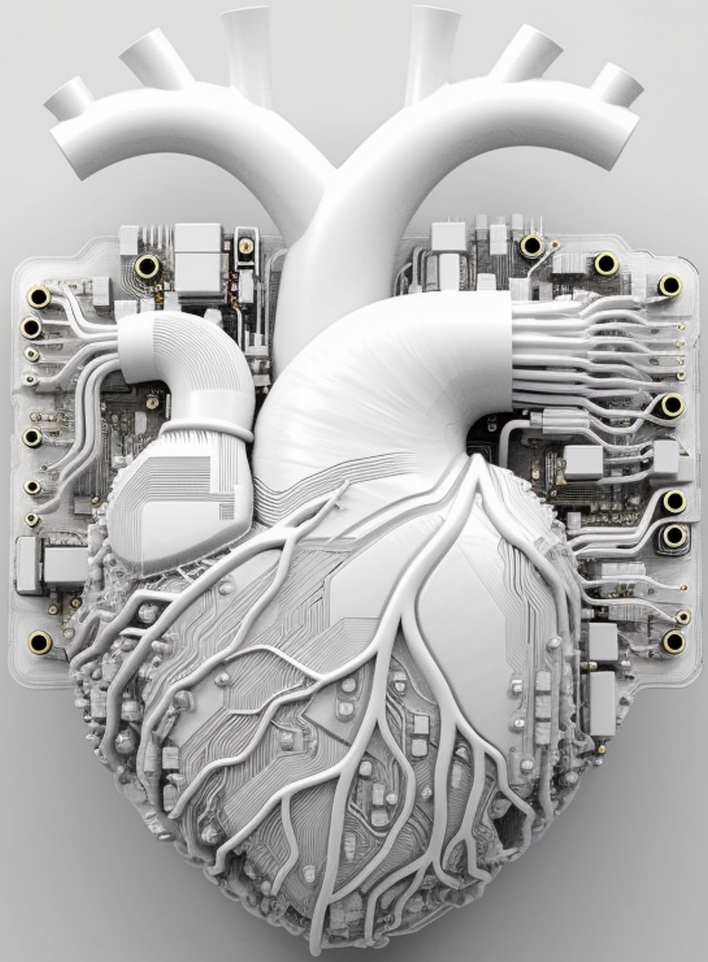
UNIVERSITAT DE BARCELONA

Automatic Cardiac Segmentation of Complex Morphologies, Modalities and Tissues

Carlos Martín Isla

AUTOMATIC CARDIAC SEGMENTATION OF COMPLEX MORPHOLOGIES, MODALITIES AND TISSUES USING DEEP LEARNING



Carlos Martín-Isla

UNIVERSITAT DE BARCELONA

UNIVERSITAT DE BARCELONA

Doctoral Program on Mathematics and Computer Science

AUTOMATIC CARDIAC SEGMENTATION OF COMPLEX MORPHOLOGIES, MODALITIES AND TISSUES USING DEEP LEARNING

Carlos Martín Isla

Supervision

Dr. Karim Lekadir

Dr. Sergio Escalera Guerrero

Tutorship

Dr. Alex Haro Provinciale

Tête-à-tête sombre et limpide
Qu'un cœur devenu son miroir!
Puits de Vérité, clair et noir,
Où tremble une étoile livide.

Charles Baudelaire, *Les fleurs du mal.*

Acknowledgements

This adventure, like an unexpected and fresh whirlwind, would not have been possible without certain whims of fate. I cannot be more grateful to my supervisor Dr. Karim Lekadir. From him I have learned to integrate a new vision, both in the research process and in the vital one. To distill, and capture and adapt objectives. He is able to do it like no one else. I also have to thank him for the experience of participating in the creation of a great laboratory like BCN-AIM from scratch, a certainly enriching experience. Secondly, I want to thank Dr. Sergio Escalera, as my co-supervisor, for his human quality, for helping me and listening whenever I needed it, and for being a brilliant researcher with incredible intuition. To both of them, of course, I am extremely grateful for their technical and methodological guidance throughout my PhD period. And last but not least, I thank you for the opportunity to work both as an assistant professor during these years and to contribute to the creation of various courses and masterclasses both at the University of Barcelona and at the Université Paris-Saclay.

Secondly, I have to especially thank my fellow adventurers Víctor M. Campello and Cristian Morcillo Izquiero for forming a team from the beginning, for long hours in the laboratory and mutual support. Without them none of this would have been possible. To Katharina F. Heil, as one of the most capable, caring and committed people I have ever met in my life. To the rest of the BCN-AIM research group, especially Kaisar Kushibar, Vien Ngoc and Richard Osuala, both on a technical and personal level.

Thirdly, to my relatives. To my parents, Nieves and Juan Carlos for having instilled in me that effort is necessary. To my late grandfather Pedro, for giving me the gift of curiosity.

And finally, to my friends and loved ones, for turning in this whirlwind in the same direction as me.

Abstract

Cardiovascular diseases (CVDs) continue to take a significant toll on global health, highlighting the need for more accurate and efficient diagnostic tools. This thesis, titled "Automatic Cardiac Segmentation of Complex Morphologies, Modalities, and Tissues Using Deep Learning," delves into complex medical imaging and artificial intelligence (AI) technologies necessary to perform advanced and cutting-edge cardiovascular diagnostics.

The groundwork of this work is laid by emphasizing the critical importance of early, precise, and personalized CVD assessment by means of machine learning (ML) and deep learning (DL), in order to evolve from qualitative visual assessments and basic quantitative measures into advanced, quantitative, data-driven insights. The importance of accurate delineation of cardiac structures for a correct assessment of their status and function is crucial to move forward in that direction.

The first chapter delves into the right ventricle segmentation within magnetic resonance imaging (MRI) images, highlighting the challenges posed by complex shapes and ill-defined borders. It introduces the M&Ms-2 challenge, a substantial dataset encompassing diverse pathologies, multiple views, and various scanners. The chapter discusses the success of nnU-Net and underscores the value of multi-view approaches, indicating the need for comprehensive cardiac segmentation algorithms.

In the second chapter, the focus shifts to late gadolinium enhancement MRI (LGE-MRI) segmentation, crucial for quantifying scar tissue in cardiac patients. The proposed solution leverages generative adversarial networks to create synthetic images, enhancing segmentation accuracy in the presence of scar tissue. Results reveal the potential of multi-sequence model training with synthetic images and data augmentation to outperform traditional methods.

The third chapter addresses the segmentation of pathological tissue, specifically scar tissue and edema, within multi-modal cardiac MRI images. The chapter introduces a two-staged approach, involving a stacked BCDU-net for accurate myocardium segmentation and multi-modal pathological region segmentation. Anatomically constrained synthetic data augmentation enriches the model's performances. This thesis represents a pioneering effort to enhance cardiac deep learning-driven segmentation. By tackling the complexities of morphologies, MRI modalities and pathological tissues, this research contributes valuable insights, algorithms, and datasets to such task.

Contents

List of Figures **13**

List of Tables **15**

1 Introduction **17**
1.1 Background 17
1.1.1 Traditional segmentation in cardiac imaging 17
1.1.2 Clinical importance of cardiac imaging segmentation 18
1.1.3 Limitations of traditional cardiac imaging segmentation methods 23
1.2 Deep Learning in Cardiac Imaging Segmentation 24
1.2.1 Overview 24
1.2.2 Towards big data: Deep Learning versus Traditional Methods . 25
1.2.3 Recent developments in Deep Learning for cardiac magnetic resonance imaging segmentation 28
1.3 Objectives and contributions 29
1.3.1 Research Questions 29
1.3.2 Contributions of this Thesis 30
1.4 Other Work by the Author 31

2 Segmentation of complex morphologies: Abnormal Right Ventricle segmentation in MRI **35**
2.1 Abstract 36
2.2 Introduction 36
2.3 Challenge framework 40
2.3.1 Data preparation 40
2.3.2 Model training and validation 43
2.3.3 Model evaluation 44

2.4 Participating methods . . . 45
2.4.1 Backbone architectures . . . 45
2.4.2 Data augmentation . . . 49
2.5 Results . . . 50
2.5.1 Team Ranking . . . 51
2.5.2 Results per Pathology . . . 52
2.5.3 Results per Cardiac Region . . . 52
2.5.4 Clinical Measurements . . . 54
2.5.5 Qualitative results . . . 56
2.6 Discussion . . . 57
2.6.1 Summary of the challenge results . . . 58
2.6.2 Analysis of Pathologies . . . 58
2.6.3 Single- versus Multi-view Models . . . 58
2.6.4 Impact on clinical indices . . . 59
2.6.5 Further considerations . . . 59
2.6.6 Future work . . . 60
2.7 Conclusions . . . 60

3 Segmentation of complex modalities: Myocardium, Left Ventricle and Rigth Ventricle segmentation in LGE 63
3.1 Abstract . . . 64
3.2 Introduction . . . 64
3.3 Method . . . 66
3.3.1 Dataset . . . 66
3.3.2 Increasing training sample . . . 67
3.3.3 CNN-based LGE segmentation . . . 69
3.4 Results . . . 70
3.5 Conclusions . . . 71
3.6 Acknowledgements . . . 72

4 Segmentation of complex tissues: Scar and edema segmentation in LGE 75
4.1 Abstract . . . 76
4.2 Introduction . . . 76
4.3 Materials and methods . . . 77
4.3.1 Dataset . . . 77
4.3.2 Proposed Method . . . 78

4.3.3 Data augmentation strategy 81
4.3.4 Post-processing 86
4.4 Results 87
4.4.1 Protocol and Metrics of the challenge 87
4.4.2 Ablation study 88
4.4.3 Challenge results 89
4.5 Discussion 90

5 Conclusions **93**
5.1 Summary 93
5.2 Limitations 96
5.3 Future perspectives 97
5.3.1 Validation 97
5.3.2 Efficient data collection 97
5.3.3 Trustworthy AI 98

Bibliography **101**

List of Figures

1.1 Input variables type distribution in reviewed literature. 20
1.2 Distribution of image-based diagnostic application using machine learning per disease, per modality. 23
1.3 Cardiac function assessment using scintiphotography. 26
1.4 U-net architecture applied to cardiac imaging segmentation. 28

2.1 Visual appearance of short-axis and long-axis views of pathological and healthy subjects. 39
2.2 Distribution per pathology and scanner along train, validation, and test sets. 41
2.3 Data collection and pre-processing pipeline. 43
2.4 Weighted average DSC per pathology. 52
2.5 Statistical difference according to the Mann-Whitney U rank test for DSC scores between seen and unseen pathologies. 53
2.6 Average performance of the top 5 ranked methods in SA from basal to the apical regions. 53
2.7 Number of not segmented slices at basal region. 54
2.8 Prediction examples for some of the presented methods. 57

3.1 Example of three rotations of the myocardial wall. 67
3.2 Examples of synthetic LGE-MRI images. 68
3.3 Detailed architecture of the CNN model used for LGE segmentation. . 70
3.4 Three segmentation examples as obtained by using different training combinations. 72

4.1 Overview of the proposed stacked network. 79
4.2 Overview of the proposed SPADE generative model. 82

4.3 Style modifications. 84
4.4 Epicardial contour warping between a pair of subjects. 84
4.5 Morphological operations involved in the scar rotation process. 85
4.6 Morphological operations involved in the scar and edema dilation and erosion process. 85
4.7 Segmentation examples combining different sets of training data, showing the improvement of SPADE synthesis. 88
4.8 Improvement offered by applying post-processing on the outputs of the localization and segmentation networks. 91

List of Tables

1.1 Selected studies using cardiac image-based AI analysis 22
1.2 Selection of cardiac structural and functional analysis software incorporating AI. 24

2.1 Automatic CMR segmentation challenges in figures 36
2.2 Number of studies per pathology in each dataset partition 42
2.3 Average specifications for the images acquired in the different centers. 42
2.4 List and details of the participating teams in the challenge. 45
2.5 Characteristics of participating models. 46
2.6 DSC and HD and inference time for the final submissions. 51
2.7 Clinical metrics for the 15 participating methods 55
2.8 Number of patients above different RVEF error thresholds. 56

3.1 MS-CMRSeg sequences details. 66
3.2 Average and standard deviation for the Dice score of segmentation results over the five labeled LGE volumes. 69
3.3 Average and standard deviation for results over the test set. 71

4.1 2D Dice score of the proposed method. 89
4.2 3D Dice score for the final testing set of 20 subjects. 90

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 Traditional segmentation in cardiac imaging

Cardiac imaging segmentation enables the accurate and quantitative analysis of various cardiac structures, such as the myocardium, ventricles, blood vessels or pathological tissue. By delineating these structures, clinicians and researchers can extract useful information about their shape, size, motion, and function, which can aid in the diagnosis and treatment of various cardiovascular diseases.

A variety of segmentation methods have been widely used in cardiac imaging in the past, and they include techniques such as thresholding, region-growing, edge-based methods, and deformable models. These methods are based on mathematical models and algorithms that aim to partition the image into different regions of interest.

Furthermore, segmentation of cardiac structures is essential for the development and evaluation of various cardiac imaging modalities. By comparing the segmentation results obtained from different imaging modalities, researchers can assess their accuracy and consistency, and optimize their protocols and parameters to improve quality and efficiency.

In the following subsections, we will discuss in more detail the clinical importance of cardiac imaging segmentation, and how advanced segmentation methods, such as deep learning-based approaches, can address the limitations of traditional methods and enable more accurate, efficient, and clinically relevant analysis of cardiac structures.

1.1.2 Clinical importance of cardiac imaging segmentation

Despite significant advances in diagnosis and treatment, cardiovascular disease (CVD) remains the most common cause of morbidity and mortality worldwide, accounting for approximately one third of annual deaths [175, 143]. Early and accurate diagnosis is key to improving CVD outcomes. Cardiovascular imaging has a pivotal role in diagnostic decision making. Current image analysis techniques are mostly reliant on qualitative visual assessment of images and crude quantitative measures of cardiac structure and function. In order to optimise the diagnostic value of cardiac imaging, there is need for more advanced image analysis techniques that allow deeper quantification of imaging phenotypes. In recent years, the development of big data and availability of high computational power have driven exponential advancement of artificial intelligence (AI) technologies in medical imaging. Machine learning (ML) approaches to image-based diagnosis rely on algorithms/models that *learn* from past clinical examples through identification of hidden and complex imaging patterns. Existing work already demonstrates the incremental value of image-based cardiovascular diagnosis with ML for a number of important conditions such as coronary artery disease (CAD) and heart failure (HF). The superior diagnostic performance of AI image analysis has the potential to substantially alleviate the burden of cardiovascular disease through facilitation of faster and more accurate diagnostic decision making.

Cardiac imaging segmentation is essential for the accurate and quantitative analysis of various cardiac structures, which is crucial for the diagnosis, treatment, and management of cardiovascular diseases. Cardiac imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound are widely used in clinical practice to visualize the anatomy and function of the heart. However, the interpretation of these images can be challenging due to the complex and dynamic nature of cardiac structures, and their variability among different patients and diseases.

By enabling, via automatic segmentation, the extraction of precise and quantitative measurements of various cardiac structures, such as the myocardium, ventricles, and blood vessels. These measurements can provide valuable information about the morphology, function, and hemodynamics of the heart and blood vessels, which can aid in the diagnosis and treatment of various cardiovascular diseases.

Image-based cardiac diagnosis with machine learning: diagnostic applications

To analise the impact of automatic segmentation in cardiac diagnostic applications, an organised, pre-defined literature search of two electronic databases (Google Scholar, Scopus) was conducted. We included studies using a well-defined ML technique for cardiac image analysis using echocardiography, cardiac magnetic resonance, cardiac computed tomography, or single photon emissioncomputed tomography (SPECT). Our search strategy comprised a series of title and whole text searches with search terms combined using Boolean operators. Search results were filtered by subject area, limiting to entries from Cardiology, Computer Science and Engineering fields. We review in detail various achievements in the diagnosis of a wide range of cardiac diseases using image-based ML methods.Approximately 95% of these studies rely on some segmentation method to extract cardiac biomarkers for diagnostic and risk assessment tasks. The preferred imaging modality for these types of applications is cardiac magnetic resonance imaging due to diverse factors, namely:

- **Image quality**: Cardiac MRI provides detailed, high-resolution images of the heart. This allows accurate visualization of cardiac structures, as well as abnormalities and diseases that may be present. These high-quality images are essential for accurate analysis and proper interpretation by artificial intelligence algorithms.

- **Precise quantification**: Cardiac MRI offers precise quantification capabilities of various cardiac parameters, such as ventricular volume and ejection fraction, and other relevant markers. Functional parameters such as blood flow, myocardial perfusion or tissue viability can be also assessed via cardiac MRI. These quantitative measurements are valuable for monitoring heart disease and early detection of subtle changes in heart function.

- **Multi-view data**: Cardiac MRI provides images in multiple spatial planes, allowing a comprehensive and detailed evaluation of the heart from different perspectives. Artificial intelligence can take advantage of this multiplanar information to improve precision and decision-making in medicine.

Before a ML model can be built for image-based diagnosis estimation, it is necessary to suitably define the imaging inputs. Imaging inputs may be the raw imaging data (*i.e.* pixel intensities), conventional cardiac indices (and other transformed quantitative image parameters) or radiomics features extracted from the image. Figure 1.1 shows

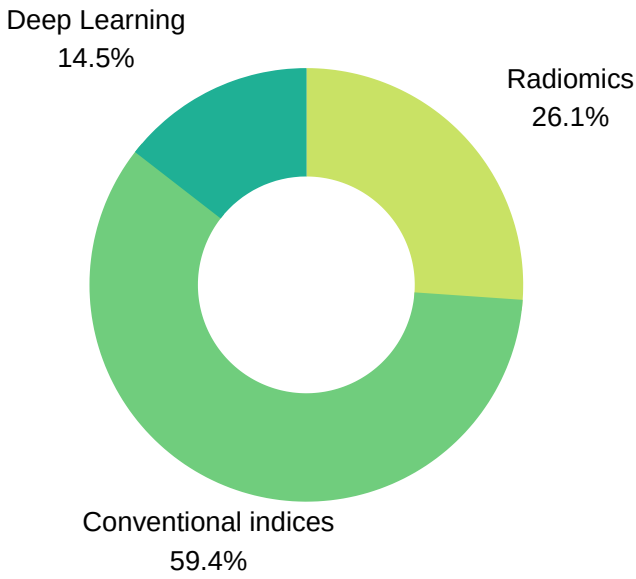


Figure 1.1: Input variables type distribution in reviewed literature. As seen in the pie chart, conventional indices are the predominant features for training ML models, followed by radiomics and deep learning techniques.

that, despite the exponential growth of Deep Learning, conventional indexes are the preferred alternative for AI-powered automatic image-based cardiac diagnostic applications. However, advanced texture biomarkers such as radiomic quantitative features are gaining popularity in recent years due to an improved expressivity and increased assessment capabilities.

Conventional indexes

Conventional imaging indices include measures commonly used in routine clinical image analysis such as ventricular volumes in end diastole/systole and ventricular ejection fractions. Estimation of these clinical indices requires prior contouring of the endocardial and epicardial boundaries of the relevant cardiac chambers as well as necrotic, oedemic and scar tissue among others. Deep learning approaches have been used to develop automated/semi-automated contouring tools for more efficient and reproducible segmentation of cardiac chambers.

Radiomics

Radiomics analysis is the process of converting digital images to minable data. Analysis of the data through application of various statistical and mathematical processes allows quantification of various shape and textural characteristics of the image, referred to as radiomics features. Radiomics analysis quantifies more advanced and complex characteristics of the cardiac structures than is visually perceptible. Introduced in 2012 [92, 90], the radiomics paradigm was, for a long time, mostly exploited in oncology. Recently, a number of works have shown the promise of radiomics combined with ML for image-aided diagnosis of CVD.

Deep Learning

Whole raw images may also be used as the input for the ML model, without any pre-processing or calculation of hand-crafted input imaging features. In this case, the optimal features for predicting the cardiac diagnoses are self-learned automatically by the ML techniques based on the image or volume, as opposed to a priori definition by the AI scientist, that always requires a segmentation of the region of interest.

Review of image-based cardiac diagnostic publications

Table 1.1 shows an extensive set of publications using AI for cardiac diagnostic on a heterogeneous set of cardiac diseases and imaging modalities. One can see that only about 10% of the analyzed publications rely on Deep Learning for classification tasks. The reasoning behind this is based on three premises, namely i) Small inter-class variability: the differences between a pathological and a healthy subject are subtle compared to a regular natural imaging problem i.e. training a Deep Learning model to distinguish between dogs and cats, ii) inevitably, this leads to an increase in terms of number of samples required to successfully train a Deep Learning diagnosis model, in a domain where data is often scarce and difficult to collect, and iii) none of the AI-based cardiac imaging algorithms will be applied in clinical routine if there is not a certain degree of explainability and interpretability.

Therefore, conventional and advanced, quantitative biomarkers solve these problems by integrating human knowledge and compressing the information while endowing the models with a certain degree of interpretability.

Table 1.1: Selected studies using cardiac image-based AI analysis .

Publication	Modality	Biomarker	ML technique	Diagnostic	Sample size	Performance
[12]	MRI	Radiomics	LR	MI	180	ACC = 0.92
[186]	MRI	Conventional	ANN	MI	299	AUC = 0.94
[93]	MRI	Radiomics	SVM	MI	50	AUC = 0.84
[120]	MRI	Conventional	SVM / RF	MI / HCM	45	ACC = 0.94
[144]	MRI	Conventional	DT / CL / SVM	MI	200	ACC = 0.95
[96]	MRI	Conventional	PLS	MI	200	ACC = 0.98
[48]	echo	Qualitative	SVM	MI	242	ACC = 0.97
[121]	echo	Conventional	ANN	MI / AP	91	ACC = 0.95
[171]	echo	Conventional	BN / DT / CL / SVM	MI	42	ACC = 0.91
[157]	echo	Radiomics	DT / ANN / SVM	MI	160	ACC = 0.94
[1]	echo	Radiomics	CL	MI	17	ACC = 0.91
[172]	echo	Radiomics	SVM	MI	800	ACC = 0.99
[161]	echo	Conventional	CL	MI	120	ACC = 0.87
[108]	CT	Radiomics	RF / CL / ANN	MI	87	ACC = 0.78
[109]	CT	Radiomics	DT	MI	30	ACC = 0.97
[179]	CT	Conventional	SVM / RF	MI	170	ACC = 0.85
[148]	SPECT	Conventional	BN	MI / CAD	728	ACC = 0.78
[64]	MRI	Conventional	BN	HCM / DCM / ARV / MYO	83	AUC = 0.79
[11]	MRI	Radiomics	RF / LR	HCM	62	AUC = 0.95
[178]	MRI	Conventional	RF	MI / HCM / DCM / ARV	100	ACC = 0.86
[39]	MRI	Radiomics	SVM	MI / HCM / DCM / ARV	100	ACC = 0.92
[73]	MRI	Conventional	RF	MI / HCM / DCM / ARV	100	ACC = 0.92
[85]	MRI	Conventional	RF	MI / HCM / DCM / ARV	100	ACC = 0.96
[25]	MRI	Deep Learning	VAE	HCM	737	ACC = 1.00
[190]	MRI	Conventional	LR	MI / HCM / DCM / ARV	100	ACC = 0.94
[126]	MRI	Conventional	LR	HHD / HCM	224	ACC = 0.67
[125]	MRI	Radiomics	SVM	HHD / HCM	224	ACC = 0.86
[154]	MRI	Deep Learning	CNN	MI / HCM / DCM / ARV	100	ACC = 0.78
[120]	MRI	Conventional	SVM / RF	MI / HCM	45	ACC = 0.94
[29]	MRI	Conventional	CL	CHD	60	ACC = 0.89
[124]	echo	Conventional	SVM / RF / ANN	HCM / ATHCM	139	ACC = 0.91
[170]	echo	Radiomics	ANN / GA	HCM / DCM	90	ACC = 0.95
[105]	echo	Deep Learning	CNN	HCM / DCM	927	AUC = 0.84
[136]	echo / MRI	Conventional	SVM	DCM	69	ACC = 0.94
[26]	echo	Radiomics	SVM	DCM / ASD	439	ACC = 0.98
[107]	echo	Deep Learning	CNN / GAN	HCM	772	ACC = 0.92
[184]	echo	Deep Learning	CNN	HCM / CA / PH	14,035	AUC = 0.93
[7]	SPECT	Conventional	LB	CAD	1,181	AUC = 0.94
[24]	SPECT	Deep Learning	CNN	CAD	1,160	AUC = 0.81
[122]	SPECT	Conventional	ANN	CAD	1,365	AUC = 0.75
[123]	SPECT	Conventional	ANN	CAD	106	AUC = 0.96
[66]	SPECT	Conventional	ANN	CAD	65	AUC = 0.74
[167]	SPECT	Conventional	DT / GA	CAD	267	ACC = 0.83
[46]	SPECT	Qualitative	SVM	CAD	267	ACC = 0.92
[13]	SPECT	Deep Learning	ANN / CL	CAD	173	AUC = 0.80
[153]	SPECT	Conventional	ANN	CAD	109	AUC = 0.88
[82]	PET	Conventional	N/A	CAD / MACE	1,234	AUC = 0.72
[47]	CT	Conventional	N/A	CAD	352	AUC = 0.84
[68]	CT	Conventional	GBRT	CAD	252	AUC = 0.75
[89]	echo / SCI	Qualitative	ANN	CAD	327	ACC = 0.80
[185]	echo	Radiomics	SVM	CAD	61	AUC = 0.88
[21]	echo	Qualitative	SVM	CAD	228	ACC = 0.99
[87]	CT	Radiomics	N/A	ATH	60	AUC = 0.91
[194]	CT	Deep Learning	CNN	ATH	163	ACC = 0.80
[177]	CT	Radiomics	DT	ATH	164	ACC = 0.86
[176]	CT	Deep Learning	CNN	ATH	250	ACC = 0.72
[76]	CT	Conventional	CL	ATH	615	ACC = 0.74
[77]	CT	Conventional	CL	ATH	249	ACC = 0.83
[54]	echo	Radiomics	ANN	HVD	120	ACC = 0.93
[119]	echo	Radiomics	SVM / CL	HVD	102	ACC = 0.99
[120]	MRI	Conventional	SVM / RF	MI / HCM / HF	45	ACC = 0.77
[10]	MRI	Radiomics	LR	HF	79	AUC = 0.85
[162]	echo	Conventional	CL	HHD / HFePF	100	ACC = 0.81
[152]	echo	Conventional	CL / SVM	HFePF	397	AUC = 0.76

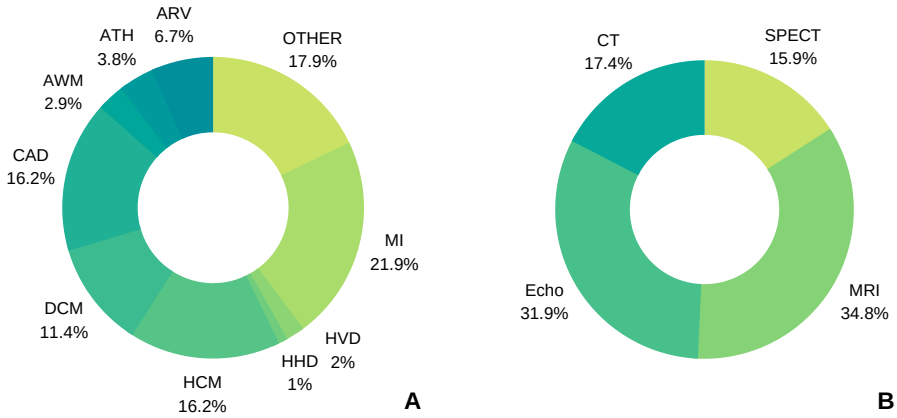


Figure 1.2: Distribution of image-based diagnostic application using machine learning (A) per disease, (B) per modality.

1.1.3 Limitations of traditional cardiac imaging segmentation methods

Both clinical imaging indices and radiomics require prior delineation of the involved cardiac structures, but while computer-aided cardiac imaging segmentation methods have been widely used for many years, they have several limitations that can affect their accuracy and reliability. One major limitation is their sensitivity to image artifacts, such as noise, motion, and partial volume effects. These artifacts can distort the image intensity and geometry, which can lead to errors in the segmentation results. Another limitation is their dependence on manual or semi-automatic procedures, which can be time-consuming, subjective, and thus prone to inter- and intra-observer variability. Manual segmentation methods also require extensive training and expertise, which can limit their widespread adoption and repeatability. Moreover, traditional segmentation methods often rely on hand-crafted and heuristic rules, which may not fully capture the complex and dynamic nature of cardiac imaging data. This can result in segmentation errors, especially in the presence of structural abnormalities or disease. Finally, traditional methods may not be able to handle large datasets or complex imaging protocols, which are becoming increasingly common in clinical practice and research. This can limit their scalability and generalizability, and hinder their integration with other data sources and analysis tools. These limitations highlight the need for alternative and more advanced methods for cardiac imaging segmentation,

Table 1.2: Selection of cardiac structural and functional analysis software incorporating AI.

Name	Producer	Modality
CMRtools	Cardiovascular Imaging Solutions	MRI
suiteHEART	NeoSoft	MRI
CVI42	Circle Cardiovascular Imaging	MRI / CT
Medis Suite	Medis	MRI / CT
iNtuition	Terarecon	MRI / CT
Segment	Medviso	MRI / CT / SPECT
syngo.via	Siemens	MRI / CT / SPECT
IntelliSpace Portal	Philips	MRI / CT / echo
VevoLAB	Visualsonics	echo
QLAB	Philips	echo
TOMTEC	Philips	echo

such as deep learning-based approaches. By leveraging the power of deep neural networks, these methods can overcome many of the limitations of traditional methods, and enable more accurate, efficient, and robust segmentation of cardiac structure: many automatic or semi-automatic tools have been developed (see Table 1.2 for examples of existing tools).

1.2 Deep Learning in Cardiac Imaging Segmentation

1.2.1 Overview

One can think that Deep Learning and, in particular, Neural Networks capable of learning from unstructured data as images, is a recent concept. But nothing could be more untrue. In the 1960s we find what can be considered the embryo of a modern Convolutional Neural Network. The Magnetic Ink Character Recognition system used handmade filters to identify numeric characters printed with magnetic ink on bank checks. The filters in this system were designed to identify specific characteristics of the numbers, such as the presence or absence of certain lines and their orientation, and were based on the mathematical concept of convolution: the magnetic ink used to print the numerical codes was convolved with a system of magnets in a similar way that an image is convolved with a kernel nowadays.

Two decades later Kunihiko Fukushima created the first vision Neural Network

with hierarchical layers, the Neocognitron [57]. While the principles and concepts introduced by the Neocognitron have laid the foundations for the advancement of artificial intelligence in the field of visual perception, it is not until 1989 when Yan LeCunn et al. [94] proposed backpropagation as an efficient solution for the task of training vision Neural Networks in a supervised setup. The same year the term coined by Richard E. Bellman in the domain of dynamic programming, *the curse of dimensionality*, was also considered by Yan LeCunn in [95]: the number of free parameters on a Neural Network has to be carefully controlled to improve the generalization capablility of vision Neural Networks, where the input dimensionality tends to be high. This is particular relevant in the biomedical imaging field, since the number of available examples is normally low when compared to another domain such as natural images. Decades later, with the exponential growth of computational power and the advent of well-optimized modern Convolutional Neural Networks, Generative Models show how data synthesis can play a crucial role in such domains where data scarcity continues being the main impediment for multiple applications of Deep Learning.

1.2.2 Towards big data: Deep Learning versus Traditional Methods

These early steps, primarily applied to the classification task that had formed the core of supervised learning for decades, suffered from a lack of expansion proportional to the lack of expressivity related to the type of tasks that such systems were capable to solve for a long time. To be honest, an artificial system that can learn the differences between cats and dogs when trained with thousands of examples is quite impressive, but the clinical community would never accept software that simply classifies a cardiac image as pathological or healthy, with nothing more to add that a simple probability or *logit*. And even if the system were to be shown to have superior efficacy than the human, there would be obscurity to resolve when such a system failed from a technical, legal or moral point of view, among others.

Despite this, thanks to the efforts of the scientific community that has endeavored to collect large amounts of data and release it to promote the advancement of research in the field of artificial intelligence, and decades after the first vision Neural Network trained by backpropagation was presented, when it was thought that the field of artificial intelligence had reached a plateau, one of the most interesting phenomena in computer science of all time occurred and, we could say, that in modern science in general. It happened in the year 2012, promoted by the ImageNet [52] competition. At

this time, AlexNet [88] improved the best existent classification method on ImageNet by a robust 10%. On the following years, Convolutional Neural Networks such as GoogLeNet (2014) [160], closed the gap between the human performance and such models in classification tasks.

Suddenly an explosion of new architectures expanded the capabilities of Deep Learning models simultaneously: these architectures could easily be adapted to increasing complexity tasks such as for example, detection, segmentation, or generation, being the final two the main topics covered along this work.

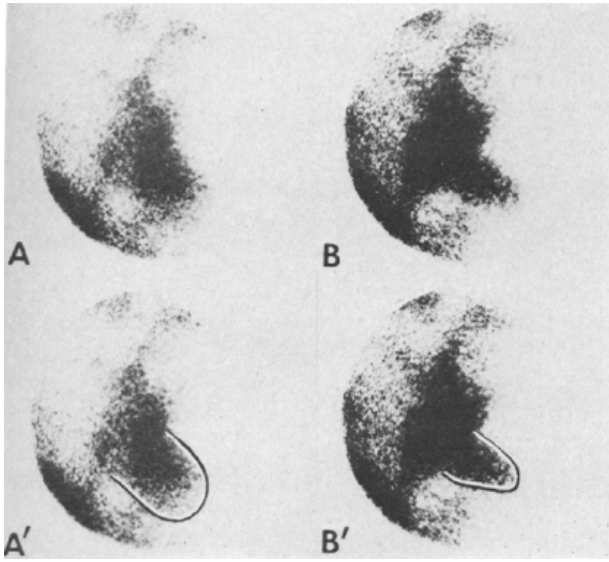


Figure 1.3: Cardiac function assessment using scintiphotography.[156]

On the other hand, the field of modern medicine has much deeper roots in time. In 1628, *Excercitatio anatomica de motu cordis et sanguinis in animalibus* [69], William Harvey examined the structure of the heart through dissections. These studies allowed him to understand the function of the atria and ventricles, as well as the one-way flow of blood through the heart valves. Centuries later, and more aligned with the chronology of the supervised learning methods expansion over the previous decades, the field of medical imaging has experienced growth of a similar magnitude, assisted, in the same sense, by the growth of digital computation technologies. Long gone is the first method that made it possible to explore the human body in a non-invasive way. Wilhelm Conrad Roentgen discovered X-rays in 1895 and, in 1918, a method to estimate the

size of the heart using X-rays was described in [15]. Since then, multiple medical imaging techniques with different qualities have been developed, such as ultrasound, computerized tomography or magnetic resonance. In the field of cardiology, it was not until 1958 when biplane cinefluorography was used to determine ventricular volumes [40] and, later in 1971 when the use of the ECG-triggered scintiphotography [156] allowed to compute a basic cardiac clinical index such as the Left Ventricular Ejection Fraction while invasively validating the findings in a human cohort. In both cases the ventricular silhouette was manually delineated and end-diastolic and end-systolic volumes were calculated using the equation for determining the volume of a prolate spheroid, as shown in Figure 1.3. These are the very first times that functional cardiac assessment was achieved solely by imaging techniques and one can see how from the very beginning delineation is a necessary component for the purpose. The task can be easily covered by Deep Learning segmentation techniques, thus allowing this way to benefit from their qualities, such as robustness, immediacy, or invariability of these.

However, before developing AI-based methods, multiple classical image-processing techniques were proposed in order to compute cardiac contours at a pixel level, leaving aside the mathematical modeling of the volume of this. In 1972, Chow et al. [45] proposed a primitive technique based on dynamic thresholding of cineangiographic images using Gaussian Mixture Models. While it was an attempt to take profit from the growing computation power, the results were not sufficient to incorporate such techniques in clinical practice. Over the following decades, segmentation methods developed in the field of computer vision were rapidly adopted into bioinformatics, and by extension, into cardiac image analysis. Techniques such as *region growing*, *clustering* or *active contours*, in conjunction with more advanced imaging acquisition protocols, started producing more refined results until the arrival of Deep Learning. It is important to note that classical techniques still have their place in certain scenarios and may be more suitable in situations with limited resource requirements, small data sets, or when human interpretation and expert knowledge are critical: Deep Learning can require large training data sets and significant computational and human resources.

In the same way that in the field of natural imaging there was a great effort on the part of the community to prepare sets of images that would pave the way for research, the field of medical imaging is not exempt from such efforts. Regarding cardiac imaging, datasets such as RVSC [132], LVSC [158], and Sunnybrook [138] were curated, manually annotated and released between the years of 2008 and 2012, just before the Convolutional Neural Network renaissance, thus allowing a fair comparison between a diverse set of classical imaging techniques proposed by the community.

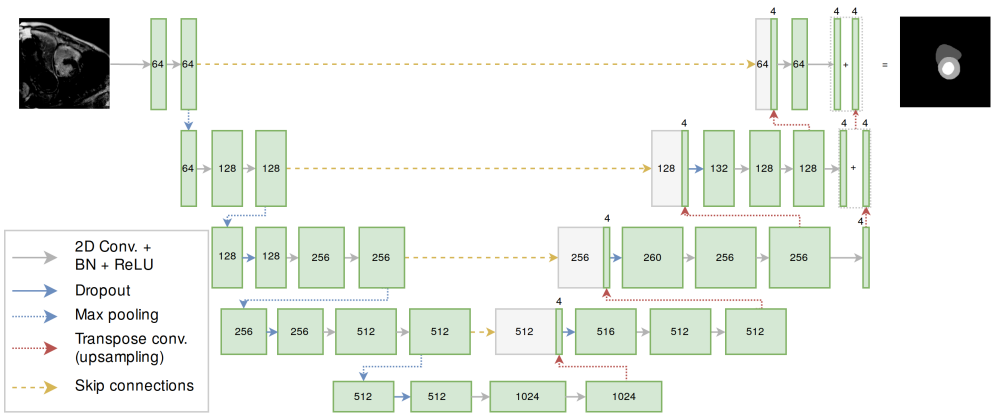


Figure 1.4: U-net architecture applied to cardiac imaging segmentation.

1.2.3 Recent developments in Deep Learning for cardiac magnetic resonance imaging segmentation

The early application of Deep Learning in cardiac magnetic resonance segmentation showed promising results. Tran et al.[168] applied a Fully Convolutional Network (FCN) architecture and demonstrated improved performance compared to previous methods in RVSC, LVSC, and Sunnybrook datasets. The architecture was constructed on top of a regular Convolutional Neural Network, by simply replacing the last layer of a classification Convolutional Neural Network, the system was capable of generating cardiac segmentation maps with unprecedented quality, very close to human capability.

Around the same time, the U-Net architecture [145], which incorporated a symmetric decoding path, gained popularity in the biomedical imaging segmentation community. A scheme of such architecture is presented in Figure 1.4.

In recent CMR segmentation challenges such as the Automated Cardiac Diagnosis Challenge (ACDC) [23] and the M&Ms challenge [32], the U-Net architecture emerged as the dominant choice. Baumgartner et al. [16] demonstrated that U-Net outperformed FCN in all proposed segmentation tasks in the ACDC. Additionally, Isensee et al. [73] showcased the early stages of the nnU-Net framework, which automatically optimized preprocessing, network architecture, training, inference, and post-processing without manual intervention, in the ACDC challenge. The top three participants in the MMs challenge [32] also utilized nnU-Net for their segmentation approaches.

There are currently new architectures based on the Vision Transformer, a recent AI technique that uses attention mechanisms to process images, breaking with the

convolutional paradigm that has dominated the field of computer vision in recent decades. These techniques have logically also been applied to cardiac segmentation, in some cases refining the existent results [38]. In parallel, the development of generative models has made it possible to generate highly realistic synthetic images, which is particularly useful in the field of medical imaging, where data can be scarce. [112, 2]

1.3 Objectives and contributions

Bearing in mind that throughout the development of Deep Learning-based cardiac image segmentation techniques, the lag with the natural image field is mainly related to data scarcity, the work in this thesis is aimed at exploring and alleviating this effect by attacking it from the two main related flanks: data collection and image synthesis that will be defined in detail in the following subsections.

1.3.1 Research Questions

RQ1 (Chapter 2) Can the generation of a diverse set of cardiac data can improve the right ventricular segmentation task in cardiac magnetic resonance?

Most of the open-source data sets for cardiac segmentation are i) based on multi-structure cardiac segmentation and ii) centered on pathological groups related to Left Ventricle dysfunction.

Objectives:

- Collect a diverse dataset of cardiac magnetic resonance imaging data that includes a wide range of patient demographics, image acquisition parameters, and cardiac pathologies.
- Design and implement a competition.
- Quantitatively compare the performance of the participating methods on the generated dataset.
- Investigate the potential benefits of the diverse dataset and augmented data in reducing bias and improving the segmentation performance across different patient populations and imaging protocols.

RQ2 (Chapter 3) How to develop DL models to improve the segmentation of complex cardiac modalities in the presence of annotation scarcity.

State-of-the-art cardiac imaging segmentation is often presented for conventional Steady State Free Precision Cine sequences, where the cardiac tissue tends to be homogeneous. Segmentation performance is degraded in the presence of a contrast agent.

Objectives:

- Investigate the limitations of current segmentation techniques when applied to complex cardiac modalities such as Late Gadolinium Enhancement.
- Develop a methodology for generating realistic synthetic data representative of Late Gadolinium Enhancement from Steady State Free Precision acquisitions, where annotation is a common procedure.
- Develop a LGE-specific augmentation technique.
- Assess the effectiveness of such techniques for training segmentation algorithms on Late Gadolinium Enhancement.

RQ3 (Chapter 4) How to develop DL models to improve the segmentation of multi-modal complex cardiac pathological tissues.

Segmentation of pathological tissue such as scar tissue and edema is a complex problem due to data scarcity and inter-operator variability in terms of annotation. Tissue viability assessment requires multiple imaging acquisitions, thus highlighting the need for multi-modal algorithms.

Objectives:

- Investigate the limitations of current segmentation techniques when applied to multi-sequence segmentation.
- Develop a methodology for generating conditional and multi-modal realistic synthetic data representative of complex cardiac tissues such as scar and edema.
- Assess the effectiveness of using multi-modal DL algorithms in conjunction with synthetic data on complex cardiac pathological tissue segmentation tasks.

1.3.2 Contributions of this Thesis

[114] Carlos Martín-Isla, Víctor M Campello, Cristian Izquierdo, Kaisar Kushibar, Carla Sendra-Balcells, Polyxeni Gkontra, Alireza Sojoudi, Mitchell J Fulton,

Tewodros Weldebirhan Arega, Kumaradevan Punithakumar, et al. Deep learning segmentation of the right ventricle in cardiac mri: The m&ms challenge. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[35] Víctor M Campello, Carlos Martín-Isla, Cristian Izquierdo, Steffen E Petersen, Miguel A González Ballester, and Karim Lekadir. Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhancement cardiac mri. In *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Revised Selected Papers 10*, pages 290–299. Springer, 2020.

[113] Carlos Martín-Isla, Maryam Asadi-Aghbolaghi, Polyxeni Gkontra, Victor M Campello, Sergio Escalera, and Karim Lekadir. Stacked bcdu-net with semantic cmr synthesis: Application to myocardial pathology segmentation challenge. In *Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1*, pages 1–16. Springer, 2020.

1.4 Other Work by the Author

[115] Carlos Martin-Isla, Victor M Campello, Cristian Izquierdo, Zahra Raisi-Estabragh, Bettina Baeßler, Steffen E Petersen, and Karim Lekadir. Image-based cardiac diagnosis with machine learning: a review. *Frontiers in cardiovascular medicine*, 7:1, 2020.

[140] Zahra Raisi-Estabragh, Carlos Martin-Isla, Louise Nissen, Liliana Szabo, Sergio Escalera, Simon Winther, Morten Bøttcher, Karim Lekadir, and Steffen Erhard Petersen. Radiomics analysis enhances the diagnostic performance of cmr stress perfusion: A proof-of-concept study using the dan-nicad dataset short title: Radiomics analysis for cmr perfusion. *Frontiers in Cardiovascular Medicine*, 10:1141026, 2023.

[53] Matthias Eisenmann, Annika Reinke, Vivienn Weru, Minu D Tizabi, Fabian Isensee, Tim J Adler, Sharib Ali, Vincent Andrearczyk, Marc Aubreville, Ujjwal Baid, et al. Why is the winner the best? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19955–19966, 2023.

[134] Esmeralda Ruiz Pujadas, Zahra Raisi-Estabragh, Liliana Szabo, Cristian Izquierdo Morcillo, Víctor M Campello, Carlos Martin-Isla, Hajnalka Vago, Bela Merkely, Nicholas C Harvey, Steffen E Petersen, et al. Atrial fibrillation prediction by combining ecg markers and cmr radiomics. *Scientific Reports*, 12(1):18876, 2022.

[133] Esmeralda Ruiz Pujadas, Zahra Raisi-Estabragh, Liliana Szabo, Celeste McCracken, Cristian Izquierdo Morcillo, Víctor M Campello, Carlos Martín-Isla, Angelica M Atehortua, Hajnalka Vago, Bela Merkely, et al. Prediction of incident cardiovascular events using machine learning and cmr radiomics. *European Radiology*, pages 1–13, 2022.

[100] Lei Li, Fuping Wu, Sihan Wang, Xinzhe Luo, Carlos Martín-Isla, Shuwei Zhai, Jianpeng Zhang, Yanfei Liu, Zhen Zhang, Markus J Ankenbrand, et al. Myops: A benchmark of myocardial pathology segmentation combining three-sequence cardiac magnetic resonance images. *Medical Image Analysis*, 87:102808, 2023.

[33] Victor M Campello, Polyxeni Gkontra, Cristian Izquierdo, Carlos Martin-Isla, Alireza Sojoudi, Peter M Full, Klaus Maier-Hein, Yao Zhang, Zhiqiang He, Jun Ma, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. *IEEE Transactions on Medical Imaging*, 40(12):3543–3554, 2021.

[81] Smriti Joshi, Richard Osuala, Carlos Martín-Isla, Victor M Campello, Carla Sendra-Balcells, Karim Lekadir, and Sergio Escalera. nn-unet training on cyclegan-translated images for cross-modal domain adaptation in biomedical imaging. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part II*, pages 540–551. Springer, 2022.

[37] Victor Manuel Campello, Tian Xia, Xiao Liu, Pedro Sanchez, Carlos Martín-Isla, Steffen Erhard Petersen, Santi Seguí, Sotirios Tsaftaris, and Karim Lekadir. Cardiac aging synthesis from cross-sectional data with conditional generative adversarial networks. *Frontiers in Cardiovascular Medicine*, page 2693, 2022.

[150] Carla Sendra-Balcells, Víctor M Campello, Carlos Martín-Isla, David Vilades Medel, Martín Luís Descalzo, Andrea Guala, José F Rodríguez Palomares, and Karim Lekadir. Multi-center, multi-vendor automated segmentation of left ventricular anatomy in contrast-enhanced mri. *arXiv preprint arXiv:2110.07360*, 2021.

[151] Carla Sendra-Balcells, Víctor M Campello, Carlos Martín-Isla, David Viladés, Martín L Descalzo, Andrea Guala, José F Rodríguez-Palomares, and Karim Lekadir. Domain generalization in deep learning for contrast-enhanced imaging. *Computers in Biology and Medicine*, 149:106052, 2022.

[78] Cristian Izquierdo Morcillo, Guillem Casas Masnou, Carlos Martin-Isla, Víctor Manuel Campello, Andrea Guala, Polyxeni Gkontra, José F Rodriguez-Palomares, and Karim Lekadir. Radiomics-based classification of left ventricular non-compaction, hypertrophic cardiomyopathy, and dilated cardiomyopathy in cardiovascular magnetic resonance. *Frontiers in Cardiovascular Medicine, 2021, vol. 8*, 2021.

[139] Zahra Raisi-Estabragh, Cristian Izquierdo, Victor M Campello, Carlos Martin-Isla, Akshay Jaggi, Nicholas C Harvey, Karim Lekadir, and Steffen E Petersen. Cardiac magnetic resonance radiomics: basic principles and clinical perspectives. *European Heart Journal-Cardiovascular Imaging*, 21(4):349–356, 2020.

[34] Víctor M Campello, Carlos Martín-Isla, Cristian Izquierdo, Andrea Guala, José F Rodríguez Palomares, David Viladés, Martín L Descalzo, Mahir Karakas, Ersin Çavuş, Zahra Raisi-Estabragh, et al. Minimising multi-centre radiomics variability through image normalisation: a pilot study. *Scientific Reports*, 12(1):12532, 2022.

Chapter 2

Segmentation of complex morphologies: Abnormal Right Ventricle segmentation in MRI

This chapter was published as:

Carlos Martín-Isla, Víctor M Campello, Cristian Izquierdo, Kaisar Kushibar, Carla Sendra-Balcells, Polyxeni Gkontra, Alireza Sojoudi, Mitchell J Fulton, Tewodros Weldebirhan Arega, Kumaradevan Punithakumar, Lei Li, Xiaowu Sun, Yasmina Al Khalil, Di Liu, Sana Jabbar, Sandro Queirós, Francesco Galati, Moona Mazher, Zheyao Gao, Marcel Beetz, Lennart Tautz, Christoforos Galazis, Marta Varela, Markus Hullebrand, Vicente Grau, Xiahai Zhuang, Domenec Puig, Maria A Zuluaga, Hassan Mohy-ud-Din, Dimitris Metaxas, Marcel Breeuwer, Rob J van der Geest, Michelle Noga, Stephanie Bricq, Mark E Rentschler, Andrea Guala, Steffen E Petersen, Sergio Escalera, José F Rodríguez Palomares, Karim Lekadir. "Deep Learning Segmentation of the Right Ventricle in Cardiac MRI: The M&ms Challenge". In: *IEEE Journal of Biomedical and Health Informatics* (2023).

Doi: 10.1109/JBHI.2023.3267857.

2.1 Abstract

In recent years, several deep learning models have been proposed to accurately quantify and diagnose cardiac pathologies. These automated tools heavily rely on the accurate segmentation of cardiac structures in MRI images. However, segmentation of the right ventricle is challenging due to its highly complex shape and ill-defined borders. Hence, there is a need for new methods to handle such structure's geometrical and textural complexities, notably in the presence of pathologies such as Dilated Right Ventricle, Tricuspid Regurgitation, Arrhythmogenesis, Tetralogy of Fallot, and Inter-atrial Communication. The last MICCAI challenge on right ventricle segmentation was held in 2012 and included only 48 cases from a single clinical center. As part of the 12$^{\text{th}}$ Workshop on Statistical Atlases and Computational Models of the Heart (STACOM 2021), the M&Ms-2 challenge was organized to promote the interest of the research community around right ventricle segmentation in multi-disease, multi-view, and multi-center cardiac MRI. Three hundred sixty CMR cases, including short-axis and long-axis 4-chamber views, were collected from three Spanish hospitals using nine different scanners from three different vendors, and included a diverse set of right and left ventricle pathologies. The solutions provided by the participants show that nnU-Net achieved the best results overall. However, multi-view approaches were able to capture additional information, highlighting the need to integrate multiple cardiac diseases, views, scanners, and acquisition protocols to produce reliable automatic cardiac segmentation algorithms.

Table 2.1: Automatic CMR segmentation challenges in figures

Challenge	Year	Cases	Number of scanners	Target Regions	Multiview	Techniques used	Number of pathologies	Stratified by pathology
M&Ms-2	2021	360	9	RV	✓	Deep Learning	8	✓
RVSC	2012	48	1	RV	X	Atlas-based	6	X
M&Ms	2020	375	5	LV/RV/MYO	X	Deep Learning	6	X
ACDC	2017	150	1	LV/RV/MYO	X	Deep Learning	5	✓
LVSC [158]	2011	200	-	MYO	X	Atlas-based	1	X
Sunnybrook [138]	2009	45	1	LV/MYO	X	Atlas-based	4	✓

2.2 Introduction

The role of the right ventricle (RV) in circulation has historically been overshadowed by that of the left ventricle (LV). For years, RV dysfunction was thought to not contribute significantly to cardiac output and pressures, while LV was considered the key player in

cardiac hemodynamics [67]. This led to RV receiving limited attention, and often being described as the "forgotten ventricle" [142]. However, in the past few decades, the misconception regarding the lack of impact of the RV dysfunction in cardiac function has changed [5, 30, 117, 50, 51, 84]. A significant amount of research has progressively demonstrated the pivotal role of RV in cardiac function, and its implication and prognostic value in high-burden diseases, such as heart failure and/or pulmonary hypertension [28, 55, 62], dilated cardiomyopathy [65], tricuspid regurgitation [141], tetralogy of fallot [104], to name a few.

Given the prognostic significance of RV, the clinical interest has shifted in recent years from a simple visual inspection of the RV from cardiac magnetic resonance imaging (CMR), the reference modality for RV assessment [118], to extracting quantitative RV parameters by first segmenting the structure. Despite this renewed interest of the medical community to quantitatively assess the RV [80], the artificial intelligence community has lagged in providing fully automated solutions for RV segmentation from CMR, that are as accurate as for LV [32], and in benchmarking deep learning (DL) algorithms, the current state-of-the-art in medical imaging.

More precisely, the last challenge focused on RV segmentation using CMR data was the Right Ventricle Segmentation Challenge Dataset (RVSC) [132]. Prior to the RVSC, challenges solely focused on the myocardium and LV (Table 2.1). Despite its significance, the RVSC challenge was organized back in 2012 when DL was still in its early development and not yet adopted for CMR segmentation [22]. Therefore, none of the seven participants in the challenge used DL. Three approaches were atlas-based, two prior-based, and the other two based on cardiac motion without needing prior information. The best semi-automated methods achieved a dice accuracy of 80% and a Hausdorff distance of 1 cm. At the same time, automated approaches demonstrated a similar performance at the expense of higher computational costs. At those times, this performance level was competitive, but it is now considered far from what the current state-of-the-art DL-based models could achieve.

The early application of DL in CMR segmentation using a Fully Convolutional Network (FCN) by Tran [169] showed improved results compared to prior CMR segmentation methods in RVSC, LVSC, and Sunnybrook datasets. At the same time, the U-Net[145] architecture, which added a symmetric decoding path to the FCN architecture, started gaining inertia along the biomedical imaging segmentation community. However, Lieman et al. [101] shown that there was no statistical difference in CMR segmentation performance between the two architectures, with FCN outperforming U-Net in LV volume prediction using a large sample size of 1,143 subjects. This was

further validated by Bai et al. [14] in a large-scale study using 4,875 cases for the bi-ventricular segmentation task.

In recent CMR segmentation challenges such as the Automated Cardiac Diagnosis Challenge (ACDC) [22] and the MMs challenge [32], the U-Net architecture has emerged as the dominant choice. In the ACDC, research by Baumgartner et al. [17] showed that U-Net outperformed the Fully Convolutional Network (FCN) in all proposed segmentation tasks, except for the RV end diastolic average symmetric surface distance. The early stages of the nnU-Net framework [75], which is capable of optimizing preprocessing, network architecture, training, inference, and post-processing automatically without manual intervention, were also demonstrated by Isensee et al. [72] in such challenge. The Top-3 participants in the MMs challenge [32] used nnU-Net.

Nonetheless, both aforementioned challenges were focused on cardiac multi-structure segmentation, and the best performance was achieved for the LV and the myocardium. The reduced accuracy in the RV segmentation can be explained by the additional challenges posed by the complex geometry and appearance of the RV. These include its irregular shape, the heterogeneity in the appearance and thickness of its free wall, and its complex trabeculations [132]. As a result, several works have been recently proposed to improve RV segmentation. [195, 127, 106, 180, 98, 27, 70, 8, 14, 83]. Nonetheless, the scarcity of relevant public CMR data has resulted in the vast majority of current state-of-the-art methods using the data provided by the RVSC challenge which comprises solely 48 cases from a single clinical center. Moreover, while the cohort includes diverse pathologies, the considered diseases are not directly related to the RV. Lastly, the complementary long-axis 4-chamber views, particularly helpful for improving RV apical and basal slices segmentation, were not released. Other relevant works using larger datasets, such as that of Chen et al. [43] based on 145 cases, although important, rely on private cohorts and, therefore, do not allow for benchmarking.

In response to the gap in public datasets and evaluation frameworks for computational approaches focused on automated RV segmentation from CMR, the Multi-Disease, Multi-View & Multi-Center Right Ventricular Segmentation Challenge (M&Ms-2) was organized as part of the Statistical Atlases and Computational Models of the Heart (STACOM) Workshop held in conjunction with the MICCAI 2021 Conference. This is the first work to provide a public multi-center, multi-disease, multi-view CMR dataset, associated contours, and an evaluation framework to benchmark DL algorithms for RV segmentation. Moreover, the dataset complements the dataset of the challenge's first edition [32], a reference dataset for multi-structure segmentation, by providing multi-view information and other diseases relevant to RV dysfunction. In total, the

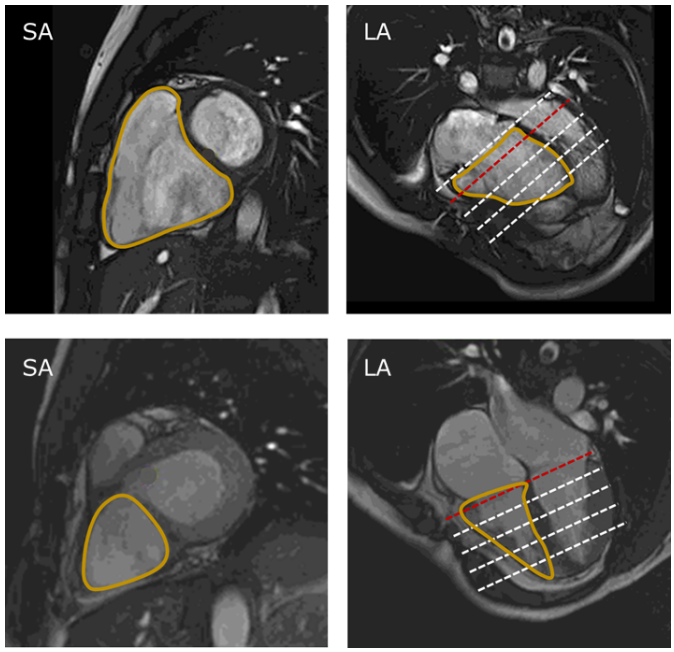


Figure 2.1: Visual appearance of short-axis (SA) and long-axis (LA) views of pathological (upper row) and healthy (lower row) subjects. Dashed lines (white) correspond to the projection of SA slices into the LA view. The red dashed line corresponds to the projection of the SA slice shown in the first column. The yellow line corresponds to ground truth delineations.

M&Ms-2 challenge dataset comprises CMR data from 360 participants originating from three Spanish hospitals. The data were acquired by nine different scanners from three different vendors (Siemens, Philips, and General Electric). The dataset was built in close collaboration with clinicians and accounts for seven different pathologies, while it also includes a control group of 75 healthy participants. It should be noted that the short-axis studies were annotated using the same Standard Operation Procedure (SOP) as previous reference challenges, while the complementary long-axis 4-chamber acquisitions for precise basal and apical delineation were also made publicly available.

In this paper, we present and discuss the results of the M&Ms-2 challenge in detail. The obtained results show the challenging nature of the task of automatically segmenting the RV from CMR images and the promise of the proposed solutions. Moreover, the findings of the challenge highlight the need for further research to build tools that can integrate multi-view cardiac information for the RV segmentation task in the presence of a diverse set of pathologies.

2.3 Challenge framework

2.3.1 Data preparation

A total of three clinical centers from Spain contributed to this challenge by providing several CMR studies with different left and right ventricular pathologies, namely:

Dilated Left Ventricle (DLV): LV is considered dilated when the LV end-diastolic volume measured in CMR is >214mL (>105mL/m2) in men or 179mL (>96mL/m2) in women.

Dilated Right Ventricle (DRV): RV is considered dilated when RV end-diastolic volume measured in CMR is >250mL (>121mL/m2) in men or 201mL (>112mL/m2) in women.

Hypertrophic cardiomyopathy (HCM) is an inherited heart disease defined by increased LV wall thickness (>15mm in one or more LV myocardial segments) that cannot be explained by abnormal loading conditions. In CMR, left ventricular mass typical values are 62-176g in men and 56-140g in women, and right ventricular mass typical values are 25-57g in men and 50-56g in women.

Arrhythmogenic cardiomyopathy (ARR) , inherited heart disease with a loss of myocytes and fibrofatty replacement of right ventricular myocardium; biventricular involvement is often observed. Diagnosis includes global RV dilatation and regional wall motion abnormalities with or without a decreased ejection fraction.

Tetrology of Fallot (FALL) is characterized by the following four features: a nonrestrictive ventricular septal defect, overriding aorta; right ventricle outflow tract obstruction and/or branch pulmonary artery stenosis; and consequent RV hypertrophy.

Inter-atrial communication (CIA) , a defect in the septum that separates the two atria. CMR is rarely required but may be useful for assessment of RV volume overload, identification of inferior sinus venous defect in the long-axis 4-chamber view, quantification of pulmonary to systemic flow ratio, and evaluation of pulmonary venous connection.

Tricuspid regurgitation (TRI) consists of the insufficiency of the tricuspid valve, causing blood flow from the RV to the right atrium during systole. In CMR, TRI appears as one or more flow jets emanating from the tricuspid valve and projecting into the RV. Jets are often holosystolic and readily apparent on the long-axis 4-chamber view.

In total, 360 studies were included. Images were acquired with different scanners, field strengths, and resolutions for both short-axis (SA) and long-axis 4-chamber (LA) views.

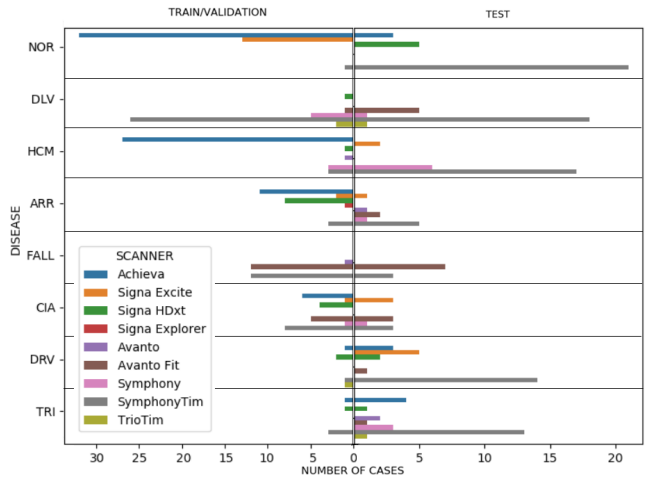


Figure 2.2: Distribution per pathology and scanner along train, validation, and test sets.

Most images were acquired from scanners with magnetic strength of 1.5T and a small fraction of 3.0T. The specific vendors are 1) Siemens (Siemens Healthineers, Germany) – including Avanto (AVA), Avanto Fit (AVF), Symphony (SYM), SymphonyTim (SYT), and TrioTim (TRT) scanners; 2) Philips (Philips Healthcare, Netherlands) – including Achieva (ACH) scanners; and 3) General Electric (GE, GE Healthcare, USA) – including Signa Excite (EXC), Signa Explorer (EXP), and Signa HDxt (HDXT) scanners. More specific details on the collected studies are given in Table 2.3.

The subjects included in this multi-disease study were selected among groups of the aforementioned cardiovascular diseases and healthy volunteers (NOR). The distribution of pathologies within the dataset partitions and scanners are specified in Table 2.2 and Figure 2.2.

Each CMR imaging study was annotated manually by an expert clinician from the corresponding center, with clinical experience ranging from 3 to over 10 years. The annotation process involved marking the short-axis and long-axis 4-chamber views at both end-diastolic (ED) and end-systolic (ES) phases, which correspond to the phases used to calculate clinically relevant biomarkers such as ejection fraction and myocardial mass, for cardiac diagnosis and monitoring. Furthermore, the basal slice of the RV at ED/ES was inferred from the position of the tricuspid annulus as defined on the long-axis 4-chamber view at ED/ES. The apical slice was defined as the last slice with a detectable ventricular cavity. Three main regions were provided: the left and right ventricular cavities and the left ventricle myocardium (MYO). However, the

Table 2.2: Number of studies per pathology in each dataset partition

	Number of studies		
Pathology	Training	Validation	Test
Normal subjects	40	5	30
Dilated Left Ventricle	30	5	25
Hypertrophic Cardiomyopathy	30	5	25
Congenital Arrhythmogenesis	20	5	10
Tetralogy of Fallot	20	5	10
Interatrial Communication	20	5	10
Dilated Right Ventricle	0	5	25
Tricuspidal Regurgitation	0	5	25
Total	160	40	160

Table 2.3: Average specifications for the images acquired in the different centers.

Center*	Vendor	Model	In-plane res. (mm) (SA/LA)	In-plane dim. (pixels)	Slice thickness (mm)	Number of slices	Field Strength (T)	Number of studies
A	Philips	Achieva	1.18/1.19	332±32/288 ± 38	10	10	1.5	88
B	GE	Signa Excite	1.40/1.58	270±28/258 ± 6	9.8	12	1.5	27
B	GE	Signa HDxt	0.98/1.08	420±124/420 ± 124	10	12	1.5/3.0	25
B	GE	Signa Explorer	0.78/0.78	512±0/512 ± 0	10	13	1.5	1
C	Siemens	Avanto	1.21/1.15	232±24/240 ± 20	14	9	1.5	5
C	Siemens	Avanto Fit	1.13/1.24	234±24/234 ± 24	9.9	11	1.5	37
C	Siemens	Symphony	1.27/1.27	232±24/240 ± 18	9.7	10	1.5	21
C	Siemens	SymphonyTim	1.34/1.24	230±36/238 ± 26	9.7	12	1.5	151
C	Siemens	TrioTim	1.15/1.20	234±24/238 ± 18	8.6	13	3.0	5

* A: Clínica Sagrada Familia, B: Hospital Universitari Dexeus, C: Hospital Vall d'Hebron.

evaluation was performed exclusively on the RV. Two additional researchers performed a detailed revision of the provided segmentation to reduce inter-observer and inter-center variability in the contours, particularly in the apical and basal regions. Discrepancies were resolved by consensus between the observers. Such observers applied the same SOP across all CMR datasets to obtain the final ground truth. To generate consistent annotations, we chose to apply the SOP that was already used by the ACDC and M&Ms challenges with an additional rule (d) as follows:

a) The LV and RV cavities, including the papillary muscles, must be completely covered.

b) No interpolation of the myocardial boundaries must be performed at the basal region.

c) The RV must have a larger volume at the ED time frame compared to ES.

d) Additionally, long-axis view is used as a reference to delimit the basal and apical regions, as stated above.

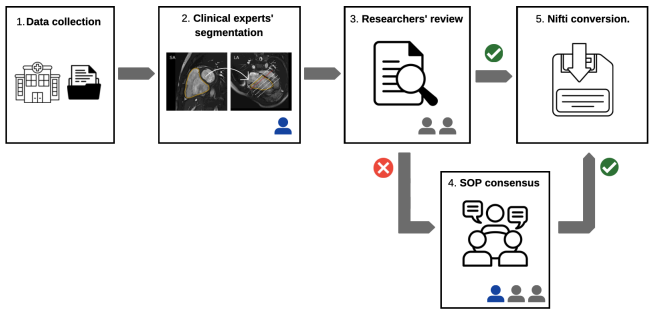


Figure 2.3: Data collection and pre-processing pipeline.

Clinical delineations and subsequent corrections were performed using the cvi42 software (Circle Cardiovascular Imaging Inc., Calgary, Alberta, Canada). All studies were provided in DICOM format, and contours were extracted in cvi42 workspace format (.cvi42ws). In-house software was then used to create the contours and transform the images into NIFTI format, and this final file format was delivered to the challenge participants. The inter-view correspondence was preserved during pre-processing. Figure 2.3 presents the data collection and pre-processing pipeline.

2.3.2 Model training and validation

The 360 CMR studies were divided into training, validation, and testing, as detailed in Table 2.2. The participants received the 160 training cases with annotations for short and long-axis views and 40 validation cases without annotation on May 10^{th}, 2021. Two pathologies, DRV and TRI, were excluded from the training dataset to test the generalization capability of the models to unseen pathologies. In order to optimize the models, the participants were allowed to automatically inspect their models' performance against 40 validation CMR cases, *i.e.* 5 from each of the pathologies, and publish their validation scores using the Codalab platform [131]. A maximum of 20 submissions per team were allowed during the validation process. Note that it was not permitted to use any external datasets or pre-trained models during training.

2.3.3 Model evaluation

The testing phase started on July 1st, 2021, and concluded on July 20th, 2021. The participants were forced to evaluate their models remotely to ensure the unseen test set was hidden from the segmentation methods. The organizers' GPU server infrastructure with five NVIDIA 3090 RTX GPUs was provided to evaluate the submissions. The participants were asked to assess their models by submitting their trained models to the Codalab platform and executing them using a Docker[1] image.

To assess the quality of the automatic segmentations (P) against the ground truth (G), two measures were used:

(i) Dice similarity coefficient (DSC) – degree of overlapping of two volumes:

$$DSC(P,G) = \frac{2|P \cap G|}{|P| + |G|} \tag{2.1}$$

(ii) Hausdorff distance (HD) – largest disagreement between the contours, useful for identifying small outliers:

$$HD(P,G) = \max\left\{\sup_{p\in P} d(p,G), \sup_{g\in G} d(g,P)\right\} \tag{2.2}$$

where sup represents the supremum, inf the infimum, and

$$d(a,B) = \inf_{b\in B} d(a,b) \tag{2.3}$$

quantifies the distance from a point $a \in$X to the subset $B \subseteq$X.

These metrics were computed for the RV segmentation from both SA and LA views, resulting in 4 measures for each cardiac phase. If one participant had a prediction missing for a specific subject, a zero value was assumed for DSC. A distance of $50mm$ was considered for HD, $10mm$ above the maximum HD distance computed across all participants and cases.

To obtain the final ranking, HD was min-max normalized across all subjects ($\widehat{HD}$) to get a number between 0 and 1 for ED and ES phases in both SA and LA views independently. Due to the difference in dimensionality between SA and LA views, a weighted average was performed. The weighted metric, M, was obtained as follows:

$$M = \frac{0.75(DSC_{SA} + \widehat{HD}_{SA}) + 0.25(DSC_{LA} + \widehat{HD}_{LA})}{2} \tag{2.4}$$

where DSC and HD are the average of the corresponding metrics in ED and ES:

[1] https://www.docker.com/

$$DSC = \frac{DSC_{ED} + DSC_{ES}}{2} \; and \; HD = \frac{HD_{ED} + HD_{ES}}{2} \tag{2.5}$$

The normalized metrics returned a performance between 0 and 1, being 1 the value that a team would obtain if it had perfect results for every metric.

2.4 Participating methods

More than 120 teams registered to download the M&Ms-2 training dataset, 17 submitted a solution for the final testing phase, and 15 teams presented their methodology as a paper to the STACOM Workshop (see Table 2.4 for the participant details). Table 2.5 summarizes the main features of the submitted techniques, which are described in more detail in the following subsections.

Table 2.4: List and details of the participating teams in the challenge.

Team	Institution	Location
P1	University of Colorado Boulder	Boulder, USA
P2	ImViA Laboratory, Université Bourgogne Franche-Comté	Dijon, France
P3	Dept. of Radiology and Diagnostic Imaging, University of Alberta	Edmonton, Canada
P4	School of Data Science, Fudan University	Shanghai, China
P5	Department of Radiology, Leiden University Medical Center	Leiden, Netherlands
P6	Eindhoven University of Technology	Eindhoven, Netherlands
P7	Department of Computer Science, Rutgers University	Piscataway, USA
P8	Department of Electrical Engineering, Syed Babar Ali School of Science and Engineering	Lahore, Pakistan
P9	Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho	Braga, Portugal
P10	Data Science Department, EURECOM	Sophia Antipolis, France
P11	Department of Computer Engineering and Mathematics, University Rovira i Virgili	Tarragona, Spain
P12	School of Data Science, Fudan University	Shanghai, China
P13	Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford	Oxford, UK
P14	Charité - Universitatsmedizin Berlin,	Berlin, Germany
P15	Department of Computing, Imperial College London	London, UK

2.4.1 Backbone architectures

There is a degree of diversity in the backbone architectures employed by the various participants (as depicted in Table 2.5). This subsection will provide a comprehensive overview of the various architectures implemented by the participants.

nnU-Net architectures

Six teams used the nnU-Net [74] framework as their baseline segmentation models (P1–P6). The nnU-Net framework includes 2D, 3D and cascaded 2D/3D U-Net [146] architectures. The choice of base architecture for a specific segmentation problem is left to the user. In the case of 3D short-axis (SA) volumes, the variations among P1-P6

Table 2.5: Characteristics of participating models. Spatial Augmentation includes rotations, flipping, scaling, and deformations. Intensity augmentation includes Gaussian noise, brightness, gamma, and contrast.

	Architecture			Data Augmentation		
Method	Backbone	Additional Features	Multiview	Spatial	Intensity	Other
P1[58]	nnU-Net	Deformable Bayesian Convolutions	X	✓	✓	
P2[6]	nnU-Net	Dropout + Batch Normalization	X	✓	✓	MRI-Specific
P3[135]	nnU-Net	Default configuration	X	✓	✓	
P4[99]	nnU-Net	Cross-view ROI detection LA → SA	✓	✓	✓	
P5[159]	nnU-Net	Spatial and temporal Multi-channel input	X	✓	✓	Label propagation
P6[3]	nnU-Net	ROI detection, Intensity-based Multi-channel input	X	✓	✓	SPADE Synthesis
P7[103]	DLA	Cross-view refinement network	✓	✓	✓	Histogram Matching
P8[79]	U-Net	Shared Bottleneck between views	✓	✓	X	
P9[137]	xU-Net	3D Unit + 2D Unit with cross-view mid-fusion	✓	✓	✓	Test Time Augmentation
P10[59]	U-Net	OoD detection and refinement (Convolutional Autoencoder)	X	X	X	
P11[116]	U-Net	Single 2D network. Expansion, depth-wise, projection block	X	✓	X	Test Time Augmentation
P12[61]	U-Net	Tranformer encoder in the bottleneck, cross-view consistency loss	✓	✓	✓	
P13[19]	AttU-Net	ROI detection, cross-over Attention	✓	✓	✓	Histogram Matching, Fourier
P14[165]	U-Net	Multi-view 3D mesh reconstruction	✓	X	X	
P15[60]	MPFP+ViT	Multi-scale Feature Pyramid, Geometric Spatial Transformer	✓	✓	✓	In-painting

models were primarily in terms of the input dimensionality, with some additional minor modifications to the base architecture. All of these methods produced separate models for each view.

P1 adopted a 2D nnU-Net for both SA and LA views and replaced its bottleneck convolutions with deformable Bayesian convolutions. Deformable convolutions enable an increased and adaptable receptive field without requiring additional convolution layers, while Bayesian convolutions improve generalisability and training speed.

P2 used a 3D nnU-Net for the SA view and a 2D nnU-Net for LA views and added batch normalisation instead of the default nnU-Net instance normalisation. P2 also added dropout of 0.2 to the intermediate layers of the network.

P3 used an ensemble of 2D and 3D nnU-Nets for the SA segmentation task and a regular 2D nnU-Net for the LA view. The default nnU-Net architectures were used.

P4 trained a default 2D nnU-Net for LA views and used its output to delimit SA views along the z axis and trained a default 3D nnU-Net with the extracted region of interest.

P5 and P6 used multi-channel late fusion approaches in their independent default 2D nnU-Nets for both SA and LA views.

P5 used stacks of three registered CMR consecutive images to train a three input channel 2D nnU-Net. While spatial and temporal information were porposed to generate the SA stacks, the LA stacks only incorporated temporal information.

P6 used six filtered versions of each 2D image as input for 2D nnU-Net with six input channel. The images feeded to this network were pre-processed extracting the region of interest by means of a regression CNN that delimited them to their bounding box.

U-Net architectures

Seven participants (P8–P14) constructed their architectures on top of a traditional U-Net.

P8 generated a multi-view SA-LA model consisting of two 2D U-Net structures with a shared bottleneck. Each of the out-of-plane 2D SA slices belonging to the same subject received the same complementary LA view and their features were concatenated in the bottleneck, training simultaneously both SA and LA views in a single end-to-end model.

P9 combined a 2D U-Net with a 3D U-Net in a unified model. In order to achieve this goal, both views are centered around the mean position of their original centroids. Moreover, both images are rotated to align their axes, where the LAx image is rotated to make its Y-axis match the Z-axis of the SAx stack. To take advantage of the complementary spatial context offered by both aligned views, a set of 3 cross-view modules were placed at the end of the three lowest levels in the compression path. Each cross-view module concatenated SA and LA information and retrieved a new set of spatially significant features using a 1x1 convolution layer. At inference time, SA and LA views were reoriented to their original pose.

P10 implemented for each view two 2D U-Nets and a 2D autoencoder. The architecture used in the implementation of the U-Net networks corresponds to the best methods presented in [32] and [22], while the architecture used in the implementation of the autoencoder network can be found in [20]. While the segmentation network used pairs of input images and their manual delineations, the autoencoder was trained to reconstruct delineations of the training set. The autoencoder loss was used as a quality control measure, being backpropagated to the U-Net when a poor quality was detected. At inference time, the best segmentation network was selected for each subject, taking in consideration the quality assessment of the autoencoder.

P11 used a single 2D U-Net for both views and replaced the standard convolutional blocks of its decoder with expansion, depth-wise, and projection blocks. These blocks extract helpful information with less computational complexity and thus allowed to increase the number of channels in the decoding stage. Channels are then combined via depthwise convolutions and finally collapsed to the original depth in the projection

stage. Additionally, P11 added residual blocks to the standard U-Net skip connections.

P12 proposed co-training a pair of 2D U-Nets end-to-end. The main modification of the backbone U-Net architecture used in each branch consist of the addition of an transformer module to the bottleneck that established self-attention mechanisms on high-level convolutional features. At training time, a SA slice and its complementary LA view were simultaneously fed to the paired U-Net. The segmentations obtained were then mapped between views using their complementary affine transformations. The final loss consisted of a combination of per-view standard DSC score and the co-segmentation SA to LA and LA to SA inter-view DSC scores.

P13 used Attention U-Net [128] as backbone. Initially, two 2D Attention U-Nets were utilized to extract the heart's location in both LA and SA views. The information from both views was then combined into one volume. For LA segmentation, the cropped LA slice and three mid-cavity SA slices were joined together. For SA segmentation, the cropped SA slice is combined with the cropped LA slice, allowing access to additional anatomy information in the basal and apical heart regions. Finally, each volume is processed as a multi-channel input through a separate Attention U-Net to produce the final segmentation masks for each view.

P14 used independent 2D U-Nets for LA and SA views and combined them into a 3D deformable model to improve quantification and volumetry. An initial 3D deformable model was triangulated directly from the SA segmentation contour points obtained from the network. SA apical and basal planes were estimated from the obtained LA segmentation and used to reconstruct the final SA volume.

Other architectures

P7 used 2D Deep Layer Aggregation (DLA) networks as a backbone for both SA and LA views. Being the backbone the same presented in [182]. The implementation consists of two stages: initially, two individual networks were employed to segment the SA and LA images independently. In the following stage, the results are then jointly refined using two additional networks. Four networks were trained independently in total, all having a similar structure except for the refinement networks, whose input comprised the original image, the respective 2D segmentation, and the aligned segmentation obtained from the complementary SA/LA view. Both stages were trained independently.

P15 propose a new hybrid 2D/3D geometric spatial Transformer Multi-Pass feature pyramid to simultanenously segment SA and LA views. The architecture consists of 2D SA/LA feature pyramid [102], independent 3D (SA) and 2D (LA) branches and finally

a geometric spatial transformer (GST). The feature pyramid receives individual 2D in-plane complementary slices for both the SA and LA as inputs and extracts features at different downsampling levels. Then, the SA features are regrouped in a 3D SA stack, and a segmentation is obtained by means of a simple 3D convolutional residual block. LA features pyramids follow the same procedure on its 2D counterpart.

The GST takes as input the pre-computed affine matrix and the complementary LA and SA views. After projecting SA volume to its complementary LA view, both are concatenated and merged via a 2D convolutional block to obtain a refined LA prediction.

2.4.2 Data augmentation

Data augmentation (DA) is a widely utilized technique that helps to enhance the performance of deep learning algorithms through improved generalization and regularization. Its utilization in the medical imaging field has been well documented[44], and it has been consistently shown that incorporating DA can greatly benefit segmentation tasks in cardiovascular magnetic resonance imaging[41, 32].

All participants in the challenge (except P10 and P14) used some form of data augmentation to enhance their models. Specifically, two kinds of data augmentations were considered: (1) spatial transformations to increase sample size through flipping, rotation, scaling, or deformation of the original images; (2) intensity-driven techniques, which maintain the spatial configuration of the anatomical structures but modify their image appearance. Both augmentation families seem particularly relevant for the M&Ms-2: while spatial transformations can reduce the gap between seen and unseen anatomies and pathologies, intensity-driven techniques are useful in the presence of heterogeneous imaging protocols and scanner vendors. Two teams performed data augmentation using only spatial transformations (P8, P11). Nine teams utilized intensity-based augmentations using standard image transformations such as blurring, change in brightness and contrast, or addition of Gaussian noise (P1–P9, P12–P13, P15). P3, P6 and P7 added histogram matching to their pool of intensity transformations. Additionally, P2 used MRI-specific augmentations such as random bias fields, random ghosting, and random motion artifacts to increase the textural variability of the images.

P5 and P6 added more sophisticated augmentations to their pipeline, and both methods used multi-channel inputs.

P5 registered temporal (SA and LA views) and spatial (z-axis SA view) and propagated the label information to unlabeled temporal phases to increase the training

set. As described in the previous subsection, triplets of consecutive unlabeled images were effectively used to pretrain each SA and LA multi-channel net, taking as ground truth a registered label from an annotated cardiac phase. Since the propagated masks are not as accurate as the manual segmentations, the network was fine-tuned using the real labeled images and the adjacent registered cardiac phases.

P6 applied advanced image synthesis by using Generative Adversarial Networks (GANs). In particular, P6 used the method proposed in SPADE [129] to increase the number of samples per vendor and per cardiac region in an anatomically consistent way. The augmentation consisted of morphological manipulations of the segmentation masks to obtain synthetic images with the desired RV cavity shape. Multi-channel augmentations were then applied on top of synthesis, as a stack of intensity transformed channels and the the original (real or synthetic) image. The transformed images were obtained using Laplacian, posterization, and edge-preserving filters.

On the other hand, P6 also proposed two data balancing strategies: (1) For SA stacks, the mid-ventricular slices cover most of the 3D volume, generating unbalance between basal, mid-ventricular and apical regions when using a 2D segmentation model. Approaches such as [111] alleviated this effect using balanced batches of the different short axial regions i.e. apical, basal and mid-ventricular regions. Following the same principle, P6 generated synthetic basal samples from randomly deformed segmentations.

(2) Since the provided dataset is acquired using 9 different scanners with a different number of samples per scanner and vendor, it is appropriate to consider some degree of unbalance related to domain shifts. Approaches such as [49, 149] tried to minimize the domain shift negative effects using domain adversarial training. P6 instead identified a set of outliers for each vendor based on the computed RV cardiac indices. Then, each vendor was synthetically augmented up to 1000 times, incorporating a 50% of outliers and a 50% of regular cases.

Finally, P13 added Fourier Domain Adaptation [181] to alleviate vendor differences.

2.5 Results

As shown in Table 2.2, a diverse testing set integrating nine scanners and eight cardiac pathologies was prepared for evaluating the final submissions with a total of 160 subjects. We show the obtained results per team, per cardiac region, per pathology and per clinical indices. Additionally, we show some qualitative results.

To understand and analyse the participating methods in this challenge, we have performed the following experimental comparisons. Firstly, we rank the participants

exactly as it was presented during the challenge workshop. Secondly, we further dissect the results to emphasize different aspects and qualities of cardiac segmentation, such as pathological groups, cardiac regions or clinical indices. Thirdly, we perform a qualitative comparison of the approaches of the participants.

2.5.1 Team Ranking

The results of the challenge, as displayed in Table 2.6, present the evaluation of all participants using two relevant segmentation metrics (DSC and HD) for both SA and LA acqusitions. Additionally, the average inference time is included in terms of volumes per second for SA acquisitions and images per second for LA acquisitions . The inference time for methods using an unified model whose inference time could not be computed independently (P8, P9 and P15) for each view present a single inference time . Lastly, a Welch's t-test was conducted to determine statistical significance between participants' evaluation performance.

Table 2.6: DSC and HD and inference time for the final submissions of all participants. HD is measured in millimeters. Volume Error is measured in mililitres. Inference time is measured in seconds per volume.Boldface numbers are the best results for each column. Blue numbers represent results are not significantly different compared to the top-performing method for each column (p-value > 0.01 for Welch's t-test)

	SA					LA				
	ED		ES			ED		ES		
Method	DSC	HD	DSC	HD	Inference (s)	DSC	HD	DSC	HD	Inference (s)
P1	0.934	**9.610**	0.910	10.032	1.72	**0.935**	6.227	0.904	**5.935**	0.34
P2	0.932	10.078	0.910	**9.782**	0.86	**0.935**	**6.028**	**0.905**	6.188	0.11
P3	**0.940**	10.122	**0.914**	9.987	1.8	0.931	6.337	0.904	5.976	0.42
P4	0.933	10.563	0.907	10.050	2.22	0.930	6.246	0.902	6.097	0.54
P5	0.937	10.879	0.913	10.300	2.43	**0.935**	6.056	0.903	6.031	0.17
P6	0.927	9.941	0.897	10.307	2.74	0.907	8.444	0.883	7.265	0.56
P7	0.932	10.517	0.903	10.880	4.11	0.923	7.371	0.902	6.019	1.23
P8	0.923	11.258	0.897	11.062	2.23	0.910	7.757	0.882	6.933	-
P9	0.924	11.327	0.898	11.447	2.89	0.922	7.173	0.900	6.391	-
P10	0.916	11.681	0.890	11.347	2.12	0.923	7.846	0.894	6.970	0.34
P11	0.909	15.275	0.880	14.606	0.67	0.888	9.323	0.854	8.347	0.18
P12	0.844	15.495	0.821	16.750	2.34	0.887	9.733	0.851	9.659	0.42
P13	0.873	16.682	0.791	18.499	3.12	0.852	11.325	0.829	9.591	0.30
P14	0.883	17.024	0.838	17.803	4.27	0.839	13.303	0.809	13.716	0.67
P15	0.852	19.430	0.821	19.117	1.54	0.814	18.629	0.781	17.198	-

2.5.2 Results per Pathology

Figure 2.4 summarizes the average DSC per pathology according to equations (2.4) and (2.5). This dissection is particular relevant since accurate segmentation of different pathologies is critical for several clinical applications, including diagnosis, treatment planning, and monitoring disease progression.

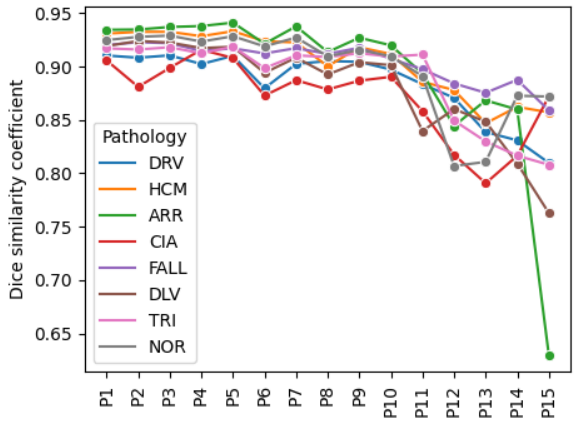


Figure 2.4: Weighted average DSC per pathology according to equations (2.4) and (2.5).

In order to evaluate the ability of the proposed methods to generalize to new, unseen pathological groups, subjects with Dilated Right Ventricle and Tricuspidal Regurgitation were omitted from the training phase. A Mann-Whitney U rank test, with a significance level of 0.05, was conducted for each participant to compare their segmentation DSC scores for known and unknown pathologies. The results of this analysis are presented in Figure 2.5 in an organized manner, separated by imaging view and cardiac phase.

2.5.3 Results per Cardiac Region

The examination of various segments of the heart, including the apical, basal, and mid-ventricular regions, is crucial for determining the individual impact each region may have on the segmentation error. To illustrate such impact, Figure 2.6 shows the average performance of P1–P5 in SA volumes from basal to apical planes. Further analysis is presented in relation to the detection of the basal plane, whose contribution to segmentation accuracy is greater than another regions: detection rate of the basal

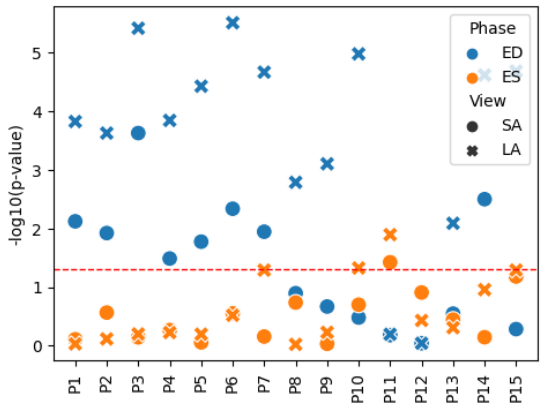


Figure 2.5: Statistical difference according to the Mann-Whitney U rank test for DSC scores between seen and unseen pathologies. The red dashed line stands for the 0.05 significance threshold.

plane, as shown in Figure 2.7, presents the number of subjects per participant where there was a disagreement regarding the manual delineation in the detected first basal slice.

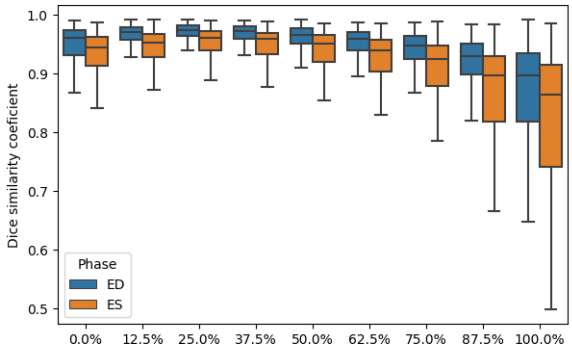


Figure 2.6: Average performance of the top 5 ranked methods in SA from basal (0%) to the apical (100%) regions.

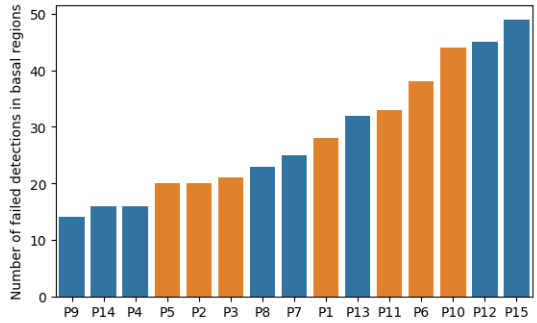


Figure 2.7: Number of not segmented slices at basal region. In blue, multi-view approaches. In orange, non-multi-view approaches.

2.5.4 Clinical Measurements

In the assessment of cardiac function, clinical metrics such as End-Diastolic and End-Systolic Volumes, and Right Ventricle Ejection Fraction (RVEF) are commonly utilized indices. However, geometrical metrics, such as DSC and HD, may not always correlate with these indices. This lack of correlation is attributed to the scalar, rather than spatial, nature of the clinical indices, which can result in good estimations of volumes and RVEF even when the contour is not accurately defining the cardiac structures. For such reason, the beforementioned clinical measurements are presented in Table 2.7, in term of i) correlation (corr), ii) mean average error (mae), and iii) bias. Note that, may be the case where ED volume is not accurately predicted. In such case the RVEF, defined as $(Vol_{ED} - Vol_{ES})/Vol_{ED}$ can increase considerably or be infinite. In such cases a RVEF of 100% was considered.

Outliers play a crucial role on integrating automatic segmentation methods in clinical practice, as a single missed case or a significant discrepancy in a few instances can have a greater impact than a small average improvement that may not make a noticeable difference in diagnostic tasks. In Table 2.8, the number of cases exhibiting an RV ejection fraction above various thresholds is presented, alongside the number of cases in which computation was not feasible due to a missing segmentation in some of the cardiac phases.

Table 2.7: Clinical metrics for the 15 participating methods. Boldface numbers are the best results for each column. Blue numbers represent results that are not significantly different compared to the top-performing method for each column (p-value > 0.01 for Welch's t-test)

	Volume ED		Volume ES		RVEF		
Method	corr. val.	mae mL	corr. val.	mae mL	corr. val.	mae %	bias %±σ
P1	0.949	11.94	0.967	7.63	0.878	4.81	-0.31±6.9
P2	0.952	11.14	0.967	7.63	0.873	4.54	0.89±6.7
P3	**0.963**	**10.16**	0.967	7.63	0.891	4.4	0.65±6.2
P4	0.958	11.07	0.965	8.3	0.87	4.67	**-0.02±6.8**
P5	0.955	10.83	**0.97**	**7.22**	**0.892**	**4.36**	0.18 ±6.5
P6	0.915	13.49	0.936	9.16	0.864	4.77	-0.14±7.5
P7	0.951	11.61	0.964	8.52	**0.892**	4.64	-0.63±6.4
P8	0.95	11.94	0.954	9.04	0.855	5.26	0.36±8.0
P9	0.954	12.12	0.959	8.93	0.871	4.74	0.41 ±6.6
P10	0.944	14.55	0.93	10.98	0.764	6.65	0.6±9.7
P11	.913	15.79	0.917	11.01	0.772	4.96	1.38 ±9.3
P12	0.744	32.18	0.823	15.24	0.491	11.45	-7.88±15.0
P13	0.883	21.37	0.865	17.06	0.674	9.06	1.42±13.2
P14	0.898	16.99	0.886	13.00	0.671	7.74	-1.1 ±11.7
P15	0.732	23.10	0.825	14.87	0.55	11.7	4.19±18.1

Table 2.8: Number of patients above different RVEF error thresholds.

	RV Ejection Fraction Mean Average Error				
Method	$\geq 5\%$	$\geq 10\%$	$\geq 15\%$	$\geq 20\%$	Missing
P1	60	19	6	3	0
P2	59	16	4	2	1
P3	55	16	5	2	0
P4	57	14	4	3	0
P5	47	17	4	1	0
P6	49	15	7	3	0
P7	52	19	5	2	0
P8	58	22	9	6	0
P9	67	12	1	1	0
P10	59	34	13	8	1
P11	59	20	6	3	0
P12	116	70	36	14	1
P13	87	49	29	18	1
P14	79	37	15	11	2
P15	67	34	23	23	16

2.5.5 Qualitative results

Figure 2.8 provides some visual examples from different teams to discuss the possible limitations and strengths of the implemented methods. In the first row, complex basal regions for short-axis views are correctly captured by various multi-view approaches. All of these examples were not segmented by the top 5 non-multi-view strategies. In the second row, a pathological subject with a high degree of remodeling in the RV is not correctly segmented by the best-performing methods, capturing the surrounding tissue instead of the cardiac structure. P10 captured the cardiac structure as well as the surrounding tissue. P13 and P8 delineated only the cardiac structure with different degrees of accuracy. These methods merged SA and LA views in their networks without additional cross-view affine projections. Finally, the last two rows show highly remodeled right ventricular cavities correctly segmented by top-performing methods.

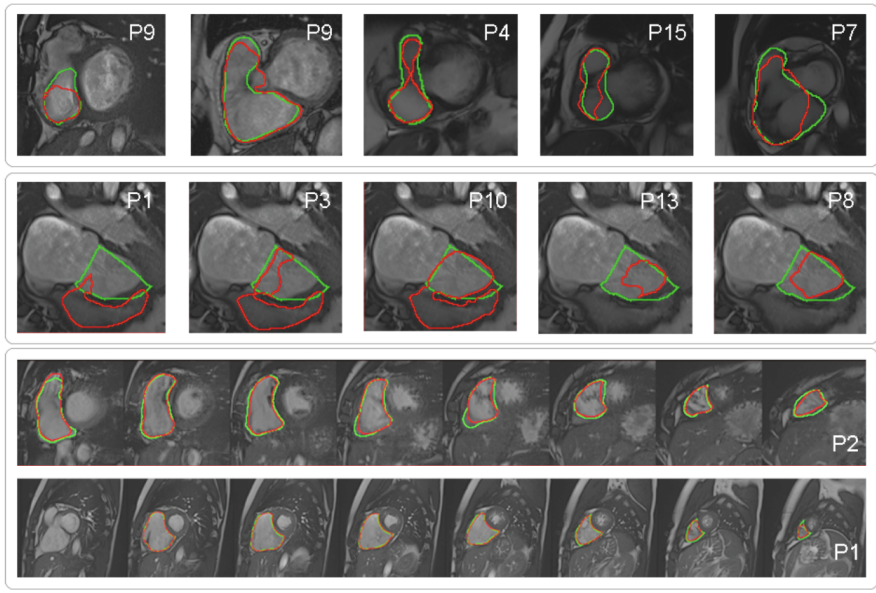


Figure 2.8: Prediction examples for some of the presented methods. The first row shows satisfactory segmentations at conflictive basal regions for SA images that were missed by non-multi-view approaches but correctly captured by multi-view methods. The second row shows a pathological subject with severe right ventricular dilation that was only correctly captured by multi-view methods. The last two rows show pathological subjects from unseen pathologies correctly segmented by top-ranked methods. Color correspondence: ground truth (green), prediction (red).

2.6 Discussion

This study presents a comprehensive evaluation of various automatic deep learning-based methods for multi-disease, multi-view, and multi-center right ventricular segmentation in cardiac magnetic resonance imaging (CMR). The 15 participants employed a diverse range of methodologies, including the choice of backbone architecture, number of stages, multi-view fusion, and data augmentation strategies. In addition to a large training sample of 160 cases, the authors were given 20 opportunities to optimize the parameters and characteristics of their models during the validation process using a well-stratified validation set of 40 cases. A Codalab-based automatic submission system was provided to allow for public comparison of performance and promote fair and dynamic competition between participants.

2.6.1 Summary of the challenge results

It can be concluded that the performance of the different proposals, and in particular for P1-P5, is relatively comparable. Statistical analysis has shown limited significant differences between the methods, with no clear advantage for any of the participants.

From a general point of view, our study supports several observations found in the previous edition of the challenge and other studies based on different CMR datasets. Specifically, the results confirm that end-diastolic segmentations are more accurate than end-systolic segmentations for the right ventricle. Additionally, the accuracy of segmentation decreases in the basal regions that are susceptible to under-segmentation and also is impacted in the apical regions due to their smaller size relative to the rest of the ventricular cavity.

The accuracy of segmentation is more stable across cardiac phases in comparison to previous challenges such as MMs-1 or ACDC, with an improvement of 0.042 in average DSC over MMs-1 and a comparable performance with ACDC (+0.004 average DSC), despite being MMs-2 a heterogeneous cohort.

2.6.2 Analysis of Pathologies

One of the relevant aspects of the challenge consists on evaluating the generalization capacity of the proposed methods to new, unseen pathologies. For this reason, the participants trained their models without access to subjects belonging to the Dilated Right Ventricle and Tricuspidal Regurgitation groups. Figure 2.4 shows that unseen pathologies perform consistently worse with exception of Inter-atrial communication. It is remarkable that three out of the ten subjects belonging to this group had a closure device visible in the basal region of the image.

We investigated in more detail the statistical differences between both, seen and unseen groups by analyzing independently the two annotated cardiac phases and the two views available in each study. The results in Figure 2.5 present some degree of statistical significance between seen and unseen pathologies for both SA and LA end-diastolic phases. This finding reveals the need for including diverse cardiac morphologies to improve model generalisability.

2.6.3 Single- versus Multi-view Models

Regarding multi-view approaches, eight out of fifteen participants utilized the complementary information between views. Although a definitive conclusion cannot be

drawn on the general benefits, the evaluation suggests that multi-view methods have the potential to improve basal plane detection in certain circumstances. Specifically, participants P9, P14, and P4 achieved a lower number of not-segmented basal slices. Additionally, some of the multi-view approaches presented a better RVEF stability. The solution proposed by P9 obtained the lowest number of cases with a RVEF error $\geq$ 10%, improving significantly the results obtained by P1-P5, as expressed in Table 2.8. Further research is required to incorporate multi-view techniques into thoroughly optimized frameworks such as nnU-Net.

2.6.4 Impact on clinical indices

We also assesed the participating methods by computing the clincal indices derived from the generated segmentations. The results were consistent with the ranking presented in Table 2.6, with almost any statistical difference between the Top-10 ranked methods. Interestingly, the multi-view approach P9 presented a more consistent Ejection Fraction across patients, with fewer cases with RVEF error greater than 10%. This point is specially relevant for diagnostic tasks.

2.6.5 Further considerations

Due to the high heterogeneity of the presented dataset, one could argue that there are many sensible parameters affecting the segmentation performance. Different image dimensions, in-plane resolutions or field strengths may be critical parameters for a DL segmentation algorithms.

Field strength: only five out of twelve samples obtained using 3T scanners were included in the test set. Despite the small sample size, there were no substantial differences in the segmentation performance between 1.5T and 3T acquisitions.

In-plane resolution: In Table 2.3, we presented a wide range of in-plane resolutions and volume dimensions directly related to the acquisition scanners. When comparing the average performance of P1-P5 across different scanners, we obtain a stable DSC of 0.912$\pm$0.016 for long-axis 4-chamber images and 0.922$\pm$0.011 for short-axis volumes. Interestingly, the learning methods were able to generalise correctly to the heterogeneous set of scanners, resolutions and protocols present during the training stage.

2.6.6 Future work

In addition to the analyses and results presented in this paper, we also provide the M&Ms-2 dataset open-access for the community, which can be downloaded from the M&Ms-2 website[2]. In conjunction with M&Ms-1, it represents the most heterogeneous dataset ever compiled in CMRs image analysis, comprising CMRs from various imaging protocols and cardiology units. It also includes a wide range of cardiovascular diseases and multi-view information. It is anticipated that the scientific community will embrace the dataset as a comprehensive resource to support a wide range of automated cardiac imaging research initiatives, including automatic pathology assessment, multi-scanner and multi-view image registration, multi-structure segmentation, cardiac imaging quantification, strain and motion analysis, and image synthesis. Further efforts will focus on incorporating 2-chamber and 3-chamber long-axis views to fully leverage the multi-view aspect of cardiac magnetic resonance studies. The integration of diverse disease characterization with these various views will also be pivotal in facilitating automatic evaluation and diagnosis.

2.7 Conclusions

To summarize, the key conclusions are:

1. The main findings correlate with the obtained results in previous CMR segmentation challenges: end-systolic phase and basal and apical cardiac regions are more conflictive than their counterparts.

2. nnU-Net based approaches proved to be more effective overall. Additional effort is required to incorporate complex models into optimized frameworks such as nnU-Net for a fair evaluation of different architectural proposals.

3. Further research is needed regarding generalisation: it is essential to develop methods that can generalize well across a wide range of pathologies and patient populations.The results highlight the need to integrate a variety of cardiac diseases, centers, scanners, and acquisition protocols to generate robust DL approaches in the biomedical imaging analysis domain.

4. Regarding multi-view methods, it cannot be definitively concluded that they bring a significant improvement to the CMR RV segmentation problem. However,

[2] www.ub.edu/mnms-2

further study is necessary in order to perform a conclusive assessment of their impact and potential.

Ethical Approval

The study was approved by the ethics committee of the three centers involved: Vall d'Hebron Hospital, Sagrada Familia Hospital, and Dexeus Hospital. Written informed consent was obtained from all participants.

Acknowledgment

This work was partly funded by the European Union Horizon 2020 research and innovation programme under grant agreement no. 825903 (euCanSHare project). This work has been partially supported by the Spanish project PID2019-105093GB-I00 and by ICREA under the ICREA Academia programme. K. Lekadir is supported by the Ramon y Cajal Program of the Spanish Ministry of Economy and Competitiveness under grant no. RYC-2015- 17183. A. Guala has received funding from the Spanish Ministry of Science, Innovation and Universities (IJC2018-037349-I) and from "la Caixa" Foundation (LCF/BQ/PR22/11920008).

Chapter 3

Segmentation of complex modalities: Myocardium, Left Ventricle and Rigth Ventricle segmentation in LGE

This chapter was published as:

Víctor M Campello, Carlos Martín-Isla, Cristian Izquierdo, Steffen E Petersen, Miguel A González Ballester, Karim Lekadir. "Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhancement cardiac MRI". In: *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges. STACOM 2019. Lecture Notes in Computer Science, vol 12009. Springer, Cham* (2020).

Doi: 10.1007/978-3-030-39074-7_31.

3.1 Abstract

Accurate segmentation of the cardiac boundaries in late gadolinium enhancement magnetic resonance images (LGE-MRI) is a fundamental step for accurate quantification of scar tissue. However, while there are many solutions for automatic cardiac segmentation of cine images, the presence of scar tissue can make the correct delineation of the myocardium in LGE-MRI challenging even for human experts. As part of the Multi-Sequence Cardiac MR Segmentation Challenge, we propose a solution for LGE-MRI segmentation based on two components. First, a generative adversarial network is trained for the task of modality-to-modality translation between cine and LGE-MRI sequences to obtain extra synthetic images for both modalities. Second, a deep learning model is trained for segmentation with different combinations of original, augmented and synthetic sequences. Our results based on three magnetic resonance sequences (LGE, bSSFP and T2) from 45 different patients show that the multi-sequence model training integrating synthetic images and data augmentation improves in the segmentation over conventional training with real datasets. In conclusion, the accuracy of the segmentation of LGE-MRI images can be improved by using complementary information provided by non-contrast MRI sequences.

3.2 Introduction

Late gadolinium enhancement magnetic resonance imaging (LGE-MRI) is widely used to assess presence, location and extent of regional scar or fibrotic tissue in the myocardium. Whilst LGE-MRI is a well-established technique and key to many cardiovascular magnetic resonance (CMR) examinations there are challenges in quantification and interpretation due to a number of factors. Image analysis depends on image quality which can be affected by suboptimal CMR acquisition. Correct inversion times (TI) need to be identified and then TI require appropriate adjustments to allow good 'nulling' of remote, unaffected myocardium. This ensures optimal contrast between scar/fibrosis (bright) and normal, remote myocardium (dark). Timing after contrast administration is important to allow not only sufficient wash-out of contrast agent (gadolinium chelate) from the remote myocardium but also from the blood pool. Images acquired too early will leave the blood pool bright which makes differentiating subendocardial infarct from blood pool challenging.

In the existing literature, two main families of techniques have been proposed to automatically segment LGE-MRI data. The first one segments directly the LGE-MRI

images by using different techniques such as graph-cuts [4], atlas-based registration [91], or more recently Convolutional Neural Networks (CNNs) [183]. However, these techniques generally lack robustness due to the limited availability of LGE-MRI datasets for training. As a result, the second family of techniques has considered exploiting other cardiac MRI sequences to provide additional signals for guiding more robustly the segmentation process. For instance, some researchers [174, 164] proposed to segment first cine-MRI images and to propagate the obtained contours into the LGE-MRI images through image registration. Similarly but by using additional sequences, the authors in [192] implemented an atlas-based segmentation approach combining information from balanced-Steady State Free Processing (bSSFP), LGE and T2 sequences. However, these techniques are highly dependent on the image registration step, which is challenging due to the inherent differences between the cardiac MRI sequences.

In addition, in order to improve segmentation and increase the model robustness over unseen data, image synthesis has been proposed recently. The most common model combines generative adversarial networks (GANs) with a cycle-consistency constrain for image-to-image translation and two segmentation networks, one for each image domain, trained end-to-end in order to benefit from a combined loss function. This model has been applied for cross-modality segmentation improvement [188, 31], domain adaptation across scanners [31] or across modalities [42] and segmentation of an unlabeled target modality using only the source ground truth [71, 187]. Alternatively, a GAN can be trained to generate synthetic images from masks according to some conditional value, like the dataset style, as in the case of retinal fundus images for vessel segmentation [189].

In this paper, we propose an approach to circumvent the need for image registration, while addressing the lack of LGE-MRI images for training. Concretely, we implement a CNN-based approach that is capable of learning key properties of the cardiac structures simultaneously from multiple cardiac MRI sequences. Furthermore, image synthesis and data augmentation are used to generate new examples that take into account both the global appearance of LGE-MRI data and the local appearance of scar tissues. With this approach, direct deep learning based segmentation of LGE-MRI is enabled without the need for inter-sequence image registration and while exploiting the richness of multi-sequence cardiac MRI.

Table 3.1: MS-CMRSeg sequences details.

	bSSFP	LGE	T2
Number of patients	45	45	45
Segmented patients	35	5	35
Number of slices	8 – 12	10 – 18	3 – 7
Slice thickness (mm)	8 – 13	5	12 – 20
TR/TE (ms)	2.7/1.4	3.6/1.8	2000/90
In-plane resolution (mm)	1.25×1.25	0.75×0.75	1.35×1.35

3.3 Method

3.3.1 Dataset

Data description

The LGE-MRI dataset used in this paper was provided as part of the Multi-Sequence Cardiac Magnetic Resonance Segmentation Challenge (MS-CMRSeg). It consists of 45 patients from Shanghai Renji Hospital that were scanned using three MRI sequences: bSSFP, LGE and T2. Ground truth segmentations of the left ventricle (LV), right ventricle (RV) and myocardium (MYO) were provided for some of the cases according to the distribution in Table 4.1 (second row). Even though all sequences were acquired and selected for the end-diastolic cardiac phase, there were differences in the shape of the cardiac boundaries consistently between the three sequences for the same patient. Moreover, the slices were not aligned between the sequences in the direction of the ventricular axis, which further complicates the application of image registration. Note that all patients in the sample suffer from cardiomyopathies and that every LGE-MRI image presents a scar of variable size within the myocardial wall.

Data pre-processing

As a first step, intensity bias correction was applied to all sequences to correct for potential artifacts and the intensity histograms of all images were matched to a common one to obtain coherent appearances across images. Furthermore, before the training process, all images were interpolated and cropped so that they had a pixel size of 256×256 and the same resolution. They were also normalised such that the mean intensity and the standard deviation equal 0.5, thus ensuring most of the input values to be positive in between 0 and 1 for convenience in later representation of the images.

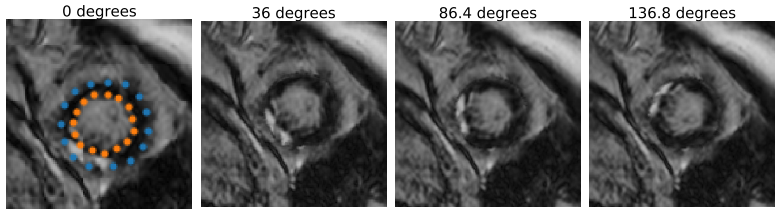


Figure 3.1: Example of three rotations of the myocardial wall with respect to the whole image by using the landmarks provided in the leftmost image. This shows the changes in the location of the scar tissues

3.3.2 Increasing training sample

Before describing the CNN model implemented in this paper for LGE-MRI segmentation, this section presents two methods used to increase the number of training data and obtain higher LGE-MRI variability.

Data augmentation

By using the provided segmentations, a set of 50 landmarks were evenly placed around the epicardium and endocardium. With these, the myocardium and left ventricle were rotated relative to the rest of the image, as shown in the examples in Figure 3.1, in order to obtain an augmented dataset with varying locations of the scar tissues. Since the contour of the epicardium is not perfectly round in general, a Gaussian filter of size 3×3 was applied around the outer boundary to smooth the transition between the rotated and fixed regions, thus preventing image intensity discontinuities. A total of twenty 7.2 degrees rotations were applied. Thus, the LGE-MRI dataset was multiplied by a factor of 20 and the location of the scar in the myocardium ranged between the initial position and 144 degrees clockwise. This augmentation technique increases the variability in the scar locations within the myocardial wall that was otherwise very low due to the small number of patients available for training. Furthermore, further data augmentations were obtained by applying small rotations of the input images up to 15 degrees before training.

Image synthesis

The rationale behind the proposed image synthesis is that there are many more segmented cine-MRI datasets available open-access or in clinical registries for training

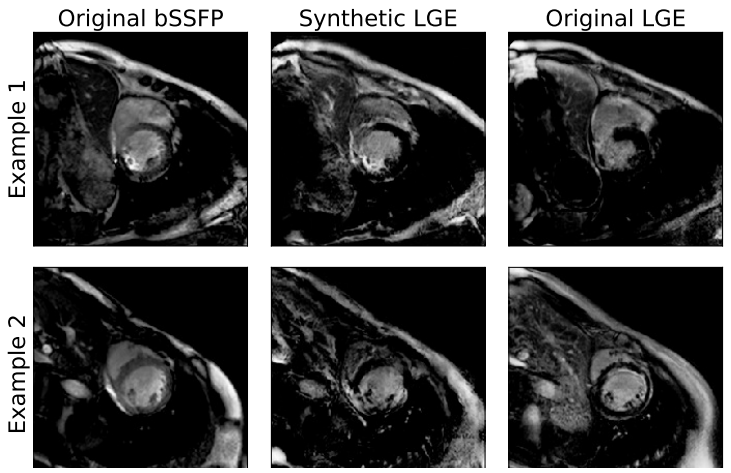


Figure 3.2: Examples of synthetic LGE-MRI images. The leftmost column are the original cine images, the central column shows the transformed images to the LGE domain and the rightmost column is the most similar slice from the real LGE sequences, since they were not registered/aligned.

CNN models. Thus, to increase the number of annotated LGE-MRI cases for training, image synthesis from cine-MRI images sequences is proposed. To achieve this, the CycleGAN method was implemented using the PyTorch library provided at this link[2].

This method translates images from one domain to another without the need for image registration or for the sequences to be from the same patients. It consists of a pair of generators G_{LGE}, G_{bSSFP} and a pair of discriminators D_{LGE}, D_{bSSFP} that have opposed goals. The generator G_{LGE} (G_{bSSFP}) transforms the bSSFP (resp. LGE) sequence into a realistic LGE (bSSFP) image, while the discriminator D_{LGE} (D_{bSSFP}) attempts to distinguish between real and fake LGE (bSSFP) sequences. To achieve a good image translation between the two sequences, the loss function contains two terms: (1) an adversarial loss for each target domain that accounts for the similarity between the generated and real images, and (2) a cycle consistency loss that ensures that the transformed image $G_{LGE}(X)$ ($G_{bSSFP}(Y)$) is transformed back to X (Y) through G_{bSSFP} (G_{LGE}).

For the training of the CycleGAN model, all slices from the 45 patients for the LGE and bSSFP sequences were used during 200 epochs. The training took 12 hours on a NVIDIA 1080 GPU with a batch size of 1. The Adam optimizer was used with learning rate of 2×10^{-4}, with first and second moment decay rates of 0.5 and 0.999,

[2]https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

Table 3.2: Average and standard deviation for the Dice score of segmentation results over the five labeled LGE volumes.

	LV		MYO		RV	
	avg.	std.	avg.	std.	avg.	std.
model trained w. bSSFP	0.503	0.406	0.370	0.301	0.515	0.434
model trained w. synthetic LGE	0.809	0.116	0.688	0.145	0.820	0.065

respectively. Some examples for the generated images are shown in Figure 3.2.

In order to evaluate the quality of the generated images, two segmentation models (like the one described in the next subsection) were trained using the bSSFP images and the synthetic LGE images separately. The obtained results are presented in Table 3.2. In particular, the synthetic LGE images, that are anatomically similar to the original bSSFP, provide more information for the task of LGE segmentation.

3.3.3 CNN-based LGE segmentation

Once a large set of training sample was obtained from the original, augmented and synthetic images, a modified U-Net architecture [145] was used for the image segmentation by integrating two techniques: (1) a deep supervision term in the upsampling path as proposed in [73] that will act as lower-resolution masks that are convolved to condition the final predictions; and (2) a reduction of the number of filters after each upsampling operation to match the number of labels as proposed by [16]. Each image in the dataset was provided as a single channel input, thus forcing the model to differentiate between sequences with a unique set of weights. Additionally, in order to avoid overfitting given the sample size, dropout was used after every max pooling and upsampling operations, except for the high level features in the architecture, as shown in Figure 3.3.

During training, 20% of the patients for each dataset was reserved for validation and early stopping. With a batch size of 8 images, this model took less than 36 hours to achieve the best accuracy on the validation set after almost 90 epochs on a NVIDIA TITAN X GPU. The Adam optimizer was used with a learning rate of 10^{-4}, with first and second moment decay rates equal to 0.9 and 0.99, respectively.

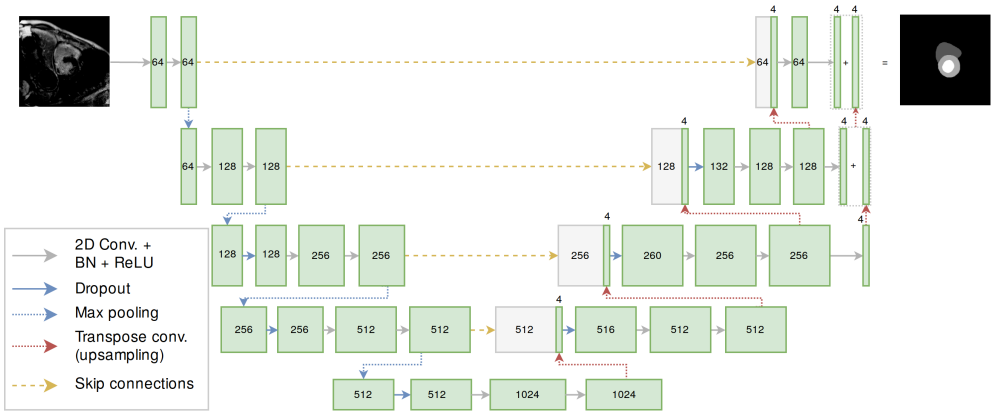


Figure 3.3: Detailed architecture of the CNN model used for LGE segmentation. The numbers in the boxes correspond to the number of channels. Convolution operations have a kernel size of 3×3 and stride of 1, while transpose convolutions have a kernel size of 4×4 and stride of 2.

3.4 Results

In order to define the best trained CNN model for LGE-MRI segmentation, various training sets were used by varying the input sequences and combinations of image synthesis and scar augmentation, as follows:

1. LGE sequences only;
2. LGE and bSSFP sequences;
3. All sequences (LGE, bSSFP and T2);
4. All sequences plus MYO and LV rotations in LGE sequences;
5. Number 1 plus synthetic LGE sequences;
6. Number 2 plus synthetic LGE sequences;
7. Number 3 plus synthetic LGE sequences;
8. Number 4 plus synthetic LGE sequences.

When evaluated on the validation set, the training set number 8 resulted in the best segmentations, showing the added value of image synthesis and data augmentation for LGE-MRI segmentation. Thus, we applied the corresponding CNN model to the test dataset composed of 40 LGE-MRI cases. The obtained segmentations were sent

Table 3.3: Average and standard deviation for results over the test set.

	LV		MYO		RV	
	avg.	std.	avg.	std.	avg.	std.
Dice score	0.898	0.045	0.810	0.061	0.866	0.051
Jaccard index	0.817	0.072	0.685	0.084	0.768	0.078
Surface distance (mm)	2.0	0.8	1.8	0.5	2.3	0.9
Hausdorff distance (mm)	11	4	12	4	16	7

to the organizers of MS-CMRSeg Challenge for evaluation. The obtained results are summarized in Table 3.3, showing average dice scores of 90% (LV), 87% (MYO) and 81% (RV).

Two remarks are important to note regarding the results reported in Table 3.3: (1) Despite the high variability in the LGE-MRI datasets, especially in the presence, extent and location of the scar tissues, relatively consistent results are obtained with standard deviations for the dice scores around 5%. (2) Despite the availability of only 5 LGE-MRI volumes for training, the proposed approach was able to achieve comparable results to very recent deep learning techniques, which reported dice scores of 0.915 ± 0.052 (LV), 0.812 ± 0.105 (MYO) and 0.882 ± 0.084 (RV) based on 5 times more training cases (25 LGE-MRI images). [183]. This indicates the value of the proposed inter-sequence synthesis and scar augmentation for generating richer training samples.

Finally, for visual illustration, Figure 3.4 shows three segmentation examples as obtained in this study. Model number 3 (second column) introduces errors that are corrected when adding synthetic images (model number 7 in the third column). The last column shows that the segmentation further improves when integrating the scar tissue augmentation as proposed in this paper (model 8).

3.5 Conclusions

This paper proposed to address the limited availability of training samples for LGE-MRI segmentation by enriching the CNN models using two complimentary methods. Firstly, since samples of annotated cine-MRI sequences are more commonly available, image synthesis of LGE-MRI images was implemented using a CycleGAN approach, thus obtaining a larger number of LGE-MRI cases during training. Secondly, we performed LGE-specific data augmentation through shape-guided rotations of the myocardium, which increases the variability related to the location of the scar tissues in

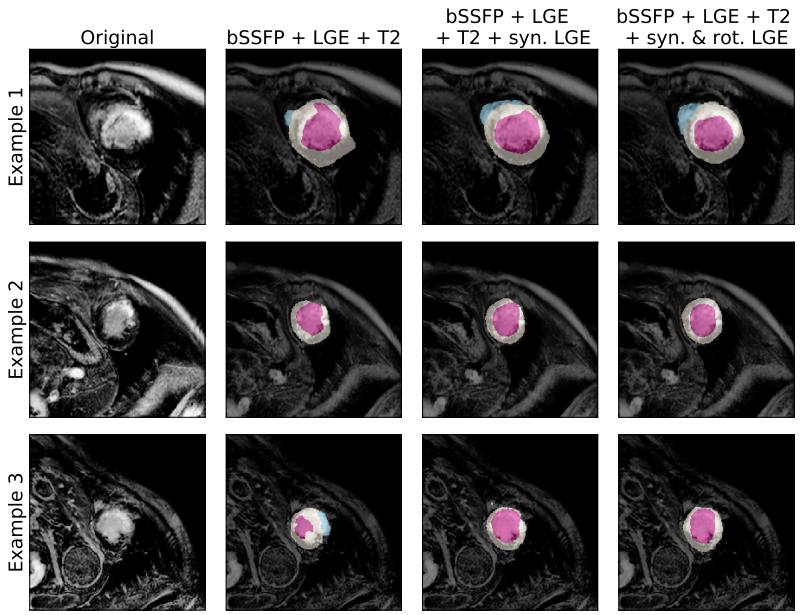


Figure 3.4: Three segmentation examples as obtained by using different training combinations, showing the improvement achieved by integrating inter-sequence image synthesis (column 3) and scar tissue augmentation (column 4) during training.

the myocardium. The validation shows consistent results across the datasets, indicating the potential of this approach for enhancing the richness and generalization of LGE-specific CNNs.

Future work include the extension of the image synthesis to take into account local cardiac motion abnormality for synthesizing scar tissue, as well as the use of elastic deformations of the myocardium and scar to augment non-rigidly the LGE-MRI examples. Furthermore, extensive validation will be performed to assess in detail the relative importance of the different steps and sequences (bSSFP, T2) in enriching the CNN models for LGE segmentation.

3.6 Acknowledgements

This work was partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 825903 (euCanSHare project). SEP acts as a paid consultant to Circle Cardiovascular Imaging Inc., Calgary, Canada and Servier. SEP acknowledges support from the National Institute for Health Research (NIHR) Cardiovascular Biomedical Research Centre at Barts, from the SmartHeart

EPSRC programme grant (EP/P001009/1) and the London Medical Imaging and AI Centre for Value-Based Healthcare. SEP and KL acknowledge support from the CAP-AI programme, London's first AI enabling programme focused on stimulating growth in the capital's AI Sector.

Chapter 4

Segmentation of complex tissues: Scar and edema segmentation in LGE

This chapter was published as:

Carlos Martín-Isla, Maryam Asadi-Aghbolaghi, Polyxeni Gkontra, Victor M Campello, Sergio Escalera, Karim Lekadir. "Stacked BCDU-Net with semantic CMR synthesis: Application to myocardial pathology segmentation challenge". In: *Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images. MyoPS 2020. Lecture Notes in Computer Science, vol 12554. Springer, Cham* (2020).

Doi: 10.1007/978-3-030-65651-5_1.

4.1 Abstract

Accurate segmentation of pathological tissue, such as scar tissue and edema, from cardiac magnetic resonance images (CMR) is fundamental to the assessment of the severity of myocardial infarction and myocardial viability. There are many accurate solutions for automatic segmentation of cardiac structures from CMR. On the contrary, a solution has not as yet been found for the automatic segmentation of myocardial pathological regions due to their challenging nature. As part of the Myocardial Pathology Segmentation combining multi-sequence CMR (MyoPS) challenge, we propose a fully automatic pipeline for segmenting pathological tissue using registered multi-sequence CMR images sequences (LGE, bSSFP and T2). The proposed approach involves a two-staged process. First, in order to reduce task complexity, a two-stacked BCDU-net is proposed to a) detect a small ROI based on accurate myocardium segmentation and b) perform inside-ROI multi-modal pathological region segmentation. Second, in order to regularize the proposed stacked architecture and deal with the under-represented data problem, we propose a synthetic data augmentation pipeline that generates anatomically meaningful samples. The outputs of the proposed stacked BCDU-NET with semantic CMR synthesis are post-processed based on anatomical constrains to refine output segmentation masks. Results from 25 different patients demonstrate that the proposed model improves 1-stage equivalent architectures and benefits from the addition of synthetic anatomically meaningful samples. A final ensemble of 15 trained models show a challenge Dice test score of 0.665 ± 0.143 and 0.698 ± 0.128 for scar and scar+edema, respectively.

4.2 Introduction

Myocardial viability assessment is key in the diagnosis of patients suffering from myocardial infarction and ischemic heart disease, among others. Cardiovascular magnetic resonance (CMR) is a well-established imaging technique that provides anatomical and functional information of the heart. Multiple sequences with different properties can be acquired, registered and combined to obtain a complete viability assessment. Late gadolinium enhancement magnetic resonance imaging (LGE-MRI) is widely used to assess presence, location and extent of regional scar or fibrotic tissue in the myocardium. T2-weighted CMR images are able to identify edema and acute or recent myocardial ischemic injury, and have been employed to distinguish acute coronary syndrome (ACS) from non-ACS as well as acute from chronic myocardial infarction. On the other

hand, balanced - Steady State Free Precession (bSSFP) cine sequence presents clear boundaries for the cardiac anatomical regions, often unclear in the first two modalities due the presence of pathological regions.

LGE and T2-weighted are well-established techniques to many CMR examinations, but there are challenges in their quantification and interpretation due to a variety of factors. First, image analysis depends on image quality which can be affected by CMR acquisition protocol. Suboptimal parameters such as inversion time (TI), repetition time (TR), echo time (TE) need to be correctly identified in order to maximize the difference in intensity curves between pathological and non pathological regions, but also to minimize inter-subject acquisitions variability. Additionally, timing after contrast administration in LGE is important to allow sufficient wash-out of the contrast agent. On top of that, the variability in morphology and texture of infarcted, edemic areas and the combination of both leads to a difficult automation of the process. For this reason, manual and automated techniques with no user interaction for infarct borders detection often results in significant within-patient variability [86, 166, 56, 163].

In order to explore the complementary nature of existing modalities for the purpose of myocardial pathology segmentation, the MyoPS challenge is proposed. It includes a challenging data distribution of 45 multi-modality subjects with the goal of doing an accurate automatic infarcted and edemic regions segmentation.

In this work, we propose a challenge solution based on a stacked BCDU-NET late fusion architecture including localisation and segmentation stages. Additionally, we tackle the insufficent training size by means of state-of-the-art generative adversarial models [63, 173]. To do so, we propose an image synthesis strategy based on Semantic Image Synthesis with Spatially-Adaptive Normalization[130]. The results demonstrate that the proposed model improves 1-stage equivalent architectures and benefits from the addition of synthetic anatomically meaningful samples.

4.3 Materials and methods

4.3.1 Dataset

A set of 45 cases of multi-sequence CMR are collected for the challenge. Each case refers to a patient with three CMR sequences, i.e., LGE, T2 and bSSFP CMR. All clinical data have got institutional ethic approval and have been anonymized. The data released have been pre-processed using the MvMM method [193, 191] to align the three-sequence CMR into a common space and to resample them into the same spatial

resolution.

The provided gold standard labels of interest for the challenge are LV myocardial edema (label 1220) and LV myocardial scars (label 2221). Additional annotations of cardiac structures are provided: left ventricular (LV) blood pool (label 500), right ventricular blood pool (label 600) and LV normal myocardium (label 200). Thus, the evaluation of the test data will be focused on the myocardial pathology segmentation, i.e., scars and edema. The inter-observer variation of manual scar segmentation, in terms of Dice, was 0.5243±0.1578, which gives an insight of the difficulty of the task.

4.3.2 Proposed Method

An overview of the proposed automated segmentation method is presented in Figure 4.1. The approach consists of two stacked segmentation networks. In brief, after preprocessing, we employ a computationally efficient U-Net [147] on the bSSFP CMR to localize the rounded shape of myocardium which includes the LV normal myocardium, LV myocardial edema and scar tissue. Subsequently, the bSSFP, T2-weighted and LGE CMR are cropped using the bounding box of the localized myocardium. Histogram normalization is then applied on the cropped part of imgages. During the second stage, the cropped multi-sequence CMR is passed to a higher capacity model, the BCDU-Net [9], to segment the myocardium scar and edema. The output is finally post-processed based on anatomical constrains to refine output segmentation masks. The individual stages are explained in detail in the following sections.

Preprocessing

Before the training process, all images were cropped so that they had a pixel size of 256 × 256. Furthermore, all images were normalised between 0 and 1 within the Region Of Interest (ROI) for each independent modality.

Localization Network

The pathological tissue is located within LV blood pool and LV normal myocardium. Therefore, we first employ a network to localize the myocardial ROI, i.e. a binary segmentation, using cine-MRI as the input modality. Cine-MRI was chosen over the other modalities for this task because it is the most accurate for myocardial boundary detection due to its clear structure definition and lack of appearance of pathological regions. This task will reduce the search space when dealing with scar and edema segmentation by the stacked network. To do that, the myocardium, edema, and scar

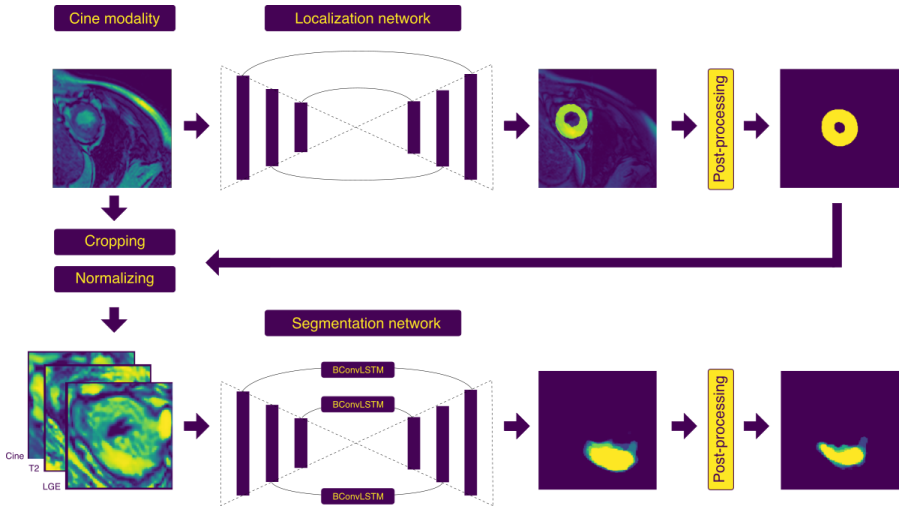


Figure 4.1: Overview of the proposed stacked network.

labels are considered as the foreground, and the other labels (left ventricular blood pool, right ventricular blood pool) as the background. U-Net, [147], is a popular convolutional network architecture for fast and precise segmentation of images which is built upon the Fully Convolutional Network (FCN). The main advantages of this network is that is capable to work well with few training samples, and the network has the potential to make use of the global location and context information at the same time.

This symmetric network is separated in three parts of encoding (contracting), Bottleneck, and decoding (expanding) paths. The encoding path is composed of 4 blocks. In each block we have two 3×3 convolutional layers followed by one 2×2 Max Pooling function and ReLU. In each block, the number of feature maps are doubled, and the size of feature get half. The contracting path aims at progressively capturing context of the input image and increasing the dimension of feature representation block by block. These coarse contextual information are then transferred into the decoding path through skip connections. The output of the last block of the encoder is first passed to the bottleneck which is built by two 3×3 convolutional layers. At the end of bottleneck we have a high dimensional image representation with high semantic information.

The decoding path is composed of four blocks. Each block starts with performing a deconvolution (up-sampling) over the output of previous layer. The corresponding

feature maps in the encoding path are then copied to this layer, and are then concatenated with the output of deconvolutional layer. These features are then go through one 3×3 convolutional layers. In each block of the decoder, the size of the feature maps gradually increases and the number of feature maps gradually decreases. The target of decoder in U-Net is to enable precise localisation by using transposed convolutions and recovering the size of the segmentation. Since that data is imbalanced and most of the pixels have background label, we use the weighted binary cross entropy loss to train the network.

In our U-net implementation, for efficiency purposes, the number of classes is used as the number of feature maps in the deconvolutions of the decoding path, as shown in [18, 36]. It is also worth mentioning that we do not need a very accurate segmentation result here, since we just crop the smallest bounding box around the myocardium with a small margin of 10 pixels.

Normalisation

The output of the localisation network provides the approximate location of the myocardial region. Therefore, by considering the fact that the myocardial infarcted and edemic regions are within such ROI, we can ignore unwanted background information by finding the smallest bounding box with a small margin around the myocardium. Moreover, an histogram equalisation is applied by modality, avoiding the effect of unuseful background pixels in the pixel histogram redistribution.

Segmentation

We exploit the BCDU-Net [9] to segment the myocardial scar and edema from the normalized myocardium of the three input modalities. The BCDU-Net is an extension of U-Net by including bidirectional convolutional LSTM (BConvLSTM) [155] in the skip connection and reusing feature maps with densely convolutions. The output features of the deconvolutional layer contain more semantic information while the features extracted by the corresponding encoding layer have higher resolution. To combine these two kinds of features, the authors replaced the simple concatenation of the skip connection with nonlinear functions, i.e. BConvLSTM in the BCDU-Net which resulted in more precise segmentation output.

Moreover, the idea of densely connected convolutions is utilized in the bottleneck of the BCDU-Net. By having a sequence of convolutional layers, the network may learn redundant features, therefore, in the bottleneck of the BCDU-Net, features which

are learned in each block are passed forward to the next block. The dense blocks help the method to enhance information flow and learn a diverse set of features based on the collective knowledge gained by previous layers. Furthermore, the convergence speed of the network is accelerated by employing Batch Normalization (BN) after the up-convolution filters.

Like U-Net, the encoding path of the BCD-Net includes four steps. Each step consists of two 3×3 convolutional filters followed by a 2×2 max pooling function and ReLU. The depth of feature maps are doubled at each step and the size of each feature map get half. There are two states of BConvLSTM in the skip connection of the BCDU-Net. The second state receives the output of the previous deconvolutional function and the input data of the first one its corresponding feature maps in the encoding path. The output of the second BConvLSTM is then passed to the two 3×3 convolutional filters. Like original U-Net, the decoding path doubles the size of each feature map and halves the number of feature channels layer by layer to reach the original size of the input image after the final layer. To train the network, we use Dice score-based loss.

We propose to combine the three input modalities with a late fusion approach. In other words, the network is trained separately for the three modalities and before the last convolutional layer after the last deconvolutional layer, the three networks are merged.

Implementation Details

All trainings were performed on a NVIDIA 1080 GPU with a batch size of 8. The Adam optimization function with learning rate equal to $1e-4$ was used to train both networks. Each network is trained with 50 as the number of epochs. The input size was 256×256 for both localization and segmentation networks.

4.3.3 Data augmentation strategy

Online augmentation

A series of common augmentation techniques were applied to each batched image independently. For the first stacked u-net, these augmentations included random rotations between -15º and 15º and random scaling and offsets of a maximum of 30 pixels. For the second stacked u-net the offset augmentation is avoided due to the fact that images were already center-cropped.

Offline augmentation

The rationale behind the proposed image synthesis is the insufficient training sample size. Low number of images, variability in modality acquisitions, in location and extent of pathological regions can cause loss of generalisation in CNN-based segmentation algorithms. Thus, in an effort to increase the number of annotated multi-sequence images, semantic image synthesis from annotated mask to multi-sequence CMR is performed in such way that new multi-modality images can be generated from altered versions of real annotations. To achieve this, the Semantic Image Synthesis with Spatially-Adaptive Normalization (SPADE) method [130] was implemented using the PyTorch library provided at this link†. Previous methods [173] directly feed the semantic layout as input to the deep network, which is then processed through stacks of convolution, normalization, and nonlinearity layers. In [130], is shown that this is suboptimal as the normalization layers tend to wash away semantic information, desired for accurate pathology tissue and cardiac structure generation. To address the issue, SPADE uses the input semantic annotation for modulating the activations in normalization layers through a spatially-adaptive, learned transformation. A general overview of the SPADE multi-modality generative model is represented in Figure 4.2.

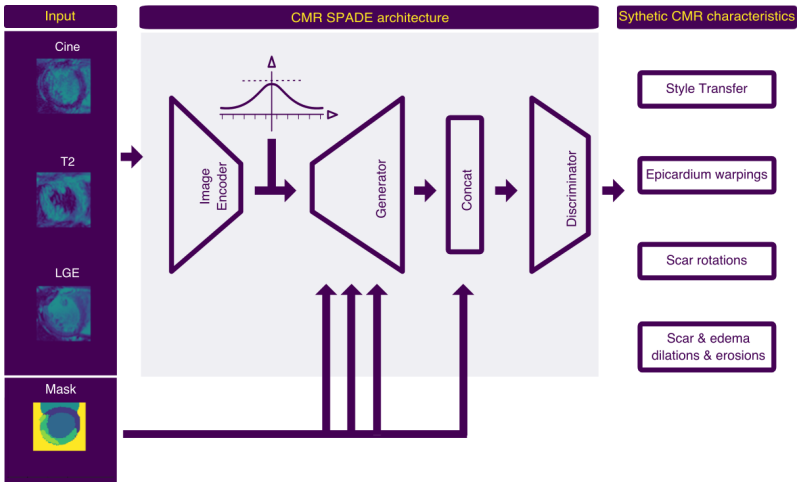


Figure 4.2: Overview of the proposed SPADE generative model.

Two SPADE models were generated. For the training/validation subset, a model with 71 training images (17 subjects) was used and 31 validation images (8 subjects) were kept aside. For the final model, all the subjects were used to train an additional

SPADE model.

Both models were trained during 45 epochs with a morphological augmentation consisting of warping epicardium contours between pairs of subjects. Both trainings took 24 hours on a NVIDIA 1080 GPU with a batch size of 2. The Adam optimizer was used with learning rate of $2x10e-4$, with first and second moment decay rates of 0 and 0.9, respectively. The Variational Autoencoder (VAE) was generated with a latent dimension of 200. Once the models were trained, a set of morphological operations were defined in order to generate different versions of real annotations. The resulting anatomical consistent annotations were used then to feed the SPADE models and generate synthetic multi-modality images with controlled characteristics:

Style transfer. By training the SPADE with a Variational Autoencoder (VAE), the style of the images can be transferred, generating a variety of images with different pathology appearances for the same morphology. The encoder and generator of our SPADE architecture form a VAE, in which the encoder tries to capture the style of the image, while the generator combines the encoded style and the segmentation mask information via the SPADEs to reconstruct the original image. The encoder also serves as a style guidance network at test time to capture the style of target images. For training the VAE, KL-Divergence loss term was used. Every training image was used to generate a set of latent representations of size 200. The latter were used alone -with random linear combinations and scaling factors- or in conjunction with the methods described below in order to produce the final synthetic multi-modality images. The effect of this technique is shown in Figure 4.3, where an original image in first row is transferred to two additional pseudo-random styles, rows 2 and 3.

Epicardium warpings. As shown in Figure 4.4, a set of 8 equidistant landmarks were placed in the epicardial contour of the source and target annotations. Epicardial contours were then warped between pairs of training subjects by means of piecewise affine transformations.

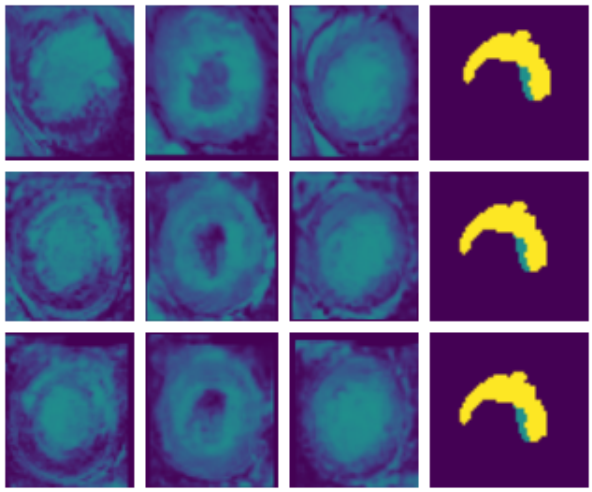


Figure 4.3: Style modifications.

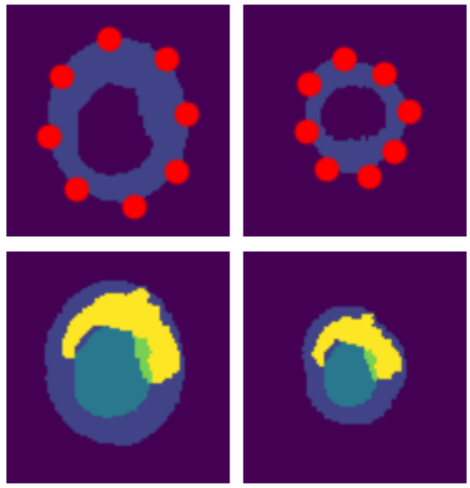


Figure 4.4: Epicardial contour warping between a pair of subjects.

Scar and edema rotations. As shown in Figure 4.7, scar, edema and myocardium labels were combined in a binary mask. The epicardium was then converted to a circular shape, rotated and reconverted to the original shape taking profit of the same

Figure 4.5: Morphological operations involved in the scar rotation process.

Figure 4.6: Morphological operations involved in the scar and edema dilation and erosion process.

technique used in the *Epicardium warpings* section. This set of transformations was then also applied to the original labels, generating a rotated version of the scar and edema within the myocardium. To ensure that the generated segmentations were not too far from the distribution seen by the SPADE generator while covering the label space, the rotation was fixed to four possible values of [-30º, -20º,20º,30º].

Scar and edema dilations and erosions. A set of random complementary dilations and erosions with a random kernel radius from 1 to 3 pixels were applied to the training annotations. By fixing one of them for the scar label and applying the opposite one for the edema label, we avoid an empty gap between both. Random deletion of edemic labels is also included in this stage. In Figure 4.6 shows the effect of an eroded scar and dilated edema.

Offline datasets A group of datasets is generated by means of the augmentation strategies described above. More precisely, for each of the transformable labels, i.e. non-empty annotations, the original images are used up to three times to keep the training size relatively small. This methodology leads to the creation of a set of four

datasets, one per type of augmentation, i.e. style transfer alone, pathology rotations, epicardial warping and pathology dilation/erosion. It should be noted that the resulting datasets contain the same amount of real and synthetic data. Additionally, for all datasets, random style transfers are applied after the annotation manipulation in the synthesis stage. In total, each dataset contains 415 images. A fifth dataset is generated by combining all individual four datasets. This dataset consists of 1660 images and is used to train and validate the models. The same procedure is repeated for the final ensemblea using the SPADE trained over all the training data. This leads to datasets of 597 and 2388 images, for the partial augmentations and the addition, respectively.

4.3.4 Post-processing

The myocardium, scar and edema-scar segmentations produced from the stacked networks were morphologically processed to satisfy certain anatomical constraints. In short axis CMR, the shape of the myocardium closely resembles that of a ring throughout the apex-base slices. Therefore, slices for which the automatically segmented myocardium is a partial ring must be detected and corrected. To this end, the skeleton of the myocardium was calculated for each slice. Subsequently, spur skeleton branches, i.e. branches consisting of pixels with only one neighboring pixel, were iteratively pruned. For non-complete rings, iterative pruning results in the removal of the entire skeleton. In such cases, the missing arc of the partial ring was completed by adding a circular ring whose thickness is equal to the maximum thickness of the detected myocardium. To construct the ring, the centroid of the convex hull of the detected myocardial region was used as its center. The thickness of the myocardium was given by the distance of the skeleton points to the closest non-myocardial pixel and the maximum among all points was considered. The corrected myocardium was subsequently used to refine the scar segmentation, while an additional step was necessary in the case of the edema-scar region. More precisely, edema can be noticed in the myocardium, but also in the LV blood pool close to the border with the myocardium. Therefore, an extended myocardial mask was created, which contained neighboring LV regions where edema could be localized. In order to achieve this, an artificial ring was constructed by using the myocardium skeleton and the distance of every pixel to it. Pixels belonging to the myocardium or the region enclosed by it were considered to belong to the extended myocardial mask if they were within a distance smaller than a threshold from the skeleton points. This threshold is defined as the maximum myocardium thickness plus a small margin of 6 pixels to account for errors in the myocardium segmentation.

As a first step in the process of refining the scar tissue, 3D components smaller than 100 voxels were considered to be artifacts and were, therefore, excluded from the segmentation mask. Despite good localization of the scar region by the network, we observed a tendency to underestimate the scar region and to produce multiple disconnected components instead of one continuous region. To tackle this issue, the components were connected by using their convex hull in cases where the output of the network consisted of more than one connected components. The area of the convex hull inside an eroded version of the extended myocardium was eliminated. For the erosion, a disk element with radius equal to 20% of the maximum myocardium radius was used. Furthermore, morphological closing of the image with a disk object of radius equal to 90% of the myocardium maximum thickness was performed to enlarge the component's border without losing the form of the original shape boundary in cases where only one component was observed. Lastly, areas outside the corrected myocardium and the joined edema-scar mask regions were excluded from the final scar segmentation.

In the case of the refinement of the joined edema-scar mask, 3D components of size smaller than 300 voxels were considered as artifacts. In addition, regions of edema-scar outside the extended myocardial area were excluded from the final segmentation by performing element-wise multiplication of the artificial extended myocardium region mask with the edema-scar segmentation.

4.4 Results

4.4.1 Protocol and Metrics of the challenge

In order to train our models and generate the ablation study, the training set is divided in two partitions. From the original 25 subjects, 8 of them are kept aside for validation, with the aim of preserving a large pool of subjects in the validation stage. The decision is motivated by the variability in image quality and the presence of difficult cases that may lead to a sub-optimal model selection. Moreover, this allows us to have a sufficient validation size to evaluate the post-processing algorithm. For the same reason, we avoided to preserve a test partition that leads to a conflict between validation and testing results and generates additional uncertainty when selecting the best method. After model generation, selection, evaluation and post-processing, 3D Dice scores are computed to select the final models taking into consideration the post-processing gains. For all the experiments, 2D Dice score is used as objective loss function, except for the localisation U-net, where the selected loss is binary weighted cross-entropy.

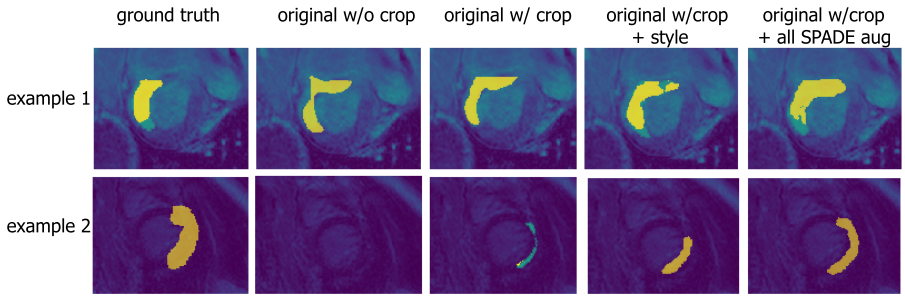


Figure 4.7: Segmentation examples combining different sets of training data, showing the improvement of SPADE synthesis.

4.4.2 Ablation study

We performed a detailed ablation study in order to quantify the effect of every component of the proposed methodology individually. The results in terms of 2D Dice score (mean ± standard deviation), which is the accuracy evaluation metric used in the loss function of this work, are summarized in Table 4.1. In brief, our first experiment involved segmenting the scar and scar+segmentation using solely the original data without performing inter-stage normalization or offline augmentation. This resulted in a Dice score equal to 0.202 ± 0.286 and 0.170 ± 0.253 for scar and scar+edema, respectively. The low accuracy demonstrates the extremely challenging nature of the task and the need for incorporating a ROI-based normalization between stages and novel augmentation strategies. To test our assumption, we added the inter-stage cropping and normalization step to enhance the contrast between scar and edema and the rest of the tissue within the myocardial ROI where the pathological tissue is expected to localized. The mean dice score increased by 24.70% for scar and 33.80% for scar+edema.

We then compared the improvement offered by any of the four types of offline augmentation, i.e. style transfer alone, pathology rotations, epicardial warping and pathology dilation/erosion. Style transfer produced an improvement in terms of Dice by 9% and 14.4% for scar and scar+edema, respectively. The effects of epicardium warping and scar and edema rotation, were lower than that of style-transfer, but yet non-negligable. More precisely, the mean dice increased by 4.1% for scar and 7.8% for scar+edema in the case of epicardium warping. Similarly, when scar and edema rotation

Table 4.1: 2D Dice score (mean ± standard deviation) of the proposed method for scar and scar+edema for different data.

Data	Scar	Scar + Edema
Original data	0.202 ± 0.286	0.170 ± 0.253
Original data + cropping and normalizing	0.449 ± 0.261	0.508 ± 0.243
Style transfer	0.548 ± 0.250	0.640 ± 0.192
Epicardium warping	0.490 ± 0.260	0.586 ± 0.222
Scar and edema rotation	0.466 ± 0.241	0.554 ± 0.224
Scar and edema dilation and erosion	0.458 ± 0.299	0.600 ± 0.224
All spade	0.518 ± 0.286	0.617 ± 0.253

were applied the offered improvement was 1.7% for scar and 4.6% for scar+edema. Interestingly, scar and edema dilation and erosion did not provide any significant improvement in the scar tissue, but offered a 10.4% mean improvement in Dice for the scar+edema region. Subsequently, we combined the four types of data-augmentation. We observed a Dice score of 0.518 ± 0.286 and 0.617 ± 0.253 for scar and scar+edema, respectively. This indicates that for the case of pathological tissue segmentation the most effective augmentation type is style transfer, while morphological augmentations have a more limited effect. We speculate that this might be related to the highly irregular shape of the pathological tissue. However, these types of morphological augmentations might be important in other more regular structures. In this work, to account for possible variability found in the test sample non present in the training set, for the final model, we decided to use the combination of all augmentation types, presented as "All spade" in Table 4.1. Nonetheless, future work will focus on using the style transfer only for pathological tissue segmentation.

Lastly, we evaluated the improvement offered by applying post-processing on the outputs of the localization and segmentation networks. A visual example of the improvement can be seen in Figure 4.8. Post-processing produces a continuous scar region, while both edema and scar after post-processing are localized within the myocardial area and in the close vicinity of left ventricle, as physiologically expected.

4.4.3 Challenge results

In order to obtain the final predictions, two ensembles are generated. For the first ensemble, a set of 5 models is generated with 10 consecutive training samples and 5 consecutive validation samples, with a roll factor of 5. For the second ensemble, a set of 15 models is generated with 22 consecutive training samples and 3 consecutive

Table 4.2: 3D Dice score for the final testing set of 20 subjects.

Data	Scar	Scar + Edema
5 models ensemble	0.625±0.255	0.677±0.146
5 models ensemble + post-processing	0.635±0.281	0.692±0.143
15 models ensemble	0.636±0.243	0.687±0.131
15 models ensemble + post-processing	0.665±0.241	0.698±0.128

validation subjects, with a roll factor of 2, making the validation set to share one subject between consecutive models in the case of the 15 models ensemble.

The confidence maps of each one of the 5 models are averaged together. The final predictions of the 20 unseen test subjects provided by the challenge organization are defined as the maximum average probability of each pixel belonging to each class, maximizing the expected results and reducing the variance. The same procedure was applied to the 15 models ensemble. After that, post-processing, as described in Section 2.4, is applied to further enhance the model's output. The effect of the ensemble size can be observed in Table 4.2. The bigger ensemble obtained better results due to the bigger training sizes. The effect of the low validation size was noticeable as a noisier validation curve, and attenuated by means of a greater regularisation power, with an overall improved accuracy. The quantitative effect of post-processing is also appreciated. The 15 models ensemble captured a greater number of non-trivial unconnected components. In combination with the convex hull process described in Section 2.4, for the 15 models ensemble the post-processing generated an improvement in accuracy of 2.9% for scar and 1.1% for scar+edema, respectively.

4.5 Discussion

This work proposes a novel approach to address automatic multi-sequence CMR pathology segmentation. The method is based on a two-staged process and leverages advanced state-of-the-art deep learning techniques. CMR pathology segmentation is a particularly challenging task even for the expert clinician due to the large variability in imaging quality and morphology of pathological regions. To tackle this limitation, we focus on reducing the task complexity. To this end, a localisation U-net is used to localize the myocardial ROI. Subsequently, the detected ROI is used to partially address the problem of intra- and inter-subject variability in signal intensity by using the bounding box of the ROI to crop the CMR images and perform a refined normalisation within the cropped region. The normalised CMR are then fed to a BCDU-net in

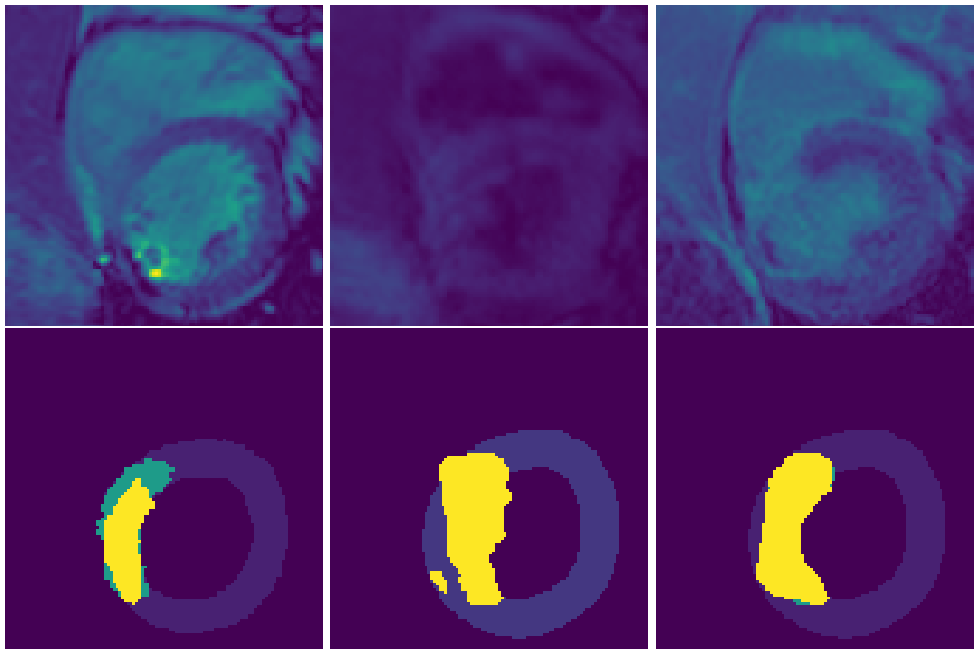


Figure 4.8: Improvement offered by applying post-processing on the outputs of the localization and segmentation networks. On the top row, a slice from the bSSFP (left), T2-weighted (middle) and LGE (right) CMR are provided for one subject of the training dataset used as validation subject during training. On the bottom row, the corresponding manual segmentations for myocardium, scar and edema (left), the combined output of the two networks before (middle) and after (right) post-processing are provided. Post-processing permits to connect the two disconnected components produced by the network and constrain the segmentation within the myocardial area and neighboring LV area.

order to perform the pathologic tissue segmentation. BCDU-net effectiveness has been previously demonstrated and is related to the bidirectional flow of the gradient. In addition, we address the problem of insufficient training examples by means of multi-modality semantic image synthesis using morphological and style transformations. This approach increases the variability of the training samples in terms of the location of the infarcted and edemic tissues within the myocardium, as well as, in terms of their appearance. The validation shows the effect of the stacked architecture with inter-stage normalisation, giving an insight about the importance of standarisation for multi-modality medical imaging acquisitions. Moreover, consistent results across the different semantic manipulations and their respective synthesis, indicate the potential of this set of transformations for enriching and improving generalization of multi-modality cardiac pathology segmentation algorithms. Future work includes the implementation of an end-to-end model as well as the exploration of the generated synthetic data in detail with the aim of enhancing interpretability and quality of the image synthesis methods.

Acknowledgement.

This work was partly funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 825903 (euCanSHare project). This work has been partially supported by the Spanish project PID2019-105093GB-I00 (MINECO/FEDER, UE) and CERCA Programme/Generalitat de Catalunya.). This work is partially supported by ICREA under the ICREA Academia programme. KL is supported by the Ramon y Cajal Program of the the Spanish Ministry of Economy and Competitiveness under grant no. RYC-2015-17183.

Chapter 5

Conclusions

5.1 Summary

Based on the research questions presented in this thesis, the following conclusions can be drawn:

Chapter 1 introduced a high-level summary of the topic of study in this thesis: the necessity of research on cardiac MRI segmentation is identified and developed. A systematic review of literature on image-based automatic classification methods in cardiology is also presented, highlighting the need for accurate cardiac segmentation methods as a starting point for most of the automatic and human cardiac diagnosis procedures.

This chapter was partially based on a published research by the author in *Frontiers in Cardiovascular Medicine* in January 2020 (Impact Factor 5.846, Q1).

Chapter 2 identified the necessity of in-depth assessment for the right ventricular automatic segmentation task on Cine Steady State Free Precision cardiac magnetic resonance imaging modality. The collection and curation of a large group of multi-view, multi-centre studies within the scope of a balanced set of pathological subjects allowed to assess a wide range of aspects related to the proposed task.

Firstly, the spin-off nature of the organized challenge revealed that the creation of this diverse dataset, carefully oriented to the diversification of right ventricular pathologies, obtained an improvement of 4.2% segmentation accuracy in terms of Dice Score when compared to the last edition of the competition.

Secondly, the statistical analysis was degraded in pathological groups not present in the training set. Interestingly, the best-performing methods suffered from this performance drop, particularly in the end-diastolic phase.

Both results highlight the need to integrate a variety of cardiac diseases, centers, scanners, and acquisition protocols to generate robust Deep Learning approaches with greater generalization capacity in the presence of complex morphologies.

As for the multi-view component of the study, taking into consideration that the manual annotation procedure was guided by complementary views, some degree of improvement was expected when incorporating the complementary information in the automatic counterpart.

Due to the organic development of the competition, there was no explicit position on the use of multi-view approaches by the organizing body. This fact divided the participating methods into two subgroups:

- Best performing methods focused on optimizing their algorithms through the well-known nn-Unet framework, thus neglecting, in general, the use of multi-view information in a tight schedule such as that of competitions.

- By focusing their methodology on the multi-view component of the collected data, the participating methods that opted for this type of strategy did not have the necessary time to integrate them into the nn-Unet framework, with the subsequent loss of accuracy related to non-exhaustive fine-tuning.

This fact highlights the need for a structured ablation study to extract a final conclusion in terms of the validity of multi-view approaches.

This work accomplished the quality standards necessary to be published in *Journal of Biomedical and Health Informatics* in July 2023 (Impact Factor 7.021, Q1).

Chapter 3 introduced image synthesis and specific Late Gadolinium Enhancement augmentations, scar tissue rotations, in response to the second research question and the premises exposed in Chapter 1. An incremental ablation study was performed for the task of Late Gadolinium Enhancement bi-ventricular and myocardial automatic delineation. Taking into account one of the most relevant factors in the field of automated biomedical analysis, data scarcity, a set of 45 studies integrating three domains (bSSFP, LGE and T2) was used to train a segmentation network and test it in the second domain, where the scarcity of annotated images was even more accentuated: only 5 annotated studies were available in this modality.

During the development of this investigation it was notorious how the main particularity of the LGE modality (injection of contrast agent to accentuate myocardial scar tissue) was also the main source of error when transferring the information of the annotated domains (bSSFP, T2) to the target domain (LGE).

Two data augmentation strategies, one classic (scar rotations) and the other based on Unpaired Image-to-Image Translation (CycleGAN), were implemented for two different purposes.

The first increased the number of images annotated in the target domain, providing variability at the scar tissue location level, while the second increased the number of images annotated in the target domain, providing variability at the subject level, showing that a problem-specific data augmentation can be particularly relevant in some scenarios. Both strategies were favorable, concluding that the proposed image synthesis pipeline is a useful tool to regularize segmentation neural networks in the absence of a sufficient number of training examples for complex cardiac imaging modalities such as LGE.

This work was published in the conference proceedings of the *Statistical Atlases and Computational Models of the Heart, held in Conjunction with Medical Image Computing and Computer Assisted Interventions* in January 2020. (QUALIS A1). This work also obtained the third position in the *Multi-sequence Cardiac MR Segmentation Challenge (MS-CMRseg)* competition.

Chapter 4. The previous work based on image synthesis was an effective technique to mitigate the effect of data scarcity and pathological tissue location variability in LGE cardiac studies. In response to the third research question and the knowledge developed in Chapter 3, and with the intention of deepening and expanding the pool of image synthesis protocols, proven beneficial for increasing the generalization capacity of cardiac segmentation neural networks, this chapter shows how multi-modal Semantic Image Synthesis with Spatially-Adaptive Normalization, in conjunction with multi-modal fusion strategies enhances the performance of such networks. A multi-modal segmentation network was trained with a large set of augmented images for the task of myocardial scar tissue and edema segmentation, being such myocardial pathological tissues present in LGE and T2 modalities, respectively. Different multi-modal function strategies such as multi-stream early and late fusion strategies were implemented and compared.Complementing these techniques, our method implemented a generative network responsible of the generation of a variety of segmentation-conditioned multi-modal images: ground truth segmentations were modified manipulating the morphology

of the myocardial ring and the extent and location of the pathological regions. The synthetic images, along with the manipulated segmentations, were then used to train a second neural network, leading to a stacked architecture of multi-modal synthesis plus multi-modal segmentation.

All the proposed generative techniques, including style transfer, contour warpings and scar and edema manipulations, offered an improvement over regular augmentation techniques in terms of Dice Score when used to train the second network, thus conditioning the synthesis to the particularities that involve complex tissue segmentation tasks.

This work was published in the conference proceedings of the *Statistical Atlases and Computational Models of the Heart, held in Conjunction with Medical Image Computing and Computer Assisted Interventions* in December 2020. (QUALIS A1). This work also obtained the Best Paper Award mention in the *Myocardial Pathology Segmentation Challenge (MyoPS)* competition.

5.2 Limitations

In this section, the key limitations of the presented research are outlined, providing a perspective on the inherent challenges and constraints. Understanding these limitations is essential for guiding future research directions.

Data Variability: The research heavily relies on the quality and diversity of available data. Limitations may arise from the variability in the datasets used, such as differences in image resolution, patient demographics, and acquisition protocols. Chapter 2 highlights this phenomenon: a well-curated diverse dataset improved the State-of-the-Art segmentation accuracies for the RV segmentation task. On the other hand, the results presented in Chapters 3 and 4 rely on less diverse datasets. Therefore, it is necessary to collect more data to validate the methots proposed in these chapters.

Ground Truth Accuracy: The accuracy of manual annotations or ground truth used for training and evaluation can introduce potential limitations. Human annotators may have variability in defining complex boundaries or regions of interest, affecting the model's performance.

Computational Resources: Deep learning models, especially when dealing with large medical image datasets, such as multi-modal and multi-view studies, can be com-

putationally intensive. The work's applicability in resource-constrained environments or low-resource settings may be limited.

Clinical Validation and Integration: While the models offer promising segmentation results, their clinical utility and impact on patient outcomes require rigorous validation through external clinical studies, which are not covered in the thesis. Additionally, integrating the developed algorithms into existing clinical workflows and electronic health record systems may present technical and practical challenges that need to be addressed.

5.3 Future perspectives

5.3.1 Validation

Immediate future work related to the content of this thesis is directly incremental with respect to Section 5.2. Validation of the proposed techniques in increasingly large datasets or complex multi-center setups is necessary to verify the benefits of the proposed techniques in real-world scenarios.

On the other hand, the multi-modal nature of clinical assessment on a daily basis suggests that the integration of different imaging modalities is crucial for the execution of differential diagnoses and personalized medicine. In this context, the proposed algorithms will be validated in other modalities, such as CMR perfusion.

5.3.2 Efficient data collection

As discussed, large and diverse datasets are essential to develop robust AI algorithms capable of accurately interpreting MRI scans. Efficient data collection is a key point for accomplishing such a task. The in-house M&Ms-2 multi-center dataset prepared and released to the community, used as a backbone to develop the study presented in Chapter 2 was locally collected in the involved centers, which created a bottleneck and, therefore, a significant time expenditure for the study's development.

Technologies such as XNAT [110], which stands for "Extensible Neuroimaging Archive Toolkit", could significantly speed up the collection and curation process. XNAT is an open-source software platform designed for the management and sharing of medical imaging and related data. The implementation of this framework in conjunction with clinical partners will accelerate the process of data acquisition and curation in future studies.

5.3.3 Trustworthy AI

The successful integration of AI into clinical practice hinges on adhering to guidelines and fundamental principles. The FUTURE-AI initiative, as outlined in [97], assesses these aspects through six key building blocks: Fairness, Universality, Traceability, Usability, Robustness, and Explainability. In the context of this thesis, we emphasize the significance of certain building blocks and introduce additional crucial aspects for establishing trustworthy AI.

Fairness: AI models can be influenced by specific patterns within training data, such as variations in scanner manufacturers, skin tones, gender, ethnicity, and other factors. Identifying such bias is necessary for a better understanding of the AI models, and developing fair AI solutions that perform equitably across different subgroups is essential to include them in the clinical routine.

Usability: This foundation within the FUTURE-AI framework encompasses several concepts, among which we can highlight the concept of *Efficiency*. Throughout the methods proposed in this thesis, notably dense algorithms with different subnetworks have been demonstrated. Computational refinement of the proposed pipelines is necessary to i) efficiently integrate them into clinical practice and ii) accelerate the research process surrounding these methods.

Robustness: AI models must handle various sources of variation in medical images, such as scanner manufacturer, clinical center differences, or inter-observer annotation variability. These variations should be analyzed and, where possible, accounted for using generalizable methodologies.

Interpretability: AI tools are often seen as "black boxes" due to their opaque decision-making processes. However, clinical decisions require reasoning and understanding of potential model failures. While segmentation is one of the most understandable topics in medical AI, interpretation mechanisms for decision-making in segmentation neural networks would benefit processes such as quality control or final human decision-making in complex cases.

Bibliography

[1] N Agani, SAR Abu-Bakar, and SH Sheikh Salleh. Application of texture analysis in echocardiography images for myocardial infarction tissue. *Jurnal Teknologi*, 46(1):61–76, 2007.

[2] Yasmina Al Khalil, Sina Amirrajab, Cristian Lorenz, Jürgen Weese, Josien Pluim, and Marcel Breeuwer. On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images. *Medical Image Analysis*, 84:102688, 2023.

[3] Yasmina Al Khalil, Sina Amirrajab, Josien Pluim, and Marcel Breeuwer. Late fusion u-net with gan-based augmentation for generalizable cardiac mri segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 360–373. Springer, 2021.

[4] Xenia Alba, Rosa M Figueras i Ventura, Karim Lekadir, Catalina Tobon-Gomez, Corné Hoogendoorn, and Alejandro F Frangi. Automatic cardiac LV segmentation in MRI using modified graph cuts with smoothness and interslice constraints. *Magnetic resonance in medicine*, 72(6):1775–1784, 2014.

[5] Myriam Amsallem, Olaf Mercier, Yukari Kobayashi, Kegan Moneghetti, and Francois Haddad. Forgotten no more: a focused update on the right ventricle in cardiovascular disease. *JACC: Heart Failure*, 6(11):891–903, 2018.

[6] Tewodros Weldebirhan Arega, François Legrand, Stéphanie Bricq, and Fabrice Meriaudeau. Using mri-specific data augmentation to enhance the segmentation of right ventricle in multi-disease, multi-center and multi-view cardiac mri. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 250–258. Springer, 2021.

[7] Reza Arsanjani, Yuan Xu, Damini Dey, Vishal Vahistha, Aryeh Shalev, Rine Nakanishi, Sean Hayes, Mathews Fish, Daniel Berman, Guido Germano, et al. Improved accuracy of myocardial perfusion spect for detection of coronary artery disease by machine learning in a large population. *Journal of Nuclear Cardiology*, 20(4):553–562, 2013.

[8] Michael R Avendi, Arash Kheradvar, and Hamid Jafarkhani. Automatic segmentation of the right ventricle from cardiac mri using a learning-based approach. *Magnetic resonance in medicine*, 78(6):2439–2448, 2017.

[9] Reza Azad, Maryam Asadi-Aghbolaghi, Mahmood Fathy, and Sergio Escalera. Bi-directional ConvLSTM u-net with densley connected convolutions. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, October 2019.

[10] Bettina Baeßler, Christian Luecke, Julia Lurz, Karin Klingel, Arijit Das, Maximilian von Roeder, Suzanne de Waha-Thiele, Christian Besler, Karl-Philipp Rommel, David Maintz, et al. Cardiac mri and texture analysis of myocardial t1 and t2 maps in myocarditis with acute versus chronic symptoms of heart failure. *Radiology*, 292(3):608–617, 2019.

[11] Bettina Baeßler, Manoj Mannil, David Maintz, Hatem Alkadhi, and Robert Manka. Texture analysis and machine learning of non-contrast t1-weighted mr images in patients with hypertrophic cardiomyopathy—preliminary results. *European journal of radiology*, 102:61–67, 2018.

[12] Bettina Baeßler, Manoj Mannil, Sabrina Oebel, David Maintz, Hatem Alkadhi, and Robert Manka. Subacute and chronic left ventricular myocardial scar: accuracy of texture analysis on nonenhanced cine mr images. *Radiology*, 286(1):103–112, 2017.

[13] Hassan Bagher-Ebadian, Hamid Soltanian-Zadeh, Saeed Setayeshi, and Stephen T Smith. Neural network and fuzzy clustering approach for automatic diagnosis of coronary artery disease in nuclear medicine. *IEEE Transactions on Nuclear Science*, 51(1):184–192, 2004.

[14] Wenjia Bai, M. Sinclair, G. Tarroni, O. Oktay, Martin Rajchl, G. Vaillant, A. Lee, Nay Aung, E. Lukaschuk, Mihir M. Sanghvi, F. Zemrak, K. Fung, J. Paiva, Valentina Carapella, Y. Kim, Hideaki Suzuki, Bernhard Kainz, P. Matthews, S. Petersen, S. Piechnik, S. Neubauer, Ben Glocker, and D. Rueckert. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20, 2018.

[15] CR Bardeen. Determination of the size of the heart by means of the x-rays. *American Journal of Anatomy*, 1918. Query date: 2023-06-05.

[16] Christian F Baumgartner, Lisa M Koch, Marc Pollefeys, and Ender Konukoglu. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 111–119. Springer, 2017.

[17] Christian F Baumgartner, Lisa M Koch, Marc Pollefeys, and Ender Konukoglu. An exploration of 2d and 3d deep learning techniques for cardiac mr image segmentation. In *Statistical Atlases and Computational Models of the Heart.*

ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8, pages 111–119. Springer, 2018.

[18] Christian F. Baumgartner, Lisa M. Koch, Marc Pollefeys, and Ender Konukoglu. An exploration of 2d and 3d deep learning techniques for cardiac MR image segmentation. In *Lecture Notes in Computer Science*, pages 111–119. Springer International Publishing, 2018.

[19] Marcel Beetz, Jorge Corral Acero, and Vicente Grau. A multi-view crossover attention u-net cascade with fourier domain adaptation for multi-domain cardiac mri segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 323–334. Springer, 2021.

[20] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011*, 2018.

[21] Göksu Bozdereli Berikol, Oktay Yildiz, and İ Türkay Özcan. Diagnosis of acute coronary syndrome with a support vector machine. *Journal of medical systems*, 40(4):84, 2016.

[22] O. Bernard, A. Lalande, Clément Zotti, F. Cervenansky, X. Yang, P. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. González Ballester, Gerard Sanroma, S. Napel, S. Petersen, Georgios Tziritas, Elias Grinias, Mahendra Khened, V. Kollerathu, G. Krishnamurthi, M. Rohé, X. Pennec, Maxime Sermesant, Fabian Isensee, P. Jäger, Klaus Maier-Hein, Peter M. Full, I. Wolf, S. Engelhardt, Christian F. Baumgartner, L. Koch, J. Wolterink, I. Išgum, Yeonggul Jang, Y. Hong, Jay Patravali, S. Jain, O. Humbert, and Pierre-Marc Jodoin. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37:2514–2525, 2018.

[23] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018.

[24] Julian Betancur, Lien-Hsin Hu, Frederic Commandeur, Tali Sharir, Andrew J Einstein, Mathews B Fish, Terrence D Ruddy, Philipp A Kaufmann, Albert J Sinusas, Edward J Miller, et al. Deep learning analysis of upright-supine high-efficiency spect myocardial perfusion imaging for prediction of obstructive coronary artery disease: A multicenter study. *Journal of Nuclear Medicine*, 60(5):664–670, 2019.

[25] Carlo Biffi, Ozan Oktay, Giacomo Tarroni, Wenjia Bai, Antonio De Marvao, Georgia Doumou, Martin Rajchl, Reem Bedair, Sanjay Prasad, Stuart Cook, et al. Learning interpretable anatomical features through deep generative models:

Application to cardiac remodeling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 464–471. Springer, 2018.

[26] Sneha Borkar and MN Annadate. Supervised machine learning algorithm for detection of cardiac disorders. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pages 1–4. IEEE, 2018.

[27] Gregory Borodin and Olga Senyukova. Right ventricle segmentation in cardiac mr images using u-net with partly dilated convolution. In *International Conference on Artificial Neural Networks*, pages 179–185. Springer, 2018.

[28] Lena Bosch, Carolyn S.P. Lam, Lingli Gong, Siew Pang Chan, David Sim, Daniel Yeo, Fazlur Jaufeerally, Kui Toh Gerard Leong, Hean Yee Ong, Tze Pin Ng, Arthur Mark Richards, Fatih Arslan, and Lieng H. Ling. Right ventricular dysfunction in left-sided heart failure with preserved versus reduced ejection fraction. *European Journal of Heart Failure*, 19(12):1664–1671, jun 2017.

[29] Jan L Bruse, Maria A Zuluaga, Abbas Khushnood, Kristin McLeod, Hopewell N Ntsinjana, Tain-Yen Hsia, Maxime Sermesant, Xavier Pennec, Andrew M Taylor, and Silvia Schievano. Detecting clinically meaningful shape clusters in medical image data: metrics analysis for hierarchical clustering applied to healthy and pathological aortic arches. *IEEE Transactions on Biomedical Engineering*, 64(10):2373–2383, 2017.

[30] Malcolm I Burgess, Nesrin Mogulkoc, Rowland J Bright-Thomas, Paul Bishop, Jim J Egan, and Simon G Ray. Comparison of echocardiographic markers of right ventricular function in determining prognosis in chronic pulmonary disease. *Journal of the American Society of Echocardiography*, 15(6):633–639, 2002.

[31] Jinzheng Cai, Zizhao Zhang, Lei Cui, Yefeng Zheng, and Lin Yang. Towards cross-modal organ translation and segmentation: A cycle-and shape-consistent generative adversarial network. *Medical image analysis*, 52:174–184, 2019.

[32] Víctor M Campello, Polyxeni Gkontra, Cristian Izquierdo, Carlos Martín-Isla, Alireza Sojoudi, Peter M Full, Klaus Maier-Hein, Yao Zhang, Zhiqiang He, Jun Ma, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge. *IEEE Transactions on Medical Imaging*, 40(12):3543–3554, 2021.

[33] Victor M Campello, Polyxeni Gkontra, Cristian Izquierdo, Carlos Martin-Isla, Alireza Sojoudi, Peter M Full, Klaus Maier-Hein, Yao Zhang, Zhiqiang He, Jun Ma, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. *IEEE Transactions on Medical Imaging*, 40(12):3543–3554, 2021.

[34] Víctor M Campello, Carlos Martín-Isla, Cristian Izquierdo, Andrea Guala, José F Rodríguez Palomares, David Viladés, Martín L Descalzo, Mahir Karakas, Ersin

Çavuş, Zahra Raisi-Estabragh, et al. Minimising multi-centre radiomics variability through image normalisation: a pilot study. *Scientific Reports*, 12(1):12532, 2022.

[35] Víctor M Campello, Carlos Martín-Isla, Cristian Izquierdo, Steffen E Petersen, Miguel A González Ballester, and Karim Lekadir. Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhancement cardiac mri. In *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Revised Selected Papers 10*, pages 290–299. Springer, 2020.

[36] Víctor M. Campello, Carlos Martín-Isla, Cristian Izquierdo, Steffen E. Petersen, Miguel A. González Ballester, and Karim Lekadir. Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhancement cardiac MRI. In *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*, pages 290–299. Springer International Publishing, 2020.

[37] Victor Manuel Campello, Tian Xia, Xiao Liu, Pedro Sanchez, Carlos Martín-Isla, Steffen Erhard Petersen, Santi Seguí, Sotirios Tsaftaris, and Karim Lekadir. Cardiac aging synthesis from cross-sectional data with conditional generative adversarial networks. *Frontiers in Cardiovascular Medicine*, page 2693, 2022.

[38] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*, pages 205–218. Springer, 2023.

[39] Irem Cetin, Gerard Sanroma, Steffen E Petersen, Sandy Napel, Oscar Camara, Miguel-Angel Gonzalez Ballester, and Karim Lekadir. A radiomics approach to computer-aided diagnosis with cardiac cine-mri. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 82–90. Springer, 2017.

[40] CB CHAPMAN, O BAKER, J REYNOLDS, and FJ BONTE. Use of biplane cinefluorography for measurement of ventricular volume. *Circulation*, 1958. Query date: 2023-06-05.

[41] Chen Chen, Wenjia Bai, Rhodri H Davies, Anish N Bhuva, Charlotte H Manisty, Joao B Augusto, James C Moon, Nay Aung, Aaron M Lee, Mihir M Sanghvi, et al. Improving the generalizability of convolutional neural network-based segmentation on cmr images. *Frontiers in cardiovascular medicine*, 7:105, 2020.

[42] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. *arXiv preprint arXiv:1901.08211*, 2019.

[43] Jun Chen, Heye Zhang, Weiwei Zhang, Xiuquan Du, Yanping Zhang, and Shuo Li. Correlated regression feature learning for automated right ventricle segmentation. *IEEE journal of translational engineering in health and medicine*, 6:1–10, 2018.

[44] Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5):545–563, 2021.

[45] CK Chow and T Kaneko. Automatic boundary detection of the left ventricle from cineangiograms. *Computers and biomedical research*, 1972. Query date: 2023-06-05.

[46] Marcin Ciecholewski. Support vector machine approach to cardiac spect diagnosis. In *International workshop on combinatorial image analysis*, pages 432–443. Springer, 2011.

[47] Adriaan Coenen, Young-Hak Kim, Mariusz Kruk, Christian Tesche, Jakob De Geer, Akira Kurata, Marisa L Lubbers, Joost Daemen, Lucian Itu, Saikiran Rapaka, et al. Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography–based fractional flow reserve: result from the machine consortium. *Circulation: Cardiovascular Imaging*, 11(6):e007217, 2018.

[48] Domenico Conforti and Rosita Guido. Kernel-based support vector machine classifiers for early detection of myocardial infarction. *Optimization Methods and Software*, 20(2-3):401–413, 2005.

[49] Jorge Corral Acero, Vaanathi Sundaresan, Nicola Dinsdale, Vicente Grau, and Mark Jenkinson. A 2-step deep learning method with domain adaptation for multi-centre, multi-vendor and multi-disease cardiac magnetic resonance segmentation. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges: 11th International Workshop, STACOM 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers 11*, pages 196–207. Springer, 2021.

[50] Pascal de Groote, Alain Millaire, Claude Foucher-Hossein, Olivier Nugue, Xavier Marchandise, Gérard Ducloux, and Jean-Marc Lablanche. Right ventricular ejection fraction is an independent predictor of survival in patients with moderate heart failure. *Journal of the American College of Cardiology*, 32(4):948–954, 1998.

[51] J Mauricio Del Rio, Loreta Grecu, and Alina Nicoara. Right ventricular function in left heart disease. In *Seminars in Cardiothoracic and Vascular Anesthesia*, volume 23, pages 88–107. SAGE Publications Sage CA: Los Angeles, CA, 2019.

[52] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[53] Matthias Eisenmann, Annika Reinke, Vivienn Weru, Minu D Tizabi, Fabian Isensee, Tim J Adler, Sharib Ali, Vincent Andrearczyk, Marc Aubreville, Ujjwal Baid, et al. Why is the winner the best? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19955–19966, 2023.

[54] Atta Elalfi, Mohamed Eisa, and Hosnia Ahmed. Artificial neural networks in medical images for diagnosis heart valve diseases. *International Journal of Computer Science Issues (IJCSI)*, 10(5):83, 2013.

[55] Nowell M. Fine, Libo Chen, Paul M. Bastiansen, Robert P. Frantz, Patricia A. Pellikka, Jae K. Oh, and Garvan C. Kane. Outcome prediction by quantitative right ventricular function assessment in 575 subjects evaluated for pulmonary hypertension. *Circulation: Cardiovascular Imaging*, 6(5):711–721, sep 2013.

[56] Andrew S. Flett, Jonathan Hasleton, Christopher Cook, Derek Hausenloy, Giovanni Quarta, Cono Ariti, Vivek Muthurangu, and James C. Moon. Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. *JACC: Cardiovascular Imaging*, 4(2):150–156, February 2011.

[57] Kunihiko Fukushima, Sei Miyake, and Takayuki Ito. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE transactions on systems, man, and cybernetics*, (5):826–834, 1983.

[58] Mitchell J Fulton, Christoffer R Heckman, and Mark E Rentschler. Deformable bayesian convolutional networks for disease-robust cardiac mri segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 296–305. Springer, 2021.

[59] Francesco Galati and Maria A Zuluaga. Using out-of-distribution detection for model refinement in cardiac image segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 374–382. Springer, 2021.

[60] Christoforos Galazis, Huiyi Wu, Zhuoyu Li, Camille Petri, Anil A Bharath, and Marta Varela. Tempera: Spatial transformer feature pyramid network for cardiac mri segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 268–276. Springer, 2021.

[61] Zheyao Gao and Xiahai Zhuang. Consistency based co-segmentation for multi-view cardiac mri using vision transformer. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 306–314. Springer, 2021.

[62] Stefano Ghio, Pier Luigi Temporelli, Catherine Klersy, Anca Simioniuc, Bruna Girardi, Laura Scelsi, Andrea Rossi, Mariantonietta Cicoira, Franco Tarro Genta, and Frank L. Dini. Prognostic relevance of a non-invasive evaluation of right ventricular function and pulmonary artery pressure in patients with chronic heart failure. *European Journal of Heart Failure*, 15(4):408–414, apr 2013.

[63] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[64] Vanathi Gopalakrishnan, Prahlad G Menon, and Shobhit Madan. cmri-bed: A novel informatics framework for cardiac mri biomarker extraction and discovery applied to pediatric cardiomyopathy classification. *Biomedical engineering online*, 14(2):S7, 2015.

[65] Ankur Gulati, Tevfik F. Ismail, Andrew Jabbour, Francisco Alpendurada, Kaushik Guha, Nizar A. Ismail, Sadaf Raza, Jahanzaib Khwaja, Tristan D.H. Brown, Kishen Morarji, Emmanouil Liodakis, Michael Roughton, Ricardo Wage, Tapesh C. Pakrashi, Rakesh Sharma, John-Paul Carpenter, Stuart A. Cook, Martin R. Cowie, Ravi G. Assomull, Dudley J. Pennell, and Sanjay K. Prasad. The prevalence and prognostic significance of right ventricular systolic dysfunction in nonischemic dilated cardiomyopathy. *Circulation*, 128(15):1623–1633, oct 2013.

[66] Levent A Guner, Nese Ilgin Karabacak, Ozgur U Akdemir, Pinar Senkul Karagoz, Sinan A Kocaman, Atiye Cengel, and Mustafa Unlu. An open-source framework of neural networks for diagnosis of coronary artery disease from myocardial perfusion spect. *Journal of Nuclear Cardiology*, 17(3):405–413, 2010.

[67] Franc
ois Haddad, Ramona Doyle, Daniel J. Murphy, and Sharon A. Hunt. Right ventricular function in cardiovascular disease, part II. *Circulation*, 117(13):1717–1731, apr 2008.

[68] Donghee Han, Ji Hyun Lee, Asim Rizvi, Heidi Gransar, Lohendran Baskaran, Joshua Schulman-Marcus, Bríain ó Hartaigh, Fay Y Lin, and James K Min. Incremental role of resting myocardial computed tomography perfusion for predicting physiologically significant coronary artery disease: a machine learning approach. *Journal of Nuclear Cardiology*, 25(1):223–233, 2018.

[69] W Harvey. Exercitatio anatomica de motu cordis et sanguinis in animalibus. *Frankfurt am Main*, 1928. Query date: 2023-06-05.

[70] Kaiwen Huang, Lei Xu, Yingliang Zhu, and Penghui Meng. Au-snake based deep learning network for right ventricle segmentation. *Medical Physics*, 2022.

[71] Yuankai Huo, Zhoubing Xu, Hyeonsoo Moon, Shunxing Bao, Albert Assad, Tamara K Moyo, Michael R Savona, Richard G Abramson, and Bennett A Landman. Synseg-net: Synthetic segmentation without target modality ground truth. *IEEE transactions on medical imaging*, 38(4):1016–1025, 2018.

[72] Fabian Isensee, Paul Jaeger, Peter M. Full, Ivo Wolf, S. Engelhardt, and Klaus Maier-Hein. Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features. *ArXiv*, abs/1707.00587, 2017.

[73] Fabian Isensee, Paul F Jaeger, Peter M Full, Ivo Wolf, Sandy Engelhardt, and Klaus H Maier-Hein. Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features. In *International workshop on statistical atlases and computational models of the heart*, pages 120–129. Springer, 2017.

[74] Fabian Isensee, Paul F. Jager, Simon A. A. Kohl, J. Petersen, and Klaus Maier-Hein. Automated design of deep learning methods for biomedical image segmentation. *arXiv: Computer Vision and Pattern Recognition*, 2019.

[75] Fabian Isensee, Jens Petersen, Simon A. A. Kohl, P. Jäger, and Klaus Maier-Hein. nnu-net: Breaking the spell on successful medical image segmentation. *ArXiv*, abs/1904.08128, 2019.

[76] Ivana Išgum, Annemarieke Rutten, Mathias Prokop, and Bram van Ginneken. Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease. *Medical physics*, 34(4):1450–1461, 2007.

[77] Ivana Išgum, Bram van Ginneken, and Marco Olree. Automatic detection of calcifications in the aorta from ct scans of the abdomen1: 3d computer-aided diagnosis. *Academic radiology*, 11(3):247–257, 2004.

[78] Cristian Izquierdo Morcillo, Guillem Casas Masnou, Carlos Martin-Isla, Víctor Manuel Campello, Andrea Guala, Polyxeni Gkontra, José F Rodriguez-Palomares, and Karim Lekadir. Radiomics-based classification of left ventricular non-compaction, hypertrophic cardiomyopathy, and dilated cardiomyopathy in cardiovascular magnetic resonance. *Frontiers in Cardiovascular Medicine, 2021, vol. 8*, 2021.

[79] Sana Jabbar, Syed Talha Bukhari, and Hassan Mohy-ud Din. Multi-view sa-la net: A framework for simultaneous segmentation of rv on multi-view cardiac mr images. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 277–286. Springer, 2021.

[80] Nicholas Jones, Andrew T. Burns, and David L. Prior. Echocardiographic assessment of the right ventricle–state of the art. *Heart, Lung and Circulation*, 28(9):1339–1350, sep 2019.

[81] Smriti Joshi, Richard Osuala, Carlos Martín-Isla, Victor M Campello, Carla Sendra-Balcells, Karim Lekadir, and Sergio Escalera. nn-unet training on cyclegan-translated images for cross-modal domain adaptation in biomedical imaging. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part II*, pages 540–551. Springer, 2022.

[82] Luis Eduardo Juarez-Orozco, Remco JJ Knol, Carlos A Sanchez-Catasus, Octavio Martinez-Manzanera, Friso M van der Zant, and Juhani Knuuti. Machine learning in the integration of simple variables for identifying patients with myocardial ischemia. *Journal of Nuclear Cardiology*, pages 1–9, 2018.

[83] Saeed Karimi-Bidhendi, Arghavan Arafati, Andrew L Cheng, Yilei Wu, Arash Kheradvar, and Hamid Jafarkhani. Fully-automated deep-learning segmentation of pediatric cardiovascular magnetic resonance of patients with complex congenital heart diseases. *Journal of cardiovascular magnetic resonance*, 22(1):80, 2020.

[84] Kalliopi Keramida, George Lazaros, and Petros Nihoyannopoulos. Right ventricular involvement in hypertrophic cardiomyopathy: Patterns and implications. *Hellenic Journal of Cardiology*, 61(1):3–8, 2020.

[85] Mahendra Khened, Varghese Alex, and Ganapathy Krishnamurthi. Densely connected fully convolutional network for short-axis cardiac cine mr image segmentation and heart diagnosis using random forest. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 140–151. Springer, 2017.

[86] Igor Klem, Einar Heiberg, Lowie Van Assche, Michele A. Parker, Han W. Kim, John D. Grizzard, Håkan Arheden, and Raymond J. Kim. Sources of variability in quantification of cardiovascular magnetic resonance infarct size - reproducibility among three core laboratories. *Journal of Cardiovascular Magnetic Resonance*, 19(1), August 2017.

[87] Márton Kolossváry, Júlia Karády, Bálint Szilveszter, Pieter Kitslaar, Udo Hoffmann, Béla Merkely, and Pál Maurovich-Horvat. Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign. *Circulation: Cardiovascular Imaging*, 10(12):e006843, 2017.

[88] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[89] Matjaž Kukar, Igor Kononenko, Ciril Grošelj, Katarina Kralj, and Jure Fettich. Analysing and improving the diagnosis of ischaemic heart disease with machine learning. *Artificial intelligence in medicine*, 16(1):25–50, 1999.

[90] Virendra Kumar, Yuhua Gu, Satrajit Basu, Anders Berglund, Steven A Eschrich, Matthew B Schabath, Kenneth Forster, Hugo JWL Aerts, Andre Dekker, David Fenstermacher, et al. Radiomics: the process and the challenges. *Magnetic resonance imaging*, 30(9):1234–1248, 2012.

[91] Tanja Kurzendorfer, Christoph Forman, Michaela Schmidt, Christoph Tillmanns, Andreas Maier, and Alexander Brost. Fully automatic segmentation of left

ventricular anatomy in 3-D LGE-MRI. *Computerized Medical Imaging and Graphics*, 59:13–27, 2017.

[92] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud GPM Van Stiphout, Patrick Granton, Catharina ML Zegers, Robert Gillies, Ronald Boellard, André Dekker, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*, 48(4):441–446, 2012.

[93] Andrés Larroza, María P López-Lereu, José V Monmeneu, Jose Gavara, Francisco J Chorro, Vicente Bodí, and David Moratal. Texture analysis of cardiac cine magnetic resonance imaging to detect nonviable segments in patients with chronic myocardial infarction. *Medical physics*, 45(4):1471–1480, 2018.

[94] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[95] Yann LeCun et al. Generalization and network design strategies. *Connectionism in perspective*, 19(143-155):18, 1989.

[96] Karim Lekadir, Xènia Albà, Marco Pereañez, and Alejandro F Frangi. Statistical shape modeling using partial least squares: application to the assessment of myocardial infarction. In *Statistical Atlases and Computational Models of the Heart*, pages 130–139. Springer, 2015.

[97] Karim Lekadir, Richard Osuala, Catherine Gallin, Noussair Lazrak, Kaisar Kushibar, Gianna Tsakou, Susanna Aussó, Leonor Cerdá Alberich, Kostas Marias, Manolis Tsiknakis, et al. Future-ai: guiding principles and consensus recommendations for trustworthy artificial intelligence in medical imaging. *arXiv preprint arXiv:2109.09658*, 2021.

[98] Jingcong Li, Zhu Liang Yu, Zhenghui Gu, Hui Liu, and Yuanqing Li. Dilated-inception net: multi-scale feature aggregation for cardiac right ventricle segmentation. *IEEE Transactions on Biomedical Engineering*, 66(12):3499–3508, 2019.

[99] Lei Li, Wangbin Ding, Liqin Huang, and Xiahai Zhuang. Right ventricular segmentation from short-and long-axis mris via information transition. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 259–267. Springer, 2021.

[100] Lei Li, Fuping Wu, Sihan Wang, Xinzhe Luo, Carlos Martín-Isla, Shuwei Zhai, Jianpeng Zhang, Yanfei Liu, Zhen Zhang, Markus J Ankenbrand, et al. Myops: A benchmark of myocardial pathology segmentation combining three-sequence cardiac magnetic resonance images. *Medical Image Analysis*, 87:102808, 2023.

[101] Jesse Lieman-Sifry, Matthieu Le, Felix Lau, Sean Sall, and Daniel Golden. Fastventricle: cardiac segmentation with enet. In *Functional Imaging and Modelling of the Heart: 9th International Conference, FIMH 2017, Toronto, ON, Canada, June 11-13, 2017, Proceedings*, pages 127–138. Springer, 2017.

[102] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[103] Di Liu, Zhennan Yan, Qi Chang, Leon Axel, and Dimitris N Metaxas. Refined deep layer aggregation for multi-disease, multi-view & multi-center cardiac mr segmentation. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 315–322. Springer, 2021.

[104] Leo Lopez, Meryl S. Cohen, Robert H. Anderson, Andrew N. Redington, David G. Nykanen, Daniel J. Penny, John E. Deanfield, and Benjamin W. Eidem. Unnatural history of the right ventricle in patients with congenitally malformed hearts. *Cardiology in the Young*, 20(S3):107–112, dec 2010.

[105] Allen Lu, Ehsan Dehghan, Gopalkrishna Veni, Mehdi Moradi, and Tanveer Syeda-Mahmood. Detecting anomalies from echocardiography using multi-view regression of clinical measurements. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1504–1508. IEEE, 2018.

[106] Gongning Luo, Ran An, Kuanquan Wang, Suyu Dong, and Henggui Zhang. A deep learning network for right ventricle segmentation in short-axis mri. In *2016 Computing in Cardiology Conference (CinC)*, pages 485–488. IEEE, 2016.

[107] Ali Madani, Jia Rui Ong, Anshul Tibrewal, and Mohammad RK Mofrad. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. *npj Digital Medicine*, 1(1):59, 2018.

[108] Manoj Mannil, Jochen von Spiczak, Robert Manka, and Hatem Alkadhi. Texture analysis and machine learning for detecting myocardial infarction in noncontrast low-dose computed tomography: unveiling the invisible. *Investigative radiology*, 53(6):338–343, 2018.

[109] Manoj Mannil, Jochen von Spiczak, Urs J Muehlematter, Arjun Thanabalasingam, Dagmar I Keller, Robert Manka, and Hatem Alkadhi. Texture analysis of myocardial infarction in ct: Comparison with visual analysis and impact of iterative reconstruction. *European journal of radiology*, 113:245–250, 2019.

[110] DS Marcus, T Olsen, M Ramaratnam, and RL Buckner. Xnat: a software framework for managing neuroimaging laboratory data. In *Proceedings of the 12th annual meeting of the organization for human brain mapping, Florence*, 2006.

[111] Jorge Mariscal-Harana, Naomi Kifle, Reza Razavi, Andrew P King, Bram Ruijsink, and Esther Puyol-Antón. Improved ai-based segmentation of apical and basal slices from clinical cine cmr. In *Statistical Atlases and Computational Models of the Heart. Multi-Disease, Multi-View, and Multi-Center Right Ventricular Segmentation in Cardiac MRI Challenge: 12th International Workshop, STACOM 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Revised Selected Papers*, pages 84–92. Springer, 2022.

[112] Carlos Martín-Isla, Maryam Asadi-Aghbolaghi, Polyxeni Gkontra, Victor M Campello, Sergio Escalera, and Karim Lekadir. Stacked bcdu-net with semantic cmr synthesis: Application to myocardial pathology segmentation challenge. In *Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1*, pages 1–16. Springer, 2020.

[113] Carlos Martín-Isla, Maryam Asadi-Aghbolaghi, Polyxeni Gkontra, Victor M Campello, Sergio Escalera, and Karim Lekadir. Stacked bcdu-net with semantic cmr synthesis: Application to myocardial pathology segmentation challenge. In *Myocardial Pathology Segmentation Combining Multi-Sequence Cardiac Magnetic Resonance Images: First Challenge, MyoPS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings 1*, pages 1–16. Springer, 2020.

[114] Carlos Martín-Isla, Víctor M Campello, Cristian Izquierdo, Kaisar Kushibar, Carla Sendra-Balcells, Polyxeni Gkontra, Alireza Sojoudi, Mitchell J Fulton, Tewodros Weldebirhan Arega, Kumaradevan Punithakumar, et al. Deep learning segmentation of the right ventricle in cardiac mri: The m&ms challenge. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[115] Carlos Martin-Isla, Victor M Campello, Cristian Izquierdo, Zahra Raisi-Estabragh, Bettina Baeßler, Steffen E Petersen, and Karim Lekadir. Image-based cardiac diagnosis with machine learning: a review. *Frontiers in cardiovascular medicine*, 7:1, 2020.

[116] Moona Mazher, Abdul Qayyum, Abdesslam Benzinou, Mohamed Abdel-Nasser, and Domenec Puig. Multi-disease, multi-view and multi-center right ventricular segmentation in cardiac mri using efficient late-ensemble deep learning approach. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 335–343. Springer, 2021.

[117] Shamir R Mehta, John W Eikelboom, Madhu K Natarajan, Rafael Diaz, Cheelong Yi, Raymond J Gibbons, and Salim Yusuf. Impact of right ventricular involvement on mortality and morbidity in patients with inferior myocardial infarction. *Journal of the American College of Cardiology*, 37(1):37–43, 2001.

[118] Luc L. Mertens and Mark K. Friedberg. Imaging the right ventricle—current state of the art. *Nature Reviews Cardiology*, 7(10):551–563, aug 2010.

[119] Hanie Moghaddasi, Saeed Nourian, and Saee Nourian. Automatic assessment of mitral regurgitation severity based on extensive textural features on 2d echocardiography videos. *Computers in biology and medicine*, 73:47–55, 2016.

[120] Alejandra Moreno, Jefferson Rodriguez, and Fabio Martínez. Regional multiscale motion representation for cardiac disease prediction. In *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pages 1–5. IEEE, 2019.

[121] Akifumi Nakada, Akira Shiozaki, Yutaka Hirano, Hisakazu Uehara, and Tohru Masuyama. Using neural networks in color kinesis image process to automate diagnosis of cardiac disease. *Electronics and Communications in Japan (Part II: Electronics)*, 89(2):46–54, 2006.

[122] Kenichi Nakajima, Takashi Kudo, Tomoaki Nakata, Keisuke Kiso, Tokuo Kasai, Yasuyo Taniguchi, Shinro Matsuo, Mitsuru Momose, Masayasu Nakagawa, Masayoshi Sarai, et al. Diagnostic accuracy of an artificial neural network compared with statistical quantitation of myocardial perfusion images: a japanese multicenter study. *European journal of nuclear medicine and molecular imaging*, 44(13):2280–2289, 2017.

[123] Kenichi Nakajima, Koichi Okuda, Satoru Watanabe, Shinro Matsuo, Seigo Kinuya, Karin Toth, and Lars Edenbrandt. Artificial neural network retrained to detect myocardial ischemia using a japanese multicenter database. *Annals of nuclear medicine*, 32(5):303–310, 2018.

[124] Sukrit Narula, Khader Shameer, Alaa Mabrouk Salem Omar, Joel T Dudley, and Partho P Sengupta. Machine-learning algorithms to automate morphological and functional assessments in 2d echocardiography. *Journal of the American College of Cardiology*, 68(21):2287–2295, 2016.

[125] Ulf Neisius, Hossam El-Rewaidy, Shiro Nakamori, Jennifer Rodriguez, Warren J Manning, and Reza Nezafat. Radiomic analysis of myocardial native t1 imaging discriminates between hypertensive heart disease and hypertrophic cardiomyopathy. *JACC: Cardiovascular Imaging*, 2019.

[126] Ulf Neisius, Lana Myerson, Ahmed S Fahmy, Shiro Nakamori, Hossam El-Rewaidy, Gargi Joshi, Chong Duan, Warren J Manning, and Reza Nezafat. Cardiovascular magnetic resonance feature tracking strain analysis for discrimination between hypertensive heart disease and hypertrophic cardiomyopathy. *PloS one*, 14(8):e0221061, 2019.

[127] Mostafa Ghelich Oghli, Ali Mohammadzadeh, Raheleh Kafieh, and Saeed Kermani. A hybrid graph-based approach for right ventricle segmentation in cardiac mri by long axis information transition. *Physica Medica*, 54:103–116, 2018.

[128] O. Oktay, Jo Schlemper, Loïc Le Folgoc, M. J. Lee, M. Heinrich, K. Misawa, K. Mori, Steven G. McDonagh, N. Hammerla, Bernhard Kainz, Ben Glocker, and

D. Rueckert. Attention u-net: Learning where to look for the pancreas. *ArXiv*, abs/1804.03999, 2018.

[129] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[130] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2019.

[131] Adrien Pavao, Isabelle Guyon, Anne-Catherine Letournel, Xavier Baró, Hugo Escalante, Sergio Escalera, Tyler Thomas, and Zhen Xu. *CodaLab Competitions: An open source platform to organize scientific challenges*. PhD thesis, Université Paris-Saclay, FRA., 2022.

[132] Caroline Petitjean, Maria A. Zuluaga, Wenjia Bai, Jean-Nicolas Dacher, Damien Grosgeorge, Jérôme Caudron, Su Ruan, Ismail Ben Ayed, M. Jorge Cardoso, Hsiang-Chou Chen, Daniel Jimenez-Carretero, Maria J. Ledesma-Carbayo, Christos Davatzikos, Jimit Doshi, Guray Erus, Oskar M.O. Maier, Cyrus M.S. Nambakhsh, Yangming Ou, Sébastien Ourselin, Chun-Wei Peng, Nicholas S. Peters, Terry M. Peters, Martin Rajchl, Daniel Rueckert, Andres Santos, Wenzhe Shi, Ching-Wei Wang, Haiyan Wang, and Jing Yuan. Right ventricle segmentation from cardiac mri: A collation study. *Medical Image Analysis*, 19(1):187 – 202, 2015.

[133] Esmeralda Ruiz Pujadas, Zahra Raisi-Estabragh, Liliana Szabo, Celeste McCracken, Cristian Izquierdo Morcillo, Víctor M Campello, Carlos Martín-Isla, Angelica M Atehortua, Hajnalka Vago, Bela Merkely, et al. Prediction of incident cardiovascular events using machine learning and cmr radiomics. *European Radiology*, pages 1–13, 2022.

[134] Esmeralda Ruiz Pujadas, Zahra Raisi-Estabragh, Liliana Szabo, Cristian Izquierdo Morcillo, Víctor M Campello, Carlos Martin-Isla, Hajnalka Vago, Bela Merkely, Nicholas C Harvey, Steffen E Petersen, et al. Atrial fibrillation prediction by combining ecg markers and cmr radiomics. *Scientific Reports*, 12(1):18876, 2022.

[135] Kumaradevan Punithakumar, Adam Carscadden, and Michelle Noga. Automated segmentation of the right ventricle from magnetic resonance imaging using deep convolutional neural networks. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 344–351. Springer, 2021.

[136] Esther Puyol-Antón, Bram Ruijsink, Bernhard Gerber, Mihaela Silvia Amzulescu, Hélène Langet, Mathieu De Craene, Julia A Schnabel, Paolo Piro, and Andrew P King. Regional multi-view learning for cardiac motion analysis: Application to identification of dilated cardiomyopathy patients. *IEEE Transactions on Biomedical Engineering*, 66(4):956–966, 2018.

[137] Sandro Queirós. Right ventricular segmentation in multi-view cardiac mri using a unified u-net model. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 287–295. Springer, 2021.

[138] P Radau, Y Lu, K Connelly, G Paul, A Dick, and G Wright. Evaluation framework for algorithms segmenting short axis cardiac mri. *The MIDAS Journal-Cardiac MR Left Ventricle Segmentation Challenge*, 49, 2009.

[139] Zahra Raisi-Estabragh, Cristian Izquierdo, Victor M Campello, Carlos Martin-Isla, Akshay Jaggi, Nicholas C Harvey, Karim Lekadir, and Steffen E Petersen. Cardiac magnetic resonance radiomics: basic principles and clinical perspectives. *European Heart Journal-Cardiovascular Imaging*, 21(4):349–356, 2020.

[140] Zahra Raisi-Estabragh, Carlos Martin-Isla, Louise Nissen, Liliana Szabo, Sergio Escalera, Simon Winther, Morten Bøttcher, Karim Lekadir, and Steffen Erhard Petersen. Radiomics analysis enhances the diagnostic performance of cmr stress perfusion: A proof-of-concept study using the dan-nicad dataset short title: Radiomics analysis for cmr perfusion. *Frontiers in Cardiovascular Medicine*, 10:1141026, 2023.

[141] Bushra S. Rana, Shaun Robinson, Rajeevan Francis, Mark Toshner, Martin J. Swaans, Sharad Agarwal, Ravi de Silva, Amer A. Rana, and Petros Nihoyannopoulos. Tricuspid regurgitation and the right ventricle in risk stratification and timing of intervention. *Echo Research &amp Practice*, 6(1):R26–R40, mar 2019.

[142] Vera H. Rigolin, Paul A. Robiolio, John S. Wilson, J. Kevin Harrison, and Thomas M. Bashore. Impact of right ventricular involvement on mortality and morbidity in patients with inferior myocardial infarction. *Catheterization and Cardiovascular Diagnosis*, 35:18–28, 1995.

[143] Hannah Ritchie and M Roser. Our world in data. *Causes of death, published February*, 2018.

[144] Marc-Michel Rohé, Nicolas Duchateau, Maxime Sermesant, and Xavier Pennec. Combination of polyaffine transformations and supervised learning for the automatic diagnosis of lv infarct. In *Statistical Atlases and Computational Models of the Heart*, pages 190–198. Springer, 2015.

[145] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[146] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, pages 234–241. Springer International Publishing, 2015.

[147] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, pages 234–241. Springer International Publishing, 2015.

[148] Jarosław P Sacha, Lucy S Goodenday, and Krzysztof J Cios. Bayesian learning for cardiac spect image interpretation. *Artificial Intelligence in Medicine*, 26(1-2):109–143, 2002.

[149] Cian M Scannell, Amedeo Chiribiri, and Mitko Veta. Domain-adversarial learning for multi-centre, multi-vendor, and multi-disease cardiac mr image segmentation. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges: 11th International Workshop, STACOM 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers 11*, pages 228–237. Springer, 2021.

[150] Carla Sendra-Balcells, Víctor M Campello, Carlos Martín-Isla, David Vilades Medel, Martín Luís Descalzo, Andrea Guala, José F Rodríguez Palomares, and Karim Lekadir. Multi-center, multi-vendor automated segmentation of left ventricular anatomy in contrast-enhanced mri. *arXiv preprint arXiv:2110.07360*, 2021.

[151] Carla Sendra-Balcells, Víctor M Campello, Carlos Martín-Isla, David Viladés, Martín L Descalzo, Andrea Guala, José F Rodríguez-Palomares, and Karim Lekadir. Domain generalization in deep learning for contrast-enhanced imaging. *Computers in Biology and Medicine*, 149:106052, 2022.

[152] Sanjiv J Shah, Daniel H Katz, Senthil Selvaraj, Michael A Burke, Clyde W Yancy, Mihai Gheorghiade, Robert O Bonow, Chiang-Ching Huang, and Rahul C Deo. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*, 131(3):269–279, 2015.

[153] Takayuki Shibutani, Kenichi Nakajima, Hiroshi Wakabayashi, Hiroshi Mori, Shinro Matsuo, Hiroto Yoneyama, Takahiro Konishi, Koichi Okuda, Masahisa Onoguchi, and Seigo Kinuya. Accuracy of an artificial neural network for detecting a regional abnormality in myocardial perfusion spect. *Annals of nuclear medicine*, 33(2):86–92, 2019.

[154] Gerard Snaauw, Dong Gong, Gabriel Maicas, Anton van den Hengel, Wiro J Niessen, Johan Verjans, and Gustavo Carneiro. End-to-end diagnosis and segmentation learning from cardiac magnetic resonance imaging. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 802–805. IEEE, 2019.

[155] Hongmei Song, Wenguan Wang, Sanyuan Zhao, Jianbing Shen, and Kin-Man Lam. Pyramid dilated deeper ConvLSTM for video salient object detection. In *Computer Vision – ECCV 2018*, pages 744–760. Springer International Publishing, 2018.

[156] H William Strauss, Barry L Zaret, Peter J Hurley, TK Natarajan, and Bertram Pitt. A scintiphotographic method for measuring left ventricular ejection fraction in man without cardiac catheterization. *The American journal of cardiology*, 28(5):575–580, 1971.

[157] Vidya K Sudarshan, U Rajendra Acharya, EYK Ng, Ru San Tan, Siaw Meng Chou, and Dhanjoo N Ghista. An integrated index for automated detection of infarcted myocardium from cross-sectional echocardiograms using texton-based features (part 1). *Computers in biology and medicine*, 71:231–240, 2016.

[158] Avan Suinesiaputra, B. Cowan, Ahmed O. Al-Agamy, M. A. Alattar, N. Ayache, A. Fahmy, A. Khalifa, Pau Medrano-Gracia, M. Jolly, A. H. Kadish, D. Lee, Ján Margeta, S. Warfield, and A. Young. A collaborative resource to build consensus for automated left ventricular segmentation of cardiac mr images. *Medical image analysis*, 18 1:50–62, 2014.

[159] Xiaowu Sun, Li-Hsin Cheng, and Rob J Geest. Right ventricle segmentation via registration and multi-input modalities in cardiac magnetic resonance imaging from multi-disease, multi-view and multi-center. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 241–249. Springer, 2021.

[160] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[161] Mahdi Tabassian, Martino Alessandrini, Lieven Herbots, Oana Mirea, Efstathios D Pagourelias, Ruta Jasaityte, Jan Engvall, Luca De Marchi, Guido Masetti, and Jan D'hooge. Machine learning of the spatio-temporal characteristics of echocardiographic deformation curves for infarct classification. *The international journal of cardiovascular imaging*, 33(8):1159–1167, 2017.

[162] Mahdi Tabassian, Imran Sunderji, Tamas Erdei, Sergio Sanchez-Martinez, Anna Degiovanni, Paolo Marino, Alan G Fraser, and Jan D'hooge. Diagnosis of heart failure with preserved ejection fraction: machine learning of spatiotemporal variations in left ventricular deformation. *Journal of the American Society of Echocardiography*, 31(12):1272–1284, 2018.

[163] Qian Tao, Sebastiaan R.D. Piers, Hildo J. Lamb, and Rob J. van der Geest. Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment. *Journal of Magnetic Resonance Imaging*, 42(2):390–399, November 2014.

[164] Qian Tao, Sebastiaan RD Piers, Hildo J Lamb, and Rob J van der Geest. Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment. *Journal of Magnetic Resonance Imaging*, 42(2):390–399, 2015.

[165] Lennart Tautz, Lars Walczak, Chiara Manini, Anja Hennemuth, and Markus Hüllebrand. 3d right ventricle reconstruction from 2d u-net segmentation of sparse short-axis and 4-chamber cardiac cine mri views. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 352–359. Springer, 2021.

[166] Holger Thiele, Mathias J.E. Kappl, Stefan Conradi, Josef Niebauer, Rainer Hambrecht, and Gerhard Schuler. Reproducibility of chronic and acute infarct size measurement by delayed enhancement-magnetic resonance imaging. *Journal of the American College of Cardiology*, 47(8):1641–1645, April 2006.

[167] Cuong To and Tuan D Pham. Analysis of cardiac imaging data using decision tree based parallel genetic programming. In *2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis*, pages 317–320. IEEE, 2009.

[168] P. Tran. A fully convolutional neural network for cardiac segmentation in short-axis mri. *ArXiv*, abs/1604.00494, 2016.

[169] Phi Vu Tran. A fully convolutional neural network for cardiac segmentation in short-axis mri.

[170] Du-Yih Tsai, Yongbum Lee, Masaru Sekiya, and Masaki Ohkubo. Medical image classification using genetic-algorithm based fuzzy-logic approach. *Journal of Electronic Imaging*, 13(4):780–789, 2004.

[171] Kathrin Ungru, Daniel Tenbrinck, Xiaoyi Jiang, and Jörg Stypmann. Automatic classification of left ventricular wall segments in small animal ultrasound imaging. *Computer methods and programs in biomedicine*, 117(1):2–12, 2014.

[172] K Sudarshan Vidya, EYK Ng, U Rajendra Acharya, Siaw Meng Chou, Ru San Tan, and Dhanjoo N Ghista. Computer-aided diagnosis of myocardial infarction using ultrasound images with dwt, glcm and hos methods: a comparative study. *Computers in biology and medicine*, 62:86–93, 2015.

[173] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional GANs. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, June 2018.

[174] Dong Wei, Ying Sun, Ping Chai, Adrian Low, and Sim Heng Ong. Myocardial segmentation of late gadolinium enhanced MR images by propagation of contours from cine MR images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 428–435. Springer, 2011.

[175] Elizabeth Wilkins, L Wilson, Kremlin Wickramasinghe, Prachi Bhatnagar, Jose Leal, Ramon Luengo-Fernandez, R Burns, Mike Rayner, and Nick Townsend. European cardiovascular disease statistics 2017. 2017.

[176] Jelmer M Wolterink, Tim Leiner, Bob D de Vos, Robbert W van Hamersvelt, Max A Viergever, and Ivana Išgum. Automatic coronary artery calcium scoring in cardiac ct angiography using paired convolutional neural networks. *Medical image analysis*, 34:123–136, 2016.

[177] Jelmer M Wolterink, Tim Leiner, Richard AP Takx, Max A Viergever, and Ivana Išgum. An automatic machine learning system for coronary calcium scoring in clinical non-contrast enhanced, ecg-triggered cardiac ct. In *Medical Imaging 2014: Computer-Aided Diagnosis*, volume 9035, page 90350E. International Society for Optics and Photonics, 2014.

[178] Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. Automatic segmentation and disease classification using cardiac cine mr images. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 101–110. Springer, 2017.

[179] Ken CL Wong, Michael Tee, Marcus Chen, David A Bluemke, Ronald M Summers, and Jianhua Yao. Regional infarction identification from cardiac ct images: a computer-aided biomechanical approach. *International journal of computer assisted radiology and surgery*, 11(9):1573–1583, 2016.

[180] Hao Yang, Zexiong Liu, and Xuan Yang. Right ventricle segmentation in short-axis mri using a shape constrained dense connected u-net. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 532–540. Springer, 2019.

[181] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020.

[182] Fisher Yu, Dequan Wang, Evan Shelhamer, and Trevor Darrell. Deep layer aggregation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, June 2018.

[183] Qian Yue, Xinzhe Luo, Qing Ye, Lingchao Xu, and Xiahai Zhuang. Cardiac segmentation from LGE MRI using deep neural network incorporating shape and spatial priors. *arXiv preprint arXiv:1906.07347*, 2019.

[184] Jeffrey Zhang, Sravani Gajjala, Pulkit Agrawal, Geoffrey H Tison, Laura A Hallock, Lauren Beussink-Nelson, Mats H Lassen, Eugene Fan, Mandar A Aras, ChaRandle Jordan, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation*, 138(16):1623–1635, 2018.

[185] Ling Zhang, Andreas Wahle, Zhi Chen, John J Lopez, Tomas Kovarnik, and Milan Sonka. Predicting locations of high-risk plaques in coronary arteries in patients receiving statin therapy. *IEEE transactions on medical imaging*, 37(1):151–161, 2017.

[186] Nan Zhang, Guang Yang, Zhifan Gao, Chenchu Xu, Yanping Zhang, Rui Shi, Jennifer Keegan, Lei Xu, Heye Zhang, Zhanming Fan, et al. Deep learning for diagnosis of chronic myocardial infarction on nonenhanced cardiac cine mri. *Radiology*, 291(3):606–617, 2019.

[187] Yue Zhang, Shun Miao, Tommaso Mansi, and Rui Liao. Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 599–607. Springer, 2018.

[188] Zizhao Zhang, Lin Yang, and Yefeng Zheng. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9242–9251, 2018.

[189] He Zhao, Huiqi Li, Sebastian Maurer-Stroh, Yuhong Guo, Qiuju Deng, and Li Cheng. Supervised segmentation of un-annotated retinal fundus images by synthesis. *IEEE transactions on medical imaging*, 38(1):46–56, 2018.

[190] Qiao Zheng, Hervé Delingette, and Nicholas Ayache. Explainable cardiac pathology classification on cine mri with motion characterization by semi-supervised learning of apparent flow. *Medical image analysis*, 2019.

[191] Xiahai Zhuang. Multivariate mixture model for cardiac segmentation from multi-sequence MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 581–588. Springer International Publishing, 2016.

[192] Xiahai Zhuang. Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[193] Xiahai Zhuang. Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2933–2946, December 2019.

[194] Majd Zreik, Robbert W van Hamersvelt, Jelmer M Wolterink, Tim Leiner, Max A Viergever, and Ivana Išgum. A recurrent cnn for automatic detection and classification of coronary artery plaque and stenosis in coronary ct angiography. *IEEE transactions on medical imaging*, 2018.

[195] Maria A Zuluaga, M Jorge Cardoso, and Sebastien Ourselin. Automatic right ventricle segmentation using multi-label fusion in cardiac mri. *arXiv preprint arXiv:2004.02317*, 2020.