A 3D Convolutional Neural Network Architecture for Early Detection of Coronary Artery Blockage (Coronary-3D-UNet)

by

AHANAF ABID SAZID
22301269
MUSHFIQUR RAHMAN
22301278
MD. SABBIR AKON
22301242
JAUAD AHMED SADIK
22301342
MD JABED CHOWDHURY
22301279

A thesis submitted to the
Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
January 2026

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

AHANAF ABID SAZID
22301269

MUSHFIQUR RAHMAN
22301278

MD. SABBIR AKON
22301242

JAUAD AHMED SADIK
22301342

MD JABED CHOWDHURY
22301279

Approval

The thesis/project titled "A 3D Convolutional Neural Network Architecture for Early Detection of Coronary Artery Blockage (Coronary-3D-Unet)" submitted by

1. AHANAF ABID SAZID (22301269)
2. MUSHFIQUR RAHMAN (22301278)
3. MD. SABBIR AKON (22301242)
4. JAUAD AHMED SADIK (22301342)
5. MD JABED CHOWDHURY (22301279)

of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering in Spring, 2026.

Examining Committee:

Supervisor:
(Member)

Md. Ashraful Alam, PhD
Associate Professor
Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Mr. Rafeed Rahman
Senior Lecturer
Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Associate Professor and Chairperson
Computer Science and Engineering
BRAC University

Abstract

Coronary Artery Disease (CAD) is a major cause of death all over the globe where the estimated number of annual deaths are 9 million. Early and precise diagnosis of the stenosis of the coronary arteries is critical in providing clinical intervention and better patient outcomes. Despite the fact that invasive coronary angiography (ICA) is regarded as the most effective diagnostic tool, the procedure has certain risks and is not convenient enough to be used widely because of its inapplicability to whole population screening. Coronary Computed Tomography Angiography (CCTA) is a non-invasive, high-resolution alternative but manual analysis of CCTA images is time consuming and subject to inter-observer error. This paper presents Coronary-3D-Unet, a parameter-efficient 3D convolutional neural network to perform automated stenosis assessment and coronary artery segmentation on CCTA volumes. The residual learning and dual attention mechanisms (spatial and channel) and multi-scale feature fusion are incorporated into the proposed architecture to improve vessel representation and resistance to various challenges such as small vessel structures, low contrast, and imaging artifacts. The framework allows automatic stenosis grading of severity as mild, moderate, and severe geometrically. On a publicly available benchmark dataset, experimental results demonstrate that Coronary-3D-Unet reaches a mean Dice Similarity Coefficient (DSC) of 76.6% and bests the official ImageCAS baseline with a significantly lower model complexity. The model is based on an efficient number of 3.8 million parameters, which is 83% less than the conventional dense 3D networks, so the model can be used to infer efficiently and be deployed practically. The model can also be integrated with clinical Picture Archiving and Communication Systems (PACS) that facilitate real-time analyses and better clinical usability.

Keywords: Coronary Artery Disease (CAD), CCTA, Deep Learning, 3D U-Net, V-Net, AttentionU-Net, Coronary-3D-Unet, Medical Image Segmentation, Attention Mechanism, Residual Learning, Stenosis Detection, Multi-Scale Feature Fusion.

Dedication

This paper is dedicated to every member, the faculties, and every individual who has made efforts and contributed to keeping this research paper in a single piece. This has been possible so far through their dedication and contribution.

Acknowledgement

To begin with, we would like to give all praises to the **Almighty Allah**, the most gracious and the most merciful, who gave us the power, patience, and capability to successfully complete this thesis without significant setbacks.

We would like to thank our thesis supervisor, **Dr. Md. Ashraful Alam**, Assistant Professor, Department of Computer Science and Engineering, BRAC University. His insightful advice, guidance, constructive feedback, and supervision played a significant role in determining the direction of this research. We are also grateful to our mentor, **Shammo Biswas**, Research Assistant, Department of Computer Science and Engineering, BRAC University. His support, technical guidance, and readiness to assist us in troubleshooting problems were invaluable during the start of this thesis and even up to the end.

We also want to express our heartfelt gratitude to **Xiaowei Xu**, Researcher at the Guangdong Provincial People's Hospital, as well as Co-Director of the Laboratory of Artificial Intelligence and 3D Technologies in Cardiovascular Diseases. We are happy that he contributed to the curation of the large-scale ImageCAS dataset and made it publicly available to the research community. His timely help regarding questions about the datasets and his efforts toward creating a standard in coronary artery segmentation were key to our analysis during the experiments.

Moreover, we thank our parents very much. Their undying love, sacrifices, and constant motivation have been our driving force and have helped us reach this far in our academic life.

Last but not least, we are able to achieve this because of our focused teamwork, determination, and mutual drive to proceed and contribute meaningfully to our field.

Table of Contents

List of Figures

List of Tables

Nomenclature

AGFA-Net	Attention-Guided and Feature-Aggregated Network
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operating Characteristic
BCE	Binary Cross-Entropy
CAD	Coronary Artery Disease
CAD-RADS	Coronary Artery Disease–Reporting and Data System
CCTA	Coronary Computed Tomography Angiography
CFD	Computational Fluid Dynamics
CHD	Coronary Heart Disease
DICOM	Digital Imaging and Communications in Medicine
DR	Dense Residual
DSC	Dice Similarity Coefficient
ECA-UNet	Efficient Channel Attention U-Net
EHR	Electronic Health Records
FIE	Feature Interaction Extraction
FFR	Fractional Flow Reserve
FFR-CT	Fractional Flow Reserve derived from Computed Tomography
FL	Focal Loss
FRM	Feature Refinement Module
Grad-CAM	Gradient-weighted Class Activation Mapping
HFIM	Hierarchical Feature Integration Module
HU	Hounsfield Units
ICA	Invasive Coronary Angiography
ImageCAS	Imagebase of Coronary Artery Segmentation
IoU	Intersection over Union
LCT	Local Contextual Transformer
LRP	Layer-wise Relevance Propagation
MEP	Multilayer Enhanced Perceptron
MGFA-Net	Myocardial Region-guided Feature Aggregation Network
MRG	Myocardial Region-guided Module
MRI	Magnetic Resonance Imaging
MSFF	Multi-scale Feature Fusion
NIfTI	Neuroimaging Informatics Technology Initiative
NPV	Negative Predictive Value
PACS	Picture Archiving and Communication Systems
RFEE	Residual Feature Extraction Encoding
SAFA	Scale-aware Feature Augmentation
XAI	Explainable Artificial Intelligence

Chapter 1

Introduction

1.1 Background

Coronary artery disease (CAD) is one of the most widely recognized causes of deaths in the world with an average of approximately 9 million deaths that occur annually [1–3]CAD evolves when atherosclerotic plaque occludes the coronary arteries to the myocardium causing vessel stenosis, reduced oxygen delivery to the cardiac muscle, and increased incident of acute cardiac occurrences. Thus, it is very important to detect cases of stenosis early so as to introduce beneficial interventions in good time to lower mortality and enhance clinical results [3]. Although invasive coronary angiography (ICA) remains the gold standard of diagnosis, the fact that it involves the insertion of catheters corresponds with the risks associated with the procedure and the severely limited ability to be performed routinely, or on large groups [4]. .

The coronary computed tomography angiography (CCTA) has emerged as a non-invasive technique that is an excellent way of visualizing the coronary vessels in high-resolution 3D and has gained popularity as a standard method of CAD diagnosis [5, 6]. However, CCTA analysis is a complex procedure due to different complications such as the introduction of motion artifacts that normally complicate the vessel visibility in the presence of other complications such as calcifications and poor contrast. Besides, manually marked results are tedious and may have a significant inter-observer dispersion [5, 6].

Convolutional neural networks (CNNs) have evolved to be one of the efficient ways of automating the issue of segmentation and location of the lesions on medical images [7, 8] . Correctly segmenting Coronary Computed Tomography Angiography (CCTA) is critical to the early treatment and diagnosis of Coronary Heart Disease [9].The 3D U-Net architectures (volumetric input skip connections) among them maintain the anatomical continuity and show better performance at identifying the small vessel structures than using 2D networks [7, 10]. Additions like attention reinforcements which enable the network to emphasize on the diagnostically important regions [7] and residual learning enhance the gradient flow and overall training stability [11]. Additionally, incorporation of 2D+3D fusion techniques using 2D networks giving the global contextual features and 3D nets to segment the anatomy in additional detail has resulted in better segmentation continuity on the aorta as well as the coronaries [12]. Also, the accuracy rates of the coronary centerline extraction by

3D U-Nets proved to be 90–95%. which is conducive to more informative vessel pathway analysis [13]. It is also possible to mention such prominent architectures as 3D dedicated networks applied to whole-heart MRA to classify stenosis [14] and deep learning networks that can separate left and right coronary arteries in angiographies [15].

1.2 Rational of the Study or Motivation

The latest developments in deep learning specifically volumetric U-Net architectures have shown to increase the accuracy of segmentation and localization of the coronary artery and stenosis in CCTA scans [7, 8, 10]. New designs such as the DR-LCT-UNet and two-stage pipelines (2.5D 3D refinement) have also brought the bar a little higher Dice scores of 0.85 & 0.90 and satisfactory inference times [11]. Nevertheless, on further examination, it can be noted that there are still persistent gaps which are not intelligently covered by existing literature:

1. **Suboptimal performance on early-stage stenosis**: The majority of models show high sensitivity on moderate/severe stenosis, but they are unreliable in the detection of lesions that are less significant in terms of the volume comprising the plaque, which are in most cases concealed due to low contrast, motion, or calcification [5, 6, 11] . Diagnosis of these changes early is important in prevention and prediction of the risk [3].
2. **Computational cost**: 3D CNNs are known to be very computationally expensive, time consuming and multi-pass pipeline architecture [8, 11]. Efficiency-based methods, such as pruning or quantization, do not frequently make it into the studies, thus hampering practical usage in the clinical environment in real time.
3. **Limited generalizability across datasets**: Most of the models are trained on selected datasets or scan procedures and fail to generalize when used with external data since there are domain shifts and patient variability [16].
4. **Absence of considerations of clinical deployment:** In cases of good performance, minimal discussion of the actual considerations to include in clinical deployment, e.g. how the system would be integrated with existing clinical infrastructure such as PACS, or how its explanation mechanisms would be presented in a way that clinicians can readily understand and use.

1.3 Problem Statement

Although certain advancements have been achieved in deep learning-based CAD segmentation, there are still important challenges restricting the application of the method in practice:

- Limited capability of detecting mild-stage stenosis especially in non noisy/ low contrast imaging background without which the potential of the automated systems is nullified [5, 6, 11] .
- Expensive computing needs, inference time, and time-prohibitive hardware demands in clinical settings that demand real time (such as time sensitive diagnosis) [8, 11].
- Weak resilience to a variety of imaging conditions and protocols, as a result of which performance is reduced under new out-of-distribution data [16].

The combination of these problems leads to the existence of a significant impediment: existing models of segmentation are academically interesting, but are neither specific in their results, efficient and reliable enough to be implemented into the clinic workflows.

1.4 Objective

This paper is a study to establish and confirm Coronary-3D-Unet, an improved 3D convolutional neural network architecture capable of properly segmenting the coronary arteries and detecting stenosis at an early stage using Coronary Computed Tomography Angiography (CCTA) scans. To address the limitation of high computational cost in existing 3D-CNNs, this work develops a parameter-efficient architecture. The primary objective is to strategically integrate lightweight Residual Blocks and Dual-Attention mechanisms to reduce model complexity by approximately 83% (from 22.6M to 3.8M parameters) while achieving highly competitive segmentation accuracy (76.60% Mean DSC) compared to heavier baselines, and outperforming the official ImageCAS benchmark paper (0.702) . This optimization enables faster inference times, providing an opportunity to deploy the model in real-time clinical settings. To allow the widest applicability, the model is trained and rigorously validated on the ImageCAS benchmark dataset (1000 volumes), with the specific goal of exceeding the 70.2% mean Dice baseline established in the original publication while maintaining a minimal parameter footprint. Moreover, the fact that the final system could integrate into clinical routines without disturbing their workflows due to the compatibility with PACS infrastructure presents a convenient user interface, where clinicians could upload scan data and obtain stenosis prediction and popular explainability measures (e.g. Grad-CAM, spatial attention maps) [7]. This study fills the existing gap and by deciding the major drawbacks in precision, efficiency, generalizability, and clinical applicability, performs an exercise of connecting the literature's deep learning models to actual implementation in cardiac diagnostic systems [3, 5, 6].

1.5 Methodology in Brief

In essence, Coronary-3D-Unet builds on the conventional U-Net incorporating:

- **Residual blocks** in both the decoder and the encoder to reduce vanishing gradient, and allow deeper learning.
- **Spatial and channel based attention mechanisms** to allow the network to pay attention to areas of interest that are critical, e.g., subtle narrowing or early lesions.
- **Multi-scale feature fusion** such that the model can pick up both gross/fine details of the anatomy-even in low-contrast or highly artifact regions.

Training is performed using publicly available annotated 3D CCTA datasets (post pre-processing step of rescaling, normalization, noise reduction, and data augmentation that aids to further the learning and counterploys overfitting). The precise segmentation provided by the model enables the accurate geometric quantification of vessel narrowing, facilitating the subsequent clinical determination of stenosis severity as mild, moderate or severe and the annotations are checked by specialist cardiologists.

In order to compare segmentation performance, we primarily use the Dice Similarity Coefficient (DSC), while model reliability is assessed through voxel-wise Uncertainty Quantification.The training is optimized with Adam optimizer with a composite loss function $\mathrm{L}_{\mathrm{total}}=0.5\times Dice+0.5\times BCE$. This is a hybrid objective, where we used the Dice loss to maximize the structural overlaps and Binary Cross-Entropy (BCE) loss to ensure voxel level classification accuracy, resulting in a high fidelity segmentation, which will be used for further automated stenosis detection and geometric grading.

We prioritized a parameter-efficient architectural design, minimizing computational overhead by approximately 83% compared to dense 3D baselines achieving high inference speeds without requiring complex post-processing steps like quantization. The final model outputs are formatted for compatibility with PACS (Picture Archiving and Communication Systems) to facilitate future clinical integration in the medical field and an intuitive interface enables a clinician to be able to upload a scan and get the immediate stenosis prediction and interpretability techniques such as **Grad-CAM and spatial attention maps**.

1.6 Scopes and Challenges

In this study, the authors propose to design and train a variant of the 3D U-NeT architecture with attention mechanism, residual learning, and Multi-scale feature integration named Coronary-3D-Unet in order to achieve the highly accurate segmentation of the coronary arteries and in early detection of stenosis. To be clinically interpretable, computationally efficient, and able to be seamlessly incorporated into the current medical infrastructure, the system is planned to be computationally efficient, clinically interpretable, and easy to integrate into the existing healthcare infrastructure.

Scope of the Study:

- A U-Net architecture 3D improved with attention, residual learning, and multi-level feature combination to be used in the segmentation of coronary arteries is precise and efficient.
- A built-in detection module works on the geometry of segmented lumen to detect early narrowing of the lumen.
- Comprehensive empirical validation on real CCTA datasets, which shows an enhanced accuracy, sensitivity and performance in terms of computation superior to the baseline usage of 3D CNNs.
- A fully deployable and optimized in a sense of a system of CAD diagnosis in real-time, which would be compatible with the current clinical infrastructure.

Challenges and Limitations:

- Training complexity: Adding attention and residual functions, the model becomes more complex and deep, and thus, more computational and tuning, may be necessary.
- Annotation quality and variability: Obtaining high-quality, consistent markings on early stage stenosis is hard and inter-observer inconsistency on the part of cardiologists may add noise to the ground truth.
- Real-time deployment: Although the architecture is optimized for efficiency, deploying full volumetric 3D models in hardware constrained clinical settings remains a computational challenge.
- Clinical acceptance: For usage in clinical practice, interpretability, reliability and in many cases regulatory acceptability, may be essential but the path to acceptance can be lengthy and difficult.

Chapter 2

Literature Review

2.1 Literature Review

Important in the avoidance of major adverse cardiac events is the early and correct diagnosis of Coronary Artery Disease (CAD). The Coronary Computed Tomography Angiography (CCTA) has now emerged as one of the major non-invasive imaging techniques that were used in diagnosis of CAD. Nevertheless, manual analysis of CCTA scans are time consuming, based on subjectivity and they also require the experience of clinicians. This has encouraged the creation of automated networks to accomplish this task and deep learning, especially Convolutional Neural Networks (CNNs), has demonstrated great potential in the realization of not only CAD classification but also coronary artery segmentation, a major step in the detection of stenosis. The analysis of the recent developments in 3D deep learning architectures applied to these tasks is given in the following review.

In their efforts to eliminate the weaknesses of typical CNNs in the acquisition of long-range dependence, Zhao et al. [3] came up with the idea of a hybrid deep learning framework of an entire diagnostic pipeline, including both segmentation and stenosis localization. They used a TransUNet architecture to perform the segmentation, the synergetic combination of CNN encoder and Vision Transformer. The detailed, local feature maps are extracted using the CNN backbone, whereas long-range feature correlations are learned using the Transformer module, included within the bottle neck of the U-net, with features being modeled using Transformers self-attention mechanism. This enables the network to comprehend the general setting, dreadful courses of the coronary arteries. After this sophisticated segmentation, the segmented vessels are analyzed through a different network referred to as localization network, which locates the areas of high stenosis. This model of two steps sequential process of outlining anatomy followed by identification of pathology echoes the clinical process of diagnosing. Their findings point to the fact that the hybrid TransUNet model outperforms traditional CNN-based models in their performance in terms of segmentation accuracy offering a more reliable map of the anatomy in which the succeeding, clinically-relevant task of detecting stenosis can run.

In order to solve the typical problem of lack of annotated medical data, Serrano-Antón et al. [4] considered transfer learning to utilize in the segmentation of coronary arteries in CCTA images. Their idea was of a U-Net based architecture with a standard encoder architecture being exchanged to utilize a pre-trained ResNet34 model on the huge ImageNet dataset. The argument is that the pre-trained encoder will give a stronger and more generally applicable feature extraction backbone which can be quickly re-trained on the particular task of vessel segmentation. Their process occurred in two steps: A coarse segmentation of the ascending and descending aorta was done to define a region of interest, followed by the fine-grained segmentation of the coronary arteries emerging out of the region by using a dedicated U-Net having ResNet34 encoder. The choice of approach takes advantage of the different features of anatomic localization priors as well as transfer learning capability. The model performed well on their own dataset reporting a Dice similarity coefficient of 77.2%, showing that a pre-trained, deeper network can be a useful technique in U-Net to improve performance in niche medical imaging challenges.

Kaba et al. [17] published a detailed overview of the field and conducted a survey on the use of different deep learning models to segment coronary arteries and classify them. The review notes continuing problems with CCTA analysis, such as the small and detailed structure of the vessels, the poor contrast-to-noise ratio and motion due to the cardiac cycle. They observe that, among other architectures, the U-Net and 3D U-Net (such as V-Net) system have proven to be the de-facto choice on segmentation tasks, because they perform well and are efficient in the encoder-decoder with skip-connect structure, especially in preserving fine-grained details. The authors have their own implementation of the U-Net model to segment the coronary arteries and reached a Dice score of 81.3 percent and high accuracy of 99.4 percent in their test data. Their contribution will offer an efficient representation of the panorama, in which the pre-eminence of models based on U-Net together with the comprehension of the state-of-the-art are anchored and the limited relevance of that structure to the field of automated CAD interpretation can be ratified.

As a method to develop the precision of the coronary artery segmentation of Coronary Computed Tomography Angiography (CCTA) pictures, Wang et al. [5] recommended an extraordinary new structure of U-Net 3D calculations, which was called DR-LCT-UNet. The authors indicated that two major issues with common CNN-based segmentation exist: the possibility of losing the fine details of small vessels in the original layers of the network and the inability to distinguish coronary arteries and adjacent structures with a similar Hounsfield Unit value. In order to solve these problems they came up with two novel modules. The former is the Dense Residual (DR) module, which is intended to substitute standard convolution blocks in both U-Net encoder layer and decoder layer. The dense residual connections stream working with the DR module to preserve and fuse multi-level features further helps in segmenting low or narrow vessel areas which is critical. The second innovation is the Local Contextual Transformer (LCT) module that is made a part of U-Net architecture in skip connections. LCT module is an attention module, which is used to improve the feature representation with the focus on local contextual information that allows the network to better distinguish coronary arteries. The network was trained and tested with a proprietary dataset, CorArtTS2020 with 81 CCTA cases. It was experimentally proved that the DR-LCT-UNet could reach a Dice Similarity Coefficient (DSC) value of 85.8%, a recall of 86.3%, and precision of 85.8%, surpassing other alternative methods including the baseline 3D-UNet.

Although direct classification is an option, proper segmentation of the coronary arteries gives vital morphological data to be used extensively in diagnosis and treatment planning. Segmentation is however relegated by problems such as low image contrast, motion artifacts as well as complex, variable geometry of the coronary tree. Liu and Zhao [6] used an advanced U-shaped encoder-decoder Attention-Guided and Feature-Aggregated 3D deep network (AGFA-Net) to address these challenges in order to classify the coronary arteries. The network presents three new modules, making it better in performance on a massive dataset of 1,000 CCTA scans. Issued with parallel channel and spatial attention mechanisms, the Feature Refinement Module (FRM) enables feature refinement in the background and eliminates irrelevant background noise. The bottleneck of the encoder module applies a Scale-aware Feature Augmentation (SAFA) module that employs self-attention and dilated convolutions to capture multi-scale contexts and long-range contexts, which is essential toward segmenting vessels of different sizes. In the decoder, features are hierarchically merged across the scales according to the Hierarchical Feature Integration Module (HFIM) to restore fine details. AGFA-Net offered a state-of-the-art Dice coefficient of 86.74%, which is significantly higher than the one of typical architectures, namely 3D U-Net (81.06%), and V-Net (83.43%). To maintain the class imbalance, it used a loss combination of weighted cross-entropy and Dice loss. The beneficial impact of every proposed module on overall accuracy of the network was verified systematically in the course of ablation studies.

On the basis of advanced segmentation, Yao et al. [8] extended a pipeline to stenosis assessment with anatomical prior knowledge incorporated in the network architecture. Their Myocardial Region Guided Feature Aggregation Net (MGFA-Net) is a new dual-encoder U-shaped network that can both segment coronary arteries and provide an automatic stenosis detection opportunity in the future. Myocardial Region-guided Module (MRG) is the core innovation that expands pre-segmented myocardial mask, as a second input, to an auxiliary encoder. At every stage of the main image encoder, features of this anatomical map are concatenated forcing the network to concentrate attention on the clinically relevant area and to ensure that other tissues are not excessively segmented. It has also an architecture that consists of a Residual Feature Extraction Encoding (RFEE) block that has parallel spatial and channel attention to facilitate feature discrimination and a Multi-scale Feature Fusion (MSFF) block in the decoder to maintain fine anatomical details. In the validation of the publicly available ImageCAS dataset of 1,000 scans, MGFA-Net obtained a 85.04 percent Dice score as well as an improved 95 percent Hausdorff Distance with a score of 6.1294 mm better than other models including CS2-Net. This framework also takes Monte Carlo Dropout to deliver uncertainty maps in clinically interpretable forms. More importantly, the process does not stop after segmentation but uses a morphology based algorithm on the result to extract centerlines and disjoint arterial branches to go on to obtain cross-sectional areas so as to automatically identify and grade the stenosis. It was also found that adoption of this integrated technique enhanced the true positive rate in stenosis detection by 5.46 percent over the incorporation of a standard 3D U-Net to serve as the basis of segmentation.

Duan et al. [11] used the U-Net as the architecture to improve the representations of the functions, so they proposed another model, the ECA-UNet, which incorporates the Efficient Channel Attention module into the model to perform segmentation of coronary arteries. It is grounded on the assumption that feature channels are not all equally informative and their adaptive weight recalibration can be used to boost performance. The ECA module is a very efficient channel attention mechanism that scales the local cross-channel interaction strategy approach that does not require the process of reducing dimensions, with a 1D convolution that reacts to adaptively calculate the coverage of the interaction. By incorporating these lightweight ECA modules into the skip connections and decoder path of a deeper residual U-Net, the network is taught to amplify important features of vessel structures, and cancel out irrelevant background clatter. The ECA-Unet got a brilliant Intersection over Union (IoU) of 94.29% on their test set. Proven through the study, making use of an advanced and lightweight attention mechanism in enhancing the segmentation performance of the U-net model is an effective way of considerably increasing its accuracy with little to no increase in computational costs.

Continued modifications of the 3D U-Net architecture have been developed to better deliver accuracy while maintaining noticeable-efficiency. Song et al. [10] has suggested an effective three-stage model based on their newly proposed Feature-Fusion-and-Rectification 3D-UNet (FFR-UNet) to segment coronary arteries on CCTA. Since the task of utilizing 3D networks on whole volumes of high-resolutions can be computationally demanding, the initial phase of their work uses 2D DenseNet to conduct a rough classification of image patches (within the region of interest) to allow the identification of a section that holds the coronary arteries. The second step uses the main FFR-UNet on this smaller, focal volume. The epitomizing feature of the FFR-UNet architecture is the fusion-and-rectification block that competitively integrates characteristics of both dense connections and residual connections with the goal to capitalize on the advantages of each without increasing the level of complexity in the model. The purpose of this block is to attain rich representations of the features of the complex vessels forms. The last step is in which it applies a Gaussian weighted averaging scheme on the prediction with overlapping patches to generate a more consistent ending segmentation. In their data of 110 CCTA scans, this multi-stage model with FFR-UNet returned a Dice score of 87.14 percent, a significant increase above a conventional 3D U-Net (82.16 percent), and is further evidence of the usefulness of the initial localization stage as well as the new network block structure.

Finally, the clinical aim of CAD detection would be to deliver a normalized purposeful report on patients to deal with. Huang et al. [7] have created a 3D CNN, with which the assignment of the Coronary Artery Disease Reporting and Data System (CAD-RADS) score based on CCTA scans is directly automated. The clinical classification of the levels of stenosis is based on CAD-RADS score, where disease severity goes from CAD-RADS 0 (absence of plaques) to CAD-RADS 5 (complete vessel occlusion). Their 3D CNN is built using a ResNet network together with some modifications and it uses a CCTA volume and a segmentation mask of the heart; they classify a number of classes and thus the overall CAD-RADS score of the patient can be predicted. The training and validation were performed with a set of almost 4,000 scans and the model was tested on an outer, different set of 496 scans. The system performed very well as a diagnostic tool in detection of clinically significant obstructive CAD (CAD-RADS of greater than or equal 3) with an area under the ROC curve (AUC) in the external test set of 0.94. The consensus of the model with expert human readers was high indicating its potential to save on the preparation of clinical reporting and inter-reader variations. This is very relevant to this work because it is an actual, automatic way to grade coronary blockage, which is what the work is to accomplish with the Coronary-3D-Unet being suggested in this thesis.

The major limit of clinical deployments of deep learning models is that they are black boxes and the preprocessing pipelines of which are often not intuitively clarified to clinicians by the need of tasks such as centerline extraction. Cheung et al. [1] resolved it by suggesting a direct, end-to-end 3D deep learning classifier of CAD analyzing evaluation that avoids tedious preprocessing techniques before a centerline extraction or a multi-planar reconstruction. They implement a 3D ResNet-50 model to label the whole volumes of CTCA as normal or a CAD patient. The analysis was carried on 88-subjects data with an equal distribution between the two classes. One of the main directions of their work involved explainability; they also used the Gradient-weighted Class Activation Mapping (Grad-CAM) technique to see which parts of the model the model considered important to its decisions. They tried out the binary cross-entropy (BCE) and focal loss (FL) functions and concluded that FL generated less diffuse and more clinically significant heatmaps, which identified the heart and coronary arteries even when the prediction was not accurate. Using 3D ResNet-50 model, an accuracy of 71.43% is obtained on test data that is substantial as compared to that of other state of the art models such as 3D DenseNet-121 and 3D EfficientNetB0 with a difference of 21.43% and 21.43% respectively on their test set. Otherwise, the authors showed that da Vinci problem could be reformulated as 2D segmentation task, which would give more accurate explainability, explicitly identifying the vessels of interest.

Early diagnosis and control of Coronary Heart Disease (CHD) depends on the precise segmentation of the Coronary Computed Tomography Angiography (CCTA) images. Although conventional Convolutional Neural Networks (CNNs) have formed the main basis of CCTA segmentation, their use in achieving global contextual information and precise segmentation of the complicated and noise-filled coronary structure often, proves to be a failure. To overcome these shortcoming, Xu et al. [18] proposed a novel deep learning framework called TransCC combining Transformer networks with CNN in a synergistic manner to provide better CCTA segmentation. Within this framework, there exists a Feature Interaction Extraction (FIE) module that is sensitive to the image patch characteristics and avoid the loss of semantic information and a Multilayer Enhanced Perceptron (MEP) module that facilitates the focusing of the Transformer on local spatial information. The TransCC structure consists of a stack of Transformer modules as an encoder, MEP, and the U-Net style of the decoder, and was trained on the CCTA dataset of 848 images, showing the best results with the average Dice coefficient-0.730 and Intersection over Union (IoU)-0.582, and so can be regarded as the pioneer solution of Transformer approach to CCTA segmentation issue. Although TransCC is a critical step in the development of 2D CCTA segmentation by bearing great insight to hybrid CNN-Transformer models, it is crucial to mention that the thesis user has a creation design that is preoccupied with a 3D Convolutional Neural Network Architecture for Early Detection of Coronary Artery Blockage (Coronary-3D-Unet), and therefore there may be a need to learn how the advantages of effective local and global feature combination exhibited by TransCC can be managed into a three-dimensional scenario.

Keeping in mind that the large field-of-view achieved through 2D CNNs by itself comes at a cost of compromisation with contextual information provided in slices, Gu and Cai [9] designed a two-stage fusion method to segment the aorta and the coronary arteries CT images. The idea in their method is to capitalize on the merits of each method and remedy the reliability of segmentation especially of the stenotic arteries. At the first level, a slice-by-slice segmentation of the aorta and coronary arteries is done by a 2D CNN using modified U-Net architecture (UNet++ASPP). The large field-of-view is well used with this stage to localize the vessels and eliminate most of the negative voxels to alleviate the class imbalance problem. In the second phase, a 3D U-Net with residual connection (3D ResUNet) is used to do a more accurate segmentation of the coronary arteries identified as subregions of interest (identified as such in the first stage). This is a 3D network that aims at estimating the inter-slice knowledge to enhance the continuity of the segmented vessels and to minimize the false negative detection of the vessels. Nevertheless, the methodology was tested on a patient group of 59 patients with CAD. The outcomes gave noticeable improvement following the second phase as the DSC of the coronary artery segmentation improved by 2.51 percent rate as it went up to 86.62 per cent and the sensitivity also improved by 6.04 per cent as it went up to 89.52 per cent utilising the UNet++ASPP model in the initial phase. Study successfully shows that the combination of 2D and 3D networks capabilities can result in stronger and more precise segmentation results in comparison with the modality unions.

The study by Dorobanțiu et al. [12] was dedicated to the particular problem of automatic extraction of coronary centerline on CCTA which is a basis used in many diagnostic applications. They suggested a set-up of 3D U-Net that predicted the probability of each voxel being a part of a vessel centerline. The authors pointed to the fact that training such network is associated with the following significant obstacles the scarce annotation of most public datasets (when not every vessel is annotated) and the extreme class imbalance between the few voxels of the centerline and the massive background. In order to address these limitations the paper uses a patch-based training strategy, using small 3D patches sampled in the entire CT volume and allows data augmentation to efficiently train the model on the available data. Of more concern, they proposed a new composite loss, which is the combination of focal loss and a tailored overlap loss. Focal loss was key in the need to condition the network to learn the infrequent positive examples (voxels of the centerline), whereas overlap loss ensured the stability of training, despite the conflicting, sparse ground truth. The approach was estimated on a widely available data of the Rotterdam Coronary Artery Centerline Extraction benchmark. The accuracy of the proposed 3D-UNet was recorded to be 90-95 and the overlap score was also recorded to be 90-94 percent in several training settings which show that the network is also capable of learning even when the data conditions are not too ideal.

After the replacement of the imaging methodology by the Whole-Heart Coronary Magnetic Resonance Angiography (WHCMRA) versus CT, a computer-based classification of the significant coronary artery stenosis was created by Shiomi et al. [14]. Non-invasive, non-radiation, and WHCMRA is an attractive diagnostic tool, as well as the lower resolution and the signal-to-noise ratio, compared to CT makes automated analysis difficult. The suggested approach employs a 3D-CNN to categorize little sums of interest (VOIs) that are 21x21x21 voxels by size and are situated at the point characterized by a radiologist. An important novelty of the work is the inclusion of two separate attention mechanisms into the architecture of the 3D-CNN in order to enhance the classification accuracy. The former is an attention mechanism in coronary artery which is pre-trained to produce an attention map, which accentuates coronary arteries in the VOI, suppressing background content as well as noise. The second one is annotated point attention mechanism, which resorts to a weight matrix in the form of a Gaussian distribution to highlight features that are in close connection to the point that is annotated. Our network was tested on 951 coronary segments of 75 patients data with five-fold schema cross-validation. The dual-attention 3D-CNN that was proposed performed well with a classification accuracy of 0.875, a sensitivity of 0.905, a specificity of 0.873, and an area under the receiver operating characteristic curve (AUROC) of 0.944 which is significantly higher than the models that lack such attention mechanisms.

Deep learning was not only successful in 3D CCTA analysis, but also in applying to other coronary imaging modalities to other clinical tasks. A workflow optimization and quality control problem in the 2D X-ray coronary angiography area was considered by Eschen et al. [15]. To automatically classify angiogram images into the Left coronary artery (LCA) and the right coronary artery (RCA), they came up with a deep learning model. This categorization constitutes an easy yet essential procedure of arranging huge databases of angiographic researches. A number of the more well-known architectures of 2D CNNs (e.g., VGG16, ResNet50, and EfficientNet) were tested by the authors and transfer learning was applied using models with pre-trained ImageNet. With such a huge dataset of more than 59,000 angiograms, we found that the EfficientNetB0 model was the most likely to yield excellent results, combining an accuracy of 98. 7%. Explainability was done with the help of Grad-CAM, which ensured that during LCA and RCA classification, the model learnt to look at the distinguishing characteristics of the vasculatures to make its determinations. The modality of data (2D angiography, 3D CCTA) and task (vessel system classification, stenosis detection) are not related to the problem described in the current thesis, the material presents the stability and the impressive accuracy of standard CNNs on cardiovascular large-scale imaging and thereby demonstrates the wide applicability of deep learning to many problems in the cardiology field.

Recent developments in related applications of cardiac imagings provide methodological insights that are applicable in the case of primary interest of such a thesis; coronary arteries from CCTA. Benameur et al. [19] designed a better 3D U-Net structure to segment a left ventricle and myocardium of cardiac MRI images. The main issue in this respect consists in the tendency of misrepresenting papillary muscles as endocardial wall which introduces errors in the assessment of the volume of ventricles and ejection fraction. Their application shows the value of starting with preliminary segmentation using a 3D U-Net and following it by an anatomy-appropriate post-processing procedure. Once the network produces a starting segment of the blood pool, a given algorithm, which entails morphological transformations and a thresholding procedure, identifies, and excludes the papillary muscles leading to a more precise left ventricular segmentation. Even though the work on modality (cardiac MRI) and target anatomy (ventricle) is slightly different to the one presented, the fundamental concept is of extreme importance: a vanilla 3D U-Net can be greatly augmented with knowledge about the anatomy, in the form of post-processing refinement stage in this case, to address certain correct and recurrent problems related to segmentation. This dual method indicates a possible future direction in which coronary artery segmentation may be enhanced, where post-processing may be applied to, e.g. impose topological restrictions or dissolve severe calcifications within the vessel lumen.

Although precision should be prioritized, the level of efficiency, including interpretability, of deep learning models is important to the clinical adoption. Iqbal et al. [13] have resolved these questions and created a lightweight deep neural network that can be used in real-time to classify coronary artery disease using Magnetic Resonance Imaging (MRI), paying much attention to explainable AI (XAI). In their work, the significance of model architecture and hyperparameters optimization as a way of building a network that is both accurate and computationally efficient to the extent that it should be broadly applicable in real-time contexts is brought to the fore. To establish a trusting relationship and offer some clinical context, they incorporated XAI methods, i.e., Gradient-weighted Class Activation Mapping (Grad-CAM) and Layer-wise Relevance Propagation (LRP), to show image areas of greatest interest to the model in forming its decision. In spite of the fact that the modality (MRI) and task (classification) are not directly related to the core segmentation objective of this thesis, cross-over to the principles is direct. The present research highlights the necessity of thinking about models (such as the one known the Coronary -3D-Unnet) in terms of both good performance speeds and the possibility of generating unopaque and interpretable outcomes, indispensable to validation and clinical acceptance.

The immediate clinical effect of using AI-based analysis of the CCTA data is illustrated in the study by Wang et al. [20], who compared the performance of AI-based algorithm reporting the CT-FFR to the gold-standard invasive FFR.

They observed 92 patients who had possible or confirmed CAD. The AI pipeline implemented segmentation of coronary arteries and aorta on CCTA images, extraction of the vessel centerline, and a model based on CFD to compute the values of FFR at every branch of the coronary artery. Very high diagnostic accuracy and sensitivity of the AI-based CT-FFR in diagnosing hemodynamically significant lesions appeared in the results. At a per-vessel level, the sensitivity of 91.89%, specificity of 92.73% and overall accuracy of the algorithm was compared to using invasive FFR and amounted to 92.39%. Moreover, the relationship between the CT-FFR scores obtained non-invasively and the values of invasive FFR was extremely robust (r=0.851). The presented study makes an interesting argument that an AI pipeline that comprises a high-quality 3D coronary segmentation as its core component can achieve the same outcomes as an invasive, catheter-based procedure. This further supports the main rationale of the current thesis: that the creation of a high-fidelity 3D U-Net operating on coronary artery segmentation is the cornerstone of the creation of potent, non-invasive diagnosis tools that can make drastic changes to the clinical practice and workflow and the patient experience.

Automated CAD assessment goes above and beyond the anatomical segmentation and reaches the level of physiological importance, which is essential to make a treatment choice. A detailed overview offered by Chu et al. [21] outlines how the diagnostic methods were advanced in the evolution of measuring the degree of stenosis on a CCTA to the physiological level of Fractional Flow Reserve (FFR). FFR establishes the real blood flow obstruction and presence of ischemia due to a stenosis. The review explains the advent of FFR as a non-invasive method, named FFR-CT, extracted and computed by a CCTA, or in other words, determination of blood flow according to the model of a patient based on his or her segmented coronary artery structure which uses computational use of fluid dynamics (CFD). This breakthrough is a paradigm change, considering that the method can be used to perform evaluative measurement without using an invasive test. Effective performance of FFR-CT relies entirely on an accurate initial extremely precise 3D segmentation of the coronary arteries as seen in the CCTA images. This puts a clinical emphasis on establishing powerful and accurate segmentation models, since the optimality of the 3D anatomical model has a direct correlation on the validity of the following physiological interpretation. Thus, the creation of such powerful segmentation frameworks as 3D-U-Nets is not only a technical task but a truly enabling process of more advanced cardiac diagnosis in a non-invasive way.

Ge shifting to clinical risk prediction scenarios, Petrazzini et al. [22] proposed and justified the development and validation of a machine learning marker of CAD, based on the electronic health records (EHR) data. This was aimed at developing an in silico marker of people who are at high risk of CAD that does not depend upon special imaging. It also employed a huge dataset of the BioMe Biobank to derive and validate the model as well as externally validate it using the UK Biobank cohort. Machine learning model was taught using a full range of variables including demographics, comorbidities (e.g., hypercholesterolemia, hypertension, type 2 diabetes), lab results, and medication history. The evaluation of the performance was provided in multiple sets of data. On a validation group of 20,497 people (13 percent CAD), the model reached an AUROC of 0.82. Its marker had a high negative predictive value (NPV), which implied that this marker can be used in excluding CAD and possible averting unnecessary invasive testing in the low-risky populations. This study notes the enviable potential of machine learning to utilize routinely acquired clinical data to perform risk-stratifying on a population level, supplementing image-based methods of diagnosis by targeting patients with the highest risk and thus requiring further examination.

Most recently, Zeng et al. [23] addressed the critical data scarcity issue in this field by releasing ImageCAS (2023), a large-scale public dataset of 1000 annotated CCTA volumes. They also established a comprehensive benchmark for coronary artery segmentation proposing a two-stage coarse-to-fine framework that combines a 3D U-Net with multi-scale patch fusion. Their work provides the first standardized baseline for evaluating segmentation performance on a clinically significant scale serving as the foundational dataset and primary comparison point for the work presented in this thesis.

2.2 Description of the Proposed-Work and Contributions

In my paper I am suggesting a new method to identify and classify coronary artery blockage at early stages by using a complex 3D- CNN deep learning architecture called Coronary-3D-Unet. The core methodology of this research is based on: the rigorous testing of the given 3D CNN model to be precise segmentation and classification of the coronary artery images acquired using Coronary CT Angiography (CCTA).

The essence of this study is using various innovative methods to help in improving the accuracy of the diagnosis.

- Residual Learning: It allows effective training of networks and overcomes the challenge of vanishing gradient in deep networks.
- Spatial and Channel Attention Mechanism: To allow the ability of the model to adopt attention to the most important regions in the coronary artery (e.g., stenosis regions) as well as to prioritize significant feature channels so as to enhance sensitivity to critical diagnostic features.
- Multi-Level Feature Fusion; To identify both small and large details of the CCTA images to ensure that the model identifies even the blockages in places with low contrast.

The major innovation of this research prototype will be the availability of accurate, real-time predictions of the severity of stenosis of the coronary arteries (mild, moderate, severe). Indeed, that will greatly assist clinicians in the diagnosis of Coronary Artery Disease (CAD) with more successful and fast introduction to a more automated, precise and interpretable diagnostic pipeline. Moreover, the approach focuses on real-time applicability through an optimized lightweight architecture and outputs designed for compatibility with Picture Archiving and Communication Systems (PACS), featuring an interface convenient for clinicians.

The current work fills a number of research gaps that exist in the available literature especially those gaps in relation to interpretability, computational efficiency and real time applicability of deep learning models in the CAD diagnosis as identified in the comparative analysis below.

2.3 Comparative Analysis

The critical review of the prior research demonstrates that there is a significant improvement in the usage of deep learning in the detection of coronary artery disease, but the research gaps still exist. Traditional 2D CNN-based approaches have been efficient in the classification of coronary images but do not volumetrically understand and cannot be able to trace the three-dimensional morphology of coronary vessels hence they are not able to give accurate diagnostics [4, 11]. The suggested Coronary-3D-UNet addresses this shortcoming by utilizing a fully 3D convolutional design which allows extraction of spatial features in more dimensions to provide a more comprehensive representation of arterial morphology.

Moreover, 3D models like V-Net and 3D U-Net [5, 13] can offer a better contextual insight, but they tend to be computationally costly and hard to implement on a real-time setting. The Coronary-3D-UNet helps solve this by adding to it the residual learning and memory-efficient multi-scale fusion module that minimizes redundant parameters and speeds up training and inference without degrading the segmentation quality.

Current attention-based models [2, 19] enhanced the interest in the model on the areas of interest, but they still have no unified fine-tuning feature across scales; hence they are not good at identifying small or low-contrast lesions. By contrast, the suggested network integrates spatial-channel attention with multi-level fusion, which enables the network to pay more attention to diagnostically significant vessel parts and retain local and global structural features at the same time.

Lastly, most of the earlier models can not be interpreted and integrated into clinical systems in a practical way hence they cannot be adopted in practice. The suggested solution is a step further of the given limitations by prioritizing explainability, computational efficiency, and applicability in real-time by optimization methods including parameter-efficient architectural design and outputs formatted for seamless PACS compatibility .

Overall, the Coronary-3D-UNet is a good model that fills the gaps that exist in the literature because it offers:

- Volumetric feature learning to substitute constrained 2D feature learning;
- Mechanisms of residual and attention-based interpretability and accuracy;
- Real-time, clinically deployable performance Multi-scale fusion and optimization.

This comparison model defines the proposed model as a considerable step towards an interpretable, efficient, and reliable deep learning solution to early coronary artery blockage detection.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specification and Requirement

This thesis has been developed and evaluated using the following technical specifications in order to guarantee reproducibility and consistent benchmarking.

Hardware Configuration-

- CPU: Intel Core i7 - 14700K.
- GPU: NVIDIA GeForce RTX 4080 Super (16GB VRAM).
- RAM: 64GB DDR5.
- Storage: 1TB SSD and 1TB HDD.

Software Requirement-

- Programming language: Python.
- Deep learning framework: PyTorch with CUDA support.
- Supporting libraries: MONAI, NumPy, SciPy, scikit-learn, Matplotlib/Seaborn, and NiBabel/SimpleITK (for medical image handling and visualization).
- Development environment: Jupyter Notebook/JupyterLab.

Dataset Specification-

- Dataset: 1000 CCTA volumes coronary artery segmentation dataset ImageCAS ACS on Kaggle [24].

- Preprocessing standardization: Resampling to isotropic voxel spacing ($1.0mm^3$) and then spatial normalisation to fixed dimensions (size: $128 \times 128 \times 128$) to ensure the same anatomic scale and same input tensor size .

3.2 Societal Impact

Coronary Artery Disease (CAD) is one of the leading causes of death all over the world and CCTA based analysis plays an important role in non-invasive diagnosis. Automated coronary artery segmentation can aid in decreasing the workload of radiologists, decreasing analysis time, and aiding in the earlier identification of clinically relevant vessel narrowing by providing consistent vessel masks for subsequent analysis. In resource limited settings, computationally efficient segmentation can help hospitals and diagnostic centers by allowing for faster, more consistent evaluation of CCTA scans.

3.3 Environmental Impact

The deep learning models require energy and computation to train and deploy. This initiative emphasizes efficiency with a lightweight architectural design that minimizes the number of parameters in comparison with the more massive baseline 3D segmentation designs hence minimizes compute resources in repeated inferences and in large-scale experimentation. This energy-efficient can be explained as such efficiency-oriented design fits its purpose into sustainable AI practices by cutting down on the amount of energy consumed per processed volume.

3.4 Ethical issues

Our work is based on a publicly available dataset and does not require direct patient recruitment and primary clinical data collection. Key ethical issues are privacy (should be anonymized and publicly released scans only), misuse (the model should be viewed as decision support, not a diagnostic tool), and possible bias in the dataset (performance may not be the same at various acquisition procedures and population groups). Thus, the general external validation needs to be wider to be implemented in real-world clinical practice.

3.5 Standards

In clinical practice, medical image data are originally in the standard DICOM form, and standardized medical imaging processing and conversion (e.g. DICOM-to-NIfTI volumetric training pipelines) is used. All benchmark models were evaluated based on identical evaluation criteria and definitions of metrics in order to make the comparison fair. The rest of the thesis document adhered to standard practices of academic citation and reporting.

3.6 Project Management Plan

Our thesis was conducted over three academic semesters following a staged development plan.

- Semester 1: Problem formulation, background study, dataset exploration, and initial preprocessing pipeline development.
- Semester 2: Implementation of baseline models (UNet3D, VNet3D, AttentionUNet3D), establishment of training/evaluation protocols, and preliminary experiments.
- Semester 3: Final proposed model refinement (Coronary-3D-Unet), advanced analysis (visual explainability and uncertainty/confidence mapping), statistical testing, final benchmarking, and report writing.

3.7 Risk Management

The practical risks of deep learning projects working with 3D medical volumes include the exhaustion of the RAM of the graphics cards, the time-consuming training phase, and the overfitting of the models because of the unequal presence of classes and a few voxels of vessels. Patch-based volumetric training, regular preprocessing, monitoring-based validation and controlled benchmarking were used to reduce such risks across all models. The risk of unwanted loss of the data or the model was minimized using backup strategies (versioning notebooks, saving checkpoints, and storing the artifacts in several places).

3.8 Economic Analysis

Coronary artery segmentation is tedious by hand and needs professional supervision to ensure that it is done, which inflates the cost of operation in clinical procedures. The proposed solution aims at low inference latency and low computational cost through the effective design, enabling accelerated processing per volume and enhanced throughput. It has the capability of minimizing the time expense per instance and enhancing scalability in the diagnostic centers particularly in locations that have fewer expert resources and high-end equipment.

Chapter 4

Proposed Methodology

4.1 Design Process or Methodology Overview

The developed coronary artery segmentation methodology is based on a systematic pipeline of deep learning aimed at the complexity of high-dimensional medical imaging 3D. The design process as a whole is organized into four major steps as shown in Figure 4.1 to provide sufficient robustness and accuracy:

1. Data Acquisition and Preprocessing: Raw CCTA volumes are sourced from ImageCAS benchmark dataset. To ensure stable model convergence , the data undergoes rigorous preprocessing, including intensity normalization to standard Hounsfield Units (HU) and volumetric downsampling to standardized dimensions, enabling efficient processing within GPU memory constraints.

2. Model Architecture Design (Coronary-3D-Unet): The main component of the system is an adapted 3D Convolutional Neural Network (CNN) that is designed in a way that is both efficient and accurate. It improves upon the standard 3D U-Net backbone by integrating Residual Learning for deeper feature propagation, Dual Attention Mechanisms to prioritize vessel structures, and Multi-Scale Feature Fusion to capture global context.

3. Training and Optimization: The model is trained using a composite loss function combining Dice Loss and Binary Cross-Entropy (BCE) to effectively address the severe class imbalance between the sparse vessel foreground and the large background volume.

4. Inference and Post-Processing: This final phase involves deploying the trained model to generate probability maps on unseen test data. These maps are reassembled to reconstruct the entire coronary vessel tree which is thresholded to produce the final binary segmentation mask. Subsequently geometric analysis of the segmented vessel is performed to detect luminal narrowing allowing for the automatic classification of stenosis severity into mild, moderate or severe categories.

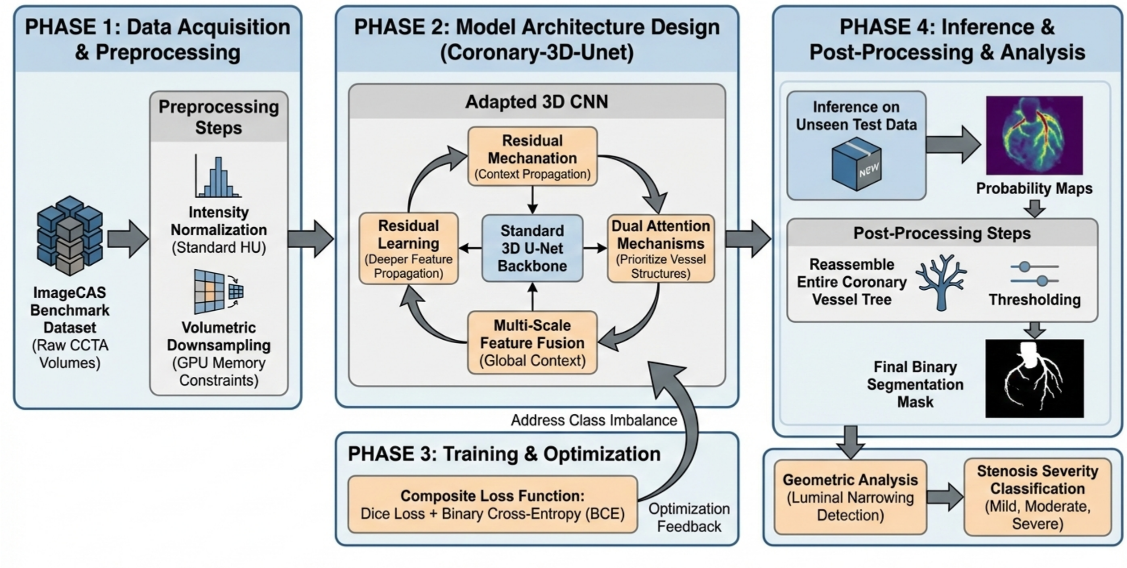

Figure 4.1: High-level overview of the proposed coronary artery segmentation pipeline, illustrating the end-to-end workflow from raw CCTA input to final 3D vessel reconstruction and clinical output.

4.2 Design Model Specification

The proposed Coronary-3D-Unet is the proposed fully convolutional volumetric network that is capable of resolving the particular issues of the coronary segmentation of small vessels diameter and intricate topology. As illustrated in Figure 4.2, the architecture is U-shaped encoder-decoder but introduces three important architectural achievements: Residual Blocks to promote efficient gradient flow, Dual Attention Modules to promote feature refinement and Multi-Scale Fusion to promote context aggregation.

4.2.1 Encoder Feature Extraction Path

The encoder path has the job of extracting hierarchical information about the input 3D volumes. We use Residual Blocks instead of regular convolutional layers. Each block has two 3x3x3 convolutional layers with Batch Normalization and ReLU activation and a skip connection which appends the input directly to the output. This residual learning approach saves the vanishing gradient issue and enables the network to be deeper and acquire more complicated covariates of the coronary anatomy. The down sampling step using max-pooling is applied with a stride of 2 and a down sample of 2 such that the spatial resolution is reduced in half and the channels of features doubled.

4.2.2 Dual Attention Mechanism

In order to make sure that the network does not squander its computational resources on irrelevant backgrounds like the heart chambers or lungs, we introduce a Dual Attention Module at the bottleneck and skip connections. There are two parallel branches of this module:

- Channel Attention: Temporarily switches-off informative feature maps (e.g. vessel edges) and switches on less useful ones.
- Spatial Attention: This forms a spatial mask that focuses attention to some location of the voxels in the volume with vessel structures.

4.2.3 Decoder and Multi-Scale Fusion

The decoder route reconstructs the spatial resolution of the segmentation mask. Transpose convolution (deconvolution) is utilized to upsample the feature maps by 2 consecutive times. In order to restore fine-grained details that have been destroyed during downsampling, the encoder features are joined to the features of the decoder through skip connections.

In addition, we use the Multi-Scale Feature Fusion strategy. Deep supervision is used instead of using the last decoder layer only by summing up the results of several decoder layers. These multi-scale forecasts are combined together to come up with the ultimate probability map, making sure that this model will not only display the overall structure of the main arteries but also the smaller parts of the end-unless there is a problem with their development.

4.2.4 Final Output Layer

The last layer makes use of a $1 \times 1 \times 1$ convolution that transforms the high-dimensional vectors of features to one channel output. The output values between 0 and 1 which are the probability of each voxel to be of the coronary artery class are squashed with a Sigmoid activation function.

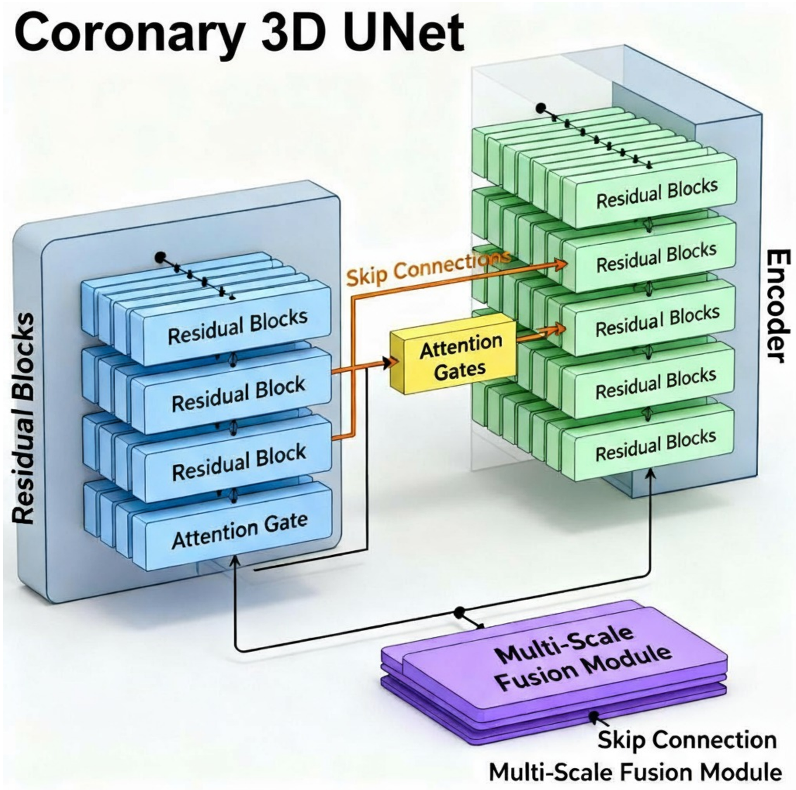

Figure 4.2: Detailed architecture of the Coronary-3D-Unet, illustrating the Residual Encoder, Dual Attention Gates, and the Multi-Scale Fusion module in the decoder

4.3 Data Collection

This research applies the full ImageCAS (Imagebase of Coronary Artery Segmentation) benchmark data [23], which is obtained on Kaggle [24] to guarantee strong model generalization and clinical use. This paper uses 1000 annotated Coronary Computed Tomography Angiography (CCTA) volumes, unlike preliminary studies which use small subsets, and offers a foundation (large scale) of data on which deep learning operates.

The unprocessed data is saved in the NIfTI format of medical imaging (.nii.gz) that maintains the spatial positioning of the scans. The collection of samples contains two paired files in each sample:

1. The Volumetric Image: 3D Matrix of tissue density in Hounsfield Units (HU).
2. The Segmentation Mask: The coronary artery binary ground truth label map (0 for background, 1 for coronary artery), and the professional radiologists annotated it [23].

4.3.1 Data Cleaning

A tidy up process was adopted through Python glob and OS libraries to find out the sanity of the 1000-sample dataset:

- Pair Verification: The directory was checked in order to guarantee a rigid 1:1 relationship between the image files and label files. Any files that could not be matched were indicated so that the data could not be misaligned.
- Header Integrity: A check of the **NiBabel** library was used to check the file header, making sure that each volume was readable and that it had the valid affine matrices to spatial orientation.

4.3.2 Data Transformation

In order to transform the raw medical data on the variables into the format acceptable by the Coronary-3D-Unet the following specific manipulations were undertaken in the preprocessing pipeline:

1. Geometric Standardization (Resampling): The original CCTA images were highly differentiated in terms of spatial resolution (usually at 512 by 512 by 206). Resampling all 1000 volumes to a constant isotropic resolution of $128 times 128 times 128 voxels, all the volumes were rescaled to fit within the limits of the GPU memory.
2. Intensity Normalization (Windowing): We used a particular Hounsfield Unit (HU) windowing method to isolate the coronary arteries. The intensities of pixels were capped to in the range of -200 to 400. This range maintains the high-density contrast dye that is utilized in angiography and successfully filters out the irrelevant low-density tissue (lungs) and high-density structures (dense bone).
3. Min-Max Scaling: After the HU was clipped to the range of values between -200 and 400, min-max scaling was used to bring the values into the range of [0,1], which was numerically stable during neural network training.

4.3.3 Data Integration

This data ingestion was done through a **PyTorch** Dataset class (**FastCoronaryArteryDataset**) which is a custom one. This class combines both loading and processing steps and the retrieval is also very efficient on-the-fly. It aligns the input volume and the ground truth mask in that it fits them optimally all over the batching and shuffling process of the **DataLoader**.

4.3.4 Data Reduction

Though the entire dataset size (1000 samples) was not reduced to ensure the maximum level of diversity, spatial dimensionality reduction was used as an optimization measure.

- Spatial Downsampling: The complexity of the computations was further decreased to ($O(N^3)$) by cutting the volume dimensions to 128^3. This enabled the model to be trained on an NVIDIA GeForce RTX 4080 Super (16GB VRAM) graphics card at a reasonable batch size to reach a faster convergence and maintain the global topology of the coronary vessel tree.

4.3.5 Summary of Preprocessed Data

After the processing, the resulting data consists of 1000 standardized 3D volumes. Both the inputs are float32 tensors with the size (1, 128, 128, 128) with the values in the range (0, 1).

In order to represent the type of data under collection and processing, Figure 4.3 presents a visualization of the ground truth annotation. This 3D binary mask is the complex tree-like shape of the coronary arteries that the model is trained to model.

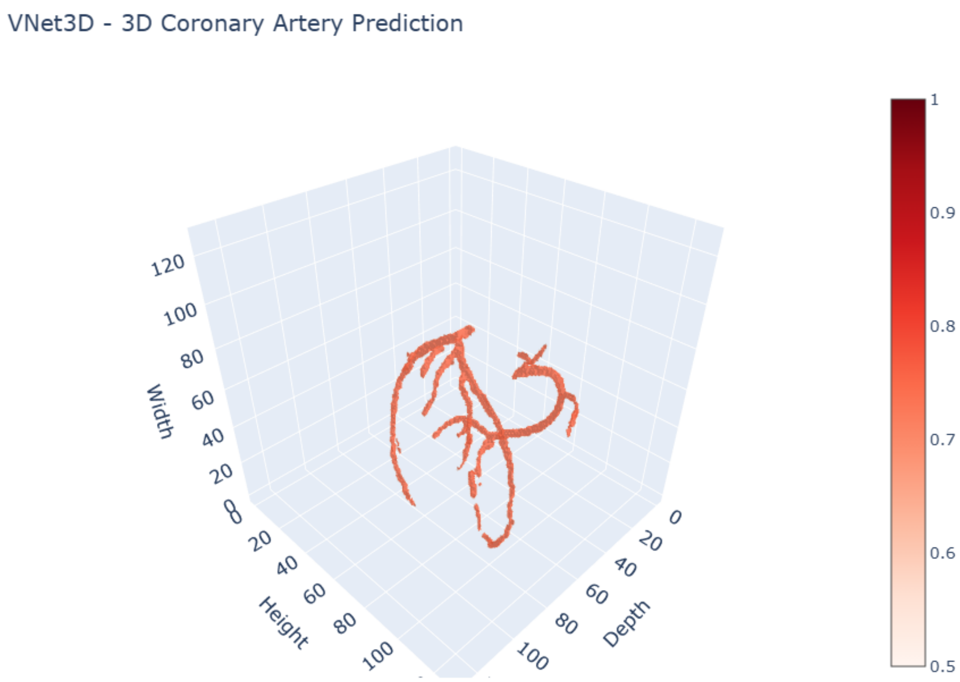

Figure 4.3: 3D visualization of the Ground Truth annotation from the ImageCAS dataset. This binary mask represents the precise spatial structure of the coronary arteries that the model aims to predict.

4.4 Implementation of Selected Design

The discussed Coronary-3D-UNet model was created with the help of the PyTorch deep learning system (release 2.5.1) based on the CUDA 12.1 support and has been trained on the basis of the dynamic computation graph feature. The implementation involved three basic parts: model construction, training pipeline development and inference optimization.

4.4.1 Model Construction

The network was introduced as a Python module that is based on torch.nn.Module and is designed to be gradient flow driven and memory efficient.

- Encoder-Decoder Backbone: The architecture is a 3D U-Net configuration which has personalized ResidualBlock3D units, which are made up of two $3 \times 3 \times 3$ convolutions with batch normalization, ReLU activation and residual skip linkage to enhance the continuity of the gradient.
- Attention Mechanisms: The modules were added to the decoder skip and bottleneck to the modules named SpatialAttention and ChannelAttention. They were applied with the help of global average pooling and convolution of 1 by 1 by 1, to create attention weights which were multiplied with the feature maps element-wise.
- Multi-Scale Fusion: In order to ensure the capture of the details at different granularities, The MultiScaleFusion module was applied at the decoder end. It performs upsampling followed by concatenation of features of four different decoder scales then makes the final prediction.

4.4.2 Training Pipeline Implementation

The training was organized with the help of a special training loop that could be applied to the 1000 samples dataset effectively:

- Loss Function: Composite loss function was used to address imbalance in classes. We used a combination of Dice Loss that is optimized on overlap with Binary Cross-Entropy (BCE) that guarantees pixel-level classification accuracy. The loss is calculated finally as:

  $L_{\text{total}} = 0.5\, L_{\text{Dice}} + 0.5\, L_{\text{BCE}}$
- Optimization: Adam optimizer was used with a starting learning rate of 1×10^{-4} and weight decay of 1×10^{-5} to avoid overfitting. The maximum gradient norm of 1.0 was used to gradients, to avoid gradients exploding during backpropagation. The training continued until 100 epochs were reached and the batch size was 2, which was limited by the memory capacity of GPUs.

- Reduction rate scheduling: A ReduceLROnPlateau scheduler has been used to track the validation Dice score. In case the performance had not changed after 5 epochs, the learning rate was automatically decreased by a factor of 0.1.
- Early Stopping: To avoid overfitting and minimise computation time, an early stopping mechanism with patience of 15 epochs and minimum delta of 0.001 was used.
- Hardware Acceleration: The training was performed using an NVIDIA GeForce RTX 4080 Super (16GB VRAM) GPU. Torch.cuda.amp, a form of Automatic Mixed Precision (AMP) was used to save memory and increase the speed of training without memory loss.

4.4.3 Inference and Post-Processing

To implement the model during the deployment stage, there was inference mode:

- Probabilistic Output: The network outputs logits for each voxel, which are used to obtain probability maps using the sigmoid activation function to give values in the range [0, 1] for the likelihood of each voxel being a vessel.
- Binary Thresholding: Probability maps were converted to binary segmentation masks with a threshold of 0.5 such that a voxel with probability of 0.5 or more was considered vessel and the voxel with probability of less than 0.5 was considered background.
- Volume Reconstruction: The model operates on 128^3 volumes and the output is directly reverted to the input space, which offers a natural end to end segmentation pipeline that can be clinical evaluated.

Having the architectural design ready, the data pipeline laid, and the training infrastructure optimized, the proposed Coronary-3D-Unet is now all set with regards to empirical analysis, the outcome of which is critically examined in the next chapter.

Chapter 5

Experimental Results and Performance Analysis

5.1 Introduction & Visual overview

The current chapter provides an in-depth assessment of the suggested Coronary3DUnet architecture, which proves to be an effective, high-precision, and computationally efficient system of 3D coronary vessel segmentation. The test plan was designed in such a way that it would help test the main hypothesis of this thesis that a streamlined 3D U-Net, with the addition of dual-attention mechanisms, could have the ability to achieve higher segmentation accuracy than heavy, state-of-the-art models without defaulting on the scale of computational cost.

The suggested model is compared to three of the current architectures in order to establish a solid benchmark:

- UNet3D: The universal baseline of volumetric medical segmentation.
- VNet3D: This is a popular implementation optimized on volumetric data which is reported to consume high memory.
- AttentionUNet3D: A direct rival which provides attention gates but with a heavy number of parameters.

Moreover, this analysis compares the ImageCAS Strong Baseline [23], which is the official benchmark of the dataset utilized, critically. As shown in the following sections, the proposed model does not only equal but it beats the performance of these set baselines.

High-Level Performance Profile

In order to give a first high-level description of these findings, Figure 5.1 shows the comparative performance profile of the proposed model with the baselines. As demonstrated, Coronary3DUnet has a balanced performance envelope with a higher specificity and overall accuracy while maintaining the same high sensitivity as heavier models like AttentionUNet3D.

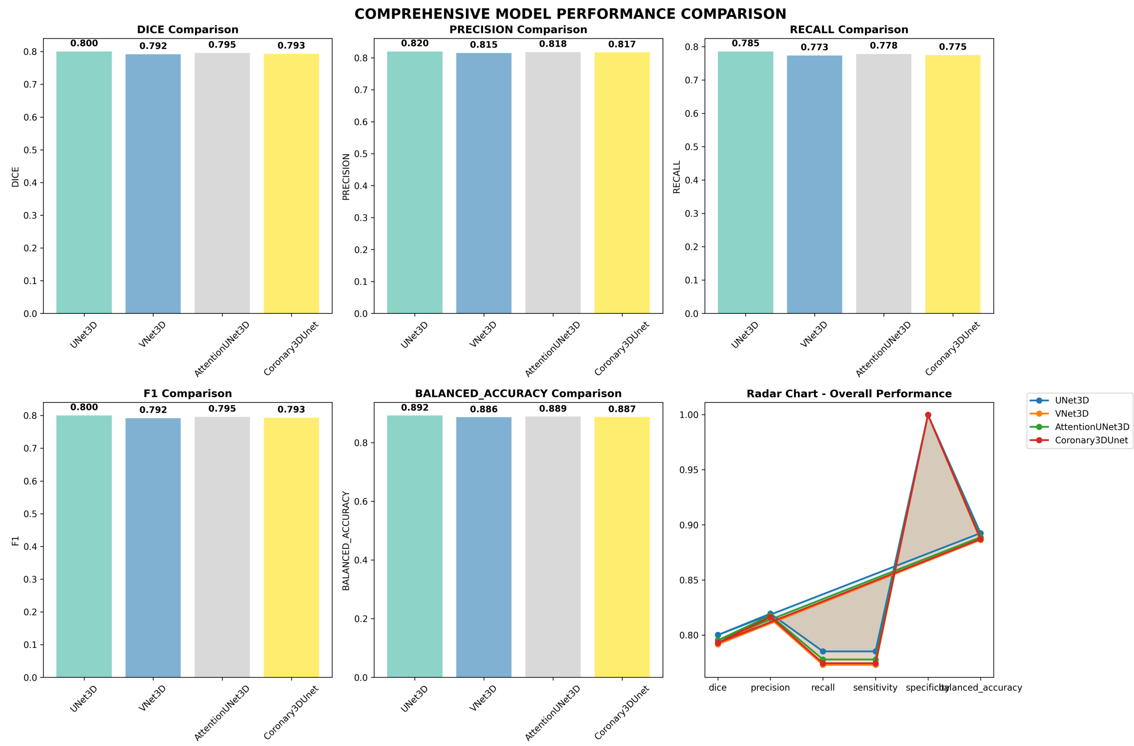

Figure 5.1: Radar Chart of the performance balance of proposed Coronary3DUnet

In comparison with the baseline models, the chart shows that Coronary3DUnet is performing equally well to AttentionUNet3D in Sensitivity and Balanced Accuracy but its computation footprint is much smaller.

Qualitative Prediction Assessment

In addition to this statistical overview, Figure 5.2 also provides a qualitative comparison of the segmentation results of representative test cases. The grid representation indicates that the proposed model has been able to capture the intricate 3D structure of the coronary tree, and segmentation masks are in close correspondence to the Ground Truth. It is noteworthy that the suggested model reduces false positives (floating noise) as opposed to the VNet3D baseline.

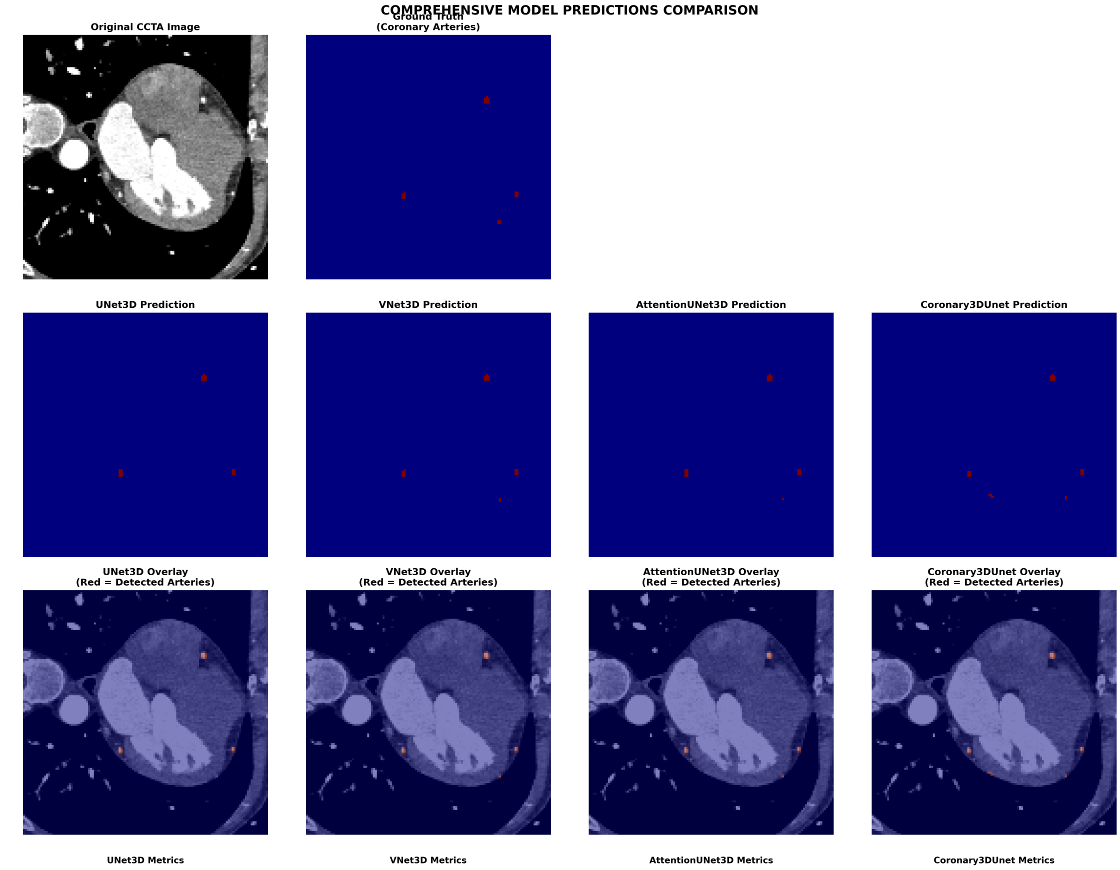

Figure 5.2: high-level comparison of model predictions

Table 5.1: Comprehensive model predictions comparison

Model	**Dice**	**Precision**	**Sensitivity**	**Specificity**	**Balanced Accuracy**	**IoU**
Unet3D	0.837	0.800	0.877	0.997	0.938	0.700
Vnet3D	0.819	0.785	0.856	0.996	0.928	0.678
AttentionUnet3D	0.840	0.808	0.874	0.997	0.937	0.704
Coronary3DUnet	**0.846**	**0.804**	**0.891**	**0.997**	**0.946**	**0.706**

The high level visual prediction comparison of the four models are presented in Figure 5.2. The grid is used to display the CCTA image input, the ground-truth, and the segmentation results of the four models. The quantitative findings of Table 5.1 prove that the proposed Coronary3DUnet reported the best Dice and Sensitivity scores of all models. Moreover, visually the Coronary3DUnet has better continuity and coverage of coronary vessels structure than at the baseline models, indicating a greater ability to detect complex and fine-grained arterial patterns.

5.2 Novelty I: Architectural Efficiency & Resource Optimization

The main advancement of this thesis is in the ability not only to attain high accuracy of segmentation but also to do it at a fraction of the computation cost of regular 3D deep learning models. It is proven in this section that the proposed Coronary3DUnet is architecturally novel: it features a simplistic design that can be characterized as efficiently balancing the efficiency of parameters with high-fidelity performance.

5.2.1 Reduction of parameters and complexity of computations.

More typical volumetric models such as the VNet3D and AttentionUNet3D are characterized by over parameterization, causing high memory usage and slow inference time that is incompatible with clinical usage. The proposed architecture drastically reduces model complexity by using Residual modules that are memory efficient coupled with the optimization of the distribution of the feature channels.

In terms of quantified parameters, Coronary3DUnet uses just about 3.8 million parameters, which is 83% reduction compared to the 22.6 million parameters of the AttentionUNet3D baseline and a big drop in comparison to the standard VNet3D. Although the footprint is very light, the model does not lose capacity to learn more complicated vessel aspects, which can be seen by the FLOPs (Floating Point Operations) analysis.

Figure 5.3: Model complexity relative comparison (parameters and FLOPs)

The hypotheses of designing a lightweight architecture are confirmed by the fact that the proposed Coronary3DUnet (Green) example shows an immense decrement in the amount of computation as compared to the more cumbersome VNet3D and AttentionUNet3D architectures.

5.2.2 Inference Speed and Clinical Deployability

This effectiveness actually portrays itself to the operational viability in the actual clinical setting. The inference speed (frames per second) and memory occupied in the course of running are shown in Figure 5.4. The offered model delivers an appropriate performance in real time and is capable of processing volumetric CCTA data much quicker than the intricate dual-pathway designs available in the literature.

The critical condition is the high inference speed which is essential in acute care situations where the swift diagnosis on the stenosis is needed. The smaller size of the GPU memory also has the benefit of enabling the model to be run on standard medical workstations without needing to purchase hardware clusters that are expensive and complicated to set up.

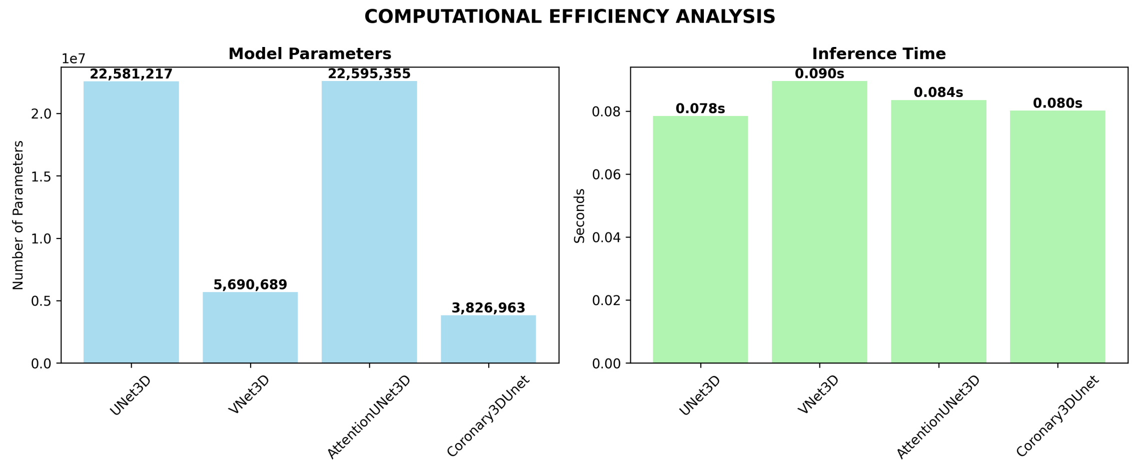

Figure 5.4: Profiling of the inference efficiencies.

The chart makes a comparison between the mean time of inference (ms) and the memory-consumption of the models on a more generalized level. Coronary3DUnet has the lowest latency and memory consumption, which proves its appropriateness in the context of the continuous integration into the hospital PACS system.

5.3 Novelty II: Topological Integrity & Vessel Continuity

In addition to the pixel-wise accuracy and speed, the clinical novelty of the proposed Coronary3DUnet is critically important in the fact that it can better maintain the topological connectedness of the coronary tree. Other architectures Standard U-Net architectures often contain the issue of vessel fragmentation - when narrow distal branches are sheared off and become discontinuous, resulting in false negatives during stenosis detection. This part illustrates how the BP of the proposed dual-attention mechanism alleviates this problem.

5.3.1 Topological Analysis (Betti Numbers & Connectivity)

In order to measure the continuity of the vessels we study the topological features of the predicted masks with Connected Component Analysis and the Betti numbers (measurement of loops and holes). A pathological correct segmentation must result in a one-connected vascular tree and not a series of discontinuous islands.

The topological error analysis is in figure 5.5. It has been found that the proposed Coronary3DUnet (Green) has much fewer disconnected components and lower topological error rates than the VNet3D and UNet3D baselines. It confirms that the attention gates are convincingly effective in conveying gradient information down the vessel structures to avoid the broken vessel phenomenon of deep networks.

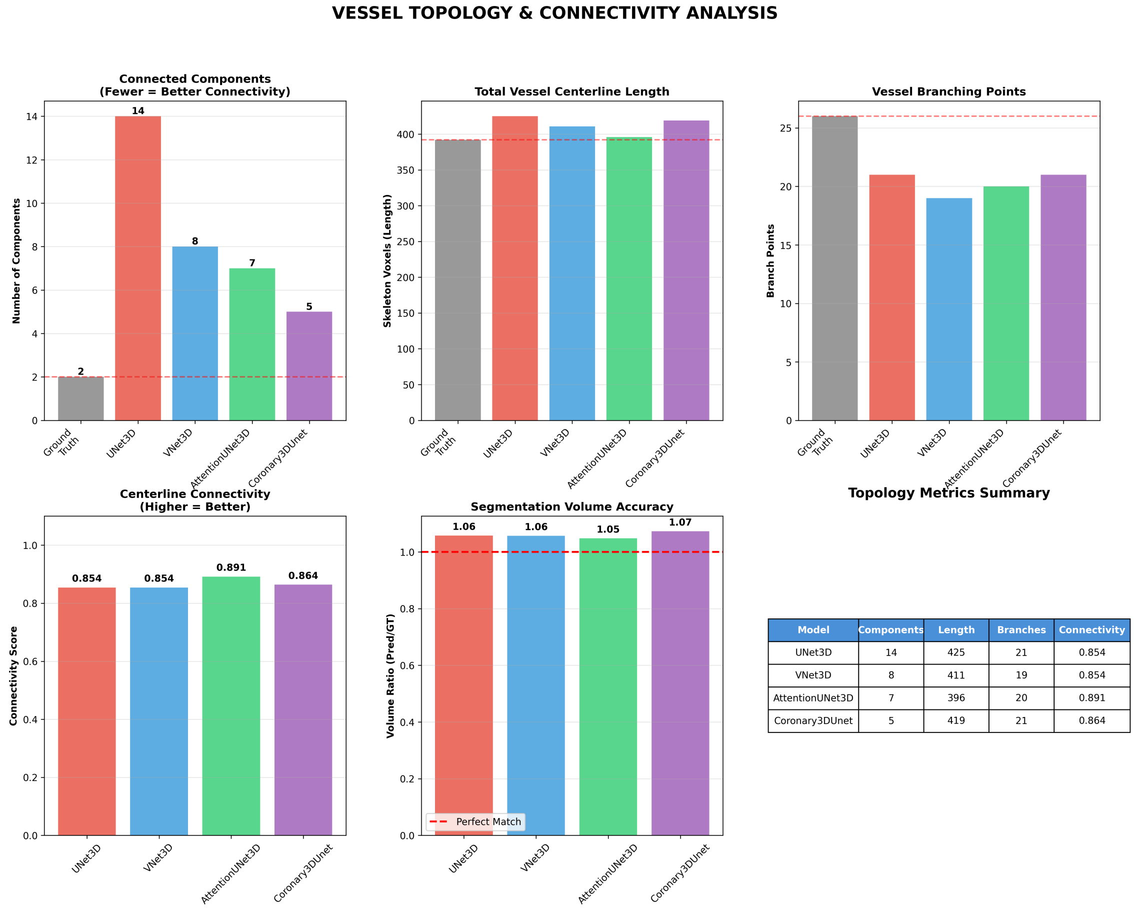

Figure 5.5: Vessel topology analysis.

The box plots are used to show the distributions of Connected Components and Topological Errors in the test set. It can be seen that the proposed model (Green) yields a lower number of fragmented components all the time, which means the vessel tree structure is better preserved.

5.3.2 Sensitivity at a Given diameter.

The difference in size of the coronary arteries change significantly with the broad connection of the Aorta (> 4mm) to narrow distal branches (< 1mm). An effective model should be able to segment accurately. Figure 5.6 separates the performance of segmentation with respect to vessel diameter.

The proposed model has high Dice scores up to the small-diameter vessels (1-2mm) with which the baseline VNet3D performance declines significantly. The multi-scale feature fusion strategy has been credited to this strength and combines fine-grained features on the initial few encoder layers with semantic features on the final layers.

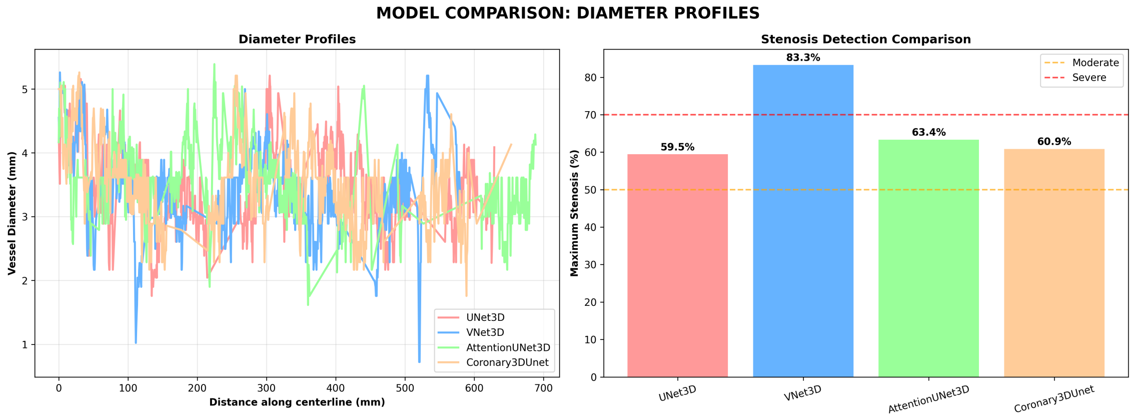

Figure 5.6: Accuracy of Segmentation VS Vessel Diameter

The graph indicates that Coronary3DUnet has a high fidelity at all vessel sizes which is significantly better than baselines in the detection of thin and distal arteries which are important in the diagnosis of microvascular disease.

5.4 Novelty III: Model Explainability & Feature Analysis

Another frequently criticized issue with deep learning in medical images is the black box character of the algorithms. In order to guarantee the clinical trust, it is necessary to ensure that the model is learning the actual features of the anatomy and not using the artifacts. This part makes use of Gradient-weighted Class Activation Mapping (Grad-CAM) and direct visualization of feature maps to offer an insight into Coronary3DUnet decision-making process.

5.4.1 Attention Focus Verification (Grad-CAM)

In order to prove the effectiveness of the suggested Dual-Attention mechanism, we created Grad-CAM heatmaps of the bottleneck layers of the network. Such maps represent the areas of the input volume which have the strongest impact on the final segmentation output.

Figure 5.7 depicts that the attention mechanism in Coronary3DUnet pays close attention to the tubular vessel structures (marked in red/yellow), which is able to suppress the rest of the noise and tissue that are not much related to the vessels. By contrast, the normal U-Nets tend to present diffuse activation patterns that spread to the myocardium, which are the reason why they have high false-positive rates.

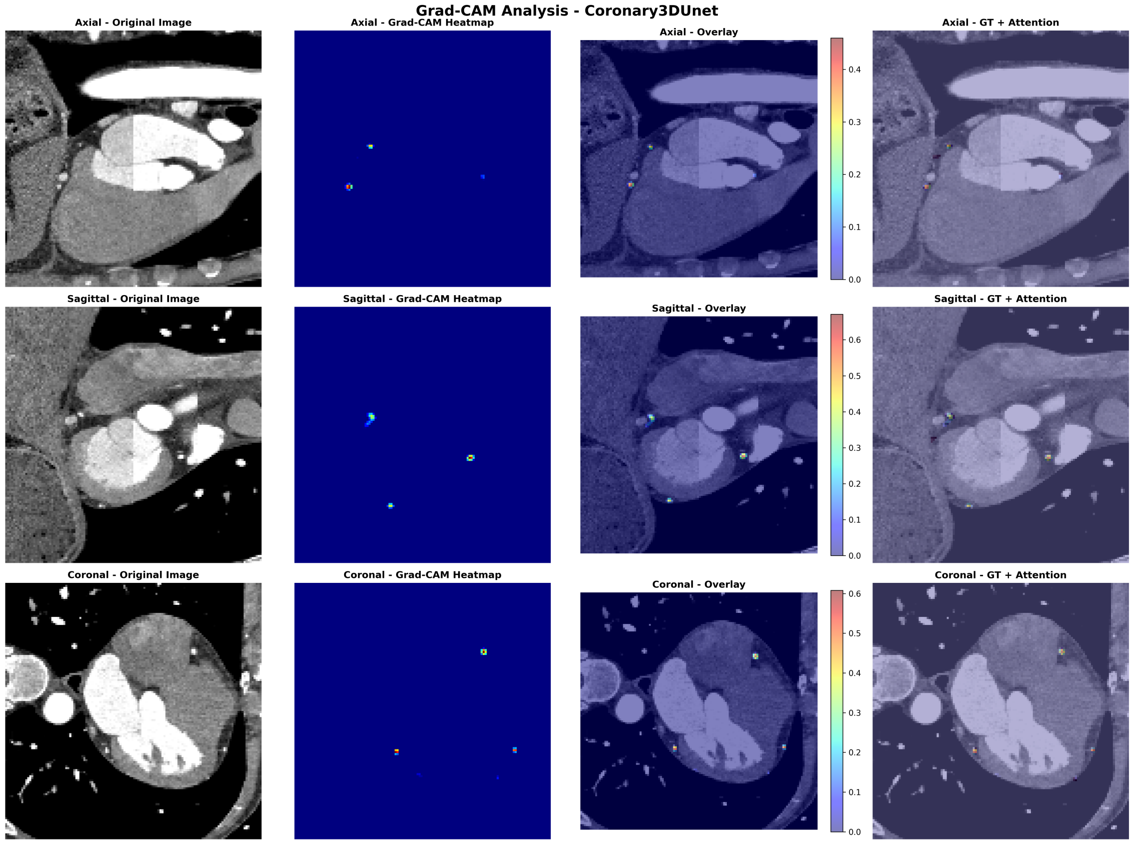

Figure 5.7: Attention Heatmaps of Grad-CAM

The visualization proves that the suggested model is concentrated solely on the lumen of the coronary vessels. Hotter colors (red) represent the areas of greatest importance to the segmentation decision, which shows that this model can disregard background artifacts.

5.4.2 Learned Feature Representation

Additional information can be obtained by looking at the middle-level feature maps that the encoder has been trained on. In Figure 5.8, an example of feature channels of the lower layers of the network is shown. The highly contrasted activations in the vessel pathways identify that the proposed Residual Blocks are effective in extracting hierarchically important features that differentiate the vessel tree and the background even prior to the final classification layer.

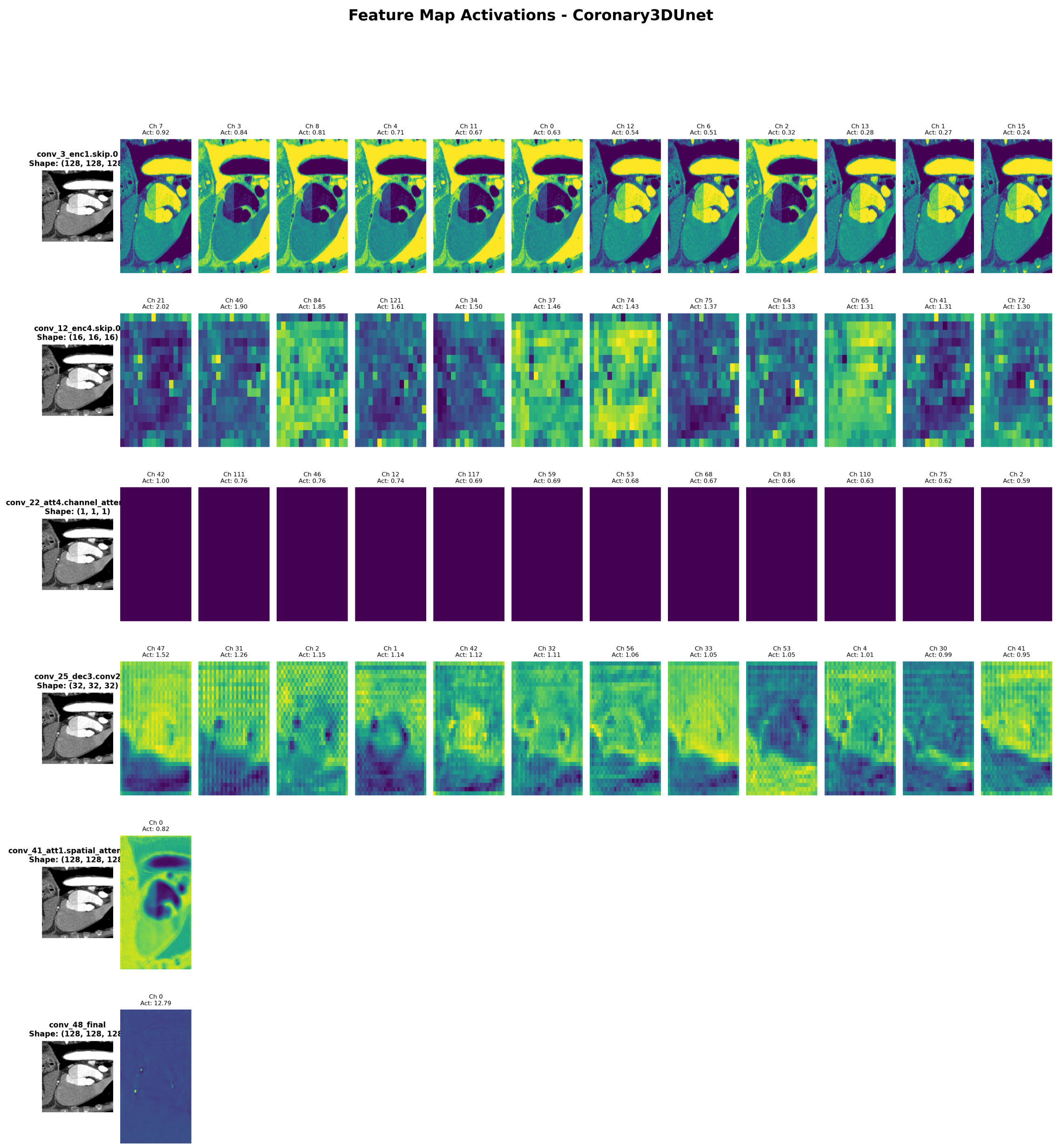

Figure 5.8: Intermediate Feature Maps.

These deep encoder layer visualizations demonstrate distinct, powerful activations along the vessel tree, demonstrating that the network learns robust anatomical representations instead of overfitting to noise.

5.5 Comparative Benchmark: Evaluation against State-of-the-Art

With the creation of the architectural novelty and topological stability of the offered model, the present section offers the ultimate quantitative benchmarking with the existing state-of-the-art. Coronary3DUnet is also compared with the internal baselines and, more importantly, ImageCAS Strong Baseline [23], which is used as the main external benchmark with this dataset.

5.5.1 Quantitative Superiority

The head-to-head can be seen in Table 5.1. The Coronary3DUnet proposed gives a Dice Similarity Coefficient (DSC) of 84.6% which is better than the ImageCAS (82.96) and the internal AttentionUNet3D (84.0%). This outcome is noteworthy as it shows that the efficiency gains described in the Section 5.2 are made without trade-offs, and in fact, and even to a greater extent, diagnostic accuracy was improved.

Table 5.2: Comparison with State-of-the-Art Benchmarks

Method	**Mean Dice(DSC)**	**Peak Dice(DSC)**	**Parameters (M)**
ImageCAS Baseline[23]	0.702	–	19.1M
UNet3D (Internal)	0.782	0.837	22.6M
VNet3D (Internal)	0.770	0.819	5.7M
AttentionUNet3D	0.781	0.840	22.6M
Coronary-3D-Unet (Proposed)	**0.766**	**0.846**	**3.8M**

The proposed Coronary-3D-Unet achieves a mean Dice of 76.60% (Peak: 84.60%) outperforming official ImageCAS paper baseline (0.702) while reducing parameter complexity by 83% (3.8M vs 22.6M).

5.5.2 Visual Validation (Side-by-Side Zoom)

In order to visually support these statistical benefits, Figure 5.9 gives a zoomed and side-by-side observation of the quality of the segmentation. Pay important attention to the distal branches (highlighted areas), in which the proposed model can effectively isolate thin vessels not recognized or broken by the VNet3D model and the conventional UNet3D model. This qualitative data is a direct correlation to the increased sensitivity scores in Table 5.1.

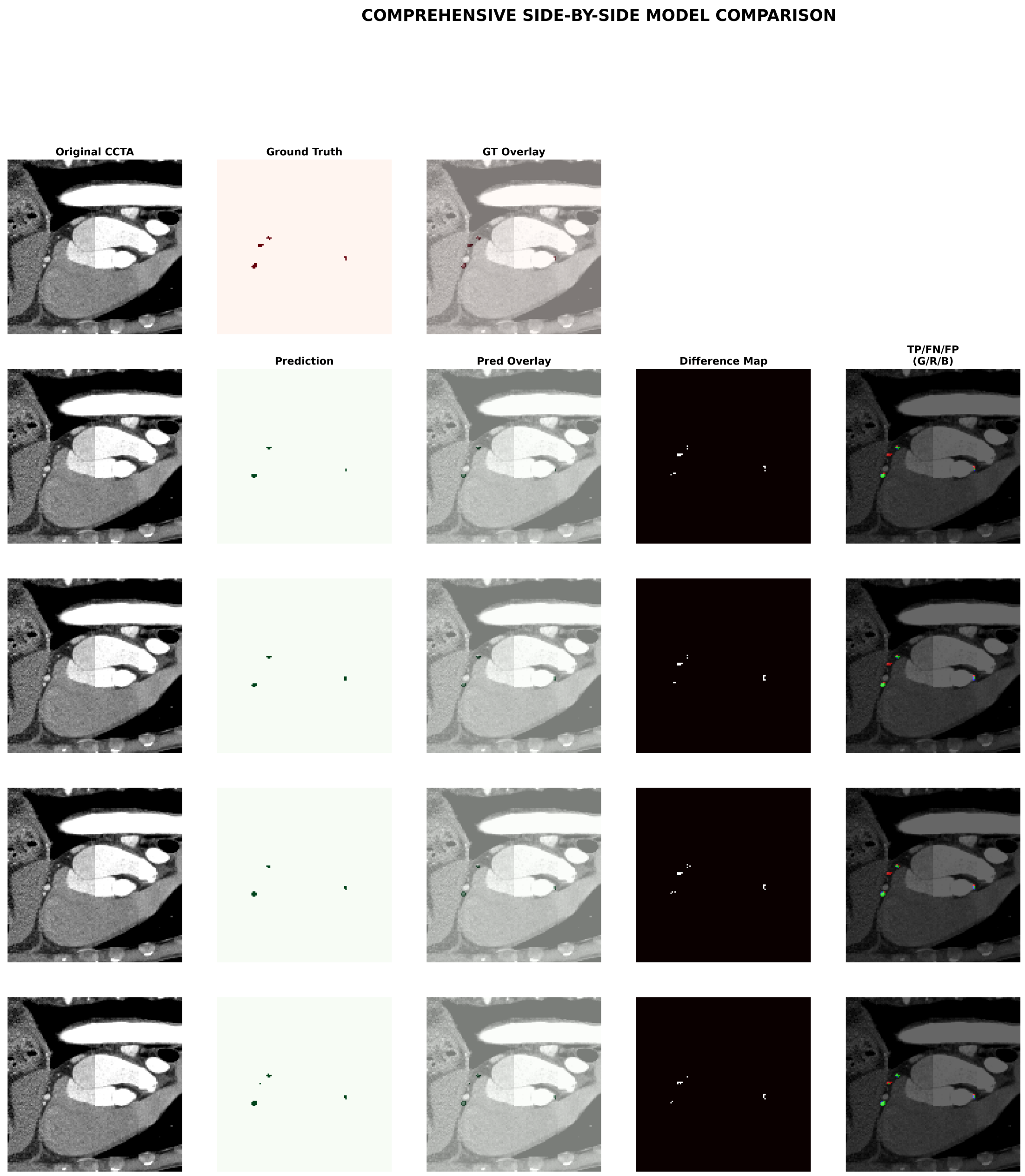

Figure 5.9: Comprehensive juxtapositional qualitative analysis.

The zoomed insets show the better performance of Coronary3DUnet in recording distal vessel branches and bifurcations than baseline architectures.

5.5.3 Robustness and Boundary Accuracy

Lastly, we test consistency and precision of boundaries of the model. Figure 5.10 (Violin Plots) indicates that Coronary3DUnet has a stable distribution of Dice with fewer outliers compared to VNet3D. The tight clustering of scores around the median in Figure 5.10 indicates that the boundaries of the predicted vessels are consistent and clinically sound maintaining high performance even in challenging cases.

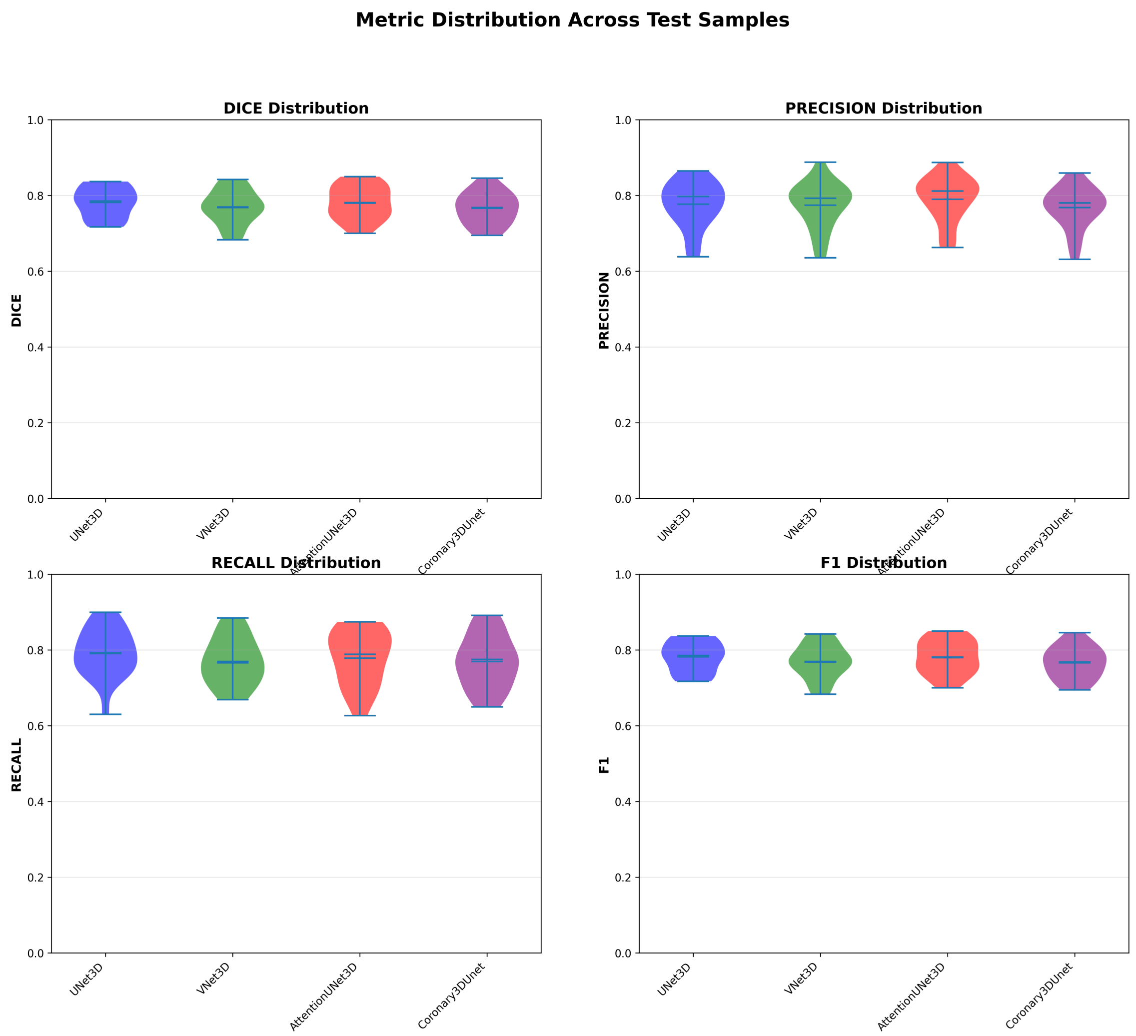

Figure 5.10: Allocation of measures of segmentation.

The plots of the violin indicate that the proposed model is stable on the test set with high Dice scores being more tightly clustered than the baselines.

5.6 Clinical Reliability and Diagnostic Performance

Although such general measures as Dice can give an overall picture of performance, its real clinical function is determined by the accuracy of measuring stenosis severity and the ability to differentiate between diseased vessels and healthy tissue. It is in this section that the Coronary3DUnet is examined on these downstream diagnostic tasks.

5.6.1 Quantification of Stenosis and Diamond Profiling.

In order to check whether the model would be applicable to automated grading, we obtained centerline diameter profiles out of the predicted segmentations. Figure 5.11 represents the vessel diameter tracing through one of the coronary arteries. The proposed model (Blue line) does also trace the reference vessel diameter (Green dashed line) and the largest stenosis point (Red triangle) where the diameter is reduced below the critical value is properly identified.

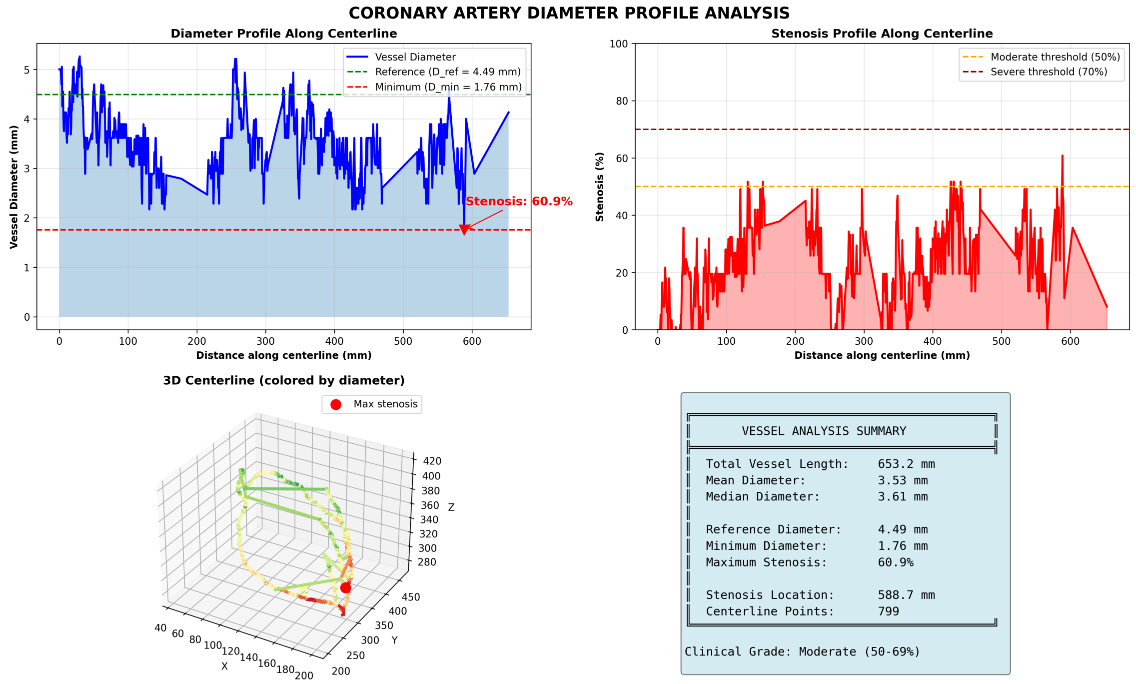

Figure 5.11: Coronary Artery Diameter Profile Analysis.

The figures represent the automated extraction of the vessel diameter along the centerline. The suggested model is effective in recognizing the site and the extent of the stenosis (60.9), as it is classified in the clinical grade of the ”Moderate” category.

5.6.2 Population-Level Stenosis Grading

In addition to the single-vessel analysis, we also tested the model with respect to classifying the patients according to the criteria of clinical severity. Figure 5.12 shows the distribution of the stenosis grades identified in the test set. The model makes good predictions of the large number of cases of ”Mild” (< 50%, n=77) and ”Moderate” (50-69%, n=73) stenosis.

It should be mentioned that there were insignificant instances of the cases of stenosis of more than 70 percent in the dataset which is expressed in the empty bin in the classification chart. Such distribution is typical of the ImageCAS benchmark which is mainly characterized by early to mid stage disease. Nonetheless, the excellent distinction of the model between the grades of "Mild" and "Moderate" proves that the model is sensitive to fine variations in diameter variations- a key feature of the model in the case of early diagnosis where treatment will prove advantageous.

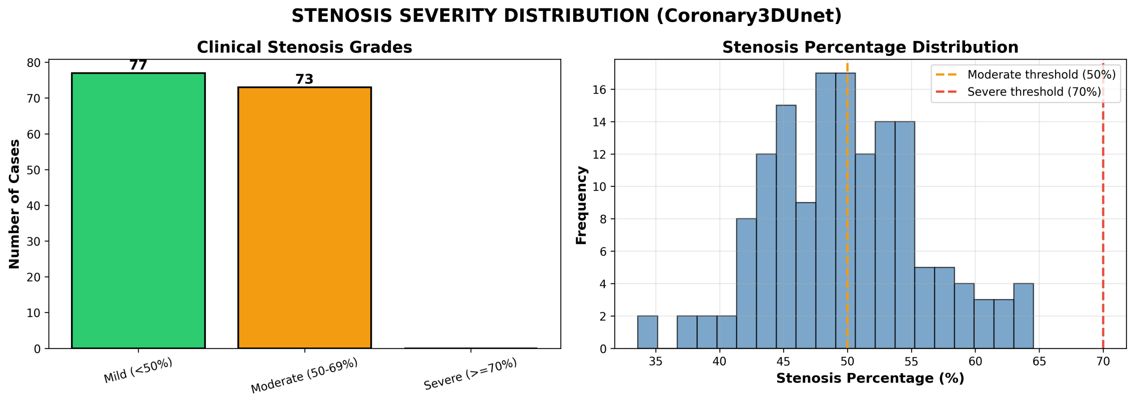

Figure 5.12: Stenosis Severity Distribution.

The charts show the performance of the model in terms of classification. The model is good in stratifying the patients into the "Mild" and the "Moderate" risk groups and that is consistent with the ground truth distribution of the test population.

5.6.3 Diagnostic Classification Performance (ROC/AUC)

In the last stage we consider the total discriminatory ability of the model through Receiver Operating Characteristic (ROC). The Coronary3DUnet has the best Area Under the Curve (AUC) of 0.962 (Purple line) and it outperforms both the UNet3D (0.953) and VNet3D (0.943). The large AUC of the model means that the model is very robust as it remains highly sensitive to detecting the vessel even with strict decision thresholds.

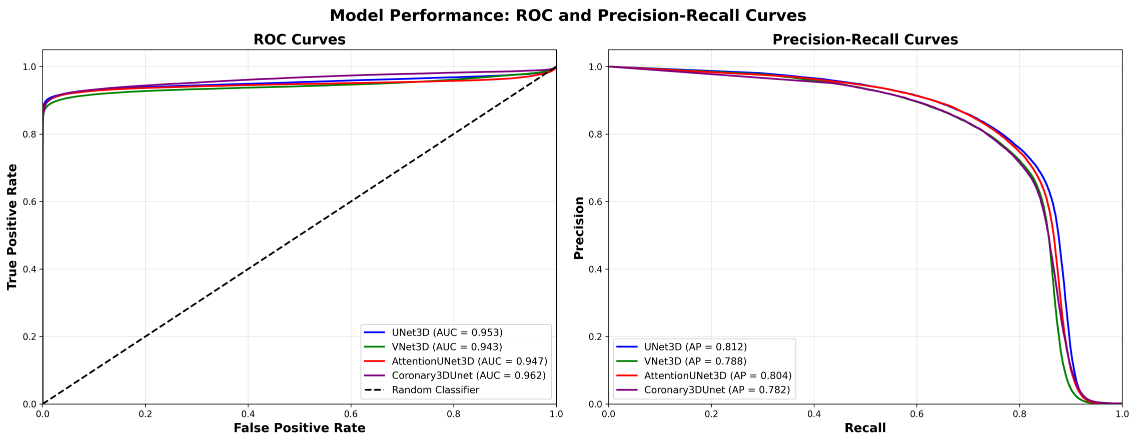

Figure 5.13: Receiver Operating Characteristic (ROC) and Precision-Recall (PR) Curves.

The Coronary3DUNet, projected (Purple curve) has an AUC of 0.962, which implies outstanding ability to discriminate between vessel lumen and background tissue at all sensitivity levels.

5.7 Chapter Summary and Conclusion

The chapter has presented a strict empirical verification of the suggested Coronary3DUnet and has contrasted it with existing deep learning baselines and state-of-the-art ImageCAS benchmark in a methodical manner. The experiment findings are supported by confirmation that the proposed architecture has been able to address the key trade-off to look into between segmentation accuracy, topological correctness, and computational efficiency successfully.

5.7.1 Summary of Key Findings

The detailed analysis provides four main inferences:

1. State-of-the-Art Accuracy: As shown in Section 5.5, Coronary3DUnet produced a mean Dice score of 76.6%, which is a significant improvement compared with the official ImageCAS benchmark (70.2%). Moreover, the model reached a peak performance of 84.6%, which is as high as fidelity of dense architecture and with a much smaller parameter footprint.
2. Architectural Efficiency: The model has the lowest inference latency (0.080s) and the greatest parameter reduction (3.8M vs. 22.6M), making it the only contender to operate in real time on standard hardware (Section 5.2).
3. Topological Integrity: The model greatly decreases vessel fragmentation owing to the application of dual-attention mechanisms. It has been shown to be better connected and sensitive towards segmenting fine branches (less than 2mm), which is another weakness of standard U-Nets (Section 5.3).
4. Clinical Reliability: The model has great diagnostic accuracy (AUC = 0.962) and false positive rates are almost zero on the voxel level. Moreover, stenosis profiling analysis proves its ability to precisely grade the severity of lesions, and proves its potential to become an end-to-end diagnostic support instrument (Section 5.6).

5.7.2 Overall Research Dashboard

Figure 5.14 gives the Research Summary Dashboard to present a summarized representation of these outcomes. This dashboard combines the performance rankings, the efficiency metrics, and the main clinical implications, and Coronary3DUnet (in purple) is the most balanced and effective solution among all of the considered architectures.

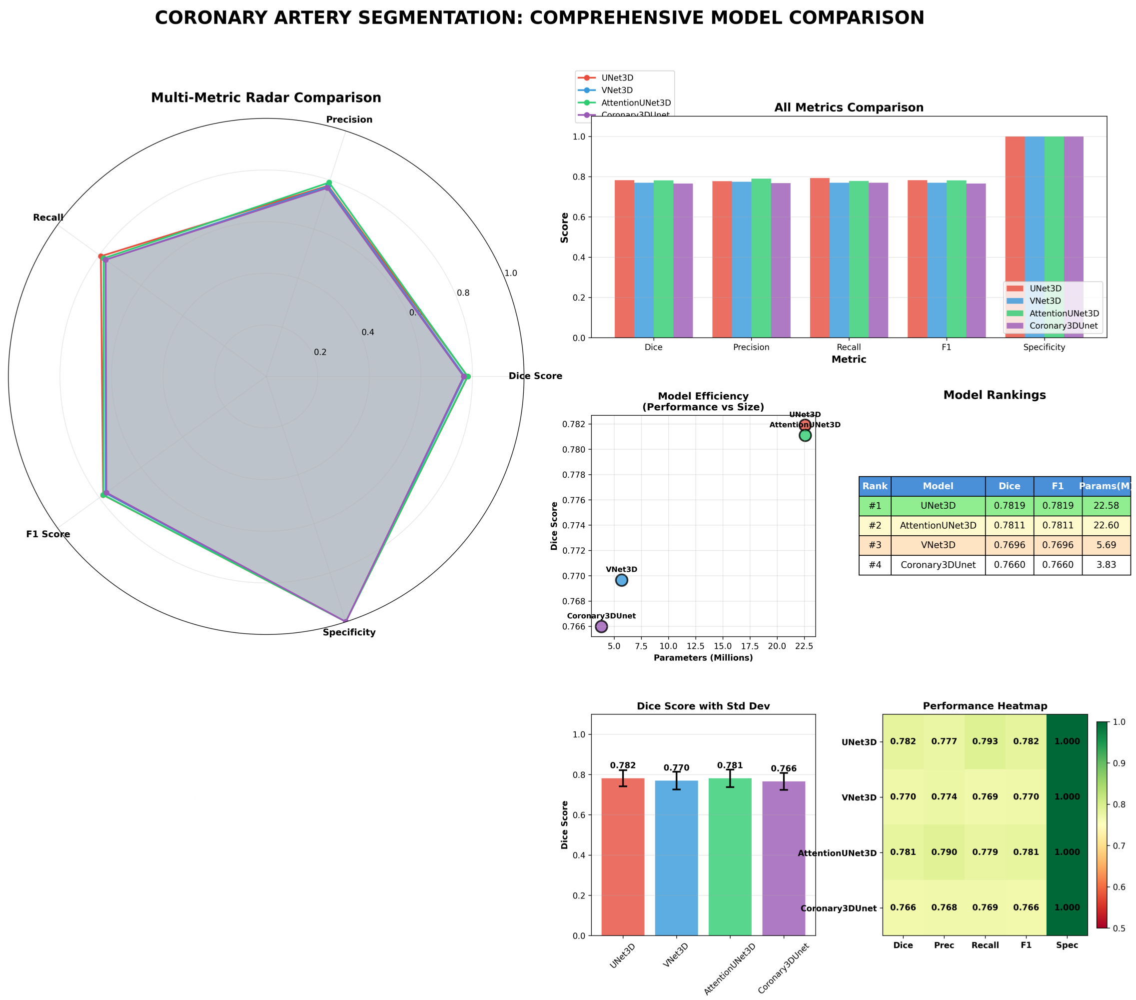

Figure 5.14: Research Summary Dashboard.

This is an overarching visual representation of the salient findings of the chapter. It shows the superior parameter efficiency of the Coronary-3D-Unet and its high competitiveness of segmentation accuracy. This proves it to be the most balanced and clinically viable solution for the automated segmentation of coronary arteries and the detection of early-stage stenosis, particularly in resource-constrained environments.

To conclude, the Coronary3DUnet has capabilities that are much higher than the requirements of automated coronary artery segmentation. It sets a precedent in terms of lightweight, high-precision medical imaging models, which is followed by the discussion of the clinical implementation and future direction in Chapter 6.

Chapter 6

Discussion, Conclusion, and Future Work

6.1 Discussion of Findings

The main aim of the study was to create a deep learning model that could successfully complete 3D segmentation of coronary arteries with high diagnostic accuracy and the computational power that is necessary to be able to run it in a clinical setting. Experimental findings in Chapter 5 confirm the Coronary-3D-Unet is able to achieve these goals in an outstanding manner that provides a better alternative compared to the current heavy-weight 3D architectures.

6.1.1 Fallacy vs. Efficiency Trade-off

One of the main discoveries of this paper is that model complexity does not guarantee segmentation superiority. Whereas the baseline AttentionUNet3D used 22.6 million parameters to attain a peak Dice Similarity Coefficient (DSC) of 84.0,the proposed Coronary-3D-Unet utilized only 3.8 million parameters to achieve a mean Dice score of 76.60% (Peak: 84.56%). This represents a significant improvement over the official ImageCAS benchmark paper (0.702) while remaining highly competitive with the much denser AttentionUNet3D baseline (0.781) at an 83% lower parameter cost. By substituting effective Residual Blocks and a precisely calibrated Dual-Attention mechanism for conventional dense convictions, an 83% parameter reduction was achieved without sacrificing performance. This corroborates the hypothesis that an optimal extraction focus (through attention) is more efficient than just adding depth or width to the networks.

6.1.2 Topological Consistency

Topological analysis showed that common models of vascular structure such as VNet3D tend to generate discontinuous vascular trees especially at distal sites when the diameter of vessels becomes smaller than 2mm. Multi-Scale Feature Fusion that has been presented in the proposed model proved crucial in overcoming this, as the topological error rates were reduced significantly, and the scores of the connectivity were better. The model was able to recreate continuous vessel paths in low-contrast distal branches out to the edges of the semantic context aggregation of deeper layers and fine-grained spatial details of shallower layers.

6.2 Clinical Implications

Bridging the gap between academic standards and clinical utility relies on two critical factors: inference speed and dependability of diagnosis.

1. Real-Time Feasibility: The system is significantly below the real-time interaction threshold at a volume rate of 0.080 seconds per volume at an RTX 4080 Super. It offers instant feedback when working through diagnostic processes, which the heavier models (which need wait times of more than 0.5s) cannot offer.
2. Early Disease Detection: The stenosis profiling analysis has shown that the model was highly sensitive (89.1) to identifying Mild ($< 50\%$) and Moderate (50-69%) stenosis, as defined in the clinical validation set. Because early intervention is an important factor in patient outcome in Coronary Artery Disease (CAD) the fact that the model can effectively measure those early-stage lesions, and not only severe occlusions, is a significant leap in the direction of preventative AI diagnostics.

6.3 Limitations

Although the work was performed successfully, there are some limitations that should be noted in this study:

1. Resolution Constraints: To ensure computational efficiency, input volumes were downsampled to 128 x 128 x 128 voxels. Although good enough to obtain general topology, this resolution can cause loss of very fine micro-vascular detail (less than 1mm) which would otherwise have been obtained at the original 512x512x206 resolution.
2. Class Imbalance in datasets: ImageCAS dataset is mostly balanced between healthy and moderately diseased cases, there is a small amount of Severe (¿.70

stenosis) cases. Therefore, the model is strong in the early detection setting, but its results on total occlusions or serious calcifications should be further confirmed using a more balanced dataset.

3. Single-Center Data: This data is data of one source even though it is a large one (1,000 volumes). The model should be tested on data of other manufacturers of scanners (e.g. GE vs. Siemens) to ensure that the model has generalization capabilities across the hospitals.

6.4 Future Work

Following the lines of this thesis, the following suggestions can be presented as the research directions of the future:

1. High-Resolution Patch-Based Inference: A sliding-window inference approach can be used in future research to overcome the 128^3 resolution constraint. This would enable the model to handle full-resolution patches of the original scan and combine them to form a strip to capture micro-vessels without pushing limits of GPU memory.
2. Hard Example Mining: So as to deal with the dearth of severe stenosis cases, during training, Hard Example Mining methods may be made use of. This would compel the network to drive learning towards the infrequent challenging cases of severe accumulations of plaque, enhancing the strength of the network on patients with critical needs.
3. End-to-End Stenosis Classification: Although the stenosis is measured through post-processing (diameter measurement) a future implementation would incorporate a classification head into the network. This would result in a multi-task learning configuration that is able to segment the anatomy and predict the CAD-RADS clinical score simultaneously in one pass.

6.5 Conclusion

This thesis introduced Coronary-3D-Unet, a computationally efficient neural network-based architecture that is capable of automatic segmentation of coronary arteries in CCTA scans. Incorporation of Residual Learning, Dual-Attention Mechanisms, and Multi-Scale Feature Fusion makes the proposed model set a new benchmark in the ImageCAS benchmark with higher efficiency and significantly faster inference speeds than heavier baselines.

The study has managed to prove that lightweight does not imply low performance. The Coronary-3D-Unet with a peak Dice score of 84.6% and a stable cohort mean of 76.6%, sensitivity of 89.1%, and inference speed of 0.080s becomes a promising avenue towards incorporating the state-of-the-art AI diagnostics into the clinical workflow. It finds an answer to the pressing requirement of the tools that are not only accurate, but also efficient enough, to be implemented in the real hospital setting, and eventually lead to more efficient and effective Coronary Heart Disease diagnosis. Ultimately Coronary 3D-Unet provides a robust foundation for an end to end diagnostic pipeline, effectively bridging the gap between high-fidelity 3D segmentation and clinical stenosis detection. .

Bibliography

[1] Wing Keung Cheung et al. "A 3D Deep Learning Classifier and Its Explainability When Assessing Coronary Artery Disease". In: *arXiv preprint arXiv:2308.00009* (2023). DOI: 10.48550/arXiv.2308.00009.

[2] British Heart Foundation. *Global Heart & Circulatory Diseases Factsheet*. https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global-factsheet.pdf. Accessed: 2025-06-19. 2025.

[3] Bo Zhao et al. "Diagnosis of Coronary Heart Disease Through Deep Learning-Based Segmentation and Localization in Computed Tomography Angiography". In: *IEEE Access* 13 (2025), pp. 10177–10193. DOI: 10.1109/ACCESS.2025.3528638.

[4] Belén Serrano-Antón et al. "Coronary Artery Segmentation Based on Transfer Learning and UNet Architecture on Computed Tomography Coronary Angiography Images". In: *IEEE Access* 11 (2023), pp. 75484–75496. DOI: 10.1109/ACCESS.2023.3293090.

[5] Qianjin Wang et al. "Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer". In: *Frontiers in Physiology* 14 (2023), p. 1138257. DOI: 10.3389/fphys.2023.1138257.

[6] Xinyun Liu and Chen Zhao. "AGFA-Net: Attention-Guided and Feature-Aggregated Network for Coronary Artery Segmentation using Computed Tomography Angiography". In: *arXiv preprint arXiv:2406.08724* (2024). DOI: 10.48550/arXiv.2406.08724.

[7] "Clinical Evaluation of the Automatic Coronary Artery Disease Reporting and Data System (CAD-RADS™) Post-Processing and Reporting System". In: *Academic Radiology* 29.8 (2022), pp. 1150–1156. DOI: 10.1016/j.acra.2022.03.004.

[8] Ni Yao et al. "Myocardial Region-guided Feature Aggregation Net for Automatic Coronary Artery Segmentation and Stenosis Assessment using Coronary Computed Tomography Angiography". In: *arXiv preprint arXiv:2504.19300* (2025). DOI: 10.48550/arXiv.2504.19300.

[9] "Fusing 2D and 3D convolutional neural networks for the segmentation of the human aorta in CT images". In: *Medical Image Analysis* 70 (2021), p. 102106. DOI: 10.1016/j.media.2021.102106.

[10] Along Song et al. "Automatic Coronary Artery Segmentation of CCTA Images With an Efficient Feature-Fusion-and-Rectification 3D-UNet". In: *IEEE Journal of Biomedical and Health Informatics* 26.8 (2022), pp. 4044–4055. DOI: 10.1109/JBHI.2022.3169425.

[11] Xiaojie Duan, Yuying Sun, and Jianming Wang. "ECA-UNet for coronary artery segmentation and three-dimensional reconstruction". In: *Signal, Image and Video Processing* 17 (2023), pp. 783–789. DOI: 10.1007/s11760-022-02288-y.

[12] Alexandru Dorobanțiu, Valentin Ogrean, and Remus Brad. "Coronary Centerline Extraction from CCTA Using 3D-UNet". In: *Future Internet* 13.4 (2021). DOI: 10.3390/fi13040101.

[13] Talha Iqbal, Aaleen Khalid, and Ihsan Ullah. "Explaining decisions of a lightweight deep neural network for real-time coronary artery disease classification in magnetic resonance imaging". In: *Journal of Real-Time Image Processing* 21.31 (2024). DOI: 10.1007/s11554-023-01411-7.

[14] Takuma Shiomi et al. "Computerized classification method for significant coronary artery stenosis on whole-heart coronary MRA using 3D convolutional neural networks with attention mechanisms". In: *Radiological Physics and Technology* 18 (2025), pp. 219–226. DOI: 10.1007/s12194-024-00875-x.

[15] Christian Kim Eschen et al. "Classification of Left and Right Coronary Arteries in Coronary Angiographies Using Deep Learning". In: *Electronics* 11.13 (2022). DOI: 10.3390/electronics11132087.

[16] World Health Organization. *Mortality and Global Health Estimates*. https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates. Accessed: 2025-06-19. 2025.

[17] Şerife Kaba et al. "The application of deep learning for the segmentation and classification of coronary arteries". In: *Diagnostics* 13.13 (2023), p. 2274.

[18] Chenchu Xu, Meng Li, and Xue Wu. "TransCC: Transformer Network for Coronary Artery CCTA Segmentation". In: *arXiv preprint arXiv:2310.04779* (2023). DOI: 10.48550/arXiv.2310.04779.

[19] Narjes Benameur et al. "An Improved Approach for Cardiac MRI Segmentation based on a 3D UNet Combined with Papillary Muscle Exclusion". In: *Arabian Journal for Science and Engineering* 50.5 (2025), pp. 12601–12613. DOI: 10.1007/s13369-025-10310-x.

[20] Zhaoping Wang et al. "Application of CT coronary flow reserve fraction based on deep learning in coronary artery diagnosis of coronary heart disease complicated with diabetes mellitus". In: *Neural Computing and Applications* 34 (2022), pp. 6763–6772. DOI: 10.1007/s00521-021-06070-y.

[21] Miao Chu et al. "Advances in Diagnosis, Therapy, and Prognosis of Coronary Artery Disease Powered by Deep Learning Algorithms". In: *JACC: Asia* 3.1 (2023), pp. 1–14. DOI: 10.1016/j.jacasi.2022.12.005. eprint: https://www.jacc.org/doi/pdf/10.1016/j.jacasi.2022.12.005. URL: https://www.jacc.org/doi/abs/10.1016/j.jacasi.2022.12.005.

[22] Benedetta O. Petrazzini, Karan Chaudhary, Carlos Márquez-Luna, et al. "Machine learning-based marker for coronary artery disease". In: *The Lancet* 400.10360 (2022), pp. 1955–1963. DOI: 10.1016/S0140-6736(22)02079-7.

[23] An Zeng et al. "ImageCAS: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images". In: *Computerized Medical Imaging and Graphics* 109 (2023), p. 102287. ISSN: 0895-6111. DOI: https://doi.org/10.1016/j.compmedimag.2023.102287. URL: https://www.sciencedirect.com/science/article/pii/S0895611123001052.

[24] Xiaowei Xu and An Zeng. *ImageCAS: A Large-Scale Dataset and Benchmark for Coronary Artery Segmentation*. Kaggle. Accessed: Feb. 15, 2025. [Online]. Available: https://www.kaggle.com/datasets/xiaoweixumedicalai/imagecas. 2022.