



Towards Automated Segmentation of Cardiac Scars through Multimodal AI

Farheen Ramzan

Main Supervisor: Richard H. Clayton

Second Supervisor: Chen (Cherise) Chen

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor
of Philosophy

in the

University of Sheffield
Faculty of Engineering
School of Computer Science

July 10, 2026

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on [Academic Integrity](#). This work has not previously been presented for an award at this, or any other, university.

Abstract

Myocardial scarring following myocardial infarction (MI) is a major determinant of adverse cardiac outcomes, including ventricular arrhythmias, heart failure, and sudden cardiac death. Accurate quantification of scar burden and spatial distribution is therefore essential for risk stratification and treatment planning. Late gadolinium enhancement cardiac magnetic resonance (LGE-CMR) provides a non-invasive reference standard for visualizing infarcted myocardium; however, clinical assessment remains largely dependent on manual segmentation, which is time-consuming, subjective, and difficult to scale. This limits its applicability in time-sensitive clinical workflows where results must be delivered rapidly. Existing automated approaches show promise but are typically developed under simplified assumptions, including homogeneous imaging protocols, fully annotated datasets, and reliance on single-modality models; conditions rarely encountered in real-world clinical practice.

This thesis investigates how multimodal artificial intelligence (AI) can enable robust and clinically applicable myocardial scar segmentation by progressively addressing these real-world constraints. First, an automated deep learning framework for multi-sequence CMR is introduced, demonstrating that integrating complementary contrasts yields more accurate and consistent segmentation of infarcted tissue than single-sequence approaches. Second, to mitigate limited annotations and underrepresentation of rare scar patterns, a clinically guided scar synthesis strategy is proposed, by incorporating anatomical and pathological priors to generate physiologically plausible training data and improve generalization.

Third, recognizing that infarct remodeling involves both structural and electrophysiological changes, a multimodal framework is developed that integrates 12-lead electrocardiogram (ECG) signals with LGE-CMR. A temporal-aware fusion mechanism explicitly models acquisition asynchrony, enabling coherent integration of electrical and anatomical information. Finally, a domain-generalized segmentation framework is introduced to address heterogeneous imaging protocols, missing modalities, and multi-center variability, ensuring stable performance under real-world deployment conditions.

Collectively, this work advances myocardial scar segmentation from a constrained image analysis task to a clinically grounded, multimodal, and robust learning paradigm, providing a foundation for scalable and clinically deployable AI-driven cardiac assessment.

Acknowledgments

I thank Allah for His countless blessings, for granting me strength and wisdom, and for guiding me through this journey. In difficult moments, He gave me hope and steadiness when I felt uncertain, and I am deeply grateful for His mercy and guidance throughout.

This thesis reflects not only years of research, but also the support and encouragement of many people along the way.

I would like to express my sincere gratitude to my supervisors, Prof. Richard H. Clayton and Dr. Chen Chen, for their unwavering support. Professor Clayton's clarity of vision and guidance profoundly shaped both this work and my development as a researcher, helping me stay focused even when the path felt uncertain. Dr. Chen Chen's energy, encouragement, and kindness were invaluable, especially during some of the most challenging periods of my PhD. Together, their mentorship helped me grow into a more independent and confident researcher.

I am also grateful to our clinical collaborators at Nottingham University Hospitals NHS Trust, Yusuf Kiberu, Shahnaz Jamil-Copley, and Nikesh Jathanna, for their valuable clinical insight and feedback.

Beyond academia, I owe everything to my family. My parents have always supported me unconditionally, giving me the freedom to pursue my ambitions and believe in myself. My mother is my safe space; her strength, love, and unwavering belief in me have carried me through every challenge. I am especially grateful to my uncles, who helped raise me alongside my parents and have always been there through thick and thin; their support has played an important role in my life and success.

To my sisters, thank you for being my constant source of light; lifting my spirits and motivating me to keep striving to do better. I am also deeply grateful to my friends for their constant support, for always listening, and for helping me stay optimistic and grounded during difficult times.

I would like to sincerely acknowledge the Commonwealth Scholarship Commission for awarding me this scholarship. Their financial support and the opportunity they provided made this PhD possible.

Finally, I thank all my teachers and professors, from school to university, whose guidance has shaped my academic journey.

This thesis may carry my name, but it reflects the support, belief, and strength of all those who stood beside me along the way.

Contents

1	Introduction	1
1.1	Clinical Motivation	1
1.2	Technical Problem Definition	4
1.3	Research Gap	6
1.4	Central Thesis Hypothesis	7
1.5	Thesis Outline and Contributions	8
1.6	Publications Arising from This Thesis	9
2	Background	10
2.1	Overview and Scope of This Chapter	10
2.2	Cardiac Anatomy and Physiology Relevant to Imaging	11
2.2.1	Left Ventricular Structure and Myocardial Fiber Organization	11
2.2.2	Coronary Perfusion Territories	11
2.2.3	Electrical Conduction System and the His–Purkinje Network	13
2.2.4	Structure-Function Coupling in Normal and Infarcted Myocardium	14
2.3	Cardiac Arrhythmias and the Structural Substrate of Disease	14
2.4	Myocardial Infarction and Post-Infarction Scar Formation	15
2.4.1	Scar Formation	15
2.4.2	Dense Scar and the Gray Zone	15
2.4.3	Remodeling and Ventricular Dilation	15
2.4.4	Why Scar Morphology Is Heterogeneous	15
2.4.5	Implications for Imaging and Segmentation	15
2.5	Clinical Importance of Scar Quantification and Localization	16
2.5.1	Risk Stratification for Sudden Cardiac Death	16
2.5.2	Guidance for Catheter Ablation	16
2.5.3	Prediction of Functional Recovery	16
2.5.4	Limitations of Ejection Fraction as a Biomarker	17
2.5.5	Need for Reproducible Spatial Quantification	17
2.6	Fundamentals of Magnetic Resonance Imaging	17
2.7	Late Gadolinium Enhancement Imaging (LGE)	18
2.7.1	Contrast Mechanism and Tissue Characterization	18
2.7.2	Imaging Protocol and Acquisition Considerations	18
2.7.3	Strengths and Limitations of LGE for Scar Assessment	19
2.7.4	Relevance to Automated Analysis	19
2.8	Multi-Sequence CMR for Tissue Characterization	19

2.9	Electrocardiography (ECG) as a Functional Marker of Infarction	20
2.9.1	Physiological Basis of the ECG Signal	20
2.9.2	Structure–Function Relationship	20
2.9.3	Implications for Multimodal Modeling	21
2.10	Public and Clinical Datasets for Myocardial Scar Analysis	21
2.11	Traditional Image Analysis Methods	22
2.12	Machine Learning and Deep Learning for Cardiac Segmentation	22
2.13	Generative Modeling and Data Augmentation in Medical Imaging	23
2.14	Multimodal Learning and Structure–Function Modeling	23
2.15	Loss Functions for Imbalanced Medical Segmentation	24
2.15.1	Mathematical Formulation of Loss Functions	24
2.16	Evaluation of Myocardial Scar Segmentation	26
2.16.1	Mathematical Formulation of Evaluation Metrics	26
2.16.2	Interpretation of Metrics for Scar Segmentation	27
2.17	The American Heart Association 17-Segment Model	27
3	Scar Segmentation from Multi-Sequence CMR	29
3.1	Introduction	29
3.2	Background	30
3.2.1	Myocardial Scar Segmentation in Cardiac MRI	30
3.2.2	Traditional and Learning-Based Approaches	31
3.2.3	Self-Configuring Networks and Architectural Design Choices	31
3.2.4	Role of Multi-Sequence CMR	32
3.2.5	Class Imbalance and Loss Function Design	32
3.2.6	Motivation for This Work	32
3.3	Methods	33
3.3.1	Dataset and Imaging Modalities	34
3.3.2	Problem Formulation	34
3.3.3	End-to-End Learning: Baseline Network	34
3.3.4	Cascaded Learning: Coarse-to-Fine Segmentation (CNN^{C2F})	34
3.3.5	Loss Function Design and Optimization Strategy	35
3.3.6	Implementation Details	36
3.4	Results and Discussion	37
3.4.1	Experimental Setup and Implementation Details	37
3.4.2	Quantitative Evaluation of Architectural Strategies	37
3.4.3	Effect of Multi-Sequence CMR Integration	38
3.4.4	Impact of Loss Function Design	38
3.4.5	Qualitative Evaluation	39
3.4.6	Summary of Findings	39
3.5	Conclusion	40
3.6	Limitations and Future Work	41
4	Clinically-Guided Scar Synthesis in LGE	43
4.1	Introduction	43
4.2	Background	45
4.2.1	Clinical Interpretation of Myocardial Scar in LGE Imaging	45

4.2.2	Learning-Based Myocardial Scar Segmentation	46
4.2.3	Data Scarcity and Domain Shift in Cardiac Imaging	46
4.2.4	Synthetic Data Generation and Pathology Augmentation	46
4.2.5	Conditional Diffusion Models for Lesion and Scar Synthesis	47
4.2.6	Limitations of Data-Driven Mask Generation	47
4.2.7	Motivation for This Work	47
4.3	Methods	48
4.3.1	Design Rationale and Architectural Choices	49
4.3.2	Joint Modelling for Enhanced Scar Synthesis and Segmentation	51
4.3.3	SMILE: Scar Masks Incorporating cLinical knowledgE	53
4.3.4	Summary of the CLAIM Framework	54
4.4	Experiments	54
4.4.1	Datasets	55
4.4.2	Experimental Setup	55
4.4.3	Methods for Comparison	56
4.5	Results and Discussion	57
4.5.1	Visual Quality Assessment on Image Synthesis	57
4.5.2	Effectiveness of CLAIM for Scar Segmentation	59
4.5.3	Clinical Interpretation of Results	63
4.6	Conclusion	65
4.7	Limitations and Future Directions	66
4.7.1	Dependence on Anatomical Registration and Atlas Accuracy	66
4.7.2	Scar Volume Distribution Bias	66
4.7.3	Scope of Pathological Modeling	66
4.7.4	Computational Complexity and Training Overhead	67
4.7.5	Opportunities for Multimodal and Temporal Integration	67
4.7.6	Toward Domain-Generalized and Clinically Deployable Systems	67
5	Multimodal ECG-LGE Scar Segmentation	68
5.1	Introduction	69
5.2	Background	70
5.2.1	Myocardial Scar as a Structural and Electrophysiological Phenomenon	70
5.2.2	Electrocardiographic Manifestations of Myocardial Scar	71
5.2.3	Automated Myocardial Scar Segmentation from LGE MRI	71
5.2.4	Multimodal Learning in Cardiac Imaging	72
5.2.5	Temporal Misalignment and Clinical Data Heterogeneity	72
5.2.6	Motivation for This Work	73
5.3	Methods	73
5.3.1	Design Rationale and Architectural Considerations	73
5.3.2	Multi-modal Network Architecture	76
5.3.3	TAFF: Temporal-Aware Feature Fusion	76
5.3.4	AHA-17 Segment Model as Anatomical Prior	77
5.3.5	Loss Functions and Training Strategy	78
5.4	Experiments and Results	78
5.4.1	Datasets and Data Preparation	78
5.4.2	Implementation Details	79

5.4.3	Evaluation Metrics and Experimental Protocol	80
5.4.4	Quantitative Results and Ablation Studies	80
5.4.5	Qualitative Analysis and Model Interpretability	80
5.5	Discussion and Analysis	81
5.5.1	Multimodal Integration Substantially Improves Scar Segmentation Performance	81
5.5.2	Challenges of Naïve ECG-MRI Fusion	82
5.5.3	Impact of Anatomical Priors on Physiological Alignment	83
5.5.4	Role of Temporal Conditioning in Multimodal Fusion	83
5.5.5	Qualitative Analysis and Model Interpretability	84
5.6	Ablation Study: Analysis of Multimodal Fusion Strategies	84
5.6.1	Unimodal Baselines and the Effect of Training Data	85
5.6.2	Why Naïve Fusion Strategies Fail	85
5.6.3	Effect of Attention-Based Multimodal Fusion on Clinical Scar Quantification	86
5.6.4	Effect of Attention-Based Multimodal Fusion on Distribution of Segmentation Performance and Scar Volume Estimation	86
5.6.5	Benefits and Limitations of Attention-Based Fusion	89
5.6.6	Why Temporal-Aware Gated Fusion Is Necessary	90
5.6.7	Summary of Fusion Strategy Insights	91
5.7	Results: Qualitative Visualization of Representative Cases	91
5.7.1	Cases with Clear Improvement Using TAFF Multimodal Fusion	91
5.7.2	Moderate Improvements and Partial Corrections	93
5.7.3	Failure and Challenging Cases	93
5.7.4	Interpretability Through ECG Attention and Anatomical Context	96
5.7.5	Summary of Qualitative Findings	98
5.8	Limitations and Future Work	98
5.9	Conclusion	99
6	Robust Scar Segmentation from Incomplete CMR	101
6.1	Introduction	101
6.2	Background	103
6.2.1	Clinical Motivation: Multi-Sequence CMR in Myocardial Pathology Assessment	103
6.2.2	Challenges in Real-World Multi-Center CMR Data	104
6.2.3	Limitations of Existing Multi-Sequence Segmentation Approaches	104
6.2.4	Learning Under Missing Data and Domain Shift	105
6.2.5	Rationale for Investigated Approaches	105
6.2.6	Positioning of This Chapter	106
6.3	Methodology	106
6.3.1	Overall Framework	107
6.3.2	M ² FTrans: Modality-Masked Fusion Transformer	107
6.3.3	Mean Teacher-Based Semi-Supervised Learning	109
6.3.4	Domain-Randomizing Augmentation with RandConv	110
6.3.5	Missing-Modality Simulation	110
6.3.6	Overall Training Objective	110

6.4	Results and Discussion	111
6.4.1	Dataset and Experimental Protocol	111
6.4.2	Implementation Details	112
6.4.3	Fully Supervised Learning under Missing-Modality Conditions	113
6.4.4	Semi-Supervised Learning under Incomplete Multi-Center Data	116
6.4.5	Evaluation on the Challenge Test Set (Center D)	118
6.4.6	Comparison with Existing Multi-Center Segmentation Approaches	120
6.5	Conclusion	125
7	Discussion and Conclusions	127
7.1	Summary and Main Findings	127
7.1.1	Chapter 3: Multi-Sequence CMR for Automated Scar Segmentation	127
7.1.2	Chapter 4: Clinically Guided Scar Synthesis	128
7.1.3	Chapter 5: Multimodal Structure-Function Integration	128
7.1.4	Chapter 6: Robustness Under Real-World Multi-Center Settings	129
7.1.5	Synthesis Across Chapters	129
7.2	Limitations of This Work	129
7.2.1	Dependence on retrospective datasets and constrained pathological diversity	130
7.2.2	Annotation uncertainty and observer variability	130
7.2.3	Architectural complexity and computational burden	130
7.2.4	Limited exploration of predictive uncertainty and reliability calibration	131
7.2.5	Limited linkage to downstream clinical outcomes	131
7.3	Future Directions	131
7.3.1	Expansion to broader cardiac pathologies and longitudinal data	131
7.3.2	Uncertainty-aware and probabilistic segmentation	132
7.3.3	Toward causal and physiologically grounded multimodal fusion	132
7.3.4	Self-supervised learning and foundation model paradigms	132
7.3.5	Extension to 3D Scar Modeling and Volumetric Inference	132
7.3.6	Prospective validation and workflow integration	133
7.3.7	Standardization, interpretability, and regulatory readiness	133
7.4	Conclusions	133
	Appendices	156
	A List of Abbreviations	157
	B Permissions for Content Reuse	159

List of Figures

1.1	(a) Illustration of coronary artery blockage leading to myocardial infarction (Image by Blausen Medical Communications, Inc. , via Wikimedia Commons, licensed under CC BY 3.0), (b) cross-sectional view showing infarcted left ventricular myocardium (Image by Patrick J. Lynch , via Wikimedia Commons, licensed under CC BY 2.5), and (c) representative LGE-CMR slice with segmented scar region (red).	2
1.2	LGE-CMR images of chronic septal and apical myocardial infarction (arrows) in short-axis (A) and four-chamber (B) views, showing hyperenhanced regions with variable morphology and extent. Figure source: [1], reproduced under the terms of the Creative Commons Attribution License CC BY 2.5.	2
1.3	Representative LGE-CMR images from different patients showing increasing mean transmural extent of myocardial infarction, with corresponding schematic overlays. Scar severity is assessed by mean transmural involvement, highlighting the need for spatially aware analysis. Figure source: [2], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.	3
1.4	Examples of patients with varying myocardial scar burden and ventricular function on CMR, demonstrating that differences in scar distribution are associated with distinct clinical outcomes despite similar global metrics. Figure source: [3], reproduced with permission of the rights holder, © Elsevier 2012.	4
2.1	(a) Illustration of sectional anatomy of the human heart (Image by Blausen Medical Communications, Inc. , via Wikimedia Commons, licensed under CC BY 3.0), (b) heart wall (Image by BruceBlaus. , via Wikimedia Commons, licensed under CC BY 3.0), and (c) Myocardial fiber orientation and direction of rotation (Figure source: [4], reproduced under the terms of the Creative Commons Attribution License CC BY 3.0).	12
2.2	Coronary circulation, with coronary arteries labeled in red text and other landmarks in blue text. Image by Patrick J. Lynch , via Wikimedia Commons, licensed under CC BY 3.0.	13
2.3	Illustration of the electrical conduction system of the heart (a) Image by Blausen Medical Communications, Inc. , via Wikimedia Commons, licensed under CC BY 3.0, (b) Image by Madhero88 , via Wikimedia Commons, licensed under CC BY 3.0.	13

2.4	An ECG showing pardee waves indicating acute myocardial infarction in the inferior leads II, III and aVF with reciprocal changes in the anterolateral leads. Image by Glenlarsen , via Wikimedia Commons (public domain).	20
2.5	Standard segmental myocardial representation using the AHA 17-segment model across basal, mid-cavity, and apical levels, with corresponding coronary territories. Figure source: [5], reproduced under the terms of the Creative Commons Attribution NonCommercial License CC BY-NC 4.0.	28
3.1	A schematic diagram of the framework proposed for myocardial scar segmentation. (1) End-to-End pipeline (orange region) takes $\mathbf{X} : \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\}$ as input and predicts $\hat{\mathbf{P}} : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}$ as output in one step. (2) Cascaded pipeline (green region), the first stage is a coarse segmentation network (CSN) that takes $\mathbf{X} : \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\}$ as input and predicts $\hat{\mathbf{P}}_1 : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}\}$ as output. The second stage is a fine segmentation network (FSN) that takes $\mathbf{X} \oplus \hat{\mathbf{P}}_1$ as input and predicts the final segmentation maps denoted by $\hat{\mathbf{P}}_2 : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}$, where $\oplus$ denotes channel-wise concatenation. Figure source: [6], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.	33
3.2	Visualizing predictions for three consecutive slices for a test CMR from MyOPS dataset: (row 1) LGE images, (row 2) and (row 3) predictions obtained with cascaded and end-to-end pipeline overlaid on LGE, respectively. Colored regions represent respective class LV , RV , MYO , edema and scar . Figure source: [6], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.	40
4.1	Overview of CLAIM. (a) Training of the conditional diffusion model using real pathological LGE images. (b) Generation of synthetic pathological images with diverse and anatomically consistent scar patterns using clinically guided masks produced by SMILE (Sec. 4.3.3). The generated images are used to fine-tune a scar segmentation network, whose feedback further improves synthesis quality through joint training. Figure source: [7], reproduced with permission of the rights holder, Springer Nature.	48
4.2	Overview of SMILE (Scar Masks Incorporating cLinical knowledgE). Figure source: [7], reproduced with permission of the rights holder, Springer Nature.	53
4.3	Visualization of real scar masks and synthetic masks generated by DiffMask [8] and SMILE (ours). Figure source: [7], reproduced with permission of the rights holder, Springer Nature.	58
4.4	Visualization of synthetic pathological images generated by different methods from normal images and scar masks. Difference images highlight intensity changes in scar regions. Figure source: [7], reproduced with permission of the rights holder, Springer Nature.	59
4.5	Scar segmentation performance across different synthetic data configurations on in-domain and out-of-domain datasets. Figure source: [7], reproduced with permission of the rights holder, Springer Nature.	61

4.6	Comparing the relative performance of our best method with baseline on EMIDEC test dataset. (a) A scatter plot showing dice score for each patient. (b) A box plot showing the distribution of dice scores.	62
4.7	Bull's-eye plots illustrating regional scar volume distribution across AHA 17 myocardial segments on the EMIDEC test dataset. Figure source: [7], reproduced with permission of the rights holder, Springer Nature.	63
5.1	Overview of the proposed (a) multi-modal network combining LGE-MRI with AHA-based spatial priors and ECG features via separate encoders, and (b) the Temporal-Aware Feature Fusion (TAFF) module, which adaptively integrates ECG and MRI representations modulated by the ECG–MRI acquisition interval. Figure source: [9], reproduced with permission of the rights holder, © 2026 IEEE.	74
5.2	Result Visualization: ground truth (a) and segmentation predictions of left ventricular blood pool , healthy myocardium , and scar overlaid on a 2D LGE slice using unimodal (image-only) (b) and the proposed multimodal approach (c). Scar volume distribution are also plotted on the AHA-17 segments accordingly (d-f); and cross-lead (g) and temporal lead-wise (h) attention maps for the corresponding ECG input using the ECG attention network to support model explainability. Figure source: [9], adapted with permission of the rights holder, © 2026 IEEE.	82
5.3	Distribution of Dice scores across time-interval bins, comparing models with and without time conditioning. Figure source: [9], reproduced with permission of the rights holder, © 2026 IEEE.	83
5.4	Predicted vs. ground-truth scar volume of test cases (baseline vs. multimodal).	87
5.5	Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of DICE SCORE	88
5.6	Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of PRECISION	88
5.7	Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of SENSITIVITY	89
5.8	Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of SPECIFICITY	89
5.9	Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of ABSOLUTE VOLUME DIFFERENCE	90

5.10	Qualitative example (1): Representative success case showing improved scar segmentation using the proposed TAFF multimodal framework. The AHA-17 bull's-eye plots illustrate reduced regional scar volume error for the multimodal prediction compared to the baseline. From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the left ventricular blood pool , healthy myocardium , and scar are visualized in their respective colors.	92
5.11	Qualitative example (4): Clear improvement case where multimodal fusion refines fragmented baseline predictions. Anatomical priors constrain predictions to physiologically plausible regions. From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the left ventricular blood pool , healthy myocardium , and scar are visualized in their respective colors.	93
5.12	Qualitative example (3): Improved boundary consistency and regional scar volume estimation achieved through ECG-guided fusion. The multimodal prediction aligns more closely with the ground truth across apical AHA segments. From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the left ventricular blood pool , healthy myocardium , and scar are visualized in their respective colors.	94
5.13	Qualitative example (2): TAFF Multimodal fusion corrects under-segmentation observed in the image-only baseline, particularly in mid-ventricular segments. ECG attention highlights clinically relevant leads associated with the affected myocardial regions. From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the left ventricular blood pool , healthy myocardium , and scar are visualized in their respective colors.	95
5.14	Qualitative example (5): Challenging case showing reduced scar volume error in selected AHA segments. ECG attention is weaker but still contributes to stabilizing segmentation. From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the left ventricular blood pool , healthy myocardium , and scar are visualized in their respective colors.	96

5.15	Qualitative example (6): Failure case illustrating limitations of multimodal fusion when both LGE contrast and ECG signals are ambiguous. While regional errors persist, anatomically implausible predictions are avoided due to the anatomical prior. From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the left ventricular blood pool , healthy myocardium , and scar are visualized in their respective colors.	97
6.1	Overview of the framework based on M ² FTrans for myocardial pathology segmentation from incomplete multi-sequence CMR. Complete and incomplete samples are jointly processed during training.	107
6.2	Illustration of data usage within a training batch. Samples with complete modalities contribute full supervision, while synthetically or naturally incomplete samples contribute incomplete-modality supervision.	108
6.3	Qualitative comparison of predictions under different incomplete modality combinations. Colored regions represent LV , RV , Myo , edema and scar	116
6.4	Qualitative comparison of supervised and semi-supervised training strategies. Mean T = mean teacher, RC = RandConv. Colored regions represent LV , RV , Myo , edema and scar	118

List of Tables

2.1	Reading map linking Chapter 2 sections to the contribution chapters where each is used. Background that is specific to a single contribution is presented in the corresponding chapter rather than here.	11
2.2	Public benchmark datasets used in this thesis. Detailed dataset splits and pre-processing pipelines are given in the experimental sections of the contribution chapters.	21
3.1	Results obtained on the MyOPS test dataset with models trained using the end-to-end (baseline) and the cascaded pipeline strategies with varied CMR modalities as input, where $Phase_1$ and $Phase_2$ represent CSN and FSN of CNN^{C2F} , respectively. Here, the standard loss: L_{DiceCE} in our experiments is used. Table source: [6], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.	37
3.2	Results obtained with MyOPS test dataset by training end-to-end (S_2) and cascaded (S_6) networks using different loss functions. Table source: [6], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.	39
4.1	Characteristics of the LGE-MRI datasets used in this study. Table source: [7], reproduced with permission of the rights holder, Springer Nature.	55
4.2	Scar segmentation results on EMIDEC dataset [10]. P: 57 real pathological cases. N': (28×2) synthetic pathological cases generated from 28 normal cases. N'': (28×5) synthetic pathological cases generated from 28 normal cases. P': (57×1) synthetic pathological cases generated from 57 real pathological cases. CLAIM: without joint training, CLAIM (J): with joint training. Table source: [7], reproduced with permission of the rights holder, Springer Nature.	60
5.1	Scar segmentation results on the test set: image-only baseline vs. proposed multimodal pipeline with module ablations. Best scores in bold. ($AP = Anatomical Prior$). All improvements over the baseline are statistically significant (paired t-test, $p < 0.001$). Table source: [9], reproduced with permission of the rights holder, © 2026 IEEE.	81
5.2	Comparison of different multimodal fusion strategies for ECG-MRI scar segmentation.	85
5.3	Scar segmentation performance comparison between the image-only baseline and the multimodal attention-based model.	86

6.1	MyoPS++ dataset: multi-center cohort with heterogeneous modality and annotation availability [11–14].	111
6.2	Performance of fully supervised models trained on centers B/C under simulated missing-modality cases.	113
6.3	Semi-supervised performance using all centers (including partially labeled data).	116
6.4	Performance on the MyoPS++ test set (center D) as reported by the evaluation platform.	119
6.5	Comparison of problem settings across existing methods.	121
6.6	Comparison of methodological strategies.	121
6.7	Performance comparison with consideration of problem complexity.	122
6.8	Comparison of strategies for handling missing modality and incomplete supervision.	123
7.1	Methodological progression of the thesis, showing how each chapter addresses a specific limitation of prior automated scar segmentation approaches and contributes toward clinically deployable multimodal AI.	135
7.2	Progressive increase in data realism and supervision complexity across thesis chapters, moving from controlled benchmark settings toward heterogeneous clinical scenarios.	135
7.3	Evolution of learning paradigms across the thesis, highlighting the transition from conventional supervised segmentation toward multimodal and robustness-oriented learning strategies.	136
7.4	Progression toward clinical translation across thesis chapters, illustrating how methodological developments improve interpretability, robustness, and practical applicability in real-world cardiac imaging workflows.	137
A.1	List of abbreviations used throughout the thesis.	157

Chapter 1

Introduction

The subject of this thesis is the development of robust and clinically applicable artificial intelligence (AI) methods for myocardial scar segmentation from cardiac magnetic resonance imaging (CMR). Given the clinical importance of accurately characterizing myocardial scarring, this chapter introduces the clinical context, technical challenges, and key contributions of the thesis.

1.1 Clinical Motivation

Myocardial infarction (MI), commonly known as a heart attack, occurs when coronary artery occlusion leads to prolonged ischemia and irreversible cardiomyocyte death. As illustrated in Figure 1.1a, blockage of a coronary vessel deprives downstream myocardium of oxygen, initiating a cascade of tissue necrosis and inflammatory repair. This process ultimately results in fibrotic scar formation that replaces viable myocardium with non-contractile tissue.

The structural consequences of this remodeling are shown schematically in Figure 1.1b, where infarction alters the geometry and mechanical integrity of the left ventricular wall. Unlike healthy myocardium, scar tissue cannot actively contribute to contraction and instead disrupts both mechanical synchrony and electrical conduction. These changes create a substrate for ventricular arrhythmias, adverse ventricular remodeling, and progressive heart failure [15, 16]. Consequently, accurate characterization of myocardial scar burden and spatial distribution is essential for risk stratification, therapy planning, and long-term prognostic assessment [15, 16].

Cardiac magnetic resonance (CMR), particularly late gadolinium enhancement (LGE) imaging, is regarded as the clinical reference standard for non-invasive visualization of infarcted myocardium [17, 18]. By exploiting differences in gadolinium contrast kinetics between viable and fibrotic tissue, LGE provides high spatial resolution and strong histopathological agreement for detecting myocardial fibrosis [18, 19]. As illustrated in Figure 1.1c, infarcted regions appear as *hyperenhanced areas* embedded within the myocardial wall, often exhibiting irregular geometry, diffuse boundaries and heterogeneous distribution across imaging planes—features that mirror the biological variability of infarct healing.

Late gadolinium enhancement imaging enables direct visualization of infarcted myocardium as regions of hyperenhancement caused by increased extracellular volume in fibrotic tissue. As shown in Figure 1.2, scar appearance varies substantially across imaging planes, with

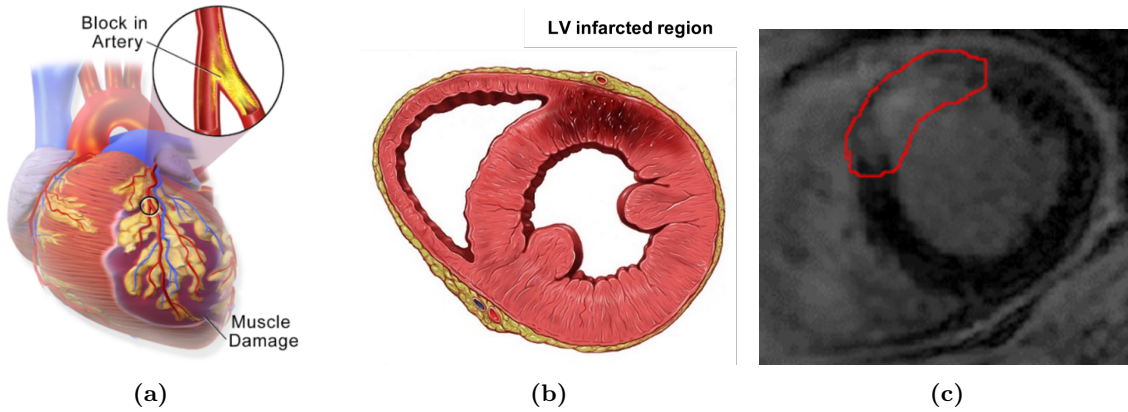


Figure 1.1: (a) Illustration of coronary artery blockage leading to myocardial infarction (Image by [Blaussen Medical Communications, Inc.](#), via Wikimedia Commons, licensed under CC BY 3.0), (b) cross-sectional view showing infarcted left ventricular myocardium (Image by [Patrick J. Lynch](#), via Wikimedia Commons, licensed under CC BY 2.5), and (c) representative LGE-CMR slice with segmented scar region (red).

differences in shape, thickness, and spatial continuity even within the same patient. This variability reflects the heterogeneous biological processes underlying infarct healing and illustrates why scar segmentation cannot rely solely on simple intensity-based assumptions.

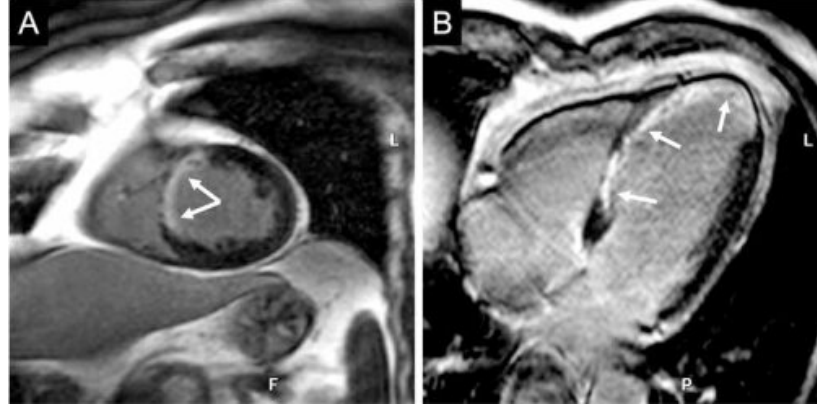


Figure 1.2: LGE-CMR images of chronic septal and apical myocardial infarction (arrows) in short-axis (A) and four-chamber (B) views, showing hyperenhanced regions with variable morphology and extent. Figure source: [1], reproduced under the terms of the Creative Commons Attribution License CC BY 2.5.

Clinical interpretation extends beyond detection to quantification of infarct transmuralty, defined as the proportion of wall thickness replaced by scar. Transmurality of the infarct correlates strongly with myocardial viability and recovery following revascularization [20, 21]. Figure 1.3 demonstrates how increasing transmural extent is associated with progressively reduced likelihood of contractile improvement, reinforcing that scar assessment must capture graded structural damage rather than perform binary classification.

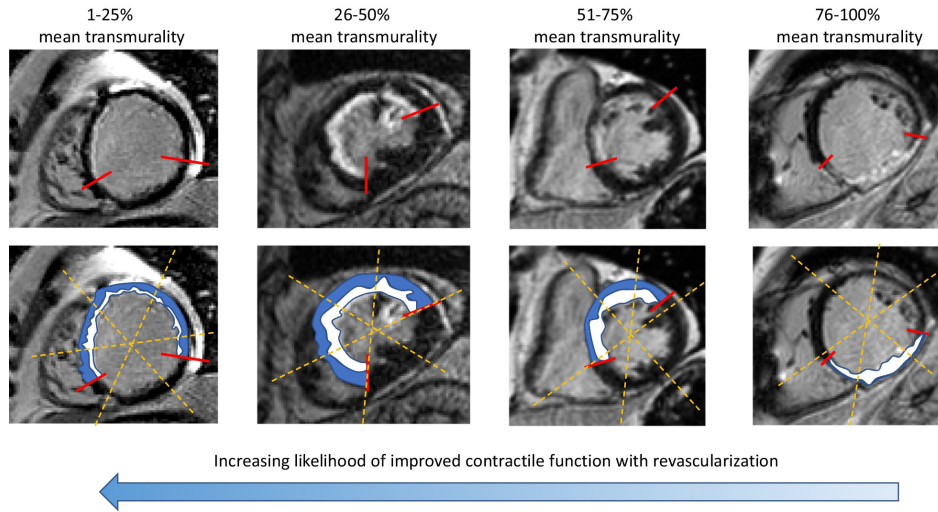


Figure 1.3: Representative LGE-CMR images from different patients showing increasing mean transmural extent of myocardial infarction, with corresponding schematic overlays. Scar severity is assessed by mean transmural involvement, highlighting the need for spatially aware analysis. Figure source: [2], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.

Despite the apparent visual contrast in LGE images, defining scar boundaries remains challenging in practice. Conventional semi-automated approaches rely on signal-intensity thresholding methods such as the standard deviation (SD) or full-width-at-half-maximum (FWHM) techniques. These methods depend on manually selected reference regions and empirically chosen thresholds, making them sensitive to noise, acquisition variability, and observer input, and leading to substantial inter-study variability [22, 23]. Such heuristic approaches treat scar as a purely intensity-driven phenomenon, ignoring the broader anatomical and physiological context visible in the images.

The clinical implications of accurate scar characterization are highlighted in Figure 1.4, where patients with differing scar distributions demonstrate markedly different functional outcomes and arrhythmic risks despite similar global metrics such as ejection fraction. Numerous studies have shown that scar extent and heterogeneity provide incremental prognostic value beyond conventional functional indices [3, 24, 25]. These examples emphasize that the spatial pattern of fibrosis—not just its presence—plays a critical role in determining patient-specific risk.

Taken together, these observations reveal both the diagnostic power and the analytical bottleneck of LGE-CMR. While it provides uniquely rich insight into myocardial pathology, extracting reproducible quantitative information remains labor-intensive and methodologically fragile. Automating myocardial scar segmentation is therefore not merely an efficiency improvement but a prerequisite for scalable, objective, and clinically actionable cardiac imaging analysis. The heterogeneity, graded severity, and context dependence evident across Figures 1.2–1.4 underscore that myocardial scar segmentation is fundamentally more complex than conventional anatomical segmentation tasks, motivating the need for computational approaches that can operate reliably under realistic clinical conditions.

Crucially, these challenges arise not only from intrinsic image complexity but also from

Risk	Delayed-enhancement CMR		Cine CMR diastole systole		Follow-up
Patient A Risk Concordance: - LVEF 33% - Scar 22%					ICD Shock for VT
Patient B Risk Discordance: - LVEF 28% - No Scar					No events
Patient C Risk Discordance: - LVEF 46% - Scar 17%					Sudden Cardiac Death at 18 months of follow-up
Patient D Risk Discordance: - LVEF 49% - Scar 13%					ICD Shock for VT

Figure 1.4: Examples of patients with varying myocardial scar burden and ventricular function on CMR, demonstrating that differences in scar distribution are associated with distinct clinical outcomes despite similar global metrics. Figure source: [3], reproduced with permission of the rights holder, © Elsevier 2012.

variability in acquisition protocols, patient populations, and data completeness across clinical settings, highlighting the need for segmentation methods that are robust to real-world heterogeneity rather than optimized for controlled conditions.

1.2 Technical Problem Definition

Automated myocardial scar segmentation aims to identify and delineate infarcted tissue within the left ventricular myocardium from CMR data. In principle, this can be formulated as a voxel-wise classification problem, where each voxel is labeled as viable or infarcted myocardium. Convolutional encoder-decoder architectures such as U-Net [26] and its self-configuring extension nnU-Net [27] have achieved state-of-the-art performance across a wide range of biomedical segmentation tasks, including cardiac imaging [12, 28, 29].

However, myocardial scar segmentation differs fundamentally from classical organ segmentation and presents several domain-specific challenges that complicate direct application of these methods.

- **Severe class imbalance:** Infarct regions typically occupy only a small proportion of the myocardium, often less than 10–20% of the ventricular volume. This imbalance

biases learning toward the dominant healthy class and can lead to under-segmentation of clinically important scar regions [30].

- **High morphological variability:** Scar appearance varies widely in location, transmural extent, and spatial continuity depending on infarct territory and healing dynamics [31, 32]. Unlike anatomical structures with consistent geometry, infarcts lack a stable shape prior, making model generalization difficult.
- **Ambiguous tissue boundaries:** Scar–myocardium interfaces are often indistinct due to partial volume effects, signal inhomogeneity, motion artifacts, and variability in inversion time selection during LGE acquisition [19, 33]. These factors produce gradual intensity transitions rather than sharp edges, violating assumptions commonly exploited in segmentation networks.
- **Protocol and domain variability:** Differences in scanner vendors, field strengths, acquisition parameters, and reconstruction pipelines substantially alter image appearance across institutions [34, 35]. Models trained on single-center datasets often fail to generalize under such distribution shifts, a well-recognized limitation in medical AI [36, 37].
- **Limited annotated datasets:** High-quality scar annotations require expert manual segmentation and are therefore scarce and expensive to obtain. This restricts dataset size and diversity, increasing susceptibility to overfitting and reducing robustness to rare infarct phenotypes [38, 39].
- **Multimodal pathophysiology:** Infarct remodeling is not purely structural; it also manifests through electrophysiological alterations routinely captured by electrocardiography (ECG). Clinical interpretation integrates both imaging and electrical information when localizing infarction, yet most automated approaches rely solely on imaging. This highlights the inherently multimodal nature of the problem, requiring integration of heterogeneous anatomical and functional signals [40–42].

Collectively, these challenges indicate that myocardial scar segmentation is fundamentally a problem of robust inference under real-world variability, rather than a conventional image segmentation task defined under controlled conditions. Instead of addressing each limitation in isolation, this thesis interprets them as manifestations of a broader requirement for robustness across heterogeneous data sources, incomplete modalities, and domain shifts.

Within this unified perspective, limited structural contrast and ambiguous boundaries motivate the integration of complementary multi-sequence CMR information (Chapter 3). Data scarcity and underrepresentation of rare infarct phenotypes are addressed through clinically guided scar synthesis (Chapter 4). The multimodal nature of infarct remodeling motivates ECG–CMR fusion mechanisms (Chapter 5). Finally, variability in acquisition protocols and incomplete multimodal datasets necessitate domain-resilient learning strategies capable of maintaining performance under realistic clinical conditions (Chapter 6).

Through this structured progression, the thesis systematically transitions from methods designed for idealized settings toward a unified framework for robust and clinically deployable myocardial scar segmentation.

1.3 Research Gap

Despite rapid advances in deep learning for medical image analysis, existing approaches to myocardial scar segmentation remain constrained by simplifying assumptions that do not reflect the realities of clinical imaging workflows. Many methods are developed and evaluated on curated datasets acquired under controlled conditions, implicitly assuming consistent imaging protocols, complete data availability, and well-represented pathology. In practice, however, cardiac imaging is heterogeneous, multimodal, and often incomplete. The following limitations highlight the disconnect between research settings and real-world deployment.

- Single-Sequence Reliance:** A large proportion of automated scar segmentation methods operate exclusively on LGE images, leveraging their strong contrast between viable and infarcted myocardium [22]. While effective for visualizing fibrosis, LGE alone does not capture complementary structural and functional information available in other routinely acquired sequences such as cine CMR or parametric mapping. These additional sequences provide context regarding myocardial motion, wall thickness, and anatomical boundaries, which can aid interpretation in regions where LGE contrast is ambiguous. Reliance on a single sequence therefore limits the model’s ability to disambiguate subtle pathology and reduces robustness to acquisition variability [28, 43]. Clinically, interpretation is rarely performed in isolation; radiologists integrate multiple sequences to form a coherent assessment, an integration that most automated methods fail to emulate.
- Data Scarcity and Pathological Imbalance:** High-quality annotations of myocardial scar require expert segmentation and are therefore scarce. Public datasets remain relatively small compared to those available in other domains of computer vision, and they frequently underrepresent uncommon infarct phenotypes or complex remodeling patterns. Conventional augmentation strategies—such as geometric transformations, cropping, or intensity perturbations—can increase apparent dataset size but do not introduce new pathological variation [44]. As a result, models may generalize poorly to rare but clinically important cases. Generative modeling approaches, including generative adversarial networks (GANs) [45] and diffusion-based methods [46], offer mechanisms for synthesizing realistic data distributions. However, most existing implementations emphasize visual realism rather than embedding clinically meaningful constraints, leaving pathology-aware synthesis insufficiently explored in cardiac applications.
- Lack of Multimodal Physiological Integration:** Myocardial infarction is a structural and electrophysiological disease. While CMR visualizes anatomical consequences, ECG captures complementary information about conduction abnormalities and infarct localization routinely used in clinical diagnosis [47, 48]. Multimodal machine learning frameworks provide principled strategies for integrating heterogeneous signals [41, 49]. Nevertheless, the majority of scar segmentation pipelines remain image-centric, effectively discarding physiological information that could help resolve uncertainty in imaging-based inference. This disconnect contrasts with clinical reasoning, where imaging and electrophysiology are interpreted jointly rather than independently.
- Poor Generalization Across Sites and Incomplete Data:** Medical AI systems often experience performance degradation when applied outside their training domain,

a phenomenon widely attributed to dataset shift and domain variability [37]. Differences in scanner vendors, field strengths, acquisition protocols, and reconstruction pipelines alter intensity distributions and spatial characteristics, challenging models trained on homogeneous datasets [34, 35]. Domain generalization strategies, including meta-learning and invariant feature learning [50], have been proposed to mitigate these effects. However, myocardial scar segmentation studies are frequently validated within single-center cohorts, limiting their ability to accommodate the variability encountered in multi-center clinical practice [51]. Additionally, real-world datasets are often incomplete, with missing sequences or asynchronous multimodal acquisitions, conditions rarely modeled explicitly during training.

Taken together, these limitations reveal a central issue: automated myocardial scar segmentation has largely been studied under assumptions that simplify the problem but constrain clinical applicability. Many approaches implicitly rely on clean data, consistent acquisition protocols, complete modality availability, and sufficiently represented pathology. Relaxing these assumptions, and designing methods that remain robust when they no longer hold, is essential for translating automated scar analysis from controlled research environments into reliable, real-world clinical deployment.

1.4 Central Thesis Hypothesis

This thesis advances the hypothesis that robust and clinically scalable myocardial scar segmentation is fundamentally a problem of learning under real-world variability, rather than optimizing performance under idealized conditions. Reliable deployment in clinical settings requires a unified framework that integrates complementary imaging information, incorporates clinically grounded pathological priors, leverages multimodal physiological signals, and remains resilient to domain shift and incomplete data.

While automated scar segmentation has demonstrated promising results in controlled environments, much of the existing literature relies on simplifying assumptions regarding data completeness, protocol consistency, and pathological representation. This thesis contends that such assumptions limit translational applicability. By systematically relaxing these constraints, it seeks to reconstruct the segmentation pipeline to function under the heterogeneity, incompleteness, and variability inherent to clinical practice.

Specifically, this work hypothesizes that:

- Integrating complementary multi-sequence CMR provides richer structural context and improves segmentation of infarcted myocardium compared to single-sequence approaches.
- Embedding clinically informed anatomical and morphological priors into generative augmentation frameworks enables realistic expansion of underrepresented scar phenotypes, improving generalization beyond conventional augmentation strategies [45, 46].
- Fusing electrophysiological information from ECG [47, 48] with imaging features captures structure–function relationships that are inaccessible to unimodal models and enhances patient-specific scar localization [49].

- Explicit modeling of domain variability and modality incompleteness improves robustness and stability of segmentation performance across heterogeneous, multi-center clinical datasets [37, 51].

Collectively, these hypotheses propose that automated myocardial scar segmentation should be reframed not as a static image analysis problem, but as a multimodal and domain-aware inference task grounded in clinical anatomy and physiology. The thesis evaluates this proposition through a structured sequence of methodological developments designed to bridge the gap between research-grade performance and clinically deployable systems.

1.5 Thesis Outline and Contributions

This thesis is structured as a methodological progression in which each chapter addresses a key limitation identified in existing approaches, moving incrementally toward a segmentation framework that reflects the realities of clinical data.

Chapter 2: Background. This chapter provides the clinical, imaging, and computational background relevant to myocardial scar segmentation. It reviews cardiac anatomy and physiology, the mechanisms of myocardial infarction and scar formation, and the principles of LGE-CMR imaging. In addition, it summarizes existing datasets, traditional and deep learning-based segmentation methods, and key challenges including data imbalance, multimodal integration, and domain variability. This foundation contextualizes the limitations of current approaches and motivates the methodological developments presented in subsequent chapters.

Chapter 3: From Limited Contrast to Multi-Sequence Integration. The thesis begins by revisiting the assumption that scar segmentation can be performed reliably using a single imaging sequence. Building on established architectures such as U-Net [26] and nnU-Net [27], this chapter investigates how the integration of complementary CMR sequences enhances structural context, improves boundary segmentation, and reduces ambiguity in regions where LGE contrast alone is insufficient.

Chapter 4: From Data Scarcity to Clinically-Guided Synthesis. Recognizing that annotated infarct datasets are both limited and pathologically imbalanced, Chapter 4 introduces a clinically informed scar synthesis framework. Drawing on advances in generative modeling [45, 46], the approach embeds anatomical and pathological priors to generate realistic infarct morphologies, enabling controlled expansion of underrepresented scar patterns and improving model generalization.

Chapter 5: From Image-Only Analysis to Multimodal Structure–Function Fusion. While infarction manifests structurally in imaging, it also produces measurable electrophysiological changes. Chapter 5 therefore extends the segmentation paradigm beyond imaging alone by integrating ECG [47, 48] and LGE-CMR data using a temporally aware fusion mechanism grounded in multimodal learning principles [49]. This enables the model to capture structure–function relationships more consistent with clinical reasoning.

Chapter 6: From Controlled Datasets to Domain-Resilient Learning. Finally, the thesis addresses the challenge of robustness under real-world clinical variability. Motivated by dataset shift, heterogeneous acquisition protocols, and incomplete multimodal data [37, 51], Chapter 6 develops domain-resilient learning strategies that enable stable and reli-

able segmentation performance across multi-center datasets and missing modality scenarios, reflecting the conditions encountered in practical clinical deployment.

Through this sequence, myocardial scar segmentation is progressively reframed from a constrained image analysis task into a clinically grounded, multimodal, and generalizable inference problem, aligning methodological development with the practical requirements of real-world cardiac imaging.

Overall, the thesis reframes myocardial scar segmentation as a problem of robust, multimodal, and clinically grounded inference, where methodological design is driven not by idealized assumptions but by the variability and incompleteness inherent to real-world cardiac imaging.

1.6 Publications Arising from This Thesis

A list of published works is given below in chronological order:

- Ramzan, Farheen, Chen (Cherise) Chen, and Richard H. Clayton. "Improving Myocardial Scar Segmentation with End-to-End and Cascaded CNN using Hybrid Loss and Multi-Modality CMR Imaging." *2024 Computing in Cardiology Conference (CinC)*. IEEE, 2024.
- Ramzan, Farheen, et al. "CLAIM: Clinically-guided LGE augmentation for realistic and diverse myocardial scar synthesis and segmentation." *International Conference on AI in Healthcare*. Cham: Springer Nature Switzerland, 2025.
- Ramzan, Farheen, et al. "Seeing Beyond the Image: ECG and Anatomical Knowledge-Guided Myocardial Scar Segmentation from Late Gadolinium-Enhanced Images." *Accepted for oral presentation at the IEEE International Symposium on Biomedical Imaging (ISBI 2026), arXiv preprint arXiv:2511.14702*, 2026.

Chapter 2

Background

2.1 Overview and Scope of This Chapter

Automated analysis of myocardial scar sits at the intersection of *cardiovascular biology*, *medical imaging*, and *machine learning*. Understanding why this problem remains challenging requires more than a description of segmentation algorithms; it demands an appreciation of the disease processes being measured, the physical principles underlying image formation, the characteristics of available clinical data, and the assumptions embedded within computational models. This chapter establishes the multidisciplinary foundations necessary for the work presented in the remainder of the thesis.

The chapter is intentionally scoped to material that recurs across more than one contribution chapter. Background that is specific to a single contribution is presented at the start of the corresponding chapter, so that each contribution chapter is self-contained and can be read on its own. A reader interested in a particular contribution should therefore be able to treat this chapter as a reference rather than as a linear prerequisite.

The chapter is organised into four groups. Sections 2.2–2.5 cover the *clinical foundations*: cardiac anatomy, the pathophysiology of myocardial infarction and post-infarction scar formation, and the clinical importance of accurate scar quantification. Sections 2.6–2.9 cover the *imaging and functional modalities* used in this thesis: the physical basis of cardiac magnetic resonance imaging, late gadolinium enhancement, complementary CMR sequences, and the electrocardiogram as a functional marker of infarction. Sections 2.10–2.15 cover the *computational background*: publicly available datasets, traditional image analysis, deep learning approaches to cardiac segmentation, and the loss functions used to train segmentation networks under class imbalance. Sections 2.16–2.17 cover the *evaluation background*: segmentation metrics and the AHA 17-segment model used for anatomically meaningful reporting.

Table 2.1 provides a reading map that links each contribution chapter to the sections of this chapter that are most relevant to it. The contributions of the thesis, presented in subsequent chapters, are positioned as responses to the challenges identified across these foundations, aiming to advance myocardial scar segmentation toward robust, multimodal, and clinically deployable solutions.

Table 2.1: Reading map linking Chapter 2 sections to the contribution chapters where each is used. Background that is specific to a single contribution is presented in the corresponding chapter rather than here.

Contribution chapter	Primary Chapter 2 sections	Additional back-ground
Ch. 3: Scar Segmentation from Multi-Sequence CMR	2.2, 2.4, 2.7, 2.12, 2.15, 2.16	—
Ch. 4: Clinically-Guided Scar Synthesis in LGE	2.4, 2.7, 2.12, 2.16	In-chapter: generative modelling
Ch. 5: Multimodal ECG-LGE Scar Segmentation	2.2, 2.7, 2.9, 2.12	In-chapter: multi-modal learning
Ch. 6: Robust Scar Segmentation from Incomplete CMR	2.2, 2.7, 2.8, 2.12	In-chapter: multi-centre variability

2.2 Cardiac Anatomy and Physiology Relevant to Imaging

Quantitative assessment of myocardial scar requires an understanding of the anatomical and physiological organization of the heart structures that determine how infarction develops and how pathology appears in both cardiac magnetic resonance and electrocardiography. This section focuses on features directly influencing infarct formation, scar morphology, and their measurable signatures.

2.2.1 Left Ventricular Structure and Myocardial Fiber Organization

The left ventricle (LV) is the principal chamber affected by ischemic injury and therefore the primary focus of scar analysis. Its wall is composed of a three-dimensional continuum of cardiomyocytes arranged in helically oriented fiber layers that rotate gradually from a right-handed helix at the endocardium to a left-handed helix at the epicardium [52, 53].

This structural organization strongly shapes infarct morphology. Following coronary occlusion, necrosis progresses as a **wavefront phenomenon** from the subendocardium toward the epicardium, producing scars with variable transmural extent rather than uniform wall involvement [54]. Because fibers are obliquely oriented, the transition between viable and infarcted tissue is rarely sharp in imaging, contributing to partial-volume effects and ambiguous scar boundaries in LGE-CMR [19]. These features complicate voxel-level classification and make scar segmentation fundamentally more difficult than segmentation of compact anatomical organs.

2.2.2 Coronary Perfusion Territories

Myocardial viability is governed by coronary arterial supply. The left anterior descending (LAD), left circumflex (LCx), and right coronary artery (RCA) perfuse distinct but partially overlapping myocardial territories. Infarction therefore follows vascular distributions rather than geometric shapes, producing crescentic or patchy regions aligned with perfusion beds [55].

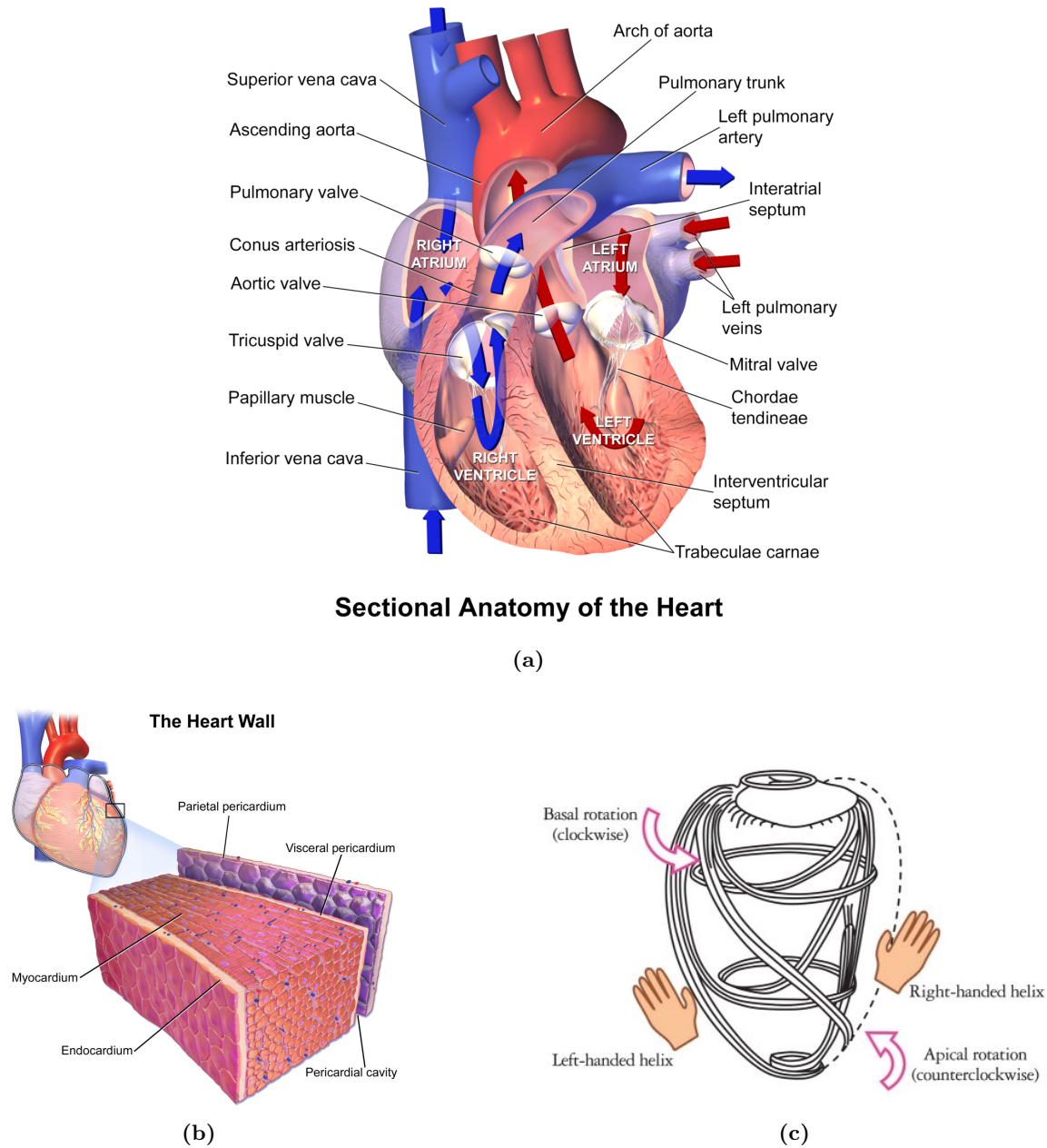


Figure 2.1: (a) Illustration of sectional anatomy of the human heart (Image by [Blaussen Medical Communications, Inc.](#), via [Wikimedia Commons](#), licensed under [CC BY 3.0](#)), (b) heart wall (Image by [BruceBlaus.](#), via [Wikimedia Commons](#), licensed under [CC BY 3.0](#)), and (c) Myocardial fiber orientation and direction of rotation (Figure source: [4], reproduced under the terms of the [Creative Commons Attribution License CC BY 3.0](#)).

This vascular dependence introduces large inter-patient variability in scar location, size, and geometry. Even infarctions affecting the same artery may differ substantially depending on collateral flow, occlusion duration, and reperfusion timing [18]. For computational

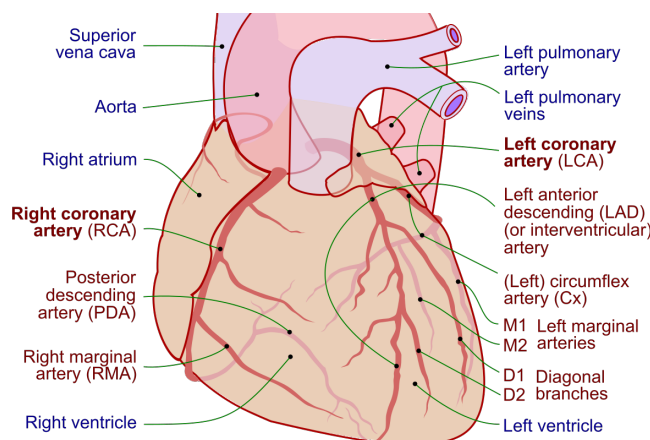


Figure 2.2: *Coronary circulation, with coronary arteries labeled in red text and other landmarks in blue text. Image by [Patrick J. Lynch](#), via Wikimedia Commons, licensed under CC BY 3.0.*

analysis, this variability limits the usefulness of rigid anatomical priors.

2.2.3 Electrical Conduction System and the His–Purkinje Network

Electrical activation originates in the sinoatrial node and propagates through the atrioventricular node into the His–Purkinje system, enabling rapid and coordinated ventricular depolarization [56].

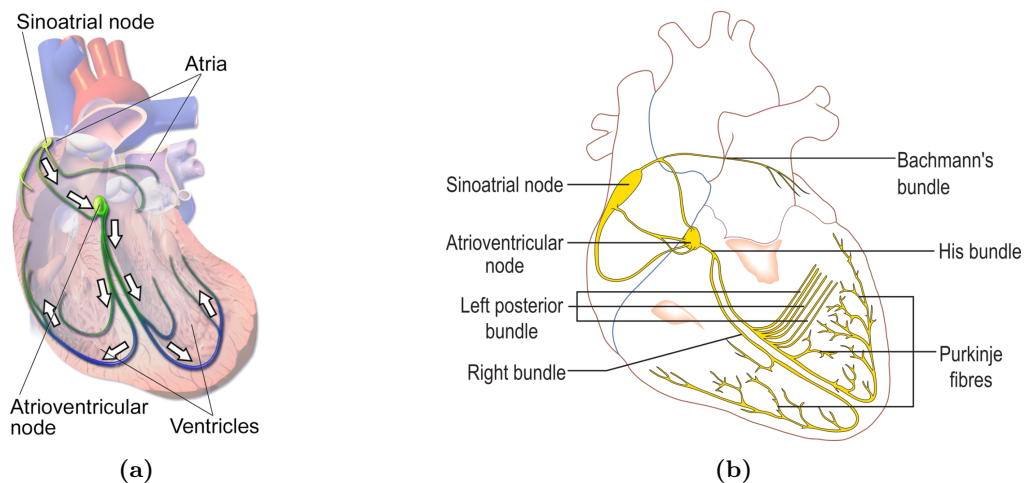


Figure 2.3: *Illustration of the electrical conduction system of the heart (a) Image by [Blaussen Medical Communications, Inc.](#), via Wikimedia Commons, licensed under CC BY 3.0, (b) Image by [Madhero88](#), via Wikimedia Commons, licensed under CC BY 3.0.*

Following myocardial infarction, fibrotic replacement disrupts conduction pathways and creates zones of slow conduction interspersed with surviving myocyte bundles. These hetero-

geneous regions can form the substrate for re-entrant ventricular tachycardia, where electrical wavefronts circulate around scar tissue [57]. Such electrophysiological abnormalities may not be directly visible in structural imaging but manifest in ECG signals, underscoring the complementary nature of anatomical and functional assessment.

2.2.4 Structure-Function Coupling in Normal and Infarcted Myocardium

In the healthy heart, myocardial architecture, perfusion, electrical activation and contraction are tightly coupled, producing predictable relationships between structure and function. Infarction disrupts this coupling: fibrotic tissue eliminates contractility while simultaneously altering conduction and mechanical stress distribution [15].

For imaging-based analysis, this decoupling presents a central challenge. Regions with similar appearance on CMR may differ in electrophysiological significance or arrhythmic risk, meaning that structural imaging alone provides an incomplete representation of disease. This limitation motivates the integration of ECG-derived functional information with imaging data, forming the conceptual basis for multimodal approaches explored later in this thesis.

2.3 Cardiac Arrhythmias and the Structural Substrate of Disease

Myocardial infarction is not solely a structural injury but also an electrophysiological disorder. Following ischemic necrosis, the replacement of viable myocardium with fibrotic tissue disrupts normal electrical propagation, creating a substrate for ventricular arrhythmias, particularly ventricular tachycardia (VT), a major cause of sudden cardiac death in ischemic heart disease [58, 59].

In the post-infarction ventricle, sustained monomorphic VT is most commonly driven by re-entry, in which electrical wavefronts circulate through regions of slow conduction surrounding non-conductive scar tissue [57, 60]. Surviving strands of myocytes within fibrotic regions can form conducting channels that connect areas of viable myocardium, allowing activation to re-enter previously excited tissue and sustain arrhythmic circuits [61]. These circuits are inherently dependent on the geometry and spatial continuity of scar, not simply its total volume.

Post-infarction tissue is rarely composed of dense scar alone. A heterogeneous border region containing mixtures of fibrosis and viable myocardium, often referred to as the gray zone, exhibits altered conduction properties and has been strongly associated with arrhythmia susceptibility and adverse clinical outcomes [3, 62, 63]. Because this tissue represents a continuum rather than a discrete boundary, its segmentation is intrinsically uncertain in imaging, contributing to variability in both manual and automated segmentation [19].

The clinical relevance of myocardial scar therefore lies not only in its presence but in its spatial organization. Scar distribution and heterogeneity provide prognostic information beyond global functional measures such as ejection fraction and are increasingly used to guide catheter ablation and risk stratification [64, 65]. This dual role — as both structural abnormality and electrical substrate — motivates approaches that integrate anatomical imaging with functional information, forming the basis for multimodal analysis strategies developed later in this thesis.

2.4 Myocardial Infarction and Post-Infarction Scar Formation

This section describes the pathological processes underlying myocardial infarction and subsequent scar formation, providing context for the tissue characteristics that automated segmentation methods must identify. The morphology observed in cardiac magnetic resonance is the end result of a dynamic biological sequence rather than a static lesion, and this evolution directly influences imaging appearance and segmentation complexity.

2.4.1 Scar Formation

Myocardial infarction begins with prolonged ischemia caused by coronary artery occlusion, which deprives cardiomyocytes of oxygen and metabolic substrates. Sustained ischemia leads to irreversible cellular injury and necrosis, typically progressing from the subendocardium toward the epicardium in a wavefront pattern [54]. This acute phase is followed by an inflammatory response, during which immune cells clear necrotic debris and initiate tissue repair. Over subsequent weeks, fibroblast activation and extracellular matrix deposition replace the damaged myocardium with collagen-rich fibrotic tissue, forming a permanent scar [66].

2.4.2 Dense Scar and the Gray Zone

The healed infarct rarely consists of homogeneous fibrosis alone. Dense non-viable scar is often surrounded by a transitional region containing interstitial fibrosis and surviving myocytes, commonly referred to as the gray zone or border zone. This heterogeneous tissue exhibits altered mechanical and electrical properties and has been associated with adverse remodeling and arrhythmia susceptibility [19, 62]. From an imaging perspective, this continuum produces gradual signal transitions rather than discrete tissue classes.

2.4.3 Remodeling and Ventricular Dilation

Beyond the infarcted region, myocardial infarction triggers global ventricular remodeling characterized by wall thinning at the infarct site, dilation of the ventricular cavity, and compensatory changes in remote myocardium [15]. These structural adaptations alter ventricular geometry and contribute to patient-specific variability in scar shape and distribution.

2.4.4 Why Scar Morphology Is Heterogeneous

Because scar formation reflects vascular territories, healing dynamics, and remodeling processes, infarcts are highly variable in size, transmural, and spatial configuration. They often appear irregular, patchy, or partially transmural rather than forming compact, well-defined regions. This biological variability explains why myocardial scar cannot be reliably characterized using simple geometric assumptions.

2.4.5 Implications for Imaging and Segmentation

The pathological evolution of infarction leads directly to several challenges observed in CMR-based analysis. Signal characteristics represent a spectrum of tissue states rather than a binary distinction between healthy and infarcted myocardium; scar boundaries are inherently

diffuse due to partial-volume effects and heterogeneous composition; and ventricular remodeling introduces inter-patient anatomical variability that complicates spatial standardization. Together, these factors make myocardial scar segmentation substantially more difficult than segmentation of well-defined anatomical structures, and explain why late gadolinium enhancement imaging reveals complex intensity patterns that computational approaches must accommodate through models robust to ambiguity, heterogeneity, and variability in scar representation.

2.5 Clinical Importance of Scar Quantification and Localization

The clinical relevance of myocardial scar extends far beyond its visual detection. Scar burden, distribution, and heterogeneity carry prognostic and therapeutic implications that directly influence patient management. Accurate and reproducible quantification is therefore not merely a technical objective, but a clinical necessity.

2.5.1 Risk Stratification for Sudden Cardiac Death

Sudden cardiac death in ischemic heart disease is frequently driven by ventricular arrhythmias arising from scar-related substrates. While left ventricular ejection fraction (LVEF) has historically been used to guide implantable cardioverter-defibrillator (ICD) therapy, it provides only a global measure of systolic function and does not directly capture arrhythmogenic substrate [59, 67]. Multiple studies have demonstrated that the extent and heterogeneity of myocardial scar detected by LGE-CMR provide incremental prognostic value beyond LVEF alone, with gray-zone burden and scar distribution in particular associated with ventricular tachyarrhythmias and adverse outcomes [3, 64]. Spatial scar characterization may therefore improve identification of patients at risk for life-threatening arrhythmias.

2.5.2 Guidance for Catheter Ablation

In patients with sustained ventricular tachycardia, catheter ablation targets critical components of re-entrant circuits often located within or adjacent to scar tissue, and pre-procedural imaging can assist in identifying potential conducting channels and defining substrate boundaries [65, 68]. The success of substrate-based ablation depends on accurate spatial representation of scar geometry: overestimation may lead to unnecessary ablation of viable tissue, whereas underestimation risks incomplete substrate modification. Reliable segmentation is therefore essential for integrating imaging with electroanatomical mapping systems and for supporting precision-guided interventions.

2.5.3 Prediction of Functional Recovery

The transmural extent of infarction correlates strongly with the likelihood of contractile recovery following revascularization [18, 20]. Segments with limited transmural involvement often regain function, whereas fully transmural scars are unlikely to recover, so quantifying scar at a regional level informs decisions regarding revascularization and viability assessment.

Accurate segmentation of infarct borders is particularly important when assessing partially viable myocardium.

2.5.4 Limitations of Ejection Fraction as a Biomarker

Although LVEF remains central to clinical guidelines, it reflects global pump function rather than regional structural pathology, and patients with similar ejection fractions may exhibit markedly different scar burdens and arrhythmic risk profiles [3]. LVEF may also improve or decline due to factors unrelated to scar extent, such as loading conditions or medical therapy. Structural scar characterization provides complementary information that may better reflect arrhythmogenic potential and long-term prognosis.

2.5.5 Need for Reproducible Spatial Quantification

Despite its clinical value, scar assessment is often performed manually or semi-manually, introducing inter-observer variability and limiting scalability [19]. Differences in threshold selection, contouring practices, and interpretation of ambiguous regions can substantially alter reported scar burden and transmural extent. For scar quantification to meaningfully inform clinical decisions—particularly those involving device implantation or invasive procedures—it must be reproducible, standardized, and spatially precise, and automated methods offer the potential to reduce subjectivity, improve consistency across centers, and enable large-scale risk modeling.

In summary, myocardial scar quantification influences risk stratification, procedural planning, and therapeutic decision-making. Its clinical importance lies not only in detecting fibrosis, but in accurately representing its spatial distribution and heterogeneity, providing the central clinical justification for the development of automated, multimodal, and generalizable scar segmentation frameworks pursued in this thesis.

2.6 Fundamentals of Magnetic Resonance Imaging

Cardiac magnetic resonance provides a non-invasive means of characterizing myocardial structure, function, and tissue composition. Unlike imaging modalities based on ionizing radiation, MRI exploits the magnetic properties of hydrogen protons to generate contrast that reflects underlying tissue characteristics. In the presence of a strong external magnetic field, radiofrequency pulses perturb proton alignment, and the subsequent relaxation of protons back to equilibrium produces signals whose amplitude and timing depend on tissue composition and molecular environment [69, 70]. Two relaxation mechanisms dominate image contrast: T1 (longitudinal recovery) and T2 (transverse decay). Different tissues exhibit distinct T1 and T2 properties, allowing MRI to generate contrast sensitive to edema, fibrosis, fat content, and other pathological changes. Spatial encoding is achieved through magnetic field gradients, and the acquired signals are mapped into frequency space (k-space) and reconstructed into images via Fourier transformation [71]. As a result, MR images are not direct photographs of anatomy but reconstructions shaped by both tissue properties and acquisition choices.

Cardiac imaging presents additional technical challenges due to the heart’s continuous motion and the need for synchronization with the cardiac cycle. Electrocardiographic gating

is typically used to acquire data at consistent phases of the heartbeat; in practice, a volumetric CMR dataset is constructed as a stack of two-dimensional slices, each acquired over several heartbeats, with only partial data captured per cycle and subsequently combined into a complete image or volume. Adjacent slices within a stack may therefore originate from different cardiac cycles, introducing inconsistencies due to beat-to-beat variability, arrhythmias, or imperfect gating. Breath-holding or motion correction strategies are employed to reduce respiratory artifacts [72], yet residual motion and physiological variability remain.

The physics of MRI directly influences how myocardial scar is represented computationally. Image intensities are not absolute physical quantities but depend on sequence parameters and reconstruction choices, and identical tissue types may exhibit different signal characteristics across scanners or protocols. Variability encountered in automated scar segmentation is therefore rooted not only in pathology but also in measurement physics, and the resulting distribution shifts across scanners and centres motivate the domain-aware learning strategies discussed later in this chapter.

2.7 Late Gadolinium Enhancement Imaging (LGE)

LGE imaging is the principal CMR technique used to visualize myocardial infarction and fibrosis, and it serves as the reference standard for non-invasive scar assessment in clinical practice. LGE exploits differences in contrast agent distribution between healthy and damaged myocardium to highlight regions of irreversible injury.

2.7.1 Contrast Mechanism and Tissue Characterization

Gadolinium-based contrast agents are extracellular and do not enter intact cardiomyocytes. In normal myocardium, gadolinium rapidly washes in and out through the vascular and interstitial spaces. In infarcted or fibrotic tissue, disruption of cell membranes and expansion of extracellular volume allow greater accumulation and delayed washout of contrast agent [18, 73]. This shortens T1 relaxation time in scarred regions, causing them to appear hyper-enhanced (bright) relative to nulled normal myocardium on inversion-recovery sequences. By carefully selecting the inversion time to suppress signal from healthy tissue, LGE provides strong contrast between viable and non-viable myocardium.

2.7.2 Imaging Protocol and Acquisition Considerations

LGE images are typically acquired 10–20 minutes after contrast administration to allow sufficient redistribution of gadolinium. Inversion-recovery gradient echo sequences are used to maximize T1-weighted contrast, and imaging is synchronized to the cardiac cycle to minimize motion artifacts [74]. Despite standardized principles, acquisition parameters (including inversion time selection, spatial resolution, and sequence implementation) vary across scanners and institutions, influencing intensity appearance and contrast-to-noise ratio and contributing to inter-study variability.

2.7.3 Strengths and Limitations of LGE for Scar Assessment

LGE-CMR has demonstrated strong correlation with histopathology and is widely used for assessing infarct presence, extent, and transmural extent [18]. Its high spatial resolution and three-dimensional visualization make it particularly valuable for regional scar analysis. However, several characteristics complicate quantitative interpretation: signal intensity represents a continuum influenced by tissue heterogeneity rather than discrete classes; partial-volume effects blur transitions between viable myocardium and fibrosis; image contrast depends on acquisition settings, making absolute intensity thresholds unreliable across datasets [19]; and manual segmentation is sensitive to observer choices such as threshold selection or region-of-interest definition. These limitations contribute to variability in scar quantification and present challenges for automated segmentation systems that must generalize across heterogeneous imaging conditions.

2.7.4 Relevance to Automated Analysis

LGE provides uniquely rich information about myocardial injury but does not directly encode tissue labels; instead, it produces contrast patterns that must be interpreted within anatomical and clinical context. Automated methods must therefore learn to distinguish subtle intensity variations associated with fibrosis while remaining robust to acquisition-dependent variability, and must integrate complementary imaging sequences, incorporate clinical priors, and address domain differences that are central themes in subsequent chapters.

2.8 Multi-Sequence CMR for Tissue Characterization

While LGE is the reference standard for visualizing chronic myocardial scar, clinical CMR examinations typically include multiple complementary sequences that emphasize different physical and physiological properties of tissue. Balanced steady-state free precession (bSSFP), commonly used for cine imaging, provides high signal-to-noise ratio and strong contrast between blood and myocardium, and is primarily used to assess ventricular volumes, wall motion, and global function [72, 75]. Although bSSFP does not directly highlight fibrosis, it provides detailed anatomical segmentation of endocardial and epicardial borders, offering spatial context that can constrain scar localization within the myocardial wall and reduce false positives near the blood pool. T2-weighted sequences are sensitive to increased water content and are commonly used to detect myocardial edema in acute or subacute infarction, where T2 hyperintensity may extend beyond the region of irreversible necrosis and help differentiate acute from chronic injury [76].

These contrasts are not redundant but complementary: boundaries that are ambiguous in one sequence may be clearer in another, and anatomical wall thickness in particular is often more precisely defined on cine bSSFP than on LGE. Many automated scar segmentation approaches nonetheless rely solely on LGE input, whereas clinicians typically incorporate multiple sequences to contextualize enhancement patterns. Integrating multi-sequence inputs can therefore allow models to exploit complementary feature spaces and improve robustness to contrast variability, though differences in spatial resolution, slice alignment, and acquisition timing introduce practical challenges for co-registration and fusion. These trade-offs are examined systematically in Chapter 6.

2.9 Electrocardiography (ECG) as a Functional Marker of Infarction

While CMR provides detailed structural information about myocardial tissue, ECG captures the electrical behavior of the heart in real time. Because myocardial infarction alters both structure and conduction, ECG offers complementary functional information that reflects the electrophysiological consequences of scar formation.

2.9.1 Physiological Basis of the ECG Signal

The surface ECG records voltage differences generated by the propagation of electrical activation through the myocardium. Depolarization and repolarization waves produce characteristic deflections (P waves, QRS complexes, and T waves) whose morphology and timing reflect atrial and ventricular activation patterns [77]. The recorded signal is also influenced by factors external to cardiac electrophysiology, including electrode placement, torso geometry, and patient-specific anatomy, so variations in lead positioning or anatomical differences can alter signal amplitude and morphology independently of underlying pathology. The QRS complex, representing ventricular depolarization, is particularly relevant to myocardial infarction: its duration, amplitude, and morphology depend on the integrity of conduction pathways and the distribution of viable myocardium, and infarction disrupts this coordinated activation, producing conduction delays, abnormal Q waves, ST-segment changes in acute settings, and altered R-wave progression in chronic scar [78].

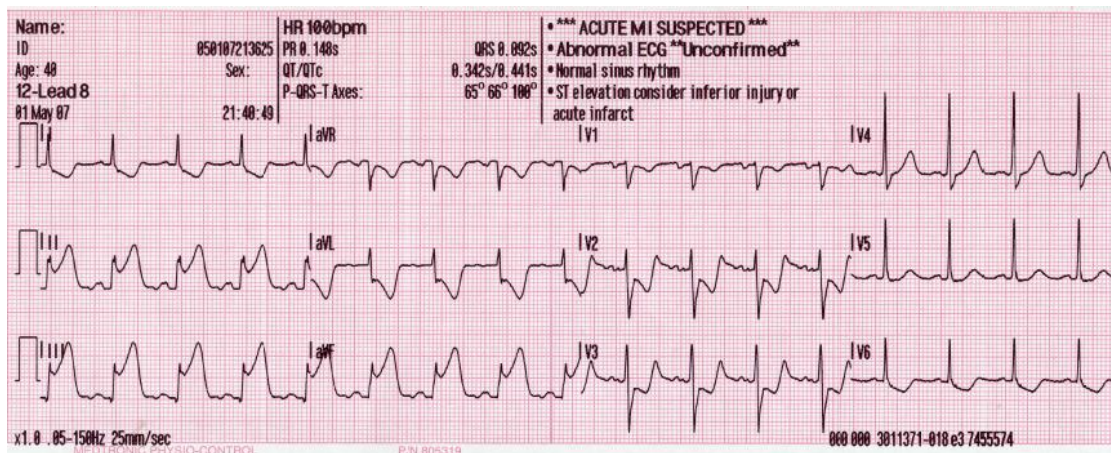


Figure 2.4: An ECG showing pardee waves indicating acute myocardial infarction in the inferior leads II, III and aVF with reciprocal changes in the anterolateral leads. Image by [Glenlarsen](#), via Wikimedia Commons (public domain).

2.9.2 Structure–Function Relationship

Myocardial scar alters electrical conduction velocity and anisotropy, producing delayed activation, fractionated signals, and potential re-entry circuits. These electrophysiological changes influence ECG morphology but are not directly visualized by imaging alone; conversely,

structural imaging does not capture dynamic conduction patterns. CMR and ECG therefore provide a complementary but partial view: CMR gives spatially resolved structural information, while ECG reflects integrated electrical function across the myocardium, and neither modality alone fully characterizes the arrhythmogenic substrate.

2.9.3 Implications for Multimodal Modeling

In clinical practice, physicians routinely interpret ECG findings in conjunction with imaging results to infer infarct location and severity. Incorporating ECG signals into scar segmentation frameworks therefore offers an opportunity to encode structure–function coupling, though the integration is non-trivial: ECG and imaging are often acquired at different time points, differ in dimensionality and spatial representation, and encode distinct but correlated physiological phenomena. These characteristics motivate the multimodal learning strategies for aligning heterogeneous anatomical and electrical signals that are developed in Chapter 5.

2.10 Public and Clinical Datasets for Myocardial Scar Analysis

The development of automated myocardial scar segmentation methods depends critically on the availability of curated imaging datasets with expert annotations. Unlike natural image analysis, medical imaging research is constrained by privacy considerations, acquisition costs, and the need for specialist interpretation, so datasets for myocardial infarction and scar analysis are relatively small, heterogeneous, and often collected for specific research challenges rather than routine clinical deployment. Two categories of dataset are used throughout this thesis: publicly available benchmarks released through international segmentation challenges, and retrospective institutional cohorts. The main public benchmarks are the Myocardial Pathology Segmentation (MyoPS) challenge and its multi-centre extension MyoPS++ [79, 80], which provide multi-sequence CMR (LGE, T2-weighted, and bSSFP) with expert annotations of myocardial structure and pathology, and the EMIDEC dataset [10], which focuses on delayed enhancement MRI for myocardial infarction assessment. A high-level summary is given in Table 2.2; the specific dataset splits, preprocessing pipelines, and inclusion criteria used in each contribution chapter are described in the corresponding experimental sections.

Table 2.2: *Public benchmark datasets used in this thesis. Detailed dataset splits and preprocessing pipelines are given in the experimental sections of the contribution chapters.*

Dataset	Modalities	Annotations	Used in
MyoPS [79]	LGE, T2, bSSFP	LV, myocardium, scar, edema	Ch. 3
MyoPS++ [80]	LGE, T2, bSSFP (multi-centre)	LV, myocardium, scar, edema	Ch. 6
EMIDEC [10]	LGE	LV, myocardium, scar, MVO	Ch. 4

Several recurring characteristics of these datasets influence algorithm design across the

contribution chapters. Sample sizes are limited because expert annotation of LGE-CMR is labour-intensive. Scar regions typically occupy a small fraction of myocardial volume, producing severe class imbalance. Acquisition protocols differ across scanners and centres, altering image appearance. Modality availability is often incomplete: not all patients undergo identical imaging protocols, so multi-sequence datasets frequently contain missing sequences. Finally, manual segmentation shows non-negligible inter-observer variability, particularly in heterogeneous border zones. In addition to public benchmarks, retrospective institutional cohorts are used in this thesis where paired ECG and CMR are required (Chapter 5); these cohorts introduce realistic variability in scanner vendors and acquisition parameters at the cost of smaller sample sizes and non-standardised annotation practices. These constraints mean that myocardial scar segmentation must be addressed not only as a pattern-recognition task but also as a data-limited learning problem, motivating the clinically guided data synthesis, flexible multimodal fusion, and domain-resilient learning strategies developed in the remainder of this thesis.

2.11 Traditional Image Analysis Methods

Early attempts to quantify myocardial scar relied on deterministic image analysis techniques rooted in signal intensity modelling. The most widely used approaches include the **standard deviation (SD) thresholding method**, where scar is defined as myocardium exceeding a specified number of standard deviations above a reference remote region, and the **full-width-at-half-maximum (FWHM) method**, which identifies enhanced tissue relative to the peak signal within an infarct core [19, 81]. Region-growing algorithms and atlas-based segmentation strategies were also explored to incorporate spatial continuity and anatomical priors [82]. While computationally simple and clinically interpretable, these techniques assume consistent intensity distributions across scans, are sensitive to threshold selection under noise and partial-volume effects, and struggle with heterogeneous infarcts containing mixtures of fibrosis, edema, and viable myocardium. As a result, they often produce inconsistent quantification and fail to capture the complex morphology of post-infarction remodelling.

2.12 Machine Learning and Deep Learning for Cardiac Segmentation

The emergence of machine learning enabled data-driven modelling of myocardial pathology. Early supervised classifiers incorporated handcrafted features derived from intensity, texture, and spatial context, but their generalization remained limited by feature design choices. Deep learning, particularly convolutional neural networks (CNNs), transformed medical image segmentation by enabling hierarchical feature learning directly from data. Architectures such as U-Net introduced encoder-decoder structures capable of integrating global context with fine spatial resolution [83], and rapidly became the dominant paradigm for cardiac segmentation. More recently, the **nnU-Net framework** demonstrated that much of segmentation performance depends on systematic configuration of preprocessing, architecture scaling, and training strategies rather than novel architectural design [84]; by automatically adapting to dataset characteristics, nnU-Net has established a strong reproducible baseline across diverse biomedical problems and is used as such in this thesis.

Despite these advances, existing deep-learning approaches to scar segmentation are typically developed under simplifying assumptions that do not hold in routine clinical practice: fully supervised datasets with reliable annotations, complete and aligned imaging modalities, consistent acquisition protocols, and adequate representation of pathological variability. Under these assumptions, most models rely exclusively on LGE imaging, ignore complementary structural and functional information, are trained on small single-centre cohorts with severe class imbalance, and degrade when applied to data from different scanners or centres [85]. Each contribution chapter of this thesis develops its own methodological response to a specific subset of these limitations and describes the relevant prior work in more detail.

2.13 Generative Modeling and Data Augmentation in Medical Imaging

Data augmentation is routinely used to improve generalization in deep learning, with common transformations including rotation, scaling, elastic deformation, and intensity perturbation [44]. These strategies increase the effective size of the dataset and promote invariance to geometric variation, but they do not create new pathological patterns; in myocardial scar segmentation, where rare infarct morphologies may be underrepresented, geometric transformations alone cannot introduce genuinely novel spatial distributions of fibrosis. Learned generative models offer a complementary route. Generative adversarial networks (GANs) introduced a framework for learning data distributions through adversarial training between a generator and a discriminator [45], and have been applied to medical imaging for data synthesis, domain adaptation, and augmentation, though they often lack explicit anatomical constraints and can produce visually plausible but clinically implausible pathology. Denoising diffusion probabilistic models (DDPMs) represent a more recent class of generative models that learn to iteratively reverse a noise process [46], offering improved training stability and better coverage of complex data distributions than GANs, and enabling conditional generation guided by structural inputs such as anatomical priors. Embedding clinical knowledge (myocardial segment boundaries, typical infarct territories, transmural progression patterns) can further constrain generation toward clinically realistic distributions, motivating the clinically guided synthesis framework introduced in Chapter 4.

2.14 Multimodal Learning and Structure–Function Modeling

Multimodal learning seeks to integrate complementary sources of information to obtain a more comprehensive representation of disease than is possible using a single modality alone [86]. In cardiovascular medicine this is particularly relevant because myocardial infarction manifests simultaneously as a structural abnormality visible on imaging and as a functional disturbance observable in electrical activity: CMR visualizes the spatial distribution of fibrosis, while ECG reflects the integrated electrical consequences of these structural changes, and clinicians routinely correlate the two (for example, associating Q-wave abnormalities or conduction delays with infarct localization seen on imaging). Fusion strategies in the multimodal-learning literature are commonly categorised as early fusion (concatenation at the input level), intermediate fusion (interaction between modality-specific features during

representation learning), or late fusion (combination of independent predictions), with intermediate fusion often most appropriate for medical applications since it preserves modality-specific structure while enabling cross-modal interaction. Unlike multimodal datasets in other machine-learning domains, however, clinical modalities are rarely perfectly aligned: ECG and CMR are often acquired at different time points, encode fundamentally different physical quantities, vary in signal quality, and may be missing entirely for some patients. These characteristics motivate fusion mechanisms that reflect clinically meaningful structure–function coupling and adapt to incomplete or asynchronous data rather than assuming deterministic alignment; the specific architecture developed for this thesis is described in Chapter 5.

2.15 Loss Functions for Imbalanced Medical Segmentation

Automated myocardial scar segmentation is a markedly imbalanced learning problem: infarcted tissue often constitutes only a small fraction of the left ventricular myocardium, so naïve optimization strategies bias models toward predicting background tissue while neglecting clinically important pathological regions. Standard voxel-wise losses such as cross-entropy treat each voxel independently and assign equal weight to all classes, which leads to dominance of the background class during training. In myocardial scar segmentation, this imbalance is compounded by small and irregular lesion volumes, ambiguous transition zones between scar and viable tissue, and variability in scar prevalence across patients, and can produce models that achieve high overall accuracy while failing to detect clinically meaningful scar regions.

2.15.1 Mathematical Formulation of Loss Functions

Let V_P and V_G denote the predicted and ground truth segmentation volumes, respectively. Let $p_i \in [0, 1]$ denote the predicted probability for voxel i , and $g_i \in \{0, 1\}$ the corresponding ground truth label.

Cross-Entropy Loss The cross-entropy loss (L_{CE}) is defined as:

$$\mathcal{L}_{CE} = - \sum_i [g_i \log(p_i) + (1 - g_i) \log(1 - p_i)] \quad (2.1)$$

This loss operates at the voxel level and is sensitive to class imbalance, as background voxels dominate the summation.

Dice Loss The Dice loss (L_{Dice}), derived from the Dice similarity coefficient, is given by:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum_i p_i g_i}{\sum_i p_i + \sum_i g_i} \quad (2.2)$$

This formulation emphasizes overlap between prediction and ground truth, making it inherently robust to class imbalance.

Intersection over Union (IoU) Loss The IoU loss (L_{IOU}), also known as the Jaccard loss, is defined as:

$$\mathcal{L}_{IOU} = 1 - \frac{\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i - \sum_i p_i g_i} \quad (2.3)$$

Compared to Dice, IoU imposes a stricter penalty on mismatches between prediction and ground truth.

Focal Loss The focal loss (L_{Focal}) addresses class imbalance by down-weighting easy examples:

$$\mathcal{L}_{Focal} = - \sum_i (1 - p_i)^\gamma g_i \log(p_i) \quad (2.4)$$

where $\gamma > 0$ is a focusing parameter that increases the contribution of hard-to-classify voxels.

Tversky Loss The Tversky loss (L_{Tv}) [87] is a generalisation of the Dice loss that allows false positives and false negatives to be weighted differently:

$$\mathcal{L}_{Tv} = 1 - \frac{\sum_i p_i g_i}{\sum_i p_i g_i + \alpha \sum_i p_i (1 - g_i) + \beta \sum_i (1 - p_i) g_i} \quad (2.5)$$

where α and β weight the contribution of false positives and false negatives respectively. When $\alpha = \beta = 0.5$, the denominator collapses to $\frac{1}{2}(\sum_i p_i + \sum_i g_i)$ and the expression becomes the standard Dice loss, so Dice can be viewed as a special case of Tversky. Choosing $\beta > \alpha$ places a heavier penalty on false negatives, which is desirable in our setting since missing scar tissue is clinically more costly than slightly over-segmenting it.

Boundary Loss The boundary loss (L_{BD}) incorporates spatial distance information using a signed distance map ϕ_G of the ground truth:

$$\mathcal{L}_{BD} = \sum_i p_i \phi_G(i) \quad (2.6)$$

This encourages alignment of predicted boundaries with anatomical contours.

Hausdorff Distance Loss The Hausdorff distance loss (L_{HD}) penalizes the maximum boundary deviation:

$$\mathcal{L}_{HD} = \max_{x \in \partial V_P} \min_{y \in \partial V_G} d(x, y) \quad (2.7)$$

where ∂V_P and ∂V_G denote predicted and ground truth boundaries, and $d(\cdot, \cdot)$ is the Euclidean distance. Differentiable approximations are typically used in practice.

Hybrid Loss Functions Given the complementary strengths of different loss functions, hybrid formulations are commonly employed:

$$\mathcal{L}_{Hybrid} = \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{CE} + \lambda_3 \mathcal{L}_{BD} + \lambda_4 \mathcal{L}_{HD} \quad (2.8)$$

where λ_i are weighting coefficients that balance region accuracy, voxel-wise classification, and boundary precision. Because myocardial scar segmentation involves detecting small, heterogeneous regions embedded within large volumes of normal myocardium, no single loss function adequately captures all aspects of the problem, and the composite objectives adopted in later chapters combine region-based, focal, and boundary-aware terms accordingly.

2.16 Evaluation of Myocardial Scar Segmentation

Evaluating myocardial scar segmentation requires metrics that capture both geometric accuracy and clinical relevance. Unlike segmentation of well-defined anatomical structures, scar assessment involves small, heterogeneous regions whose spatial distribution carries diagnostic and prognostic significance.

2.16.1 Mathematical Formulation of Evaluation Metrics

Let V_P and V_G denote the predicted and ground truth segmentation volumes, respectively. Let TP , FP , and FN denote true positives, false positives, and false negatives.

Dice Similarity Coefficient (DSC) The Dice Similarity Coefficient measures overlap between prediction and ground truth:

$$\text{DSC} = \frac{2|V_P \cap V_G|}{|V_P| + |V_G|} \quad (2.9)$$

It is the most widely used metric for medical image segmentation due to its robustness to class imbalance.

Intersection over Union (IoU) The IoU (Jaccard Index) is defined as:

$$\text{IoU} = \frac{|V_P \cap V_G|}{|V_P \cup V_G|} \quad (2.10)$$

It provides a stricter measure of overlap compared to DSC.

Precision and Recall Precision and recall quantify detection performance:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \quad (2.11)$$

Precision reflects false positive control, while recall measures sensitivity to scar detection.

Hausdorff Distance (HD) The Hausdorff Distance measures the maximum boundary deviation:

$$HD(V_P, V_G) = \max \left\{ \sup_{x \in \partial V_P} \inf_{y \in \partial V_G} d(x, y), \sup_{y \in \partial V_G} \inf_{x \in \partial V_P} d(x, y) \right\} \quad (2.12)$$

where ∂V_P and ∂V_G denote boundaries and $d(\cdot, \cdot)$ is the Euclidean distance.

Average Surface Distance (ASD) To mitigate sensitivity to outliers in HD, the Average Surface Distance is defined as:

$$ASD = \frac{1}{|\partial V_P|} \sum_{x \in \partial V_P} \min_{y \in \partial V_G} d(x, y) \quad (2.13)$$

This provides a more stable estimate of boundary agreement.

2.16.2 Interpretation of Metrics for Scar Segmentation

Each metric captures a different aspect of segmentation quality: DSC and IoU measure volumetric overlap but may overlook small boundary errors; precision penalises over-segmentation; recall penalises missed scar regions, which are often clinically critical; HD highlights worst-case boundary errors, relevant for transmural estimation; and ASD reflects average contour agreement and is less sensitive to outliers. Because myocardial scars are often small and irregular, DSC alone can be misleading, as high overlap may still correspond to clinically significant boundary inaccuracies. Pixel-wise agreement therefore does not guarantee clinical usefulness: small spatial discrepancies may have minimal effect on overlap metrics yet significantly alter interpretation in subendocardial or transmural regions. Beyond geometric similarity, clinically meaningful evaluation also requires assessing scar burden and its spatial distribution. Scar volume error, defined as $\Delta V = (|V_P| - |V_G|)/|V_G|$, quantifies the relative volumetric discrepancy; regional transmural (the proportion of myocardial wall thickness occupied by scar within predefined segments, such as those of the AHA 17-segment model introduced in the next section) captures spatial extent; and spatial consistency (agreement in infarct localization across slices and regions) reflects anatomical plausibility. These measures relate directly to clinical tasks such as viability assessment and arrhythmia risk stratification.

2.17 The American Heart Association 17-Segment Model

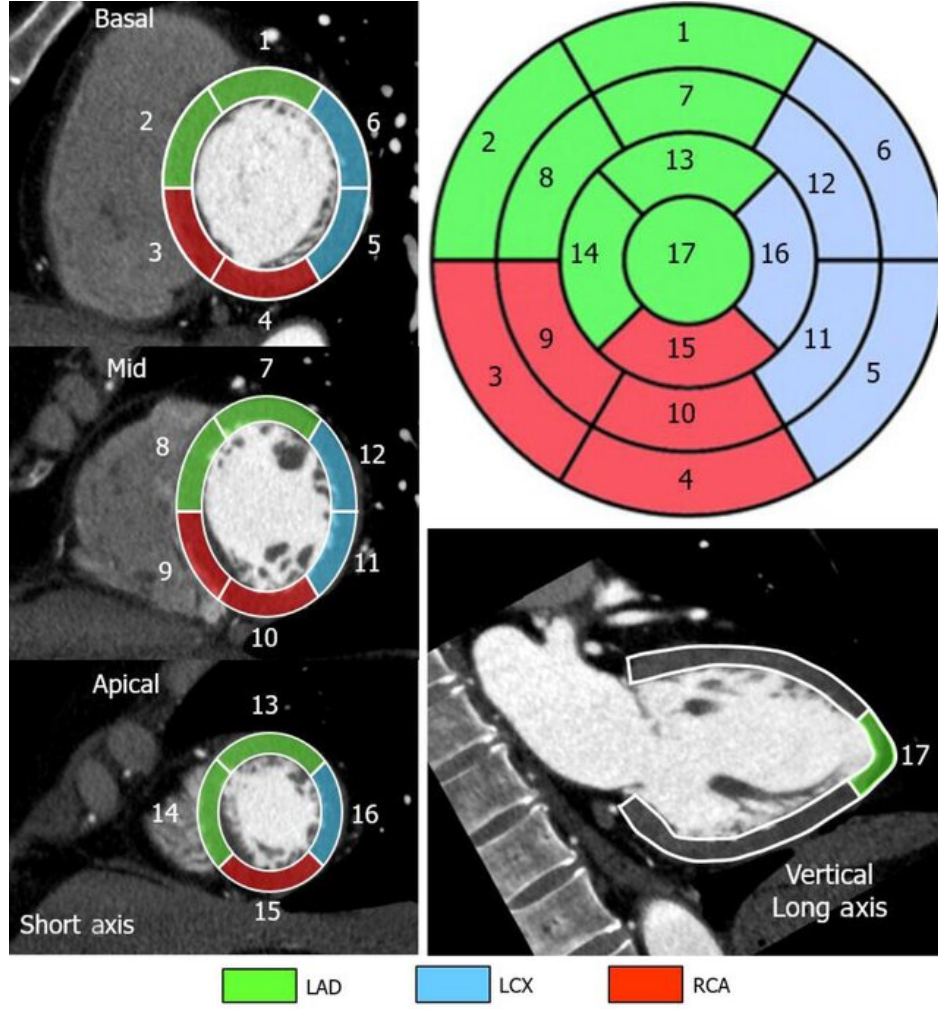
To standardize the description of myocardial structure and pathology, the American Heart Association (AHA) introduced a 17-segment model that divides the left ventricle into anatomically consistent regions aligned with coronary perfusion territories [55]. As illustrated in Figure 2.5, the model partitions the left ventricle into three short-axis levels (basal segments 1–6, mid-cavity segments 7–12, and apical segments 13–16) along with a distinct apical cap (segment 17). Each segment corresponds to a reproducible anatomical region, enabling consistent regional analysis independent of imaging orientation, and the segmental organization reflects coronary artery territories typically associated with the left anterior descending (LAD), left circumflex (LCX), and right coronary artery (RCA), as shown in the figure. This mapping enables direct interpretation of imaging-derived scar patterns in relation to underlying vascular supply and provides a shared anatomical reference for aligning heterogeneous data sources such as LGE-MRI and ECG.

While conventional metrics such as Dice or Hausdorff distance quantify geometric agreement, they do not explicitly capture whether segmentation errors occur in clinically critical regions. The AHA model enables segment-wise evaluation, where performance can be assessed within anatomically and clinically meaningful units. Let S_k denote the set of voxels belonging to the k -th AHA segment, where $k \in \{1, \dots, 17\}$. The segment-wise Dice Similarity Coefficient is defined as

$$\text{DSC}_k = \frac{2|(V_P \cap S_k) \cap (V_G \cap S_k)|}{|V_P \cap S_k| + |V_G \cap S_k|}, \quad (2.14)$$

and an overall segment-level score may be computed as

$$\overline{\text{DSC}}_{\text{seg}} = \frac{1}{17} \sum_{k=1}^{17} \text{DSC}_k. \quad (2.15)$$



DOI: 10.4329/wjr.v14.i2.30 Copyright ©The Author(s) 2022.

Figure 2.5: Standard segmental myocardial representation using the AHA 17-segment model across basal, mid-cavity, and apical levels, with corresponding coronary territories. Figure source: [5], reproduced under the terms of the Creative Commons Attribution NonCommercial License CC BY-NC 4.0.

In addition, segment-wise scar burden may be estimated as $B_k = |V \cap S_k|/|S_k|$, where V denotes either the predicted or ground truth scar volume, providing a clinically interpretable estimate of the proportion of each segment affected by scar. Such measures identify region-specific failures that may be obscured by global scores, and this segmental framework is used in later chapters both for region-wise evaluation and, in the multimodal setting, for grounding latent features from heterogeneous modalities in a shared anatomical reference.

Chapter 3

Scar Segmentation from Multi-Sequence CMR

This chapter contains material from:

- Ramzan, Farheen, Chen (Cherise) Chen, and Richard H. Clayton. "Improving Myocardial Scar Segmentation with End-to-End and Cascaded CNN using Hybrid Loss and Multi-Modality CMR Imaging." *2024 Computing in Cardiology Conference (CinC)*. IEEE, 2024.

In this chapter, a fundamental challenge in automated cardiac image analysis is addressed: the accurate and reliable segmentation of myocardial scars from CMR imaging. While deep learning methods have shown promise in automating cardiac segmentation tasks, myocardial scar segmentation remains particularly difficult due to subtle contrast differences, heterogeneous morphology, and strong class imbalance. This chapter establishes the methodological foundation of the thesis by systematically investigating how architectural design choices, multi-sequence CMR inputs, and loss function formulations influence automated scar segmentation performance.

3.1 Introduction

As discussed in Chapter 2, myocardial scarring following infarction represents irreversible tissue injury and a key determinant of ventricular dysfunction and arrhythmic risk. Accurate identification and quantification of scar are therefore essential for clinical applications such as risk stratification and catheter ablation planning [88, 89].

CMR provides a non-invasive means of assessing myocardial structure and viability, with different imaging sequences offering complementary information about infarcted tissue. In particular, LGE highlights chronic fibrosis, T2-weighted imaging reflects edema and acute injury, and bSSFP cine sequences provide high-contrast anatomical context. Together, these modalities form a rich but heterogeneous representation of post-infarction pathology.

Despite these advances, automated segmentation of myocardial scar from CMR remains challenging. Scar regions are typically small, irregularly shaped, and exhibit low contrast

with surrounding myocardium, particularly in heterogeneous border zones. These challenges are compounded by variability in image quality, acquisition protocols, and patient-specific anatomy. Manual segmentation remains the clinical standard but is time-consuming, subjective, and difficult to scale, motivating the development of automated methods.

Deep learning approaches, particularly convolutional neural networks such as U-Net [26] and self-configuring frameworks such as nnU-Net [27], have achieved strong performance in many medical image segmentation tasks. However, accurate and robust segmentation of myocardial scar remains an open problem [12, 29]. Key challenges include severe class imbalance, ambiguous scar boundaries, and the reliance on single-sequence inputs that fail to capture the full spectrum of infarct pathology.

Clinical practice routinely integrates information from multiple CMR sequences to interpret myocardial injury, suggesting that multi-sequence approaches may improve automated segmentation. However, important questions remain unresolved: how best to combine complementary imaging sequences, whether additional architectural complexity such as cascaded pipelines provides consistent benefits, and how training strategies can be designed to address extreme class imbalance.

This chapter addresses these questions through a systematic investigation of automated myocardial scar segmentation using multi-sequence CMR. A standardized framework based on nnU-Net [27] is employed to compare end-to-end and cascaded segmentation pipelines, evaluate the contribution of different CMR sequence combinations, and assess the impact of loss function design under severe class imbalance.

The contributions of this chapter are threefold:

- First, a reproducible and extensible baseline is established for automated myocardial scar segmentation using multi-sequence CMR.
- Second, a detailed empirical analysis of architectural and training strategies is provided, clarifying when anatomical priors and cascaded pipelines are beneficial and when simpler end-to-end approaches are preferable.
- Third, effective loss function combinations are identified that improve segmentation accuracy for scars in the presence of severe class imbalance.

3.2 Background

3.2.1 Myocardial Scar Segmentation in Cardiac MRI

As outlined in Chapter 2, myocardial scarring following infarction represents irreversible tissue injury and a key determinant of ventricular dysfunction and arrhythmic risk. Accurate segmentation of scar extent and spatial distribution is therefore essential for clinical applications such as viability assessment, risk stratification, and catheter ablation planning [18, 88].

LGE-CMR, introduced in Chapter 2, is widely regarded as the reference standard for visualizing infarcted myocardium. However, automated scar segmentation from LGE images remains challenging. Scar regions are typically small, irregularly shaped, and exhibit intensity overlap with surrounding myocardium and blood pool, particularly in heterogeneous border zones [90]. These challenges are further compounded by partial volume effects, motion artifacts, and inter-subject variability in scar morphology.

Manual segmentation remains the clinical standard but is time-consuming, subjective, and difficult to scale. Substantial inter-observer variability has been reported even among expert readers, particularly for diffuse or subendocardial scars [19]. These limitations motivate the development of automated methods capable of providing accurate and reproducible scar quantification.

3.2.2 Traditional and Learning-Based Approaches

As discussed in Chapter 2, early automated methods relied on thresholding-based techniques such as n -standard deviation (n -SD) and full width at half-maximum approaches [81, 91, 92]. While computationally efficient, these methods are highly sensitive to noise, intensity inhomogeneity, and user-defined reference regions, limiting their robustness across datasets and scanners.

Subsequent approaches introduced machine learning techniques based on hand-crafted features and classical classifiers, including random forests and support vector machines [93, 94]. Although these methods improved robustness relative to pure thresholding, they remained constrained by feature design and limited generalization to unseen data.

Deep learning approaches, particularly convolutional neural networks such as U-Net [26], have significantly advanced medical image segmentation. However, extending these models to myocardial scar segmentation has proven more difficult than for anatomical structures due to the small size, irregular morphology, and heterogeneous appearance of scar tissue.

Recent benchmark challenges, including MyoPS [12, 29], highlight the persistent difficulty of this task. Despite the use of advanced architectures, segmentation performance for scar regions remains substantially lower than for anatomical structures, indicating that current methods struggle to capture the complexity of infarcted tissue.

3.2.3 Self-Configuring Networks and Architectural Design Choices

One limitation of many deep learning approaches is their reliance on manually designed architectures and hyperparameters, which may not generalize across datasets. Self-configuring frameworks such as nnU-Net [27] address this by automatically adapting network architecture, preprocessing, and training protocols to dataset characteristics, providing a strong and reproducible baseline.

Recent studies have adopted nnU-Net for cardiac segmentation tasks, including myocardial scar and edema segmentation [12, 95]. However, its performance remains constrained by input representations and training strategies. In particular, it remains unclear whether additional architectural complexity, such as cascaded pipelines or explicit incorporation of anatomical priors, consistently improves scar segmentation performance.

Cascaded frameworks first segment cardiac anatomy and subsequently use this information to guide pathology segmentation [96]. While intuitively appealing, such approaches increase computational complexity and may propagate errors from earlier stages. A systematic comparison between cascaded and end-to-end pipelines for myocardial scar segmentation remains limited.

3.2.4 Role of Multi-Sequence CMR

Building on the imaging principles described in Chapter 2, clinical CMR protocols include multiple complementary sequences, each capturing distinct aspects of myocardial pathology. LGE emphasizes chronic fibrosis, T2-weighted imaging reflects edema and acute injury, and bSSFP cine imaging provides high-contrast anatomical context [95, 97].

Although several studies suggest that integrating multi-sequence information can improve myocardial tissue characterization [13, 14, 98, 99], most existing scar segmentation approaches rely on LGE alone. This is largely due to challenges in multimodal alignment, increased model complexity, and limited availability of fully annotated multi-sequence datasets.

Furthermore, there is no consensus on how multi-sequence information should be integrated. Early fusion strategies concatenate modalities at the input level, while late fusion combines independent predictions. Each approach introduces trade-offs, including sensitivity to misregistration and increased model complexity. A systematic evaluation of sequence combinations and fusion strategies specifically for myocardial scar segmentation remains limited.

3.2.5 Class Imbalance and Loss Function Design

A fundamental challenge in myocardial scar segmentation is extreme class imbalance: scar regions typically occupy only a small fraction of the myocardial volume (on the order of 5–20%). Standard loss functions such as cross-entropy tend to bias predictions toward majority classes, leading to poor sensitivity for small or diffuse scars [100].

To address this, several loss functions have been proposed, including Dice loss, focal loss, and focal Tversky loss [101, 102]. While these approaches improve sensitivity to small structures, their relative effectiveness for myocardial scar segmentation—particularly in combination with multi-sequence inputs—remains insufficiently explored.

3.2.6 Motivation for This Work

In contrast to prior studies that consider individual design choices in isolation, this work provides a systematic investigation of automated myocardial scar segmentation using multi-sequence CMR. Building on the robust baseline provided by nnU-Net [27], end-to-end and cascaded segmentation pipelines are compared, the contribution of complementary CMR sequences is evaluated, and the impact of different loss function formulations is assessed under conditions of severe class imbalance.

The selection of these pipeline strategies reflects a central methodological question in cardiac image analysis. End-to-end approaches enable holistic feature learning directly from imaging data, potentially capturing complex interactions between anatomy and pathology. In contrast, cascaded approaches explicitly incorporate anatomical priors by separating structural and pathological segmentation, which may improve localization in challenging regions. Comparing these paradigms allows assessment of whether explicit structural guidance provides advantages over integrated representation learning for myocardial scar segmentation.

The use of multi-sequence CMR is motivated by the complementary nature of different imaging contrasts. LGE provides direct visualization of fibrosis, T2-weighted imaging reflects edema and acute injury, and bSSFP offers precise anatomical context. While clinical interpretation routinely integrates these modalities, their combined impact on automated scar segmentation remains insufficiently understood. Evaluating different sequence combinations

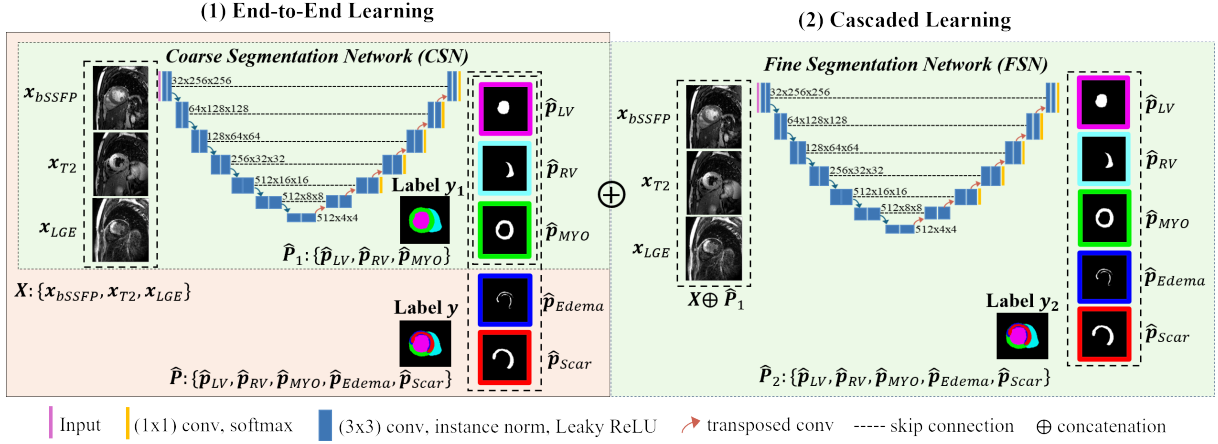


Figure 3.1: A schematic diagram of the framework proposed for myocardial scar segmentation. (1) End-to-End pipeline (orange region) takes $\mathbf{X} : \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\}$ as input and predicts $\hat{\mathbf{P}} : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}$ as output in one step. (2) Cascaded pipeline (green region), the first stage is a coarse segmentation network (CSN) that takes $\mathbf{X} : \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\}$ as input and predicts $\hat{\mathbf{P}}_1 : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}\}$ as output. The second stage is a fine segmentation network (FSN) that takes $\mathbf{X} \oplus \hat{\mathbf{P}}_1$ as input and predicts the final segmentation maps denoted by $\hat{\mathbf{P}}_2 : \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}$, where $\oplus$ denotes channel-wise concatenation. Figure source: [6], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.

therefore enables assessment of whether complementary contrasts provide meaningful gains over single-sequence inputs.

Finally, the investigation of loss function design is motivated by the severe class imbalance inherent in myocardial scar segmentation, where scar regions occupy only a small fraction of the myocardium. Standard optimization objectives often bias predictions toward majority classes, limiting sensitivity to small or diffuse lesions. Exploring alternative loss formulations allows assessment of how optimization strategies influence the detection of clinically relevant scar regions under realistic data conditions.

By grounding these analyses in a standardized experimental framework and benchmarking against established challenge datasets, this work identifies which design choices meaningfully improve segmentation performance and which introduce unnecessary complexity.

3.3 Methods

To investigate automated myocardial scar segmentation using deep learning under a standardized and reproducible experimental framework, two learning paradigms are considered: an end-to-end segmentation pipeline and a cascaded coarse-to-fine segmentation framework. An overview of both approaches is illustrated in Fig. 3.1.

3.3.1 Dataset and Imaging Modalities

All experiments were conducted using the MyoPS 2020 Challenge dataset [13, 98] described in Chapter 2, which provides multi-sequence CMR images acquired from the same subjects. Each case consists of three complementary imaging modalities: LGE, T2-weighted imaging, and bSSFP cine imaging. These sequences capture distinct but complementary aspects of myocardial pathology and anatomy.

Expert annotated ground truth labels are provided for five clinically relevant structures: left ventricular (LV) blood pool, right ventricular (RV) blood pool, healthy myocardium (MYO), myocardial edema, and myocardial scar. This dataset enables a comprehensive evaluation of both anatomical and pathological segmentation performance under a unified framework.

3.3.2 Problem Formulation

Given a multi-sequence CMR input

$$\mathbf{X} = \{\mathbf{x}_{bSSFP}, \mathbf{x}_{T2}, \mathbf{x}_{LGE}\},$$

the objective is to predict segmentation maps

$$\hat{\mathbf{P}} = \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\}.$$

The primary focus is accurate segmentation of myocardial scar, which is challenging due to its small spatial extent, heterogeneous appearance, and severe class imbalance.

3.3.3 End-to-End Learning: Baseline Network

As a baseline, a fully end-to-end segmentation strategy is adopted based on the 2D nnU-Net framework [27], a self-configuring convolutional neural network that automatically adapts its preprocessing, architecture, training schedule, and post-processing steps to the given dataset.

In this setting, the three CMR modalities are concatenated channel-wise and fed directly into the network. The model jointly learns to segment all cardiac substructures and pathological regions in a single forward pass. Formally, the end-to-end model with parameters θ is defined as

$$\hat{\mathbf{P}} = \text{Baseline}(\mathbf{X}; \theta).$$

This approach serves as a strong reference point, as nnU-Net has demonstrated state-of-the-art performance across a wide range of biomedical segmentation tasks. However, it does not explicitly encode prior knowledge about the hierarchical relationship between cardiac anatomy and pathology.

3.3.4 Cascaded Learning: Coarse-to-Fine Segmentation (CNN^{C2F})

To investigate whether explicit incorporation of anatomical priors improves scar segmentation, a cascaded coarse-to-fine segmentation framework is designed, denoted CNN^{C2F} . This approach leverages the hierarchical structure of cardiac anatomy and pathology by decomposing the segmentation task into two sequential stages.

The underlying motivation is that accurate localization of the myocardium may provide useful spatial context for identifying pathological regions such as scar and edema, which are confined within the myocardial wall.

Coarse Segmentation Network (CSN)

In the first stage, a Coarse Segmentation Network (CSN) is trained to segment the major cardiac structures: LV blood pool, RV blood pool, and the complete myocardium. The complete myocardium encompasses healthy myocardium, scar, and edema regions:

$$MYO_{complete} = MYO_{healthy} \cup Scar \cup Edema.$$

The CSN takes the multi-sequence input $\mathbf{X}$ and predicts coarse structural maps

$$\hat{\mathbf{P}}_1 = \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}\},$$

with trainable parameters θ_1 :

$$\hat{\mathbf{P}}_1 = CSN(\mathbf{X}; \theta_1).$$

This stage focuses on learning robust anatomical representations that are relatively invariant to pathological heterogeneity.

Fine Segmentation Network (FSN)

In the second stage, a Fine Segmentation Network (FSN) refines the segmentation by explicitly incorporating the coarse anatomical predictions as prior knowledge. The input to FSN is constructed by channel-wise concatenation of the original CMR images and the predicted coarse maps:

$$\mathbf{X} \oplus \hat{\mathbf{P}}_1.$$

The FSN predicts the final segmentation maps

$$\hat{\mathbf{P}}_2 = \{\hat{\mathbf{p}}_{LV}, \hat{\mathbf{p}}_{RV}, \hat{\mathbf{p}}_{MYO}, \hat{\mathbf{p}}_{Edema}, \hat{\mathbf{p}}_{Scar}\},$$

parameterized by θ_2 :

$$\hat{\mathbf{P}}_2 = FSN(\mathbf{X} \oplus \hat{\mathbf{P}}_1; \theta_2).$$

Both CSN and FSN employ nnU-Net as their backbone, benefiting from its automated configuration and robust training protocol. The cascaded design explicitly encodes a coarse-to-fine hierarchy, allowing to study whether anatomical guidance improves pathological segmentation.

3.3.5 Loss Function Design and Optimization Strategy

Accurate myocardial scar segmentation is severely affected by class imbalance, as scar regions occupy only a small fraction of the myocardium. To address this challenge, a range of loss functions and hybrid loss formulations are systematically evaluated, that combine region-based, boundary-aware, and class-balanced objectives.

Let V_P and V_G denote the predicted and ground truth segmentation volumes, respectively. The following loss functions and combinations [103] are considered by combining existing

loss functions (refer to Chapter 2) such as Dice loss (L_{Dice}), cross-entropy loss (L_{CE}), focal loss (L_{Focal}), Tversky loss (L_{Tv}), boundary loss (L_{BD}), Hausdorff distance loss (L_{HD}) and intersection-over-union loss (L_{IOU}):

$$L_{DiceCE}(V_G, V_P) = L_{Dice} + L_{CE} \quad (3.1)$$

$$L_{DiceHD}(V_G, V_P) = L_{Dice} + L_{HD} \quad (3.2)$$

$$L_{FocalTv}(V_G, V_P) = L_{Focal} + L_{Tv} \quad (3.3)$$

$$L_{DiceFocalTv}(V_G, V_P) = L_{Dice} + L_{Focal} + L_{Tv} \quad (3.4)$$

$$L_{IOUCE}(V_G, V_P) = L_{IOU} + L_{CE} \quad (3.5)$$

$$L_{H_1}(V_G, V_P) = \alpha(L_{Dice} + L_{CE}) + (1 - \alpha)L_{BD} \quad (3.6)$$

$$L_{H_2}(V_G, V_P) = \alpha(L_{Dice} + L_{Focal}) + (1 - \alpha)L_{BD} \quad (3.7)$$

$$L_{H_3}(V_G, V_P) = \alpha(L_{FocalTv} + L_{Focal}) + (1 - \alpha)L_{BD} \quad (3.8)$$

$$L_{H_4}(V_G, V_P) = \alpha(L_{Dice} + L_{FocalTv}) + (1 - \alpha)L_{BD} \quad (3.9)$$

Here, $\alpha = \alpha_0(1 - \frac{epoch}{epoch_{max}})$ is a scheduling parameter that decays over training epochs. This design emphasizes region-based losses during early training to stabilize convergence, while progressively increasing the influence of boundary-aware loss terms to refine scar contours in later stages.

Each network stage is optimized independently using its corresponding loss function, enabling fine-grained control over learning dynamics in both end-to-end and cascaded settings.

3.3.6 Implementation Details

All models are trained in a 5-fold cross-validation setting as prescribed by the nnU-Net framework. Preprocessing, data augmentation, learning rate scheduling, and post-processing steps are automatically configured by nnU-Net to ensure reproducibility and fairness across experiments.

This systematic methodological design enables a controlled comparison of architectural strategies, modality combinations, and loss formulations, providing the empirical foundation for the analyses presented in the subsequent sections.

Table 3.1: Results obtained on the MyOPS test dataset with models trained using the end-to-end (baseline) and the cascaded pipeline strategies with varied CMR modalities as input, where $Phase_1$ and $Phase_2$ represent CSN and FSN of CNN^{C2F} , respectively. Here, the standard loss: L_{DiceCE} in our experiments is used. Table source: [6], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.

Strategy	Model Type	CMR Modality (Input)			Cardiac Structure (Output)					Dice (Scar)	Dice (Edema + Scar)	Dice (Average)
		LGE	T2	bSSFP	LV	RV	MYO	Edema	Scar			
S_1	End-to-End	✓	✗	✗	✓	✓	✓	✓	✓	0.6425	0.5894	0.6160
S_2	End-to-End	✓	✓	✓	✓	✓	✓	✓	✓	0.7247	0.7095	0.7171
S_3	Cascaded	$Phase_1$	✗	✓	✓	✗	✓	✗	✗	0.6051	0.6458	0.6255
		$Phase_2$	✓	✓	✗	✗	✗	✓	✓			
S_4	Cascaded	$Phase_1$	✗	✗	✓	✓	✓	✗	✗	0.6087	0.6445	0.6266
		$Phase_2$	✓	✓	✗	✓	✓	✓	✓			
S_5	Cascaded	$Phase_1$	✓	✗	✗	✓	✓	✗	✗	0.6196	0.6014	0.6105
		$Phase_2$	✓	✗	✗	✓	✓	✓	✓			
S_6	Cascaded	$Phase_1$	✓	✓	✓	✓	✓	✗	✗	0.7076	0.7215	0.7146
		$Phase_2$	✓	✓	✓	✓	✓	✓	✓			
Zhai et al. [12] <i>MyOPS Winner</i>	Cascaded	$Phase_1$	✓	✓	✓	✓	✓	✗	✗	0.6723	0.7314	0.7019
		$Phase_2$	✓	✓	✓	✓	✓	✓	✓			

3.4 Results and Discussion

This section presents a comprehensive quantitative and qualitative evaluation of the proposed segmentation strategies. The impact of architectural design (end-to-end versus cascaded learning), multi-sequence CMR integration, and loss function selection on myocardial scar segmentation performance is analyzed using the MyoPS 2020 test dataset.

3.4.1 Experimental Setup and Implementation Details

All models were trained for a maximum of $epoch_{max} = 1000$ epochs using a batch size of 5 and a patch size of 256×256 . Stochastic Gradient Descent (SGD) was employed with an initial learning rate of 0.01, which was decayed using the polynomial learning rate schedule

$$\left(1 - \frac{epoch}{epoch_{max}}\right)^{0.9}.$$

Input images were normalized using z-score normalization, and extensive data augmentation was applied on-the-fly, including random rotations, scaling, elastic deformations, and intensity perturbations. These training settings were kept consistent across all experiments to ensure fair comparison between different strategies.

3.4.2 Quantitative Evaluation of Architectural Strategies

Table 3.1 summarizes the segmentation performance obtained using different training strategies and modality combinations. Dice scores for myocardial scar, combined edema+scar, and the average across all predicted structures are reported.

End-to-End vs. Cascaded Learning

Among all evaluated strategies, the end-to-end model trained using all three CMR modalities (S_2) achieved the highest Dice score for scar segmentation (0.7247) as well as the highest

overall average Dice score (0.7171). This result indicates that jointly learning anatomical and pathological structures within a single optimization process is advantageous for capturing the subtle appearance variations associated with myocardial scar.

In contrast, cascaded strategies (S_3 - S_6), which explicitly separate anatomical segmentation from pathology refinement, generally yielded inferior scar Dice scores. Although the best cascaded configuration (S_6) achieved competitive performance for edema+scar segmentation (Dice = 0.7215), it did not surpass the end-to-end approach for scar-only segmentation.

These findings suggest that while coarse anatomical priors are intuitively appealing, segmentation errors introduced in the first stage of cascaded pipelines may propagate to subsequent stages, limiting the model’s ability to accurately delineate small and heterogeneous scar regions. This observation aligns with prior studies noting error accumulation in multi-stage segmentation frameworks [12].

Comparison with the MyoPS Challenge Winner

Notably, the proposed end-to-end approach outperformed the MyoPS 2020 winning method [12] in terms of scar Dice score, despite the latter relying on a carefully engineered cascaded strategy. This result highlights the effectiveness of a standardized, self-configuring framework when combined with appropriate modality integration and loss function design.

3.4.3 Effect of Multi-Sequence CMR Integration

A consistent trend across Table 3.1 is the performance improvement obtained by incorporating complementary CMR sequences. Models trained using only LGE imaging (S_1) achieved substantially lower scar Dice scores (0.6425) compared to those leveraging all three modalities.

The inclusion of T2-weighted imaging provides sensitivity to myocardial edema, while bSSFP cine images offer high-contrast anatomical segmentation of ventricular boundaries. When integrated effectively, these complementary cues allow the network to better localize pathology within the myocardial wall and disambiguate scar from surrounding structures. This finding supports the hypothesis that multi-sequence CMR contains synergistic information that cannot be fully exploited by single-modality models.

3.4.4 Impact of Loss Function Design

To further investigate optimization behavior under severe class imbalance, a range of loss functions and hybrid loss formulations are evaluated using the two best performing strategies: the end-to-end model (S_2) and the cascaded model (S_6). Results are reported in Table 3.2.

For scar segmentation, the end-to-end model achieved its best performance (Dice = 0.7257) when trained with the hybrid $L_{DiceFocalTv}$ loss. This loss combines region overlap, class imbalance weighting, and asymmetric penalization, enabling the model to better capture small and irregular scar regions.

For edema+scar segmentation, the cascaded model achieved its highest Dice score (0.7295) using the L_{DiceHD} loss, which incorporates boundary distance information. This suggests that boundary-aware losses are particularly beneficial for capturing diffuse pathological regions such as edema, where precise segmentation of spatial extent is critical.

Table 3.2: Results obtained with MyOPS test dataset by training end-to-end (S_2) and cascaded (S_6) networks using different loss functions. Table source: [6], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.

Loss Function	Dice (Scar)		Dice (Edema + Scar)	
	S_2	S_6	S_2	S_6
L_{DiceCE}	0.7247	0.7076	0.7095	0.7215
L_{DiceHD}	0.7221	0.7111	0.7061	0.7295
$L_{FocalTv}$	0.7182	0.7117	0.7021	0.7225
$L_{DiceFocalTv}$	0.7257	0.7073	0.7082	0.7236
L_{IOUCE}	0.7236	0.7115	0.7090	0.7238
L_{H_1}	0.7148	0.7084	0.6992	0.7243
L_{H_2}	0.7205	0.7062	0.7055	0.7224
L_{H_3}	0.7231	0.7140	0.7094	0.7265
L_{H_4}	0.7254	0.7101	0.7118	0.7272

* Top two best Dice scores per model (**1st** and **2nd**) are highlighted.

* Gray region represents the overall best segmentation performance.

* L_{DiceCE} is used for the first stage of the cascaded model (S_6).

Overall, these results demonstrate that no single loss function is universally optimal; instead, loss design should be tailored to the specific target structure and segmentation objective.

3.4.5 Qualitative Evaluation

Figure 6.2 presents qualitative results for a representative test case, visualizing predictions from both cascaded and end-to-end pipelines overlaid on LGE images across apical, mid-ventricular, and basal slices.

The end-to-end model produces smoother and more spatially consistent scar segmentations, particularly in the mid-ventricular and apical regions where scar morphology is most heterogeneous. In contrast, cascaded predictions exhibit occasional discontinuities and under-segmentation of scar regions, consistent with the quantitative findings.

It is important to note that ground truth annotations are not available for the MyoPS test set, limiting objective verification of qualitative results. Nevertheless, the predicted scar patterns appear anatomically plausible and clinically meaningful, suggesting that the proposed framework may substantially reduce the burden of manual segmentation and inter-observer variability in clinical workflows.

3.4.6 Summary of Findings

The experimental results demonstrate that:

- End-to-end segmentation consistently outperforms cascaded approaches for myocardial scar segmentation.
- Integrating complementary CMR sequences significantly enhances segmentation accuracy.

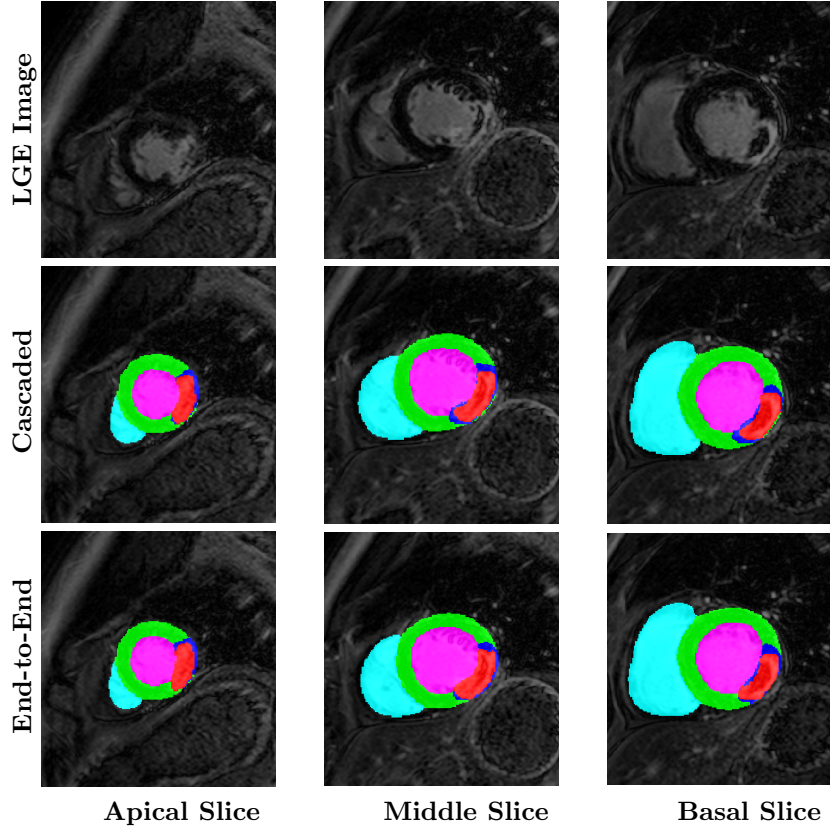


Figure 3.2: Visualizing predictions for three consecutive slices for a test CMR from MyOPS dataset: (row 1) LGE images, (row 2) and (row 3) predictions obtained with cascaded and end-to-end pipeline overlaid on LGE, respectively. Colored regions represent respective class LV, RV, MYO, edema and scar. Figure source: [6], reproduced under the terms of the Creative Commons Attribution License CC BY 4.0.

- Hybrid loss functions tailored to class imbalance and boundary precision yield measurable performance gains.
- Self-configuring frameworks such as nnU-Net provide a strong and reproducible foundation for automated cardiac pathology segmentation.

These findings establish a robust baseline for automated myocardial scar segmentation.

3.5 Conclusion

This chapter established a standardized and reproducible framework for automated myocardial scar segmentation using a self-configuring 2D CNN backbone (nnU-Net). Within this framework, key methodological choices are systematically investigated that substantially influence segmentation performance: (i) architectural strategy (end-to-end versus cascaded coarse-to-fine learning), (ii) the role of multi-sequence CMR inputs (LGE, T2-weighted, and

bSSFP), and (iii) loss function design to mitigate severe class imbalance and improve sensitivity to small, heterogeneous scar regions.

Across all evaluated strategies, our results demonstrate that an end-to-end formulation provides the most effective and reliable solution for scar segmentation on the MyoPS test benchmark. In contrast to the prevailing trend toward cascaded pipelines in prior work, the cascaded approach did not yield superior scar segmentation in our experiments. A key contributing factor is the error propagation between stages: inaccuracies in the coarse anatomical predictions can restrict or distort the subsequent pathology refinement, ultimately limiting scar sensitivity and boundary precision. These findings emphasize that explicit anatomical priors must be incorporated carefully to avoid introducing brittle dependencies between stages.

It is further shown that multi-sequence CMR integration provides consistent performance gains over single-sequence learning. Leveraging complementary contrasts from T2-weighted and bSSFP cine imaging alongside LGE improves localization of pathology within the myocardial wall and reduces ambiguity at scar boundaries, supporting the hypothesis that multi-sequence inputs better represent the structural and pathological variability present in real infarct cases.

Finally, loss function design is identified as a critical driver of performance in the presence of extreme class imbalance. The strongest scar segmentation performance was achieved when the end-to-end model was optimized with a composite objective combining Dice-based overlap (to encourage region consistency) with focal Tversky weighting (to emphasize small structures and penalize false negatives). This combination improves sensitivity to rare scar pixels while maintaining stable optimization and anatomically plausible segmentation outputs.

In summary, this chapter provides a principled baseline for automated scar segmentation and clarifies which design choices are most impactful for performance on multi-sequence CMR. Ultimately, advancing reliable automated scar segmentation is an important step toward scalable myocardial scar quantification and patient specific treatment planning for ventricular tachycardia.

3.6 Limitations and Future Work

Despite the strong performance achieved by the proposed framework, several limitations remain and motivate directions for future research.

First, although the use of multi-sequence CMR improves scar segmentation accuracy, this study relies on datasets in which all imaging modalities are well-aligned and consistently available. In real-world clinical practice, multi-sequence acquisitions may be incomplete, misaligned, or affected by varying image quality and protocol differences across scanners and institutions. Future work should investigate strategies for handling missing or partially corrupted modalities, such as modality-agnostic networks, cross-modality synthesis, or uncertainty-aware fusion mechanisms.

Second, the experiments in this chapter are conducted using 2D slice-based segmentation. While this choice enables efficient training and fair comparison across strategies, myocardial scarring is inherently a 3D structure with spatial continuity across slices. Extending the proposed framework to 3D architectures may further improve anatomical consistency and volumetric scar quantification, albeit at the cost of increased computational complexity and

memory requirements.

Third, although a wide range of loss functions are explored to address class imbalance, the loss formulations remain hand-crafted and static. Adaptive or data-driven loss weighting strategies, such as curriculum learning, uncertainty-based weighting, or task-aware dynamic loss scheduling, could further enhance sensitivity to rare scar patterns, particularly in small or diffuse infarcts.

In addition, while cascaded architectures did not outperform end-to-end learning in this study, the incorporation of anatomical priors remains a promising direction. Future work may explore more flexible mechanisms for injecting structural information, such as attention-based conditioning, soft anatomical constraints, or latent-space regularization, rather than rigid multi-stage pipelines that are susceptible to error propagation.

Finally, the evaluation in this chapter focuses on benchmark datasets with expert annotations. Comprehensive clinical validation, including inter-observer variability analysis and prospective testing, is required to assess the translational impact of automated scar segmentation in routine clinical workflows. Integrating model outputs into clinical decision making pipelines, such as guiding catheter ablation or supporting risk stratification, represents an important avenue for future investigation.

Overall, addressing these limitations will be critical for advancing automated myocardial scar segmentation from controlled experimental settings toward robust, generalizable, and clinically deployable systems. The subsequent chapters of this thesis directly tackle several of these challenges by introducing clinically guided data augmentation, multimodal integration with electrophysiological signals, and domain generalized learning under missing data and heterogeneous clinical sites.

Chapter 4

Clinically-Guided Scar Synthesis in LGE

This chapter contains material from:

- Ramzan, Farheen, et al. "CLAIM: Clinically-guided LGE augmentation for realistic and diverse myocardial scar synthesis and segmentation." *International Conference on AI in Healthcare*. Cham: Springer Nature Switzerland, 2025.

In this chapter, a fundamental limitation is addressed in data-driven myocardial scar segmentation: the scarcity and imbalance of clinically representative LGE datasets. Although deep learning models have demonstrated strong performance when trained on curated cohorts, their generalization remains constrained by the limited diversity of scar morphologies and distributions encountered during training. In real-world clinical practice, however, myocardial scarring exhibits substantial variability in location, extent, and pattern, with many clinically relevant presentations being underrepresented or entirely absent from available labeled data. To bridge this gap, a clinically-guided scar synthesis framework is introduced for LGE CMR that leverages clinical knowledge of scar morphology and spatial distribution to generate realistic and diverse pathological images. By augmenting training data with clinically plausible synthetic scars rather than relying solely on conventional augmentation techniques, this chapter aims to enhance model robustness and improve segmentation performance on rare and challenging scar presentations.

4.1 Introduction

Automated analysis of LGE CMR imaging has become an increasingly important component of modern cardiac assessment, offering detailed characterization of myocardial injury and fibrosis. In particular, deep learning based segmentation of myocardial scar has shown promise for supporting clinical decision making by enabling reproducible and time-efficient segmentation of pathological tissue. Despite these advances, the performance and reliability of current segmentation models remain highly dependent on the availability and diversity of annotated training data, which poses a major limitation in real-world clinical deployment.

LGE imaging enables visualization of myocardial scar by exploiting delayed contrast washout in regions of fibrosis or necrosis, making it a gold-standard modality for the evaluation of myocardial infarction, cardiomyopathies, and other structural heart diseases [2, 104]. However, accurate identification of hyperenhanced regions is challenging due to ambiguous scar boundaries, partial volume effects, and variations in acquisition protocols across scanners and clinical centers. These technical challenges are further exacerbated by the intrinsic heterogeneity of myocardial scar, which may vary substantially in spatial location, transmural extent, shape, and burden across patients. As a result, scar segmentation remains a particularly demanding task even for state-of-the-art deep learning based methods.

While deep learning approaches have achieved notable success in a wide range of medical image segmentation problems, their application to myocardial scar segmentation has been hindered by limited access to large-scale, consistently annotated datasets [12, 105]. Manual scar annotation requires considerable clinical expertise and is subject to inter-observer variability, making large-scale labeling both time-consuming and expensive. Moreover, privacy regulations and institutional constraints often restrict data sharing across centers [106], leading to fragmented datasets that inadequately capture the full spectrum of clinically observed scar patterns. Consequently, models trained on such datasets may perform well on in-distribution samples while failing to generalize to rare scar presentations.

Synthetic data generation has emerged as a promising strategy for alleviating data scarcity and enhancing the generalization of deep learning models in medical imaging [107]. Recent progress in generative modeling has enabled the synthesis of pathological images across a range of imaging modalities and clinical applications, including tumor generation in microscopy [108], computed tomography [109, 110], and magnetic resonance imaging [111], as well as lesion synthesis in endoscopic and ultrasound imaging [112, 113]. By augmenting training datasets with synthetic samples, these approaches aim to improve model robustness while reducing reliance on extensive manual annotation.

Nevertheless, generating clinically meaningful pathological images remains a non-trivial challenge. Many generative models suffer from unintended modifications to anatomically normal regions, commonly referred to as data hallucination, or fail to capture sufficient diversity in lesion morphology and spatial distribution. More recent work has explored conditional diffusion models as a means of addressing these limitations, enabling controlled pathology synthesis by conditioning on anatomical context or lesion masks [8, 114, 115]. In particular, inpainting-based diffusion strategies have shown improved preservation of background anatomy by explicitly incorporating uncorrupted regions into the diffusion process [116]. Despite these advances, most existing approaches rely on lesion masks derived from limited training data, which may encode dataset specific biases and lack explicit clinical interpretability.

A key observation motivating this work is that myocardial scar formation follows well-established anatomical and clinical patterns that are not explicitly leveraged by most data-driven synthesis methods. In clinical practice, scar distribution is commonly interpreted with respect to standardized left ventricular segmentation schemes, such as the AHA 17-segment model [55]. Incorporating such domain knowledge into the data generation process offers an opportunity to constrain synthetic pathology to clinically plausible regions while encouraging controlled diversity beyond what is observed in available datasets.

Building on this insight, this chapter introduces a clinically guided framework for myocar-

dial scar synthesis and segmentation, termed CLAIM (**C**linically-Guided **L**GE **A**ugmentation for Realistic and Diverse **M**yo**c**ardial Scar Synthesis and Segmentation). The proposed approach leverages conditional diffusion model for LGE image synthesis while explicitly embedding clinical knowledge into the scar generation process. Central to the framework is the SMILE module (Scar Mask generation guided by cLinical knowledgE), which uses the AHA 17-segment model to guide the spatial placement and distribution of synthetic scar patterns within the left ventricular myocardium. This design enables the generation of anatomically consistent yet diverse pathological examples that align with clinically observed scar characteristics.

In addition, the proposed framework adopts a joint optimization strategy in which the scar synthesis model and the segmentation network are trained concurrently. This joint formulation allows the segmentation task to benefit from increasingly realistic synthetic data while simultaneously encouraging the generator to produce scar patterns that are informative for downstream segmentation. The effectiveness of the proposed approach is evaluated using pathological and normal cases from the EMIDEC dataset [10], with further validation on an independent out-of-domain dataset [117, 118] to assess robustness under domain shift. The impact of clinically guided synthesis is also investigated in conjunction with nnU-Net-based segmentation models [27].

The main contributions of this chapter can be summarized as follows:

- A clinically guided framework for myocardial scar synthesis is introduced that integrates anatomical prior knowledge into a conditional diffusion-based generation process.
- The use of standardized left ventricular segmentation (AHA 17-segment model) is proposed to generate anatomically interpretable and spatially diverse synthetic scar patterns.
- It is demonstrated that augmenting training data with clinically guided synthetic LGE images improves myocardial scar segmentation performance in both in-domain and out-of-domain settings.
- A comprehensive evaluation of the proposed framework is provided, highlighting its robustness and practical utility for data limited cardiac imaging applications.

4.2 Background

4.2.1 Clinical Interpretation of Myocardial Scar in LGE Imaging

LGE CMR imaging is widely regarded as the gold standard for non-invasive characterization of myocardial fibrosis and scarring. Following gadolinium administration, damaged myocardial tissue exhibits delayed washout kinetics, resulting in hyperenhanced signal intensity relative to viable myocardium. Numerous studies have demonstrated strong associations between LGE-derived scar burden and clinical outcomes, including ventricular arrhythmias, adverse remodeling, and mortality [31, 32, 119–123].

Crucially, myocardial scar is not a homogeneous entity. Ischemic scars typically follow coronary perfusion territories and present as subendocardial or transmural enhancement,

whereas non-ischemic etiologies such as dilated or hypertrophic cardiomyopathy often exhibit mid-wall, epicardial, or patchy enhancement patterns [124, 125]. Scar morphology, extent, and spatial distribution therefore encode clinically meaningful information related to disease etiology and prognosis, underscoring the importance of anatomically consistent scar representation in automated analysis systems.

To support standardized reporting and interpretation, the AHA 17-segment model provides a clinically accepted framework for partitioning the left ventricular myocardium into anatomically meaningful regions [126]. This model is routinely used for regional assessment of myocardial viability and perfusion, yet remains underutilized in computational approaches to image synthesis and segmentation.

4.2.2 Learning-Based Myocardial Scar Segmentation

Deep learning has transformed medical image segmentation, with CNN such as U-Net and its variants becoming de facto standards [26, 27, 127–130]. In cardiac imaging, these architectures have been adapted for segmentation of ventricular structures and myocardial tissue across multiple modalities [28, 131]. For LGE scar segmentation, several supervised learning approaches have reported encouraging results [12, 105].

However, these methods are heavily dependent on dataset size, annotation quality, and cohort diversity. Scar regions typically occupy a small fraction of the myocardium, resulting in severe class imbalance and sensitivity to annotation noise. Moreover, inter-observer variability in scar segmentation remains substantial, particularly in cases with diffuse fibrosis or low contrast-to-noise ratio [19]. Consequently, models trained on limited datasets often fail to generalize to rare scar phenotypes or data acquired under different imaging protocols.

4.2.3 Data Scarcity and Domain Shift in Cardiac Imaging

The challenges of data scarcity and domain shift are particularly acute in cardiac imaging. Variations in scanner vendors, magnetic field strength, pulse sequences, and contrast timing introduce substantial appearance variability across datasets [11–14]. Privacy regulations and institutional policies further restrict multi-center data sharing, leading to fragmented datasets that inadequately capture population-level variability [106]. These factors collectively limit the scalability and clinical robustness of supervised learning approaches.

As a result, there has been growing interest in methods that reduce reliance on large annotated datasets, including weak supervision, self-supervised learning, and data augmentation. Among these, synthetic data generation has emerged as a particularly promising avenue for expanding pathological diversity without additional annotation cost.

4.2.4 Synthetic Data Generation and Pathology Augmentation

Synthetic data augmentation has been extensively explored in medical imaging to address limited training data. Traditional augmentation strategies, such as spatial transformations and intensity perturbations, improve robustness to minor variations but do not introduce novel pathological structures. GANs have been widely adopted for lesion and pathology synthesis, enabling the generation of visually realistic images across a range of modalities [109, 132–135]. However, GAN-based methods are often prone to mode collapse, training instability, and limited controllability over lesion characteristics.

Diffusion-based generative models have recently emerged as a powerful alternative, offering improved stability and superior sample diversity [46, 136]. These models have been successfully applied to medical image synthesis, reconstruction, and anomaly detection [137, 138]. Their iterative denoising formulation makes them particularly well suited for high-fidelity image generation in data-limited settings.

4.2.5 Conditional Diffusion Models for Lesion and Scar Synthesis

Conditional diffusion models extend the generative framework by incorporating auxiliary conditioning signals, such as anatomical context or lesion masks, to guide the synthesis process. In pathology synthesis, this is commonly implemented using inpainting strategies, where pathological regions are generated within a fixed anatomical background [116]. This approach has been shown to better preserve normal tissue structure while allowing targeted pathology manipulation.

Recent studies have demonstrated the effectiveness of conditional diffusion for lesion synthesis in brain imaging, abdominal imaging, and ultrasound [8, 113–115]. While these methods improve controllability and reduce background artifacts, they typically rely on lesion masks derived from existing datasets. As a result, the diversity and clinical relevance of synthesized lesions remain constrained by dataset specific biases.

4.2.6 Limitations of Data-Driven Mask Generation

A critical bottleneck in conditional pathology synthesis lies in the generation of conditioning masks. Existing approaches commonly perturb or resample lesion masks from training data, implicitly encoding the same anatomical biases and underrepresentation present in the original dataset [8]. Moreover, these masks are rarely constrained by explicit clinical knowledge, raising concerns about semantic validity and anatomical plausibility.

In the context of myocardial scar synthesis, this limitation is particularly consequential. Scar distribution follows well-established anatomical and etiological patterns that are central to clinical interpretation. Ignoring these patterns during synthesis risks generating lesions that are visually plausible but clinically implausible, reducing the utility of synthetic data for training robust segmentation models.

4.2.7 Motivation for This Work

The preceding literature highlights a fundamental gap between advances in generative modeling and the clinical requirements of myocardial scar synthesis. While diffusion-based methods provide powerful tools for high-quality image synthesis, existing approaches lack mechanisms to incorporate standardized anatomical and pathological priors. This disconnect limits both the interpretability of synthetic data and its effectiveness for downstream tasks.

Motivated by these observations, this chapter introduces a clinically guided scar synthesis framework that explicitly integrates the AHA 17-segment model into a conditional diffusion-based generation process. By embedding clinically meaningful anatomical constraints into scar mask generation, the proposed approach aims to produce synthetic LGE images with anatomically consistent, diverse, and clinically interpretable scar patterns. This design directly addresses key limitations of prior work by reducing dataset bias, enhancing pathological diversity, and improving the robustness of deep learning based scar segmentation models.

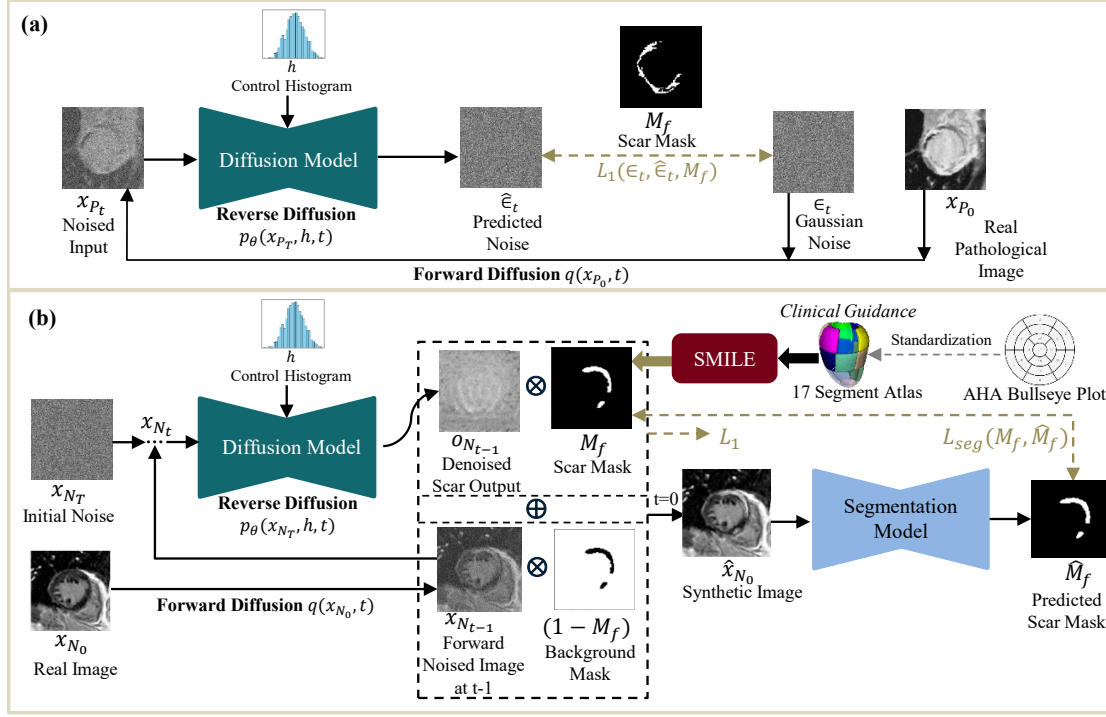


Figure 4.1: Overview of CLAIM. (a) Training of the conditional diffusion model using real pathological LGE images. (b) Generation of synthetic pathological images with diverse and anatomically consistent scar patterns using clinically guided masks produced by SMILE (Sec. 4.3.3). The generated images are used to fine-tune a scar segmentation network, whose feedback further improves synthesis quality through joint training. Figure source: [7], reproduced with permission of the rights holder, Springer Nature.

4.3 Methods

This section presents CLAIM (Clinically-Guided LGE Augmentation for Realistic and Diverse Myocardial Scar Synthesis and Segmentation), a clinically guided framework for myocardial scar synthesis and segmentation in LGE CMR. The overarching goal of CLAIM is to alleviate the data scarcity and dataset bias that limit the generalization of deep learning based scar segmentation models. Unlike conventional data augmentation pipelines, which operate purely at the image level, CLAIM integrates clinical prior knowledge, generative modeling, and task-driven optimization into a unified framework.

At a high level, CLAIM consists of three tightly coupled components: (i) a conditional diffusion-based image synthesis model for generating realistic pathological LGE images, (ii) a clinically guided scar mask generation module (SMILE: Scar Mask generation guided by cLinical knowldgE), and (iii) a deep scar segmentation network. These components are jointly optimized through an alternating training strategy that enables mutual reinforcement between scar synthesis quality and segmentation accuracy. An overview of the framework is illustrated in Fig. 4.1, highlighting the flow of information between real data, synthetic data generation, and segmentation-driven feedback.

4.3.1 Design Rationale and Architectural Choices

The design of CLAIM is driven by the dual objectives of generating clinically plausible myocardial scar patterns and improving the robustness of downstream scar segmentation in data-limited settings. Rather than treating image synthesis and segmentation as independent stages, the proposed framework was deliberately structured to integrate clinical knowledge, generative modeling, and task-driven learning into a unified system. This section articulates the rationale behind the key architectural and methodological choices made in the development of CLAIM.

Choice of Diffusion-Based Generative Modeling

A central design decision in CLAIM is the adoption of diffusion-based generative models for pathological image synthesis. Alternative generative approaches, most notably GANs, have been widely used for medical image synthesis. However, GAN-based methods are known to suffer from training instability, mode collapse, and limited controllability over the generated pathology, particularly in scenarios involving small and heterogeneous targets such as myocardial scars.

Diffusion models offer several advantages that align with the requirements of myocardial scar synthesis. First, their iterative denoising formulation enables stable training and improved coverage of the data distribution, reducing the risk of generating repetitive or unrealistic scar patterns. Second, diffusion models naturally support conditional generation and inpainting, allowing pathological regions to be synthesized while preserving background anatomy. This property is particularly important for LGE imaging, where subtle intensity differences between healthy and scarred myocardium must be preserved with high fidelity. Finally, diffusion models provide a principled mechanism for incorporating spatial constraints, making them well suited for controlled pathology synthesis.

Conditional Inpainting versus Full Image Generation

Another key design choice is the use of a conditional inpainting strategy rather than full-image synthesis. In CLAIM, the diffusion model is explicitly constrained to operate within predefined scar regions, while non-pathological regions are preserved through forward-diffused background conditioning. This choice was motivated by both anatomical and practical considerations.

Full-image generation approaches risk introducing unintended alterations to background myocardial tissue, blood pools, or surrounding structures, which can degrade anatomical realism and reduce the usefulness of synthetic data for segmentation training. In contrast, inpainting-based synthesis ensures that only the pathological regions are modified, maintaining consistency with the original subject-specific anatomy. This is particularly important in cardiac imaging, where small spatial inconsistencies can significantly impact segmentation performance. By constraining synthesis to scar regions, CLAIM achieves fine-grained control over pathological appearance while minimizing the risk of background hallucination.

Incorporation of Clinical Anatomical Priors

A defining feature of CLAIM is the explicit incorporation of clinical anatomical knowledge through the AHA 17-segment model. Rather than relying solely on data-driven lesion mask generation, SMILE embeds standardized myocardial segmentation into the synthesis process. This design choice reflects the observation that myocardial scar distribution is not arbitrary, but follows well-established anatomical and etiological patterns that are central to clinical interpretation.

Using the AHA 17-segment model provides several advantages. It enables controllable and interpretable scar placement, facilitates systematic exploration of rare or underrepresented pathological patterns, and aligns synthetic data generation with clinical reporting conventions. Importantly, this approach decouples pathological diversity from dataset-specific biases, allowing CLAIM to generate scar patterns beyond those observed in the available training data. The use of a standardized anatomical model ensures that generated scars remain semantically meaningful and clinically grounded.

Joint Optimization of Synthesis and Segmentation

A further critical design decision is the joint optimization of the diffusion-based synthesis model and the scar segmentation network. In conventional pipelines, synthetic data are generated offline and subsequently used to train segmentation models. This decoupled approach assumes that visually realistic images necessarily improve segmentation performance, an assumption that does not always hold in practice.

By contrast, CLAIM integrates segmentation feedback directly into the synthesis process, enabling task-aware generative modeling. The segmentation network provides an additional supervisory signal that encourages the diffusion model to generate pathological patterns that are informative for segmentation, particularly in challenging scenarios such as small or low-contrast scars. This joint training paradigm aligns the objectives of synthesis and analysis, reducing the risk of generating visually plausible but clinically uninformative data. While joint optimization introduces additional training complexity, the observed improvements in segmentation robustness justify this design choice.

Selection of Segmentation Backbone

The segmentation component of CLAIM is implemented using nnU-Net [27] (refer back to Chapter 3), a widely adopted self-configuring segmentation framework. This choice was motivated by the desire to ensure a strong and fair baseline, minimizing the influence of architectural tuning on performance comparisons. By using nnU-Net, the proposed framework can be evaluated primarily on the basis of data augmentation and synthesis strategy rather than segmentation network design.

Moreover, nnU-Net’s robustness across diverse datasets and imaging modalities makes it a suitable candidate for assessing generalization under domain shift. The use of a standardized segmentation backbone also facilitates reproducibility and enables direct comparison with existing literature.

Trade-offs and Design Constraints

The design of CLAIM reflects a series of trade-offs between controllability, realism, and computational complexity. While clinically guided mask generation and joint training introduce additional computational overhead, they provide substantial gains in interpretability and segmentation performance. Similarly, reliance on anatomical registration introduces sensitivity to registration quality, but this is mitigated by the use of robust registration algorithms and by the clinical relevance of the resulting anatomical constraints.

Overall, the architectural choices embodied in CLAIM represent a deliberate balance between methodological rigor and practical applicability. By integrating clinical knowledge with modern generative modeling and task-driven learning, the proposed framework addresses key limitations of existing pathology synthesis approaches and provides a scalable foundation for robust myocardial scar segmentation.

4.3.2 Joint Modelling for Enhanced Scar Synthesis and Segmentation

Diffusion-Based Pathology Synthesis

The generative backbone of CLAIM is based on a conditional denoising diffusion probabilistic model, adapted from LeFusion [8]. Diffusion models define a Markov chain that gradually corrupts an input image with Gaussian noise (forward diffusion) and subsequently learns to reverse this process to recover a clean sample (reverse diffusion) [46, 136]. Compared to adversarial generative models, diffusion models offer improved training stability, reduced mode collapse, and superior sample diversity [139, 140], making them well suited for medical image synthesis.

Let x_{P_0} denote a pathological LGE image containing myocardial scar. The forward diffusion process $q(x_{P_t} | x_{P_0})$ progressively adds Gaussian noise $\epsilon_t \sim \mathcal{N}(0, I)$ over T time steps:

$$x_{P_t} = \sqrt{\alpha_t}x_{P_0} + \sqrt{1 - \alpha_t}\epsilon_t,$$

where α_t controls the noise schedule. The reverse diffusion process $p_\theta(x_{P_{t-1}} | x_{P_t})$ is parameterized by a U-Net architecture [26] and trained to predict the injected noise ϵ_t .

To restrict synthesis to pathological regions, diffusion is conditioned on a scar mask M_f and a lesion texture control histogram h derived from real data, following LeFusion [8]. Concretely, h is a fixed-length vector built by binning the grayscale LGE intensities of voxels that fall inside the scar mask of a real pathological case. It therefore captures the intensity profile of scar tissue on its own, decoupled from the surrounding myocardium. Because this work targets scar as a single lesion class, h is computed only over scar voxels. At training time, h is taken directly from the ground-truth scar region of the sample being denoised and passed to the U-Net alongside M_f as an extra condition. It is injected into the intermediate feature maps through cross-attention [141], which nudges the noise prediction towards textures whose intensity distribution is consistent with h . The reverse step can therefore be written as

$$o_{t-1} \sim p_\theta(\hat{x}_t, h, t).$$

At inference, this same slot can be filled either with a histogram taken from a real reference scar or with one sampled from the empirical distribution of h over the training set, giving

control over the appearance of the synthesised lesion without needing any explicit scar-type labels. A scar-focused noise prediction loss is then applied:

$$L_1(\epsilon_t, \hat{\epsilon}_t; M_f) = \|M_f \odot \epsilon_t - M_f \odot \hat{\epsilon}_t\|_2^2. \quad (4.1)$$

This formulation ensures that the diffusion model manipulates only the scar regions while preserving the background myocardium and surrounding anatomy.

Inpainting-Based Background Preservation

During reverse diffusion, CLAIM adopts an inpainting-based formulation to prevent unintended alterations to non-scar regions. At each diffusion step, synthesized scar content is combined with forward-noised background regions:

$$\hat{x}_{N_{t-1}} = o_{N_{t-1}} \odot M_f + x_{N_{t-1}} \odot (1 - M_f),$$

where $o_{N_{t-1}} \sim p_\theta$ denotes the synthesized scar prediction and $x_{N_{t-1}}$ is obtained via forward diffusion from the original image. This approach explicitly enforces anatomical consistency and has been shown to reduce background hallucination in conditional diffusion models [116].

Integration of Segmentation Feedback

A key innovation of CLAIM is the integration of a scar segmentation network into the diffusion training loop. The segmentation model, based on nnU-Net [27], is initially pretrained using real pathological cases to establish a strong baseline. During joint training, synthetic pathological images generated by the diffusion model are used to fine-tune the segmentation network online.

Given a synthetic image $\hat{x}_{N_0}$ and corresponding scar mask M_f , the segmentation network predicts $\hat{M}_f$. Segmentation performance is optimized using:

$$L_{\text{seg}}(M_f, \hat{M}_f) = L_{\text{Dice}}(M_f, \hat{M}_f) + L_{\text{WCE}}(M_f, \hat{M}_f), \quad (4.2)$$

where L_{Dice} mitigates class imbalance and L_{WCE} penalizes false negatives in small scar regions.

This joint optimization strategy serves two purposes. First, it improves segmentation robustness by exposing the model to a broader and more diverse set of scar patterns. Second, it indirectly regularizes the diffusion model by encouraging synthesis of scars that are informative for segmentation, rather than merely visually plausible.

Training Schedule and Alternating Optimization

Training proceeds in an alternating fashion. At each iteration, the diffusion model generates a batch of synthetic pathological images using scar masks produced by SMILE. These images are used to update the segmentation network, after which the updated segmentation performance influences subsequent diffusion updates. This alternating scheme prevents either component from dominating the optimization process and promotes gradual convergence toward clinically meaningful synthesis and segmentation.

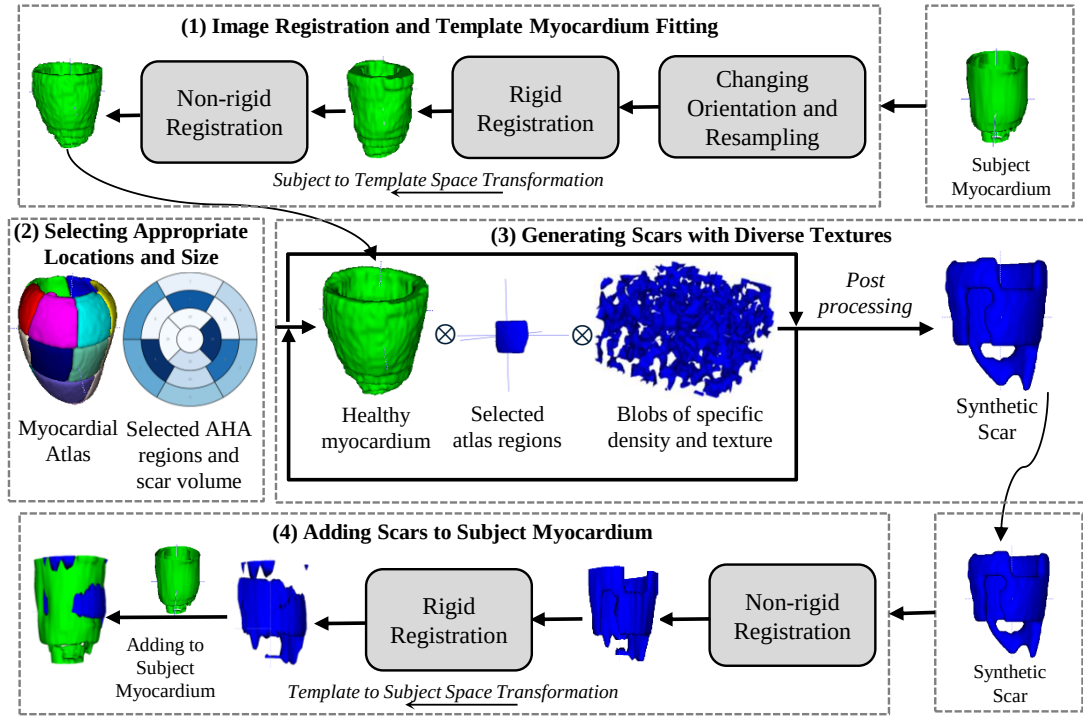


Figure 4.2: Overview of SMILE (Scar Masks Incorporating cLinical knowledgeE).
Figure source: [7], reproduced with permission of the rights holder, Springer Nature.

4.3.3 SMILE: Scar Masks Incorporating cLinical knowledgeE

Motivation for Clinically Guided Mask Generation

Existing pathology synthesis frameworks typically rely on lesion masks derived from training datasets or random perturbations thereof. Such approaches implicitly encode dataset biases and lack explicit guarantees of clinical plausibility. In myocardial scar synthesis, this limitation is particularly problematic, as scar distribution follows well-established anatomical and etiological patterns that are central to clinical interpretation.

SMILE addresses this limitation by embedding clinical knowledge directly into the scar mask generation process. By leveraging the AHA 17-segment model [126], SMILE ensures that synthesized scars are spatially consistent with standardized left ventricular anatomy.

Image Registration and Myocardium Fitting

To enable consistent mapping between subject-specific anatomy and the AHA template, myocardial masks are registered to a reference atlas [142]. The registration pipeline includes orientation normalization, isotropic resampling, rigid alignment, and non-rigid deformation using the fast symmetric forces Demons algorithm [143]. This process establishes voxel-wise correspondence between subject myocardium and AHA segments, enabling anatomically controlled scar placement.

Region Selection and Scar Extent Control

Scar regions are selected by sampling one or more adjacent AHA segments across basal, mid-ventricular, and apical slices. Adjacency constraints are enforced using region connectivity graphs to prevent anatomically implausible disjoint scars. Scar extent is controlled by sampling region-specific volumes from predefined distributions, allowing systematic variation in scar burden across subjects.

Scar Texture and Morphological Modeling

Within selected regions, three-dimensional scar blobs are generated with controllable anisotropy, porosity, and texture. Morphological erosion with randomly selected kernels introduces boundary irregularity, mimicking clinically observed scar heterogeneity. Post-processing operations, including Gaussian smoothing and hole filling, ensure spatial coherence and remove isolated artifacts.

Transformation to Subject Space and Integration

Generated scar masks are transformed back to subject space using inverse non-rigid and rigid registration. Scars may be inserted into healthy myocardium or combined with existing pathological masks to augment scar extent. This flexibility enables SMILE to support both de novo pathology synthesis and controlled augmentation of existing scars.

Advantages of SMILE over Data-Driven Mask Generation

By incorporating standardized anatomical priors, SMILE offers several advantages over purely data-driven mask generation methods. It enables controllable scar placement, systematic exploration of rare pathological patterns, and improved interpretability of synthetic data. Importantly, SMILE decouples pathological diversity from dataset bias, allowing CLAIM to generate clinically meaningful scars beyond those observed in available datasets.

4.3.4 Summary of the CLAIM Framework

In summary, CLAIM integrates clinically guided scar mask generation, diffusion-based image synthesis, and segmentation-driven feedback into a unified framework for myocardial scar augmentation and segmentation. By embedding clinical knowledge directly into the generative process and aligning synthesis objectives with downstream segmentation performance, CLAIM addresses key limitations of existing pathology synthesis approaches and provides a robust foundation for training generalizable scar segmentation models.

4.4 Experiments

This section describes the datasets, experimental design, and comparison protocols used to evaluate the proposed CLAIM framework. The experimental setup is designed to assess three key aspects of the proposed approach: (i) the realism and diversity of synthesized myocardial scar patterns, (ii) the impact of clinically guided synthetic data on downstream scar segmentation performance, and (iii) the robustness and generalizability of the trained

Table 4.1: Characteristics of the LGE-MRI datasets used in this study. Table source: [7], reproduced with permission of the rights holder, Springer Nature.

Dataset	Subjects		Pixel Spacing (mm)	Slice Thickness (mm)	Slice Distance (mm)	Scar Volume (mean $\pm$ std)
	Normal	Pathological				
EMIDEC [10]	33	67	1.25 $\times$ 1.25, 2 $\times$ 2	8	10	23.68 $\pm$ 15.81
Private [117, 118]	0	80	0.89 $\times$ 0.89, 1.7 $\times$ 1.7	8-10	10	14.33 $\pm$ 13.62

segmentation models under domain shift. To this end, experiments were conducted on both a publicly available dataset and an independent private clinical dataset acquired at a different institution.

4.4.1 Datasets

Public Dataset (In-Domain)

The EMIDEC dataset [10] (introduced in Chapter 2) is a publicly available benchmark dataset for myocardial infarction analysis using delayed enhancement MRI. It comprises 100 short-axis LGE-MRI scans, including 67 pathological cases and 33 normal cases. Each scan is accompanied by expert annotations for multiple anatomical and pathological structures, including the LV cavity, healthy myocardium (MYO), myocardial infarction, and persistent microvascular obstruction (PMO). This study focuses exclusively on the MI class, as it represents the primary target pathology for scar synthesis and segmentation. EMIDEC exhibits moderate inter-subject variability in scar size and distribution, making it suitable for controlled in-domain evaluation while still reflecting clinically relevant challenges.

Private Dataset (Out-of-Domain)

To evaluate robustness and generalization under domain shift, furthermore, a private dataset was utilized which consists of 80 pathological LGE-MRI scans provided by Nottingham University Hospitals [117, 118]. All scans were acquired using clinical imaging protocols distinct from EMIDEC (introduced in Chapter 2) and were manually annotated by experienced cardiologists. This dataset contains only pathological cases and exhibits notable differences in image resolution, acquisition parameters, and scar burden compared to EMIDEC. Importantly, this dataset is used *exclusively* for testing and is never seen during training, thereby enabling a rigorous assessment of out-of-domain performance.

A summary of dataset characteristics, including subject counts, spatial resolution, slice parameters, and scar volume statistics, is provided in Table 4.1.

4.4.2 Experimental Setup

From the 67 pathological EMIDEC cases, 57 subjects were used for training the image synthesis models, while the remaining 10 pathological cases were reserved for in-domain segmentation evaluation. The 33 normal cases were split into 28 subjects for image synthesis model training and 5 subjects for validation during joint training. This split ensures that both pathological and non-pathological images are available for clinically guided scar synthesis, while maintaining strict separation between training and evaluation data.

Synthetic images generated by CLAIM were used to train segmentation models under different real and synthetic data composition settings, allowing assessment of the contribution of synthetic data to downstream performance. Segmentation evaluation was conducted on two test sets: (i) the 10 held-out pathological EMIDEC cases (in-domain evaluation), and (ii) the 80 pathological cases from the private dataset (out-of-domain evaluation).

Preprocessing and Data Augmentation

For image synthesis, LGE images were cropped around the heart region using bounding boxes derived from the corresponding myocardial masks. Image intensities were rescaled to the range $[-1, 1]$ to stabilize diffusion model training. Random flipping was applied as data augmentation during synthesis model training to increase variability without altering anatomical plausibility.

For segmentation, images were normalized using z-score normalization on a per-volume basis. Data augmentation was applied on the fly during training using the standard nnU-Net augmentation pipeline [27], including random rotations, scaling, elastic deformations, intensity perturbations, and mirroring. This augmentation strategy was applied consistently across all segmentation experiments to ensure fair comparison.

Training Configuration

The image synthesis models were trained for 50,000 epochs using the Adam optimizer with a learning rate of 1×10^{-4} , following the training protocol adopted in LeFusion [8]. Segmentation models were trained for 1000 epochs using SGD with an initial learning rate of 1×10^{-2} . The learning rate was decayed using the polynomial learning rate schedule:

$$\text{lr}_{\text{epoch}} = \text{lr}_0 \left(1 - \frac{\text{epoch}}{\text{epoch}_{\text{max}}} \right)^{0.9}.$$

All models were trained until convergence, and the best-performing checkpoints were selected based on validation performance.

4.4.3 Methods for Comparison

To comprehensively evaluate the proposed approach, CLAIM is compared against state-of-the-art methods for scar mask generation, image synthesis, and segmentation.

Scar Mask Generation

For scar mask generation, SMILE is compared with DiffMask, the mask generation strategy used in LeFusion [8]. DiffMask employs a control sphere and boundary smoothing operations to generate lesion masks with varying size and location. While effective for generic lesion synthesis, DiffMask does not incorporate explicit anatomical or clinical constraints, limiting its ability to generate semantically meaningful myocardial scar distributions.

Image Synthesis

For pathological image synthesis, CLAIM is compared with LeFusion [8], a strong diffusion-based baseline for lesion-focused synthesis. This comparison allows to isolate the effect of clinically guided mask generation and joint training from the underlying diffusion architecture.

CLAIM Variants (Ours)

To study the contribution of joint optimization, two variants of the proposed framework are evaluated. CLAIM denotes the version in which only the diffusion-based generator is trained, without segmentation feedback. CLAIM (J) denotes the full framework with joint training of the diffusion model and the segmentation network. This comparison enables assessment of whether segmentation-aware feedback improves synthesis quality and downstream performance.

Segmentation Baseline

As a reference point, a baseline scar segmentation model was trained using only real pathological EMIDEC cases, without any synthetic data augmentation. This baseline represents standard supervised learning and provides a lower bound for evaluating the benefit of clinically guided synthetic data.

Together, these comparisons allow to disentangle the effects of clinical guidance, diffusion-based synthesis, and joint training, providing a comprehensive evaluation of the proposed CLAIM framework.

4.5 Results and Discussion

This section presents a comprehensive evaluation of the proposed CLAIM framework, focusing on both qualitative and quantitative analyses. Firstly, the visual realism and anatomical plausibility of the synthesized scar masks and pathological images is assessed. Then, the effectiveness of CLAIM is evaluated for downstream myocardial scar segmentation under both in-domain and out-of-domain settings. Finally, predicted scar volume distributions are analyzed using the clinically adopted AHA 17-segment model to gain deeper insight into regional performance characteristics and remaining challenges.

4.5.1 Visual Quality Assessment on Image Synthesis

Effectiveness of SMILE with Clinical Knowledge

To evaluate the realism and anatomical plausibility of synthesized scar masks, real scar annotations were visually compared with synthetic masks generated by DiffMask., as used in LeFusion [8], and by the proposed SMILE module. Representative examples are shown in Fig. 4.3. This qualitative comparison reveals clear differences in the spatial distribution, morphology, and anatomical consistency of the generated scar masks.

The scar masks generated by SMILE exhibit substantial diversity in shape, size, and spatial location while remaining confined to anatomically plausible regions of the left ventricular

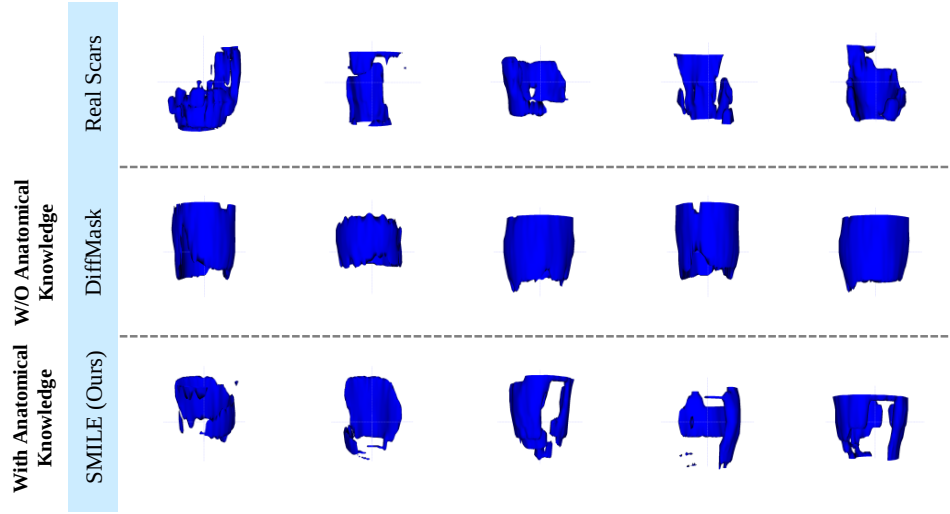


Figure 4.3: Visualization of real scar masks and synthetic masks generated by DiffMask [8] and SMILE (ours). Figure source: [7], reproduced with permission of the rights holder, Springer Nature.

myocardium. Importantly, the synthesized scars follow clinically meaningful patterns that resemble real infarct distributions, including partial-thickness and transmural configurations distributed across basal, mid-ventricular, and apical segments. This behavior reflects the explicit incorporation of standardized anatomical priors via the AHA 17-segment model.

In contrast, DiffMask relies on a generic spherical control mechanism combined with boundary smoothing operations to regulate lesion size and placement. While such an approach is effective for organs with relatively homogeneous and enclosed geometry, such as the lungs or kidneys, it is poorly suited to the complex hollow cylindrical structure of the left ventricular myocardium. As a result, DiffMask-generated scars frequently appear overly large, spatially diffuse, or anatomically implausible, occasionally spanning unrealistic portions of the myocardium or covering nearly the entire myocardial wall. Similar failure modes are also evident in the myocardial examples reported in the original LeFusion work (Fig. A2 in [8]).

The observed differences highlight a key limitation of data-driven mask generation approaches that do not incorporate explicit clinical knowledge. By contrast, SMILE leverages decades of clinical consensus embedded in the AHA segmentation scheme, enabling the generation of scar masks that are not only visually realistic but also clinically interpretable. This qualitative improvement in mask generation provides a strong foundation for downstream pathological image synthesis and segmentation.

Effectiveness of Joint Training

Next, the impact of the proposed joint training paradigm on pathological image synthesis quality is investigated. Fig. 4.4 presents synthetic pathological images generated from normal input images using different variants of the proposed framework, along with corresponding difference images highlighting intensity changes within scar regions.

Across all methods, background myocardial and non-cardiac regions remain well pre-

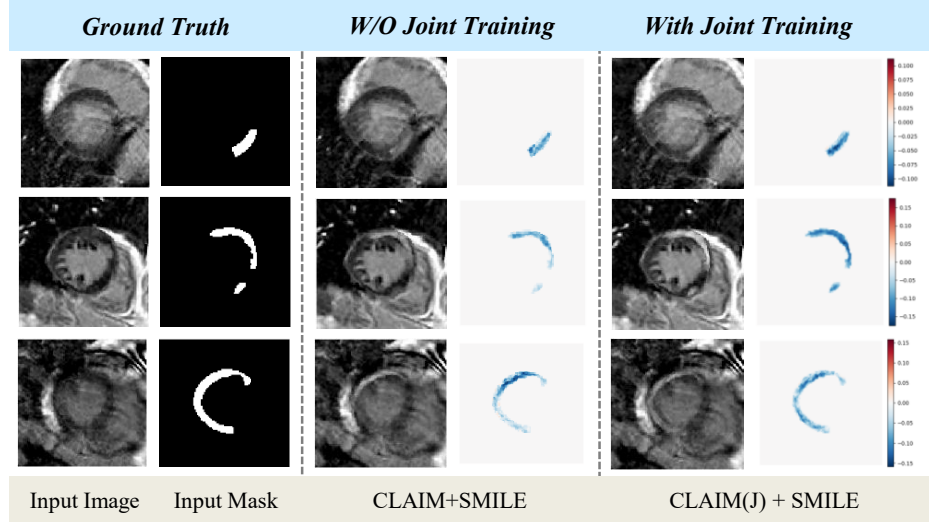


Figure 4.4: Visualization of synthetic pathological images generated by different methods from normal images and scar masks. Difference images highlight intensity changes in scar regions. Figure source: [7], reproduced with permission of the rights holder, Springer Nature.

served, indicating that the inpainting-based diffusion formulation effectively constrains synthesis to the specified scar regions. However, notable differences emerge in the intensity and contrast characteristics of the synthesized scars. Models trained without joint optimization tend to generate scars with weaker contrast enhancement, particularly for small or thin scar regions. This behavior suggests that diffusion-only training may prioritize visual smoothness over pathological salience.

In contrast, CLAIM with joint training consistently produces brighter and more clearly delineated scar regions, closely resembling the contrast characteristics observed in real LGE images. The improvement is particularly pronounced for small or subtle scar regions, which are often the most challenging to model and segment. This observation supports the hypothesis that segmentation-aware feedback encourages the diffusion model to generate pathological patterns that are not only visually plausible but also informative for the downstream task.

These results demonstrate that joint optimization acts as a form of task-driven regularization for the generative model, aligning image synthesis objectives with clinically relevant segmentation performance rather than purely perceptual realism.

4.5.2 Effectiveness of CLAIM for Scar Segmentation

To assess the utility of CLAIM for downstream scar segmentation, nnU-Net based segmentation models were trained under different real and synthetic data configurations. All segmentation backbones were identical, ensuring that observed performance differences could be attributed solely to data augmentation strategy and training paradigm. Comparisons were made against the baseline nnU-Net trained only on real data and against LeFusion-based augmentation using both DiffMask and SMILE.

Table 4.2: Scar segmentation results on EMIDEC dataset [10]. *P*: 57 real pathological cases. *N'*: (28×2) synthetic pathological cases generated from 28 normal cases. *N''*: (28×5) synthetic pathological cases generated from 28 normal cases. *P'*: (57×1) synthetic pathological cases generated from 57 real pathological cases. CLAIM: without joint training, CLAIM (J): with joint training. Table source: [7], reproduced with permission of the rights holder, Springer Nature.

Methods	Scar Mask	Dataset	Dice Score	Precision	Sensitivity	Specificity
Baseline	—	P	58.89	70.99	62.35	99.35
LeFusion [8]	DiffMask	P+N’	59.50	68.48	62.68	99.02
LeFusion	SMILE		59.33	69.71	63.97	98.98
CLAIM (Ours)			62.04	73.28	63.65	99.03
CLAIM (J) (Ours)			63.07	74.13	63.52	99.11
LeFusion [8]	DiffMask[8]	P+N’’	60.43	68.96	63.40	99.04
LeFusion	SMILE		60.09	70.22	64.34	99.03
CLAIM (Ours)			62.98	73.73	63.51	99.08
CLAIM (J) (Ours)			63.27	73.77	64.32	99.07
LeFusion [8]	DiffMask[8]	P+P’+N’’	60.91	71.95	62.34	99.04
LeFusion	SMILE		62.63	71.14	65.69	98.99
CLAIM (Ours)			62.87	73.61	64.05	98.99
CLAIM (J) (Ours)			63.53	73.88	64.71	99.07

In-Domain Evaluation on EMIDEC

Quantitative segmentation results on the in-domain EMIDEC test set are summarized in Table 4.2. Across all training configurations, models trained with clinically guided synthetic data outperform the baseline model trained exclusively on real pathological cases. In particular, CLAIM (J) consistently achieves the highest Dice scores and precision values, demonstrating improved overlap and reduced false-positive predictions.

As the amount of synthetic data increases, segmentation performance improves steadily, highlighting the benefit of augmenting limited real datasets with diverse synthetic examples. Notably, replacing DiffMask with SMILE within the LeFusion framework already yields measurable performance gains, underscoring the importance of anatomically meaningful mask generation. However, the largest improvements are observed when both SMILE and joint training are employed, confirming that clinical guidance and segmentation-aware synthesis provide complementary benefits.

Sensitivity and specificity metrics indicate that the proposed approach achieves a favorable balance between detecting small scar regions and avoiding over-segmentation. The consistent improvement across multiple metrics suggests that CLAIM enhances both boundary accuracy and overall robustness.

Out-of-Domain Generalization

Robustness under domain shift was evaluated using the independent private dataset comprising 80 pathological LGE-MRI scans [117, 118]. Fig. 4.5 compares segmentation performance on in-domain EMIDEC data and out-of-domain private data across different synthetic data configurations.

On the in-domain dataset, CLAIM consistently outperforms the baseline and LeFusion-based methods as synthetic data volume increases. On the out-of-domain dataset, a similar

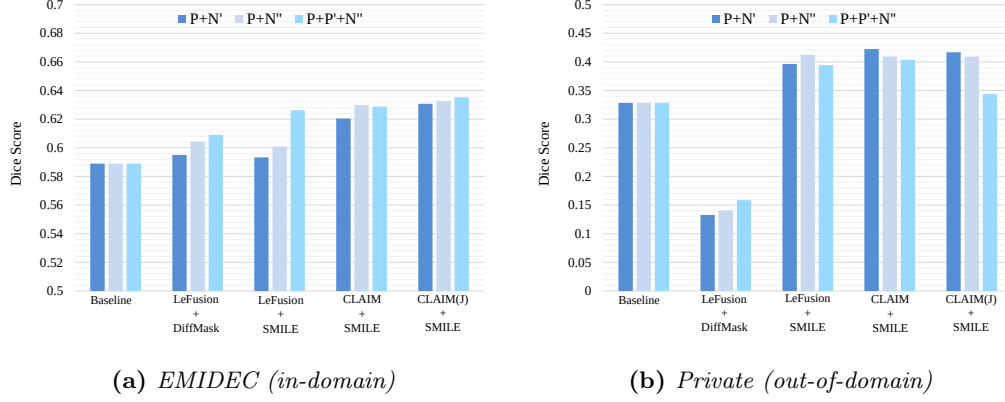


Figure 4.5: Scar segmentation performance across different synthetic data configurations on in-domain and out-of-domain datasets. Figure source: [7], reproduced with permission of the rights holder, Springer Nature.

trend is observed initially: moderate amounts of clinically guided synthetic data significantly improve generalization compared to the baseline. However, performance begins to degrade as the volume of synthetic data increases further.

This phenomenon can be attributed to differences in scar volume distribution between the EMIDEC and private datasets (Table 4.1). Excessive augmentation using synthetic scars derived from EMIDEC statistics may inadvertently introduce new biases, particularly when synthetic scars disproportionately represent larger or more extensive infarcts. Interestingly, the joint training variant does not consistently outperform its non-joint counterpart on out-of-domain data, suggesting that tighter coupling between synthesis and segmentation may amplify dataset-specific characteristics.

These findings highlight an important practical consideration: while synthetic data can substantially improve generalization, its distribution must be carefully controlled to avoid overfitting to source-domain pathology characteristics. Future work may address this limitation through adaptive re-sampling, domain-aware synthesis, or volume-matched augmentation strategies.

Comparison with Baseline

To further contextualize the performance gains achieved by the proposed framework, a direct comparison is conducted between our best-performing model and a baseline segmentation model trained exclusively on real pathological cases. The baseline represents a conventional supervised learning scenario in which no synthetic data augmentation is employed, thereby providing a meaningful reference for assessing the robustness and practical utility of the proposed approach. Since the baseline model is trained only on authentic pathological images, this comparison specifically evaluates whether incorporating clinically guided synthetic data improves segmentation performance on *real* pathological cases rather than introducing artifacts or overfitting to synthetic patterns.

Fig. 4.6 presents a case-wise comparison of segmentation performance between the baseline and the proposed CLAIM (J) model on 10 held-out pathological cases from the EMIDEC

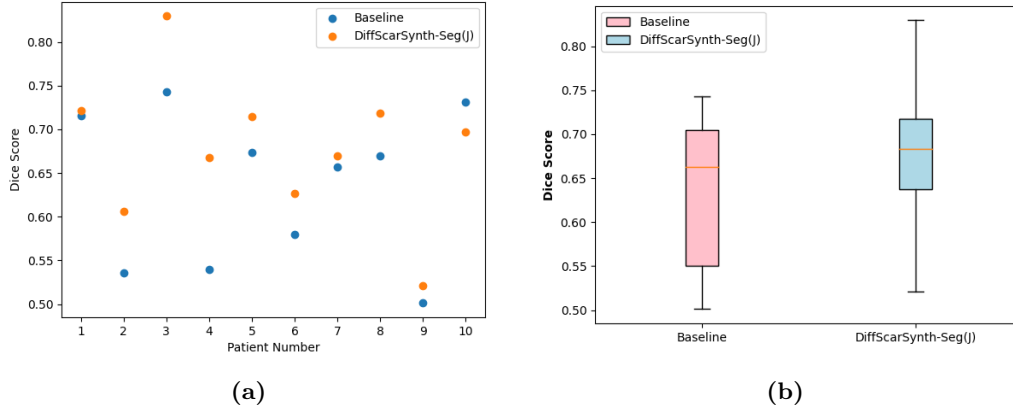


Figure 4.6: Comparing the relative performance of our best method with baseline on EMIDEC test dataset. (a) A scatter plot showing dice score for each patient. (b) A box plot showing the distribution of dice scores.

dataset. The scatter plot in Fig. 4.6a illustrates the Dice scores achieved by both methods for each individual case. It is evident that the proposed model outperforms the baseline in 9 out of 10 cases, indicating a consistent improvement across the majority of subjects rather than isolated gains driven by a small subset of cases. This observation suggests that the performance improvement is systematic and robust, rather than being attributable to favorable outliers.

In addition, the Dice score distribution shown in Fig. 4.6b further highlights the stability of the proposed approach. Compared to the baseline, CLAIM (J) exhibits a higher median Dice score and reduced performance variability, indicating more reliable segmentation across cases with varying scar size, morphology, and contrast characteristics. The improved consistency is particularly important in clinical settings, where unpredictable failure cases can limit trust in automated systems. Collectively, these results demonstrate that clinically guided synthetic data augmentation enhances segmentation robustness on real pathological data, even when evaluated under strictly supervised test conditions.

Evaluation of Predicted Scar Volume Distribution

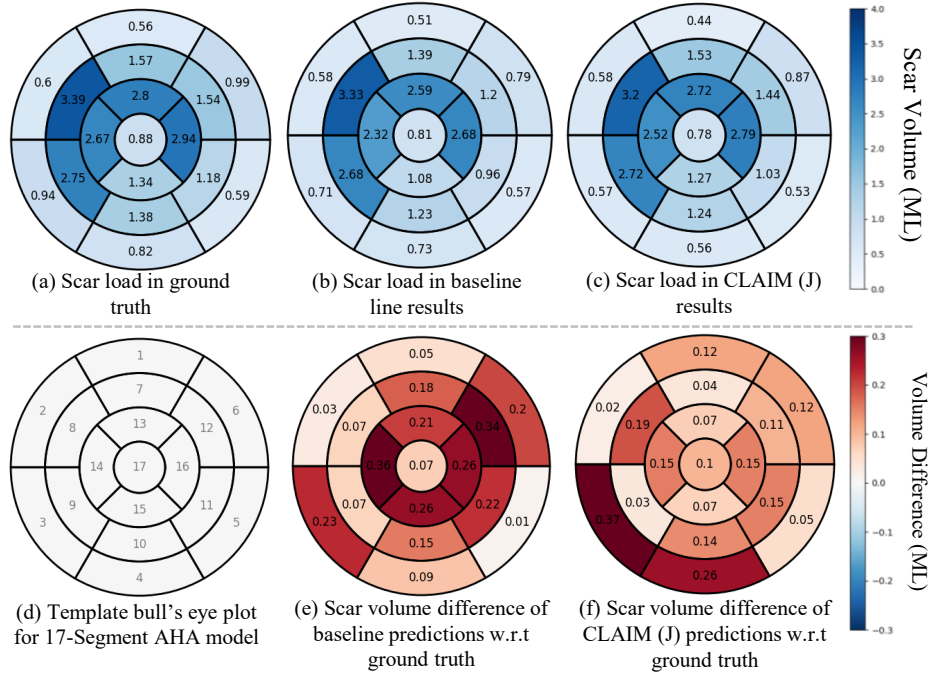
Beyond voxel-wise overlap metrics, segmentation performance is further evaluated by analyzing predicted scar volume distributions across the AHA 17 myocardial segments. Fig. 4.7 presents bull’s-eye plots summarizing average scar volume per segment for the ground truth, baseline segmentation, and CLAIM (J) predictions.

The baseline model tends to underestimate scar volume in mid-ventricular and apical regions, which are known to be challenging due to thinner myocardial walls and partial volume effects. In contrast, CLAIM (J) exhibits closer alignment with ground truth volumes across these regions, indicating improved sensitivity to anatomically subtle scar patterns. Differences between predicted and ground truth volumes provide insight into region-specific failure modes and highlight areas where additional targeted augmentation may be beneficial.

This regional analysis demonstrates that CLAIM not only improves global segmentation metrics but also enhances anatomical fidelity at a clinically meaningful level. Such insights

are critical for guiding future refinement of synthetic data generation strategies toward more balanced and representative pathology distributions.

While quantitative metrics such as Dice score, precision, and sensitivity provide objective measures of segmentation performance, their clinical relevance must be interpreted in the context of how myocardial scar information is used in practice. This subsection discusses the clinical implications of the observed results, focusing on anatomical fidelity, regional performance, and potential impact on clinical decision-making.



Standardized Nomenclature for Tomographic Imaging of the Heart

<u>Basal (B)</u>	<u>Middle (M)</u>	<u>Apical (A)</u>
1 B anterior	7 M anterior	13 A anterior
2 B anteroseptal	8 M anteroseptal	14 A septal
3 B inferoseptal	9 M inferoseptal	15 A inferior
4 B inferior	10 M inferior	16 A lateral
5 B inferolateral	11 M inferolateral	17 Apex
6 B anterolateral	12 M anterolateral	

Figure 4.7: Bull's-eye plots illustrating regional scar volume distribution across AHA 17 myocardial segments on the EMIDEC test dataset. Figure source: [7], reproduced with permission of the rights holder, Springer Nature.

4.5.3 Clinical Interpretation of Results

Anatomical Consistency and Clinical Plausibility

One of the primary clinical requirements for automated myocardial scar segmentation is anatomical plausibility. Scar patterns that violate known myocardial anatomy or present unrealistic spatial distributions may undermine clinician trust, regardless of numerical per-

formance metrics. The qualitative and quantitative results presented in this chapter indicate that CLAIM generates and segments scars that are anatomically consistent with standardized left ventricular segmentation, as defined by the AHA 17-segment model.

The improved alignment between predicted and ground-truth scar distributions across basal, mid-ventricular, and apical regions suggests that the proposed framework captures clinically meaningful spatial patterns rather than merely optimizing voxel-wise overlap. This is particularly important for LGE imaging, where regional scar localization is routinely used for reporting, prognosis, and therapeutic planning.

Regional Performance and Its Clinical Significance

The regional scar volume analysis using AHA bull’s-eye plots reveals that CLAIM achieves notable improvements in the mid-ventricular and apical segments compared to the baseline model. From a clinical perspective, these regions are often more challenging to assess due to thinner myocardial walls, partial volume effects, and increased sensitivity to motion and imaging artifacts. Improved performance in these segments therefore represents a meaningful advance rather than a marginal numerical gain.

Accurate detection of apical and mid-wall scars is particularly relevant for risk stratification, as small or patchy scars in these regions may be associated with arrhythmogenic substrates that are not readily apparent on global assessments. The ability of CLAIM to better capture such patterns suggests potential utility in supporting more nuanced clinical interpretation of LGE findings.

Balancing Sensitivity and Over-Segmentation

In clinical practice, the consequences of segmentation errors are asymmetric. Over-segmentation of scar tissue may lead to overestimation of disease burden, potentially influencing clinical decisions such as ICD eligibility or ablation planning. Conversely, under-segmentation may obscure clinically relevant pathology, particularly in early or subtle disease.

The results demonstrate that CLAIM improves Dice score and precision without a corresponding loss in specificity, indicating a favorable balance between sensitivity and false-positive control. This balance is clinically desirable, as it reduces the likelihood of exaggerated scar burden while maintaining adequate detection of true pathological regions. The improved consistency observed across cases further supports the reliability of the proposed approach in heterogeneous clinical scenarios.

Implications for Clinical Workflow Integration

Beyond segmentation accuracy, the integration of automated tools into clinical workflows requires consistency, interpretability, and alignment with established reporting standards. By explicitly leveraging the AHA 17-segment model during both synthesis and evaluation, CLAIM aligns closely with existing cardiology workflows. This alignment facilitates intuitive interpretation of results and supports potential integration into structured reporting systems.

Moreover, the use of clinically guided synthetic data augmentation addresses a practical barrier to deployment: the limited availability of annotated LGE datasets that capture the full spectrum of scar presentations. By enhancing model robustness through anatomically

informed synthesis, CLAIM reduces sensitivity to rare or underrepresented scar patterns that may otherwise lead to unreliable predictions in routine clinical use.

Caveats and Clinical Readiness

It is important to note that, while the results demonstrate promising improvements, CLAIM is not intended to replace expert clinical judgment. Variability in image quality, acquisition protocols, and patient-specific anatomy continues to pose challenges for fully automated systems. The observed performance gains should therefore be interpreted as enabling more reliable decision support rather than definitive diagnostic outputs.

Nevertheless, the consistent improvements observed across regional analyses and out-of-domain evaluations suggest that clinically guided generative augmentation represents a viable pathway toward more robust and generalizable scar segmentation systems. In this context, CLAIM provides a meaningful step toward clinically deployable tools that respect both anatomical knowledge and real-world data constraints.

4.6 Conclusion

This chapter introduced CLAIM, a clinically guided framework for myocardial scar synthesis and segmentation that addresses fundamental limitations arising from data scarcity and pathological imbalance in LGE cardiac MRI. By jointly modeling diffusion-based image synthesis and scar segmentation, CLAIM moves beyond conventional data augmentation strategies and establishes a task-aware generative paradigm in which synthesized pathology is explicitly optimized for downstream clinical utility.

Central to the proposed framework is SMILE, a clinically informed scar mask generation module that embeds standardized anatomical knowledge through the AHA 17-segment model. By leveraging this well-established clinical representation, SMILE enables controlled generation of myocardial scar patterns that are anatomically consistent, spatially diverse, and clinically interpretable. This design allows systematic variation in scar location, extent, and morphology, overcoming the biases and limitations inherent to data-driven mask generation approaches.

In addition, CLAIM incorporates a joint training strategy in which feedback from a scar segmentation network is used to guide the diffusion model toward generating pathologies that are not only visually realistic but also informative for segmentation. This coupling aligns the objectives of synthesis and analysis, resulting in synthetic data that meaningfully enhances segmentation performance. Extensive experiments on both public and private datasets demonstrate that the proposed approach improves the performance of state-of-the-art segmentation models, including nnU-Net [27], across in-domain and out-of-domain evaluation settings.

Overall, this chapter demonstrates that embedding clinical knowledge into generative modeling, combined with task-aware joint optimization, offers a powerful and scalable solution for myocardial scar analysis in data-limited scenarios. Beyond its immediate application to LGE scar segmentation, the proposed framework provides a general blueprint for clinically guided pathology synthesis that can be extended to other cardiac pathologies, imaging modalities, and downstream tasks.

4.7 Limitations and Future Directions

While the proposed CLAIM framework demonstrates strong performance in synthesizing clinically plausible myocardial scar patterns and improving downstream segmentation robustness, several limitations remain that warrant discussion. Acknowledging these limitations not only contextualizes the current results but also highlights important directions for future research.

4.7.1 Dependence on Anatomical Registration and Atlas Accuracy

A core component of CLAIM is the use of the AHA 17-segment model to guide scar mask generation through anatomical registration. While this design enables clinically interpretable and anatomically consistent synthesis, it introduces a dependency on the accuracy of myocardial segmentation and registration. Imperfect registration, particularly in cases with abnormal ventricular geometry, severe remodeling, or poor image quality, may lead to misalignment between the subject anatomy and the AHA template. Such misalignments can propagate to the synthesized scar masks and, consequently, affect image synthesis quality.

Future work could explore more robust or adaptive registration strategies, including learning-based registration methods or uncertainty-aware atlas mapping. Alternatively, soft anatomical constraints that allow probabilistic rather than deterministic region assignment may further improve robustness in challenging cases.

4.7.2 Scar Volume Distribution Bias

Although CLAIM enables controlled variation in scar size and spatial distribution, the synthetic data generation process is still influenced by the scar volume statistics of the source dataset. As observed in out-of-domain experiments, differences in scar volume distributions between datasets can lead to suboptimal generalization when large amounts of synthetic data are used. This highlights an inherent tension between increasing data diversity and introducing new forms of dataset bias.

Addressing this limitation will require more principled control over synthetic data distributions. Potential future directions include volume-aware sampling strategies, adaptive reweighting of synthetic samples, or conditioning the diffusion model on explicit volumetric priors. Incorporating dataset-specific calibration or domain-adaptive synthesis mechanisms may further improve generalization across institutions.

4.7.3 Scope of Pathological Modeling

In its current form, CLAIM focuses primarily on myocardial infarction-related scarring as observed in LGE imaging. While this represents a clinically important pathology, myocardial fibrosis encompasses a broader spectrum of disease etiologies, including non-ischemic cardiomyopathies with diffuse or patchy enhancement patterns. Extending the proposed framework to model such pathologies would require additional clinical priors and potentially different anatomical constraints.

Future research may investigate the extension of SMILE to support pathology-specific scar morphologies or multi-class scar synthesis, enabling simultaneous modeling of ischemic and non-ischemic fibrosis. Such extensions would further enhance the clinical relevance and applicability of the framework.

4.7.4 Computational Complexity and Training Overhead

The joint optimization strategy employed by CLAIM introduces additional computational overhead compared to conventional offline data augmentation pipelines. Training diffusion-based models is inherently resource-intensive, and alternating optimization between synthesis and segmentation further increases training time. While this overhead is acceptable in research and offline training scenarios, it may limit rapid experimentation or deployment in resource-constrained environments.

Future work could explore more efficient training strategies, such as reduced diffusion step schedules, knowledge distillation from diffusion models to lighter generators, or partial joint training schemes in which segmentation feedback is applied selectively. These approaches may help balance performance gains with computational efficiency.

4.7.5 Opportunities for Multimodal and Temporal Integration

CLAIM currently operates solely on LGE imaging, focusing on anatomical scar representation. However, myocardial scarring is fundamentally linked to electrophysiological abnormalities, which are captured by complementary modalities such as ECG. Integrating multimodal information during scar synthesis or segmentation could enable more patient-specific and physiologically informed modeling.

Future directions include extending CLAIM to incorporate ECG-derived features or temporal information as additional conditioning signals for diffusion-based synthesis. Such multimodal integration may enable the generation of scar patterns that are not only anatomically plausible but also consistent with observed electrical activity, laying the groundwork for more holistic cardiac assessment.

4.7.6 Toward Domain-Generalized and Clinically Deployable Systems

Finally, while CLAIM improves robustness under domain shift, achieving true domain generalization remains an open challenge. Variability in acquisition protocols, scanner vendors, and patient populations continues to pose barriers to widespread clinical deployment. Future research may combine clinically guided synthesis with domain generalization techniques, such as style-invariant representations or test-time adaptation, to further enhance reliability across diverse clinical settings.

In summary, CLAIM represents a significant step toward clinically informed generative modeling for myocardial scar analysis. Addressing the limitations outlined above offers a clear roadmap for future research, with the potential to further improve robustness, scalability, and clinical impact.

Chapter 5

Multimodal ECG-LGE Scar Segmentation

This chapter contains material from:

- Ramzan, Farheen, et al. "Seeing Beyond the Image: ECG and Anatomical Knowledge-Guided Myocardial Scar Segmentation from Late Gadolinium-Enhanced Images." *Accepted for oral presentation at the IEEE International Symposium on Biomedical Imaging (ISBI 2026)*, arXiv preprint arXiv:2511.14702, 2026.

In this chapter, a fundamental limitation of image-only approaches to myocardial scar segmentation is addressed: the reliance on anatomical information alone to characterize a pathology that manifests through both structural and electrophysiological alterations. While LGE CMR provides high-resolution visualization of myocardial fibrosis; scar segmentation remains challenging in cases with subtle contrast differences, diffuse scarring, or ambiguous boundaries. In clinical practice, cardiologists routinely complement imaging findings with electrocardiographic information, as myocardial scarring disrupts electrical conduction and gives rise to characteristic electrophysiological signatures. Despite this, most deep learning-based scar segmentation methods operate exclusively on imaging data, overlooking complementary physiological cues that may improve robustness and patient specificity.

To bridge this gap, this chapter introduces a multimodal cardiac scar segmentation framework that integrates 12-lead ECG signals with LGE imaging to jointly model the anatomical and electrophysiological manifestations of myocardial scarring. By fusing learned representations from ECG and LGE at the feature level, the proposed approach captures cross-modal correlations that are often inaccessible to unimodal models. This multimodal formulation aims to reduce ambiguity in scar segmentation, enhance segmentation performance in challenging cases, and move toward more holistic and clinically meaningful characterization of myocardial pathology.

5.1 Introduction

Accurate segmentation of myocardial scar from LGE CMR plays a central role in the assessment of myocardial viability, disease etiology, and arrhythmic risk. The presence, extent, and spatial distribution of myocardial scar have been shown to correlate strongly with adverse clinical outcomes, including ventricular arrhythmias and sudden cardiac death. As a result, reliable quantification of scar burden is essential for guiding therapeutic decisions such as revascularization, catheter ablation, and implantable cardioverter–defibrillator placement. Despite its clinical importance, automated segmentation of myocardial scar from LGE MRI remains a challenging task due to heterogeneous enhancement patterns, low signal-to-noise ratio, partial volume effects, and substantial variability in acquisition protocols and patient anatomy [12, 105].

While deep learning-based segmentation models, including U-Net [26] and its self-configuring variant nnU-Net [27], have demonstrated strong performance across a wide range of medical imaging tasks, their effectiveness for myocardial scar segmentation is limited by the intrinsic ambiguity of LGE imaging. In many cases, scar boundaries are poorly defined, particularly in diffuse or patchy fibrosis, mid-wall scarring, or regions affected by motion artifacts. Under such conditions, image-only models must rely solely on local intensity patterns, which may be insufficient to distinguish scarred myocardium from normal tissue with confidence. These limitations motivate the incorporation of complementary information beyond anatomical imaging alone.

In clinical practice, myocardial scarring is not assessed using imaging in isolation. ECG provides a non-invasive window into the electrophysiological consequences of myocardial injury, as fibrotic tissue alters electrical conduction pathways and gives rise to characteristic ECG signatures. Established markers such as pathological Q-waves, fragmented QRS complexes, and abnormal repolarization patterns have been associated with the presence and localization of myocardial scar [144, 145]. Importantly, ECG reflects functional abnormalities that may persist even when structural changes are subtle or difficult to delineate on imaging, suggesting that electrophysiological signals can provide complementary cues for scar characterization.

Despite the routine clinical integration of ECG and imaging information, most automated scar segmentation approaches rely exclusively on CMR data. Existing multimodal frameworks in cardiac imaging primarily focus on combining multiple MRI sequences, such as cine, T1/T2 mapping, and LGE, to improve anatomical segmentation [13, 29]. These approaches do not directly leverage electrophysiological information and are therefore limited in their ability to capture the functional impact of myocardial scarring. Joint learning from ECG and MRI remains underexplored, in part due to the heterogeneity of these modalities, differences in dimensionality, and challenges associated with temporal alignment.

Temporal misalignment represents a particularly important barrier to ECG–MRI integration in real-world clinical settings. In referral-based healthcare systems such as the National Health Service (NHS) in the United Kingdom, ECGs are often acquired days, weeks, or even months prior to MRI examinations. Even when ECG and MRI are acquired in close temporal proximity, simultaneous recordings within the MRI scanner are confounded by magnetohydrodynamic (MHD) effects, which distort ECG waveforms and complicate interpretation [146]. As a result, naïve fusion strategies that assume temporal synchrony between modalities may introduce noise or misleading correlations, ultimately degrading model performance.

To address these challenges, this work proposes a multimodal cardiac scar segmentation framework that explicitly integrates electrophysiological and anatomical information while accounting for temporal discrepancies between ECG and LGE modalities. The proposed approach incorporates two key design elements. First, a Temporal-Aware Feature Fusion (TAFF) mechanism is introduced that adaptively modulates the contribution of ECG-derived and MRI-derived features based on the acquisition time difference between the two modalities. By embedding temporal awareness into the fusion process, TAFF enables the model to down-weight potentially stale or less informative ECG signals while still exploiting complementary physiological cues when appropriate. Second, the standardized 17-segment AHA model [126] as an anatomical prior is incorporated to provide spatial context and guide alignment between ECG-derived representations and left ventricular myocardial regions.

Together, these components enable a physiologically grounded approach to myocardial scar segmentation that more closely reflects clinical reasoning. Rather than treating ECG and MRI as independent data streams, the proposed framework learns shared representations that capture cross-modal correlations between electrical conduction abnormalities and structural scar patterns.

The resulting model demonstrates substantial improvements over state-of-the-art image-only baselines, increasing the average Dice score for scar segmentation while achieving high precision and sensitivity. These results suggest that integrating electrophysiological and anatomical knowledge allows the model to effectively “see beyond the image,” reducing ambiguity and improving robustness in challenging segmentation scenarios.

The main contributions of this chapter are summarized as follows:

- A novel Temporal-Aware Feature Fusion (TAFF) framework is proposed that explicitly accounts for acquisition time differences between ECG and MRI, enabling adaptive and context-aware multimodal integration for myocardial scar segmentation.
- The 17-segment AHA model is incorporated as an anatomical prior to guide multimodal feature alignment and enhance spatial interpretability of ECG-informed scar predictions.
- To the best of our knowledge, this work represents one of the first systematic efforts to leverage ECG signals for guiding MRI-based myocardial scar segmentation, bridging electrophysiological and structural domains to enable more physiologically consistent cardiac image analysis.

5.2 Background

5.2.1 Myocardial Scar as a Structural and Electrophysiological Phenomenon

Myocardial scarring represents the end stage of a variety of pathological processes, including ischemic injury, inflammation, and cardiomyopathic remodeling. Structurally, scar tissue is characterized by replacement fibrosis, altered extracellular matrix composition, and loss of viable myocytes. Functionally, these structural changes disrupt normal electrical conduction pathways, leading to conduction slowing, block, and electrical heterogeneity. As a result, myocardial scar is intrinsically a multimodal phenomenon, manifesting simultaneously in anatomical imaging and electrophysiological measurements [147].

LGE CMR imaging has become the reference standard for non-invasive visualization of myocardial scar. Gadolinium-based contrast agents accumulate preferentially in fibrotic tissue due to altered wash-in and wash-out kinetics, resulting in hyperenhanced signal intensity relative to healthy myocardium. Numerous studies have demonstrated strong associations between LGE-derived scar burden and adverse outcomes, including ventricular arrhythmias, heart failure progression, and mortality [18, 123, 148]. Beyond global scar burden, the spatial distribution and transmural extent of scarring provide clinically meaningful information regarding disease etiology and prognosis.

However, myocardial scarring also exerts profound effects on cardiac electrophysiology. Fibrotic tissue alters action potential propagation, creating regions of slow conduction and unidirectional block that can serve as substrates for re-entrant arrhythmias [47, 48]. These electrophysiological alterations are reflected in surface electrocardiograms, which capture the integrated electrical activity of the heart. Importantly, ECG abnormalities may persist even when scar is subtle or difficult to delineate on imaging, highlighting the complementary nature of structural and functional assessments.

5.2.2 Electrocardiographic Manifestations of Myocardial Scar

The relationship between myocardial scar and ECG features has been extensively studied in clinical electrophysiology. Classical markers such as pathological Q-waves have long been associated with transmural myocardial infarction, reflecting electrically silent scar regions [149, 150]. More recent studies have identified fragmented QRS complexes as indicators of heterogeneous conduction through fibrotic myocardium, correlating with both scar presence and arrhythmic risk [144, 151–153]. Repolarization abnormalities, including ST-segment changes and T-wave inversions, have also been linked to underlying myocardial injury [154–158].

Beyond qualitative interpretation, quantitative ECG analyses have been shown to correlate with scar location and extent. Studies combining ECG mapping with imaging or invasive electrophysiological studies have demonstrated that specific lead patterns can suggest regional scar involvement, particularly in ischemic cardiomyopathy [144, 159, 160]. These findings underpin the routine clinical practice of interpreting ECGs alongside imaging data to inform diagnosis and management.

Despite this rich clinical literature, most automated image analysis pipelines do not exploit ECG information. This disconnect highlights a gap between how cardiologists reason about myocardial scar and how machine learning models are typically designed.

5.2.3 Automated Myocardial Scar Segmentation from LGE MRI

Deep learning has substantially advanced automated segmentation in medical imaging, with CNNs such as U-Net and its variants becoming the dominant paradigm [26, 27, 127–130]. In cardiac imaging, these models have been successfully applied to segment ventricular structures and myocardium across multiple modalities [28, 131]. For myocardial scar segmentation in LGE MRI, several supervised approaches have reported promising results [12, 105].

However, scar segmentation remains particularly challenging due to several factors. Scar regions often occupy a small fraction of the myocardium, resulting in severe class imbalance. Boundaries between scar and healthy tissue are frequently ill-defined, especially in cases of

diffuse fibrosis or low contrast-to-noise ratio. Furthermore, inter-observer variability in scar annotation is substantial, reflecting the inherent ambiguity of the task [19]. As a result, image-only models are prone to failure in clinically challenging cases, even when trained on carefully curated datasets.

These limitations motivate the incorporation of additional sources of information that can help disambiguate scar regions beyond local image intensity patterns.

5.2.4 Multimodal Learning in Cardiac Imaging

Multimodal learning has attracted increasing interest in medical image analysis as a means of integrating complementary sources of information. In cardiac imaging, most multimodal approaches focus on combining multiple MRI sequences, such as cine MRI, T1/T2 mapping, and LGE, to improve anatomical segmentation or tissue characterization [13, 161, 162]. While these methods demonstrate improved performance over single-sequence models, they remain confined to imaging modalities and do not incorporate physiological signals.

Efforts to integrate ECG with imaging have largely focused on correlation studies or post hoc analysis rather than joint learning. For example, ECG features have been correlated with LGE-derived scar burden or arrhythmic risk, but these analyses typically operate at the patient or regional level rather than enabling pixel-wise image segmentation [144, 152]. To the best of our knowledge, no prior work has attempted end-to-end learning from ECG and MRI for spatially resolved scar mapping.

Joint ECG–MRI learning presents several unique challenges. ECG signals are one-dimensional time series, while MRI data are high-dimensional spatial volumes. Aligning representations across these modalities requires careful architectural design. Moreover, ECG and MRI are rarely acquired simultaneously in clinical practice, introducing temporal misalignment that complicates naïve fusion strategies.

5.2.5 Temporal Misalignment and Clinical Data Heterogeneity

Temporal heterogeneity is a defining characteristic of real-world clinical data. In referral-based healthcare systems such as NHS [163, 164], ECGs are often obtained during outpatient visits, emergency presentations, or routine monitoring, whereas MRI examinations may occur weeks or months later. Disease progression, treatment effects, and physiological variability can all influence the relationship between electrical and structural manifestations of myocardial scar over time.

Even in research settings, simultaneous ECG–MRI acquisition is problematic. ECGs recorded inside MRI scanners are affected by MHD effects, which distort waveform morphology and complicate feature extraction [146, 165–169]. As a result, assuming temporal synchrony between ECG and MRI can introduce spurious correlations and degrade model performance.

These considerations underscore the need for multimodal fusion strategies that explicitly account for temporal uncertainty rather than treating all modalities as equally informative at all times.

5.2.6 Motivation for This Work

The limitations of existing approaches highlight a critical gap in current cardiac image analysis pipelines: the absence of physiologically informed, temporally aware multimodal learning frameworks. While ECG and LGE MRI provide complementary views of myocardial scar, effectively integrating these modalities requires models that can adaptively weigh their contributions based on reliability, relevance, and temporal proximity.

Motivated by these observations, this work introduces a multimodal scar segmentation framework that integrates ECG-derived electrophysiological information with LGE-derived anatomical features while explicitly modeling acquisition time differences. By incorporating a TAFF mechanism and grounding feature alignment in the standardized AHA 17-segment model [126], the proposed approach seeks to bridge the gap between clinical reasoning and automated image analysis. This design enables physiologically consistent scar segmentation that reflects both structural damage and its functional consequences, laying the foundation for more robust and clinically meaningful cardiac assessment.

5.3 Methods

This section presents the proposed multimodal learning framework for myocardial scar segmentation from LGE CMR, augmented with electrophysiological information derived from ECG signals. An overview of the architecture is illustrated in Fig. 5.1a, with the proposed Temporal-Aware Feature Fusion (TAFF) mechanism detailed in Fig. 5.1b. The design of the framework is motivated by the need to integrate heterogeneous modalities with distinct spatial, temporal, and physiological characteristics, while maintaining anatomical interpretability and robustness under real-world clinical acquisition constraints.

At a high level, the framework consists of three principal components: (i) an LGE-based segmentation branch augmented with anatomical priors derived from the AHA 17-segment model, (ii) an ECG autoencoder that learns compact electrophysiological representations from 12-lead ECG signals, and (iii) a temporal-aware fusion mechanism that adaptively integrates LGE and ECG features while accounting for differences in acquisition time between modalities.

5.3.1 Design Rationale and Architectural Considerations

The design of the proposed multimodal myocardial scar segmentation framework is guided by the clinical reality that myocardial scarring manifests through both anatomical disruption and electrophysiological dysfunction. Rather than treating multimodal integration as a generic feature fusion problem, the architecture is deliberately structured to reflect how cardiologists reason about myocardial pathology, while accounting for the practical constraints of real-world clinical data acquisition. This section outlines the rationale behind the key architectural and methodological choices underlying the proposed framework.

Why Integrate ECG with LGE CMR?

LGE CMR provides high-resolution anatomical characterization of myocardial scar, yet its interpretation is often ambiguous in cases of diffuse fibrosis, low contrast-to-noise ratio, or

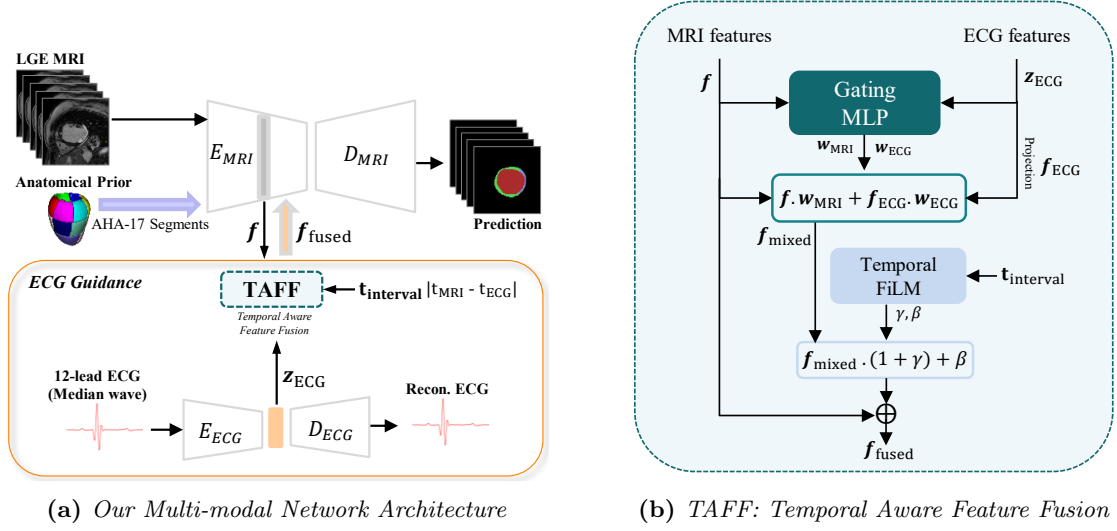


Figure 5.1: Overview of the proposed (a) multi-modal network combining LGE-MRI with AHA-based spatial priors and ECG features via separate encoders, and (b) the Temporal-Aware Feature Fusion (TAFF) module, which adaptively integrates ECG and MRI representations modulated by the ECG–MRI acquisition interval. Figure source: [9], reproduced with permission of the rights holder, © 2026 IEEE.

partial volume effects. In contrast, ECG captures the functional consequences of myocardial injury, reflecting altered electrical conduction pathways caused by fibrotic tissue. These electrophysiological changes may persist even when structural abnormalities are subtle or spatially heterogeneous.

In clinical practice, ECG and imaging findings are interpreted jointly rather than independently. Motivated by this paradigm, the proposed framework integrates ECG-derived features with LGE-based representations to provide complementary information that can disambiguate scar boundaries and improve robustness. Importantly, the goal is not to replace imaging-based segmentation with ECG-derived localization, but to use ECG as a contextual signal that informs and refines image-based predictions.

Separation of Modality-Specific Encoders

A key architectural decision is the use of separate encoders for ECG and LGE CMR rather than a shared or early-fusion representation. ECG and MRI differ fundamentally in their dimensionality, statistical structure, and noise characteristics. ECG signals are one-dimensional time series governed by electrophysiological dynamics, whereas LGE CMR data are high-dimensional spatial volumes encoding tissue contrast.

Early fusion of raw modalities risks entangling modality-specific noise and obscuring meaningful representations. By employing dedicated encoders, the framework allows each modality to be processed in a manner suited to its intrinsic properties. The ECG autoencoder [170] learns compact electrophysiological embeddings that preserve signal morphology, while the MRI encoder [27] focuses on spatial and textural patterns relevant to scar segmentation. This separation provides a stable foundation for subsequent multimodal interaction.

Mid-Level Feature Fusion versus Late or Bottleneck Fusion

The point at which multimodal fusion is performed has a significant impact on model performance and interpretability. Late fusion strategies, which combine modality-specific predictions, limit cross-modal interaction and fail to influence spatial feature learning. Conversely, fusion at the bottleneck level often sacrifices spatial resolution, which is critical for accurate scar segmentation.

In this work, fusion is performed at an intermediate level of the MRI encoder, where features retain sufficient spatial detail while encoding semantically meaningful information. Empirical evaluation confirmed that mid-level fusion offers a favorable balance between anatomical fidelity and representational richness. This design choice allows ECG-derived information to influence spatial feature refinement without overwhelming fine-grained image cues.

Adaptive Gating for Robust Multimodal Integration

Another central design consideration is the use of adaptive gating to regulate the contribution of each modality. In clinical datasets, the informativeness of ECG signals varies substantially across patients and acquisition conditions. A fixed fusion strategy risks over-reliance on noisy or uninformative ECG features, potentially degrading segmentation performance.

The gating mechanism within TAFF allows the network to dynamically adjust the relative weighting of LGE CMR and ECG features on a per-sample basis. This adaptive behavior enables the model to emphasize ECG information when it provides meaningful complementary cues and to default to LGE-driven segmentation when ECG signals are unreliable. Such flexibility is essential for robust multimodal learning in heterogeneous clinical populations.

Temporal Awareness as a Design Constraint

Temporal misalignment between ECG and MRI acquisitions represents a defining challenge in real-world multimodal cardiac analysis. Unlike controlled research settings, clinical workflows rarely guarantee simultaneous acquisition, and disease progression or intervention effects may alter the relationship between electrophysiological and structural findings over time.

Rather than treating temporal discrepancy as noise, the proposed framework explicitly models acquisition time difference through the TAFF mechanism. By conditioning feature modulation on the ECG–LGE time interval, the model learns to attenuate stale or potentially misleading information while preserving useful physiological context when temporal proximity is high. This design reflects a clinically grounded assumption: the relevance of ECG information diminishes as temporal separation increases, but does not vanish entirely.

Role of the AHA-17 Model in Multimodal Alignment

Incorporating the AHA 17-segment model as an anatomical prior serves both interpretability and alignment purposes. ECG signals reflect global and regional electrical activity but lack explicit spatial resolution. By embedding standardized myocardial segmentation into the LGE CMR branch, the framework provides a spatial scaffold onto which ECG-derived information can be meaningfully projected.

This anatomical grounding constrains multimodal interaction to physiologically plausible regions and aligns learned representations with clinical reporting standards. Unlike purely

data-driven spatial embeddings, the AHA-based prior ensures that model outputs remain interpretable and consistent with established cardiology practice.

Balancing Model Complexity and Clinical Applicability

Finally, the architecture reflects a deliberate balance between representational power and practical feasibility. While multimodal fusion and temporal modeling introduce additional complexity, the framework avoids unnecessary architectural depth or overly specialized components. The use of established backbones [27, 170], lightweight gating networks, and explicit priors facilitates reproducibility and reduces barriers to clinical translation.

Overall, the design of the proposed multimodal framework prioritizes physiological relevance, robustness under clinical variability, and interpretability. These considerations position the approach as a principled step toward patient-specific and clinically meaningful myocardial scar mapping.

5.3.2 Multi-modal Network Architecture

The proposed multimodal architecture is designed to jointly model anatomical and electrophysiological manifestations of myocardial scar. The MRI branch follows an encoder–decoder [27] paradigm, where an image encoder E_{MRI} processes 2D LGE MRI slices concatenated with anatomical prior maps derived from the AHA 17-segment model [126] (see Sec. 5.3.4). The encoded feature representations are subsequently decoded by an image decoder D_{MRI} to produce pixel-wise scar segmentation maps.

In parallel, an ECG autoencoder [170] is employed to extract compact electrophysiological representations from 12-lead ECG signals. The ECG encoder E_{ECG} maps median ECG waveforms into a latent feature space $\mathbf{z}_{ECG} \in \mathbb{R}^{B \times 64}$, where B denotes the batch size. This latent representation is learned in an unsupervised manner via signal reconstruction using an ECG decoder D_{ECG} . The use of an autoencoding objective encourages the ECG encoder to retain physiologically meaningful information while reducing sensitivity to noise and inter-patient variability.

To enable interaction between modalities, multimodal fusion is performed at the mid-level of the MRI encoder. Specifically, an intermediate feature map $\mathbf{f} \in \mathbb{R}^{B \times C \times H_{mid} \times W_{mid}}$ is extracted, which balances spatial resolution with semantic abstraction. Empirically, it was observed that fusing at this level yields superior performance compared to fusion at the bottleneck, as mid-level features preserve spatial detail critical for accurate scar segmentation while remaining expressive enough to incorporate high-level electrophysiological cues. The ECG latent representation $\mathbf{z}_{ECG}$ is projected into the same spatial feature space and fused with $\mathbf{f}$ using the proposed TAFF module before being passed to the remaining encoder and decoder stages.

5.3.3 TAFF: Temporal-Aware Feature Fusion

A central challenge in ECG–LGE integration lies in the temporal misalignment between modalities. In clinical workflows, ECGs are often acquired days, weeks, or months before MRI examinations, and the reliability of electrophysiological information may degrade as the temporal gap increases. To explicitly account for this, the TAFF module is proposed,

which adaptively integrates LGE and ECG features based on both modality relevance and acquisition time difference.

As illustrated in Fig. 5.1b, TAFF consists of two key stages: modality gating and temporal modulation. First, a gating multi-layer perceptron (MLP) learns modality-specific weights $w = [w_{MRI}, w_{ECG}] \in \mathbb{R}^{B \times 2}$, constrained such that $\sum_i w_i = \mathbf{1}$. These weights determine the relative contribution of MRI features $\mathbf{f}$ and projected ECG features $\mathbf{f}_{ECG} \in \mathbb{R}^{B \times C \times H_{mid} \times W_{mid}}$, yielding a mixed representation:

$$\mathbf{f}_{mixed} = \mathbf{f} \times w_{MRI} + \mathbf{f}_{ECG} \times w_{ECG}.$$

The ECG branch first produces a global embedding $\mathbf{f}_{ECG}^{1D} \in \mathbb{R}^{B \times D_{ECG}}$ that summarises the whole-heart electrophysiological signal. Since this embedding carries no spatial information of its own, it needs to be expanded before it can be mixed with the MRI feature map. This is done in two steps. A learned linear layer with a SiLU activation first maps $\mathbf{f}_{ECG}^{1D}$ from D_{ECG} to C channels, so that it lines up with the channel dimension of the MRI features at the fusion stage. The resulting vector in $\mathbb{R}^{B \times C}$ is then reshaped to $\mathbb{R}^{B \times C \times 1 \times 1}$ and broadcast across the spatial dimensions, giving $\mathbf{f}_{ECG} \in \mathbb{R}^{B \times C \times H_{mid} \times W_{mid}}$. In effect, every spatial location receives the same projected ECG embedding, which is a deliberate choice. A surface ECG is a global measurement of cardiac electrical activity and does not carry spatially localised information that could be assigned to specific voxels. Before the two streams are mixed, $\mathbf{f}_{ECG}$ is passed through a group normalisation so that its scale is comparable to the instance-normalised MRI features.

To incorporate temporal information, the mixed features are further modulated using a Temporal Feature-wise Linear Modulation (FiLM) layer inspired by [171]. The absolute time interval between ECG and MRI acquisitions, $t_{interval} = |t_{MRI} - t_{ECG}|$, is provided as input to a dedicated MLP that predicts affine scale γ and shift β parameters. These parameters condition the fused features on temporal proximity, enabling the network to attenuate or amplify ECG influence depending on its expected relevance.

Residual connections are employed to stabilize training and allow the network to default to MRI-only representations when ECG information is unreliable. The final fused feature representation is given by:

$$\mathbf{f}_{fused} = \mathbf{f} + \mathbf{f}_{mixed} \times (1 + \gamma) + \beta.$$

This formulation ensures smooth integration of multimodal information while preserving robustness under varying temporal conditions.

5.3.4 AHA-17 Segment Model as Anatomical Prior

To provide anatomically informed spatial guidance, the standardized 17-segment AHA model [126] is incorporated as an explicit anatomical prior. The AHA model divides the left ventricular myocardium into basal, mid-ventricular, and apical regions, offering a clinically interpretable framework for localizing myocardial pathology.

Following established geometric implementations [172], slice-wise and angular partitioning are applied to the 3D myocardium segmentation to generate 17-channel binary masks, each corresponding to a specific AHA segment. These masks are concatenated with the corresponding MRI slices and provided as input to the MRI encoder. By embedding anatomical

context directly into the learning process, the network is encouraged to learn region-aware representations and produce anatomically consistent scar segmentations.

5.3.5 Loss Functions and Training Strategy

The entire network is optimized using a composite loss function:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \lambda_{ECG}\mathcal{L}_{ecg} + \lambda_{ent}\mathcal{L}_{ent},$$

where $\mathcal{L}_{seg}$ denotes the segmentation loss, $\mathcal{L}_{ecg}$ the ECG reconstruction loss, and $\mathcal{L}_{ent}$ an entropy-based regularization term.

The segmentation loss $\mathcal{L}_{seg}$ consists of a combination of Dice loss and cross-entropy loss, balancing overlap accuracy with pixel-wise classification. The ECG reconstruction loss $\mathcal{L}_{ecg}$ is defined as the mean squared error between reconstructed and ground-truth ECG signals, encouraging faithful preservation of electrophysiological information.

To ensure stable multimodal learning, the coefficient λ_{ECG} is gradually increased during the initial training phase following a quadratic warm-up schedule, preventing premature dominance of ECG features. Additionally, an entropy-based regularization term is introduced on the gating weights:

$$\mathcal{L}_{ent} = - \sum_i w_i \log w_i,$$

which discourages early gating collapse and promotes balanced modality exploration during training [173]. A small weighting coefficient $\lambda_{ent} = 3 \times 10^{-3}$ ensures that this regularization stabilizes training without overpowering the primary objectives.

Together, these design choices enable robust, temporally aware multimodal learning that integrates electrophysiological and anatomical information for physiologically consistent myocardial scar segmentation.

5.4 Experiments and Results

This section describes the datasets, preprocessing steps, experimental protocol, and evaluation strategy used to assess the effectiveness of the proposed multimodal cardiac scar mapping framework. Particular emphasis is placed on clinical realism, robustness under temporal misalignment, and systematic evaluation of the proposed TAFF mechanism and anatomical priors.

5.4.1 Datasets and Data Preparation

Clinical Dataset

Experiments were conducted using a private dataset provided by Nottingham University Hospitals [117, 118], consisting of paired LGE CMR scans and corresponding 12-lead ECG recordings. This dataset reflects a real-world clinical setting, where data are collected as part of routine patient care rather than under tightly controlled research protocols.

To ensure physiologically meaningful multimodal learning, cases in which ECG and MRI acquisitions occurred within a 90-day interval were retained. This threshold balances data availability with clinical relevance, as larger temporal gaps may reflect disease progression or

treatment effects that weaken cross-modal correspondence. After applying this criterion, a total of 103 high-quality LGE–ECG pairs were retained for analysis.

ECG Digitization and Quality Control

ECG recordings were originally available in paper-based format. These were digitized using the open-source toolkit proposed by Fortune et al. [174]. Manual bounding boxes were applied to isolate individual leads, followed by automated waveform extraction. Given the sensitivity of multimodal learning to ECG signal quality, all digitized ECGs underwent visual inspection, and cases with incomplete leads, excessive noise, or failed digitization were excluded. This conservative quality control process ensured reliable electrophysiological representations at the cost of reduced dataset size, reflecting the practical trade-offs encountered in clinical research.

LGE Annotation and Preprocessing

LGE CMR scans were annotated by trained cardiologists, with labels provided for left ventricular blood pool, healthy myocardium, and myocardial scar. An additional quality assurance step was performed to ensure annotation consistency, particularly in regions with ambiguous enhancement. LGE preprocessing followed the standardized nnU-Net pipeline [27], including resampling to a common resolution, intensity normalization, and spatial cropping based on myocardial bounding boxes. This ensured fair comparison with image-only baselines and minimized confounding effects due to preprocessing choices.

ECG Representation

For ECG preprocessing, median waveforms were extracted from digitized signals using the approach proposed by Hong et al. [175]. This representation reduces beat-to-beat variability while preserving characteristic waveform morphology associated with myocardial pathology. The resulting standardized ECG inputs were used for both ECG autoencoder training and multimodal fusion.

The dataset was split at the patient level into training, validation, and test sets using a 7:1:2 ratio to prevent data leakage and ensure unbiased evaluation.

5.4.2 Implementation Details

The ECG branch of the network was implemented using the dual-attention ECG architecture proposed by Chen et al. [170], selected for its ability to capture both temporal and inter-lead dependencies. To improve generalization and stabilize training, the ECG encoder was first pretrained on the large-scale PTB-XL dataset [176], which contains diverse ECG morphologies across a wide range of cardiac conditions. Pretraining was performed using the AdamW optimizer with cosine learning rate decay and weight decay of 5×10^{-5} , followed by fine-tuning on the digitized clinical ECGs.

The ECG–MRI acquisition interval $t_{interval} = |t_{MRI} - t_{ECG}|$, measured in days, was normalized to the range $[0, 1]$ and provided as input to the TAFF module. To prevent premature collapse of multimodal fusion during early training, separate optimizers were employed for the MRI branch, ECG branch, and fusion module. The MRI segmentation backbone was

trained using SGD with an initial learning rate of 10^{-2} , polynomial decay, momentum of 0.99, and weight decay of 3×10^{-5} . Fusion gate parameters were optimized using AdamW with weight decay 10^{-4} and a synchronized learning rate schedule to promote stable cross-modal adaptation.

5.4.3 Evaluation Metrics and Experimental Protocol

Segmentation performance was evaluated using Dice similarity coefficient, precision, sensitivity, and absolute scar volume difference. Dice score measures spatial overlap, while precision and sensitivity provide complementary perspectives on false positives and false negatives. Scar volume difference was included to assess clinical relevance, as over- or under-estimation of scar burden can directly influence treatment decisions.

All reported improvements over the baseline were assessed for statistical significance using paired t -tests on test-set cases, with a significance threshold of $p < 0.001$.

5.4.4 Quantitative Results and Ablation Studies

Table 5.1 summarizes the quantitative segmentation results comparing the unimodal image-only baseline with the proposed multimodal framework and several ablated variants. The image-only baseline, trained using LGE MRI alone, achieved a Dice score of 0.6149, reflecting the difficulty of scar segmentation under challenging contrast conditions.

In contrast, the proposed multimodal approach integrating ECG and MRI with anatomical priors achieved a substantial performance gain, increasing the Dice score to 0.8463 while simultaneously improving precision (0.9115) and sensitivity (0.9043). The marked reduction in scar volume difference further indicates improved quantitative reliability, highlighting the clinical relevance of the proposed fusion strategy.

The ablation study provides insight into the contribution of individual components. Removing temporal conditioning from the fusion mechanism resulted in a significant performance drop, particularly when anatomical priors were not used, underscoring the importance of modeling acquisition time differences. Incorporating the AHA-17 anatomical prior consistently improved performance across both unimodal and multimodal settings, demonstrating its value as a spatial constraint. The best performance was achieved only when ECG fusion, anatomical priors, and temporal conditioning were jointly enabled, confirming the complementary nature of these design elements.

5.4.5 Qualitative Analysis and Model Interpretability

Figure 5.2 presents qualitative comparisons between the unimodal baseline and the proposed multimodal approach. The image-only model frequently exhibits fragmented or incomplete scar predictions, particularly in regions with subtle enhancement. In contrast, the multimodal model produces more spatially coherent scar segmentations that better align with expert annotations.

To support interpretability, scar volume distributions across AHA-17 segments are visualized, demonstrating improved regional consistency. Additionally, cross-lead and temporal attention maps from the ECG encoder reveal physiologically meaningful patterns, with increased attention on leads associated with the inferred scar location. These visualizations

Table 5.1: Scar segmentation results on the test set: image-only baseline vs. proposed multimodal pipeline with module ablations. Best scores in bold. (AP = Anatomical Prior). All improvements over the baseline are statistically significant (paired t -test, $p < 0.001$). Table source: [9], reproduced with permission of the rights holder, © 2026 IEEE.

Model	Input Modality	w/ AP	Dice	Pre.	Sen	Vol. Diff.
Baseline (unimodal)	LGE	x	0.6149	0.7315	0.7072	4.436
Ours (multimodal)	LGE+ECG	✓	0.8463	0.9115	0.9043	1.971
Ablation study	LGE+ECG (w/o time conditioning)	x	0.5916	0.7351	0.6398	6.256
	LGE+ECG (w/o time conditioning)	✓	0.8234	0.8234	0.8900	2.039
	LGE	✓	0.8204	0.9090	0.8408	2.196
	LGE+ECG (w/ time conditioning)	✓	0.8463	0.9115	0.9043	1.971

provide evidence that the model is not merely leveraging ECG as a generic regularizer, but is actively learning clinically relevant electrophysiological–anatomical associations.

Overall, the experimental results demonstrate that temporally aware multimodal fusion of ECG and MRI, guided by anatomical priors, yields significant improvements in myocardial scar segmentation accuracy, robustness, and interpretability compared to state-of-the-art image-only approaches.

5.5 Discussion and Analysis

This section provides a detailed analysis of the experimental results, with particular emphasis on understanding the contributions of multimodal integration, anatomical priors, and temporal conditioning to myocardial scar segmentation performance. All models were implemented using the same nnU-Net backbone [27] to ensure a fair comparison and isolate the impact of multimodal design choices.

5.5.1 Multimodal Integration Substantially Improves Scar Segmentation Performance

As summarized in Table 5.1, the proposed multimodal framework significantly outperforms the image-only baseline across all evaluation metrics. When trained using only LGE MRI, the baseline model achieves a Dice score of 0.6149, reflecting the intrinsic difficulty of myocardial scar segmentation under conditions of variable contrast, motion artifacts, and ambiguous scar boundaries. These challenges are well documented in the literature and are particularly pronounced for small or diffuse scar regions.

By incorporating ECG-derived electrophysiological features, anatomical priors, and temporal-aware fusion, the proposed multimodal approach achieves a Dice score of 0.8463, corresponding to an improvement of over 37% relative to the baseline. This gain is accompanied by substantial increases in both precision and sensitivity, indicating that the model not only captures a larger fraction of true scar tissue but also reduces false-positive predictions. The

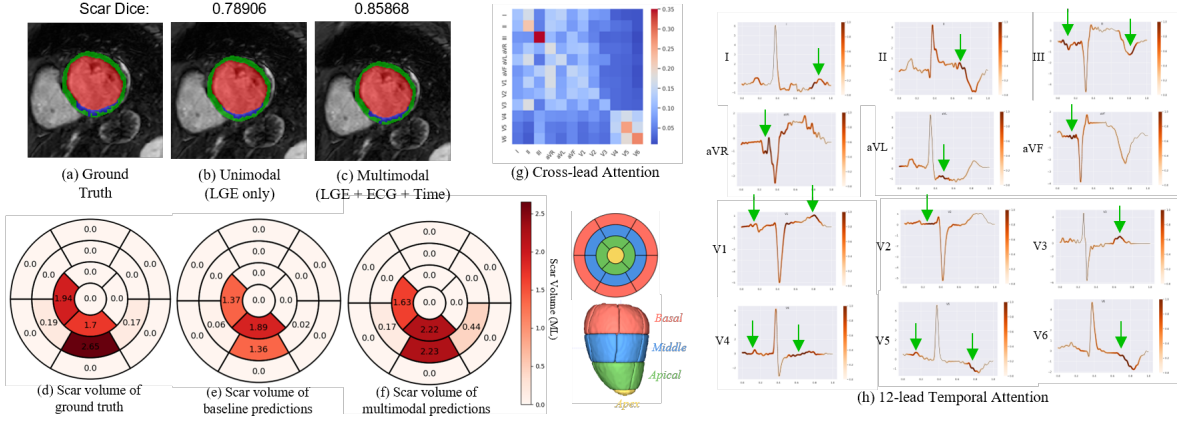


Figure 5.2: *Result Visualization: ground truth (a) and segmentation predictions of left ventricular blood pool, healthy myocardium, and scar overlaid on a 2D LGE slice using unimodal (image-only) (b) and the proposed multimodal approach (c). Scar volume distribution are also plotted on the AHA-17 segments accordingly (d-f); and cross-lead (g) and temporal lead-wise (h) attention maps for the corresponding ECG input using the ECG attention network to support model explainability. Figure source: [9], adapted with permission of the rights holder, © 2026 IEEE.*

improved scar volume agreement further underscores the clinical relevance of the proposed method, as accurate estimation of scar burden is critical for downstream clinical decision-making.

These results demonstrate that physiological and anatomical information provides complementary cues that are not readily available from imaging alone. ECG features offer global and regional indicators of myocardial injury, while anatomical priors constrain predictions to physiologically plausible regions. Together, they enable more accurate localization and segmentation of scar boundaries, as qualitatively illustrated in Fig. 5.2.

5.5.2 Challenges of Naïve ECG-MRI Fusion

The ablation results highlight the inherent difficulty of directly fusing ECG and MRI features without appropriate structural and temporal guidance. When both the AHA-17 anatomical prior and temporal conditioning are removed, the multimodal model achieves a Dice score of only 0.5916, which is lower than the image-only baseline. This counterintuitive degradation underscores that multimodal learning is not inherently beneficial and can be detrimental if modality interactions are not carefully designed.

ECG signals encode global electrical activity rather than explicit spatial information, and without anatomical grounding, the model struggles to associate electrophysiological patterns with specific myocardial regions. In this setting, ECG features may introduce noise or spurious correlations that interfere with image-based feature learning. This observation reinforces the need for principled fusion strategies that respect the fundamental differences between modalities, rather than treating multimodal integration as a simple concatenation problem.

5.5.3 Impact of Anatomical Priors on Physiological Alignment

Introducing the AHA-17 anatomical prior leads to a marked improvement in performance, increasing the Dice score from 0.5916 to 0.8234. This gain can be attributed to two complementary effects. First, the anatomical prior acts as a strong spatial regularizer, constraining scar predictions to the myocardial wall and reducing anatomically implausible false positives. This is particularly important in LGE MRI, where enhancement artifacts or blood pool signal can confound image-only models.

Second, the AHA-17 segmentation provides a clinically meaningful spatial framework that facilitates alignment between ECG-derived electrophysiological information and myocardial regions. ECG abnormalities are often interpreted in terms of regional myocardial involvement, and embedding this regional structure into the learning process allows the model to establish more meaningful correspondences between electrical signatures and anatomical locations. The resulting improvement in sensitivity, from 0.8408 to 0.8900, indicates enhanced detection of true scar regions, particularly in anatomically challenging segments [177].

5.5.4 Role of Temporal Conditioning in Multimodal Fusion

While anatomical priors provide spatial alignment, temporal conditioning addresses a distinct and equally important challenge: variability in acquisition timing between ECG and MRI. Incorporating acquisition time difference into the TAFF module further improves the Dice score from 0.8234 to 0.8463, demonstrating that temporal awareness enhances multimodal integration.

This improvement suggests that the relevance of ECG-derived features depends on their temporal proximity to the MRI scan. When ECGs are acquired closer in time, electrophysiological patterns are more likely to reflect the structural state of the myocardium. Conversely, ECGs obtained further from the MRI acquisition may be less informative due to disease progression or therapeutic interventions. Temporal conditioning enables the model to adaptively modulate the influence of ECG features, down-weighting stale or less reliable signals while preserving useful physiological context.

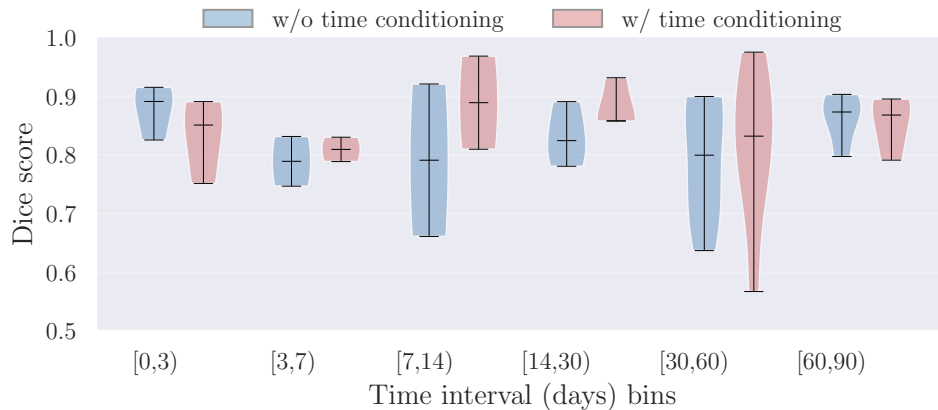


Figure 5.3: Distribution of Dice scores across time-interval bins, comparing models with and without time conditioning. Figure source: [9], reproduced with permission of the rights holder, © 2026 IEEE.

Figure 5.3 further illustrates this effect by showing the distribution of Dice scores across time-interval bins. Models equipped with temporal conditioning exhibit more stable performance across varying acquisition intervals, whereas models without time awareness show increased variability and performance degradation for larger time gaps. This highlights the importance of explicitly modeling temporal uncertainty in real-world multimodal clinical data.

5.5.5 Qualitative Analysis and Model Interpretability

Figure 5.2 provides qualitative insight into how multimodal integration influences segmentation outcomes. The image-only baseline captures the general pattern of scar distribution but consistently underestimates scar extent, particularly in mid-ventricular and apical regions. In contrast, the proposed multimodal model produces predictions that more closely match both the spatial distribution and magnitude of the ground truth scar burden across the 17 AHA segments.

The region-wise scar volume analysis reveals that the multimodal model achieves more uniform performance across myocardial segments, reducing regional bias that is commonly observed in image-only approaches. This improvement is particularly relevant for clinical applications, as underestimation of scar in specific regions may lead to inaccurate risk stratification or suboptimal treatment planning.

To enhance interpretability, attention maps from the ECG encoder are visualized alongside segmentation results. Temporal lead-wise attention maps indicate increased focus on clinically relevant ECG components, such as Q waves, QRS fragmentation, and abnormal ST-T segments, which are known markers of myocardial scar [178, 179]. Cross-lead attention patterns further reveal physiologically meaningful relationships, with heightened importance assigned to leads II, aVR, V2, and V4, as well as correlated activation among V5, V6, and lead II. These observations provide qualitative evidence that the model learns clinically interpretable electrophysiological features rather than treating ECG input as an opaque auxiliary signal.

Overall, the discussion highlights that performance gains arise not from multimodality alone, but from the careful integration of physiological signals, anatomical structure, and temporal context. The proposed framework demonstrates that when these elements are jointly modeled, multimodal learning can substantially enhance the robustness, accuracy, and interpretability of myocardial scar segmentation.

5.6 Ablation Study: Analysis of Multimodal Fusion Strategies

Effectively integrating electrophysiological information from ECG with anatomical representations from LGE-MRI is a non-trivial challenge due to fundamental differences in spatial structure, signal semantics, and acquisition timing. To understand how different fusion strategies influence myocardial scar segmentation performance, a systematic comparison of unimodal baselines and multiple multimodal fusion paradigms is conducted. All models employ the same MRI [27] and ECG [170] backbone to ensure a fair comparison, differing only in their fusion strategy and training configuration. The quantitative results are summarized in Table 5.2.

Table 5.2: Comparison of different multimodal fusion strategies for ECG-MRI scar segmentation.

Model	Fusion Strategy	Input Modality	Avg. Dice
Unimodal	–	LGE	0.6549
Unimodal (finetuned)	–	LGE	0.6712
Multimodal	Concatenation	LGE + ECG + Time	0.6329
Multimodal	Residual Concat	LGE + ECG + Time	0.6211
Multimodal	Attention	LGE + ECG + Time	0.6774
Multimodal	Multi-head Attention	LGE + ECG + Time	0.6576

5.6.1 Unimodal Baselines and the Effect of Training Data

Firstly, unimodal baselines are established using nnU-Net trained under different data regimes. When trained on the clinical dataset, achieves a 0.6549 Dice score, demonstrating the importance of real-world data for accurate scar segmentation. Fine-tuning on a synthetic dataset generated by CLAIM (see chapter 4) provides a modest additional gain (Dice 0.6712), suggesting that synthetic finetuning offers non-negligible benefits.

These unimodal baselines form a strong reference point for evaluating the contribution of multimodal fusion strategies.

5.6.2 Why Naïve Fusion Strategies Fail

While Section 5.5.2 established the fundamental challenges of multimodal ECG–MRI integration from a representation perspective, this section examines how these challenges manifest in practice through specific naïve fusion strategies.

Simple multimodal fusion approaches are evaluated, including direct concatenation and residual concatenation of ECG features with MRI feature maps. Despite incorporating ECG signals and acquisition time as additional inputs, both approaches underperform the unimodal baseline, achieving Dice scores of 0.6329 and 0.6211 respectively.

This degradation highlights the practical consequences of misaligned multimodal representations. ECG signals encode global electrophysiological behavior without explicit spatial correspondence, whereas MRI features are spatially localized. Directly injecting ECG features into spatial MRI representations forces the network to learn correspondences that are weakly defined or physiologically ambiguous. In this setting, the additional modality does not provide complementary information in a structured manner, but instead introduces noise and destabilizes training.

These results extend the observations in Section 5.5.2 by demonstrating that the failure of multimodal fusion is not only theoretical but also strongly dependent on how fusion is implemented. Naïve strategies such as feature concatenation fail to respect the structural and functional differences between modalities, leading to degraded segmentation performance.

These findings emphasize that multimodal learning is not inherently beneficial and that poorly designed fusion mechanisms can be actively harmful.

5.6.3 Effect of Attention-Based Multimodal Fusion on Clinical Scar Quantification

While Dice score provides a useful summary of spatial overlap, clinical decision-making often depends more directly on accurate quantification of scar burden. To assess the impact of multimodal fusion on clinically relevant metrics, the image-only baseline is compared with the attention-based multimodal model in terms of precision, sensitivity, specificity, and scar volume estimation. The results are summarized in Table 5.3.

Table 5.3: *Scar segmentation performance comparison between the image-only baseline and the multimodal attention-based model.*

Model	Precision	Sensitivity	Specificity	Pred. Vol.	Target Vol.	Abs. Vol. Diff.
Unimodal	0.8246	0.6882	0.9996	15.53	20.55	5.44
Multimodal (Attention)	0.8457	0.6889	0.9996	14.43	20.55	6.27

The multimodal attention-based model achieves higher precision (0.8457 vs. 0.8246) while maintaining comparable sensitivity and specificity to the image-only baseline. This improvement in precision indicates a reduction in false-positive scar predictions, suggesting that electrophysiological cues from ECG help suppress anatomically implausible scar regions that may arise from imaging artifacts or noise in LGE MRI. The consistently high specificity for both models reflects the strong anatomical constraints imposed by the myocardium, but the multimodal model demonstrates improved selectivity within these constraints.

However, despite improved precision, the multimodal attention model does not yield better scar volume estimation. As shown in Table 5.3, the absolute scar volume difference increases from 5.44 ml for the baseline to 6.27 ml for the attention-based multimodal model. This finding highlights an important limitation of static attention-based fusion: while attention can refine spatial localization and reduce false positives, it does not guarantee accurate volumetric agreement when electrophysiological information is weakly aligned or temporally inconsistent with structural imaging.

This discrepancy between spatial overlap metrics and volumetric accuracy further motivates the need for more principled fusion mechanisms. In particular, accurate scar quantification requires not only selective feature weighting, but also explicit modeling of modality reliability and temporal relevance. These observations directly inform the design of the proposed TAFF, which introduces adaptive gating and temporal conditioning to better regulate the contribution of ECG features and improve both segmentation accuracy and scar burden estimation.

5.6.4 Effect of Attention-Based Multimodal Fusion on Distribution of Segmentation Performance and Scar Volume Estimation

To further characterize model behavior beyond aggregate metrics, the distribution of segmentation accuracy and scar volume estimation errors across individual test cases is analyzed. Figures 5.4– 5.9 provide a comprehensive comparison between the unimodal (image-only) baseline and the multimodal attention-based model using scatter plots, violin plots, and box plots for Dice score, precision, sensitivity, specificity, and absolute scar volume difference.

Scar Volume Agreement Across the Full Dynamic Range

Figure 5.4 shows predicted scar volume plotted against ground-truth volume, with the dashed diagonal indicating perfect agreement. Both models exhibit a tendency to underestimate scar volume, particularly for larger scars, reflecting the difficulty of delineating diffuse or heterogeneous enhancement patterns in LGE MRI. However, the multimodal model demonstrates tighter clustering around the identity line for moderate scar volumes, indicating improved consistency in volumetric estimation across a broad range of cases.

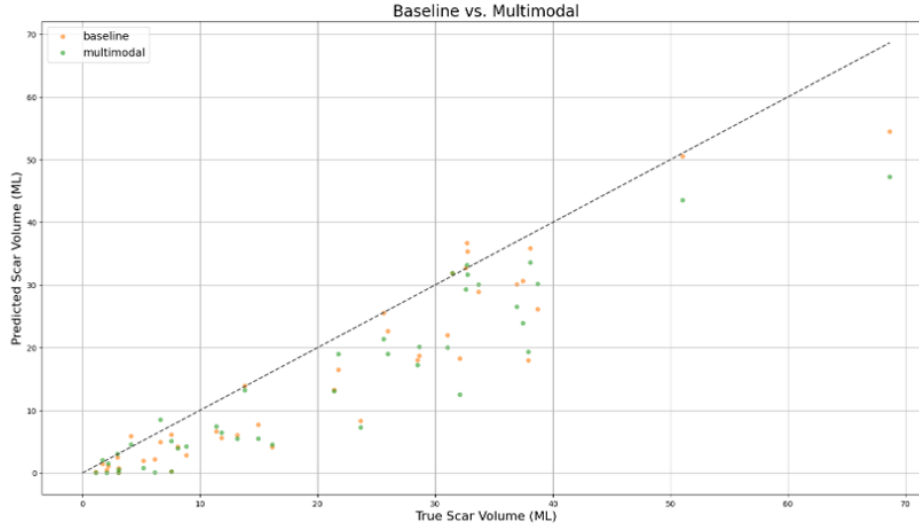


Figure 5.4: *Predicted vs. ground-truth scar volume of test cases (baseline vs. multimodal).*

Dice Score Distribution and Robustness

The Dice score distributions in Figures 5.5 reveal that the attention-based multimodal model achieves a slightly higher median Dice score and a more compact interquartile range. Notably, the baseline exhibits a heavier lower tail, corresponding to severe failure cases. The multimodal model suppresses these low-performing outliers, indicating improved robustness in challenging cases where image contrast alone is insufficient.

At higher scar volumes, deviations increase for both methods, but attention based multimodal predictions show reduced dispersion relative to the baseline, suggesting that ECG-derived information stabilizes volumetric estimates even when absolute errors remain non-negligible.

Precision, Sensitivity, and Specificity Trade-offs

Figures 5.6 and 5.7 present the distributions of precision and sensitivity, respectively. Precision shows a consistent upward shift for the multimodal model, reflecting a reduction in false-positive scar predictions. Sensitivity distributions largely overlap, indicating that improved precision is achieved without sacrificing the detection of true scar regions.

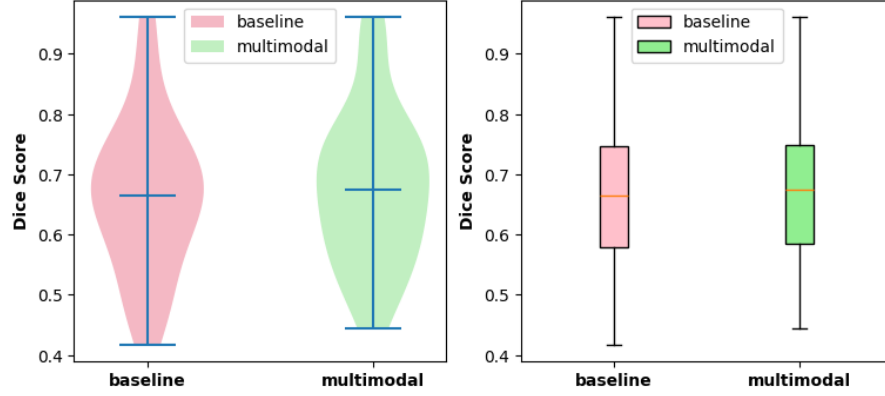


Figure 5.5: *Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of DICE SCORE*

Specificity distributions in Figure 5.8 are uniformly high for both models, consistent with strong anatomical constraints imposed by the myocardium. The multimodal model exhibits slightly reduced variance, suggesting more stable background suppression across subjects.

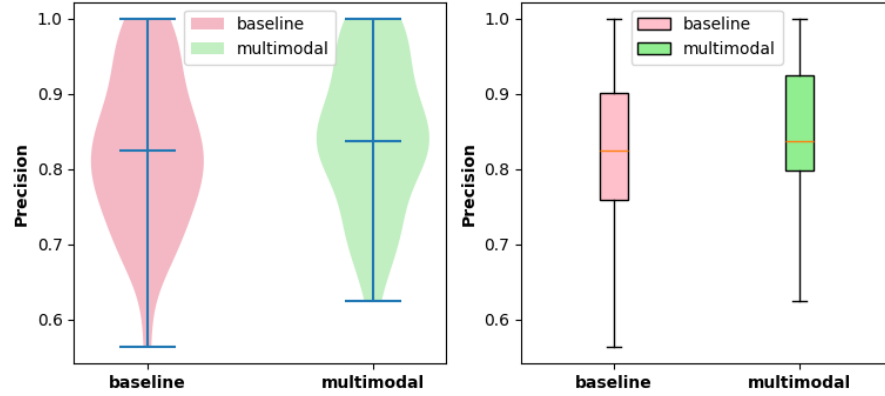


Figure 5.6: *Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of PRECISION*

Absolute Scar Volume Error Distribution

Figure 5.9 further illustrates the distribution of absolute scar volume differences. While median volume errors are comparable between models, the multimodal approach exhibits a longer upper tail, indicating that a subset of cases experience larger volumetric discrepancies. These cases often correspond to extensive or diffuse scarring, where electrophysiological

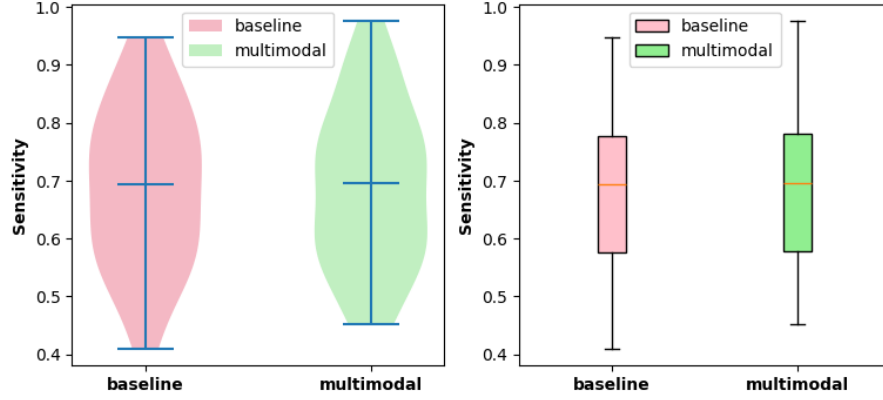


Figure 5.7: *Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of SENSITIVITY*

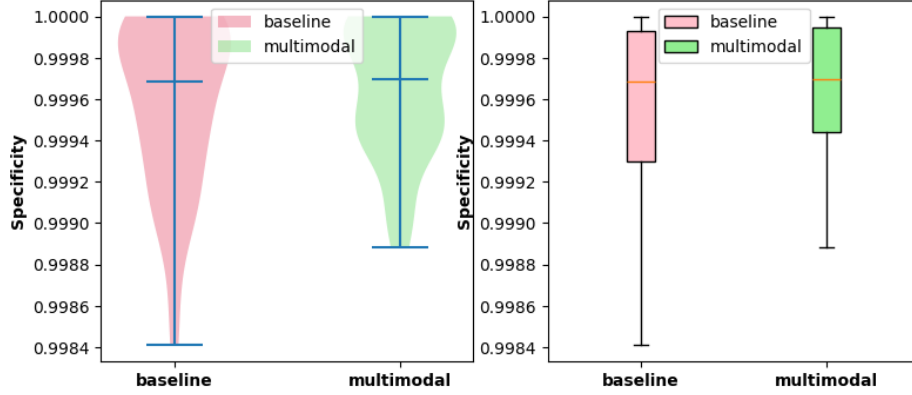


Figure 5.8: *Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of SPECIFICITY*

signals may reflect global conduction abnormalities that are difficult to localize spatially.

Overall, Figures 5.4–5.9 demonstrate that the attention based-multimodal model provides more consistent segmentation performance across patients, particularly by reducing extreme failures and improving precision. However, improvements in volumetric accuracy remain uneven, highlighting the limitations of static attention-based fusion and motivating the development of more principled temporal-aware fusion strategies.

5.6.5 Benefits and Limitations of Attention-Based Fusion

To mitigate the limitations of naïve fusion, attention-based strategies were explored to allow the model to selectively weight ECG features. Single-head attention improves performance

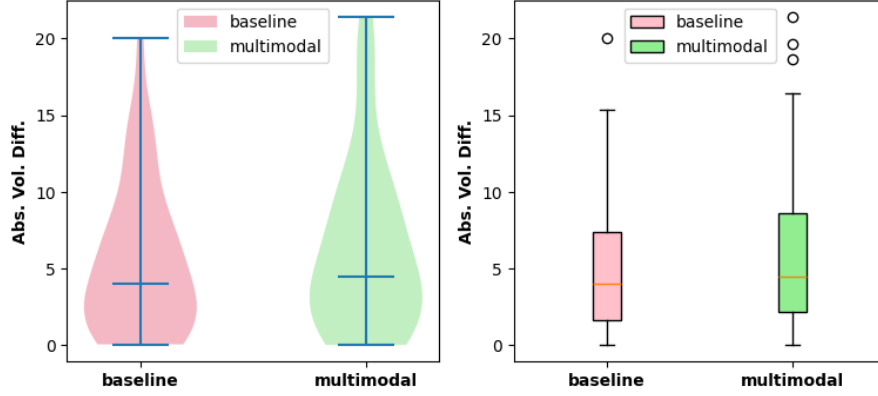


Figure 5.9: *Distributional comparison between the image-only baseline and the attention-based multimodal model across segmentation accuracy in terms of ABSOLUTE VOLUME DIFFERENCE*

to a Dice score of 0.6774, outperforming both unimodal baselines and simple fusion methods. This improvement indicates that attention mechanisms partially resolve modality mismatch by enabling soft selection of informative ECG features.

However, multi-head attention does not yield further gains, achieving a Dice score of 0.6576. While multi-head attention increases representational capacity, it also introduces additional parameters that can be difficult to optimize in limited-data regimes. Without explicit constraints on modality reliability or temporal alignment, attention mechanisms may overfit spurious correlations or attend to temporally inconsistent signals.

These results suggest that attention alone is insufficient for robust ECG–MRI fusion in clinical settings.

5.6.6 Why Temporal-Aware Gated Fusion Is Necessary

The comparative results in Table 5.2 motivate the design of the proposed TAFF framework. Unlike static attention, TAFF explicitly incorporates the acquisition time difference between ECG and MRI as a conditioning signal, reflecting the clinically grounded assumption that electrophysiological relevance decays over time.

By modulating fusion weights based on temporal proximity, TAFF enables the network to down-weight stale or unreliable ECG features while preserving their contribution when temporally consistent. In addition, the use of gated residual fusion provides a mechanism for soft modality selection, allowing the network to rely predominantly on MRI features when ECG information is weak or contradictory.

This design addresses three critical requirements for effective multimodal fusion in clinical data: selective integration, anatomical grounding, and temporal awareness. The superior performance of TAFF observed in subsequent experiments confirms that these requirements are not optional refinements, but essential components for achieving robust and physiologically meaningful multimodal scar segmentation.

5.6.7 Summary of Fusion Strategy Insights

Overall, this analysis demonstrates that effective ECG–MRI fusion requires more than architectural complexity. Naïve fusion strategies fail due to modality mismatch, attention-based methods offer partial improvements but remain limited, and only temporally aware, gated fusion mechanisms provide consistent and clinically plausible gains. These findings directly inform the design rationale of the proposed framework and underscore the importance of incorporating domain knowledge into multimodal learning architectures.

5.7 Results: Qualitative Visualization of Representative Cases

To complement the quantitative evaluation and ablation studies, a detailed qualitative analysis of representative test cases is presented to better understand how the proposed multimodal framework behaves in practice. Figures 5.10–5.15 illustrate six representative examples, including both successful predictions and challenging or failure cases. For each case, (i) the segmentation outputs of the image-only baseline and the proposed multimodal model, (ii) the corresponding AHA-17-based scar volume differences with respect to the ground truth, (iii) the learned ECG cross-lead attention maps, and (iv) the lead-wise ECG waveforms that contribute to the fusion process, are visualized.

These visualizations allow to analyze not only segmentation accuracy, but also the physiological plausibility and interpretability of the model’s predictions.

5.7.1 Cases with Clear Improvement Using TAFF Multimodal Fusion

Figures 5.10, 5.11, and 5.12 present cases where the proposed multimodal model substantially outperforms the image-only baseline. In these examples, the baseline model either underestimates scar extent or fails to capture thin or spatially heterogeneous scar regions, particularly along the subendocardial or lateral myocardial walls (highlighted by yellow arrows).

In contrast, the multimodal predictions demonstrate improved segmentation of scar boundaries, especially in regions that are ambiguous in LGE alone. This improvement is reflected quantitatively in the AHA-17 bull’s-eye plots, where the multimodal model exhibits scar volume differences that are closer to zero across affected segments, particularly in the mid-ventricular and apical regions. The baseline predictions, by comparison, show systematic underestimation in these segments.

The corresponding ECG attention maps reveal that, in these cases, the model assigns higher weights to clinically relevant leads, such as Leads II, III, aVF, and precordial leads V4–V6, which are known to be informative for inferior and lateral wall abnormalities. The temporal attention further highlights characteristic ECG features, including Q waves and abnormal ST–T morphology, which align with the scar locations identified in the LGE images. These observations suggest that the multimodal framework is able to leverage electrophysiological cues to disambiguate visually subtle scar patterns, resulting in more anatomically and physiologically consistent segmentation.

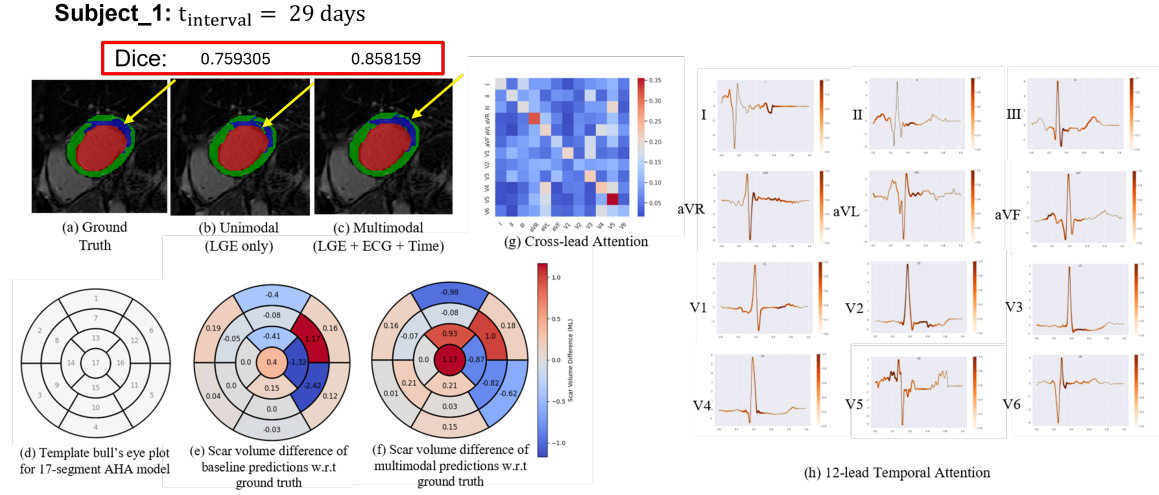


Figure 5.10: *Qualitative example (1): Representative success case showing improved scar segmentation using the proposed TAFF multimodal framework. The AHA-17 bull's-eye plots illustrate reduced regional scar volume error for the multimodal prediction compared to the baseline.*

From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the left ventricular blood pool, healthy myocardium, and scar are visualized in their respective colors.

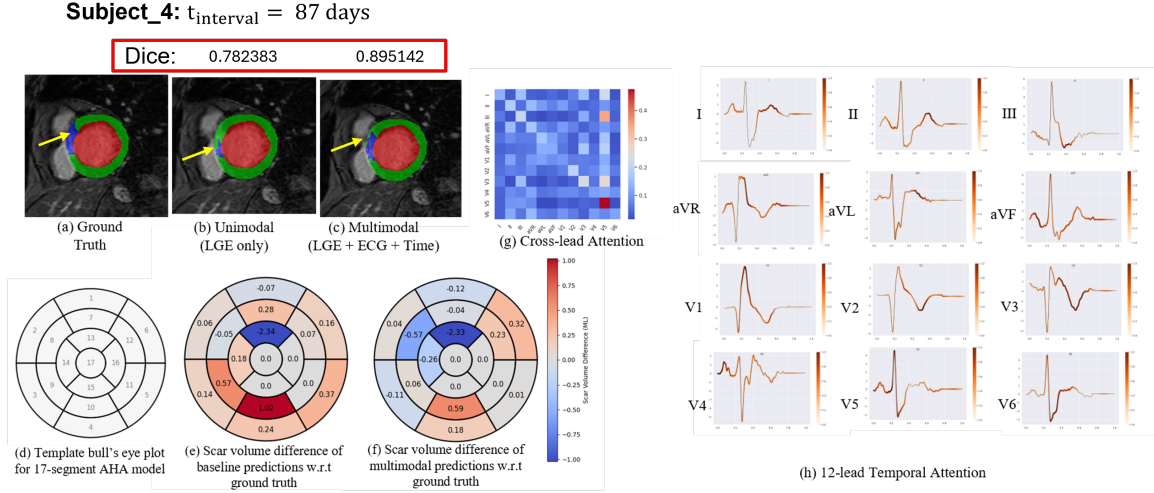


Figure 5.11: *Qualitative example (4): Clear improvement case where multimodal fusion refines fragmented baseline predictions. Anatomical priors constrain predictions to physiologically plausible regions.*

From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the *left ventricular blood pool*, *healthy myocardium*, and *scar* are visualized in their respective colors.

5.7.2 Moderate Improvements and Partial Corrections

Figure 5.13 illustrates a case where the multimodal approach yields moderate but meaningful improvements over the baseline. In this example, both models identify the general scar location; however, the baseline predictions tend to produce fragmented scar regions and imprecise boundaries. The multimodal model refines these predictions by improving spatial coherence and reducing false negatives along the myocardial border.

The AHA-17-based analysis shows that while some regional volume discrepancies persist, particularly in apical segments, the multimodal predictions reduce the overall error magnitude and produce a more balanced regional distribution. The ECG attention maps in these cases are less sharply peaked than in the strongest success cases, indicating weaker but still informative electrophysiological signals. This suggests that even when ECG evidence is subtle or partially noisy, the fusion mechanism can still provide a stabilizing effect that improves segmentation consistency.

5.7.3 Failure and Challenging Cases

Figures 5.14 and 5.15 highlight representative failure cases, where neither the baseline nor the multimodal model fully recovers the ground truth scar distribution. In this example, the scar exhibits highly diffuse enhancement with low contrast relative to surrounding myocardium, making it challenging to identify from LGE alone. Additionally, the ECG waveforms show relatively weak or non-specific abnormalities, resulting in diffuse attention patterns across

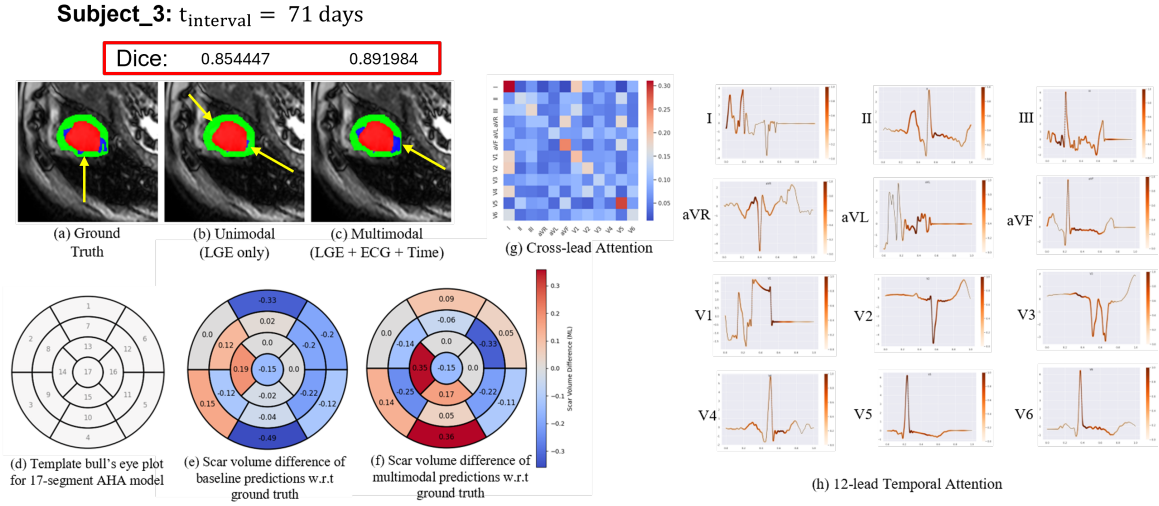


Figure 5.12: *Qualitative example (3): Improved boundary consistency and regional scar volume estimation achieved through ECG-guided fusion. The multimodal prediction aligns more closely with the ground truth across apical AHA segments.*

From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the *left ventricular blood pool*, *healthy myocardium*, and *scar* are visualized in their respective colors.

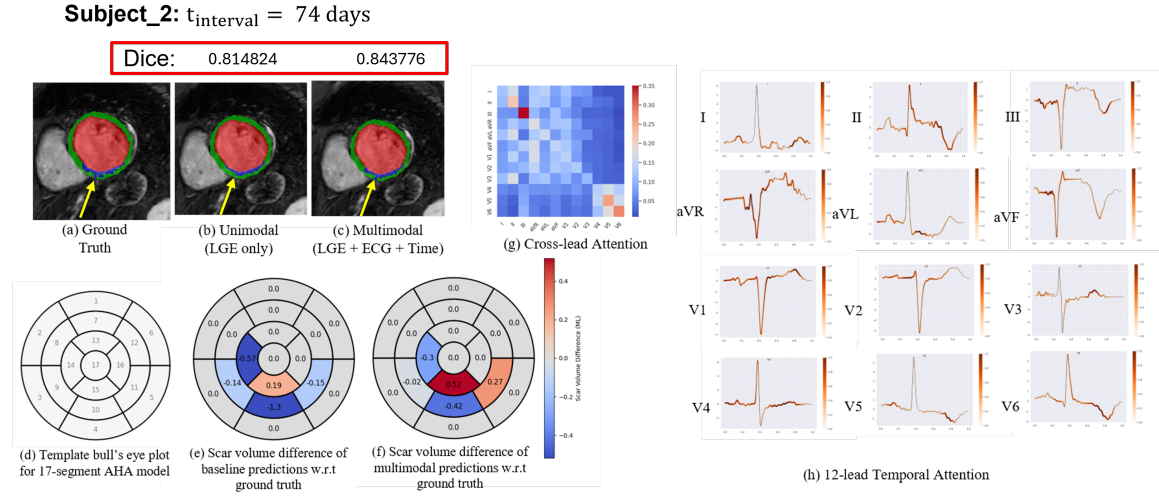


Figure 5.13: *Qualitative example (2): TAFF Multimodal fusion corrects under-segmentation observed in the image-only baseline, particularly in mid-ventricular segments. ECG attention highlights clinically relevant leads associated with the affected myocardial regions.*

From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d) template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the *left ventricular blood pool*, *healthy myocardium*, and *scar* are visualized in their respective colors.

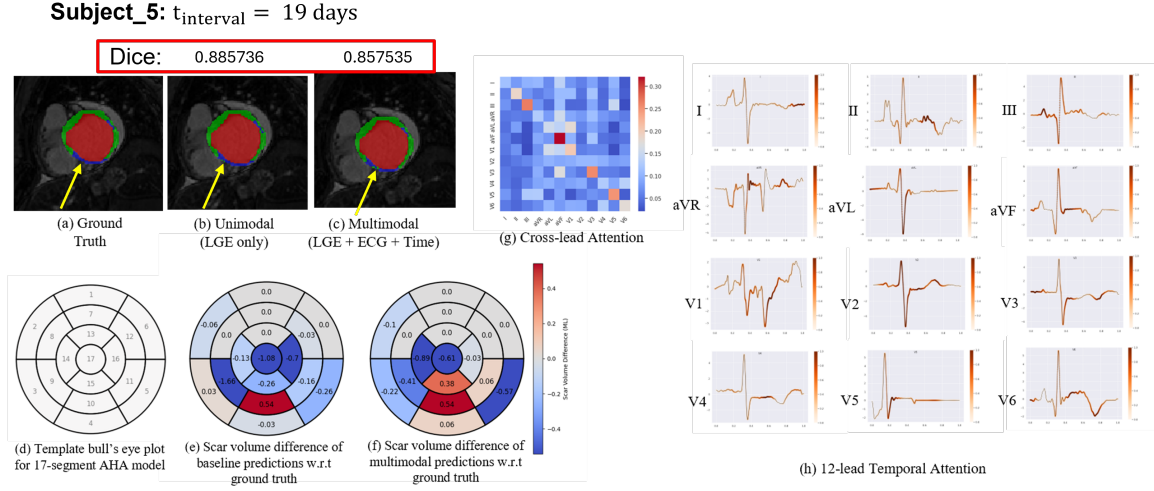


Figure 5.14: *Qualitative example (5): Challenging case showing reduced scar volume error in selected AHA segments. ECG attention is weaker but still contributes to stabilizing segmentation.*

From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the *left ventricular blood pool*, *healthy myocardium*, and *scar* are visualized in their respective colors.

leads.

Although the multimodal model marginally improves regional volume estimates in some AHA segments, it also introduces overcorrection in others, as reflected by alternating positive and negative volume differences in the bull's-eye plot. This behavior indicates that when both imaging contrast and electrophysiological signals are ambiguous, the fusion mechanism may struggle to assign reliable weights, limiting its corrective capability.

Importantly, however, these failure cases do not exhibit catastrophic errors or anatomically implausible predictions. The segmentation remains constrained within the myocardial wall, highlighting the stabilizing role of the anatomical prior even when multimodal cues are weak.

5.7.4 Interpretability Through ECG Attention and Anatomical Context

Across all cases, the ECG attention maps provide valuable insights into the decision-making process of the model. Successful cases are characterized by focused attention on clinically meaningful leads and waveform segments, whereas failure cases show more diffuse or inconsistent attention. This behavior aligns with clinical intuition, reinforcing the interpretability of the proposed framework.

Furthermore, the integration of the AHA-17 anatomical prior enables region-wise analysis that goes beyond global Dice scores. By examining scar volume discrepancies at the segment level, a clearer understanding was gained of where the model succeeds and fails, and how electrophysiological information contributes differently across myocardial regions.

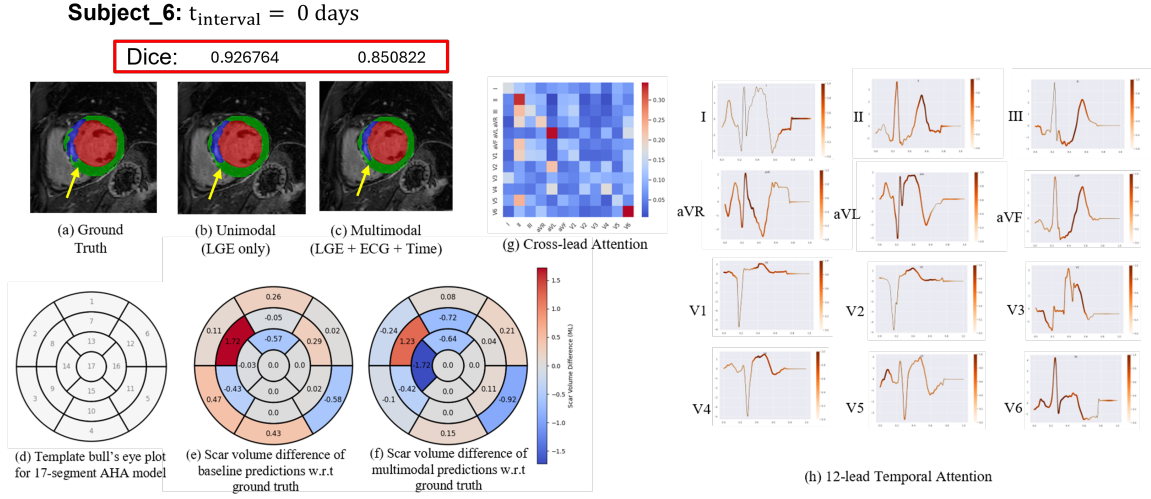


Figure 5.15: *Qualitative example (6): Failure case illustrating limitations of multimodal fusion when both LGE contrast and ECG signals are ambiguous. While regional errors persist, anatomically implausible predictions are avoided due to the anatomical prior.*

From left to right: (a) ground truth, (b) image-only baseline segmentation, (c) multimodal segmentation, (d), template bulls' eye plot, (e and f) bulls' eye plots depicting scar volume differences of unimodal and multimodal predictions with respect to the ground truth, (g) ECG cross-lead attention map, and (h) lead-wise ECG temporal attention maps. In (a–c), the *left ventricular blood pool*, *healthy myocardium*, and *scar* are visualized in their respective colors.

5.7.5 Summary of Qualitative Findings

Overall, the qualitative analysis supports the quantitative results and demonstrates that the proposed multimodal framework:

- Improves scar boundary segmentation and regional consistency in cases with subtle or heterogeneous enhancement.
- Leverages ECG-derived electrophysiological cues in a physiologically meaningful manner, as evidenced by interpretable attention patterns.
- Maintains anatomically plausible predictions even in challenging or failure cases due to the incorporation of anatomical priors.
- Exhibits limitations in cases with weak imaging contrast and non-specific ECG signals, highlighting directions for future improvement.

These findings underscore the value of integrating imaging, electrophysiological, anatomical, and temporal information for robust myocardial scar segmentation, while also clarifying the conditions under which multimodal fusion provides the greatest benefit.

5.8 Limitations and Future Work

While the proposed multimodal framework demonstrates clear advantages over image-only baselines, several limitations remain that warrant careful consideration. Addressing these limitations provides important insight into the current performance ceiling of ECG–MRI fusion and outlines concrete directions for future research.

A primary limitation of the proposed approach lies in the quality and consistency of the ECG signals. In this study, ECGs were digitized from paper-based recordings, a process that inevitably introduces noise, signal distortion, and occasional loss of waveform fidelity despite careful manual quality control. In addition, ECG signals are inherently sensitive to variations in electrode placement and patient-specific anatomical factors, which can influence signal amplitude and morphology independently of underlying pathology. Variability in both digitization quality and acquisition conditions may therefore partially explain why the observed ECG-driven performance gains, while statistically significant, remain modest in some cases. In particular, subtle electrophysiological markers associated with myocardial scar may be attenuated or obscured during digitization or confounded by acquisition variability, limiting the informativeness of ECG-derived features. Future work could benefit from more robust ECG digitization pipelines, standardized acquisition protocols, or, where possible, direct access to digitally acquired ECG signals.

Another limitation arises from the imbalanced distribution of ECG–MRI acquisition intervals across the dataset. As illustrated in Fig. 5.3, some time-interval bins contain substantially fewer samples than others, which may bias both training and evaluation. Although the TAFF mechanism improves robustness across varying temporal gaps, its effectiveness is constrained by the availability of representative training data at larger time intervals. Expanding the dataset to include a more balanced distribution of acquisition times would allow more reliable learning of temporal dependencies and enable finer-grained analysis of temporal effects.

Despite these challenges, Fig. 5.3 provides important insight into the behavior of the proposed temporal conditioning strategy. The model equipped with time conditioning consistently exhibits higher median Dice scores across most time bins compared to the non-temporal variant. Notably, more pronounced gains are observed in cases with acquisition intervals of approximately one week to one month. This suggests that temporal conditioning is particularly beneficial when the ECG–MRI gap is large enough to introduce potential misalignment, but not so large that electrophysiological information becomes entirely obsolete. These findings support the hypothesis that explicitly modeling temporal uncertainty enables the network to down-weight outdated or less reliable ECG features while still exploiting residual physiological information.

At the same time, performance improvements at very short time intervals are less pronounced, indicating that when ECG and MRI are acquired in close temporal proximity, the added benefit of temporal modulation may be limited. In such cases, ECG-derived features may already be highly informative, reducing the need for aggressive temporal weighting. Conversely, at extremely long intervals, electrophysiological changes may reflect disease progression or intervention effects that are no longer aligned with the MRI findings. These observations highlight the complex, non-linear relationship between temporal proximity and multimodal relevance, suggesting that more expressive temporal modeling strategies may be required.

Several avenues for future work emerge from these findings. First, extending the proposed framework from 2D slice-based processing to full 3D volumetric modeling may improve spatial consistency and enable better exploitation of inter-slice contextual information, particularly for capturing complex scar geometries. Second, more advanced temporal modeling approaches, such as continuous-time embeddings or uncertainty-aware temporal weighting, could further enhance robustness under large acquisition gaps. Third, integrating additional clinical metadata, such as intervention history or symptom onset timing, may provide complementary context that helps disambiguate temporal relationships between modalities.

Finally, while this chapter focuses on ECG–MRI fusion for myocardial scar segmentation, the proposed temporal-aware multimodal framework is not limited to this specific application. The underlying principles of adaptive modality weighting and anatomically grounded fusion may be extended to other multimodal cardiac analysis tasks, such as arrhythmia risk stratification or therapy planning, where heterogeneous and temporally misaligned data are the norm. Addressing the limitations identified here will be critical for translating multimodal learning approaches into robust, clinically deployable systems.

5.9 Conclusion

This chapter investigated the integration of electrophysiological and anatomical information for myocardial scar segmentation, addressing key limitations of image-only approaches that rely exclusively on LGE CMR. While LGE imaging provides detailed structural characterization of myocardial fibrosis, its interpretation is often challenged by heterogeneous enhancement patterns, low contrast, and acquisition related artifacts. By incorporating complementary ECG-derived electrophysiological cues and standardized anatomical priors, the proposed framework moves toward a more physiologically grounded approach to cardiac scar mapping.

A multimodal segmentation architecture was introduced that combines LGE-CMR with 12-lead ECG signals, guided by the AHA-17 left ventricular segmentation model to provide anatomically interpretable spatial context. Central to the framework is the TAFF mechanism, which explicitly models the acquisition time difference between ECG and MRI. This temporal awareness enables adaptive modulation of multimodal feature contributions, allowing the network to exploit informative electrophysiological signals while mitigating the impact of temporal misalignment commonly encountered in real-world clinical workflows.

Comprehensive evaluation on a clinically acquired dataset demonstrated that the proposed multimodal approach consistently outperforms a strong image-only baseline based on nnU-Net. Improvements were observed not only in Dice similarity coefficient but also in precision, sensitivity, and scar volume, indicating enhanced boundary accuracy and more reliable quantification of scar burden. Ablation studies further confirmed that meaningful performance gains arise only when anatomical priors, adaptive multimodal fusion, and temporal conditioning are jointly employed, underscoring the importance of principled fusion design.

Beyond quantitative performance, qualitative analysis and attention-based interpretability revealed that the model leverages clinically relevant ECG features and aligns them with anatomically plausible myocardial regions. These findings suggest that the proposed framework does not merely aggregate heterogeneous inputs, but learns physiologically meaningful associations between electrical conduction abnormalities and structural scar patterns.

In summary, this chapter demonstrates that integrating anatomical, functional, and temporal information enables more robust and clinically consistent myocardial scar segmentation than imaging alone. The proposed multimodal framework represents a step toward patient-specific and physiology-aware cardiac image analysis and lays the foundation for future extensions to richer multimodal settings, larger-scale clinical cohorts, and domain-generalized cardiac modeling.

Chapter 6

Robust Scar Segmentation from Incomplete CMR

This chapter addresses a clinically important but underexplored challenge in automated myocardial scar segmentation: learning under the combined effects of missing imaging modalities, incomplete annotations, and multi-center variability. While the preceding chapters focused on improving segmentation performance under more controlled settings; real-world CMR datasets are frequently heterogeneous, partially labeled, and only partially observed. Such conditions can substantially degrade model performance and limit translation into practice. This chapter therefore investigates myocardial scar segmentation under incomplete multi-center CMR and evaluates a set of strategies designed to improve robustness, including missing-modality-aware fusion, hinge-based supervised optimization, semi-supervised teacher-student learning, and domain-randomizing augmentation.

6.1 Introduction

CMR imaging plays a central role in the assessment of myocardial infarction, where the presence, extent, and spatial distribution of ventricular scarring directly inform treatment planning for life-threatening arrhythmias such as ventricular tachycardia. Modern CMR protocols routinely acquire multiple complementary sequences, each emphasizing different aspects of myocardial structure and pathology. bSSFP cine imaging provides high-contrast visualization of cardiac anatomy and motion, enabling accurate segmentation of ventricular boundaries. T2-weighted imaging highlights myocardial edema and acute injury, while LGE imaging remains the clinical reference standard for identifying chronic infarct scar. Together, these sequences offer a rich but heterogeneous representation of myocardial pathology.

Despite these advances, coronary heart disease remains a leading cause of mortality worldwide, with sudden cardiac death accounting for a substantial proportion of global deaths [180]. Myocardial infarction, typically caused by acute coronary occlusion, leads to irreversible myocardial necrosis and subsequent scar formation [181]. These scars form a critical arrhythmogenic substrate that can sustain re-entrant ventricular tachycardia, making accurate scar characterization essential for patient-specific risk stratification and ablation planning. Automated segmentation of myocardial pathology from CMR therefore represents a key enabling technology for scalable and reproducible clinical decision support.

In recent years, deep learning methods have shown considerable promise for automating myocardial segmentation, driven in part by the availability of multi-sequence CMR datasets and benchmarking initiatives such as MyoPS and EMIDEC [10–12]. Numerous approaches have been proposed that integrate information from LGE, T2-weighted, and cine sequences using convolutional neural networks and attention-based fusion strategies [182–187]. These methods demonstrate strong performance under curated experimental conditions and highlight the benefits of multi-sequence fusion for myocardial pathology segmentation.

However, a critical limitation of existing work is that most methods are developed and evaluated under relatively controlled assumptions that do not reflect real-world clinical data. In particular, many approaches are trained on data from a single center or a small number of harmonized sites, leaving their robustness to cross-center variability largely untested [12, 105]. In practice, CMR acquisition protocols vary substantially across institutions due to differences in scanner vendors, field strengths, sequence implementations, and reconstruction pipelines, leading to pronounced domain shifts. In addition, most multi-sequence segmentation frameworks implicitly assume that all imaging sequences are available at both training and inference time. This assumption is often violated in clinical workflows, where sequences may be missing due to time constraints, motion artifacts, or protocol differences.

Learning under such incomplete and heterogeneous conditions presents a significant challenge. When one or more modalities are absent, conventional multi-sequence models can fail or produce unstable predictions, particularly when missing modalities correspond to the primary source of pathological contrast. For example, myocardial scar is predominantly visible in LGE, while edema is primarily captured in T2-weighted imaging. Missing modalities therefore do not simply reduce redundant information but may remove direct evidence for specific pathological structures. Furthermore, incomplete annotations introduce additional ambiguity during training, as unlabeled regions may correspond either to background or to missing labels.

Recent work in multimodal learning has explored strategies for handling missing modalities, including modality dropout, adaptive fusion, and transformer-based architectures that dynamically reweight available inputs [188, 189]. These approaches have shown promise in other medical imaging domains, such as brain tumor segmentation, by enabling flexible integration of heterogeneous inputs. However, their reliability under combined conditions of modality incompleteness and domain shift remains an open question. In addition, studies in general computer vision have shown that models trained on incomplete inputs can exhibit pathological overconfidence, producing predictions with high certainty despite reduced evidence [190]. Such behavior is particularly concerning in safety-critical clinical applications.

Against this backdrop, this chapter investigates myocardial pathology segmentation under a more clinically realistic setting defined by the joint presence of missing imaging modalities, incomplete annotations, and multi-center variability. Rather than assuming idealized conditions, the aim is to understand how these factors affect segmentation performance and to evaluate strategies that may improve robustness under such constraints. To this end, a modality-masked fusion transformer (M²FTrans) [189] is adopted as the core framework, as it enables flexible learning from arbitrary subsets of available sequences. Building on this backbone, several complementary strategies are examined, including domain-randomizing augmentation, hinge-based supervised optimization, and semi-supervised teacher–student learning.

Importantly, this chapter does not assume that methods successful under controlled settings will necessarily translate to heterogeneous clinical data. Instead, it provides a structured evaluation of their behavior under incomplete multi-center conditions. In particular, it is investigated whether semi-supervised learning can effectively leverage partially labeled and heterogeneous data, or whether its performance is limited by the combined effects of domain shift and modality missingness.

The main contributions of this chapter are summarized as follows:

- A unified experimental framework is established for myocardial pathology segmentation under the combined challenges of missing imaging modalities, incomplete annotations, and multi-center variability.
- A modality-masked fusion transformer is adapted and evaluated under both complete and incomplete modality configurations to enable flexible multi-sequence learning.
- The effects of domain-randomizing augmentation and a hinge-based max-margin supervised objective are investigated for improving robustness to domain shift and reduced modality evidence.
- The effectiveness of semi-supervised learning under incomplete multi-center CMR is critically assessed, showing that pseudo-label-based approaches do not consistently outperform fully supervised training in this setting.
- A comparison with recent multi-center myocardial pathology segmentation studies is provided to contextualize the difficulty of the proposed setting relative to prior work.

By explicitly studying myocardial scar segmentation under incomplete and heterogeneous multi-center CMR, this chapter aims to bridge the gap between high-performing research models and the practical requirements of clinically deployable systems.

6.2 Background

6.2.1 Clinical Motivation: Multi-Sequence CMR in Myocardial Pathology Assessment

CMR imaging has become a cornerstone of non-invasive assessment for ischemic heart disease, offering high soft-tissue contrast and the ability to characterize myocardial structure, function, and tissue composition within a single examination. In the context of myocardial infarction, CMR enables clinicians to identify infarcted myocardium, quantify scar burden, assess myocardial edema, and evaluate ventricular remodeling; factors that are closely linked to arrhythmic risk and long-term prognosis.

Different CMR sequences are designed to emphasize distinct physiological and pathological properties. Cine bSSFP imaging provides high signal-to-noise ratio and excellent blood–myocardium contrast, making it well suited for delineating ventricular anatomy and cardiac function. T2-weighted imaging is sensitive to myocardial water content and is commonly used to identify edema associated with acute ischemic injury and inflammation. LGE imaging exploits contrast kinetics to visualize regions of irreversible myocardial injury and remains the clinical reference standard for scar detection and viability assessment.

In clinical practice, these sequences are interpreted jointly, with clinicians integrating information across modalities to localize pathology, distinguish acute from chronic injury, and guide therapeutic decisions such as catheter ablation. Consequently, automated segmentation systems must not only process heterogeneous image inputs but also account for the complementary and modality-specific nature of myocardial pathology representation.

6.2.2 Challenges in Real-World Multi-Center CMR Data

Despite the availability of recommended imaging protocols, substantial variability exists in how CMR studies are acquired across institutions. Differences in scanner vendors, magnetic field strengths, pulse sequence implementations, contrast administration, inversion time selection, and reconstruction pipelines can lead to pronounced differences in image appearance. Even within a single center, protocol updates and scanner upgrades can introduce additional variability over time.

These factors contribute to *domain shift*, where the statistical properties of images encountered during deployment differ from those observed during training. Deep learning models trained on limited or homogeneous datasets are therefore prone to performance degradation when applied to external data, particularly for tasks such as myocardial scar segmentation that rely on subtle intensity variations and fine structural detail.

A further challenge is *modality incompleteness*. In routine clinical workflows, not all CMR sequences are acquired for every patient. Certain sequences may be omitted due to time constraints, patient condition, or motion artifacts, while retrospective datasets may only contain a subset of modalities. Importantly, this missingness is often structured rather than random, reflecting clinical priorities or acquisition limitations. Models that assume complete multi-sequence input are therefore poorly aligned with real-world deployment scenarios.

A further challenge, closely related to both domain shift and modality incompleteness, is *incomplete annotation availability*. In multi-center clinical datasets, annotations are often only partially available, with certain structures or pathologies labeled in some centers but not others. For example, edema annotations may be restricted to specific sites, while other centers provide labels only for ventricular structures and scar. This introduces ambiguity during training, as unlabeled regions may correspond either to true background or to missing labels. Conventional supervised learning frameworks assume complete and consistent annotations across the dataset, and may therefore struggle to effectively utilize partially labeled data. Addressing this form of supervision incompleteness is essential for scaling segmentation methods to realistic multi-center settings where exhaustive annotation is rarely feasible.

6.2.3 Limitations of Existing Multi-Sequence Segmentation Approaches

Many state-of-the-art myocardial segmentation methods leverage multi-sequence CMR by combining features from LGE, T2-weighted, and cine images using convolutional neural networks or attention-based fusion mechanisms. While these approaches achieve strong performance on benchmark datasets, they typically rely on two key assumptions: (i) that all modalities are consistently available during both training and inference, and (ii) that the training and testing data distributions are similar.

When these assumptions are violated, performance can degrade significantly. In particular, models trained under complete-modality assumptions may fail when one or more

sequences are missing at inference time. Furthermore, models trained on data from a limited number of centers may learn spurious correlations tied to site-specific appearance characteristics rather than robust anatomical or pathological features. These limitations restrict the clinical applicability of otherwise high-performing segmentation frameworks.

6.2.4 Learning Under Missing Data and Domain Shift

Learning from incomplete data has become an active area of research in multimodal deep learning. In medical imaging, approaches such as modality dropout, generative imputation, and adaptive fusion have been proposed to address missing-modality scenarios. Transformer-based architectures, in particular, offer a flexible framework for integrating heterogeneous inputs by modeling long-range dependencies and dynamically weighting available modalities.

However, robustness to missing data alone is insufficient for clinical deployment. Models must also maintain reliable behavior under domain shift and reduced evidence. Prior studies have shown that deep learning models can exhibit overconfident predictions when presented with incomplete or corrupted inputs, raising concerns about their reliability in safety-critical settings. This is particularly relevant in cardiac imaging, where the absence of a modality may remove key diagnostic information.

Domain generalization provides a complementary perspective by aiming to train models that perform well on unseen domains without requiring adaptation at deployment. In the context of CMR, this involves exposing models to diverse variations in image appearance during training, such as changes in intensity, contrast, and noise, to encourage invariant feature learning. Nevertheless, the combined impact of domain shift and modality incompleteness remains insufficiently explored, particularly for myocardial pathology segmentation.

These challenges motivate the need for approaches that can explicitly handle modality incompleteness, leverage partially labeled data, and remain robust under domain variability. The following subsection outlines the rationale behind the specific strategies investigated in this chapter.

6.2.5 Rationale for Investigated Approaches

Given the combined challenges of modality incompleteness, incomplete annotations, and multi-center variability, several strategies were selected for investigation based on their potential to address different aspects of this problem.

First, a modality-masked fusion transformer (M²FTrans) [189] was chosen as the primary architecture due to its ability to operate under arbitrary subsets of available modalities. Unlike conventional fusion approaches that assume complete inputs, modality-masked attention enables controlled interaction between available sequences, making it a suitable backbone for studying missing-modality scenarios.

Second, a hinge-based max-margin supervised objective was introduced to explicitly address performance degradation under incomplete modality inputs. The motivation for this design is that conventional training objectives do not penalize large discrepancies between complete and incomplete modality conditions, which can lead to unstable behavior when modalities are missing at inference time.

Third, semi-supervised learning based on teacher-student consistency (Mean Teacher [191] and FixMatch [192]) was incorporated to investigate whether partially labeled multi-

center data can be effectively utilized. This is particularly relevant for CMR datasets such as MyoPS++, where annotations are incomplete across centers. Semi-supervised learning provides a natural framework for leveraging such data, although its effectiveness under combined domain shift and modality incompleteness remains unclear.

Finally, domain-randomizing augmentation (RandConv [193]) was employed to improve robustness to cross-center variability. By introducing controlled perturbations in image appearance, RandConv encourages the model to learn features that are less sensitive to scanner- and protocol-specific characteristics, thereby supporting domain generalization.

Together, these strategies were selected not as a single unified solution, but as complementary approaches targeting different failure modes associated with incomplete and heterogeneous multi-center CMR data. Their combined evaluation enables a systematic investigation of which components contribute to improved robustness under realistic conditions.

6.2.6 Positioning of This Chapter

Against this background, this chapter focuses on myocardial pathology segmentation under conditions that more closely reflect real clinical practice, where data are heterogeneous, partially observed, and incompletely labeled. Unlike prior work that evaluates segmentation under idealized settings, this chapter investigates how segmentation performance is affected when missing modalities, incomplete annotations, and multi-center variability occur simultaneously.

To study this problem, a modality-aware segmentation framework is employed together with robustness-oriented training strategies, including domain-randomizing augmentation, hinge-based supervised objectives, and semi-supervised learning. Rather than assuming that these strategies will necessarily improve performance, the aim is to evaluate their behavior under realistic conditions and to identify both their strengths and limitations. In doing so, this chapter contributes to a more grounded understanding of how myocardial scar segmentation methods perform beyond controlled benchmark settings.

6.3 Methodology

This chapter investigates a set of strategies for myocardial pathology segmentation under incomplete multi-sequence and multi-center CMR. To this end, a modality-masked fusion transformer (M²FTrans) [189] is adopted as the primary backbone due to its ability to operate under arbitrary subsets of available modalities. Building on this architecture, three additional components are examined: (i) a hinge-based supervised objective to improve robustness under missing-modality conditions, (ii) semi-supervised learning to leverage partially labeled data, and (iii) domain-randomizing augmentation using RandConv [193] to mitigate cross-center variability.

Importantly, these components are not assumed to provide uniform improvements. Rather, they are evaluated within a unified framework to understand their behavior under realistic deployment conditions, where modality availability, annotation completeness, and data distributions vary across centers. An overview of the framework is shown in Fig. 6.1.

6.3.1 Overall Framework

The framework is designed to process three complementary CMR sequences; LGE, T2-weighted, and bSSFP, while remaining functional when one or more sequences are absent. During training, both complete and incomplete samples are included within each batch. Incomplete samples arise from two sources: (i) naturally missing modalities in certain centers, and (ii) synthetic modality masking applied during training.

This joint training strategy allows the model to learn consistent representations across varying modality availability. Furthermore, incorporating data from multiple centers introduces natural domain variability, enabling the model to be exposed to realistic inter-site differences. The data usage and supervision strategy are illustrated in Fig. 6.2.

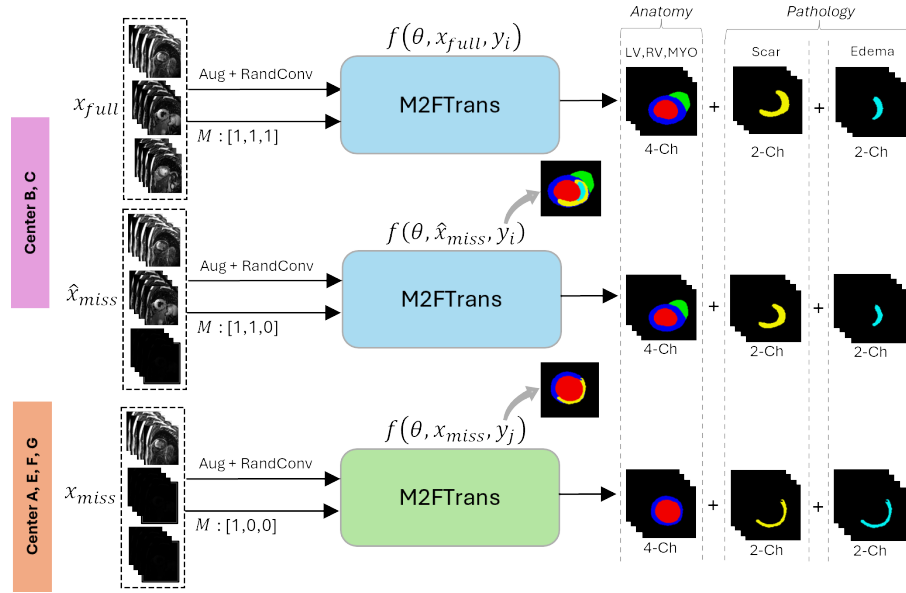


Figure 6.1: Overview of the framework based on $M^2FTrans$ for myocardial pathology segmentation from incomplete multi-sequence CMR. Complete and incomplete samples are jointly processed during training.

6.3.2 $M^2FTrans$: Modality-Masked Fusion Transformer

The modality-masked fusion transformer ($M^2FTrans$) [189] is employed as the core architecture due to its ability to handle missing-modality inputs through masked attention. Unlike conventional fusion approaches that assume complete inputs, $M^2FTrans$ restricts interactions between modalities based on their availability, preventing the propagation of spurious features from missing inputs.

The architecture consists of four main components: sequence-specific encoders, a modality-masked fusion transformer, spatial weight attention, and a channel-wise fusion transformer.

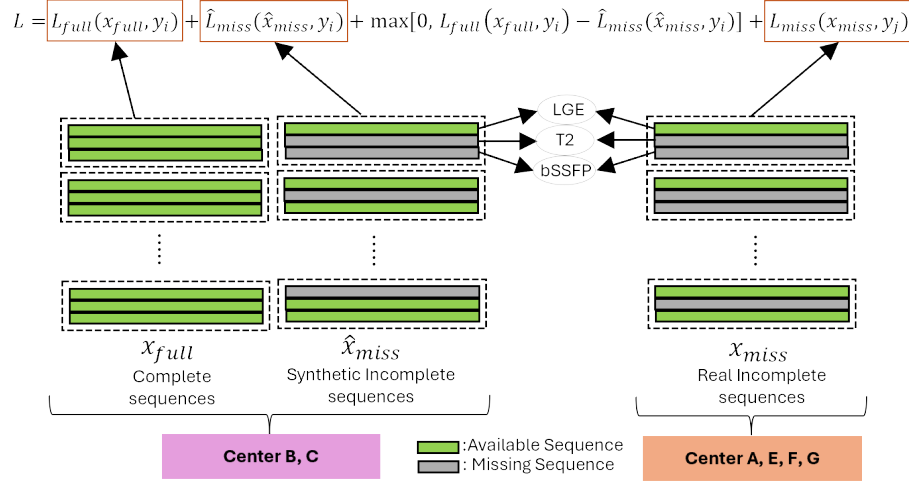


Figure 6.2: Illustration of data usage within a training batch. Samples with complete modalities contribute full supervision, while synthetically or naturally incomplete samples contribute incomplete-modality supervision.

Sequence-Specific Feature Extraction

Each modality $s \in \{\text{LGE}, \text{T2}, \text{bSSFP}\}$ is processed by a dedicated encoder E_s based on a 2D U-Net architecture. This design allows modality-specific feature learning tailored to the contrast characteristics of each sequence. A shared regularization decoder D_{reg} is applied to encourage consistent representations across modalities.

The regularization loss is defined as:

$$L_{reg} = \sum_{s \in \{\text{LGE}, \text{T2}, \text{bSSFP}\}} [L_{Dice}(D_{reg}(E_s(x_s)), y) + L_{WCE}(D_{reg}(E_s(x_s)), y)], \quad (6.1)$$

where x_s denotes the input image for modality s and y is the ground truth segmentation.

Modality-Masked Fusion Transformer

To model interactions between modalities, a transformer module is used with learnable fusion tokens [194]. Sequence-specific tokens are prevented from attending to missing modalities using masking, while fusion tokens aggregate information from all available modalities.

This design ensures that the fused representation is conditioned only on observed inputs, improving stability under incomplete modality configurations.

Spatial Weight Attention

Given that different modalities contribute differently across spatial locations, spatial weight attention [195] is applied to reweight modality contributions. This enables the model to emphasize modality-specific informative regions, such as scar in LGE or edema in T2-weighted images.

Channel-Wise Fusion Transformer

A channel-wise fusion transformer is employed to reduce redundancy and balance contributions across modalities. Grouped attention is applied along the channel dimension, ensuring that no single modality dominates the fused representation.

Decoding with Deep Supervision

The fused features are decoded using a five-stage decoder D_f with deep supervision at multiple levels. The segmentation loss is defined as:

$$L_{seg} = L_{Dice}(D_f(\text{Concat}(E_s(x_s))), y) + L_{WCE}(D_f(\text{Concat}(E_s(x_s))), y), \quad (6.2)$$

and the deep supervision loss is:

$$L_{ds} = \sum_{l=1}^5 \left[L_{Dice}(U_{2^l-1}(\hat{F}_f^l), y) + L_{WCE}(U_{2^l-1}(\hat{F}_f^l), y) \right]. \quad (6.3)$$

Supervised Loss with Max-Margin Constraint

The total supervised loss is:

$$L_t = L_{reg} + L_{seg} + L_{ds}. \quad (6.4)$$

To reduce performance degradation under missing modalities, a hinge-based max-margin objective [196] is introduced:

$$L_{sup} = L_{full} + L_{miss} + \max(0, L_{full} - L_{miss}). \quad (6.5)$$

This encourages consistent performance between complete and incomplete modality settings.

6.3.3 Mean Teacher-Based Semi-Supervised Learning

To utilize partially labeled data, semi-supervised learning based on the Mean Teacher framework [191] is incorporated. A teacher model is maintained as an exponential moving average of the student model:

$$\theta'_t = \alpha \theta'_{t-1} + (1 - \alpha) \theta_t. \quad (6.6)$$

Consistency Loss

For unlabeled samples, pseudo labels $\hat{y}$ generated by the teacher are used for supervision:

$$L_{ureg} = \sum_{s \in S} \text{MSE}(D_{reg}(E_s(x_s)), \hat{y}), \quad (6.7)$$

$$L_{useg} = \text{MSE}(D_f(\text{Concat}(E_s(x_s))), \hat{y}), \quad (6.8)$$

$$L_{uds} = \sum_{l=1}^5 \text{MSE}(U_{2^{l-1}}(\hat{F}_f^l), \hat{y}), \quad (6.9)$$

$$L_{con} = L_{ureg} + L_{useg} + L_{uds}. \quad (6.10)$$

In addition to standard Mean Teacher training, a FixMatch-inspired strategy [192] is also explored, where pseudo labels are generated from weak augmentations and enforced on strongly augmented inputs under a confidence threshold.

6.3.4 Domain-Randomizing Augmentation with RandConv

To improve robustness to cross-center appearance variability, domain-randomizing augmentation was incorporated using RandConv [193]. Multi-center CMR datasets often exhibit substantial differences in image appearance due to variations in scanner vendors, acquisition protocols, reconstruction pipelines, contrast settings, and intensity characteristics. Such differences can lead models to overfit site-specific appearance patterns rather than learning features that generalize across centers.

RandConv addresses this problem by applying randomly generated convolutional filters to input images during training, thereby producing diverse appearance transformations while preserving the underlying anatomical structure. Unlike conventional intensity augmentation alone, this strategy introduces broader texture and contrast variation, effectively exposing the model to a wider range of image styles. In this chapter, RandConv is used as a domain-randomizing mechanism to encourage invariant feature learning under multi-center distribution shifts.

The motivation for including RandConv is not to simulate any single target domain exactly, but rather to reduce sensitivity to scanner- and protocol-specific image characteristics. This is particularly relevant in the present setting, where the model is trained on heterogeneous centers and evaluated on an unseen center. By increasing appearance diversity during training, RandConv is expected to improve robustness to cross-center variation and support more stable segmentation performance under realistic deployment conditions.

6.3.5 Missing-Modality Simulation

To explicitly train the model for incomplete inputs, missing modalities are simulated using a binary mask $M \in \{0, 1\}^3$. Missing modalities are replaced with zero inputs, allowing the model to learn stable behavior across varying modality combinations.

6.3.6 Overall Training Objective

The final objective combines supervised and semi-supervised components:

$$L = L_{sup} + W_c \cdot L_{con}, \quad (6.11)$$

where W_c is a ramped consistency weight [197].

Supervised learning is primarily driven by fully labeled centers (B/C), while partially labeled or unlabeled centers contribute through the consistency objective. RandConv augmentation is applied during training to increase appearance diversity and improve robustness

to cross-center domain shift. This setup enables a unified investigation of segmentation performance under incomplete multi-center CMR conditions.

6.4 Results and Discussion

This section evaluates the investigated framework under three key deployment conditions: (i) multi-center domain shift, (ii) incomplete modality availability at inference, and (iii) incomplete annotation availability during training. The goal is not only to report segmentation performance, but also to understand how different strategies behave under these combined constraints. Both quantitative metrics (Dice, sensitivity, precision, Hausdorff distance) and qualitative analysis are used to identify strengths and failure modes.

6.4.1 Dataset and Experimental Protocol

The experiments were conducted on the MyoPS++ dataset [11–14]¹, a multi-center and multi-sequence CMR cohort designed to reflect realistic clinical variability. The dataset contains 250 subjects collected across 7 centers, with heterogeneous availability of imaging modalities and annotations.

Table 6.1 summarizes the distribution of imaging sequences and ground truth labels across centers.

Table 6.1: *MyoPS++ dataset: multi-center cohort with heterogeneous modality and annotation availability [11–14].*

Center	Num. Patients	Sequences	Ground Truth Labels
A	81	LGE	LV, Myo, scar
B	50	LGE, T2, bSSFP	LV, RV, Myo, edema, scar
C	45	LGE, T2, bSSFP	LV, RV, Myo, edema, scar
D	50	LGE, T2, bSSFP	LV, RV, Myo, edema, scar
E	7	LGE, bSSFP	LV, RV, Myo, scar
F	9	LGE, bSSFP	LV, RV, Myo, scar
G	8	LGE, bSSFP	LV, RV, Myo, scar

The dataset exhibits several forms of incompleteness that are representative of real-world clinical deployment:

Missing modalities

- Center A contains only LGE images (no T2 or bSSFP).
- Centers E, F, and G contain LGE and bSSFP, but no T2-weighted imaging.
- Only centers B, C, and D provide all three modalities (LGE, T2, bSSFP).

¹http://www.zmic.org.cn/care_2024/track4/

Incomplete annotations

- Edema annotations are available only in centers B, C, and D.
- Some anatomical labels (e.g., RV) are absent in certain centers.
- As a result, large portions of the dataset are only partially labeled, requiring strategies that can utilize incomplete supervision.

Multi-center domain variability

- Data are collected from different institutions across the UK, France, and China.
- Variations in scanner vendors, acquisition protocols, and reconstruction pipelines introduce domain shift between centers.

Training and evaluation split Training data are drawn from centers A, B, C, E, F, and G, while validation and testing are performed on center D. This setup enforces an explicit *unseen-center evaluation*, where the model must generalize to a new domain with full modality availability.

Evaluation groups To isolate the effect of modality availability, validation data are grouped as follows:

- **LGE-only (center A):** evaluates performance under severe modality incompleteness.
- **LGE + bSSFP (centers E/F/G):** evaluates intermediate modality availability.
- **Full modalities (centers B/C):** represents the most informative setting.

This grouping enables a controlled analysis of how segmentation performance varies with available modality evidence, while still reflecting realistic acquisition patterns.

Overall, the MyoPS++ dataset provides a challenging testbed for studying myocardial pathology segmentation under the combined effects of missing modalities, incomplete annotations, and multi-center variability. This combination of modality, annotation, and domain incompleteness makes the task substantially more challenging than standard multi-sequence segmentation benchmarks.

6.4.2 Implementation Details

All experiments were conducted using consistent preprocessing and training settings unless otherwise specified. Input images were center-cropped and normalized using z-score normalization. Models were trained for $epoch_{max} = 1000$ epochs with a patch size of 256×256 .

Optimization was performed using AdamW [198] with an initial learning rate of 2×10^{-4} and a weight decay of 10^{-4} . A warm-up strategy was applied during the early training phase, followed by a polynomial learning-rate decay defined as:

$$\text{lr}(e) = \text{lr}_0 \cdot \left(1 - \frac{e}{epoch_{max}}\right)^{0.9}.$$

Due to the dataset structure, only centers B and C provide both *complete modalities* and *complete annotations* (including edema). Accordingly, the training setup is defined as follows:

- **Centers B/C:** used as fully labeled data for supervised learning (L_{sup}),
- **Centers A/E/F/G:** used as partially labeled or unlabeled data for the semi-supervised consistency objective.

For semi-supervised training, the teacher model is updated as an exponential moving average (EMA) of the student model with a decay factor of 0.999. To balance labeled and unlabeled samples during training, a batch size of 24 was used with a 1:1 ratio between labeled and unlabeled data.

Standard on-the-fly data augmentation was applied, including rotation, flipping, and intensity perturbations. In addition, RandConv [193] was used to increase appearance diversity and improve robustness to cross-center variability, particularly differences in image contrast and intensity distributions.

6.4.3 Fully Supervised Learning under Missing-Modality Conditions

A fully supervised regime is first evaluated by training the model using only data from centers B and C (95 labeled samples), where all imaging modalities and complete annotations are available. During training, missing-modality scenarios are simulated by randomly dropping modalities and providing the corresponding modality mask to the model. The trained model is then evaluated across validation groups with varying modality availability to assess robustness under incomplete and cross-center conditions. This setup provides a controlled baseline in which training is performed under ideal conditions, while evaluation reflects the incomplete and heterogeneous nature of real-world multi-center CMR data. Table 6.2 summarizes segmentation performance across validation groups with different modality availability. Several important observations can be made.

Table 6.2: Performance of fully supervised models trained on centers B/C under simulated missing-modality cases.

Methods	Dice Score				
	Scar				Edema
	Center A	Center E,F,G	Center B,C	Average	Center B,C
Supervised	0.2471	0.1350	0.5195	0.3005	0.3206
Supervised + RandConv	0.2667	0.2094	0.6115	0.3625	0.5632
Supervised + RandConv + Hinge loss	0.2871	0.2674	0.6444	0.3996	0.5922

Baseline behavior under domain shift and missing modalities. The purely supervised model achieves reasonable performance on the full-modality validation set (centers B/C), with a scar Dice of 0.5195. However, performance degrades substantially in incomplete-modality settings, dropping to 0.2471 for LGE-only (center A) and further to 0.1350 for LGE+bSSFP (centers E/F/G).

This degradation reflects two interacting factors: (i) *evidence reduction* due to missing modalities, and (ii) *domain shift* between training and evaluation centers. Notably, performance is lowest for centers E/F/G despite having more modalities than center A, suggesting that domain variability can outweigh the benefit of additional inputs when the model is not explicitly trained for cross-center generalization.

Impact of domain generalization via RandConv. Introducing RandConv leads to consistent improvements across all modality groups. In the full-modality setting (B/C), Dice improves from 0.5195 to 0.6115, while in LGE+bSSFP centers (E/F/G), performance increases from 0.1350 to 0.2094. Improvements are also observed in LGE-only cases (center A).

These results indicate that RandConv effectively reduces sensitivity to center-specific appearance characteristics. By exposing the model to diverse image styles during training, it encourages the learning of more invariant representations that generalize better across unseen domains. Importantly, the relative gains are more pronounced in cross-center settings, supporting its role as a domain generalization mechanism [193].

Impact of hinge loss under reduced modality evidence. The addition of the hinge-based max-margin objective further improves performance across all groups, yielding the best results in Table 6.2. Scar Dice increases to 0.6444 on centers B/C, and to 0.2871 and 0.2674 for centers A and E/F/G, respectively.

Unlike RandConv, which primarily addresses domain variability, the hinge loss targets instability caused by missing modalities. By penalizing large performance discrepancies between complete and incomplete modality inputs, it encourages the model to maintain more consistent predictions as evidence is reduced. The larger relative improvements observed in incomplete-modality settings suggest that this objective is particularly effective in mitigating degradation under missing inputs.

Modality-dependent performance trends. Across all configurations, a consistent performance hierarchy is observed:

- Full modalities (centers B/C) achieve the highest performance,
- LGE+bSSFP (centers E/F/G) provide intermediate performance,
- LGE-only (center A) results in the lowest performance.

This trend reflects the modality-specific nature of myocardial pathology. While LGE provides primary contrast for scar detection, additional modalities contribute complementary anatomical and contextual information. However, even with missing-modality-aware training, the performance gap is not eliminated, indicating that certain information cannot be recovered when key modalities are absent.

Interaction between modality availability and performance. A clear interaction is observed between the number and type of available modalities and segmentation performance. As shown in Table 6.2, performance is consistently highest when all modalities are available (centers B/C), and decreases as modality information becomes more limited.

Interestingly, the performance drop is not strictly proportional to the number of available modalities. For example, segmentation performance in LGE+bSSFP centers (E/F/G) is lower than in LGE-only center A for the purely supervised baseline. This suggests that modality availability alone does not determine performance; instead, the interaction between modality combinations and domain shift plays a critical role.

In particular, models trained on full-modality data may learn modality-specific dependencies that do not generalize well when modalities are missing or when modality combinations differ from those seen during training. This indicates that robustness requires not only access to multiple modalities, but also explicit training strategies that account for varying modality configurations.

These observations highlight that modality incompleteness and domain variability should not be treated independently, as their combined effect can lead to non-intuitive performance behavior.

Edema segmentation. For edema segmentation, which is only available in centers B/C, performance improves substantially with both RandConv and hinge loss, increasing from 0.3206 (supervised) to 0.5922 (Supervised+RandConv+Hinge). This suggests that robustness-oriented training benefits not only scar segmentation but also pathology that is strongly modality-dependent (in this case, T2-weighted imaging).

Qualitative Analysis Under Incomplete-Modality Inputs

Figure 6.3 illustrates representative segmentation outputs under different modality combinations. The qualitative results provide further insight into model behavior:

- **Dependence on LGE for scar detection:** Scar segmentation quality degrades noticeably when LGE is absent, consistent with its role as the primary imaging modality for infarct visualization.
- **Role of bSSFP in anatomical stabilization:** The presence of bSSFP improves delineation of ventricular boundaries and myocardium, reducing structural errors and indirectly supporting scar localization.
- **T2 dependence for edema:** Edema predictions are unreliable when T2-weighted imaging is missing, reflecting the modality-specific nature of edema contrast.

Summary of findings. Overall, these results demonstrate that robustness under incomplete multi-center CMR can be improved through carefully designed supervised strategies. In particular, domain-randomizing augmentation improves generalization across centers, while hinge-based training enhances stability under missing modalities. Together, these components establish a strong supervised baseline against which more complex strategies, such as semi-supervised learning, can be evaluated.

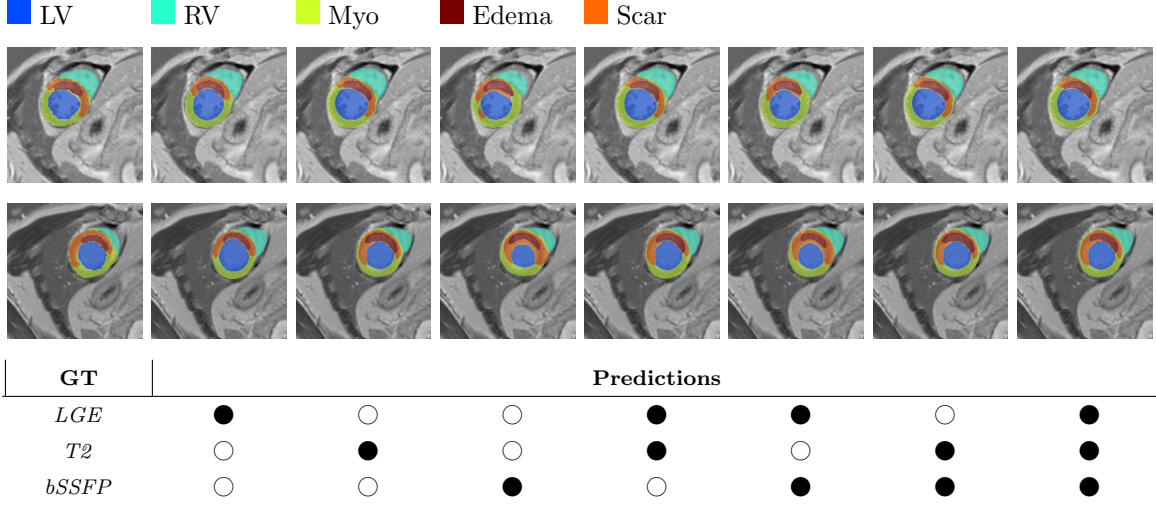


Figure 6.3: Qualitative comparison of predictions under different incomplete modality combinations. Colored regions represent *LV*, *RV*, *Myo*, *edema* and *scar*.

6.4.4 Semi-Supervised Learning under Incomplete Multi-Center Data

The use of semi-supervised learning is next investigated to determine whether partially labeled data from additional centers can improve segmentation performance under incomplete multi-center CMR conditions. This is particularly relevant for the MyoPS++ dataset, where annotations are unevenly distributed: edema labels are only available in centers B, C, and D, and some anatomical structures are missing in other centers. As a result, a large portion of the dataset cannot be directly used for fully supervised learning.

Two pseudo-labeling strategies are evaluated:

- **Mean Teacher:** consistency-based learning using noise-perturbed inputs [191];
- **FixMatch-style:** pseudo-labeling using weak-to-strong augmentation with confidence thresholding [192].

Table 6.3 summarizes the segmentation performance across validation groups.

Table 6.3: Semi-supervised performance using all centers (including partially labeled data).

Method	Dice Score				
	Scar				Edema
	Center A	Center E,F,G	Center B,C	Average	Center B,C
FixMatch + RandConv + MSE	0.2202	0.2406	0.5715	0.3441	0.5538
FixMatch + RandConv + Dice + CE	0.2399	0.2573	0.5983	0.3652	0.5597
Mean Teacher + RandConv + Dice + CE	0.1083	0.2424	0.6441	0.3316	0.6120
Mean Teacher + RandConv + MSE	0.0913	0.0303	0.6453	0.2556	0.6184

Overall behaviour of semi-supervised learning. Across all configurations, semi-supervised learning does not consistently outperform the strongest fully supervised baseline (Supervised + RandConv + Hinge). While certain configurations achieve competitive performance

in full-modality settings (centers B/C), these improvements do not generalize reliably to incomplete-modality or cross-center scenarios.

This suggests that, in this setting, the benefits of leveraging additional unlabeled or partially labeled data are limited by the quality and consistency of the pseudo-labels.

Interaction between pseudo-labeling and modality availability. A key observation is that the effectiveness of semi-supervised learning depends strongly on modality availability. Mean Teacher achieves the highest Dice score on centers B/C (0.6453), where all modalities are present and annotations are complete. However, its performance degrades sharply in LGE-only (center A) and LGE+bSSFP (centers E/F/G) settings, reaching as low as 0.0303 in some cases.

This indicates that pseudo-label quality is highly dependent on the richness of modality information. When modalities are missing, the teacher model is forced to generate pseudo-labels from incomplete evidence, which can lead to unreliable supervision and error propagation during training.

Interaction between pseudo-labeling and domain shift. In addition to modality incompleteness, domain variability further affects pseudo-label reliability. Since the teacher model is primarily trained on centers B/C, its predictions may not generalize well to unseen centers with different appearance characteristics. As a result, pseudo-labels generated for centers A and E/F/G may be systematically biased or noisy.

This interaction between domain shift and pseudo-labeling introduces a compounding effect: incorrect pseudo-labels can reinforce model errors, particularly in regions where pathology is subtle or poorly defined.

Comparison between Mean Teacher and FixMatch. FixMatch-style training demonstrates more stable performance across modality groups compared to Mean Teacher. In particular, the use of Dice + CE as the consistency objective yields better results than MSE-based consistency, suggesting that segmentation-aware losses are more appropriate for structured prediction tasks.

However, even with these improvements, FixMatch does not surpass the best supervised baseline. This indicates that confidence-based filtering alone is insufficient to address the challenges posed by missing modalities and domain shift.

Failure modes of semi-supervised learning. Several failure modes can be identified from Table 6.3:

- **Over-reliance on full-modality data:** Methods perform well on centers B/C but fail to generalize to incomplete-modality settings.
- **Error amplification:** Incorrect pseudo-labels can be reinforced during training, particularly in low-contrast or small regions such as scar.
- **Sensitivity to modality combinations:** Performance varies significantly depending on which modalities are available, indicating that pseudo-labeling does not adequately compensate for missing information.

Qualitative Analysis

Figure 6.4 provides qualitative comparisons between supervised and semi-supervised models. In some cases, semi-supervised learning improves spatial continuity and reduces fragmentation of predicted regions, particularly when RandConv is applied. This suggests that additional data can provide useful regularization.

However, failure cases remain evident. Scar regions may be underestimated or fragmented, and false positives may appear in anatomically implausible locations. These errors are consistent with the quantitative observations and highlight the limitations of pseudo-labeling under incomplete and heterogeneous data conditions.

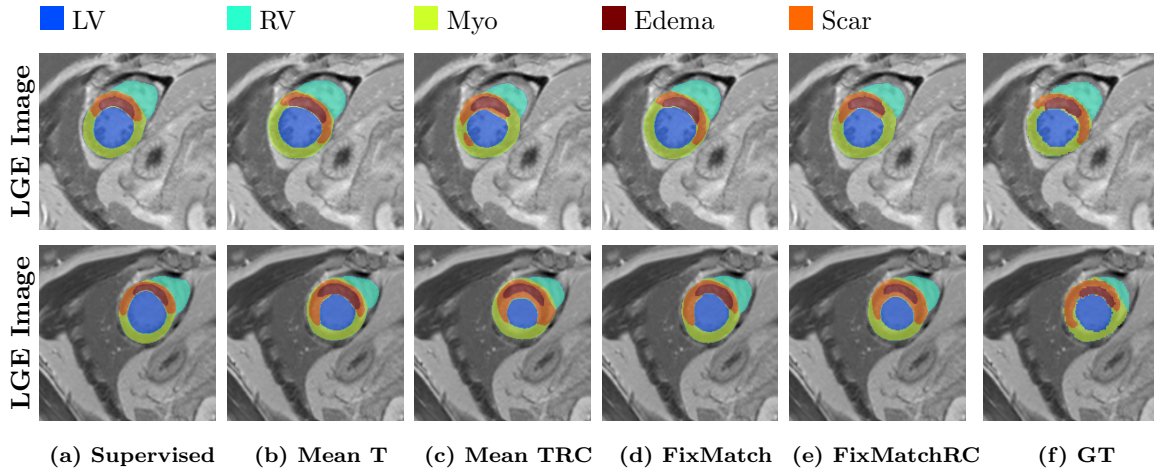


Figure 6.4: Qualitative comparison of supervised and semi-supervised training strategies. Mean T = mean teacher, RC = RandConv. Colored regions represent LV, RV, Myo, edema and scar.

Summary of findings. Overall, these results indicate that semi-supervised learning is not inherently robust under the combined challenges of missing modalities, incomplete annotations, and multi-center variability. While pseudo-labeling can be effective in controlled settings with consistent modality availability, its benefits diminish when applied to realistic clinical data with heterogeneous and partially observed inputs.

These findings suggest that improving robustness under incomplete multi-center CMR may require alternative strategies beyond standard pseudo-label-based semi-supervised learning.

6.4.5 Evaluation on the Challenge Test Set (Center D)

To assess real-world generalization, models were evaluated on the official MyoPS++ test set from center D (25 cases), which represents an unseen domain during training. This evaluation reflects a realistic deployment scenario in which models trained on heterogeneous multi-center data must generalize to a new clinical site with different acquisition characteristics.

Performance is reported using the official challenge metrics, including Dice similarity coefficient (DSC), sensitivity (SEN), precision (PRE), and Hausdorff distance (HD), for both

scar and combined edema+scar segmentation. Results are summarized in Table 6.4.

Table 6.4: Performance on the MyoPS++ test set (center D) as reported by the evaluation platform.

Training Configuration	Scar				Edema + Scar			
	DSC	SEN	PRE	HD	DSC	SEN	PRE	HD
All Centers + Modality Masking	0.547	0.563	0.589	101.0	0.591	0.538	0.720	130.2
All Centers (No Masking)	0.571	0.513	0.696	80.5	0.614	0.558	0.750	501.4
B/C Only + Modality Masking	0.582	0.528	0.696	72.4	0.615	0.541	0.780	125.0
Ensemble	0.598	0.557	0.693	54.7	0.622	0.568	0.751	124.6

Generalization to an unseen center. Since center D contains all three modalities (similar to centers B/C), models trained with strong supervision on B/C and explicit missing-modality simulation demonstrate good generalization. The *B/C Only + Modality Masking* configuration achieves the best single-model Dice for scar (0.582), indicating that high-quality supervision with complete modalities provides a strong foundation when the target domain also contains complete inputs.

However, performance does not fully match in-domain validation results, suggesting that domain shift between centers still affects generalization, even when modality configurations are consistent.

Effect of training with heterogeneous data. Models trained on all centers exhibit different behavior compared to those trained only on B/C. While they sometimes achieve higher sensitivity (e.g., detecting more pathological regions), this is often accompanied by reduced precision and increased Hausdorff distance. For example, the *All Centers (No Masking)* configuration achieves higher precision for edema+scar (0.750) but exhibits a substantially larger Hausdorff distance (501.4), indicating spatial inconsistencies.

This reflects a key trade-off: incorporating heterogeneous data can improve detection but may introduce noise due to domain variability and imperfect supervision.

Trade-off between sensitivity and precision. Across configurations, a clear trade-off between sensitivity and precision is observed. Models trained on more diverse data tend to be more sensitive, capturing a larger proportion of pathological pixels, but at the cost of increased false positives. In contrast, models trained on high-quality labeled data (B/C) produce more precise predictions but may miss subtle regions.

This trade-off is consistent with earlier observations in the semi-supervised setting, where pseudo-label noise and domain variability affect calibration and decision thresholds.

Effect of missing-modality-aware training. The inclusion of missing-modality simulation during training appears to contribute positively to generalization. Even though center D provides full modalities, models trained with simulated missing inputs maintain stable performance, suggesting that exposure to varying modality conditions improves robustness rather than harming performance in complete-modality settings.

Ensembling for robustness. No single training configuration dominates across all metrics, reflecting the complementary strengths of different strategies. The ensemble achieves the best overall performance, with the highest Dice scores for both scar (0.598) and edema+scar (0.622), and the lowest Hausdorff distance for scar (54.7).

This improvement suggests that ensemble agreement acts as an implicit regularizer, reducing extreme false positives and false negatives that arise from domain shift and variability in training regimes. In the context of multi-center CMR, where data distributions are inherently heterogeneous, combining diverse models appears to be an effective strategy for improving robustness.

Summary of findings. Overall, evaluation on the unseen center highlights several key insights:

- High-quality supervision with complete modalities provides strong baseline performance when modality configurations are matched,
- Training with heterogeneous data introduces a trade-off between sensitivity and precision,
- Missing-modality-aware training does not hinder, and may support, generalization to complete-modality settings,
- Ensembling improves robustness by combining complementary model behaviors.

These results reinforce the importance of balancing data diversity, supervision quality, and robustness-oriented training strategies when designing segmentation models for real-world multi-center clinical deployment.

6.4.6 Comparison with Existing Multi-Center Segmentation Approaches

To contextualize the findings of this chapter, a comparison was performed with recent state-of-the-art methods from the MyoPS++ challenge [199–208]. These methods represent a diverse range of strategies for addressing multi-center variability, including domain generalization, architectural improvements, and ensemble learning.

Unlike prior work, which typically addresses either domain shift *or* missing modalities, this chapter investigates the combined challenge of **multi-center variability, missing sequences, and incomplete annotations** within a unified framework.

Comparison of Problem Formulation

Table 6.5 highlights a key distinction: most existing methods assume *complete modality availability*, even when trained on multi-center data.

This comparison reveals that only a limited number of methods explicitly address missing modalities, and even fewer consider incomplete annotations. In contrast, the proposed framework jointly addresses all three challenges, reflecting a more realistic clinical deployment setting.

Table 6.5: Comparison of problem settings across existing methods.

Method	Missing Modalities	Multi-Center	Partial Labels
MixUp DG [199]	×	✓	×
Covariance Alignment [200]	×	✓	×
Two-stage Ensemble [201]	×	✓	×
ME-UNet [202]	×	✓	×
RGU-Mamba [203]	×	✓	×
Improved nnUNet [204]	×	✓	×
SimAM-UNet [206]	×	✓	×
Fusion Network [207]	×	✓	×
Partial Modality Ensemble [205]	✓	✓	×
GenSegNet [208]	✓	✓	✓
Ours	✓	✓	✓

Comparison of Methodological Strategies

Existing approaches can be broadly categorized into four groups:

Table 6.6: Comparison of methodological strategies.

Category	Methods
Domain Generalization	MixUp [199], Covariance Alignment [200]
Architecture Design	ME-UNet [202], RGU-Mamba [203], Improved nnUNet [204], SimAM-UNet [206]
Fusion-based Models	Progressive Fusion [207]
Ensemble-based	Two-stage Ensemble [201], Partial Modality Ensemble [205]
Semi-supervised / Generative	GenSegNet [208]
Ours	Unified: Fusion + DG + Missing Modality + Semi-supervised + Ensemble

A key observation is that prior methods typically focus on a *single axis of robustness*. For example, domain generalization methods improve cross-center performance but assume complete inputs, while partial-modality methods often rely on ensemble heuristics or normalization strategies without explicitly modeling missing data during training.

Role of ensembling in existing work. Several approaches [201, 205] employ ensembling as a primary mechanism for improving robustness. These methods typically combine predictions from multiple independently trained models, often with different architectures or normalization strategies. While effective, such approaches rely on *post-hoc aggregation* and do not fundamentally address missing-modality learning or domain shift during training.

Ensembling in the proposed framework. In contrast, ensembling in this work is not used as a standalone solution, but as a **complementary mechanism** built on top of models trained under different robustness-oriented regimes:

- Supervised models trained with modality masking,
- Domain-generalized models using RandConv,
- Models trained with hinge loss for stability under missing evidence.

This results in ensembles that capture **diverse failure modes** (e.g., sensitivity vs precision, domain-specific bias vs missing-modality robustness), rather than simply averaging similar models.

Key distinction. The key difference is that:

- Existing methods: *ensemble for performance gain*
- Proposed approach: *ensemble for robustness across heterogeneous conditions*

Unified framework. Overall, the proposed framework integrates:

- Modality-masked fusion for incomplete inputs,
- RandConv for domain generalization,
- Semi-supervised learning for incomplete annotations,
- Hinge loss for stability under missing evidence,
- Ensembling for combining complementary robustness properties.

This unified design enables the model to address multiple real-world challenges simultaneously, rather than optimizing for a single idealized setting.

Performance Comparison under Realistic Clinical Constraints

Table 6.7 summarizes performance on the MyoPS++ challenge benchmark (center D), explicitly accounting for differences in problem formulation and data completeness across methods.

Table 6.7: *Performance comparison with consideration of problem complexity.*

Method	Missing Modalities	Partial Labels	Dice (Scar)	Dice (Scar+Edema)
SimAM-UNet [206]	×	×	0.612	0.657
MixUp DG [199]	×	×	0.608	0.642
Two-stage Ensemble [201]	×	×	0.599	0.660
Improved nnUNet [204]	×	×	0.6089	0.5622
Covariance Alignment [200]	×	×	0.581	0.628
ME-UNet [202]	×	×	0.569	0.636
Fusion Network [207]	×	×	0.5486	0.6081
RGU-Mamba [203]	×	×	0.6575	0.6855
GenSegNet [208]	✓	✓	0.609	0.662
Ours (Ensemble)	✓	✓	0.599	0.631

Performance under ideal vs realistic settings. Several methods achieve higher Dice scores, particularly RGU-Mamba [203], which benefits from fully supervised training and complete modality availability. However, as shown in Table 6.7, the majority of these approaches operate under *idealized conditions*, where all imaging sequences and annotations are consistently available.

In contrast, the proposed framework is designed for a more challenging and clinically realistic setting, where both **modalities and annotations are incomplete**. Under these conditions, the model must simultaneously handle reduced input evidence, label sparsity, and cross-center variability, making the segmentation task inherently more difficult.

Interpreting performance in context. Despite these additional challenges, the proposed ensemble achieves competitive performance (Dice 0.599 for scar and 0.631 for scar+edema), remaining close to methods trained under more favorable conditions. This indicates that robustness-oriented training does not lead to catastrophic performance degradation, but instead enables stable predictions across heterogeneous scenarios.

Importantly, performance differences should be interpreted alongside problem assumptions. Methods trained with complete modalities may rely on modality-specific cues (e.g., scar visibility in LGE), which are not available in incomplete settings. As a result, their performance may degrade significantly when deployed in real-world clinical workflows where sequences are missing.

Comparison with GenSegNet. Among the compared methods, only GenSegNet [208] addresses a comparable level of problem complexity by explicitly considering both missing modalities and incomplete supervision. However, the underlying strategy differs fundamentally in how modality absence (i.e., modality amputation) is handled.

GenSegNet approaches modality amputation by *reconstructing* missing sequences through synthetic generation, followed by segmentation on the completed input set. While this can recover missing information, it effectively transforms the problem from *learning under incomplete evidence* to *learning from reconstructed inputs*. As a result, segmentation performance becomes dependent on the accuracy of the generated modalities. In cases where synthetic reconstruction is imperfect, particularly for subtle or heterogeneous pathology, errors may propagate and amplify in downstream predictions.

In contrast, the proposed framework treats modality amputation as a first-class learning condition rather than a deficiency to be corrected. Instead of attempting to reconstruct missing inputs, the model is explicitly trained to operate under incomplete modality configurations using modality-masked fusion. This enables the network to learn representations that are inherently robust to missing evidence. The hinge-based objective further enforces consistency between complete and incomplete modality settings, encouraging stable performance as modality information is progressively removed. In addition, RandConv-based augmentation improves robustness to cross-center variability without introducing synthetic inputs into the inference pipeline.

Table 6.8: Comparison of strategies for handling missing modality and incomplete supervision.

Method	Handling of Missing Modality	Requires Synthesis	Robustness Mechanism	Reliability Risk
GenSegNet [208]	Reconstruction via synthetic generation	✓	Pseudo-labeling	High (error propagation)
Ours	Direct learning under missing modalities	×	Hinge loss + RandConv	Low (no synthetic dependency)

This comparison highlights a key distinction: generation-based approaches aim to *restore*

missing inputs, whereas the proposed method is designed to *operate reliably despite their absence*. By explicitly modeling modality amputation during training, the proposed framework avoids reliance on potentially unreliable synthetic reconstructions and provides a more stable and clinically applicable solution.

This distinction is particularly important in clinical settings, where missing modalities are inherent to the acquisition process and cannot always be reliably reconstructed.

Critical Comparative Analysis

The expanded comparison across all methods reveals several important insights regarding the current state of multi-center myocardial pathology segmentation.

1. Performance is strongly tied to assumptions. The highest-performing method (RGU-Mamba [203]) achieves superior Dice scores; however, it assumes complete modality availability and fully supervised training. This simplifies the learning problem and does not reflect realistic clinical scenarios where missing sequences and incomplete annotations are common.

2. Most methods do not address missing modalities. Despite being evaluated on a dataset with inherent modality incompleteness, the majority of methods implicitly assume complete multi-sequence inputs. Only two approaches [205, 208] explicitly address missing modalities, highlighting a significant gap between benchmark evaluation and real-world deployment requirements.

3. Domain generalization alone is insufficient. Methods such as MixUp [199] and covariance alignment [200] improve robustness to cross-center variability but do not address missing data. This indicates that domain generalization and modality robustness are complementary challenges that must be addressed jointly.

4. Ensemble methods improve performance but not necessarily robustness. Several approaches [201, 205] rely heavily on ensembling to boost performance. While effective, these methods primarily combine predictions from similar models and do not explicitly enforce robustness during training. In contrast, the proposed framework generates diversity through different robustness-driven training strategies, resulting in ensembles that capture complementary behaviors.

5. Semi-supervised learning remains an open challenge. GenSegNet [208] explores pseudo-labeling to address incomplete annotations. However, results from this chapter indicate that pseudo-label quality degrades under domain shift and missing modalities, limiting its effectiveness. This highlights the need for more reliable semi-supervised strategies in heterogeneous medical datasets.

6. Unified robustness is largely unexplored. A key observation is that existing methods address individual challenges in isolation:

- Domain shift (e.g., MixUp, RGU-Mamba),

- Architecture improvements (e.g., ME-UNet, nnUNet variants),
- Missing modalities (e.g., partial ensemble),
- Semi-supervision (e.g., GenSegNet).

In contrast, the proposed framework integrates all these aspects within a single training pipeline, enabling robustness across multiple real-world constraints simultaneously.

Summary. Although the proposed method does not achieve the highest Dice score under ideal conditions, it operates under a more challenging and realistic problem setting. This makes the comparison more meaningful from a clinical deployment perspective, where robustness to missing data and domain variability is critical. Overall, this comparison highlights that:

- Existing methods focus on isolated challenges,
- Performance comparisons must consider problem assumptions,
- The proposed framework addresses a more complex and realistic setting,
- Robustness under incomplete and heterogeneous data remains an open research problem.

6.5 Conclusion

This chapter investigated myocardial pathology segmentation under *incomplete multi-center CMR*, focusing on a clinically realistic setting where imaging data are inherently imperfect. In particular, three key challenges were addressed: (i) **missing imaging sequences**, (ii) **incomplete annotations**, and (iii) **cross-center domain variability**. While prior work has often studied these factors in isolation, this chapter examined their *combined impact*, reflecting the conditions under which segmentation models must operate in practice.

To address modality amputation, a modality-aware framework based on M²FTrans [189] was adopted and extended with robustness-oriented training strategies. Modality-masked fusion enabled the model to learn directly from arbitrary subsets of available sequences, treating missing modalities as an inherent characteristic of the data rather than a deficiency to be corrected. A hinge-based objective was introduced to enforce consistency between complete and incomplete modality settings, encouraging stable performance as input evidence is reduced. In addition, RandConv [193] was employed to improve generalization across centers by exposing the model to diverse appearance variations during training.

A further objective of this chapter was to investigate the role of semi-supervised learning under incomplete annotation settings. While Mean Teacher [191] and FixMatch [192] enabled the use of partially labeled data, the results demonstrate that their effectiveness is limited in the presence of modality amputation and domain shift. In particular, pseudo-labeling approaches were found to be sensitive to variations in modality availability, leading to unstable behavior when input evidence is incomplete. This highlights an important insight: semi-supervised learning does not inherently confer robustness when both input data and supervision are heterogeneous.

In contrast, the proposed supervised framework with explicit robustness constraints demonstrated more consistent performance across modality configurations and centers. The combination of modality-masked learning, hinge-based regularization, and domain-randomizing augmentation resulted in improved stability under missing modalities and better generalization to unseen domains. Furthermore, ensemble models trained under complementary regimes provided additional gains, suggesting that diversity in training conditions can be leveraged to improve robustness in multi-center settings.

Comparison with existing approaches revealed that most state-of-the-art methods operate under simplified assumptions, such as complete modality availability and full supervision. While such methods can achieve higher Dice scores under ideal conditions, they do not explicitly address modality amputation as a core challenge. In contrast, the proposed framework maintains competitive performance while operating under a **strictly more realistic and constrained setting**. Notably, generation-based approaches such as GenSegNet [208] attempt to reconstruct missing modalities using synthetic data, whereas the proposed method directly models missing inputs, avoiding reliance on potentially unreliable reconstructions and reducing the risk of error propagation.

Overall, the findings of this chapter suggest a shift in perspective: rather than attempting to recover missing information, segmentation models should be designed to operate reliably in its absence. Achieving high accuracy under idealized conditions does not necessarily translate to robustness in clinical practice, where incomplete and heterogeneous data are the norm. By explicitly modeling modality amputation and systematically analyzing its interaction with supervision and domain shift, this work provides new insights into the limitations of existing approaches and highlights the importance of robustness-oriented design.

These findings underscore a broader transition from performance-centric evaluation toward reliability under real-world constraints, which is essential for translating segmentation models from research benchmarks to clinically deployable systems.

Chapter 7

Discussion and Conclusions

This thesis investigated automated myocardial scar segmentation from CMR imaging through a multimodal artificial intelligence framework designed to reflect the realities of clinical practice. Across four technical chapters, the work systematically addressed key limitations that constrain translational deployment of deep learning systems for scar analysis, including limited annotation availability, underrepresentation of rare infarct phenotypes, incomplete multimodal acquisition, and cross-center variability. Rather than optimizing segmentation performance under idealized research conditions, the thesis progressively relaxed simplifying assumptions and reconstructed the segmentation pipeline to operate under heterogeneous, data-constrained clinical environments.

This chapter synthesizes the principal findings, interprets their implications for the broader field of cardiac AI, critically evaluates remaining limitations, and outlines directions for future research.

7.1 Summary and Main Findings

The central novel contribution of this thesis lies not in a single architectural innovation, but in reframing myocardial scar segmentation as a multimodal, domain-aware inference problem rather than a narrowly defined image segmentation task. Each chapter addressed a distinct but interconnected constraint, forming a structured progression toward clinically robust automated scar analysis.

7.1.1 Chapter 3: Multi-Sequence CMR for Automated Scar Segmentation

Chapter 3 established the methodological foundation by examining how architectural design choices and multi-sequence CMR inputs influence segmentation performance. Using a standardized experimental framework built on a self-configuring convolutional architecture, this chapter compared end-to-end and cascaded learning paradigms across combinations of LGE, T2-weighted, and bSSFP sequences.

A central finding was that end-to-end learning consistently outperformed cascaded pipelines for scar segmentation, despite the latter’s popularity in prior literature. Cascaded approaches, which separate anatomical and pathological segmentation into sequential stages, were shown to propagate early-stage errors that degrade final scar segmentation. In contrast, joint opti-

mization of anatomy and pathology enabled the network to learn shared representations that better captured the spatial coupling between myocardial structure and infarct distribution. This result challenges the implicit assumption that anatomical decomposition necessarily improves pathological segmentation and suggests that tightly integrated representations are more suitable for complex disease phenotypes.

The chapter also demonstrated that complementary imaging contrasts provide measurable benefit. While LGE remains the gold standard for infarct visualization, T2-weighted and bSSFP sequences contributed structural context that improved boundary segmentation and reduced false positives, particularly in regions with ambiguous enhancement. Additionally, hybrid loss formulations combining region-based and imbalance-aware objectives proved critical for stabilizing training under severe class imbalance.

Conceptually, Chapter 3 showed that improving segmentation performance is not solely a matter of architectural depth or parameter count, but of aligning model inputs and optimization strategies with the structural realities of myocardial pathology.

7.1.2 Chapter 4: Clinically Guided Scar Synthesis

Chapter 4 addressed one of the most fundamental bottlenecks in cardiac AI: limited and imbalanced annotated datasets. Instead of relying on generic augmentation strategies, this chapter introduced a clinically guided scar synthesis framework that embedded anatomical priors, standardized segmental representations, and known infarct morphologies into the generative process.

A key insight was that synthetic data improves segmentation only when it respects physiological plausibility. Naïve synthetic augmentation risks introducing unrealistic features or reinforcing spurious correlations. By contrast, the proposed clinically grounded synthesis acted as structured regularization, expanding the pathological manifold in a controlled manner. Quantitative improvements were most pronounced for rare and underrepresented scar phenotypes, suggesting that diversity—not merely volume—is the limiting factor in segmentation generalization.

Beyond performance gains, this chapter reframed augmentation as a clinically meaningful design decision rather than a technical afterthought. It demonstrated that embedding domain knowledge into generative modeling bridges the gap between statistical realism and biological plausibility.

7.1.3 Chapter 5: Multimodal Structure–Function Integration

Chapter 5 expanded the segmentation paradigm beyond imaging by integrating electrophysiological information from 12-lead ECG signals. Recognizing that infarction manifests through both structural fibrosis and electrical remodeling, this chapter proposed a multimodal framework capable of capturing structure–function relationships.

The introduction of a Temporal-Aware Feature Fusion mechanism was central to this contribution. Because ECG and CMR are often acquired non-simultaneously in routine practice, naïve fusion risks physiologically inconsistent integration. By explicitly modeling acquisition timing, the framework ensured that ECG contributions were weighted according to temporal relevance. The resulting performance improvements confirmed that ECG provides complementary information for scar localization, particularly in cases with ambiguous

imaging contrast.

Importantly, integration of the AHA 17-segment model enhanced interpretability by aligning electrical abnormalities with anatomically defined myocardial regions. This chapter demonstrated that multimodal AI can approximate aspects of clinical reasoning, where imaging and electrophysiological signals are interpreted jointly rather than independently.

7.1.4 Chapter 6: Robustness Under Real-World Multi-Center Settings

Chapter 6 advanced the problem of robustness by addressing segmentation under realistic multi-center deployment conditions, characterized by heterogeneous imaging protocols, domain shift, and incomplete modality availability. Moving beyond controlled assumptions, this chapter explicitly formulated segmentation as a domain-resilient learning problem.

A unified framework integrating modality-aware learning, domain generalization strategies, and robustness-driven training objectives was developed to ensure consistent performance across diverse clinical settings. By jointly modeling variability in acquisition, contrast differences, and missing inputs, the approach enabled the model to generalize beyond the training distribution without relying on strict protocol alignment.

Experimental evaluation across multi-center datasets demonstrated improved stability and generalization under varying imaging conditions, including unseen distributions and incomplete inputs. The results highlight that robustness is not an emergent property of standard training pipelines, but must be explicitly enforced through architectural and learning design.

This chapter reinforces a central thesis of this work: clinically deployable segmentation systems must be resilient by design, maintaining reliable performance under real-world imperfections, rather than optimizing solely for idealized benchmark conditions.

7.1.5 Synthesis Across Chapters

Taken together, the findings of this thesis demonstrate that robust myocardial scar segmentation emerges from aligning algorithmic design with clinical reality. Each chapter addressed a distinct limitation—limited contrast, limited data, limited physiological observability, and limited robustness—yet collectively they form a unified progression toward clinically deployable multimodal AI systems.

The overarching lesson is that segmentation performance is not maximized through architectural complexity alone, but through principled integration of complementary modalities, domain knowledge, physiological signals, and robustness-aware training strategies. By reframing scar segmentation as a multimodal and domain-resilient inference problem, this thesis moves the field closer to scalable and trustworthy clinical deployment.

7.2 Limitations of This Work

While this thesis advances automated myocardial scar segmentation through multimodal integration, pathology-aware augmentation, and domain-resilient learning, important limitations remain. These limitations are not peripheral; they delineate the boundaries of the current contribution and define the research agenda required to translate multimodal cardiac AI into routine clinical practice.

7.2.1 Dependence on retrospective datasets and constrained pathological diversity

A primary limitation across all chapters is the reliance on retrospective, benchmark-oriented datasets such as MyoPS [12] and MyoPS++. These datasets provide valuable standardization and reproducibility, but they capture only a subset of real-world pathology. The majority of cases reflect post-infarction ischemic scarring, with limited representation of non-ischemic cardiomyopathies, infiltrative disease, mixed etiologies, or early remodeling stages.

Consequently, although the proposed frameworks demonstrate strong performance within these structured datasets, their generalization to broader and more heterogeneous disease spectra remains insufficiently validated. Real-world clinical populations include variable scar maturation stages, overlapping pathologies, and evolving remodeling patterns that may not conform to the distributions learned during training. This limitation is particularly relevant for the generative augmentation strategies in Chapter 4 and the multimodal integration mechanisms in Chapter 5, where the diversity of underlying pathology fundamentally shapes model behavior.

Furthermore, public datasets often exhibit partial harmonization of imaging protocols and annotation conventions. While this reduces experimental confounding, it likely underestimates the degree of heterogeneity encountered in routine multi-center deployment.

7.2.2 Annotation uncertainty and observer variability

The segmentation labels used throughout this thesis are derived from expert manual segmentations, which themselves contain variability and ambiguity. Myocardial scar boundaries are inherently uncertain in regions affected by partial volume effects, signal inhomogeneity, or borderline enhancement. As a result, the models trained in this work learn to approximate expert consensus rather than an objective pathological ground truth.

This limitation affects the interpretation of quantitative metrics such as Dice score and Hausdorff distance. High numerical agreement does not necessarily imply pathological correctness, particularly in regions where clinical interpretation is itself subjective. Future work would benefit from incorporating probabilistic annotation frameworks, consensus modeling across multiple experts, or uncertainty-aware supervision strategies that explicitly model ambiguity rather than collapsing it into binary labels.

7.2.3 Architectural complexity and computational burden

The multimodal and transformer-based architectures introduced in later chapters increase representational capacity but also computational demands. Multiple encoders, cross-modal attention modules, and deep supervision layers substantially increase memory usage and training time. While inference remains feasible for offline analysis, such architectural complexity may limit accessibility in resource-constrained environments or hinder integration into time-sensitive workflows.

In addition, increased architectural flexibility introduces additional hyperparameters and design decisions that may reduce reproducibility across institutions. Future research should explore model compression, parameter sharing strategies, lightweight fusion mechanisms, and knowledge distillation approaches to maintain robustness while reducing computational overhead.

7.2.4 Limited exploration of predictive uncertainty and reliability calibration

Although Chapter 6 explicitly addressed robustness to missing modalities and domain shift, systematic uncertainty quantification was not deeply integrated into the modeling framework. Preliminary observations suggest that models can exhibit overconfident predictions in incomplete-modality scenarios, particularly when pathological distributions deviate from training data.

In high-stakes clinical settings, calibrated uncertainty estimates are as important as raw segmentation accuracy. Without explicit reliability modeling—such as Bayesian approximations, ensemble uncertainty estimation, or conformal prediction—automated outputs may convey misleading confidence. Incorporating principled uncertainty quantification would strengthen trustworthiness and support human–AI collaboration in borderline cases.

7.2.5 Limited linkage to downstream clinical outcomes

While the proposed segmentation methods are motivated by applications such as ventricular tachycardia ablation planning and risk stratification, the thesis does not directly evaluate their impact on downstream clinical endpoints. Improvements in segmentation accuracy were not correlated with arrhythmia recurrence, procedural success, or long-term prognosis.

Thus, the clinical significance of performance gains—although plausible and theoretically motivated—remains inferential rather than empirically validated. Establishing such links would require prospective studies integrating segmentation outputs into treatment planning workflows and outcome prediction models.

7.3 Future Directions

The findings of this thesis open several avenues for further investigation, many of which extend beyond incremental model refinement toward broader questions of clinical integration, data representation, and physiological interpretability.

7.3.1 Expansion to broader cardiac pathologies and longitudinal data

The present work focuses primarily on post-infarction myocardial scarring, reflecting the composition of currently available datasets. Future research should extend these frameworks to encompass a wider spectrum of cardiomyopathies, including hypertrophic, dilated, and inflammatory conditions, where fibrosis may be diffuse, patchy, or mechanistically distinct. Such settings present fundamentally different segmentation challenges and would test the adaptability of pathology-aware learning strategies.

Incorporating longitudinal imaging represents an equally important direction. Modeling temporal evolution of scar and remodeling could enable prediction of disease progression, treatment response, or arrhythmogenic risk. Rather than static segmentation, this would shift the problem toward dynamic characterization of myocardial health over time.

7.3.2 Uncertainty-aware and probabilistic segmentation

Reliable clinical deployment requires not only accurate predictions but calibrated confidence estimates. Integrating uncertainty quantification—through Bayesian neural approximations, ensemble-based inference, or evidential learning—would allow models to express ambiguity in regions where pathology is intrinsically ill-defined. Such mechanisms could support decision support workflows by flagging uncertain cases for expert review, thereby promoting human–AI collaboration rather than full automation.

Uncertainty-aware training may also mitigate annotation noise by allowing soft probabilistic supervision instead of enforcing hard labels in diagnostically ambiguous regions.

7.3.3 Toward causal and physiologically grounded multimodal fusion

While Chapter 5 demonstrated that ECG–CMR fusion improves segmentation, the current approach remains primarily data-driven and correlational. Future work could explore more mechanistic integration of structure and function by incorporating electrophysiological simulations, conduction modeling, or biophysical constraints derived from cardiac physiology. Such hybrid models may improve interpretability and enable reasoning that more closely mirrors clinical understanding of arrhythmogenic substrates.

The integration of temporally richer data sources, including ambulatory ECG monitoring or stress imaging, could further strengthen multimodal characterization of myocardial remodeling.

7.3.4 Self-supervised learning and foundation model paradigms

The reliance on curated labeled datasets remains a central bottleneck. Emerging self-supervised and foundation model approaches offer an opportunity to learn generalized cardiac representations from large-scale unlabeled imaging and signal repositories. Pretraining strategies could enhance robustness to domain shift, reduce annotation requirements, and provide transferable feature spaces adaptable to new institutions or imaging protocols.

Such paradigms may ultimately enable cardiac-specific foundation models capable of supporting a wide range of downstream clinical tasks beyond scar segmentation alone.

7.3.5 Extension to 3D Scar Modeling and Volumetric Inference

The current work primarily considers slice-based or pseudo-3D representations of myocardial scar, reflecting common practices in both clinical annotation and deep learning frameworks. However, myocardial scar is inherently a three-dimensional structure, with spatial continuity across slices that carries important information about transmural, connectivity, and potential arrhythmogenic pathways.

Extending segmentation frameworks to fully three-dimensional inference represents an important direction for future research. Volumetric models could better capture spatial context, improve consistency across slices, and enable more accurate quantification of scar burden and distribution. In particular, 3D representations may enhance the identification of conducting channels and complex scar geometries that are not fully apparent in slice-wise analysis.

At the same time, 3D modeling introduces additional challenges, including increased computational cost, memory constraints, and sensitivity to inter-slice misalignment arising from acquisition variability. Addressing these challenges may require hybrid approaches that combine 2D efficiency with 3D contextual reasoning, as well as improved data preprocessing and alignment strategies.

Advancing toward robust 3D scar modeling would align automated segmentation more closely with clinical interpretation, where volumetric assessment and spatial continuity are central to understanding myocardial structure and arrhythmogenic substrate.

7.3.6 Prospective validation and workflow integration

A crucial next step is moving from retrospective evaluation to prospective clinical validation. Embedding automated segmentation into visualization platforms, electrophysiology planning systems, or reporting pipelines would allow assessment of usability, time efficiency, and influence on clinical decision-making. Studies comparing automated and manual segmentation in procedural planning or risk stratification contexts are necessary to establish tangible clinical benefit.

This transition from algorithmic validation to workflow integration represents a key milestone on the path toward adoption.

7.3.7 Standardization, interpretability, and regulatory readiness

Finally, advancing toward deployment requires attention to interpretability, reproducibility, and regulatory considerations. Aligning model outputs with established clinical frameworks, such as the AHA 17-segment model, can enhance transparency and facilitate clinician trust. Standardized evaluation across diverse populations and institutions will be essential to demonstrate fairness and generalizability, both of which are increasingly central to regulatory approval pathways for medical AI systems.

7.4 Conclusions

This thesis addressed a fundamental challenge in contemporary cardiac imaging: enabling automated, robust, and clinically meaningful segmentation of myocardial scars under realistic data constraints. Rather than optimizing models within controlled benchmark settings, the work reframed the problem to reflect the variability, incompleteness, and multimodal nature of clinical data.

The central finding is that accurate and reliable scar segmentation cannot be achieved through architectural refinement alone. Instead, performance depends on alignment between algorithmic design and the underlying clinical and physiological characteristics of the problem. Across a series of methodological developments, it is shown that incorporating complementary imaging contrasts, embedding clinically grounded priors into data augmentation, integrating electrophysiological signals, and explicitly accounting for domain variability each contribute to improving robustness and generalization.

A key contribution of this work is the demonstration that multimodal integration must be structured and physiologically meaningful. Simply combining modalities is insufficient and can be detrimental if relationships between structure and function are not appropriately

modeled. When guided by temporal and anatomical context, however, multimodal learning enables representations that more closely reflect clinical reasoning and improve segmentation performance.

Equally, the results highlight that robustness to real-world constraints, including missing modalities, heterogeneous acquisition protocols, and limited annotated data, is essential for translation. Methods that perform well under idealized conditions may fail in clinical settings unless variability is treated as a first-class design consideration. Approaches such as clinically guided augmentation, modality-aware learning, and domain-diversifying strategies demonstrate that this gap can be narrowed.

Taken together, this thesis advances myocardial scar segmentation from a constrained image analysis task toward a clinically grounded, multimodal, and domain-resilient learning problem. It shows that effective automation requires not only accuracy, but also reliability, interpretability, and adaptability to imperfect data.

While challenges remain, particularly in prospective validation, multimodal alignment, and scalable deployment, the frameworks developed here provide a foundation for future research and clinical translation. As cardiac imaging increasingly integrates with artificial intelligence, such approaches will be critical for enabling robust, trustworthy, and clinically useful systems capable of operating across diverse patient populations and healthcare environments.

Ultimately, this work argues that clinically meaningful AI must be designed not for idealized datasets, but for the variability, uncertainty, and imperfection inherent in real-world medicine.

Table 7.1: *Methodological progression of the thesis, showing how each chapter addresses a specific limitation of prior automated scar segmentation approaches and contributes toward clinically deployable multimodal AI.*

Chapter	Primary Challenge	Methodological Strategy	Conceptual Advancement
Chapter 3	Limited structural context from single-sequence imaging	Joint end-to-end segmentation using multi-sequence CMR (LGE, T2, bSSFP) within a self-configuring nnU-Net framework	Demonstrated that tightly coupled anatomical–pathological learning outperforms cascaded pipelines, emphasizing holistic feature representation
Chapter 4	Scarce annotations and underrepresentation of rare infarct morphologies	Clinically guided diffusion-based scar synthesis (SMILE) embedding anatomical and pathological priors	Reframed data augmentation as pathology-aware regularization, enabling controlled expansion of clinically plausible data distributions
Chapter 5	Lack of physiological information in image-only segmentation	Multimodal ECG–CMR integration via Temporal-Aware Feature Fusion (TAFF) aligned with the AHA 17-segment model	Showed that incorporating electrophysiological signals enables structure–function inference consistent with clinical reasoning
Chapter 6	Performance degradation under missing modalities and multi-center variability	Modality-masked fusion transformer (M ² FTrans) with semi-supervised learning, hinge-based robustness objectives, and domain-diversifying augmentation	Introduced reliability-aware learning capable of maintaining segmentation performance across incomplete and heterogeneous clinical datasets

Table 7.2: *Progressive increase in data realism and supervision complexity across thesis chapters, moving from controlled benchmark settings toward heterogeneous clinical scenarios.*

Chapter	Input Modalities	Data Source	Learning Paradigm	Center Variability	Handling of Clinical Imperfections
Chapter 3	Multi-sequence CMR (LGE, T2, bSSFP)	Public benchmark dataset (MyoPS)	Fully supervised learning	Single-center-like	Assumes complete, well-aligned imaging inputs
Chapter 4	LGE + synthetic pathology	Public dataset augmented with clinically guided synthesis	Supervised learning with generative augmentation	Single-center-like	Addresses limited pathological diversity via realistic synthetic scars
Chapter 5	LGE-CMR + 12-lead ECG	Private clinical cohort	Multimodal supervised learning	Single-center clinical	Models temporal misalignment between modalities
Chapter 6	Multi-sequence CMR (variable availability)	Public multi-center dataset (MyoPS++)	Semi-supervised and domain-resilient learning	Multi-center (7 institutions)	Explicitly models missing modalities, protocol variability, and domain shift

Table 7.3: *Evolution of learning paradigms across the thesis, highlighting the transition from conventional supervised segmentation toward multimodal and robustness-oriented learning strategies.*

Chapter	Learning Paradigm	Key Techniques	Purpose in the Overall Framework
Chapter 3	Supervised representation learning	Hybrid objective functions (Dice, Focal-Tversky, boundary-aware losses) within end-to-end nnU-Net training	Stabilize optimization under severe class imbalance and enable accurate learning of small, heterogeneous scar regions
Chapter 4	Generative-informed supervised learning	Clinically guided diffusion-based scar synthesis integrated with segmentation supervision	Expand pathological diversity while constraining generated data to anatomically and physiologically plausible distributions
Chapter 5	Multimodal supervised learning	Attention-based ECG–CMR fusion with temporal conditioning and anatomical priors	Integrate complementary physiological signals to model structure–function relationships beyond image-only inference
Chapter 6	Semi-supervised and robustness-aware learning	Mean Teacher consistency training, FixMatch-style pseudo-labeling, hinge-based robustness objectives, and RandConv domain augmentation	Leverage partially labeled multi-center data and improve stability under missing modalities and distribution shift

Table 7.4: *Progression toward clinical translation across thesis chapters, illustrating how methodological developments improve interpretability, robustness, and practical applicability in real-world cardiac imaging workflows.*

Chapter	Level of Automation	Interpretability Mechanisms	Operational Robustness	Implications for Clinical Use
Chapter 3	Fully automated segmentation pipeline	Anatomically consistent multi-sequence feature learning	Limited tolerance to domain variability	Provides baseline automation for myocardial scar segmentation, reducing manual workload but requiring controlled imaging conditions
Chapter 4	Automated segmentation with clinically guided data enrichment	Pathology-aware synthesis grounded in anatomical priors	Improved generalization to rare scar patterns	Enhances recognition of under-represented infarct morphologies, supporting more comprehensive disease characterization
Chapter 5	Automated multimodal inference	ECG-CMR fusion aligned with AHA 17-segment model for physiological interpretability	Moderate resilience to modality timing differences	Enables patient-specific structure-function mapping consistent with electrophysiological assessment workflows
Chapter 6	Automation designed for heterogeneous environments	Explicit modeling of missing modalities and cross-center variability	High robustness to incomplete data and domain shift	Moves toward deployment-ready systems capable of operating across multi-center datasets and variable acquisition protocols

Bibliography

- [1] A. G. Hays, M. Schar, and S. Kelle, “Clinical applications for cardiovascular magnetic resonance imaging at 3 tesla,” *Current cardiology reviews*, vol. 5, no. 3, pp. 237–242, 2009.
- [2] E. R. Jenista, D. C. Wendell, C. F. Azevedo, I. Klem, R. M. Judd, R. J. Kim, and H. W. Kim, “Revisiting how we perform late gadolinium enhancement cmr: insights gleaned over 25 years of clinical practice,” *Journal of Cardiovascular Magnetic Resonance*, vol. 25, no. 1, p. 18, 2023.
- [3] I. Klem, J. W. Weinsaft, T. D. Bahnson, D. Hegland, H. W. Kim, B. Hayes, M. A. Parker, R. M. Judd, and R. J. Kim, “Assessment of myocardial scarring improves risk stratification in patients evaluated for cardiac defibrillator implantation,” *Journal of the American College of Cardiology*, vol. 60, no. 5, pp. 408–420, 2012.
- [4] S. Nakatani, “Left ventricular rotation and twist: Why should we learn?” *Journal of Cardiovascular Ultrasound*, vol. 19, no. 1, p. 1, 2011.
- [5] S. Yoshihara, “Acute coronary syndrome on non-electrocardiogram-gated contrast-enhanced computed tomography,” *World Journal of Radiology*, vol. 14, no. 2, p. 30, 2022.
- [6] F. Ramzan, C. C. Chen, and R. H. Clayton, “Improving myocardial scar segmentation with end-to-end and cascaded cnn using hybrid loss and multi-modality cmr imaging,” 2024.
- [7] F. Ramzan, Y. Kiberu, N. Jathanna, S. Jamil-Copley, R. H. Clayton, and C. Chen, “Claim: Clinically-guided lge augmentation for realistic and diverse myocardial scar synthesis and segmentation,” in *International Conference on AI in Healthcare*. Springer, 2025, pp. 279–292.
- [8] H. Zhang, Y. Liu, J. Yang, S. Wan, X. Wang, W. Peng, and P. Fua, “Lefusion: Controllable pathology synthesis via lesion-focused diffusion models,” *arXiv preprint arXiv:2403.14066*, 2024.
- [9] F. Ramzan, C. Chen, R. H. Clayton *et al.*, “Seeing beyond the image: Ecg and anatomical knowledge-guided myocardial scar segmentation from late gadolinium-enhanced images,” in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2026, accepted. Preprint available at arXiv:2511.14702.

- [10] A. Lalande, Z. Chen, T. Decourselle, A. Qayyum, T. Pommier, L. Lorgis, E. de La Rosa, A. Cochet, Y. Cottin, D. Gin hac *et al.*, “Emidec: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac mri,” *Data*, vol. 5, no. 4, p. 89, 2020.
- [11] X. Zhuang, “Multivariate mixture model for myocardial segmentation combining multi-source images,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 12, pp. 2933–2946, 2019.
- [12] L. Li, F. Wu, S. Wang, X. Luo, C. Martín-Isla, S. Zhai, J. Zhang, Y. Liu, Z. Zhang, M. J. Ankenbrand *et al.*, “Myops: A benchmark of myocardial pathology segmentation combining three-sequence cardiac magnetic resonance images,” *Medical Image Analysis*, vol. 87, p. 102808, 2023.
- [13] J. Qiu, L. Li, S. Wang, K. Zhang, Y. Chen, S. Yang, and X. Zhuang, “Myops-net: Myocardial pathology segmentation with flexible combination of multi-sequence cmr images,” *Medical image analysis*, vol. 84, p. 102694, 2023.
- [14] W. Ding, L. Li, J. Qiu, S. Wang, L. Huang, Y. Chen, S. Yang, and X. Zhuang, “Aligning multi-sequence cmr towards fully automated myocardial pathology segmentation,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 12, pp. 3474–3486, 2023.
- [15] M. A. Pfeffer and E. Braunwald, “Ventricular remodeling after myocardial infarction. experimental observations and clinical implications.” *Circulation*, vol. 81, no. 4, pp. 1161–1172, 1990.
- [16] G. W. Stone, H. P. Selker, H. Thiele, M. R. Patel, J. E. Udelson, E. M. Ohman, A. Maehara, I. Eitel, C. B. Granger, P. L. Jenkins *et al.*, “Relationship between infarct size and outcomes following primary pci: patient-level analysis from 10 randomized trials,” *Journal of the American College of Cardiology*, vol. 67, no. 14, pp. 1674–1683, 2016.
- [17] M. Bernard, A. Jacquier, and F. Kober, “Cardiovascular magnetic resonance in ischemic heart disease,” *Future Cardiology*, vol. 10, no. 4, pp. 487–496, 2014.
- [18] R. J. Kim, E. Wu, A. Rafael, E.-L. Chen, M. A. Parker, O. Simonetti, F. J. Klocke, R. O. Bonow, and R. M. Judd, “The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction,” *New England Journal of Medicine*, vol. 343, no. 20, pp. 1445–1453, 2000.
- [19] A. S. Flett, J. Hasleton, C. Cook, D. Hausenloy, G. Quarta, C. Ariti, V. Muthurangu, and J. C. Moon, “Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance,” *JACC: cardiovascular imaging*, vol. 4, no. 2, pp. 150–156, 2011.
- [20] K. M. Choi, R. J. Kim, G. Gubernikoff, J. D. Vargas, M. Parker, and R. M. Judd, “Transmural extent of acute myocardial infarction predicts long-term improvement in contractile function,” *Circulation*, vol. 104, no. 10, pp. 1101–1107, 2001.

- [21] W. J. Rogers Jr, C. M. Kramer, G. Geskin, Y.-L. Hu, T. M. Theobald, D. A. Vido, S. Petruolo, and N. Reichek, "Early contrast-enhanced mri predicts late functional recovery after reperfused myocardial infarction," *Circulation*, vol. 99, no. 6, pp. 744–750, 1999.
- [22] L. Jada, R. J. Holtackers, B. Martens, H. M. Nies, C. M. Van De Heyning, R. M. Botnar, J. E. Wildberger, T. F. Ismail, R. Razavi, and A. Chiribiri, "Quantification of myocardial scar of different etiology using dark-and bright-blood late gadolinium enhancement cardiovascular magnetic resonance," *Scientific Reports*, vol. 14, no. 1, p. 5395, 2024.
- [23] Y. Mikami, L. Kolman, S. X. Joncas, J. Stirrat, D. Scholl, M. Rajchl, C. P. Lydell, S. G. Weeks, A. G. Howarth, and J. A. White, "Accuracy and reproducibility of semi-automated late gadolinium enhancement quantification techniques in patients with hypertrophic cardiomyopathy," *Journal of Cardiovascular Magnetic Resonance*, vol. 16, no. 1, p. 85, 2014.
- [24] S.-A. Lee, Y. E. Yoon, J.-E. Kim, J. J. Park, I.-Y. Oh, C.-H. Yoon, J.-W. Suh, J.-S. Kim, E. J. Chun, Y.-S. Cho *et al.*, "Long-term prognostic value of late gadolinium-enhanced magnetic resonance imaging in patients with and without left ventricular dysfunction undergoing coronary artery bypass grafting," *The American Journal of Cardiology*, vol. 118, no. 11, pp. 1647–1654, 2016.
- [25] R. Kittayaphong, T. Boonyasirinant, V. Chaithiraphan, A. Maneesai, P. Saiviroonporn, S. Nakyen, P. Thanapiboonpol, A. Yindeengam, and S. Udompanturak, "Prognostic value of late gadolinium enhancement in hypertensive patients with known or suspected coronary artery disease," *The international journal of cardiovascular imaging*, vol. 26, no. Suppl 1, pp. 123–131, 2010.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [27] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [28] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?" *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.
- [29] X. Zhuang and L. Li, *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*. Springer Nature, 2020, vol. 12554.
- [30] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24–33, 2019.

- [31] J. D. Schuijf, T. A. Kaandorp, H. J. Lamb, R. J. van der Geest, E. P. Viergever, E. E. van der Wall, A. de Roos, and J. J. Bax, "Quantification of myocardial infarct size and transmural by contrast-enhanced magnetic resonance imaging in men," *The American journal of cardiology*, vol. 94, no. 3, pp. 284–288, 2004.
- [32] S. G. Elis, C. I. Henschke, T. Sandor, J. Wynne, and R. A. Kloner, "Relation between the transmural extent of acute myocardial infarction and associated myocardial contractility two weeks after infarction," *The American journal of cardiology*, vol. 55, no. 11, pp. 1412–1416, 1985.
- [33] P. Kellman, H. Xue, L. J. Olivieri, R. R. Cross, E. K. Grant, M. Fontana, M. Ugander, J. C. Moon, and M. S. Hansen, "Dark blood late enhancement imaging," *Journal of Cardiovascular Magnetic Resonance*, vol. 18, no. 1, p. 77, 2016.
- [34] A. H. Curiale, F. D. Colavecchia, and G. Mato, "Automatic quantification of the lv function and mass: A deep learning approach for cardiovascular mri," *Computer methods and programs in biomedicine*, vol. 169, pp. 37–50, 2019.
- [35] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu *et al.*, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.
- [36] Y. Yang, H. Zhang, J. W. Gichoya, D. Katabi, and M. Ghassemi, "The limits of fair medical imaging ai in real-world generalization," *Nature medicine*, vol. 30, no. 10, pp. 2838–2848, 2024.
- [37] L. Goetz, N. Seedat, R. Vandersluis, and M. van der Schaar, "Generalization—a key challenge for responsible ai in patient-facing clinical applications," *NPJ Digital Medicine*, vol. 7, no. 1, p. 126, 2024.
- [38] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [39] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [40] L. Heiliger, A. Sekuboyina, B. Menze, J. Egger, and J. Kleesiek, "Beyond medical imaging-a review of multimodal deep learning in radiology," *Authorea Preprints*, 2022.
- [41] I. U. Haq, M. Mhamed, M. Al-Harbi, H. Osman, Z. Y. Hamd, and Z. Liu, "Advancements in medical radiology through multimodal machine learning: a comprehensive overview," *Bioengineering*, vol. 12, no. 5, p. 477, 2025.
- [42] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vol. 3, p. 100004, 2019.

- [43] M. Schwab, M. Paminger, C. Kremser, M. Haltmeier, and A. Mayr, “Deep learning pipeline for fully automated myocardial infarct segmentation from clinical cardiac mr scans,” *Radiology Advances*, vol. 2, no. 4, p. umaf023, 2025.
- [44] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [45] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [46] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [47] D. G. Strauss, R. H. Selvester, J. A. Lima, H. Arheden, J. M. Miller, G. Gerstenblith, E. Marban, R. G. Weiss, G. F. Tomaselli, G. S. Wagner *et al.*, “Ecg quantification of myocardial scar in cardiomyopathy patients with or without conduction defects: correlation with cardiac magnetic resonance and arrhythmogenesis,” *Circulation: Arrhythmia and Electrophysiology*, vol. 1, no. 5, pp. 327–336, 2008.
- [48] P. S. Cuculich, J. Zhang, Y. Wang, K. A. Desouza, R. Vijayakumar, P. K. Woodard, and Y. Rudy, “The electrophysiological cardiac ventricular substrate in patients after myocardial infarction: noninvasive characterization with electrocardiographic imaging,” *Journal of the American College of Cardiology*, vol. 58, no. 18, pp. 1893–1902, 2011.
- [49] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 2, pp. 423–443, 2018.
- [50] A. G. Khoei, Y. Yu, and R. Feldt, “Domain generalization through meta-learning: a survey,” *Artificial Intelligence Review*, vol. 57, no. 10, p. 285, 2024.
- [51] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martin-Isla, A. Sojoudi, P. M. Full, K. Maier-Hein, Y. Zhang, Z. He, J. Ma *et al.*, “Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3543–3554, 2021.
- [52] D. D. Streeter Jr, H. M. Spotnitz, D. P. Patel, J. Ross Jr, and E. H. Sonnenblick, “Fiber orientation in the canine left ventricle during diastole and systole,” *Circulation research*, vol. 24, no. 3, pp. 339–347, 1969.
- [53] R. Greenbaum, S. Y. Ho, D. Gibson, A. Becker, and R. Anderson, “Left ventricular fibre architecture in man.” *Heart*, vol. 45, no. 3, pp. 248–263, 1981.
- [54] K. A. Reimer and R. B. Jennings, “The wavefront phenomenon of ischemic cell death,” *Circulation*, vol. 56, no. 5, pp. 786–794, 1977.
- [55] M. Cerqueira, N. Weissman, V. Dilsizian, and A. Jacobs, “Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the cardiac imaging committee of the council on clinical

- cardiology of the american heart association,” *Circulation*, vol. 105, no. 4, pp. 539–542, 2002.
- [56] J. Jalife, “Ventricular fibrillation: mechanisms of initiation and maintenance,” *Annual Review of Physiology*, vol. 62, pp. 25–50, 2000.
 - [57] J. M. e. a. de Bakker, “Reentry as a cause of ventricular tachycardia in patients with chronic ischemic heart disease,” *Circulation*, vol. 88, pp. 915–926, 1993.
 - [58] D. P. Zipes and H. J. J. Wellens, “Sudden cardiac death,” *Circulation*, vol. 111, no. 4, pp. e30–e31, 2005.
 - [59] R. J. Myerburg and M. J. Junttila, “Sudden cardiac death: epidemiology, transient risk, and intervention assessment,” *Circulation*, vol. 129, no. 4, pp. e18–e23, 2014.
 - [60] M. E. Josephson, *Clinical Cardiac Electrophysiology: Techniques and Interpretations*, 4th ed. Lippincott Williams & Wilkins, 2008.
 - [61] J. J. Fenoglio *et al.*, “Myocardial infarction: a morphologic study of the arrhythmogenic substrate,” *The American Journal of Cardiology*, vol. 37, no. 2, pp. 214–220, 1976.
 - [62] A. Schmidt *et al.*, “Infarct tissue heterogeneity by magnetic resonance imaging identifies enhanced cardiac arrhythmia susceptibility in patients with left ventricular dysfunction,” *Circulation*, vol. 115, no. 15, pp. 2006–2014, 2007.
 - [63] K.-H. Wu *et al.*, “Late gadolinium enhancement by cardiovascular magnetic resonance heralds an adverse prognosis in nonischemic cardiomyopathy,” *Journal of the American College of Cardiology*, vol. 51, no. 25, pp. 2414–2421, 2008.
 - [64] M. Disertori *et al.*, “Myocardial fibrosis assessment by lge-cmr and arrhythmic risk in cardiomyopathy: a systematic review and meta-analysis,” *Journal of the American College of Cardiology*, vol. 69, no. 20, pp. 2588–2598, 2017.
 - [65] A. S. Jadidi *et al.*, “Substrate mapping and ablation of ventricular tachycardia: role of imaging,” *Heart Rhythm*, vol. 10, no. 4, pp. e1–e10, 2013.
 - [66] N. G. Frangogiannis, “Inflammation in cardiac injury and repair,” *Current Opinion in Cardiology*, vol. 30, no. 3, pp. 240–245, 2015.
 - [67] J. J. e. a. Goldberger, “Risk stratification for sudden cardiac death: a plan for the future,” *Circulation*, vol. 118, no. 14, pp. 1497–1518, 2008.
 - [68] J. Riesmeier *et al.*, “Cmr integration into ventricular tachycardia ablation planning and guidance,” *Europace*, vol. 21, no. 3, pp. e1–e12, 2019.
 - [69] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley-Liss, 1999.
 - [70] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences*. Elsevier, 2004.

- [71] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. IEEE Press, 2000.
- [72] F. e. a. Grothues, “Interstudy reproducibility of right ventricular volumes, function, and mass with cardiovascular magnetic resonance,” *American Heart Journal*, vol. 147, no. 2, pp. 218–223, 2002.
- [73] R. J. e. a. Kim, “Relationship of mri delayed contrast enhancement to irreversible injury, infarct age, and contractile function,” *Circulation*, vol. 100, no. 19, pp. 1992–2002, 1999.
- [74] P. Kellman and A. E. Arai, “Dark-blood late enhancement imaging,” *Journal of Cardiovascular Magnetic Resonance*, vol. 14, pp. 1–14, 2012.
- [75] M. e. a. Markl, “Cardiovascular magnetic resonance imaging: basic principles and clinical applications,” *Journal of Magnetic Resonance Imaging*, vol. 36, no. 5, pp. 1015–1036, 2012.
- [76] H. e. a. Abdel-Aty, “T2-weighted cardiovascular magnetic resonance in acute cardiac disease,” *Journal of the American College of Cardiology*, vol. 59, no. 13, pp. 1188–1197, 2012.
- [77] G. S. Wagner, *Marriott’s Practical Electrocardiography*, 11th ed. Lippincott Williams & Wilkins, 2008.
- [78] A. Bayés de Luna, *Clinical Electrocardiography: A Textbook*. Wiley-Blackwell, 2012.
- [79] X. e. a. Zhuang, “Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge,” in *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, 2019, pp. 2160–2172.
- [80] C. e. a. Chen, “Myops++: A multi-center cardiac mri dataset for myocardial pathology segmentation,” *Medical Image Analysis*, vol. 73, p. 102154, 2021.
- [81] L. C. Amado, B. L. Gerber, S. N. Gupta, D. W. Rettmann, G. Szarf, R. Schock, K. Nasir, D. L. Kraitchman, and J. A. Lima, “Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model,” *Journal of the American College of Cardiology*, vol. 44, no. 12, pp. 2383–2389, 2004.
- [82] J. e. a. Lötjönen, “Fast and robust multi-atlas segmentation of cardiac mri,” *Medical Image Analysis*, vol. 8, no. 3, pp. 245–258, 2004.
- [83] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*, 2015, pp. 234–241.
- [84] F. e. a. Isensee, “nnu-net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, vol. 18, pp. 203–211, 2021.
- [85] M. Raghu and E. Schmidt, “Generalization in medical imaging ai,” *Nature Machine Intelligence*, vol. 3, pp. 389–396, 2021.

- [86] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: a survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019.
- [87] S. S. M. e. a. Salehi, “Tversky loss function for image segmentation using 3d fully convolutional deep networks,” in *International Workshop on Machine Learning in Medical Imaging*, 2017.
- [88] W. G. Stevenson, “Ventricular scars and ventricular tachycardia,” *Transactions of the American Clinical and Climatological Association*, vol. 120, p. 403, 2009.
- [89] H. Tanner, G. Hindricks, M. Volkmer, S. Furniss, V. Kühlkamp, D. Lacroix, C. De Chillou, J. Almendral, D. Caponi, K.-H. KUCK *et al.*, “Catheter ablation of recurrent scar-related ventricular tachycardia using electroanatomical mapping and irrigated ablation technology: results of the prospective multicenter euro-vt-study,” *Journal of cardiovascular electrophysiology*, vol. 21, no. 1, pp. 47–53, 2010.
- [90] L. J. Juan, A. M. Crean, and B. J. Wintersperger, “Late gadolinium enhancement imaging in assessment of myocardial viability: techniques and clinical applications,” *Radiologic Clinics*, vol. 53, no. 2, pp. 397–411, 2015.
- [91] Q. Tao, S. R. Piers, H. J. Lamb, and R. J. van der Geest, “Automated left ventricle segmentation in late gadolinium-enhanced mri for objective myocardial scar assessment,” *Journal of Magnetic Resonance Imaging*, vol. 42, no. 2, pp. 390–399, 2015.
- [92] E. Heiberg, J. Sjögren, M. Ugander, M. Carlsson, H. Engblom, and H. Arheden, “Design and validation of segment-freely available software for cardiovascular image analysis,” *BMC medical imaging*, vol. 10, no. 1, p. 1, 2010.
- [93] Y. Li, C. P. Ho, M. Toulemonde, N. Chahal, R. Senior, and M.-X. Tang, “Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model,” *IEEE transactions on medical imaging*, vol. 37, no. 5, pp. 1081–1091, 2017.
- [94] D. Conforti and R. Guido, “Kernel-based support vector machine classifiers for early detection of myocardial infarction,” *Optimization Methods and Software*, vol. 20, no. 2-3, pp. 401–413, 2005.
- [95] Y. Wu, Z. Tang, B. Li, D. Firmin, and G. Yang, “Recent advances in fibrosis and scar segmentation from cardiac mri: a state-of-the-art review and future perspectives,” *Frontiers in Physiology*, vol. 12, p. 709230, 2021.
- [96] X. Zhuang and J. Shen, “Multi-scale patch and multi-modality atlases for whole heart segmentation of mri,” *Medical image analysis*, vol. 31, pp. 77–87, 2016.
- [97] R. Y. Kwong, M. Jerosch-Herold, and B. Heydari, *Cardiovascular magnetic resonance imaging*. Springer, 2019.
- [98] X. Zhuang, “Multivariate mixture model for myocardial segmentation combining multi-source images,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 12, pp. 2933–2946, 2018.

- [99] Y. Liu, M. Zhang, Q. Zhan, D. Gu, and G. Liu, “Two-stage method for segmentation of the myocardial scars and edema on multi-sequence cardiac magnetic resonance,” in *Myocardial pathology segmentation combining multi-sequence CMR challenge*. Springer, 2020, pp. 26–36.
- [100] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *International Workshop on Deep Learning in Medical Image Analysis*. Springer, 2017, pp. 240–248.
- [101] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [102] N. Abraham and N. M. Khan, “A novel focal tversky loss function with improved attention u-net for lesion segmentation,” in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*. IEEE, 2019, pp. 683–687.
- [103] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, “Loss odyssey in medical image segmentation,” *Medical Image Analysis*, vol. 71, p. 102035, 2021.
- [104] Y. Kiberu, N. Jathanna, N. Narayanan, A. P. Vanezis, B. Erhayiem, A. Graham, and S. Jamil-Copley, “Myocardial scar imaging: Viability beyond revived,” *Current Cardiovascular Imaging Reports*, vol. 17, no. 10, pp. 107–114, 2024.
- [105] A. Lalande, Z. Chen, T. Pommier, T. Decourselle, A. Qayyum, M. Salomon, D. Ginhaç, Y. Skandarani, A. Boucher, K. Brahim *et al.*, “Deep learning methods for automatic evaluation of delayed enhancement-mri. the results of the emidec challenge,” *Medical Image Analysis*, vol. 79, p. 102428, 2022.
- [106] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, “Secure and robust machine learning for healthcare: A survey,” *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 156–180, 2020.
- [107] N. Savage, “Synthetic data could be better than real data,” 2023.
- [108] I. Horvath, J. Paetzold, O. Schoppe, R. Al-Maskari, I. Ezhov, S. Shit, H. Li, A. Ertürk, and B. Menze, “Metgan: generative tumour inpainting and modality synthesis in light sheet microscopy,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 227–237.
- [109] C. Han, Y. Kitamura, A. Kudo, A. Ichinose, L. Rundo, Y. Furukawa, K. Umemoto, Y. Li, and H. Nakayama, “Synthesizing diverse lung nodules wherever massively: 3d multi-conditional gan-based ct image augmentation for object detection,” in *2019 International Conference on 3D Vision (3DV)*. IEEE, 2019, pp. 729–737.
- [110] Q. Jin, H. Cui, C. Sun, Z. Meng, and R. Su, “Free-form tumor synthesis in computed tomography images via richer generative adversarial network,” *Knowledge-Based Systems*, vol. 218, p. 106753, 2021.

- [111] B. Billot, D. N. Greve, O. Puonti, A. Thielscher, K. Van Leemput, B. Fischl, A. V. Dalca, J. E. Iglesias *et al.*, “Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining,” *Medical image analysis*, vol. 86, p. 102789, 2023.
- [112] A. K. Pishva, V. Thambawita, J. Torresen, and S. A. Hicks, “Repolyp: A framework for generating realistic colon polyps with corresponding segmentation masks using diffusion models,” in *2023 IEEE 36th international symposium on computer-based medical systems (CBMS)*. IEEE, 2023, pp. 47–52.
- [113] Y.-C. Chou, G. Y. Li, L. Chen, M. Zahiri, N. Balaraju, S. Patil, B. Hicks, N. Schnittke, D. O. Kessler, J. Shupp *et al.*, “Ultrasound image synthesis using generative ai for lung ultrasound detection,” *arXiv preprint arXiv:2501.06356*, 2025.
- [114] Q. Chen, X. Chen, H. Song, Z. Xiong, A. Yuille, C. Wei, and Z. Zhou, “Towards generalizable tumor synthesis,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 11 147–11 158.
- [115] H. Tian, W. Lei, L. Dai, H. Chen, and X. Zhang, “Lesiondiffusion: Towards text-controlled general lesion synthesis,” *arXiv preprint arXiv:2503.00741*, 2025.
- [116] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, “Repaint: Inpainting using denoising diffusion probabilistic models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 461–11 471.
- [117] N. Jathanna *et al.*, “The nottingham ischaemic cardiovascular magnetic resonance resource (notis cmr): a prospective paired clinical and imaging scar database—protocol,” *Journal of Cardiovascular Magnetic Resonance*, vol. 25, no. 1, p. 69, 2023.
- [118] “Creation of a cardiac magnetic resonance (CMR) scar segmentation tool using multi-vendor scans,” <https://www.hra.nhs.uk/planning-and-improving-research/applications-summaries/research-summaries/creation-of-a-cardiac-magnetic-resonance-cmr-scar-segmentation-tool-using-multi-vendor-scans/>, accessed: 2025-6-6.
- [119] E. C. Perin, G. V. Silva, R. Sarmiento-Leite, A. L. Sousa, M. Howell, R. Muthupillai, B. Lambert, W. K. Vaughn, and S. D. Flamm, “Assessing myocardial viability and infarct transmural extent with left ventricular electromechanical mapping in patients with stable coronary artery disease: validation by delayed-enhancement magnetic resonance imaging,” *Circulation*, vol. 106, no. 8, pp. 957–961, 2002.
- [120] R. J. Kim, T. S. Albert, J. H. Wible, M. D. Elliott, J. C. Allen, J. C. Lee, M. Parker, A. Napoli, and R. M. Judd, “Performance of delayed-enhancement magnetic resonance imaging with gadoversetamide contrast for the detection and assessment of myocardial infarction: an international, multicenter, double-blinded, randomized trial,” *Circulation*, vol. 117, no. 5, pp. 629–637, 2008.
- [121] R. O’Hanlon, A. Grasso, M. Roughton, J. C. Moon, S. Clark, R. Wage, J. Webb, M. Kulkarni, D. Dawson, L. Sulaiman *et al.*, “Prognostic significance of myocardial fibrosis in hypertrophic cardiomyopathy,” *Journal of the American College of Cardiology*, vol. 56, no. 11, pp. 867–874, 2010.

- [122] A. Hughes, O. Okasha, A. Farzaneh-Far, F. Kazmirczak, P. S. Nijjar, P. Velangi, M. Akçakaya, C. M. Martin, and C. Shenoy, “Myocardial fibrosis and prognosis in heart transplant recipients,” *Circulation: Cardiovascular Imaging*, vol. 12, no. 10, p. e009060, 2019.
- [123] A. Gulati, A. Jabbour, T. F. Ismail, K. Guha, J. Khwaja, S. Raza, K. Morarji, T. D. Brown, N. A. Ismail, M. R. Dweck *et al.*, “Association of fibrosis with mortality and sudden cardiac death in patients with nonischemic dilated cardiomyopathy,” *Jama*, vol. 309, no. 9, pp. 896–908, 2013.
- [124] N. Mewton, C. Y. Liu, P. Croisille, D. Bluemke, and J. A. Lima, “Assessment of myocardial fibrosis with cardiovascular magnetic resonance,” *Journal of the American College of Cardiology*, vol. 57, no. 8, pp. 891–903, 2011.
- [125] S. Kuruvilla, N. Adenaw, A. B. Katwal, M. J. Lipinski, C. M. Kramer, and M. Salerno, “Late gadolinium enhancement on cmr predicts adverse cardiovascular outcomes in non-ischemic cardiomyopathy: a systematic review and meta-analysis,” *Circulation. Cardiovascular imaging*, vol. 7, no. 2, p. 250, 2013.
- [126] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, M. S. Verani, and American Heart Association Writing Group on Myocardial Segmentation and Registration for Cardiac Imaging, “Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association: A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association,” *Circulation*, vol. 105, no. 4, pp. 539–542, Jan. 2002.
- [127] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.
- [128] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, “Swin-unet: Unet-like pure transformer for medical image segmentation,” in *European conference on computer vision*. Springer, 2022, pp. 205–218.
- [129] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, “Unet 3+: A full-scale connected unet for medical image segmentation,” in *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. Ieee, 2020, pp. 1055–1059.
- [130] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang *et al.*, “Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers,” *Medical Image Analysis*, vol. 97, p. 103280, 2024.
- [131] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, G. Vaillant, A. M. Lee, N. Aung, E. Lukaschuk, M. M. Sanghvi *et al.*, “Automated cardiovascular magnetic resonance image analysis with fully convolutional networks,” *Journal of cardiovascular magnetic resonance*, vol. 20, no. 1, p. 65, 2018.

- [132] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski, “Medical image synthesis for data augmentation and anonymization using generative adversarial networks,” in *International workshop on simulation and synthesis in medical imaging*. Springer, 2018, pp. 1–11.
- [133] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification,” *Neurocomputing*, vol. 321, pp. 321–331, 2018.
- [134] B. Cao, H. Zhang, N. Wang, X. Gao, and D. Shen, “Auto-gan: self-supervised collaborative learning for medical image synthesis,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 10 486–10 493.
- [135] Y. Skandarani, P.-M. Jodoin, and A. Lalande, “Gans for medical image synthesis: An empirical study,” *Journal of Imaging*, vol. 9, no. 3, p. 69, 2023.
- [136] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” *arXiv preprint arXiv:2011.13456*, 2020.
- [137] J. Wolleb, F. Bieder, R. Sandkühler, and P. C. Cattin, “Diffusion models for medical anomaly detection,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2022, pp. 35–45.
- [138] F. Khader, G. Müller-Franzes, S. Tayebi Arasteh, T. Han, C. Haarburger, M. Schulze-Hagen, P. Schad, S. Engelhardt, B. Baeßler, S. Foersch *et al.*, “Denoising diffusion probabilistic models for 3d medical image generation,” *Scientific Reports*, vol. 13, no. 1, p. 7303, 2023.
- [139] P. Dhariwal and A. Nichol, “Diffusion models beat gans on image synthesis,” *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.
- [140] A. Q. Nichol and P. Dhariwal, “Improved denoising diffusion probabilistic models,” in *International conference on machine learning*. PMLR, 2021, pp. 8162–8171.
- [141] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 10 684–10 695.
- [142] W. Bai, W. Shi, A. de Marvao, T. J. Dawes, D. P. O’Regan, S. A. Cook, and D. Rueckert, “A bi-ventricular cardiac atlas built from 1000+ high resolution mr images of healthy subjects and an analysis of shape and motion,” *Medical image analysis*, vol. 26, no. 1, pp. 133–145, 2015.
- [143] Z. Tang, P. Xue, P. Yang, D. Jia, and E. Dong, “An effective non-rigid image registration method based on active demons algorithm,” in *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2016, pp. 124–129.
- [144] G. Bazoukis, S. Garcia-Zamora, G. Çinier, S. Lee, E. E. Gul, J. Álvarez-García, G. Miñana, M. İ. Hayiroğlu, G. Tse, T. Liu *et al.*, “Association of electrocardiographic markers

- with myocardial fibrosis as assessed by cardiac magnetic resonance in different clinical settings,” *World Journal of Cardiology*, vol. 14, no. 9, p. 483, 2022.
- [145] T. Konno, K. Hayashi, N. Fujino, R. Oka, A. Nomura, Y. Nagata, A. Hodatsu, K. Sakata, H. Furusho, M. Takamura *et al.*, “Electrocardiographic qrs fragmentation as a marker for myocardial fibrosis in hypertrophic cardiomyopathy,” *Journal of Cardiovascular Electrophysiology*, vol. 26, no. 10, pp. 1081–1087, 2015.
 - [146] L. Li, J. Camps, Z. J. Wang, M. Beetz, A. Banerjee, B. Rodriguez, and V. Grau, “Toward enabling cardiac digital twins of myocardial infarction using deep computational models for inverse inference,” *IEEE transactions on medical imaging*, vol. 43, no. 7, pp. 2466–2478, 2024.
 - [147] A. J. Connolly and M. J. Bishop, “Computational representations of myocardial infarct scars and implications for arrhythmogenesis,” *Clinical Medicine Insights: Cardiology*, vol. 10, pp. CMC–S39 708, 2016.
 - [148] A. Saraste, S. Nekolla, and M. Schwaiger, “Contrast-enhanced magnetic resonance imaging in the assessment of myocardial infarction and viability,” *Journal of Nuclear Cardiology*, vol. 15, no. 1, pp. 105–117, 2008.
 - [149] A. Wasserman, G. Bren, A. Ross, D. Richardson, R. Hutchinson, and J. Rios, “Prognostic implications of diagnostic q waves after myocardial infarction,” *Circulation*, vol. 65, no. 7, pp. 1451–1455, 1982.
 - [150] D. Rovai, G. Di Bella, G. Rossi, M. Lombardi, G. D. Aquaro, A. L’Abbate, and A. Pingitore, “Q-wave prediction of myocardial infarct location, size and transmural extent at magnetic resonance imaging,” *Coronary artery disease*, vol. 18, no. 5, pp. 381–389, 2007.
 - [151] R. Sadeghi, V.-R. Dabbagh, M. Tayyebi, S. R. Zakavi, and N. Ayati, “Diagnostic value of fragmented qrs complex in myocardial scar detection: systematic review and meta-analysis of the literature,” *Polish Heart Journal (Kardiologia Polska)*, vol. 74, no. 4, pp. 331–337, 2016.
 - [152] M. Homsy, L. Alsayed, B. Safadi, J. Mahenthiran, and M. K. Das, “Fragmented qrs complexes on 12-lead ecg: a marker of cardiac sarcoidosis as detected by gadolinium cardiac magnetic resonance imaging,” *Annals of Noninvasive Electrocardiology*, vol. 14, no. 4, pp. 319–326, 2009.
 - [153] M. K. Das, M. A. Michael, H. Suradi, J. Peng, A. Sinha, C. Shen, J. Mahenthiran, and R. J. Kovacs, “Usefulness of fragmented qrs on a 12-lead electrocardiogram in acute coronary syndrome for predicting mortality,” *The American journal of cardiology*, vol. 104, no. 12, pp. 1631–1637, 2009.
 - [154] M. Reindl, S. J. Reinstadler, H.-J. Feistritz, L. Niess, C. Koch, A. Mayr, G. Klug, and B. Metzler, “Persistent t-wave inversion predicts myocardial damage after st-elevation myocardial infarction,” *International journal of cardiology*, vol. 241, pp. 76–82, 2017.

- [155] J. V. Nable and W. Brady, "The evolution of electrocardiographic changes in st-segment elevation myocardial infarction," *The American journal of emergency medicine*, vol. 27, no. 6, pp. 734–746, 2009.
- [156] M. J. Zema, "Q wave, st segment, and t wave myocardial infarction. useful clinical distinction," *The American journal of medicine*, vol. 78, no. 3, pp. 391–398, 1985.
- [157] J. N. de Alencar, V. F. de Andrade Matos, M. K. Scheffer, S. P. Felicioni, M. F. N. De Marchi, and M. Martínez-Sellés, "St segment and t wave abnormalities: A narrative review," *Journal of Electrocardiology*, vol. 85, pp. 7–15, 2024.
- [158] H. Roukoz and K. Wang, "St elevation and inverted t wave as another normal variant mimicking acute myocardial infarction: the prevalence, age, gender, and racial distribution," *Annals of Noninvasive Electrocardiology*, vol. 16, no. 1, pp. 64–69, 2011.
- [159] B. Wieslander, K. C. Wu, Z. Loring, L. G. Andersson, T. F. Frank, G. Gerstenblith, G. F. Tomaselli, R. G. Weiss, G. S. Wagner, M. Ugander *et al.*, "Localization of myocardial scar in patients with cardiomyopathy and left bundle branch block using electrocardiographic selvester qrs scoring," *Journal of electrocardiology*, vol. 46, no. 3, pp. 249–255, 2013.
- [160] A. Bisson, J. Pucheux, C. Andre, A. Bernard, B. Pierre, D. Babuty, L. Fauchier, and N. Clementy, "Localization of left ventricular lead electrodes in relation to myocardial scar in patients undergoing cardiac resynchronization therapy," *Journal of the American Heart Association*, vol. 7, no. 21, p. e009502, 2018.
- [161] J. Liu, A. Wei, L. Cao, X. He, and C. Tang, "Contrastive trustworthy prototype learning for multi-modality myocardial pathology segmentation," *Applied Soft Computing*, vol. 173, p. 112909, 2025.
- [162] S. Li, X. Li, P. Wang, K. Liu, B. Wei, and J. Cong, "An enhanced visual state space model for myocardial pathology segmentation in multi-sequence cardiac mri," *Medical Physics*, vol. 52, no. 6, pp. 4355–4370, 2025.
- [163] R. Saunders, M. Caterino, and P. Somaiya, "A single-patient-use ecg system for cardiothoracic surgery admissions in the uk: a cost-consequence analysis," *Frontiers in Public Health*, vol. 11, p. 1027977, 2023.
- [164] O. A. Oke and N. Cavus, "A systematic review on the impact of artificial intelligence on electrocardiograms in cardiology," *International journal of medical informatics*, vol. 195, p. 105753, 2025.
- [165] G. Nijm, S. Swiryn, A. Larson, and A. V. Sahakian, *Characterization of the magnetohydrodynamic effect as a signal from the surface electrocardiogram during cardiac magnetic resonance imaging*. IEEE, 2006.
- [166] D. Abi-Abdallah, V. Robin, A. Drochon, and O. Fokapu, "Alterations in human ecg due to the magnetohydrodynamic effect: a method for accurate r peak detection in the presence of high mhd artifacts," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 1842–1845.

- [167] T. S. Gregory, “Magnetohydrodynamics present in physiological signals and real-time electrocardiography during magnetic resonance imaging,” 2015.
- [168] J. W. Krug and G. Rose, “Magnetohydrodynamic distortions of the ecg in different mr scanner configurations,” in *2011 Computing in Cardiology*. IEEE, 2011, pp. 769–772.
- [169] T. Frauenrath, K. Fuchs, M. A. Dieringer, C. Özerdem, N. Patel, W. Renz, A. Greiser, T. Elgeti, and T. Niendorf, “Detailing the use of magnetohydrodynamic effects for synchronization of mri with the cardiac cycle: a feasibility study,” *Journal of Magnetic Resonance Imaging*, vol. 36, no. 2, pp. 364–372, 2012.
- [170] C. Chen, L. Li, M. Beetz, A. Banerjee, R. Gupta, and V. Grau, “Large language model-informed ecg dual attention network for heart failure risk prediction,” *IEEE transactions on big data*, 2025.
- [171] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, “Film: Visual reasoning with a general conditioning layer,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.
- [172] A. Suinesiaputra, P. Ablin, X. Alba, M. Alessandrini, J. Allen, W. Bai, S. Cimen, P. Claes, B. R. Cowan, J. D’hooge *et al.*, “Statistical shape modeling of the left ventricle: myocardial infarct classification challenge,” *IEEE journal of biomedical and health informatics*, vol. 22, no. 2, pp. 503–515, 2017.
- [173] L. Chlon, M. Chlon, and M. M. Awada, “Robust multimodal learning via entropy-gated contrastive fusion,” *arXiv preprint arXiv:2505.15417*, 2025.
- [174] J. D. Fortune, N. E. Coppa, K. T. Haq, H. Patel, and L. G. Tereshchenko, “Digitizing ecg image: A new method and open-source software code,” *Computer methods and programs in biomedicine*, vol. 221, p. 106890, 2022.
- [175] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, “Encase: An ensemble classifier for ecg classification using expert features and deep neural networks,” in *2017 Computing in cardiology (cinc)*. IEEE, 2017, pp. 1–4.
- [176] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, “Ptb-xl, a large publicly available electrocardiography dataset,” *Scientific data*, vol. 7, no. 1, pp. 1–15, 2020.
- [177] J. T. Ortiz-Pérez, J. Rodríguez, S. N. Meyers, D. C. Lee, C. Davidson, and E. Wu, “Correspondence between the 17-segment model and coronary arterial anatomy using contrast-enhanced cardiac magnetic resonance imaging,” *JACC: Cardiovascular Imaging*, vol. 1, no. 3, pp. 282–293, 2008.
- [178] M. K. Das, H. Suradi, W. Maskoun, M. A. Michael, C. Shen, J. Peng, G. Dandamudi, and J. Mahenthiran, “Fragmented wide qrs on a 12-lead ecg: a sign of myocardial scar and poor prognosis,” *Circulation: Arrhythmia and Electrophysiology*, vol. 1, no. 4, pp. 258–268, 2008.

- [179] A. H. Kashou, H. Basit, and A. Malik, “St segment,” in *StatPearls [Internet]*. StatPearls Publishing, 2023.
- [180] Y. Li, Y. Xu, W. Li, J. Guo, K. Wan, J. Wang, Z. Xu, Y. Han, J. Sun, and Y. Chen, “Cardiac mri to predict sudden cardiac death risk in dilated cardiomyopathy,” *Radiology*, vol. 307, no. 3, p. e222552, 2023.
- [181] G. W. Reed, J. E. Rossi, and C. P. Cannon, “Acute myocardial infarction,” *The Lancet*, vol. 389, no. 10065, pp. 197–210, 2017.
- [182] S. Zhai, R. Gu, W. Lei, and G. Wang, “Myocardial edema and scar segmentation using a coarse-to-fine framework with weighted ensemble,” 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*. Springer, 2020, pp. 49–59.
- [183] C. Martín-Isla, M. Asadi-Aghbolaghi, P. Gkontra, V. M. Campello, S. Escalera, and K. 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*. Springer, 2020, pp. 1–16.
- [184] J. Zhang, Y. Xie, Z. Liao, J. Verjans, and Y. Xia, “Efficientseg: A simple but efficient solution 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*. Springer, 2020, pp. 17–25.
- [185] H. Yu, S. Zha, Y. Huangfu, C. Chen, M. Ding, and J. Li, “Dual attention u-net for multi-sequence cardiac mr images segmentation,” 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*. Springer, 2020, pp. 118–127.
- [186] H. Jiang, C. Wang, A. Chartsias, and S. A. Tsiftaris, “Max-fusion u-net for multi-modal pathology segmentation with attention and dynamic resampling,” 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*. Springer, 2020, pp. 68–81.
- [187] D. Li, Y. Peng, Y. Guo, and J. Sun, “Taunet: a triple-attention-based multi-modality mri fusion u-net for cardiac pathology segmentation,” *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 2489–2505, 2022.
- [188] Y. Zhang, N. He, J. Yang, Y. Li, D. Wei, Y. Huang, Y. Zhang, Z. He, and Y. Zheng, “mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 107–117.

- [189] J. Shi, L. Yu, Q. Cheng, X. Yang, K.-T. Cheng, and Z. Yan, “M²ftrans: Modality-masked fusion transformer for incomplete multi-modality brain tumor segmentation,” *IEEE Journal of Biomedical and Health Informatics*, 2023.
- [190] H. Ma, Q. Zhang, C. Zhang, B. Wu, H. Fu, J. T. Zhou, and Q. Hu, “Calibrating multimodal learning,” vol. 202, pp. 23 429–23 450, 2023.
- [191] A. Tarvainen and H. Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,” *Advances in neural information processing systems*, vol. 30, 2017.
- [192] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, “Fixmatch: Simplifying semi-supervised learning with consistency and confidence,” *Advances in neural information processing systems*, vol. 33, pp. 596–608, 2020.
- [193] Z. Xu, D. Liu, J. Yang, C. Raffel, and M. Niethammer, “Robust and generalizable visual representation learning via random convolutions,” *arXiv preprint arXiv:2007.13003*, 2020.
- [194] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
- [195] Y. Liu, H. Wang, Z. Chen, K. Huangliang, and H. Zhang, “Transunet+: Redesigning the skip connection to enhance features in medical image segmentation,” *Knowledge-Based Systems*, vol. 256, p. 109859, 2022.
- [196] Y. Gong, Y. Jia, T. Leung, A. Toshev, and S. Ioffe, “Deep convolutional ranking for multilabel image annotation,” *arXiv preprint arXiv:1312.4894*, 2013.
- [197] S. Laine and T. Aila, “Temporal ensembling for semi-supervised learning,” *arXiv preprint arXiv:1610.02242*, 2016.
- [198] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” *arXiv preprint arXiv:1711.05101*, 2017.
- [199] L. Chen, Y. Tu, P. Bai, and L. Pan, “Domain generalization in myocardial pathology segmentation with mixup augmentation,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 34–45.
- [200] J. Dong, K. Wan, B. Lin, and M. Yang, “Domain generalized myocardial pathology segmentation via data statistics modeling and feature covariance alignment,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 46–54.
- [201] I. Margolis, L. D. Toso, S. Buoso, and S. Kozzerke, “Two-stage weighted ensemble method for myocardial edema and scar segmentation,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 55–65.

- [202] Z. Zhu, Y. Lin, and M. Yang, “Me-unet: Enhancing mamba for myocardial pathology segmentation in multi-center multi-sequence cmr images,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 66–76.
- [203] J. Gao, Y. Cai, Z. Zhao, X. Lan, Q. Huang, L. Lan, and T.-Q. Li, “Rgu-mamba: An u-mamba network with region-based training optimized for domain generalization applied to myocardial scar and edema segmentation,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 77–86.
- [204] M. Tang, N. Li, and L. Pan, “Improved nn-unet: Generalizable multi-scale attention-driven segmentation of multi-sequence myocardial pathology,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 96–105.
- [205] Z. Luo and G. Wang, “Multi-model ensemble and region specific normalization for myocardial pathology segmentation with partial modalities,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 106–115.
- [206] Y. Wu, X. Yang, Y. Liu, J. Zheng, and M. Yang, “Multi-task framework for myocardial pathological segmentation on multi-sequence cardiac magnetic resonance,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 116–125.
- [207] J. Wang, X. Gong, T. Jin, R. Gao, S. Zheng, and W. Chen, “Progressive multi-channel fusion network for myocardial pathology segmentation on multi-modality cmr images,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 192–201.
- [208] H. Lin, N. Tavakoli, F. Schiffrers, S. López-Tapia, D. Kim, and A. K. Katsaggelos, “Gensegnet: Leveraging synthetic sequences and pseudo labels for multi-sequence myocardial pathology segmentation,” in *MICCAI Challenge on Comprehensive Analysis and Computing of Real-World Medical Images*. Springer, 2024, pp. 227–239.

Appendices

Appendix A

List of Abbreviations

Table A.1: *List of abbreviations used throughout the thesis.*

Abbreviation	Definition
AHA	American Heart Association
ASD	Average Surface Distance
AI	Artificial Intelligence
bSSFP	Balanced Steady-State Free Precession
CE	Cross-Entropy
CMR	Cardiac Magnetic Resonance
CNNs	Convolutional Neural Networks
CSN	Coarse Segmentation Network
DDPMs	Denoising Diffusion Probabilistic Models
DSC	Dice Similarity Coefficient
ECG	Electrocardiography
FiLM	Feature-wise Linear Modulation
FSN	Fine Segmentation Network
FWHM	Full Width at Half Maximum
GANs	Generative Adversarial Networks
HD	Hausdorff Distance
ICD	Implantable Cardioverter–Defibrillator
IoU	Intersection over Union
LAD	Left Anterior Descending
LCx	Left Circumflex
LGE	Late Gadolinium Enhancement
LV	Left Ventricle
LVEF	Left Ventricular Ejection Fraction
MHD	Magnetohydrodynamic

Abbreviation	Definition
MI	Myocardial Infarction
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
MyoPS	Myocardial Pathology Segmentation
NHS	National Health Service
PMO	Persistent Microvascular Obstruction
PRE	Precision
RCA	Right Coronary Artery
RF	Radiofrequency
RV	Right Ventricle
SD	Standard Deviation
SEN	Sensitivity
SGD	Stochastic Gradient Descent
TAFF	Temporal-Aware Feature Fusion
T2	T2-Weighted Magnetic Resonance Imaging
VT	Ventricular Tachycardia

Appendix B

Permissions for Content Reuse

Permissions for Reproduced Figures

Figures reproduced from Wikimedia Commons in this thesis (Chapters 1 and 2) are licensed under Creative Commons Attribution licenses (CC BY [2.5](#), [3.0](#), or [4.0](#)). These licenses permit reuse, distribution, and adaptation of the material, provided appropriate attribution is given to the original authors and sources. All such figures have been properly credited in their respective captions.

Elsevier

Permission to reuse material from Elsevier has been obtained via RightsLink. The licensed content is reproduced in this thesis in accordance with the terms and conditions specified in the license agreement. Figure 1.4 in this thesis is reproduced from the source [3] under this license. (License number: 6238891481160)

Computing in Cardiology (CinC)

Papers published in Computing in Cardiology are distributed under an open-access model, where copyright is retained by the authors and the work is shared under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#) license¹. This permits reuse, distribution, and reproduction of the material with appropriate attribution. This applies to the content presented in Chapter 3, where previously published work has been incorporated. The original work is cited at the beginning of the corresponding chapter.

Springer Nature

In accordance with the Springer Nature author reuse policy, authors retain the right to reuse their published work, including the Version of Record, in whole or in part, within their thesis².

¹<https://cinc.org/cinc-papers-on-line/>

²<https://www.springer.com/de/rights-permissions/obtaining-permissions/882>

This applies to the contents presented in Chapter 4, where previously published work has been reused. The original work is cited at the beginning of the corresponding chapter.

IEEE

In accordance with IEEE reuse policies, authors are permitted to reuse their published work in their thesis without requesting additional permission, provided that appropriate citation and copyright acknowledgement are included³. This applies to the content presented in Chapter 5, which is based on an accepted paper at the IEEE International Symposium on Biomedical Imaging (ISBI 2026). The material has been extended with additional results and is cited in the corresponding chapter.

IEEE Copyright Notice

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of University of Sheffield's products or services. Internal or personal use of this material is permitted. For reuse of any IEEE copyrighted material beyond this thesis, permission must be obtained via IEEE RightsLink⁴.

³<https://journals.ieeeauthorcenter.ieee.org/choose-a-publishing-agreement/avoid-infringement-upon-ieee-copyright/>

⁴<https://www.ieee.org/publications/rights/rights-link>