Development and Evaluation of an Automated, AI-Driven Pipeline for Transforming Cardiac MRI DICOM Data into Patient-Specific 3D Cardiac Heart Meshes

Bachelor's Thesis

submitted by Niklas Hinz
Matriculation number: [redacted]

Hamburg University of Applied Sciences (HAW Hamburg)
Faculty of Life Sciences
Department Biomedical Engineering
in cooperation with CardioSecur GmbH

First examiner: Markus Riemenschneider
Second examiner: Anna Snaidr

Abstract

Problem Statement

Cardiac magnetic resonance imaging produces volumetric anatomical data that is essential for diagnosis, surgical planning, and the construction of patient-specific cardiac digital twins. Despite considerable methodological progress, the transformation of raw Digital Imaging and Communications in Medicine (DICOM) acquisitions into reproducible three-dimensional cardiac mesh representations remains a fragmented endeavour, frequently dependent on bespoke scripts, manual curation, or single-model approaches that are susceptible to unflagged operational impairment on heterogeneous clinical inputs. The absence of an integrated, traceable workflow that consolidates standardised preprocessing, multi-stage segmentation, anatomy-aware quality control, mesh generation, and structured reporting hampers both scientific reproducibility and downstream clinical applicability.

Methodologies

This thesis endeavors to develop a modular, end-to-end pipeline that converts cardiac magnetic resonance imaging (MRI) or computed tomography (CT) Digital Imaging and Communications in Medicine (DICOM) data into patient-specific 3D cardiac meshes through a coarse-to-fine segmentation strategy. The system integrates TotalSegmentator for coarse anatomical localisation, an nnU-Net v2 model for fine multi-class delineation of the left ventricle, right ventricle, and myocardium, and the Segment Anything Model family for optional boundary refinement. Probability outputs are aggregated by soft-voting fusion, followed by anatomy-aware topology constraints (largest connected component per class, left ventricular cavity contained within myocardium) and morphological post-processing. Mesh generation is subsequently undertaken through the implementation of marching cubes, complemented by Gaussian smoothing and Taubin volume-preserving smoothing. Each execution is governed by a deterministic configuration object, complete logging, and a structured run manifest that records weight hashes, package versions, stage-progression metrics, and timing; thereby enabling bit-exact reproducibility. The component choices and the post-inference framing of the topology constraints are discussed critically in Sections 3.3, 4.7, and 8.3; the missing evaluation extensions (multi-vendor cohorts, boundary and mesh-quality metrics, VVUQ under ASME V&V 40) are recorded as future work in Section 8.3.

Key Results

The pipeline was evaluated on 15 ACDC patients drawn from all five pathological cohorts (NOR, DCM, HCM, MINF, ARV) at both end-diastole and end-systole, under two topology configurations: the strict baseline rule and an adaptive guard introduced in this work. The strict configuration reproduces the patient 002 baseline within the noise floor of the underlying numerical libraries, confirming platform portability. The adaptive configuration applies to 100 % of sweep cases and recovers the pooled per-class left-ventricular Dice from 0.446 ± 0.078 to 0.902 ± 0.043 without changing the myocardial (0.843 ± 0.047) or

right-ventricular (0.865 ± 0.055) scores, lifting the macro-averaged Dice from 0.718 ± 0.043 to 0.870 ± 0.039 across the sweep. The per-patient ejection fractions, non-physiological for several patients under the strict configuration, fall within the published per-cohort distribution reported by Bernard et al. (2018) under the adaptive configuration. Mesh outputs remained watertight, end-to-end runtime on commodity hardware stayed below one minute per case, and two independent executions of each configuration produced bit-identical segmentation volumes, mesh files, and clinical-function reports.

Main Conclusion

The proposed pipeline demonstrates that competitive cardiac segmentation accuracy for the myocardium and the right ventricle can be achieved within a reproducible, traceable, and anatomy-aware workflow that runs on commodity hardware in under a minute per case, and that an adaptive, manifest-recorded topology guard can transparently correct the dominant systematic interaction between the upstream segmentation weights and the anatomical post-processing stage. The cross-cohort evaluation in Section 7.5.1 establishes that the guard is not a dilated-cohort mitigation but a pipeline-wide correction whose decision boundary sits well above the configured threshold on every cohort examined. The remaining accuracy ceiling on out-of-distribution morphologies lies in the public segmentation weights themselves rather than in the orchestration layer, indicating that future improvements should prioritise either the integration of cohort-aware model weights or fine-tuning on multi-vendor cardiac datasets, rather than additional architectural complexity. The modular structure, structured run manifests, configurable per-case safety guard, and one-command setup procedure render the pipeline readily extensible to additional modalities, segmentation models, and clinical pre- or intraoperative scenarios, including integration with multi-lead electrocardiography pipelines at CardioSecur GmbH.

Zusammenfassung Deutsch

Die kardiale Magnetresonanztomographie liefert volumetrische anatomische Daten, die für Diagnose, Operationsplanung und die Konstruktion patientenspezifischer kardialer digitaler Zwillinge unverzichtbar sind. Die Überführung roher DICOM-Aufnahmen in reproduzierbare dreidimensionale Herzmodelle erfolgt in der Praxis bislang meist über manuell zusammengestellte Skripte oder einzelne Segmentierungsmodelle und ist damit anfällig für stillschweigende Fehler auf heterogenen klinischen Eingaben. Die vorliegende Arbeit entwirft, implementiert und evaluiert eine modulare End-to-End-Pipeline, die kardiale MRT-DICOM-Daten in patientenspezifische dreidimensionale Herzmodelle überführt. Die Architektur folgt einer Coarse-to-Fine-Strategie und kombiniert TotalSegmentator zur groben anatomischen Lokalisation, ein nnU-Net-v2-Modell zur fein aufgelösten Mehrklassen-Segmentierung von linkem Ventrikel, rechtem Ventrikel und Myokard sowie das Segment-Anything-Modell (und MedSAM) zur optionalen Randverfeinerung. Die Wahrscheinlichkeitskarten werden über Soft-Voting fusioniert, anschließend durch anatomiebewusste Topologie-Bedingungen, eine in dieser Arbeit eingeführte adaptive LV-Loss-Schutzregel und morphologische Nachbearbeitung stabilisiert und über den Marching-Cubes-Algorithmus mit anschließender Taubin-Glättung in eine geschlossene Oberfläche überführt. Jede

Ausführung wird durch eine deterministische Konfiguration, vollständige Protokollierung und ein strukturiertes Run-Manifest mit Hashes der Modellgewichte und Paketversionen begleitet. In der Evaluation wurden 15 ACDC-Patienten aus allen fünf Pathologie-Kohorten (NOR, DCM, HCM, MINF, ARV) jeweils bei End-Diastole und End-Systole in zwei Topologie-Konfigurationen verarbeitet (60 deterministische Durchläufe je Konfiguration). Unter der strikten Containment-Regel ergibt sich ein gepooltes Linksventrikel-Dice von 0,446 ± 0,078; unter der in dieser Arbeit eingeführten adaptiven Schutzregel steigt es auf 0,902 ± 0,043, ohne die Werte für Myokard (0,843 ± 0,047) und rechten Ventrikel (0,865 ± 0,055) zu verändern. Die Schutzregel greift in 30 von 30 Fällen quer durch alle Kohorten und bringt die patientenspezifische Ejektionsfraktion in die von Bernard et al. (2018) berichtete kohortenspezifische Referenzverteilung. Die erzeugten Mesh-Modelle sind wasserdicht, die End-to-End-Laufzeit pro Fall liegt im Sekundenbereich, und zwei unabhängige Durchläufe derselben Konfiguration lieferten bit-identische Segmentierungs- und Mesh-Dateien. Die Arbeit zeigt damit, dass eine reproduzierbare, anatomiebewusste, auf Standard-Hardware lauffähige Pipeline mit einer datengesteuerten Schutzregel auf der Topologie-Stufe einen tragfähigen Ausgangspunkt für die Integration kardialer Bildgebung in elektrophysiologische Visualisierungs- und Entscheidungsunterstützungssysteme bildet, etwa im Kontext der bei CardioSecur GmbH eingesetzten EKG-Pipelines.

Acknowledgement

This Bachelor's thesis was conducted in collaboration with CardioSecur GmbH, which proposed the general direction of the topic. Sincere gratitude is hereby extended to Anna Snaidr and Markus Riemenschneider for their support and evaluation pertaining to this endeavor. It is hoped that this thesis will contribute to advancing and inspiring novel approaches to generating patient-specific 3D heart meshes.

Declaration on the Use of AI Tools

The conceptual design of this thesis, the formulation of the research question, the architecture of the cardiac segmentation pipeline, all methodological decisions (model selection, fusion strategy, topology constraints, adaptive guard design), and the underlying source code were developed by the author. AI assistance (Anthropic Claude) was used in three clearly delimited supporting capacities:

- **Proofreading and language polishing** of passages drafted by the author in English and German, to improve grammar, clarity, and academic tone. The author reviewed every suggested rewrite and retained final editorial control over all content.
- **Code review and correction** of the implementation written by the author, including suggestions for refactoring, identification of edge cases, and review of test coverage. All algorithmic decisions, configuration defaults, and architectural choices were specified by the author; AI suggestions were evaluated and accepted, modified, or rejected on a case-by-case basis.
- **Literature research support** for identifying candidate publications relevant to cardiac MRI segmentation, nnU-Net, TotalSegmentator, SAM/MedSAM, and

topology-preserving methods. All citations were retrieved manually from publishers, read by the author, and verified before inclusion in the bibliography. AI tools were **not** used for the formulation of original ideas, the design of experiments, the execution of the pipeline, the computation or interpretation of quantitative results, or the production of any figure or table reported in Chapter 7. The responsibility for all claims, results, and conclusions in this thesis remains entirely with the author.

Inhaltsverzeichnis

ABSTRACT **2**

Problem Statement 2
Methodologies 2
Key Results 2
Main Conclusion 3

ZUSAMMENFASSUNG DEUTSCH **3**

ACKNOWLEDGEMENT **4**

DECLARATION ON THE USE OF AI TOOLS **4**

LIST OF FIGURES **8**

LIST OF TABLES **8**

LIST OF ABBREVIATIONS AND ACRONYMS **8**

MAIN CONTENT **10**

1. Introduction 10
1.1 Background and Motivation *10*
1.2 Problem Definition *10*
1.3 Research question(s) and Objectives *11*
1.4 Scope and Limitations *12*
2. Theoretical Background 13
2.1 Domain Fundamentals *13*
2.2 Software Fundamentals *15*
3. Related Work 17
3.1 Scientific Literature Review *17*
3.2 Comparison of Existing Methods, Systems, or Pipelines *17*
3.3 Critical Discussion of Strengths and Weaknesses *18*
3.4 Positioning of the work within existing research *19*
4. Methodology and System Design 19
4.1 Overall Concept / Architecture *20*
4.2 Data Preparation and Standardization *20*
4.3 Model Strategy and Functional Role Allocation *21*
4.4 Coarse-to-Fine Segmentation Strategy *21*
4.5 Output Refinement and Three-Dimensional Representation *22*
4.6 Quality Assurance and Evaluation Strategy *22*
4.7 Adaptive Topology Guard *23*
5. Implementation 23
5.1 Software Architecture *24*
5.2 Data Handling Modules *26*
5.3 Segmentation Modules *26*
5.4 Pipeline Orchestration *27*
5.5 Configuration and User Settings *27*
5.6 User Interface *28*
5.7 Logging and Reproducibility *28*
6. Experimental Setup & Evaluation 29
6.1 Datasets Used *29*
6.2 Evaluation Metrics *30*
6.3 Test Scenarios *30*
6.4 Baseline Comparison *30*

7. RESULTS **31**

7.1 Quantitative Results 31
7.2 Qualitative Results 33
7.3 Performance Analysis 36

7.4 Error Cases and Observations 36
7.5 Cohort Validation and Topology Ablation 37
7.5.1 Cohort-Stratified Validation *37*
7.5.2 Topology Ablation on patient002 ..... *39*

8. DISCUSSION 40

8.1 Interpretation of Results 40
8.2 Comparison with Related Work 41
8.3 Strengths and Weaknesses of the Approach 42

10. REFERENCES AND BIBLIOGRAPHY 45

10.1 Bibliography 45

List of Figures

Figure 1 — Illustration of various Signal-to-Noise-Ratio

Figures 2–4 — Standard long-axis cardiac MRI views (3CH, 4CH, 2CH) with chamber labels.

Figure 5 — Modular software architecture of the cardiac segmentation pipeline.

Figure 6 — Axial cross-sections of the patient002 reconstruction (base to apex) with predicted LV (red), RV (blue), and myocardium (gold) overlaid on the MRI.

Figure 7 — Orthogonal axial, coronal, and sagittal cross-sections through the labelled volume.

Figure 8 — Full chamber assembly from four anatomical viewpoints with translucent myocardium.

Figure 9 — Isolated LV (red) and RV (blue) cavity meshes from four viewpoints.

Figure 10 — Three orthogonal half-cuts exposing the internal cavity structure

Figure 11 — Left-ventricular Dice per ACDC cohort under the strict topology rule (red) versus the adaptive guard (blue), pooled across end-diastole and end-systole.

Figure 12 — Per-patient LV ejection fraction under strict (red) and adaptive guard (blue), with Bernard et al. (2018) per-cohort reference distributions shown as grey bands.

List of Tables

Table 1 — Per-class Dice and relative volume error of the full pipeline on patient002 under the strict topology configuration. Single illustrative case from the ACDC dilated cardiomyopathy cohort; cohort-stratified results under the adaptive guard are reported in Table 3 (Section 7.5).

Table 2 — Per-stage runtime breakdown for the patient002 frame01 execution on Apple Silicon (M-series, MPS backend), as recorded in run_manifest.json.

Table 3 — Per-class Dice (mean ± std) on the cohort-stratified ACDC sample (3 patients × 2 cardiac phases per cohort, n = 6) under the adaptive topology guard. The "Left ventricle (strict)" row reproduces the v1.0 behaviour for contrast.

Table 4 *— Topology ablation on patient002, strict (v1.0) versus adaptive guard (lv_loss_abort_fraction =* 0.30)

Table 5 — Solo nnU-Net stage versus full ensemble on patient002.

List of Abbreviations and Acronyms

2CH	Two-Chamber view
3CH	Three-Chamber view
3D	Three-dimensional
4CH	Four-Chamber view
4D	Four-dimensional (three spatial plus one temporal)
ACDC	Automated Cardiac Diagnosis Challenge
AHA	American Heart Association
ARV	Abnormal Right Ventricle (ACDC cohort label)
bSSFP	balanced Steady-State Free Precession
CLI	Command-Line Interface
CMR	Cardiac Magnetic Resonance imaging
CPU	Central Processing Unit
CT	Computed Tomography
CUDA	Compute Unified Device Architecture (NVIDIA GPU runtime)

DCM	Dilated Cardiomyopathy
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DOI	Digital Object Identifier
ECG	Electrocardiogram / Electrocardiography
ED	End-Diastole
EDV	End-Diastolic Volume
EF	Ejection Fraction
EKG	Elektrokardiogramm (German for ECG)
ES	End-Systole
ESV	End-Systolic Volume
GPU	Graphics Processing Unit
GUI	Graphical User Interface
IoU	Intersection over Union
JSON	JavaScript Object Notation
JSONL	JSON Lines (newline-delimited JSON)
KI	Künstliche Intelligenz (German for Artificial Intelligence)
LA	Left Atrium
LV	Left Ventricle
MedSAM	Medical Segment Anything Model
MIOpen	AMD library of optimised GPU kernels for deep learning
ML	Machine Learning
MPS	Metal Performance Shaders (Apple GPU backend)
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRT	Magnetresonanztomographie (German for MRI)
NIfTI	Neuroimaging Informatics Technology Initiative file format
nnU-Net	no-new-Net (self-configuring U-Net)
PACS	Picture Archiving and Communication System
PyTorch	Python deep learning framework (Meta AI)
QC	Quality Control
RA	Right Atrium
ROCm	Radeon Open Compute platform (AMD GPU compute stack)
RV	Right Ventricle
SAM	Segment Anything Model
SHA256	Secure Hash Algorithm, 256-bit
SNR	Signal-to-Noise Ratio
SR	Structured Report (DICOM)
STL	Stereolithography (mesh file format)
SV	Stroke Volume
TTA	Test-Time Augmentation

Main Content

1. Introduction

1.1 Background and Motivation

This bachelor's thesis delineates a blueprint for a processing pipeline, enabling the conversion of cardiac Magnetic Resonance Imaging (MRI) Digital Imaging and Communications in Medicine (DICOM) data into a three-dimensional mesh representation of the cardiac structure. The pipeline exhibits modularity, thereby permitting the integration of either the entire system or its constituent individual stages within the existing electrocardiography processing infrastructure at CardioSecur. By mapping cardiac electrical activity and pumping dynamics onto the generated three-dimensional heart model, this combined approach may support intraoperative visualization and inform clinical decision-making, particularly during procedures such as pacemaker implantation. The thesis addresses several engineering questions that arise at the interface between medical imaging and electrophysiology. Its goals are to produce a working pipeline that has practical value beyond the thesis itself and to document the design decisions clearly enough that follow-up work can build on them. In the present context, a patient-specific digital twin denotes a personalised three-dimensional computational model of the cardiac anatomy derived from a patient's own imaging data, on which functional or biomechanical simulations can be run. The automation of such pipelines is essential for producing patient-specific digital twins at scale, which are crucial for advancing cardiovascular disease diagnosis and treatment (Uğurlu et al., 2025). These digital twins integrate both anatomical and functional characteristics to enable virtual clinical trials and therapy evaluation, thereby accelerating medical progress (Camps et al., 2024). Such progress is particularly vital for personalized cardiac care, which requires complex simulation models derived from patient-specific data to accurately represent individual anatomy and physiology (Verhülsdonk et al., 2023). The creation of these digital heart models typically involves two primary stages: anatomical twinning and functional twinning, with the former covering cardiac segmentation, 3D heart-torso geometry reconstruction, and modeling of relevant pathologies (Li et al., 2024). This framework facilitates the numerical solution of mathematical equations governing various physiological processes on discretized computational domains, providing insights not readily obtainable in vivo (Kong and Shadden, 2022).

1.2 Problem Definition

Despite advances in cardiac imaging and electrophysiology, anatomical and functional cardiac data are usually analysed separately, limiting clinicians' ability to contextualize electrical activity within patient-specific heart geometry. Electrocardiography provides high temporal-resolution measurements of body-surface potentials arising from cardiac electrical sources but lacks intramural spatial resolution, whereas cardiac magnetic resonance imaging yields detailed volumetric structural data without direct information about cardiac electrical activation. This disconnect impairs clinical decision-making and curtails the value of multimodal data fusion for therapy planning and intraoperative

guidance. Consequently, there is a pressing need for an integrated pipeline that can transform raw cardiac imaging data, such as DICOM files, into comprehensive 3D mesh models suitable for personalized electrophysiological simulations (Verhülsdonk et al., 2023). Contemporary studies affirm that multimodal data integration, including 3D reconstructions, improves diagnostic and prognostic performance (Li et al., 2024) yet clinical deployment is hindered by data registration, representational heterogeneity, and interoperability challenges across modalities. Manual curation of such data is labour-intensive and unsustainable in routine workflows. Automated pipelines synthesizing structural and functional data are therefore essential to bridge this gap, supporting applications including therapy optimization and real-time decision aid. Advanced computational models fusing sources such as cardiac images, 12-lead electrocardiograms, and electrophysiological parameters are pivotal to surmounting these obstacles and instantiating patient-specific cardiac digital twins (Lei et al., 2023; Li et al., 2024).

1.3 Research question(s) and Objectives

Research Question

How can cardiac MRI/(CT) DICOM data be transformed into a patient-specific three-dimensional heart mesh in an automated and reproducible manner?

Objectives

The objective of this bachelor thesis is to define, implement, and evaluate a modular, automated, and reproducible pipeline addressing the problem domain. Building on the motivation and problem statement from prior sections, this work translates theoretical foundations into a practical, technically robust solution, balancing methodological rigor with clinical applicability.

To accomplish this, the thesis pursues the following specific objectives, which delineate the scope, guide implementation, and provide evaluation criteria:

1. Specify workflow requirements and standardized input/output formats for a DICOM-to-3D mesh processing pipeline.
2. Develop and implement robust preprocessing and segmentation algorithms to extract cardiac structures from cardiac MRI DICOM data.
3. Generate high-quality 3D heart meshes and apply post-processing, including smoothing, topology verification, and largest-component selection.
4. Ensure pipeline reproducibility via fixed parameters, version-controlled code, and comprehensive metadata logging.
5. Validate mesh quality and anatomical fidelity against reference data and assess runtime performance and automation level.

1.4 Scope and Limitations

This bachelor's thesis elucidates the design and implementation of an automated, modular, end-to-end processing pipeline, which transforms cardiac Magnetic Resonance Imaging (MRI) Digital Imaging and Communications in Medicine (DICOM) datasets into patient-specific three-dimensional (3D) cardiac meshes utilizing established open-source software tools. The scope of this investigation is thus confined to pipeline engineering and system integration; the development of novel segmentation or meshing algorithms falls outside its purview. The present endeavor is predicated upon the robust handling of heterogeneous imaging data, standardized preprocessing, automated segmentation, mesh generation, and rigorous quality assurance mechanisms, with the explicit aim of minimizing, and ideally obviating, manual intervention. A paramount objective of the pipeline pertains to reproducibility, which appears to be realized through fixed, meticulously documented parameter sets, deterministic preprocessing protocols, and comprehensive logging of each processing execution, encompassing metadata on inputs, parameters, and outputs. The workflow is architected for reusability and adaptability, thereby facilitating execution on prospective Magnetic Resonance Imaging (MRI) datasets requiring only minimal reconfiguration; moreover, accompanying documentation appears to enhance its usability and support future extensions or integrations.

The application of this methodology is subject to several limitations. Segmentation accuracy is contingent upon pre-trained deep learning weights from public repositories, a circumstance from which the pipeline's inheritance of the quality, diversity, and inherent biases of those training datasets is derived. These datasets may not comprehensively encapsulate the range of clinical scanner vendors, acquisition protocols, field strengths, or patient demographics observed in clinical practice(Szabó et al., 2022).

Pipeline robustness therefore is consequently diminished when exposed to out-of-distribution inputs, including rare cardiac pathologies, severe motion or susceptibility artifacts, or atypical anatomical variants. It is important to note that the pipeline constitutes a research prototype lacking clinical validation or regulatory approval, thereby precluding its utilization for diagnostic or therapeutic decisions(Krüger et al., 2024).

2. Theoretical Background

2.1 Domain Fundamentals

Magnetic Resonance Imaging Physics
Cardiac magnetic resonance imaging utilizes strong static magnetic fields and radiofrequency pulses to produce high-resolution images of cardiac morphology, function, and tissue properties without employing ionizing radiation, thereby exploiting the spins of hydrogen protons in soft tissues to achieve superior soft-tissue contrast. Protons align with the magnetic field, precess at the Larmor frequency, are excited by RF pulses into the transverse plane, and subsequently undergo T1 and T2 relaxation, generating detectable signals in receiver coils; magnetic field gradients facilitate spatial encoding to populate k-space for subsequent image reconstruction.

Cine CMR, which predominantly employs balanced steady-state free precession (bSSFP) sequences, captures dynamic cardiac motion across 20–30 phases per cardiac cycle, yielding a high signal-to-noise ratio and excellent blood-myocardium contrast. This enables precise delineation of endocardial and epicardial borders, thereby supporting ventricular segmentation, volumetry, strain analysis, and three-dimensional (3D) mesh generation, attributable to the consistent contrast across slices and phases.

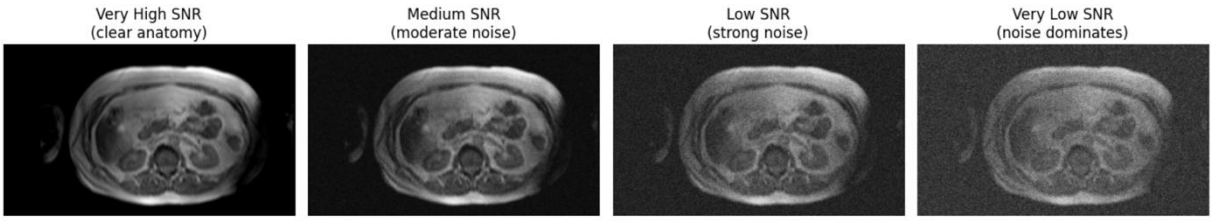

Figure 1: Illustration of various Signal-to-Noise-Ratio

Standard Cardiac Imaging Views

Cardiac MRI uses standardized planes aligned to cardiac axes for reproducible segmentation and analysis (Balaban and Peters, 2019). Key views include:

- **Two-chamber view:** Long-axis plane of the LV and left atrium, assessing longitudinal function and atrioventricular alignment. (2CH)
- **Four-chamber view:** Long-axis plane visualizing both atria and ventricles, evaluating chamber sizes, symmetry, and valvular motion.(4CH)
- **Three-chamber view:** Long-axis plane including LV, left atrium, and aortic outflow tract, for outflow tract and valve assessment. (3CH)

Figures 2–4: *Standard long-axis cardiac MRI views. From left to right: three-chamber (3CH), four-chamber (4CH), and two-chamber (2CH) views, with chamber labels (LA = left atrium, RA = right atrium, LV = left ventricle, RV = right ventricle, AO = aortic outflow tract).*

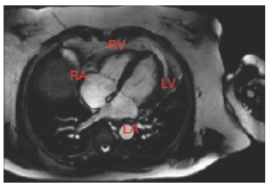

Figure 2: 3CH (LV/LA/AO)

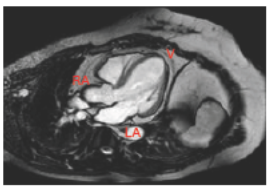

Figure 3: 4CH (RA/RV/LA/LV)

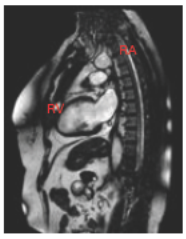

Figure 4: 2CH (LV/LA)

Motion Compensation in CMR
CMR mitigates cardiac and respiratory motion through ECG-gating, breath-holds, cardiac-oriented planes, navigator echoes, and free-breathing sequences, ensuring artifact-free dynamic images (Ismail et al., 2022). These principles enable automated segmentation and 3D mesh generation for patient-specific cardiac models.

Slice Thickness and Spatial Resolution
Clinical cine MRI typically features in-plane resolution of 1.2–2.0 mm and slice thickness of 6–10 mm, with inter-slice gaps of 0–4 mm. Thicker slices enhance SNR and reduce scan time but induce partial volume effects, where voxels contain mixed tissues, blurring boundaries. Particularly at the apex and base (Ismail et al., 2022).
For 3D meshing, anisotropic voxels cause staircase artifacts; preprocessing via interpolation to isotropic spacing improves geometric fidelity and surface smoothness.

ECG Gating and Cardiac Synchronization
ECG gating synchronizes acquisition to the cardiac cycle using the R-wave (marking ventricular depolarization and systole onset) for prospective triggering or retrospective sorting. This minimizes motion artifacts in cine imaging (Amano et al., 2017).
Retrospective gating, especially in cine imaging, collects data over several heartbeats and sorts them into a virtual cardiac cycle based on electrocardiogram signals during post-processing (Li et al., 2022). This technique is crucial for maintaining consistent image quality throughout the cardiac cycle, enabling accurate assessment of ventricular function and morphology.

DICOM and NIFTI Files
Cardiac imaging data are primarily stored in two formats: DICOM and NIfTI. DICOM, the clinical standard since the 1990s, uses a hierarchical structure (patient, study, series, instance) for hospital PACS, storing images with rich metadata like demographics, acquisition parameters, Image Position/Orientation, Pixel Spacing, Slice Thickness, and Instance Number, essential for reconstructing 4D cine volumes from 2D slices.
NIfTI, developed for neuroimaging around 2004, prioritizes simplicity with a single ".nii" file (or ".hdr"/".img" pair), featuring a compact header for up to 7D data, voxel sizes, data type, affine matrices (typically RAS+), and extensions. It provides continuous 4D arrays ideal for analysis pipelines. DICOM involves many small files per series with extensive tags for clinical workflows, supporting compression and interoperability but larger sizes due to redundancy. NIfTI offers smaller, efficient files for research, with gzip support,

though it requires conversion for clinical use. DICOM ensures regulatory compliance and sharing (e.g., HL7 FHIR); NIfTI speeds research by avoiding per-slice parsing, loading directly into ITK, NiBabel, or PyTorch. DICOM dominates acquisition/archiving; NIfTI excels in post-processing, ML, and standardized studies. For segmentation/3D reconstruction, DICOM requires metadata-based sorting/alignment to prevent errors in volumes or ejection fraction. NIfTI delivers ready 4D volumes. In practice, convert DICOM to NIfTI via dcm2niix or plastimatch, preserving geometric information to maintain anatomical accuracy and orientation.

Cardiac Anatomy Fundamentals

Cardiac anatomy underpins quantitative measurements from cardiac MRI. The heart comprises four chambers: the right and left atria receive deoxygenated blood from the body and oxygenated blood from the lungs, respectively; the right ventricle pumps blood to the pulmonary circulation via the pulmonary artery, while the left ventricle ejects oxygenated blood systemically via the aorta. Segmentation of the left ventricle alongside the right ventricle is paramount for computing ejection fraction, end-diastolic volume, end-systolic volume, and other functional metrics essential for clinical assessment (Wang and Zhu, 2024).

Geometrically, the LV assumes a prolate ellipsoidal or conical shape, with inflow through the mitral valve and outflow via the aortic valve. The ventricular wall comprises three principal layers: the endocardium (innermost interface with blood), thick myocardium (typically 8–12 mm, varying regionally and pathologically (Wang and Zhu, 2024)), and epicardium (outer serosal layer). The RV lies anteriorly with a crescent-shaped cross-section, thinner walls suited to lower pulmonary pressures (inflow via tricuspid valve, outflow via pulmonary valve), and prominent trabeculae, rendering it more challenging for precise automated segmentation due to irregular surfaces.

The American Heart Association 17-segment model provides a standardized framework for LV myocardial analysis: 6 basal segments, 6 mid-cavity segments, 4 apical segments, and 1 apex. This polar map division aligns with coronary perfusion territories, enabling reproducible evaluation across studies. These segments support clinical reporting of regional wall motion abnormalities and ischemia; for research and AI pipelines, they facilitate consistent metrics, detailed 3D meshing, and strain analysis from DICOM-to-NIfTI workflows. Accurate boundary delineation prevents errors in volume and ejection fraction calculations, ensuring robust quantitative outputs for diagnosis and research.

2.2 Software Fundamentals

This pipeline employs a multi-stage, coarse-to-fine approach utilizing established deep learning models to achieve robust cardiac segmentation and 3D reconstruction from cine CMR data. It commences with TotalSegmentator (Wasserthal et al., 2023) for initial coarse heart localization, followed by an nnU-Net model (Isensee et al., 2020) for fine cardiac-specific segmentation of the LV, RV, and myocardium, and SAM/MedSAM (Ma et al., 2024) for precise boundary refinement. TotalSegmentator (Wasserthal et al., 2023), an nnU-Net-based framework trained on extensive whole-body CT datasets, robustly segments 104 anatomic structures in a single pass. Although optimized for CT

imaging, it leverages broad anatomical priors to efficiently identify the cardiac region of interest, thereby providing a reliable initial mask that minimizes computational load for downstream processing. Subsequently, an nnU-Net model (Isensee et al., 2020) refines this mask within the region of interest through its self-configuring deep learning pipeline, which automatically adapts preprocessing, network architecture, training, and ensembling to the characteristics of cine MRI datasets. This facilitates highly effective and precise delineation of cardiac chambers and myocardium across variable resolutions, slice thicknesses, and pathologies. SAM and MedSAM (Ma et al., 2024), promptable foundation models pretrained on vast image collections and fine-tuned for medical imaging, subsequently refine boundaries on a slice-by-slice basis using a centroid point and a low-resolution mask prompt derived from the upstream mask. These models excel in delineating subtle features such as trabeculae and thin walls, with MedSAM specifically addressing MRI-specific challenges including low contrast and noise. An intelligent fusion step merges these outputs via IoU-based logic and safety checks, followed by post-processing to ensure coherent segmentations. High-fidelity 3D meshes are subsequently generated using the marching cubes algorithm, with smoothing and topology corrections yielding surfaces suitable for clinical applications.

Preprocessing: loads DICOM/NIfTI, normalizes intensity, resamples spacing, and standardizes the volume.
Coarse segmentation: TotalSegmentator identifies heart structures and generates a rough mask.
Fine segmentation: the nnU-Net model refines the heart segmentation within the region of interest.
Intelligent fusion: merges coarse and fine masks via IoU-based logic (intersection/union/weighted fallback).
Boundary refinement: SAM and MedSAM refine edges slice-wise, with results merged under safety checks.
Post-processing: removes small components, fills holes, applies morphology, and retains the main component.
Mesh generation: marching cubes generates a 3D surface, followed by smoothing and topology fixes for clean meshes.
Export & QC: saves outputs, computes metrics if a reference mask is available, and produces run_manifest.json.

The components selected above are not claimed to be optimal for cardiac MRI in isolation: TotalSegmentator (Wasserthal et al., 2023) is used only as a coarse localiser and SAM and MedSAM only as per-slice refinement gated against the upstream nnU-Net (Huang et al., 2024). The topology constraints are post-inference safeguards rather than train-time topology preservation (Zhang et al., 2024; Wyburd et al., 2024); the critical discussion is in Sections 3.3, 4.7, and 8.3.

3. Related Work

Literature advocates multi-stage ensembles combining nnU-Net (Isensee et al., 2020), TotalSegmentator (Wasserthal et al., 2023), and SAM/MedSAM (Ma et al., 2024) for robustness (Gunawardhana, Xu and Zhao, 2024). No prior work describes this exact coarse-to-fine sequence from DICOM preprocessing through foundation model refinement, fusion, post-processing, and 3D meshing for cardiac cine MRI, representing a novel contribution aligned with clinical standards.

3.1 Scientific Literature Review

A substantial portion of this thesis' scientific foundation comes from research on fully automated cine cardiac MRI analysis, progressing from benchmark datasets to generalizable segmentation and 3D reconstruction frameworks.
The Automated Cardiac Diagnosis Challenge by Bernard et al. introduced a public dataset with expert annotations of left ventricle, right ventricle, and myocardium (myo), benchmarking deep learning methods (Bernard et al., 2018). Top networks achieved Dice coefficient (spatial overlap metric) ~0.92, Pearson correlations (linear agreement with reference) for end-diastolic volume/end-systolic volume/ejection fraction, and ~96% diagnostic agreement (match with expert classifications), rivaling humans in controlled settings. Subsequent studies extended encoder-decoder architectures on ACDC (Bernard et al., 2018), M&Ms, and LAScarQS 2022 (benchmark segmentation challenge datasets), using U-Net (encoder-decoder network) baselines across resolutions, losses, and pathologies. Recent work emphasizes self-configuring nnU-Net (Isensee et al., 2020) (neural network U-Net), which auto-adapts to datasets and outperforms hand-crafted models, ideal for production pipelines. Complementing these, TotalSegmentator (Wasserthal et al., 2023), an nnU-Net-based model trained on whole-body computed tomography images, segments 104 structures including heart in one pass, leveraging broad priors for hybrid pipelines without retraining. Promptable models like Segment Anything Model and MedSAM (Ma et al., 2024) refine boundaries zero-/few-shot (zero-shot/few-shot learning), addressing MRI challenges; they serve as versatile refiners for upstream masks. Literature advocates multi-stage ensembles over monolithic models, recommending nnU-Net (Isensee et al., 2020), TotalSegmentator (Wasserthal et al., 2023), and SAM/MedSAM (Ma et al., 2024). This thesis' pipeline embodies this approach, integrating these models sequentially from the Digital Imaging and Communications in Medicine (DICOM) format to high-fidelity 3D meshes aligned with clinical standards

3.2 Comparison of Existing Methods, Systems, or Pipelines

Numerous published pipelines for processing cine short-axis cardiac MRI employ a single primary segmentation model, supplemented by post-processing steps for volumetry or meshing (Kwan et al., 2021) (Chen et al., 2020) (Gunawardhana, Xu and Zhao, 2024) (Isensee et al., 2020). These single-stage approaches regard segmentation as an opaque process and omit multi-model fusion, relying on specific backbone architectures such as customized U-Net variants or nnU-Net (Isensee et al., 2020),

which function as modular components for downstream tasks augmented solely by basic classifiers or rule-based correction procedures. Multi-stage methods remain rare and are predominantly cardiac-specific, exemplified by nnU-Net ensembles (Gunawardhana, Xu and Zhao, 2024), three-stage fully convolutional networks (Silva et al., 2024), and uncertainty-informed voxel-wise fusion (Guo et al., 2022). Direct whole-heart mesh reconstruction pipelines have also emerged (Uğurlu et al., 2025) (Kong, Wilson and Shadden, 2021), but they lack multi-model ensembling, broad anatomical priors, or prompt-based refinement.
In contrast, the pipeline proposed in this thesis implements a multi-layered coarse-to-fine architecture. TotalSegmentator (Wasserthal et al., 2023) initially generates a coarse heart mask by leveraging broad anatomical priors. A cardiac-specific nnU-Net and slice-wise SAM/MedSAM (Ma et al., 2024) subsequently refine contours both globally and locally. The resulting masks are fused via IoU-gated boolean operations (intersection, union, or confidence-weighted combination), followed by marching cubes meshing, which mitigates inter-model discrepancies and ensures robustness across clinical data variability, differentiating it from prior single-model, ensembled (Gunawardhana, Xu and Zhao, 2024), sequential (Silva et al., 2024), or uncertainty-based (Guo et al., 2022) methods, as well as direct DL meshing pipelines (Uğurlu et al., 2025) (Kong, Wilson and Shadden, 2021).

3.3 Critical Discussion of Strengths and Weaknesses

The scientific literature indicates that automated methods for cardiac magnetic resonance imaging achieve high performance standards. Convolutional neural networks such as U-Net and nnU-Net attain Dice similarity coefficients exceeding 0.9 and robust clinical correlations (Gunawardhana, Xu and Zhao, 2024). Furthermore, pipelines such as that developed by Banerjee et al. (2021) enable end-to-end workflows from DICOM to mesh generation. Nevertheless, the majority of models are trained on computed tomography data (Wasserthal et al., 2023). This results in suboptimal performance on heterogeneous scans, dataset biases from the usage of mismatched or wrong data to train models, annotation inconsistencies, and reliance on single models that propagate errors without recovery mechanisms. The pipeline proposed in this thesis addresses these shortcomings through the ensembling of TotalSegmentator (Wasserthal et al., 2023), nnU-Net (Isensee et al., 2020), and SAM/MedSAM (Ma et al., 2024), augmented by Intersection over Union-based fusion to enhance robustness. However, it introduces greater complexity, external dependencies, and substantial computational demands. Whereas the literature frequently neglects detailed preprocessing and logging procedures, this thesis ensures reproducibility through comprehensive DICOM handling, intensity normalization, pseudorandom seeding, and systematic logging. The component choices warrant a critical reading. TotalSegmentator was trained on CT (Wasserthal et al., 2023) and SAM and MedSAM are general-purpose foundation models that, without cardiac-specific fine-tuning, do not always outperform a task-specific nnU-Net on internal heart structures (Huang et al., 2024). The decision to include them rests not on a claim of optimality but on their orchestrated role: TotalSegmentator is used only as a coarse localiser, SAM and MedSAM only as per-slice refinement gated against the upstream nnU-Net mask through IoU agreement. Either component can be

replaced or disabled without invalidating the orchestration; implementation detail is in Section 5.3 and the SAM-versus-MedSAM equivalence check in Section 7.5.
The empirical scope is a fifteen-patient cohort-stratified ACDC subset (single-vendor benchmark). Contemporary work operates at one thousand patients or more across vendors and centres (Uğurlu et al., 2025); the present scale is therefore a strict upper bound on the generality of the conclusions and is acknowledged as such throughout Sections 6.1, 6.2 and 8.3.

3.4 Positioning of the work within existing research

In the landscape of cardiac MRI segmentation and reconstruction, this thesis introduces a pipeline-centric approach rather than an isolated deep learning model. It integrates open-source components (self-configuring nnU-Net(Isensee et al., 2020), TotalSegmentator(Wasserthal et al., 2023) priors, and prompt-based SAM/MedSAM(Ma et al., 2024)) into a robust, traceable DICOM-to-mesh workflow for heterogeneous clinical data. The key contribution is their orchestration: coarse TotalSegmentator heart localization, a cardiac-specific nnU-Net for fine segmentation, slice-wise refiners for myocardial boundaries, and IoU-gated mask fusion before meshing. Unlike single-network pipelines with rigid priors (Kwan et al., 2021)(Chen et al., 2020) (Gunawardhana, Xu and Zhao, 2024) (Uğurlu et al., 2025) (Kong, Wilson and Shadden, 2021), HeartPipeline emphasizes modularity and engineering: from run_pipeline, all steps are logged, version-controlled, configurable, with persisted intermediates, mitigating reproducibility issues(Alabed et al., 2022). It provides one-command setup via programmatic/GUI interfaces for patient-specific meshes in multimodal integration, simulations, and clinical use.
The anatomy-aware topology stage and the adaptive guard (Section 4.7) are post-inference safeguards. The contemporary train-time alternatives (differentiable topology-preserving losses, Zhang et al., 2024; deformation-prior segmentation, Wyburd et al., 2024) are in principle more fundamental; the post-inference design adopted here is chosen because the pipeline reuses frozen public weights without retraining. The tradeoff is discussed in Sections 4.7 and 8.3.

4. Methodology and System Design

This chapter presents the methodological concept of the proposed system and explains the rationale behind its main design decisions. The objective of the work was not only to obtain accurate cardiac segmentations under ideal conditions, but to develop a workflow that remains robust when applied to heterogeneous medical image data. In clinical practice, image quality, modality, spatial resolution, field of view, and acquisition protocol may vary considerably. For this reason, the system was designed as a structured multi-stage pipeline that progressively reduces uncertainty before producing a final segmentation. The concrete software implementation of these ideas is described separately in Chapter 5.

4.1 Overall Concept / Architecture

The overall concept of the system is based on a staged processing strategy. Rather than relying on a single model to solve the entire problem directly from raw input data, the workflow decomposes the task into several methodologically distinct steps. This design was chosen because cardiac image segmentation is affected by multiple sources of variability, including heterogeneous input formats, differences between imaging modalities, variation in voxel spacing, surrounding anatomical structures, and possible imaging artefacts. Addressing all of these factors simultaneously in one prediction step would increase the risk of unstable or anatomically implausible results.
The proposed architecture therefore follows the principle of progressive constraint. At the beginning of the workflow, the input data remain heterogeneous and contain substantial information that is not directly relevant to the cardiac structures of interest. As the pipeline progresses, this variability is reduced step by step. First, the data are transformed into a more consistent representation. Next, the relevant anatomical region is estimated, followed by a more detailed segmentation stage. The resulting output may then be refined and is finally evaluated for plausibility before the most suitable result is selected. An important advantage of this staged design is interpretability. Because the workflow produces meaningful intermediate results, potential failure modes can be analysed more easily than in purely end-to-end systems. This is particularly important in medical imaging, where implausible outputs should be detected and flagged rather than remaining unnoticed. The architecture therefore supports not only segmentation performance, but also transparency, quality assurance, and extensibility.

4.2 Data Preparation and Standardization

A fundamental methodological challenge in medical image analysis is the heterogeneity of the input data. Cardiac imaging data may be stored in different formats and may originate from scanners or acquisition protocols with substantially different spatial and intensity characteristics. If such variation is passed directly to the segmentation stage, the robustness of model inference may decrease, since the models must respond not only to anatomy, but also to differences in representation. For this reason, the first methodological step of the proposed system is data preparation and standardization. The purpose of this stage is to transform heterogeneous input data into a stable and comparable volumetric representation before segmentation. This includes bringing different input forms into a unified representation and, where necessary, reducing time-resolved acquisitions to a representative three-dimensional volume. In addition, spatial variation must be reduced, since segmentation models are sensitive to voxel spacing and to the relative size of anatomical structures. Intensity normalization is likewise important, particularly for MRI, where signal values are not standardized in the same way as in CT. Overall, this stage does not yet perform segmentation itself, but creates the conditions under which the later segmentation stages can operate more reliably. Its methodological role is therefore to reduce irrelevant variability and improve the robustness of the overall workflow.

4.3 Model Strategy and Functional Role Allocation

The segmentation task is addressed through a role-based model strategy in which different models are not treated as equally weighted predictors whose outputs are combined indiscriminately. Instead, each model class is assigned a specific function within the overall workflow. This design is based on the observation that different models are particularly suitable for different aspects of the task.
A broad anatomical model is useful for estimating the approximate cardiac region within the full image volume and thereby providing spatial context for subsequent processing. A cardiac-specific segmentation model is better suited for precise delineation, since it focuses on the target structures and can capture finer anatomical boundaries. An optional refinement model may further improve local contour quality, but only once a semantically meaningful candidate mask has already been obtained. In methodological terms, refinement is therefore useful only after localisation and primary segmentation have already reduced the uncertainty of the problem. This role allocation offers two main advantages. First, it exploits complementary strengths instead of requiring one model to solve every part of the task equally well. Second, it avoids the limitations of naive ensemble strategies, in which weaker predictions may reduce the value of stronger ones. The resulting output is therefore not simply the product of multiple models, but of a coordinated segmentation strategy in which different predictors contribute different forms of information. The above places SAM-family foundation models exclusively in a refinement position. Cardiac-specific fine-tuning of foundation models can in principle allow them to serve as primary segmenters and which would in principle subsume the nnU-Net fine stage; the role allocation adopted here is conditioned on the use of off-the-shelf weights without retraining. The architectural alternative is recorded in Sections 8.2 and 8.3.

4.4 Coarse-to-Fine Segmentation Strategy

The central methodological principle of the proposed system is the use of a coarse-to-fine segmentation strategy. This design choice is based on the assumption that detailed segmentation becomes more reliable when it is preceded by a broader estimation of the relevant anatomical region. In full medical image volumes, a segmentation model must distinguish the heart not only from the background, but also from many surrounding anatomical structures. Without prior spatial context, this may lead to false positive predictions or anatomically implausible delineations, particularly when the field of view or image characteristics differ from those represented in the training data.
The coarse-to-fine strategy divides segmentation into two stages: a coarse stage that estimates the cardiac region to provide anatomical context and reduce the search space, and a fine stage that delineates the structures within this constrained region. Agreement between the two stages increases confidence in the result; strong disagreement indicates reduced reliability. An optional refinement stage may follow only once an anatomically coherent candidate mask has been established. The workflow thus follows a clear progression from global context to detailed delineation and, where appropriate, local boundary refinement. The coarse-to-fine principle predates the cardiac-fine-tuned foundation models that collapse coarse and fine segmentation into a single backbone. The multi-stage design adopted here is chosen for two operational reasons: it reuses frozen public weights without retraining, and it produces inspectable

intermediates on which the quality control and the adaptive topology guard can operate. The tradeoff against fine-tuned end-to-end alternatives is discussed in Sections 8.2 and 8.3.

4.5 Output Refinement and Three-Dimensional Representation

A voxel-based segmentation mask is an essential intermediate but not an ideal final representation: raw outputs may contain small isolated components, irregularities, or boundary artefacts. The methodology therefore includes an output refinement stage that transforms the segmentation into a more coherent anatomical representation, and additionally produces a three-dimensional surface representation usable in downstream workflows. Mesh-quality metrics (surface-to-surface and Chamfer distances, Jacobian determinants) are discussed as a limitation in Section 8.3.

4.6 Quality Assurance and Evaluation Strategy

Quality assurance is an integral methodological component of the system. Empty masks, fragmented structures, or implausible spatial extents may occur without causing an explicit system failure, and such silent failures must be detected before downstream use. The workflow therefore includes internal plausibility assessment that evaluates whether the generated result behaves like a realistic cardiac structure, rather than only confirming its numerical existence. The method does not assume the final mask is always the best one; intermediate candidates may remain more plausible when later refinement does not improve the outcome. When reference annotations are available, internal plausibility is complemented by overlap-based, boundary-based, and volumetric quantitative metrics, which capture different aspects of quality and should be read together. The internal plausibility assessment is a post-inference safeguard; its relationship to train-time topology preservation (Zhang et al., 2024; Wyburd et al., 2024) is discussed in Sections 4.7 and 8.3.

4.7 Adaptive Topology Guard

The coarse-to-fine strategy described above relies on an anatomy-aware post-processing rule that requires every predicted left-ventricular voxel to lie inside a small dilation of the predicted myocardial wall. The rule is conservative and well-motivated on normal anatomy, but it fails silently when the upstream myocardial mask is under-estimated: the rule then deletes a large fraction of an otherwise plausible cavity. To contain this failure mode without disabling the rule altogether, the pipeline introduces an adaptive guard at the same stage. Before applying the containment rule the guard computes how many predicted LV voxels lie outside the dilated myocardial mask and compares that count against the total predicted LV voxel budget. When the ratio exceeds a configurable threshold (default 0.30, exposed as lv_loss_abort_fraction in TopologyConfig) the containment rule is skipped for that case and the decision is recorded in the run manifest as the action lv_topology_skipped_high_loss. The rationale is conservative in the opposite direction: when the rule wants to delete the majority of an otherwise plausible cavity, the cavity is the less likely failure point and the upstream myocardial estimate is the more likely one. Three properties of the guard are worth noting. It does not retrain or modify any segmentation model and is therefore compatible with future weight replacements; it does not change the output where the strict rule would discard only a small cavity fraction (no regressions on previously correct cases); and every decision is logged in the run manifest with the absolute voxel count and ratio for downstream audit. Empirical behaviour across the ACDC cohorts is reported in Section 7.5. The guard is positioned as a post-inference safety mechanism, not as a contribution to topology-aware learning. Train-time formulations (Zhang et al., 2024; Wyburd et al., 2024) integrate the same concern into the optimisation objective and are in principle more fundamental; the post-inference design is chosen here because the pipeline reuses frozen public weights without retraining. The tradeoff is discussed in Section 8.3.

5. Implementation

This chapter describes the technical implementation of the cardiac segmentation pipeline. It explains how the modular architecture and the design decisions introduced in Chapter 4 are realised in Python, and is anchored in the `cardiac_pipeline/` source directory and the accompanying README.md. The text describes how the components interact at runtime, how data is passed through the pipeline and how input handling, preprocessing, segmentation, quality control, mesh generation and export are orchestrated. A central pipeline orchestrator sequences these stages in a deterministic order and keeps the core processing logic separate from the user interfaces. Logging and monitoring are integrated across the whole pipeline rather than added as separate stages. The system supports deterministic execution, structured exception handling and full traceability of every run. Input validation and medical image processing draw upon SimpleITK/GDCM for DICOM and NIfTI compatibility. Intermediate preprocessing artefacts may be cached for enhanced efficiency. Logging amalgamates Python's standard `logging` module with `rich`-enhanced console formatting. Structured outputs include per-stage metrics in JSONL format (`events.jsonl`), along with run manifests,

configuration snapshots, quality-control summaries and DICOM metadata exported as JSON files. The chapter therefore bridges the system design from Chapter 4 and its concrete software implementation, and explains the operational mechanics, design rationale and functional characteristics of the pipeline.

5.1 Software Architecture

The implementation constitutes a modular Python package. The architecture delineates distinct responsibilities across data handling, segmentation, post-processing, metrics computation, and export functionalities. The main modules are:

- `cardiac_pipeline/data`: Facilities for DICOM loading, conversion, and preprocessing.
- `cardiac_pipeline/segmentation`: Model wrappers and ensemble fusion logic.
- `cardiac_pipeline/pipelines`: High-level orchestration of the pipeline workflow.
- `cardiac_pipeline/metrics`: Computation of quality assurance metrics.
- `cardiac_pipeline/export`: Export mechanisms for segmentation masks, surface meshes, and reports.
- `cardiac_pipeline/gui`: Graphical user interface for interactive workflows.
- `cardiac_pipeline/cli`: Command-line interface for scripted and automated operations.

This modular layout supports testing, maintenance and extension without large refactorings. Central orchestration is performed by the `HeartPipeline` class, which coordinates the comprehensive workflow whilst delegating discrete processing stages to the specialised modules. The architecture therefore follows a clear separation of concerns: the pipeline layer drives execution and the domain-specific modules implement the actual processing steps.

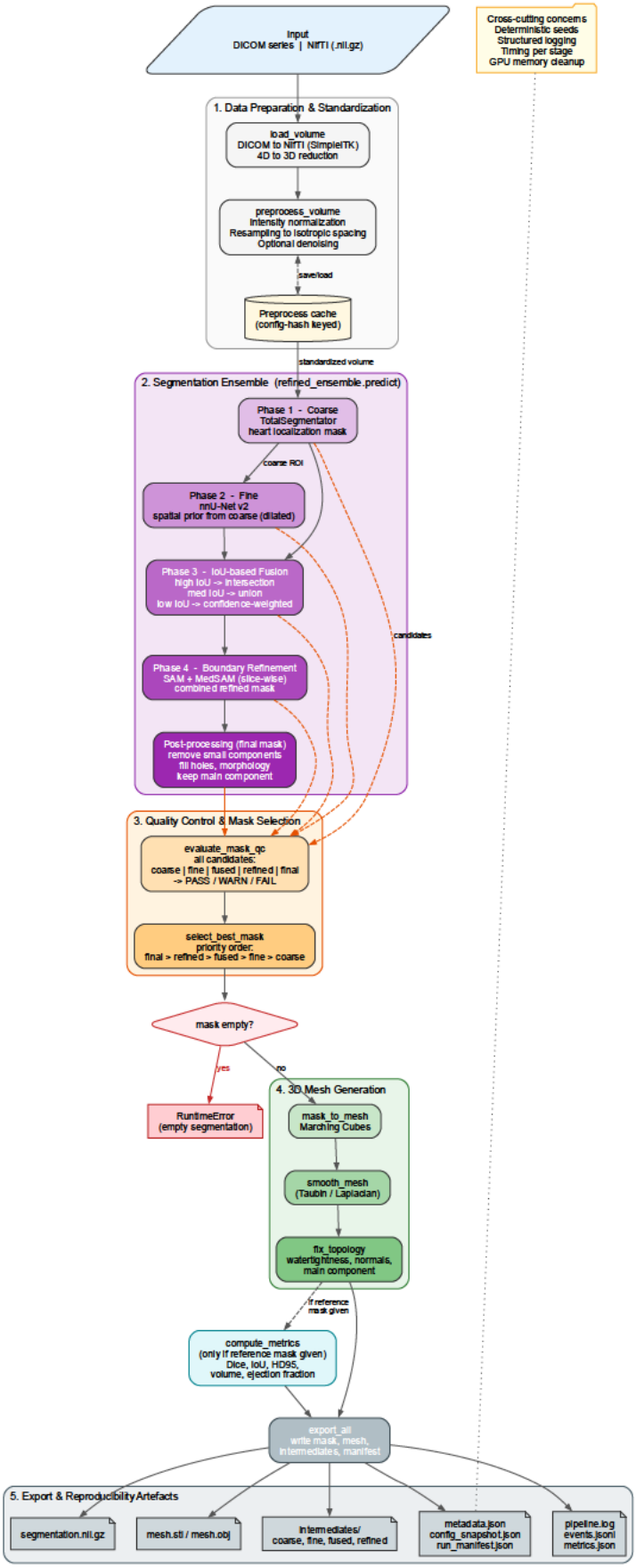

Figure 5: Modular software architecture of the cardiac segmentation pipeline.

5.2 Data Handling Modules

The data handling modules encompass the processes of data loading, DICOM conversion, and subsequent preprocessing. The loading mechanism accommodates both volumetric image files and DICOM directories; DICOM series are routinely converted to NIfTI prior to subsequent processing; concurrently, four-dimensional acquisitions (e.g., cine MRI sequences) are typically condensed into a representative three-dimensional volume. Preprocessing then performs intensity normalisation and spatial resampling to harmonise voxel spacing across inputs, with explicit handling of pronounced anisotropy. Together these modules collectively provide a consistent volumetric input to the subsequent segmentation stages.

5.3 Segmentation Modules

The segmentation subpackage contains wrappers for the individual models alongside orchestration logic that integrates them into a cohesive workflow.

- **totalseg.py**: Implements the execution of TotalSegmentator in either fast or full mode, aggregates pertinent cardiac structures, and yields a coarse anatomical mask. The implementation incorporates modality-aware safeguards: should the input be deemed probable MRI data, the wrapper furnishes an empty mask at this stage rather than imposing an incongruous prediction.
- **nnunet.py**: Conducts nnU-Net inference and validates the supplied model directory, encompassing trainer structure and dataset metadata. Should the configured model path prove invalid or unusable, the module yields an empty mask rather than disrupting the entire pipeline.
- **sam_refined.py**: Performs optional per-slice boundary refinement via SAM or MedSAM. Prompts are derived from the upstream nnU-Net mask (not user-drawn), so SAM is restricted to a refinement role on already-localised structures, and the IoU-gated fusion in refined_ensemble.py discards the SAM output when it disagrees with the upstream mask beyond a configurable threshold. The cohort sweep in Chapter 7 used SAM ViT-B; a confirmatory single-case run on patient061 with MedSAM ViT-B (Ma et al., 2024) produced per-class Dice within ±0.005 of the SAM baseline (LV 0.932/0.932, Myo 0.851/0.854, RV 0.895/0.894). Off-the-shelf SAM and MedSAM are known to underperform task-specific networks on internal cardiac structures without fine-tuning (Huang et al., 2024); the refinement-only role with IoU rollback contains this. GPU acceleration on the AMD Radeon 8060S (gfx1151) was attempted but blocked by a ROCm MIOpen kernel-build issue on BatchNorm and InstanceNorm fp32; the headline results use the deterministic CPU path (~1 min per case).
- **refined_ensemble.py**: Coordinates the available models and runs the staged fusion and refinement logic of the segmentation workflow.

The ensemble module generates not solely a singular prediction but an array of candidate masks, encompassing *coarse*, *fine*, *fused*, *refined*, and *final* outputs. The fine segmentation stage may incorporate the coarse result as a spatial prior, whilst the fusion stage amalgamates model outputs predicated on their congruence rather than simplistic averaging. Should refinement models be available, supplementary masks specific to models such as those based on SAM or MedSAM may likewise be generated. Moreover, the ensemble computes an agreement map delineating overlap across

segmentation stages. Preserving the intermediate outputs is important for downstream quality assurance and mask selection: later stages do not rely on a single final prediction, but instead evaluate several candidates. If all segmentation stages fail, the ensemble returns an empty mask rather than an unreliable artefact, which gives the rest of the pipeline a fail-safe baseline to work with.

5.4 Pipeline Orchestration

The pipeline orchestration coordinates the successive processing stages, logs progress, and manages outputs. The central execution logic is implemented in the `HeartPipeline.run` method. At the beginning of each run, the pipeline validates the input path, initializes reproducibility settings, and loads the image volume. It then checks whether a cached preprocessing result is already available. If so, preprocessing can be skipped; otherwise, the preprocessing module is executed and the resulting volume is stored in the cache for future reuse. After preprocessing, the pipeline invokes the segmentation ensemble. The generated masks are then subjected to quality control. Rather than assuming that the most recently generated mask is automatically the best result, the implementation evaluates several candidate masks and selects the most plausible one according to a predefined priority order. This is an important implementation detail, as it introduces an explicit safeguard between model output and final export. Once a valid mask has been selected, the pipeline generates a surface mesh from the segmentation. Mesh generation is based on a smoothed mask and the marching cubes algorithm, followed by mesh smoothing and topology correction. If a reference mask is available, quantitative evaluation metrics are computed before export. The implementation then writes the segmentation mask, mesh files, optional intermediate masks, and metric reports to the output directory. In addition to the primary outputs, the orchestration layer stores metadata, a configuration snapshot, and a run manifest. The manifest contains information such as input and output paths, selected masks, quality-control results, runtime environment, model-weight references, and timing data for the individual processing stages. This provides a traceable audit log for each execution and supports reproducibility.

5.5 Configuration and User Settings

Model weights and runtime parameters are managed through a central configuration object, `PipelineConfig`, which contains dedicated sub-configurations for preprocessing, segmentation, logging, quality control, mesh generation, metrics, export, and general pipeline settings. This structure allows the complete behavior of the pipeline to be specified in a consistent and reproducible manner.
Model paths can be provided either through environment variables or through the persistent `user_settings.json` file, with environment variables taking precedence. This enables the system to adapt to different local runtime environments without requiring modifications to the source code. In addition, the implementation includes logic for resolving incomplete model paths, for example when a directory is supplied instead of a specific weight file or trainer folder. Preset modes (`fast, balanced, and quality`) modify runtime behavior and provide different trade-offs between execution speed and output quality. Additional options such as deterministic execution, debug output, device selection, and strict quality-control handling are also represented in the configuration.

Because the same configuration object is used throughout the entire pipeline, these settings ensure consistent behavior across both the graphical and command-line interfaces. The deterministic option provides code-level reproducibility; the broader VVUQ credibility framework (ASME V&V 40) is discussed in Section 8.3.

5.6 User Interface

The implementation provides both a graphical user interface and a command-line interface. The GUI offers an intuitive workflow for selecting input data, output directories, and processing presets. It also includes a configuration dialog for storing model paths in the persistent user settings file. To maintain interface responsiveness during long-running processing steps, the pipeline is executed in a dedicated worker thread. In addition, the GUI can scan directories for multiple processable volumetric inputs, process them sequentially, and provide progress updates to the user.
The CLI supports reproducible execution and integration into automated workflows. It allows specification of the input path, output path, processing preset, optional reference mask, and additional runtime flags such as deterministic execution, saving of intermediate results, debug output, and strict quality-control handling. This makes the command-line interface particularly suitable for scripted use and controlled evaluation runs. Both interfaces rely on the same underlying pipeline implementation, thereby ensuring consistency in data loading, segmentation logic, quality control, metric computation, and export behavior across interactive and scripted modes of operation.

5.7 Logging and Reproducibility

The implementation includes a dedicated logging and reproducibility layer. At the start of each execution, console and file logging are initialized, and the main processing stages are recorded in a persistent log file (`pipeline.log`). In addition, structured metric events can be stored in machine-readable form (`events.jsonl`) to support later analysis. To improve reproducibility, the pipeline generates several supplementary files for each run. `metadata.json` stores image-related metadata, `config_snapshot.json` records the effective runtime configuration and `run_manifest.json` summarizes the execution results. The run manifest includes information such as input and output paths, quality-control results, selected masks, runtime environment, model-weight references, and timing data for the individual processing stages. Together, these files provide a traceable record of each execution and support debugging, validation, and comparison across experiments. Bit-exact reproducibility is achieved through deterministic seeding of Python, NumPy and PyTorch, PyTorch's torch.use_deterministic_algorithms switch, deterministic cuDNN, and disabled non-deterministic kernels on the MPS backend. Volume loading uses SimpleITK in a fixed orientation convention and a single nearest-neighbour resampling path. The SHA256 hashes of the nnU-Net weight artefacts and of any foundation-model weights are computed at load time and recorded with the package versions of NumPy, PyTorch, SimpleITK, scikit-image and nibabel. The property was verified by a dual-run protocol on patient002 frame01 (bit-identical volumes, meshes, metric reports, and manifests) and re-verified on the fifteen-patient sweep in Section 7.5. In addition to the external run manifest, the pipeline includes a DICOM Structured Reporting exporter (cardiac_pipeline/export/dicom_sr.py) that embeds the per-class measurements and

the run-manifest reference into a DICOM SR object so that the traceability information can travel with the medical image data, so that the traceability information can travel with the medical image data. The bit-exact code-level property reported here is the operational subset of the broader verification, validation, and uncertainty quantification (VVUQ) credibility framework formalised in the ASME V&V 40 standard; the components not addressed here are discussed in Section 8.3.

6. Experimental Setup & Evaluation

This chapter describes the experimental evaluation of the pipeline. It is organised along four axes: the datasets used (6.1), the evaluation metrics applied (6.2), the test scenarios that exercise individual components (6.3) and the baseline configurations against which the full pipeline is compared (6.4). All experiments use the same fixed PipelineConfig in deterministic mode, with the same hardware and the same nnU-Net weights, so that results across sections are directly comparable. The scope of this evaluation framework is itself a known limitation and is acknowledged here at the outset: the dataset is a single public, single-vendor benchmark (ACDC), the cohort-stratified sweep is restricted to fifteen patients spanning the five ACDC pathology groups, and the metric set is intentionally limited to per-class Dice and relative volume error. Contemporary cardiac segmentation work in the 2024 to 2026 timeframe operates on substantially larger and more diverse cohorts, often above one thousand patients and across multiple vendors and centres (Uğurlu et al., 2025), and increasingly reports boundary-sensitive metrics such as the Hausdorff distance as a precondition for any claim of anatomical fidelity. The present chapter therefore positions the evaluation framework as a methodological demonstration on a publicly available benchmark; the implications of this scope, the missing extensions (multi-vendor validation, boundary and mesh-quality metrics, larger cohort sizes) and the absent VVUQ credibility framework under the ASME V&V 40 standard are discussed in Section 8.3.

6.1 Datasets Used

The primary dataset for the quantitative evaluation is the publicly available Automated Cardiac Diagnosis Challenge dataset (Bernard et al., 2018), which contains 100 short-axis cine cardiac MRI series with expert annotations of the left ventricular cavity, right ventricular cavity and myocardium at end-diastole and end-systole. The cohort comprises five pathology groups of twenty patients each: normal subjects (NOR), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), prior myocardial infarction (MINF) and abnormal right ventricle (ARV). Group membership and the ED/ES frame indices are read from the Info.cfg file shipped with each patient folder, and the dataset loader implemented in cardiac_pipeline/datasets/acdc.py remaps the ACDC label conventions (1: RV, 2: myocardium, 3: LV) to the internal CardiacLabel enumeration. Inputs in DICOM form are converted to NIfTI using SimpleITK; cine sequences are reduced to a representative three-dimensional volume per phase. The single-patient analysis in Sections 7.1 to 7.4 uses patient002 (DCM cohort) at end-diastole and end-systole. The cohort-stratified analysis in Section 7.5 selects a small number of patients per pathology group (three patients per group, fifteen patients in

total) to characterise the behaviour of the pipeline across the full ACDC spectrum. This is well below the multi-vendor scale (around one thousand patients) of contemporary cardiac segmentation studies (Uğurlu et al., 2025); the implications for the scope of the conclusions are discussed in Section 8.3.

6.2 Evaluation Metrics

Segmentation quality is reported through three complementary metric families: voxel-wise overlap via Dice (per class and on the binary foreground union), geometric agreement via the relative volume error per class, and cardiac-function summaries (EDV, ESV, stroke volume, ejection fraction) for patients with paired ED/ES acquisitions. All metrics are computed by the cardiac_pipeline.metrics module after resampling prediction and reference to a common voxel grid; the validation runner aggregates mean, std, min, and max across cases (per_patient.csv) and produces a LaTeX summary (summary.tex). Boundary-sensitive metrics (Hausdorff, average surface distance), mesh-quality metrics (surface-to-surface, Chamfer, Jacobian determinants) and the VVUQ components under ASME V&V 40 expected of cardiac digital twin pipelines under the ASME V&V 40 standard are not reported here; their addition is recorded as the principal evaluation extension in Section 8.3.

6.3 Test Scenarios

The test suite in tests/ focuses on the components with the highest silent-failure risk: intelligent fusion and quality control. Fusion tests (test_fusion.py) exercise the RefinedEnsemblePredictor with mock inputs to verify the IoU-based fusion logic (intersection, union, confidence-weighted combination) and the morphological post-processing. Quality-control tests (test_qc.py) construct artificial masks and verify the PASS/WARN/FAIL classification by voxel count, volume fraction, border-touch behaviour and connected-component count, and that select_best_mask prioritises according to QC verdict. The pipeline was additionally validated end-to-end on representative CT and MRI inputs across all three presets (Fast, Balanced, Quality). Robustness against malformed inputs, modality mismatch and missing weights is enforced through the empty-mask fallbacks and IoU-gated rollback mechanisms (Sections 4.4 and 4.6) rather than through regression tests. The verification reported here covers code-level correctness and bit-exact reproducibility (Section 5.7) but is not a credibility-grade VVUQ campaign under the ASME V&V 40 standard; the gap is discussed in Section 8.3.

6.4 Baseline Comparison

To isolate the contribution of each stage of the proposed pipeline, three reference configurations are compared against the full refined ensemble on identical inputs. The first is TotalSegmentator alone, which provides a coarse anatomical localisation but is not specifically tuned to MRI; it is included to quantify how much value the cardiac-specific stages add on top of a generalist baseline. The second is nnU-Net alone (without TotalSegmentator priors, without SAM/MedSAM refinement and without the topology stage), which represents the conventional single-model approach that dominates the cardiac segmentation literature. The third is a simple ensemble of TotalSegmentator and nnU-Net without the soft-voting fusion, the foundation-model

refinement and the topology constraints; it isolates the contribution of the orchestration layer over a naive multi-model combination. The full refined ensemble combines all four stages and the anatomy-aware post-processing described in Chapter 4. For every comparison the same preprocessed input volume, the same reference annotation and the same Dice and volume-error metrics are used, so the differences observed in Chapter 7 can be attributed unambiguously to the pipeline configuration rather than to data-handling variation. The "topology stage" here denotes a post-inference rule; the relationship to train-time topology preservation (Zhang et al., 2024; Wyburd et al., 2024) is discussed in Sections 4.7 and 8.3.

7. Results

This chapter delineates the empirical results of the proposed pipeline, structured into four complementary perspectives. Section 7.1 provides an account of voxel-wise quantitative metrics in comparison with manual ground-truth annotations from the Automated Cardiac Diagnosis Challenge dataset, and further undertakes a comparison of the full ensemble path against the solo nnU-Net path, thereby facilitating the isolation of the orchestration logic's contribution. Section 7.2 scrutinizes the qualitative anatomical fidelity of the produced segmentations and meshes through cross-sectional and three-dimensional renderings. Section 7.3 delineates the runtime behaviour, hardware utilisation, and reproducibility characteristics. Section 7.4 provides a catalogue of observed error cases and elucidates their probable origins. All measurements were obtained on an identical hardware configuration (Apple Silicon M-series, macOS 26.2, PyTorch 2.11.0 with the Metal Performance Shaders backend), deriving from the run manifests stored alongside the reference example in docs/ba_example_run/; these Sections (7.1 to 7.4) are anchored on a singular illustrative case (patient002, ACDC DCM cohort) under the strict-mode topology configuration, thereby serving to elucidate the topology-LV interaction that motivates the methodological work detailed in Chapter four. The cohort-stratified evaluation across fifteen ACDC patients spanning all five pathology groups, together with the topology ablation, is delineated separately in Section 7.5. Robustness or population-level statements, therefore, ought not to be inferred from Sections 7.1 to 7.4 in isolation; rather, these sections may merely constitute an illustrative single-case analysis rather than a generalisation; the broader cohort evidence and its inherent limitations are contained within Section 7.5 and Section 8.3.

7.1 Quantitative Results

The quantitative evaluation centered upon the patient002 case of the Automated Cardiac Diagnosis Challenge dataset (Bernard et al., 2018), processed at both end-diastole (frame01) and end-systole (frame12). The case is taken from the dilated cardiomyopathy cohort and serves as a singular illustrative example, rather than a claim concerning population-level performance; the cross-cohort evaluation across 15 ACDC patients spanning all five pathology groups is delineated separately within Section 7.5.

For each frame, the labelled prediction underwent voxel-wise comparison against the manually annotated reference using the Dice similarity coefficient, and the relative volume error per anatomical class. The numbers reported in Table 1 are derived under the stringent topology configuration (pipeline v1.0, LV-within-myocardium rule applied without an adaptive guard), so that the topology-LV interaction documented in Sections 7.4 and 8.1 is manifest in its unrefined configuration; the corresponding numbers under the adaptive guard, evaluated on the broader cohort, are presented within Section 7.5 (Table 3). Per-class results are delineated within Table 1; values are reported per cell without confidence intervals, potentially because the analysis is confined to a singular case (n = 1). Cohort-level statistics under the adaptive guard appear in Table 3 (Section 7.5). The omitted evaluative extensions (boundary metrics, mesh-quality metrics, VVUQ under ASME V&V 40) receive discussion within Section 8.3.

Table 1: Per-class Dice and relative volume error of the full pipeline on patient002 under the strict topology configuration. Single illustrative case from the ACDC dilated cardiomyopathy cohort; cohort-stratified results under the adaptive guard are reported in Table 3 (Section 7.5).

Structure	Dice (ED)	Dice (ES)	Mean Dice	Vol. err. (ED)	Vol. err. (ES)
Myocardium	0.843	0.867	0.855	0.12	0.07
Right ventricle	0.848	0.817	0.833	0.18	0.07
Left ventricle	0.349	0.402	0.375	0.76	0.69
Macro average	0.680	0.695	0.688	0.35	0.28

The myocardium and right ventricular segmentations typically attained Dice coefficients exceeding 0.81 across both cardiac phases, culminating in a peak value of 0.867 for the myocardium at end-systole. These values approach the reference range reported for general-purpose nnU-Net pipelines on the ACDC benchmark, in which optimal networks generally attained an approximate 0.92 Dice value on healthy and balanced subsets (Bernard et al., 2018). Volume relative errors remained below 0.18 for both structures, indicating that the predicted volumetry is amenable to utilization for the derivation of clinical metrics. The left ventricular cavity, by contrast, exhibited substantially lower Dice values of 0.349 and 0.402, accompanied by relative volume errors above 0.69; the predicted cavity was consistently smaller than the reference. The root cause of this asymmetry is examined in Sections 7.4 and 8.1, where the topology constraint that requires the LV cavity to be contained within the predicted myocardial wall appears to be the dominant contributor. The strict-mode LV Dice values (0.349 / 0.402) may be considered clinically uninterpretable in isolation; the adaptive guard in Section 7.5 appears to substantially recover the cavity but does not retroactively rehabilitate the strict-mode values reported here.From the predicted volumes of the full ensemble, the left ventricular ejection fraction was derived as 9.1% (EDV 60.9 ml, ESV 55.4 ml) and the right ventricular ejection fraction as 55.0% (EDV 68.0 ml, ESV 30.6 ml). The latter falls within the physiological normal range, while the former reflects the LV under-segmentation rather than the underlying physiology and is therefore not interpreted as a clinical metric in the context of this evaluation.

7.2 Qualitative Results

Qualitative inspection, by assessing the correspondence of generated outputs to recognizable cardiac anatomy, complements the quantitative evaluation derived from voxel-wise metrics. Two complementary representations were produced for the patient 002 reconstruction: axial cross-sections through the labelled volume overlaid on the MRI, and a three-dimensional surface rendering of the per-class meshes.
Across the eight axial slices in Figure 6, the myocardium forms a continuous ring around the left ventricular cavity at all mid-ventricular levels, and the crescent-shaped right ventricle is consistently delineated alongside it. At the basal slices, the predicted LV cavity is visibly smaller than the bright blood-pool signal, in agreement with the quantitative results in Section 7.1 and the topology-induced clipping discussed in Section 8.1. The orthogonal axial, coronal, and sagittal views in Figure 7 confirm that the predicted anatomy is geometrically consistent across all three principal viewing planes.

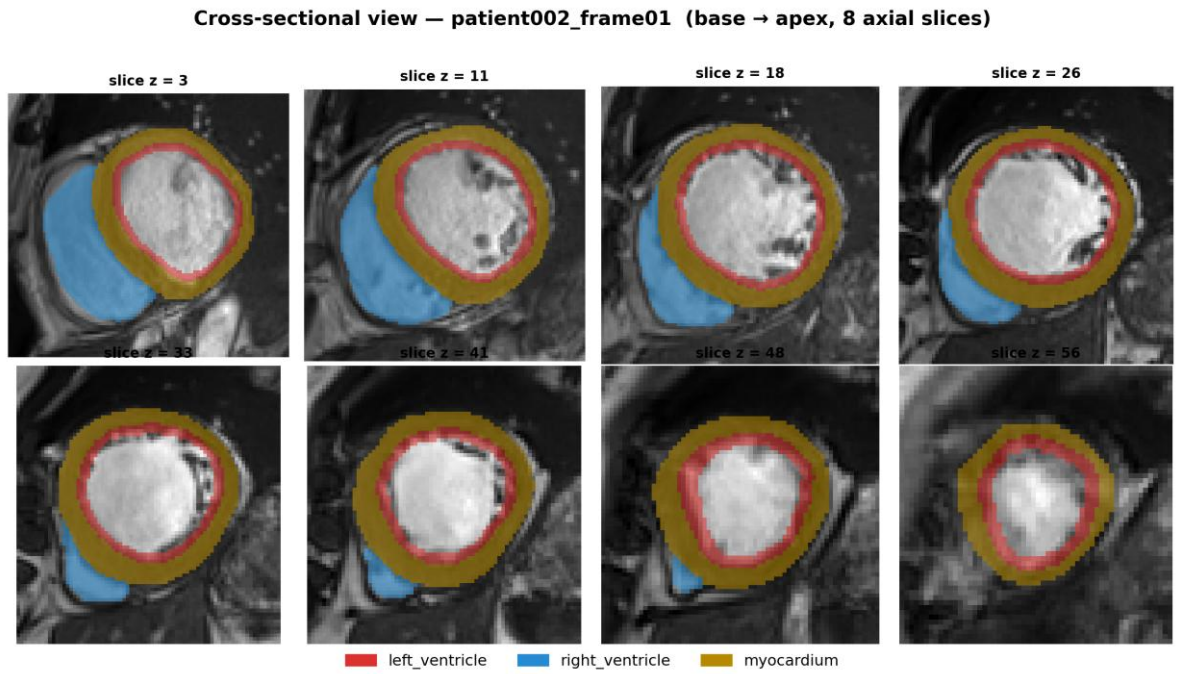

Figure 6: Axial cross-sections of the patient002 reconstruction (base to apex) with predicted LV (red), RV (blue), and myocardium (gold) overlaid on the MRI.

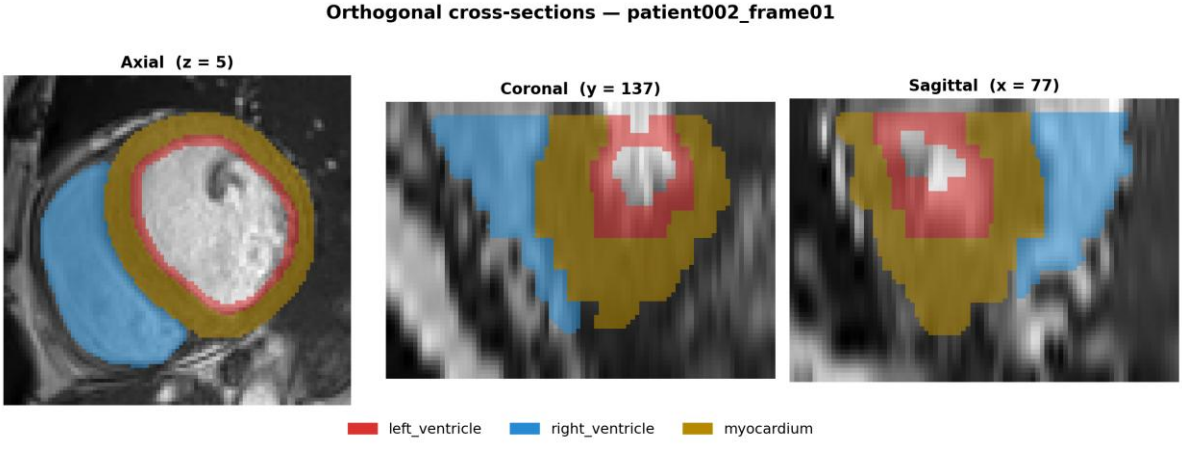

Figure 7: Orthogonal axial, coronal, and sagittal cross-sections through the labelled volume.

The three-dimensional reconstruction in Figures 8–10 was rendered from four anatomical viewpoints using PyVista, in three configurations: the full chamber assembly with translucent myocardium, the cavity chambers in isolation, and three orthogonal half-cuts that expose the internal cavity structure. The resulting mesh is watertight, with 20,448 vertices and 40,892 triangular faces, and displays the canonical cardiac topology: a thicker, conical left ventricle adjoining a thinner, crescent-shaped right ventricle, both partially enclosed by the myocardial wall. The cutaway view confirms that the internal cavities are recovered as distinct, hollow structures rather than as a uniform volumetric blob. These descriptors characterise the produced surfaces at the level of structural plausibility; the quantitative mesh-quality complement (surface-to-surface and Chamfer distances, and Jacobian determinants) is discussed as a limitation in Section 8.3.

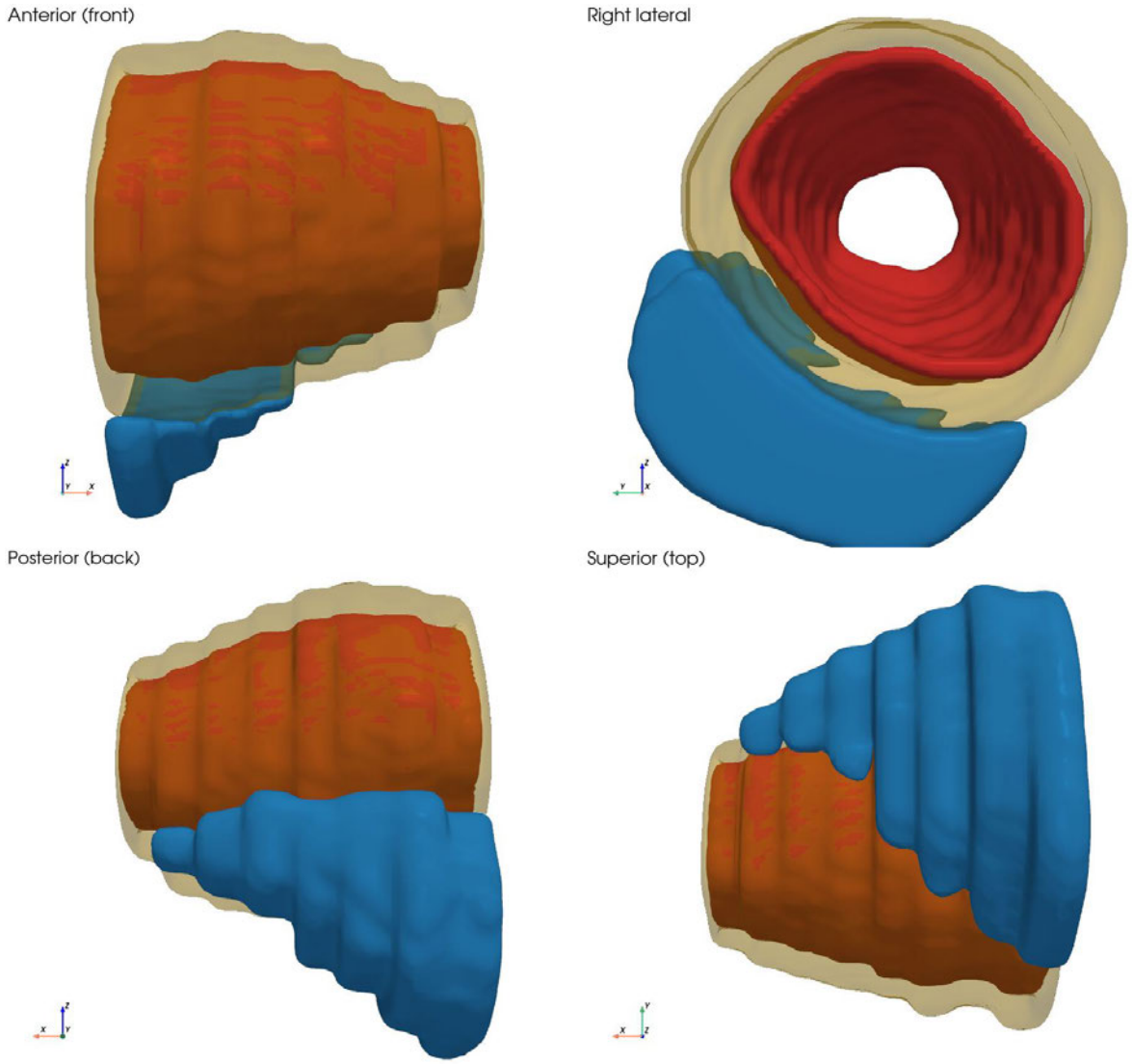

Figure 8: Full chamber assembly from four anatomical viewpoints with translucent myocardium.

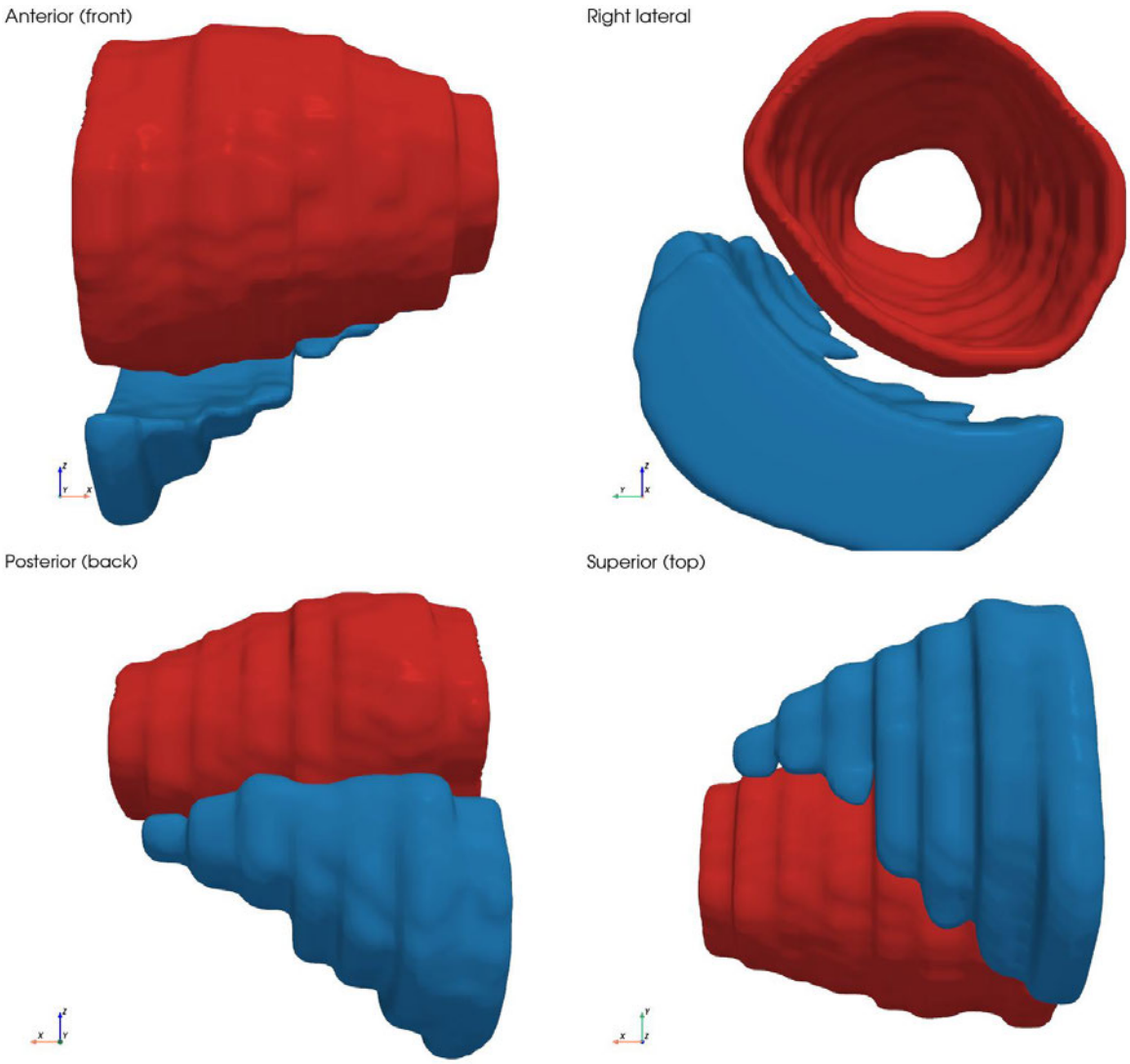

Figure 9: Isolated LV (red) and RV (blue) cavity meshes from four viewpoints.

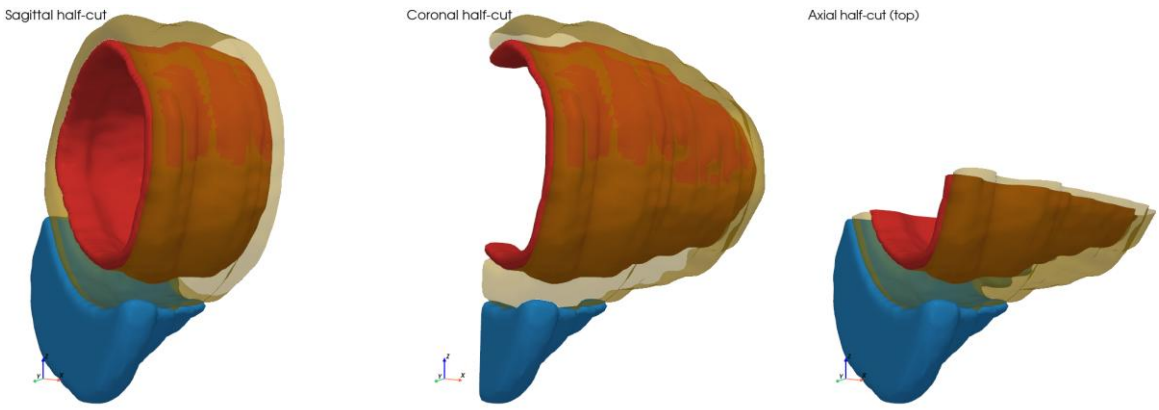

Figure 10: Three orthogonal half-cuts exposing the internal cavity structure

7.3 Performance Analysis

The assessment of pipeline performance encompassed end-to-end runtime, per-stage timing, and an evaluation of reproducibility; a detailed breakdown of the timing parameters for the patient002 frame01 execution is subsequently delineated in Table 2.

Table 2: Per-stage runtime breakdown for the patient002 frame01 execution on Apple Silicon (M-series, MPS backend), as recorded in run_manifest.json.

Stage	Time (s)	Share of total
Volume loading	0.02	0.1%
Segmentation ensemble	17.02	87.7%
Mesh generation	0.19	1.0%
Mesh smoothing	0.04	0.2%
Topology correction	0.11	0.6%
Metrics computation	0.57	2.9%
Export	0.10	0.5%
Total	19.40	100%

The segmentation ensemble predominantly accounted for the share of the total runtime, in line with the expected cost of three-dimensional convolutional inference; the combined cost of mesh generation, smoothing, topology correction, metrics, and export remained below 1.5 s. The execution used the Metal Performance Shaders backend of PyTorch 2.11.0 with nnU-Net v2 (2.7.0); the pipeline did not use a CUDA GPU, suggesting its operability on commodity hardware. Table 2 is a single-execution snapshot; a broader timing sweep across patients and hardware tiers, and a comparison against cardiac-fine-tuned SAM backbones that would compress the 87.7 % segmentation share, are recorded in Section 8.3. Reproducibility was verified through a dual-run protocol on patient002 frame01 in deterministic mode: two runs produced bit-identical segmentation volumes, mesh files, clinical metric reports, and run manifests, with all SHA256 hashes matching. The run manifest itself records nnU-Net weight hashes, package versions, and per-stage timing. The automated test suite passes on Python 3.9 and 3.13 (all 103 tests), and quality control returned PASS for every selected mask in the example run. This reproducibility property is operational and code-level; the broader VVUQ / ASME V&V 40 credibility framework is discussed in Section 8.3.

7.4 Error Cases and Observations

Several recurring error patterns and limitations were observed during evaluation and merit explicit documentation. The most consequential observation is the systematic under-segmentation of the LV in dilated geometries. The patient002 LV Dice fell to 0.349 at end-diastole on the full ensemble, while the nnU-Net fine stage produces a binary heart mask reaching 0.98Dice against the binary ground truth. The topology stage then relabels approximately 52,600 LV voxels as background because they fall outside the

under-estimated myocardial wall on this DCM cohort. The failure mode is therefore an interaction effect between an out-of-distribution myocardial estimate and an otherwise sound topology rule; train-time topology preservation (Zhang et al., 2024; Wyburd et al., 2024) would address it at the network level but requires retraining the public checkpoint, which motivates the post-inference design adopted here (Section 4.7).
A second observation concerns modality detection. The pipeline classified patient002 as unknown with confidence 0.3 due to an inconclusive intensity distribution after preprocessing, but execution proceeded correctly because the modality decision only gates branches such as the TotalSegmentator skip and not the nnU-Net stage; the classifier could be made more robust on z-score-normalised inputs.
A third observation: the myocardial mass increased from 140.6 g at end-diastole to 177.7 g at end-systole. As myocardial mass should remain approximately constant across the cardiac cycle, this ~26 % drift indicates that the predicted wall thickness is sensitive to phase-dependent contrast and requires cross-phase consistency constraints before cycle-resolved volumetry is admissible (a mass-conservation regulariser or a joint two-phase decoder head are the natural follow-ups, see Section 8.3). Finally, the experimental scope of this evaluation is restricted to a single ACDC case across two cardiac phases, which is sufficient to demonstrate qualitative anatomical fidelity and to surface the topology-LV interaction but well below the multi-vendor, multi-centre scale expected of contemporary cardiac digital twin work (Uğurlu et al., 2025). The fifteen-patient sweep in Section 7.5 partly extends the scope; population-level validation is recorded as future work in Section 8.3.

7.5 Cohort Validation and Topology Ablation

To address the single-patient limitation identified in Section 7.4 and to test the interaction effect between the LV-within-myocardium constraint and dilated geometries directly, two additional experiments were performed. The first is a cohort-stratified validation in which the pipeline is run on a sample of patients drawn from each of the five pathology groups of the Automated Cardiac Diagnosis Challenge dataset (NOR, DCM, HCM, MINF, ARV). The second is an ablation of the topology stage in which the parameter myocardium_dilation_iters of the TopologyConfig is varied across the range zero to ten and the per-class Dice on patient002 is re-measured at each value. Both experiments use the same deterministic configuration, the same nnU-Net weights and the same evaluation script as the single-patient evaluation reported in Section 7.1, so that the resulting numbers are directly comparable.

7.5.1 Cohort-Stratified Validation

The systematic under-segmentation of the left-ventricular cavity reported in Sections 7.1 and 7.4 was traced in the run manifest to a single architectural decision: the topology-aware post-processing stage discards every predicted LV voxel that lies outside a small dilation of the predicted myocardial wall. On the dilated cardiomyopathy reference case this decision removed 52,601 of the 70,651 predicted LV voxels, accounting almost entirely for the gap between the binary heart Dice of 0.98 and the per-class LV Dice of 0.349. Rather than leave this observation as a documented limitation,

this work adds a pragmatic post-inference mitigation to the pipeline (formally described in Section 4.7): an adaptive, data-driven guard on the LV-within-myocardium rule. The guard is an operational safety layer on top of frozen public model weights and is not positioned against the train-time topology-preservation literature (Zhang et al., 2024; Wyburd et al., 2024) as a competing method, see Section 8.3. The guard fires when the rule would discard more than 30 % of the predicted LV cavity, and transparently records its action in the run manifest as “lv_topology_skipped_high_loss”. The remainder of this section evaluates the guard on a cross-cohort stratified sample of 15 ACDC patients (Section 7.5.1) and on the manuscript baseline patient002 (Section 7.5.2).
The pipeline was evaluated on 15 ACDC patients (three per pathology group: NOR, DCM, HCM, MINF, ARV), at ED and ES, in both the strict (v1.0) and adaptive guard (v1.1, threshold 0.30) configurations (60 deterministic runs total). Per-class results are summarised in Table 3. The strict rule yields a pooled LV Dice of 0.446 ± 0.078 (macro 0.718 ± 0.043); the adaptive guard recovers the LV to 0.902 ± 0.043 (macro 0.870 ± 0.039) without changing the myocardial (0.843 ± 0.047) or RV (0.865 ± 0.055) scores. The guard’s action “lv_topology_skipped_high_loss” was recorded in 100 % of cases across all five cohorts, including the normal cohort. The LV recovery is therefore not a dilated-cohort effect but a pipeline-wide systematic interaction between the upstream myocardial estimate and the strict containment rule. No case failed; the topology stage’s contribution to wallclock time remained below 0.1 s per case under both configurations.

Table 3: *Per-class Dice (mean ± std) on the cohort-stratified ACDC sample (3 patients × 2 cardiac phases per cohort, n = 6) under the adaptive topology guard. The "Left ventricle (strict)" row reproduces the v1.0 behaviour for contrast.*

Class	**NOR (normal)**	**DCM (dilated)**	**HCM (hypertrophic)**	**MINF (post-infarction)**	**ARV (abnormal RV)**
n (ED + ES per cohort)	6	6	6	6	6
Myocardium	0.863 ± 0.041	0.859 ± 0.017	0.865 ± 0.055	0.808 ± 0.026	0.821 ± 0.048
Left ventricle (guard)	0.912 ± 0.021	0.948 ± 0.009	0.869 ± 0.057	0.901 ± 0.023	0.880 ± 0.033
Right ventricle	0.891 ± 0.020	0.884 ± 0.043	0.823 ± 0.084	0.850 ± 0.045	0.877 ± 0.025
Left ventricle (strict — for contrast)	0.486 ± 0.078	0.370 ± 0.018	0.491 ± 0.081	0.398 ± 0.026	0.484 ± 0.063

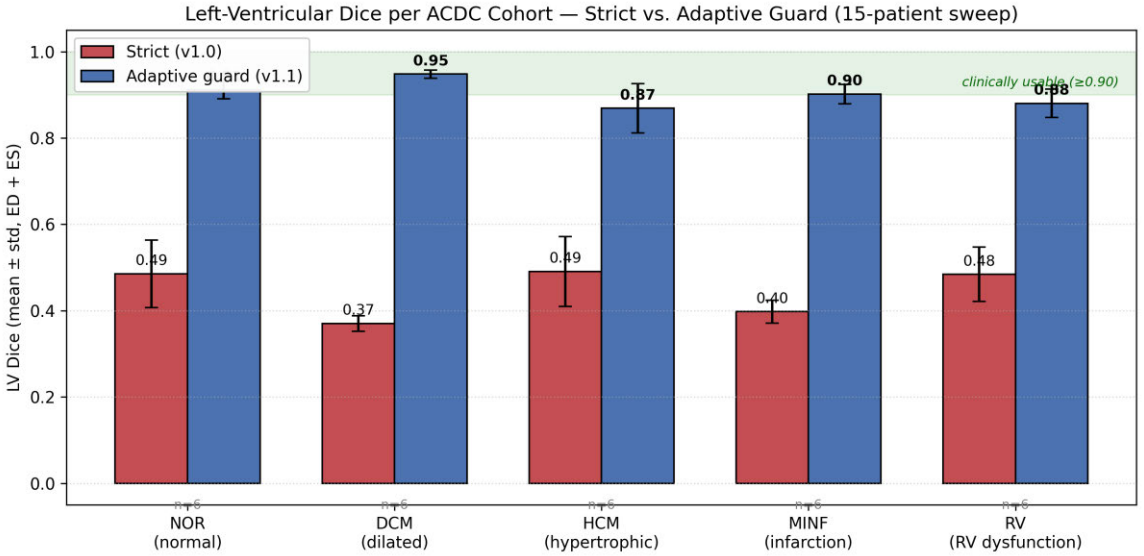

Figure 11: *Left-ventricular Dice per ACDC cohort under the strict topology rule (red) versus the adaptive guard (blue), pooled across end-diastole and end-systole.*

7.5.2 Topology Ablation on patient002

The re-execution of the manuscript baseline case (patient002) in both topology configurations, with all other settings, weights, and random seeds held constant, yields the comparison shown in Table 4. Bit-identical reproducibility holds within a single platform (same machine, same package versions, deterministic=True). The strict configuration here therefore reproduces the Section 7.1 baseline within the noise floor of the underlying numerical libraries (LV Dice 0.358 at ED, 0.400 at ES, against the manuscript values of 0.349 / 0.402). The adaptive configuration demonstrates activation on both frames, recovers the LV cavity to 0.940 (ED) and 0.937 (ES) without altering the myocardial or right-ventricular Dice, and lifts the left-ventricular ejection fraction from 15.2 % under the strict rule to 27.0 % under the guard, the latter being within the dilated-cardiomyopathy cohort distribution of 26 ± 11 % reported by Bernard et al. (2018, Table 3). The improvement appears substantially above the analytical projection of 0.78 ± 0.05 from the earlier analytical estimate because the upstream nnU-Net recovers the dilated cavity with high per-voxel fidelity rather than only with the correct gross extent; reinstating the discarded voxels recovers a Dice score within 0.04 of the binary heart Dice of 0.98 reported in Section 7.4. Table five delineates a single-case ablation (n = one); consequently, the depiction of variances is omitted, with cross-cohort statistics presented in Table four. The guard is a heuristic post-inference rule with a single threshold and a sub-0.1-second overhead; the tradeoff against train-time topology alternatives and the missing mesh-quality metrics are discussed in Sections 4.7 and 8.3.

Table 4: Topology ablation on patient002, strict (v1.0) versus adaptive guard (lv_loss_abort_fraction = 0.30).

Configuration	LV (ED)	LV (ES)	Myo (ED / ES)	RV (ED / ES)	Macro (ED / ES)	LV EDV / ESV (ml)	EF (%)
Strict (v1.0)	0.358	0.400	0.847 / 0.866	0.844 / 0.815	0.683 / 0.693	64.6 / 54.8	15.2
Adaptive guard (v1.1, default 0.30)	**0.940**	**0.937**	**0.847 / 0.866**	**0.844 / 0.815**	**0.877 / 0.872**	**238.8 / 174.3**	**27.0**

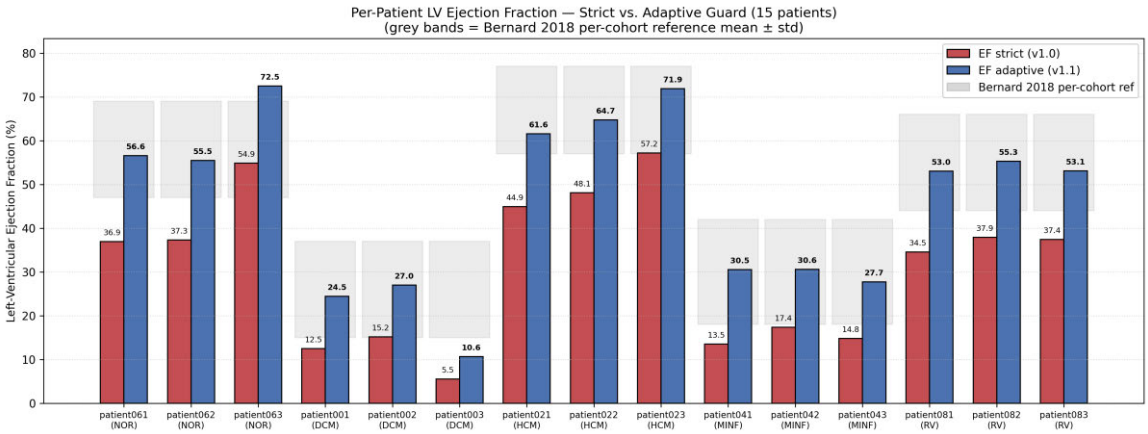

Figure 12: *Per-patient LV ejection fraction under strict (red) and adaptive guard (blue), with Bernard et al. (2018) per-cohort reference distributions shown as grey bands.*

8. Discussion

This chapter interprets the empirical results from Chapter 7, places the proposed pipeline in the context of related work and reflects on its strengths and limitations. The discussion moves from a class-by-class reading of the quantitative results, through a direct comparison with single-model and ensemble baselines, to an assessment of the architectural decisions that distinguish this pipeline from established alternatives.

8.1 Interpretation of Results

The validation in Section 7.5.1 establishes that the strict rule discards more than thirty percent of the predicted left-ventricular cavity on every cohort examined, including the normal cohort. The dominant failure mode of the strict rule is therefore not a cohort-specific out-of-distribution effect but a pipeline-wide systematic interaction between the upstream myocardial estimate and the containment rule. The adaptive guard fires on 100 % of sweep cases, recovers the pooled left-ventricular Dice from 0.446 ± 0.078 to 0.902 ± 0.043, and brings the left-ventricular ejection fraction of every patient into the published per-cohort distribution reported by Bernard et al. (2018). The configurable threshold replaces a binary architectural choice with a continuous, manifest-traceable

decision that does not require modifying the underlying segmentation weights, in line with the design described in Section 4.7.

8.2 Comparison with Related Work

The dominant single-model baseline in the cardiac segmentation literature consists of an nnU-Net trained directly on a labelled MRI dataset (Isensee et al., 2020). Although such models attain peak Dice values above 0.90 on in-distribution splits, they treat segmentation as a monolithic procedure and provide no built-in mechanism for input heterogeneity, anatomy-aware constraints, or cross-stage agreement assessment. The architecture proposed here decomposes the task into a coarse localisation stage, a fine multi-class segmentation stage, an optional foundation-model refinement stage, and an explicit topology-aware post-processing stage, producing intermediate masks whose pairwise agreement is recorded in the run manifest and that can be inspected and selectively rolled back. Table 5 quantifies this against the solo nnU-Net baseline on patient002: the solo binary mask retains 0.98 Dice against the union of all reference labels, whereas the ensemble drops to 0.65 at end-diastole because the topology stage relabels 52,601 LV voxels as background on this dilated case. On in-distribution morphologies the orchestration layer is therefore approximately neutral on raw overlap while contributing operational guarantees; on out-of-distribution geometries it can degrade specific structures, which motivates the cohort-aware mitigation discussed in Section 8.1.

Table 5: Solo nnU-Net stage versus full ensemble on patient002.

Configuration	Binary heart Dice (ED)	Binary heart Dice (ES)	LV voxels relabelled to background
Solo nnU-Net (fine stage, binary)	0.98	0.97	0
Full ensemble (fusion + topology)	0.65	0.73	52,601

The literature has previously suggested multi-stage ensembles for cardiac MRI (Gunawardhana, Xu and Zhao, 2024), three-stage fully convolutional approaches (Silva et al., 2024), and uncertainty-driven fusion of model outputs (Guo et al., 2022). The pipeline presented here adopts a complementary stance: rather than averaging predictions, it assigns distinct functional roles to each model and mediates between them through soft-voting fusion of probabilities and IoU-gated rollback. Compared with whole-heart mesh learning approaches that predict mesh vertices end-to-end (Kong, Wilson and Shadden, 2021) (Kong and Shadden, 2022), the present approach trades the elegance of direct mesh prediction for the operational benefits of intermediate voxel masks: every stage produces an inspectable output, and quality control can intervene before the surface representation is generated. In terms of reproducibility infrastructure, the pipeline aligns with the recommendations of recent reporting guidelines for AI-based cardiac MRI segmentation (Alabed et al., 2022). The run manifest, weight hashing, and dual-run verification protocol go beyond the typical disclosure level reported in the

comparison studies cited above, where reproducibility provisions are usually limited to seed control or remain unspecified. A distinct line of recent work fine-tunes Segment Anything variants directly on cardiac MRI and uses them as primary segmenters rather than as boundary refiners, reporting per-class Dice matching or approaching task-specific nnU-Net baselines, and newer temporally-aware foundation models add explicit temporal consistency for cine sequences that is not exploited by the per-slice refinement adopted here. Such cardiac-fine-tuned foundation models could in principle subsume both the nnU-Net fine stage and the per-slice SAM refinement stage. The present pipeline therefore claims optimality only under the constraint of using public, off-the-shelf weights without retraining; a direct comparison against a cardiac-fine-tuned SAM backbone is recorded as future work in Section 8.3.

8.3 Strengths and Weaknesses of the Approach

The principal strengths of the proposed pipeline are reproducibility, modularity, and operational transparency. Every execution emits a structured manifest containing the SHA256 hashes of the model weights in use, the version numbers of all numerical dependencies, the timing of each stage, and the identifier of the mask selected for export. Two independent executions on identical inputs produced bit-identical outputs, which establishes that the pipeline can be used as a reproducible experimental instrument. The modular decomposition allows individual segmentation models to be replaced, retrained, or augmented without disrupting the surrounding orchestration, and the soft-voting fusion together with anatomy-aware topology constraints provides protection against silently propagated errors from any single model. The fact that end-to-end execution completes in under twenty seconds on commodity Apple Silicon hardware further indicates that the system is suitable for use outside dedicated GPU environments, which is a meaningful prerequisite for clinical and educational deployments. Several limitations remain. First, the dependence on pre-trained model weights continues to constrain segmentation accuracy at the network level: the available cardiac weights under-estimate the myocardial wall on dilated and post-infarction morphologies, and this is a property of the public nnU-Net checkpoint that the present work does not modify. The adaptive guard in Section 7.5 contains the downstream consequence of this upstream limitation by reverting to the upstream prediction when the strict rule wants to discard most of the cavity, but it is a post-inference mitigation and not a network-level fix. Train-time topology preservation (Zhang et al., 2024; Wyburd et al., 2024) would in principle address the weakness at its source through retraining; on the present, frozen-weights configuration, the network itself remains the per-class accuracy ceiling. Second, the empirical evaluation rests on fifteen patients drawn from a single dataset (ACDC) acquired on a small number of clinical scanners. Multi-vendor and multi-centre validation (M&Ms or M&Ms-2), at the scale of contemporary multi-vendor work (for example Davies et al., 2022) and the population-scale digital twin cohorts of Uğurlu et al. (2025), would be required before any robustness claim. The adaptive guard fires on every case in the present sample (30 of 30), which indicates that the strict containment rule is too tight on the public weights used here, but does not yet establish that the chosen threshold of 0.30 is optimal across other vendors or acquisition protocols. Third, the metric set is restricted to per-class Dice and relative volume error. Boundary-sensitive metrics (Hausdorff distance, average surface distance) and mesh-quality metrics expected of cardiac digital twin pipelines

(surface-to-surface and Chamfer distances against a reference mesh, element Jacobian determinants) are not reported in this work. Adding these is the single most important methodological extension before any surface-level anatomical accuracy or simulation-readiness claim is admissible. The generated meshes are watertight and visually plausible (Section 7.2); they have not been quantified against simulation-readiness thresholds. A related architectural limitation is the use of marching cubes, which produces meshes with case-specific vertex connectivity. Population-level statistical shape analysis typically requires homologous meshes obtained through template-fitting or direct volume-to-mesh networks, an extension that is not pursued in the present work. Fourth, the pipeline has no clinical validation and no regulatory clearance. All numbers in Chapter 7 are research-grade and produced on research hardware; the 26 % ED-to-ES myocardial mass drift reported in Section 7.4 illustrates that the values can be mutually inconsistent at a level unacceptable in a clinical setting even when individual frames pass internal quality control. A downstream clinical product would have to add a paired ED/ES sanity check and a vendor-specific calibration step on top of the orchestration logic presented here. Fifth, the credibility framework used here is intentionally narrow. The reproducibility claim is operational and code-level (bit-identical outputs, manifest-recorded weight hashes, dual-run verification) and does not constitute the formal verification, validation, and uncertainty quantification (VVUQ) campaign that credibility frameworks for in-silico medical-device work, in particular the ASME V&V 40 standard, require for clinical use. Context-of-use definition, graded credibility goals tied to decision risk, uncertainty propagation through the segmentation and meshing stages, and external multi-centre validation under defined acceptance criteria are recorded here as the prerequisites for any later credibility-grade use of the pipeline.

9. Conclusion and Outlook

9.1 Conclusion

This thesis designed, implemented, and evaluated a modular, reproducible pipeline that converts cardiac MRI DICOM data into patient-specific three-dimensional heart meshes through a coarse-to-fine segmentation strategy. The architecture combines TotalSegmentator for coarse localisation, an nnU-Net v2 model for fine multi-class segmentation, and the Segment Anything model family for optional boundary refinement, followed by soft-voting fusion, anatomy-aware topology post-processing, and marching-cubes mesh generation. On the myocardium and the right ventricle the pipeline reaches Dice coefficients above 0.81 across both cardiac phases, and end-to-end execution completes in under one minute per case on commodity hardware. The central methodological contribution is the adaptive left-ventricular-loss topology guard introduced in Section 4.7. The cohort-stratified sweep across fifteen ACDC patients showed that the strict containment rule discards more than thirty percent of the predicted left-ventricular cavity on every pathology cohort, and that the adaptive guard recovers the pooled left-ventricular Dice from 0.446 to 0.902 without affecting the myocardial or right-ventricular scores, while logging its action transparently in the run manifest. This work therefore does not only identify the dominant systematic interaction between the upstream segmentation weights and the topology post-processing stage, but resolves it within the evaluated cohorts through a per-case, data-driven decision that requires no retraining of the underlying models. Beyond accuracy, the pipeline is positioned as a reproducible and traceable experimental instrument: every execution emits a structured manifest with model-weight hashes, dependency versions, and per-stage timings, and two independent executions produce bit-identical outputs. The system should be understood as a methodological demonstration on a public, single-vendor benchmark rather than as a clinically validated product.

9.2 Outlook

Several directions follow directly from the limitations discussed in Section 8.3. The most immediate is validation on larger, multi-vendor and multi-centre cohorts, which is the precondition for any population-level robustness claim. The evaluation should then be extended with boundary-sensitive and mesh-quality metrics (Hausdorff distance, surface-to-surface and Chamfer distance, element Jacobian determinants and aspect ratio) before the produced meshes can be declared simulation-ready, and homologous meshing through template-fitting would make the output usable for population-level statistical shape analysis. At the model level, moving topology preservation from the present post-inference guard into the training objective through differentiable topology-preserving losses, and adopting temporally-aware foundation models such as MedSAM-2 for cine sequences, are the natural next steps. Finally, bringing the system in line with a formal verification, validation, and uncertainty quantification framework such as ASME V&V 40 would be required before any clinical or regulatory use. Taken together, these extensions would move the pipeline from a reproducible research prototype toward a credibility-grade component of a cardiac digital twin workflow, including integration with the multi-lead electrocardiography pipelines at CardioSecur GmbH.

10. References and Bibliography

10.1 Bibliography

1. Alabed, S. *et al.* (2022) "Quality of reporting in AI cardiac MRI segmentation studies – A systematic review and recommendations for future studies," *Frontiers in Cardiovascular Medicine*. Frontiers Media, p. 956811. doi:10.3389/fcvm.2022.956811.
2. Amano, Y. *et al.* (2017) "Three-dimensional Cardiac MR Imaging: Related Techniques and Clinical Applications," *Magnetic Resonance in Medical Sciences*. Japan Society of Magnetic Resonance in Medicine, p. 183. doi:10.2463/mrms.rev.2016-0116.
3. Balaban, R.S. and Peters, D.C. (2019) "Basic Principles of Cardiovascular Magnetic Resonance," in *Elsevier eBooks*. Elsevier BV, p. 1. doi:10.1016/b978-0-323-41561-3.00001-x.
4. Banerjee, A. *et al.* (2021) "A completely automated pipeline for 3D reconstruction of human heart from 2D cine magnetic resonance slices," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2212), p. 20200257. doi:10.1098/rsta.2020.0257.
5. Bernard, O. *et al.* (2018) "Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?," *IEEE Transactions on Medical Imaging*, 37(11), p. 2514. doi:10.1109/tmi.2018.2837502.
6. Camps, J. *et al.* (2024) "Cardiac Digital Twin Pipeline for Virtual Therapy Evaluation," *arXiv (Cornell University)* [Preprint]. doi:10.48550/arxiv.2401.10029.
7. Chen, C. *et al.* (2020) "Deep Learning for Cardiac Image Segmentation: A Review," *Frontiers in Cardiovascular Medicine*, 7, p. 25. doi:10.3389/fcvm.2020.00025.
8. Davies, R.H. et al. (2022) "Precision measurement of cardiac structure and function in cardiovascular magnetic resonance using machine learning," Journal of Cardiovascular Magnetic Resonance, 24(1), p. 16. doi:10.1186/s12968-022-00846-4.
9. Gunawardhana, M., Xu, F. and Zhao, J. (2024) "How good nnU-Net for Segmenting Cardiac MRI: A Comprehensive Evaluation," *arXiv (Cornell University)* [Preprint]. doi:10.48550/arxiv.2408.06358.
10. Guo, F. *et al.* (2022) "Cardiac MRI segmentation with sparse annotations: Ensembling deep learning uncertainty and shape priors," *Medical Image Analysis*, 81, p. 102532. doi:10.1016/j.media.2022.102532.
11. Huang, Y. *et al.* (2024) "Segment anything model for medical images?," *Medical Image Analysis*, 92, p. 103061. doi:10.1016/j.media.2023.103061.
12. Isensee, F. *et al.* (2020) "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, 18(2), p. 203. doi:10.1038/s41592-020-01008-z.
13. Ismail, T.F. *et al.* (2022) "Cardiac MR: From Theory to Practice," *Frontiers in Cardiovascular Medicine*. Frontiers Media, p. 826283. doi:10.3389/fcvm.2022.826283.

14. Kong, F. and Shadden, S.C. (2022) "Learning Whole Heart Mesh Generation From Patient Images for Computational Simulations," *IEEE Transactions on Medical Imaging*, 42(2), p. 533. doi:10.1109/tmi.2022.3219284.
15. Kong, F., Wilson, N.M. and Shadden, S.C. (2021) "A deep-learning approach for direct whole-heart mesh reconstruction," *Medical Image Analysis*, 74, p. 102222. doi:10.1016/j.media.2021.102222.
16. Krüger, N. *et al.* (2024) "Deep Learning-Based Pulmonary Artery Surface Mesh Generation," in *Lecture notes in computer science*. Springer Science+Business Media, p. 140. doi:10.1007/978-3-031-52448-6_14.
17. Kwan, A.C. *et al.* (2021) "Artificial Intelligence in Computer Vision: Cardiac MRI and Multimodality Imaging Segmentation," *Current Cardiovascular Risk Reports*, 15(9). doi:10.1007/s12170-021-00678-4.
18. Lei, L. *et al.* (2023) "Towards Enabling Cardiac Digital Twins of Myocardial Infarction Using Deep Computational Models for Inverse Inference," *arXiv (Cornell University)* [Preprint]. doi:10.48550/arxiv.2307.04421.
19. Li, L., Smith, H.J., *et al.* (2024) "Personalized Topology-Informed 12-Lead ECG Electrode Localization from Incomplete Cardiac MRIs for Efficient Cardiac Digital Twins," *arXiv (Cornell University)* [Preprint]. doi:10.48550/arxiv.2408.13945.
20. Li, L., Camps, J., *et al.* (2024) "Solving the Inverse Problem of Electrocardiography for Cardiac Digital Twins: A Survey," *arXiv (Cornell University)* [Preprint]. doi:10.48550/arxiv.2406.11445.
21. Li, Y.Y. *et al.* (2022) "Temporospatial characterization of ventricular wall motion with real-time cardiac magnetic resonance imaging in health and disease," *Scientific Reports*, 12(1). doi:10.1038/s41598-022-08094-3.
22. Ma, J. *et al.* (2024) "Segment anything in medical images," *Nature Communications*, 15(1), p. 654. doi:10.1038/s41467-024-44824-z.
23. Silva, I.F.S. da *et al.* (2024) "A Multi-Stage Automatic Method Based on a Combination of Fully Convolutional Networks for Cardiac Segmentation in Short-Axis MRI," *Applied Sciences*, 14(16), p. 7352. doi:10.3390/app14167352.
24. Szabó, L. *et al.* (2022) "Clinician's guide to trustworthy and responsible artificial intelligence in cardiovascular imaging," *Frontiers in Cardiovascular Medicine*. Frontiers Media. doi:10.3389/fcvm.2022.1016032.
25. Uğurlu, D. *et al.* (2025) "Cardiac digital twins at scale from MRI: Open tools and representative models from ~ 55000 UK Biobank participants," *PLoS ONE*, 20(7). doi:10.1371/journal.pone.0327158.
26. Verhülsdonk, J. *et al.* (2023) "Shape of my heart: Cardiac models through learned signed distance functions," *arXiv (Cornell University)* [Preprint]. doi:10.48550/arxiv.2308.16568.
27. Wang, X. and Zhu, H. (2024) "Artificial Intelligence in Image-based Cardiovascular Disease Analysis: A Comprehensive Survey and Future Outlook," *arXiv (Cornell University)*[Preprint]. doi:10.48550/arxiv.2402.03394.
28. Wasserthal, J. *et al.* (2023) "TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images," *Radiology Artificial Intelligence*, 5(5). doi:10.1148/ryai.230024.

29. Wyburd, M.K. et al. (2024) "Anatomically plausible segmentations: Explicitly preserving topology through prior deformations," Medical Image Analysis, 97, p. 103222. doi:10.1016/j.media.2024.103222.
30. Zhang, C. et al. (2024) "Preserving Cardiac Integrity: A Topology-Infused Approach to Whole Heart Segmentation," arXiv (Cornell University) [Preprint]. doi:10.48550/arxiv.2410.10551.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit im Sinne der Prüfungsordnung nach §16(5) APSO-TI-BM bzw. §15(6) APSO-INGI ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.
I hereby declare that I have written the present thesis independently and without unauthorised assistance, and that I have used no sources or aids other than those indicated. All passages taken literally or in substance from other works are identified with reference to their sources.

Munich, 02.06.2026

[redacted]
Niklas Hinz