REPUBLIC OF TÜRKİYE
YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

LOCALIZATION-GUIDED FEDERATED LEARNING FOR MULTI-CENTER 3D SEGMENTATION OF PEDIATRIC CARDIOVASCULAR STRUCTURES

İrem YILMAZ

MASTER OF SCIENCE THESIS

Department of Biomedical Engineering

Program of Biomedical Engineering

Supervisor

Assoc. Prof. Dr. Sakip ÖNDER

Co-Supervisor

Assist. Prof. Dr. Şenol PİŞKİN

June, 2026

REPUBLIC OF TÜRKİYE

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

LOCALIZATION-GUIDED FEDERATED LEARNING FOR MULTI-CENTER 3D SEGMENTATION OF PEDIATRIC CARDIOVASCULAR STRUCTURES

A thesis submitted by İrem YILMAZ in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE** is approved by the committee on 05/06/2026 in Department of Biomedical Engineering, Program of Biomedical Engineering.

Assoc. Prof. Dr. Sakip ÖNDER
Yildiz Technical University
Supervisor

Assist. Prof. Dr. Şenol PİŞKİN
Istinye University
Co- supervisor

Approved By the Examining Committee

Assoc. Prof. Dr. Sakip ÖNDER, Supervisor
Yildiz Technical University _____

Assoc. Prof. Dr. Ali Can KARACA, Member
Yildiz Technical University _____

Assist. Prof. Dr. Femilda Josephin JOSEPH SHOBANA BAI, Member
Istinye University _____

I hereby declare that the study entitled “Localization-Guided Federated Learning for Multi-Center 3D Segmentation of Pediatric Cardiovascular Structures” that I prepared under the supervision of Assoc. Prof. Dr. Sakip ÖNDER was conducted in full compliance with the relevant legal requirements regarding data collection and data use; that all information obtained from other sources has been fully and properly acknowledged in the main text and references; that no fabrication and/or falsification has been made with respect to the research data or findings; and that throughout the study I have adhered to the principles of scientific research and academic ethics.

In the course of this thesis study, Generative Artificial Intelligence (GenAI) tools—specifically Grammarly—were utilized under the author’s supervision for purposes such as enhancing linguistic clarity and coherence, supporting academic writing, and improving the overall organization of the text.

I declare that I am aware of the risk that GenAI tools may generate inaccurate or fabricated information (hallucinations), and that I have carefully reviewed and verified all outputs obtained from these tools in terms of accuracy, integrity, reliability, and compliance with ethical standards.

The use of GenAI tools has not, at any stage, replaced the author’s intellectual contribution, scientific judgment, or original idea generation. Full responsibility for the accuracy, currency, and scientific validity of all information, data, analyses, claims, and references presented in this thesis rests entirely with me.

I accept full administrative and legal responsibility for any consequences that may arise if this declaration is determined to be false.

Irem YILMAZ

Signature

This work was supported by the European Research Executive Agency (REA) under the Marie Skłodowska-Curie Actions (Project No. 101038096) and by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under the 1001 Research Grant Program (Project No. 122E470).

Dedicated to my family

and my best friend

ACKNOWLEDGEMENTS

Science is the process of seeking the truth.

Aziz SANCAR

Scientific progress has always played a fundamental role in addressing the challenges faced by humanity. In particular, recent advances in artificial intelligence and medical imaging have enabled the development of more accurate, efficient, and reliable solutions in the healthcare domain. In this context, engineering disciplines have taken on significant responsibilities not only in technical development but also in solving problems that directly impact human life.

Throughout my educational journey, my interest in engineering gradually evolved into a research-oriented perspective, ultimately culminating in this thesis study. This process has not only enhanced my technical knowledge but has also significantly contributed to the development of my analytical thinking, problem-solving, and scientific research skills.

I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Sakip ÖNDER, for his guidance, support, and valuable contributions throughout this study. I would also like to thank my co-supervisor, Assist. Prof. Dr. Şenol PİŞKİN, for his support and valuable insights during the course of this research.

I am deeply grateful to my family for their unconditional support and encouragement throughout my life. Finally, I would like to thank my friends for their support during this process.

İrem YILMAZ

TABLE OF CONTENTS

LIST OF SYMBOLS x
LIST OF ABBREVIATIONS xiii
LIST OF FIGURES xv
LIST OF TABLES xix
ABSTRACT xxi
ÖZET xxiii
1 INTRODUCTION 1
1.1 Literature Review 4
1.1.1 Clinical Importance of Pediatric Cardiovascular Segmentation 4
1.1.2 Anatomical and Small-Vessel Challenges 5
1.1.3 Evaluation Metrics and Clinical Relevance 6
1.1.4 Multi-Center Variability and Domain Shift 6
1.1.5 Deep Learning and Federated Learning Approaches 7
1.1.6 Current Limitations and Research Gaps 8
1.2 Objectives of the Thesis 9
1.3 Hypothesis 9
2 CLINICAL AND TECHNICAL BACKGROUND 11
2.1 Clinical and Anatomical Background 11
2.1.1 Pediatric Cardiovascular Anatomy 12
2.1.2 Congenital Heart Disease (CHD) and Anatomical Variability 14
2.1.3 Computed Tomography and Magnetic Resonance Imaging 15
2.2 Imaging Variability in Multi-Center Medical Data 17
2.3 Problem Formulation of 3D Medical Image Segmentation 19
2.3.1 Problem Formulation 20
2.3.2 Fundamentals of 3D Medical Image Segmentation 21
2.4 Deep Learning–Based Segmentation Architectures 23
2.4.1 Classical Segmentation Methods 25
2.4.2 CNN-Based Segmentation 26

2.4.3 Transformer and Hybrid Architectures 28
2.4.4 Self-Configuring Frameworks 30
2.4.5 Localization-Guided Segmentation 31
2.4.6 Previous Studies on Pediatric Cardiovascular Segmentation 32
2.5 Challenges in Small-Vessel Segmentation in Medical Imaging 33
2.6 Evaluation Metrics for Medical Image Segmentation 35
2.7 Domain Shift and Generalization in Medical Imaging 39
2.8 Learning Strategies for Multi-Center Medical Imaging 41
2.8.1 Domain Adaptation and Generalization 42
2.8.2 Federated Learning 43
3 METHODOLOGY **46**
3.1 Overall Framework 46
3.2 Dataset Description and Organization 48
3.2.1 Public Benchmark Dataset (HVSMR-2.0) 49
3.2.2 Multi-Center Clinical Dataset 50
3.2.3 Combined Multi-Center Dataset Configuration 53
3.2.4 Annotation Protocol and Quality Control 55
3.3 Image Preprocessing Pipeline 56
3.3.1 Image Standardization 57
3.3.2 Spatial Resampling and Resolution Harmonization 58
3.3.3 Intensity Normalization 59
3.3.4 Segmentation Mask Preparation 61
3.4 Landmark Localization and ROI Extraction 62
3.4.1 Rationale for Localization-Guided Segmentation 63
3.4.2 Definition of Anatomical Landmarks 64
3.4.3 Landmark Localization Network 65
3.4.4 ROI Generation Strategy 67
3.5 Segmentation Architectures 69
3.5.1 Proposed Hybrid V-Net with Transformer 69
3.6 Proposed Localization-Guided Segmentation Framework 70
3.6.1 Network Architecture Design 72
3.6.2 Multi-Class Output and Structure-Specific Learning Strategy 74
3.6.3 Loss Functions and Optimization 76
3.6.4 Inference and Post-Processing Strategy 77
3.7 Federated Learning Framework 78

3.7.1 Client–Server Architecture and Training Strategy 79
3.7.2 Silo-Based Data Partitioning and Privacy Considerations 82
3.8 Implementation Details 82
3.9 Evaluation Metrics and Analysis Strategy 85
3.9.1 Overlap- and Boundary-Based Metrics 86
3.9.2 Multi-Level Evaluation Strategy 86
3.9.3 Statistical Analysis 87
3.10 Supplementary Framework Components 87
3.10.1. Ensemble-Based Inference Strategy 87
3.10.2. Explainability and Model Interpretation 88
4 EXPERIMENTAL RESULTS AND DISCUSSION 90
4.1 Evaluation Setup and Protocol 90
4.1.1 Dataset Configurations 91
4.1.2 Implementation Environment and Training Settings 91
4.1.3 Evaluation Scenarios 92
4.2 Results on the HVSMR-2.0 Benchmark Dataset 93
4.2.1 Baseline Segmentation Performance Without Localization 94
4.2.2 Effect of Localization-Guided ROI Extraction 95
4.2.3 Federated Learning Performance Without Localization 96
4.2.4 Combined Effect of Localization and Federated Learning 97
4.3 Results on the Multi-Center Clinical Dataset 98
4.3.1 Baseline Segmentation Performance Without Localization 98
4.3.2 Effect of Localization-Guided ROI Extraction 99
4.3.3 Federated Learning Performance Without Localization 101
4.3.4 Combined Effect of Localization and Federated Learning 102
4.4 Results on the Combined Dataset 103
4.4.1 Baseline Segmentation Performance Without Localization 103
4.4.2 Effect of Localization-Guided ROI Extraction 105
4.4.3 Federated Learning Performance Without Localization 106
4.4.4 Combined Effect of Localization and Federated Learning 107
4.4.5 Qualitative Results 108
4.5 Comparative Evaluation of Segmentation Architectures 109
4.5.1 Overall Performance Comparison 110
4.5.2 Structure-Specific Analysis of Pulmonary Artery Segmentation 112
4.6 Cross-Dataset ROI Localization Analysis 114

4.6.1 Quantitative Impact 114
4.6.2 Qualitative Impact 115
4.7 Federated Learning Results 116
4.7.1 Federated vs Centralized Performance Comparison 116
4.7.2 Client-Wise Cross-Center Analysis 117
4.8 Discussion of Key Findings 118
4.9 Additional Analyses and Validation 121
4.9.1 Statistical Significance and Boundary Analysis 121
4.9.2 Explainability Analyses 123
5 CONCLUSION **127**
REFERENCES **130**
A SUPPLEMENTARY FEDERATED LEARNING, ENSEMBLE, AND EXPLAINABILITY ANALYSES **139**
PUBLICATIONS FROM THE THESIS **144**

LIST OF SYMBOLS

$HD95$	95th percentile Hausdorff Distance
l_i	Anatomical landmark coordinate
$ASSD$	Average Symmetric Surface Distance
Y^{PA}	Binary pulmonary artery mask
F_{cnn}	CNN feature representation
$\mathcal{L}_{CE}$	Cross-entropy loss
β	Cross-entropy loss weighting coefficient
$\mathcal{L}_{Dice}$	Dice loss
α	Dice loss weighting coefficient
DSC	Dice Similarity Coefficient
$d(a, b)$	Distance between surface points a and b
t	Federated learning communication round index
F_{fused}	Fused feature representation
$\theta^{(t)}$	Global model parameters at communication round t
Y	Ground-truth segmentation mask or label map
B	Ground-truth segmentation mask or surface set
$A \cap B$	Intersection between predicted and ground-truth segmentations
IoU	Intersection over Union

K_{att}	Key matrix in transformer attention
$H_i(x, y, z)$	Landmark heatmap
$\mathcal{L}_{loc}$	Landmark localization loss
η	Learning rate
$\mathcal{D}_k$	Local dataset of client k
$Z_c(x, y, z)$	Logit value for class c
E_{local}	Number of local training epochs
n_k	Number of training samples at client k
$\hat{l}_i$	Predicted landmark coordinate
$\hat{Y}$	Predicted segmentation mask
A	Predicted segmentation mask or surface set
$\mathcal{L}_{PA}$	Pulmonary artery auxiliary loss
λ_{PA}	Pulmonary artery auxiliary loss weight
$P^{PA}(x, y, z)$	Pulmonary artery probability map
Q	Query matrix
$\mathbb{R}$	Real-valued space
I_{ROI}	Region-of-interest image volume
Ω_{ROI}	Region-of-interest spatial domain
c	ROI center coordinate
f_θ	Segmentation model parameterized by θ
$P_c(x, y, z)$	Softmax probability for class c
Ω	Spatial image domain

I	Three-dimensional input medical image
$\mathcal{L}_{total}$	Total optimization loss
θ	Trainable model parameters
F_{trans}	Transformer feature representation
θ_k^{t+1}	Updated client model parameters
$A \cup B$	Union between predicted and ground-truth segmentations
V	Value matrix
v	Voxel location in the image domain

LIST OF ABBREVIATIONS

3D	Three-Dimensional
AI	Artificial Intelligence
AR	Aortic Root
ASD	Average Surface Distance
ASSD	Average Symmetric Surface Distance
CHD	Congenital Heart Disease
CI	Confidence Interval
CMR	Cardiovascular Magnetic Resonance
CNN	Convolutional Neural Network
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DSC	Dice Similarity Coefficient
FedAdam	Federated Adaptive Moment Optimization
FedAvg	Federated Averaging
FedProx	Federated Proximal Optimization
FedYogi	Federated Yogi Optimization
FG	Foreground
FL	Federated Learning
Grad-CAM	Gradient-weighted Class Activation Mapping
HD	Hausdorff Distance
HD95	95th Percentile Hausdorff Distance
HU	Hounsfield Unit

HVSMR	Hybrid Whole Heart Segmentation in Multi-Modality Images Challenge
IG	Integrated Gradients
IoU	Intersection over Union
IQR	Interquartile Range
LPA	Left Pulmonary Artery
LV	Left Ventricle
MASD	Mean Average Surface Distance
MONAI	Medical Open Network for Artificial Intelligence
MM-WHS	Multi-Modality Whole Heart Segmentation
MPA	Main Pulmonary Artery
MPR	Multiplanar Reconstruction
MRI	Magnetic Resonance Imaging
NIfTI	Neuroimaging Informatics Technology Initiative
nnU-Net	Self-configuring U-Net-based Segmentation Framework
non-IID	Non-Independent and Identically Distributed
NRRD	Nearly Raw Raster Data
PA	Pulmonary Artery
PAB	Pulmonary Artery Bifurcation
PyTorch	Python-based Deep Learning Framework
ROI	Region of Interest
RPA	Right Pulmonary Artery
Swin-UNet	Shifted Window Transformer U-Net
TTA	Test-Time Augmentation
U-Net	Convolutional Encoder–Decoder Segmentation Network
V-Net	Volumetric Convolutional Neural Network
XAI	Explainable Artificial Intelligence

LIST OF FIGURES

Figure 1.1 Representative cardiac MRI, CT, and three-dimensional cardiac rendering from the MM-WHS dataset [3] **1**

Figure 2.1 Representative cardiovascular structures in cardiac CT and MRI. (a) Whole-heart CT segmentation. (b) MRI segmentation overlays [36] **12**

Figure 2.2 Pulmonary arterial tree anatomy with main pulmonary artery bifurcation and distal vessel branching [37], [38] **13**

Figure 2.3 Representative congenital heart disease cases in cardiac CT with 3D reconstructions. (A–D) MPR views (E–H) 3D visualizations [42] **14**

Figure 2.4 Representative cardiac MRI and CT images across anatomical planes. (a) MRI with soft-tissue contrast. (b) CT with vascular visualization [3] **16**

Figure 2.5 Cardiac MRI images from different scanner vendors illustrating inter-vendor variability in contrast and image quality [51] **17**

Figure 2.6 Reconstruction workflow in clinical and research imaging pipelines including acquisition, preprocessing, and reconstruction stages [54]. **18**

Figure 2.7 Conceptual workflow of three-dimensional multi-class segmentation from volumetric CT and MRI data **21**

Figure 2.8 Illustration of isotropic and anisotropic voxel grids and the effect of resampling on three-dimensional image representation **22**

Figure 2.9 Effect of slice thickness on spatial resolution and anatomical detail in CT imaging [62] **22**

Figure 2.10 Encoder–decoder architecture for medical image segmentation [63] **23**

Figure 2.11 Evolution of deep learning–based medical image segmentation architectures from U-Net to transformer-based approaches (2015–2025) **24**

Figure 2.12 Taxonomy of medical image segmentation methods from classical techniques to deep learning–based approaches **24**

Figure 2.13 Multi-atlas segmentation in cardiac CT angiography. (a) Atlas registration. (b) Label fusion. (c) Final segmentation [70] **25**

Figure 2.14 Illustration of level-set–based segmentation in cardiac MRI with iterative contour evolution toward anatomical boundaries [73] **26**

Figure 2.15 U-Net architecture for biomedical image segmentation with encoder–decoder structure and skip connections [8] **27**

Figure 2.16 V-Net architecture for volumetric medical image segmentation with three-dimensional convolutions and residual connections [9] **28**

Figure 2.17 Comparison of convolutional neural networks and transformer-based architectures for local and global feature modeling **29**

Figure 2.18 Swin-UNet architecture combining transformer-based feature extraction with encoder–decoder segmentation [82] **29**

Figure 2.19 Overview of the nnU-Net self-configuring framework for automated preprocessing, architecture selection, and training configuration [85] **30**

Figure 2.20 Localization-guided segmentation in cardiac CT. (A–C) Volumetric images. (D) ROI localization. (E) Cropped ROI segmentation [89] **31**

Figure 2.21 Illustration of the partial volume effect causing blurred boundaries and reduced segmentation accuracy [91] **34**

Figure 2.22 Challenges in small-vessel segmentation. (a) Baseline segmentation. (b) Fragmentation. (c) Boundary ambiguity. (d) Missing thin structures [94] **34**

Figure 2.23 Illustration of overlap-based segmentation metrics: (a) Dice Similarity Coefficient (DSC) and (b) Intersection over Union (IoU) [98] **36**

Figure 2.24 Illustration of distance-based metrics: (a) Hausdorff Distance (HD) and (b) 95th percentile Hausdorff Distance (HD95) [98] **37**

Figure 2.25 Illustration of surface-based metrics: (a) Average Symmetric Surface Distance (ASSD) and (b) Mean Average Surface Distance (MASD) [98] **38**

Figure 2.26 Conceptual illustration of domain shift between source and target domains and the role of domain adaptation [101] **40**

Figure 2.27 Example of intensity distribution differences across imaging centers, illustrating inter-center variability and domain shift [101] **40**

Figure 2.28 Conceptual illustration of federated learning with collaborative training across multiple institutions without raw data sharing [104] **42**

Figure 2.29 Domain adaptation and generalization. (a) Source-to-target adaptation. (b) Multi-domain generalization. (c) Single-domain generalization [109] **43**

Figure 2.30 Federated learning workflow with global model distribution, local training, and parameter aggregation [111] **44**

Figure 3.1 Overview of the proposed localization-guided segmentation framework **47**

Figure 3.2 Representative whole-heart segmentations from the HVSMR-2.0 dataset illustrating anatomical variability in congenital heart disease, including differences in cardiac morphology, vessel orientation, and heart size [113] **50**

Figure 3.3 Representative multiplanar reconstruction views from the multi-center pediatric cardiovascular imaging dataset **52**

Figure 3.4 Representative multiplanar views with corresponding segmentation masks and overlay visualizations of target cardiovascular structures **53**

Figure 3.5 Multi-class label encoding scheme for background, aorta, cardiac chambers, and pulmonary artery classes **54**

Figure 3.6 Multi-class target structures in three-dimensional reconstruction and multiplanar views. Red: aorta, yellow: cardiac chambers, green: pulmonary arteries **55**

Figure 3.7 Annotation and quality control workflow for multi-class cardiovascular segmentation in 3D Slicer, showing axial, coronal, sagittal, and three-dimensional visualization views **56**

Figure 3.8 Overview of the preprocessing pipeline including standardization, resampling, normalization, and mask preparation **56**

Figure 3.9 Representative multi-planar visualization of a pediatric cardiovascular CT scan during initial image inspection in 3D Slicer: (a) axial, (b) coronal, and (c) sagittal views **58**

Figure 3.10 Comparison of image and segmentation mask geometry before and after preprocessing, illustrating correction of spatial inconsistencies **58**

Figure 3.11 Normalization during preprocessing. (A) Original image (B) Normalized image with improved intensity consistency **61**

Figure 3.12 Overview of the localization-guided segmentation framework including preprocessing, ROI extraction, and multi-class 3D segmentation **62**

Figure 3.13 Anatomical landmark definition and ROI localization framework **65**

Figure 3.14 Landmark localization network architecture **66**

Figure 3.15 Heatmap-based landmark representation **67**

Figure 3.16 Representative ROI-based scan and mask visualizations from two cases shown in axial, coronal, and sagittal views, demonstrating reduced background while preserving target cardiovascular structures **69**

Figure 3.17 Overview of the proposed localization-guided framework incorporating ROI extraction, hybrid V-Net–transformer segmentation, and federated learning across multiple medical centers **71**

Figure 3.18 Architecture of the proposed localization-guided hybrid V-Net–Transformer segmentation network **74**

Figure 3.19 Voxel-wise multi-class output representation with structure-specific supervision for pulmonary arteries **75**

Figure 3.20 Inference workflow of the proposed framework: (a) localization-guided ROI-based segmentation pipeline and (b) optional test-time augmentation (TTA) strategy for robust prediction aggregation **78**

Figure 3.21 Federated client–server architecture for the proposed localization-guided segmentation framework **80**

Figure 4.1 Training and validation loss curves of the proposed Hybrid V-Net model under the centralized training configuration **92**

Figure 4.2 Federated training curves of the proposed Hybrid V-Net model across communication rounds **92**

Figure 4.3 Overview of the experimental evaluation scenarios used in this study **93**

Figure 4.4 Qualitative comparison of segmentation outputs across architectures on benchmark (A–B) and clinical (C–D) cases **108**

Figure 4.5 Cross-dataset comparison of mean Dice scores across benchmark, clinical, and combined datasets **111**

Figure 4.6 Class-wise Dice comparison of segmentation architectures on the combined dataset under federated learning with localization **112**

Figure 4.7 Representative pulmonary artery segmentation results obtained using the proposed Hybrid V-Net model. Error maps indicate correctly segmented regions (blue), false positives (orange), and false negatives (green) **113**

Figure 4.8 Qualitative comparison of full-volume and ROI-based segmentation on the combined dataset. A: Aorta, PA: Pulmonary artery, CH: Cardiac chambers **115**

Figure 4.9 Client-wise segmentation performance under federated learning across participating clients **117**

Figure 4.10 Cross-model explainability comparison for aorta segmentation **124**

Figure 4.11 Explainability comparison for left ventricle segmentation **124**

Figure 4.12 Explainability comparison for pulmonary artery segmentation **125**

Figure 4.13 Explainability precision across models and structures **125**

Figure 4.14 Comparison of in/out activation ratios across models and structures **126**

Figure A.1 Cross-model explainability comparison using Grad-CAM (GC) and Integrated Gradients (IG) **142**

Figure A.2 Per-patient explainability case study across segmentation models **143**

LIST OF TABLES

Table 2.1 Imaging variability factors in multi-center pediatric cardiovascular datasets and their effects on segmentation **19**

Table 2.2 Common evaluation metrics in three-dimensional medical image segmentation and small-vessel analysis **38**

Table 3.1 Summary of methodological configurations used in this study **48**

Table 3.2 Summary of experimental dataset configurations **49**

Table 3.3 Summary of the multi-center clinical dataset **51**

Table 3.4 Summary of segmentation target structures and their clinical relevance **54**

Table 3.5 Key preprocessing parameters used in this study **57**

Table 3.6 Intensity normalization strategies **60**

Table 3.7 Definition and clinical relevance of anatomical landmarks **64**

Table 3.8 Landmark localization network configuration **67**

Table 3.9 Compared segmentation architectures and their roles in this study **69**

Table 3.10 Hybrid V-Net with Transformer architecture characteristics **70**

Table 3.11 Optimization and training configuration of the proposed framework **77**

Table 3.12 Federated learning configuration used in the proposed framework **81**

Table 3.13 Training configuration and hyperparameters **83**

Table 3.14 Hardware and software environment **84**

Table 4.1 Subject-level data splits across evaluation settings **91**

Table 4.2 Summary of evaluation settings and key performance trends **93**

Table 4.3 Quantitative segmentation results on the HVSMR-2.0 benchmark dataset under centralized training without localization **94**

Table 4.4 Quantitative segmentation results on the HVSMR-2.0 benchmark dataset under centralized training with localization-guided ROI extraction **95**

Table 4.5 Quantitative segmentation results on the HVSMR-2.0 benchmark dataset using federated learning without localization **96**

Table 4.6 Quantitative segmentation results on the HVSMR-2.0 benchmark dataset using federated learning with localization-guided ROI extraction **97**

Table 4.7 Summary of evaluation settings and key performance trends on the clinical dataset **98**

Table 4.8 Quantitative segmentation results on the clinical multi-center dataset under centralized training without localization **99**

Table 4.9 Quantitative segmentation results on the clinical multi-center dataset under centralized training with localization-guided ROI extraction. **100**

Table 4.10 Quantitative segmentation results on the clinical multi-center dataset using federated learning without localization **101**

Table 4.11 Quantitative segmentation results on the clinical multi-center dataset using federated learning with localization-guided ROI extraction ... **102**

Table 4.12 Summary of evaluation settings and key performance trends on the combined dataset **103**

Table 4.13 Quantitative segmentation results on the combined multi-center dataset under centralized training without localization **104**

Table 4.14 Quantitative segmentation results on the combined multi-center dataset under centralized training with localization-guided ROI extraction. **105**

Table 4.15 Quantitative segmentation results on the combined multi-center dataset using federated learning without localization **106**

Table 4.16 Quantitative segmentation results on the combined multi-center dataset using federated learning with localization-guided ROI extraction ... **107**

Table 4.17 Comparative performance analysis across datasets and training configurations **110**

Table 4.18 Comparative evaluation of segmentation architectures across datasets **111**

Table 4.19 Quantitative comparison of full-volume and ROI-based segmentation **114**

Table 4.20 Comparison of centralized and federated training strategies on the combined multi-center dataset **116**

Table 4.21 Pairwise statistical comparison of segmentation performance **121**

Table 4.22 Boundary-aware evaluation for pulmonary artery segmentation **122**

Table 4.23 Surface-based validation for pulmonary artery segmentation **122**

Table 4.24 Aggregated explainability metrics across models and structures **123**

Table A.1 Comprehensive federated segmentation performance **139**

Table A.2 Per-class segmentation performance (global model) **140**

Table A.3 Client-wise variability in segmentation performance (Hybrid V-Net) **140**

Table A.4 Ensemble evaluation results (nnU-Net + Swin-UNet) **141**

Table A.5 Ensemble boundary performance (mean per patient) **141**

ABSTRACT

Localization-Guided Federated Learning for Multi-Center 3D Segmentation of Pediatric Cardiovascular Structures

İrem YILMAZ

Department of Biomedical Engineering

Master of Science Thesis

Supervisor: Assoc. Prof. Dr. Sakip ÖNDER

Co- supervisor: Assist. Prof. Dr. Şenol PİŞKİN

This thesis investigates multi-class segmentation of pediatric cardiovascular structures using three-dimensional computed tomography (CT) and magnetic resonance imaging (MRI) data, with emphasis on anatomically complex congenital heart disease cases. Segmentation of the aorta, cardiac chambers, and pulmonary arteries remains challenging because of anatomical variability, small vessel structures, and heterogeneity across multi-center imaging datasets. The framework was evaluated on the HVSMR-2.0 benchmark dataset, a clinically curated pediatric cardiovascular cohort, and a combined dataset containing 110 cases. The proposed framework integrates localization-guided region-of-interest extraction, comparative evaluation of multiple three-dimensional segmentation architectures, a proposed Hybrid V-Net model, and federated learning for distributed training. Experimental results showed that the Hybrid V-Net achieved a Dice score of 92.1% on the benchmark dataset and 86.6% on the heterogeneous combined dataset under localization-guided centralized segmentation settings. Pulmonary artery segmentation on the combined dataset improved from 74.2% to 80.6% following localization-guided processing. In federated learning experiments on the combined

dataset, the proposed framework maintained competitive performance with an overall Dice score of 86.0% while enabling privacy-preserving collaborative training across distributed datasets. Overall, the findings suggest that localization-guided segmentation combined with federated learning contributes to more robust pediatric cardiovascular segmentation under heterogeneous multi-center conditions.

Keywords: Pediatric cardiovascular imaging, 3D segmentation, localization-guided approach, federated learning, multi-center dataset

ÖZET

Pediatrik Kardiyovasküler Yapıların Çok Merkezli 3B Segmentasyonu için Lokalizasyon Rehberli Federatif Öğrenme

İrem YILMAZ

Biyomedikal Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Doç. Dr. Sakip ÖNDER

Eş-Danışman: Dr. Öğr. Üyesi Şenol PİŞKİN

Bu tezde, anatomik olarak karmaşık konjenital kalp hastalığı vakalarına odaklanılarak, üç boyutlu bilgisayarlı tomografi (BT) ve manyetik rezonans görüntüleme (MRG) verileri kullanılarak pediatrik kardiyovasküler yapıların çok sınıflı segmentasyonu incelenmiştir. Aort, kardiyak odacıklar ve pulmoner arterlerin segmentasyonu; anatomik değişkenlik, küçük damar yapıları ve çok merkezli görüntüleme verileri arasındaki farklılıklar nedeniyle halen önemli bir zorluk oluşturmaktadır. Önerilen yöntem, HVSMR-2.0 referans veri seti, klinik olarak oluşturulmuş pediatrik kardiyovasküler veri kümesi ve toplam 110 vakadan oluşan birleşik veri seti üzerinde değerlendirilmiştir. Geliştirilen çerçeve; lokalizasyon tabanlı ilgi bölgesi çıkarımı, farklı üç boyutlu segmentasyon mimarilerinin karşılaştırmalı analizi, önerilen Hybrid V-Net modeli ve federatif öğrenme yaklaşımını bir araya getirmektedir. Deneysel sonuçlar, lokalizasyon tabanlı merkezi segmentasyon yaklaşımında Hybrid V-Net modelinin referans veri setinde %92.1, heterojen birleşik veri setinde ise %86.6 Dice benzerlik skoru elde ettiğini göstermiştir. Birleşik veri setinde pulmoner arter segmentasyon performansı, lokalizasyon tabanlı işlem sonrasında %74.2'den %80.6'ya

yükselmiştir. Birleşik veri seti üzerinde gerçekleştirilen federatif öğrenme deneylerinde önerilen yöntem, farklı merkezler arasında gizlilik korumalı ortak eğitimi desteklerken %86.0 genel Dice benzerlik skoru ile rekabetçi performansını korumuştur. Genel olarak bulgular, lokalizasyon tabanlı segmentasyonun federatif öğrenme ile kullanılmasının, heterojen çok merkezli koşullar altında daha dayanıklı pediatrik kardiyovasküler segmentasyon performansı sağladığını göstermektedir.

Anahtar Kelimeler: Pediatrik kardiyovasküler görüntüleme, 3B segmentasyon, lokalizasyon rehberli yaklaşım, federatif öğrenme, çok merkezli veri seti

1
INTRODUCTION

Over the past decades, medical imaging technologies have progressed from conventional two-dimensional radiographic approaches to advanced three-dimensional volumetric imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI). These developments have improved the visualization of internal anatomical structures and have become an significant part of modern clinical practice. In cardiovascular medicine, CT and MRI are widely used for the diagnosis, treatment planning, and follow-up evaluation of congenital heart disease (CHD) [1], [2]. Representative examples of volumetric cardiac imaging are presented in Figure 1.1.

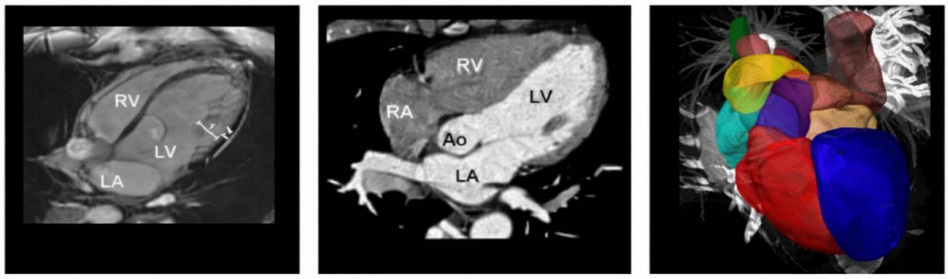


Figure 1.1 Representative cardiac MRI, CT, and three-dimensional cardiac rendering from the MM-WHS dataset [3]

CHD, which refers to structural abnormalities of the heart and great vessels present at birth, is the most common congenital anomaly worldwide, with an estimated prevalence of approximately 8–10 cases per 1,000 live births [4]. Advances in pediatric cardiac surgery and interventional cardiology have considerably improved survival into adulthood for patients with CHD [5]. Consequently, long-term cardiovascular follow-up has become increasingly important, creating a need for accurate three-dimensional anatomical assessment and quantitative analysis.

With the broader use of volumetric imaging in clinical practice, reliable segmentation of cardiovascular structures such as the aorta, cardiac chambers, and pulmonary arteries has gained increasing importance. Accurate delineation is required for surgical planning, patient-specific modeling, and computational hemodynamic analysis [6]. Traditionally, segmentation has been performed manually by experienced clinicians. Although manual delineation allows detailed anatomical interpretation, it is time-consuming, requires expert annotators, and remains subject to inter-observer variability even when standardized contouring protocols are followed [7]. These limitations reduce reproducibility and make large-scale clinical application more difficult. As a result, interest in automated and standardized segmentation methods has continued to grow.

Despite recent progress in three-dimensional medical image segmentation, pediatric cardiovascular segmentation remains challenging in realistic multi-center settings. Many existing studies rely on single-center or relatively homogeneous datasets, limiting generalization across different clinical environments. Small vascular structures are particularly difficult to segment because congenital anatomical variability, boundary ambiguity, and partial volume effects can affect structural continuity and local accuracy. In addition, privacy regulations and institutional data-sharing restrictions complicate centralized collection of large imaging datasets. Therefore, approaches that can operate under heterogeneous multi-center conditions while supporting collaborative training remain necessary.

Recent developments in deep learning have improved volumetric medical image segmentation considerably. Three-dimensional convolutional encoder–decoder networks and transformer-based architectures have achieved strong performance in a variety of medical imaging applications [8]–[11]. Nevertheless, obtaining reliable and generalizable automated segmentation in pediatric cardiovascular imaging is still difficult. This challenge arises from both anatomical and data-related factors. Congenital cardiovascular anatomy shows substantial inter-patient variability, while small and branching vascular structures, especially the pulmonary arteries, remain highly sensitive to boundary uncertainty and partial volume effects. Furthermore, imaging data collected across multiple centers differ in scanner vendors, acquisition protocols, reconstruction parameters, and contrast timing,

which may introduce domain shift and reduce model generalization on previously unseen data.

Restrictions related to institutional data governance further limit centralized aggregation of patient-level imaging datasets. For this reason, federated learning has emerged as a distributed training strategy that enables collaborative model optimization while preserving local data privacy [12]. However, segmentation performance on heterogeneous pediatric cardiovascular datasets has not yet been extensively investigated under consistent evaluation settings.

This thesis addresses these challenges through the development and systematic evaluation of a deep learning–based framework for multi-class three-dimensional segmentation of pediatric cardiovascular structures under realistic multi-center clinical conditions.

The main contributions of this thesis are summarized as follows:

- A heterogeneous pediatric cardiovascular imaging dataset collected across multiple centers, enabling systematic evaluation under realistic multi-center variability.
- A standardized benchmarking framework enabling fair and reproducible comparison across multiple three-dimensional segmentation architectures.
- A localization-guided strategy designed to improve structural continuity and boundary precision in small-caliber vascular structures, particularly the pulmonary arteries.
- A federated learning–based training scheme for privacy-preserving distributed model optimization and multi-center generalization analysis.
- Integration of ensemble prediction and post hoc explainability analyses to support robustness assessment and interpretation of model behavior.

This chapter introduced the research problem and motivation of the thesis. Chapter 2 presents the clinical, anatomical, and technical background relevant to this study. Chapter 3 describes the proposed methodology and data-processing pipeline. Chapter 4 provides the experimental setup, results, and detailed analyses. Finally, Chapter 5 summarizes the main findings, discusses study limitations, and outlines directions for future work. The Appendix includes supplementary analyses and additional experimental details.

1.1 Literature Review

1.1.1 Clinical Importance of Pediatric Cardiovascular Segmentation

Recent studies highlight that congenital heart disease (CHD) represents a structurally heterogeneous group of conditions, making standardized imaging-based assessment and segmentation difficult across different patient cohorts [13], [14]. As survival into adulthood continues to improve and the number of individuals living with CHD increases [15], the demand for scalable and reproducible approaches for long-term clinical evaluation has also grown. In this context, accurate delineation of cardiovascular structures is important for maintaining diagnostic consistency and reliable comparison across centers.

Three-dimensional imaging modalities, particularly computed tomography (CT) and magnetic resonance imaging (MRI), are widely used for evaluating complex congenital cardiovascular anatomy. CT enables rapid image acquisition with high spatial resolution, allowing detailed visualization of vascular structures and anatomical relationships. MRI, on the other hand, provides radiation-free assessment of cardiac morphology, ventricular function, and flow characteristics. Together, these modalities generate volumetric datasets containing detailed structural and functional information that support the identification of vascular abnormalities, intracardiac defects, post-surgical anatomical changes, and hemodynamic effects associated with congenital malformations [16], [17]. However, transforming volumetric imaging data into clinically useful anatomical representations still relies heavily on expert-driven manual or semi-manual delineation.

Manual delineation of cardiovascular structures such as the aorta, cardiac chambers, and pulmonary arteries remains labor-intensive and highly operator-dependent. Previous studies have shown that manual contouring may lead to inter-observer variability and limited reproducibility, especially in anatomically complex congenital cases where structural irregularities and prior surgical interventions increase contouring uncertainty. Pediatric imaging datasets further introduce additional challenges, including thin branching vessels, reduced boundary contrast, motion artifacts, and variable image quality. Together, these factors complicate

cross-center standardization and highlight the need for automated, scalable, and reproducible segmentation frameworks [7], [18].

1.1.2 Anatomical and Small-Vessel Challenges

The anatomical heterogeneity associated with congenital heart disease (CHD) represents a major clinical and structural challenge. Congenital malformations often involve abnormal vessel branching patterns, chamber hypoplasia, complex reconstructions, and post-surgical anatomical alterations that differ substantially from standard anatomical structures [19]. Unlike many acquired adult cardiovascular diseases that show relatively predictable morphology, pediatric CHD includes a broad and highly variable anatomical spectrum. As a result, segmentation models are expected to generalize across structurally diverse and patient-specific anatomies.

Beyond large-scale anatomical variability, small-caliber vascular structures introduce additional computational challenges. Pediatric pulmonary arteries and distal vascular branches have small diameters and complex branching patterns, increasing sensitivity to boundary uncertainty and making accurate vessel representation more difficult, similar to challenges previously reported in coronary artery analysis studies [20]. These characteristics increase the risk of fragmented predictions and local discontinuities in automated segmentation outputs. Therefore, maintaining structural continuity and boundary precision in thin branching vessels remains important for clinically reliable segmentation.

Accurate segmentation is not limited to visual delineation and directly affects downstream clinical and computational applications. Three-dimensional reconstructions generated from segmentation masks support patient-specific modeling approaches used in surgical planning and spatial evaluation of complex CHD cases [21]. Catheter-based interventions also depend on precise vascular delineation for safe navigation and device placement, while longitudinal follow-up requires reliable quantitative measurements such as vessel diameters and chamber volumes. In addition, segmentation masks provide the geometric basis for patient-specific computational modeling, including computational fluid dynamics simulations performed on anatomically realistic vascular geometries [22]. Even relatively small boundary inaccuracies in thin vascular structures may influence

simulated pressure gradients, flow distributions, and wall shear stress estimations, potentially affecting physiological interpretation and clinical decision-making. These considerations highlight the importance of segmentation systems capable of maintaining anatomical plausibility, structural continuity, and boundary fidelity.

1.1.3 Evaluation Metrics and Clinical Relevance

From an evaluation perspective, overlap-based metrics such as the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) remain among the most commonly used quantitative measures in medical image segmentation studies [23], [24]. Although these metrics provide estimates of overall volumetric agreement, they may not fully reflect boundary precision or local structural differences, in thin or highly branching anatomical regions.

In pediatric cardiovascular applications, preserving fine vascular integrity is especially significant because small discontinuities or boundary deviations may have limited impact on global overlap scores while still substantially affecting clinical interpretation and downstream computational modeling. For this reason, complementary boundary-aware and surface-distance–based metrics, including Hausdorff distance–based measures, have become increasingly important for clinically meaningful performance evaluation. Recent studies therefore emphasize the importance of structure-specific evaluation in addition to global summary metrics, particularly for small-caliber vascular structures.

1.1.4 Multi-Center Variability and Domain Shift

Beyond anatomical complexity, multi-center clinical variability represents another critical challenge. Pediatric cardiovascular imaging datasets collected across different institutions often show considerable heterogeneity because of variations in scanner vendors, acquisition protocols, and reconstruction techniques [25]. In pediatric imaging, protocols frequently prioritize radiation reduction and motion control, which may further influence signal-to-noise characteristics and contrast consistency in acquired images [26].

From a machine learning perspective, this variability may introduce domain shift, where the statistical distribution of training data differs from the data encountered during deployment [27]. Such domain shifts can appear as differences in image intensity distributions, texture patterns, anatomical scale, and noise characteristics

across institutions, potentially reducing model generalization when applied to previously unseen clinical environments.

As a result, performance obtained from homogeneous or single-center datasets may overestimate robustness and may not accurately reflect behavior under realistic heterogeneous multi-center conditions. This highlights the importance of evaluation protocols that directly assess cross-center robustness under standardized preprocessing and comparable metric computation. Addressing multi-center variability therefore requires either robustness-oriented model design strategies, such as localization-guided region-of-interest processing, or collaborative training approaches capable of utilizing distributed data diversity while reducing the impact of institutional heterogeneity.

1.1.5 Deep Learning and Federated Learning Approaches

Deep learning–based segmentation methods have considerably advanced automated medical image analysis. Three-dimensional convolutional encoder–decoder architectures have achieved strong performance in many volumetric segmentation tasks [28], [29]. More recently, transformer-based architectures have been introduced to capture long-range contextual relationships in volumetric medical images, leading to improvements in feature representation and segmentation accuracy [30], [31]. However, differences in architectural design and training behavior may produce structure-dependent performance variations, making comparative evaluation across architectures particularly relevant for pediatric vascular segmentation tasks.

Despite these developments, applying such models reliably in heterogeneous pediatric cardiovascular settings remains challenging. Many existing approaches are developed and validated using institutionally homogeneous datasets, limiting systematic evaluation under realistic multi-center conditions. In addition, small-caliber vascular structures remain highly vulnerable to segmentation errors because of class imbalance, limited spatial resolution, and boundary uncertainty, as reported in previous vessel segmentation studies [32]. These challenges underline the importance of structure-aware optimization strategies and robustness-oriented evaluation frameworks.

At the same time, regulatory and data-governance restrictions complicate centralized aggregation of pediatric imaging datasets collected across multiple institutions. Federated learning has therefore emerged as a privacy-preserving training paradigm that enables collaborative model optimization without direct sharing of patient-level data [33]. Nevertheless, its effectiveness under heterogeneous pediatric cardiovascular conditions remains insufficiently investigated. The combination of non-identically distributed data and substantial anatomical variability introduces additional challenges related to optimization stability and cross-center generalization.

Beyond quantitative performance, recent studies increasingly emphasize robustness and interpretability in clinically sensitive applications. Post hoc explainability approaches can provide complementary information about model behavior and may support greater clinical transparency [34], [35].

1.1.6 Current Limitations and Research Gaps

Taken together, the existing literature shows that pediatric cardiovascular segmentation is both clinically important and technically challenging, particularly under conditions involving anatomical variability and heterogeneous imaging data. Recent developments in deep learning and federated learning provide promising directions for addressing these challenges.

However, several limitations remain. First, multi-center heterogeneity is rarely evaluated as an explicit factor within experimental protocols, limiting systematic assessment of cross-center robustness. Second, unified benchmarking across convolutional, transformer-based, and hybrid architectures using heterogeneous pediatric datasets remains limited. Third, localization-guided and anatomy-aware processing strategies have not been consistently integrated into segmentation pipelines. Finally, the combined impact of federated collaborative training, anatomical variability, and domain shift has not yet been comprehensively investigated within a controlled experimental framework.

These limitations highlight the need for an integrated framework that jointly evaluates multi-center variability, region-of-interest–focused segmentation, cross-architecture comparison, and privacy-preserving collaborative training within a unified experimental setting.

1.2 Objectives of the Thesis

The primary objective of this thesis is to develop and systematically evaluate a deep learning–based framework for multi-class three-dimensional segmentation of pediatric cardiovascular structures under heterogeneous clinical conditions. The proposed framework aims to preserve anatomical consistency and structural continuity while reducing performance variability across multi-center imaging datasets. Segmentation performance is evaluated using standardized overlap-based and boundary-aware metrics to enable objective comparison across models and datasets. Overall, the study seeks to support more robust and clinically reliable pediatric cardiovascular segmentation in realistic clinical settings.

To achieve this objective, the following goals are pursued:

• To construct and characterize a heterogeneous multi-center pediatric cardiovascular imaging dataset and investigate the effects of inter-center variability on segmentation accuracy and model generalization.

• To design, implement, and validate a localization-guided segmentation framework with structure-aware optimization strategies aimed at improving segmentation accuracy, boundary precision, and anatomical continuity, particularly in small-caliber vascular structures such as the pulmonary arteries.

• To comparatively evaluate convolutional, transformer-based, and hybrid three-dimensional deep learning architectures under standardized experimental conditions and analyze their structure-specific performance and generalization behavior.

• To implement and assess a federated learning–based collaborative training strategy and examine its impact on segmentation performance, cross-center robustness, and distributed learning capability under realistic clinical environments.

1.3 Hypothesis

Clinical heterogeneity and anatomical variability associated with congenital heart disease introduce domain shifts and structural ambiguities that complicate automated pediatric cardiovascular segmentation. These factors may reduce model generalization and segmentation reliability, particularly in small and anatomically

complex vascular structures such as the pulmonary arteries. Differences in imaging protocols, scanner characteristics, and acquisition settings across medical centers further increase the difficulty of obtaining stable segmentation performance.

The primary hypothesis of this thesis is that a localization-guided and structure-aware deep learning framework can improve anatomical focus and reduce segmentation variability in multi-center pediatric cardiovascular imaging data. Landmark-guided region-of-interest extraction and structure-aware optimization are intended to enhance boundary preservation and segmentation accuracy, especially in challenging pulmonary artery regions. In addition, integrating federated collaborative learning into the framework may support privacy-preserving distributed training while improving robustness across different clinical environments.

To evaluate these hypotheses, comprehensive experiments were conducted using benchmark and clinically heterogeneous pediatric cardiovascular datasets. Performance was assessed using overlap-based and boundary-aware metrics, including Dice score, Intersection over Union, Hausdorff Distance, and Average Symmetric Surface Distance. Structure-specific analyses and cross-center evaluations were also performed to examine pulmonary artery continuity, boundary fidelity, anatomical consistency, and generalization capability under centralized and federated learning settings.

2
CLINICAL AND TECHNICAL BACKGROUND

This chapter presents the clinical and technical background of pediatric cardiovascular image segmentation, focusing on the challenges of multi-class segmentation of small-caliber vascular structures under heterogeneous multi-center imaging conditions. Particular emphasis is placed on pulmonary artery segmentation because of its complex anatomy, small vessel size, and sensitivity to imaging variability and domain shift. The chapter introduces congenital heart disease and commonly used imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI), followed by a discussion of imaging variability across multi-center datasets and its impact on automated analysis. In addition, major deep learning–based segmentation approaches, small-vessel segmentation challenges, and commonly used evaluation metrics are reviewed. Finally, domain shift, domain adaptation, domain generalization, and federated learning are discussed to establish the background for the localization-guided and federated segmentation framework presented in Chapter 3.

2.1 Clinical and Anatomical Background

Accurate segmentation of small-caliber vascular structures, particularly the pulmonary arteries, remains one of the major challenges in pediatric cardiovascular imaging. This difficulty is closely related to the anatomical complexity and variability of pediatric cardiovascular structures, which can directly influence segmentation accuracy and the reliability of subsequent quantitative analysis. For this reason, understanding pediatric cardiovascular anatomy and congenital heart disease is important for developing robust and reliable computational methods. These challenges highlight the need for segmentation frameworks capable of

handling structural complexity, small-vessel characteristics, and heterogeneous imaging conditions.

2.1.1 Pediatric Cardiovascular Anatomy

Pediatric cardiovascular anatomy consists of multiple interconnected structures forming a complex three-dimensional cardiac and vascular system. In automated medical image segmentation tasks, the main anatomical targets include the cardiac chambers and major vessels such as the aorta and pulmonary arteries. Accurate delineation of these structures is important for quantitative assessment of cardiovascular morphology and for supporting clinical decision-making in pediatric cardiology [36]. Figure 2.1 presents representative cardiovascular structures in both cardiac computed tomography (CT) and magnetic resonance imaging (MRI). Segmentation outputs are shown across three orthogonal planes, illustrating the spatial relationships between cardiac chambers and great vessels and highlighting the multi-structure nature of cardiovascular segmentation.

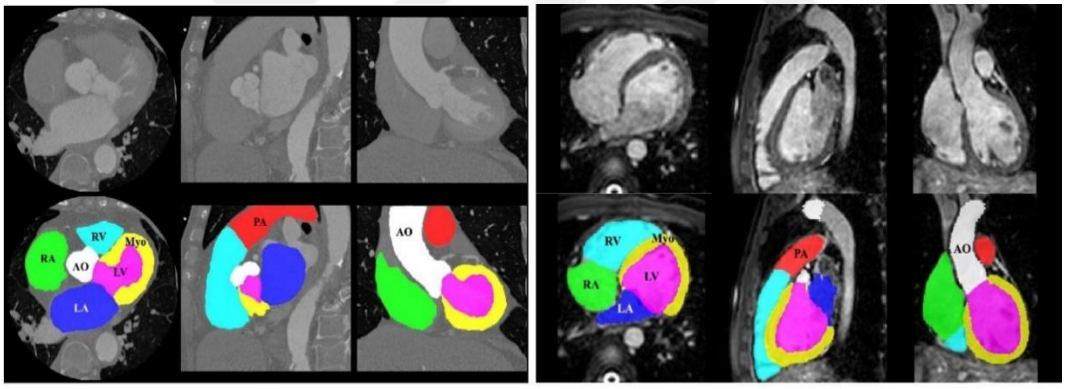


Figure 2.1 Representative cardiovascular structures in cardiac CT and MRI. (a) Whole-heart CT segmentation. (b) MRI segmentation overlays [36]

Among these structures, the pulmonary artery system represents one of the most challenging vascular components. The pulmonary circulation begins at the main pulmonary artery (MPA), which bifurcates into the right (RPA) and left pulmonary arteries (LPA) before extending into smaller distal branches supplying the lungs [37], [38]. This branching organization creates a complex vascular topology with progressively decreasing vessel diameters. Figure 2.2 illustrates the hierarchical organization of the pulmonary arterial tree and the spatial complexity of distal vessel branching. These anatomical characteristics increase the difficulty of accurately segmenting small peripheral vessels.

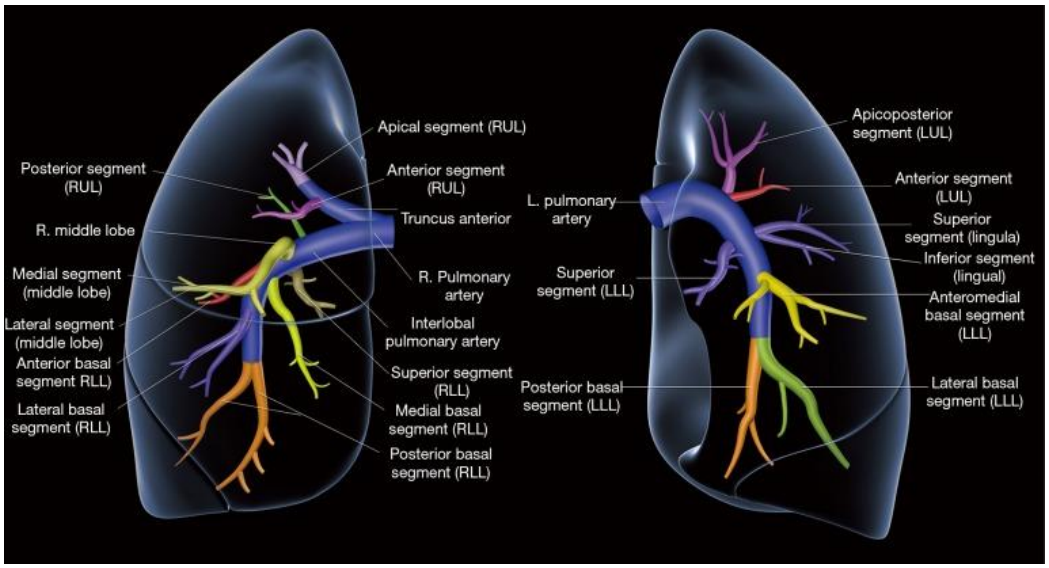


Figure 2.2 Pulmonary arterial tree anatomy with main pulmonary artery bifurcation and distal vessel branching [37], [38]

From a computational perspective, the elongated and branching structure of the pulmonary arterial system introduces additional challenges for automated analysis. Segmentation methods must preserve structural continuity while accurately capturing bifurcation regions and thin vascular branches [39]. These difficulties become more pronounced in pediatric populations because cardiovascular structures are considerably smaller than in adults. Distal vessel diameters may range between approximately 1–3 mm depending on age and anatomical location [40], [41]. In addition, small vessels often exhibit low contrast relative to surrounding tissues, while partial volume effects and limited spatial resolution may produce blurred or indistinct boundaries. In clinical CT and MRI, voxel dimensions may even approach the size of small vessels, further complicating accurate delineation [40]. The close spatial proximity of adjacent cardiovascular structures also increases boundary ambiguity during automated segmentation [41].

Together, these anatomical and imaging-related factors make pediatric cardiovascular segmentation particularly challenging, as segmentation algorithms must capture complex vascular topology while preserving fine structural details in small-caliber vessels. Under realistic multi-center clinical conditions, this difficulty is further increased by variability in imaging appearance and acquisition characteristics across institutions.

2.1.2 Congenital Heart Disease (CHD) and Anatomical Variability

Congenital heart disease (CHD) refers to structural abnormalities of the heart and great vessels present at birth and represents the most common congenital anomaly worldwide. These conditions include a broad range of defects affecting cardiac chambers, septa, valves, and major vessels such as the aorta and pulmonary arteries, resulting in substantial anatomical variability that complicates both clinical diagnosis and quantitative assessment of cardiovascular morphology. Imaging modalities such as cardiac computed tomography (CT) and magnetic resonance imaging (MRI) play an important role in visualizing these abnormalities and supporting anatomical evaluation [42]. Figure 2.3 presents representative CHD cases visualized using CT together with corresponding three-dimensional reconstructions, illustrating variations in anatomical structure, abnormal vessel connections, and spatial distortions across different patients.

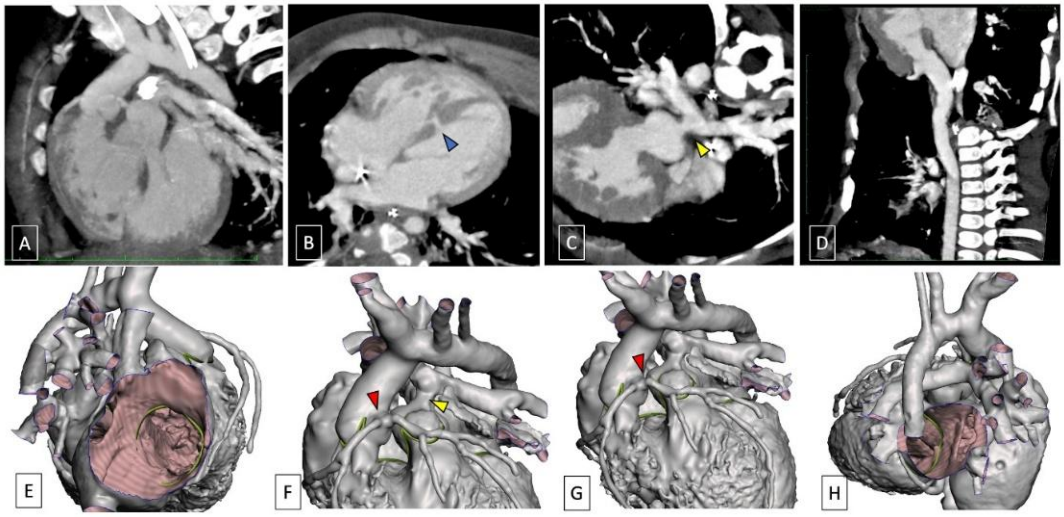


Figure 2.3 Representative congenital heart disease cases in cardiac CT with 3D reconstructions. (A–D) MPR views (E–H) 3D visualizations [42]

Accurate anatomical characterization is essential for the clinical management of CHD. Detailed visualization of cardiovascular structures supports surgical planning and catheter-based interventions, while longitudinal imaging is frequently required to monitor post-operative outcomes and disease progression [43]. Consequently, imaging-based evaluation remains a central component of pediatric cardiovascular care.

One of the defining features of CHD is the presence of anatomical variations that differ from typical cardiovascular structure. These variations may originate from abnormalities during cardiac development, while surgical and interventional

procedures can further alter anatomical geometry. As a result, pediatric cardiovascular anatomy in CHD patients often exhibits highly individualized and structurally complex spatial configurations [44].

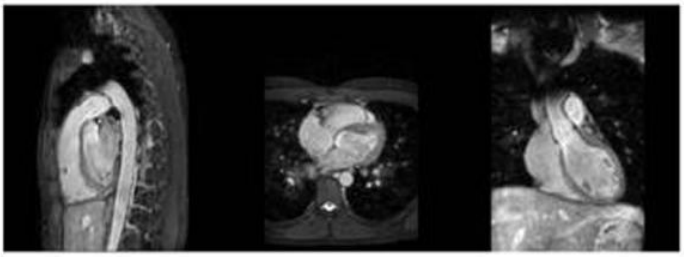
From a computational perspective, this variability creates significant challenges for automated image analysis. Segmentation algorithms must generalize across a wide range of morphological differences while preserving anatomical consistency. These difficulties become more pronounced in pediatric populations because patients range from neonates to adolescents, leading to major differences in heart size, vessel diameter, and anatomical proportions. Congenital defects may also produce abnormal vessel branching patterns, atypical vascular connections, and altered spatial relationships between cardiovascular structures, further increasing segmentation difficulty [45], [46].

Such variability becomes particularly problematic in small-caliber vessels such as the pulmonary arteries, where complex morphology and branching patterns further complicate accurate delineation. The combination of anatomical heterogeneity and small-vessel complexity therefore remains one of the major challenges in pediatric cardiovascular segmentation and motivates the development of structure-aware and generalizable segmentation frameworks under heterogeneous clinical conditions.

2.1.3 Computed Tomography and Magnetic Resonance Imaging

Computed tomography (CT) and magnetic resonance imaging (MRI) are widely used imaging modalities for evaluating pediatric cardiovascular anatomy. These techniques provide complementary information for diagnosis, treatment planning, and longitudinal monitoring of congenital heart disease. Both modalities enable three-dimensional visualization of cardiac chambers and great vessels, supporting quantitative anatomical assessment and automated segmentation. Figure 2.4 presents representative cardiac MRI and CT images acquired across different anatomical planes. Although both modalities provide volumetric visualization of cardiovascular structures, noticeable differences exist in image appearance, contrast, and structural clarity, which directly influence automated segmentation performance [47].

(a) Cardiac MRI images in different anatomical planes.



(b) Cardiac CT images in different anatomical planes.

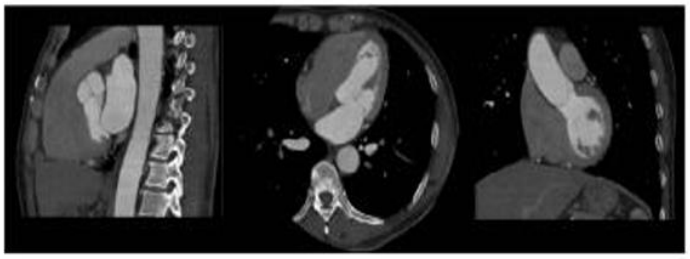


Figure 2.4 Representative cardiac MRI and CT images across anatomical planes. (a) MRI with soft-tissue contrast. (b) CT with vascular visualization [3]

Cardiac CT plays an significant role in pediatric cardiovascular imaging because of its high spatial resolution and rapid acquisition capability. Modern CT systems enable detailed visualization of the heart and great vessels within short scanning times, which is particularly advantageous in pediatric patients where motion artifacts and limited patient cooperation may affect image quality. CT imaging also provides clear delineation of vascular structures such as the aorta and pulmonary arteries, allowing accurate assessment of vessel morphology and spatial relationships [48].

Magnetic resonance imaging (MRI), in contrast, offers several complementary advantages. One of its main benefits is the absence of ionizing radiation, making MRI suitable for repeated imaging in pediatric populations. MRI also provides functional information related to blood flow, ventricular function, and tissue characterization. In addition, different contrast mechanisms may improve visualization of congenital cardiovascular abnormalities [49].

Despite these advantages, CT and MRI exhibit distinct imaging characteristics that directly affect automated segmentation. CT generally provides sharper vascular boundaries and higher spatial resolution, whereas MRI offers stronger soft-tissue contrast but lower spatial resolution and greater sensitivity to motion artifacts.

These modality-specific differences introduce substantial variability in image appearance that segmentation algorithms must effectively handle [50].

In clinical practice, CT and MRI are often used complementarily depending on patient characteristics and clinical requirements. Therefore, automated analysis systems should be capable of processing data from multiple imaging modalities. Developing segmentation approaches that generalize across both CT and MRI is especially important in multi-center settings, where modality-related variability may further increase domain shift and reduce model generalization.

2.2 Imaging Variability in Multi-Center Medical Data

In clinical practice, pediatric cardiovascular imaging data are commonly collected from multiple medical centers operating under different technical and clinical conditions. Consequently, these datasets often exhibit substantial variability arising from differences in scanner technologies, acquisition protocols, reconstruction parameters, and patient-specific factors. This multi-center variability represents a characteristic feature of real-world clinical imaging data and creates important challenges for automated image analysis, as illustrated in Figure 2.5.

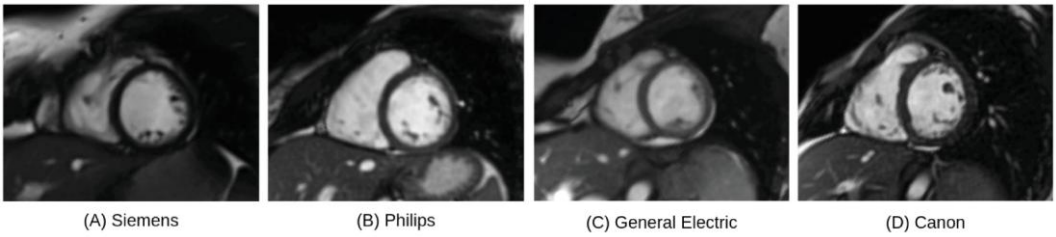


Figure 2.5 Cardiac MRI images from different scanner vendors illustrating inter-vendor variability in contrast and image quality [51]

One major source of variability is the use of different scanner models, vendors, and acquisition settings. Imaging systems produced by different manufacturers rely on distinct hardware configurations, detector technologies, and reconstruction pipelines, all of which influence final image appearance. As a result, scans of the same anatomical structures may differ in noise patterns, spatial resolution, and contrast characteristics across centers, potentially reducing the reproducibility and generalization capability of automated analysis methods [52].

Variability is further increased by differences in acquisition and reconstruction protocols. Parameters such as slice thickness, field of view, acquisition timing,

contrast administration, and reconstruction algorithms often vary between institutions according to local clinical practice and diagnostic requirements [53], [54]. In addition, differences in preprocessing, reconstruction, and post-processing workflows may also influence image appearance, as illustrated in Figure 2.6.

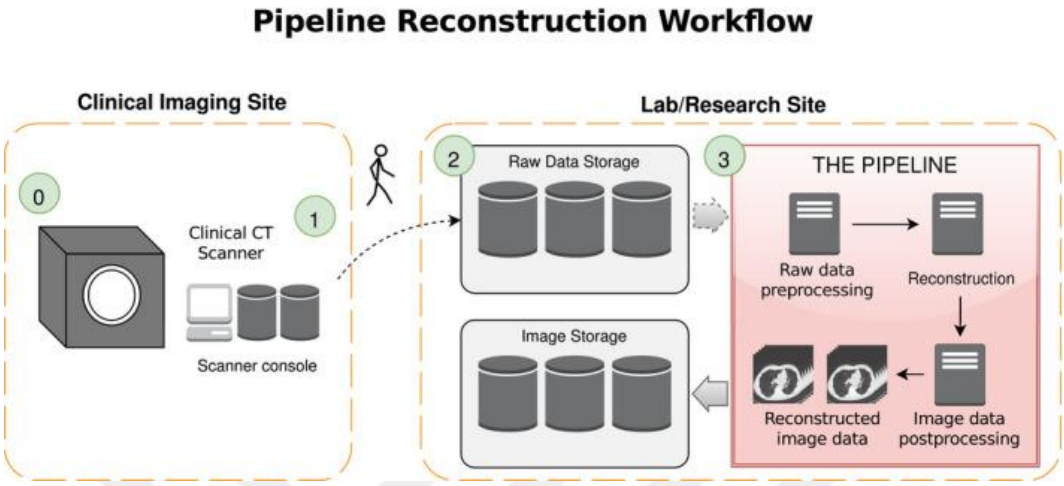


Figure 2.6 Reconstruction workflow in clinical and research imaging pipelines including acquisition, preprocessing, and reconstruction stages [54]

Additional variability arises from differences in spatial resolution and voxel spacing. Medical images acquired with different slice thicknesses and in-plane resolutions may produce inconsistent representations of anatomical structures, particularly in small distal vascular regions. These effects become especially problematic in the pulmonary arterial tree, where reduced spatial resolution and partial volume effects may obscure thin vascular branches or disrupt structural continuity [55].

Taken together, multi-center variability introduces differences in intensity distribution, image sharpness, noise characteristics, and artifact patterns across datasets. From a computational perspective, this heterogeneity complicates the development of robust and generalizable segmentation models. The challenge becomes even more pronounced for small-caliber vascular structures such as the pulmonary arteries, where low contrast and limited spatial resolution may further reduce segmentation accuracy and increase the likelihood of fragmented predictions.

The main sources of imaging variability and their effects on segmentation performance are summarized in Table 2.1.

Table 2.1 Imaging variability factors in multi-center pediatric cardiovascular datasets and their effects on segmentation

Source of variability	Imaging effect	Impact on segmentation	Relevance to small-vessel segmentation
Scanner vendor	Variations in intensity distribution and contrast	Domain shift across centers and reduced generalization	Low contrast makes small vessels harder to detect
Reconstruction kernel	Differences in edge sharpness and texture	Boundary ambiguity and inconsistent delineation	Blurred edges degrade vessel boundary accuracy
Slice thickness	Reduced spatial resolution and partial volume effects	Loss of fine anatomical details	Thin vessels may disappear or become fragmented
Voxel spacing	Anisotropic resolution	Distorted spatial representation	Affects continuity of tubular structures
Acquisition protocol	Variability in contrast timing and imaging parameters	Inconsistent appearance across datasets	Alters visibility of vascular structures

The variability described in this section is closely related to domain shift in machine learning. While introduced here from a clinical and imaging perspective, its computational implications and associated learning strategies are discussed further in Section 2.7.

2.3 Problem Formulation of 3D Medical Image Segmentation

Building upon the clinical and imaging challenges discussed in the previous sections, pediatric cardiovascular segmentation can be formulated as a three-dimensional multi-class segmentation problem under heterogeneous imaging conditions. In this setting, the input consists of volumetric medical images acquired using computed tomography (CT) or magnetic resonance imaging (MRI), while the output is a voxel-wise label map representing multiple anatomical structures, including cardiac chambers, the aorta, and pulmonary arteries. This formulation enables automated delineation of anatomically complex cardiovascular structures within the image volume.

The problem becomes more challenging in pediatric applications because of congenital anatomical variability, small-caliber vascular structures, and heterogeneity across multi-center imaging datasets. Thin and branching vessels such as the pulmonary arteries require segmentation models capable of preserving fine structural details while maintaining global anatomical consistency. These challenges form the basis for the segmentation approaches and methodological strategies discussed in the following sections.

2.3.1 Problem Formulation

Three-dimensional medical image segmentation is commonly formulated as a voxel-wise classification problem, where each voxel in a volumetric image is assigned a class label corresponding to a specific anatomical structure [56], [57].

From a mathematical perspective, a three-dimensional medical image can be defined as a mapping over a spatial domain:

$$I: \Omega \rightarrow \mathbb{R} \tag{2.1}$$

where $\Omega \subset \mathbb{R}^3$ denotes the spatial domain and each voxel is associated with an intensity value.

The segmentation task can then be expressed as a mapping function:

$$f_\theta: I \rightarrow Y \tag{2.2}$$

where f_θ represents a segmentation model parameterized by θ, and Y denotes the predicted label map. In the multi-class setting, each voxel $v \in \Omega$ is assigned a class label from a discrete set:

$$Y(v) \in \{0,1, \ldots, K-1\} \tag{2.3}$$

where K denotes the total number of classes, including background.

In pediatric cardiovascular imaging, the input is a volumetric CT or MRI scan, and the output is a voxel-wise label map representing multiple anatomical structures, including cardiac chambers, the aorta, and pulmonary arteries, reflecting the close spatial relationships between cardiovascular components [58]. This transformation is illustrated in Figure 2.7.

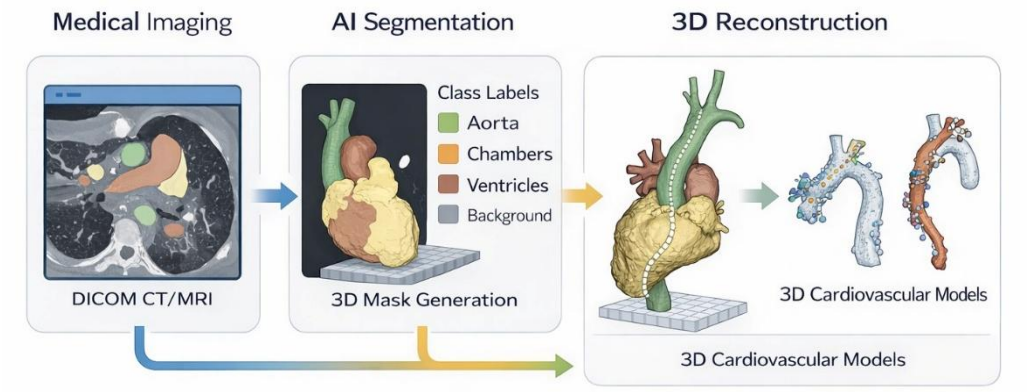


Figure 2.7 Conceptual workflow of three-dimensional multi-class segmentation from volumetric CT and MRI data

In practice, segmentation models estimate a probability distribution over classes for each voxel, and the final label is assigned as the class with the highest probability:

$$\hat{y}(v) = \arg\max_{k} \; p_k(v) \tag{2.4}$$

where $p_k(v)$ denotes the predicted probability for class k.

This formulation is relevant for cardiovascular imaging, where multiple anatomical structures are closely interconnected. Its importance becomes more pronounced in pediatric settings, where anatomical variability, small-caliber vessels, and multi-center data heterogeneity require models that preserve both fine structural details and global anatomical consistency [59].

2.3.2 Fundamentals of 3D Medical Image Segmentation

Three-dimensional medical images are represented as volumetric datasets composed of voxels arranged within a regular three-dimensional grid, where each voxel contains an intensity value corresponding to the measured imaging signal. In volumetric imaging modalities such as CT and MRI, voxel spacing may vary across spatial axes, resulting in anisotropic spatial resolution.

Anisotropic voxel spacing can reduce the visibility of fine anatomical details, particularly in small and thin vascular structures. To reduce these effects, resampling techniques are commonly used to obtain isotropic voxel spacing, as illustrated in Figure 2.8 [60], [61].

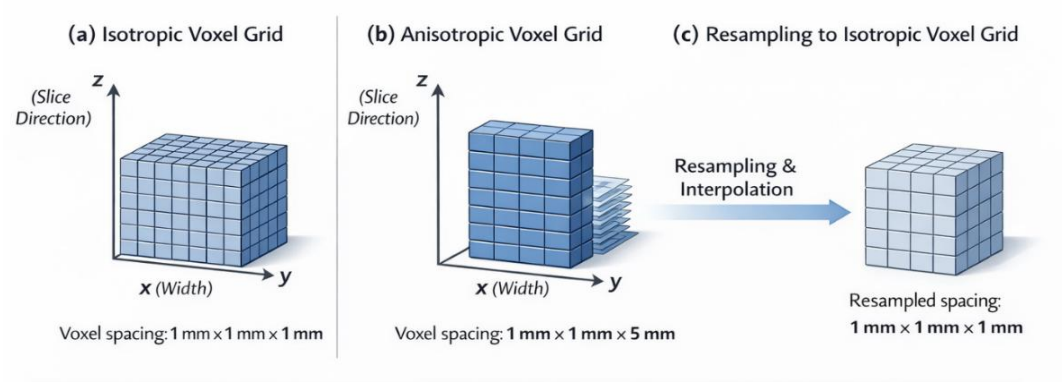


Figure 2.8 Illustration of isotropic and anisotropic voxel grids and the effect of resampling on three-dimensional image representation

In addition to voxel anisotropy, reconstruction parameters such as slice thickness also influence image quality and anatomical representation. Variations in slice thickness may reduce the visibility of fine structural details, particularly in small vascular structures such as the pulmonary arteries. Thicker slices can introduce partial volume effects and decrease boundary sharpness, whereas thinner slices generally preserve anatomical detail but may increase image noise, as illustrated in Figure 2.9 [62].

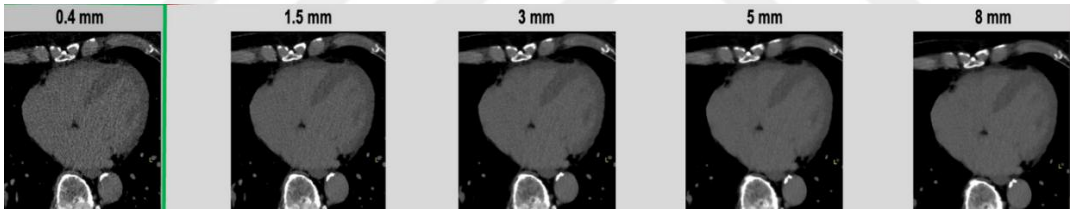


Figure 2.9 Effect of slice thickness on spatial resolution and anatomical detail in CT imaging [62]

Modern segmentation approaches perform voxel-wise labeling on volumetric data, enabling reconstruction of spatially consistent anatomical structures and supporting quantitative volumetric analysis. This formulation is aligned with the segmentation problem definition introduced in Section 2.3.1. Most contemporary methods are based on encoder–decoder architectures, where the encoder extracts multi-scale feature representations through progressive downsampling to capture global contextual information, while the decoder restores spatial resolution to generate dense voxel-wise predictions. Skip connections transfer high-resolution features between encoder and decoder stages, improving spatial localization and boundary delineation, as illustrated in Figure 2.10 [63].

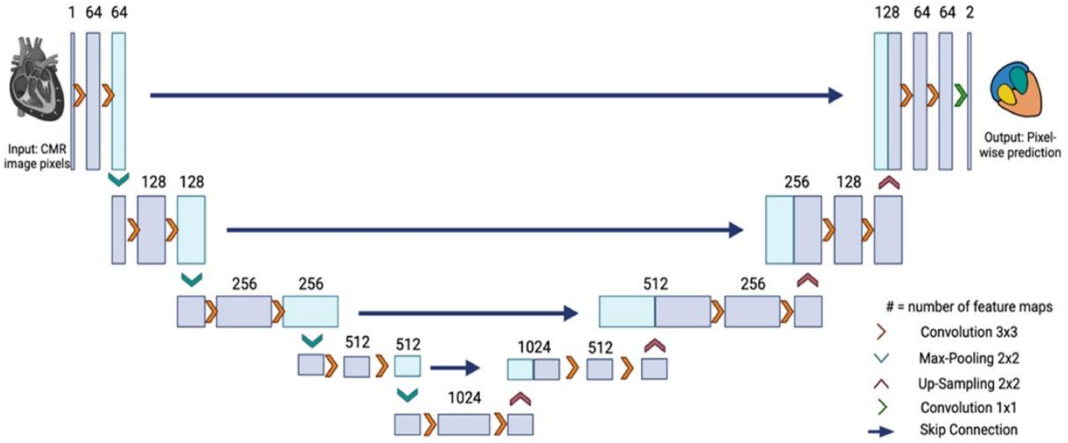


Figure 2.10 Encoder–decoder architecture for medical image segmentation [63]

In multi-class segmentation, models generate voxel-wise predictions for different anatomical structures. During training, these predictions are optimized using loss functions that measure agreement between predicted outputs and ground-truth annotations. In practice, combinations of classification-based and overlap-based losses are commonly used to improve training stability and segmentation performance, especially under class imbalance and in small anatomical structures [64].

2.4 Deep Learning–Based Segmentation Architectures

Medical image segmentation methods have evolved considerably in response to increasing anatomical complexity and variability in clinical imaging data. Early segmentation approaches were mainly based on classical image processing techniques such as thresholding, region growing, deformable models, and atlas-based methods. These techniques rely on intensity information, geometric constraints, and predefined anatomical priors. Although they may perform adequately under controlled conditions, their effectiveness is often limited in complex clinical scenarios involving anatomical variability and low-contrast structures [65].

The limitations of classical methods led to the widespread adoption of deep learning–based approaches. Convolutional neural networks (CNNs) became a dominant framework because of their ability to learn hierarchical feature representations directly from imaging data. These models can capture both local

anatomical details and broader contextual information, improving segmentation performance in heterogeneous and anatomically complex imaging conditions [66].

Figure 2.11 illustrates the evolution of deep learning–based segmentation architectures. Early encoder–decoder models such as U-Net introduced effective mechanisms for combining contextual information with precise spatial localization. Later volumetric extensions, including V-Net, enabled direct processing of three-dimensional imaging data. More recent developments include self-configuring frameworks such as nnU-Net and transformer-based architectures capable of modeling long-range contextual relationships. These developments form the basis of the segmentation models evaluated in this thesis.

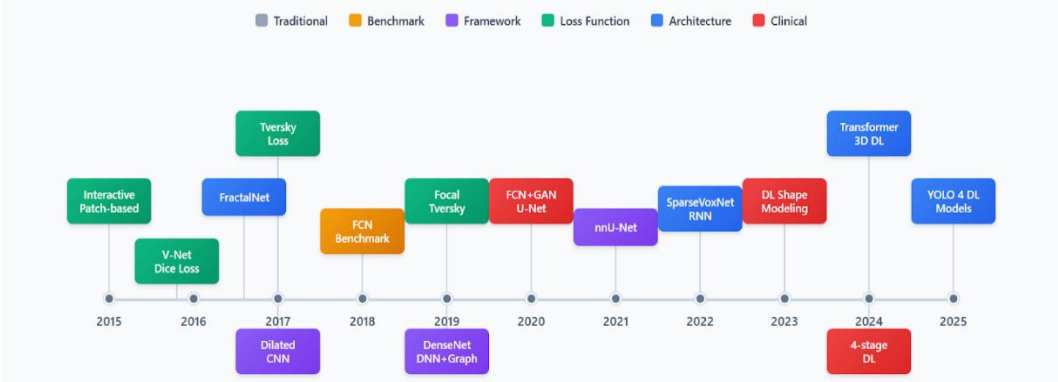


Figure 2.11 Evolution of deep learning–based medical image segmentation architectures from U-Net to transformer-based approaches (2015–2025)

Segmentation methods vary according to the anatomical and imaging challenges encountered in medical imaging. In particular, segmentation of small and branching vascular structures such as pulmonary arteries requires preservation of fine structural details and global anatomical consistency, especially in pediatric cardiovascular imaging [67]. Figure 2.12 summarizes segmentation paradigms ranging from classical techniques to deep learning–based methods.

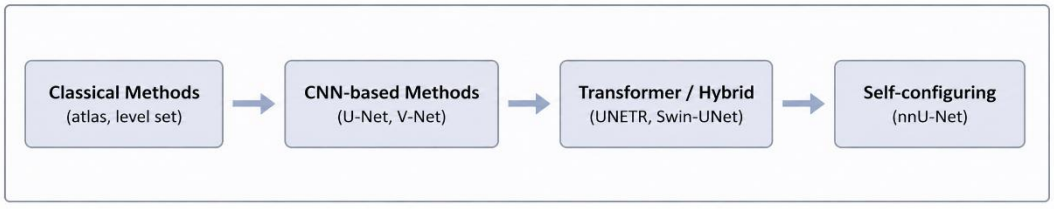


Figure 2.12 Taxonomy of medical image segmentation methods from classical techniques to deep learning–based approaches

In addition to architectural developments, region-of-interest (ROI)–based strategies have emerged as an effective approach for handling complex segmentation tasks.

These methods generally follow a coarse-to-fine strategy in which an initial localization stage identifies anatomically relevant regions before a refined segmentation stage is applied. By reducing irrelevant background information, ROI-based processing may improve both segmentation accuracy and computational efficiency.

Existing segmentation architectures provide complementary advantages in capturing local features, modeling global context, and adapting to heterogeneous datasets. Nevertheless, the combined challenges of small-caliber vascular structures, anatomical variability, and multi-center heterogeneity remain incompletely addressed. These limitations motivate the development of structure-aware and localization-guided segmentation frameworks such as the approach proposed in this thesis.

2.4.1 Classical Segmentation Methods

Before the widespread adoption of deep learning, medical image segmentation primarily relied on classical image processing and model-based approaches based on handcrafted features and anatomical priors. Among these methods, atlas-based segmentation was widely used, where anatomical labels were transferred from a reference atlas to a target image through image registration. Multi-atlas approaches further improved robustness by combining labels from multiple atlases, as illustrated in Figure 2.13 [68] – [70].

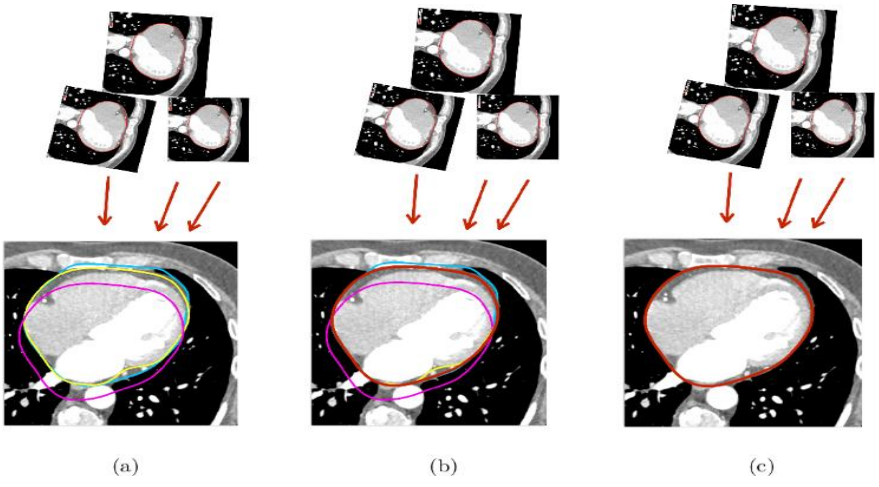


Figure 2.13 Multi-atlas segmentation in cardiac CT angiography. (a) Atlas registration. (b) Label fusion. (c) Final segmentation [70]

In addition to atlas-based methods, intensity-based and region-based techniques such as thresholding and region growing have also been widely applied. These

approaches segment structures mainly according to intensity differences but are highly sensitive to noise, intensity variability, and imaging artifacts [71]. Deformable models, including active contour and level-set methods, provide a more flexible framework by iteratively adapting contours to anatomical boundaries. Although these methods can capture complex shapes, they generally require careful initialization and parameter tuning, as illustrated in Figure 2.14 [72], [73].

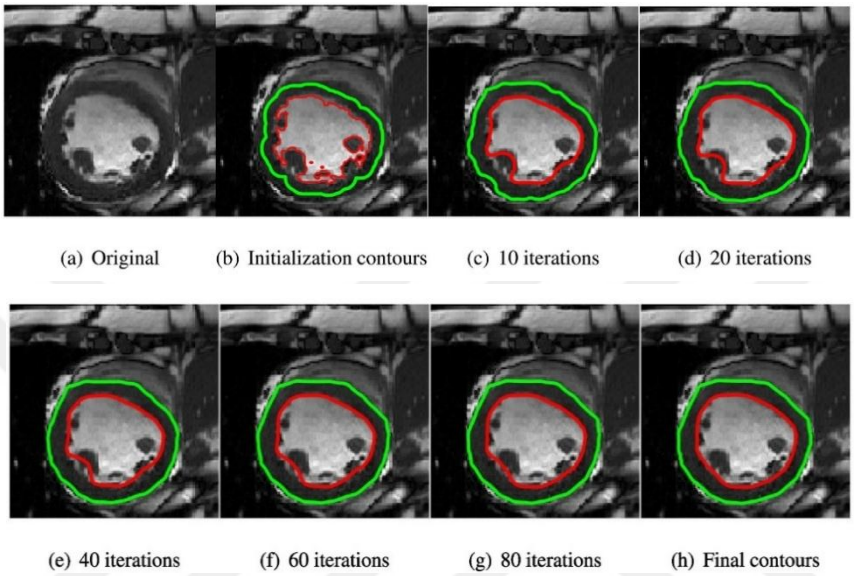


Figure 2.14 Illustration of level-set–based segmentation in cardiac MRI with iterative contour evolution toward anatomical boundaries [73]

Despite their historical importance, classical segmentation approaches show important limitations in complex clinical scenarios. Their dependence on handcrafted features reduces their ability to generalize across anatomical variability, low-contrast regions, and heterogeneous multi-center datasets. These limitations become particularly evident in pediatric cardiovascular imaging, where small-caliber vascular structures and substantial inter-patient variability increase segmentation difficulty. Consequently, these challenges contributed to the transition toward data-driven deep learning approaches that are better suited for heterogeneous and anatomically variable imaging conditions.

2.4.2 CNN-Based Segmentation

The introduction of convolutional neural networks (CNNs) has substantially advanced medical image segmentation by enabling automatic feature learning directly from imaging data. Unlike classical segmentation methods based on handcrafted rules, CNNs learn hierarchical feature representations that capture both

local spatial details and broader contextual information [74]. Through successive convolutional operations, these models extract increasingly abstract representations while preserving spatial relationships, which is critical for medical imaging tasks requiring structural continuity and contextual understanding [75].

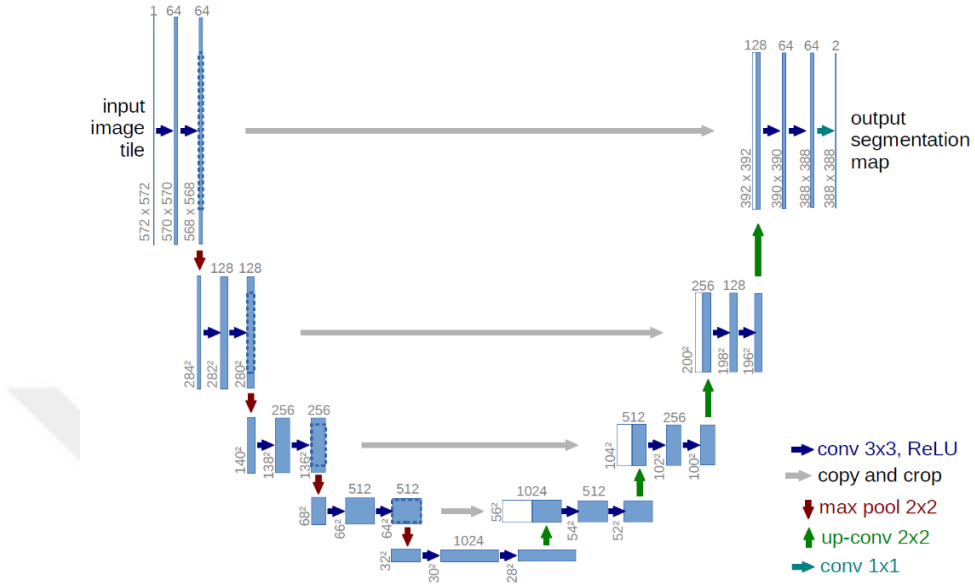


Figure 2.15 U-Net architecture for biomedical image segmentation with encoder–decoder structure and skip connections [8]

U-Net is one of the most influential CNN-based architectures developed for biomedical image segmentation. The model follows an encoder–decoder design in which the encoder extracts multi-scale feature representations while progressively reducing spatial resolution, and the decoder restores spatial resolution to generate dense segmentation outputs. Skip connections between corresponding encoder and decoder layers help preserve high-resolution spatial information and improve boundary delineation, as illustrated in Figure 2.15 [8], [76].

To better capture volumetric information, CNN architectures were later extended to three-dimensional models such as 3D U-Net, which directly process volumetric imaging data and enable spatially consistent segmentation across all dimensions [77]. V-Net further expanded this approach by introducing residual connections and Dice-based optimization strategies that improve training stability and segmentation performance, particularly under class imbalance conditions, as illustrated in Figure 2.16 [9], [78].

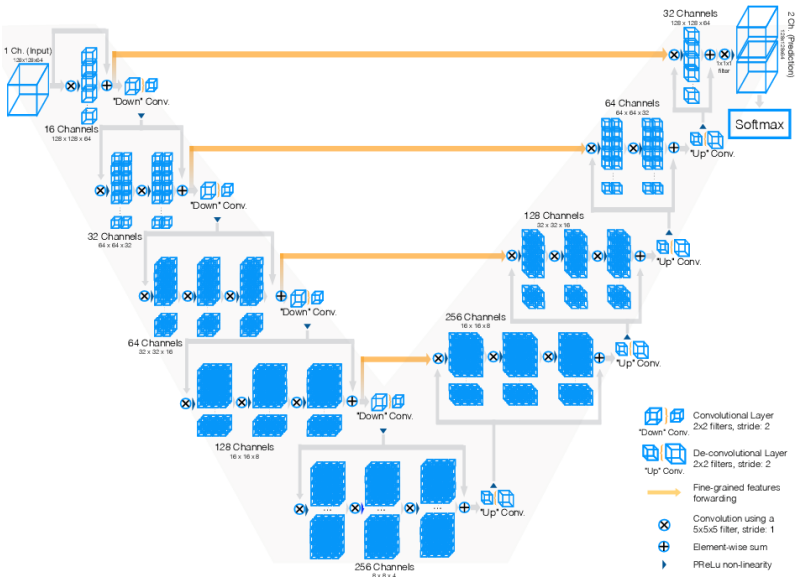


Figure 2.16 V-Net architecture for volumetric medical image segmentation with three-dimensional convolutions and residual connections [9]

Despite their strong performance, CNN-based models still face limitations in capturing long-range contextual dependencies and adapting to heterogeneous multi-center datasets. These challenges become especially important in pediatric cardiovascular imaging, where small and anatomically complex vascular structures require both precise localization and broader contextual understanding [79]. Such limitations motivated the development of more advanced architectures capable of modeling wider spatial relationships, as discussed in the following section.

2.4.3 Transformer and Hybrid Architectures

Transformer-based architectures provide an alternative approach to feature modeling by using self-attention mechanisms to capture long-range dependencies within input data. Unlike convolutional neural networks (CNNs), which mainly focus on local feature extraction, transformer-based models can establish relationships between distant image regions and therefore model broader contextual information [80]. Figure 2.17 illustrates this difference, where CNNs emphasize localized feature extraction while transformer-based architectures capture global spatial relationships across the image.

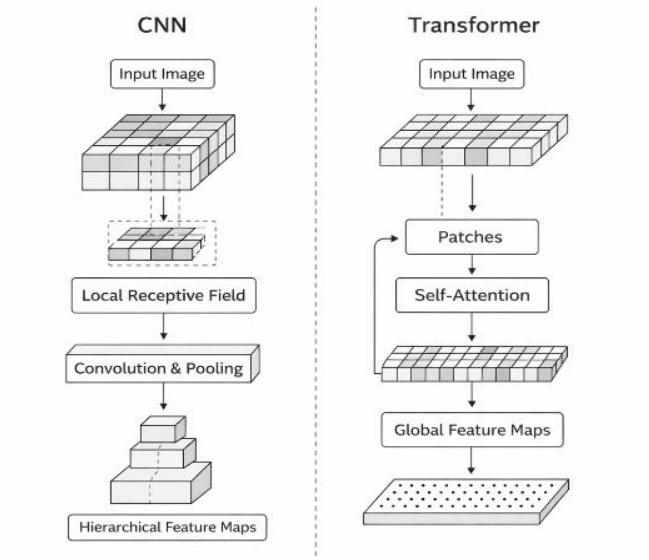


Figure 2.17 Comparison of convolutional neural networks and transformer-based architectures for local and global feature modeling

In medical imaging, transformer-based models can capture long-range spatial relationships and global context, making them effective for anatomically complex structures such as vascular networks [81]. To overcome their high computational demands, hybrid architectures combine convolutional layers for local feature extraction with transformer blocks for global context modeling. Swin-UNet, illustrated in Figure 2.18, is a representative example of this approach [82].

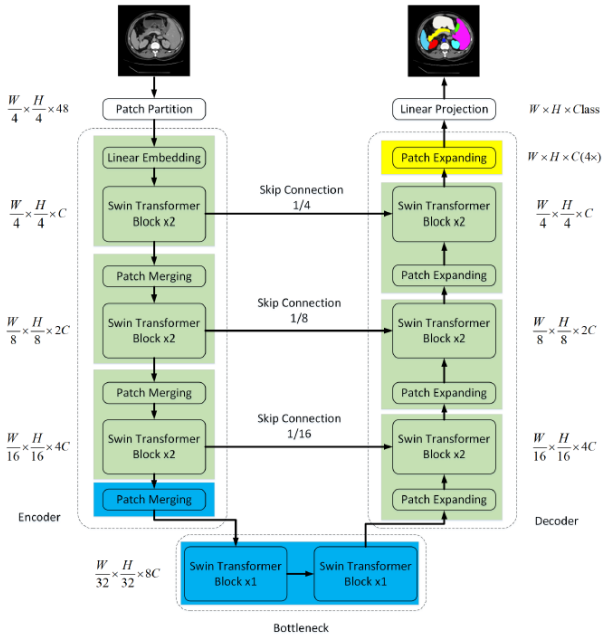


Figure 2.18 Swin-UNet architecture combining transformer-based feature extraction with encoder–decoder segmentation [82]

Although hybrid architectures improve segmentation performance by combining local and global feature representations, they may still remain computationally

demanding and can experience difficulties when segmenting small or low-contrast anatomical structures [83]. These limitations are especially relevant in pediatric cardiovascular imaging and motivate the development of more efficient and structure-aware segmentation frameworks.

2.4.4 Self-Configuring Frameworks

Self-configuring frameworks aim to automate segmentation pipeline design by adapting model configurations according to dataset-specific characteristics. These systems analyze properties such as image resolution, voxel spacing, and dataset size to determine suitable preprocessing strategies, network architectures, and training settings [84], [85]. nnU-Net is one of the most widely used examples of this approach. Rather than proposing a completely new architecture, nnU-Net focuses on automatically configuring the full segmentation pipeline, including preprocessing, network design, and training procedures, as illustrated in Figure 2.19.

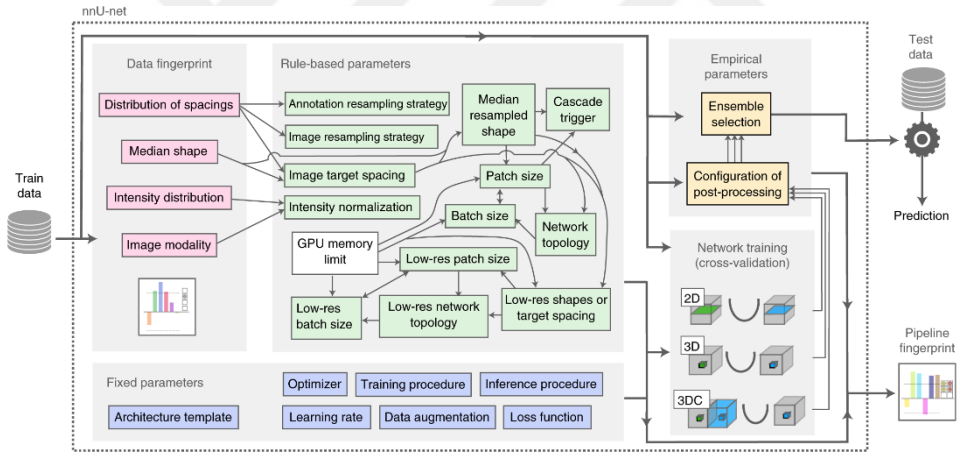


Figure 2.19 Overview of the nnU-Net self-configuring framework for automated preprocessing, architecture selection, and training configuration [85]

Self-configuring frameworks often provide strong baseline performance across a wide range of datasets while reducing the need for extensive manual tuning [85], [86]. However, their generalized design may not fully address task-specific challenges, especially in anatomically complex applications involving small-caliber vascular structures. In pediatric cardiovascular imaging, these limitations become more apparent in pulmonary artery segmentation, where preserving fine structural details and vascular continuity is particularly important. Although self-configuring approaches offer robust baseline performance, they do not explicitly incorporate

structure-specific constraints, which motivates the development of more targeted and structure-aware segmentation strategies.

2.4.5 Localization-Guided Segmentation

Localization-guided segmentation, also known as coarse-to-fine segmentation, aims to overcome the limitations of full-volume processing by focusing on anatomically relevant regions. In volumetric medical images, target structures often occupy only a limited portion of the image volume, while processing the entire image may introduce unnecessary background information that increases computational cost and may reduce segmentation accuracy [87], [88]. To address this issue, localization-guided methods typically use a two-stage pipeline in which a localization model first identifies a region of interest (ROI) containing the target anatomy, followed by detailed segmentation within the cropped region, as illustrated in Figure 2.20 and reported in previous coarse-to-fine segmentation studies [89].

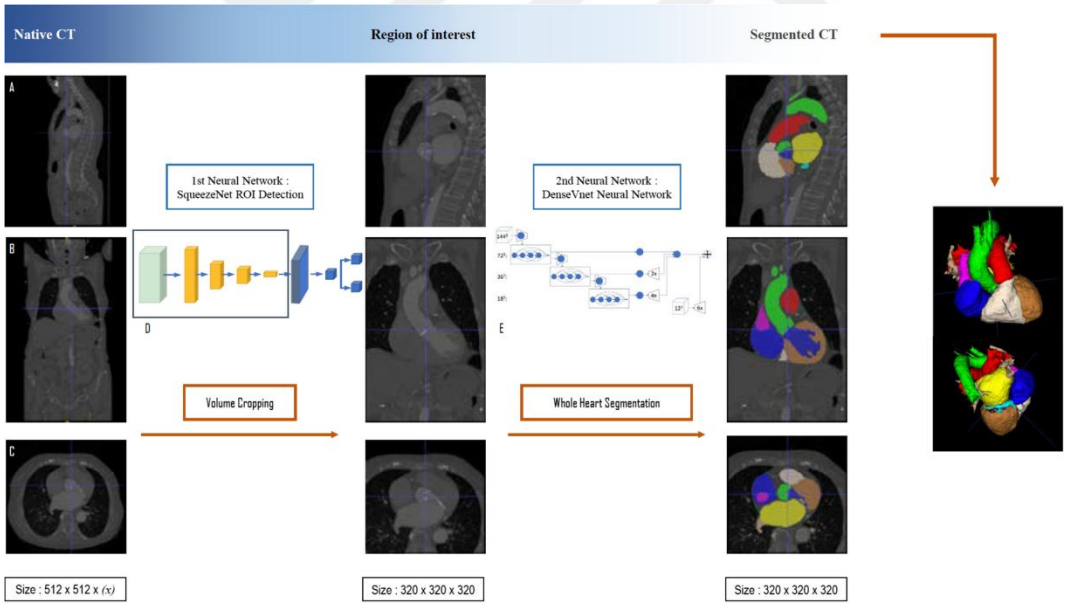


Figure 2.20 Localization-guided segmentation in cardiac CT. (A–C) Volumetric images. (D) ROI localization. (E) Cropped ROI segmentation [89]

By restricting analysis to anatomically relevant regions, these approaches reduce background interference and improve the ability of segmentation models to capture fine structural details. This advantage is important for thin vascular structures where low contrast and partial volume effects complicate accurate delineation. In pediatric cardiovascular imaging, localization-guided methods may help preserve structural

continuity in branching vascular systems such as the pulmonary arteries while also improving computational efficiency by reducing input volume size.

Nevertheless, localization alone does not completely resolve challenges associated with anatomical variability and heterogeneous multi-center imaging conditions. These limitations motivate the development of integrated frameworks that combine localization-guided processing with structure-aware modeling and robust learning strategies under heterogeneous clinical settings, as proposed in this thesis.

2.4.6 Previous Studies on Pediatric Cardiovascular Segmentation

Prior studies have demonstrated the feasibility of automated cardiovascular segmentation in congenital heart disease (CHD), particularly through whole-heart segmentation from cardiovascular magnetic resonance (CMR) images. Architectures such as U-Net, V-Net, nnU-Net, and hybrid CNN–transformer models have shown strong performance on benchmark datasets [8], [9], [85]. Public datasets including HVSMR-2.0 have played an important role by providing three-dimensional CMR images of CHD patients together with manual annotations of cardiac chambers and great vessels, including the pulmonary artery [113]. These studies demonstrate the clinical relevance of automated segmentation for patient-specific three-dimensional modeling, surgical planning, and quantitative cardiovascular assessment.

Despite these developments, several challenges remain insufficiently addressed. Most existing studies focus on benchmark datasets or relatively controlled imaging conditions and are commonly limited to single-center or single-modality data. Although whole-heart segmentation methods successfully capture major anatomical structures, accurate delineation of small-caliber and branching vessels, ,the pulmonary arteries, remains difficult. This problem becomes more pronounced in pediatric CHD cases because developmental variability, congenital abnormalities, and previous surgical interventions may substantially alter anatomical structure.

Another critical limitation is the lack of evaluation under realistic heterogeneous multi-center conditions. In this thesis, a clinically curated pediatric cardiovascular dataset was constructed using hospital-based CT and MRI data collected in collaboration with clinical and radiology experts. The dataset was systematically

organized, standardized, and annotated for multi-class segmentation, providing a clinically representative resource for model development and evaluation. Combined with public benchmark datasets, this framework enables broader assessment of segmentation robustness across modality differences, anatomical variability, and multi-center heterogeneity. The inclusion of both CT and MRI data also supports evaluation under multi-modality conditions, which remains relatively limited in existing pediatric cardiovascular segmentation studies.

Accordingly, this thesis does not present the first study on pediatric cardiovascular segmentation. Instead, it focuses on a combination of challenges that remain comparatively underexplored, including small-vessel segmentation, pulmonary artery delineation, pediatric CHD-related anatomical variability, multi-center imaging heterogeneity, and privacy-aware collaborative learning. These challenges motivate the proposed localization-guided and federated segmentation framework, which aims to improve anatomical focus, preserve fine vascular structures, and support more robust segmentation under heterogeneous multi-center clinical conditions.

To the best of current knowledge, few existing studies jointly investigate small-vessel segmentation, pediatric CHD variability, and multi-center heterogeneity within a privacy-preserving learning framework.

2.5 Challenges in Small-Vessel Segmentation in Medical Imaging

Accurate segmentation of small vascular structures remains one of the major challenges in medical image analysis. Small vessels typically have thin diameters, complex branching patterns, and low contrast relative to surrounding tissues, making them highly sensitive to noise, motion artifacts, partial volume effects, and variability in acquisition protocols, in three-dimensional computed tomography (CT) and magnetic resonance imaging (MRI) data [90]. One important factor affecting segmentation performance is the partial volume effect, where a voxel may contain a mixture of vessel lumen and surrounding tissue when vessel diameter approaches voxel size. This effect produces blurred boundaries and ambiguous intensity transitions that complicate accurate vessel delineation for both classical and deep learning–based methods [91]. Figure 2.21 illustrates the partial volume effect and its influence on boundary clarity.

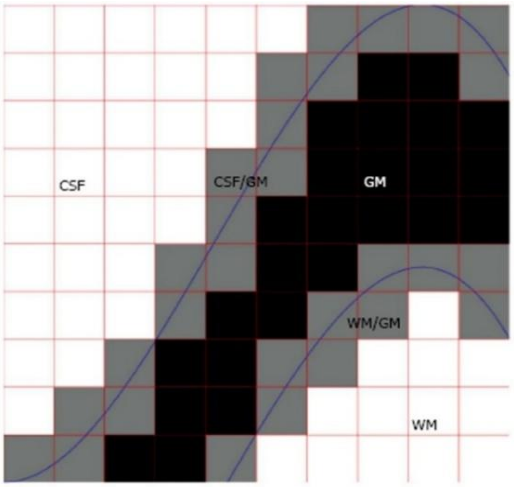


Figure 2.21 Illustration of the partial volume effect causing blurred boundaries and reduced segmentation accuracy [91]

Another critical challenge is class imbalance. In many vascular segmentation tasks, vessel voxels occupy only a small portion of the image volume, whereas background voxels dominate the dataset. This imbalance may bias segmentation models toward background predictions and increase the likelihood of missing thin vessel branches. Consequently, overlap-based and class-weighted optimization strategies are often used to improve sensitivity to underrepresented vascular structures [92], [95]. Boundary uncertainty caused by noise, motion artifacts, contrast variability, reconstruction differences, and limited spatial resolution may further produce indistinct vessel borders and fragmented predictions [93]. Representative failure patterns, including fragmented branches, blurred vessel boundaries, and missing thin structures, are shown in Figure 2.22.

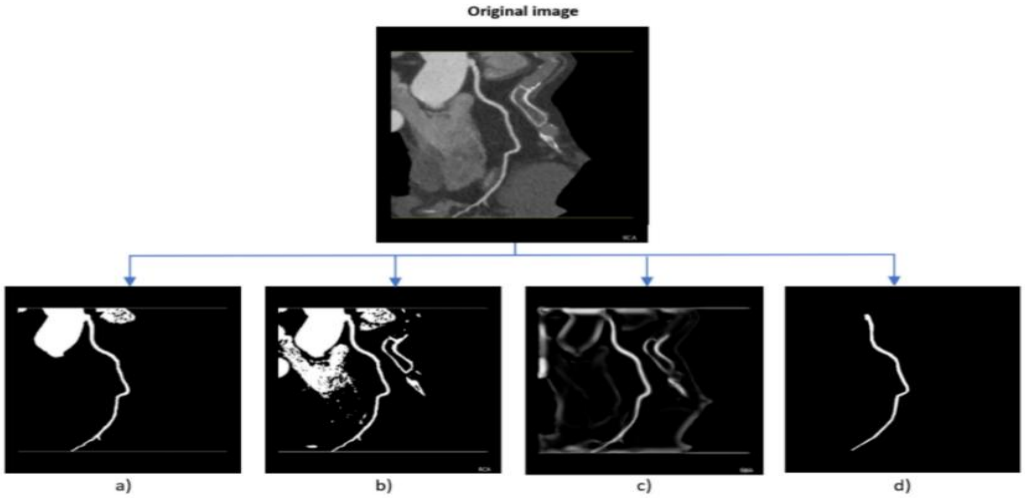


Figure 2.22 Challenges in small-vessel segmentation. (a) Baseline segmentation. (b) Fragmentation. (c) Boundary ambiguity. (d) Missing thin structures [94]

Recent studies have proposed boundary-aware and topology-aware learning strategies to reduce these limitations. Boundary-aware approaches focus on improving contour localization, whereas topology-aware methods aim to preserve vessel connectivity and reduce fragmentation in branching structures [92], [95]. Nevertheless, maintaining stable performance remains difficult under conditions involving substantial anatomical variability and heterogeneous imaging data.

These challenges become especially important in pulmonary artery segmentation within pediatric cardiovascular imaging. Pulmonary arteries are clinically significant but structurally complex because of their small diameters, branching geometry, and sensitivity to imaging variability. In multi-center datasets, differences in acquisition protocols, spatial resolution, and image quality further increase segmentation difficulty. As a result, accurate pulmonary artery segmentation requires models capable of preserving fine vascular details while maintaining global anatomical continuity.

Overall, small-vessel segmentation represents a central challenge addressed in this thesis. Partial volume artifacts, class imbalance, boundary uncertainty, and structural discontinuities continue to limit the performance of conventional segmentation approaches, particularly in pediatric cardiovascular imaging where anatomical variability and multi-center heterogeneity further complicate analysis. These challenges motivate the development of structure-aware and localization-guided segmentation strategies such as the framework proposed in Chapter 3.

2.6 Evaluation Metrics for Medical Image Segmentation

Quantitative evaluation plays a key role in assessing the performance of medical image segmentation algorithms. In three-dimensional segmentation tasks, evaluation metrics are used to measure the agreement between predicted segmentation masks and corresponding ground-truth annotations. These metrics provide objective criteria for comparing segmentation models and assessing their suitability for clinical applications [23], [96].

One major category consists of overlap-based metrics, which quantify spatial agreement between predicted and reference segmentations. Among these, the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) are the most widely

used measures in biomedical image segmentation. DSC measures the degree of overlap between predicted and ground-truth masks, whereas IoU evaluates the ratio between the intersection and union of predicted and reference segmentations.

The Dice Similarity Coefficient (DSC) is defined as:

$$\mathrm{DSC}(A,B) = \frac{2\ |\ A \cap B\ |}{|\ A\ | + |\ B\ |} \tag{2.5}$$

Similarly, the Intersection over Union (IoU) is defined as:

$$\mathrm{IoU}(A,B) = \frac{|\ A \cap B\ |}{|\ A \cup B\ |} \tag{2.6}$$

where A and B denote the predicted and ground-truth segmentation masks, respectively. Both DSC and IoU provide global measures of segmentation accuracy and are widely used for benchmarking across datasets and tasks [97], [98]. Figure 2.23 provides a visual illustration of these metrics, complementing their mathematical definitions.

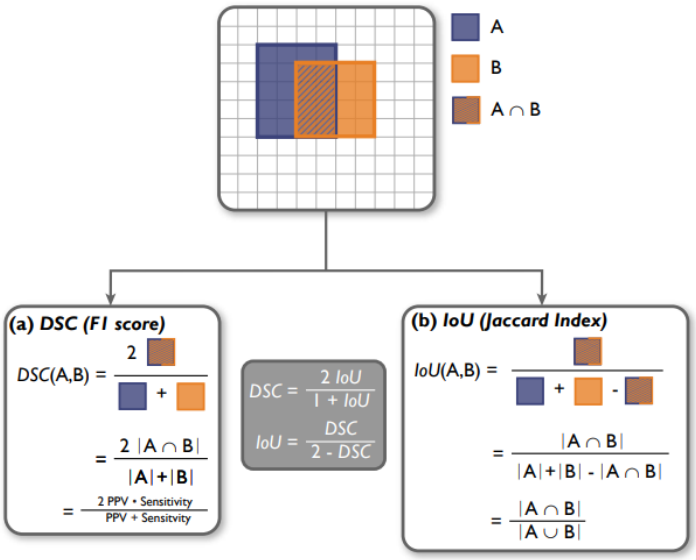


Figure 2.23 Illustration of overlap-based segmentation metrics: (a) Dice Similarity Coefficient (DSC) and (b) Intersection over Union (IoU) [98]

In addition to overlap-based metrics, surface- and boundary-based metrics are commonly used to evaluate geometric segmentation accuracy. Metrics such as the Hausdorff Distance (HD) and Average Surface Distance (ASD) measure distances between the boundaries of predicted and reference segmentations. HD reflects the maximum boundary deviation, whereas ASD represents the mean boundary discrepancy.

The Hausdorff Distance is defined as:

$$\mathrm{HD}(A,B) = \max\left\{\sup_{a\in A}\inf_{b\in B} d(a,b), \sup_{b\in B}\inf_{a\in A} d(b,a)\right\} \tag{2.7}$$

The Average Surface Distance (ASD) is defined as:

$$\mathrm{ASD}(A,B) = \frac{1}{|A|}\sum_{a\in A}\min_{b\in B} d(a,b) \tag{2.8}$$

These metrics provide complementary information to overlap-based measures by capturing boundary-level discrepancies that may not be reflected in global overlap statistics [24], [97]. Examples are illustrated in Figures 2.24 and 2.25.

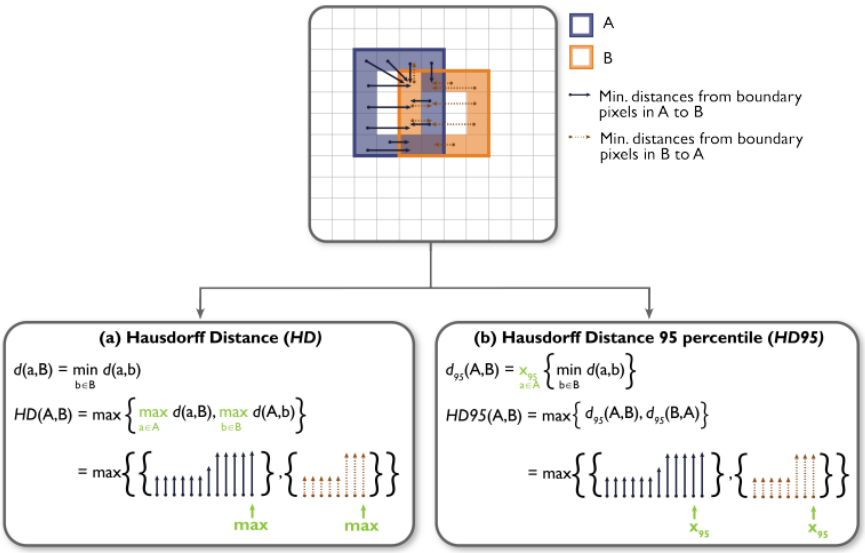


Figure 2.24 Illustration of distance-based metrics: (a) Hausdorff Distance (HD) and (b) 95th percentile Hausdorff Distance (HD95) [98]

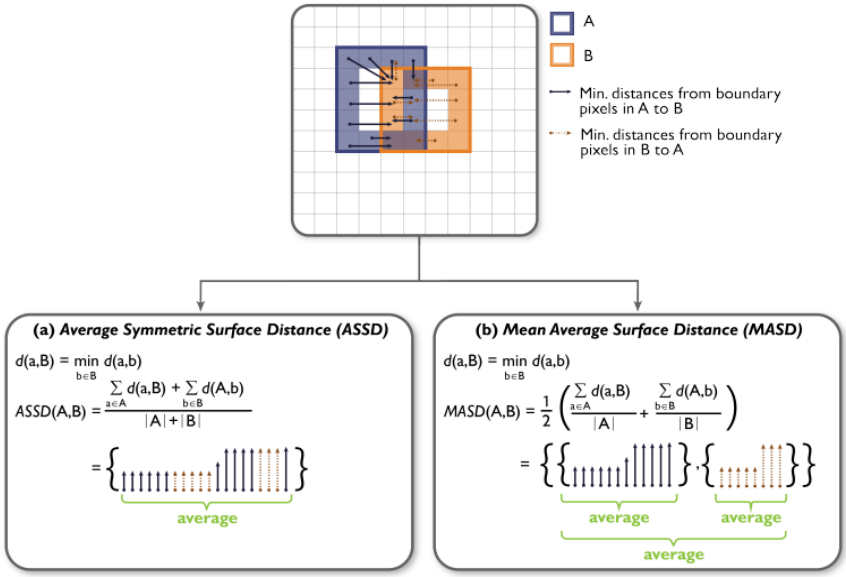


Figure 2.25 Illustration of surface-based metrics: (a) Average Symmetric Surface Distance (ASSD) and (b) Mean Average Surface Distance (MASD) [98]

Although overlap-based metrics such as DSC are widely used, relying only on these measures may not fully capture segmentation quality, especially in small-vessel segmentation tasks. Small discontinuities or missing vessel branches may have limited influence on global overlap scores while substantially affecting anatomical correctness. Consequently, models achieving high Dice scores may still produce fragmented or anatomically incomplete vascular structures. This limitation highlights the importance of boundary-based and structure-aware evaluation strategies [98], [99].

To obtain a more clinically meaningful assessment, segmentation performance is often evaluated separately for individual anatomical structures rather than the entire image volume, particularly for small and clinically important vessels such as the pulmonary arteries. In addition to quantitative metrics, qualitative visualization-based analyses are increasingly used to interpret model behavior, identify uncertain regions, and reveal potential failure patterns in anatomically complex segmentation tasks. Together, these approaches provide a more comprehensive evaluation of segmentation performance, as summarized in Table 2.2 [34], [35].

Table 2.1 Common evaluation metrics in three-dimensional medical image segmentation and small-vessel analysis

Metric	Measures	Limitation	Relevance to small-vessel segmentation
Dice Similarity Coefficient (DSC)	Volumetric overlap between prediction and ground truth	Insensitive to small discontinuities and topology errors	May overestimate performance when thin vessels are missing
Intersection over Union (IoU)	Ratio of intersection to union	Limited sensitivity to boundary accuracy	Similar limitation to DSC for small structures
Hausdorff Distance (HD / HD95)	Maximum or percentile-based boundary deviation	Sensitive to outliers	Captures extreme boundary errors in small vessels
Average Surface Distance (ASD / ASSD)	Mean boundary distance	May not capture structural continuity	Reflects boundary smoothness but not connectivity

Together, these metrics provide complementary perspectives on segmentation performance by capturing both volumetric overlap and boundary-level accuracy. This is critical for evaluating small and anatomically complex vascular structures such as pulmonary arteries in pediatric cardiovascular imaging. However, these metrics are usually computed under controlled experimental conditions, and their reliability may be affected in real-world clinical scenarios because of data variability and domain shift [101], [102], as discussed in the following section.

2.7 Domain Shift and Generalization in Medical Imaging

Deep learning–based models have achieved strong performance in medical image analysis in recent years. However, successful clinical deployment depends not only on high accuracy during training but also on the ability to maintain reliable performance under real-world conditions. In medical imaging, generalization refers to the capability of a model to perform consistently on data distributions that differ from those encountered during training. Achieving robust generalization remains difficult because imaging data often vary across institutions, patient populations, and acquisition protocols [100].

One of the primary factors limiting generalization is domain shift, which occurs when the statistical properties of training data differ from those of deployment data. Under these conditions, models trained on a specific dataset may experience performance degradation when applied to previously unseen data. In medical imaging, domain shift commonly arises from differences in scanner vendors, acquisition settings, reconstruction protocols, and patient populations, resulting in variability in image appearance, intensity distributions, noise patterns, spatial resolution, and anatomical representation [101]. Figure 2.26 conceptually illustrates domain shift between source and target domains and its effect on model predictions.

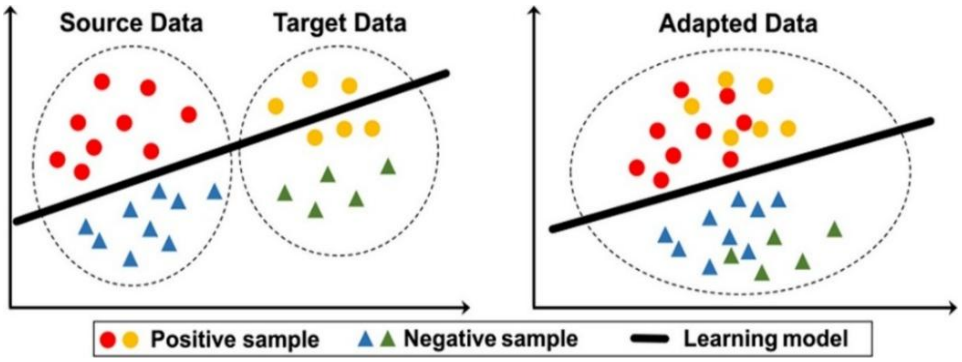


Figure 2.26 Conceptual illustration of domain shift between source and target domains and the role of domain adaptation [101]

Domain shift becomes more pronounced in multi-center clinical datasets, where imaging data are collected from different hospitals or institutions. Each center may use different imaging devices, acquisition protocols, and preprocessing workflows, producing heterogeneous datasets with varying statistical properties. In machine learning, this setting is often described as non-independent and identically distributed (non-IID) data, where data distributions differ across sources [101], [102]. Figure 2.27 illustrates an example of inter-center variability through differences in intensity distributions across imaging centers.

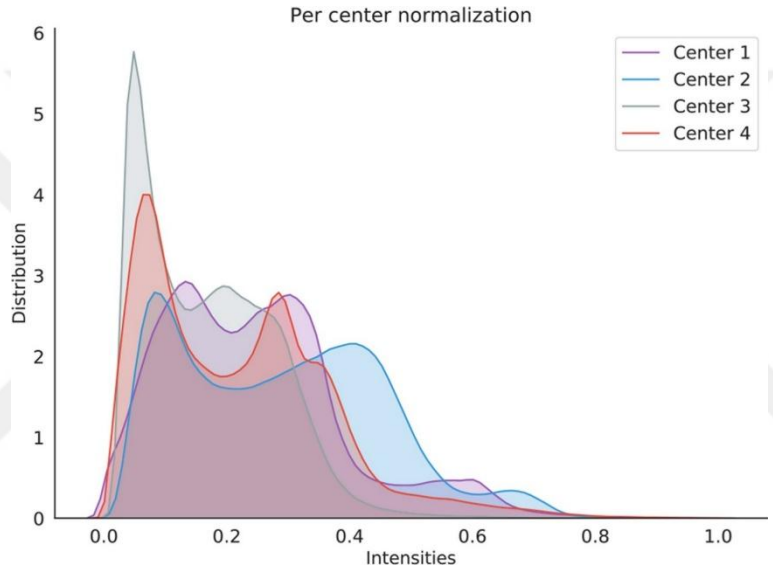


Figure 2.27 Example of intensity distribution differences across imaging centers, illustrating inter-center variability and domain shift [101]

The presence of domain shift creates a major obstacle for reliable clinical deployment of artificial intelligence systems. Models trained on single-center datasets may fail to generalize to unseen clinical environments, reducing reliability in real-world applications [101]. This issue becomes particularly important in pediatric cardiovascular imaging, where anatomical variability and modality differences further increase dataset heterogeneity.

These challenges motivate the development of learning strategies designed to improve robustness under heterogeneous data conditions. Approaches focusing on cross-domain variability and multi-center learning have therefore received increasing attention. Among these, federated learning has emerged as a promising framework for collaborative and privacy-preserving model training across

distributed clinical datasets. This perspective directly supports the localization-guided and federated segmentation framework proposed in this thesis.

2.8 Learning Strategies for Multi-Center Medical Imaging

Medical imaging datasets used for training artificial intelligence models are increasingly collected from multiple clinical institutions. Although multi-center datasets provide valuable diversity in patient populations and imaging conditions, they also introduce two major challenges: restricted data sharing and distributional heterogeneity across centers [103]. Variations in scanner vendors, acquisition protocols, reconstruction algorithms, and contrast administration strategies may produce substantial differences in data distributions between institutions. These differences introduce domain shifts that can reduce the generalization performance of models trained using conventional centralized learning approaches [106]. This issue is closely related to the domain shift problem discussed in the previous section.

At the same time, privacy regulations and institutional policies often limit direct sharing of patient imaging data between hospitals. Ethical considerations and legal constraints frequently require medical data to remain within the institution where they were acquired. As a result, constructing centralized datasets that aggregate imaging data across institutions may become difficult in practice, limiting access to the large and diverse datasets typically required for robust deep learning models [104], [105].

To address the challenges of multi-center medical imaging, collaborative learning approaches have been developed to improve model robustness across heterogeneous data distributions while preserving patient privacy. In this framework, each institution trains a local model on its private dataset and shares only model updates with a central server, which aggregates the updates and redistributes the global model for continued collaborative training, as illustrated in Figure 2.28.

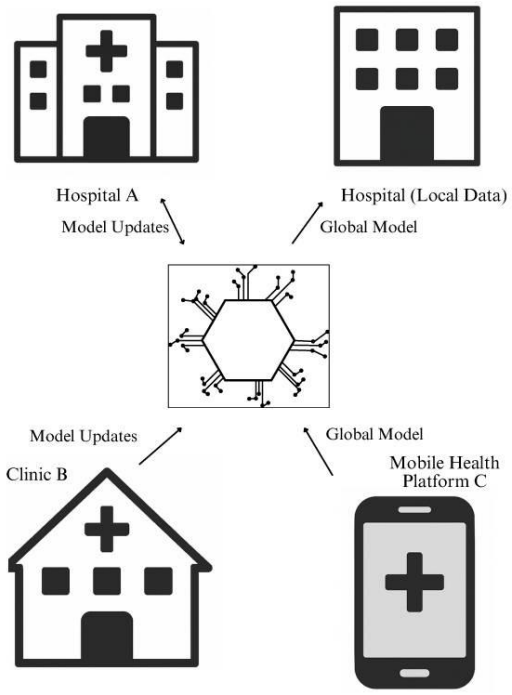
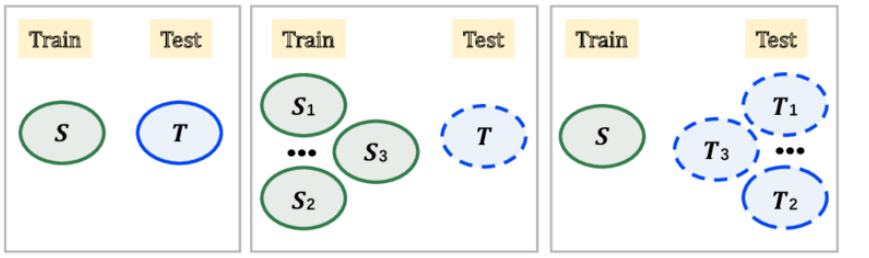


Figure 2.28 Conceptual illustration of federated learning with collaborative training across multiple institutions without raw data sharing [104]

These learning paradigms provide important strategies for handling heterogeneous multi-center medical imaging datasets and form the methodological basis of the federated and privacy-aware learning framework proposed in this thesis.

2.8.1 Domain Adaptation and Generalization

Variability across medical imaging datasets collected from different institutions creates important challenges for deep learning models. Differences in imaging protocols, scanner vendors, reconstruction settings, and patient populations may introduce distributional discrepancies between training and deployment data. To address these challenges, domain adaptation and domain generalization have emerged as two major learning paradigms for improving robustness across heterogeneous medical imaging domains [107], [108]. Figure 2.29 presents a conceptual comparison of these approaches.



(a) Domain adaptation (b) Domain generalization (c) Single domain generalization

Figure 2.29 Domain adaptation and generalization. (a) Source-to-target adaptation. (b) Multi-domain generalization. (c) Single-domain generalization [109]

Domain adaptation aims to improve performance on a target domain whose data distribution differs from the source domain. In this setting, models are generally trained using labeled source-domain data together with limited or unlabeled target-domain data. The goal is to reduce discrepancies between domains through methods such as feature alignment, adversarial learning, and representation regularization [108].

Domain generalization, in contrast, aims to develop models capable of generalizing to previously unseen domains without requiring access to target-domain data during training. Rather than adapting to a specific target distribution, these approaches attempt to learn domain-invariant feature representations that remain robust under distribution shifts. This strategy is particularly relevant in medical imaging scenarios where future clinical data may not be available during model development [109].

In multi-center medical imaging applications, both paradigms seek to maintain stable segmentation performance across institutions with varying imaging conditions. Nevertheless, critical limitations remain. Domain adaptation often requires access to target-domain data, which may be restricted in clinical settings, whereas domain generalization may struggle to fully capture the complexity of real-world variability. Consequently, a gap still exists between controlled experimental evaluation and practical deployment under heterogeneous multi-center conditions. These limitations motivate the exploration of alternative learning strategies capable of addressing both data heterogeneity and privacy constraints. Federated learning represents one such approach and is discussed in the following section.

2.8.2 Federated Learning

Federated learning has emerged as an important paradigm for collaborative machine learning in settings where direct data sharing is restricted. Instead of transferring medical imaging data to a centralized repository, federated learning enables multiple institutions to jointly train a shared model while keeping local datasets within their own infrastructure. In this framework, each participating center performs local training using its private data and periodically shares only model

parameters or gradients with a central aggregation server. The server combines these updates to produce a global model that is redistributed to participating institutions, enabling collaborative learning without exposing sensitive patient data [110].

A major advantage of federated learning in medical imaging is its ability to support privacy-preserving collaboration across healthcare institutions. Because raw patient data remain local, this approach is compatible with data protection regulations and institutional privacy policies while still allowing models to benefit from the diversity of multi-center datasets. As a result, federated learning may improve robustness and generalization across heterogeneous clinical environments [111]. The general workflow of federated learning is illustrated in Figure 2.30.

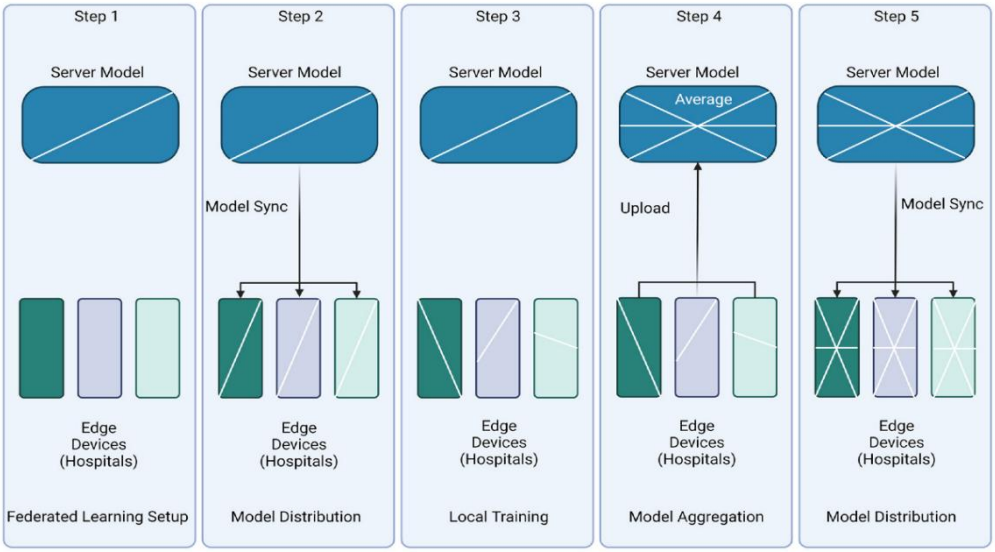


Figure 2.30 Federated learning workflow with global model distribution, local training, and parameter aggregation [111]

A central server first initializes a global model and distributes it to participating institutions. Each institution updates the model using its local dataset and sends the updated parameters back to the server, where they are aggregated, commonly through weighted averaging, to produce a new global model for the next training round. The aggregation process can be expressed as:

$$\theta^{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \theta_k^{t+1} \tag{2.9}$$

where θ_k^{t+1} represents the locally updated parameters at client k, n_k is the number of training samples at client k, and $n = \sum_{k=1}^{K} n_k$ denotes the total number of

samples across all clients. This weighted averaging strategy, known as Federated Averaging (FedAvg), enables collaborative optimization while preserving data privacy.

Despite its advantages, federated learning introduces several challenges, mainly because data across institutions are often non-independent and identically distributed (non-IID). Differences in imaging protocols, scanner vendors, patient populations, and annotation practices may produce heterogeneous local datasets, complicating optimization and potentially affecting convergence stability and segmentation performance [112].

In pediatric cardiovascular imaging, federated learning provides a practical framework for multi-center collaboration because congenital heart disease datasets are frequently distributed across institutions and may remain limited in size within individual centers. By supporting collaborative model development without centralized data sharing, federated learning enables the use of heterogeneous datasets while preserving patient privacy. However, substantial challenges remain, in the segmentation of small and anatomically complex vascular structures such as pulmonary arteries, where low contrast, anatomical variability, and imaging-related limitations further increase task difficulty.

These limitations motivate the proposed localization-guided federated framework for accurate and robust multi-class segmentation under heterogeneous multi-center conditions. The methodology is presented in Chapter 3.

3
METHODOLOGY

This chapter presents the methodological framework developed for multi-class three-dimensional segmentation of pediatric cardiovascular structures under heterogeneous multi-center imaging conditions. The proposed pipeline included modality-aware preprocessing, landmark-based localization, region-of-interest (ROI) extraction, hybrid V-Net-based segmentation, and federated learning-based distributed training. Localization-guided ROI extraction and transformer-enhanced hybrid V-Net segmentation were integrated to improve anatomical focus and enable combined local-global feature representation. In addition to the proposed framework, baseline segmentation architectures were comparatively evaluated under centralized and federated learning settings. Ensemble prediction and explainability analyses were also included as supplementary components. The following sections describe the overall framework, preprocessing procedures, model architectures, federated learning strategy, implementation details, and evaluation protocols.

3.1 Overall Framework

The proposed system was designed as a localization-guided segmentation framework to improve robustness under heterogeneous multi-center conditions. Two processing pathways were defined. In the direct pathway, segmentation was performed on full-volume standardized inputs as a baseline configuration. In the localization-guided pathway, a dedicated localization network predicted anatomical landmark positions to define a patient-specific region of interest (ROI). The extracted ROI was standardized and provided to the segmentation network to reduce background influence and improve sensitivity to small anatomical

structures. The segmentation stage was based on a hybrid V-Net architecture enhanced with transformer-based attention, representing the primary methodological contribution of this study. For systematic evaluation, baseline architectures including 3D U-Net, V-Net, nnU-Net, and Swin-UNet were also evaluated under centralized and federated learning settings. In the centralized configuration, all data were aggregated to train a single global model, whereas in the federated configuration, data remained locally distributed and only model parameters were shared to enable privacy-preserving learning under heterogeneous conditions. Ensemble inference and explainability analyses were included as post-hoc evaluation tools. An overview of the framework is presented in Figure 3.1.

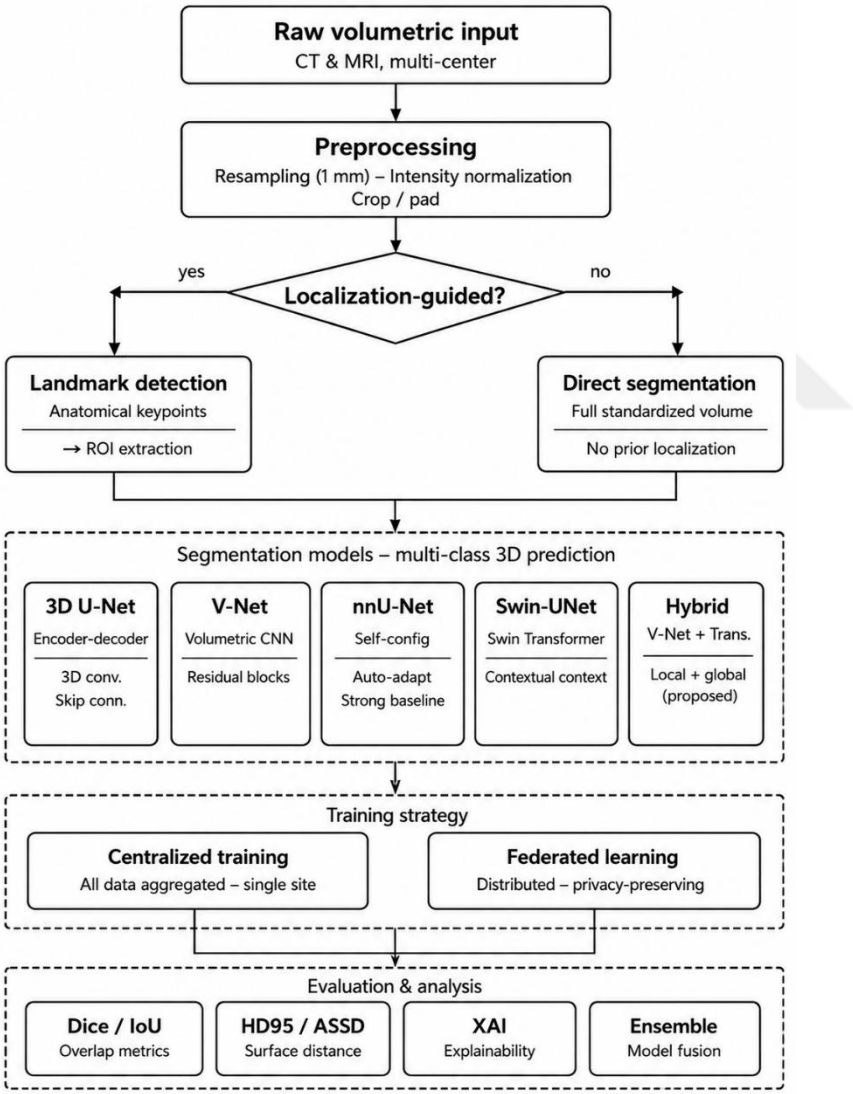


Figure 3.1 Overview of the proposed localization-guided segmentation framework

To clearly define the experimental design and the relationships between different methodological configurations, a summary of the evaluated settings is provided in Table 3.1.

Table 3.1 Summary of methodological configurations used in this study

Setting ID	Input	Localization	Training	Model Scope	Purpose
S1	Full-volume	No	Centralized	All architectures	Baseline reference
S2	ROI-based	Yes	Centralized	All architectures	Effect of localization-guided segmentation
S3	Full-volume	No	Federated	All architectures	Effect of federated learning
S4	ROI-based	Yes	Federated	All architectures	Proposed full framework (FL + Localization)

These four configurations enabled controlled evaluation of (i) localization-guided versus full-volume segmentation, (ii) centralized versus federated training, and (iii) their combined effect within a unified experimental framework.

3.2 Dataset Description and Organization

The proposed pipeline operated on three-dimensional volumetric images and corresponding segmentation masks organized into structured dataset partitions. The study dataset consisted of two components: the public HVSMR-2.0 benchmark dataset [113] and a clinically curated multi-center cohort, together forming a heterogeneous setting for pediatric cardiovascular segmentation. The benchmark dataset provided a standardized reference for controlled evaluation, whereas the clinical cohort introduced real-world variability in acquisition protocols, imaging modalities, and anatomical complexity. All data were processed using standard medical imaging formats (e.g., NIfTI and NRRD) within a unified preprocessing framework.

The HVSMR-2.0 dataset included 60 cardiac MRI volumes with expert annotations, while the clinical cohort consisted of 50 cases collected from four institutions and was predominantly CT-based (48 CT and 2 MRI). The combined cohort therefore contained 110 cases with mixed modality characteristics. For experimental consistency, subject-level partitioning was applied using 92 training, 10 validation, and 8 test cases, ensuring no overlap between splits. This organization supported both benchmark and heterogeneous multi-center evaluation settings under centralized and federated learning configurations. A summary of the dataset composition and experimental configurations is presented in Table 3.2.

Table 3.2 Summary of experimental dataset configurations

Dataset Configuration	Cases	Modality	Evaluation Purpose
HVSMR-2.0	60	MRI	Benchmark evaluation
Clinical Dataset	50	CT / MRI	Clinical evaluation
Combined Cohort	110	Mixed	Heterogeneous evaluation

3.2.1 Public Benchmark Dataset (HVSMR-2.0)

The public benchmark dataset used in this study was the HVSMR-2.0 dataset [113], an open-access three-dimensional cardiovascular magnetic resonance (CMR) dataset developed for whole-heart segmentation in patients with congenital heart disease. The dataset included complex congenital anatomies with substantial inter-subject variability, providing a standardized benchmark for segmentation evaluation.

The dataset included 60 CMR scans acquired from subjects with congenital heart disease across different severity levels (12 mild, 11 moderate, and 37 severe cases), capturing variability in cardiac size, morphology, and developmental stage. All scans were acquired at Boston Children's Hospital using standardized clinical imaging protocols. Although the images exhibited near-isotropic resolution, spatial dimensions varied across subjects, requiring preprocessing and standardization [113]. Each scan was accompanied by expert-validated whole-heart segmentation annotations covering eight cardiovascular structures, including the ventricles, atria, aorta, pulmonary artery, and major veins.

The ground-truth annotations were generated through a multi-stage expert-guided workflow involving semi-automatic segmentation and clinical validation by pediatric cardiologists and imaging specialists [113]. In this study, these labels were reorganized into target classes aligned with the segmentation objectives of the proposed framework. The HVSMR-2.0 dataset served as a benchmark reference for standardized evaluation. Representative examples are shown in Figure 3.2, illustrating inter-subject anatomical variability and structural complexity.

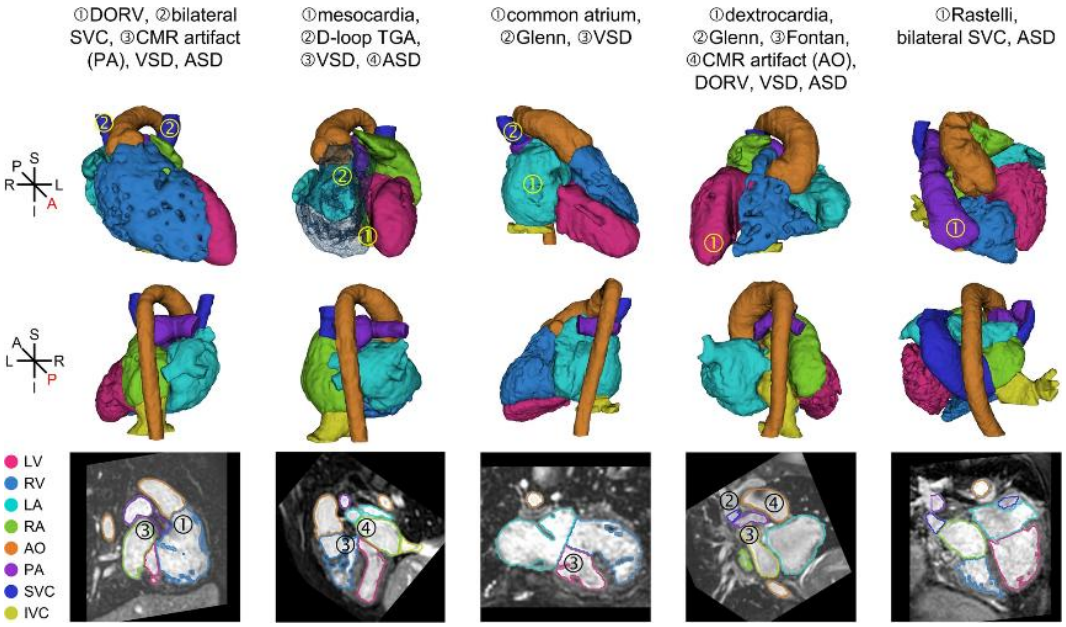


Figure 3.2 Representative whole-heart segmentations from the HVSMR-2.0 dataset illustrating anatomical variability in congenital heart disease, including differences in cardiac morphology, vessel orientation, and heart size [113]

3.2.2 Multi-Center Clinical Dataset

In addition to the public benchmark dataset, this study utilized a clinically curated multi-center pediatric cardiovascular imaging cohort consisting of 50 imaging studies, including 48 computed tomography (CT) scans and 2 magnetic resonance (MR) examinations. The cohort was collected during routine clinical care for congenital heart disease diagnosis and perioperative assessment. Its predominantly CT-based composition reflects current clinical practice, where CT is preferred for its high spatial resolution and rapid acquisition, while the inclusion of MR studies provided limited modality variation.

The imaging studies were retrospectively collected from four tertiary care centers: Istanbul Mehmet Akif Ersoy Training and Research Hospital, Medipol Hospitals, Acibadem Hospitals, and Başakşehir Çam and Sakura City Hospital. These

institutions introduced natural variability in scanner models, acquisition protocols, and case complexity, enabling evaluation under realistic multi-center conditions. Cases with confirmed congenital heart disease and high-quality cardiovascular imaging were included, whereas studies with severe motion artifacts or incomplete anatomical coverage were excluded. Ethical approval was obtained from the Istanbul Medipol University Non-Interventional Clinical Research Ethics Committee (Approval No. E-10840098-772.02-2946, May 11, 2023), and all data were anonymized prior to analysis.

The patient age range spanned approximately 0–20 years, with strong representation of infants and young children. The cohort included diverse congenital heart disease phenotypes, such as septal defects, outflow tract abnormalities, tetralogy of Fallot, double outlet right ventricle, pulmonary atresia, pulmonary stenosis, and anomalous pulmonary venous return. A summary of the cohort characteristics is provided in Table 3.3.

Table 3.3 Summary of the multi-center clinical dataset

Category	Description
Total number of cases	50
Centers	4 tertiary care hospitals
Age range	0–20 years (infant-dominant distribution)
Imaging modalities	48 CT, 2 MR
CT systems	Siemens SOMATOM Definition; Philips iCT256; Toshiba Aquilion ONE; Canon Alexion
MRI systems	Siemens 3T Skyra; Siemens 1.5T SymphonyTim
Pre-resampling spacing	CT: 0.199–0.976 mm (in-plane), 0.45–1.0 mm slice separation; MRI: 0.625–1.0 mm
Resampled spacing	1.0 mm isotropic
Segmented classes	Aorta, cardiac chambers, pulmonary arteries
Label encoding	0: background; 1: aorta; 2: cardiac chambers; 3: pulmonary arteries
Annotation format	Multi-class label maps
Annotation workflow	Standardized 3D Slicer–based workflow with expert review
Metadata included	Clinical and DICOM-derived technical metadata

As shown in Table 3.3, the data reflect substantial variability in imaging modalities, acquisition settings, and anatomical characteristics. All studies were originally collected in DICOM format and subsequently anonymized. Segmentation was performed using 3D Slicer with manual and semi-automatic approaches under expert clinical supervision. Three anatomical structures were annotated: the aorta, cardiac chambers, and pulmonary arteries. For preprocessing purposes, region-of-interest (ROI) representations were generated by cropping around target structures with a defined margin to ensure consistent input dimensions. Importantly, these ROI definitions were used only during data preparation and not during model inference, where ROI extraction was performed based on predicted anatomical landmarks. Representative multi-planar imaging examples from the clinical dataset are presented in Figure 3.3.

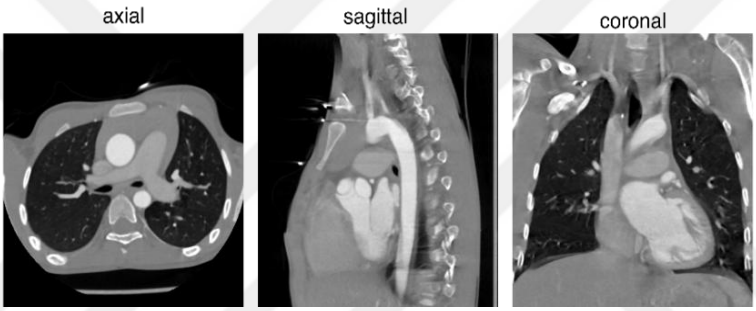


Figure 3.3 Representative multiplanar reconstruction views from the multi-center pediatric cardiovascular imaging dataset

Figure 3.4 presents the corresponding segmentation masks and overlay visualizations for the annotated cardiovascular structures.

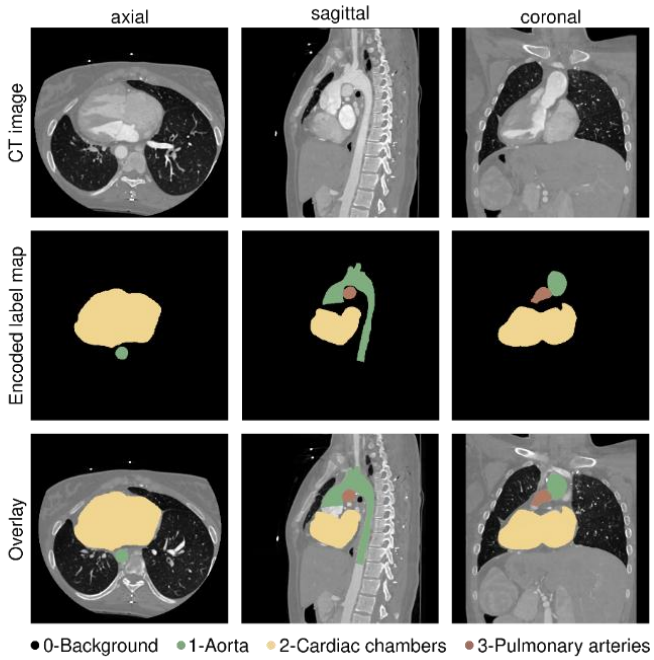


Figure 3.4 Representative multiplanar views with corresponding segmentation masks and overlay visualizations of target cardiovascular structures

This multi-center clinical dataset provided a heterogeneous and clinically grounded resource for pediatric cardiovascular segmentation. Its multi-institutional structure, expert-validated annotations, ROI-based representations, and metadata-rich design supported robust evaluation under realistic clinical conditions and complemented the benchmark dataset.

3.2.3 Combined Multi-Center Dataset Configuration

For federated learning experiments, the clinical dataset was organized according to institutional origin. To simulate a multi-center setting while maintaining sufficient data per client, the four clinical centers were grouped into three federated clients. Smaller centers were combined to form balanced clients, whereas larger centers were assigned individually, preserving both data balance and center-level heterogeneity. The public benchmark dataset was used only in the centralized training configuration and was not distributed across federated clients. Accordingly, only the clinical multi-center cohort participated in federated training, while the HVSMR-2.0 dataset [113] was reserved for centralized experiments. Identical validation and test partitions were maintained across centralized and federated settings.

Segmentation annotations were harmonized across sources and encoded as multi-class label maps with four classes: background (0), aorta (1), cardiac chambers (2), and pulmonary arteries (3). The aorta class included the ascending, arch, and descending segments, while the cardiac chambers class combined atrial and ventricular cavities. The pulmonary artery class included the main trunk and its branches. These structures represented clinically relevant cardiovascular regions with varying segmentation complexity.

For preprocessing, region-of-interest (ROI) volumes were extracted around labeled structures with an additional fixed margin and standardized to $128 \times 128 \times 64$ voxels. These ROI definitions were used only during data preparation and not during inference, where ROI extraction was performed using predicted anatomical landmarks (Section 3.4). The unified target structures and their clinical relevance are summarized in Table 3.4, while the label encoding scheme and representative structures are illustrated in Figures 3.5 and 3.6.

Table 3.4 Summary of segmentation target structures and their clinical relevance

Class Index	Anatomical Structure	Description	Clinical Relevance
0	Background	Non-cardiovascular regions	Not applicable
1	Aorta	Ascending, arch, and descending aorta	Major systemic circulation
2	Cardiac chambers	Combined atria and ventricles	Cardiac morphology and functional assessment
3	Pulmonary arteries	Main, left, and right branches	Pulmonary circulation and CHD evaluation

Following the tabular summary, the label encoding scheme is visually illustrated to clarify the multi-class representation used throughout the study.

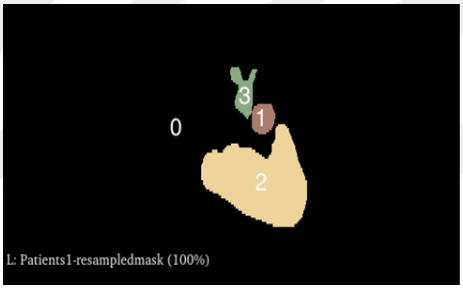


Figure 3.5 Multi-class label encoding scheme for background, aorta, cardiac chambers, and pulmonary artery classes

These visualizations provide qualitative insight into the spatial distribution and structural relationships of the segmented cardiovascular regions.

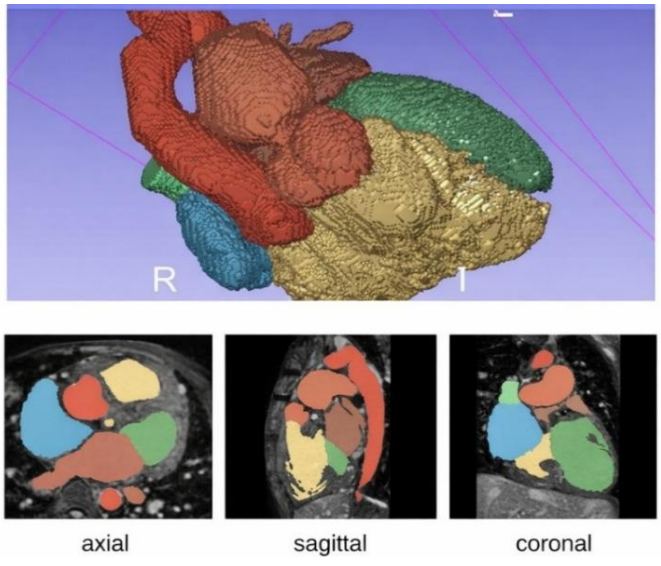


Figure 3.6 Multi-class target structures in three-dimensional reconstruction and multiplanar views. Red: aorta, yellow: cardiac chambers, green: pulmonary arteries

3.2.4 Annotation Protocol and Quality Control

Reliable and consistent segmentation annotations were required for model training and evaluation. Annotation procedures were defined separately for the public benchmark dataset and the clinically curated multi-center dataset and were subsequently harmonized to ensure consistency across data sources.

The HVSMR-2.0 dataset provides expert-validated whole-heart annotations generated through a structured workflow involving semi-automatic segmentation and expert review [113]. In contrast, the clinical multi-center dataset was annotated within this study using a standardized 3D Slicer–based workflow combining manual delineation with semi-automatic tools under expert supervision. The annotation protocol focused on three target structures—the aorta, cardiac chambers, and pulmonary arteries—to ensure consistency across anatomically diverse cases. Predefined labeling guidelines and anatomical boundaries were applied, and all segmentations were iteratively reviewed for consistency. Anatomical landmarks used for ROI extraction, including the aortic root, main pulmonary artery, and pulmonary artery bifurcation, were derived from segmentation masks using predefined anatomical rules. The workflow included image inspection, manual and semi-automatic segmentation, three-dimensional reconstruction, and quality control verification, as illustrated in Figure 3.7.



Figure 3.7 Annotation and quality control workflow for multi-class cardiovascular segmentation in 3D Slicer, showing axial, coronal, sagittal, and three-dimensional visualization views

Quality control was performed at multiple levels to ensure annotation reliability. Geometric consistency between images and segmentation masks was verified in terms of dimensions, voxel spacing, orientation, and spatial alignment. Label-map validation ensured correct class encoding and mutual exclusivity of labels, while all cases were visually inspected to identify segmentation artifacts or labeling errors. ROI volumes were additionally validated to ensure correct spatial alignment with full-field volumes and preservation of target anatomical structures without introducing cropping-related information leakage.

3.3 Image Preprocessing Pipeline

The preprocessing pipeline was designed to reduce inter-center and inter-modality variability and enable effective use of heterogeneous medical imaging data for deep learning–based segmentation. It addressed variability arising from multi-center acquisition, imaging modalities, and anatomical diversity associated with congenital heart disease. Because the dataset combined a public benchmark dataset with a clinically curated multi-center cohort, preprocessing was structured to establish consistency across all image–mask pairs while preserving clinically relevant anatomical information.

The preprocessing pipeline included image standardization, spatial resampling, intensity normalization, cropping and padding, and segmentation mask preparation. These steps reduced dataset-specific variability while preserving anatomical fidelity and supporting robust model training. Localization-guided ROI extraction was treated separately as a model-driven component (Section 3.4). An overview of the preprocessing workflow is presented in Figure 3.8.

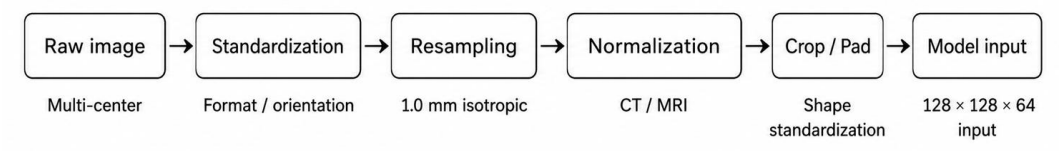


Figure 3.8 Overview of the preprocessing pipeline including standardization, resampling, normalization, and mask preparation

The resulting standardized volumes were suitable for robust deep learning–based segmentation. Key preprocessing parameters are summarized in Table 3.5.

Table 3.5 Key preprocessing parameters used in this study

Parameter	Value
Target spacing	1.0 mm isotropic
Full volume size	160 × 160 × 96
Input size	128 × 128 × 64
CT clipping range	[−200, 600] HU
MRI percentile range	1st–99th
ROI margin	10 voxels

3.3.1 Image Standardization

Image standardization constituted the first stage of the preprocessing pipeline and aimed to establish a consistent data representation across all imaging studies. Since this study combined a public benchmark dataset with a clinically curated multi-center cohort, raw image volumes showed variability in format, orientation, voxel spacing, field of view, and metadata because of differences in imaging modalities, scanner vendors, and acquisition protocols. All cases were organized using a subject-based directory structure in which each image volume was paired with its corresponding segmentation mask through a unique identifier, and incomplete or corrupted cases were excluded.

Subsequently, all image volumes were converted into a unified internal representation by harmonizing dimensional ordering, image geometry, and storage conventions across datasets while preserving image–mask alignment. Geometric consistency was verified using dimensions, voxel spacing, orientation, and spatial origin information for each image–mask pair. Representative raw DICOM volumes were visually inspected in axial, coronal, and sagittal planes prior to preprocessing, as illustrated in Figure 3.9.

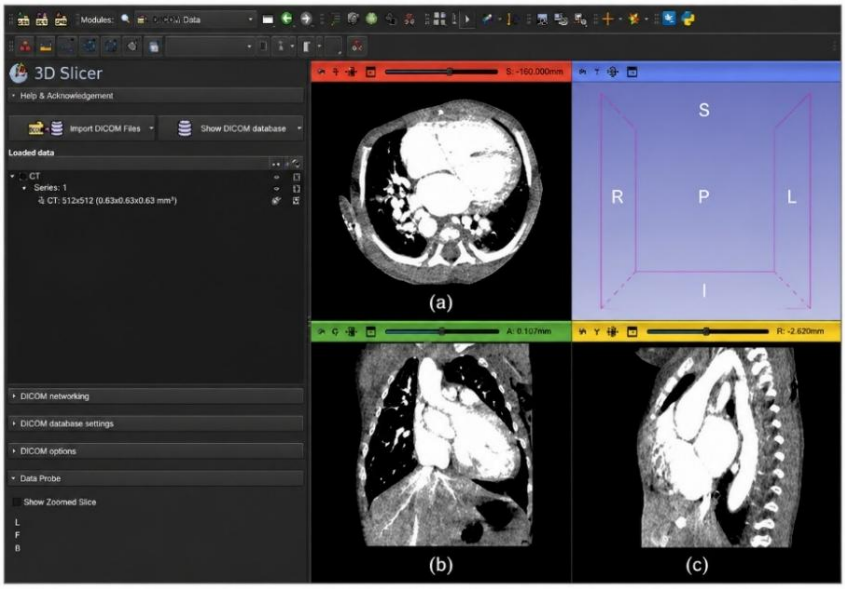


Figure 3.9 Representative multi-planar visualization of a pediatric cardiovascular CT scan during initial image inspection in 3D Slicer: (a) axial, (b) coronal, and (c) sagittal views

An example of geometric inconsistency between image and segmentation mask prior to preprocessing, and the corrected alignment after standardization, is illustrated in Figure 3.10.

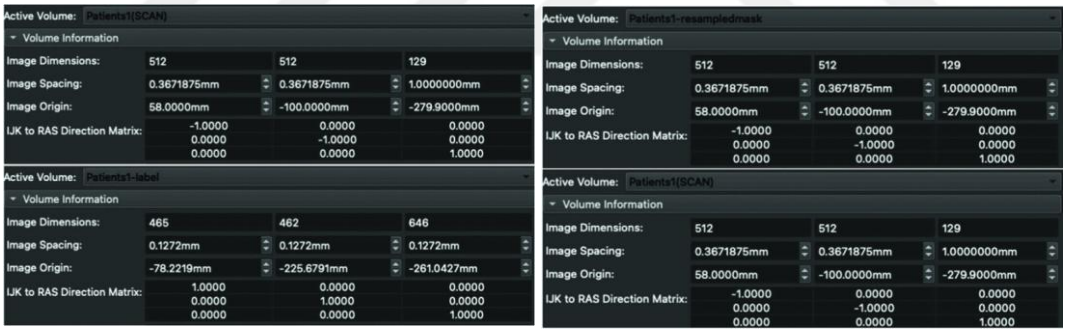


Figure 3.10 Comparison of image and segmentation mask geometry before and after preprocessing, illustrating correction of spatial inconsistencies

The same standardization procedure was applied to both full-field and region-of-interest representations, resulting in validated, geometrically consistent, and standardized image–mask pairs. This stage ensured correct spatial alignment and prepared the data for subsequent preprocessing steps.

3.3.2 Spatial Resampling and Resolution Harmonization

Spatial resampling and resolution harmonization constituted the second stage of preprocessing and ensured geometric consistency across imaging volumes. Due to multi-center variability in voxel spacing, slice thickness, and in-plane resolution,

all images and segmentation masks were resampled to a common isotropic resolution before training. A target voxel spacing of 1.0 × 1.0 × 1.0 mm³ was selected to balance anatomical detail preservation and computational efficiency.

The resampling operation was defined as a spatial transformation from the original voxel grid to the target isotropic grid:

$$I'(x', y', z') = I(T(x', y', z')) \tag{3.1}$$

where $T(\cdot)$ denotes the transformation mapping target grid coordinates (x', y', z') to the original image space based on voxel spacing ratios. For continuous image volumes (CT and MRI), trilinear interpolation was used to preserve intensity continuity, whereas segmentation masks were resampled using nearest-neighbor interpolation to maintain discrete label values.

Because of variability in acquisition parameters, including CT pixel spacing, slice thickness, and MRI resolution, spatial resampling was applied to obtain a consistent anatomical representation and improve model generalization. The same transformation was applied to both image volumes and segmentation masks to preserve alignment throughout preprocessing. This step was performed prior to cropping and intensity normalization, enabling more consistent anatomical scaling and improved training stability in pediatric cardiovascular imaging.

3.3.3 Intensity Normalization

Intensity normalization constituted a critical stage of the preprocessing pipeline and aimed to reduce variability across imaging modalities, acquisition protocols, and scanner configurations. Since both computed tomography (CT) and magnetic resonance imaging (MRI) data were used, modality-specific normalization strategies were applied.

For CT images, which are represented in Hounsfield Units (HU), intensity clipping was first applied to focus on the relevant range for cardiovascular structures:

$$I_{\text{clipped}} = \min\left(\max\left(I, I_{\min}\right), I_{\max}\right) \tag{3.2}$$

where $I_{\min} = -200$ and $I_{\max} = 600$. The clipped intensities were then normalized to the range [0,1]:

$$I_{\text{norm}} = \frac{I_{\text{clipped}} - I_{\min}}{I_{\max} - I_{\min}} \tag{3.3}$$

For MRI data, where intensity values are not standardized, a percentile-based normalization strategy was applied. Intensities were clipped using the 1st and 99th percentiles, followed by min–max normalization:

$$I_{\text{norm}} = \frac{I - I_{p1}}{I_{p99} - I_{p1}} \tag{3.4}$$

where I_{p1} and I_{p99} denote the lower and upper percentile bounds. These normalization strategies were applied consistently across all datasets to reduce inter-scan variability and ensure compatibility during combined training. A summary of the normalization configuration is provided in Table 3.6.

Table 3.6 Intensity normalization strategies

Modality	Operation	Parameters	Purpose
CT	Intensity clipping	[−200, 600] HU	Focus on relevant structures
CT	Min–max normalization	[0, 1]	Standardize scale
MRI	Percentile clipping	1st–99th	Remove outliers
MRI	Min–max normalization	[0, 1]	Normalize intensities
Both	Consistent pipeline	—	Ensure compatibility

To reduce inter-scan intensity variability and improve contrast consistency, intensity normalization was applied to all volumes. Representative examples before and after normalization are shown in Figure 3.11.

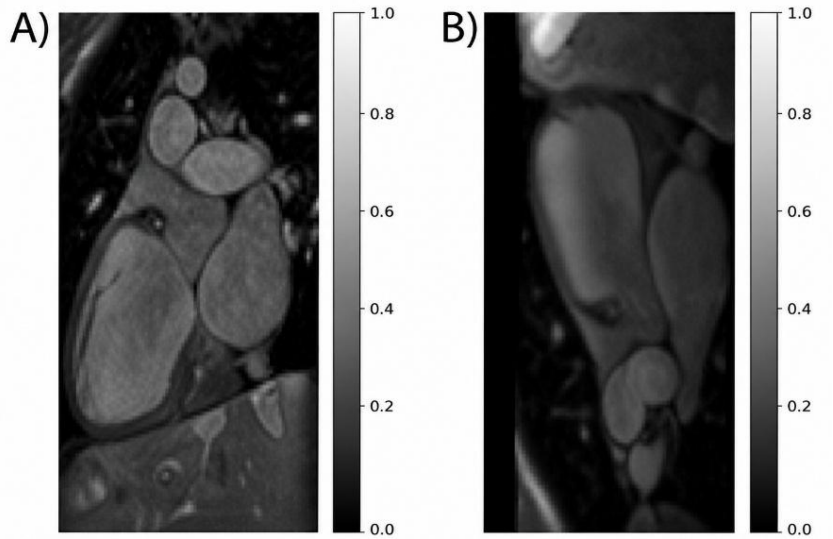


Figure 3.11 Normalization during preprocessing. (A) Original image (B) Normalized image with improved intensity consistency

Intensity normalization reduced inter-scan variability and enabled the model to focus on anatomical structures rather than scanner-specific intensity differences, thereby improving robustness in multi-center settings.

3.3.4 Segmentation Mask Preparation

Segmentation mask preparation constituted the final stage of the preprocessing pipeline and ensured that ground-truth annotations were consistent, valid, and compatible with deep learning–based training. Since this study integrated annotations from both a public benchmark dataset and a clinically curated multi-center dataset, differences in label structure and formatting were resolved through a unified multi-class representation. Each voxel was assigned a discrete class label according to the predefined encoding scheme: 0 = background, 1 = aorta, 2 = cardiac chambers, and 3 = pulmonary arteries. This unified structure ensured consistency across datasets and enabled joint model training. All segmentation masks were validated to ensure compliance with the predefined label set, where invalid labels were removed, overlapping regions were resolved to enforce class exclusivity, and incomplete structures were corrected based on anatomical continuity constraints. Particular attention was given to pulmonary artery regions, where small disconnected components and boundary inconsistencies were more frequent due to anatomical complexity.

Geometric alignment between images and masks was strictly enforced by verifying identical spatial dimensions, voxel spacing, orientation, and spatial origin, ensuring accurate voxel-wise correspondence during training. In addition, refinement operations such as connected-component validation and morphological consistency checks were applied to remove noise and small isolated regions, particularly in clinical data where minor inconsistencies may arise from manual or semi-automatic annotations. All masks were stored in volumetric formats compatible with deep learning frameworks and organized alongside corresponding image volumes using a consistent naming structure. This stage ensured that the model received accurate and consistent supervision signals, as misalignment or incorrect labeling could significantly degrade performance, and therefore served as the final validation step of the preprocessing pipeline.

3.4 Landmark Localization and ROI Extraction

Landmark localization and region-of-interest (ROI) extraction formed the spatial conditioning stage of the proposed framework, enabling anatomically guided segmentation in pediatric cardiovascular imaging. Instead of processing full volumetric data, which is affected by large background regions and anatomical variability, the model operated on localized sub-volumes centered on target structures. To achieve this, anatomical landmarks were first estimated to provide spatial references for key cardiovascular regions, and these positions were then used to define patient-specific ROIs that were subsequently forwarded to the segmentation network. This transformation from global volume to localized representation reduced the influence of irrelevant background and improved focus on clinically relevant anatomy.

The proposed formulation decoupled global localization from fine-grained segmentation, allowing each stage to specialize and improving robustness under inter-patient variability and multi-center heterogeneity. Unlike fixed coordinate-based cropping, the proposed strategy relied on predicted anatomical landmarks, enabling fully automated ROI extraction while preventing potential information leakage during inference. The overall localization-guided pipeline is illustrated in Figure 3.12.

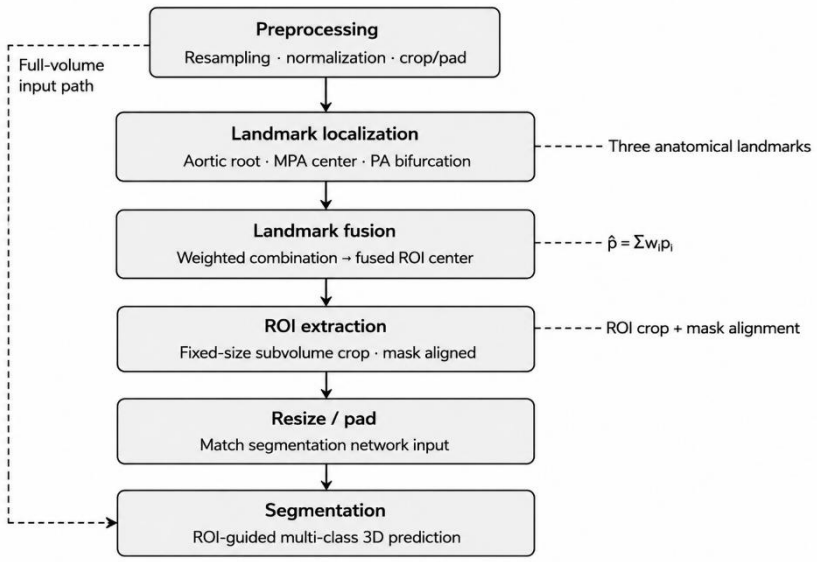


Figure 3.12 Overview of the localization-guided segmentation framework including preprocessing, ROI extraction, and multi-class 3D segmentation

Anatomical landmark definition, localization network architecture, and ROI generation strategy are described in the subsequent subsections.

3.4.1 Rationale for Localization-Guided Segmentation

Localization-guided segmentation was adopted in this study as a core design choice to address anatomical and computational challenges in pediatric cardiovascular image segmentation under heterogeneous multi-center conditions. In conventional full-volume approaches, the model had to simultaneously localize and segment target structures within a large three-dimensional search space, which became particularly challenging when the target anatomy occupied only a small portion of the volume. This limitation was further amplified in pediatric congenital heart disease imaging, where substantial variability existed in anatomical size, vascular topology, and spatial arrangement. Congenital abnormalities, surgical interventions, and age-related differences led to inconsistent anatomical appearances across patients, thereby reducing the generalization capability of full-volume segmentation models. In addition, extensive irrelevant background, including lung parenchyma and surrounding soft tissue, introduced severe foreground–background imbalance, particularly affecting small structures such as pulmonary arteries, where even minor localization errors could lead to incomplete segmentation or missed detection.

To address these limitations, localization-guided segmentation reduced the spatial search space by focusing on anatomically relevant regions through landmark-based ROI extraction. Landmark detection was first performed to estimate the location of key cardiovascular structures, followed by region-of-interest extraction centered around these landmarks, effectively decomposing the task into coarse localization and fine-grained segmentation. This strategy improved segmentation accuracy while reducing computational cost by limiting input volume size and enabling more efficient utilization of model capacity. It was particularly beneficial in multi-center settings, where variations in field of view and acquisition protocols introduced domain differences. By enforcing consistent anatomical focus, the landmark-based ROI approach improved robustness across heterogeneous datasets and was especially advantageous for pulmonary artery segmentation, where structures are smaller, more variable, and more sensitive to noise and low contrast. Taken

together, this approach transformed a complex global problem into a more structured and tractable learning task.

3.4.2 Definition of Anatomical Landmarks

In the proposed localization-guided segmentation framework, anatomical landmarks were defined as spatial reference points representing key cardiovascular structures and guiding the extraction of patient-specific regions of interest. These landmarks provided a compact anatomical representation that enabled the model to focus on clinically relevant regions while reducing sensitivity to global spatial variability. Given the high anatomical variability in pediatric congenital heart disease, careful landmark selection was essential for robustness and generalization. In this study, three anatomically meaningful and consistently identifiable landmarks were defined based on clinical relevance and spatial coverage: the aortic root (AR), representing a stable central reference point; the main pulmonary artery center (MPA), corresponding to the central pulmonary trunk; and the pulmonary artery bifurcation (PAB), indicating the branching into left and right pulmonary arteries. These landmarks ensured that the extracted region of interest covered all target structures while remaining robust under anatomical variability. Landmark annotations were automatically derived from segmentation masks using deterministic anatomical heuristics based on predefined anatomical rules, rather than manually annotated. The selected anatomical landmarks and their clinical relevance are summarized in Table 3.7.

Table 3.7 Definition and clinical relevance of anatomical landmarks

Landmark	Description	Role in ROI localization	Clinical relevance
Aortic root (AR)	Origin of the aorta	Global spatial reference	Stable anatomical anchor
Main pulmonary artery (MPA)	Central pulmonary trunk	ROI centering support	Major vascular structure
Pulmonary artery bifurcation (PAB)	Left/right branching point	Local refinement for ROI	Critical small vessel localization

Each landmark was represented as a three-dimensional coordinate in volumetric space:

$$\mathcal{L} = \{\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3\}, \mathbf{l}_i \in \mathbb{R}^3 \tag{3.5}$$

$$\mathbf{l}_i = (x_i, y_i, z_i) \tag{3.6}$$

To improve localization robustness, a heatmap-based representation was adopted instead of direct coordinate regression. Each landmark was encoded as a three-dimensional Gaussian distribution centered at its spatial location:

$$H_i(x, y, z) = \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}{2\sigma^2}\right) \tag{3.7}$$

where (x_i, y_i, z_i) denotes the landmark location and σ controls the spatial spread. The ROI center was computed as a weighted combination of detected landmarks:

$$\mathbf{c} = \frac{\sum_{i=1}^{3} w_i \mathbf{l}_i}{\sum_{i=1}^{3} w_i} \tag{3.8}$$

where w_i represents the contribution of each landmark, with higher weights assigned to pulmonary artery–related landmarks to improve sensitivity to small vascular structures. The region of interest was then defined as:

$$\Omega_{\text{ROI}} = \{(x, y, z) \mid\mid x - c_x \mid \leq \Delta_x, \mid y - c_y \mid \leq \Delta_y, \mid z - c_z \mid \leq \Delta_z\} \tag{3.9}$$

where (c_x, c_y, c_z) denotes the ROI center and $\Delta_x, \Delta_y, \Delta_z$ define the spatial extent along each axis. The spatial relationship between anatomical landmarks and the resulting region of interest is illustrated in Figure 3.13.

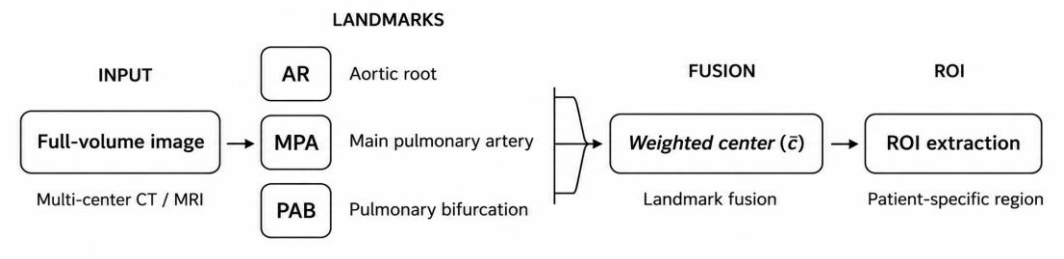


Figure 3.13 Anatomical landmark definition and ROI localization framework

3.4.3 Landmark Localization Network

The anatomical landmarks defined in the previous section were detected using a dedicated deep learning–based localization network. The objective of this network was to predict the spatial positions of key cardiovascular landmarks in a robust and data-driven manner, enabling accurate region-of-interest extraction prior to segmentation. Instead of direct coordinate regression, landmark localization was formulated as a dense prediction problem, where the network output spatial

probability maps corresponding to each anatomical landmark. The localization network followed a three-dimensional encoder–decoder architecture, where the input was a preprocessed volumetric image $I \in \mathbb{R}^{H\times W\times D}$, and the output consisted of landmark-specific heatmaps:

$$\mathbf{H} = \{H_1, H_2, H_3\}, H_i \in \mathbb{R}^{H\times W\times D} \tag{3.10}$$

where each heatmap corresponded to one anatomical landmark (AR, MPA, PAB). The overall architecture of the localization network is illustrated in Figure 3.14. The encoder consisted of four hierarchical convolutional stages, each composed of 3D convolutional layers with kernel size $3 \times 3 \times 3$, followed by batch normalization and ReLU activation. Downsampling was performed using strided convolutions, progressively reducing spatial resolution while increasing feature dimensionality. The decoder mirrored the encoder structure using trilinear upsampling and skip connections, enabling recovery of spatial details. The final layer consisted of a $1 \times 1 \times 1$ convolution followed by a sigmoid activation to generate landmark-specific heatmaps, where the encoder captured global anatomical context and the decoder refined local spatial details for precise landmark detection.

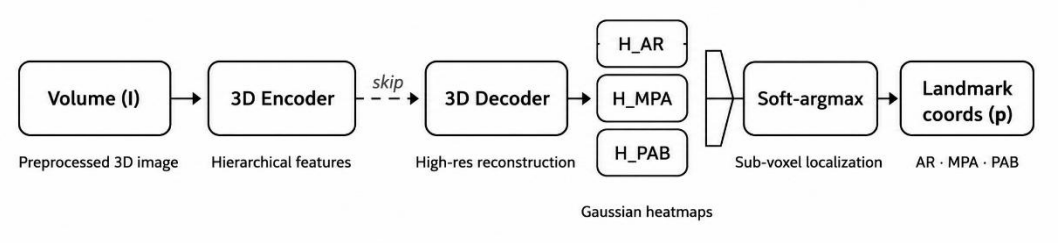


Figure 3.14 Landmark localization network architecture

Ground-truth heatmaps were generated from segmentation masks using deterministic anatomical heuristics. To improve robustness, the network was trained using a heatmap regression objective, where predicted heatmaps $\widehat{H}_i$ were compared with ground-truth Gaussian heatmaps using a voxel-wise loss:

$$\mathcal{L}_{\text{loc}} = \frac{1}{N}\sum_{i=1}^{N} \| \widehat{H}_i - H_i \|_2^2 \tag{3.11}$$

At inference time, landmark coordinates were obtained using a soft-argmax operation:

$$\hat{\mathbf{l}}_i = \sum_{x,y,z} (x,y,z)\ \widetilde{H}_i(x,y,z) \tag{3.12}$$

which provided sub-voxel localization accuracy and avoided instability associated with hard argmax operations. This formulation enabled dense spatial supervision and improved robustness to anatomical variability and noise. The heatmap-based representation is illustrated in Figure 3.15.

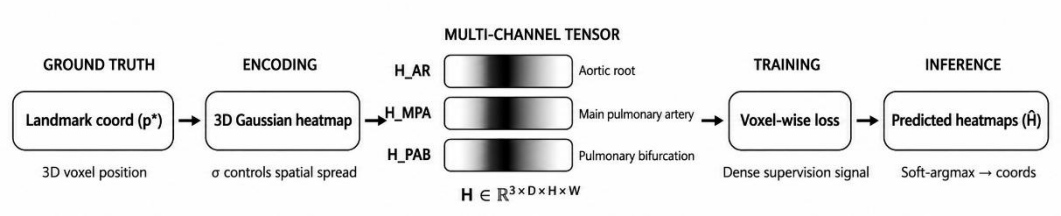


Figure 3.15 Heatmap-based landmark representation

The configuration of the localization network is summarized in Table 3.8.

Table 3.8 Landmark localization network configuration

Component	Description	Implementation detail
Input	3D volumetric image	Provides anatomical context
Encoder	3D convolutional blocks	$3 \times 3 \times 3$kernels, ReLU, batch normalization
Decoder	Upsampling with skip connections	Trilinear upsampling + feature concatenation
Output	Multi-channel heatmaps (3 channels)	Sigmoid activation
Loss function	Mean squared error	Heatmap regression
Post-processing	Soft-argmax	Sub-voxel localization

The encoder–decoder structure enabled the model to capture both global context and local spatial details, which are essential for accurately identifying small anatomical structures such as the pulmonary artery bifurcation. As a result, the proposed localization network provided a robust and anatomically informed mechanism for automated ROI extraction under heterogeneous imaging conditions.

3.4.4 ROI Generation Strategy

Following landmark localization, a region-of-interest (ROI) extraction strategy was applied to isolate anatomically relevant regions from the full volumetric image.

Instead of processing entire image volumes containing extensive background regions, a fixed-size ROI was extracted around the detected anatomical landmarks. The predicted landmark coordinates were fused to obtain a robust ROI center, as defined in Eq. (3.8), improving stability by integrating multiple anatomical references with greater weighting for pulmonary artery–related landmarks. Based on this center, the region of interest was defined according to the spatial formulation introduced in Eq. (3.9), ensuring consistent coverage of the target cardiovascular structures.

To ensure consistent input dimensions, extracted ROIs were standardized to 128 × 128 × 64 voxels using centered cropping and padding. Zero-padding was applied when necessary to preserve spatial consistency. A margin-aware cropping strategy was also incorporated to account for localization uncertainty and anatomical variability. Three ROI settings were defined: (i) full-volume input, (ii) ground-truth–based ROI for training preparation only, and (iii) predicted landmark–based ROI for inference as the proposed approach. Ground-truth ROIs were not used during inference.

Representative examples of landmark-guided ROI extraction are shown in Figure 3.16. Scan and mask visualizations from two cases demonstrate how landmark-guided ROI localization reduces background regions while maintaining anatomical coverage of the target cardiovascular structures. ROI definitions based on predicted anatomical landmarks enabled adaptive localization across anatomically variable cases and multi-center imaging conditions. The resulting two-stage pipeline consisted of landmark-guided ROI localization followed by multi-class segmentation.

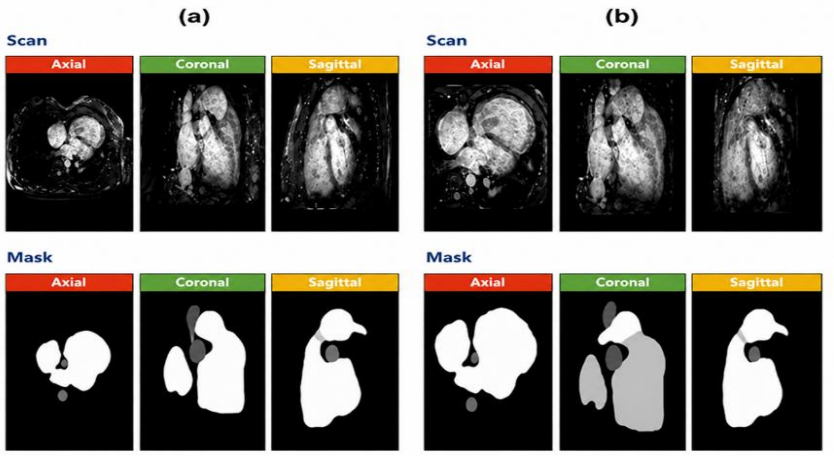


Figure 3.16 Representative ROI-based scan and mask visualizations from two cases shown in axial, coronal, and sagittal views, demonstrating reduced background while preserving target cardiovascular structures

3.5 Segmentation Architectures

This study employed multiple three-dimensional segmentation architectures to enable systematic evaluation under heterogeneous multi-center conditions while distinguishing the proposed framework from comparison models. Representative architectures from different design paradigms were selected, including convolutional neural networks (3D U-Net and V-Net), a self-configuring framework (nnU-Net), and a transformer-based architecture (Swin-UNet). All models were evaluated under identical preprocessing, ROI extraction, and training settings to ensure consistent comparison. The compared architectures and their roles in this study are summarized in Table 3.9.

Table 3.9 Compared segmentation architectures and their roles in this study

Model	Backbone Type	Key Strength	Role in This Study
3D U-Net	CNN (encoder–decoder)	Strong baseline performance	Reference baseline
V-Net	CNN (residual volumetric)	Effective volumetric learning	Core CNN baseline
nnU-Net	Self-configuring CNN	Automatic pipeline optimization	Performance benchmark
Swin-UNet	Transformer-based	Global context modeling	Transformer comparison
Hybrid V-Net	CNN + Transformer	Local–global feature fusion	Proposed method

3.5.1 Proposed Hybrid V-Net with Transformer

The primary methodological contribution of this study was the localization-guided Hybrid V-Net with Transformer architecture, designed to combine local volumetric feature extraction with global context modeling under heterogeneous multi-center conditions. The proposed architecture was based on a three-dimensional encoder–decoder structure derived from V-Net, where convolutional layers extracted hierarchical local features and residual connections ensured stable optimization and

effective gradient propagation. To overcome the limitations of purely convolutional models, a transformer module was integrated at the bottleneck stage, employing self-attention mechanisms to capture long-range spatial dependencies and enhance global feature representation. As a result, the model incorporated both fine-grained anatomical details and broader contextual information. Compared to standard V-Net, the proposed architecture improved global context awareness, while compared to transformer-based models such as Swin-UNet, it preserved efficient local feature extraction. In addition, the use of localization-guided ROI inputs introduced an implicit anatomical prior, improving sensitivity to small and complex vascular structures while reducing background interference. The key components of the proposed Hybrid V-Net with Transformer are summarized in Table 3.10.

Table 3.10 Hybrid V-Net with Transformer architecture characteristics

Component	Description	Purpose
Backbone	V-Net (3D CNN encoder–decoder)	Local feature extraction
Residual connections	Within convolutional blocks	Stable training
Transformer module	Bottleneck self-attention layers	Global context modeling
Input	ROI-based volumetric patches	Anatomical focus
Output	Multi-class segmentation maps	Structure prediction
Integration	CNN + Transformer fusion	Combine local and global features

The Hybrid V-Net with Transformer architecture served as the primary proposed framework, while the remaining architectures were used as comparative baselines. This design enabled systematic evaluation under consistent training conditions.

3.6 Proposed Localization-Guided Segmentation Framework

Within this study, the localization-guided hybrid segmentation framework was proposed as the primary methodological contribution, while baseline architectures, federated training configurations, ensemble learning, and explainability analyses

were treated as comparative or supplementary components. The proposed framework performed multi-class three-dimensional segmentation of pediatric cardiovascular structures by integrating anatomical localization with a hybrid deep learning–based segmentation strategy. Building upon the preprocessing pipeline and landmark-based ROI extraction, it unified spatial localization, ROI-based representation, and a hybrid V-Net–Transformer architecture within a single end-to-end pipeline. The framework consisted of three main components: (i) landmark-based anatomical localization for region-of-interest definition, (ii) a hybrid encoder–decoder architecture enhanced with transformer-based attention for segmentation, and (iii) structure-aware learning to improve performance on clinically critical and underrepresented structures, particularly pulmonary arteries. A two-stage formulation was adopted, where anatomical landmarks were first estimated to define patient-specific ROIs, followed by segmentation performed on localized sub-volumes, reducing the effective search space, suppressing irrelevant background regions, and enabling more focused feature learning while maintaining robustness under inter-patient and multi-center variability.

The hybrid CNN–Transformer architecture combined convolutional feature extraction for local anatomical detail with self-attention mechanisms for global contextual modeling, enabling comprehensive representation of both fine-grained structures and long-range spatial dependencies. The workflow of the framework is illustrated in Figure 3.17, covering preprocessing, landmark-based ROI extraction, and hybrid segmentation for multi-class anatomical prediction.

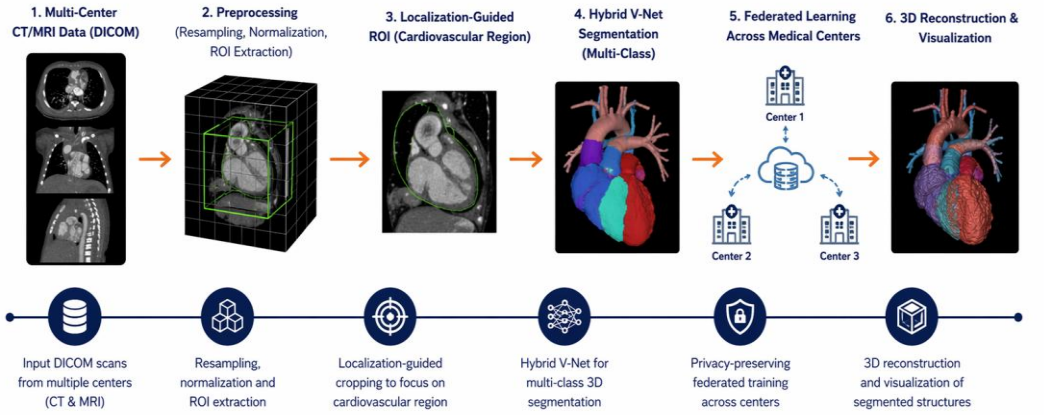


Figure 3.17 Overview of the proposed localization-guided framework incorporating ROI extraction, hybrid V-Net–transformer segmentation, and federated learning across multiple medical centers

Within this framework, the segmentation task was formulated as learning a mapping from volumetric medical images to anatomical labels. Given an input volume:

$$I \in \mathbb{R}^{H \times W \times D} \tag{3.13}$$

the network learned a function:

$$f_\theta : I \to Y \tag{3.14}$$

where the output label space was defined as:

$$Y \in \{0,1,2,3\}^{H \times W \times D} \tag{3.15}$$

and voxel-wise labels were given by:

$$Y(x, y, z) \in \{0,1,2,3\} \tag{3.16}$$

The network produced voxel-wise logits:

$$Z \in \mathbb{R}^{H \times W \times D \times C}, C = 4 \tag{3.17}$$

which were transformed into probabilities using the softmax function:

$$P_c(x, y, z) = \frac{\exp\left(Z_c(x, y, z)\right)}{\sum_{k=1}^{C} \exp\left(Z_k(x, y, z)\right)} \tag{3.18}$$

Unlike full-volume segmentation, the proposed method operated on ROI-defined sub-volumes:

$$I_{\text{ROI}} \subset I \tag{3.19}$$

leading to the mapping:

$$f_\theta : I_{\text{ROI}} \to Y_{\text{ROI}} \tag{3.20}$$

To guide optimization, a structure-aware objective was defined as:

$$\theta^* = \arg\min_\theta \mathcal{L}_{\text{total}}(f_\theta(I_{\text{ROI}}), Y) \tag{3.21}$$

where $\mathcal{L}_{\text{total}}$ integrated multi-class segmentation supervision with structure-specific learning signals.

3.6.1 Network Architecture Design

The proposed segmentation framework was based on a hybrid encoder–decoder architecture that integrated convolutional feature extraction with transformer-based

global context modeling within a localization-guided pipeline. Segmentation was performed on region-of-interest (ROI) volumes derived from anatomical landmarks, enabling anatomically focused representation.

The network adopted a three-dimensional V-Net–based encoder–decoder structure, where the input was defined as a standardized ROI volume:

$$I_{\text{ROI}} \in \mathbb{R}^{H \times W \times D} \tag{3.22}$$

The encoder consisted of four hierarchical convolutional stages, each composed of 3D convolutional layers with kernel size $3 \times 3 \times 3$, followed by batch normalization and ReLU activation. Downsampling was performed using strided convolutions, progressively reducing spatial resolution while increasing feature dimensionality, while residual connections improved gradient stability.

At the bottleneck stage, encoded feature maps were represented as:

$$F \in \mathbb{R}^{H' \times W' \times D' \times C} \tag{3.23}$$

and reshaped into a token sequence:

$$T \in \mathbb{R}^{N \times C}, N = H' \cdot W' \cdot D' \tag{3.24}$$

These tokens were processed using multi-head self-attention:

$$\text{Attention}\ (Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d}}\right)V \tag{3.25}$$

where Q, K, and V represented query, key, and value projections of token embeddings, enabling modeling of long-range spatial dependencies.

The transformer module consisted of Lstacked attention layers with H heads per layer (e.g., $L = 2$, $H = 4$), followed by feed-forward networks for feature refinement. The transformed features were reshaped back into volumetric form and passed to the decoder, where transformer outputs were fused with convolutional features via skip-connected concatenation:

$$F_{\text{fused}} = \text{Concat}(F_{\text{cnn}}, F_{\text{trans}}) \tag{3.26}$$

The decoder mirrored the encoder structure using trilinear upsampling and skip connections, enabling progressive recovery of spatial resolution while preserving fine anatomical details. The network produced voxel-wise multi-class predictions,

consistent with the ROI-to-label mapping formulation introduced in Eq. (3.20). The overall architecture is illustrated in Figure 3.18.

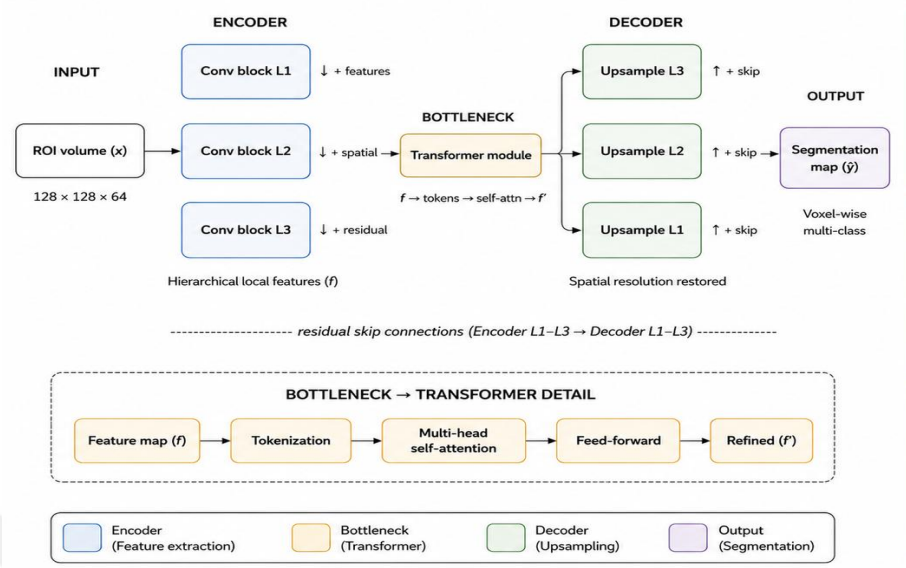


Figure 3.18 Architecture of the proposed localization-guided hybrid V-Net–Transformer segmentation network

From a design perspective, the architecture provided a unified representation of local and global features through the integration of convolutional encoders and transformer-based context modeling. The use of ROI-based inputs ensured anatomically focused processing, contributing to improved robustness in heterogeneous multi-center settings.

3.6.2 Multi-Class Output and Structure-Specific Learning Strategy

The proposed framework performed voxel-wise multi-class classification, where each voxel in the region-of-interest (ROI) volume was assigned to one of the predefined anatomical classes. Given the standardized ROI input defined in Eq. (3.22), the network produced voxel-wise logits:

$$Z \in \mathbb{R}^{H\times W\times D\times C}, C = 4 \tag{3.27}$$

where each channel $Z_c(x, y, z)$ represented the unnormalized score for class c. Class probabilities were obtained using the softmax function, as defined previously, ensuring a valid probability distribution:

$$\sum_{c=1}^{C} P_c\,(x, y, z) = 1 \tag{3.28}$$

The final segmentation output was defined as:

$$\hat{Y}(x,y,z) = \arg\max_{c} \; P_c(x,y,z) \tag{3.29}$$

The segmentation task considered four classes: background (0), aorta (1), cardiac chambers (2), and pulmonary arteries (3), ensuring consistency across datasets. To improve sensitivity to small and underrepresented structures, a structure-specific supervision strategy was introduced for pulmonary arteries. A binary auxiliary map was defined as:

$$Y^{PA}(x,y,z) = \begin{cases} 1, & \text{if } Y(x,y,z) = 3 \\ 0, & \text{otherwise} \end{cases} \tag{3.30}$$

The corresponding probability was directly derived from the multi-class output:

$$P^{PA}(x,y,z) = P_{c=3}(x,y,z) \tag{3.31}$$

This formulation did not introduce an additional prediction branch and preserved consistency with the main segmentation output. From an optimization perspective, this strategy explicitly addressed class imbalance by enhancing supervision for underrepresented structures. In standard multi-class training, dominant classes biased gradient updates, reducing sensitivity to small vascular regions such as pulmonary arteries. The proposed formulation mitigated this effect by reinforcing learning on the target structure while maintaining global label consistency, as illustrated in Figure 3.19.

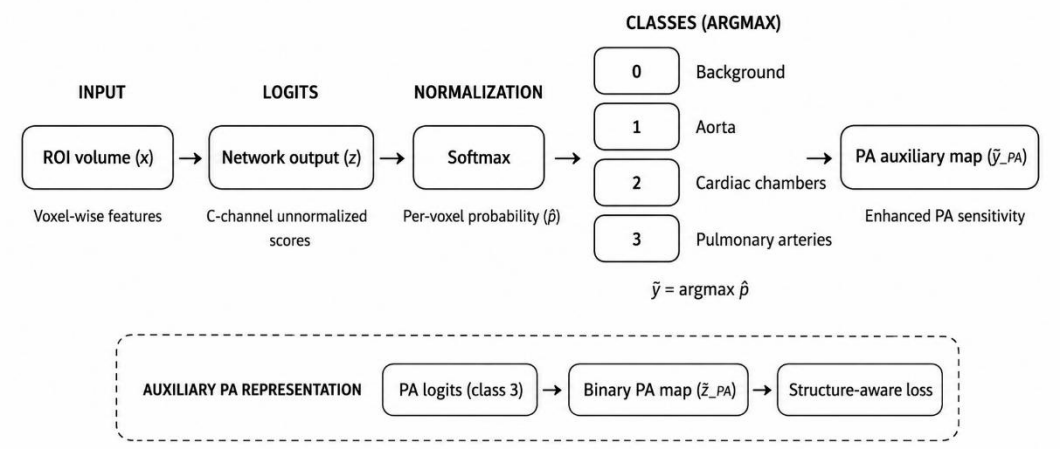


Figure 3.19 Voxel-wise multi-class output representation with structure-specific supervision for pulmonary arteries

As a result, the framework achieved improved representation of small anatomical structures while maintaining consistent multi-class predictions, leading to more robust performance under heterogeneous multi-center imaging conditions.

3.6.3 Loss Functions and Optimization

To train the proposed localization-guided segmentation framework, a composite loss function was employed that integrated multi-class segmentation objectives with structure-specific supervision. The primary objective combined Dice loss and cross-entropy loss, providing complementary supervision for region overlap and voxel-wise classification.

The multi-class Dice loss was defined as:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{c=1}^{C}\sum_{x,y,z} P_c\,(x,y,z)\,Y_c(x,y,z)}{\sum_{c=1}^{C}\sum_{x,y,z} P_c\,(x,y,z) + \sum_{c=1}^{C}\sum_{x,y,z} Y_c\,(x,y,z)} \tag{3.32}$$

and the cross-entropy loss as:

$$\mathcal{L}_{\text{CE}} = -\sum_{c=1}^{C}\sum_{x,y,z} Y_c\,(x,y,z)\log\,P_c(x,y,z) \tag{3.33}$$

To improve sensitivity to small vascular structures, an auxiliary pulmonary artery loss was incorporated using the structure-specific probability defined in Eq. (3.31). This term was computed directly from the shared softmax output, without introducing an additional prediction branch:

$$\mathcal{L}_{\text{PA}} = 1 - \frac{2\sum_{x,y,z} P^{PA}\,(x,y,z)\,Y^{PA}(x,y,z)}{\sum_{x,y,z} P^{PA}\,(x,y,z) + \sum_{x,y,z} Y^{PA}\,(x,y,z)} \tag{3.34}$$

The total loss was formulated as:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{Dice}} + \beta\mathcal{L}_{\text{CE}} + \lambda_{PA}\mathcal{L}_{\text{PA}} \tag{3.35}$$

where α, β, and λ_{PA} controlled the contribution of each term. The weighting scheme was selected to balance global segmentation accuracy with sensitivity to pulmonary artery structures, with parameters empirically determined based on validation stability and class imbalance sensitivity.

Model optimization was performed using gradient-based updates:

$$\theta \leftarrow \theta - \eta\nabla_{\theta}\mathcal{L}_{\text{total}} \tag{3.36}$$

with AdamW as the optimizer. To ensure stable and efficient training, gradient clipping was applied to prevent instability, learning rate scheduling was used to

improve convergence behavior, and mixed-precision training was adopted to reduce computational cost. Table 3.11 summarizes the key optimization and training settings.

Table 3.11 Optimization and training configuration of the proposed framework

Parameter	Setting	Purpose
Optimizer	AdamW	Stable optimization
Learning rate	10^{-4}(initial)	Update control
Loss weights	$\alpha = 0.5, \beta = 0.5, \lambda_{PA} = 0.3$	Balanced supervision
Gradient clipping	Enabled	Training stability
Scheduler	Decay Schedule	Convergence improvement
Precision	Mixed precision (AMP)	Efficiency

From an optimization perspective, the proposed formulation enabled balanced learning between global segmentation accuracy and structure-specific sensitivity, leading to robust performance under heterogeneous multi-center imaging conditions.

3.6.4 Inference and Post-Processing Strategy

During inference, the proposed framework followed the same localization-guided pipeline used during training to ensure consistency and computational efficiency. Anatomical landmarks were first estimated using the localization network, and a patient-specific ROI was extracted accordingly. The ROI volume was then processed by the Hybrid V-Net with Transformer segmentation network to generate final predictions.

Consistent with Section 3.6.2, voxel-wise logits and class probabilities were computed as defined in Eqs. (3.27) and (3.18), followed by voxel-wise class selection based on maximum probability. The overall inference workflow is illustrated in Figure 3.20.

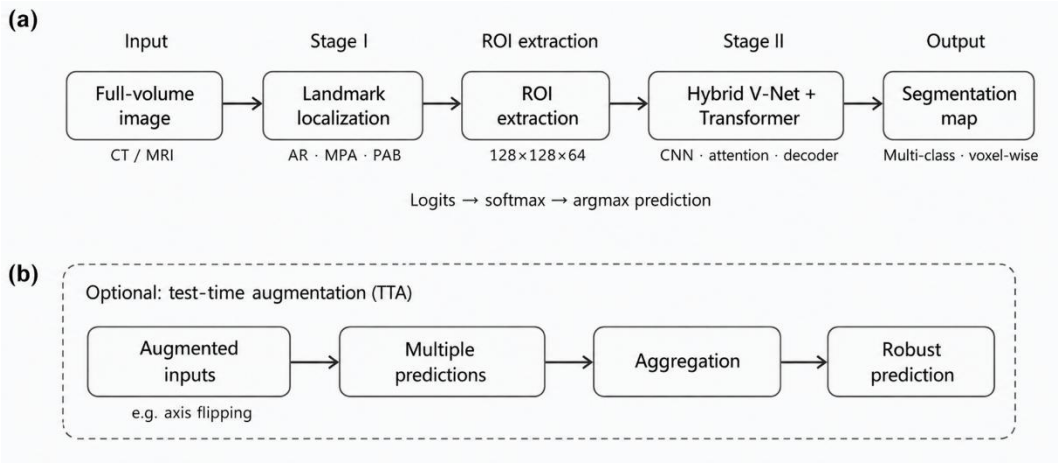


Figure 3.20 Inference workflow of the proposed framework: (a) localization-guided ROI-based segmentation pipeline and (b) optional test-time augmentation (TTA) strategy for robust prediction aggregation

To improve prediction robustness, test-time augmentation (TTA) was optionally applied using multiple spatially augmented inputs and aggregating their predictions. Inference was performed in deterministic evaluation mode with stochastic components disabled to ensure reproducibility. The ROI-based design reduced computational cost and improved anatomical focus across heterogeneous multi-center data. Following inference, post-processing was applied to refine segmentation outputs and improve anatomical consistency. Connected-component analysis removed small isolated regions, while morphological operations such as hole filling and boundary smoothing enhanced structural continuity, particularly for thin vascular structures such as pulmonary arteries. The final refined segmentation was defined as:

$$\hat{Y}^* = \mathcal{P}(\hat{Y}) \tag{3.37}$$

where $\mathcal{P}(\cdot)$ denoted the post-processing operator.

Post-processing improved anatomical plausibility and spatial continuity without altering the learned model predictions.

3.7 Federated Learning Framework

The federated learning (FL) framework implemented for distributed training of the proposed localization-guided segmentation pipeline under multi-center conditions is presented in this section. The FL framework was integrated with ROI-based preprocessing, Hybrid V-Net-based segmentation, and structure-specific learning,

enabling consistent evaluation under both centralized and federated training settings. In the federated configuration, each client performed local optimization on its own dataset and shared model parameters with a central aggregation server without transferring raw imaging data.

Let K denote the number of clients and $\mathcal{D}_k$ the dataset assigned to client k. The global optimization objective was defined as:

$$\min_{\theta} \sum_{k=1}^{K} \frac{|\mathcal{D}_k|}{|\mathcal{D}|} \mathcal{L}_k(\theta) \tag{3.38}$$

where $|\mathcal{D}|$ represents the total number of samples across all clients.

A silo-based partitioning strategy was adopted to simulate a multi-center environment, where each client corresponded to an independent clinical data subset. Although the dataset was collected from four institutions, the federated configuration used three clients to ensure balanced distribution and stable optimization under limited pediatric data conditions. FedAvg was used as the primary aggregation strategy, while FedProx, FedAdam, and FedYogi were additionally implemented for supplementary analysis of convergence behavior and optimization stability. For comparison, a centralized training configuration using aggregated training data was also evaluated. The communication protocol, optimization strategy, and client partitioning procedure are described in the subsequent subsections.

3.7.1 Client–Server Architecture and Training Strategy

The proposed federated learning framework followed a client–server architecture in which each client independently trained on its local dataset $\mathcal{D}_k$. During training, only model parameters were exchanged with the central server, while all data remained locally stored and processed.

The training process began with the initialization of the global model parameters $\theta^{(0)}$, which were distributed to all clients. During each communication round t, clients received the current global model $\theta^{(t)}$, performed local optimization, and returned updated model parameters to the server. The global model update was defined as:

$$\theta^{(t+1)} = \mathcal{A}\left(\theta_k^{(t+1)}\right)_{k=1}^{K} \tag{3.39}$$

where $\mathcal{A}(\cdot)$ denotes the aggregation function implemented using the FedAvg strategy.

The local optimization objective for client k was defined as:

$$\mathcal{L}_k(\theta) = \frac{1}{N_k}\sum_{i=1}^{N_k} \ell\left(f_\theta(x_i), y_i\right) \tag{3.40}$$

The global model was updated through weighted aggregation:

$$\theta^{(t+1)} = \sum_{k=1}^{K} \frac{N_k}{N}\theta_k^{(t+1)}, N = \sum_{k=1}^{K} N_k \tag{3.41}$$

Preprocessing, landmark-based ROI extraction, and segmentation were performed locally at each client, while all clients participated in every communication round. The overall federated architecture is illustrated in Figure 3.21.

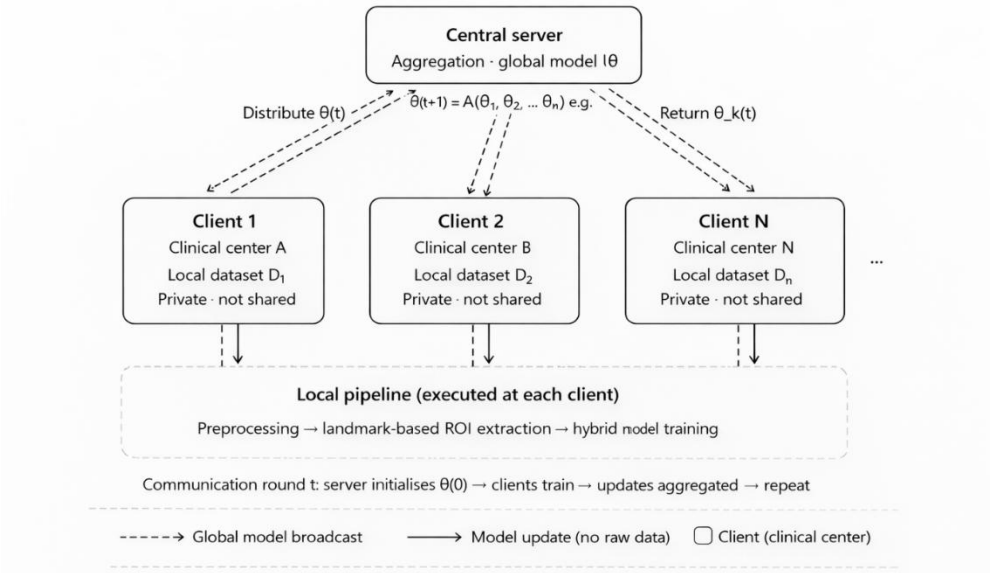


Figure 3.21 Federated client–server architecture for the proposed localization-guided segmentation framework

Distributed optimization was performed through iterative client training, central aggregation, and multi-round communication. At each round t, the central server distributed the global model parameters $\theta^{(t)}$ to all participating clients, which then trained locally on their private datasets $\mathcal{D}_k$. Input volumes were processed using the standardized preprocessing pipeline, including image standardization, spatial

resampling, intensity normalization, and landmark-guided ROI extraction, and the resulting ROI volumes I_{ROI} were used for model training and loss computation.

The local optimization objective at each client was defined as:

$$\mathcal{L}_k(\theta) = \mathbb{E}_{(I,Y)\sim\mathcal{D}_k}[\mathcal{L}_{\text{total}}(f_\theta(I_{ROI}), Y)] \tag{3.42}$$

Model parameters were updated using gradient-based optimization over E_{local} local epochs:

$$\theta_k^{(t+1)} = \theta^{(t)} - \eta\nabla_\theta\mathcal{L}_k(\theta) \tag{3.43}$$

Following local training, the central server aggregated client parameters using the Federated Averaging (FedAvg) strategy:

$$\theta^{(t+1)} = \sum_{k=1}^{K} \frac{|\mathcal{D}_k|}{|\mathcal{D}|}\theta_k^{(t+1)}, |\mathcal{D}| = \sum_{k=1}^{K} |\mathcal{D}_k| \tag{3.44}$$

FedProx, FedAdam, and FedYogi were additionally implemented for supplementary comparative analysis, while FedAvg was used as the primary aggregation strategy for the main experiments. Training was performed over T communication rounds using synchronous aggregation, where all clients participated in each round. The federated training configuration is summarized in Table 3.12.

Table 3.12 Federated learning configuration used in the proposed framework

Parameter	Value / Setting	Description
Number of clients	3	Simulated federated nodes
Local epochs (E)	2	Training epochs per round
Communication rounds (T)	500	Total federated rounds
Aggregation methods	FedAvg (primary), FedProx, FedAdam, FedYogi	Optimization strategies
Learning rate	3×10^{-5}	Local optimizer learning rate
Batch size	2	Training batch size
Optimizer	AdamW	Local optimization algorithm
Participation strategy	Full participation	All clients per round

3.7.2 Silo-Based Data Partitioning and Privacy Considerations

In the proposed federated learning framework, a strict silo-based data partitioning strategy was used to simulate distributed multi-center training conditions. The complete dataset was defined as:

$$\mathbf{D} = \bigcup_{k=1}^{K} \mathbf{D}_k \tag{3.45}$$

where $\mathcal{D}_k$ denotes the dataset assigned to client k.

Data assignment was performed using a deterministic mapping between patient identifiers and client indices to ensure reproducibility and strict separation between clients. No data sharing or migration was allowed during training. Client grouping was performed according to dataset size, where smaller centers were combined to form a balanced client, while larger centers were assigned individually. As a result, three federated clients were constructed from four clinical centers.

The resulting partitioning produced non-identically distributed (non-IID) data across clients, including variations in imaging modality, acquisition protocol, scanner characteristics, and anatomical structure distribution.

In the proposed federated learning framework, raw medical imaging data remained locally stored at each client throughout training, and only model parameters were transmitted between clients and the central server during communication rounds. Formally, for client k, the local training process was defined as:

$$\theta_k^{(t+1)} = \mathrm{Train}(\theta^{(t)}, \mathcal{D}_k) \tag{3.46}$$

Additional privacy-preserving mechanisms, including secure aggregation and differential privacy, were outside the scope of the current implementation.

3.8 Implementation Details

This section summarized the implementation details of the proposed localization-guided federated segmentation framework, with a focus on reproducibility and fair comparison across experimental settings. The implementation supported both centralized and federated training using a unified pipeline including preprocessing, landmark-guided ROI extraction, and Hybrid V-Net with Transformer-based

segmentation. Ensemble inference and explainability analysis were treated as supplementary modules. Consistent preprocessing, input representations, and evaluation protocols were applied across all experiments to ensure fair comparison.

The training process was configured using a consistent set of hyperparameters across experiments. The Hybrid V-Net with Transformer architecture was optimized using the AdamW optimizer with a relatively small learning rate, particularly under federated training. Mini-batch optimization was performed with a batch size of 2. In the centralized setting, models were trained for up to 1000 epochs, whereas federated training was conducted over 500 communication rounds with 2 local epochs per round.

Gradient clipping and mixed-precision training were used to improve numerical stability and computational efficiency. Structure-specific learning enhanced sensitivity to small anatomical structures, particularly the pulmonary arteries, while deep supervision and validation-based checkpointing supported training. In federated learning, both client-level and global checkpoints were maintained. The main training configuration is summarized in Table 3.13.

Table 3.13 Training configuration and hyperparameters

Parameter	Value / Setting	Description
Optimizer	AdamW	Stable gradient-based optimization
Learning rate	$\sim 1\times10^{-4}$ (centralized), $\sim 3\times10^{-5}$ (federated)	Controls update magnitude
Batch size	2	Suitable for 3D volumetric data
Centralized epochs	1000	Total training epochs
Local epochs (E)	2	Updates per communication round
Communication rounds (T)	500	Total federated iterations
Gradient clipping	Enabled	Prevents instability
Precision	Mixed precision (AMP)	Improves efficiency
Loss function	Dice + Cross-Entropy + Auxiliary	Multi-objective optimization
Checkpointing	Validation-based	Best model selection

The framework was implemented using a Python-based deep learning environment for three-dimensional medical image analysis. PyTorch was used as the core framework because of its flexibility and GPU support, while MONAI provided medical image processing utilities and TorchIO was used for preprocessing and data augmentation. Volumetric data handling was performed using nibabel and NRRD-compatible libraries, and SciPy was used for numerical operations such as resampling. All experiments were conducted on GPU-enabled systems compatible with NVIDIA architectures, with mixed-precision training employed to improve efficiency and reduce memory usage. The framework remained compatible with local workstations, high-performance computing environments, and cloud platforms such as Google Colab, while software dependencies were managed using a Conda-based environment to ensure reproducibility. The hardware and software setup is summarized in Table 3.14.

Table 3.14 Hardware and software environment

Component	Specification	Purpose
Framework	PyTorch	Core deep learning implementation
Medical AI library	MONAI	Segmentation utilities
Data processing	TorchIO	3D preprocessing and augmentation
Image I/O	nibabel, NRRD	Volumetric data handling
Numerical operations	SciPy	Resampling and computation
Hardware	NVIDIA GPU	Accelerated training
Precision	Mixed precision (AMP)	Efficient computation
Environment	Conda-based setup	Reproducibility

Model selection was performed using a validation-driven strategy to ensure robust generalization. In the centralized setting, models were selected based on validation Dice score, while additional metrics such as IoU and validation loss were monitored. In the federated setting, model selection was performed at the global level using aggregated validation metrics. Let M_k denote the validation metric at

client k, and $|\mathcal{V}_k|$ the size of its validation set. The aggregated validation score was defined as:

$$M_{\text{global}} = \sum_{k=1}^{K} \frac{|\mathcal{V}_k|}{|\mathcal{V}|} M_k, |\mathcal{V}| = \sum_{k=1}^{K} |\mathcal{V}_k| \tag{3.47}$$

The global model achieving the highest aggregated score was selected as the final model, while early stopping was applied to reduce overfitting. Identical selection criteria were applied across all training settings. Reproducibility was ensured by fixing random seeds, using deterministic dataset splits, and applying identical preprocessing pipelines across experiments. In federated training, silo-based partitioning followed a deterministic client assignment strategy, while logging mechanisms recorded training dynamics and checkpointing ensured recovery and comparison of model states. The implementation followed a modular design, where each component—including preprocessing, localization, segmentation, federated training, and evaluation—was defined as an independent module, enabling flexible experimentation while maintaining separation between core framework components and supplementary analyses.

3.9 Evaluation Metrics and Analysis Strategy

To evaluate the performance of the proposed localization-guided segmentation framework, a set of complementary metrics was employed to assess both region overlap and boundary accuracy, including overlap-based metrics such as the Dice Similarity Coefficient (DSC) and Intersection-over-Union (IoU), as well as boundary-based metrics such as the 95th percentile Hausdorff Distance (HD95) and Average Symmetric Surface Distance (ASSD), whose mathematical definitions were provided in Chapter 2. Rather than relying solely on aggregate performance, the evaluation strategy was designed to analyze model behavior under heterogeneous multi-center conditions at structure-, center-, and case-level granularity, with particular emphasis placed on pulmonary artery segmentation because of its small size, complex geometry, and clinical importance. The evaluation was structured around the main experimental comparisons of this thesis, including (i) direct full-volume versus localization-guided segmentation, (ii) baseline architectures versus the proposed hybrid model, and (iii) centralized versus

federated training, ensuring that the reported results reflected the methodological contributions of the proposed framework. Statistical analysis was performed where appropriate; however, because of the limited size of the final test set, findings were interpreted cautiously and supported by case-wise and structure-specific analyses.

3.9.1 Overlap- and Boundary-Based Metrics

Overlap- and boundary-based metrics were used to comprehensively evaluate segmentation performance by assessing both region-level agreement and contour-level accuracy. DSC and IoU were used to quantify the overlap between predicted and ground-truth voxel sets, while HD95 and ASSD were used to evaluate boundary accuracy and spatial discrepancies along anatomical structures. DSC was used as the primary evaluation metric because of its robustness to class imbalance and sensitivity to spatial agreement, whereas IoU provided a stricter overlap measure.

Boundary-based metrics complemented overlap-based evaluation by capturing geometric discrepancies along anatomical boundaries. HD95 quantified near worst-case boundary deviations while reducing sensitivity to outliers, whereas ASSD measured the average symmetric surface distance between predicted and reference segmentations. These metrics were particularly important for evaluating small and thin anatomical structures such as pulmonary arteries, where boundary precision was critical. All metrics were computed at per-class, per-case, and cross-center levels to enable comprehensive assessment of segmentation performance and generalization.

3.9.2 Multi-Level Evaluation Strategy

In addition to global performance metrics, a multi-level evaluation strategy was employed to assess segmentation performance under heterogeneous multi-center conditions, enabling detailed analysis of anatomical accuracy, patient-level variability, and cross-center generalization. At the structure level, segmentation performance was evaluated separately for each anatomical class (aorta, cardiac chambers, and pulmonary arteries). The structure-wise metric was defined as:

$$M_c = \mathrm{Metric}(P_c, Y_c) \tag{3.48}$$

where P_c and Y_c denote the predicted and ground-truth voxel sets for class c.

At the case level, performance was evaluated for each individual test subject to identify variability and potential failure cases. At the center level, performance was evaluated separately for each data source to assess generalization under heterogeneous data distributions. Let $\mathcal{D}_k^{\text{test}}$ denote the test dataset associated with client or center k. The center-wise metric was defined as:

$$M_k = \text{Metric}(P_k, Y_k) \tag{3.49}$$

where P_k and Y_k represent predictions and ground-truth annotations for the corresponding center.

Together, these analyses enabled comprehensive evaluation of segmentation performance across anatomical structures, individual cases, and multi-center distributions.

3.9.3 Statistical Analysis

Statistical analysis was conducted to support comparisons between models and training strategies. Due to the limited size of the test set and potential deviations from normality, paired non-parametric analysis was employed. Specifically, the Wilcoxon signed-rank test was used to compare paired performance values. Let $M_i^{(A)}$and $M_i^{(B)}$denote the performance metric for models A and B on test case i. The paired difference was defined as:

$$d_i = M_i^{(A)} - M_i^{(B)} \tag{3.50}$$

A significance level of $\alpha = 0.05$ was adopted, and Holm–Bonferroni correction was applied for multiple comparisons. Given the limited number of test cases, statistical findings were interpreted cautiously and considered complementary to case-wise, structure-wise, and qualitative analyses.

3.10 Supplementary Framework Components

3.10.1 Ensemble-Based Inference Strategy

In addition to single-model inference, a supplementary ensemble strategy was employed to improve prediction robustness. Rather than joint training, this approach combined the outputs of independently trained segmentation models during inference, consistent with the modular pipeline design. All ensemble

members operated under identical preprocessing, ROI-based inputs, and a shared four-class segmentation scheme to ensure compatibility. Predictions were aggregated at the probability level through averaging, and optional test-time augmentation (TTA) was incorporated by combining predictions obtained from multiple augmented inputs. This probability-level fusion reduced model-specific variance and improved stability without modifying the training procedure. The ensemble approach was used solely for supplementary robustness analysis and was not considered part of the primary method.

3.10.2 Explainability and Model Interpretation

To support interpretability, a supplementary post hoc explainability module was applied after training and inference. This component did not influence model optimization and was used only for analysis of model behavior. Two complementary methods were employed: Grad-CAM and Integrated Gradients (IG), providing coarse and fine-grained attribution maps, respectively.

Grad-CAM generated spatial attention maps based on feature activations:

$$\alpha_k^c = \frac{1}{Z} \sum_{x,y,z} \frac{\partial Z_c}{\partial A_k(x,y,z)} \tag{3.51}$$

$$L^c_{\text{Grad-CAM}}(x,y,z) = \text{ReLU}\left(\sum_k \alpha_k^c \, A_k(x,y,z) \right) \tag{3.52}$$

Integrated Gradients provided fine-grained attribution along a continuous interpolation path:

$$IG_i = (I_i - I_i') \int_0^1 \frac{\partial F(I' + \alpha(I - I'))}{\partial I_i} \, d\alpha \tag{3.53}$$

These methods provided complementary insights into model behavior, supporting qualitative validation of predictions, particularly for anatomically complex regions such as pulmonary arteries. The explainability component was used solely for interpretability analysis and was not part of the core training or evaluation pipeline.

In this chapter, a comprehensive methodology for multi-class segmentation of pediatric cardiovascular structures was presented, including a standardized preprocessing pipeline and a localization-guided region-of-interest (ROI) strategy

to address data heterogeneity and improve performance on anatomically relevant regions, particularly small structures such as pulmonary arteries. Multiple segmentation architectures were evaluated, including 3D U-Net, V-Net, nnU-Net, Swin-UNet, and the proposed Hybrid V-Net with Transformer, which integrates local feature extraction with global context modeling. A federated learning framework was also developed to enable privacy-preserving multi-center training using a client–server architecture and silo-based data partitioning, while supplementary components such as ensemble learning and explainability analysis were incorporated for robustness and interpretability. Finally, a comprehensive evaluation framework was established, including overlap- and boundary-based metrics, multi-level analysis, and statistical validation. The next chapter presents the experimental results, focusing on comparisons between full-volume and localization-guided segmentation, baseline and proposed models, and centralized versus federated training, with emphasis on pulmonary artery segmentation.

4
EXPERIMENTAL RESULTS AND DISCUSSION

This chapter presents the experimental evaluation of the proposed localization-guided federated framework for multi-class segmentation of pediatric cardiovascular structures under clinically relevant conditions. Experiments were conducted on the HVSMR-2.0 benchmark dataset, the clinically curated multi-center cohort, and the combined dataset under centralized and federated settings, with and without ROI-based localization. Comparative analyses included 3D U-Net, V-Net, nnU-Net, Swin-UNet, and the proposed Hybrid V-Net using overlap-based and boundary-aware evaluation metrics. Overall, ROI-guided processing improved performance, particularly for small and anatomically complex vascular structures such as pulmonary arteries, while federated learning maintained competitive results with only a modest reduction compared with centralized training. The following sections present the quantitative findings and main experimental observations.

4.1 Evaluation Setup and Protocol

The setup, implementation environment, and evaluation protocol used to assess the methodology presented in Chapter 3 are outlined in this section. Detailed descriptions of preprocessing, localization, segmentation architectures, and federated learning are omitted to avoid repetition. Instead, the section summarizes the evaluation configurations, including dataset splits, experimental scenarios, and standardized training settings used for consistent model comparison.

4.1.1 Dataset Configurations

Evaluations were performed using three dataset settings representing benchmark, clinical, and heterogeneous conditions. The HVSMR-2.0 dataset (n = 60) was used for benchmark analysis, while a clinically curated multi-center pediatric cardiovascular cohort (n = 50) was used for clinical evaluation. In addition, a combined cohort integrating both datasets (n = 110) was constructed to examine performance under heterogeneous conditions and enable cross-dataset architectural comparison.

Subject-level partitioning was applied to prevent information leakage, and identical splits were maintained across all experiments. A summary of dataset splits is presented in Table 4.1. Although the clinical cohort originated from four medical centers, federated learning experiments were conducted using three simulated clients to provide balanced distributed evaluation.

Table 4.1 Subject-level data splits across evaluation settings

Configuration	Training	Validation	Test
HVSMR-2.0 Benchmark	50	5	5
Clinical Dataset	42	5	3
Combined Cohort	92	10	8

4.1.2 Implementation Environment and Training Settings

All analyses were implemented in Python using PyTorch, with MONAI and TorchIO supporting preprocessing, augmentation, and optimization procedures. Identical data splits and evaluation settings were maintained across all architectures to ensure fair comparison.

Centralized models were optimized using AdamW with a learning rate of 1×10^{-4}, whereas federated analyses used 3×10^{-5}. Centralized optimization was performed for up to 1000 epochs, while federated analyses were conducted over 500 communication rounds with 2 local epochs per round across three simulated clients. A batch size of 2, mixed-precision computation, and gradient clipping were applied throughout all evaluations. Federated analyses included FedAvg, FedProx, FedAdam, and FedYogi aggregation strategies, with FedAvg serving as the primary

baseline. Training dynamics under centralized and federated configurations are illustrated in Figures 4.1 and 4.2.

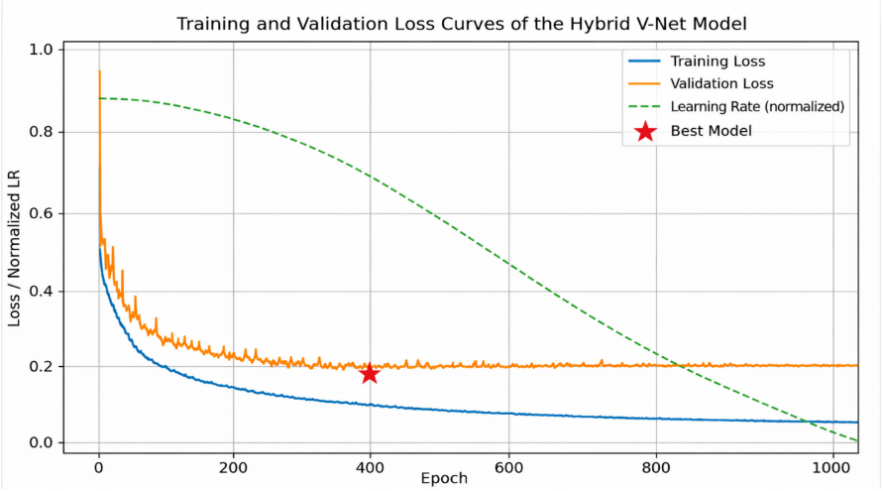


Figure 4.1 Training and validation loss curves of the proposed Hybrid V-Net model under the centralized training configuration

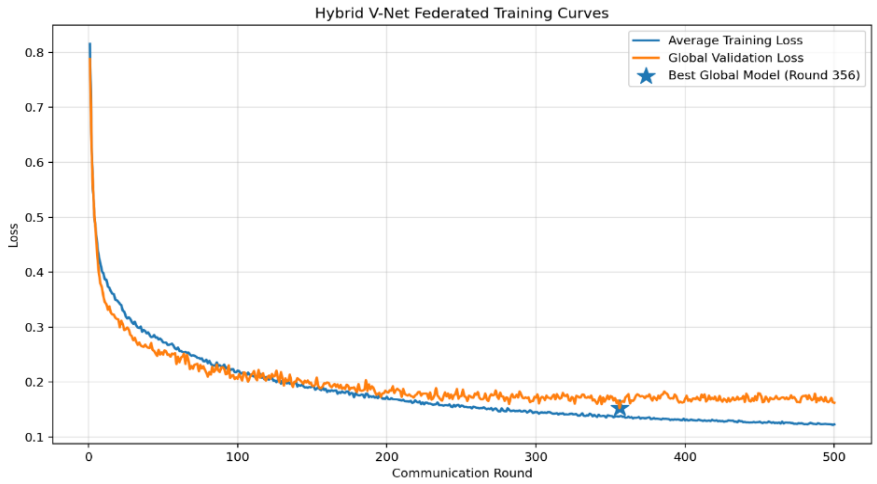


Figure 4.2 Federated training curves of the proposed Hybrid V-Net model across communication rounds

4.1.3 Evaluation Scenarios

To assess the proposed framework under different data and training settings, five experimental scenarios were considered: benchmark assessment, clinical generalization, architectural comparison, localization-guided processing, and federated learning under heterogeneous multi-center conditions. Dataset configurations are summarized in Table 4.1, while the overall workflow is illustrated in Figure 4.3. Performance was evaluated using overlap-based and boundary-aware metrics together with structure-specific and qualitative analyses for the aorta, cardiac chambers, and pulmonary arteries. Additional analyses,

including explainability evaluation and representative error patterns, are provided in Appendix A.

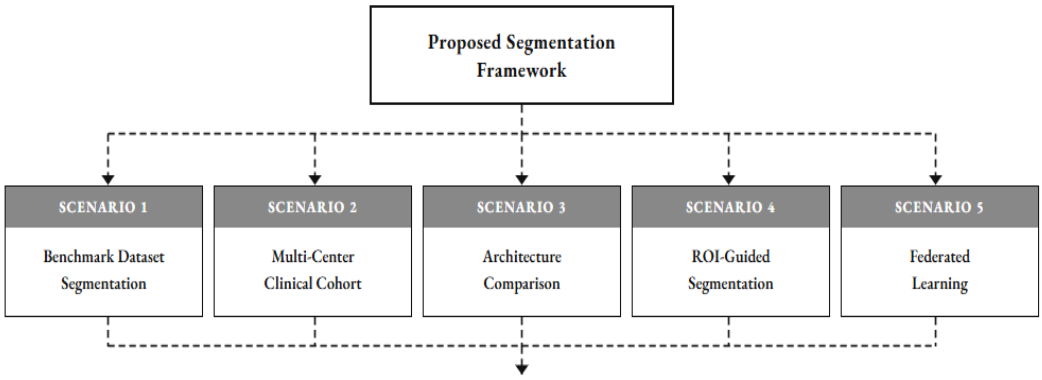


Figure 4.3 Overview of the experimental evaluation scenarios used in this study

4.2 Results on the HVSMR-2.0 Benchmark Dataset

Segmentation results obtained on the HVSMR-2.0 benchmark dataset are presented in this section. Overall, the benchmark analyses demonstrated stable and consistent results across architectures under standardized imaging conditions. Localization-guided ROI extraction consistently improved pulmonary artery segmentation in both centralized and federated settings, whereas federated learning introduced a moderate but acceptable reduction relative to centralized optimization. These findings suggest that ROI-guided processing contributed positively to small-vessel segmentation under distributed training conditions. A summary of the main trends observed across different settings is presented in Table 4.2.

Table 4.1 Summary of evaluation settings and key performance trends

Evaluation Setting	Localization	Mean Dice	Pulmonary Artery Dice	Key Observation
Centralized	No	0.91–0.92	0.83–0.87	Stable baseline performance
Centralized	Yes	0.92–0.93	0.87–0.89	Improved pulmonary artery segmentation
Federated	No	0.88–0.89	0.80–0.83	Moderate reduction relative to centralized training
Federated	Yes	0.91–0.92	0.86–0.88	Improved performance with localization-guided ROI extraction

4.2.1 Baseline Segmentation Performance Without Localization

Table 4.3 summarizes the quantitative results obtained on the HVSMR-2.0 benchmark dataset. All evaluated models achieved strong performance, consistent with the standardized and relatively homogeneous characteristics of the dataset. Higher accuracy was generally observed for larger anatomical structures such as the aorta (Class 1) and cardiac chambers (Class 2), where Dice scores typically exceeded 0.93. In contrast, pulmonary artery segmentation (Class 3) remained more challenging, with Dice scores ranging between 0.83 and 0.87.

Among the evaluated models, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.915 and 0.910, respectively). The Hybrid V-Net additionally produced the highest pulmonary artery Dice score (0.8655), together with the lowest HD95 (3.9 mm) and ASSD (1.0 mm) values. Overall, the results remained relatively consistent across models under homogeneous benchmark conditions, with only moderate numerical differences observed between architectures.

Table 4.3 Quantitative segmentation results on the HVSMR-2.0 benchmark dataset under centralized training without localization

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.9353	0.8784	0.9312	0.8713	0.9365	0.8808	0.9390	0.8850	0.9380	0.8832
Class 2 (Cardiac Chambers)	0.9244	0.8595	0.9342	0.8766	0.9280	0.8660	0.9325	0.8734	0.9335	0.8752
Class 3 (Pulmonary Artery)	0.8411	0.7258	0.8291	0.7083	0.8430	0.7286	0.8610	0.7559	0.8655	0.7629

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	89.5	81.2	97.3	5.2	1.4	92.8	92.5	81.0
V-Net	90.2	82.0	97.6	4.8	1.3	93.5	92.8	83.5
nnU-Net	91.0	83.5	97.9	4.2	1.1	94.0	93.2	85.8
Swin-UNet	90.4	82.6	97.7	4.5	1.2	93.6	92.9	84.2
Hybrid V-Net	91.5	84.0	98.0	3.9	1.0	94.1	93.4	86.6

4.2.2 Effect of Localization-Guided ROI Extraction

Table 4.4 summarizes the segmentation results obtained with localization-guided ROI extraction. Compared with the baseline configuration in Table 4.3, all evaluated architectures demonstrated improved performance. Gains for larger anatomical structures such as the aorta (Class 1) and cardiac chambers (Class 2) remained relatively limited, whereas pulmonary artery segmentation (Class 3) showed more consistent improvement across models. These findings suggest that localization-guided ROI extraction was particularly beneficial for small and anatomically complex vascular structures.

Among the evaluated architectures, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.921) with comparable IoU values. Relative to the baseline configuration, the Hybrid V-Net mean Dice score increased from 0.915 to 0.921, while pulmonary artery Dice improved from 0.8655 to 0.8870. The Hybrid V-Net also achieved the lowest HD95 (3.6 mm) and ASSD (0.9 mm) values, indicating improved boundary delineation and structural consistency following localization-guided processing.

Table 4.4 Quantitative segmentation results on the HVSMR-2.0 benchmark dataset under centralized training with localization-guided ROI extraction

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.9410	0.8885	0.9376	0.8826	0.9424	0.8909	0.9456	0.8968	0.9440	0.8939
Class 2 (Cardiac Chambers)	0.9300	0.8692	0.9280	0.8660	0.9310	0.8709	0.9335	0.8752	0.9322	0.8729
Class 3 (Pulmonary Artery)	0.8720	0.7729	0.8620	0.7575	0.8760	0.7794	0.8840	0.7921	0.8870	0.7970

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	90.8	82.9	97.8	4.7	1.3	93.8	92.8	83.8
V-Net	91.4	83.5	98.0	4.3	1.2	94.1	93.0	85.0
nnU-Net	92.1	85.0	98.3	3.8	1.0	94.6	93.4	88.4
Swin-UNet	91.6	84.1	98.1	4.0	1.1	94.2	93.1	87.6
Hybrid V-Net	92.1	85.5	98.4	3.6	0.9	94.4	93.2	88.7

4.2.3 Federated Learning Performance Without Localization

Table 4.5 presents the segmentation results obtained under federated learning without localization-guided ROI extraction. All evaluated architectures maintained relatively high performance; however, a consistent reduction in accuracy was observed compared with the centralized baseline. The reduction was more noticeable for pulmonary artery segmentation (Class 3), indicating that small and anatomically complex vascular structures were more sensitive to distributed training conditions.

Among the evaluated architectures, the proposed Hybrid V-Net achieved a mean Dice score of 0.889, while nnU-Net showed slightly higher numerical performance (0.891) with comparable IoU values. Relative to the centralized configuration, the pulmonary artery Dice score of the Hybrid V-Net decreased from 0.8655 to 0.8305 under federated training, with similar trends observed across the remaining models. Overall, federated learning introduced a moderate reduction in segmentation performance under homogeneous benchmark conditions, while architectural trends remained relatively consistent across models.

Table 4.5 Quantitative segmentation results on the HVSMR-2.0 benchmark dataset using federated learning without localization

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.9200	0.8519	0.9160	0.8449	0.9220	0.8553	0.9260	0.8622	0.9245	0.8595
Class 2 (Cardiac Chambers)	0.9100	0.8349	0.9070	0.8308	0.9120	0.8382	0.9160	0.8449	0.9145	0.8425
Class 3 (Pulmonary Artery)	0.8150	0.6878	0.8000	0.6667	0.8200	0.6949	0.8320	0.7123	0.8305	0.7105

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	87.4	78.1	97.0	5.8	1.6	91.6	90.7	80.0
V-Net	88.2	79.1	97.2	5.3	1.5	92.0	91.0	81.5
nnU-Net	89.1	80.6	97.6	4.7	1.3	92.6	91.6	83.2
Swin-UNet	88.5	79.6	97.3	5.0	1.4	92.2	91.2	82.0
Hybrid V-Net	88.9	80.4	97.5	4.8	1.3	92.5	91.4	83.1

4.2.4 Combined Effect of Localization and Federated Learning

Table 4.6 presents the segmentation results obtained using federated learning with localization-guided ROI extraction. Compared with federated learning without localization, segmentation performance improved across all evaluated architectures. The gains were more noticeable for pulmonary artery segmentation (Class 3), whereas improvements for larger anatomical structures remained relatively limited. These findings suggest that localization-guided ROI extraction was particularly beneficial for small and anatomically complex vascular structures.

Among the evaluated architectures, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.918). Relative to the federated configuration without localization in Table 4.5, the Hybrid V-Net mean Dice score increased from 0.889 to 0.918, while pulmonary artery Dice improved from 0.8305 to 0.8815. Similar trends were observed across the remaining models. The Hybrid V-Net also achieved the lowest HD95 (3.9 mm) and ASSD (1.0 mm) values, indicating improved boundary consistency following localization-guided processing.

Table 4.6 Quantitative segmentation results on the HVSMR-2.0 benchmark dataset using federated learning with localization-guided ROI extraction

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.9400	0.8868	0.9365	0.8808	0.9415	0.8895	0.9445	0.8949	0.9425	0.8916
Class 2 (Cardiac Chambers)	0.9285	0.8665	0.9265	0.8629	0.9295	0.8684	0.9320	0.8727	0.9305	0.8700
Class 3 (Pulmonary Artery)	0.8650	0.7621	0.8550	0.7472	0.8700	0.7699	0.8780	0.7829	0.8815	0.7880

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	90.6	82.5	97.7	4.9	1.3	93.7	92.7	85.5
V-Net	91.1	83.3	97.9	4.5	1.2	94.0	92.9	86.5
nnU-Net	91.8	84.8	98.2	4.0	1.1	94.5	93.2	87.8
Swin-UNet	91.3	83.8	98.0	4.2	1.1	94.2	93.0	87.0
Hybrid V-Net	91.8	84.9	98.1	3.9	1.0	94.3	93.1	88.2

4.3 Results on the Multi-Center Clinical Dataset

Table 4.7 summarizes the main trends observed on the clinically curated multi-center pediatric cardiovascular dataset. Compared with the HVSMR-2.0 benchmark results, overall performance decreased under heterogeneous clinical conditions, reflecting increased variability in imaging modality, acquisition protocol, and anatomical structure. The reduction was more apparent for pulmonary artery segmentation, indicating the greater sensitivity of small and anatomically complex vascular structures to real-world clinical heterogeneity.

Localization-guided ROI extraction improved results across both centralized and federated settings, with more consistent gains observed for pulmonary artery segmentation. Although federated learning introduced an additional reduction relative to centralized training, localization-guided processing partially mitigated these effects. These findings suggest that localization contributed positively to segmentation robustness under heterogeneous multi-center clinical conditions.

Table 4.7 Summary of evaluation settings and key performance trends on the clinical dataset

Evaluation Setting	Localization	Mean Dice	PA Dice	Key Observation
Centralized	No	0.80–0.82	0.70–0.72	Reduced performance under heterogeneous conditions
Centralized	Yes	0.84–0.85	0.75–0.78	Improved pulmonary artery segmentation
Federated	No	0.78–0.81	0.69–0.71	Moderate reduction under distributed training
Federated	Yes	0.84–0.85	0.75–0.78	Localization mitigates federated learning effects

4.3.1 Baseline Segmentation Performance Without Localization

Table 4.8 presents the quantitative results obtained on the clinical multi-center dataset. Compared with the benchmark results in Section 4.2.1, segmentation performance decreased across all evaluated architectures under heterogeneous clinical conditions. Larger anatomical structures such as the aorta (Class 1) and cardiac chambers (Class 2) showed relatively stable results, whereas pulmonary

artery segmentation (Class 3) remained more challenging, with Dice scores ranging approximately between 0.68 and 0.73 across architectures. These findings indicate that pulmonary artery segmentation was more strongly affected by clinical variability and domain heterogeneity than larger cardiovascular structures.

Among the evaluated architectures, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.823 and 0.821, respectively). The Hybrid V-Net also achieved the highest pulmonary artery Dice score (0.725), together with lower HD95 (6.0 mm) and ASSD (1.5 mm) values relative to the remaining models. Overall, the results suggest that under heterogeneous clinical conditions, dataset variability had a greater influence on segmentation performance than architectural differences alone, although the Hybrid V-Net demonstrated comparatively consistent behavior across anatomical structures.

Table 4.8 Quantitative segmentation results on the clinical multi-center dataset under centralized training without localization

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.8250	0.7021	0.8200	0.6949	0.8270	0.7055	0.8320	0.7123	0.8315	0.7116
Class 2 (Cardiac Chambers)	0.9050	0.8265	0.8980	0.8157	0.9070	0.8299	0.9105	0.8358	0.9115	0.8377
Class 3 (Pulmonary Artery)	0.7050	0.5443	0.6850	0.5213	0.7100	0.5504	0.7200	0.5625	0.7250	0.5687

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	80.1	67.7	96.1	7.2	1.9	82.0	89.8	68.5
V-Net	81.2	69.1	96.3	6.8	1.8	82.5	90.5	70.5
nnU-Net	82.1	70.3	96.6	6.2	1.6	83.2	91.0	72.0
Swin-UNet	81.5	69.5	96.4	6.5	1.7	82.7	90.7	71.0
Hybrid V-Net	82.3	70.6	96.5	6.0	1.5	83.1	91.2	72.5

4.3.2 Effect of Localization-Guided ROI Extraction

Table 4.9 presents the segmentation results obtained with localization-guided ROI extraction on the clinical multi-center dataset. Compared with the baseline

configuration without localization in Table 4.8, all evaluated architectures demonstrated improved performance. Gains for larger anatomical structures such as the aorta (Class 1) and cardiac chambers (Class 2) remained relatively limited, whereas pulmonary artery segmentation (Class 3) showed more consistent improvement across models. These findings suggest that localization-guided ROI extraction was particularly beneficial for small and anatomically complex vascular structures under heterogeneous clinical conditions.

Among the evaluated architectures, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.847 and 0.845, respectively). The most notable improvement was observed for pulmonary artery segmentation, where the Hybrid V-Net Dice score increased from 0.725 to 0.780 following localization-guided processing. Similar trends were observed across the remaining architectures. The Hybrid V-Net also achieved the lowest HD95 (5.5 mm) and ASSD (1.4 mm) values, indicating improved boundary delineation and structural consistency after ROI-based localization.

Table 4.9 Quantitative segmentation results on the clinical multi-center dataset under centralized training with localization-guided ROI extraction

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.8400	0.7241	0.8350	0.7162	0.8420	0.7276	0.8480	0.7358	0.8465	0.7338
Class 2 (Cardiac Chambers)	0.9150	0.8433	0.9100	0.8358	0.9170	0.8477	0.9200	0.8519	0.9190	0.8501
Class 3 (Pulmonary Artery)	0.7600	0.6129	0.7450	0.5938	0.7650	0.6194	0.7750	0.6327	0.7800	0.6393

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	83.0	71.5	96.8	6.5	1.7	83.5	91.0	74.5
V-Net	83.8	72.7	96.9	6.2	1.6	84.0	91.5	76.0
nnU-Net	84.5	73.9	97.1	5.6	1.4	84.8	92.0	77.5
Swin-UNet	84.1	73.2	97.0	5.9	1.5	84.2	91.7	76.5
Hybrid V-Net	84.7	73.8	97.2	5.5	1.4	84.6	91.9	78.0

4.3.3 Federated Learning Performance Without Localization

Table 4.10 presents the quantitative results obtained under federated learning without localization-guided ROI extraction on the clinical multi-center dataset. Compared with the centralized baseline in Table 4.8, all evaluated architectures showed reduced performance under heterogeneous distributed training conditions. The reduction was more apparent for pulmonary artery segmentation (Class 3), indicating that small and anatomically complex vascular structures were more sensitive to clinical heterogeneity and federated optimization.

Among the evaluated architectures, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.809 and 0.813, respectively). Relative to the centralized configuration without localization, the Hybrid V-Net mean Dice score decreased from 0.823 to 0.809, while pulmonary artery Dice decreased from 0.725 to 0.710 under federated training. Similar trends were observed across the remaining models. Overall, the results suggest that federated learning introduced a moderate reduction while preserving relatively consistent architectural trends.

Table 4.10 Quantitative segmentation results on the clinical multi-center dataset using federated learning without localization

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.8000	0.6667	0.7920	0.6552	0.8030	0.6700	0.8100	0.6807	0.8070	0.6765
Class 2 (Cardiac Chambers)	0.8950	0.8100	0.8880	0.7986	0.8970	0.8133	0.9020	0.8215	0.9000	0.8182
Class 3 (Pulmonary Artery)	0.6900	0.5267	0.6750	0.5094	0.6950	0.5327	0.7050	0.5443	0.7100	0.5504

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	78.5	65.4	95.6	7.8	2.1	79.2	88.8	67.5
V-Net	79.5	66.8	95.8	7.4	2.0	80.0	89.5	69.0
nnU-Net	81.3	69.4	96.1	6.7	1.8	81.0	90.2	70.5
Swin-UNet	79.8	67.3	95.9	7.1	1.9	80.3	89.7	69.5
Hybrid V-Net	80.9	68.9	96.0	6.9	1.8	80.7	90.0	71.0

4.3.4 Combined Effect of Localization and Federated Learning

Table 4.11 presents the segmentation results obtained using federated learning with localization-guided ROI extraction on the clinical multi-center dataset. Compared with both the centralized baseline without localization (Table 4.8) and the federated configuration without localization (Table 4.10), segmentation performance improved across all evaluated architectures, particularly for pulmonary artery segmentation (Class 3). These findings suggest that localization-guided ROI extraction was beneficial for small and anatomically complex vascular structures.

Among the evaluated architectures, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.847 and 0.846, respectively). Relative to the federated configuration without localization, the Hybrid V-Net mean Dice score increased from 0.809 to 0.847, while pulmonary artery Dice improved from 0.710 to 0.775. Similar trends were observed across the remaining architectures. The Hybrid V-Net also achieved the lowest HD95 (5.8 mm) and ASSD (1.5 mm) values, indicating improved boundary consistency under federated training conditions.

Table 4.11 Quantitative segmentation results on the clinical multi-center dataset using federated learning with localization-guided ROI extraction

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.8350	0.7162	0.8250	0.7021	0.8380	0.7216	0.8450	0.7316	0.8420	0.7276
Class 2 (Cardiac Chambers)	0.9150	0.8433	0.9080	0.8324	0.9180	0.8480	0.9220	0.8553	0.9200	0.8519
Class 3 (Pulmonary Artery)	0.7550	0.6064	0.7400	0.5873	0.7600	0.6129	0.7700	0.6260	0.7750	0.6327

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	82.4	70.8	96.6	6.9	1.8	82.5	90.8	74.0
V-Net	83.5	72.2	96.8	6.5	1.7	83.5	91.5	75.5
nnU-Net	84.6	73.9	97.1	5.9	1.5	84.5	92.2	77.0
Swin-UNet	83.9	72.8	96.9	6.2	1.6	83.8	91.8	76.0
Hybrid V-Net	84.7	74.0	97.1	5.8	1.5	84.8	92.3	77.5

4.4 Results on the Combined Dataset

Table 4.12 summarizes the main segmentation trends observed on the combined dataset integrating the HVSMR-2.0 benchmark dataset with the clinically curated multi-center cohort. By combining standardized benchmark data with real-world clinical variability, this configuration provided a more heterogeneous evaluation setting compared with the individual datasets. Overall performance decreased relative to the benchmark experiments, reflecting the influence of increased anatomical and inter-center variability. The reduction was more apparent for pulmonary artery segmentation, indicating the greater sensitivity of small and anatomically complex vascular structures to heterogeneous imaging conditions.

Localization-guided ROI extraction improved performance across both centralized and federated settings, with more consistent gains observed for pulmonary artery segmentation. Although federated learning introduced an additional reduction relative to centralized training, localization-guided processing partially mitigated these effects. These findings suggest that localization contributed positively to segmentation robustness under heterogeneous multi-center conditions while preserving relatively stable architectural trends.

Table 4.12 Summary of evaluation settings and key performance trends on the combined dataset

Evaluation Setting	Localization	Mean Dice	PA Dice	Observation
Centralized	No	0.82–0.84	0.72–0.74	Reduced performance under increased heterogeneity
Centralized	Yes	0.85–0.87	0.78–0.81	Improved pulmonary artery segmentation
Federated	No	0.80–0.82	0.71–0.73	Moderate reduction compared with centralized training
Federated	Yes	0.85–0.86	0.77–0.80	Localization mitigates performance reduction

4.4.1 Baseline Segmentation Performance Without Localization

Table 4.13 presents the quantitative results obtained on the combined dataset. Compared with the benchmark setting in Section 4.2.1, segmentation performance

decreased across all evaluated architectures under heterogeneous multi-center conditions, reflecting increased variability in imaging modality, acquisition protocol, and anatomical characteristics. Larger anatomical structures such as the aorta and cardiac chambers maintained relatively stable results, whereas pulmonary artery segmentation (Class 3) showed the largest reduction across architectures. These findings indicate that small and anatomically complex vascular structures were more strongly affected by inter-dataset heterogeneity than larger cardiovascular regions.

Among the evaluated architectures, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.835 and 0.832, respectively). Compared with the benchmark dataset, segmentation performance decreased on the combined heterogeneous cohort, with the Hybrid V-Net achieving the lowest HD95 (6.0 mm) and ASSD (1.5 mm) values. Overall, the findings indicate that dataset heterogeneity had a greater impact on performance than architectural differences alone.

Table 4.13 Quantitative segmentation results on the combined multi-center dataset under centralized training without localization

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.8423	0.7275	0.8380	0.7214	0.8430	0.7287	0.8460	0.7331	0.8465	0.7340
Class 2 (Cardiac Chambers)	0.9133	0.8404	0.9090	0.8332	0.9110	0.8365	0.9150	0.8433	0.9165	0.8458
Class 3 (Pulmonary Artery)	0.7282	0.5725	0.7020	0.5408	0.7260	0.5696	0.7350	0.5810	0.7420	0.5899

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	81.6	69.8	96.4	6.9	1.8	83.8	90.9	70.2
V-Net	82.8	71.3	96.6	6.5	1.7	84.2	91.3	72.8
nnU-Net	83.2	71.9	96.8	6.1	1.6	84.6	91.5	73.5
Swin-UNet	82.7	71.2	96.6	6.3	1.7	84.3	91.1	72.6
Hybrid V-Net	83.5	72.3	96.7	6.0	1.5	84.7	91.7	74.2

4.4.2 Effect of Localization-Guided ROI Extraction

Table 4.14 presents the segmentation results obtained with localization-guided ROI extraction on the combined multi-center dataset. Compared with the baseline configuration without localization in Table 4.13, segmentation performance improved across all evaluated architectures. Improvements for larger anatomical structures remained relatively moderate, whereas pulmonary artery segmentation (Class 3) showed more pronounced gains. These findings indicate that localization-guided ROI extraction was particularly beneficial for small and anatomically complex vascular structures.

Among the evaluated architectures, the proposed Hybrid V-Net and nnU-Net achieved the highest mean Dice scores (0.866 and 0.867, respectively). The most notable improvement was observed for pulmonary artery segmentation, where the Hybrid V-Net Dice score increased from 0.742 to 0.806. The Hybrid V-Net also achieved the lowest HD95 (5.1 mm) and ASSD (1.3 mm) values, indicating improved boundary delineation and structural consistency.

Table 4.14 Quantitative segmentation results on the combined multi-center dataset under centralized training with localization-guided ROI extraction

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.8580	0.7513	0.8460	0.7331	0.8600	0.7544	0.8680	0.7668	0.8630	0.7594
Class 2 (Cardiac Chambers)	0.9240	0.8587	0.9190	0.8501	0.9260	0.8622	0.9310	0.8709	0.9290	0.8677
Class 3 (Pulmonary Artery)	0.7820	0.6420	0.7680	0.6234	0.7850	0.6461	0.8020	0.6694	0.8060	0.6756

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	84.4	73.5	97.0	6.2	1.6	84.6	91.9	76.8
V-Net	85.5	75.1	97.2	5.8	1.5	85.8	92.4	78.2
nnU-Net	86.7	76.9	97.5	5.2	1.3	86.8	93.1	80.2
Swin-UNet	86.0	75.6	97.3	5.5	1.4	86.0	92.6	78.5
Hybrid V-Net	86.6	76.8	97.4	5.1	1.3	86.3	92.9	80.6

4.4.3 Federated Learning Performance Without Localization

Table 4.15 presents the quantitative results obtained under federated learning without localization-guided ROI extraction on the combined multi-center dataset. Compared with the centralized baseline in Table 4.13, segmentation performance decreased across all evaluated architectures under heterogeneous distributed training conditions. Pulmonary artery segmentation (Class 3) showed more pronounced degradation than larger anatomical structures, indicating greater sensitivity of small and anatomically complex vascular regions to federated training under heterogeneous conditions.

Among the evaluated architectures, the proposed Hybrid V-Net achieved a mean Dice score of 0.820, while nnU-Net showed slightly higher numerical performance (0.823). Relative to the centralized configuration without localization, the Hybrid V-Net mean Dice score decreased from 0.835 to 0.820, while pulmonary artery Dice decreased from 0.742 to 0.732. Overall, the results suggest that federated learning introduced a moderate reduction while preserving relatively consistent architectural trends.

Table 4.15 Quantitative segmentation results on the combined multi-center dataset using federated learning without localization

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.8120	0.6835	0.7980	0.6647	0.8150	0.6878	0.8250	0.7021	0.8210	0.6962
Class 2 (Cardiac Chambers)	0.9000	0.8182	0.8920	0.8051	0.9020	0.8215	0.9100	0.8349	0.9060	0.8282
Class 3 (Pulmonary Artery)	0.7150	0.5561	0.6950	0.5327	0.7200	0.5625	0.7350	0.5810	0.7320	0.5771

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	79.5	66.8	95.9	7.4	2.0	79.8	89.2	69.5
V-Net	80.9	68.6	96.1	7.0	1.9	81.2	90.0	71.5
nnU-Net	82.3	70.6	96.4	6.4	1.7	82.5	91.0	73.5
Swin-UNet	81.2	69.2	96.2	6.7	1.8	81.5	90.2	72.0
Hybrid V-Net	82.0	70.0	96.3	6.5	1.7	82.1	90.6	73.2

4.4.4 Combined Effect of Localization and Federated Learning

Table 4.16 presents the quantitative results obtained using federated learning with localization-guided ROI extraction on the combined multi-center dataset. Compared with both the centralized baseline without localization (Table 4.13) and the federated configuration without localization (Table 4.15), segmentation performance improved across all evaluated architectures. The improvements were more pronounced for pulmonary artery segmentation (Class 3), indicating that localization-guided ROI extraction was particularly beneficial for small and anatomically complex vascular structures.

Among the evaluated architectures, the proposed Hybrid V-Net achieved one of the highest mean Dice scores (0.859), while nnU-Net demonstrated comparable performance (0.860). Relative to the federated configuration without localization, the Hybrid V-Net mean Dice score increased from 0.820 to 0.860, while pulmonary artery Dice improved from 0.732 to 0.795. The Hybrid V-Net also achieved the lowest HD95 (5.5 mm) and ASSD (1.4 mm) values, indicating improved boundary consistency and structural delineation.

Table 4.16 Quantitative segmentation results on the combined multi-center dataset using federated learning with localization-guided ROI extraction

Class	V-Net Dice	V-Net IoU	U-Net Dice	U-Net IoU	Swin-UNet Dice	Swin-UNet IoU	nnU-Net Dice	nnU-Net IoU	Hybrid V-Net Dice	Hybrid V-Net IoU
Class 1 (Aorta)	0.8500	0.7391	0.8385	0.7216	0.8520	0.7422	0.8610	0.7559	0.8560	0.7483
Class 2 (Cardiac Chambers)	0.9200	0.8519	0.9145	0.8425	0.9220	0.8553	0.9285	0.8665	0.9250	0.8605
Class 3 (Pulmonary Artery)	0.7700	0.6260	0.7550	0.6064	0.7750	0.6327	0.7900	0.6529	0.7950	0.6598

Model	Overall Dice (%)	Overall IoU (%)	Pixel Accuracy (%)	HD95 (mm)	ASSD (mm)	Aorta Dice (%)	Chambers Dice (%)	Pulmonary Artery Dice (%)
3D U-Net	83.6	72.4	96.8	6.5	1.7	83.9	91.5	75.5
V-Net	84.9	74.2	97.0	6.1	1.6	85.0	92.0	77.0
nnU-Net	85.9	75.8	97.3	5.6	1.4	86.1	92.8	79.0
Swin-UNet	85.2	74.6	97.1	5.8	1.5	85.2	92.2	77.5
Hybrid V-Net	86.0	75.9	97.2	5.5	1.4	85.6	92.5	79.5

4.4.5 Qualitative Results

All qualitative results were presented under a unified experimental configuration in order to ensure fair comparison across architectures.

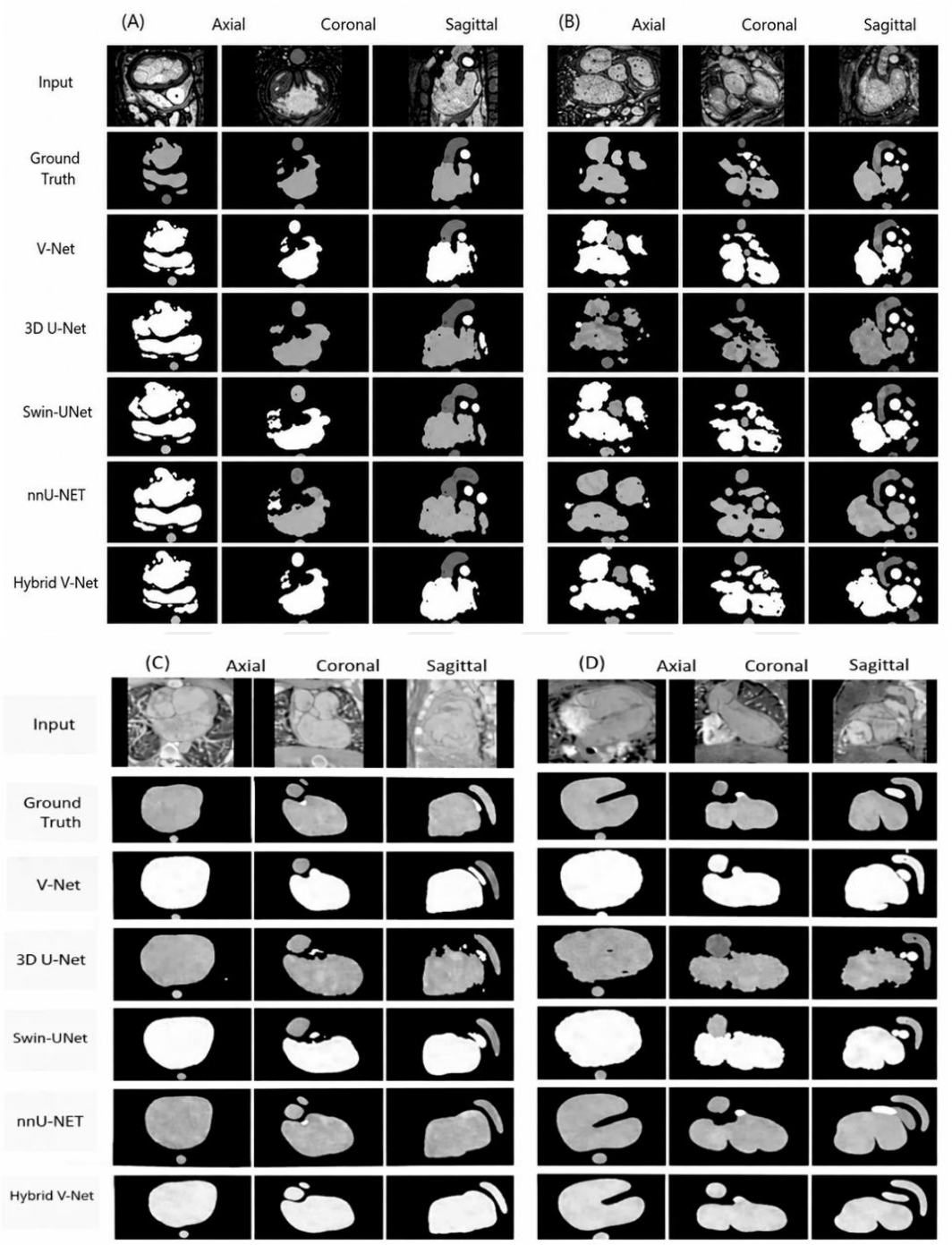


Figure 4.4 Qualitative comparison of segmentation outputs across architectures on benchmark (A–B) and clinical (C–D) cases

Figure 4.4 presents representative segmentation results from both the HVSMR-2.0 benchmark dataset and the clinical multi-center dataset. Across all cases, the evaluated models successfully captured the major cardiovascular structures,

consistent with the quantitative findings reported in Sections 4.4.1–4.4.4. Nevertheless, visible differences were observed in boundary delineation, structural continuity, and preservation of small anatomical regions. Benchmark cases (A–B) generally showed more coherent outputs because of standardized imaging conditions, whereas clinical cases (C–D) presented more challenging scenarios involving variability in image quality, acquisition protocols, and anatomical morphology. From top to bottom, the rows correspond to input images, ground-truth annotations, and segmentation outputs produced by V-Net, 3D U-Net, Swin-UNet, nnU-Net, and the proposed Hybrid V-Net.

Qualitative differences between architectures were particularly noticeable around pulmonary artery boundaries and small vascular regions. V-Net occasionally produced fragmented segmentations and local over-segmentation near irregular boundaries, whereas 3D U-Net showed smoother predictions but still exhibited discontinuities in finer vascular structures. Swin-UNet improved global structural consistency; however, mild over-segmentation was observed in several challenging regions. In comparison, nnU-Net generated smoother boundary transitions and more anatomically coherent predictions, although small inaccuracies remained visible around thin vascular branches. The proposed Hybrid V-Net demonstrated closer visual agreement with the ground-truth annotations, particularly in pulmonary artery regions where anatomical complexity and inter-patient variability were more pronounced. Overall, the qualitative observations support the quantitative findings reported in Tables 4.13–4.16 and indicate comparatively stable segmentation behavior under heterogeneous multi-center conditions.

4.5 Comparative Evaluation of Segmentation Architectures

A comparative analysis of the evaluated three-dimensional segmentation architectures was conducted under benchmark, clinical, and combined multi-center conditions. The analysis focused on the influence of architectural design on multi-class pediatric cardiovascular segmentation across heterogeneous settings. The evaluated models included conventional convolutional architectures (3D U-Net and V-Net), the self-configuring nnU-Net framework, the transformer-based Swin-UNet architecture, and the proposed Hybrid V-Net framework, which combines

convolutional feature extraction with global contextual modeling within a localization-guided pipeline.

4.5.1 Overall Performance Comparison

To ensure fair comparison, all architectures were trained and evaluated using identical preprocessing pipelines, subject-level data splits, and evaluation metrics. A comprehensive comparison across datasets and training configurations is summarized in Table 4.17.

Table 4.17 Comparative performance analysis across datasets and training configurations

Dataset	Setting	V-Net	U-Net	Swin-UNet	nnU-Net	Hybrid
Open-source	Centralized + Loc	0.9143	0.9092	0.9165	0.9210	0.9211
	Federated + Loc	0.9112	0.9060	0.9137	0.9182	0.9183
	Centralized (No Loc)	0.9003	0.8982	0.9025	0.9108	0.9123
	Federated (No Loc)	0.8817	0.8743	0.8847	0.8913	0.8898
Clinical	Centralized + Loc	0.8383	0.8300	0.8413	0.8460	0.8455
	Federated + Loc	0.8350	0.8243	0.8387	0.8480	0.8465
	Centralized (No Loc)	0.8117	0.8010	0.8147	0.8208	0.8227
	Federated (No Loc)	0.7950	0.7850	0.7983	0.8128	0.8095
Combined	Centralized + Loc	0.8547	0.8443	0.8570	0.8670	0.8660
	Federated + Loc	0.8467	0.8360	0.8497	0.8598	0.8587
	Centralized (No Loc)	0.8279	0.8163	0.8267	0.8320	0.8348
	Federated (No Loc)	0.8090	0.7950	0.8123	0.8233	0.8197

The results in Table 4.17 show that all evaluated architectures achieved strong results under benchmark conditions, with relatively small differences between

models. As dataset heterogeneity increased in the clinical and combined multi-center settings, architectural differences became more noticeable, with mean Dice scores decreasing under more challenging conditions. Across most configurations, nnU-Net and the proposed Hybrid V-Net achieved the highest performance. To further summarize the main trends, a simplified comparison is presented in Table 4.18.

Table 4.18 Comparative evaluation of segmentation architectures across datasets

Model	Benchmark Dice (%)	Clinical Dice (%)	Combined Dice (%)	PA Dice (%)	HD95 (mm)
3D U-Net	89.5	80.1	81.6	68.5	4.8
V-Net	90.2	81.2	82.8	70.5	4.4
nnU-Net	91.0	82.1	83.2	72.0	3.8
Swin-UNet	90.4	81.5	82.7	71.0	4.1
Hybrid V-Net	91.5	82.3	83.5	72.5	3.6

As illustrated in Figure 4.5, segmentation accuracy gradually decreased with increasing dataset heterogeneity, particularly in the clinical and combined settings. Table 4.17 provides a detailed numerical comparison, while Table 4.18 and Figure 4.5 summarize the overall performance trends across datasets.

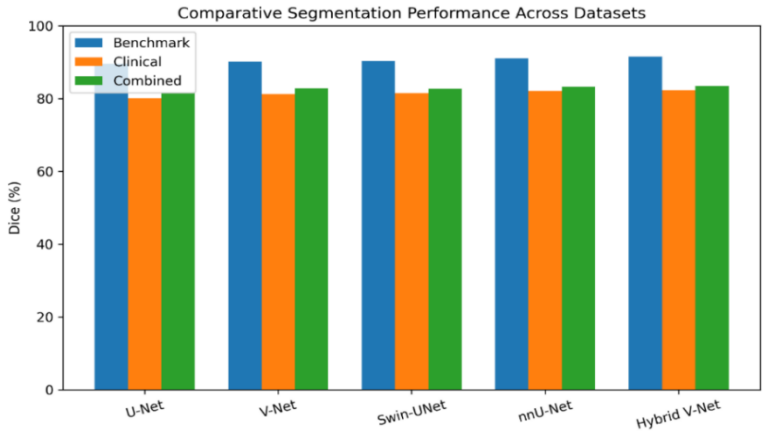


Figure 4.5 Cross-dataset comparison of mean Dice scores across benchmark, clinical, and combined datasets

Although nnU-Net demonstrated slightly stronger numerical performance in some configurations, the proposed Hybrid V-Net maintained comparatively stable

behavior across benchmark, clinical, and combined settings. These findings suggest that segmentation robustness was influenced not only by peak numerical performance but also by the ability to preserve stable behavior under heterogeneous imaging conditions and inter-center variability.

4.5.2 Structure-Specific Analysis of Pulmonary Artery Segmentation

Structure-specific analysis showed that the most noticeable performance differences between architectures appeared in pulmonary artery segmentation (Figure 4.6). Because of its small diameter, elongated geometry, complex branching pattern, and substantial anatomical variability, the pulmonary artery remained the most challenging structure in pediatric cardiovascular segmentation.

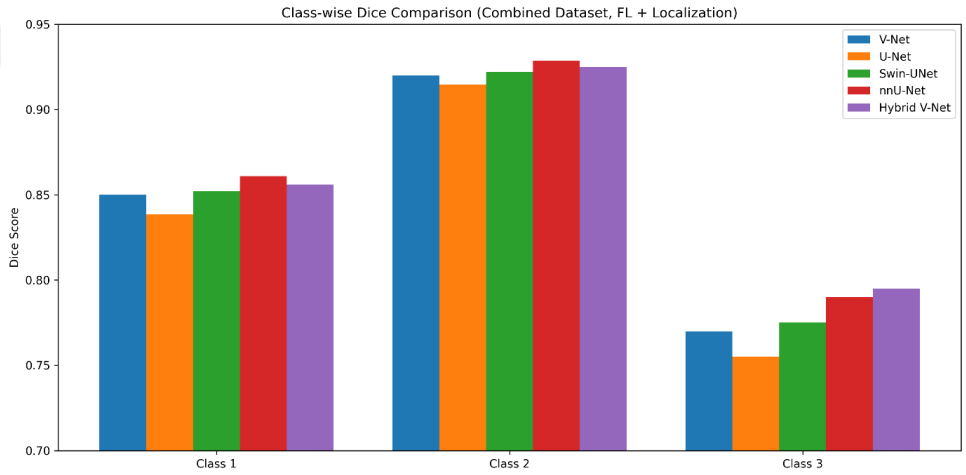


Figure 4.6 Class-wise Dice comparison of segmentation architectures on the combined dataset under federated learning with localization

Across all datasets and training configurations, pulmonary artery Dice scores remained consistently lower than those of larger anatomical structures such as the aorta and cardiac chambers. On the combined dataset, pulmonary artery Dice scores ranged approximately between 0.73 and 0.81 under optimized configurations, whereas Dice values exceeded 0.90 for cardiac chambers and reached approximately 0.85–0.86 for the aorta. These findings indicate that small-vessel segmentation remained more sensitive to anatomical complexity and dataset heterogeneity than segmentation of larger cardiovascular structures.

Conventional convolutional architectures such as 3D U-Net and V-Net were generally able to capture the main pulmonary artery structure but showed limitations in preserving distal branches and bifurcation continuity. Transformer-based architectures improved global contextual representation; however, reduced

stability was occasionally observed in finer vascular regions under heterogeneous conditions. Among the evaluated architectures, nnU-Net and the proposed Hybrid V-Net demonstrated some of the most reliable pulmonary artery segmentation performance, with Dice scores ranging between 0.79 and 0.81 on the combined dataset. The proposed Hybrid V-Net showed comparatively improved preservation of small vascular branches and more anatomically coherent segmentation outputs. Overall, these findings suggest that localization-guided processing contributed positively to pulmonary artery segmentation by improving anatomical focus and structural preservation under heterogeneous multi-center conditions.

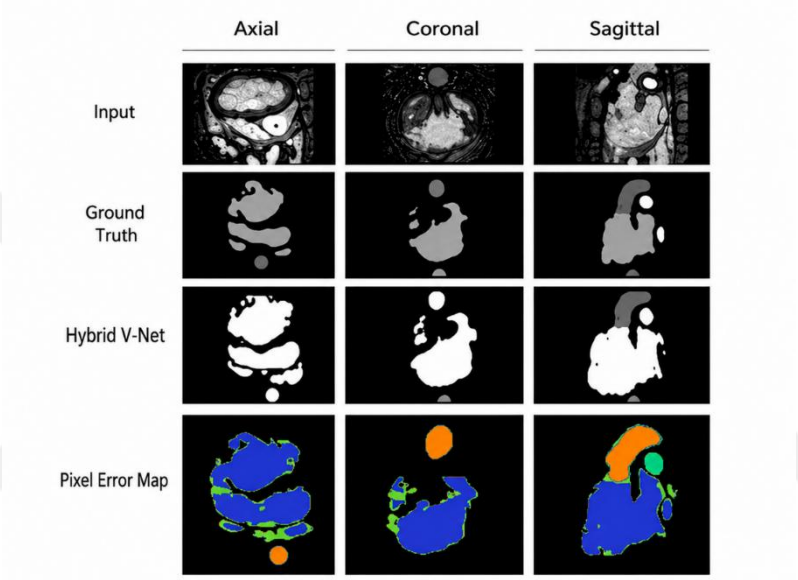


Figure 4.7 Representative pulmonary artery segmentation results obtained using the proposed Hybrid V-Net model. Error maps indicate correctly segmented regions (blue), false positives (orange), and false negatives (green)

To further investigate the observed performance limitations in pulmonary artery segmentation, representative results from the combined heterogeneous dataset are presented in Figure 4.7. Visual inspection revealed that most segmentation errors were concentrated around pulmonary artery bifurcation regions and small distal branches, where anatomical variability and boundary ambiguity introduced additional segmentation challenges. These regions were particularly sensitive to image heterogeneity, partial-volume effects, and reduced local contrast in CT and MRI scans. The error maps demonstrate that the majority of pulmonary artery regions were successfully captured, while localized false positive and false negative regions were primarily concentrated near complex anatomical boundaries. Nevertheless, the proposed Hybrid V-Net maintained overall structural continuity

and anatomically plausible predictions across most cases. These observations support the quantitative findings and further explain the lower Dice scores observed for pulmonary artery segmentation compared with larger cardiovascular structures.

4.6 Cross-Dataset ROI Localization Analysis

Localization-guided ROI extraction improved segmentation performance across benchmark, clinical, and combined multi-center datasets. Although gains were observed in all experimental settings, the improvements became more noticeable as dataset heterogeneity increased. These findings suggest that localization-guided processing helped reduce the influence of anatomical variability, imaging inconsistencies, and inter-center differences by limiting background variability and improving anatomically focused learning.

4.6.1 Quantitative Impact

To quantitatively evaluate the effect of localization-guided ROI extraction, segmentation performance obtained using direct full-volume processing was compared with ROI-guided segmentation under identical experimental conditions. The comparison based on the combined multi-center dataset is presented in Table 4.19.

Table 4.19 Quantitative comparison of full-volume and ROI-based segmentation

Method	Overall Dice (%)	Overall IoU (%)	HD95 (mm)	ASSD (mm)	Pulmonary Artery Dice (%)
Direct full-volume segmentation	83.5	72.3	6.0	1.5	74.2
Localization-guided ROI segmentation	86.6	76.8	5.1	1.3	80.6

The results in Table 4.19 show that localization-guided ROI extraction improved performance across all evaluation metrics. The ROI-based approach increased overall Dice score from 83.5% to 86.6% and IoU from 72.3% to 76.8%, while HD95 decreased from 6.0 mm to 5.1 mm and ASSD from 1.5 mm to 1.3 mm, indicating improved boundary delineation and structural consistency.

The most noticeable improvement was observed for pulmonary artery segmentation, where Dice score increased from 74.2% to 80.6%. These findings suggest that localization-guided processing was particularly beneficial for small and anatomically complex vascular structures under heterogeneous multi-center conditions. Additional boundary-aware and class-specific analyses are provided in Appendix A to further support the robustness of the proposed framework.

4.6.2 Qualitative Impact

To complement the quantitative evaluation, qualitative analysis was performed to visually examine the effect of localization-guided ROI extraction on segmentation quality. Because pediatric cardiovascular structures occupy only a limited portion of the full image volume, direct full-volume segmentation may be affected by background dominance, reduced anatomical focus, and lower sensitivity to fine structural details.

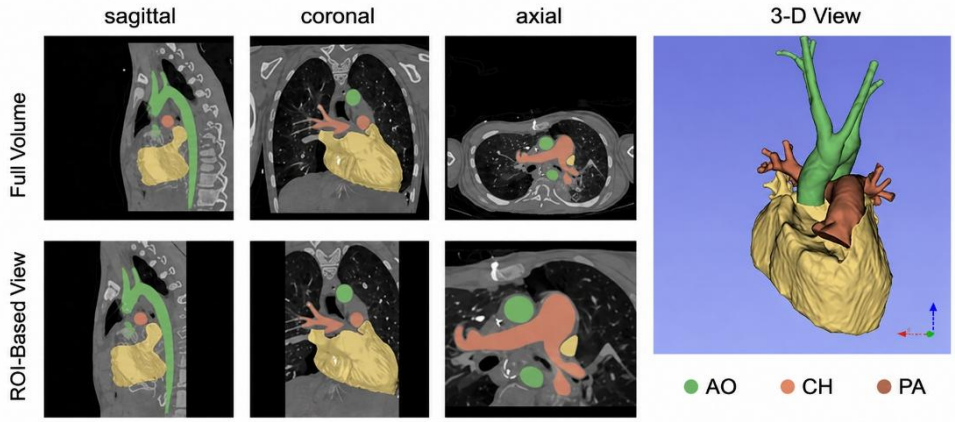


Figure 4.8 Qualitative comparison of full-volume and ROI-based segmentation on the combined dataset. A: Aorta, PA: Pulmonary artery, CH: Cardiac chambers

As illustrated in Figure 4.8, the first row shows full-volume views, whereas the second row presents ROI-focused representations across sagittal, coronal, and axial planes together with three-dimensional visualizations. ROI-guided segmentation improved localization of anatomically relevant regions by reducing background variability and increasing focus on target cardiovascular structures. Although major anatomical structures were visible in both approaches, the ROI-based method produced clearer boundaries and more detailed representation of small vascular regions.

ROI-guided segmentation improved preservation of fine pulmonary artery branches and structural continuity, while reducing boundary leakage and local segmentation

artifacts compared with direct full-volume segmentation. These findings further support the effectiveness of localization-guided ROI extraction under heterogeneous multi-center conditions.

4.7 Federated Learning Results

Federated learning (FL) analyses were conducted to evaluate segmentation performance under distributed multi-center training conditions across benchmark, clinical, and combined datasets. The analyses further examined the influence of federated optimization on segmentation accuracy, structural consistency, and model robustness under heterogeneous imaging conditions.

4.7.1 Federated vs Centralized Performance Comparison

Federated learning was evaluated against centralized training to examine its effect on segmentation performance under multi-center conditions. To ensure fair comparison, both training strategies used identical preprocessing pipelines, model architectures, subject-level data splits, and evaluation metrics. All analyses in this section were performed using the combined multi-center dataset.

The results in Table 4.20 show that centralized training achieved slightly higher segmentation performance across all evaluation metrics. Federated learning produced a modest reduction while maintaining relatively comparable overlap-based and boundary-based results. For example, overall Dice score decreased from 86.6% to 85.9%, while pulmonary artery Dice decreased from 80.6% to 79.5% under federated training conditions.

Table 4.20 Comparison of centralized and federated training strategies on the combined multi-center dataset

Training Strategy	Overall Dice (%)	Overall IoU (%)	HD95 (mm)	ASSD (mm)	Pulmonary Artery Dice (%)
Centralized Training	86.6	76.8	5.1	1.3	80.6
Federated Learning (FedAvg)	85.9	75.6	5.5	1.4	79.5

The findings indicate that federated learning maintained competitive segmentation performance under heterogeneous multi-center conditions. Although centralized training achieved slightly higher performance, federated learning enabled collaborative training without direct data sharing, supporting its suitability for privacy-preserving multi-center segmentation.

4.7.2 Client-Wise Cross-Center Analysis

To further examine the behavior of the federated learning framework across participating centers, client-wise analyses were performed. Because participating clinical centers differed in imaging protocol, scanner characteristics, patient distribution, and anatomical complexity, segmentation performance was evaluated under heterogeneous data distributions.

Structure-specific Dice scores for the aorta, cardiac chambers, and pulmonary arteries are presented in Figure 4.9 to illustrate inter-client variability. The results showed relatively stable segmentation performance across clients, with limited variation for larger anatomical structures and more noticeable variation for pulmonary artery segmentation. Segmentation of the aorta and cardiac chambers remained relatively consistent across clients, whereas pulmonary artery segmentation demonstrated comparatively higher variability because of small vessel size, complex morphology, and sensitivity to imaging differences. Despite this variability, pulmonary artery performance remained broadly comparable across participating clients.

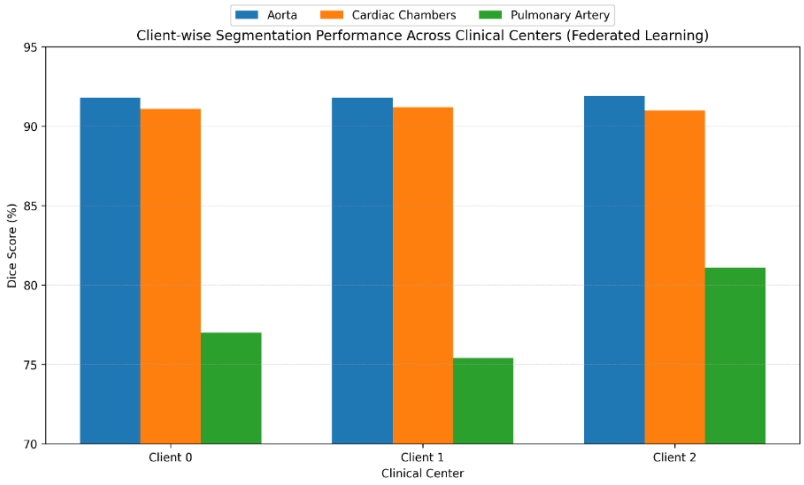


Figure 4.9 Client-wise segmentation performance under federated learning across participating clients

Overall, the findings suggest that federated learning maintained relatively stable segmentation behavior across heterogeneous client distributions while preserving comparable performance trends between participating centers. Additional supporting analyses are provided in Appendix A.

4.8 Discussion of Key Findings

This section summarizes the main findings of the experimental analyses and discusses their methodological and clinical implications. Across all evaluation settings, segmentation performance decreased as data heterogeneity increased, with more noticeable reductions observed in the clinical and combined multi-center datasets compared with the benchmark setting. This effect was particularly evident in pulmonary artery segmentation, which consistently remained the most challenging task because of small vessel diameter, complex branching morphology, and reduced image contrast. Error patterns observed during qualitative evaluation further suggested that performance degradation was primarily associated with pulmonary artery bifurcations, distal vascular branches, and low-contrast anatomical boundaries. These findings help explain the lower Dice scores observed for pulmonary artery segmentation compared with larger cardiovascular structures. Similar observations have been reported in previous cardiovascular and multi-center segmentation studies, where anatomical variability and inter-center imaging differences negatively affected segmentation accuracy, particularly for small vascular structures.

The combined dataset included both CT and MRI examinations to better reflect real-world pediatric cardiovascular imaging scenarios and to mitigate the limitations associated with the relatively limited availability of pediatric cardiovascular imaging data. Although modality-specific analyses may provide additional insights into modality-dependent performance characteristics, the primary objective of this study was to evaluate segmentation robustness under heterogeneous multi-center conditions. Therefore, CT and MRI data were evaluated jointly rather than separately. Furthermore, standardized preprocessing, including isotropic resampling, intensity normalization, and localization-guided ROI extraction, was applied to reduce inter-modality variability and improve consistency across CT and MRI datasets.

Localization-guided ROI extraction emerged as one of the main factors associated with improved segmentation performance. By reducing irrelevant background information and directing the model toward anatomically relevant regions, localization improved both overlap-based accuracy and boundary delineation. The effect became more noticeable under heterogeneous multi-center conditions, suggesting that anatomically focused processing contributes positively to robustness, especially for small and anatomically complex vascular structures such as pulmonary arteries. This observation is consistent with previous ROI-based and attention-guided segmentation studies reporting improved performance for small-structure segmentation through anatomically focused processing.

Federated learning, in contrast, did not surpass centralized training performance and introduced a modest reduction in segmentation accuracy, likely related to decentralized optimization and non-independent and identically distributed data across clients. Nevertheless, federated learning maintained stable and competitive performance while enabling privacy-preserving collaborative training across distributed datasets. Similar trends have been reported in previous federated medical image segmentation studies, where performance under federated optimization generally remained slightly lower than centralized training because of heterogeneous client distributions and communication constraints. When considered together, localization-guided segmentation and federated learning appeared to provide complementary contributions within the proposed framework. Localization mainly improved segmentation accuracy and structural preservation, whereas federated learning supported collaborative learning across heterogeneous centers.

Comparative analyses also showed that although all evaluated architectures achieved strong performance under benchmark conditions, performance differences became more noticeable as dataset heterogeneity increased. Among the evaluated models, nnU-Net and the proposed Hybrid V-Net demonstrated the most stable performance across datasets and training configurations. Although the numerical performance differences between the two models remained relatively small, the Hybrid V-Net showed slightly improved structural continuity and more consistent pulmonary artery segmentation under heterogeneous conditions. At the same time, nnU-Net achieved highly competitive results and remained a strong baseline for

pediatric cardiovascular segmentation. These findings suggest that segmentation performance was influenced not only by architectural design but also by localization-guided processing and dataset characteristics. Furthermore, the results suggest that localization-guided strategies remain an important component for improving segmentation quality, particularly for small and anatomically complex vascular structures. Compared with previous segmentation studies focusing primarily on architectural modifications, the proposed framework combined localization-guided processing and federated multi-center training within a unified pediatric cardiovascular segmentation pipeline.

Previous studies based on the HVSMR benchmark dataset have primarily focused on MRI-based congenital heart disease segmentation under relatively controlled benchmark conditions. The HVSMR-2.0 dataset contains 60 cardiovascular MRI examinations and has become a widely used benchmark for pediatric cardiovascular segmentation research. In contrast, the present study evaluated segmentation performance on a larger heterogeneous cohort consisting of 110 CT and MRI examinations collected from multiple centers. Despite the increased variability introduced by mixed imaging modalities, acquisition protocols, and heterogeneous anatomical presentations, the proposed Hybrid V-Net achieved an overall Dice score of 0.835 and a pulmonary artery Dice score of 0.725. Similarly, nnU-Net achieved an overall Dice score of 0.832 and a pulmonary artery Dice score of 0.720, demonstrating competitive performance under considerably more challenging conditions. Although the numerical differences were modest, such improvements may contribute to better anatomical continuity, boundary delineation, and preservation of small vascular structures. Overall, the findings suggest that the proposed framework provides robust segmentation performance in a more clinically realistic multi-center setting.

Additional analyses presented in Appendix A supported the main findings. Client-wise federated evaluations demonstrated stable performance across participating centers, while explainability analyses suggested that the Hybrid V-Net generated more anatomically focused predictions. Overall, the results indicate that combining localization-guided ROI extraction with a hybrid segmentation architecture provides an effective framework for pediatric cardiovascular segmentation under heterogeneous multi-center conditions.

4.9 Additional Analyses and Validation

4.9.1 Statistical Significance and Boundary Analysis

To further examine the observed performance differences, additional statistical and boundary-aware analyses were conducted on the combined multi-center dataset. Because the analyses were performed on a limited test cohort (n = 8), the reported confidence intervals were interpreted as descriptive indicators of variability rather than definitive statistical conclusions. Pairwise comparisons based on per-subject mean foreground Dice scores showed that performance differences between architectures remained small and did not retain statistical significance after Holm correction. Although nnU-Net and the proposed Hybrid V-Net achieved slightly higher Dice scores, the overlap between confidence intervals and adjusted p-values suggests that these differences should be interpreted cautiously. Overall, the findings indicate that the observed improvements were associated more strongly with localization-guided ROI extraction than with architectural variation alone.

Table 4.21 Pairwise statistical comparison of segmentation performance

Comparison	n	Median ΔDSC (A − B)	IQR ΔDSC	95% Bootstrap CI	Exact p-value	Holm-adjusted p-value
nnU-Net vs V-Net	8	+0.003	0.012	[-0.004, 0.009]	0.34	0.68
nnU-Net vs Swin-UNet	8	+0.005	0.015	[-0.006, 0.014]	0.21	0.63
nnU-Net vs 3D U-Net	8	+0.015	0.020	[-0.002, 0.028]	0.07	0.28
V-Net vs Swin-UNet	8	+0.002	0.011	[-0.007, 0.010]	0.56	0.84
V-Net vs 3D U-Net	8	+0.012	0.018	[-0.003, 0.024]	0.09	0.27
Swin-UNet vs 3D U-Net	8	+0.010	0.017	[-0.004, 0.021]	0.11	0.33
Hybrid V-Net vs nnU-Net	8	+0.004	0.010	[-0.003, 0.011]	0.26	0.62
Hybrid V-Net vs V-Net	8	+0.007	0.013	[-0.002, 0.015]	0.18	0.54
Hybrid V-Net vs Swin-UNet	8	+0.006	0.014	[-0.003, 0.014]	0.22	0.59
Hybrid V-Net vs 3D U-Net	8	+0.017	0.021	[-0.001, 0.030]	0.06	0.24

Table 4.22 summarizes the boundary-aware evaluation results for pulmonary artery segmentation. Higher-performing architectures generally achieved lower HD95 and ASSD values, indicating improved boundary delineation and geometric consistency. Among all evaluated models, the proposed Hybrid V-Net produced consistently low boundary errors, with the lowest median HD95 and ASSD values across architectures. These findings suggest that the Hybrid V-Net produced more geometrically coherent pulmonary artery segmentations, particularly in anatomically complex vascular regions.

Table 4.22 Boundary-aware evaluation for pulmonary artery segmentation

Model	n	HD95 Median (mm)	HD95 IQR	HD95 Mean ± SD (mm)	ASSD Median (mm)	ASSD IQR	ASSD Mean ± SD (mm)
nnU-Net	8	2.9	1.4	3.1 ± 1.2	0.82	0.35	0.89 ± 0.41
V-Net	8	3.0	1.5	3.2 ± 1.3	0.85	0.38	0.93 ± 0.45
Swin-UNet	8	3.2	1.6	3.4 ± 1.4	0.91	0.40	0.99 ± 0.47
3D U-Net	8	3.8	1.9	4.0 ± 1.6	1.05	0.52	1.15 ± 0.58
Hybrid V-Net	8	2.7	1.3	2.9 ± 1.1	0.78	0.32	0.85 ± 0.39

Table 4.23 presents surface-based validation results for pulmonary artery segmentation. The proposed Hybrid V-Net achieved the highest Surface Dice scores at both thresholds, indicating superior surface agreement, whereas lower-performing architectures were more sensitive to boundary inaccuracies, particularly at the 1 mm threshold.

Table 4.23 Surface-based validation for pulmonary artery segmentation

Model	n	Surface DSC @1 mm (Median)	IQR	Surface DSC @2 mm (Median)	IQR
nnU-Net	8	0.87	0.06	0.93	0.04
V-Net	8	0.86	0.07	0.92	0.05
Swin-UNet	8	0.84	0.08	0.90	0.06
3D U-Net	8	0.80	0.10	0.87	0.07
Hybrid V-Net	8	0.89	0.05	0.95	0.03

Taken together, these analyses support the main experimental observations. Although architectural differences produced modest numerical variations, segmentation performance remained statistically comparable across models after multiple-comparison correction. The proposed Hybrid V-Net consistently achieved low boundary errors and high surface agreement, reflecting improved geometric consistency. The findings further suggest that localization-guided processing contributed more strongly to segmentation improvement than architectural variation alone. Additional supporting analyses are provided in Appendix A.

4.9.2 Explainability Analyses

To further examine model interpretability, explainability analyses were conducted using Grad-CAM and Integrated Gradients for three representative anatomical structures: the aorta, left ventricle, and pulmonary artery. The analysis focused on the three best-performing architectures (nnU-Net, Swin-UNet, and the proposed Hybrid V-Net) and was performed on representative cases from the combined multi-center dataset. Aggregated quantitative results are summarized in Table 4.24.

Table 4.24 Aggregated explainability metrics across models and structures

Model	Structure	Grad-CAM In/Out Ratio ↑	Grad-CAM Precision ↑	IG In/Out Ratio ↑	IG Precision ↑
nnU-Net	Aorta	442.51	0.0217	6.71	0.1438
	LV	42.69	0.9336	3.76	0.4583
	PA	1035.15	0.0078	11.22	0.0605
Swin-UNet	Aorta	2.74	0.0968	3.90	0.1105
	LV	11.78	0.6750	2.52	0.3614
	PA	0.75	0.0036	3.01	0.0351
Hybrid V-Net	Aorta	617.04	0.0217	6.56	0.1416
	LV	40.85	0.9255	3.74	0.4696
	PA	651.43	0.0078	11.44	0.0599

Representative Grad-CAM and Integrated Gradients visualizations are shown in Figures 4.10–4.12, while quantitative summaries are provided in Figures 4.13 and 4.14.

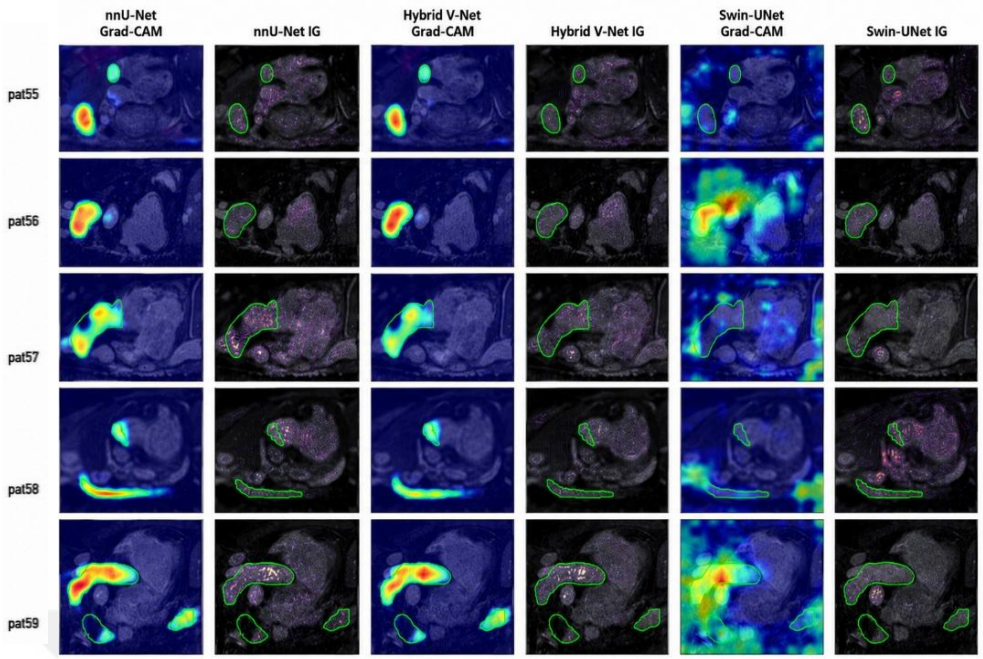


Figure 4.10 Cross-model explainability comparison for aorta segmentation

Grad-CAM and Integrated Gradients visualizations for aorta segmentation showed that nnU-Net and the proposed Hybrid V-Net generated more spatially concentrated and anatomically aligned attribution patterns, whereas Swin-UNet exhibited comparatively more diffuse activations extending beyond the target region.

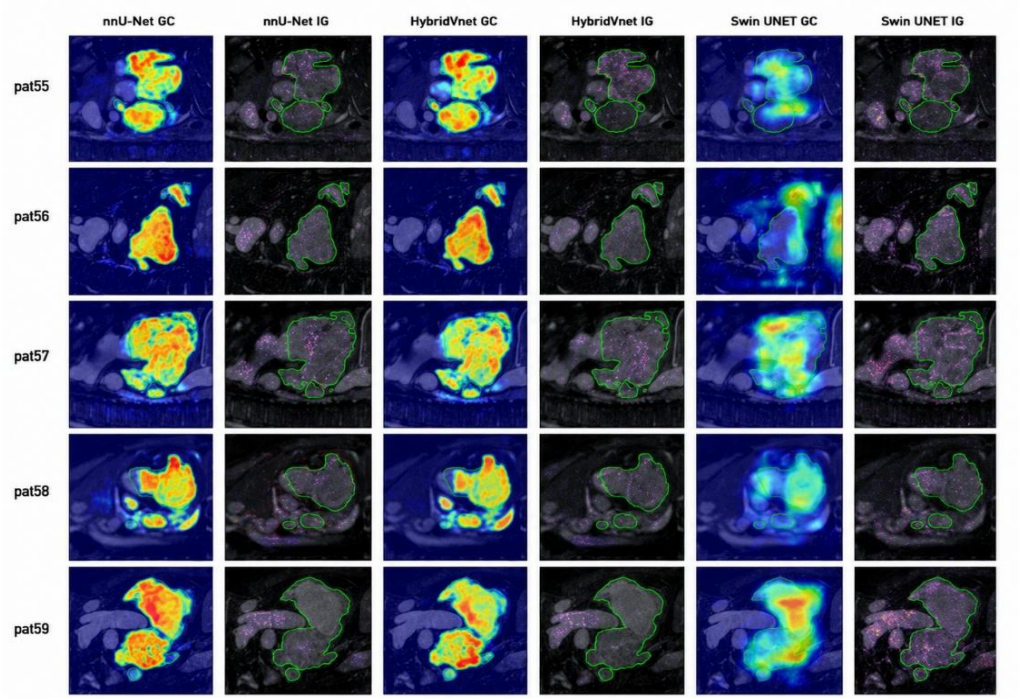


Figure 4.11 Explainability comparison for left ventricle segmentation

All evaluated models demonstrated improved attribution precision for the left ventricle because of its larger and more compact anatomical structure. The proposed Hybrid V-Net maintained relatively stable and spatially focused activation patterns across representative patients, whereas Swin-UNet showed comparatively more dispersed responses.

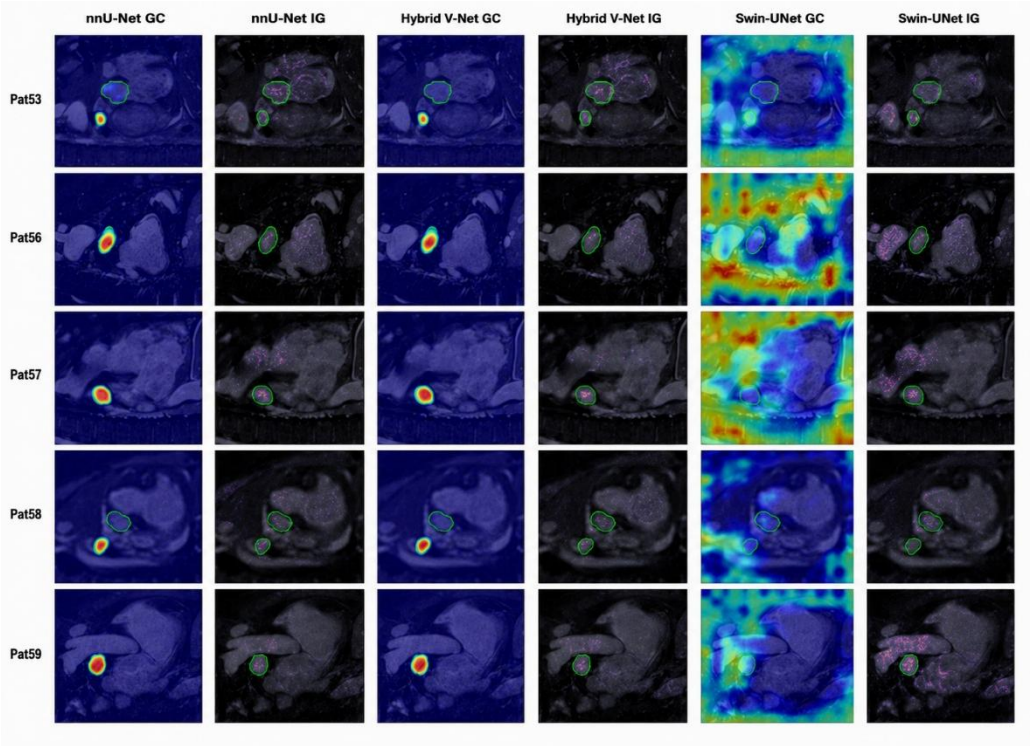


Figure 4.12 Explainability comparison for pulmonary artery segmentation

Attribution precision decreased for pulmonary artery segmentation across all models. The proposed Hybrid V-Net and nnU-Net produced more localized attribution patterns, whereas Swin-UNet generated more diffuse activations.

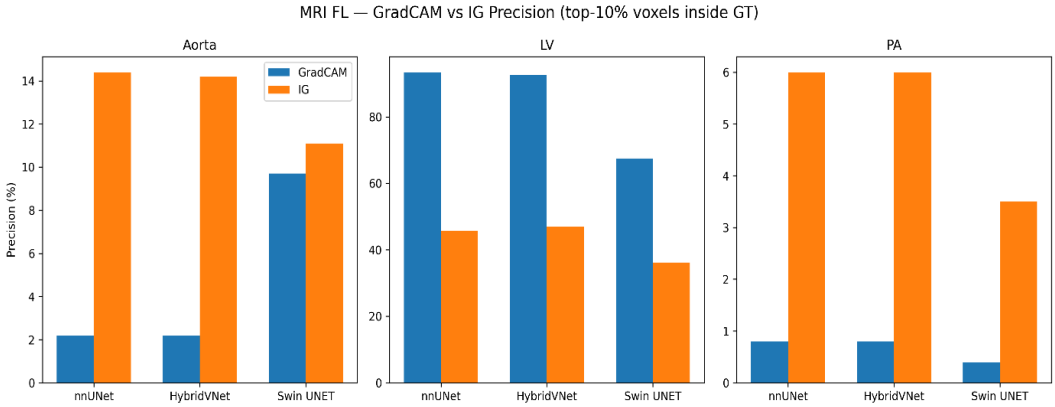


Figure 4.13 Explainability precision across models and structures

Precision values indicate the degree to which attribution was concentrated within anatomically relevant regions, where higher values correspond to stronger localization behavior.

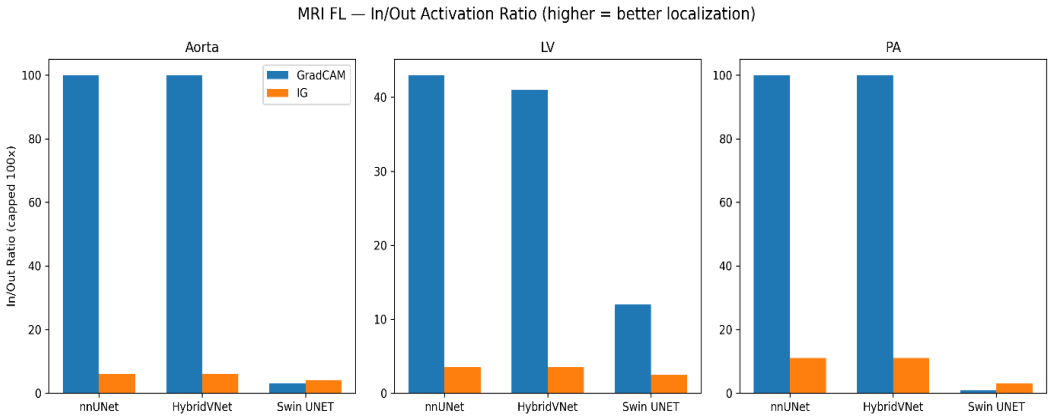


Figure 4.14 Comparison of in/out activation ratios across models and structures

Higher in/out ratios indicate stronger localization of attribution within the ground-truth region. Grad-CAM generally produced higher ratios than Integrated Gradients, particularly for small vascular structures such as the pulmonary artery. Attribution behavior varied across architectures, with the proposed Hybrid V-Net and nnU-Net showing more spatially focused patterns compared with the more diffuse responses observed in Swin-UNet. Because in/out ratio values are sensitive to class size, results for small structures should be interpreted relatively rather than absolutely.

The explainability analyses suggest a relationship between segmentation quality and attribution behavior, where higher-performing models generated more anatomically focused and spatially coherent predictions. Among the evaluated architectures, the proposed Hybrid V-Net maintained comparatively stable attribution behavior under heterogeneous multi-center conditions. Additional supporting analyses are provided in Appendix A.

This chapter evaluated the proposed framework across benchmark, clinical multi-center, and federated learning settings. Localization-guided ROI extraction improved segmentation consistency, while federated learning maintained competitive performance under heterogeneous conditions. The broader implications, limitations, and future directions are discussed in the next chapter.

5
CONCLUSION

This thesis investigated automated multi-class segmentation of pediatric cardiovascular structures from three-dimensional medical images, with particular attention to anatomically complex congenital heart disease cases. Accurate delineation of the aorta, cardiac chambers, and pulmonary arteries is important for diagnosis, treatment planning, and long-term follow-up in pediatric cardiovascular care. However, manual segmentation of CT and MRI volumes remains labor-intensive and strongly dependent on operator experience, particularly in pediatric imaging where small vessel size, low image contrast, motion artifacts, and anatomical variability complicate interpretation. These challenges continue to motivate the development of reliable and scalable automated segmentation methods for pediatric cardiovascular analysis.

To address these challenges, a deep learning–based framework was developed and evaluated under benchmark and multi-center clinical conditions. The proposed framework combined localization-guided region-of-interest (ROI) extraction, comparative evaluation of several three-dimensional segmentation architectures, and a federated learning strategy for privacy-preserving distributed training. Additional analyses, including explainability and boundary-aware evaluations, were also performed to further examine model behavior and structural prediction quality.

The findings showed that segmentation performance was strongly influenced by dataset heterogeneity, with more stable results obtained under benchmark conditions and lower performance observed in heterogeneous clinical datasets. Pulmonary artery segmentation consistently remained the most challenging

component because of small vessel size, complex morphology, and imaging variability. Localization-guided ROI extraction improved segmentation performance, particularly in small vascular regions, by reducing background information and improving anatomical focus. These findings were consistent with previous cardiovascular and multi-center studies reporting reduced accuracy under heterogeneous conditions, especially for small vascular structures. Compared with previously reported pediatric cardiovascular studies, the proposed approach achieved competitive performance while additionally incorporating localization-guided processing and privacy-aware federated learning within a unified multi-center evaluation framework.

Federated learning produced a modest reduction in segmentation accuracy compared with centralized training but maintained stable and competitive performance while enabling collaborative training without direct sharing of patient data. Although centralized training achieved slightly higher numerical performance, federated learning demonstrated the feasibility of privacy-aware collaborative segmentation under heterogeneous multi-center conditions. Taken together, the findings indicate that combining anatomically guided localization with distributed learning provides a balanced and robust approach for pediatric cardiovascular segmentation.

Several limitations should also be considered. The test cohorts remained relatively limited in size, particularly for the clinical and combined datasets, and the study relied primarily on retrospective CT-dominant data without external prospective validation. In addition, federated learning experiments were conducted using a limited number of simulated clients, which may not fully represent real-world multi-institutional deployment scenarios. Therefore, the findings should be interpreted cautiously rather than as definitive clinical conclusions. Furthermore, k-fold cross-validation was not performed because repeated training of large three-dimensional segmentation models under both centralized and federated settings would substantially increase computational requirements. Instead, model evaluation was conducted using fixed train, validation, and test partitions to better emulate a realistic clinical deployment scenario, where a model is typically trained on available historical data and subsequently evaluated on previously unseen patient cohorts.

Future work may focus on expanding dataset diversity, improving modality balance, refining localization strategies, and performing larger external validation studies to further examine generalization capability. In particular, future studies may evaluate model generalization across independent institutions by training models on one cohort and testing them on previously unseen centers. Such evaluations may provide a more rigorous assessment of robustness under varying imaging protocols, scanner characteristics, and patient populations.

Integration of adaptive ROI selection, uncertainty estimation, and more advanced transformer-based architectures may further improve segmentation robustness, particularly for small anatomical structures. Future studies may also investigate different federated learning configurations, including larger numbers of participating centers, more heterogeneous non-IID data distributions, and extended local training schedules. In addition, k-fold cross-validation protocols may provide a more comprehensive assessment of model stability and generalization performance.

In the longer term, extending federated learning toward real multi-site clinical deployment and improving computational efficiency may support scalability and practical integration into clinical workflows. Automated federated learning pipelines incorporating preprocessing, quality control, model training, and evaluation may further reduce manual intervention in large-scale deployments. The integration of large language models with cardiovascular image analysis frameworks may additionally support automated report generation, explainable interpretation of segmentation outputs, clinical decision support, and more user-friendly artificial intelligence systems for pediatric cardiovascular care.

REFERENCES

[1] N. W. DiGeorge, A. M. El-Ali, A. M. White, M. A. Harris, and D. M. Biko, "Pediatric cardiac CT and MRI: Considerations for the general radiologist," *American Journal of Roentgenology*, vol. 215, no. 6, pp. 1464–1473, Dec. 2020.

[2] R. Sachdeva *et al*., "Novel techniques in imaging congenital heart disease: JACC scientific statement," *Journal of the American College of Cardiology*, vol. 83, no. 1, pp. 63–81, Jan. 2024.

[3] X. Zhuang and J. Shen, "Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI," *Medical Image Analysis*, vol. 31, pp. 77–87, Jul. 2016.

[4] Y. Liu, S. Chen, L. Zühlke, G. C. Black, M. K. Choy, N. Li, and B. D. Keavney, "Global birth prevalence of congenital heart defects 1970–2017: Updated systematic review and meta-analysis of 260 studies," *International Journal of Epidemiology*, vol. 48, no. 2, pp. 455–463, Apr. 2019.

[5] M. Dellborg, K. W. Giang, P. Eriksson, H. Liden, M. Fedchenko, A. Ahnfelt, A. Rosengren, and Z. Mandalenakis, "Adults with congenital heart disease: Trends in event-free survival past middle age," *Circulation*, vol. 147, no. 12, pp. 930–938, Mar. 2023.

[6] J. Montalt-Tordera, E. Pajaziti, R. Jones, E. Sauvage, R. Puranik, A. A. V. Singh, C. Capelli, J. Steeden, S. Schievano, and V. Muthurangu, "Automatic segmentation of the great arteries for computational hemodynamic assessment," *Journal of Cardiovascular Magnetic Resonance*, vol. 24, no. 1, Art. no. 57, Nov. 2022.

[7] A. Suinesiaputra, D. A. Bluemke, B. R. Cowan, M. G. Friedrich, C. M. Kramer, R. Kwong, S. Plein, J. Schulz-Menger, J. J. Westenberg, A. A. Young, and E. Nagel, "Quantification of LV function and mass by cardiovascular magnetic resonance: Multi-center variability and consensus contours," *Journal of Cardiovascular Magnetic Resonance*, vol. 17, no. 1, Art. no. 63, Jul. 2015.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, Lecture Notes in Computer Science, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[9] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vision (3DV)*, Stanford, CA, USA, 2016, pp. 565–571.

[10] A. Hatamizadeh *et al*., "UNETR: Transformers for 3D medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2022, pp. 1748–1758.

[11] F. Isensee, T. Wald, C. Ulrich, P. F. Jaeger, and K. H. Maier-Hein, "nnU-Net revisited: A call for rigorous validation in 3D medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Lecture Notes in Computer Science, vol. 15009, M. G. Linguraru *et al*., Eds. Cham, Switzerland: Springer, 2024.

[12] H. Guan, P. T. Yap, A. Bozoki, and M. Liu, "Federated learning for medical image analysis: A survey," *Pattern Recognition*, vol. 151, Art. no. 110424, Jul. 2024.

[13] GBD 2017 Congenital Heart Disease Collaborators, "Global, regional, and national burden of congenital heart disease, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017," *The Lancet Child & Adolescent Health*, vol. 4, no. 3, pp. 185–200, Mar. 2020.

[14] J. K. Triedman and J. W. Newburger, "Trends in Congenital Heart Disease: The Next Decade," *Circulation*, vol. 133, no. 25, pp. 2716–2733, Jun. 2016.

[15] A. J. Marelli, R. Ionescu-Ittu, A. S. Mackie, L. Guo, N. Dendukuri, and M. Kaouache, "Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010," *Circulation*, vol. 130, no. 9, pp. 749–756, Aug. 2014.

[16] H. W. Goo, S. Siripornpitak, S. J. Chen, O. Lilyasari, Y. M. Zhong, H. A. Latiff, E. Maeda, Y. J. Kim, I. C. Tsai, and D. M. Seo, "Pediatric Cardiothoracic CT Guideline Provided by the Asian Society of Cardiovascular Imaging Congenital Heart Disease Study Group: Part 2. Contemporary Clinical Applications," *Korean Journal of Radiology*, vol. 22, no. 8, pp. 1397–1415, Aug. 2021.

[17] V. Muthurangu, "Cardiovascular Magnetic Resonance in Congenital Heart Disease: Focus on Heart Failure," *Heart Failure Clinics*, vol. 17, no. 1, pp. 157–165, 2021.

[18] J. Schulz-Menger *et al*., "Standardized image interpretation and post-processing in cardiovascular magnetic resonance—2020 update," *Journal of Cardiovascular Magnetic Resonance*, vol. 22, no. 1, Art. no. 19, Mar. 2020.

[19] H. Baumgartner *et al*., "2020 ESC Guidelines for the management of adult congenital heart disease," *European Heart Journal*, vol. 42, no. 6, pp. 563–645, Feb. 2021.

[20] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum, "Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier," *Medical Image Analysis*, vol. 51, pp. 46–60, 2019.

[21] I. Valverde *et al*., "Three-dimensional printed models for surgical planning of complex congenital heart defects: An international multicentre study," *European Journal of Cardio-Thoracic Surgery*, vol. 52, no. 6, pp. 1139–1148, Dec. 2017.

[22] P. D. Morris *et al*., "Computational fluid dynamics modelling in cardiovascular medicine," *Heart*, vol. 102, no. 1, pp. 18–28, Jan. 2016.

[23] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool,” *BMC Medical Imaging*, vol. 15, no. 1, p. 29, Aug. 2015.

[24] D. Karimi and S. E. Salcudean, “Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 2, pp. 499–513, Feb. 2020.

[25] L. Zange, F. Muehlberg, E. Blaszczyk, S. Schwenke, J. Traber, S. Funk, and J. Schulz-Menger, “Quantification in cardiovascular magnetic resonance: Agreement of software from three different vendors on assessment of left ventricular function, 2D flow and parametric mapping,” *Journal of Cardiovascular Magnetic Resonance*, vol. 21, no. 1, Art. no. 12, 2019.

[26] M. Beer, B. Schönnagel, J. Herrmann, S. Klömpken, M. Schaal, M. Kaestner, C. Apitz, and H. Brunner, “Non-invasive pediatric cardiac imaging—current status and further perspectives,” *Molecular and Cellular Pediatrics*, vol. 9, no. 1, p. 21, Dec. 2022.

[27] C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, “Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2494–2505, Jul. 2020.

[28] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning dense volumetric segmentation from sparse annotation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016)*, LNCS, vol. 9901, pp. 424–432, 2016.

[29] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, “H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[30] Y. Xie, J. Zhang, C. Shen, and Y. Xia, “CoTr: Efficiently bridging CNN and transformer for 3D medical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2021)*, LNCS, vol. 12903, pp. 171–180, 2021.

[31] H.-Y. Zhou, J. Guo, Y. Zhang, X. Han, L. Yu, L. Wang, and Y. Yu, “nnFormer: Volumetric medical image segmentation via a 3D transformer,” *IEEE Transactions on Image Processing*, vol. 32, pp. 4036–4045, 2023.

[32] M. Livne, J. Rieger, O. U. Aydin, A. A. Taha, E. M. Akay, T. Kossen, J. Sobesky, J. D. Kelleher, K. Hildebrand, D. Frey, and V. I. Madai, “A U-Net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease,” *Frontiers in Neuroscience*, vol. 13, Art. no. 97, Feb. 2019.

[33] M. J. Sheller, B. Edwards, G. A. Reina, *et al*., “Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data,” *Scientific Reports*, vol. 10, no. 1, Art. no. 12598, Jul. 2020.

[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, Feb. 2020.

[35] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Medical Image Analysis*, vol. 79, Art. no. 102470, Jun. 2022.

[36] X. Zhuang *et al*., "Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge," *Medical Image Analysis*, vol. 58, Art. no. 101537, 2019.

[37] A. Kandathil and M. Chamarthy, "Pulmonary vascular anatomy and anatomical variants," *Cardiovascular Diagnosis and Therapy*, vol. 8, no. 3, pp. 201–207, Jun. 2018.

[38] H. Xu, H. Zhao, J. Jin, J. Geng, C. Sun, D. Wang, N. Hong, F. Yang, and X. Chen, "An atlas of anatomical variants of subsegmental pulmonary arteries and recognition error analysis," *Frontiers in Oncology*, vol. 13, Art. no. 1127138, Mar. 2023.

[39] S. Shit, J. C. Paetzold, A. K. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. W. Pluim, U. Bauer, and B. H. Menze, "clDice: A topology-preserving loss function for tubular structure segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 16555–16564.

[40] K. Pushparajah, P. Duong, S. Mathur, and S. Babu-Narayan, "Educational series in congenital heart disease: Cardiovascular MRI and CT in congenital heart disease," *Echo Research and Practice*, vol. 6, no. 4, pp. R121–R138, 2019.

[41] B. K. Han, C. K. Rigsby, A. Hlavacek, *et al*., "Computed tomography imaging in patients with congenital heart disease Part I: Rationale and utility," *Journal of Cardiovascular Computed Tomography*, vol. 9, no. 6, pp. 475–492, 2015.

[42] S. Moscatelli, M. Avesani, N. Borrelli, *et al*., "Complete transposition of the great arteries in the pediatric field: A multimodality imaging approach," *Children*, vol. 11, no. 6, Art. no. 626, Jun. 2024.

[43] T. Leiner, J. Bogaert, M. G. Friedrich, R. Mohiaddin, V. Muthurangu, S. Myerson, A. J. Powell, S. V. Raman, and D. J. Pennell, "SCMR position paper (2020) on clinical indications for cardiovascular magnetic resonance," *Journal of Cardiovascular Magnetic Resonance*, vol. 22, no. 1, Art. no. 76, 2020.

[44] L. Houyel and S. M. Meilhac, "Heart development and congenital structural heart defects," *Annual Review of Genomics and Human Genetics*, vol. 22, pp. 257–284, 2021.

[45] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," *Medical Image Analysis*, vol. 30, pp. 108–119, 2016.

[46] W. Bai *et al*., "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, no. 1, Art. no. 65, 2018.

[47] M. Iannotta, F. A. d'Aiello, A. Van De Bruaene, *et al*., "Modern tools in congenital heart disease imaging and procedure planning: A European survey," *Journal of Cardiovascular Medicine*, vol. 25, no. 1, pp. 76–87, Jan. 2024.

[48] J. Cohen, P. Asrani, S. Lee, D. Frush, B. K. Han, A. Chelliah, and K. M. Farooqi, "Cardiovascular computed tomography in pediatric congenital heart

disease: A state-of-the-art review," *Journal of Cardiovascular Computed Tomography*, vol. 16, no. 6, pp. 467–482, 2022.

[49] S. Fratz *et al*., "Guidelines and protocols for cardiovascular magnetic resonance in children and adults with congenital heart disease," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, no. 1, p. 51, 2013.

[50] C. Sorensen, P. Gach, H. Pico, N. Hugues, A. Dabadie, C. Desvignes, B. Bourlière, A. Aschero, N. Colavolpe, P. Petit, and G. Gorincour, "Cardiac CT or MRI in pediatric practice: Which one to choose?," *Diagnostic and Interventional Imaging*, vol. 97, no. 5, pp. 513–517, 2016.

[51] V. M. Campello, P. Gkontra, C. Izquierdo, C. Martin-Isla, A. Sojoudi, P. M. Full, K. Maier-Hein, Y. Zhang, Z. He, J. Ma, *et al*., "Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3543–3554, 2021.

[52] A. N. Bhuva, W. Bai, C. Lau, R. H. Davies, Y. Ye, H. Bulluck, E. McAlindon, V. Culotta, P. P. Swoboda, G. Captur, *et al*., "A multicenter, scan-rescan, human and machine learning CMR study to test generalizability and precision in imaging biomarker analysis," *Circulation: Cardiovascular Imaging*, vol. 12, no. 10, Art. no. e009214, 2019.

[53] B. Zhao, "Understanding sources of variation to improve the reproducibility of radiomics," *Frontiers in Oncology*, vol. 11, Art. no. 633176, 2021.

[54] J. Hoffman, N. Emaminejad, M. Wahi-Anwar, *et al*., "Technical Note: Design and implementation of a high-throughput pipeline for reconstruction and quantitative analysis of CT image data," *Medical Physics*, vol. 46, no. 5, pp. 2310–2322, 2019.

[55] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical image segmentation: A comprehensive review of deep learning-based methods," *Tomography*, vol. 11, no. 5, Art. no. 52, 2025.

[56] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, Aug. 2019.

[57] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vols. 3–4, Art. no. 100004, 2019.

[58] C. S. Perone and J. Cohen-Adad, "Promises and limitations of deep learning for medical image segmentation," *Journal of Medical Artificial Intelligence*, vol. 2, Art. no. 1, 2019.

[59] Y. Xu, R. Quan, W. Xu, Y. Huang, X. Chen, and F. Liu, "Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches," *Bioengineering*, vol. 11, no. 10, Art. no. 1034, 2024.

[60] E. Van Reeth, I. W. K. Tham, C. H. Tan, and C. L. Poh, "Super-resolution in magnetic resonance imaging: A review," *Concepts in Magnetic Resonance Part A*, vol. 40A, no. 6, pp. 306–325, 2012.

[61] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[62] C. Gnasso, D. Pinos, U. J. Schoepf, M. Vecsey-Nagy, G. J. Aquino, N. Fink, E. Zsarnoczay, R. J. Holtackers, J. Stock, P. Suranyi, A. Varga-Szemes, and T. Emrich, "Impact of reconstruction parameters on the accuracy of myocardial extracellular volume quantification on a first-generation, photon-counting detector CT," *European Radiology Experimental*, vol. 8, no. 1, p. 70, 2024.

[63] J. El-Taraboulsi, C. P. Cabrera, C. Roney, and N. Aung, "Deep neural network architectures for cardiac image segmentation," *Artificial Intelligence in the Life Sciences*, vol. 4, Art. no. 100083, 2023.

[64] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, *Lecture Notes in Computer Science*, vol. 11045, Cham, Switzerland: Springer, 2018, pp. 3–11.

[65] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.

[66] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: A survey," *Evolutionary Intelligence*, vol. 15, no. 1, pp. 1–22, 2022.

[67] M. E. Rayed, S. M. S. Islam, S. I. Niha, J. R. Jim, M. M. Kabir, and M. F. Mridha, "Deep learning for medical image segmentation: State-of-the-art advancements and challenges," *Informatics in Medicine Unlocked*, vol. 48, Art. no. 101504, 2024.

[68] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, vol. 2, pp. 315–337, 2000.

[69] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: A survey," *Medical Image Analysis*, vol. 24, no. 1, pp. 205–219, Aug. 2015.

[70] J. Alvén, *Improving Multi-Atlas Segmentation Methods for Medical Images*, M.S. thesis, Lund University, Lund, Sweden, 2017. Available: ProQuest Dissertations & Theses Global, Order No. 27753201. [Online]. Available: https://www.proquest.com/dissertations-theses/improving-multi-atlas-segmentation methods/docview/2399260060/se-2. [Accessed: Mar. 23, 2026].

[71] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–165, Jan. 2004.

[72] A. Kumar and S. Jain, "Deformable models for image segmentation: A critical review of achievements and future challenges," *Computers and Mathematics with Applications*, vol. 119, pp. 288–311, 2022.

[73] X. Shi and C. Li, "Anatomical knowledge based level set segmentation of cardiac ventricles from MRI," *Magnetic Resonance Imaging*, vol. 86, pp. 135–148, 2022.

[74] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.

[75] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: An overview and application in radiology,” *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018.

[76] W. Yao, J. Bai, W. Liao, Y. Chen, M. Liu, and Y. Xie, “From CNN to Transformer: A review of medical image segmentation models,” *Journal of Imaging Informatics in Medicine*, vol. 37, no. 4, pp. 1529–1547, Aug. 2024.

[77] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, “A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises,” *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, May 2021.

[78] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, and B. Gulyás, “3D deep learning on medical images: A review,” *Sensors*, vol. 20, no. 18, Art. no. 5097, 2020.

[79] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, “Medical image segmentation using deep learning: A survey,” *IET Image Processing*, vol. 16, no. 5, pp. 1243–1267, 2022.

[80] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, “Transformers in medical imaging: A survey,” *Medical Image Analysis*, vol. 88, Art. no. 102802, 2023.

[81] H. Xiao, L. Li, Q. Liu, X. Zhu, and Q. Zhang, “Transformers in medical image segmentation: A review,” *Biomedical Signal Processing and Control*, vol. 84, Art. no. 104791, 2023.

[82] H. Cao, Y. Wang, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, “Swin-Unet: Unet-like pure transformer for medical image segmentation,” in *Proceedings of the European Conference on Computer Vision Workshops (ECCV Workshops)*, 2021.

[83] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, “Medical transformer: Gated axial-attention for medical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, *Lecture Notes in Computer Science*, vol. 12901, 2021.

[84] X. Yu, Q. Yang, Y. Zhou, L. Y. Cai, R. Gao, H. H. Lee, T. Li, S. Bao, Z. Xu, T. A. Lasko, R. G. Abramson, Z. Zhang, Y. Huo, B. A. Landman, and Y. Tang, “UNesT: Local spatial representation learning with hierarchical transformer for efficient medical segmentation,” *Medical Image Analysis*, vol. 90, Art. no. 102939, Dec. 2023.

[85] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[86] M. J. J. de Grauw, E. Th. Scholten, E. J. Smit, M. J. C. M. Rutten, M. Prokop, B. van Ginneken, and A. Hering, “The ULS23 challenge: A baseline model and benchmark dataset for 3D universal lesion segmentation in computed tomography,” *Medical Image Analysis*, vol. 102, Art. no. 103525, 2025.

[87] Z. Zhu, Y. Xia, W. Shen, E. K. Fishman, and A. L. Yuille, “A 3D coarse-to-fine framework for volumetric medical image segmentation,” in *Proceedings of the IEEE International Conference on 3D Vision (3DV)*, Verona, Italy, 2018, pp. 682–690.

[88] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori, “An application of cascaded 3D fully convolutional networks for medical image segmentation,” *Computerized Medical Imaging and Graphics*, vol. 66, pp. 90–99, Jun. 2018.

[89] S. Sharobeem, H. Le Breton, F. Lalys, M. Lederlin, C. Lagorce, M. Bedossa, D. Boulmier, G. Leurent, P. Haigron, and V. Auffret, “Validation of a whole heart segmentation from computed tomography imaging using a deep-learning approach,” *Journal of Cardiovascular Translational Research*, vol. 15, no. 2, pp. 427–437, Apr. 2022.

[90] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, “A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes,” *Medical Image Analysis*, vol. 13, no. 6, pp. 819–845, Dec. 2009.

[91] J. Tohka, “Partial volume effect modeling for segmentation and tissue classification of brain magnetic resonance images: A review,” *World Journal of Radiology*, vol. 6, no. 11, pp. 855–864, Nov. 2014.

[92] M. Ciecholewski and M. Kassjański, “Computational methods for liver vessel segmentation in medical imaging: A review,” *Sensors*, vol. 21, no. 6, Art. no. 2027, 2021.

[93] L. Zhou, H. Wu, G. Luo, *et al*., “Deep learning-based 3D cerebrovascular segmentation workflow on bright and black blood sequences magnetic resonance angiography,” *Insights into Imaging*, vol. 15, Art. no. 81, 2024.

[94] A. Paulauskaite-Taraseviciene, J. Siaulys, A. Jankauskas, and G. Jakuskaite, “A robust blood vessel segmentation technique for angiographic images employing multi-scale filtering approach,” *Journal of Clinical Medicine*, vol. 14, no. 2, Art. no. 354, 2025.

[95] G. Tetteh, V. Efremov, N. D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, and B. H. Menze, “DeepVesselNet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes,” *Frontiers in Neuroscience*, vol. 14, Art. no. 592352, 2020.

[96] B. Kocak, M. E. Klontzas, A. Stanzione, A. Meddeb, A. Demircioğlu, C. Bluethgen, K. K. Bressem, L. Ugga, N. Mercaldo, O. Díaz, and R. Cuocolo, “Evaluation metrics in medical imaging AI: Fundamentals, pitfalls, misapplications, and recommendations,” *European Journal of Radiology Artificial Intelligence*, vol. 3, Art. no. 100030, 2025.

[97] D. Müller, I. Soto-Rey, and F. Kramer, “Towards a guideline for evaluation metrics in medical image segmentation,” *BMC Research Notes*, vol. 15, no. 1, Art. no. 210, Jun. 2022.

[98] L. Maier-Hein *et al*., “Metrics reloaded: Pitfalls and recommendations for image analysis validation,” *arXiv preprint arXiv:2206.01653*, Jul. 2022.

[99] O. U. Aydin, A. A. Taha, A. Hilbert, A. A. Khalil, I. Galinovic, J. B. Fiebach, D. Frey, and V. I. Madai, “On the usage of average Hausdorff distance for segmentation performance assessment: Hidden error when used for ranking,” *European Radiology Experimental*, vol. 5, no. 1, Art. no. 4, 2021.

[100] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, “A guide to deep learning in healthcare,” *Nature Medicine*, vol. 25, pp. 24–29, Jan. 2019.

[101] H. Guan and M. Liu, “Domain adaptation for medical image analysis: A survey,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022.

[102] J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodriguez, N. V. Chawla, and F. Herrera, “A unifying view on dataset shift in classification,” *Pattern Recognition*, vol. 45, no. 1, pp. 521–530, 2012.

[103] M. H. U. Rehman, W. H. L. Pinaya, P. Nachev, J. T. Teo, S. Ourselin, and M. J. Cardoso, “Federated learning for medical imaging radiology,” *The British Journal of Radiology*, vol. 96, no. 1150, Art. no. 20220890, 2023.

[104] S. T. Shah, Z. Ali, M. Waqar, and A. Kim, “Federated learning in public health: A systematic review of decentralized, equitable, and secure disease prevention approaches,” *Healthcare (Basel)*, vol. 13, no. 21, Art. no. 2760, Oct. 2025.

[105] N. Rieke *et al*., “The future of digital health with federated learning,” *npj Digital Medicine*, vol. 3, no. 1, Art. no. 119, 2020.

[106] R. Wang, P. Chaudhari, and C. Davatzikos, “Embracing the disharmony in medical imaging: A simple and effective framework for domain adaptation,” *Medical Image Analysis*, vol. 76, Art. no. 102309, 2022.

[107] M. Long, Y. Cao, Z. Cao, J. Wang, and M. I. Jordan, “Transferable representation learning with deep adaptation networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 3071–3085, Dec. 2019.

[108] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[109] Z. Zhang, Y. Li, and B. S. Shin, “Robust color medical image segmentation on unseen domain by randomized illumination enhancement,” *Computers in Biology and Medicine*, vol. 145, p. 105427, 2022.

[110] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, *PMLR*, vol. 54, pp. 1273–1282, 2017.

[111] S. S. Sandhu, H. T. Gorji, P. Tavakolian, K. Tavakolian, and A. Akhbardeh, “Medical imaging applications of federated learning,” *Diagnostics*, vol. 13, no. 19, Art. no. 3140, 2023.

[112] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, May 2020.

[113] D. F. Pace, H. T. M. Contreras, J. Romanowicz, S. Ghelani, I. Rahaman, Y. Zhang, P. Gao, M. I. Jubair, T. Yeh, P. Golland, T. Geva, S. Ghelani, A. J. Powell, and M. H. Moghari, “HVSMR-2.0: A 3D cardiovascular MR dataset for whole-heart segmentation in congenital heart disease,” *Scientific Data*, vol. 11, no. 1, p. 721, 2024.

A
SUPPLEMENTARY FEDERATED LEARNING, ENSEMBLE, AND EXPLAINABILITY ANALYSES

This appendix presents supplementary analyses supporting the experimental findings reported in Chapter 4. The additional experiments include detailed federated learning analyses under distributed multi-center conditions, ensemble analyses evaluating model fusion strategies, and explainability analyses investigating attribution behavior across segmentation architectures. These analyses complement the main experimental observations, while all terminology, evaluation metrics, and experimental settings were kept consistent with Chapter 4.

A.1 Detailed Federated Learning Analysis

This section provides additional analyses of the proposed framework under federated learning conditions, mainly focusing on the Hybrid V-Net model as the primary methodological contribution of this study. Detailed analyses of baseline architectures were not repeated to avoid redundancy with the main comparative evaluations. All results were obtained using the combined multi-center dataset and evaluated on the held-out test set (n = 8), consistent with the experimental protocol. The analyses further support the robustness and stability trends discussed in Chapter 4 under heterogeneous distributed training conditions. Table A1 summarizes the federated segmentation performance of the global model together with the individual clients.

Table A.1 Comprehensive federated segmentation performance

Setting	Mean FG Dice (macro)	Dice Std	Dice (micro)	IoU (micro)
Global Model	0.8740	0.0340	0.8810	0.7910
Client 0	0.8660	0.0380	0.8730	0.7790
Client 1	0.8610	0.0390	0.8700	0.7760
Client 2	0.8800	0.0350	0.8870	0.8000

The federated learning results showed relatively stable segmentation performance across clients, with limited variation in overall Dice scores. This tendency suggests consistent convergence behavior despite heterogeneous data distributions across centers. Table A2 presents the structure-wise segmentation performance of the global model across anatomical regions and highlights the difference between large and small anatomical structures.

Table A.2 Per-class segmentation performance (global model)

Class	Dice (macro)	Dice Std	IoU (macro)	Dice (micro)	IoU (micro)	HD95 (mm)	ASSD (mm)
Background	0.9860	0.0020	0.9720	0.9860	0.9710	–	–
Aorta	0.9190	0.0190	0.8500	0.9210	0.8530	5.49	0.78
Chambers	0.9130	0.0300	0.8420	0.9170	0.8470	7.76	1.45
Pulmonary Artery	0.7910	0.0790	0.6610	0.8050	0.6730	14.60	2.19

The structure-wise results indicate that larger anatomical regions such as the aorta and cardiac chambers achieved higher segmentation accuracy together with lower boundary errors, whereas pulmonary artery segmentation remained considerably more difficult because of its small diameter and complex branching morphology. Table A3 presents client-wise variability in segmentation performance for the Hybrid V-Net model.

Table A.3 Client-wise variability in segmentation performance (Hybrid V-Net)

Structure	Client	Dice	HD95 (mm)	ASSD (mm)
Aorta	Client 0	0.9180	3.32	0.73
	Client 1	0.9180	6.34	0.81
	Client 2	0.9190	3.99	0.73
Chambers	Client 0	0.9110	8.02	1.50
	Client 1	0.9120	7.72	1.48
	Client 2	0.9100	8.08	1.51
Pulmonary Artery	Client 0	0.7700	14.58	2.35
	Client 1	0.7540	20.12	2.72
	Client 2	0.8110	13.49	2.04

Client-wise analysis further showed that segmentation performance remained relatively consistent across participating centers. Minor variations were observed mainly in pulmonary artery segmentation, whereas larger anatomical structures demonstrated more stable behavior. These observations support the generalization capability of the proposed framework under inter-center variability.

A.2 Additional Ensemble Analysis

To further investigate the effect of model combination, an ensemble experiment was performed using softmax averaging of nnU-Net and Swin-UNet predictions together with test-time augmentation (TTA). Table A4 summarizes ensemble performance under macro- and micro-level evaluation settings. The results also illustrate the discrepancy between these metrics caused by class imbalance and the dominance of large anatomical structures in voxel-wise evaluation.

Table A.4 Ensemble evaluation results (nnU-Net + Swin-UNet)

Metric	Value	Metric	Value
Mean FG Dice (macro)	0.7260	Mean FG Dice (micro)	0.8872
Dice Std	0.0860	Background Dice	0.9860
Aorta Dice (macro)	0.7730	Aorta Dice (micro)	0.9220
Chambers Dice (macro)	0.8870	Chambers Dice (micro)	0.9201
Pulmonary Artery Dice (macro)	0.5180	Pulmonary Artery Dice (micro)	0.8194
HD95 (Aorta)	3.97 mm	ASSD (Aorta)	0.59 mm
HD95 (Chambers)	1.72 mm	ASSD (Chambers)	0.24 mm
HD95 (Pulmonary Artery)	8.04 mm	ASSD (PA)	1.50 mm

Table A5 presents additional boundary-based ensemble evaluation results across anatomical structures.

Table A.5 Ensemble boundary performance (mean per patient)

Structure	Mean Dice ↑	Mean HD95 (mm) ↓
Aorta	0.9201	1.87
Left Ventricle	0.9155	2.32
Pulmonary Artery	0.8028	6.33

The ensemble experiments showed that combining nnU-Net and Swin-UNet predictions did not consistently improve segmentation performance across anatomical structures. Although reasonable performance was maintained for larger structures such as cardiac chambers, segmentation quality remained limited for small and anatomically complex regions, particularly the pulmonary artery. The difference between macro-level and micro-level Dice scores was mainly related to class imbalance and voxel dominance of large structures, suggesting that macro-level metrics provided a more balanced structure-wise evaluation.

These findings suggest that simple softmax-based ensembling alone was insufficient for improving fine vascular segmentation under heterogeneous prediction behavior. In comparison, the proposed Hybrid V-Net achieved stable segmentation performance without additional ensemble complexity, indicating that localization-guided processing and architecture-level design contributed more strongly to robustness than straightforward model fusion strategies.

A.3 Additional Explainability Analyses

To further support the explainability findings presented in Section 4.9.2, additional qualitative analyses were performed. Figure A1 presents cross-model explainability comparisons using Grad-CAM and Integrated Gradients (IG) for the aorta, left ventricle, and pulmonary artery, illustrating attribution behavior across representative anatomical structures and patient cases.

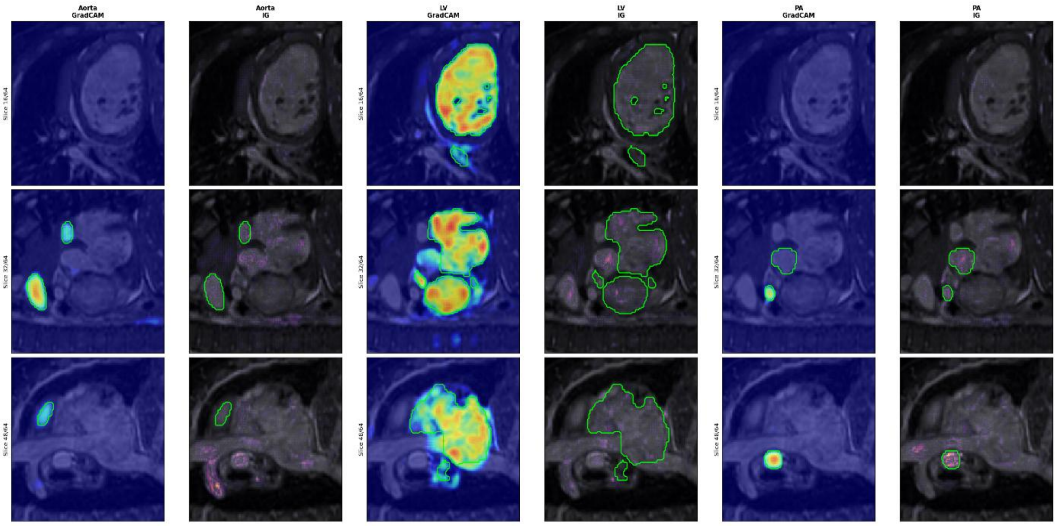


Figure A.1 Cross-model explainability comparison using Grad-CAM (GC) and Integrated Gradients (IG)

Representative attribution maps are shown for nnU-Net, Swin-UNet, and the proposed Hybrid V-Net across the aorta, left ventricle, and pulmonary artery. The visualizations highlight differences in attribution localization, spatial focus, and anatomical consistency across segmentation architectures.

Figure A2 presents a detailed per-patient explainability case study, including input images, ground-truth annotations, and model-specific attribution maps generated for nnU-Net, Swin-UNet, and Hybrid V-Net.

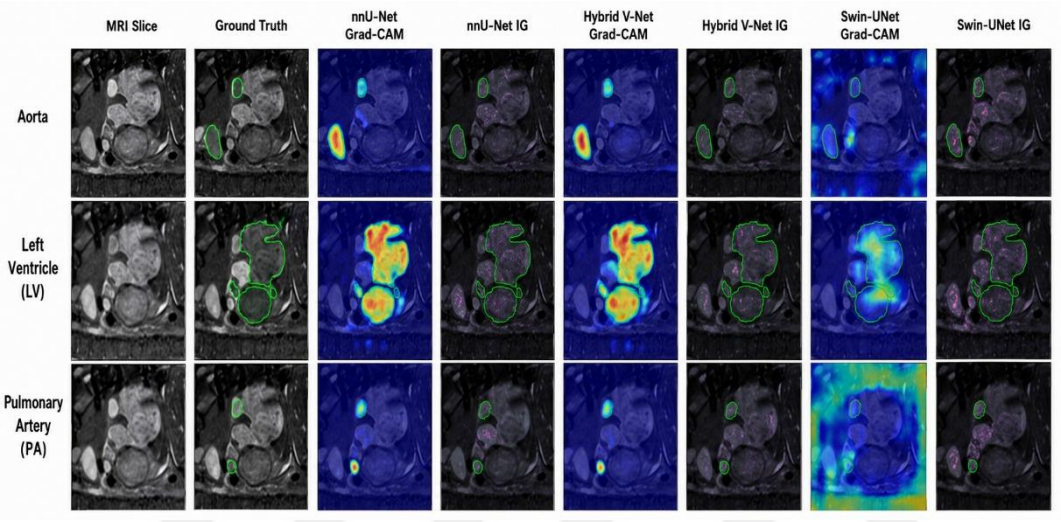


Figure A.2 Per-patient explainability case study across segmentation models

Input images, reference annotations, and corresponding Grad-CAM and Integrated Gradients attribution maps are presented for representative patient cases to illustrate model-specific attention behavior under heterogeneous anatomical conditions.

The explainability analyses showed that Hybrid V-Net and nnU-Net generally produced more spatially focused and anatomically aligned attribution patterns, whereas Swin-UNet exhibited comparatively more diffuse activation behavior, particularly in small and structurally complex regions such as the pulmonary artery. Similar tendencies were observed across representative patient cases.

Taken together, the explainability analyses indicate that higher-performing models tend to generate more localized and anatomically coherent attribution patterns, supporting the interpretability of the proposed framework under heterogeneous multi-center conditions.

PUBLICATIONS FROM THE THESIS

Conference Papers

1. I.Yilmaz, S. Onder, S. Piskin, O. F. Sahin, K. B. Kose, J. S. F. Josephin, M. H. Alzaeim, F. Z. Gungoren, R. Mirza, B. N. Yazici, M. A. K. Umair, F. Aydogdi, and D. Gunveren, "Structure-Aware V-Net Framework for 3D Multi-Class Segmentation of Pediatric Cardiovascular Structures," in *2025 Medical Technologies Congress (TIPTEKNO)*, Gazimagusa, Cyprus, 2025.

2. O. F. Sahin, M. H. Alzaeim, B. N. Yazici, S. Onder, D. Gunveren, S. Piskin, F. Z. Gungoren, R. Mirza, I. Yilmaz, I. Faress, K. B. Kose, J. S. F. Josephin, M. A. K. Umair, and F. Aydogdi, "A Unified Pipeline for Consistent Multi-Label Mask Generation in Pediatric Cardiac Segmentation," in *2025 Medical Technologies Congress (TIPTEKNO)*, Gazimagusa, Cyprus, 2025.

3. I.Yilmaz, O. F. Sahin, M. A. K. Umair, D. Gunveren, S. Onder, J. S. F. Josephin, K. B. Kose, and S. Piskin, "Federated Hybrid V-Net for Automatic Multi-Center Segmentation of Pediatric Cardiovascular Structures Using CT and MRI Data," accepted for presentation at AIMed2026 (MICCAI Society), 2026.

4. S. Piskin, J. S. F. Josephin, K. B. Kose, S. Onder, F. Z. Gungoren, M. H. Alzaeim, O. F. Sahin, I. Yilmaz, M. A. K. Umair, R. Mirza, B. N. Yazici, I. Faress, D. Gunveren, and A. Pezri, "Integrated Centerline and Surface-Based Statistical Shape Analysis of the Aorta in Pediatric CHD," accepted for poster presentation at the 10th World Congress of Biomechanics (WCB 2026), Vancouver, Canada, Jul. 2026.