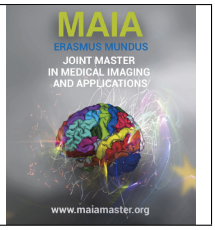
Medical Imaging and Applications

Master Thesis, June 2025



Uncertainty-Guided Active Learning Based Aortic Structure Segmentation for Optimized TAVI Access Route Planning

Mahdi Islam, Agnes Mayr, Christian Kremser, Markus Haltmeier, Enrique Almar-Munoz

Medical University of Innsbruck, Austria

Abstract

Introduction: Transcatheter Aortic Valve Implantation (TAVI) is a minimally invasive procedure for replacing a diseased aortic valve, where accurate preoperative planning is essential to minimize complications during catheter navigation. A critical component of this planning involves determining the optimal vascular access route for catheter insertion. However, due to the thin, small and convoluted structure of iliac arteries, selecting the optimal access route based solely on imaging data remains a significant challenge. Vessel segmentation enables iliac diameter measurement and route optimization but is hindered by the thin and convoluted structure of iliac arteries, making 3D annotation both difficult and time-consuming.

Methods: This study develops an efficient segmentation framework using an active learning pipeline guided by probabilistic uncertainty. The strategy intelligently queries uncertain samples for expert correction while using high-confidence predictions as pseudo-labels to enrich the training data. The resulting segmentations are then input into a complete downstream workflow for automated diameter quantification, which uses skeletonization and graph-based centerline extraction. Finally, we conduct a modality ablation study using both pre- and post-contrast CMR to systematically evaluate the impact of multi-modality on model performance.

Results: The active learning pipeline successfully trained a robust segmentation model over five iterations, achieving a final overall Dice score of 0.912 and a corresponding Mean Absolute Percentage Error (MAPE) of 4.92% for diameter quantification with minimal manual annotation. The ablation study revealed that models trained on post-contrast data significantly outperformed those trained on pre-contrast data, establishing data quality as the most critical factor for performance. The diameter quantification workflow proved effective in generating detailed profiles of the entire aorto-iliac access route.

Conclusion: This thesis successfully demonstrates that an annotation-efficient active learning framework can produce highly accurate segmentation of complex aortic structures from CMR. The work delivers a valuable end-to-end pipeline from image to clinically relevant diameter measurements, providing critical insights into the factors governing automated TAVI planning and underscoring the importance of data quality for robust performance.

Keywords: Active Learning, Aortic Segmentation, Vessel Diameter Quantification, TAVI Planning, Cardiovascular Magnetic Resonance (CMR)

1. Introduction

Aortic Stenosis (AS), a progressive narrowing of the aortic valve opening, represents the most prevalent valvular heart disease in the Western world, affecting approximately 12.4% of individuals over 75 years old, with severe AS present in about 3.4% of this population(Osnabrugge et al., 2013). This condition, often caused by age-related calcification and scarring, restricts blood flow from the heart, yet many elderly patients are considered too high-risk for traditional open-heart surgery. For these patients, Transcatheter Aortic Valve Implantation (TAVI), a minimally invasive procedure, has emerged as the gold-standard treatment(Díez, 2013). Successful TAVI planning relies on precise pre-procedural imaging to assess the aortoiliofemoral access route and size the valve prosthesis, a task convention-

ally performed using contrast-enhanced Computed Tomography Angiography (CTA) (Al-Najafi et al., 2016). A pivotal aspect of this planning involves evaluating the iliofemoral arteries to determine their suitability for catheter insertion. Specifically, arteries with diameters below **5 to 5.5 mm** are associated with a heightened risk of vascular complications, including iliac artery rupture, which can be fatal (Mach et al., 2021)

However, the reliance on CTA presents a significant clinical challenge. A substantial portion of the TAVI patient population suffers from pre-existing renal impairment, with reported prevalence rates of chronic kidney disease (CKD) as high as 70%(Ram et al., 2017). The iodinated contrast agents required for CTA are nephrotoxic and contraindicated in these patients, creating a critical need for contrast-free imaging alternatives. Cardiovascular Magnetic Resonance (CMR) has emerged as a promising radiation-free solution, capable of providing the necessary anatomical and functional details for TAVI planning(Mahon and Mohiaddin, 2021). Unenhanced CMR protocols, combining techniques like Quiescent-Interval Single-Shot MR Angiography (QISS-MRA) with 3D "whole heart" imaging, have shown particular promise for comprehensive, contrast-free TAVI guidancePamminger et al. (2020). A recognized limitation of techniques like Quiescent-Interval Single-Shot MR Angiography (QISS-MRA) is the inability to directly visualize vascular calcifications, as they depict the vessel lumen rather than the vessel wall and its associated plaque burden. Despite this, the clinical viability of a CMR-based approach remains high. Notably, CTA can be susceptible to blooming artifacts in heavily calcified vessels, which may lead to an underestimation of the true luminal diameter. Studies comparing non-contrast MRA with the CTA "gold standard" have found strong correlations in vessel diameter measurements, resulting in comparable decisions regarding transfemoral access capability. This evidence underscores the value of CMR as a robust alternative for TAVI planning, particularly for patients with contraindications to contrast-enhanced imaging (Mayr et al., 2018).

Despite its potential, the development of fully automated pipelines for CMR-based TAVI planning is hampered by several challenges. CMR imaging is susceptible to various artifacts, including motion artifacts due to cardiac and respiratory movements, as well as issues related to contrast variability. These artifacts can compromise image quality and pose significant challenges for automated analysis (Oscanoa et al., 2023). Additionally, the development of robust deep learning models for automated analysis is hindered by the limited availability of large, publicly accessible, and expertly annotated CMR datasets. This annotation bottleneck makes it difficult and time-consuming to train effective models (Tahir et al., 2023).

To address these challenges, this thesis proposes a comprehensive framework for the semi-automated analysis of both pre- and post-contrast CMR images for TAVI planning, with a focus on annotation efficiency. The contributions of this work are fourfold:

1. We introduce an **active learning pipeline** for aortic structure segmentation, designed to achieve high accuracy while significantly reducing the manual annotation effort required by traditional fully supervised methods.
2. The pipeline is guided by a robust **uncertainty estimation** strategy that intelligently queries the most informative samples for expert review and uses high-confidence predictions for pseudo-labeling, thereby optimizing the use of the limited annotation budget.
3. We present a complete workflow that extends beyond segmentation to include automated centerline extraction and a hybrid **diameter calculation strategy**, providing the critical vessel measurements necessary for access route planning.
4. Finally, leveraging a dataset of paired pre- and post-contrast scans, we conduct a **modality ablation study** to systematically evaluate the distinct and combined contributions of each imaging modality to segmentation and diameter measurement accuracy.

2. State of the Art

Accurate delineation of the aortic anatomy is critical for Transcatheter Aortic Valve Implantation (TAVI) to determine optimal prosthesis sizing and catheter access routes. While manual analysis of computed tomography angiography (CTA) has been the standard, this process is both time-consuming and prone to inter-observer variability. Consequently, recent advances in deep learning have focused on developing automated pipelines to streamline TAVI planning.

2.1. Automated Measurement and Planning from CT

The automation of TAVI planning from CT data has seen significant progress. For instance, Santaló-Corcoy introduced *TAVI-PREP*, a deep learning pipeline that extracts 22 key anatomical measurements from CT scans, demonstrating a high correlation with expert-derived values. Similarly, a fully automated system for localizing landmarks and quantifying aortic root morphology from CT was proposed by Krüger et al. (Krüger, 2023), showcasing the potential to substantially improve clinical workflow efficiency. Another research group introduced 4TAVR, a fully automated AI-based software that performs complete CT-based TAVR planning—including segmentation, landmark detection, and valve sizing—with accuracy comparable to expert clinicians (Blankenberg et al., 2024). These works highlight the maturity of deep learning applications for CT-based TAVI planning.

2.2. Aorta and Iliac Segmentation

Central to any planning pipeline is the accurate segmentation of the aorta and iliac arteries. Deep learning models like DeepLabv3+ have proven effective for this task in both CT and CMR contexts. One study successfully applied such models for real-time aortic valve detection from cardiac CT, indicating the feasibility of automated segmentation (Inomata, 2025). However, a persistent challenge remains in segmenting smaller, more intricate structures like the iliac arteries, particularly from CMR data where image quality or contrast may be limited compared to CTA. Recent architectures like CIS-UNet further improve multi-class segmentation of the aorta and its branches by integrating convolutional and transformer-based modules, enabling accurate delineation of complex vascular structures in CTA volumes (Imran et al., 2024).

2.3. Limitations of Fully Supervised Approaches

A primary limitation of the aforementioned deep learning methods is their reliance on large, expertly annotated datasets for training. While CT datasets are relatively common, high-quality annotated CMR datasets suitable for TAVI planning are scarce. This "annotation bottleneck" is a major barrier to developing robust, automated CMR-based tools and motivates the exploration of more data-efficient learning strategies. One such approach is semi-supervised learning (SSL), where models learn from a small labeled set and a larger unlabeled set. For instance, a quality-aware SSL method was proposed for CMR segmentation, which leverages downstream task performance to identify high-quality predictions for use as pseudo-labels in subsequent training (Ruijsink et al., 2020). Other approaches include recent foundation models like MedSAM2, which demonstrate strong generalization capabilities for segmenting anatomical structures and lesions across diverse imaging modalities (Ma et al., 2025). However, the segmentation performance of these foundation models on small and intricate structures, such as the iliac arteries, remains uncertain.

2.4. Active Learning for Annotation Efficiency

To more directly address the scarcity of labeled data, active learning (AL) has emerged as a particularly suitable strategy. AL aims to minimize annotation cost by intelligently querying a human expert to label only the most informative samples. Traditional uncertainty-based AL methods often use metrics like entropy or Monte Carlo dropout to identify these samples Li and Alstrøm (2020). However, these can be affected by model overconfidence. To mitigate this, more advanced frameworks have been proposed. For example, *PAAL* is a predictive-accuracy-based framework that combines an Accuracy Predictor with a Weighted Polling Strategy to select samples that are both informative and diverse, reducing annotation costs by up to 80% Shi (2024).

2.5. Hybrid and Semi-supervised Active Learning Strategies

The current frontier in data-efficient learning often involves blending AL with semi-supervised techniques. *TAAL*, for example, is a hybrid approach that uses test-time augmentation and entropy-based uncertainty to generate pseudo-labels from unlabeled data, combining them with actively queried samples Gaillochet (2023). Other methods focus on refining the selection strategy itself; *SUBAL* is a strategy that selects uncertain *pixels* within specific target regions rather than entire images, thereby improving class balance and performance in medical segmentation tasks Ma (2024). These methods demonstrate a clear trend toward combining pseudo-labeling with uncertainty-aware sample selection to achieve high performance with limited supervision.

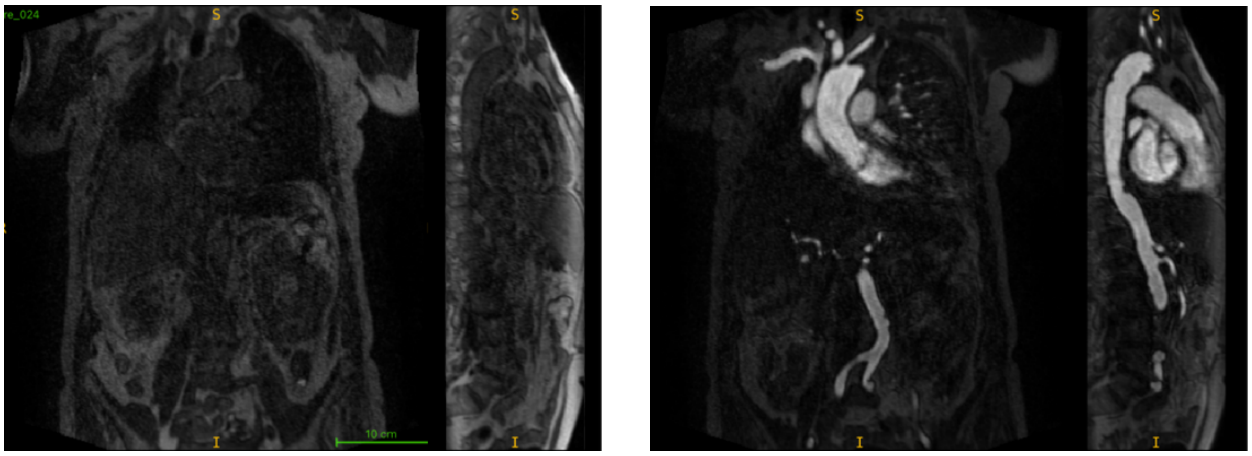
In summary, the literature reveals a strong and evolving trend toward developing automated cardiovascular planning tools that are both accurate and annotation-efficient. While CT-based pipelines have matured considerably, CMR-based solutions are less developed due to challenges in data quality, annotation availability, and the anatomical complexity of the access route. The current study builds upon these state-of-the-art advances by developing an active learning pipeline tailored for robust and efficient segmentation of aortic structures from CMR, aiming to bridge the gap in automated, contrast-free TAVI planning.

3. Dataset

3.1. Dataset Acquisition

The dataset utilized in this study was acquired at the Medical University of Innsbruck, Austria, and comprises 337 cardiovascular magnetic resonance (CMR) scans in DICOM format. This collection includes 166 pre-contrast (non-contrast enhanced) and 171 post-contrast (gadolinium-enhanced) CMR examinations. All scans were performed as part of comprehensive pre-TAVI (Transcatheter Aortic Valve Implantation) assessments.

The pre-contrast CMR scans were acquired using the non-contrast MRI protocol for TAVI guidance. These examinations were predominantly performed on 1.5-T MRI systems. Key acquisition characteristics included ECG-gating for sequences like QISS-MRA or navigator-gating for 3D whole heart acquisitions, with imaging typically synchronized to a quiescent phase of the cardiac cycle to minimize motion artifacts. The anatomical coverage for access route assessment, crucial for TAVI planning, generally extended from the supra-aortic branches to the femoral arteries. The post-contrast enhanced CMR scans involved established protocols for gadolinium-enhanced vascular imaging, complementing the non-contrast assessments. Representative samples of the pre contrast and post contrast CMR



(a) Pre-contrast CMR

(b) Post-contrast CMR

Figure 1: Coronal and sagittal views of representative samples from the dataset. Subfigure (a) displays a pre-contrast CMR scan, while subfigure (b) shows the corresponding post-contrast scan of the same patient.

of a patient are shown in Figure 1(a) and Figure 1(b) respectively.

3.2. Data Preprocessing

The initially acquired CMR scans were first converted from the original DICOM format to NIfTI format. To ensure consistency in spatial orientation, all images were then reoriented to the RAS (Right-Anterior-Superior) coordinate system. Following this, to ensure a consistent voxel size across all images, the CMR scans were resampled to a voxel spacing of $1 \times 1 \times 1$ mm³ using trilinear interpolation.

4. Methodology

This section details the methods employed in this research, encompassing Vessel Diameter Quantification, Active Learning Implementation, and the Multimodal Ablation Study. Furthermore, it describes the quantitative criteria used for evaluating the experiments.

4.1. Vessel Diameter Quantification

The target of this experiment was to develop a robust pipeline that accurately quantifies the diameters of aortic structures if their segmentation masks are given. The methodology for quantifying vessel diameters from 3D medical image segmentations involves several sequential processing steps, from preprocessed data loading to centerline extraction and diameter measurement. All 3D spatial data and coordinates are consistently processed using (Z, Y, X) array indexing, corresponding to depth, height, and width respectively, resampled to 1mm x 1mm x 1mm spacing, following an initial conversion of the images to a standardized anatomical orientation, such as RAS (Right-Anterior-Superior) as discussed in section 3.2. The subsequent subsections describe the quantification steps in detail. The full pipeline is illustrated in Figure 3.

4.1.1. Centerline Extraction

Let I_{RAS} be the 3D image data in RAS orientation and its corresponding vessel segmentation masks (e.g. aorta, iliac artery) be M_{RAS}. This binary mask M_{RAS} (in XYZ order) is also converted into ZYX order, M_{ZYX}, for subsequent processing. The core of the vessel's path, its centerline, is extracted from the binary mask M_{ZYX}.

- **Morphological Skeletonization**: The 3D binary mask M_{ZYX} is thinned to a 1-voxel wide 3D skeleton S_{ZYX} using `skeletonization` function of the 'skimage.morphology' library.

$$S_{ZYX} = \text{skeletonize}(M_{ZYX})$$

- **Largest Connected Component**: To focus on the primary vessel segment and remove smaller and irrelevant branches or noise, connected component analysis is performed on S_{ZYX}. Only the largest connected component (LCC), S_{LCC}, is retained.

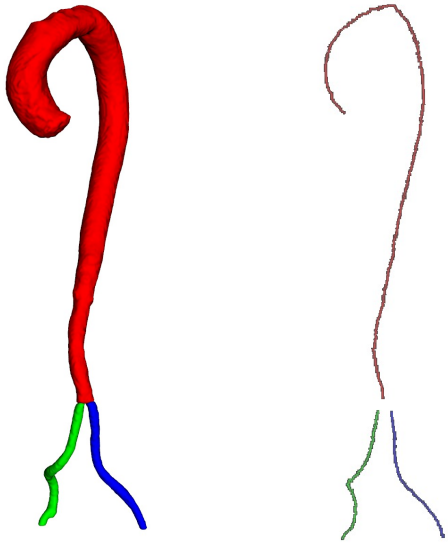
- **Graph Representation and Path Finding**: The skeleton S_{LCC} is converted into a graph representation where skeleton voxels are nodes and adjacent voxels form edges, weighted by Euclidean distance considering voxel spacing S. The `skan` library's `skeleton_to_csgraph` function is utilized for this conversion, yielding a graph $G = (V, E)$ and corresponding ZYX coordinates for each node. Endpoints $V_{EP} \subset V$ are identified as nodes with a degree of 1. To identify the principal centerline of the vessel segment, a two-step process is used:

1. First, for every distinct pair of endpoints $(v_i, v_j) \in V_{EP} \times V_{EP}$ in the graph G, the shortest path connecting them is calculated using Dijkstra's algorithm (Dijkstra, 1959). This results in a set of shortest paths, each representing the most direct route between a specific pair of endpoints.
2. Second, from this set of all calculated shortest paths, the path that has the greatest length is selected.

This selected path—the longest among all the direct routes between any two endpoints—is then defined as the principal centerline of the vessel segment.

- **Centerline Ordering**: The sequence of ZYX coordinates $C = \{(z_k, y_k, x_k)\}_{k=1}^{N_p}$ representing the N_p points along the extracted longest path is then ordered. The path is oriented such that the Z-coordinate of the first point is less than or equal to the Z-coordinate of the last point ($z_1 \leq z_{N_p}$), ensuring a consistent inferior-to-superior (or equivalent based on Z-axis definition) progression along the centerline.

The resulting ordered centerline coordinates C are saved as a NIfTI file for visualization.



(a) 3D Mask (b) Extracted Centerlines

Figure 2: Illustration of centerline extraction. (a) Shows the 3D mask of aortic structures. (b) Displays the extracted and refined centerlines from the mask.

4.1.2. Centerline Refinement

1. **Smoothing**: The raw centerline C can exhibit minor step-wise artifacts due to voxel discretizations. To mitigate this, the Y and X coordinates of the centerline points are smoothed using a 1D uniform filter with a defined window size (e.g., 5 points). The Z-coordinates, representing the primary axis of progression, are typically not smoothed.

$$y'_k = \text{uniform\_filter1d}(y_k, \text{win})$$

$$x'_k = \text{uniform\_filter1d}(x_k, \text{win})$$

Sudden jumps in smoothed Y or X coordinates exceeding a threshold are further corrected by local averaging. The smoothed centerline is $C' = \{(z_k, y'_k, x'_k)\}_{k=1}^{N_p}$.
2. **Endpoint Correction**: For a specified fraction of points at the beginning of the centerline C' (e.g., the initial 5%), their Y and X coordinates can be adjusted. For each selected centerline point (z_k, y'_k, x'_k), the Center of Mass (CoM) of the vessel mask M_{ZYX} in the YX-plane at that specific Z-slice z_k is computed. Let $M_{ZYX}(z_k, \cdot, \cdot)$ be the 2D slice. The CoM $(x_{com}(z_k), y_{com}(z_k))$ is found for the LCC in this slice. The Y and X coordinates of the centerline point are then updated: $y''_k = y_{com}(z_k)$ and $x''_k = x_{com}(z_k)$. This step aims to ensure the centerline starts robustly within the vessel lumen at its ends. The refined centerline is C''. The output of centerline extraction and refinement from a representative mask is illustrated in Figure 2

4.1.3. Diameter Calculation

In the final step, for each point $P_k = (z_k, y_k, x_k)$ on the refined centerline C'', the local vessel diameter is estimated using a hybrid approach.

1. **Distance Transform (DT) Diameter**: A 3D Euclidean Distance Transform (EDT) is computed on the binary mask M_{ZYX}, considering voxel spacing S. For each centerline point P_k, the EDT value $EDT(P_k)$ gives the shortest distance to the vessel boundary. An initial diameter estimate is $D_{DT,k} = 2 \cdot EDT(P_k)$.
2. **Ray-Casting Diameter**:
   - **Local Tangent**: A local tangent vector $\vec{T}_k$ at P_k is computed using finite differences between neighboring centerline points.
   - **Orthogonal Plane Definition**: A plane orthogonal to $\vec{T}_k$ is defined. If $\vec{T}_k$ is nearly parallel to the Z-axis (i.e., $|\vec{T}_k \cdot \hat{z}| > \cos\theta_{thresh}$), the YX-plane at z_k is used. Otherwise, two orthogonal vectors $\vec{U}_k, \vec{V}_k$ spanning the plane perpendicular to $\vec{T}_k$ are computed (e.g., using Frisvad's method).
   - **Ray Casting**: A set of N_{rays} (e.g., 8) rays are cast from P_k within this defined plane. Each ray direction $\vec{d}_{ki}$ is a linear combination of $\vec{U}_k, \vec{V}_k$ or lies in the YX plane.

- **Boundary Detection**: For each ray, originating from P_k and extending in directions $\pm\vec{d}_{ki}$, the points where the ray first exits the vessel mask M_{ZYX} are found, denoted B^+_{ki} and B^-_{ki}.
- **Ray Diameter**: The diameter along this ray is $D_{ray,ki} = \|(B^+_{ki} - B^-_{ki}) \odot S\|_2$, where $\odot$ is element-wise multiplication with voxel spacing S.

3. **Combined Diameter**: The "raw" diameter $D_{raw,k}$ at point P_k is calculated with a combination of the DT-derived diameter and the median of valid ray-casting diameters:

$$D_{raw,k} = \frac{D_{DT,k} + \text{median}(\{D_{ray,ki}\})}{2}$$

If insufficient valid ray diameters are found, then $D_{raw,k} = D_{DT,k}$ is considered.

4. **Post-processing**: Lastly, the sequence of raw diameters $\{D_{raw,k}\}$ is processed by linearly interpolating any missing values. Subsequently, a Savitzky-Golay filter (e.g., polynomial order 3, window length up to 51 points, adapted for shorter sequences) is applied to obtain the final smoothed diameter profile $\{D_{final,k}\}$.

The resulting ZYX coordinates and corresponding final diameters are stored for analysis and visualization.

4.2. Active Learning Implementation

This study employed active learning strategies with the primary goal of developing a robust aortic structure segmentation pipeline while minimizing manual annotation effort. The underlying principle of active learning is that models benefit most from training on samples they find most challenging, i.e., those for which they are most uncertain. Consequently, iterative training with the selective inclusion of new, informative samples aims to accelerate performance improvement and achieve high accuracy with fewer labeled examples (Li et al., 2024)

4.2.1. Dataset Preparation

For the active learning experiment, we exclusively used post-contrast Cardiovascular Magnetic Resonance (CMR) images, as these offer enhanced visibility and contrast of the cardiac and aortic regions. The dataset comprised 171 post-contrast CMR scans. These scans were randomly shuffled and subsequently divided into training and test sets. From these, 145 samples were allocated to the training set and 26 to a hold-out test set. The training set was further partitioned into two subsets: an iterative training pool and an iterative residual pool. Initially, the training pool contained 15 randomly selected samples from the training set. The other training samples were placed in the residual pool for subsequent active learning iterations.

4.2.2. Supervised Segmentation

The supervised segmentation iterations were done using the nnU-Net framework, renowned for its state-of-the-art performance in various medical image segmentation tasks (Isensee et al., 2021). The training configuration employed in each iteration is summarized in Table 1. No manual modifications were made to the nnU-Net hyperparameters or its inherent data augmentation strategies. A 5-fold cross-validation approach was applied during each training phase to ensure robust model generalization and effective utilization of the available data.

Table 1: nnU-Net Training Configurations.

Trainer	ResEnc M
Total epochs	300
Optimizer	SGD
Initial learning rate (lr)	1e − 2
Momentum	0.99
Weight decay	3e − 5
Lr decay schedule	PolyLR
Loss function	Dice + Cross-Entropy
Deep supervision	True
Iterations per epoch (train)	250
Iterations per epoch (val)	50
Oversample foreground (%)	0.33
Probabilistic oversampling	False
Gradient clipping norm	12

4.2.3. Iterative Training Pipeline

The initial training iteration started with 15 samples from the training pool. Following this, the model performed inference on all samples in both the residual and test pools. For each subsequent iteration, 15 new samples were strategically drawn from the residual pool and added to the training pool. Based on rankings derived from the model's uncertainty scores (detailed in Section 4.2.4), 12 **most certain** samples were selected, their model-generated segmentations serving directly as pseudo-labels for the next training round. The 3 **most uncertain** samples were chosen for the remaining spots; their predicted segmentations underwent manual expert correction before inclusion in the training pool. This iterative cycle of training, inference, uncertainty-based selection, and selective re-annotation was repeated for five iterations, after which performance gains plateaued. The complete iterative active learning pipeline is illustrated in Figure 4

4.2.4. Uncertainty Estimation

To guide the sample selection process in our active learning pipeline, we implemented a robust uncertainty estimation strategy based on the probabilistic outputs of the nnU-Net model. For each unlabeled case in the

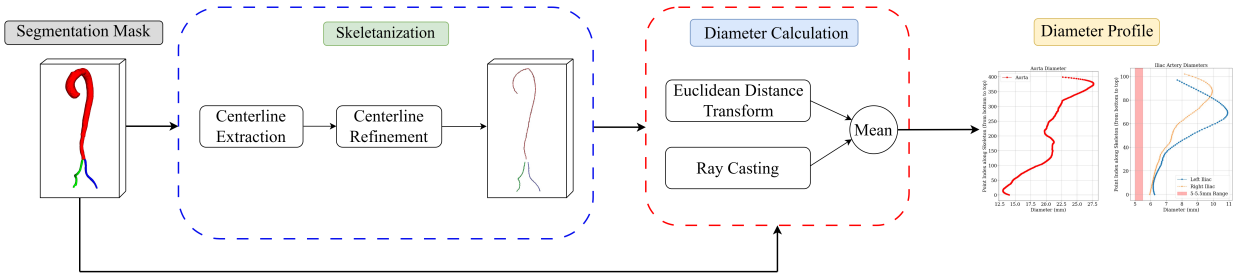


Figure 3: Overview of the automated vessel diameter quantification pipeline. The flowchart illustrates the sequence of operations from initial segmentation mask processing, through centerline extraction and refinement, to the final diameter calculation along the derived vessel centerline.

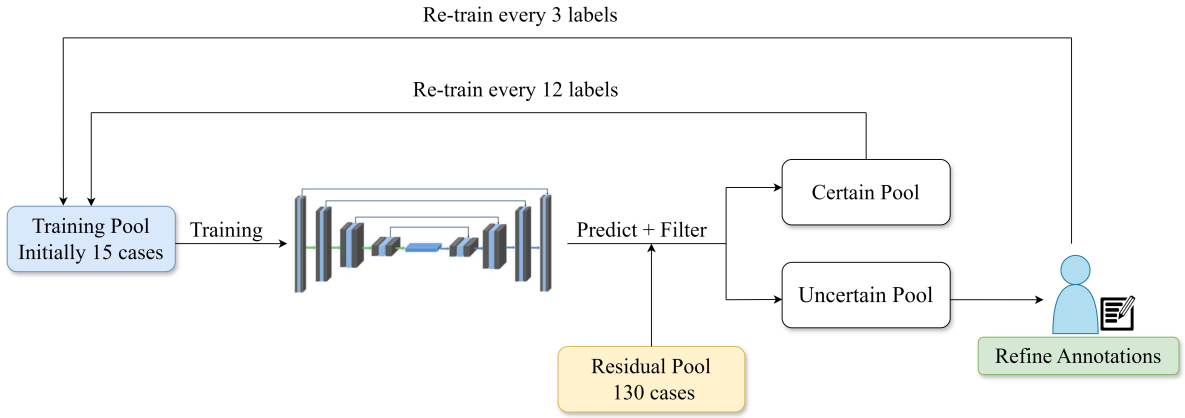


Figure 4: Overview of the iterative active learning (AL) pipeline designed for annotation-efficient segmentation. The cycle consists of training a model on the iterative pool, performing inference on the residual pool, and using an uncertainty estimation module to rank the predictions. In each iteration, new samples are selected: the most uncertain cases are chosen for manual annotation refinement, while the most certain cases are added with their model-generated pseudo-labels. This process progressively increases the iterative pool to improve model performance with minimal manual effort.

residual pool, the model generated probability maps for the foreground classes: Aorta, Left Iliac artery, and Right Iliac artery. These probabilities were used for calculating model uncertainty.

For each voxel j within a given class c prediction, we first considered only those voxels with a predicted foreground probability $p_j > 0.01$ to focus on regions with at least minimal evidence of belonging to the target structure. Several metrics were then computed for each class c in each case:

1. **Mean Voxel Probability**: The average of the predicted probabilities for class c across all relevant voxels:

$$\bar{P}_c = \frac{1}{N_c} \sum_{j=1}^{N_c} p_j$$

where N_c is the number of voxels considered for class c.

2. **Standard Deviation of Voxel Probabilities**: This measures the dispersion or variability of probability values for class c:

$$\sigma_{P_c} = \sqrt{\frac{1}{N_c} \sum_{j=1}^{N_c} (p_j - \bar{P}_c)^2}$$

3. **Mean Voxel Entropy**: Entropy is a common measure of uncertainty. For each voxel j with probability p_j for a given class, the binary entropy H_j is calculated as:

$$H_j(p_j) = -p_j \ln(p_j) - (1 - p_j) \ln(1 - p_j)$$

A small epsilon (1×10^{-6}) was used to clip probabilities to avoid issues with ln(0). The Mean Entropy for class c, $\bar{H}_c$, is the average of these voxel-wise entropies over all N_c considered voxels. Higher entropy values indicate greater uncertainty.

4. **Fraction of Uncertain Voxels**: The proportion of voxels whose predicted probability p_j for class c falls within a predefined "uncertain" range, specif-

ically $[0.4, 0.6]$:

$$F_{\text{uncertain},c} = \frac{1}{N_c} \sum_{j=1}^{N_c} \mathbb{I}(0.4 \leq p_j \leq 0.6)$$

where $\mathbb{I}(\cdot)$ is the indicator function.

5. **Fraction of Confident Voxels**: The proportion of voxels with $p_j > 0.8$ for class c.
6. **Fraction of Very Confident Voxels**: The proportion of voxels with $p_j > 0.9$ for class c. This metric, $F_{\text{very\_confident},c}$, contributes to identifying reliably segmented regions.

To derive a holistic uncertainty measure for case selection, particularly focusing on the often more challenging iliac arteries, a composite **Case Uncertainty Score** (S_{case}) was formulated. For each iliac artery class $c \in \{\text{Left Iliac, Right Iliac}\}$ in a case, an intermediate uncertainty score was calculated:

$$S_c = \bar{H}_c + \left(1 - \frac{F_{\text{very\_confident},c}}{100}\right)$$

This score combines the average entropy (direct uncertainty) with a penalty for not having a high fraction of very confidently predicted voxels (indirect uncertainty), assuming $F_{\text{very\_confident},c}$ is expressed as a percentage value (0-100). The final S_{case} for a given patient scan was then determined by averaging these scores across the iliac artery classes present and evaluated for that case:

$$S_{\text{case}} = \text{mean}(S_{\text{Left Iliac}}, S_{\text{Right Iliac}})$$

Cases with higher S_{case} values were deemed more uncertain, while those with lower scores were considered more certain. This ranking allowed for the selection of the 3 **most uncertain** and 12 **most certain** samples for each subsequent active learning iteration mentioned in 4.2.3.

4.3. Multimodal Ablation Study

We designed a comprehensive ablation study to systematically evaluate how different imaging modalities and training data compositions affect segmentation and diameter measurement performance. This study assesses models trained on unimodal (pre- or post-contrast CMR) and multimodal (combined pre- and post-contrast CMR) data.

4.3.1. Dataset Preparation for Ablation Study

For this component of the research, a specific subset of the main dataset was utilized, comprising only patients with paired pre- and post-contrast CMR scans. This initial selection yielded 161 image pairs. A subsequent filtering step was applied to ensure matching image dimensions across pairs, resulting in a final dataset of 160 pairs. Crucially, these image pairs were pre-registered, which guaranteed perfect spatial alignment and prevented the need for an explicit registration step within our pipeline.

The dataset for this ablation study was partitioned into training and test sets, all derived from the paired pre- and post-contrast CMR scans. The data used to construct the various training configurations consisted of two label types sourced from the preceding active learning stage:

- **Strong Labels:** A set of **70 paired cases** with high-quality, manually corrected segmentation masks obtained from the final pool of the active learning iterations.
- **Pseudo (Weak) Labels:** A set of **66 paired cases** with model-generated pseudo-labels, produced by running inference with the final active learning model on the remaining unlabeled cases from the residual pool.

A separate, held-out **Test Set** of **24 paired cases**, consistent with the one used to evaluate the final active learning pipeline, was used for the final performance assessment of all models.

4.3.2. Training and Ablation Configurations

Three distinct deep learning architectures were evaluated across all configurations to ensure that our findings were not model-dependent: **Mednext** (Roy et al., 2024), **3D nnU-Net**, and **UMamba** (Ma et al., 2024). All models were trained for 100 epochs across two primary experimental scenarios: unimodal and multimodal training.

Unimodal Configurations:. In this setting, models were trained using only a single modality at a time. Four distinct experiments were conducted to assess the value of each modality and the impact of adding pseudo-labeled data:

1. **Pre-contrast (Strong):** Models trained using only pre-contrast CMR images with the 70 strong labels.
2. **Pre-contrast (Strong + Pseudo):** Models trained using pre-contrast images with all available labels (70 strong + 66 pseudo).
3. **Post-contrast (Strong):** Models trained using only post-contrast CMR images with the 70 strong labels.
4. **Post-contrast (Strong + Pseudo):** Models trained using post-contrast images with all available labels (70 strong + 66 pseudo).

Multimodal Configuration:. To investigate the benefit of combining information from both modalities, a multimodal training strategy was implemented. This approach treated the pre- and post-contrast images from a single patient as two independent samples for a single-channel model, effectively doubling the number of

training instances (e.g., 70 pre-contrast images and 70 post-contrast images for the strong-only set). This strategy was evaluated under two conditions: (i) using only the 70 strong labels, and (ii) using the full set of 70 strong and 66 pseudo-labels.

These two multimodal configurations, along with the four unimodal experiments, created a total of six distinct setups for a thorough analysis of how data modality and composition affect segmentation and downstream quantification performance.

4.3.3. Evaluation Protocol

Following the training of each model, performance was assessed on a held-out test set using evaluation protocols tailored to the model's training configuration (unimodal or multimodal).

Unimodal Model Evaluation. The models trained under the unimodal configurations were evaluated exclusively on their corresponding modality. For instance, a model trained on pre-contrast data was tested only on the pre-contrast images from the test set, and likewise for the post-contrast models. This approach was designed to directly measure the performance ceiling for each individual modality.

Multimodal Model Evaluation. For the multimodally trained models, a more challenging evaluation was conducted to assess their robustness in a mixed-data environment. All images from the test set, comprising both pre- and post-contrast scans, were shuffled and presented to the trained model in a single, randomized inference session. After predictions were generated for all images, the results were segregated based on their original modality (i.e., grouped into pre-contrast and post-contrast results).

4.4. Quantitative Assessment Metrics

To rigorously evaluate the performance of our methods, quantitative metrics were employed to assess both the accuracy of vessel segmentations and the precision of derived diameter measurements.

4.4.1. Volumetric Segmentation Accuracy

The accuracy of 3D segmentations for the Aorta, Left Iliac, and Right Iliac arteries was evaluated using the Dice Similarity Coefficient (DSC). The DSC measures the volumetric overlap between an automated segmentation and a ground truth segmentation. This evaluation criteria was used in Sections 5.2.1 and 5.3

4.4.2. Centerline Diameter Measurement Accuracy

The accuracy of automatically computed vessel diameter profiles was benchmarked against ground truth profiles. To ensure consistent comparisons, diameter profiles were preprocessed: aortic profiles were typically resampled to a fixed number of points (e.g., 200), while iliac artery profiles were either analyzed at their native lengths or the predicted profile was resampled to match the ground truth length for certain metrics. The evaluation involved the following metrics:

- **Mean Absolute Error (MAE):** The average absolute difference between corresponding ground truth (D_{GT}) and predicted (D_{Pred}) diameter points:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|D_{GT,i} - D_{Pred,i}|$$

- **Mean Absolute Percentage Error (MAPE):** A relative error measure, calculated as:

$$\text{MAPE} = \frac{100\%}{N_{\text{valid}}}\sum_{i\in\text{valid}}\left|\frac{D_{GT,i} - D_{Pred,i}}{\max(D_{GT,i}, \epsilon)}\right|$$

  where ϵ (1×10^{-6}) prevents division by zero. For cases with missing or entirely invalid predicted diameter data, MAPE was set to 100 % to reflect a total error in averages.

- **Pearson Correlation Coefficient (r):** Assessed the linear correlation between ground truth and predicted diameter profiles. A value of 0 was assigned for cases with missing/invalid predictions.

- **Dynamic Time Warping (DTW) Distance:** Measured profile similarity, allowing for non-linear alignments, thus robust to minor shifts along the centerline.

- **Hausdorff Distance (HD):** Calculated on 2D point sets (normalized centerline position vs. diameter) to quantify the maximum discrepancy between profile shapes.

These metrics were computed per-case and per-structure, with results aggregated to evaluate overall diameter quantification performance.

5. Results

5.1. Vessel Diameter Profile

Diameter profiles were obtained for both ground truth and predicted segmentation masks using the Vessel Diameter Quantification pipeline. The ground truth diameter profiles are used to gain insights to for TAVI access route planning. Whereas the prediction diameter profiles were evaluated based on the criteria described in Section 4.4.2. Figure 5 shows the generated diameter profiles of the aorta and iliac arteries from a representative case obtained from it's ground truth mask.

Generally, the aortic diameter profiles (an example shown in Figure 5a) exhibited a characteristic tapering from the proximal (superior) to the distal (inferior) end.

The measured diameters typically ranged from approximately 20-30 mm proximally in the descending thoracic or upper abdominal aorta, tapering to approximately 15-25 mm more distally towards the bifurcation, with anatomical variations visible along the vessel length.

For the iliac arteries (an example shown in Figure 5b), distinct profiles were generated for the left and right branches. These profiles generally showed diameters within the expected physiological range for common and external iliac arteries, typically between 6-12 mm. The plots also allowed for the visualization of reference thresholds, such as the 5 mm limit indicated in the example. This threshold is clinically significant, as minimum diameter requirements (often in the range of 5mm to 5.5mm) are crucial for selecting a safe vascular access route for Transcatheter Aortic Valve Implantation (TAVI) delivery systems (Mach et al., 2021); diameters below such thresholds may indicate stenosis that could complicate or preclude catheter passage. The y-axis in these plots consistently represents the point index along the extracted vessel skeleton, ordered from the proximal to the distal end.

These visual outputs confirm the pipeline's ability to generate detailed diameter profiles along the primary trajectory of the target vessels across the dataset. Details of the quantitative comparison between ground truth and predicted profiles are in Section 5.2.2 and 5.3.

5.2. Active Learning Performance Evaluation

The efficacy of the implemented active learning strategy was evaluated by tracking key performance metrics over five iterations. Figure 6 illustrates the progression of both volumetric segmentation accuracy and diameter measurement precision.

5.2.1. Volumetric Segmentation Accuracy

The evolution of volumetric segmentation accuracy, assessed by the Dice Similarity Coefficient (DSC), is presented in Figure 6a. The results demonstrate a consistent improvement in segmentation performance with successive active learning iterations for all evaluated structures.

The aorta segmentation exhibited high accuracy from the outset, with DSC improving from approximately 0.926 at the first iteration to a peak of around 0.976 by the fifth iteration. The iliac arteries, which typically present greater segmentation challenges, showed more substantial gains. The average DSC for iliac arteries commenced at approximately 0.715 and rose significantly to about 0.878 after five iterations. Consequently, the overall mean DSC for all foreground classes increased from roughly 0.789 to 0.912, underscoring the benefit of the active learning approach in enhancing segmentation robustness. The most pronounced improvements were observed in the initial two to three iterations, with performance gains gradually plateauing in later stages.

5.2.2. Centerline Diameter Measurement Accuracy

The accuracy of centerline diameter measurements also benefited from the iterative training process. Figure 6b shows the reduction in Mean Absolute Percentage Error (MAPE) over the five active learning iterations, where lower MAPE values indicate higher accuracy.

Similar to the segmentation accuracy, diameter measurement precision improved for all structures. The MAPE for the aorta started at approximately 6.20% and decreased substantially to about 1.59% by the final iteration. The iliac arteries, presenting more complex geometries for diameter estimation, began with a higher average MAPE of around 18.37%, which was reduced to approximately 8.26% after five iterations. The overall MAPE across structures improved from about 12.29% to 4.92%. These trends indicate that the active learning strategy not only refined the segmentation boundaries but also contributed to more accurate downstream geometric measurements.

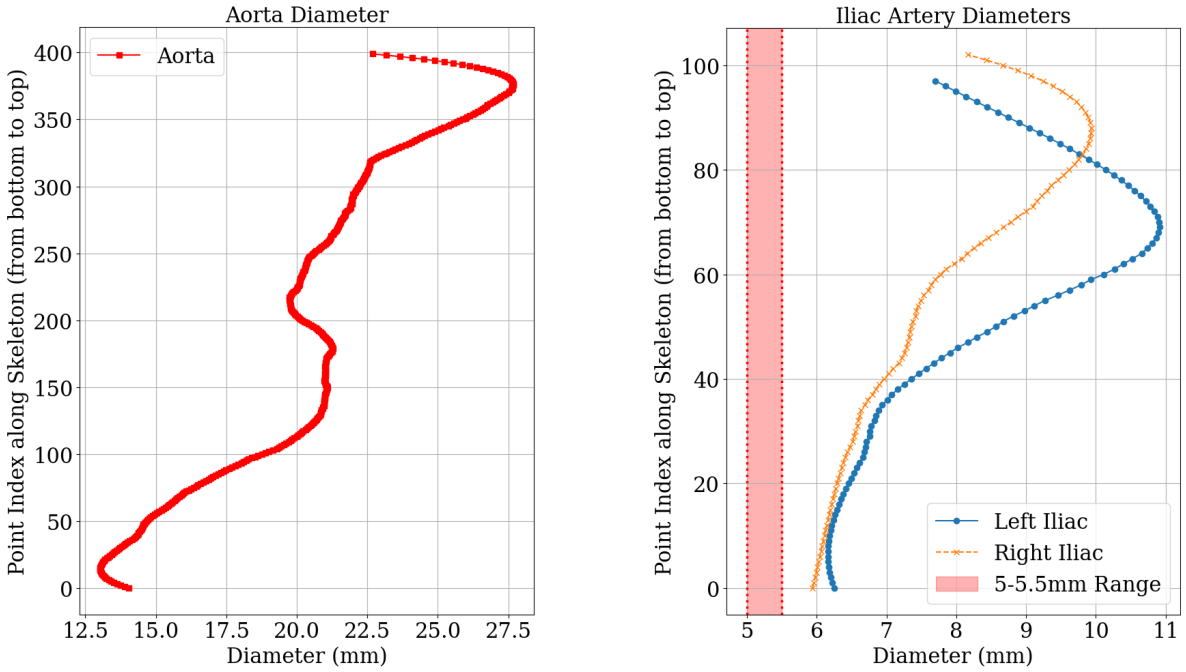
Qualitative examples of diameter profile comparisons from a representative test case, presumably from the final active learning iteration, are presented in Figure 7. These plots compare the ground truth diameter profiles with those derived from the predicted segmentations.

For the aorta (Figure 7a), the predicted diameter profile closely follows the ground truth, exhibiting excellent agreement with a reported MAPE of 1.88% and a Pearson correlation of 1.00 for this case. The left iliac artery (Figure 7b) also shows good concordance, with a MAPE of 2.19% and correlation of 0.93, though some minor deviations are visible, particularly in the distal segment. The right iliac artery (Figure 7c) demonstrates strong agreement as well, achieving a MAPE of 1.90% and a correlation of 0.99. The length ratios for all three structures are close to 1.0, indicating accurate centerline length recovery. These examples visually corroborate the quantitative improvements achieved through active learning.

5.3. Multimodal Ablation

The results of the multimodal ablation study, designed to assess the impact of imaging modality and pseudo-labeling, are presented in Table 2 for segmentation accuracy (DSC) and Table 3 for diameter quantification accuracy (MAPE).

Impact of Imaging Modality. A clear performance disparity was observed between models trained on post-contrast versus pre-contrast data. For segmentation, the highest average Dice score achieved with post-contrast data was **0.8871** (3D nnU-Net, Strong), significantly outperforming the best average score of 0.7667 achieved with pre-contrast data (UMamba, Strong + Pseudo). This trend was mirrored in diameter quantification accuracy. The best-performing post-contrast



(a) Aorta Diameter Profile

(b) Iliac Artery Diameter Profiles

Figure 5: Diameter profiles of a representative sample. (a) Shows the diameter measurements along the aorta. (b) Illustrates the diameter measurements for the left and right iliac arteries, with the 5-5.5mm reference threshold range indicated.

model achieved an average MAPE of **5.56%** (3D nnU-Net, Post Strong + Pseudo), whereas the best pre-contrast model only reached an average MAPE of 11.24% (Mednext, Pre Strong + Pseudo).

Effect of Pseudo-Labeling. The addition of pseudo-labeled data to the training set yielded varied results. For models trained on pre-contrast data, including pseudo-labels consistently led to modest improvements in average Dice scores across all architectures. However, for the high-performing post-contrast models, the impact was negligible or slightly negative; the best overall Dice score (0.8871) was achieved without pseudo-labels. A similar pattern was observed for MAPE, where the benefit of pseudo-labels was more apparent in the pre-contrast configurations.

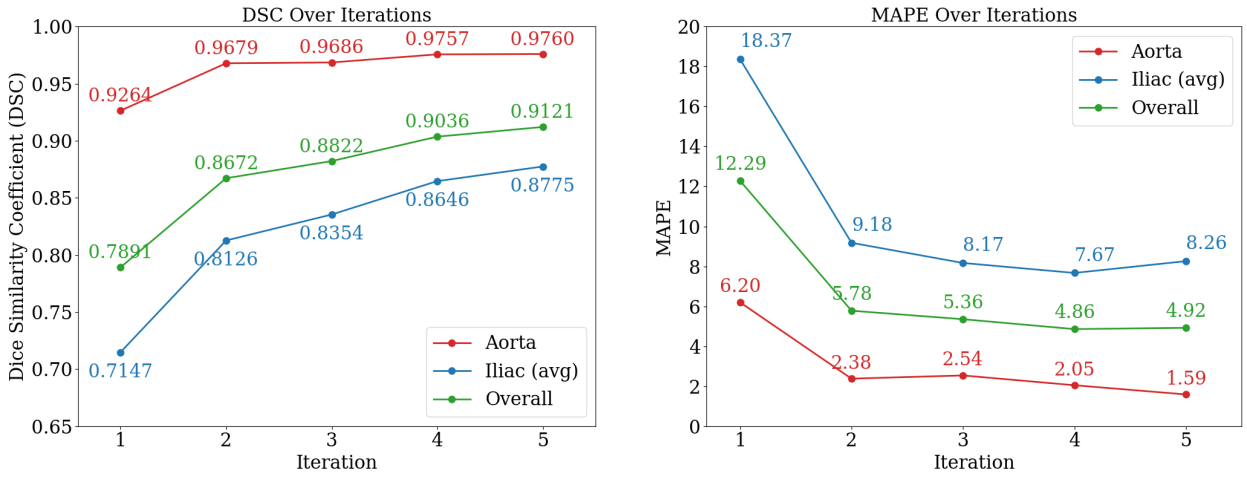
Model Architecture Comparison. No single architecture was universally superior across all metrics and configurations. The 3D nnU-Net model achieved the highest overall Dice score in the post-contrast setting. For MAPE, UMamba demonstrated the best performance for several individual vessel measurements, while the 3D nnU-Net yielded the lowest overall average MAPE. The performance differences between the models, however, were generally smaller than the performance gap between the pre- and post-contrast data configurations.

6. Discussion

6.1. Active Learning Strategy

To enhance annotation efficiency for aortic structure segmentation in this research, we implemented an uncertainty-based active learning strategy. Our iterative pipeline centered on strategically selecting new samples to maximize model performance gains with limited manual intervention. In each of the five active learning cycles, 15 samples were chosen from a residual pool: the 12 samples for which the segmentation model exhibited the highest certainty had their model-generated predictions directly incorporated as pseudo-labels. Conversely, the 3 samples deemed most uncertain by the model underwent manual expert correction before being added to the training set. These settings were fixed empirically focusing on decreasing manual annotation as much as possible while maintaining the quality of the iterative pool. This dual approach was designed to leverage the model's strong predictions while focusing expert annotation efforts on the most challenging and informative cases. The uncertainty quantification itself was based on the model's output probabilities, utilizing a composite score to rank samples, as detailed in Section 4.2.4.

During the development of this pipeline, alternative uncertainty estimation methods were considered, including a Monte Carlo (MC) ensemble technique. This



(a) (b)

Figure 6: Performance metrics across five active learning iterations. (a) Progression of Dice Similarity Coefficient (DSC) for Aorta, average Iliac arteries, and Overall segmentation. (b) Progression of Mean Absolute Percentage Error (MAPE) for Aorta, average Iliac arteries, and Overall diameter measurements.

Table 2: DSC scores for Aorta and Iliac segmentation across different configurations. For each metric, the single best result across all configurations is highlighted in **bold**. Results within 0.005 of the best score are underlined.

Configuration	Model	Aorta	Left Iliac	Right Iliac	Average
Pre Strong	Mednext	0.8779	0.6263	0.6741	0.7308
	3D nnUNET	0.8756	0.6141	0.6601	0.7215
	UMamba	0.8779	0.6613	0.6908	0.7475
Pre Strong + Pseudo	Mednext	0.8744	0.6892	0.7120	0.7621
	3D nnUNET	0.8763	0.6312	0.6998	0.7401
	UMamba	0.8806	0.6967	0.7118	0.7667
Post Strong	Mednext	0.9543	0.8122	0.8187	0.8645
	3D nnUNET	0.9664	0.8250	**0.8627**	**0.8871**
	UMamba	**0.9678**	0.8111	0.8321	0.8732
Post Strong + Pseudo	Mednext	0.9559	**0.8276**	0.8507	0.8804
	3D nnUNET	0.9626	0.8107	0.8507	0.8773
	UMamba	0.9617	0.8272	0.8260	0.8743

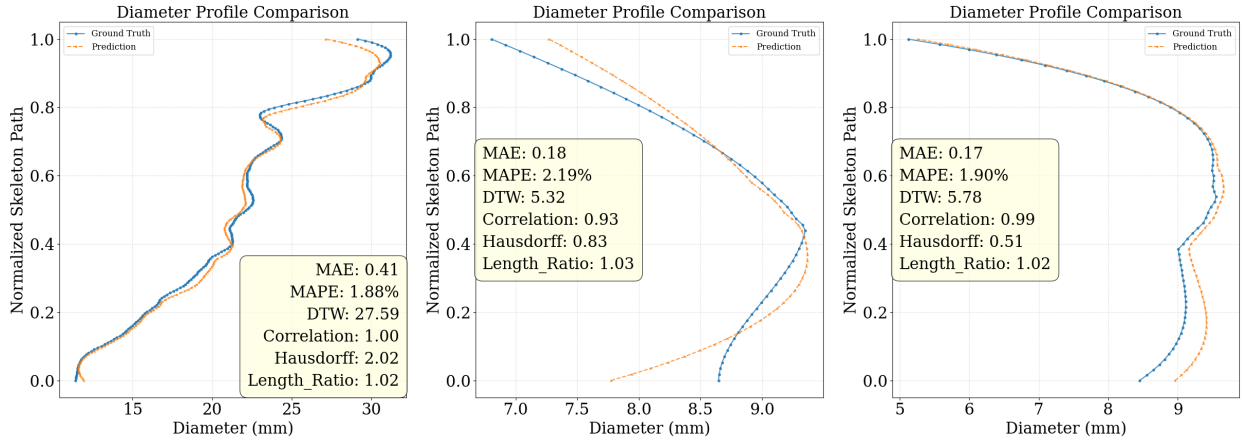
approach involved generating multiple variants of input images through minor perturbations, performing inference on each variant, and then calculating uncertainty from the variance in their predictions. While the MC ensemble method provided comparable uncertainty scores to our adopted single-model, probability-based estimation, it was found to be substantially more computationally intensive and time-consuming. Given the similar effectiveness in identifying uncertain samples but the significant advantage in resource efficiency, the direct probability-based uncertainty estimation was chosen for our active learning framework. The iterative process was halted after five cycles as performance metrics began to plateau.

6.2. Ablation Study Insights

The results from our ablation study provide several key insights into developing automated pipelines for TAVI planning.

The pronounced performance gap between models trained on post-contrast and pre-contrast data underscores the critical role of image quality. The superior vessel-to-background contrast in post-contrast CMR provides clearer anatomical boundaries, which directly translates to more accurate segmentations and, consequently, more precise downstream diameter measurements. This suggests that for achieving the highest accuracy in automated TAVI planning tools, post-contrast CMR data is strongly preferred.

The nuanced effect of pseudo-labeling highlights its situational utility. The consistent, albeit small, benefit for the weaker pre-contrast models indicates its value



(a) Aorta (b) Left Iliac Artery (c) Right Iliac Artery

Figure 7: Diameter profile comparisons between ground truth (blue, solid line with circles) and prediction (orange, dashed line with crosses) for a sample from the test set, likely after the final active learning iteration. Each sub-figure includes quantitative comparison metrics (MAE, MAPE, DTW, Correlation, Hausdorff, Length Ratio) in the yellow text box.

for bootstrapping performance in low-signal or data-poor regimes. Conversely, its limited impact on the already high-performing post-contrast models suggests a potential ceiling effect. In this scenario, pseudo-labels generated by an already accurate model may not provide sufficient new information for the model to learn from, or may even introduce subtle noise that slightly degrades performance specially for iliac structures.

Finally, while the evaluated deep learning architectures showed competitive performance, the variations between them were less significant than the impact of the input modality. This implies that for this specific clinical application, efforts to improve performance should prioritize the acquisition and use of high-quality data (i.e., post-contrast CMR) over minor optimizations between different state-of-the-art model architectures. These findings collectively guide future development by emphasizing data quality as the most dominant factor for success in this domain.

6.3. Limitations

Several limitations of this study should be acknowledged:

- **Generalizability:** The training and evaluation data were sourced from a single institution. This may limit the generalizability of our findings to different patient populations, scanner vendors, or imaging protocols.
- **Active Learning Strategy:** Our findings indicated that the benefit of pseudo-labeling was situational, offering negligible improvement for the already high-performing post-contrast models. Furthermore, an exhaustive comparison of different active learning strategies and their parameters was beyond the scope of this work due to project constraints.
- **Performance on Pre-contrast Data:** Despite training with multimodal data, the models' performance on pre-contrast CMR, for both segmentation and diameter quantification, remained substantially inferior to their performance on post-contrast data.
- **Error Propagation:** The entire quantification pipeline is sequential. As such, the accuracy of the final diameter measurements is fundamentally dependent on the quality of the upstream segmentation, meaning any segmentation inaccuracies directly propagate to the downstream task.

7. Future Work

Potential directions for future research emerging from this work include:

- **Advanced Active Learning:** Investigating more active learning strategies, such as diverse uncertainty metrics, adaptive query methods, or hybrid sample selection criteria.
- **Multimodal Integration:** Developing semi-supervised multimodal pipelines for improved integration of pre- and post-contrast CMR data, po-

Table 3: Mean Absolute Percentage Error (MAPE) for diameter quantification. For each metric, the single best result (lowest error) across all configurations is in **bold**. Results within 0.5% of the best score are underlined.

Configuration	Model	Aorta	Left Iliac	Right Iliac	Average
Pre Strong	Mednext	5.54	17.07	17.26	11.35
	3D nnUNET	6.27	20.29	25.65	14.62
	UMamba	5.78	16.96	20.33	12.21
Pre Strong + Pseudo	Mednext	5.43	15.60	18.49	11.24
	3D nnUNET	6.33	20.95	19.60	13.30
	UMamba	5.66	15.99	18.81	11.53
Post Strong	Mednext	3.01	8.90	15.20	7.53
	3D nnUNET	2.26	8.58	9.89	5.75
	UMamba	**2.08**	8.75	12.40	6.33
Post Strong + Pseudo	Mednext	2.61	8.25	10.14	5.90
	3D nnUNET	2.47	8.26	**9.03**	**5.56**
	UMamba	2.23	**7.46**	10.40	5.58

Table 4: DSC scores of multimodally trained models. For each metric, the single best result (highest score) across all configurations is in **bold**. Results within 0.005 of the best score are underlined.

		Pre-contrast Test Set			Post-contrast Test Set		
Configuration	Model	Aorta	L-Iliac	R-Iliac	Aorta	L-Iliac	R-Iliac
Strong Labels	Mednext	0.8668	0.6549	0.6680	0.9680	0.8039	0.8251
	nnUNET	0.8539	0.5907	0.6298	0.9667	0.8167	0.8132
	Umamba	0.8588	0.6348	0.6572	**0.9684**	**0.8222**	**0.8404**
Strong + Pseudo	Mednext	**0.8640**	0.5935	0.6479	0.9650	0.7848	0.7810
	nnUNET	0.8607	0.6002	**0.6543**	0.9677	0.7991	0.7954
	Umamba	0.8588	**0.6259**	0.6385	0.9660	0.8152	0.7944

tentially leveraging cross-modal learning or image synthesis.

- **Optimized Geometric Quantification:** Refining diameter measurement techniques to be less dependent on upstream segmentation accuracy, exploring joint optimization, or enhancing centerline refinement algorithms.
- **Enhanced Generalizability:** Assessing and improving model robustness through rigorous evaluation on multi-center, diverse datasets and investigating domain adaptation techniques.
- **Clinical Validation and Impact:** Pursuing prospective clinical validation of the developed tools within TAVI planning workflows and conducting longitudinal studies to correlate image-derived metrics with patient outcomes.

8. Conclusion

This thesis addressed the significant challenge of developing automated tools for TAVI access route planning from Cardiovascular Magnetic Resonance (CMR) imaging, where the scarcity of annotated data hinders the application of standard deep learning methods. To overcome this annotation bottleneck, we successfully developed and evaluated a comprehensive framework centered on an annotation-efficient active learning pipeline for aortic structure segmentation, complemented by an end-to-end workflow for subsequent vessel diameter quantification. The active learning strategy effectively utilized uncertainty estimation to guide a hybrid selection process, incorporating both expert-corrected annotations for uncertain cases and pseudo-labels for high-confidence predictions.

Our experimental results demonstrated the efficacy of the active learning approach, which iteratively improved both segmentation and diameter quantification accuracy. The subsequent ablation study yielded several critical insights. Most notably, it quantified the profound impact of data quality, revealing that models trained on post-contrast CMR data significantly outperform those trained on pre-contrast data across all evaluated architectures. Furthermore, our findings highlighted the situational utility of pseudo-labeling, which provided a modest benefit for weaker models trained on pre-contrast data but had a negligible effect on the already high-performing post-contrast models. These results suggest that while data augmentation techniques are valuable, the choice of imaging modality and the

Table 5: Mean Absolute Percentage Error (MAPE, %) for diameter quantification. For each metric, the single best result (lowest error) across all configurations is in **bold**. Results within 0.5% of the best score are underlined.

		Pre-contrast Test Set			Post-contrast Test Set		
Configuration	**Model**	**Aorta**	**L-Iliac**	**R-Iliac**	**Aorta**	**L-Iliac**	**R-Iliac**
Strong Labels	Mednext	**5.94**	18.69	19.19	2.30	11.06	12.30
	nnUNET	6.90	20.14	25.88	2.39	10.31	12.87
	Umamba	6.51	**18.67**	**24.19**	**2.22**	8.91	11.89
Strong + Pseudo	Mednext	6.54	23.43	24.07	2.39	8.77	11.32
	nnUNET	7.50	22.86	23.78	2.37	**8.23**	11.49
	Umamba	7.30	21.98	20.50	2.42	8.65	**11.25**

quality of each CMR are the most dominant factors for achieving high accuracy.

In conclusion, this work successfully demonstrates a viable active learning pipeline that reduces the manual annotation burden required for developing accurate segmentation models in CMR-based TAVI planning. A key contribution of this research is the integration of these segmentations into a complete, automated workflow that provides clinically relevant vessel diameter measurements. This downstream quantification component has potential significance, as it delivers the precise vessel dimensions necessary to assess the safety of the TAVI access route and helps clinicians identify potential contraindications, such as severe stenosis, which could lead to vascular complications. The study offers clear evidence that prioritizing high-quality, post-contrast data is critical for achieving this level of success. While acknowledging the limitations of a single-center dataset and the need for further optimization of the active learning strategy, this research contributes a valuable end-to-end framework and provides key insights that can guide the future development of more efficient and robust tools for cardiovascular diagnostics and interventional planning.

Acknowledgments

I would like to extend my deepest gratitude to my supervisors, for the I extend my sincere gratitude to the Tyrol Klinik and the Medical University of Innsbruck for graciously hosting this research and providing the critical computational infrastructure necessary for its execution. I would also like to acknowledge the MAIA master consortium and the European Commission, whose support provided the foundation for my master's education and this unique opportunity. To my family and friends, thank you for being my constant emotional support; your encouragement has been invaluable, bridging any physical distance with strength and love.

References

Al-Najafi, S., Sanchez, F., Lerakis, S., 2016. The crucial role of cardiac imaging in transcatheter aortic valve replacement (TAVR): Pre- and post-procedural assessment. Curr. Treat. Options Cardiovasc. Med. 18, 70.

Blankenberg, S., Stenzinger, A., Kiefer, P., Braun, D., Wehrum, T., Michel, C., Gigliotti, M.J., Kloeckner, R., Schnabel, J.A., Maintz, D., Isfort, P., 2024. A fully automated artificial intelligence-driven software for planning of transcatheter aortic valve replacement. JACC: Cardiovascular Imaging doi:10.1016/j.jcmg.2024.03.004.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numerische mathematik 1, 269–271.

Díez, 2013. Transcatheter aortic valve implantation (tavi): the hype and the hope. Texas Heart Institute journal 40, 298–301.

Gaillochet, M.e.a., 2023. Taal: Test-time augmentation for active learning in medical image segmentation. arXiv preprint arXiv:2301.06624 .

Imran, M., Krebs, J.R., Gopu, V.R.R., Fazzone, B., Sivaramand, V.B., Kumar, A., Viscardi, C., Heithaus, R.E., Shickel, B., Zhou, Y., Cooper, M.A., Shao, W., 2024. Cis-unet: Multi-class segmentation of the aorta in computed tomography angiography via context-aware shifted window self-attention. arXiv preprint arXiv:2401.13049 URL: https://arxiv.org/abs/2401.13049.

Inomata, Y.e.a., 2025. Automatic aortic valve extraction using deep learning with contrast-enhanced cardiac ct images. JCDD 12, 3.

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods 18, 203–211.

Krüger, A.e.a., 2023. A ct-based deep learning system for automatic assessment of aortic root morphology for tavi planning. European Radiology .

Li, B., Alstrøm, T.S., 2020. On uncertainty estimation in active learning for image segmentation. arXiv preprint arXiv:2007.06364 .

Li, D., Wang, Z., Chen, Y., Jiang, R., Ding, W., Okumura, M., 2024. A survey on deep active learning: Recent advances and new frontiers. IEEE Transactions on Neural Networks and Learning Systems .

Ma, J., Li, F., Wang, B., 2024. U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 .

Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B., 2025. Medsam2: Segment anything in 3d medical images and videos. arXiv preprint arXiv:2504.03600 .

Ma, Y.e.a., 2024. Breaking the barrier: Selective uncertainty-based active learning for medical image segmentation. arXiv preprint arXiv:2401.16298 .

Mach, M., Okutucu, S., Kerbel, T., Arjomand, A., Fatihoglu, S.G., Werner, P., Simon, P., Andreas, M., 2021. Vascular complications in tavr: Incidence, clinical impact, and management. Journal of Clinical Medicine 10. URL: https://www.mdpi.com/2077-0383/10/21/5046, doi:10.3390/jcm10215046.

Mahon, C., Mohiaddin, R., 2021. The emerging applications of cardiovascular magnetic resonance imaging in transcatheter aortic valve implantation. Clinical Radiology 76, 73.e21–73.e37. doi:https://doi.org/10.1016/j.crad.2019.11.011.

Mayr, A., Klug, G., Reinstadler, S.J., Feistritzer, H.J., Reindl, M., Kremser, C., Kranewitter, C., Bonaros, N., Friedrich, G., Feuchtner, G., Metzler, B., 2018. Is MRI equivalent to CT in the guidance of TAVR? a pilot study. Eur. Radiol. 28, 4625–4634.

Oscanoa, J.A., Middione, M.J., Alkan, C., Yurt, M., Loecher, M., Vasanawala, S.S., Ennis, D.B., 2023. Deep learning-based reconstruction for cardiac mri: A review. Bioengineering 10. URL: https://www.mdpi.com/2306-5354/10/3/334, doi:10.3390/bioengineering10030334.

Osnabrugge, R.L., Mylotte, D., Head, S.J., Van Mieghem, N.M., Nkomo, V.T., LeReun, C.M., Bogers, A.J., Piazza, N., Kappetein, A.P., 2013. Aortic stenosis in the elderly: Disease prevalence and number of candidates for transcatheter aortic valve replacement: A meta-analysis and modeling study. Journal of the American College of Cardiology 62, 1002–1012. URL: https://www.sciencedirect.com/science/article/pii/S0735109713020792, doi:https://doi.org/10.1016/j.jacc.2013.05.015.

Pamminger, M., Klug, G., Kranewitter, C., Reindl, M., Reinstadler, S.J., Henninger, B., Tiller, C., Holzknecht, M., Kremser, C., Bauer, A., Jaschke, W., Metzler, B., Mayr, A., 2020. Non-contrast MRI protocol for TAVI guidance: quiescent-interval single-shot angiography in comparison with contrast-enhanced CT. Eur. Radiol. 30, 4847–4856.

Ram, P., Mezue, K., Pressman, G., Rangaswami, J., 2017. Acute kidney injury post–transcatheter aortic valve replacement. Clin. Cardiol. 40, 1357–1362.

Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jaeger, P.F., Maier-Hein, K., 2024. Mednext: Transformer-driven scaling of convnets for medical image segmentation. URL: https://arxiv.org/abs/2303.09975, arXiv:2303.09975.

Ruijsink, B., Puyol-Anton, E., Li, Y., Bai, W., Kerfoot, E., Razavi, R., King, A.P., 2020. Quality-aware semi-supervised learning for cmr segmentation. URL: https://arxiv.org/abs/2009.00584, arXiv:2009.00584.

Santaló-Corcoy, M.e.a., 2023. Tavi-prep: A deep learning-based tool for automated measurements extraction in tavi planning. Diagnostics 13, 3181.

Shi, J.e.a., 2024. Predictive accuracy-based active learning for medical image segmentation. arXiv preprint arXiv:2405.00452 .

Tahir, A.M., Mutlu, O., Bensaali, F., Ward, R., Ghareeb, A.N., Helmy, S.M.H.A., Othman, K.T., Al-Hashemi, M.A., Abujalala, S., Chowdhury, M.E.H., Alnabti, A.D.M.H., Yalcin, H.C., 2023. Latest developments in adapting deep learning for assessing tavr procedures and outcomes. Journal of Clinical Medicine 12. URL: https://www.mdpi.com/2077-0383/12/14/4774, doi:10.3390/jcm12144774.