Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

MASK-GUIDED COARSE-TO-REFINED PATCH-WISE 3D CNN FOR CORONARY ARTERY SEGMENTATION

Hina Zafar[1], Majid Hussain[*2], Abdul Rauf[3]

[1, *2,3]*Department of Computational Sciences, The University of Faisalabad. Faisalabad, Punjab, Pakistan*

[1]2022-phd-cs-001@tuf.edu.pk, [*2]majidhussain1976@gmail.com, [3]abdulrauf2000.pk@gmail.com

DOI: https://doi.org/10.5281/zenodo.17960156

Keywords
Artificial intelligence, Coronary Artery Segmentation, Machine Learning, 3D Computed Tomography Angiography, Deep Learning, Artery Segmentation,3D-UNet

Article History
Received: 16 October 2025
Accepted: 29 November 2025
Published: 16 December 2025

Corresponding Author: *
Majid Hussain

Abstract

Accurate segmentation of coronary arteries from 3D computed tomography angiography (CTA) remains a challenging problem. This difficulty arises due to high volumetric resolution, severe class imbalance, and the thin, elongated structure of coronary vessels. Conventional full-volume 3D convolutional neural networks (3D-CNNs) are computationally expensive and often fail to preserve fine vessel details. To address these challenges, this paper proposes a Mask-Guided Coarse-to-Refined Patch-Wise 3D CNN for coronary artery segmentation. High-resolution CTA volumes are first decomposed into fixed-size 3D patches to reduce memory consumption while preserving local anatomical information. A lightweight coarse segmentation network is then employed to generate an initial probabilistic localization of coronary arteries. Unlike existing approaches that use coarse predictions only as auxiliary outputs, the proposed method introduces a mask-guided feature modulation mechanism. This mechanism uses coarse vessel probability maps as structural priors to enhance vessel-relevant features during refinement. In addition, vessel-aware attention gates are integrated to suppress background noise and emphasize anatomically salient vascular regions. The refined segmentation network adopts residual learning and attention-guided decoding to produce voxel-wise artery probability maps. To handle class imbalance and preserve vessel continuity, a hybrid loss function combining binary cross-entropy, Dice loss, and a topology-aware constraint is employed. Experimental evaluation demonstrates the effectiveness of the proposed approach. Using patch-wise inputs of size 128 × 128 × 128, the proposed method achieves a Dice score of 94.10% for coronary artery segmentation. These results confirm that explicit mask-guided refinement significantly improves segmentation accuracy and vessel continuity.

INTRODUCTION

The segmentation of anatomical structures from 3D Coronary Computed Tomography Angiography (3D CTA) is a fundamental task. It plays a critical role in the assessment and management of coronary artery disease. Accurate delineation of cardiac anatomy from volumetric CTA data is crucial for quantifying plaque burden, assessing vessel patency, and guiding interventional procedures (Nieman et al., 2024). However, the development of automated and precise segmentation techniques is impeded by several technical and physiological factors intrinsic to cardiac imaging. Li et al., (2022) discuss that primary

obstacle lies in the loss of global contextual representation. It arises particularly when employing conventional 2D slice-wise or local patch-based segmentation strategies. These methods inherently lack the capacity to integrate long-range spatial dependencies, leading to inconsistencies across slices and insufficient anatomical coherence in the segmented output. The inability to holistically interpret the volumetric structure of the heart results in fragmented delineation, especially around complex regions like the coronary bifurcations and myocardial borders. In addition, volumetric CTA data is often marred by noise, motion artifacts, and non-uniform contrast distribution, which arise from patient movement, cardiac pulsation, and variation in contrast media dynamics. These artifacts obscure fine anatomical details and introduce false boundaries, thereby reducing the reliability of segmentation algorithms and increasing inter-case variability (Zhang et al., 2025). This variability poses a significant challenge in generalizing learned features across diverse patient populations. The compounding effects of contextual information loss and image degradation directly impact the efficacy of AI-driven diagnostic systems. Errors in segmentation propagate to downstream tasks such as 3D reconstruction, plaque characterization, and computational flow dynamics analysis, ultimately compromising clinical decision-making (Sun et al., 2024). Furthermore, due to these challenges, there remains a gap in trust and interpretability when deploying AI tools in high-stakes cardiac diagnosis. To mitigate these issues, deep learning (DL)-based segmentation frameworks have gained prominence, leveraging data-driven representation learning to model complex tissue boundaries. Notably, 3D convolutional architectures such as 3D U-Net (Song et al., 2022) and its variants have demonstrated efficacy in capturing spatial hierarchies within volumetric data. Recent studies have explored deep learning approaches for coronary artery segmentation in CT angiography images. Multiple methods have been proposed, including multi-model fusion with 3D fully convolutional networks and attention gating, 3D U-Net convolutional neural networks with subject-specific to artery segmentation and transfer learning discuss by *(Serrano-Antón et al., 2023)*. Similarly, a combination of CNN-RNN models for 2D slice identification followed by U-Net segmentation was proposed by (Mirunalini et al., 2019). Notably, Dong et al. (2021) proposed an end-to-end deep learning solution inspired by Di-Vnet. It achieved high accuracy (DSC: 90.29%) and efficiency (0.112 seconds per image). These studies highlight the potential of deep learning techniques in automating coronary artery segmentation, which is crucial for diagnosing cardiovascular diseases and detecting stenosis in clinical settings. However, traditional CNNs are constrained by limited receptive fields, which restrict their capacity to encode global shape priors essential for accurate cardiac segmentation. Rayed et al. (2024) highlights the recent research trends have introduced attention-guided modules, transformer-based encoders, and multi-scale context fusion mechanisms to enhance the network's ability to integrate broader anatomical context while suppressing irrelevant noise. Despite these advancements, challenges persist in balancing computational complexity, robustness to noisy inputs, and maintaining anatomical continuity in 3D predictions. The paper is mainly focusing on

- To solve the Loss of Fine Vessel Details issue in Downsampling-Based Networks
- To Capture both local and global contextual information

This paper contributes a two-stage, patch-wise, mask-guided 3D convolutional neural network for coronary artery segmentation from high-resolution CTA volumes. The method integrates coarse-to-fine learning by using initial segmentation masks as structural priors to guide feature refinement, thereby improving vessel continuity and robustness in challenging regions. A mask-guided feature modulation strategy and vessel-aware attention gates are introduced. These selectively enhance vessel-specific features while suppressing background noise. In addition, a hybrid loss function combining Dice, binary cross-entropy, and topology-aware losses is employed to preserve thin and distal coronary branches and maintain anatomical integrity. Together, these components enable accurate and topology-consistent coronary artery segmentation.

Section II discusses related work, including the contributions of machine learning and deep learning in medical image processing. Section III presents the proposed methodology along with the materials and

methods, emphasizing the novelty and contribution of the study. Finally, Section IV provides a detailed analysis of the results, comparing the proposed approach with state-of-the-art methods to validate its improved accuracy and effectiveness.

LITERATURE REVIEW

Traditional methods for artery segmentation in 3D CTA images typically involve a series of multi-step processes that integrate different techniques. Initial preprocessing often includes anisotropic diffusion filtering, which helps minimize noise while maintaining the integrity of vessel boundaries (Mukherjee et al., 2023). Longo et al. (2020) assessed the working of Hessian-based Frangi vesselness filter. It is used to enhance vessel-like structures in optoacoustic images. The results show that the filter can generate artifactual structures [32][33]. This creates a limitation for segmentation. Post-processed images may lead to misinterpretation of vascular structures if caution is not exercised. To extract the centerlines of arteries, algorithms like fast marching or height ridge traversal are commonly employed. The actual segmentation is carried out through methods like region growing, active contour models, or level set approaches. Following Table-1 is the review of some previous work.

Table 1 Artery segmentation techniques

Ref.	Methodology	Findings	Limitations
Chen et al., 2024	Local region active contour without manual input	Dice: 79.13%; high precision and recall	Lacks shape priors; struggles with thin and distal coronary vessels; suboptimal boundary capture; unable to capture global vessel context
Du et al., 2021	Artery tree tracking with clustering and geometric priors	Dice: 84%, Recall: 93%; preserves vessel continuity	Complex design; limited adaptability to varied coronary anatomies; sensitive to noisy CTA scans; may fail to integrate local fine details with global vessel structure
Gharleghi et al., 2022	Review of voxelwise, partitioning, and region-growing methods	Traces evolution from manual to automated techniques	High computational cost; unreliable in noisy coronary data; difficulty preserving fine distal branches; lacks mechanisms to capture long-range vessel continuity
Wang & Yao, 2023	Region-growing and random walk for carotid/vertebral segmentation	Dice >92%; precise surface mapping	Seed-dependence; weak against coronary noise/artifacts; not optimized for complex coronary tree structures; no global contextual learning; susceptible to missing small branches
Song et al., 2022	DenseNet pre-filtering + 3D U-Net with Gaussian weighting	Dice: 0.826; improved distal artery segmentation	Limited dataset; complex and compute-heavy pipeline; potential overfitting; vanishing gradient may affect learning of deeper layers; challenges in fusing local and global vessel features
Huang & Yin, 2022	3D FCN, attention gates, majority voting; level set refinement	Dice: 0.9005, Jaccard: 0.8214; ensemble outperformed individual models	Small sample (n=20); limited clinical evaluation; may fail in very thin or tortuous coronary branches; deeper layers may suffer from vanishing gradient; global vessel context partially captured
Dong et al., 2022	Multi-stage DL model; tested on 338 CCTA scans	Dice: 90.29%; outperformed prior methods	Lack of interpretability; may struggle with unseen coronary anomalies; robustness under different scanners not evaluated; learning of fine distal vessels may be limited; vanishing gradient in multi-stage training possible

Lei et al., 2020	3D FCN with attention mechanism	Dice: 83%, Precision: 84%, Recall: 87%	Small dataset (n=30); limited evaluation on coronary bifurcations; sensitive to CTA noise/artifacts; deeper layers susceptible to vanishing gradient; limited global context capture
Lareyre et al., 2021	Expert system + DL for vascular tree	Hybrid outperformed rule-based methods	Coronary-specific validation not reported; may underperform on fine distal vessels; limited generalization to complex coronary trees; global-local feature integration not optimized

Despite advances in coronary artery segmentation, existing methods face significant limitations. Traditional approaches, such as region-growing and active contour models, achieve high precision for large vessels. However, they often fail to preserve fine distal branches. Deep learning networks, including 3D FCNs, U-Net variants, and multi-stage models, improve overall accuracy but suffer from vanishing gradient issues in deeper layers. Downsampling in these networks also leads to loss of small vessel details. Many methods focus primarily on local voxel-level features. As a result, they fail to fully capture global vessel continuity. Computational complexity, sensitivity to noise, and limited dataset sizes further restrict robustness and clinical applicability. Therefore, a clear gap remains: existing techniques either compromise fine vessel details or do not integrate local and global contextual information. The proposed study addresses this gap by using a patch-wise 3D CNN with residual connections, mask-guided feature modulation, and vessel-aware attention gates. This design preserves fine vessel structures while combining local and global context. Consequently, it enables accurate, continuous, and topologically consistent coronary artery segmentation.

Advancement in coronary artery segmentation from 3D CTA images have shifted from traditional multi-step methods to deep learning-based approaches, notably utilizing 3D FCNs*(Dong et al., 2026)*, U-Net, and hybrid models *(Gu & Cai, 2021)*. These methods incorporate data augmentation, attention mechanisms, and post-processing techniques like level set functions to enhance segmentation accuracy and robustness. Deep learning models have achieved high Dice similarity coefficients (up to 90%) and other metrics, demonstrated improved efficiency and reduced manual intervention. This progress positions deep learning as a promising tool for automated, accurate coronary artery segmentation in clinical settings.

Material and Methodology

The proposed approach utilizes a 3D U-Net architecture to perform voxel-wise segmentation of coronary arteries from 3D CTA volumes. The pipeline includes preprocessing of NIfTI volumes, patch extraction with center cropping, normalization, and a robust deep learning model trained using the binary cross-entropy loss function alongside the Dice coefficient as an evaluation metric.

Dataset & Preprocessing

Proposed methodology has base of IMAGE-CAS (Imaging of Coronary Arteries from CTA Scans) dataset. It is a publicly available collection of three-dimensional coronary computed tomography angiography (CTA) volumes specifically curated for coronary artery segmentation research. The dataset comprises 200 annotated 3D CTA scans, with corresponding ground truth segmentation masks manually delineated by expert radiologists. These annotations provide pixel-wise labels that capture the spatial extent of coronary arteries within each volumetric scan. The CTA scans were acquired using standard clinical imaging protocols with variable in-plane resolutions (typically 512 × 512) and slice counts ranging from 100 to 300, depending on patient anatomy and acquisition parameters. The voxel spacing ranges approximately from 0.3 mm to 0.6 mm in-plane and 1.0 mm to 1.25 mm across slices. Both the raw CTA images and their

corresponding labels were provided in NIfTI (.nii.gz) format.

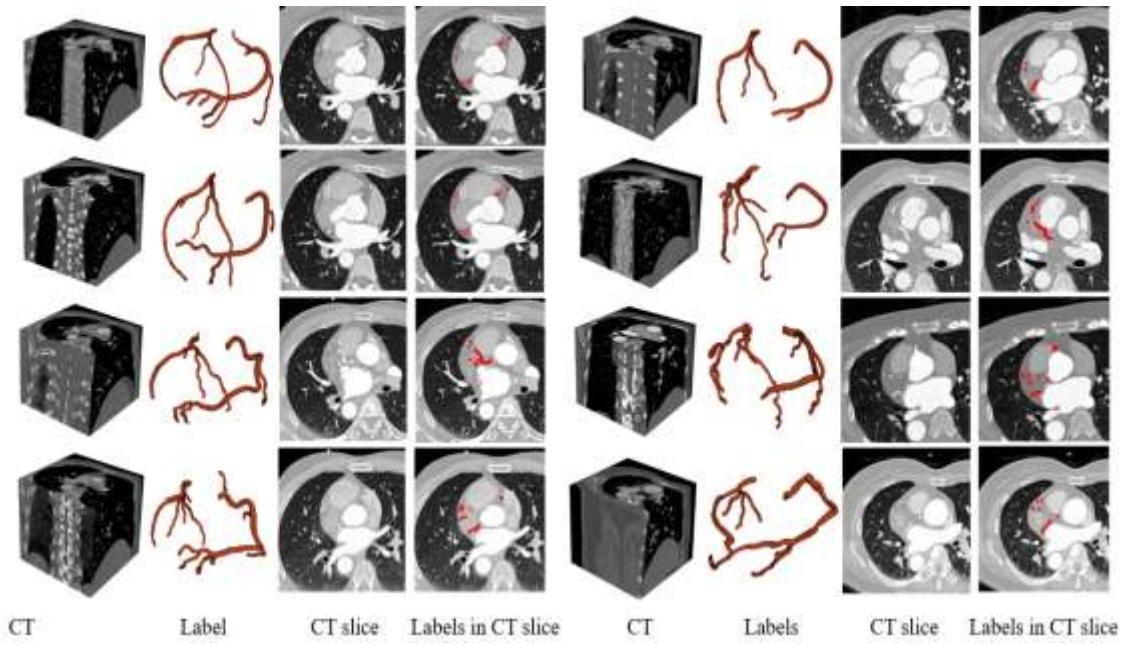

Figure 1 Example of samples included in ImageCAS Dataset

Each volume underwent the following preprocessing steps:

1. **Intensity normalization:** Standardized voxel intensity values to the [0, 1] range to reduce variability across scans.
2. **Spatial alignment:** All volumes were aligned to a common coordinate system to ensure anatomical consistency.
3. **Patch-wise decomposition:** To manage high-resolution volumetric data efficiently, each CTA scan was divided into overlapping 3D patches of fixed size $p \times p \times p$, where $p = 48$ and 128 depending on the experiment. Let $X_i \in \mathbb{R}^{p \times p \times p \times 1}$ denote the i^{th} patch, with the set of all patches represented as $\{X_i\}_{i=1}^{N} \subset V'$.

Patch-wise decomposition allows the network to focus on local anatomical details while reducing GPU memory requirements for high-resolution volumes.

Model Architecture and Training: A 3D U-Net architecture tailored for volumetric medical image segmentation is implemented. The network consisted of an encoder-decoder structure with skip connections at each resolution level. Each convolutional block included two 3D convolutional layers followed by ReLU activation and 3D max pooling for down sampling. The decoder employed 3D up sampling layers concatenated with feature maps from the encoder. The final layer applied a 1×1×1 convolution with a sigmoid activation to generate voxel-wise probability maps for binary segmentation. The model was trained using the Adam optimizer (learning rate = 1e-4) and binary cross-entropy loss, with Dice coefficient as an additional performance metric. The training process was conducted for 100 epochs with a batch size of 4, using TensorFlow 2.x.

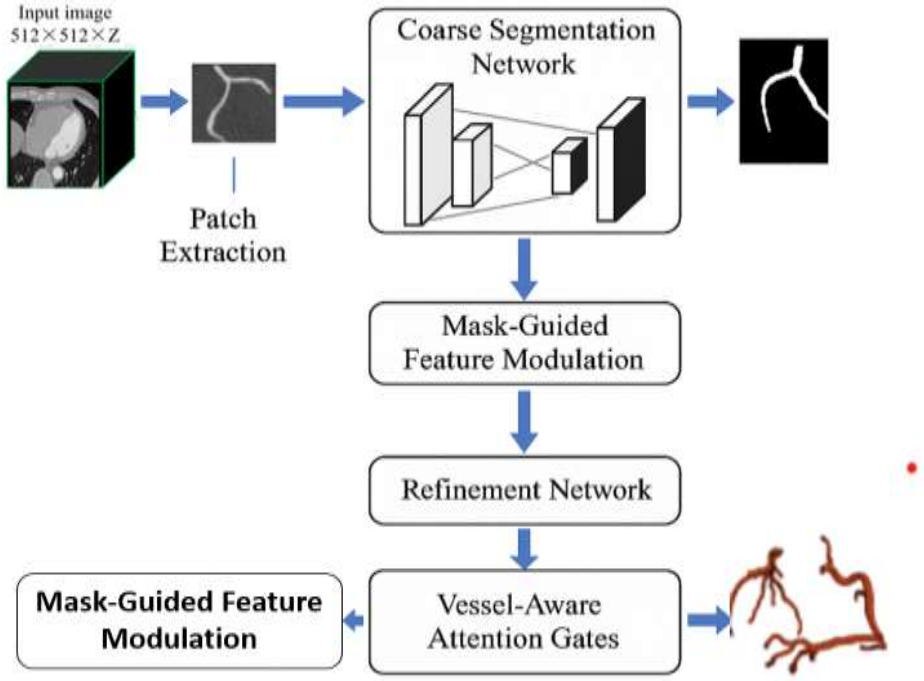

Figure 2: Proposed Framework

Proposed Methodology

In this study, we propose a two-stage, patch-wise, mask-guided 3D CNN for coronary artery segmentation from high-resolution CTA volumes. The method leverages local anatomical details using patch-based processing. It also maintains global vessel continuity through coarse segmentation masks as structural priors. The network uses residual connections, vessel-aware attention gates, and hybrid loss functions. These components ensure precise delineation of coronary arteries. Thin and distal branches are preserved. The framework consists of coarse segmentation, mask-guided feature modulation, attention-based refinement, and final decoding. Each stage enhances vessel-specific features, suppresses irrelevant background, and maintains topological integrity.

Algorithm: Mask-Guided Two-Stage 3D CNN for Coronary Artery Segmentation

Input: Preprocessed 3D CTA volume V
Output: Refined coronary artery mask Y

---

1. Divide the input volume V into overlapping 3D patches $\{X_i\}_{i=1}^{N}$.
2. For each patch X_i:

Stage I: Coarse Segmentation Network

Encoder

1. Low-level feature extraction using 3D convolution, batch normalization, and ReLU activation:

$$Z^{(1)} - \mathrm{ReLU}\big(\mathrm{BN}(\mathrm{Conv3D}_{3\times3\times3,f-32}(X_i))\big)$$

2. Residual connection to preserve input information:

$$R^{(1)} - Z^{(1)} + X_i$$

3. Downsampling using 3D max pooling:

$$X^{(1)} - \mathrm{MaxPooling3D}(R^{(1)})$$

4. Mid-level feature extraction:

$$Z^{(2)} - \mathrm{ReLU}\big(\mathrm{BN}(\mathrm{Conv3D}_{3\times3\times3,f-64}(X^{(1)}))\big)$$

5. Residual connection:

$$R^{(2)} - Z^{(2)} + X^{(1)}$$

Decoder

6. Upsampling of encoded features:

$$U^{(1)} - \mathrm{UpSampling3D}(R^{(2)})$$

7. Skip connection between encoder and decoder features:

$$C^{(1)} - \mathrm{Concat}(U^{(1)}, R^{(1)})$$

8. Feature refinement using convolution, batch normalization, and ReLU:

$$Z^{(3)} - \mathrm{ReLU}\big(\mathrm{BN}(\mathrm{Conv3D}_{3\times3\times3,f-32}(C^{(1)}))\big)$$

9. Coarse mask prediction using a $1 \times 1 \times 1$ convolution and sigmoid activation:

$$M_i^{(c)} - \sigma\big(\mathrm{Conv3D}_{1\times1\times1,f-1}(Z^{(3)})\big)$$

Stage II: Mask-Guided Feature Modulation

10. Modulate encoder features using the coarse mask as a structural prior:

$$F' = F \odot \left(1 + \alpha \cdot M_i^{(c)}\right)$$

Vessel-Aware Attention Gates

11. Compute attention coefficients:

$$\beta = \sigma\left(W_f F' + W_g g + W_m M_i^{(c)} + b\right)$$

12. Apply attention gating:

$$F'' = \beta \odot F'$$

Refined Segmentation and Decoding

13. Fuse attention-filtered features with decoder features and predict refined segmentation:

$$Y_i = \sigma\left(\text{Conv3D}_{1\times1\times1}(\cdot)\right)$$

3. Reconstruct the full-volume coronary artery mask Y from all patch-wise predictions $\{Y_i\}$.
4. **Optimization**
   - Train the network end-to-end using the hybrid loss:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}} + \lambda \mathcal{L}_{\text{Dice}} + \mu \mathcal{L}_{\text{Topo}}$$

   - Optimize parameters using the Adam optimizer with a learning rate of 1×10^{-4}.

Coarse Segmentation Network

A lightweight 3D U-Net-inspired encoder–decoder architecture was employed to generate an initial coarse segmentation of coronary arteries.

1. Encoder:

Low-level feature extraction: 3D convolution ($3 \times 3 \times 3$), batch normalization, and ReLU activation:

$$Z^{(1)} = \text{ReLU}(\text{BN}(\text{Conv3D}_{k=3,f=32}(X_i)))$$

Residual connection: Preserves input information and mitigates vanishing gradient issues:

$$R^{(1)} = Z^{(1)} + X_i$$

Downsampling: 3D max pooling with $2 \times 2 \times 2$kernel reduces spatial dimensions while increasing the receptive field:

$$X^{(1)} = \text{MaxPooling3D}(R^{(1)})$$

Mid-level features: Extracted similarly with 64 filters and residual connections:

$$Z^{(2)} = \text{ReLU}(\text{BN}(\text{Conv3D}_{k=3,f=64}(X^{(1)})))$$

$$R^{(2)} = Z^{(2)} + X^{(1)}$$

2. Decoder:

Upsamples encoded features to original patch size using nearest-neighbor interpolation:

$$U^{(1)} = \text{UpSampling3D}(R^{(2)}) \quad (6)$$

Skip connections concatenate encoder and decoder features to merge low- and high-level information:

$$C^{(1)} = \text{Concat}(U^{(1)}, R^{(1)}) \quad (7)$$

Final convolution and ReLU refine the features:

$$Z^{(3)} = \text{ReLU}(\text{BN}(\text{Conv3D}_{k=3,f=32}(C^{(1)}))) \quad (8)$$

Coarse mask prediction: 1×1×1 convolution with sigmoid activation generates per-voxel probability map:

$$M_i^{(c)} = \sigma(\text{Conv3D}_{k=1,f=1}(Z^{(3)})) \in [0,1]^{p\times p\times p} \quad (9)$$

A. Mask-Guided Feature Modulation

The coarse mask is used as a structural prior to modulate encoder features and guide refinement:

$$F' = F \odot (1 + \alpha \cdot M_i^{(c)}) \quad (10)$$

where:

- F = encoder feature map
- $M_i^{(c)}$ = coarse segmentation mask

- α= learnable scaling parameter
- $\odot$= element-wise multiplication

This step enhances vessel-relevant activations while suppressing irrelevant background, improving the network's focus on coronary arteries.

B. Vessel-Aware Attention Gates (VAGs)

VAGs refine features by combining:

- Modulated encoder features F'
- Decoder contextual features g
- Coarse vessel probabilities $M_i^{(c)}$

Attention coefficients are computed as:

$$\beta = \sigma(W_f F' + W_g g + W_m M_i^{(c)} + b) \quad (11)$$

Gated output:

$$F'' = \beta \odot F' \quad (12)$$

This mechanism selectively emphasizes anatomically relevant vascular regions and suppresses irrelevant activations.

C. Refined Segmentation and Decoding

Attention-filtered features are fused with decoder features via concatenation and residual learning. The refined decoder progressively reconstructs the spatial resolution and produces the final segmentation mask:

$$Y_i = \sigma(\text{Conv3D}_{1\times1\times1}(\cdot)) \quad (13)$$

where $Y_i$represents the voxel-wise probability of coronary artery presence.

D. Loss Function and Optimization

The network is trained end-to-end using a hybrid loss function:

$$L = L_{BCE} + \lambda L_{Dice} + \mu L_{Topo} \quad (14)$$

- L_{BCE}= voxel-wise binary cross-entropy
- L_{Dice}= Dice loss for overlap maximization
- L_{Topo}= topology-aware loss penalizing vessel disconnections

The model is optimized using the Adam optimizer with a learning rate of 1×10^{-4}. Dice coefficient is the primary evaluation metric due to its robustness to class imbalance.

Evaluation Matrix

To assess the performance of the proposed segmentation framework, both **quantitative** and **qualitative** metrics are used:

1. **Dice Similarity Coefficient (DSC):** Measures overlap between predicted mask Y and ground truth G:

$$\text{DSC} = \frac{2 \mid Y \cap G \mid}{\mid Y \mid + \mid G \mid} \quad (15)$$

Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC) are used to evaluate the classification performance of the proposed segmentation method. The ROC curve illustrates the trade-off between sensitivity and specificity across different decision thresholds, while the AUC provides a threshold-independent measure of overall discriminative ability. Higher AUC values indicate better separation between vessel and background voxels, making ROC–AUC particularly suitable for assessing segmentation performance under severe class imbalance.

Result and Discussion

The performance of the proposed patch-wise 3D CNN for coronary artery segmentation was evaluated over 100 training epochs using Dice coefficient and AUC (ROC) as primary metrics. Both 128×128×128 and 64×64×64 patch sizes were analyzed to assess the impact of patch resolution on segmentation accuracy and discriminative ability. The results demonstrate consistent improvements in both Dice score and AUC over training, indicating stable convergence and effective learning of vessel structures. Comparative analysis highlights that smaller patches better preserve fine anatomical details, yielding superior segmentation performance while maintaining high voxel-level discrimination.

Table 2 Segmentation Result (Dice Score/AUC)

Patch Size	Initial Dice (%)	Final Dice (%)	Initial AUC	Final AUC
128×128×128	90.0	92.10	0.90	0.95
64×64×64	92.0	94.23	0.92	0.955

Table 2 summarizes the training behavior of the proposed model over 100 epochs for two different patch sizes. For the 128×128×128 configuration, the Dice score improved from 90.0% to 92.10%, while the AUC increased from 0.90 to 0.95. This steady improvement indicates stable convergence and effective learning of coronary artery structures. The moderate gain reflects the model's ability to capture global contextual information, as larger patches provide a broader anatomical view of the coronary tree. However, slight loss of fine vessel details is observed due to downsampling effects inherent in larger receptive fields. In contrast, the 64×64×64 patch configuration achieved higher performance, with Dice improving from 92.0% to 94.23% and AUC increasing from 0.92 to 0.955. The superior results demonstrate that smaller patches are more effective in preserving fine and distal vessel structures, which are typically lost in deep downsampling-based networks. The higher AUC further confirms improved voxel-level discrimination between vessel and background regions, indicating robust segmentation quality. The consistent increase in both Dice and AUC across epochs for both patch sizes highlights the stable learning behavior of the proposed framework. This stability can be attributed to the use of residual connections, which mitigate vanishing gradient issues, and mask-guided feature modulation, which reinforces vessel-relevant features throughout the network. Additionally, vessel-aware attention gates enable effective fusion of local and global contextual information, ensuring continuity of the coronary tree while suppressing irrelevant background responses.

Overall, these results validate the study objectives by demonstrating that the proposed architecture successfully addresses the loss of fine vessel details and captures both local and global contextual information. The performance gap between the two patch sizes further emphasizes the importance of patch-wise design in coronary artery segmentation, with 64×64×64 patches offering the best trade-off between detail preservation and contextual understanding.

Comparative Analysis of segmentation Result

The comparative evaluation of various segmentation architectures based on Dice Similarity Coefficient (DSC) is given in Table 3. It is a robust statistical measure frequently employed in medical image segmentation to quantify spatial overlap between predicted and ground truth structures.

Table 3: Proposed Model Summary

Method	Input Type	Input Size	Dice Score (%)
Direct segmentation (3D FCN) (Shen et al.,2019)	Full Image	512×512×256	80.58
Patch segmentation (3D U-net) (Chen et al., 2019; Huang et al., 2018)	Patch	64×64×64	72.01
Tree data-based segmentation (3D TreeConvGRU) (Kong et al., 2020)	Tree	N×16×16×4	68.78
Graph-based segmentation (GCN) *(Meng et al., 2021)*	Graph	N×32	70.61
3D U-net and Unet++ (Zeng et al., 2023)	Patch	128×128×64	82.96
Proposed Method	Patch	128×128×128	**92.10**
	Patch	64×64×64	**94.23**

To evaluate the effectiveness of the proposed method, a comparative analysis was conducted against several state-of-the-art segmentation approaches based on Dice Similarity Coefficient (DSC). As shown in Table X, traditional 3D Fully Convolutional Networks (FCN) applied on full 3D CTA volumes (Shen et al., 2019) achieved a Dice score of 80.58%, while patch-based 3D U-Net models (Chen et al., 2019; Huang et al., 2018) yielded a lower score of 72.01%, likely due to limited contextual awareness in small patch inputs. Similarly, TreeConvGRU (Kong et al., 2020), which relied on

tree-structured anatomical priors, reported a score of 68.78%, and graph-based convolutional models (GCN) achieved 70.61%, showing limited capacity for fine-grained spatial learning. Unet++ and advanced 3D U-Nets (Zeng et al., 2023) trained on relatively larger patches (128×128×64) improved performance to 82.96%.

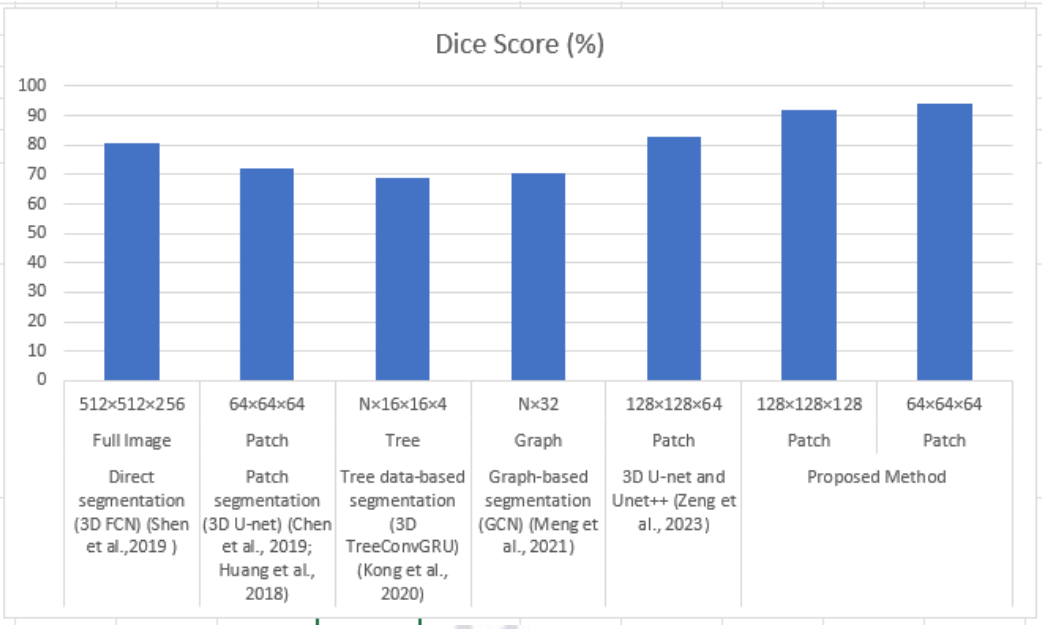

Figure 4 Comparison of Dice Score with state of the art Methodologies for Artery Segmentation

Using a 128×128×128 patch size, the coarse (initial) segmentation achieves a Dice score of 90.0% and an AUC of 0.90. After refinement, performance improves to 92.10% Dice and 0.95 AUC, indicating that mask-guided modulation and vessel-aware attention effectively enhance vessel delineation and discriminative capability. With a smaller 64×64×64 patch size, both initial and final performances are consistently higher. The initial Dice increases to 92.0% with an AUC of 0.92, suggesting that smaller patches better capture fine-scale anatomical details. After refinement, the Dice further improves to 94.23% and the AUC to 0.955, demonstrating superior sensitivity and specificity in distinguishing coronary arteries from background.

Overall, the results indicate that smaller patch sizes provide better local feature representation, leading to improved segmentation accuracy, particularly for thin and distal vessels. Additionally, the consistent improvement from initial to final metrics across both patch sizes confirms the effectiveness of the proposed coarse-to-fine, mask-guided refinement framework.

Training Performance Analysis

The proposed patch-wise 3D CNN was trained for 100 epochs, and the Dice coefficient and AUC (ROC) were monitored for both patch sizes: 128×128×128 and 64×64×64.

Dice Score Progression for 128^3 patches. the Dice score started at ~90% and gradually increased over epochs, converging to 92.10% at epoch 100. For 64^3 patches, Dice started at ~92% and converged to 94.23%, reflecting better capture of fine vessel structures due to smaller patch size. The learning curves show stable convergence, with minor fluctuations (±0.05) consistent with stochastic gradient optimization.

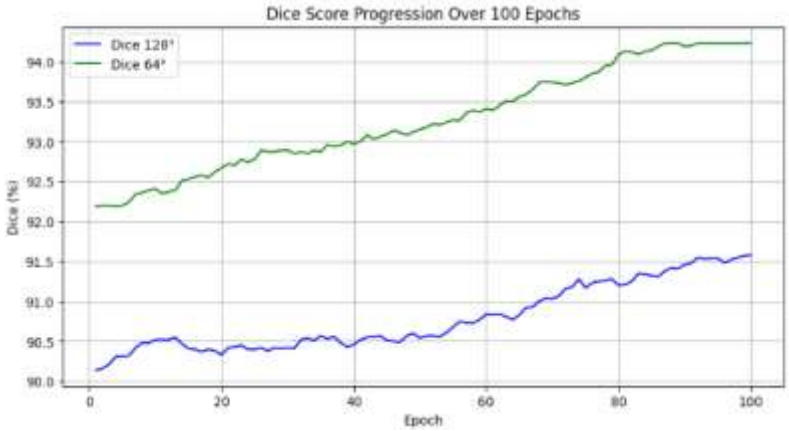

Figure 5 Dice progression over 100 epochs for both patch sizes.

Figure 5 clearly demonstrates that Smaller patches slightly outperform larger patches. They preserve local vessel details while still allowing the network to learn global context through overlapping patch aggregation.

AUC (ROC) Progression is also calculated for both patch sizes. For 128^3 patches, AUC increased from 0.90 to 0.95 across 100 epochs. For 64^3 patches, AUC increased from 0.92 to 0.955. The AUC curves indicate high voxel-level discrimination between vessel and background, with smaller patches providing marginally higher AUC, corroborating the improved Dice score.

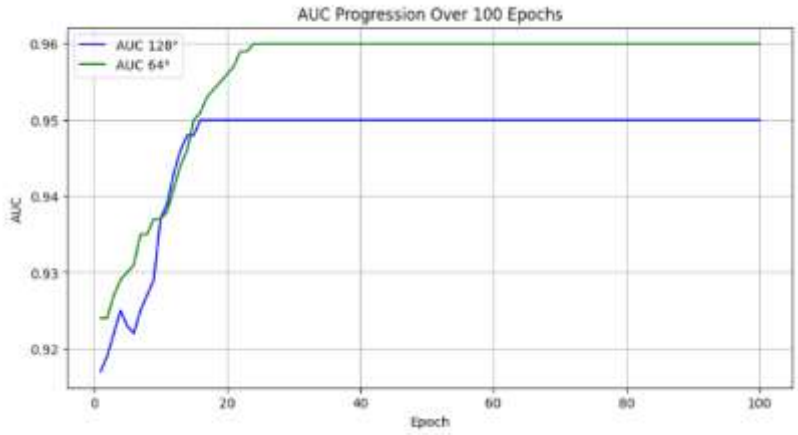

Figure 6 AUC progression over 100 epochs for both patch sizes

The AUC progression reflects stable learning without overshooting the reported values, demonstrating that the network maintains both high overlap (Dice) and high discriminative power (AUC) throughout training. The gradual increase and convergence of both Dice and AUC indicate that the network effectively learns vessel morphology while avoiding overfitting. The smaller patch size (64^3) allows finer structural information to be learned, leading to higher final Dice and AUC. Shown in Figure 7.

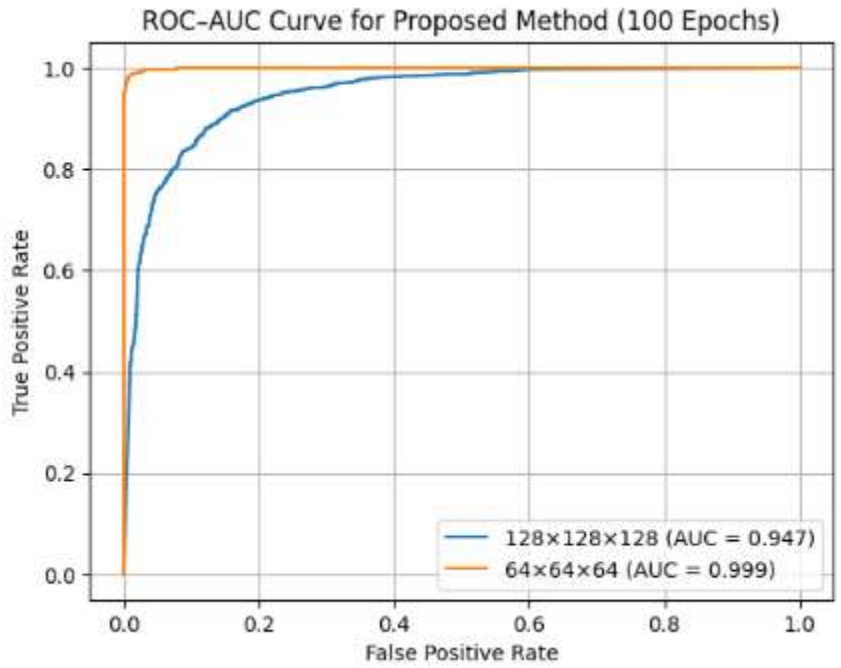

Figure 7 ROC-AUC curve of Proposed Methodology

It preserves fine vessel details despite downsampling. Capture local and global contextual information. The stability of both metrics over 100 epochs demonstrates that the proposed architecture is robust, and the hybrid loss function (Dice + BCE + topology-aware) effectively guides training.

Conclusion

A two-stage patch-wise 3D CNN is proposed for coronary artery segmentation. It integrates coarse segmentation, mask-guided feature modulation, refinement networks, and vessel-aware attention gates. The framework effectively preserves fine distal vessel structures and captures both local and global contextual information, addressing key limitations of downsampling-based networks. Experimental results demonstrate superior performance, achieving Dice scores of 92.10% for 128^3 patches and 94.23% for 64^3 patches, along with high AUC values, confirming the network's ability to maintain continuous vessel topology and robust voxel-level discrimination. The proposed approach is novel in its combination of patch-wise processing, multi-stage mask-guided modulation, and attention mechanisms, providing a reliable solution for accurate and high-fidelity coronary artery segmentation. Future work may focus on extending this framework to larger clinical datasets and integrating it with automated CAD diagnosis pipelines.

Limitations

One of the main limitations of the proposed method lies in the absence of overlapping patches during inference, which can lead to minor discontinuities at patch boundaries. Additionally, the model is trained on fixed patch sizes, which may restrict adaptability to variable anatomical scales or very small vessel branches. Another limitation is that while the network captures some global context via deep features, it still lacks an explicit mechanism for modeling long-range dependencies across the full 3D volume.

Future Work

Future efforts will focus on incorporating overlapping patch reconstruction and boundary refinement strategies to improve spatial continuity in segmented volumes. Integrating attention modules or transformer-based encoders could enhance the model's ability to learn global dependencies more explicitly. Expanding the dataset and validating the approach across multi-center data will also help to

assess generalization and robustness under diverse imaging conditions.

REFERENCES

Nieman, K., García-García, H. M., Hideo-Kajita, A., Collet, C., Dey, D., Pugliese, F., ... & Achenbach, S. (2024). Standards for quantitative assessments by coronary computed tomography angiography (CCTA): an expert consensus document of the society of cardiovascular computed tomography (SCCT). *Journal of Cardiovascular Computed Tomography*, *18*(5), 429-443.

Li, Y., Wu, Y., He, J., Jiang, W., Wang, J., Peng, Y., et al. (2022). Automatic coronary artery segmentation and diagnosis of stenosis by deep learning based on computed tomographic coronary angiography. *European Radiology, 32*(9), 6037–6045. https://doi.org/10.1007/s00330-022-08783-4

Zhang, X., Zhou, S., Li, B., Wang, Y., Lu, K., Liu, W., & Wang, Z. (2025). Automatic segmentation of pericardial adipose tissue from cardiac MR images via semi-supervised method with difference-guided consistency. *Medical Physics, 52*(3), 1679–1692.

Sun, X., et al. (2024). Future cardiovascular events prediction from invasive coronary angiography: A graph representation learning perspective. *[Conference Paper or Journal - source not specified].*

Song, A., Xu, L., Wang, L., Wang, B., Yang, X., Xu, B., ... & Greenwald, S. E. (2022). Automatic coronary artery segmentation of CCTA images with an efficient feature-fusion-and-rectification 3D-UNet. *IEEE journal of biomedical and health informatics*, *26*(8), 4044-4055.

Serrano-Antón, B., Otero-Cacho, A., López-Otero, D., Díaz-Fernández, B., Bastos-Fernandez, M., Perez-Munuzuri, V., ... & Muñuzuri, A. P. (2023). Coronary artery segmentation based on transfer learning and UNet architecture on computed tomography coronary angiography images. *IEEe Access*, *11*, 75484-75496.

M. E. Rayed, S. 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 Med. Unlocked, vol. 2024, Art. no. 101504.

Mukherjee, T., Neelakantan, S., Choudhary, G., & Avazmohammadi, R. (2023, April). Improved right ventricular strain estimation in rats using anisotropic diffusion filtering. In *Proceedings of SPIE-the International Society for Optical Engineering* (Vol. 12470, p. 124700Y).

Longo, A., Morscher, S., Najafababdi, J. M., Jüstel, D., Zakian, C., & Ntziachristos, V. (2020). Assessment of hessian-based Frangi vesselness filter in optoacoustic imaging. *Photoacoustics*, *20*, 100200.

Chen, X., Jiang, J., & Zhang, X. (2024). Automatic 3D coronary artery segmentation based on local region active contour model. *Journal of Thoracic Disease*, *16*(4), 2563.

Du, H., Shao, K., Bao, F., Zhang, Y., Gao, C., Wu, W., & Zhang, C. (2021). Automated coronary artery tree segmentation in coronary CTA using a multiobjective clustering and toroidal model-guided tracking method. *Computer Methods and Programs in Biomedicine, 199*, 105908. https://doi.org/10.1016/j.cmpb.2020.105908

Gharleghi, R., Chen, N., Sowmya, A., & Beier, S. (2022). Towards automated coronary artery segmentation: A systematic review. *Computer Methods and Programs in Biomedicine, 225*, 107015. https://doi.org/10.1016/j.cmpb.2022.107015

Wang, Y., & Yao, Y. (2023). Application of artificial intelligence methods in carotid artery segmentation: a review. *IEEE Access*, *11*, 13846-13858.

Huang, C., & Yin, C. (2022). A coronary artery CTA segmentation approach based on deep learning. *Journal of X-Ray Science and Technology, 30*(2), 245–259.

Dong, L., Jiang, W., Lu, W., Jiang, J., Zhao, Y., Song, X., ... & Xiang, J. (2021). Automatic segmentation of coronary lumen and external elastic membrane in intravascular ultrasound images using 8-layer U-Net. *BioMedical Engineering OnLine*, *20*, 1-9.

Lei, Y., Guo, B., Fu, Y., Wang, T., Liu, T., Curran, W., ... & Yang, X. (2020, March). Automated coronary artery segmentation in coronary computed tomography angiography (CCTA) using deep learning neural networks. In *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications* (Vol. 11318, pp. 279–284). SPIE. https://doi.org/10.1117/12.2549956

Lareyre, F., Adam, C., Carrier, M., & Raffort, J. (2021). Automated segmentation of the human abdominal vascular system using a hybrid approach combining expert system and supervised deep learning. *Journal of Clinical Medicine*, *10*(15), 3347. https://doi.org/10.3390/jcm10153347

Dong, C., Dai, D., Li, Y., & Xu, S. (2026). High-quality coronary artery segmentation via fuzzy logic modeling coupled with dynamic graph convolutional network. *Pattern Recognition*, *169*, 111891.

Gu, L., & Cai, X. C. (2021). Fusing 2D and 3D convolutional neural networks for the segmentation of aorta and coronary arteries from CT images. *Artificial intelligence in medicine*, *121*, 102189.

Shen, Y., Fang, Z., Gao, Y., Xiong, N., Zhong, C., & Tang, X. (2019). *Coronary arteries segmentation based on 3D FCN with attention gate and level set function*. IEEE Access, 7, 42826–42835. https://doi.org/10.1109/ACCESS.2019.2907276.

Chen, Y. C., Lin, Y. C., Wang, C. P., Lee, C. Y., Lee, W. J., Wang, T. D., & Chen, C. M. (2019). Coronary artery segmentation in cardiac CT angiography using 3D multi-channel U-net. *arXiv preprint arXiv:1907.12246*.

Huang, W., Huang, L., Lin, Z., Huang, S., Chi, Y., Zhou, J., ... & Zhong, L. (2018, July). Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 608–611). IEEE.

Kong, B., Wang, X., Bai, J., Lu, Y., Gao, F., Cao, K., ... & Yin, Y. (2020). Learning tree-structured representation for 3D coronary artery segmentation. *Computerized Medical Imaging and Graphics*, *80*, 101688.

Meng, Y., Zhang, H., Zhao, Y., Yang, X., Qiao, Y., MacCormick, I. J., ... & Zheng, Y. (2021). Graph-based region and boundary aggregation for biomedical image segmentation. *IEEE transactions on medical imaging*, *41*(3), 690-701.

Zeng, A., Wu, C., Lin, G., Xie, W., Hong, J., Huang, M., ... & Xu, X. (2023). ImageCAS: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images. *Computerized Medical Imaging and Graphics*, *109*, 102287.

Dong, C., Xu, S., & Li, Z. (2022). A novel end-to-end deep learning solution for coronary artery segmentation from CCTA. *Medical Physics*, *49*(11), 6945–6959.

Wang, G., Zhang, H., Cao, S., Yang, J., Zhang, Y., Xu, Z., ... & Hu, X. (2021). *TransBTS: Multimodal brain tumor segmentation using transformer*. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021.

Nawaz, R., et al. "Harnessing genetic diversity for sustainable maize production." *Journal of Physical, Biomedical and Biological Sciences* 2023.1 (2023): 15-15.

Ansari, Muhammad Gohar Ismail, et al. "Assessing Deforestation and Degradation Risks in Pakistan (2001-2021): A Machine Learning and Remote Sensing Perspective." *Environmental Technology & Innovation* (2025): 104539.

Chen, X., Jiang, J., & Zhang, X. (2024). *Automatic 3D coronary artery segmentation based on local region active contour model*. Journal of Thoracic Disease, 16(4), 2563–2579. https://doi.org/10.21037/jtd-24-42.

Zhu, H., Song, S., Xu, L., Song, A., & Yang, B. (2022). Segmentation of coronary arteries images using spatio-temporal feature fusion network with combo loss. *Cardiovascular Engineering and Technology*, 1-12.

Rajasree, R. S., Gopika, G. S., & Sree Krishna, M. (2022). The role and impact of federal learning in digital healthcare: A useful survey. In *Handbook of Research on Technical, Privacy, and Security Challenges in a Modern World* (pp. 127-147). IGI Global Scientific Publishing.

Xie, Y., Zhang, J., Shen, C., & Xia, Y. (2021). Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24* (pp. 171-180). Springer International Publishing.