamber across subjects, notably in pathological cases.

Next, we observe that the MCD is 1.18 ± 0.41 mm for the LV_{endo} , 1.23 ± 0.50 mm for the LV_{myo} , and 1.80 ± 0.69 mm for the RV_{endo} , all of which are smaller than the in-plane pixel spacing range of 1.8 to 2.3 mm. The HD measures were 3.44 ± 1.08 mm, 3.98 ± 1.49 mm and 7.84 ± 3.19 mm for the LV_{endo} , LV_{myo} and RV_{endo} , respectively. Although HD measures are larger than the in-plane pixel spacing, they are still within acceptable range compared to inter-observer variability. For instance, the first three columns of Table 2.3 show inter-observer variability, where the variability between observers O1 and O2 for the HD metric is 7.56 ± 5.51 mm.

When comparing our method (i.e. A_S and A_L) with B, there was a notable difference in performance between the relatively small training set (i.e. A_S) and the same training set as that of B (i.e. A_L). In Table 2.3, we note a slight improvement of the mean and standard deviation values, particularly for MCD measures. Nevertheless, improvements become more apparent in Figure 2.3, where the number of outlying subjects was drastically reduced for A_L as compared to both B and A_S . Although the overall mean and standard deviation values remained slightly better for B, we observe in Figure 2.3 that A_L was generally more robust as it reduced the number and deviation of outlying results.

Also from Figure 2.3, we note that the performance of A_S largely agrees with the ground-truth and is comparable to the results of B. We also investigated the segmentation accuracy of the LV myocardium in detail based on the AHA 17-segment model of Heller et al. [45] to report on local segmentation accuracy in terms of DSC, MCD and HD measures between manual segmentation and automatic approaches, i.e. B and A_S on the test set of size 600. We report local segmentation accuracy in Table 2.4, which shows that B and A_S consistently performed better with mid-ventricular and apical slices, respectively; however, for base slices, the performance of B and A_S varies per region.

Note that when comparing the performance of A_S versus B ($n = 600$) in Table 2.3, B yielded slightly better global results than A_S , but in breaking down the results into specific cardiac regions (basal, mid and apical), as presented in Table 2.4 we observe that our method, A_S , consistently outperformed B for all metrics in the apical region (AHA segments 13-17). A possible reason for this is an inability of the CNN method to capture small features in the image, and the inherent ability of PDMs to infer missing or noisy image data.

To provide a visual sense of the quality of our segmentations, we defined three categories based on the mean contour distance from the gold standard, i.e., excellent ($MCD < 1$ pixel), good ($1 \text{ pixel} < MCD < 2$ pixels) and bad ($MCD > 2$ pixels). We present examples of these categories in Figure 2.4, thereby showing that automated segmentation agrees well with manual segmentation both at ED and ES; further, such agreement occurs at different slice locations (i.e. apical, mid and basal regions). Finally, Table 2.5 shows the prevalence of the different categories of segmentation quality for the different approaches presented in this chapter.

2.3.2 Estimation of Cardiac Function Indexes

In this subsection, we present our work in evaluating the accuracy of cardiac function indexes derived from automated segmentation using gold standard reference ranges derived from manual segmentations. Further, we report on analysis of all available CMR images from the UKB, which to date is 40K subjects. More specifically, we calculate the following two sets of indexes: (1) *global indexes* including LVEDV, LVESV, LVSV, LVEF, LVM, RVEDV and RVESV, RSV and RVEF; and (2) *regional indexes* including the myocardium wall thickness, thickening and motion.

Note that we report the clinical indexes obtained from automated segmentation of subjects that have successfully passed the IQA and SQA modules. Table 2.6 shows the number of subjects that were included in our analysis. For example, of the given 4,620 subjects, 4,430 were deemed of good quality after IQA and SQA analyses were applied. More specifically, IQA detected 145 subjects to exclude, whereas SQA detected 105 subjects to omit; note that 60 subjects were common to both lists. Therefore, a total of 190 subjects were automatically removed before continuing with the analysis.

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

Table 2.4: Regional segmentation accuracy of LV myocardium based on the AHA 17-segment model covering 600 subjects. Values indicate mean \pm standard deviation. **Bold** indicates the cases in which our algorithm (i.e. As) outperformed algorithm B.

	DSC		MCD		HD		
	B vs M	As vs M	B vs M	As vs M	B vs M	As vs M	
Basal	1	0.82 \pm 0.03	0.84 \pm 0.02	0.97 \pm 0.7	0.95 \pm 0.37	3.52 \pm 1.00	2.29 \pm 1.71
	2	0.86 \pm 0.03	0.82 \pm 0.03	1.10 \pm 0.48	1.28 \pm 0.36	2.86 \pm 1.11	3.30 \pm 0.92
	3	0.85 \pm 0.04	0.83 \pm 0.03	1.01 \pm 0.36	1.10 \pm 0.39	3.86 \pm 1.04	2.31 \pm 0.95
	4	0.85 \pm 0.01	0.83 \pm 0.02	0.82 \pm 0.36	0.93 \pm 0.34	3.59 \pm 1.49	2.86 \pm 1.23
	5	0.83 \pm 0.03	0.85 \pm 0.01	1.10 \pm 0.34	1.16 \pm 0.44	2.94 \pm 1.41	3.12 \pm 1.16
	6	0.86 \pm 0.01	0.85 \pm 0.03	1.07 \pm 0.45	1.01 \pm 0.42	3.45 \pm 1.28	3.28 \pm 1.02
Mid	7	0.90 \pm 0.02	0.86 \pm 0.03	0.88 \pm 0.37	0.98 \pm 0.43	2.06 \pm 1.31	3.68 \pm 1.24
	8	0.91 \pm 0.03	0.86 \pm 0.02	1.14 \pm 0.42	1.20 \pm 0.45	3.42 \pm 1.26	3.72 \pm 1.33
	9	0.89 \pm 0.03	0.87 \pm 0.02	1.04 \pm 0.31	1.08 \pm 0.38	2.63 \pm 1.30	3.80 \pm 0.93
	10	0.88 \pm 0.02	0.87 \pm 0.02	1.34 \pm 0.37	1.49 \pm 0.36	2.76 \pm 1.22	3.88 \pm 1.09
	11	0.90 \pm 0.03	0.88 \pm 0.02	1.16 \pm 0.43	1.24 \pm 0.41	2.50 \pm 1.13	3.52 \pm 0.90
	12	0.90 \pm 0.04	0.88 \pm 0.03	1.03 \pm 0.33	1.06 \pm 0.44	3.00 \pm 1.27	3.65 \pm 1.19
Apical	13	0.86 \pm 0.02	0.88 \pm 0.02	1.39 \pm 0.40	1.24 \pm 0.43	5.60 \pm 1.10	4.20 \pm 1.21
	14	0.87 \pm 0.02	0.89 \pm 0.02	1.58 \pm 0.42	1.53 \pm 0.43	5.16 \pm 1.09	4.26 \pm 1.32
	15	0.88 \pm 0.02	0.90 \pm 0.02	1.76 \pm 0.48	1.56 \pm 0.46	5.60 \pm 1.11	4.31 \pm 0.95
	16	0.89 \pm 0.03	0.91 \pm 0.02	1.83 \pm 0.43	1.59 \pm 0.40	5.64 \pm 1.15	4.71 \pm 1.24
Apex	17	0.91 \pm 0.03	0.93 \pm 0.03	2.00 \pm 0.44	1.83 \pm 0.45	5.40 \pm 1.17	4.81 \pm 1.14

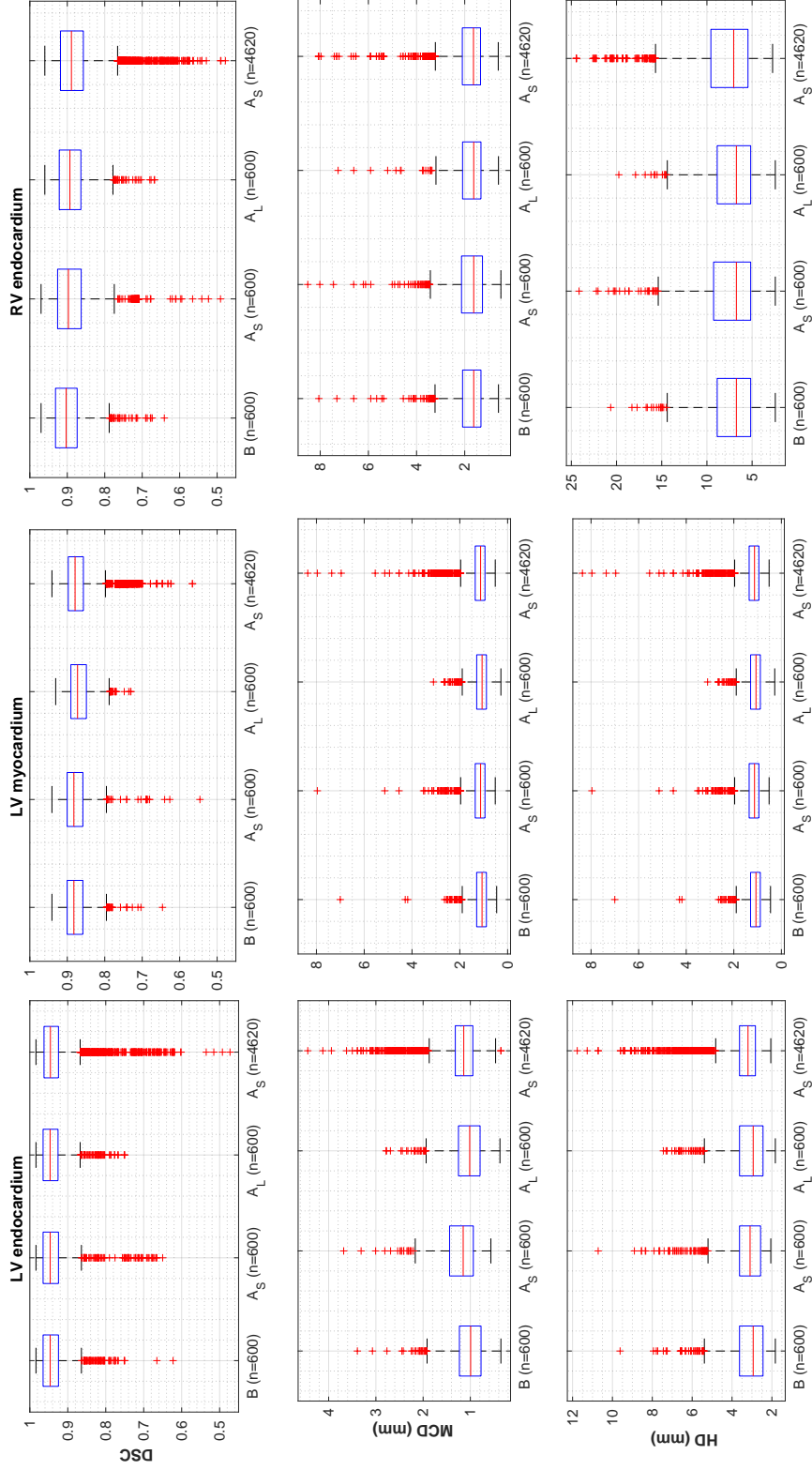


Figure 2.3: Segmentation accuracy expressed in terms of the DSC, MCD and HD measures.

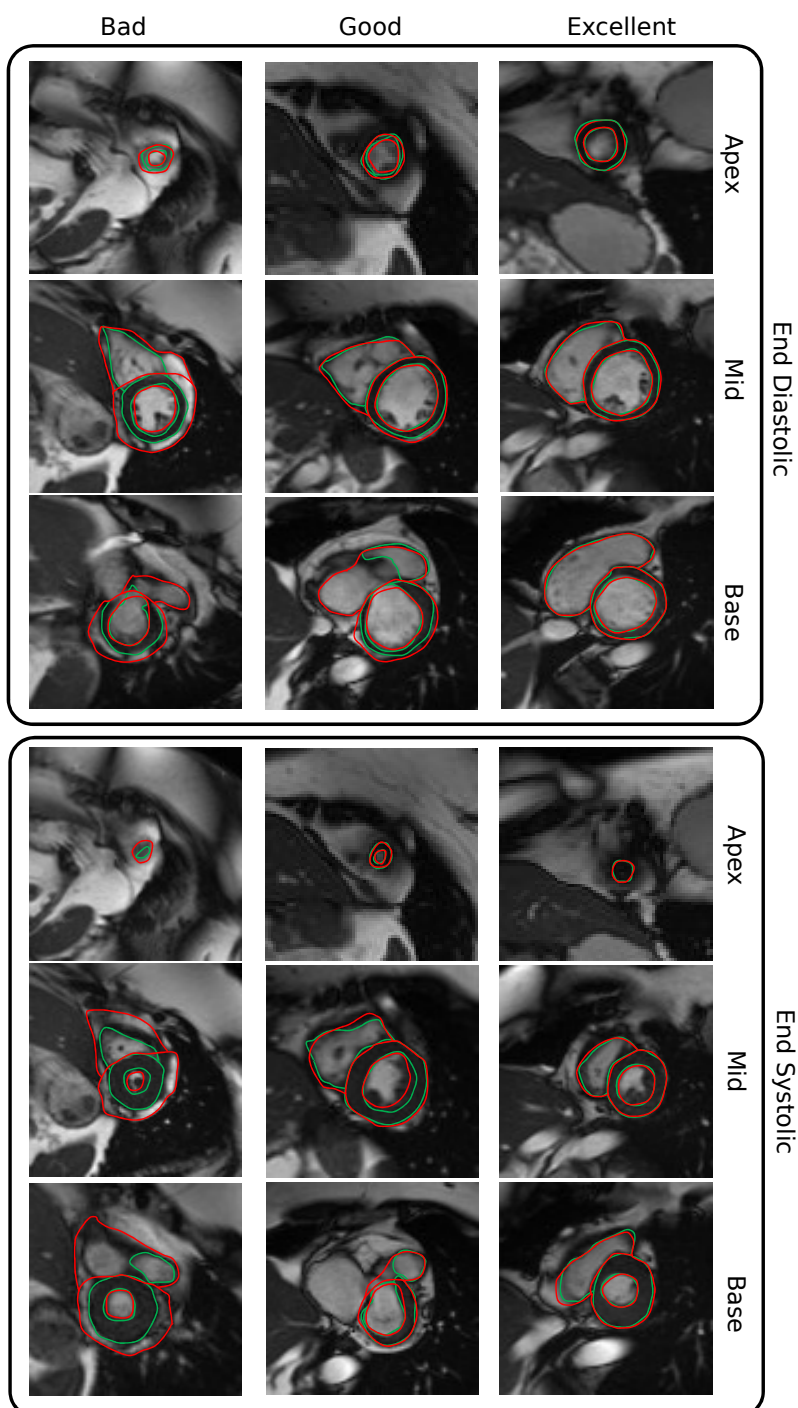


Figure 2.4: Examples segmentation results at the ED and ES phases illustrating our three degrees of quality for automated segmentation contours versus manual contours. Red: automated segmentations. Green: ground-truth segmentations.

Table 2.5: Categories of segmentation quality for the different approaches presented in this chapter.

	B (n=600)	A _S (n=600)	A _L (n=600)	A _S (n=4620)
Excellent MCD <1 pixel	84.19 %	82.14 %	84.21 %	80.67 %
Good 1 pixel <MCD <2 pixels	15.25 %	16.80 %	15.50 %	17.50 %
Bad MCD >2 pixels	0.55 %	1.05 %	0.30 %	1.82 %

Table 2.7 shows the main cardiac clinical indexes, with the two first columns representing the ventricular parameters of the healthy population obtained through automated and manual segmentations. We observe here that there was strong agreement between the two methods for computing the presented cardiac function indexes, also reported in our previous work [48].

Similarly, the computed clinical indexes for the large cohort of 4,620 subjects correlated well with the corresponding ground-truth values, as shown in columns three and four of the table; however, we note that although the mean and standard deviation values of the RV indexes for the healthy population of 800 subjects were in good agreement, for the population of 4,620 subjects, the mean and standard deviation values of the RV indexes differed slightly compared with the ground-truth values. This correlates with the larger inter-observer variability shown in Table 2.3, which is at least in part due to thinness of the RV myocardium vis-a-vis the LV [49].

Table 2.8 presents the mean absolute and relative differences between the automated and manual measurements, as well as between the automated and manual measurements computed by different expert human observers and by the built-in automated segmentation software of the scanner device (i.e. inlineVF D13A). We observe here that the absolute and relative differences for two subsets of 50 and 600 subjects matched well and were within the error range of the three expert human observers. Similarly, although the range of differences over the cohort of 4,620 subjects were not directly comparable with a small test set of only 50 subjects, the difference range still was either within that range or very close to the difference range obtained by the different expert observers. Overall, B, A_S and A_L performed substantially better than the auto-

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

Table 2.6: A summary of subjects used in our analysis after quality control measures were applied.

Datasets	Total number of subjects (n)	Detected by IQA only	Detected by SQA only	Detected by IQA & SQA	Remain for analysis
Healthy population [36]	800	0	21	0	779
All manually segmented	4,620	145	105	60	4,430
Dataset used by Bai et al. [37]	600	0	11	0	589
UKB dataset	40,000	543	450	248	39,255

Table 2.7: Summarising the differences in clinical measures derived from our proposed method and manual segmentation. Here, GT represents the ground-truth values provided by manual segmentation. Values indicate mean \pm standard deviation.

	GT (n=800)	Automated (n=800)	GT (n=4,620)	Automated (n=4,620)	Automated (n=40,000)
LVEDV (ml)	144 \pm 34	146 \pm 31	144 \pm 34	144 \pm 33	142 \pm 26
LVESV (ml)	59 \pm 18	60 \pm 18	59 \pm 20	60 \pm 23	53 \pm 14
LVSV (ml)	85 \pm 20	86 \pm 18	84 \pm 18	84 \pm 19	89 \pm 18
LVEF (%)	60 \pm 6	60 \pm 7	60 \pm 6	59 \pm 7	63 \pm 6
LVM (g)	86 \pm 24	87 \pm 23	88 \pm 23	91 \pm 23	92 \pm 18
RVEDV (ml)	154 \pm 40	154 \pm 40	152 \pm 37	160 \pm 49	165 \pm 41
RVESV (ml)	69 \pm 24	71 \pm 26	67 \pm 22	77 \pm 26	61 \pm 24
RVSV (ml)	85 \pm 20	83 \pm 21	84 \pm 18	82 \pm 24	90 \pm 27
RVEF (%)	56 \pm 6	54 \pm 7	57 \pm 6	54 \pm 11	60 \pm 9

mated segmentation obtained from the inlineVF D13A software; note that these data were retrieved for every subject from the main UKB database.

Next, in Figure 2.5, we present Bland-Altman plots (i.e. the top row of the figure) and correlation plots (i.e. the bottom row of the figure) of the ventricular parameters computed based on our proposed automated method and a manual reference covering 4,620 test subjects. The Bland-Altman plot is commonly used for analysing agreement and bias between two measurements. In Figure 2.5, the Bland-Altman plots show strong agreement and a mean difference line at nearly zero, suggesting that the clinical indexes obtained through the automated approach have little bias. Conversely, the bias between different pairs of human observers as reported by Bai et al. [37] is considerable – i.e. nearly 8 (ml) for LVEDV and LVESV, approximately 8 (g) for LVM, and approximately 15 (ml) for RVEDV and RVESV.

More specifically, Figure 2.5 presents correlation plots between the manual and automated methods for the different cardiac function indexes. The correlation coefficient (corr) measures the strength of the relationship between two sets of observations. The strength and direction of the relationship indicates the predictive power of our framework. Coefficients for all indexes ranged between 0.85 and 0.91, indicating a strong relationship between the manual and automated approaches.

To illustrate whether the values of clinical indexes computed automatically share the same distribution as those obtained via the manual approach, we visualised their

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

Table 2.8: The difference in clinical measures between the automatic and manual segmentations, as well between measurements by different human observers. M: ground truth provided by manual segmentation. VF: Automatic segmentation obtained from the automatic segmentation software inlineVF D13A. Values indicate mean \pm standard deviation.

(a) Absolute difference											
	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	B vs M (n=50)	As vs M (n=50)	B vs M (n=600)	VF vs M (n=600)	As vs M (n=600)	Al vs M (n=600)	As vs M (n=4620)	
LVEDV (ml)	6.1 \pm 4.4	8.8 \pm 4.8	4.8 \pm 3.1	4.3 \pm 4.9	5.9 \pm 4.2	6.1 \pm 5.3	12.4 \pm 18.5	7.9 \pm 9.1	6.5 \pm 5.4	9.9 \pm 7.5	
LVESV (ml)	4.1 \pm 4.2	6.7 \pm 4.2	7.1 \pm 3.8	6.5 \pm 5.4	6.8 \pm 5.1	5.3 \pm 4.9	9.2 \pm 14.8	7.0 \pm 10.0	5.1 \pm 5.0	8.2 \pm 6.3	
LVM (g)	4.2 \pm 3.2	6.6 \pm 4.9	6.5 \pm 4.8	6.4 \pm 3.5	6.0 \pm 4.4	6.9 \pm 5.5	NA	7.1 \pm 6.3	7.0 \pm 5.4	9.0 \pm 6.7	
RVEDV (ml)	11.1 \pm 7.2	6.2 \pm 4.6	8.7 \pm 5.8	8.4 \pm 6.8	10.0 \pm 5.8	8.5 \pm 7.1	NA	10.1 \pm 7.2	8.4 \pm 7.8	12.9 \pm 9.8	
RVESV (ml)	15.6 \pm 7.8	6.6 \pm 5.5	11.7 \pm 6.9	13.9 \pm 9.9	10.0 \pm 6.5	7.2 \pm 6.8	NA	8.7 \pm 9.5	7.7 \pm 6.5	12.2 \pm 9.6	
(b) Relative difference (%)											
	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	B vs M (n=50)	As vs M (n=50)	B vs M (n=600)	VF vs M (n=600)	As vs M (n=600)	Al vs M (n=600)	As vs M (n=4620)	
LVEDV	4.2 \pm 3.1	6.3 \pm 3.3	3.4 \pm 2.2	2.9 \pm 3.6	4.2 \pm 3.0	4.1 \pm 3.5	8.8 \pm 12.9	5.0 \pm 3.3	4.7 \pm 3.3	7.0 \pm 5.2	
LVESV	6.8 \pm 7.5	12.5 \pm 8.5	11.7 \pm 5.1	12.5 \pm 11.2	10.2 \pm 8.1	9.5 \pm 9.5	17.0 \pm 27.7	10.2 \pm 9.6	9.3 \pm 9.4	12.2 \pm 9.6	
LVM	4.4 \pm 3.3	6.0 \pm 3.7	6.7 \pm 4.6	8.0 \pm 4.8	6.5 \pm 4.1	8.3 \pm 7.6	NA	8.1 \pm 8.2	8.3 \pm 7.7	8.2 \pm 7.6	
RVEDV	8.0 \pm 5.0	4.2 \pm 3.1	5.7 \pm 3.6	5.7 \pm 4.3	7.3 \pm 4.2	5.6 \pm 4.6	NA	6.2 \pm 5.0	5.4 \pm 4.7	7.8 \pm 5.1	
RVESV	30.6 \pm 15.5	10.9 \pm 8.3	16.9 \pm 9.2	29.8 \pm 22.1	22.0 \pm 8.4	11.8 \pm 12.2	NA	16.1 \pm 9.7	12.4 \pm 9.0	19.4 \pm 15.0	

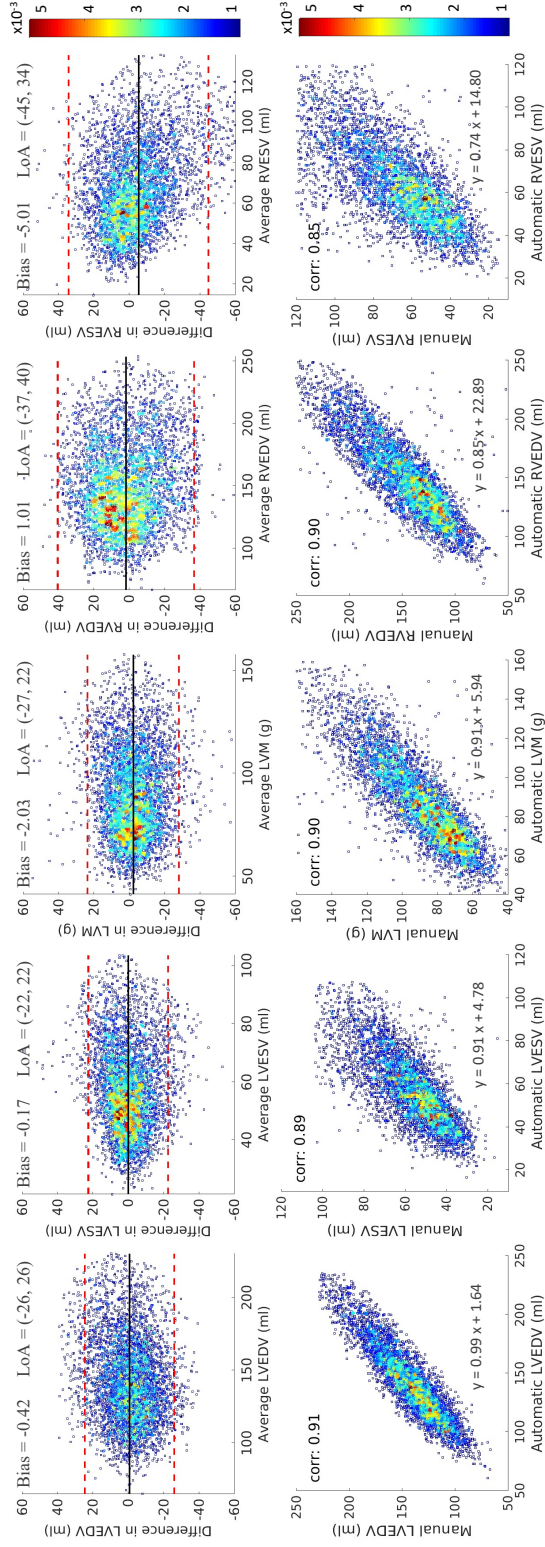


Figure 2.5: Illustrating the repeatability of various cardiac functional indexes comparing the manual and automated analysis of 4,620 subjects from the UKB cohort. The top row shows Bland-Altman plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available. The black horizontal lines denote the mean difference (i.e. bias), whereas the two red dashed lines denote limits of agreement (LoA) i.e. ± 1.96 standard deviations from the mean. The second row shows correlation plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available.

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

distributions. In Figure 2.6, we present probability distribution plots (i.e. the top row of the figure) and Q-Q plots (i.e. the bottom row of the figure) for various cardiac functional indexes computed both manually and automatically over the full cohort for which manual segmentations were available. From the plots, we observe that the distribution of the various indexes closely match those obtained from the manual segmentations—More specifically, we observe a common distribution, common location and scale, similar distributional shapes, and similar tail behaviour.

Because ground-truth manual regional (AHA-17) quantification for the subjects in this study was not available, all AHA-17 regional indexes reported in this chapter are computed using 3D techniques, in contrast to the global quantification indexes, where direct comparison with manual assessment was desirable. Nevertheless, in order to approximate a comparison with what would be a regional analysis derived from manual delineations, we generated 3D shapes by non-rigid registration of a model to all manual delineations. We used the resulting 3D shapes to perform regional quantification, and compared with our automatic results.

We computed the regional LV myocardial wall parameters in terms of thickness, thickening, and motion. Visual results can be seen on Figure 2.7, and corresponding numerical results on Table 2.9. Figure 2.7 shows the mean and standard deviation values of the regional analysis of 4,620 subjects for both the automated and manual approaches in a bulls-eye display based on the AHA 17-segment model. We observe here that the (top and bottom) panels are similar in most regions in terms of the mean and standard deviation values, thereby confirming the quality of our fully automated pipeline. Indeed, results already published in many clinical journals [50–57], primarily based on the manual delineation of a few dozen images confirm the values and ranges we have obtained and present in our bulls-eye plots.

Figures 2.8 and 2.9 show the distribution of wall thickness, thickening and motion for all AHA-17 segments in the LV myocardium. These histograms show measurements obtained from the automated segmentation applied to two cohorts (i.e. $n=4,620$ and $n=40,000$), as well as from manual delineations. The figures show excellent agreement between measurements obtained from automated segmentations from both cohorts and those derived from manual delineations.

We also performed two-sample Kolmogorov-Smirnov (K-S) tests to verify that ventricu-

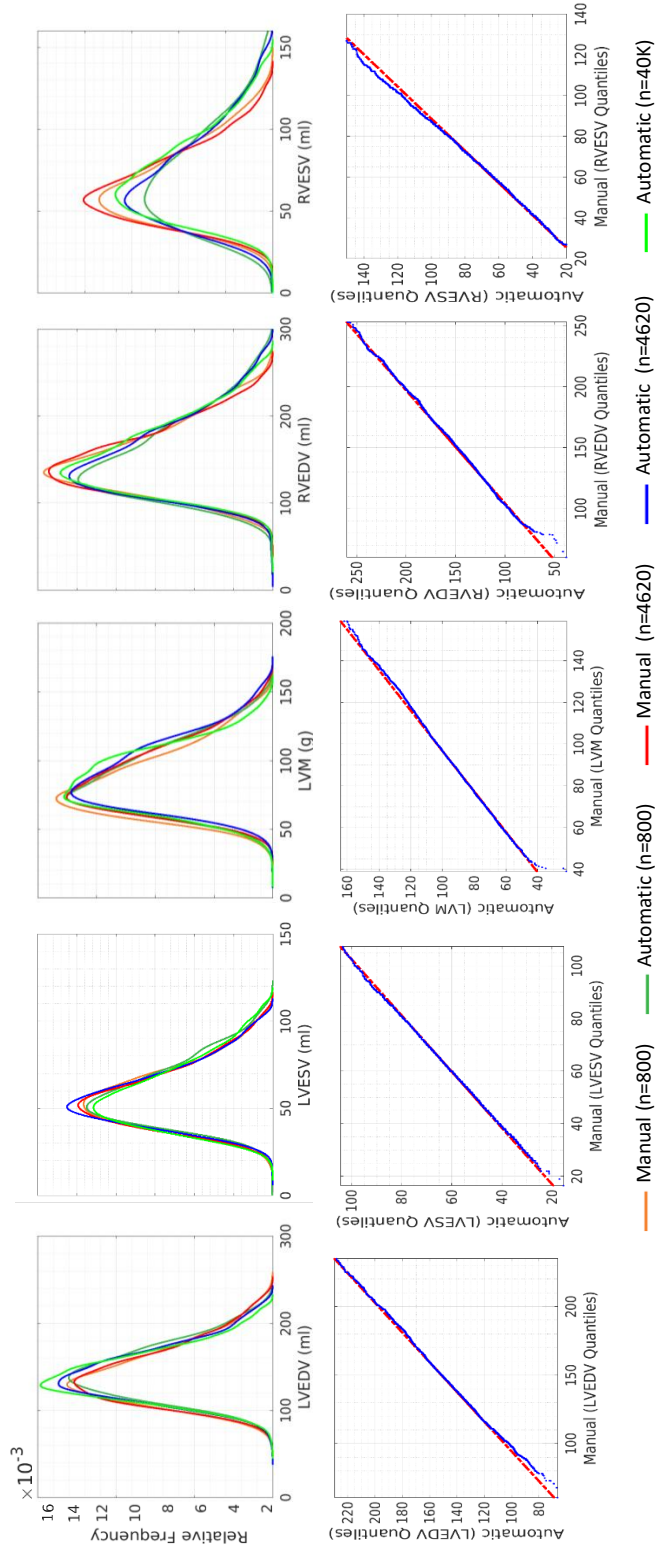


Figure 2.6: Distributions of various cardiac functional indexes comparing results of manual and automatic analyses of 4,620 subjects. The top row shows probability distribution plots, whereas the bottom row shows Q-Q plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available.

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

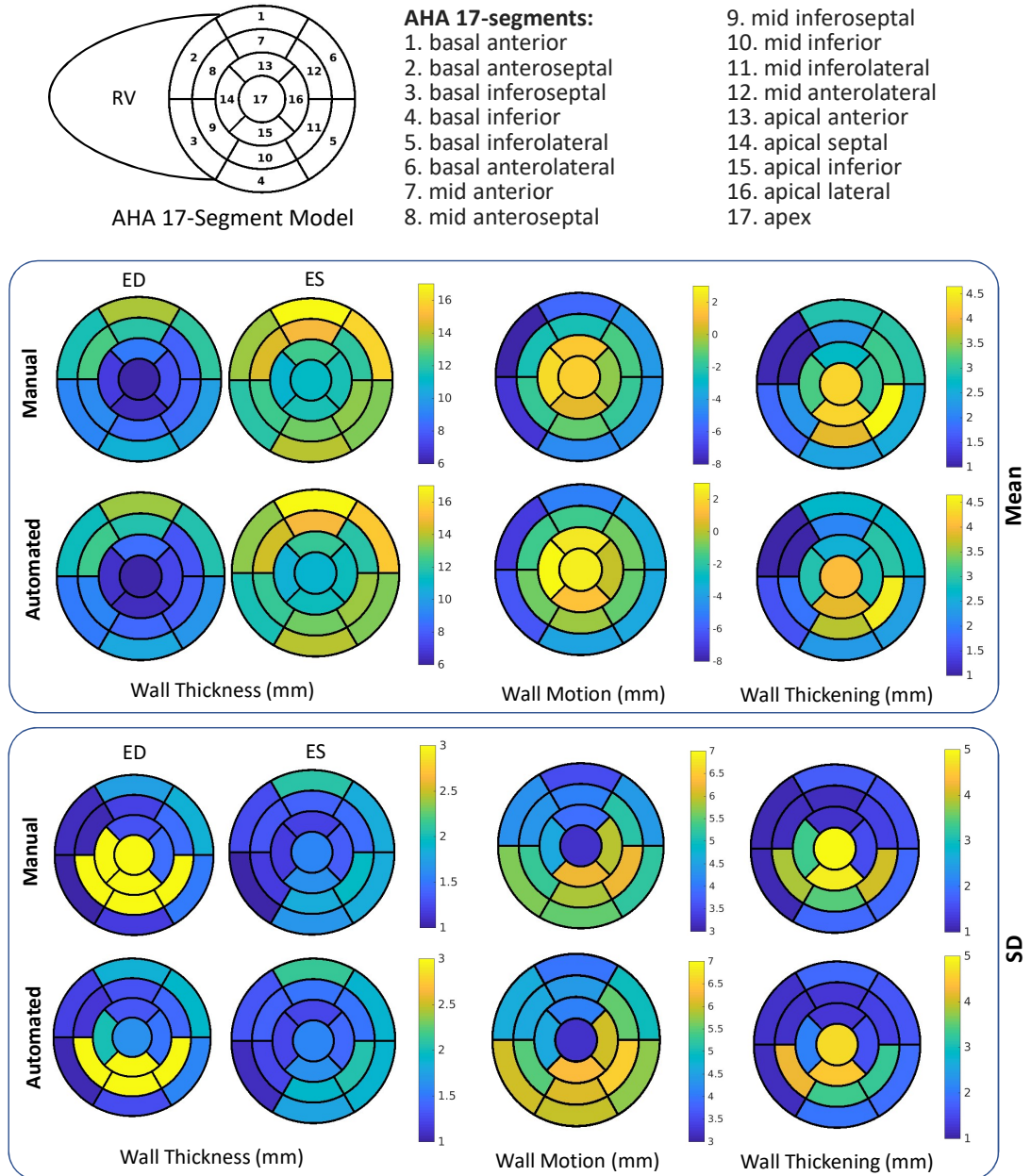


Figure 2.7: Segmental LV parameters of 4,620 subjects presented as bulls-eye displays.

lar parameters obtained through manual and automated approaches are drawn from the same distribution, under the null hypothesis that the manual and automatic methods are from the same continuous distribution in terms of clinical indexes. From our

Table 2.9: Segmental LV parameters of 4,620 subjects obtained from manual and automatic approaches. Upper rows correspond to shapes generated from the manual segmentation and lower rows to those obtained with the automatic approach.

ID	Wall Thickness at ED (mm)	Wall Thickness at ES (mm)	Wall Motion (mm)	Wall Thickening (mm)
1	12.05 ± 1.84	15.69 ± 1.83	-4.05 ± 4.39	2.95 ± 1.61
	11.72 ± 1.92	15.59 ± 1.88	-3.46 ± 4.94	2.67 ± 1.66
2	13.81 ± 1.73	17.49 ± 2.07	-5.70 ± 3.51	2.89 ± 1.74
	13.59 ± 1.85	17.49 ± 2.14	-4.91 ± 3.97	2.69 ± 1.81
3	11.63 ± 1.09	13.57 ± 1.26	-8.01 ± 4.28	1.08 ± 1.10
	11.38 ± 1.20	13.47 ± 1.33	-7.06 ± 4.63	0.87 ± 1.19
4	9.17 ± 0.97	11.87 ± 1.00	-7.20 ± 5.68	1.76 ± 1.08
	8.87 ± 1.06	11.65 ± 1.06	-6.20 ± 6.03	1.55 ± 1.18
5	10.76 ± 1.17	14.00 ± 1.67	-4.60 ± 5.53	2.41 ± 1.86
	10.45 ± 1.27	13.96 ± 1.77	-3.80 ± 5.96	2.29 ± 1.95
6	10.02 ± 1.50	13.35 ± 1.81	-4.29 ± 5.21	2.52 ± 1.82
	9.69 ± 1.58	13.26 ± 1.89	-3.60 ± 5.72	2.35 ± 1.87
7	8.14 ± 1.46	12.14 ± 1.44	-1.44 ± 5.18	3.15 ± 1.41
	7.81 ± 1.43	11.96 ± 1.48	-1.31 ± 5.50	2.94 ± 1.42
8	11.99 ± 1.21	15.03 ± 1.40	-2.33 ± 4.06	2.25 ± 1.18
	11.68 ± 1.31	14.95 ± 1.46	-1.57 ± 4.40	2.04 ± 1.26
9	12.63 ± 1.04	14.51 ± 1.32	-2.35 ± 4.37	1.01 ± 1.12
	12.37 ± 1.15	14.41 ± 1.40	-1.52 ± 4.62	0.84 ± 1.20
10	9.33 ± 3.59	12.43 ± 1.17	-1.78 ± 5.26	2.25 ± 3.83
	9.17 ± 4.16	12.28 ± 1.24	-0.93 ± 5.46	1.88 ± 4.23
11	8.53 ± 3.12	13.09 ± 1.82	-1.02 ± 5.85	3.85 ± 3.49
	8.25 ± 3.07	13.11 ± 1.91	-0.84 ± 6.11	3.66 ± 3.33
12	8.17 ± 3.76	13.55 ± 1.94	-1.72 ± 6.23	4.65 ± 4.00
	7.90 ± 3.14	13.59 ± 2.03	-1.09 ± 6.89	4.50 ± 3.30
13	7.58 ± 1.45	11.57 ± 1.37	-0.58 ± 5.91	3.10 ± 1.50
	7.32 ± 1.49	11.40 ± 1.42	-0.39 ± 6.05	2.86 ± 1.56
14	8.75 ± 1.30	12.40 ± 1.19	1.56 ± 3.91	2.75 ± 1.22
	8.46 ± 1.33	12.21 ± 1.23	2.32 ± 4.10	2.54 ± 1.28
15	6.86 ± 3.17	10.89 ± 1.17	2.07 ± 4.59	3.20 ± 3.36
	6.59 ± 2.02	10.70 ± 1.21	2.93 ± 4.44	2.91 ± 2.09
16	6.48 ± 4.66	11.58 ± 1.63	0.69 ± 6.22	4.29 ± 4.83
	6.37 ± 4.44	11.46 ± 1.63	0.41 ± 6.68	3.90 ± 4.49
17	5.97 ± 4.49	11.17 ± 1.57	1.62 ± 3.21	4.28 ± 5.11
	5.71 ± 3.63	10.94 ± 4.68	2.58 ± 3.18	4.03 ± 1.68

analysis, K-S test results on different global and regional indexes do not reject the null hypothesis of being from the same distribution at the 5% significance level.

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

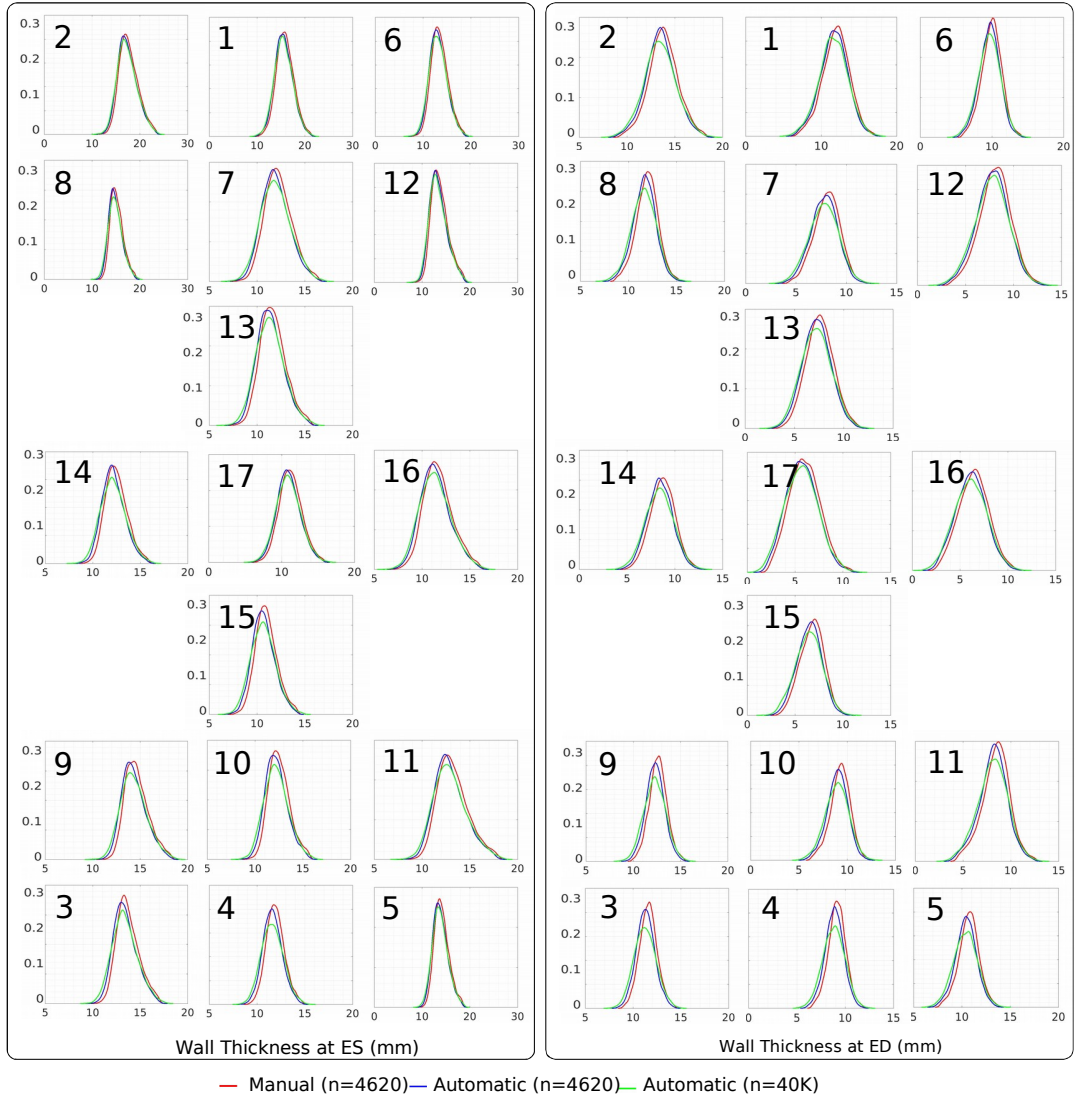


Figure 2.8: Regional analysis of LV shapes covering 40,000 subjects in terms of distribution of wall thickness at ED and ES phases. Here, red, blue and green lines indicate ground-truth values for 4,620 subjects, automated values for 4,620 subjects and automated values for 40,000 subjects, respectively. In all plots, the y-axis represents the relative frequency.

An important final note is that although our image parsing implementation performs fully in 3D, to ensure a fair comparison with both ground-truth data and the methods we compare with in this chapter, we had to convert our segmentation results

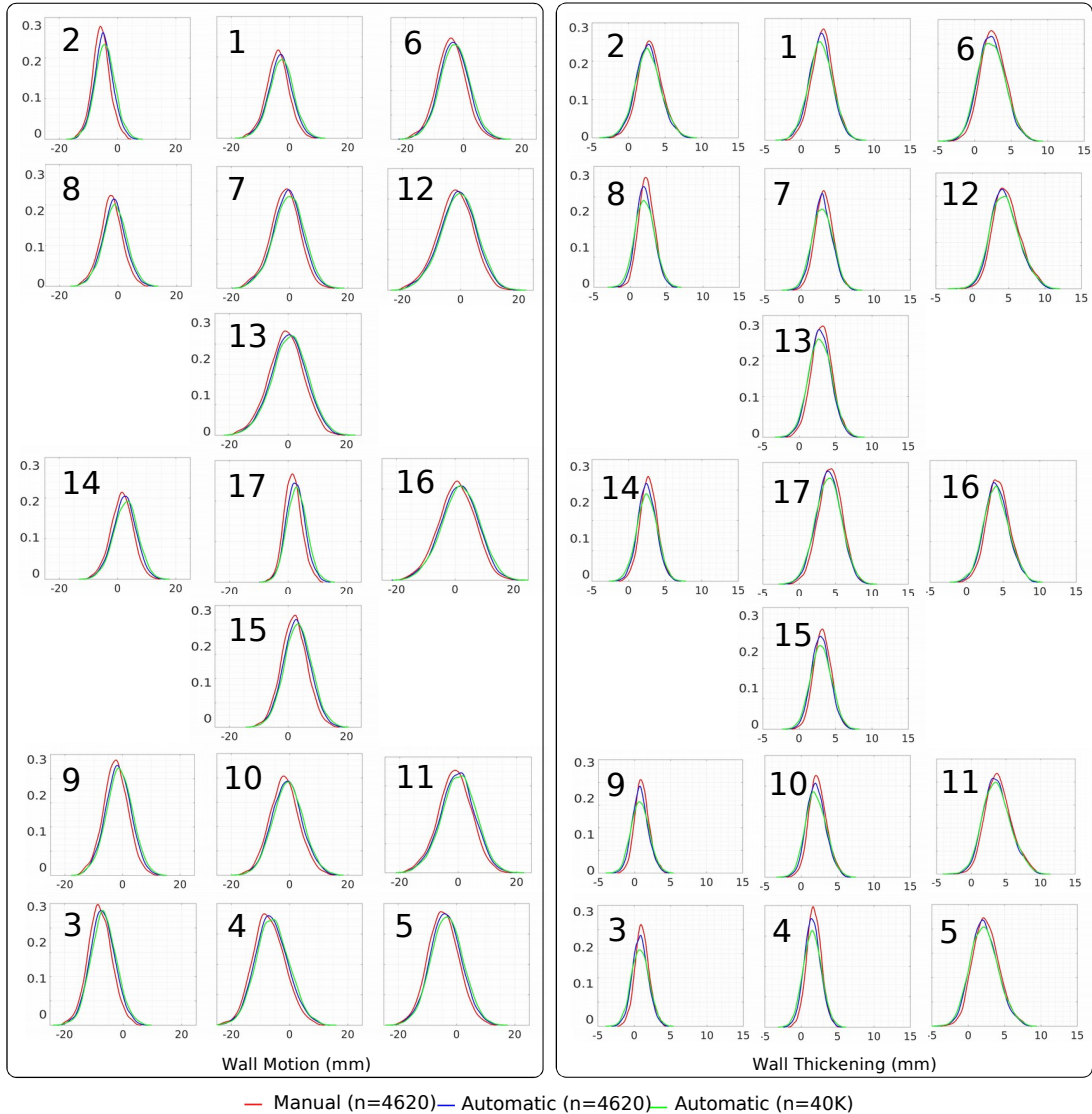


Figure 2.9: Regional analysis of LV shapes covering 40,000 subjects in terms of distribution of wall motion and thickening. Here, red, blue and green lines indicate ground-truth values for 4,620 subjects, automated values for 4,620 subjects and automated values for 40,000 subjects, respectively. In all plots, the y-axis represents the relative frequency.

to 2D contours from 3D meshes; this does not pose a problem for objective quantification of segmentation accuracy, however, given the sparse nature of CMR images, where voxel resolution along the z axis is typically on the order of 10mm , gross miscalculations may occur when approximating volumetric measurements such as ventricular

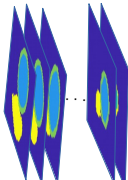

volumes and myocardial masses via simple integration methods such as Simpson’s rule. We believe that although many CNN-based methods have recently received a lot of attention, showing the capacity for image texture characterisation, most of them are restricted to handling 2D data. This simplification can introduce large biases in volume computations, and be less resilient to image artefacts such as those caused by breathing motion. In addition to our pipeline approach fully supporting 3D data, our method provides other advantages when compared to 2D CNN-based implementations. More specifically, the size of the training dataset required to achieve similar performance for an equal task differs by at least one order of magnitude between CNNs and ASM-based methods. Further, ASM implementations such as SPASM have the inherent ability to handle multi-view image volume segmentation without the need to retrain. This is particularly useful for functional CMR segmentation in which multiple views of the heart are captured as part of standard analysis protocols.

In addition, because the output of our segmentation are 3D meshes, more apt mathematical formulations can be used for volumetric computation, i.e. Green’s theorem for surface integration, and any further higher level structural analyses of the cardiac tissue. Some CNN-based methods such as those proposed by Zheng et al. [49] do take into account inter-dependencies between short-axis slices potentially resulting in more robust segmentations, even so, such CNN-based algorithms are still not globally constrained, their output is typically two dimensional in nature, their training is very costly both in time and sample size requirements, and they cannot handle dynamically changing input image views without redefinition of the architecture and re-training. We present the key differences between our implementation and the 2D CNN-based implementation method by Bai et al. [37] on Table 2.10.

2.3.3 Hardware and Computational Cost

In terms of computational cost of training and testing, method B takes approximately 10 hours to train the VGG-16 network on a Nvidia Tesla K80 GPU, and about 11 seconds to segment all 2D slices of a full cardiac cycle for one subject [37]. For our method, it takes approximately 30 minutes to train both the PDM and IAM on a Intel Xeon(R) CPU E5-1620 @3.60GHz with 32 GB of RAM, and about 15 minutes to generate the 3D shapes of a full cardiac cycle for one subject. Finally, the total

Table 2.10: Comparing one of the current state-of-the-art CNN-based methods proposed by Bai et al. with our proposed framework.

	Bai et al. [37]	Proposed pipeline
Output	 2D masks $N \approx a \times 10^3$	 3D surface mesh $N \approx a \times 10^2$
Size of training dataset (N) for equal performance		
Image slice dimensions	Must match training (cropping or re-sampling required)	Independent to image size
Image view (SAX, LAX)	Must be consistent, i.e. SAX	Independent to image view
Slice stack inter-dependency	Each slice processed independently	Slice stack handled by PDM
Training computational cost	High*	Low*
Testing computational cost	Low*	High*

*For more details see section 2.3.3.

end-to-end execution time for the 40,000 subjects using our MULTI-X platform was performed using 50 Amazon Web Service (AWS) "m4.10xlarge" machines each with 40 2.4-GHz Intel Xeon ES-2676 v3 vCPUs, and 160 GB of RAM.

2.3.4 Sub-Cohort Analysis

Thus far in this chapter we have only shown a global population analysis of the UKB. We have presented statistics on the most commonly used clinical indexes derived from CMR exams. With the exception of the "healthy" group as defined by Petersen et al. [36], introduced in this chapter on Table 2.6, and corresponding quantification results shown on the first two columns of Table 2.7, we have only presented global population statistics. We believe however that the power of population studies lies in the opportunity to define and characterise human sub-populations.

Though in this chapter our principal aim is to present the first fully automatic large-scale, global and segmental, 3D analysis of this magnitude we have included some preliminary quantification results on UKB sub-populations in this section. Based on the 40,000 subjects available, we have used patient age at the time of imaging, and patient gender (male, female), to present cardiovascular index reference ranges for these cohorts. Table 2.11 presents the arithmetic mean, and upper/lower bounds of the 95% prediction interval for each clinical index, and each age group. Each of the three age-groups span a 10-year interval, and the total age range includes patients 45 to 74 years old. Also, for each clinical index, and age-group we compute separate statistics for males and females.

Figure 2.10 shows the mean value for each of the five clinical indexes, for the three different age groups, and for males and females. Perhaps the most evident, and in some ways expected feature of these plots, is the consistent decline in cardiac volumes and cardiac mass with ageing. For the five indexes LVEDV, LVESV, LVM, RVEDV and RVESV, we see a decline of 9%, 15%, 7%, 8%, and 13% for males, and 11%, 17%, 5%, 6%, and 11% for females. As stated before, a deep analysis of sub-populations is out of the scope of this chapter, nevertheless, we hope to have shown the potential of the techniques presented in this chapter to gain insight from large population imaging studies.

Table 2.1.1: Male (M) and Female (F) ventricular reference ranges detailing mean, lower reference limit and upper reference limit by age group. Reference limits are derived by the upper and lower bounds of the 95% prediction interval for each parameter at each age group.

Age groups (years)		45-54			55-64			65-74		
		lower	mean	upper	lower	mean	upper	lower	mean	upper
Male gender (%)		43%			43%			52%		
LVEDV (ml)	M	109	170	231	102	163	223	94	154	213
	F	95	140	184	88	132	175	80	124	168
LVESV (ml)	M	31	81	130	28	75	122	25	69	113
	F	25	58	91	21	53	85	15	48	82
LVM (g)	M	71	112	152	69	109	148	66	104	142
	F	44	78	111	43	76	108	42	74	107
RVEDV (ml)	M	115	197	279	112	191	269	105	182	259
	F	73	145	218	72	141	210	72	136	200
RVESV (ml)	M	38	102	144	34	96	137	31	89	127
	F	16	66	116	15	63	110	14	59	105

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

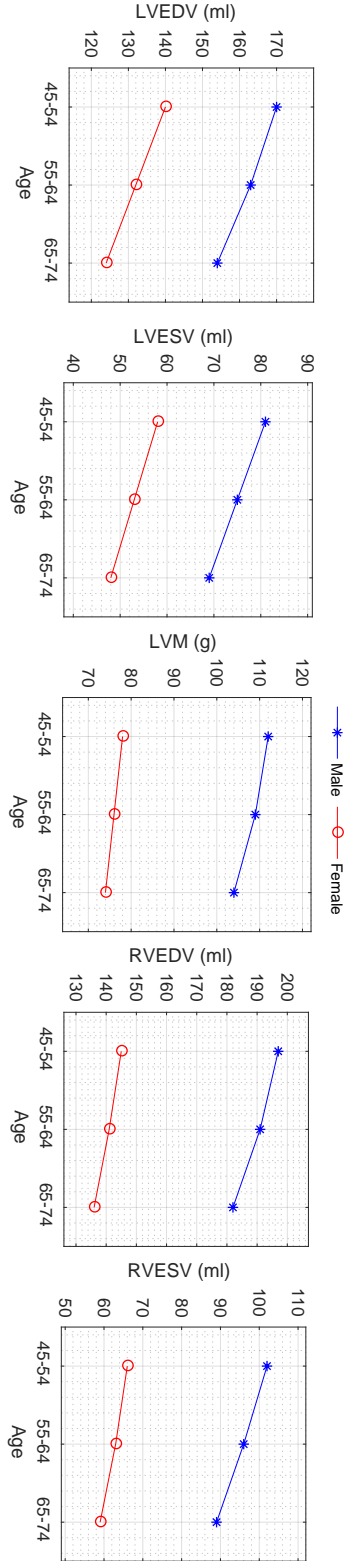


Figure 2.10: Male (blue star marker) and Female (red circle marker) clinical indexes showing their mean value per age group.

2.4 Summary

In this chapter, we presented a fully automatic framework capable of performing high-throughput end-to-end 3D cardiac image analysis of 40,000 subjects. We validated our workflow on a reference cohort of 4,620 subjects for which both manual delineations and reference functional indexes exist. Our results show that differences between our automatic workflow and the manually obtained global and regional reference indexes are within the expected variability observed in human raters. The method is fast and scalable, overcoming limitations associated with current clinical CMR image analysis routine, which is manual, time-consuming and prone to subjective errors. This pipeline has remarkable potential for improving work efficiency and assisting clinicians in diagnosing and performing large-scale clinical research. Our proposed framework includes quality analysis modules designed to detect possible erroneous results for review, allowing fully automated analysis of CMR images. Additionally, this would help decrease inter-and intra-observer variabilities, which are unavoidable when such analyses are performed manually, thereby streamlining the overall process.

Furthermore, the 3D results can provide more comprehensive anatomical and functional details because of the additional dimension compared with all the 2D segmentation methods. This proposed approach ensures the global coherence of the cardiac anatomy and naturally provides detailed surface models for applications in which full 3D anatomy is necessary; for example, in mechanical and flow simulations, intervention planning and subsequent image-guided interventions could benefit significantly from the availability of high-quality 3D mesh models. The analysis and interpretation of these spatio-temporal information from the cardiac structure and function in large-scale population imaging data can help find and understand patterns and trends across population groups, and consequently, reveal insights into key risk factors before CVDs fully develop.

In the next chapter, we will present a novel deep neural network using both CMR images and patient metadata to directly predict cardiac shape parameters. The proposed method uses the promising ability of statistical shape models to simplify shape complexity and variability together with the advantages of convolutional neural networks for the extraction of solid visual features.

2. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC VENTRICLES

CHAPTER 3

Quantitative Population Analysis of Cardiac Chambers

Accurate 3D modelling of cardiac chambers is essential for clinical assessment of cardiac volume and function, including structural, and motion analysis. Furthermore, to study the correlation between cardiac morphology and other patient information within a large population, it is necessary to automatically generate cardiac mesh models of each subject within the population. In this chapter, we introduce MCSI-Net (Multi-Cue Shape Inference Network), where we embedded statistical shape models inside a fully convolutional neural network to jointly learn phenotypic and demographic information from the cohort. In this way, we leveraged the ability of the network to learn the appearance of cardiac chambers in cine cardiac magnetic resonance (CMR) images, and generate plausible 3D cardiac shapes, by constraining the prediction using a shape prior, in the form of the statistical modes of shape variation learned a priori from a subset of the population. This, in turn, enabled the network to generalise to samples across the entire population. The motivations for utilising deep neural networks for this application are three-fold: 1) it allows the method to extract much more solid

visual features at each network layer and incorporate spatial context from neighbouring regions by using all available image views. 2) it has superior accuracy to reference shapes than the previous methods while being on average about 30 times faster during inference, producing the nearly real-time output either directly used as the final result or further input into other systems. 3) it allows multiple input types, e.g. imaging and non-imaging; different types of features out of different types of data can be combined for more informative distribution of parameters. We show that including this information provides the network with a variable prior by learning the likely distributions of shape parameters across different populations. To the best of our knowledge, this is the first work that uses such an approach for patient-specific cardiac shape generation. MCSI-Net is capable of producing accurate 3D shapes using just a fraction (about 23% to 46%) of the available image data, which is of significant importance to the community as it supports the acceleration of CMR scan acquisitions. We trained and evaluated the MCSI-Net on a large-scale dataset of 3,925 and 600 subjects, respectively, from the UK Biobank. We also present the results from analysing 40,000 subjects of the UK Biobank at 50 time-frames, totalling two million image volumes. Our results broadly show strong agreement with the reference annotations, achieving an average Dice score of 92.5% across cardiac ventricles and atria.

3.1 Introduction

According to the World Health Organisation [3], cardiovascular disease (CVD) is the most prevalent cause of death worldwide, accounting for nearly 18 million deaths each year. Identifying individuals at risk of CVDs and ensuring they receive appropriate and timely treatment can help prevent premature deaths.

Early quantitative assessment of cardiac structure, motion, and function support preventive care and early cardiovascular treatment. Therefore, fully automated analysis and interpretation of large-scale population-based cardiovascular magnetic resonance (CMR) imaging studies is of high importance. This analysis helps to identify patterns and trends across population groups, and accordingly, provides insights into key risk factors before CVDs fully develop.

UK Biobank (UKB) is currently the world’s most extensive prospective population

study [34], which contains questionnaire data, biological samples, physical measurements, CMR images, and so forth [13]. CMR is an essential element of multi-organ, multi-modality imaging visits for patients in multiple dedicated UKB imaging centres that will acquire and store imaging data from 100,000 participants by 2022. At this time, CMR scans of 40,000 subjects have been released and are available for health-related research.

We believe that 3D analysis is critical for the accurate clinical assessment of cardiac function. In this chapter, we introduce a new approach that ensures the global coherence of cardiac anatomy and naturally lends itself to any further analysis requiring the full 3D anatomy; for example, in interventional treatment planning requiring precise volumetric quantification, mechanical and flow simulations, motion analysis, and modelling the associations between cardiac structure and patient metadata (such as socio-demographic, lifestyle and environmental factors, or family history, genetic, and omics data). Though fully automatic 3D segmentation is required to facilitate such analyses, the complexity of anatomical structures, intensity and morphology variation across a population cohort, and the sparse information available from CMR images (typically on average around 12 image slices covering the full heart) make this task challenging.

In chapter 2, which already published as [48] and [58], we showed that 3D statistical shape model-based approaches have the power and potential to automatically segment cardiac structures, and generate associated cardiac function indexes. This success is attributed to the inclusion of prior knowledge of cardiac shape, within the segmentation method. These segmentation approaches typically use simple sets of features to fit a shape model through an iterative process and the goal is to minimise the Mahalanobis distance between an intensity profile sampled at a candidate position and its corresponding intensity appearance model, by deforming the shape within its range of normal variation to match the image data.

On the other hand, in the last decade, fully convolutional networks (FCN) have shown great potential in image-based pattern recognition in a variety of tasks, including cardiac segmentation. However, their output results are, by nature, 2D segmentation masks for every short axis (SAX) and long-axis (LAX) CMR slice. Although these 2D masks are sometimes extended via a further step of non-rigid registration to a 3D

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

atlas to produce a 3D cardiac shape [59], this is not efficient for learning topological shape information. Furthermore, this is based on the strong assumption that all the 2D segmentation masks are always correct and meaningful, however, in practice, there are often errors in pixel-wise segmentation approaches due to spurious false positives. For instance, Painchaud et al. [60] recently proposed a generative model based on a variational autoencoder to identify anatomically implausible results following 2D segmentation, and corrected these to fit the closest anatomically correct contours, based on the learned latent space. This provides further evidence that, conducting CMR segmentation in 2D requires a subsequent quality control step, in order to ensure that the downstream quantitative analyses are accurate. Consequently, large-scale studies would benefit from an efficient approach for reconstructing cardiac shapes in 3D, as it would remove the need for multiple sequential steps involving pixel-wise 2D segmentation, followed by quality control, and iterative registration-based 3D shape reconstruction, which can be cumbersome and time-consuming.

In order to address these limitations, in a recent study [61] published at MICCAI 2019, we proposed an approach to exploit image features obtained using deep FCNs trained on both SAX and LAX views, along with a rich shape prior learned using a statistical shape model, to directly predict the shape-space parameters required to reconstruct 3D cardiac shapes. Another significant aspect of this study was the integration of patient metadata into the process of shape prediction using a Multilayer Perceptron (MLP). This information, which is currently ignored by most cardiac segmentation or shape generation techniques, has been shown in different clinical studies to have an impact on cardiac morphology and structure [62]. To evaluate our method, in addition to comparing against manual measurements, we also compared our performance against two state-of-the-art methods that have reported the highest accuracy thus far, namely, the work by Bai et al. [37] in which the authors propose a 2D convolutional neural network (CNN)-based segmentation method and our previous work [58] presented in chapter 2 where, we analysed and reported cardiac functional indexes of 40,000 subjects of the UKB through a fully automatic quality-controlled image parsing framework.

The main contributions of this chapter are as follows:

- We extend the shape model from biventricular to four-chamber cardiac mesh

model. We segment all four cardiac chambers; namely, left ventricle (LV), right ventricle (RV), left atrium (LA), and right atrium (RA). For the LV, we segment both the endocardium and epicardium surfaces; while for the other chambers, we only segment the endocardium. This is because the myocardium is too thin to reliably distinguish epicardium from the endocardium. Consequently, there is no label available to use for the training. This was achieved by first generating a reference cohort of four-chambered cardiac meshes through the non-rigid registration of a four-chambered cardiac atlas, to a set of 3D points obtained from manual delineations. To ensure high accuracy when fitting the 3D shape to the stack of manually delineated contours, we adopted an alternating global-local non-rigid registration approach, using the Coherent Point Drift (CPD) method [63].

- We propose an innovative end-to-end deep neural network that directly predicts 3D shape parameters derived from a Principal Component Analysis (PCA) space. The network is optimised using a loss function defined in the domain of shape space parameters which weights each PCA mode of variation independently, prioritising the more significant modes and leading to more accurate shape prediction. In addition to the shape parameters, the network also learns the similarity transformation parameters required to transform the generated shape back from the normalised PCA space, to the image space. We achieved this through a multi-task learning approach, where the extracted features are used to jointly optimise a second loss function defined over the desired transformation parameters.
- We utilise a novel approach of exploiting all the CMR image views (short axis and long axis two-, three- and four-chamber CMR images) and a comprehensive range of patient metadata simultaneously, to predict 3D four-chambered cardiac shapes. The introduction of the metadata yielded a substantial positive impact on shape prediction with about 5% average improvement across all metrics. We hope this idea inspires other researchers to exploit such informative priors in their applications, to improve the performance of their models. We categorised the available patient metadata into four main groups, i.e. demographics, blood pressure, biological samples, and lifestyle. We carried out systematic experiments to understand and report the impact of different metadata categories on the

predicted cardiac shape.

- We investigate the importance of the available CMR slices towards the accuracy of the final predicted shape and evaluate the performance of our approach in scenarios where only a few SAX and/or LAX slices are available. These experiments were conducted to validate the hypothesis that 3D cardiac shapes can be predicted accurately, given sparse CMR acquisitions. This is particularly relevant to applications requiring acceleration of CMR scan acquisitions, at minimal cost to subsequent cardiac quantification accuracy.

3.2 Method

In the sections that follow, we describe the data and methods used within each step of our framework, and our corresponding design choices. These are ordered as follows — first, we describe the generation of reliable 3D reference shapes from the stack of 2D manual contours; next, the construction of the point distribution model; and finally, we provide details of the input data used to train and validate MCSI-Net, and describe its architecture.

3.2.1 Reference 3D Cardiac Shapes

To generate the reference cohort of 3D cardiac shapes, we first construct the 3D stack of 2D manual delineations by combining all the available contours from the three views (i.e. SAX, two- and four-chamber LAX slices) while exploiting the recorded orientation and position information available in the DICOM header to be aligned with their actual corresponding CMR images, as shown in Figure 3.1. Subsequently, the mean shape of a high-resolution atlas of the human heart available from a study by Hoogendoorn et al. [64] is rigidly and then non-rigidly registered to the 3D stack of manual delineations, to produce the patient-specific, four-chamber cardiac shape model. The resulting 3D mesh model comprises five structures; the Left Ventricle (LV) endo- and epicardium, Right Ventricle (RV), Left Atrium (LA) and Right Atrium (RA).

To generate the reference cardiac meshes, we used the standard Coherent Point Drift (CPD) algorithm [63] for rigid and non-rigid registration of the two point-sets i.e., vertices of the cardiac mesh and points in the 3D stack of manual delineations. For

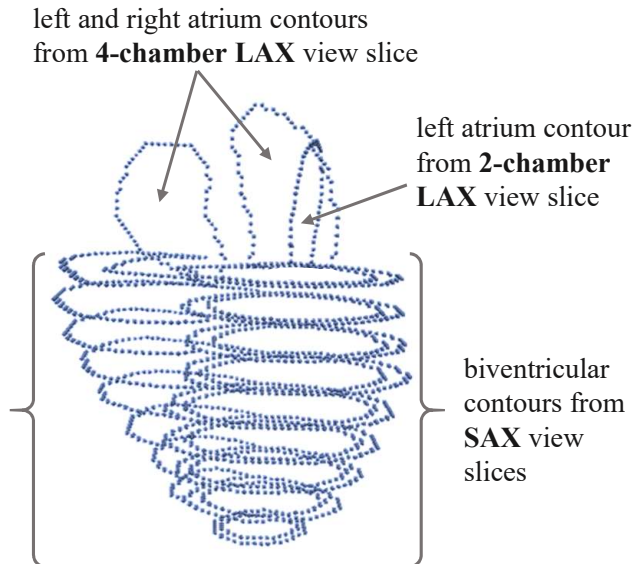


Figure 3.1: An example 3D stack of 2D manual contours on SAX, two- and four-chamber LAX view slices.

each subject comprising manual delineations, we start with rigid registration to align the two point sets and then perform non-rigid registration to deform the 3D cardiac mesh to fit the set of 3D points obtained from manual delineations.

Here, we briefly review the CPD algorithm and highlight our proposed approach to improve its overall performance in this application. CPD treats the problem of registering a source point set $\mathbf{X} \in \mathbb{R}^{n \times 3}$ to a target point set $\mathbf{Y} \in \mathbb{R}^{n \times 3}$ as one of probability density estimation. \mathbf{X} is considered to represent the centroids of a Gaussian Mixture Model (GMM), from which a transformed set of observations \mathbf{Y} are sampled. Consequently, by fitting the GMM to \mathbf{Y} , in a manner analogous to data clustering, the underlying spatial transformations that map \mathbf{X} to \mathbf{Y} and maximise the likelihood of the latter being sampled from the former, are estimated. This is achieved by maximising the log-likelihood function using the expectation-maximisation (EM) algorithm, given as follows:

$$p(\mathbf{Y}|\mathbf{X}, \mathcal{T}) = \sum_{j=1}^M \log \sum_{i=1}^N \pi_i \mathcal{N}(\mathbf{y}_j | \mathcal{T}\mathbf{x}_i, \sigma^2), \quad (3.1)$$

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

where $\mathbf{x}_{i=1\dots N} \in \mathbf{X}$ denotes a mixture component, $\mathbf{y}_{j=1\dots M} \in \mathbf{Y}$ are the observed data points, \mathcal{T} is spatial transformation parameters, π_i is a mixture coefficient and \mathcal{N} is a normal distribution function with variance σ^2 .

EM is necessary as no tractable solution exists for directly maximising Equation 3.1. The EM algorithm iteratively alternates between two steps. In the expectation (E)-step, the posterior probabilities that describe the responsibility of each mixture component \mathbf{x}_i , in describing the observed data points \mathbf{y}_j , are estimated. In the maximisation (M)-step, the posterior probabilities estimated in the preceding E-step are used to maximise Equation 3.1 with respect to the unknown spatial transformation parameters \mathcal{T} , mixture coefficients π_i , and the covariance $\sigma^2\mathbf{I}$ associated with each mixture component. CPD considers an isotropic and shared covariance across all components in the GMM (i.e. covariance of the GMM is a single scalar parameter). Both rigid and non-rigid registration are achieved by alternating between these two steps of EM, until a suitable convergence criterion is reached.

In our case, the target point set (observed data) \mathbf{Y} is given by the 3D stack of 2D manual contours (refer to Figure 3.1), while the source point set \mathbf{X} , representing the GMM centroids, is defined by vertices of the mean atlas mesh [64]. Registration of \mathbf{X} to \mathbf{Y} is achieved over three steps, starting with an initial rigid registration, where, the global rotation and translation parameters required to align \mathbf{X} to \mathbf{Y} are estimated. Subsequently, we conduct region-wise non-rigid registration, wherein, each chamber of the heart in \mathbf{X} , is registered independently to the corresponding region-specific contours in \mathbf{Y} (i.e. for example, vertices of the LV mesh in \mathbf{X} , are only registered non-rigidly to the LV contour points in \mathbf{Y}). This step is necessary as the standard non-rigid registration approach proposed in CPD does not explicitly account for multi-region point sets and enforces a global smoothing constraint on the estimated deformation field through Tikhonov regularisation. While this is necessary to prevent gross topological changes during registration, it also causes intersections between adjacent regions in the registered mesh/point set. Furthermore, such a global smoothness constraint limits the recovery of fine structural details following registration of \mathbf{Y} to \mathbf{X} , resulting in issues such as underestimation of myocardial thickness, among others. By conducting non-rigid registration in a region-wise manner, we correct for these issues as it allows region-specific parameters to be defined, to control the degree of smoothness of the deformation

field estimated for each region. The non-rigid transformation in CPD is parameterised as a linear combination of Gaussian radial basis functions. Two parameters control the degree of smoothness of the deformation field, namely, λ a weight which controls the trade-off between smoothness and registration accuracy, and β , which represents the width of the Gaussian kernel, used to parameterise the deformation.

Following the region-wise non-rigid registration, we revert to the standard global non-rigid registration formulation of CPD in order to ensure that the overall topology of the full heart meshes, and the spatial relationships between its constituent chambers are maintained. We empirically determined that registering \mathbf{X} to \mathbf{Y} in this manner, and adopting n iterations for region-wise and global non-rigid registration, provided better registration quality than using either approach alone. The values used for these parameters, for all four chambers of the heart are summarised in Table 3.1, all of which were determined empirically.

Table 3.1: Parameters used for region-wise and global registration in the CPD method.

	LV _{endo}	LV _{epi}	RV	LA	RA	global
β	2	2	2	2.5	2.5	3
λ	2.5	2.5	2.5	3	3	5
n	100	100	100	50	50	25

Typical β values are in the range [1.5, 3]. Increasing this value increases interaction between the points in the point cloud, and results in a coherent motion of larger neighbourhoods of points in the point cloud (i.e. similar displacements are estimated for larger proportions of points). Alternatively, decreasing this value reduces interaction between the points in the point cloud and conduces localised displacement of points. Similarly, typical λ values are in the range [0.1, 10]. Increasing this value produces more coherent motion. More details of the algorithm we used to achieve this effect can be found in work published by Myronenko et al. [63, 65].

Finally, all the reference shapes were quality controlled to maintain high accuracy in the generated shapes. As a first check, we computed the point-to-point distance of the generated shape to the 3D stack of 2D manual contours, and if the average error was less than half of the in-plane pixel spacing, we used the shape for the PDM construction. We then visually checked all the shapes overlaid on the stack of 2D contours to discard any sub-optimal shapes from the dataset. Ultimately, 4,525 subjects were available after

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

quality control and were randomly split into two sets of 3,925/600 for training/test, i.e. 3,925 subjects for training the neural network, and 600 subjects for evaluating performance.

In the next section, we describe how we use the generated 3D reference shapes to construct the point distribution model, as illustrated in Figure 3.2.

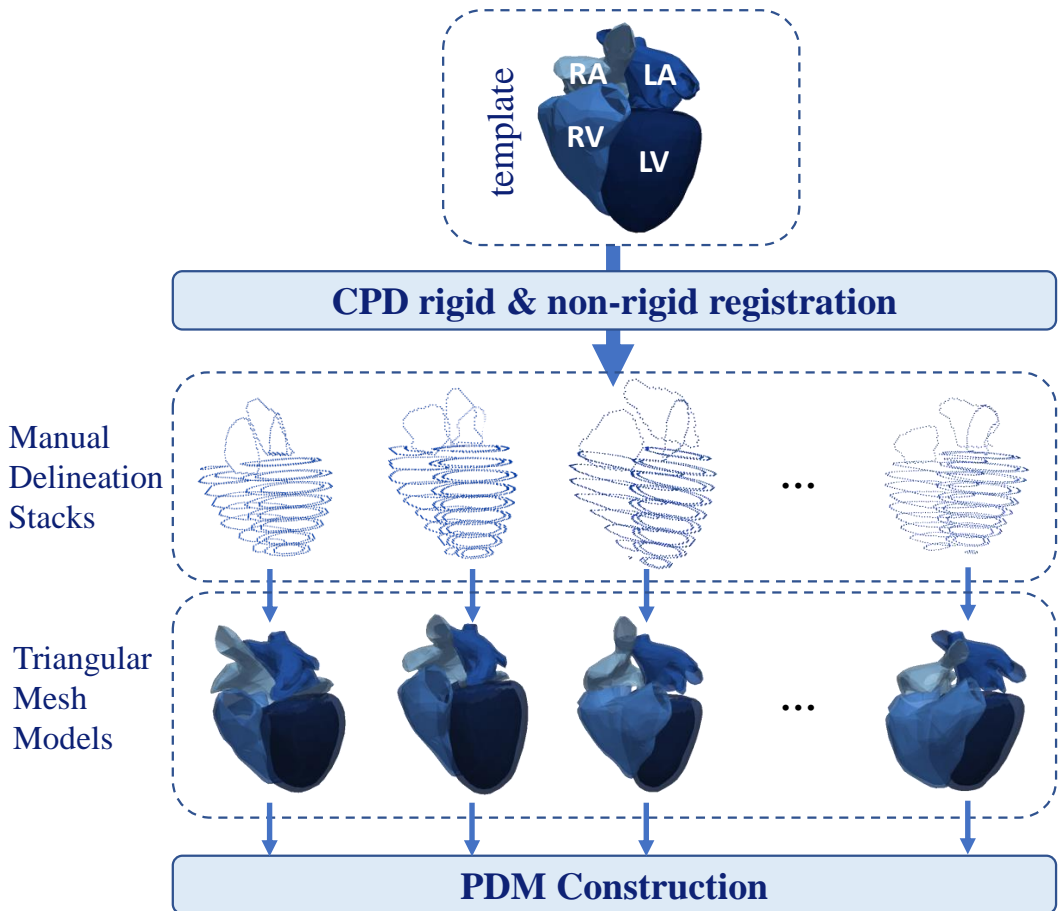


Figure 3.2: The template is registered to each stack of 2D manual delineations to produce high resolution and smooth triangular mesh models for each subject. Then all the new generated reference shapes are used to create the point distribution model.

3.2.2 Point Distribution Model (PDM)

To encode the mean and variance of the 3D cardiac shapes, we use a PCA-based PDM. We construct the PDM by applying PCA on a set of generalised Procrustes aligned

shapes.

Following Procrustes analysis, for a training set of M shapes \mathbf{z}_i , we get M new shapes \mathbf{s}_i (representing the i -th shape of the dataset) where all the nuisance pose parameters, i.e., translation $\mathbf{t} = (T_x, T_y, T_z)$, rotation $\mathbf{r} = (R_\alpha, R_\beta, R_\gamma)$ and scaling (C) were removed. Hence using these 7 transformation parameters, we can transform back the Procrustes-aligned shapes to their original (image) coordinates as follows:

$$\mathbf{z}_i = C_i \times \mathbf{s}_i \times \mathbf{r}_i + \mathbf{t}_i, \quad (3.2)$$

where $\mathbf{s}_i \in \mathbb{R}^{3N}$ represents the i -th shape as $(\mathbf{x}_1^i, \mathbf{y}_1^i, \mathbf{z}_1^i, \dots, \mathbf{x}_N^i, \mathbf{y}_N^i, \mathbf{z}_N^i)^T$ vector. The shape covariance is represented in a low-dimensional PCA space providing $l < \min(M, N)$ eigenvectors $\Phi = [\varphi_1 \varphi_2 \dots \varphi_l]$, and corresponding eigenvalues $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_l)$ computed through the Singular Value Decomposition of the covariance matrix. Thus, assuming the group of shapes follows a multi-dimensional Gaussian distribution, we can approximate any shape in the group using the shape class mean $\bar{\mathbf{s}}$ and the following linear generative model:

$$\mathbf{s} \approx \bar{\mathbf{s}} + \Phi \mathbf{b} \quad (3.3)$$

where, $\mathbf{b} \in \mathbb{R}^l$ are shape parameters restricted to $|\mathbf{b}_i| \leq \beta \sqrt{\lambda_i}$; to capture 99.7% of shape variability, we set $\beta = 3$. The shape parameters of \mathbf{s} can then be estimated as follows:

$$\mathbf{b} = \Phi_l^T (\mathbf{s} - \bar{\mathbf{s}}). \quad (3.4)$$

where the entries of \mathbf{b} are the projection coefficients of mean-centred shapes ($\mathbf{s} - \bar{\mathbf{s}}$) along the first l columns of Φ_l . Figure 3.3 shows the mean ± 3 standard deviation (SD) of the first five PCA modes variation and illustrate the variations present in the training dataset.

3.2.3 Network Architecture and Loss Function

The overall architecture of MCSI-Net is shown in Figure 3.4. The network has five inputs: SAX view images, two-, three and four-chamber LAX view images, and metadata. The network has two sets of outputs: 1) the predicted shape parameters $\mathbf{b}^P = \{b_j^P | j = 1, \dots, k\}$ that are obtained from the top branch of the network, and 2) the predicted

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

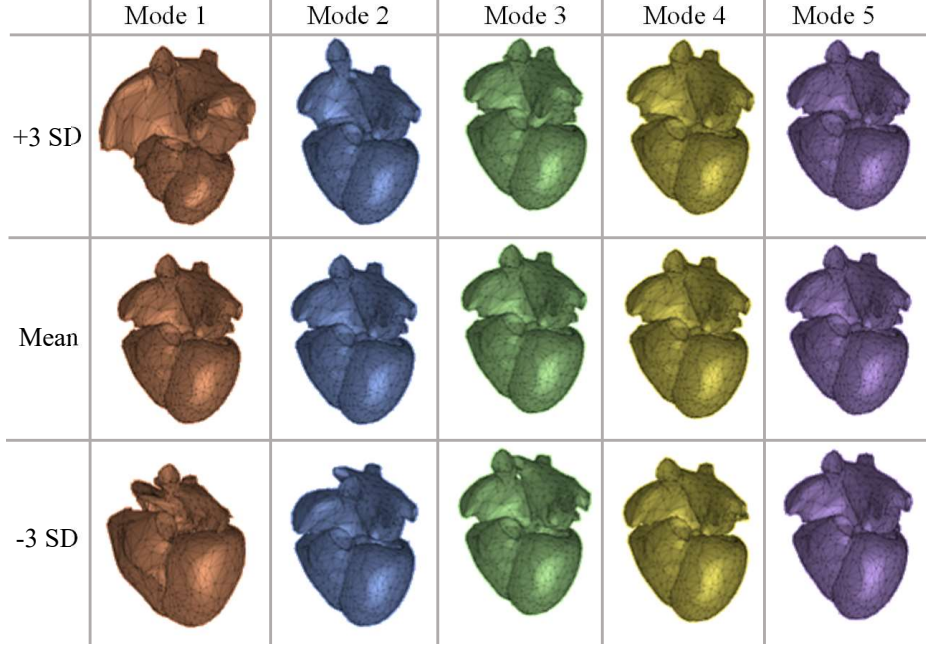


Figure 3.3: Representation of the mean ± 3 standard deviation (SD) of the first five modes of variation in the 3D shape models of four-chambered cardiac.

transformation parameters $\mathbf{t}^{\mathbf{P}} = \{T_x, T_y, T_z, R_\alpha, R_\beta, R_\gamma, C\}$ which are obtained from bottom branch of the network. The two tasks are learned through the two branches of the network simultaneously. The proposed network is trained using the following loss function:

$$\mathbf{E}(\theta) = \psi L_b + (1 - \psi) L_t \quad (3.5)$$

where

$$L_b = \sum_{j=1}^k f(b_j^P(\theta), b_j^R) \cdot w(j, k) \quad (3.6)$$

$$w(j, k) = \sqrt{\frac{k - j + 1}{k}} \quad (3.7)$$

and,

$$L_t = \sum_{l=1}^7 f(t_l^P(\theta), t_l^R) \quad (3.8)$$

where ψ is a hyper-parameter that weights the relative influence of each loss term, on the overall gradient back propagated through the network to update the constituent weights. We empirically obtained the optimal weight as 0.5 where the network equally optimises the the two terms. k is the number of shape parameters, θ denotes the network parameters, $f(\cdot)$ denotes the absolute error of the difference between the reference values (b_j^R and t_i^R) and the values predicted by the network ($b_j^P(\theta)$ and $t_i^P(\theta)$). $w(\cdot)$ denotes a function that weights the importance of the j – th mode of variation, on shape prediction, i.e. it assigns a higher weight to the first few modes of variation in the shape’s PCA space. The first few modes of variation in PCA space are critical as they capture the largest proportion of shape variability observed in the population. Correspondingly, small variations in their shape parameters (b_j^P) have a substantial impact on the predicted cardiac shape. Therefore, predicting these accurately is essential, as they have the greatest control over the final predicted shape.

Ultimately, having the mean shape, eigenvectors and predicted shape parameters, the final shape can be reconstructed using Eq. 3.3 and Eq. 3.2. We highlight that MCSI-Net uses all the available labels jointly and simultaneously. Once the network is trained, it is capable of segmenting all five sub-parts of the heart in different views simultaneously. While other state-of-the-art approaches have been proposed to segment the heart in multi-view CMR images, to the best of our knowledge, none of them exploit the contextual shape information provided by each view, to enhance cardiac shape reconstruction in 3D, as done by MCSI-Net.

As shown in Figure 3.4, we used FCNs to extract image features. We looked into recent deep learning networks that have demonstrated the best performance in regression problems [66], namely U-Net [67], VGG-16 [68], DenseNet [69], and ResNet [70].

All these FCNs are available in their standard architecture, in a Pytorch package called *torchvision*, and are used here for the purpose of feature extraction. However, as U-Net is an autoencoder, we exploited only the down-sampling path of U-Net architecture with an encoder depth of 4. We added an extra convolutional layer with the kernel size equal to the feature map dimensions to produce a vector of features. We examined the performance of the above mentioned FCNs in our application, and found that the ResNet outperforms the other architectures. We computed the segmentation accuracy using three key metrics: Dice Similarity Coefficient (DSC), Mean Contour Distance

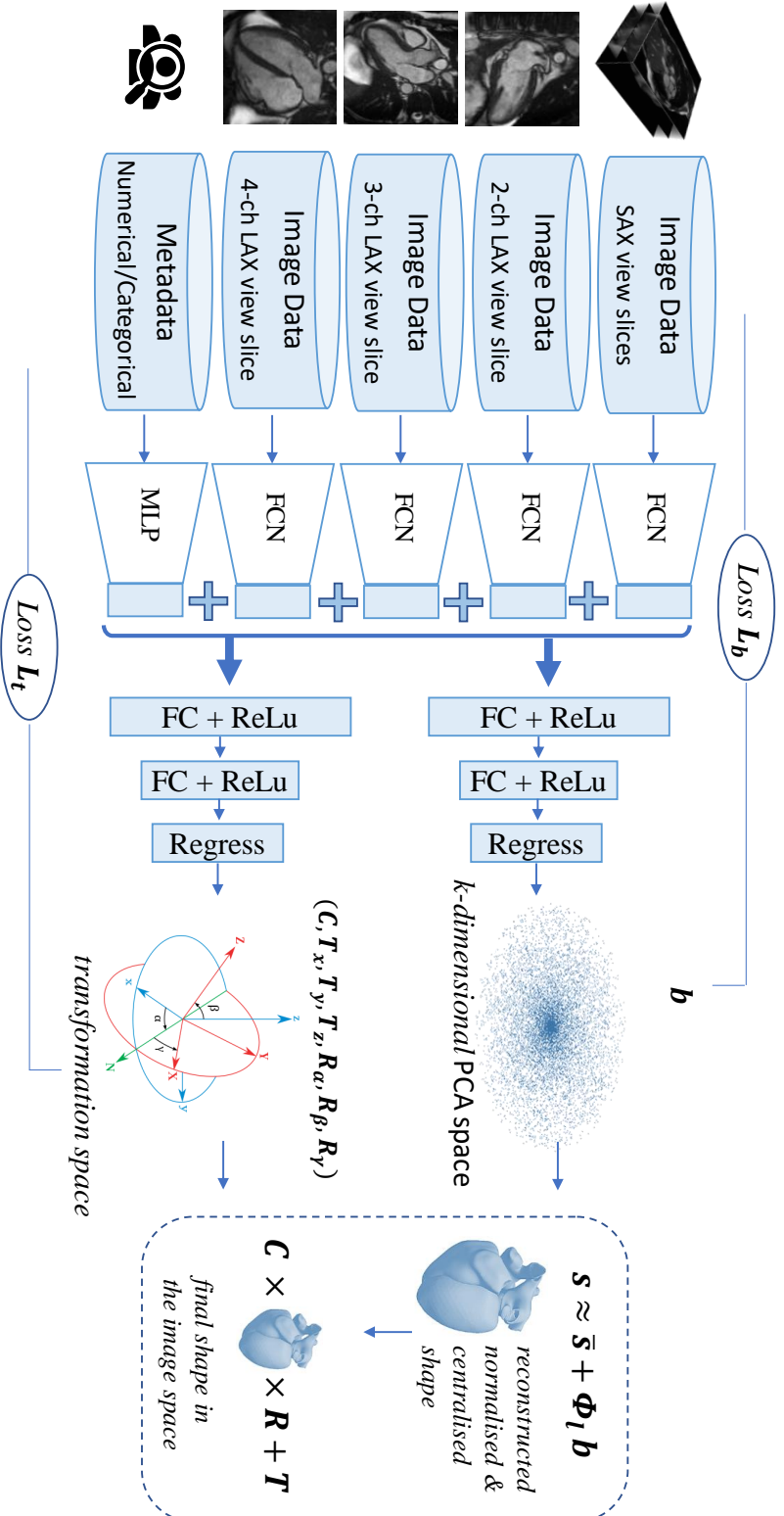


Figure 3.4: MCSI-Net extracts a high-level representation of the image from SAX and LAX views using four FCNs. It concatenates the image features together along with the output of an MLP network SAX and LAX views using four FCNs. It concatenates the image features together along with the output of an MLP network SAX and LAX views using four FCNs. It concatenates the image features together along with the output of an MLP network SAX and LAX views using four FCNs. Through a multitask learning approach, fully connected layers with ReLU activation functions produce the k parameters in PCA space and their corresponding transformation parameters which provide the 3D shape of the cardiac.

(MCD) and Hausdorff Distance (HD). DSC is between 0 and 1, with a higher DSC indicating a better match between the two shapes. MCD and HD measure the mean and maximum distance, respectively, between the manual and automatic results, with a lower value indicating a better the agreement. These metrics were defined in Equations 1.12, 1.13 and 1.14, respectively, in Section 1.5.2. Table 3.2 summarises the performance of each architecture investigated within our framework. This indicates that the best performing FCN is ResNet, which appears at the right-most column of the table. It consistently achieves the largest overlap and lowest point-to-point distance, across all sub-parts of the heart. Thus, we used ResNet as the deep feature extractor within our approach, for all subsequent experiments conducted throughout the study.

Table 3.2: Comparison of shape prediction accuracy using different FCNs in terms of DSC, MCD(mm) and HD(mm) for LV endo-/epicardium, RV, LA and RA. **Bold** indicates best performing method.

		U-Net	VGG-16	DenseNet	ResNet
LV _{endo}	DSC	0.80 ± 0.10	0.83 ± 0.08	0.89 ± 0.07	0.95 ± 0.03
	MCD	2.12 ± 1.05	1.98 ± 0.92	1.73 ± 0.69	1.04 ± 0.30
	HD	5.89 ± 2.51	4.39 ± 1.76	3.91 ± 1.15	3.10 ± 0.91
LV _{myo}	DSC	0.71 ± 0.15	0.80 ± 0.10	0.82 ± 0.09	0.89 ± 0.03
	MCD	2.31 ± 1.25	2.00 ± 0.99	1.81 ± 0.69	1.11 ± 0.39
	HD	6.11 ± 2.22	4.88 ± 1.98	4.12 ± 1.65	3.65 ± 1.21
RV _{endo}	DSC	0.76 ± 0.17	0.81 ± 0.11	0.86 ± 0.09	0.91 ± 0.05
	MCD	2.56 ± 1.76	2.22 ± 1.05	2.10 ± 1.00	1.70 ± 0.62
	HD	11.02 ± 4.01	9.13 ± 3.85	8.11 ± 3.09	6.98 ± 2.60
LA _{2ch}	DSC	0.77 ± 0.13	0.81 ± 0.10	0.89 ± 0.08	0.92 ± 0.04
	MCD	2.65 ± 1.89	2.19 ± 1.74	1.99 ± 1.63	1.67 ± 1.05
	HD	10.82 ± 7.07	8.11 ± 6.09	7.90 ± 6.11	6.26 ± 5.45
LA _{4ch}	DSC	0.78 ± 0.14	0.83 ± 0.10	0.88 ± 0.08	0.94 ± 0.03
	MCD	2.61 ± 1.81	2.18 ± 1.10	1.66 ± 0.89	1.15 ± 0.37
	HD	8.05 ± 4.19	6.65 ± 3.98	4.79 ± 2.38	4.06 ± 2.16
RA _{4ch}	DSC	0.77 ± 0.13	0.82 ± 0.09	0.87 ± 0.07	0.94 ± 0.03
	MCD	2.62 ± 2.01	1.99 ± 1.02	1.61 ± 0.65	1.10 ± 0.41
	HD	8.17 ± 4.15	7.09 ± 4.01	4.92 ± 2.60	4.16 ± 2.25

As shown in Figure 3.4, we also used an MLP to learn features from patient metadata and integrate it with the extracted image-based features. The MLP is a two-hidden-layer function denoted $g(\boldsymbol{\eta})$, where, $\boldsymbol{\eta}$ is the input (metadata feature vector

with 31 neurons). It comprises two hidden layers (with 64 and 128 neurons), and an output layer (with 256 neurons). ReLU is used in the hidden and output layers as an activation function, such that:

$$g(\boldsymbol{\eta}) = \text{ReLU}(\mathbf{c}^{(2)} + \mathbf{w}^{(2)} (\text{ReLU}(\mathbf{c}^{(1)} + \dots \mathbf{w}^{(1)} (\text{ReLU}(\mathbf{c}^{(0)} + \mathbf{w}^{(0)}\boldsymbol{\eta})))))) \quad (3.9)$$

where $\{\mathbf{c}^{(0)}, \mathbf{c}^{(1)}, \mathbf{c}^{(2)}\}$ and $\{\mathbf{w}^{(0)}, \mathbf{w}^{(1)}, \mathbf{w}^{(2)}\}$ denote biases and weights for the input and two hidden layers, respectively.

The outputs of the five sub-networks are concatenated to construct one feature vector that contains the behavioural, phenotypic, and demographic information derived from the metadata in addition to appearance information from the imaging data. This information is fed into two fully connected layers, with ReLU activation functions, so that, by minimising $E(\theta)$ from Eq. 3.5, they produce the first k parameters in PCA space, which describe the 3D shape of the cardiac chambers and the corresponding transformation parameters. To capture 99.7% of shape variability in the training dataset, we set $k = 71$ and regress only those parameters from randomly initialised weights.

3.3 Experiments and results

3.3.1 Data and Annotations

We performed train/test experiments on CMR images of 4525 subjects from the UKB using both end-diastolic (ED) and end-systolic (ES) time points. In terms of population sample size, experimental setup, and quality control, the most reliable reference annotations of cardiovascular structure and function found in the literature are those reported by Petersen et al. [36], in which CMR scans were manually delineated and analysed by a team of eight expert observers. These delineations were used to generate the reference 3D shapes, as explained in Sec. 3.2.1.

Regarding the image data, each slice (SAX and LAX views) was intensity- and spatially-normalised similar to our previous work [61]. After the pre-processing, every

slice has a size of 100×100 px and intensity values ranging between 0 and 1. As the number of SAX slices in CMR images varies typically from 7 to 15 slices, the SAX image volumes were resampled to a fixed volume size of 15 slices, using cubic B-spline interpolation.

With respect to the metadata, based on available literature and advice from our clinical collaborators, we selected a list of attributes that might directly/indirectly contribute to variations in cardiac morphology. Table 3.3 shows the summary of the metadata available for every subject in the UKB. All variables were scaled to the range $[0, 1]$, including categorical variables, which were first indexed by grouping variables and then scaled (viz. $sex \in (0, 1)$, or $smoking\ status \in (0, 0.5, 1)$).

3.3.2 Implementation and Training

The method was implemented using Python and Pytorch. The network was trained using Adam for optimising the loss function (Eq. 3.5) through 300 epochs with a learning rate of 0.001 and batch size of 8 subjects, all of which were determined empirically. There was no data augmentation. Training took ~ 24 hours on Nvidia Tesla V100 GPUs hosted by Amazon Web Service and accessed using the MULTI-X platform [38]. At test time, it took about 5 seconds to predict the shape parameters of the full cardiac cycle.

3.3.3 Accuracy of Predicted Shapes

Figure 3.5 shows a sample of our 3D cardiac shape result, at ED and ES phases, and its corresponding 2D contours (in blue) overlaid with the corresponding reference contours (in red). It confirms that the system is capable of producing accurate shape parameters to generate shapes very similar to the reference contours obtained by manual delineations. We evaluated the performance of our approach using standard metrics for assessing segmentation accuracy. These include DSC, MCD and HD between reference and predicted contours.

Table 3.4 shows the list of state-of-the-art methods that have been proposed for cardiac MR image segmentation and their corresponding average accuracy reported in terms of the three common metrics: DSC, MCD(mm), and HD(mm). The method

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

Table 3.3: Summary of the metadata (MD) of 40,000 subjects used in this study. According to the nature of the metadata, we have grouped them into four categories, i.e. MD₁: demographics, MD₂: blood pressure, MD₃: biological samples, and MD₄: lifestyle. Continuous values indicate mean \pm standard deviation.

Type	Metadata	Range
MD ₁	Sex	Male / Female
	Age (years)	57 \pm 8
	Height (cm)	170 \pm 9
	Body mass index (kg/m ²)	27 \pm 4
	Weight (kg)	78 \pm 16
	Body surface area (m ²)	1.8 \pm 0.2
	Ethnic background	White / Mixed / Other
MD ₂	Systolic blood pressure (mmHg)	139 \pm 19
	Diastolic blood pressure (mmHg)	82 \pm 11
	Diastolic brachial blood pressure (mmHg)	69 \pm 12
	Systolic brachial blood pressure (mmHg)	137 \pm 22
	Pulse rate (bpm)	69 \pm 12
	Central systolic blood pressure during PWA (mmHg)	134 \pm 21
	End systolic pressure during PWA (mmHg)	113 \pm 19
	Mean arterial pressure during PWA (mmHg)	96 \pm 14
MD ₃	HbA1c (mmol/mol)	36.1 \pm 6.7
	Cholesterol (mmol/L)	5.7 \pm 1.1
	C-reactive protein (mg/L)	2.5 \pm 4.3
	Glucose (mmol/L)	5.1 \pm 1.2
	HDL Cholesterol (mmol/L)	1.4 \pm 0.3
	IGF-1 (nmol/L)	21.3 \pm 5.6
	LDL direct Cholesterol (mmol/L)	3.5 \pm 0.8
	Triglycerides (mmol/L)	1.7 \pm 1.0
MD ₄	Smoking status	Never / Previous / Current
	Number of cigarettes smoked daily	15 \pm 8
	Sleep duration (hours/day)	7 \pm 1
	Duration of moderate activity (minutes/day)	66 \pm 75
	Duration of vigorous activity (minutes/day)	44 \pm 47
	Duration of walks (minutes/day)	61 \pm 77
	Alcohol drinker status	Never / Previous / Current
		- Never - Daily or almost daily - Three or four times a week - Once or twice a week - One to three times a month - Special occasions only
	Alcohol intake frequency	

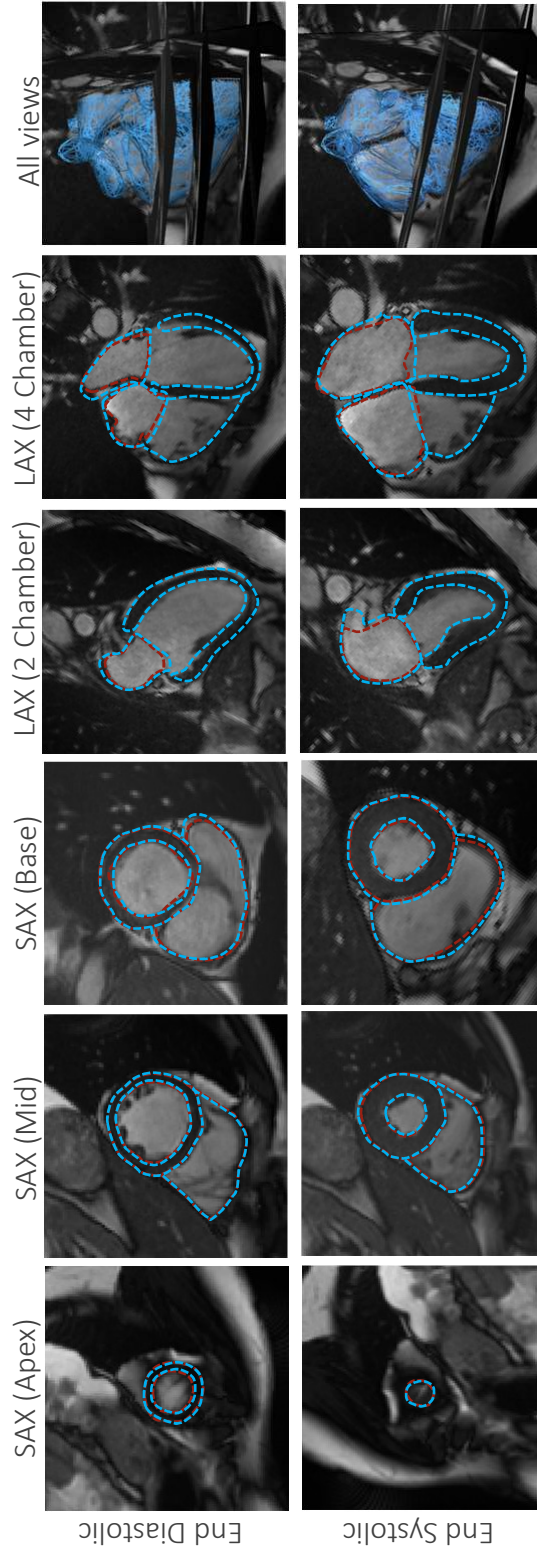


Figure 3.5: Example segmentation results for short-axis and long-axis slices at the ED and ES phases illustrating the quality for automated segmentation contours versus manual contours. Blue: automated segmentations. Red: ground-truth segmentations. The automated method segments all the time frames. However, only ED and ES frames are shown, as manual analysis only annotates ED and ES frames.

Table 3.4: Recent cardiac segmentation methods and reported accuracy compared with manual delineation in terms of the three common metrics: DSC, MCD(mm), and HD(mm), where available.

Method	Structures	Description	Average accuracy		
			DSC	MCD	HD
Bai et al. [37]	LV/RV/LA/RA	variant of U-Net	DSC = 0.91	MCD = 1.32	HD = 4.78
Attar et al. [58]	LV/RV	active shape model	DSC = 0.90	MCD = 1.31	HD = 4.75
Ruijsink et al. [71]	LV/RV	variant of U-Net	DSC = 0.89	MCD = NA	HD = NA
Zheng et al. [72]	LV/RV	variant of U-Net	DSC = 0.87	MCD = NA	HD = 7.64
Khened et al. [73]	LV/RV	variant of U-Net	DSC = 0.91	MCD = NA	HD = 5.43
Duan et al. [59]	LV/RV	FCN & registration	DSC = 0.82	MCD = NA	HD = 4.89
Tao et al. [74]	LV	standard U-Net	DSC = 0.90	MCD = 1.10	HD = NA

by [37] has shown the highest accuracy in literature thus far, for segmenting all four chambers of the heart. Thus, in addition to manual delineations and our previous work [58], we compare the performance of MCSI-Net against the fully automated CNN-based method by [37]. In Table 3.5, we present the data we used for training, testing and evaluating different methods. Here, the proposed method MCSI-Net is labelled as C, our previous work [58] in chapter 2 is labelled as A, and the method by Bai et al. [37] is labelled as B. To reasonably compare our results with those of B and A, we used the same training and testing datasets. Nearly 7% of the test set of 600 subjects are with CVDs, namely, cases with myocardial infarction, chronic ischaemic heart disease, cardiomyopathy, and heart failure.

As an additional assessment, we conducted a quantitative evaluation of human performance by measuring the inter-observer variability among the segmentations performed manually by three different clinical experts. Here, we randomly selected 50 subjects, where, each subject was analysed independently by three expert observers labelled O1, O2 and O3. We compare segmentation results on the same set of subjects to show automated versus human performance.

Tables 3.6, 3.7, and 3.8 present DSC, MCD and HD measures, respectively, that compare automated and manual segmentation results. Evaluations were performed on test sets consisting of 50 and 600 subjects which have not been used to train the PDM or MCSI-Net. Here, the group of 50 subjects is the same set used to evaluate inter-observer variability, whereas the set of 600 subjects is the same set used as a test set by [58] and [37]. Note that n is the number of subjects; we evaluated every subject at ED and ES phases, so there are 100 and 1200 image volumes in total for the dataset of 50 and 600, respectively.

In Table 3.6, the mean and standard deviations of DSC for the LV_{endo} , LV_{myo} , RV_{endo} , $LA_{2\text{ch}}$, $LA_{4\text{ch}}$ and $RA_{4\text{ch}}$ with $n = 600$ are 0.95 ± 0.03 , 0.89 ± 0.03 , 0.91 ± 0.05 , 0.92 ± 0.04 , 0.94 ± 0.03 and 0.94 ± 0.03 respectively, indicating excellent agreement between manual delineations and automated segmentation with the average DSC of 0.92. Due to the annular shape of LV_{myo} which has a larger perimeter (i.e. endo and epicardial boundary) and causes some overlap shifts, its DSC value is slightly less than that of the other cases.

In Table 3.7, we observe that the MCD is 1.04 ± 0.30 mm for the LV_{endo} , 1.11 ± 0.39

Table 3.5: Specific UKB datasets used for training and testing the methods proposed and presented in this chapter.

Label	Method	Training/Tuning Data	Test Data
A	Attrar et al. [58]	4,275 subjects from SAX view	50 subjects 600 subjects
B	Bai et al. [37]	Trained separately on: 4,275 subjects from SAX view 4,123 subjects from LAX 2CH view 4,082 subjects from LAX 4CH view	50 subjects 600 subjects
C	MCSI-Net	Trained jointly on: 3,925 subjects from SAX and LAX views	50 subjects 600 subjects
O1-O3	Three expert readers	Manual contours	50 subjects

mm for the LV_{myo} , 1.70 ± 0.62 mm for the RV_{endo} , 1.67 ± 1.05 mm for the $LA_{2\text{ch}}$, 1.15 ± 0.37 mm for the $LA_{4\text{ch}}$, and 1.10 ± 0.41 mm for the $RA_{4\text{ch}}$, all of which are smaller than the in-plane pixel spacing range of 1.8 to 2.3 mm.

In Table 3.8, the HD measures are 3.10 ± 0.91 mm, 3.65 ± 1.21 mm, 6.98 ± 2.60 mm, 6.26 ± 5.45 mm, 4.06 ± 2.16 mm and 4.16 ± 2.25 mm for the LV_{endo} , LV_{myo} , RV_{endo} , $LA_{2\text{ch}}$, $LA_{4\text{ch}}$, $RA_{4\text{ch}}$, respectively. Although HD measures are larger than the in-plane pixel spacing, they are still within an acceptable range when compared with inter-observer variability (O1-O3), or performance of methods A and B. For instance, the first three columns of Tables 3.6, 3.7, and 3.8 show inter-observer variability, where the variability between observers O1 and O2 for the HD metric is 7.56 ± 5.51 mm. Note that the performance of the proposed method (C) on LV and RV is consistently better than the other approaches (A and B) in the test set of 600 subjects, with $p < 0.01$ in paired t-tests.

3.3.4 Accuracy of Cardiac Function Indexes

Here, we report clinical cardiac functional indexes derived from manual and automated segmentation such as atrial and ventricular end-diastolic and end-systolic volume. To reproduce the reference ranges reported by Petersen et al. [36], we first extract contours corresponding to the intersection between our 3D triangular meshes and CMR image slices. Then, for the ventricular indexes calculated on SAX slices, we use Simpson’s method of integration, whereby a cardiac 3D volume can be approximated by summing the areas within 2D segmentation contours and multiplying by the inter-slice spacing. Similarly, for the atrial indexes, we calculated the volume according to the area-length method on LAX slices.

We processed and quantified all the available 40,000 subjects of the UKB; each imaged at 50 time points, i.e. in total two million image volumes. Figure 3.6 illustrates the mean and standard deviation of LV mass and the ventricular and atrial volumes of all subjects at each time point. The volumes trend matches well with the physiology of the heart where ventricles have the highest blood volume at ED, and lowest at ES phase. At the same time, the atria have the opposite behaviour where, it contains the lowest blood volume at ED and highest at ES phase. LV mass remains consistent over the cardiac cycle, which is apparent in the diagram.

Table 3.6: **Dice similarity coefficient.** Segmentation results based on the different test sets used in [37], and [36] (n=50 and 600). We compare manual with automatic methods, and error between human observers. M represents the manual ground-truth provided by [36]. A, B, and C represents the automated segmentation results from [58], [37], and our proposed MCST-Net, respectively. LV_{endo} represents LV endocardium, LV_{myo} represents LV myocardium, RV_{endo} represents RV endocardium, LA_{2ch} represents LA in two chamber LAX view, LA_{4ch} represents RA in four chamber LAX view, and RA_{4ch} represents RA in four chamber LAX view. Table values are shown as mean \pm standard deviation.

Test-set	DSC			A vs M			B vs M			C vs M		
	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	A vs M (n=50)	B vs M (n=50)	C vs M (n=50)	A vs M (n=600)	B vs M (n=600)	C vs M (n=600)	A vs M (n=600)	B vs M (n=600)	C vs M (n=600)
LV _{endo}	0.94 \pm 0.04	0.92 \pm 0.04	0.93 \pm 0.04	0.93 \pm 0.03	0.94 \pm 0.04	0.93 \pm 0.04	0.94 \pm 0.04	0.94 \pm 0.04	0.95 \pm 0.03	0.94 \pm 0.04	0.94 \pm 0.04	0.95 \pm 0.03
LV _{myo}	0.88 \pm 0.02	0.87 \pm 0.03	0.88 \pm 0.02	0.88 \pm 0.03	0.87 \pm 0.03	0.89 \pm 0.03	0.87 \pm 0.03	0.88 \pm 0.03	0.89 \pm 0.03	0.88 \pm 0.03	0.88 \pm 0.03	0.89 \pm 0.03
RV _{endo}	0.87 \pm 0.06	0.88 \pm 0.05	0.89 \pm 0.05	0.87 \pm 0.06	0.86 \pm 0.07	0.90 \pm 0.06	0.89 \pm 0.05	0.90 \pm 0.05	0.91 \pm 0.05	0.90 \pm 0.05	0.90 \pm 0.05	0.91 \pm 0.05
LA _{2ch}	0.92 \pm 0.02	0.90 \pm 0.04	0.90 \pm 0.04	NA	0.92 \pm 0.04	0.91 \pm 0.04	NA	0.93 \pm 0.05	0.91 \pm 0.04	0.93 \pm 0.05	0.92 \pm 0.04	0.92 \pm 0.04
LA _{4ch}	0.95 \pm 0.03	0.94 \pm 0.02	0.94 \pm 0.03	NA	0.95 \pm 0.02	0.94 \pm 0.03	NA	0.95 \pm 0.02	0.94 \pm 0.03	0.95 \pm 0.02	0.94 \pm 0.03	0.94 \pm 0.03
RA _{4ch}	0.95 \pm 0.02	0.95 \pm 0.02	0.95 \pm 0.02	NA	0.95 \pm 0.02	0.94 \pm 0.03	NA	0.96 \pm 0.02	0.94 \pm 0.03	0.96 \pm 0.02	0.94 \pm 0.03	0.94 \pm 0.03

Table 3.7: **Mean contour distance.** Segmentation results based on the different test sets used in [37], and [36] (n=50 and 600). We compare manual with automatic methods, and error between human observers. M represents the manual ground-truth provided by [36]. A, B, and C represents the automated segmentation results from [58], [37], and our proposed MCSI-Net, respectively. LV_{endo} represents LV endocardium, LV_{myo} represents LV myocardium, RV_{endo} represents RV endocardium, LA_{2ch} represents LA in two chamber LAX view, LA_{4ch} represents RA in four chamber LAX view, and RA_{4ch} represents RA in four chamber LAX view. Table values are shown as mean \pm standard deviation.

Test-set	MCD (mm)											
	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	A vs M (n=50)	B vs M (n=50)	C vs M (n=50)	A vs M (n=600)	B vs M (n=600)	C vs M (n=600)			
LV _{endo}	1.00 \pm 0.25	1.30 \pm 0.37	1.21 \pm 0.48	1.28 \pm 0.39	1.08 \pm 0.30	1.10 \pm 0.29	1.06 \pm 0.35	1.04 \pm 0.35	1.04 \pm 0.30			
LV _{myo}	1.16 \pm 0.34	1.19 \pm 0.25	1.21 \pm 0.36	1.20 \pm 0.34	1.18 \pm 0.31	1.15 \pm 0.30	1.13 \pm 0.34	1.14 \pm 0.40	1.11 \pm 0.39			
RV _{endo}	2.00 \pm 0.79	1.78 \pm 0.45	1.87 \pm 0.74	1.79 \pm 0.80	2.20 \pm 0.92	1.59 \pm 0.81	1.74 \pm 0.61	1.78 \pm 0.70	1.70 \pm 0.62			
LA _{2ch}	1.57 \pm 0.39	1.94 \pm 0.68	1.95 \pm 0.57	NA	1.51 \pm 0.65	1.71 \pm 0.60	NA	1.46 \pm 1.06	1.67 \pm 1.05			
LA _{4ch}	1.08 \pm 0.40	1.21 \pm 0.33	1.23 \pm 0.35	NA	0.98 \pm 0.28	1.10 \pm 0.27	NA	1.04 \pm 0.38	1.15 \pm 0.37			
RA _{4ch}	1.13 \pm 0.35	1.22 \pm 0.37	1.16 \pm 0.37	NA	1.04 \pm 0.44	1.12 \pm 0.40	NA	0.99 \pm 0.43	1.10 \pm 0.41			

Table 3.8: **Hausdorff distance.** Segmentation results based on the different test sets used in [37], and [36] (n=50 and 600). We compare manual with automatic methods, and error between human observers. M represents the manual ground-truth provided by [36]. A, B, and C represents the automated segmentation results from [58], [37], and our proposed MGSI-Net, respectively. LV_{endo} represents LV endocardium, LV_{myo} represents LV myocardium, RV_{endo} represents RV endocardium, LA_{2ch} represents LA in two chamber LAX view, LA_{4ch} represents RA in four chamber LAX view, and RA_{4ch} represents RA in four chamber LAX view. Table values are shown as mean \pm standard deviation.

Test-set	HD (mm)			HD (mm)			HD (mm)		
	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	A vs M (n=50)	B vs M (n=50)	C vs M (n=50)	A vs M (n=600)	B vs M (n=600)	C vs M (n=600)
LV _{endo}	2.84 \pm 0.70	3.31 \pm 0.90	3.25 \pm 0.96	3.21 \pm 0.97	3.46 \pm 1.05	3.05 \pm 0.90	3.15 \pm 0.96	3.16 \pm 0.98	3.10 \pm 0.91
LV _{myo}	3.70 \pm 1.16	3.82 \pm 1.07	3.76 \pm 1.21	3.91 \pm 1.20	4.06 \pm 1.16	3.72 \pm 1.15	3.90 \pm 1.29	3.92 \pm 1.37	3.65 \pm 1.21
RV _{endo}	7.56 \pm 5.51	7.35 \pm 2.19	7.14 \pm 2.20	7.41 \pm 4.11	9.02 \pm 3.54	7.30 \pm 3.00	7.21 \pm 2.62	7.25 \pm 2.70	6.98 \pm 2.60
LA _{2ch}	5.66 \pm 1.97	7.16 \pm 3.12	6.78 \pm 2.53	NA	6.59 \pm 3.64	6.72 \pm 3.60	NA	5.76 \pm 5.85	6.26 \pm 5.45
LA _{4ch}	3.89 \pm 1.85	4.29 \pm 1.97	4.06 \pm 1.44	NA	3.51 \pm 1.58	3.71 \pm 1.58	NA	4.03 \pm 2.26	4.06 \pm 2.16
RA _{4ch}	4.31 \pm 2.20	4.20 \pm 2.16	4.08 \pm 2.06	NA	4.14 \pm 2.45	4.30 \pm 2.15	NA	3.89 \pm 2.39	4.16 \pm 2.25

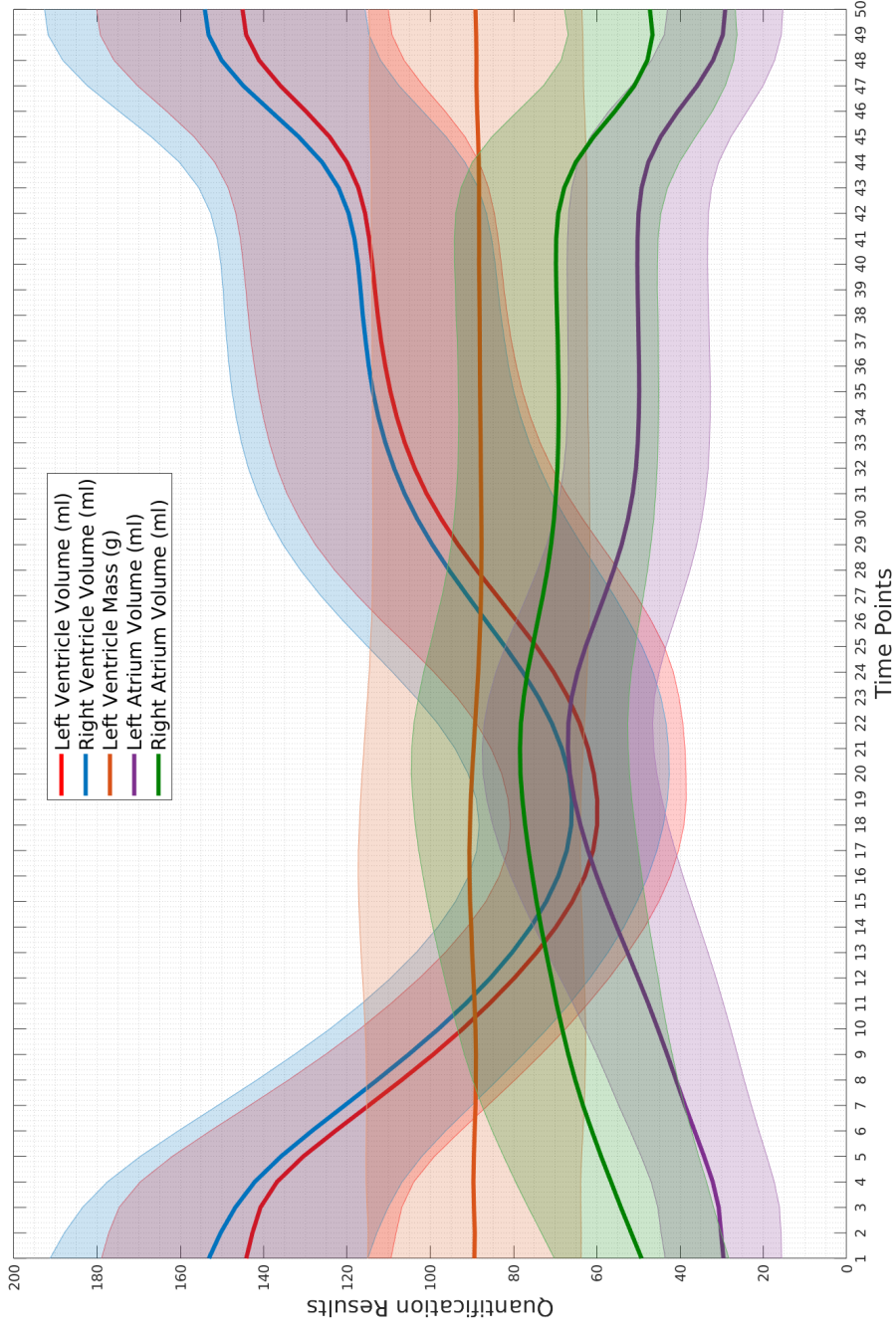


Figure 3.6: Quantification results of **2 million** CMR image volumes. Illustrating the mean (solid lines) and standard deviation (shaded area) of the ventricular and atrial volume (ml), LV mass (g) of 40k subjects of the UKB over the 50 time points.

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

Having quantified cardiac chamber volumes across all 50 time points, we report the following indexes: LVEDV, LVESV, LVSV, LVEF, LVM, RVEDV, RVESV, RVSV, RVEF, LAEDV, LAESV, LASV, LAEF, RAEDV, RAESV, RASV, and RAEF. Tables 3.9 and 3.10 show the main aforementioned cardiac functional indexes, with the first column representing the indexes derived from all the available manual segmentation. As we used part of this dataset to train our model, we compare only with the test set of 600 subjects, which is the same for all three methods of A, B and C. We also report the same indexes for the population of 40k subjects in the last column of the table. We observe here that there is excellent agreement between the clinical indexes of our proposed method and the ground truth values for all the indexes listed in the table.

Tables 3.11 and 3.12 presents the mean absolute and relative differences between the automated and manual measurements, as well as between the automated and manual measurements computed by different expert human observers. We observe here that the absolute and relative differences for the two subsets of 50 and 600 subjects matched well, and were within the error range of the three expert human observers. Similarly, although the range of differences over the cohort of 600 subjects were not directly comparable with a small test set of only 50 subjects, the difference range still was either within, or very close, to the difference range obtained by the different expert observers. We also looked into the automated values from a built-in automated segmentation software of the scanner device (i.e. *inlineVF D13A*), which was only available for the LV. Overall, A, B, and C performed substantially better than *inlineVF*, achieving an improvement of nearly 45% for all metrics. Note that these data were retrieved for every subject from the central UKB database.

Next, in Figure 3.7 and 3.8, we present Bland-Altman plots and correlation plots of the main ventricular and atrial parameters computed using our approach and the manual reference covering 600 test subjects. The Bland-Altman plot is commonly used for analysing agreement and bias between two measurements. In Figure 3.7, we have reported the mean difference (i.e. bias) and limits of agreement (LoA), i.e. ± 1.96 standard deviations from the mean. The Bland-Altman plots show strong agreement and a mean difference line at nearly zero, suggesting that the clinical indexes estimated using our approach have little bias. Conversely, the bias between different pairs of human observers as reported by Bai et al. [37] is considerable – i.e. approximately 10

Table 3.9: Summarising the differences in **ventricular** measures derived from our proposed method and manual segmentation. Here, GT represents the ground-truth values provided by manual segmentation from [36]. A, B, and C represents the quantification results derived from automated segmentation by [58], [37], and our proposed MCSI-Net, respectively. Values indicate mean \pm standard deviation.

Indexes	GT (n=4,875)	GT (n=600)	A (n=600)	B (n=600)	C (n=600)	C (n=40k)
LVEDV (ml)	144 \pm 34	143 \pm 34	144 \pm 32	148 \pm 35	144 \pm 33	144 \pm 34
LVESV (ml)	59 \pm 20	58 \pm 19	57 \pm 18	62 \pm 20	57 \pm 18	55 \pm 20
LVSV (ml)	85 \pm 19	85 \pm 20	87 \pm 20	86 \pm 20	87 \pm 19	89 \pm 20
LVEF (%)	59 \pm 6	60 \pm 6	61 \pm 7	58 \pm 6	61 \pm 6	62 \pm 7
LVM (g)	90 \pm 25	88 \pm 25	89 \pm 25	91 \pm 24	88 \pm 24	89 \pm 23
RVEDV (ml)	153 \pm 37	150 \pm 39	153 \pm 36	156 \pm 39	150 \pm 36	153 \pm 38
RVESV (ml)	67 \pm 22	65 \pm 23	63 \pm 20	68 \pm 21	64 \pm 20	66 \pm 23
RVSV (ml)	85 \pm 19	85 \pm 20	90 \pm 23	88 \pm 22	86 \pm 20	87 \pm 21
RVEF (%)	56 \pm 6	57 \pm 7	59 \pm 8	57 \pm 6	58 \pm 6	57 \pm 7

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

Table 3.10: Summarising the differences in **atrial** measures derived from our proposed method and manual segmentation. Here, GT represents the ground-truth values provided by manual segmentation from [36]. A, B, and C represents the quantification results derived from automated segmentation by [58], [37], and our proposed MCSI-Net, respectively. Values indicate mean \pm standard deviation.

Indexes	GT (n=4,875)	GT (n=600)	A (n=600)	B (n=600)	C (n=600)	C (n=40k)
LAEDV (ml)	29 \pm 14	29 \pm 14	NA	29 \pm 16	29 \pm 13	29 \pm 15
LAESV (ml)	67 \pm 21	67 \pm 21	NA	67 \pm 23	67 \pm 19	64 \pm 23
LASV (ml)	39 \pm 11	40 \pm 11	NA	39 \pm 11	38 \pm 12	34 \pm 12
LAEF (%)	59 \pm 9	59 \pm 8	NA	58 \pm 10	57 \pm 9	55 \pm 11
RAEDV (ml)	46 \pm 19	46 \pm 18	NA	49 \pm 21	44 \pm 17	49 \pm 20
RAESV (ml)	79 \pm 26	79 \pm 26	NA	82 \pm 28	77 \pm 24	77 \pm 28
RASV (ml)	33 \pm 11	33 \pm 13	NA	34 \pm 13	33 \pm 11	29 \pm 12
RAEF (%)	42 \pm 11	42 \pm 10	NA	41 \pm 11	43 \pm 10	40 \pm 12

Table 3.11: The **absolute difference** in clinical measures between the automatic and manual segmentations, as well between measurements by different human observers. M: ground truth provided by manual segmentation [36]. A, B, and C represents the clinical indexes derived from the automated segmentation by [58], [37], and our proposed MCSI-Net, respectively. Values indicate mean \pm standard deviation.

	Absolute difference									
	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)	A vs M (n=50)	B vs M (n=50)	C vs M (n=50)	A vs M (n=600)	B vs M (n=600)	C vs M (n=600)	
LVEDV	6.1 \pm 4.4	8.8 \pm 4.8	4.8 \pm 3.1	5.9 \pm 4.2	4.3 \pm 4.9	6.4 \pm 4.8	6.5 \pm 5.4	6.1 \pm 5.3	6.1 \pm 4.4	
LVESV	4.1 \pm 4.2	6.7 \pm 4.2	7.1 \pm 3.8	6.8 \pm 5.1	6.5 \pm 5.4	4.7 \pm 3.8	5.1 \pm 5.0	5.3 \pm 4.9	4.8 \pm 3.5	
LVM	4.2 \pm 3.2	6.6 \pm 4.9	6.5 \pm 4.8	6.0 \pm 4.4	6.4 \pm 3.5	5.9 \pm 4.0	7.0 \pm 5.4	6.9 \pm 5.5	6.3 \pm 4.7	
RVEDV	11.1 \pm 7.2	6.2 \pm 4.6	8.7 \pm 5.8	10.0 \pm 5.8	8.4 \pm 6.8	7.0 \pm 4.2	8.4 \pm 7.8	8.5 \pm 7.1	6.6 \pm 4.2	
RVESV	15.6 \pm 7.8	6.6 \pm 5.5	11.7 \pm 6.9	10.0 \pm 6.5	13.9 \pm 9.9	7.8 \pm 5.4	7.7 \pm 6.5	7.2 \pm 6.8	6.6 \pm 4.8	
LAEDV	3.1 \pm 2.1	2.8 \pm 2.3	2.2 \pm 1.6	NA	4.8 \pm 3.3	3.5 \pm 2.5	NA	4.6 \pm 3.0	3.6 \pm 2.7	
LAESV	4.0 \pm 2.8	4.8 \pm 4.0	3.9 \pm 3.0	NA	6.8 \pm 5.1	5.7 \pm 3.0	NA	5.2 \pm 4.3	5.3 \pm 3.1	
RAEDV	2.5 \pm 2.3	3.6 \pm 3.0	3.6 \pm 2.5	NA	5.3 \pm 6.0	5.6 \pm 3.8	NA	5.0 \pm 5.6	5.5 \pm 4.0	
RAESV	3.5 \pm 3.6	3.8 \pm 4.3	3.7 \pm 3.3	NA	5.9 \pm 5.4	6.1 \pm 3.7	NA	5.6 \pm 5.3	6.5 \pm 3.9	

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

Table 3.12: The **relative difference** in clinical measures between the automatic and manual segmentations, as well between measurements by different human observers. M: ground truth provided by manual segmentation [36]. A, B, and C represents the clinical indexes derived from the automated segmentation by [58], [37], and our proposed MCSI-Net, respectively. Values indicate mean \pm standard deviation.

	Relative difference (%)			A vs M (n=50)	B vs M (n=50)	C vs M (n=50)	A vs M (n=600)	B vs M (n=600)	C vs M (n=600)
	O1 vs O2 (n=50)	O2 vs O3 (n=50)	O3 vs O1 (n=50)						
LVEDV	4.2 \pm 3.1	6.3 \pm 3.3	3.4 \pm 2.2	4.2 \pm 3.0	2.9 \pm 3.6	4.7 \pm 3.6	4.7 \pm 3.3	4.1 \pm 3.5	4.4 \pm 3.5
LVESV	6.8 \pm 7.5	12.5 \pm 8.5	11.7 \pm 5.1	10.2 \pm 8.1	12.5 \pm 11.2	9.8 \pm 8.0	9.3 \pm 9.4	9.5 \pm 9.5	9.1 \pm 7.2
LVM	4.4 \pm 3.3	6.0 \pm 3.7	6.7 \pm 4.6	6.5 \pm 4.1	8.0 \pm 4.8	7.4 \pm 5.6	8.3 \pm 7.7	8.3 \pm 7.6	7.8 \pm 6.5
RVEDV	8.0 \pm 5.0	4.2 \pm 3.1	5.7 \pm 3.6	7.3 \pm 4.2	5.7 \pm 4.3	5.2 \pm 3.4	5.4 \pm 4.7	5.6 \pm 4.6	4.7 \pm 3.4
RVESV	30.6 \pm 15.5	10.9 \pm 8.3	16.9 \pm 9.2	22.0 \pm 8.4	29.8 \pm 22.1	18.6 \pm 16.6	12.4 \pm 9.0	11.8 \pm 12.2	11.0 \pm 9.3
LAEDV	13.3 \pm 9.8	11.0 \pm 7.8	7.8 \pm 5.3	NA	13.1 \pm 10.0	16.1 \pm 13.5	NA	13.1 \pm 10.3	14.8 \pm 11.8
LAESV	6.6 \pm 5.4	7.3 \pm 5.6	6.0 \pm 5.0	NA	7.1 \pm 6.9	9.4 \pm 5.7	NA	7.5 \pm 6.8	8.8 \pm 6.3
RAEDV	6.5 \pm 7.6	8.0 \pm 4.9	8.4 \pm 5.7	NA	12.1 \pm 14.3	15.2 \pm 13.0	NA	12.0 \pm 18.7	13.8 \pm 12.0
RAESV	4.7 \pm 6.4	5.2 \pm 6.4	4.9 \pm 4.3	NA	7.7 \pm 6.6	8.7 \pm 5.9	NA	7.4 \pm 6.9	9.1 \pm 6.2

(ml) for ventricular parameters.

Figure 3.8 presents correlation plots comparing the manual and automated methods, for different cardiac function indexes. The correlation coefficient (corr) measures the strength of the relationship between two sets of observations. The strength and direction of the relationship indicate the predictive power of our framework. Coefficients for all indexes ranged between 0.90 and 0.97, indicating a strong relationship between our approach and the manual delineations.

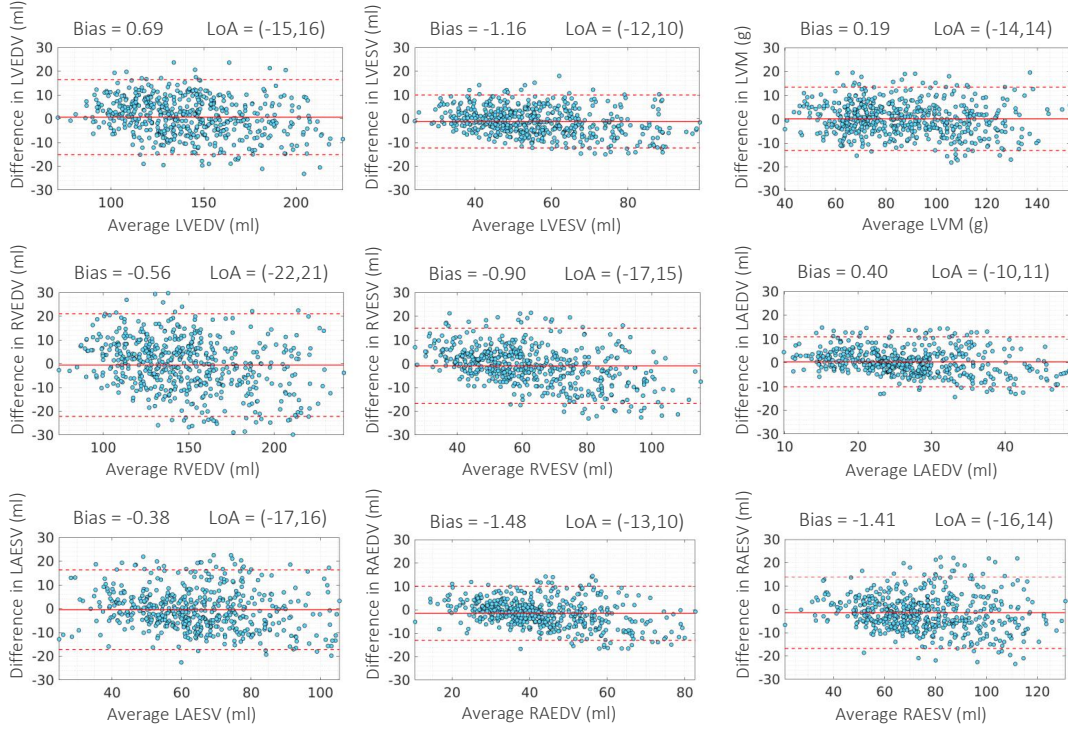


Figure 3.7: Illustrating the repeatability of various cardiac functional indexes comparing the manual and automated analysis of 600 subjects from the UKB cohort; Bland-Altman plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available. The solid horizontal lines denote the bias or the mean difference (Automatic – Manual), whereas the two dashed lines denote limits of agreement (LoA) i.e. ± 1.96 standard deviations from the mean.

Furthermore, to demonstrate whether the values of clinical indexes estimated automatically share the same distribution as those acquired via the manual approach, we

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

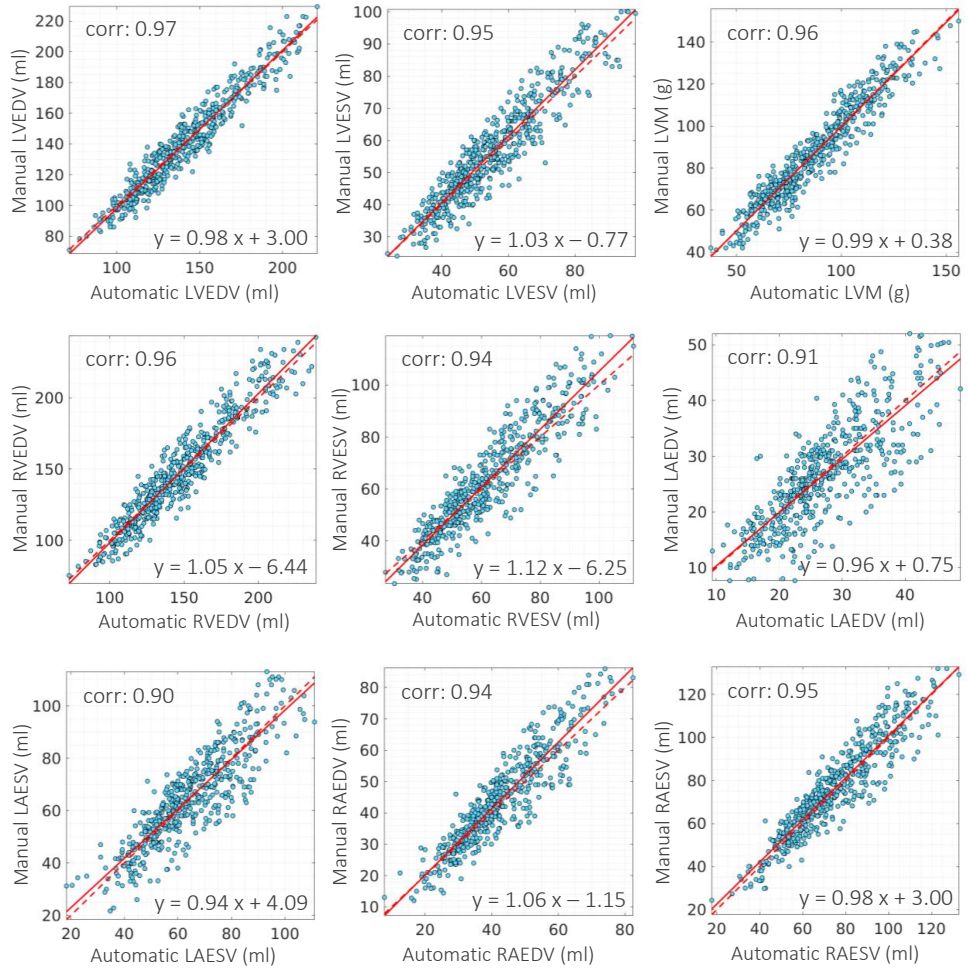


Figure 3.8: Illustrating the repeatability of various cardiac functional indexes comparing the manual and automated analysis of 600 subjects from the UKB cohort; Correlation plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available. The dashed and solid line denote the identity and linear regression lines, respectively.

visualise their distributions in Figure 3.9. Here, we present probability distribution plots for various cardiac functional indexes computed both manually and/or automatically over three sets, i.e. set of all available 4875 subjects with manual delineations, set of 600 subjects we used to test our method (not used for training), and the full cohort of 40k subjects. From these, we observe that the distribution of the various indexes estimated using our approach agrees almost exactly with those obtained from the manual segmentations. We observe the same distribution shape, location and scale, and similar tail behaviour. Furthermore, Figure 3.10, illustrates the Q-Q plots of the same indexes on the test of 600 subjects, confirming the agreement on different quantiles over all the indexes, where we see the quantiles points roughly forming a straight line.

We also performed two-sample Kolmogorov-Smirnov (K-S) tests to verify that ventricular and atrial parameters obtained through manual and automated approaches are drawn from the same distribution, under the null hypothesis that the manual and automatic methods are from the same continuous distribution in terms of clinical indexes. From our analysis, K-S test results on different indexes do not reject the null hypothesis of being from the same distribution at the 5% significance level.

3.3.5 Sub-Cohort Analysis

The proposed approach enables us to perform large-scale population-wide analysis of CMR images, demonstrated thus far on data from the UKB, with statistics on the most commonly used clinical indexes derived from CMR exams. We believe, however, that the power of population studies lies in the opportunity to define and characterise human sub-populations, and investigate the patterns and trends across different sub-populations.

Table 3.13 compares the ventricular and atrial volume and LV mass, derived from our approach, between two groups of subjects, namely, healthy subjects ($n=800$), and subjects with myocardial infarction (MI) ($n=800$). The table shows that MI is associated with increased ventricular and atrial volume and LV mass with statistical significance ($p\text{-value} < 0.001$). This is consistent with previous findings, from manual analysis of a few dozens of subjects, where cardiac remodelling and enlargement happens in patients with MI. Now we can confirm similar findings and perform similar studies, with the advantage of reliable automated analysis on really large-scale datasets such

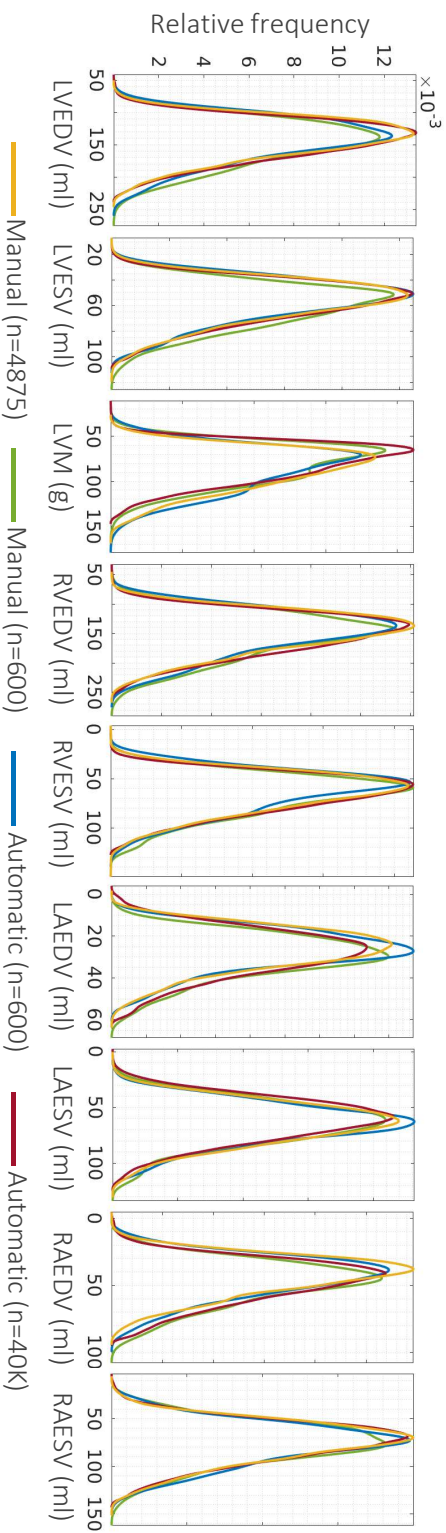


Figure 3.9: Distributions of various cardiac functional indexes comparing results of manual and automatic analyses of 600 subjects. It shows probability distribution plots for various cardiac functional indexes computed both manually and automatically in which manual segmentation was available.

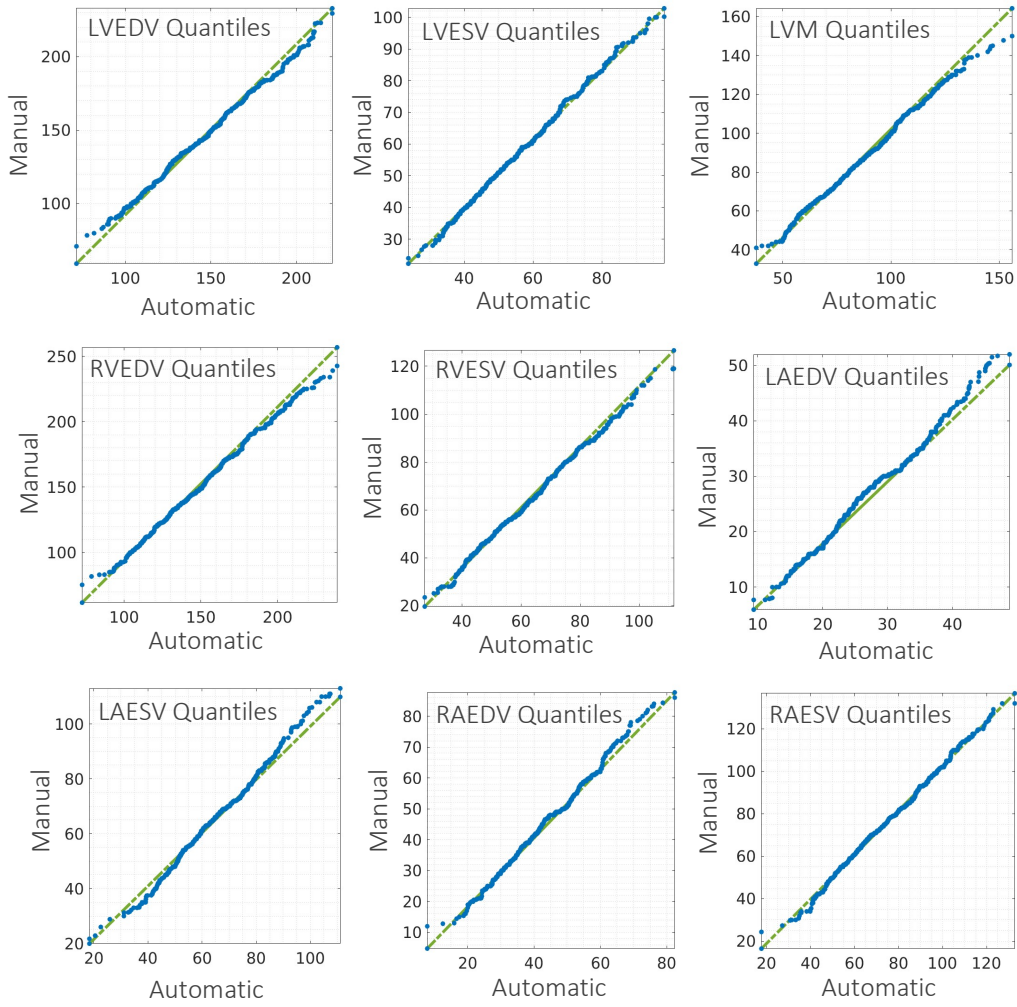


Figure 3.10: Q-Q plots for various cardiac functional indexes computed both manually and automatically of 600 subjects.

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

as UKB, where, there is tons of data available to discover different associations and patterns.

Table 3.13: Example comparison of cardiac function on large-scale datasets using automatically derived clinical indexes of 1600 subjects. Illustration of cardiac remodelling on healthy subjects vs subjects with myocardial infarction (MI).

	Healthy	MI	<i>p</i> -value
LVEDV (ml)	144 ± 34	166 ± 43	< 0.001
LVESV (ml)	59 ± 18	72 ± 33	< 0.001
LVM (g)	86 ± 24	103 ± 24	< 0.001
RVEDV (ml)	154 ± 40	163 ± 37	< 0.001
RVESV (ml)	69 ± 24	73 ± 21	< 0.001
LAEDV (ml)	27 ± 11	40 ± 24	< 0.001
LAESV (ml)	66 ± 20	75 ± 30	< 0.001
RAEDV (ml)	46 ± 19	55 ± 26	< 0.001
RAESV (ml)	80 ± 25	81 ± 31	= 0.630

3.3.6 Impact of Metadata on Shape Accuracy

Several studies ([75], [62] and [76], just to name a few) have shown the correlation between baseline features (such as lifestyle and demographics) and cardiac morphology and structure. In this subsection, we report the effect of including metadata (as shown in Table 3.3) as additional information in our network. As expected, the use of metadata alongside image information improves the network, leading to more accurate prediction in all cardiac substructures.

Table 3.14 summarises the effect of including different sets of metadata on our method. The first column shows the accuracy metrics of DSC, MCD and HD when we only used the image data (labelled as IMG). The next four columns report the accuracy of the predicted shapes using four different sets of metadata, i.e. (1) demographics, (2) blood pressure, (3) biological samples, and (4) lifestyle (labelled as MD₁ to MD₄) alongside the image data (IMG). We observe that MD₁, MD₂ and MD₄ which represent the demographics, blood pressure and lifestyle, have a considerable impact on the shape accuracy (*p*-value < 0.01), while, MD₃ does not contribute significantly (*p*-value < 0.1). The right-most column represents the most accurate results where we use all the metadata alongside the image data, achieving on average, around 5% improvement

compared with the first column (p -value < 0.001), where, we only use the image data. This can be attributed to the combined use of image and patient metadata within a single network to directly predict shape parameters. The introduction of the metadata yielded a substantial positive impact on shape prediction accuracy, across all metrics. We believe that including this information provides the network with a variable prior by allowing it to learn the likely distributions of shape parameters across different populations.

3.3.7 Contribution of Slices on Shape Accuracy

We further investigated generating accurate 3D cardiac shapes when only a fraction of the image data is available. This was considered highly relevant to this study for the following reasons: (1) Despite all the advantages of CMR scans, there is an explicit limitation of acquisition time. A routine CMR scan takes 20-60 minutes, which is time-consuming and expensive, especially in environments where resources are strained and availability of scan time is limited. Additionally, such long scan times may be infeasible entirely, for certain niche groups of patients. Besides, CMR often requires breath-holds, which could be not easy for many patients. Accordingly, it is essential to decrease the acquisition time, while maintaining the quality of subsequent quantitative analyses. Consequently, accurate cardiac quantification using fewer CMR slices, would help facilitate the acceleration of CMR scan acquisitions, improving patient comfort and reducing expenses. (2) Typical artefacts found during CMR image acquisitions are missing slices, which result in missing contours from 2D segmentation methods (such as the previously mentioned by [37]) across the heart and, as a result, sub-optimal estimation of volumetric indexes. Consequently, a 3D cardiac shape generation framework that is robust to the presence of such variations would be of significant clinical value as it would enable accurate quantification of cardiac functional indexes, despite the presence of such artefacts. To address these issues, in addition to the default image data set where we use all the available SAX and LAX image slices, we used incomplete samples to train our network, and subsequently, predicted complete cardiac shapes. We extended our training dataset by adding new cases where the image data in SAX and LAX are not all present, i.e. retaining only

- the three Basal, Mid and Apical slices (labelled as BMA) without any LAX slices,

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

Table 3.14: Comparison of shape prediction accuracy using only images (IMG) or images with different type of metadata (IMG+MD) in terms of DSC, MCD(mm) and HD(mm) for LV endo-/epicardium and RV, LA, and RA on 600 UKB subjects. **Bold** indicates best performing method.

	IMG	IMG+MD ₁	IMG+MD ₂	IMG+MD ₃	IMG+MD ₄	IMG+MD _{1,2,3,4}	
LV _{endo}	DSC	0.91 ± 0.03	0.94 ± 0.03	0.93 ± 0.03	0.92 ± 0.03	0.93 ± 0.03	0.95 ± 0.03
	MCD	1.10 ± 0.34	1.05 ± 0.31	1.07 ± 0.31	1.09 ± 0.32	1.06 ± 0.32	1.04 ± 0.30
	HD	3.29 ± 0.93	3.13 ± 0.92	3.19 ± 0.92	3.23 ± 0.93	3.20 ± 0.93	3.10 ± 0.91
LV _{myo}	DSC	0.87 ± 0.04	0.88 ± 0.04	0.88 ± 0.04	0.87 ± 0.05	0.88 ± 0.04	0.89 ± 0.03
	MCD	1.15 ± 0.40	1.13 ± 0.40	1.12 ± 0.40	1.15 ± 0.40	1.14 ± 0.40	1.11 ± 0.39
	HD	3.84 ± 1.23	3.72 ± 1.22	3.73 ± 1.23	3.82 ± 1.23	3.71 ± 0.22	3.65 ± 1.21
RV _{endo}	DSC	0.90 ± 0.05	0.91 ± 0.05	0.91 ± 0.05	0.90 ± 0.05	0.91 ± 0.05	0.91 ± 0.05
	MCD	1.75 ± 0.63	1.72 ± 0.63	1.71 ± 0.63	1.74 ± 0.63	1.72 ± 0.63	1.70 ± 0.62
	HD	7.41 ± 2.63	7.11 ± 2.60	7.07 ± 2.62	7.40 ± 2.63	7.14 ± 2.61	6.98 ± 2.60
LA _{2ch}	DSC	0.90 ± 0.04	0.91 ± 0.04	0.91 ± 0.04	0.90 ± 0.04	0.91 ± 0.04	0.92 ± 0.04
	MCD	1.75 ± 1.08	1.70 ± 1.07	1.69 ± 1.06	1.74 ± 1.08	1.72 ± 1.06	1.67 ± 1.05
	HD	6.63 ± 5.51	6.37 ± 5.47	6.38 ± 5.49	6.61 ± 5.50	6.37 ± 5.47	6.26 ± 5.45
LA _{4ch}	DSC	0.91 ± 0.04	0.93 ± 0.04	0.93 ± 0.03	0.92 ± 0.04	0.93 ± 0.04	0.94 ± 0.03
	MCD	1.20 ± 0.38	1.17 ± 0.38	1.16 ± 0.37	1.19 ± 0.38	1.17 ± 0.38	1.15 ± 0.37
	HD	4.27 ± 2.17	4.20 ± 2.17	4.22 ± 2.17	4.25 ± 2.17	4.20 ± 2.16	4.06 ± 2.16
RA _{4ch}	DSC	0.91 ± 0.04	0.93 ± 0.04	0.94 ± 0.03	0.92 ± 0.05	0.93 ± 0.04	0.94 ± 0.03
	MCD	1.17 ± 0.44	1.12 ± 0.43	1.14 ± 0.42	1.16 ± 0.44	1.12 ± 0.43	1.10 ± 0.41
	HD	4.39 ± 2.27	4.20 ± 2.27	4.22 ± 2.26	4.39 ± 2.27	4.26 ± 2.26	4.16 ± 2.25

- SAX slices (on average ten SAX slices without any LAX slices), and
- the three LAX view slices (two-, three and four-chamber view) without any SAX slices.

Using this process, we generated three new samples, from every sample in the original dataset. To keep the network architecture, input format, and structure consistent across all experiments, slices were excluded during training by replacing them with an empty slice, i.e. with zero values for all pixels.

Table 3.15 summarises the shape prediction accuracy of our approach for all cardiac structures, compared with the manual delineations, where, we do not use all the available image data. As expected, we observe that the most accurate results are obtained when we use all the available image data, reported in the right-most column of the table.

The first column shows the accuracy metrics of different cardiac structures when we only use Basal, Mid and Apical slices (BMA). We see the accuracy of our approach is satisfactory (average DSC of 0.84 for ventricles and 0.80 for atria) using three SAX slices that do not cover the atria. It is encouraging to see that the network can still predict the atrial shapes to a reasonable degree of accuracy based on the minimal cues available in the three SAX slices.

The second column presents the results associated with the SAX slices only. We see that the results become better in terms of both ventricles and atria segmentation accuracy when the network has access to more image data. Note, on average the number of slices is at least three times higher than the BMA case, but the improvement is about 2% on average. Similarly, the third column shows the results of segmentation when three LAX images are used, and we get comparable results for the ventricles (relative to BMA and SAX). However, we see higher accuracy for atria as the LAX cover the atria regions, while, in SAX and BMA there is no direct information of atria regions.

We see the addition of the three LAX slices to the three BMA slices (labelled as BMA+LAX), results in higher accuracy compared with BMA, SAX and LAX, as the visual cues exist to describe both ventricles and atria regions, covering the sagittal and transverse representation of the heart. Overall, we see the potential strength of this method to produce accurate and high-resolution 3D meshes of cardiac shape using only

a fraction (about 23% to 46%) of image data. Figure 3.11 presents three sets of sample output shapes generated from a fraction of image data, overlaid with the corresponding CMR images.

3.4 Summary

This chapter presented a fully automatic method capable of performing high-throughput end-to-end 3D cardiac MR image analysis via the simultaneous use of images and patient metadata. In this chapter, we introduced MCSI-Net, where we embedded statistical shape models within a fully convolutional neural network to jointly learn phenotypic and demographic information at scale. We leveraged the ability of the network to learn the visual features of cardiac chambers in cine CMR images, and generate plausible 3D cardiac shapes, by constraining the prediction using a shape prior, in the form of the statistical modes of shape variation learned a priori from a subset of the population. We validated our workflow on a reference cohort of 600 subjects for which both manual delineations and reference functional indexes exist, and the full dataset with 40,000 subjects. Our results show that differences between our approach and the manually obtained reference indexes are within the expected variability observed in human raters. Similar to the previous chapter, this method provides the scientific community with the first fully automated 3D analysis of cardiac chambers from the sparse CMR images. This method provides a detailed analysis of cardiac morphology from high-resolution spatio-temporal information provided by the cardiac mesh models. We also presented the positive impact of the inclusion of patient metadata on the accuracy of the predicted shapes. We believe that including this information provides the network with a variable prior by learning the likely distributions of shape parameters across different sub-cohorts. We hope this significant improvement encourages other researchers to consider all the available information such as these side metadata as priors in other applications for potentially more accurate and patient-specific models. Furthermore, we showed that the proposed method achieves comparable results using just a fraction of image data, which supports the idea of accelerating the CMR scan acquisition by capturing significantly fewer image slices. Facilitated by the architecture of the network, training of the network is performed jointly using all the available image views at the

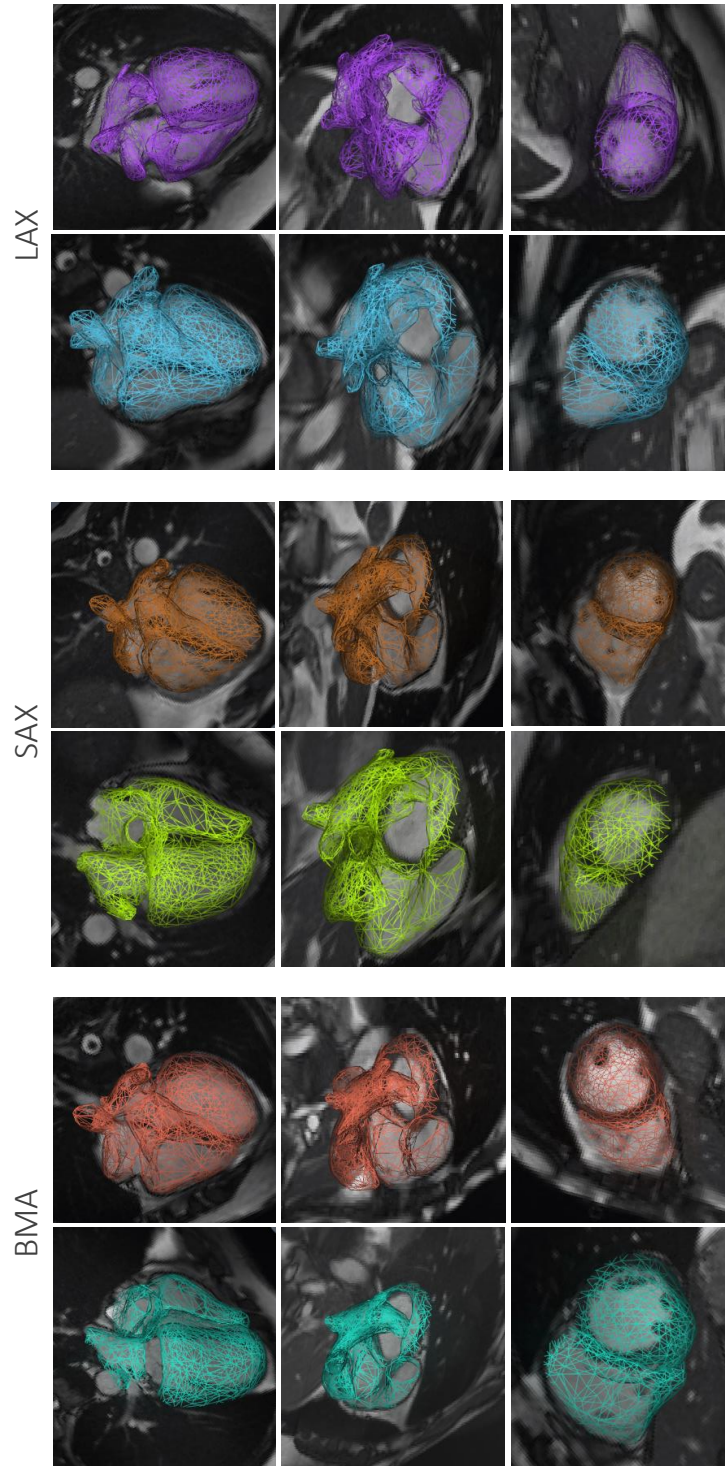


Figure 3.11: Illustration of the quality of the shapes in different views. Cardiac shapes generated from a fraction of image data using only BMA, only SAX, or only LAX. Each colour corresponds to an individual subject.

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

Table 3.15: Comparison of shape prediction accuracy using different set of images i.e. only SAX images, only LAX images, only three main SAX images (BMA: base, mid, and apex) or combination of SAX and LAX images in terms of DSC, MCD(mm) and HD(mm) for LV endo-/epicardium and RV, LA, and RA on 600 UKB subjects.

	BMA	SAX	LAX	BMA+LAX	SAX+LAX
LV _{endo}	DSC	0.89 ± 0.04	0.91 ± 0.04	0.88 ± 0.05	0.92 ± 0.05
	MCD	1.25 ± 0.34	1.14 ± 0.35	1.71 ± 0.52	1.10 ± 0.31
	HD	3.22 ± 1.01	3.19 ± 0.99	4.13 ± 1.29	3.14 ± 0.95
LV _{myo}	DSC	0.82 ± 0.04	0.85 ± 0.05	0.84 ± 0.09	0.87 ± 0.04
	MCD	1.44 ± 0.55	1.20 ± 0.43	1.94 ± 0.65	1.15 ± 0.40
	HD	4.01 ± 1.50	3.86 ± 1.41	4.62 ± 1.45	3.71 ± 1.30
RV _{endo}	DSC	0.81 ± 0.06	0.86 ± 0.04	0.82 ± 0.07	0.88 ± 0.05
	MCD	2.12 ± 0.85	1.91 ± 0.70	1.95 ± 0.80	1.86 ± 0.65
	HD	8.01 ± 2.83	7.17 ± 2.71	7.41 ± 3.05	7.06 ± 2.72
LA _{2ch}	DSC	0.80 ± 0.06	0.81 ± 0.10	0.90 ± 0.04	0.91 ± 0.04
	MCD	2.39 ± 2.08	2.24 ± 1.99	1.69 ± 1.06	1.67 ± 1.06
	HD	7.71 ± 6.61	7.13 ± 6.05	6.28 ± 5.44	6.29 ± 5.46
LA _{4ch}	DSC	0.80 ± 0.05	0.82 ± 0.08	0.93 ± 0.03	0.93 ± 0.04
	MCD	2.66 ± 1.43	2.29 ± 1.21	1.16 ± 0.36	1.15 ± 0.37
	HD	7.01 ± 4.05	6.70 ± 3.95	4.06 ± 2.14	4.05 ± 2.16
RA _{4ch}	DSC	0.80 ± 0.06	0.81 ± 0.08	0.94 ± 0.04	0.94 ± 0.03
	MCD	2.91 ± 2.01	2.32 ± 1.81	1.12 ± 0.40	1.11 ± 0.41
	HD	7.02 ± 4.10	6.81 ± 4.01	4.17 ± 2.25	4.15 ± 2.27

same time, which leads to a more informative and rich model containing the contextual and topological representation of the cardiac shape. This simultaneous use of different image views also supports the network to handle cases where there are missing imaging data from any short-axis or long-axis view slices.

In the next chapter, we will present a new method for automatically quantifying epicardial fat tissue in CMR images. This method will be built up on top of the cardiac segmentation method we have developed so far. We will then look into the distribution of epicardial fat in different sub-populations to identify the relationship of the quantified fat and other clinical variables for the first time at this scale.

3. QUANTITATIVE POPULATION ANALYSIS OF CARDIAC CHAMBERS

CHAPTER 4

Quantitative Population Analysis of Epicardial Fat Tissue

Epicardial fat tissue (EFT) is a visceral adipose tissue that surrounds the myocardium and has a significant role in cardiac function. It is necessary to better define normal reference values and the risk associated with EFT to further assess its function in cardiovascular and metabolic diseases. In this chapter, we introduce a new EFT quantification method based on our fully validated cardiac segmentation pipelines. We use the privilege information given by our cardiac segmentation method as accurate contours of the epicardium wall. Then we will use a clustering approach to identify the fat pixels in a minimised search region around the epicardial wall. To the best of our knowledge, this is the first work addressing a fully automated EFT quantification approach for a large-scale dataset of CMR images available in the UK Biobank. We evaluated the proposed method on a dataset of 200 random CMR image slices, and our results broadly show strong agreement with the reference annotations, achieving an average Dice similarity coefficient of 92.25 (%), and Hausdorff distance of 2.44 (mm). We present the results from analysing 40,000 subjects of the UK Biobank, and investigate

the association of the EFT volumes with different baseline clinical characteristics and genetic variations. Finally, we present the distribution and relationship of the EFT for three groups of Non-Diabetic, Prediabetic, and Diabetic subjects. These different sub-cohort analyses revealed significant findings which are of high importance to the community and will help to explain the role of EFT in the progress of cardiovascular diseases.

4.1 Introduction

The layers of fat on the surroundings of the heart have several physiological functions such as local distribution and regulation of vascular flow through a molecular mechanism, immune barrier, protecting the myocardium from inflammatory and pathogenic substances; local source of fatty acids for the myocardium during of high-demand moments, and thermogenic effects related to brown adipose tissue [77, 78]. However, these types of fats are also correlated to several health risk factors such as atherosclerosis, carotid stiffness, coronary artery calcification, atrial fibrillation and many others. They can release and uptake free fatty acids and to affect low glucose utilisation, which plays an essential role in metabolic syndrome and Cardiovascular Disease (CVD) [78–80].

In particular, Epicardial Fat Tissue (EFT) is the visceral adipose tissue surrounding epicardium, the outer layer of the cardiac myocardium. However, the terminologies used to define fat deposits surrounding the heart in the current literature is diverse and, to some extent, confusing. There is no consensus regarding its exact anatomical definition, location, and the gold standard modality (ultrasound, cardiac computed tomography (CT), or cardiovascular magnetic resonance (CMR)) for an *in vivo* quantification of EFT. Consequently, different studies quantify different regions or layers of cardiac fat deposited around the epicardium as EFT [81–84]. A previous study by Homsí et al. [85] also has shown that the quantities of different layers of fat are highly correlated, suggesting measuring the total fat around epicardium is sufficient to study and analyse cardiac fat structure and function.

Thus, since the standard protocol for cardiac fats is currently not adequately established in the literature, we decided to choose the anatomical definition in which EFT is the total fat located just around the epicardial wall, which also agrees with

several studies, such as Thomas et al. [86], Brinkley et al. [87], and Raggi et al. [88]. Moreover, this definition also matches well with our dataset of interest, i.e. CMR images of the UK Biobank (UKB) where there is no change in pixel intensity for different types of adipose tissue surrounding epicardium, showing a homogeneous region of fat for different layers of fat, as shown in Figure 4.1.

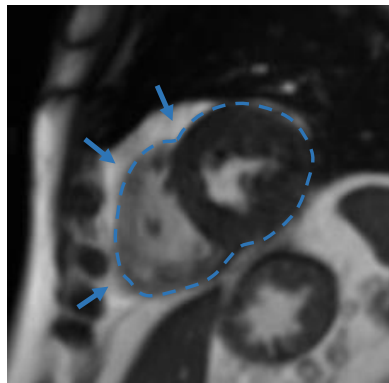


Figure 4.1: Sample CMR image showing the adipose tissue (bright regions) surrounding the epicardial boundary (dashed line).

Table 4.1 summarises the recently published attempts that have studied the EFT and its correlation with other biomarkers or baseline CVDs risk factors. As shown in the table, most of the studies perform manual delineation or semi-automatic approaches for the EFT quantification. There have been very few methods that have used fully automated approaches to quantify the EFT regions. This could be due to the fact that the EFT often has unstructured shape and distribution with extremely variant size across the population.

The method proposed by Rodrigues et al. [103], as one of the very early fully automated EFT quantification approaches, has proposed a method which uses a hybrid approach of registration and classification algorithms to perform the segmentation by classifying fat pixels using a set of extracted low-level features. In a recent study, Fulton et al. [104] has proposed an automated algorithm in which the image is transformed into the polar domain. By training two neural networks, the model estimates the inner and outer edge of the EFT for identifying the thickness of the fat at a radial location

4. QUANTITATIVE POPULATION ANALYSIS OF EPICARDIAL FAT TISSUE

Table 4.1: Summary of the most recent articles studied EFT and its association with other biomarkers.

Study	Method	Modality	Number of subjects
Wu et al. (2017) [89]	Manual	MRI	180
Woerden et al. (2018) [90]	Manual	MRI	64
Mahajan et al. (2018) [91]	Manual	MRI	26
Henningsson et al. (2020) [92]	Manual	MRI	15
Aliyari et al. (2018) [93]	Manual	CT	44
Seker et al. (2017) [94]	Manual	Echocardiography	226
Chen et al. (2017) [95]	Manual	Echocardiography	167
Philouze et al. (2018) [96]	Manual	Echocardiography	79
Wang et al. (2017) [97]	Manual	Echocardiography	98
Militello et al. (2019) [98]	Semi-Automatic	CT	145
Marwan et al. (2019) [99]	Semi-Automatic	CT	127
Mangili et al. (2016) [100]	Semi-Automatic	CT	97
He et al. (2020) [101]	Automatic	CT	40
Zhang et al. (2020) [102]	Automatic	CT	20
Rodrigues et al. (2016) [103]	Automatic	CT	20
Fulton et al. (2020) [104]	Automatic	MRI	12

in a slice. In a similar study, He et al. [101] has proposed to use a deep attention U-Net method to learn from the manual labels to segment the EFT for cardiac CT image directly. In another recent study by Zhang et al. [102], they have put forward a method for automatic segmentation and quantification of EFT from CT scans. They use two U-Nets; one to detect the pericardium, and the other to refine the inside region of the pericardium to segment the EFT.

However, the number of train and test dataset in all studies mentioned above are very limited to truly evaluate the performance of the method. The extreme morphology variations pose challenges and obstacles when attempting to segment and quantify the EFT fully automatically. Consequently, the studies have been published to date have analysed very small datasets, of which population generalisability and normative reference values are under question.

Furthermore, the variation within the structure and shape of the EFT lead to significant difference and remarkable inter/intra-observer variability in reference ranges when manually delineating the EFT regions. For instance, as reported by Li et al. [78] and Bertaso et al. [77], the difference between reference ranges of EFT is incredibly significant. This inconsistency makes those limited manually driven labels sub-optimal

for the purpose of training a reliable and generalisable machine (or deep) learning approach for fully automated EFT quantification method.

Thus, to thrive in data-poor environments where both the quantity and quality of available data is limited by the challenges and costs involved in its collection, automated image analysis methods must leverage available information as much as possible to extract the most relevant information. As explained above, this is particularly obvious in the research area of EFT quantification, where EFT regions are often quantified by manually delineating acquired 2D slices for a very limited cohort.

However, due to the importance of the function of other cardiac regions (such as ventricles, myocardium, and atria) and their correlation with other CVDs risk factors, and also, their coherent and consistent structure within different populations, quantification of cardiac chambers have received considerable attention, and there are various labels and reference contours of cardiac regions available in the community. In terms of population sample size, experimental setup, and quality control, the most reliable reference ranges for cardiovascular structure and function found in the literature are those reported by Petersen et al. [36], in which CMR scans were manually delineated and analysed by a team of eight expert observers using the commercially available cvi42 post-processing software (Version 5.1.1, Circle Cardiovascular Imaging Inc., Calgary, Canada). This rich source of information has enabled the community to develop very accurate fully automated segmentation methods of cardiac chambers, such as our proposed methods presented in previous chapters [48, 58, 61] where we successfully analysed two million CMR images of 40,000 UKB subjects. In this chapter, we use this privilege information available from the rich source of accurate automatic delineation of the cardiac ventricles to look for the fat pixels around the cardiac epicardial wall from all the available two million UKB CMR images of 40,000 subjects.

The main contributions of this chapter are as follows. First, we propose a novel approach of EFT quantification using a pre-developed and validated cardiac segmentation approach and providing volumetric reference ranges for EFT. We perform this through a fully scalable framework, capable of processing arbitrarily large population imaging studies, in a completely automatic manner. We demonstrate this by processing 40,000 subjects from the UKB study, each comprised of 50 time frames for a total of two million image volumes. Second, we look into the EFT volume and its association with

other risk factors in a very large scale dataset. Having analysed such enormous dataset of subjects have revealed interesting findings that will be discussed in the following sections.

The remainder of this chapter is organised as follows. In Section 4.2, we present our strategy for EFT quantification and an evaluation of our method both from technical and clinical perspectives. In Section 4.3, we look into the association of the EFT with different baseline clinical characteristics and genetic variations of all 40,000 UKB subjects. Finally, in Section 4.4, we present final remarks.

4.2 Method

In the sections that follow, we describe the data and methods used within our framework and our corresponding design choices.

4.2.1 Image Analysis Method

As discussed in the previous section, we propose to use the privilege information given by the advances in method developments for the accurate and fully automated segmentation of cardiac ventricles. In this approach, instead of directly segmenting the EFT regions in the CMR image volumes, we take advantage of using an automated segmentation approach to limit our region of interest to the area surrounding the epicardial boundary. This approach allows looking for fat pixels in a minimal search region and intensity variation compared with the entire CMR image volume.

Here, for the large-scale analysis of cardiac ventricles, we utilise our proposed fully automatic image parsing workflow [58] in Chapter 2, where we have extensively validated the workflow on the UKB CMR imaging study. We briefly review the pipeline; CMR images first go through the pre-processing and data organisation, then flow into both the image quality analysis and segmentation phases, which in turn communicate with one another, next, produce 3D surface meshes of cardiac ventricles of the full cardiac cycle. Finally, the 3D surface meshes fed into quality evaluation and quantification. The output 3D mesh models are comprised of two structures; the Left Ventricle (LV) and the Right Ventricle (RV). The LV is a closed water-tight mesh comprising both endo and epicardial walls. The RV is an open mesh representing only the RV

endocardium. The RV has two openings, the atrioventricular valve opening, and the pulmonary valve opening.

Having access to this modular pipeline within the in-house cloud-based image analytics framework called MULTI-X¹ [38], we developed a new module for the extraction and volumetric quantification of fat tissue around the epicardial boundary. Figure 4.2 illustrates the overview of our proposed approach to quantify the EFT regions of the UKB CMR images.

The EFT quantification method works as follows. The output 3D mesh model of the cardiac ventricle is intersected by the image slices, producing 2D contours spanned by the intersections of the mesh triangles. The 2D contours are obtained from the intersections of the epicardial wall mesh with the image slices by first propagating them to the nodes of the mesh, and projecting them onto the local surface normals. Then, the 2D patches of search regions are sampled around the intersections (2D contours) and along the local surface normals at different image planes.

Consequently, as the sampled patches around the epicardium cover a small fraction (about 1.5 %) of the entire CMR image volume, a simple clustering approach can detect the fat pixels from non-fat pixels. Here, we use k-means clustering [105] method on the grey values of a 3D volume patch surrounding the epicardial wall mesh to detect the EFT regions. k-means method is a widely used unsupervised machine learning algorithm for data clustering. Here, given an image patch of n pixels (x_1, x_2, \dots, x_n) , k-means clustering aims to partition the n pixels into k ($\leq n$) regions (clusters) $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$ by minimising the within-cluster sum of squares with the following objective function:

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2, \quad (4.1)$$

where μ_i is the mean of pixels in S_i . Details of the algorithm we used to achieve this effect can be found in the work published by Arthur and Vassilvitskii [105]. Here we set $k = 2$, and as a result of the k-means, two tissue clusters exist in the extracted patches around the epicardial wall; bright and dark regions. Blood and fat regions have similar intensity and are bright while myocardium and other neighbouring regions are dark. So

¹<https://multi-x.org>

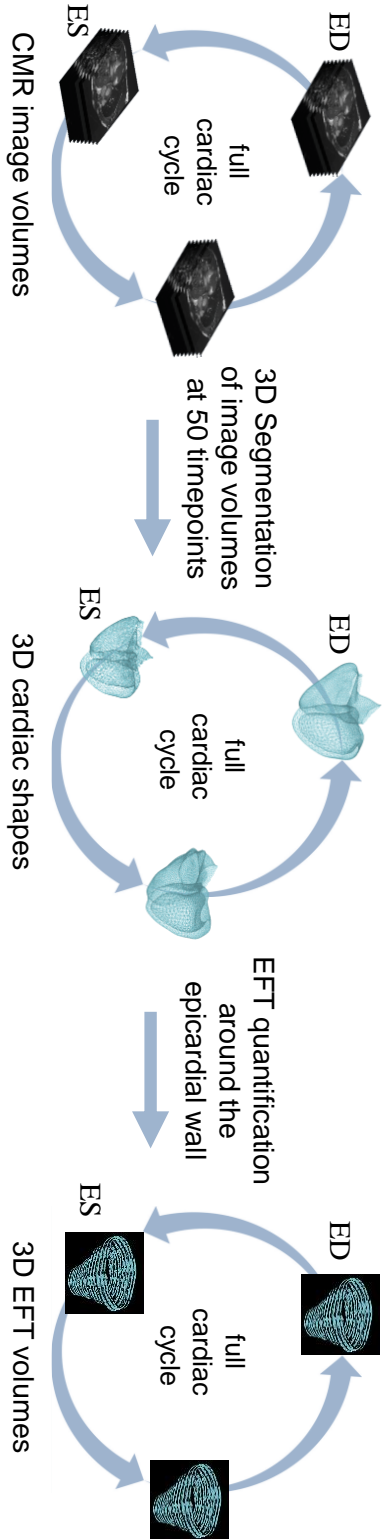


Figure 4.2: The proposed method using the CMR images produces 3D mesh models of the cardiac ventricles in a full cardiac cycle. Then EFT is quantified within the search region around the epicardial wall.

the bright areas are either fat or blood. Then the relative ordering of the tissues with respect to their spatial representation from the centre of the cardiac are as follows: the bright outer cluster is accounted as fat, while the bright inner cluster as blood. This approach can be summarised as follows:

(1) *Search region*: extracting a rectangular image patch at each intersection between the mesh and all image planes and grey values from the patches are pooled accordingly.

(2) *Clustering*: using k-means clustering operation for the image patches, grey values are classified as bright or dark based on the intensities of fat/blood and myocardium/others.

(3) *Quantification*: pixels belonging to the bright cluster are considered as EFT if they are spatially located outside the epicardial wall within the transition from cardiac centre to outside. Finally, the EFT 3D volume can be approximated by summing the areas of all the fat segments surrounding the epicardial wall and multiplying by the inter-slice spacing. EFT volume for a single subject is a median of all the EFT volumes over the full-cardiac cycle.

4.2.2 Data

We performed experiments on 2,000,000 cine CMR image volumes from the UKB using all 50 time points of 40,000 subjects. A short-axis image stack of each time points typically consists of 10 image slices with the in-plane image resolution of $1.8 \times 1.8 \text{ mm}^2$, slice thickness of 8.0 mm and slice gap of 2.0 mm. Table 4.2 shows the summary of the clinical characteristics of subjects of the UKB, including both continuous and categorical variables. In section 4.3, we will perform a thorough study of different sub-cohorts according to different clinical characteristics.

4.2.3 Method Validation

Although we extensively evaluated the performance of our fully automated segmentation method for the cardiac ventricles in Chapter 2 we also evaluated the performance of our fully automated EFT quantification method compared with the manual contours. As there are no publicly available reference manual contours available for the EFT regions, we randomly sampled 200 short-axis slices from the dataset and manually delineated the EFT regions using MATLAB *imageSegmenter* app (version R2019a).

4. QUANTITATIVE POPULATION ANALYSIS OF EPICARDIAL FAT TISSUE

Table 4.2: Summary of the clinical characteristics of 40,000 subjects of the UKB analysed in this study. Continuous values indicate mean \pm standard deviation.

Variables	Range
Female	52%
Age (years)	59 \pm 8
Height (cm)	170 \pm 9
Body mass index (BMI) (kg/m ²)	27 \pm 4
Weight (kg)	78 \pm 16
Body surface area (m ²)	1.8 \pm 0.2
Systolic blood pressure (mmHg)	139 \pm 19
Diastolic blood pressure (mmHg)	82 \pm 11
Pulse rate (bpm)	69 \pm 12
HbA1c (mmol/mol)	35.0 \pm 5.2
Cholesterol (mmol/L)	5.7 \pm 1.1
C-reactive protein (mg/L)	2.5 \pm 4.3
Glucose (mmol/L)	5.1 \pm 1.2
HDL Cholesterol (mmol/L)	1.4 \pm 0.3
IGF-1 (nmol/L)	21.3 \pm 5.6
LDL direct Cholesterol (mmol/L)	3.5 \pm 0.8
Triglycerides (mmol/L)	1.7 \pm 1.0

Figure 4.3 illustrates eight sample image slices from different subjects, different anatomical location, and different time-points in a cardiac cycle, with their corresponding manual and automatic segmentation. It confirms that the system is capable of accurately segmenting EFT regions very similar to the reference contours obtained by manual delineations.

Then, to measure the accuracy of our method, we used the standard metrics, including Dice Similarity Coefficient (DSC), Mean Contour Distance (MCD) and Hausdorff Distance (HD) between reference and predicted contours, explained in details in Section 1.5.2. Table 4.3 presents the mean and standard deviation (SD) of DSC, MCD and HD measures that compare automated and manual EFT segmentation results of 200 CMR image slices. The mean and standard deviation of DSC is 92.251 ± 6.125 , indicating excellent agreement between manual delineations and automated segmentation. We observe that the MCD is 1.534 ± 0.446 mm which is smaller than the in-plane pixel spacing range of 1.8 to 2.3 mm. The HD measure is 2.448 ± 1.336 mm, and slightly larger than the in-plane pixel spacing, however, still within an acceptable range.

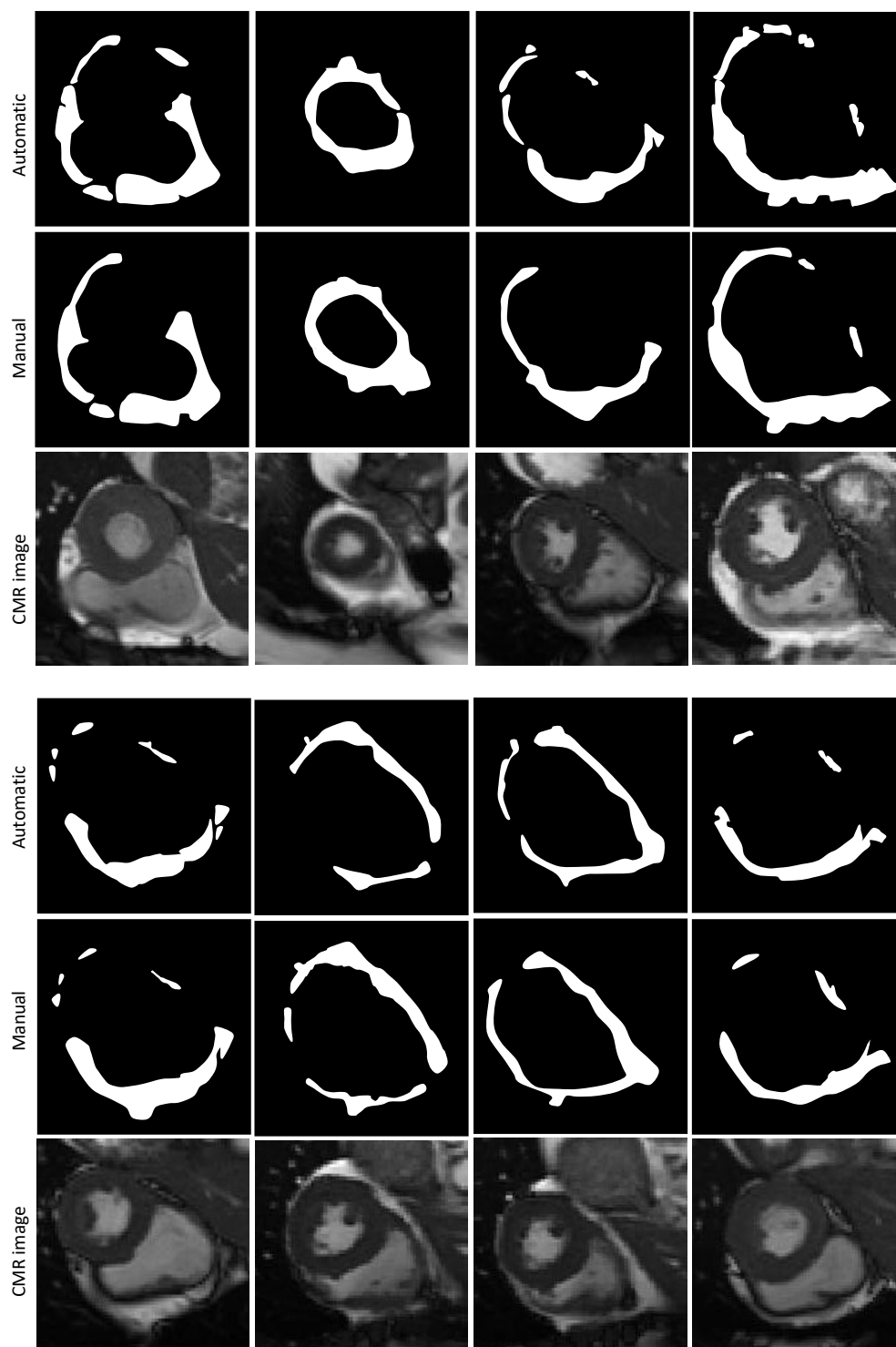


Figure 4.3: Example segmentation results of eight random short-axis slices of different subjects at random phases of a cardiac cycle illustrating the quality for automated segmentation versus manual.

Table 4.3: Segmentation accuracy. The comparison of the automated segmentation results with the reference manual contours in terms of DSC(%), MCD(mm) and HD(mm).

	DSC (%)	MCD (mm)	HD (mm)
Mean	92.251	1.534	2.448
SD	6.125	0.446	1.336

In the next section, we report and investigate all the processed and quantified 40,000 subjects of the UKB; each imaged at 50 time points, i.e. in total two million image volumes.

4.3 Statistical Analysis and Interpretation

In this section, we look into the statistical analysis and interpretation of the EFT volumes in different sub populations. In particular, we look into the sub-cohort analysis based on the Glycated Hemoglobin (HbA1c) values to identify the association of EFT with HbA1c and different clinical variables. We also perform a genome-wide association study (GWAS) for identifying associations between genetic regions (loci) and EFT volumes of 40,000 UKB subjects.

4.3.1 Association of EFT with HbA1c

HbA1c is measured primarily to determine the three-month average blood sugar level. It can be used as a diagnostic test for diabetes mellitus and as an assessment test for glycemic control. Figure 4.4 shows the distribution of the EFT volume (ml) and HbA1c (mmol/mol) for the entire population. The mean and standard deviation of the EFT volume and HbA1c of the 40,000 subjects of the UKB is 17.57 ± 6.06 and 35.04 ± 5.19 , respectively.

To better understand the association between EFT and HbA1c, we investigated the distribution of EFT volume of subjects based on their corresponding HbA1c levels. We categorise the subjects based on their HbA1c into three groups: Non-Diabetic ($\text{HbA1c} < 42$), Prediabetic ($42 \leq \text{HbA1c} < 48$), and Diabetic ($\text{HbA1c} \geq 48$). Figure 4.5 shows the distribution of EFT in three groups of Non-Diabetic, Prediabetic, and Diabetic, each contributing 95.40%, 2.73%, and 1.87% of the total number of 40,000 subjects. We observe that increases in EFT are closely associated with HbA1c level

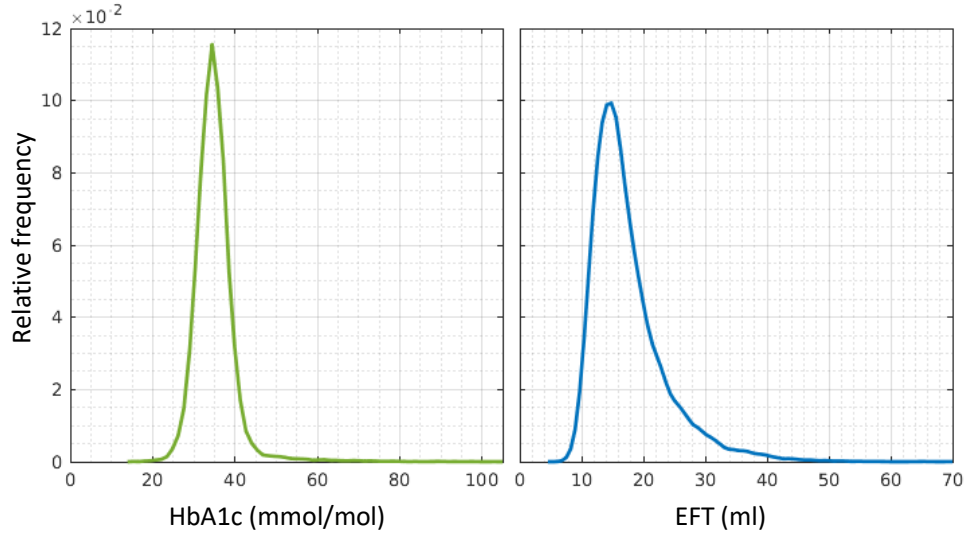


Figure 4.4: Distribution of EFT volume (ml) and HbA1c (mmol/mol) of 40,000 subjects of the UKB.

and, in general, groups with a higher amount of HbA1c have a higher amount of EFT with p -value of less than 0.0001.

4.3.2 Association of EFT with Demographics

In the previous section, we observed that the average EFT volume in Diabetic group is significantly more than Prediabetic and in Prediabetic more than Non-Diabetic. Here, we look into the EFT changes in each category stratified by some of the primary demographics data. Besides, there has been a surprising lack of consensus in the literature on the impact of gender, BMI, and age on the amount of EFT. Taking this into consideration, and to further understand the distribution of EFT volumes in three groups of Non-Diabetic, Prediabetic, and Diabetic, we also look into the sub-cohort analysis within each of the three groups based on the three main characteristics of the subjects: gender, BMI, and age. Figure 4.6 shows the EFT volumes of each group separated by gender. We observe that the female population consistently in all three groups have a significantly higher amount of EFT. In a similar plot, in Figure 4.7, we see subjects with BMI over 25 who are considered as overweight subjects, in all three groups of HbA1c levels have a higher level of EFT, indicating risk of developing additional CVDs risk factors. Correspondingly, the EFT distribution in two age groups of 45-60 and 60-75

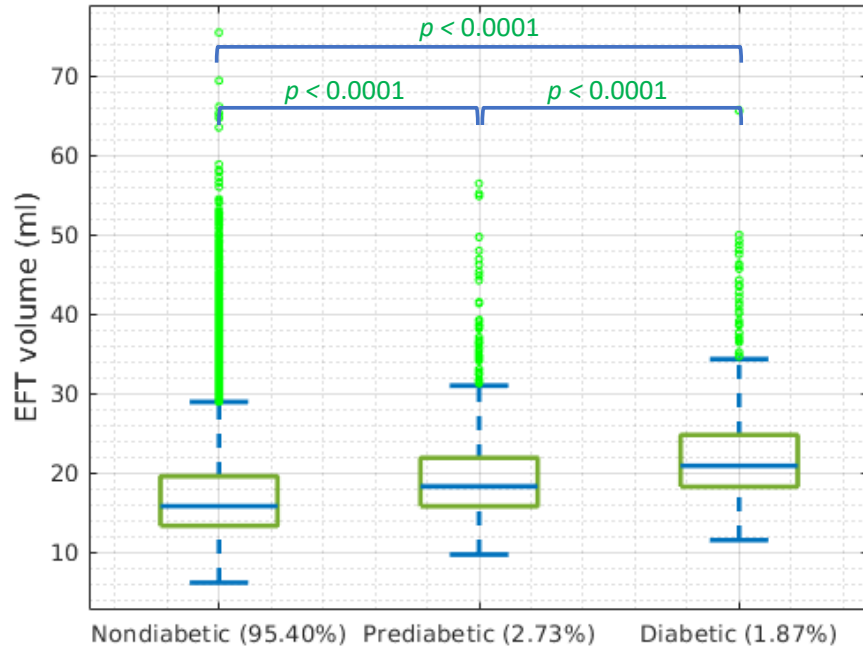


Figure 4.5: Distribution of EFT over three different groups of Non-Diabetic, Prediabetic, and Diabetic shows an extremely significant difference. Numbers in parenthesis shows the percentage of each category in the entire population.

is shown in Figure 4.8, where we see the older age group have significantly higher EFT volume in all three HbA1c groups. We notice there is a broad individual variation in the amount and distribution of EFT attributed to their demographic characteristics. We see that having the opportunity of looking into these patterns in large-scale datasets of this size, and revealing new findings or confirming the findings in the literature but with much more substantial evidence is only plausible where can run fully automated analysis and experiments at this scale.

4.3.3 Association of EFT with Cardiac Function

In this section, we look into the association of EFT volume with the main cardiac function indexes. Here, we use our automatic four-chamber CMR quantification method presented in Chapter 3 to compute the following indexes; LVEDV, LVESV, LVEF, LVM, RVEDV, RVESV, RVEF, LAEDV, LAESV, LAEF, RAEDV, RAESV, and RAEF.

Table 4.4 shows the mean and standard deviation of the above-mentioned cardiac

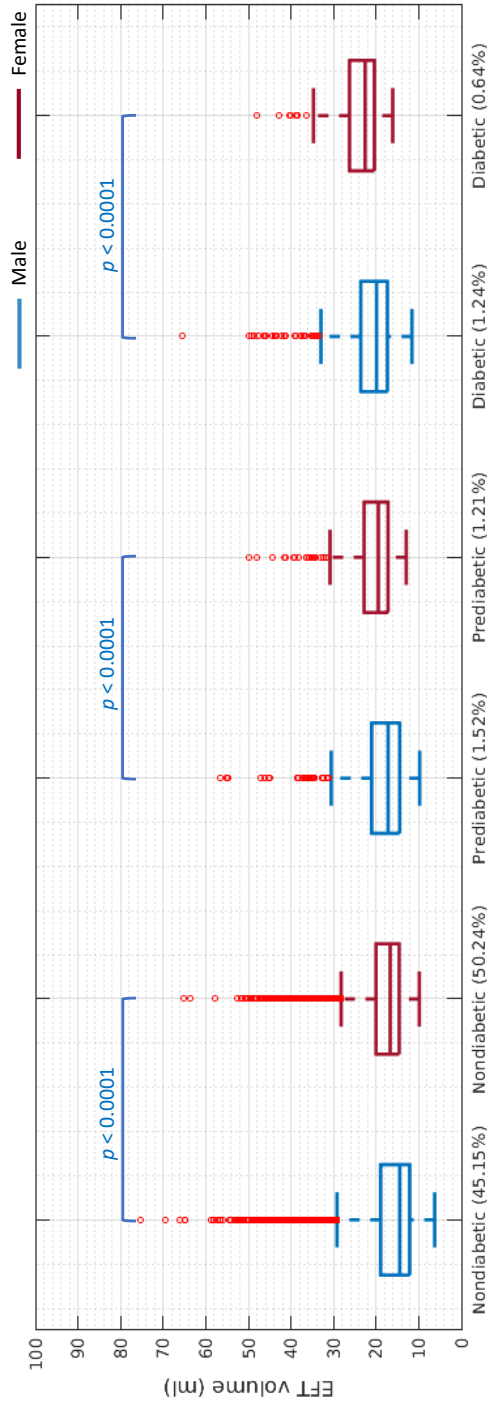


Figure 4.6: Distribution of EFT for male and female subjects. Female subjects have higher EFT compared with male. Numbers in parenthesis shows the percentage of each category in the entire population.

4. QUANTITATIVE POPULATION ANALYSIS OF EPICARDIAL FAT TISSUE

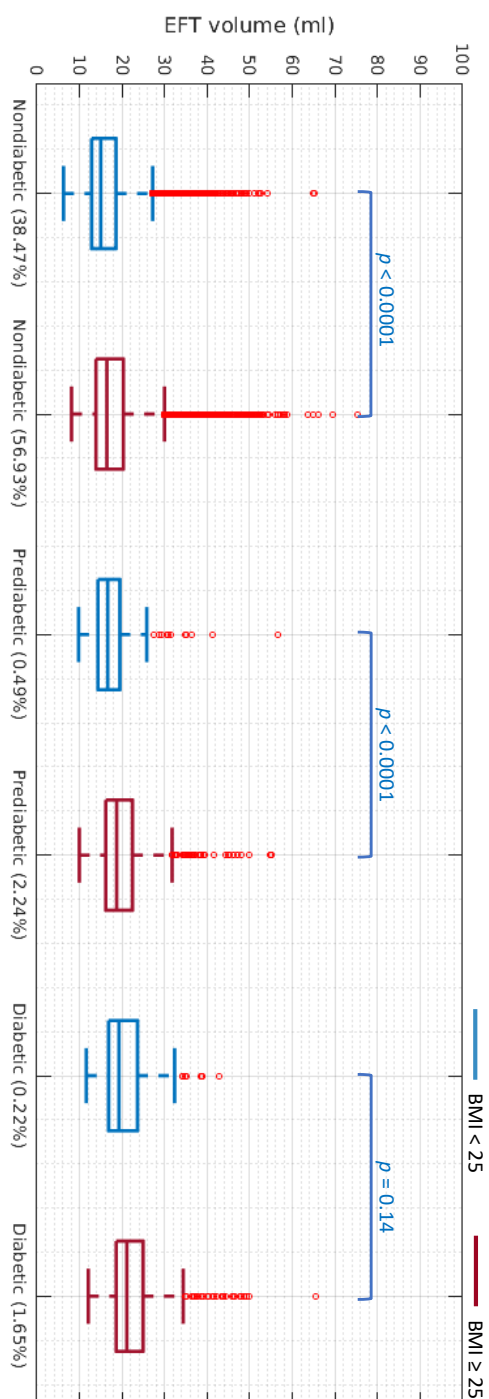


Figure 4.7: Distribution of EFT for subjects with normal BMI and overweight subjects have higher EFT. Numbers in parenthesis shows the percentage of each category in the entire population.

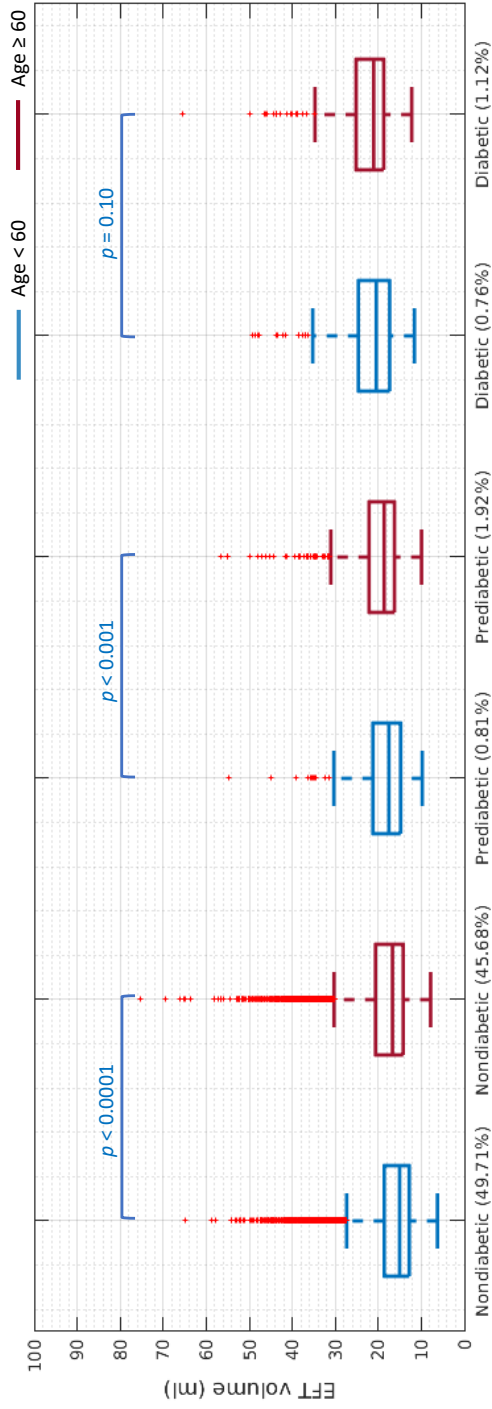


Figure 4.8: Distribution of EFT for two age groups. Older subjects have higher EFT compared with younger. Numbers in parenthesis shows the percentage of each category in the entire population.

function indexes for two groups of subjects with low (less than the median) and high (greater than the median) amount of EFT in different HbA1c levels (Non-Diabetic, Prediabetic, and Diabetic). The proportion of the subjects with low or high EFT also has been reported in each HbA1c level group. We see that LV indexes are mainly affected by the level of fat. Interestingly, in the first three columns, we see that all LV driven indexes, RVESV, and RVEF of subjects with low and high EFT are different (p -value <0.001). We do not see any effect on LA indexes while RAEDV and RAEF change significantly (p -value <0.001). In the Prediabetic group, we see a similar pattern for LV driven indexes while there is no significant change for RV. In the Prediabetic group, we also see that LVEF, RVEF, and RAEF of subjects with high EFT is significantly higher than those of low EFT. Furthermore, in the set of Diabetic subjects, LV, RV, and RA volume at both end-diastolic and end-systolic phases are significantly higher for those with less EFT (p -value <0.05).

4.3.4 Association of EFT with Genetic Variations

We performed a genome-wide association study on the quantified EFT volumes, using the genotypic data provided by the UKB, which covers more than 800,000 genetic variants, i.e. single-nucleotide polymorphism (SNPs) and short indels. We adjusted normalised EFT values for gender, height, BMI, age and diastolic blood pressure by performing linear regression against these covariates and taking the residues as the phenotype of interest. The reason is that these particular covariates have been found to correlate with cardiac function indexes, and we wish to correct for non-genetic factors affecting our phenotypes of interest. Finally, we inverse-normalised the resulting phenotypes. These values were tested against each genetic variant across the autosomes, which is encoded as 0, 1 or 2 depending on the dosage of the non-reference allele that the individual has for that particular variant. Only variants on autosomes (chromosomes 1 through 22) were tested.

In particular, we tested each genetic variant, $X_i \in \{0, 1, 2\}$, for association with EFT value z through a univariate linear additive model of genetic effects:

$$z = \beta_i X_i + \epsilon_i \tag{4.2}$$

where ϵ_i is the component not explained by the genotype, assumed to be normally

Table 4.4: Association of the low and high volumes of EFT with cardiac function indexes of all the UKB subjects, Non-Diabetic, Prediabetic, and Diabetic subjects. **Bold** indicates significance level of 5%.

	All Subjects			Non-Diabetic			Prediabetic			Diabetic		
	Low 57%	High 43%	p-value	Low 58%	High 42%	p-value	Low 35%	High 65%	p-value	Low 13%	High 87%	p-value
Female	46%	61%	<0.001	46%	62%	<0.001	27%	54%	<0.001	5%	39%	<0.001
Age	57 ± 8	60 ± 7	<0.001	57 ± 8	60 ± 7	<0.001	61 ± 7	63 ± 6	<0.001	56 ± 7	61 ± 7	<0.001
BMI	26 ± 4	27 ± 5	<0.001	26 ± 4	27 ± 4	<0.001	28 ± 5	31 ± 6	<0.001	29 ± 5	32 ± 6	<0.001
LVEDV(ml)	143 ± 35	140 ± 34	< 0.001	143 ± 35	140 ± 34	< 0.001	139 ± 38	138 ± 37	=0.713	150 ± 32	141 ± 35	= 0.017
LVESV(ml)	60 ± 19	58 ± 20	< 0.001	60 ± 19	57 ± 19	< 0.001	60 ± 27	58 ± 24	=0.148	68 ± 22	61 ± 23	= 0.010
LVEF(%)	59 ± 7	59 ± 7	< 0.001	59 ± 7	59 ± 7	< 0.001	57 ± 7	59 ± 8	= 0.004	55 ± 8	57 ± 8	= 0.034
LVM(g)	89 ± 22	89 ± 21	< 0.001	89 ± 22	88 ± 21	< 0.001	95 ± 23	93 ± 23	=0.245	104 ± 19	98 ± 22	= 0.014
RVEDV(ml)	149 ± 38	148 ± 37	=0.092	149 ± 38	149 ± 37	=0.448	144 ± 36	143 ± 36	=0.559	155 ± 31	145 ± 35	= 0.015
RVESV(ml)	69 ± 22	68 ± 21	= 0.029	69 ± 22	68 ± 22	=0.237	68 ± 22	66 ± 22	=0.094	75 ± 19	68 ± 20	= 0.003
RVEF(%)	54 ± 7	54 ± 7	= 0.021	54 ± 7	54 ± 6	=0.057	53 ± 7	55 ± 7	= 0.006	52 ± 8	53 ± 7	=0.066
LAEDV(ml)	28 ± 16	28 ± 16	=0.969	28 ± 15	28 ± 16	=0.447	30 ± 20	31 ± 19	=0.713	29 ± 17	30 ± 18	=0.637
LAESV(ml)	66 ± 22	66 ± 22	=0.932	66 ± 22	66 ± 22	=0.659	65 ± 25	68 ± 24	=0.158	67 ± 23	68 ± 24	=0.717
LAEF(%)	59 ± 11	59 ± 11	=0.113	59 ± 11	59 ± 11	= 0.014	57 ± 13	57 ± 12	=0.723	58 ± 13	57 ± 13	=0.810
RAEDV(ml)	48 ± 20	46 ± 21	< 0.001	47 ± 20	46 ± 21	< 0.001	48 ± 24	43 ± 21	= 0.001	48 ± 25	43 ± 19	= 0.018
RAESV(ml)	79 ± 26	80 ± 27	=0.175	79 ± 25	80 ± 27	< 0.001	76 ± 27	73 ± 28	=0.148	79 ± 31	72 ± 24	= 0.014
RAEF(%)	41 ± 12	43 ± 13	< 0.001	41 ± 12	43 ± 13	< 0.001	38 ± 14	41 ± 13	< 0.001	39 ± 14	41 ± 13	=0.180

distributed. The null hypothesis tested is that $\beta_i = 0$.

The outcome of GWAS is usually represented via the so-called Manhattan plots. In these plots, the horizontal axis corresponds to the genomic position of the SNP, where the different chromosomes are juxtaposed along the axis. The vertical axis measures the strength of the association, usually via $-\log_{10}(p)$, where p is the p -value of the association as derived from the linear regression in Equation 4.2. Figure 4.9 shows the GWAS results displayed in the Manhattan plot where the red and blue lines indicate the Bonferroni significance threshold ($0.05/800000$), and a less stringent threshold (10^{-5}), respectively.

We observe that the significant genetic loci found are located in chromosomes 2 and 10, where they have been mapped to genes TTN and BAG3. Both genes have been previously reported to be associated with LV phenotypes [106]. TTN encodes for protein titin, which is responsible for the sarcomere assembly of the myocytes, which determines stretching, contraction and passive stiffness of the myocardium, and protein-truncating TTN variants have also been shown to be responsible for dilated cardiomyopathy. The protein that BAG3 encodes for has been shown to have several functions, and also to be a cause for dilated cardiomyopathy [106]. More investigation of these results and unveiling the association of these two loci with the role of EFT will require further research which is beyond the scope of this work. We also acknowledge some limitations in our study. Despite being the largest GWAS of EFT phenotype, the relatively small discovery sample size translated to the discovery of two loci. However, the expected collection of 100,000 CMR images in the UKB, together with our proposed pipeline, will probably support future studies to detect more genetic loci.

4.4 Summary

EFT has several systemic effects and plays a vital role in developing both insulin resistance and CVDs. In the last couple of years, particular attention has been paid to the EFT in which non-invasive cardiac imaging techniques can be exploited to quantify those regions of interest. EFT volume has a relatively strong genetic and demographics dependence, similarly to other cardiac sub-structures. In this chapter, the anatomical and visual characteristics of the EFT were reviewed. In particular, we presen-

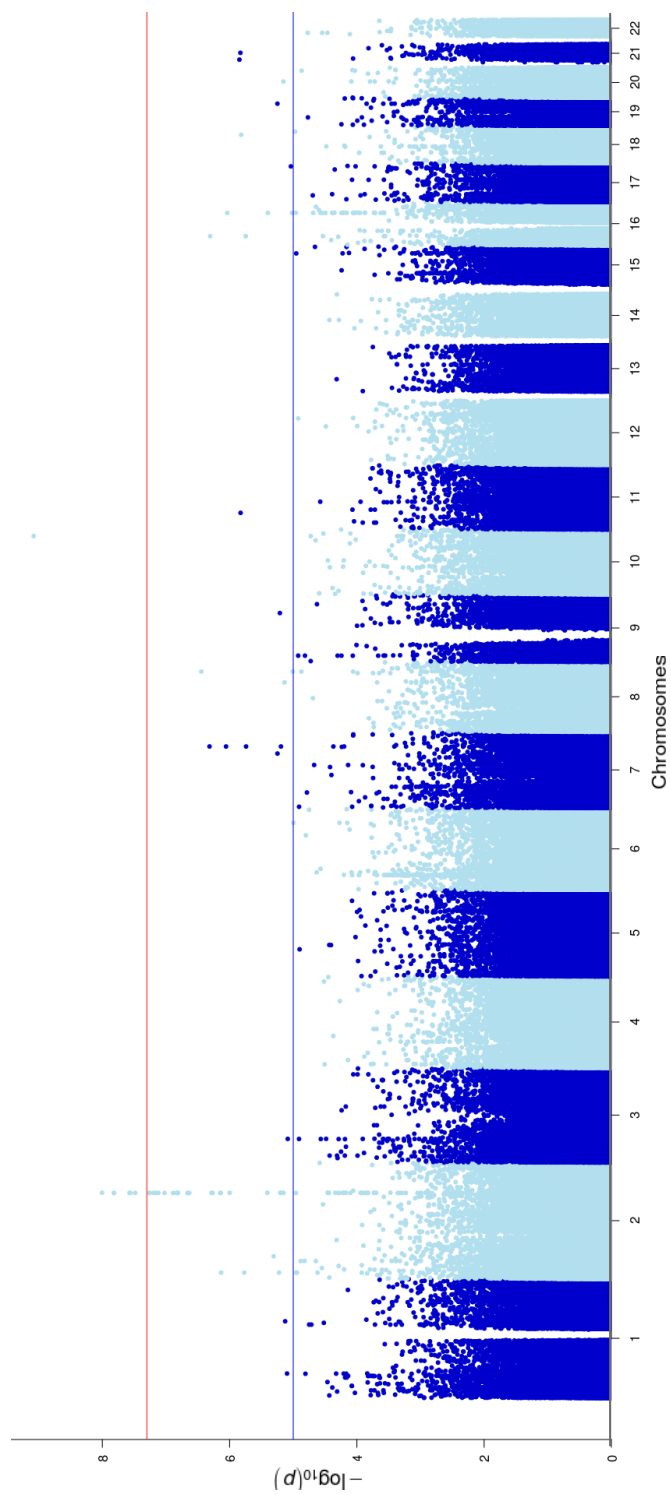


Figure 4.9: Genomic loci associated with EFT. Manhattan plot depicting the genome-wide association study results of EFT traits. The red line indicates the Bonferroni significance threshold ($p < 10^{-5}$) is highlighted in blue.

4. QUANTITATIVE POPULATION ANALYSIS OF EPICARDIAL FAT TISSUE

ted a fully automatic method capable of performing high-throughput end-to-end EFT quantification. The method was developed based on our previously developed cardiac segmentation method, which was extensively validated. We used the privilege information given by our cardiac segmentation method as accurate contours of the epicardium wall. Then we exploited a clustering approach to identify the fat pixels in a minimised search region around the epicardial wall. We validated our EFT quantification method on a reference cohort of 200 random slices for which manual delineations exist. We processed all the available CMR images of 40,000 subjects of UKB and generated the corresponding EFT volume for every single subject. To the best of our knowledge, this is the first attempt to analyse the EFT in the UKB and the first EFT quantification and analysis at this scale. All the previous studies have been published to date, have used manual delineation or semi-automatic approaches for the EFT quantification on tiny datasets, of which population generalisability and normative reference values are under question.

Having quantified EFT for the whole dataset, we looked into the EFT distribution in three main HbA1c based clinical conditions, i.e. Non-Diabetic, Prediabetic, and Diabetic. We found that EFT volume for groups with a higher level of HbA1c was significantly higher. We also observed that the older, female gender and overweight groups have significantly higher EFT than the younger, male gender, and normal weight range group. Finally, we looked into the association of the genetic variations with EFT volume in the entire population. We discovered two genes that were already known to be linked to the ventricles structure. These are significant findings which would provide insights into the patterns and trends across the diabetic, and prediabetic population at risk of CVDs before it fully develops. As future work, we would like to increase our method's robustness to handle other CMR sequences and image modalities. In addition to UKB, we would like to extend our analyses to establish reference ranges for other available cohorts and look into further sub-cohort analysis of these populations to discover more insight into the development of CVDs.

CHAPTER 5

Conclusions

In this chapter, we summarise the main achievements of this thesis which facilitate the move toward the quantitative image analysis of population cardiac magnetic resonance imaging. Moreover, we explain some of the existing limitations and introduce future directions for the development of new methods that can be built upon the frameworks proposed in this thesis.

5.1 Summary and Achievements

The motivation of this thesis was the development of fully automated workflows to cope with some of the challenges associated with the processing of large amounts of cardiac image data. In particular, we have developed algorithms and methods for the quantification of several cardiac regions in large-scale CMR studies. Our fully automatic pipelines have an excellent potential for improving work efficiency and assisting clinicians in diagnosing and performing large-scale clinical research. In the following paragraphs, we outline the main contributions of the thesis.

Firstly, in **Chapter 2**, we presented a fully automated quality-controlled workflow

capable of performing high throughput CMR image analysis and tested its performance on the UKB cardiac dataset of 40,000 subjects each imaged at 50 timepoints, i.e. two million image volumes. This study has been the first published attempt tackling the fully automatic 3D analysis of the UKB population study, providing reference ranges for all key cardiovascular functional indexes, from both left and right ventricles of the heart.

Secondly, in **Chapter 3**, we presented a new method for the accurate 3D modeling of cardiac four chambers. We proposed a novel deep neural network using both CMR images and patient metadata to directly predict cardiac shape parameters. The proposed method uses the promising ability of statistical shape models to simplify shape complexity and variability together with the advantages of convolutional neural networks for the extraction of solid visual features. The newly proposed method was capable of producing accurate 3D shapes using just a fraction (about 23% to 46%) of the available image data, which is of significant importance to the community as it supports the acceleration of CMR scan acquisitions.

Finally, in **Chapter 4**, we presented a new method for quantification of epicardial fat tissue, a marker of cardiovascular risk. Having developed fully automated methods for cardiac ventricles segmentation, we extended the method for the volumetric quantification of epicardial fat tissue. This study has been the first attempt looking into the large-scale quantification of the epicardial fat tissue and its association with other patient information which revealed new findings such as a significant difference of epicardial fat volume in different sub-cohorts and identification of two genes associated with epicardial fat.

5.2 Limitations and Further Research Directions

Here, we point out some of the improvement and challenges that require further attention for better and better fully automated analytics workflows. As future work, the robustness of methods should be increased to handle more severe morphological variations due to pathology and variable image quality, and extend its generalisability to cope with other CMR sequences and image modalities. Our work's main limitation is that the models were trained only on one dataset, limiting the model's generalis-

ability. The UKB CMR dataset does not include all the possible variability of image quality and CMR protocols existing in different clinical environments. Moreover, most of the data are from a healthy population within a limited age group of 45-75. Although we have shown that the method works well on the UKB pathological cases, there can be different pathological patterns in the clinical environment, which are not currently manifested by the UKB cohort. To ensure that the image-driven phenotypes and biomarkers are consistent across the dataset, UKB has acquired CMR images using a standard imaging protocol through the same scanner model and configurations. However, this also indicates that the trained models have inherited the UKB image patterns and features and might not generalise to other vendors or sequence datasets. It would be of interest to investigate whether we could design a large-scale dataset for training and evaluation, covering a wide range of typical CMR imaging protocols and scanner types. A possible solution to extend these datasets' generalisability is to use the proposed large-scale segmentation pipelines in this thesis to segment other CMR datasets. We observed that our methods could analyse the CMR images in a large cohort very fast, overcoming limitations associated with current clinical CMR image analysis routine, which is manual and time-consuming. Considering our pipelines may not perform the best in all cases, we then correct sub-optimal segmentation contours through a semi-automatic approach and then include the revised contours in the next round of our training phase to improve the generalisability of the model.

Further to this, we foresee including automatic feedback loops in our pipeline that would allow the automatic adjustment of segmentation parameters for image re-processing, based on quality assessment modules. Such feedback loops may include triggering new modules designed to handle low image quality or missing data imputation. Similarly, alternative segmentation techniques could be triggered upon detection of specific pathologies. We also foresee including 3D+t quality control methods to assess the 3D shape accuracy over the full cardiac cycle, allowing more generalised and comprehensive quality control of the output 3D shapes considering the inter-frame dependency along the cardiac cycle.

In addition to the essential above mentioned points for further improvement of the current quantification results and semantic interpretation of the those integrated with other metadata, the other future directions could be carried out on intelligent data

5. CONCLUSIONS

preparation for further analysis. The datasets collected by researchers and clinicians from different imaging sites are currently recorded in varying formats and protocols. This means that they cannot be processed directly by the computational techniques, such as the ones developed in this thesis, without re-organising the image files. However, achieving this wholly and consistently is a technical challenge. It could be addressed by developing deep learning methods to organise the image data based on the image content, not the image metadata recorded in the header of image file which is usually prone to subjective error during the image acquisition. For practical large-scale analysis of the data, the data organisation is an essential pre-processing step.

Besides increasing the robustness of the system, further analysis of reference ranges for specific sub-populations is of high importance. The UKB provides a wealth of patient information including, socio-demographic, lifestyle and environmental, family history, genetic, and omics data. Also looking into the modelling the relationship between cardiac morphology and function, and other available demographics, genetic, and omics data, would help improve our understanding of disease progression, potentially leading to improved treatment planning and delivery.

A potential future direction based on the results of this thesis is patient stratification. By enabling the high-throughput computation of morphological and functional information obtained from a very large population, a new patient could be examined and compared to those of a baseline statistics or intricate data-driven patterns within the population for better decision making during the diagnosis and treatment. This approach is possible through the smart combination of image-driven phenotypes and other clinical, biochemical, environmental, and genetic information. Thus, it confirms the importance of the very accurate quantification of the anatomy of interest, using techniques similar to what developed in this thesis.

In summary, large-scale medical image analytics is an emerging trend that is creating new opportunities, as well as challenges. In the future, we will see the accelerated, extensive implementation and application of large-scale analytics across healthcare-related organisations, so establishing standards and continuously improving the tools and technologies are crucial. In this context, thus, the fully automated quantification and analysis tools that have been presented in this thesis are an essential step toward the capability of achieving large-scale analytics of population CMR imaging.

5.3 Code and Results Availability

Here, we provide some more details about the source code of the methods developed in the three main chapters of this thesis and their corresponding output results.

To scale both data access and computation, we used our in-house cloud-based image analytics framework called MULTI-X ¹ [38]. It enables both secured distributed access to data storage and distributed execution of our workflows on the cloud. In particular, we used Amazon Simple Storage Service (S3) for data storage and Amazon Elastic Compute Cloud (EC2) for cloud computing.

All the input files, intermediate files, and final results generated during the analysis of 40,000 subjects of the UKB through three different workflows (presented in Chapters 2, 3, and 4) are stored in the following directory,

```
s3://results-ukbb-cardio/FULL_ANALYSIS_11350/
```

where they can be accessed through the portal with the prior permissions. Regarding the source code, each of the proposed workflows is located in a separate Amazon virtual Linux machine with their customised software and hardware configurations. The virtual machines are also accessible through the MULTI-X portal with the admin's prior guaranteed access keys. Once obtained the access, a copy of the virtual machine with all its prerequisite packages and libraries can be instantiated and immediately ready to explore, run, or modify. In the following, we provide more details about the structure and interaction of the main components of every workflow.

Chapter 2 : The virtual machine name is `UKB_LVRV_Quantification` with the following icon. The machine contains all the required files to run the workflow.

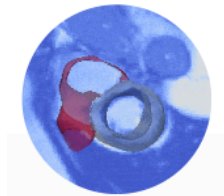


Figure 5.1 shows the architecture of the workflow addressing the issue of large-scale analysis of CMR images. It consists of several modules to analyse every single subject of the database. A workflow manager software package is required to create a modular workflow and enable processing of multiple subjects in parallel. This provides an infrastructure for the set-up, performance and monitoring of a defined sequence of tasks, regardless of their programming language. In our implement-

¹<https://multi-x.org>

5. CONCLUSIONS

ation, the Nipype package [39] has been used. It allows us to combine a heterogeneous set of software packages within a single and highly efficient workflow, processing several subjects in parallel using cloud computing resources.

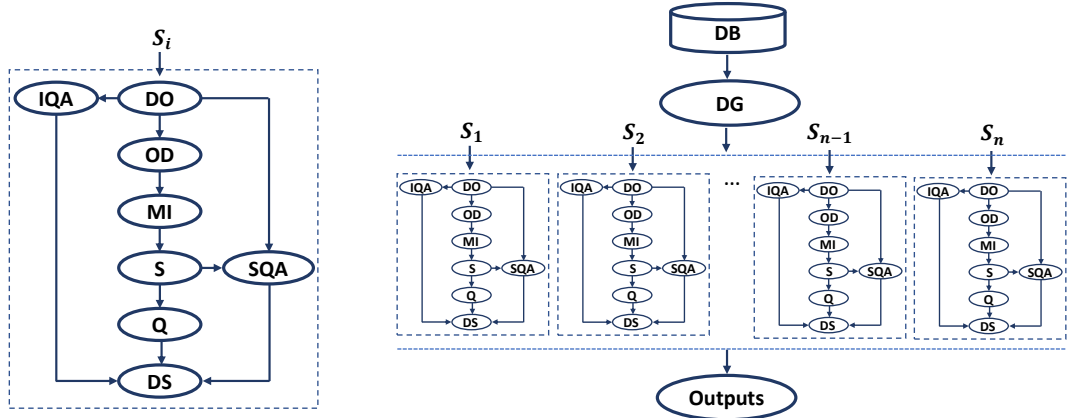


Figure 5.1: The fully automatic image parsing workflow for the analysis of cardiac ventricles in parallel. Left: The workflow includes the following modules: DO: Data Organisation, IQA: Image Quality Assessment, OD: Organ Detection, MI: Model Initialisation, S: Segmentation, SQA: Segmentation Quality Assessment, Q: Quantification, DS: Data Sink. Right: The quantitative functional analysis of a large database in parallel mode. DB: Database, DG: Data Grabber, n : number of subjects, and S_i : i^{th} subject of the dataset.

Under the following directory `/MULTIX/WORKFLOWS/`, there are three main files:

- `input.txt` This file contains the S3 full path of every subject of the dataset required to be processed. Each line corresponds to one subject; therefore, if the plan is to process 1000 subjects, this file should contain 1000 lines of the full path, as shown below.

```
s3://dataset-ukbb-cardio/11350/10xxxxx/1000215/
s3://dataset-ukbb-cardio/11350/10xxxxx/1000336/
s3://dataset-ukbb-cardio/11350/10xxxxx/1000363/
s3://dataset-ukbb-cardio/11350/10xxxxx/1000380/
.
.
```

- `main.py` This is the main python file that works as a wrapper and contains different classes for different modules of the workflow. Each module is man-

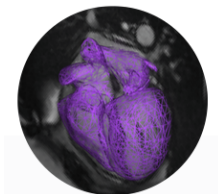
aged by a *node* of workflow with its specific inputs and outputs. After defining all the required classes (nodes of the workflow), they are connected to each other to streamline the data from the unstructured input data all the way to the final output results. The workflow nodes are connected as shown below and can simply be appended, removed or modified to change the workflow structure. Every module's corresponding files and libraries are located under this directory `/MULTIX/ANALYSE/`.

```
cardiacWorkflow = Workflow(name='cardiacPipeline')
cardiacWorkflow.connect([(organisation, iqa, [('input_path', 'input_path')]),
                        (organisation, initialisation, [('input_path', 'input_path')]),
                        (initialisation, segmentation, [('input_path', 'input_path')]),
                        (segmentation, quantification, [('input_path', 'input_path')]),
                        (segmentation, sq, [('input_path', 'input_path')]),
                        (quantification, datasink, [('input_path', 'input_path')]),
                        (iqa, datasink, [('input_path', 'input_path')]),
                        (sq, datasink, [('input_path', 'input_path')]),
                        ])
cardiacWorkflow.run()
```

- **workflow.nf** This is a Nextflow file that sets the high-level configuration of the workflow and its parallelisation details when running on the cluster of virtual Linux machines. In this file, the `main.py` and `input.txt` files, as mentioned earlier, are used. The following command starts the workflow and processes all the subjects listed in the input file using the maximum capacity of computation resources allocated in the `workflow.nf` file.

```
nextflow run workflow.nf
```

Chapter 3 : The virtual machine name is `UKB_4Chamber_SHARE` with the following icon. The machine contains all the required files and packages to generate the shape parameters within a deep learning based approach, presented in details in Chapter 3.



In the deep learning based methods, usually, the training takes a long time, the inference phase is speedy, and the produced model is a big file, each of which is the opposite of the previous strategy. Due to this difference, the pipeline structure and its execution on large scale datasets would differ from the previous approach. Figure 5.2 shows the schematic of the workflow, where, in this case, chunks

5. CONCLUSIONS

of data (green arrows) are addressed sequentially while the analysis of subjects within each chunk is performed in parallel.

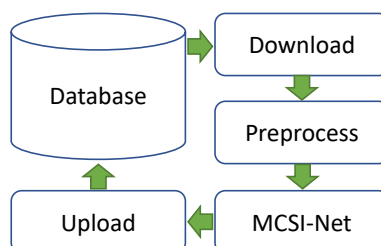


Figure 5.2: The schematic of the fully automatic workflow for the analysis of cardiac chambers. Green arrows indicate a chunk of data to be processed in every round. Chunk size, which is the number of subjects to be processed simultaneously, depends on the virtual machine’s computational resources.

Under the following directory `/MULTIX/APPS/`, there are five main files:

- `inputs.csv` This file contains the S3 full path of every single subject of the dataset required to be processed.
- `main_preprocess.py` This python file preprocesses the input image data where all images are intensity- and spatially-normalised so that every slice has a standard size and intensity values. This file produces `ROI.csv` in which the coordinates of a square fitted to the region of interest is stored. Instead of making a copy of the data and storing the cropped version of the images, we keep the images as they are, and we work with the cropped image’s coordinates to read only the region of interest which reduces the computational time and resources.

`mcsi_net.py` This file contains the network architecture. Different imaging and non-imaging data pass through multiple fully connected layers to learn the tasks and produce the model weights for our application.

`main_mcsi_net.py` This file is the main python file used fro train/test the network. This file uses the network as mentioned earlier, `mcsi_net.py`, and sets the high-level configuration and hyperparameters to train and then test the network, as shown below.

```

if __name__ == "__main__":

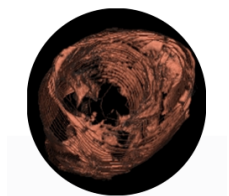
    parser = argparse.ArgumentParser(description='MCSI-Net')
    parser.add_argument('--optimiser', type=str, default='Adam')
    parser.add_argument('--lr', type=float, default=1e-3)
    parser.add_argument('--epochs', type=int, default=300)
    parser.add_argument('--batch_size', type=int, default=8)
    parser.add_argument('--dir_dataset', type=str, default='./input_data/')
    parser.add_argument('--dir_ids', type=str, default='./input_data/ids/ROI.csv')
    parser.add_argument('--percentage', type=float, default=0.75)
    parser.add_argument('--n_b_vals', type=int, default=71)
    parser.add_argument('--n_rot_vals', type=int, default=7)
    parser.add_argument('--sax_img_size', type=list, default=[100, 100, 15])
    parser.add_argument('--lax_img_size', type=int, default=100)
    parser.add_argument('--num_metadata', type=int, default=31)
    parser.add_argument('--save_dir', type=str, default='./results/')
    parser.add_argument('--save_model', type=str, default='mcsi.net')
    parser.add_argument('--train', type=bool, default=True)
    args = parser.parse_args()

```

`main_large_scale.py` This python file has been designed to manage high throughput analysis of a massive data set, for example, CMR images of the UKB where we processed more than two million image volumes. Depending on the computational and storage resources, this script sets some high-level configurations to automatically process the whole dataset using the maximum hardware capacity and our pretrained model. There are three main variables, each responsible for a specific aspect of the machine resources. `chunk_size` defines the chunk size, i.e., the number of subjects to be read and downloaded each round. This depends on the disk size of the virtual machine. `core_numbers` defines the maximum number of parallelisation threads. The higher the number of cores, the quicker CPU based tasks are completed. `batch_size` defines the number of subjects to be processed using the GPU capacity. The more powerful GPU the quicker is the analysis of the chunk. Subsequently, once all the chunks are processed, the results are uploaded to the S3, and the disk is cleaned and available for the next chunk of the data.

Chapter 4 : The virtual machine name is `UKB_EFT_Quantification` with the following icon. The machine contains all the required files and software packages to run and manage the workflow in a parallel mode for simultaneously running program tasks on multiple clusters of virtual machines.

5. CONCLUSIONS



As mentioned in Chapter 4, this module was built up on top of the workflow presented in Chapter 2. The idea was to use the epicardial boundary's accurate segmentation, as an initial starting point, to look for fat pixels in a limited search region. Figure 5.3 shows how the new EFT module is added to the existing workflow to process new subjects. It uses the output of the segmentation module to produce the EFT quantification results.

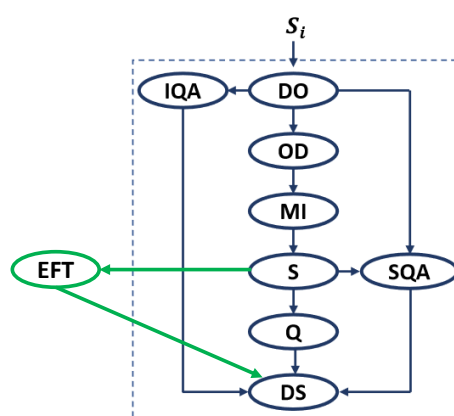


Figure 5.3: Addition of a new module to the existing workflow for the fully automatic analysis of epicardial fat tissue in a large scale dataset. We use the privilege information given by our cardiac segmentation workflow to quantify fat tissue accurately.

For the analysis of new subjects, the cardiac ventricles segmentation workflow should be extended with the following module. It returns the EFT quantification results in the same directory where all the intermediate and final segmentation results are located.

```
def eft_quantification(input_path):
    from nipy.interfaces.base import CommandLine
    mcr = '/MULTIX/APPS/v96'
    eft_quantification_exe = '/MULTIX/ANALYSE/ eft_quantification/run-EFT.sh'
    cli = CommandLine(command=eft_quantification_exe)
    cli.inputs.args = mcr + input_path

    try:
        cli.run()
    except Exception:
        pass

    return input_path

eft = Node(Function(input_names=['input_path'], output_names=['input_path'],
                  function=eft_quantification), name='EFT-Quantification-Module')
```


Under the following directory `/MULTIX/WORKFLOWS/`, there are two main files:

- `inputs.txt` Similar to previous approaches, this file contains the S3 full path of every single subject of the dataset required to be processed.
- `main.py` This is the main python file that extends the workflow with a new module and runs the workflow in the same way that cardiac biventricular segmentation workflow performs using Nextflow and Nipype package. As we already had processed the CMR images and the segmentation contours were available, we only used the output segmentation to quantify fat tissue with the following compact workflow.

```
eftWorkflow = Workflow(name='eftPipeline')
eftWorkflow.connect([
    (resource, download, [('subject_id', 'input_path'])),
    (download, eft, [('input_path', 'input_path'])),
    (eft, datasink, [('input_path', 'input_path']))
])

eftWorkflow.run()
```

However, if we want to process a new set of subjects, the workflow has the following nodes and connections.

```
cardiacWorkflow = Workflow(name='cardiacPipeline')
cardiacWorkflow.connect([
    (organisation, iqa, [('input_path', 'input_path'])),
    (organisation, initialisation, [('input_path', 'input_path'])),
    (initialisation, segmentation, [('input_path', 'input_path'])),
    (segmentation, quantification, [('input_path', 'input_path'])),
    (segmentation, sq, [('input_path', 'input_path'])),
    (segmentation, eft, [('input_path', 'input_path'])),
    (quantification, datasink, [('input_path', 'input_path'])),
    (iqa, datasink, [('input_path', 'input_path'])),
    (sq, datasink, [('input_path', 'input_path'])),
    (eft, datasink, [('input_path', 'input_path'])),
])

cardiacWorkflow.run()
```

5. CONCLUSIONS

PUBLICATIONS

List of Publications

Journal Papers

1. **Attar, Rahman**, et al. "Quantitative CMR population imaging on 20,000 subjects of the UK Biobank imaging study: LV/RV quantification pipeline and its evaluation." *Medical Image Analysis* 56 (2019): 26-42.
2. Xia, Yan, Le Zhang, Nishant Ravikumar, **Rahman Attar**, Stefan K. Piechnik, Stefan Neubauer, Steffen E. Petersen, and Alejandro F. Frangi. "Recovering from Missing Data in Population Imaging - Cardiac MR Image Imputation via Conditional Generative Adversarial Nets." *Medical Image Analysis* (2020): 101812.
3. **Attar, Rahman**, et al. "Automatic 3D+t Four-Chamber CMR Quantification of the UK Biobank: integrating anatomical and demographic priors at scale." (under review)
4. **Attar, Rahman**, et al. "Volumetric Epicardial Fat Quantification of UK Biobank Imaging Study." (under review).

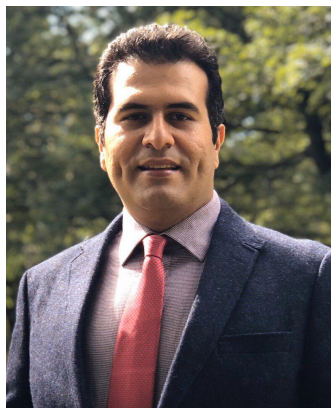
5. Diaz-Pinto, Andres, **Rahman Attar**, et al. "Seeing your Heart through your Eyes: Predicting Myocardial Infarction using Retinal Images and Demographic Data." (under review)
6. Xiang, Chen, **Rahman Attar**, et al. "Learned Generative Shape Reconstruction from Sparse, Incomplete and Noisy Point-Clouds." (under review)
7. Bonazzola, Rodrigo, Andres Diaz-Pinto, **Rahman Attar**, et al. "Unsupervised Representation Learning for Imaging Genetics Discovery via Convolutional Mesh Autoencoders." (under review)

Conference Papers

8. **Attar, Rahman**, et al. "3D Cardiac Shape Prediction with Deep Neural Networks: Simultaneous Use of Images and Patient Metadata." In MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 586-594. Springer, Cham, 2019.
9. **Attar, Rahman**, et al. "High Throughput Computation of Reference Ranges of Biventricular Cardiac Function on the UK Biobank Population Cohort." MICCAI Workshop on Statistical Atlases and Computational Models of the Heart, pp. 114-121. Springer, Cham, 2018.
10. De Vila, Milton Hoz, **Rahman Attar**, et al. "MULTI-X, a state-of-the-art cloud-based ecosystem for biomedical research." In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1726-1733. IEEE, 2018.

BIOSKETCH

Biosketch



Abdolrahman Attar Qazaani (known as Rahman Attar) was born in Tehran-Iran (1989). He graduated from the University of Guilan with a BEng degree in Computer Engineering in 2012. In 2013, he joined Computational Intelligence Laboratory (CIL) at University of Lincoln (UK) where he received a full scholarship to carry out research on image enhancement and registration for the latest generation of wireless endoscopy systems with multiple image sensors, under supervision of Prof. Shigang Yue. He was awarded a Marie Curie Fellowship to be involved in the EU FP7- EYE2E/LIVCODE

. BIOSKETCH

projects during his study. He received MPhil degree from School of Computer Science - the University of Lincoln in 2015 and since then has collaborated with CIL as a research assistant.

In 2016, he joined Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB). He was awarded a Faculty of Engineering Doctoral Academy Scholarship to undertake research in CISTIB under supervision of Prof. Alejandro Frangi. The main aim of his PhD project has been on the development of automatic image analysis methods capable of extracting quantitative imaging biomarkers from large population imaging cohorts in an accurate, robust and efficient manner which has led to some other collaborations with clinicians in the Leeds Institute of Cardiac and Metabolic Medicine (LICAMM) at School of Medicine in Leeds or other external collaborators such as Prof. Stefan Neubauer and Prof. Stefan K. Piechnik (University of Oxford) and Prof. Steffen Petersen (Queen Mary University of London).

He has completed several academic/industrial projects giving him the opportunity to publish several journal and conference papers in the field of image processing, computer vision, and machine learning. Click [here](#) to see the complete list of his publications available at Google Scholar, (currently, total citations = 185 and h-index = 8). He has served as a reviewer for several leading journals and conferences including IEEE Transactions on Medical Imaging, Medical Image Analysis, and MICCAI. Since 2010, he also has had various teaching responsibilities as a tutor, teaching assistant, and lab demonstrator in several BSc/MSc courses such as data structures and algorithms, artificial intelligence, discrete mathematics, advanced programming, and scientific software development.

BIBLIOGRAPHY

- [1] Florin Mihail Filipoiu. *Atlas of heart anatomy and development*. Springer Science & Business Media, 2013.
- [2] Paul A Iaizzo. The Visible Heart® project and free-access website Atlas of Human Cardiac Anatomy. *EP Europace*, 18(suppl_4):iv163–iv172, 2016.
- [3] WHO. Cardiovascular Diseases. https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1/, July 2020. Last checked on Jul 01, 2020.
- [4] NHS. Cardiovascular disease (CVD). <https://www.england.nhs.uk/ourwork/clinical-policy/cvd/>, February 2020. Last checked on Sep 01, 2020.
- [5] Thomas R Dawber, Gilcin F Meadors, and Felix E Moore Jr. Epidemiological approaches to heart disease: the Framingham Study. *American Journal of Public Health and the Nations Health*, 41(3):279–286, 1951.
- [6] WHO MONICA Project Principal Investigators et al. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. *Journal of clinical epidemiology*, 41(2):105–114, 1988.
- [7] Annika Rosengren, Steven Hawken, Stephanie Ôunpuu, Karen Sliwa, Mohammad Zubaid, Wael A Almahmeed, Kathleen Ngu Blackett, Chitr Sitthi-amorn, Hiroshi Sato, Salim Yusuf, et al. Association of psychosocial risk factors with risk of acute myocardial infarction in 11 119 cases and 13 648 controls from 52 countries (the INTERHEART study): case-control study. *The Lancet*, 364(9438):953–962, 2004.
- [8] Elke Hensel, Dietmar Gesch, Reiner Biffar, Olaf Bernhardt, Thomas Kocher, Christian Splieth, Gabriele Born, and Ulrich John. Study of Health in Pomerania

BIBLIOGRAPHY

- (SHIP): a health survey in an East German region. Objectives and design of the oral health section. *Quintessence international*, 34(5), 2003.
- [9] Christopher T Sempos, Diane E Bild, and Teri A Manolio. Overview of the Jackson Heart Study: a study of cardiovascular diseases in African American men and women. *The American journal of the medical sciences*, 317(3):142–146, 1999.
- [10] Diane E Bild, David A Bluemke, Gregory L Burke, Robert Detrano, Ana V Diez Roux, Aaron R Folsom, Philip Greenland, David R Jacobs Jr, Richard Kronmal, Kiang Liu, et al. Multi-ethnic study of atherosclerosis: objectives and design. *American journal of epidemiology*, 156(9):871–881, 2002.
- [11] Ronald G Victor, Robert W Haley, DuWayne L Willett, Ronald M Peshock, Patrice C Vaeth, David Leonard, Mujeeb Basit, Richard S Cooper, Vincent G Iannacchione, Wendy A Visscher, et al. The Dallas Heart Study: a population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. *The American journal of cardiology*, 93(12):1473–1480, 2004.
- [12] Tamara B Harris, Lenore J Launer, Gudny Eiriksdottir, Olafur Kjartansson, Palmi V Jonsson, Gunnar Sigurdsson, Gudmundur Thorgeirsson, Thor Aspelund, Melissa E Garcia, Mary Frances Cotch, et al. Age, gene/environment susceptibility–Reykjavik Study: multidisciplinary applied phenomics. *American journal of epidemiology*, 165(9):1076–1087, 2007.
- [13] Steffen E Petersen, Paul M Matthews, Fabian Bamberg, David A Bluemke, Jane M Francis, Matthias G Friedrich, Paul Leeson, Eike Nagel, Sven Plein, Frank E Rademakers, et al. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK biobank—rationale, challenges and approaches. *Journal of Cardiovascular Magnetic Resonance*, 15(1):46, 2013.
- [14] DP Ripley, TA Musa, LE Dobson, S Plein, and JP Greenwood. Cardiovascular magnetic resonance imaging: what the general cardiologist should know. *Heart*, 102(19):1589–1603, 2016.

- [15] Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai, and Daniel Rueckert. Deep Learning for Cardiac Image Segmentation: A Review. *Frontiers in Cardiovascular Medicine*, 7:25, 2020.
- [16] Kathleen Gilbert, Charlène Mauger, Alistair A Young, and Avan Suinesiaputra. Artificial Intelligence in Cardiac Imaging With Statistical Atlases of Cardiac Anatomy. *Frontiers in Cardiovascular Medicine*, 7, 2020.
- [17] Damini Dey, Piotr J Slomka, Paul Leeson, Dorin Comaniciu, Sirish Shrestha, Partho P Sengupta, and Thomas H Marwick. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *Journal of the American College of Cardiology*, 73(11):1317–1335, 2019.
- [18] Gustavo Carneiro, Yefeng Zheng, Fuyong Xing, and Lin Yang. Review of deep learning methods in mammography, cardiovascular, and microscopy image analysis. In *Deep Learning and Convolutional Neural Networks for Medical Image Computing*, pages 11–32. Springer, 2017.
- [19] Peng Peng, Karim Lekadir, Ali Gooya, Ling Shao, Steffen E Petersen, and Alejandro F Frangi. 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, 2016.
- [20] Dongwoo Kang, Jonghye Woo, CC Jay Kuo, Piotr J Slomka, Damini Dey, and Guido Germano. Heart chambers and whole heart segmentation techniques. *Journal of Electronic Imaging*, 21(1):010901, 2012.
- [21] Caroline Petitjean and Jean-Nicolas Dacher. A review of segmentation methods in short axis cardiac MR images. *Medical image analysis*, 15(2):169–184, 2011.
- [22] EL Ridley. Big data in radiology will drive personalized patient care. *Online: <http://www.auntminnie.com/>(accessed 2015 01 08)*, 2014.
- [23] Xènia Albà, Karim Lekadir, Corné Hoogendoorn, Marco Pereanez, Andrew J Swift, Jim M Wild, and Alejandro F Frangi. Reusability of statistical shape models for the segmentation of severely abnormal hearts. In *International Workshop*

BIBLIOGRAPHY

- on Statistical Atlases and Computational Models of the Heart*, pages 257–264. Springer, 2014.
- [24] Ramachandran S Vasani. Biomarkers of cardiovascular disease. *Circulation*, 113(19):2335–2362, 2006.
- [25] Robert E Gerszten and Thomas J Wang. The search for new cardiovascular biomarkers. *Nature*, 451(7181):949–952, 2008.
- [26] Leroy Hood and Stephen H Friend. Predictive, personalized, preventive, participatory (p4) cancer medicine. *Nature reviews Clinical oncology*, 8(3):184–187, 2011.
- [27] Albert Lardo, Zahi A Fayad, Nicolas Chronos, and Valentin Fuster. *Cardiovascular magnetic resonance: established and emerging applications*. Taylor & Francis, 2004.
- [28] Sangeeta Shah, Emanuel D Chryssos, and Hugh Parker. Magnetic Resonance Imaging: A Wealth of Cardiovascular Information. *The Ochsner Journal*, 9(4):266–277, 2009.
- [29] Thomas J Littlejohns, Jo Holliday, Lorna M Gibson, Steve Garratt, Niels Oesingmann, Fidel Alfaró-Almagro, Jimmy D Bell, Chris Boultonwood, Rory Collins, Megan C Conroy, et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature Communications*, 11(1):1–12, 2020.
- [30] Ruogu Fang, Samira Pouyanfar, Yimin Yang, Shu-Ching Chen, and SS Iyengar. Computational health informatics in the big data age: a survey. *ACM Computing Surveys (CSUR)*, 49(1):12, 2016.
- [31] Pau Medrano-Gracia, Brett R Cowan, Avan Suinesiaputra, and Alistair A Young. Challenges of cardiac image analysis in large-scale population-based studies. *Current cardiology reports*, 17(3):9, 2015.
- [32] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. 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, 2015.
- [33] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.
- [34] Steffen E Petersen, Paul M Matthews, Jane M Francis, Matthew D Robson, Filip Zemrak, Redha Boubertakh, Alistair A Young, Sarah Hudson, Peter Weale, Steve Garratt, et al. UK Biobank’s cardiovascular magnetic resonance protocol. *Journal of cardiovascular magnetic resonance*, 18(1):8, 2015.
- [35] Gregory A Roth, Catherine Johnson, Amanuel Abajobir, Foad Abd-Allah, Semaw Ferede Abera, Gebre Abyu, Muktar Ahmed, Baran Aksut, Tahiya Alam, Khurshid Alam, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *Journal of the American College of Cardiology*, 70(1):1–25, 2017.
- [36] Steffen E Petersen, Nay Aung, Mihir M Sanghvi, Filip Zemrak, Kenneth Fung, Jose Miguel Paiva, Jane M Francis, Mohammed Y Khanji, Elena Lukaschuk, Aaron M Lee, et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in caucasians from the UK Biobank population cohort. *Journal of Cardiovascular Magnetic Resonance*, 19(1):18, 2017.
- [37] Wenjia Bai, Matthew Sinclair, Giacomo Tarroni, Ozan Oktay, Martin Rajchl, Ghislain Vaillant, Aaron M Lee, Nay Aung, Elena Lukaschuk, Mihir M Sanghvi, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1):65, 2018.
- [38] Milton Hoz de Vila, Rahman Attar, Marco Pereanez, and Alejandro F Frangi. MULTI-X, a state-of-the-art cloud-based ecosystem for biomedical research. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1726–1733. IEEE, 2018.
- [39] Krzysztof Gorgolewski, Christopher D Burns, Cindee Madison, Dav Clark, Yaroslav O Halchenko, Michael L Waskom, and Satrajit S Ghosh. Nipype: a

BIBLIOGRAPHY

- flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics*, 5:13, 2011.
- [40] Vincenzo Klinké, Stefano Muzzarelli, Nathalie Lauriers, Didier Locca, Gabriella Vincenti, Pierre Monney, Christian Lu, Detlev Nothnagel, Guenter Pilz, Massimo Lombardi, et al. Quality assessment of cardiovascular magnetic resonance in the setting of the european CMR registry: description and validation of standardized criteria. *Journal of Cardiovascular Magnetic Resonance*, 15(1):55, 2013.
- [41] Le Zhang, Ali Gooya, Bo Dong, Rui Hua, Steffen E Petersen, Pau Medrano-Gracia, and Alejandro F Frangi. Automated quality assessment of cardiac MR images using convolutional neural networks. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 138–145. Springer, 2016.
- [42] Vanya V Valindria, Ioannis Lavdas, Wenjia Bai, Konstantinos Kamnitsas, Eric O Aboagye, Andrea G Rockall, Daniel Rueckert, and Ben Glocker. Reverse Classification Accuracy: Predicting Segmentation Performance in the Absence of Ground Truth. *IEEE Transactions on Medical Imaging*, 2017.
- [43] Xènia Albà, Karim Lekadir, Marco Pereañez, Pau Medrano-Gracia, Alistair A Young, and Alejandro F Frangi. Automatic initialization and quality control of large-scale cardiac MRI segmentations. *Medical image analysis*, 43:129–141, 2018.
- [44] Hans C Van Assen, Mikhail G Danilouchkine, Alejandro F Frangi, Sebastián Ordás, Jos JM Westenberg, Johan HC Reiber, and Boudewijn PF Lelieveldt. SPASM: a 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical Image Analysis*, 10(2):286–303, 2006.
- [45] Gary V Heller, Manuel D Cerqueira, Neil J Weissman, Vasken Dilsizian, Alice K Jacobs, Sanjiv Kaul, Warren K Laskey, Dudley J Pennell, John A Rumberger, Thomas Ryan, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Journal of Nuclear Cardiology*, 9(2):240–245, 2002.

- [46] Alejandro F Frangi, Wiro J Niessen, and Max A Viergever. Three-dimensional modeling for functional analysis of cardiac images, a review. *IEEE transactions on medical imaging*, 20(1):2–5, 2001.
- [47] Catalina Tobon-Gomez, Federico M Sukno, Constantine Butakoff, Marina Huguet, and Alejandro F Frangi. Automatic training and reliability estimation for 3D ASM applied to cardiac MRI segmentation. *Physics in Medicine & Biology*, 57(13):4155, 2012.
- [48] Rahman Attar, Marco Pereañez, Ali Gooya, Xènia Albà, Le Zhang, Stefan K Piechnik, Stefan Neubauer, Steffen E Petersen, and Alejandro F Frangi. High throughput computation of reference ranges of biventricular cardiac function on the UK Biobank population cohort. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 114–121. Springer, 2018.
- [49] Qiao Zheng, Hervé Delingette, Nicolas Duchateau, and Nicholas Ayache. 3D consistent & robust segmentation of cardiac images by deep learning with spatial propagation. *IEEE Transactions on Medical Imaging*, 2018.
- [50] Florian Andre, Stephanie Lehrke, Hugo A Katus, and Henning Steen. Reference values for the left ventricular wall thickness in cardiac MRI in a modified AHA 17-segment model. *Journal of Cardiovascular Magnetic Resonance*, 14(S1):P223, 2012.
- [51] Alejandro Deviggiano, Patricia Carrascosa, Macarena De Zan, Carlos Capuñay, Héctor Deschle, and Gastón A Rodríguez Granillo. Wall Thickness and Patterns of Fibrosis in Hypertrophic Cardiomyopathy Assessed by Cardiac Magnetic Resonance Imaging. *Revista Argentina de Cardiología*, 84(3), 2016.
- [52] Valentina O Puntmann, Yee Guan Yap, William McKenna, and A John Camm. Significance of maximal and regional left ventricular wall thickness in association with arrhythmic events in patients with hypertrophic cardiomyopathy. *Circulation Journal*, 74(3):531–537, 2010.
- [53] Rene E Kanza, Hiroshi Higashino, Teruhito Kido, Akira Kurata, Makoto Saito, Yoshifumi Sugawara, and Teruhito Mochizuki. Quantitative assessment of regional left ventricular wall thickness and thickening using 16 multidetector-row

BIBLIOGRAPHY

- computed tomography: comparison with cine magnetic resonance imaging. *Radiation medicine*, 25(3):119–126, 2007.
- [54] Mithun Prasad, Amit Ramesh, Paul Kavanagh, Balaji K Tamarappoo, Ryo Nakazato, James Gerlach, Victor Cheng, Louise EJ Thomson, Daniel S Berman, Guido Germano, et al. Quantification of 3D regional myocardial wall thickening from gated magnetic resonance images. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 31(2):317–327, 2010.
- [55] Florent Le Ven, Karine Bibeau, Élianne De Larochelière, Helena Tizón-Marcos, Stéphanie Deneault-Bissonnette, Philippe Pibarot, Christian F Deschepper, and Éric Larose. Cardiac morphology and function reference values derived from a large subset of healthy young Caucasian adults by magnetic resonance imaging. *European Heart Journal-Cardiovascular Imaging*, 17(9):981–990, 2015.
- [56] Aigul Baltabaeva, Maciej Marciniak, Bart Bijmens, James Moggridge, Feng J He, Tarek F Antonios, Graham A MacGregor, and George R Sutherland. Regional left ventricular deformation and geometry analysis provides insights in myocardial remodelling in mild to moderate hypertension. *European Journal of Echocardiography*, 9(4):501–508, 2007.
- [57] Ion Codreanu, Tammy J Pegg, Joseph B Selvanayagam, Matthew D Robson, Oliver J Rider, Constantin A Dasanu, Bernd A Jung, David P Taggart, Stephen J Golding, Kieran Clarke, et al. Normal values of regional and global myocardial wall motion in young and elderly individuals using navigator gated tissue phase mapping. *Age*, 36(1):231–241, 2014.
- [58] Rahman Attar, Marco Pereañez, Ali Gooya, Xènia Albà, Le Zhang, Milton Hoz de Vila, Aaron M Lee, Nay Aung, Elena Lukaschuk, Mihir M Sanghvi, et al. Quantitative CMR population imaging on 20,000 subjects of the UK Biobank imaging study: LV/RV quantification pipeline and its evaluation. *Medical image analysis*, 56:26–42, 2019.
- [59] Jinming Duan, Ghalib Bello, Jo Schlemper, Wenjia Bai, Timothy JW Dawes, Carlo Biffi, Antonio de Marvao, Georgia Doumou, Declan P O’Regan, and

- Daniel Rueckert. Automatic 3D Bi-ventricular Segmentation of Cardiac Images by a Shape-refined Multi-task Deep Learning Approach. *IEEE transactions on medical imaging*, 2019.
- [60] Nathan Painchaud, Youssef Skandarani, Thierry Judge, Olivier Bernard, Alain Lalande, and Pierre-Marc Jodoin. Cardiac Segmentation with Strong Anatomical Guarantees. *IEEE Transactions on Medical Imaging*, 2020.
- [61] Rahman Attar, Marco Pereanez, Christopher Bowles, Stefan K Piechnik, Stefan Neubauer, Steffen E Petersen, and Alejandro F Frangi. 3D Cardiac Shape Prediction with Deep Neural Networks: Simultaneous Use of Images and Patient Metadata. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 586–594. Springer, 2019.
- [62] Kathleen Gilbert, Wenjia Bai, Charlene Mauger, Pau Medrano-Gracia, Avan Suinesiaputra, Aaron M Lee, Mihir M Sanghvi, Nay Aung, Stefan K Piechnik, et al. Independent Left Ventricular Morphometric Atlases Show Consistent Relationships with Cardiovascular Risk Factors: A UK Biobank Study. *Scientific reports*, 9(1):1130, 2019.
- [63] Andriy Myronenko and Xubo Song. Point Set Registration: Coherent Point Drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2262–2275, 2010.
- [64] Corné Hoogendoorn, Nicolas Duchateau, Damián Sanchez-Quintana, Tristan Whitmarsh, Federico M Sukno, Mathieu De Craene, Karim Lekadir, and Alejandro F Frangi. A high-resolution atlas and statistical model of the human heart from multislice CT. *IEEE transactions on medical imaging*, 32(1):28–44, 2012.
- [65] Andriy Myronenko, Xubo Song, and Miguel A Carreira-Perpinán. Non-rigid point set registration: Coherent point drift. In *Advances in neural information processing systems*, pages 1009–1016, 2007.
- [66] Stéphane Lathuilière, Pablo Mesejo, Xavier Alameda-Pineda, and Radu Horaud. A comprehensive analysis of deep regression. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

BIBLIOGRAPHY

- [67] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [68] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [69] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [70] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [71] Bram Ruijsink, Esther Puyol-Antón, Ilkay Oksuz, Matthew Sinclair, Wenjia Bai, Julia A Schnabel, Reza Razavi, and Andrew P King. Fully automated, quality-controlled cardiac analysis from CMR: validation and large-scale application to characterize cardiac function. *JACC: Cardiovascular Imaging*, 13(3):684–695, 2020.
- [72] Qiao Zheng, Hervé Delingette, Nicolas Duchateau, and Nicholas Ayache. 3-D consistent and robust segmentation of cardiac images by deep learning with spatial propagation. *IEEE transactions on medical imaging*, 37(9):2137–2148, 2018.
- [73] Mahendra Khened, Varghese Alex Kollerathu, and Ganapathy Krishnamurthi. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. *Medical image analysis*, 51:21–45, 2019.
- [74] Qian Tao, Wenjun Yan, Yuanyuan Wang, Elisabeth HM Paiman, Denis P Shamonin, Pankaj Garg, Sven Plein, Lu Huang, Liming Xia, Marek Sramko, et al. Deep learning-based method for fully automatic quantification of left ventricle function from cine MR images: a multivendor, multicenter study. *Radiology*, 290(1):81–88, 2019.

- [75] Steffen E Petersen, Mihir M Sanghvi, Nay Aung, Jackie A Cooper, Jose Miguel Paiva, Filip Zemrak, Kenneth Fung, Elena Lukaschuk, Aaron M Lee, Valentina Carapella, et al. The impact of cardiovascular risk factors on cardiac structure and function: Insights from the UK Biobank imaging enhancement study. *PLoS one*, 12(10):e0185114, 2017.
- [76] Magnus T Jensen, Kenneth Fung, Nay Aung, Mihir M Sanghvi, Sucharitha Chadalavada, Jose M Paiva, Mohammed Y Khanji, Martina C de Knecht, Elena Lukaschuk, Aaron M Lee, et al. Changes in cardiac morphology and function in individuals with diabetes mellitus: the UK Biobank cardiovascular magnetic resonance substudy. *Circulation: Cardiovascular Imaging*, 12(9):e009476, 2019.
- [77] Angela Gallina Bertaso, Daniela Bertol, Bruce Bartholow Duncan, and Murilo Foppa. Epicardial fat: definition, measurements and systematic review of main outcomes. *Arquivos brasileiros de cardiologia*, 101(1):e18–e28, 2013.
- [78] Yingrui Li, Bin Liu, Yu Li, Xiaodong Jing, Songbai Deng, Yulin Yan, and Qiang She. Epicardial fat tissue in patients with diabetes mellitus: a systematic review and meta-analysis. *Cardiovascular Diabetology*, 18(1):1–10, 2019.
- [79] Dong-Hyuk Cho, Hyung Joon Joo, Mi-Na Kim, Do-Sun Lim, Wan Joo Shim, and Seong-Mi Park. Association between epicardial adipose tissue, high-sensitivity C-reactive protein and myocardial dysfunction in middle-aged men with suspected metabolic syndrome. *Cardiovascular diabetology*, 17(1):95, 2018.
- [80] Arnold CT Ng, Mark Strudwick, Rob J van der Geest, Austin CC Ng, Lisa Gillinder, Shi Yi Goo, Gary Cowin, Victoria Delgado, William YS Wang, and Jeroen J Bax. Impact of epicardial adipose tissue, left ventricular myocardial fat content, and interstitial fibrosis on myocardial contractile function. *Circulation: Cardiovascular Imaging*, 11(8):e007372, 2018.
- [81] Rahil Shahzad, Daniel Bos, Coert Metz, Alexia Rossi, Hortense Kirişli, Aad van der Lugt, Stefan Klein, Jacqueline Witteman, Pim de Feyter, Wiro Niessen, et al. Automatic quantification of epicardial fat volume on non-enhanced cardiac CT scans using a multi-atlas segmentation approach. *Medical physics*, 40(9):091910, 2013.

BIBLIOGRAPHY

- [82] Mohamed Marwan and Stephan Achenbach. Quantification of epicardial fat by computed tomography: why, when and how? *Journal of cardiovascular computed tomography*, 7(1):3–10, 2013.
- [83] Daniel Davidovich, Amalia Gastaldelli, and Rosa Sicari. Imaging cardiac fat. *European Heart Journal–Cardiovascular Imaging*, 14(7):625–630, 2013.
- [84] Tonye Teme, Bassel Sayegh, Mushabbar Syed, David Wilber, Lara Bakhos, and Mark Rabbat. Quantification of epicardial fat volume using cardiovascular magnetic resonance imaging. *Journal of Cardiovascular Magnetic Resonance*, 16(1):1–2, 2014.
- [85] Rami Homsy, Michael Meier-Schroers, Jürgen Gieseke, Darius Dabir, Julian A Luetkens, Daniel L Kuetting, Claas P Naehle, Christian Marx, Hans H Schild, Daniel K Thomas, et al. 3D-Dixon MRI based volumetry of peri-and epicardial fat. *The international journal of cardiovascular imaging*, 32(2):291–299, 2016.
- [86] E Louise Thomas, JA Fitzpatrick, SJ Malik, Simon D Taylor-Robinson, and Jimmy D Bell. Whole body fat: content and distribution. *Progress in nuclear magnetic resonance spectroscopy*, 73:56–80, 2013.
- [87] Tina E Brinkley, F-C Hsu, J Jeffrey Carr, W Gregory Hundley, David A Bluemke, Joseph F Polak, and Jingzhong Ding. Pericardial fat is associated with carotid stiffness in the Multi-Ethnic Study of Atherosclerosis. *Nutrition, Metabolism and Cardiovascular Diseases*, 21(5):332–338, 2011.
- [88] Paolo Raggi. Epicardial adipose tissue as a marker of coronary artery disease risk, 2013.
- [89] Cho-Kai Wu, Hao-Yuan Tsai, Mao-Yuan M Su, Yi-Fan Wu, Juey-Jen Hwang, Jiunn-Lee Lin, Lian-Yu Lin, and Jien-Jiun Chen. Evolutional change in epicardial fat and its correlation with myocardial diffuse fibrosis in heart failure patients. *Journal of clinical lipidology*, 11(6):1421–1431, 2017.
- [90] Gijs van Woerden, Thomas M Gorter, B Daan Westenbrink, Tineke P Willems, Dirk J van Veldhuisen, and Michiel Rienstra. Epicardial fat in heart failure

- patients with mid-range and preserved ejection fraction. *European journal of heart failure*, 20(11):1559–1566, 2018.
- [91] Rajiv Mahajan, Adam Nelson, Rajeev K Pathak, Melissa E Middeldorp, Christopher X Wong, Darragh J Twomey, Angelo Carbone, Karen Teo, Thomas Agbaedeng, Dominik Linz, et al. Electroanatomical remodeling of the atria in obesity: impact of adjacent epicardial fat. *JACC: Clinical Electrophysiology*, 4(12):1529–1540, 2018.
- [92] Markus Henningsson, Martin Brundin, Tobias Scheffel, Carl Edin, Federica Viola, and Carl-Johan Carlhäll. Quantification of Epicardial Fat using 3D Cine Dixon MRI. *BMC Medical Imaging*, 2020.
- [93] Mounes Aliyari Ghasabeh, Anneline SJM Te Riele, Cynthia A James, HS Vincent Chen, Crystal Tichnell, Brittney Murray, John Eng, Brian G Kral, Harikrishna Tandri, Hugh Calkins, et al. Epicardial fat distribution assessed with cardiac CT in arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Radiology*, 289(3):641–648, 2018.
- [94] Taner Seker, Caner Turkoglu, Hazar Harbaloglu, and Mustafa Gur. The impact of diabetes on the association between epicardial fat thickness and extent and complexity of coronary artery disease in patients with non-ST elevation myocardial infarction. *Kardiologia polska*, 75(11):1177, 2017.
- [95] Xiong Chen, Wenjun Wu, Luyin Wang, Yujuan Shi, Feixia Shen, Xuemei Gu, and Zhijun Jia. Association between 25-hydroxyvitamin D and epicardial adipose tissue in Chinese non-obese patients with type 2 diabetes. *Medical science monitor: international medical journal of experimental and clinical research*, 23:4304, 2017.
- [96] Clothilde Philouze, Philippe Obert, Stephane Nottin, Asma Benamor, Olivier Barthez, and Falah Aboukhouidir. Dobutamine stress echocardiography un-masks early left ventricular dysfunction in asymptomatic patients with uncomplicated type 2 diabetes: a comprehensive two-dimensional speckle-tracking imaging study. *Journal of the American Society of Echocardiography*, 31(5):587–597, 2018.

BIBLIOGRAPHY

- [97] Zihang Wang, Yuhong Zhang, Weiwei Liu, and Benli Su. Evaluation of epicardial adipose tissue in patients of type 2 diabetes mellitus by echocardiography and its correlation with intimal medial thickness of carotid artery. *Experimental and Clinical Endocrinology & Diabetes*, 125(09):598–602, 2017.
- [98] Carmelo Militello, Leonardo Rundo, Patrizia Toia, Vincenzo Conti, Giorgio Russo, Clarissa Filorizzo, Erica Maffei, Filippo Cademartiri, Ludovico La Grutta, Massimo Midiri, et al. A semi-automatic approach for epicardial adipose tissue segmentation and quantification on cardiac CT scans. *Computers in Biology and Medicine*, 114:103424, 2019.
- [99] Mohamed Marwan, Susanna Koenig, Kirsten Schreiber, Fabian Ammon, Markus Goeller, Daniel Bittner, Stephan Achenbach, and Michaela M Hell. Quantification of epicardial adipose tissue by cardiac CT: Influence of acquisition parameters and contrast enhancement. *European journal of radiology*, 121:108732, 2019.
- [100] Leonardo C Mangili, Otavio C Mangili, Márcio S Bittencourt, Márcio H Miname, Paulo H Harada, Leonardo M Lima, Carlos E Rochitte, and Raul D Santos. Epicardial fat is associated with severity of subclinical coronary atherosclerosis in familial hypercholesterolemia. *Atherosclerosis*, 254:73–77, 2016.
- [101] Xiuxiu He, Bang Jun Guo, Yang Lei, Tonghe Wang, Tian Liu, Walter J Curran, Long Jiang Zhang, and Xiaofeng Yang. Automatic epicardial fat segmentation in cardiac CT imaging using 3D deep attention U-Net. In *Medical Imaging 2020: Image Processing*, volume 11313, page 113132D. International Society for Optics and Photonics, 2020.
- [102] Qi Zhang, Jianhang Zhou, Bob Zhang, Weijia Jia, and Enhua Wu. Automatic Epicardial Fat Segmentation and Quantification of CT Scans Using Dual U-Nets With a Morphological Processing Layer. *IEEE Access*, 8:128032–128041, 2020.
- [103] Érick Oliveira Rodrigues, FFC Morais, NAOS Morais, LS Conci, LV Neto, and Aura Conci. A novel approach for the automated segmentation and volume quantification of cardiac fats on computed tomography. *Computer methods and programs in biomedicine*, 123:109–128, 2016.

- [104] Miranda R Fulton, Amy H Givan, Maria Fernandez-del Valle, and Jon D Klingensmith. Segmentation of epicardial adipose tissue in cardiac MRI using deep learning. In *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 11317, page 113170Q. International Society for Optics and Photonics, 2020.
- [105] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006.
- [106] Nay Aung, Jose D Vargas, Chaojie Yang, Claudia P Cabrera, Helen R Warren, Kenneth Fung, Evan Tzanis, Michael R Barnes, Jerome I Rotter, Kent D Taylor, et al. Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. *Circulation*, 140(16):1318–1330, 2019.