



Paper Type: Original Article

Development of a Deep Learning-Based Intelligent Model for the Detection of Coronary Artery Stenosis in Angiographic Images

Kamran Balani^{1*}, Rasoul Ketabi¹, Ahmad Dashti²

¹ Department of Computer Engineering, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran; kamran.balani@iau.ac.ir; rasoul.ketabi@iau.ac.ir.

² Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran; ahmad.dashti@iau.ac.ir.

Citation:

Received: 16 June 2025

Revised: 25 August 2025

Accepted: 05 November 2025

Balani, K., Ketabi, R & Dashti, A. (2025). Development of a deep learning-based intelligent model for the detection of coronary artery stenosis in angiographic images. *Annals of healthcare systems engineering*, 2(4), 250-261.

Abstract

Coronary Artery Disease (CAD) is one of the leading causes of global mortality, and its diagnosis primarily relies on the visual interpretation of angiographic images by specialists—a process that is both time-consuming and prone to human error. This study aims to develop an automated and efficient diagnostic framework based on deep learning for precise localization of vascular stenosis. In this study, the performance of two advanced object detection architectures, namely RetinaNet and EfficientDet-D3, was evaluated. The models were trained and validated using a dataset comprising 8,325 angiographic images obtained from 100 patients diagnosed with single-vessel CAD. To enhance generalizability, image preprocessing and data augmentation techniques were applied. Experimental results demonstrated that the EfficientDet-D3 model achieved a superior performance, with a mean Average Precision (mAP) of 96.6%. The RetinaNet model also exhibited strong robustness, achieving an mAP of 93.2% while maintaining high processing speed. The proposed framework, by achieving a balance between accuracy and computational efficiency, demonstrates strong potential for deployment in clinical settings as a decision support system. It may facilitate timely and accurate diagnosis by reducing reliance on manual interpretation.

Keywords: Coronary artery stenosis detection, Deep learning, Object detection, EfficientDet, RetinaNet, Coronary angiography.

1 | Introduction

Coronary Artery Disease (CAD), resulting from the narrowing of blood-supplying coronary arteries, remains one of the major challenges of global healthcare systems. According to epidemiological reports, it

✉ Corresponding Author: Kamran.Balani@iau.ac.ir

doi <https://doi.org/10.22105/ahse.v2i4.54>



Licensee System Analytics. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

accounts for more than 31% of all deaths worldwide. The pathology of this disease is primarily driven by a process known as atherosclerosis, in which the accumulation of lipid- and cholesterol-rich plaques within the arterial wall leads to progressive luminal narrowing. This obstruction, by restricting oxygenated blood flow, can result in severe outcomes such as myocardial infarction and heart failure, thereby underscoring the critical importance of early diagnosis and precise clinical management [1].

For the identification of such lesions, Coronary Angiography (CAG) using X-ray imaging is considered the primary diagnostic technique and the clinical gold standard. These images enable cardiologists to visualize vascular occlusions; however, the main challenge lies in the inherent complexity of angiographic data. In many cases, stenotic regions constitute only a small portion of the vessel, while the majority remains healthy. Consequently, accurately delineating lesion boundaries within image noise and dynamic anatomical structures is a demanding task [2].

Despite the critical importance of this diagnostic task, current clinical approaches remain heavily dependent on manual interpretation of angiographic images by specialists. This process is not only time-consuming but also inherently subjective, making it susceptible to inter-observer variability and human error. Furthermore, the shortage of expert cardiologists relative to the rapidly increasing volume of medical imaging data has created a significant bottleneck in healthcare systems. These operational limitations highlight an urgent need for automated diagnostic systems capable of reducing subjective judgment, improving diagnostic accuracy, and accelerating clinical decision-making [3].

With the advent of deep learning in medical applications, numerous studies have attempted to automate this process; however, a significant research and technical gap still persists. Most existing approaches have focused on computationally intensive and complex models such as Faster R-CNN. Although these models achieve high detection accuracy, their two-stage architecture and high inference time limit their applicability in real-time scenarios. In addition, they require substantial computational resources, which are often unavailable in resource-constrained clinical environments. Moreover, faster alternative models frequently suffer from degraded performance when applied to angiographic images, which are typically characterized by high noise levels and fine-grained vascular structures. Therefore, the lack of a framework that can effectively balance diagnostic accuracy and computational efficiency remains a critical challenge in this domain [4], [5].

To address this gap, the present study proposes an efficient object detection approach based on two advanced architectures, namely RetinaNet and EfficientDet-D3. The selection of these models directly responds to the aforementioned challenges. RetinaNet, through its focal loss function, effectively mitigates the issue of class imbalance and improves the detection of small objects in medical images. In parallel, EfficientDet-D3, with its lightweight and scalable architecture, enables fast processing of noisy images without compromising detection accuracy. These models are specifically optimized for deployment in resource-constrained environments to enable real-time detection.

In this study, a comprehensive and challenging dataset comprising 8,325 angiographic images was used to evaluate the effectiveness and robustness of the proposed framework. These images were collected from 100 patients diagnosed with single-vessel CAD who underwent CAG at the institute for complex cardiovascular disease research in Kemerovo, Russia. Model training and evaluation were conducted using a carefully designed data split (80% training, 10% validation, and 10% testing) to ensure unbiased performance assessment. The results indicate that the EfficientDet-D3 model achieved a superior performance, reaching a mean Average Precision (mAP) of 96.6%, demonstrating strong potential for integration into clinical decision-support systems.

Overall, the main scientific contributions and innovations of this study can be summarized in the development of an efficient automated diagnostic framework capable of precise localization of coronary artery stenosis. By leveraging lightweight architectures, this work achieves an optimal balance between accuracy and computational speed, making it suitable for real-time clinical applications. Moreover, the

proposed model demonstrates superior performance compared to existing methods while maintaining strong robustness in handling noisy and high-resolution angiographic images. This achievement ultimately reduces reliance on manual interpretation and facilitates improved clinical decision-making, thereby contributing to enhanced patient outcomes.

The remainder of this paper is organized as follows: Section 2 reviews related literature. Section 3 presents the materials and methods, including a description of the dataset and the deep learning architectures employed. Section 4 reports the experimental results, performance evaluation of the models, and comparative analysis with existing approaches. Finally, Section 5 provides a discussion of the findings, concluding remarks, and directions for future research.

2 | Related Work

In the literature, traditional Machine Learning (ML) approaches for coronary artery stenosis detection have typically relied on a discrete three-stage pipeline consisting of feature extraction, feature selection, and classification. These methods are often time-consuming and computationally inefficient due to their dependence on handcrafted feature engineering. However, the emergence of the deep learning paradigm has brought a fundamental transformation to this field, as such models are capable of integrating all these stages into a unified end-to-end optimization framework. In particular, CNNs have achieved significant improvements in medical image analysis by automatically learning representative features from data, thereby substantially enhancing diagnostic performance [6], [7].

In X-ray CAG, the video-based nature of the data introduces a unique set of challenges. To achieve accurate diagnosis, it is first necessary to preprocess the video sequence and extract only those frames in which the coronary arterial tree is clearly visible as candidate frames. After selecting the appropriate frames, prior studies have generally followed one of two main paradigms for the detection task: 1) image classification, in which the probability of class membership (presence or absence of disease) is estimated for the entire image, and 2) object detection and localization, where the goal is to precisely determine the location of lesions and assign labels by drawing bounding boxes around the regions of interest. In the following, studies related to coronary artery stenosis detection using deep learning-based methods are briefly reviewed.

2.1 | Object-Based Classification

In high-precision medical applications, merely detecting the presence of disease is insufficient; clinicians also require precise information about the location of lesions. The “object-based classification” approach directly addresses this need. In this method, instead of assigning a single label to the entire image, the algorithm aims to draw a bounding box around the stenotic region. The technical process typically involves generating a large number of candidate bounding boxes based on image features, clustering overlapping boxes, and ultimately selecting the most probable box as the final output. Various researchers have employed advanced architectures to tackle this challenge.

In this context, Pang et al. [8] proposed the Stenosis-DetNet network by considering the video nature of angiographic data. This network is an end-to-end model that operates directly on raw video inputs. Their main innovation was the incorporation of prior knowledge regarding the displacement of coronary vessels throughout the cardiac cycle, enabling the model to track and refine the location of stenosis across consecutive frames. They utilized the powerful ResNet architecture as the backbone for deep feature extraction.

Danilov et al. [9] conducted a comprehensive study to identify the most suitable architecture for single-stenosis lesion detection. They evaluated eight different combinations of object detectors (including SSD, Faster R-CNN, and R-FCN) and feature extraction backbones (including MobileNet-v2, ResNet50, Inception-v4, and NASNet). Their findings indicated that, although deeper and more complex models

generally achieve higher accuracy, the combination of R-FCN with ResNet101 provides the best balance between detection performance and computational speed for clinical applications.

Li et al. [10] introduced a more complex framework, STQD-Det, which leverages spatial-temporal feature sharing alongside a novel concept referred to as “quantum propagation.” This approach achieved a very high F1-score of 92.39% while processing at 25.08 frames per second. However, the authors also noted that the high computational complexity and reliance on specialized modules remain major barriers to deployment in standard hospital hardware environments.

2.2 | Image-Based Classification

In many scenarios, precise annotation of individual lesions (i.e., bounding box delineation) is not feasible due to the high level of expertise required from radiologists and the time-consuming nature of the process. Under such conditions, “image-based classification,” in which an entire image or a region of interest is labeled as either “healthy” or “diseased,” offers a more practical alternative. In this context, Moon et al. [11] proposed a two-stage system to address the challenge of low-quality angiographic images. In their approach, the algorithm first automatically identifies and selects frames in which the contrast agent fully fills the vessel. Subsequently, the pre-trained Inception-v3 model—initially trained on the ImageNet dataset—is applied to these selected frames. A key contribution of their work is the integration of Convolutional Block Attention Modules (CBAM) in parallel within the early layers of the network. These attention modules enable the model to focus on vascular structures rather than irrelevant noisy background regions.

One of the major challenges in medical deep learning is the limited availability of annotated data. To address this issue, researchers often divide original images into smaller patches (e.g., 32×32 pixels). Antczak and Liberadzki [12], in addition to releasing a dedicated dataset, proposed an innovative data augmentation strategy to mitigate data scarcity. Assuming that vascular structures exhibit geometric similarity to Bézier curves, they generated synthetic training samples based on this assumption. These artificially generated data were then used to pre-train a lightweight convolutional neural network consisting of only five layers, resulting in improved detection performance.

On the other hand, the use of overly large models on limited medical datasets often leads to overfitting. To address this issue, Ovalle-Magalan et al. [13] proposed an approach known as “network slicing.” They conducted an extensive search across different layers of VGG16, ResNet50, and Inception-v3 to determine how much of the pretrained knowledge should be retained during transfer learning. Their results indicated that, for coronary artery stenosis detection, Inception-v3 provides optimal performance when only its first three blocks are preserved and fine-tuned, while the remaining layers are retrained to adapt to the new task.

2.3 | Existing Challenges and Research Gap

Despite the progress achieved by the aforementioned approaches, each of these methods is associated with specific technical and operational limitations. Image classification-based methods, despite their relative simplicity, suffer from an inherent drawback, namely the lack of precise localization capability. These methods only indicate the presence or absence of disease within an image but fail to provide spatial information about lesion location, which is critical for interventional clinical procedures. In addition, alternative strategies often rely on complex preprocessing pipelines and synthetic data generation techniques (such as Bézier curve modeling) to compensate for data scarcity, which significantly increases computational burden and implementation complexity.

On the other hand, object detection-based methods, while addressing the localization problem, still face the persistent challenge of balancing accuracy and speed. Accurate two-stage models such as Faster R-CNN, due to their heavy architectural design, exhibit long inference times, making their deployment in real-time and intra-procedural angiography systems practically infeasible. In contrast, highly efficient models such as early versions of YOLO often experience reduced accuracy when detecting very small and subtle abnormalities in high-resolution medical images. Moreover, recent advanced and complex frameworks such

as quantum-based architectures (e.g., STQD-Det), despite their high accuracy, require specialized hardware resources that are not readily available in most clinical environments.

Therefore, a review of the existing literature reveals a clear research gap: the absence of an efficient framework that can provide the computational speed required for clinical applications without compromising the accuracy of small lesion detection. The present study aims to address this gap by proposing the use of RetinaNet and EfficientDet-D3 architectures. These models, by leveraging mechanisms such as the Feature Pyramid Network (FPN) and focal loss function, are capable of overcoming the challenges posed by noisy images and small object detection, making them well-suited candidates for deployment in real-time medical imaging systems.

3 | Proposed Method

In this study, the problem of coronary artery stenosis detection is formulated as an object detection task. The primary objective is to design a system capable of receiving raw coronary angiographic images, identifying the precise location of stenotic regions, and enclosing them within bounding boxes. To address existing challenges such as high noise levels, class imbalance, and the need for rapid processing, two advanced one-stage detection architectures, namely RetinaNet and EfficientDet-D3, have been implemented and optimized. The overall pipeline of the proposed framework is illustrated in *Fig. 1*.

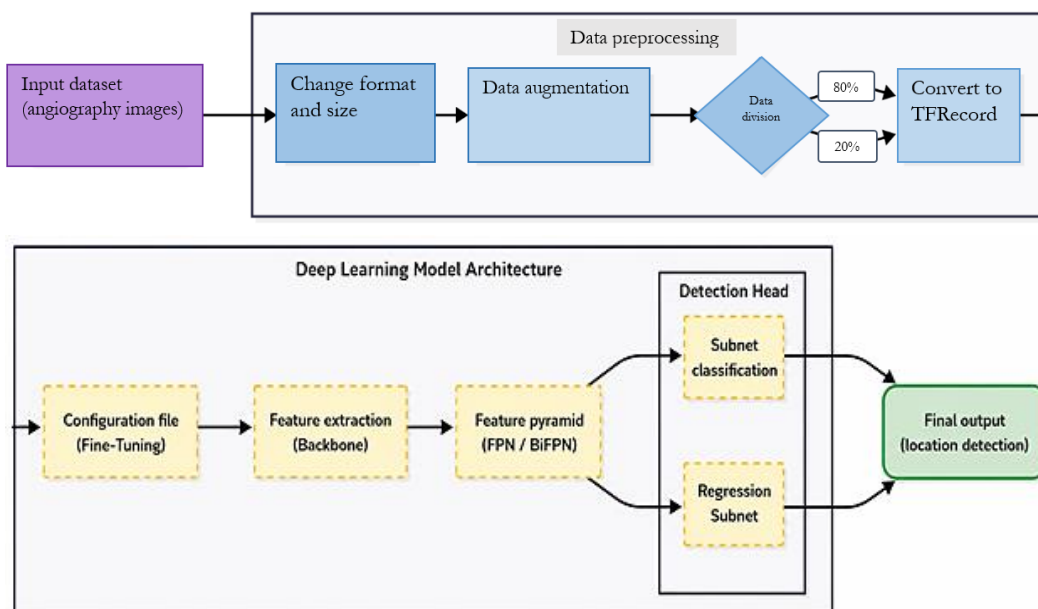


Fig. 1. Overview of the proposed model framework.

3.1 | RetinaNet Model

RetinaNet is one of the most powerful one-stage object detectors, specifically designed for the detection of small and densely distributed objects. The architecture of this network consists of two main components. The overall structure of the RetinaNet model is illustrated in *Fig. 2*.

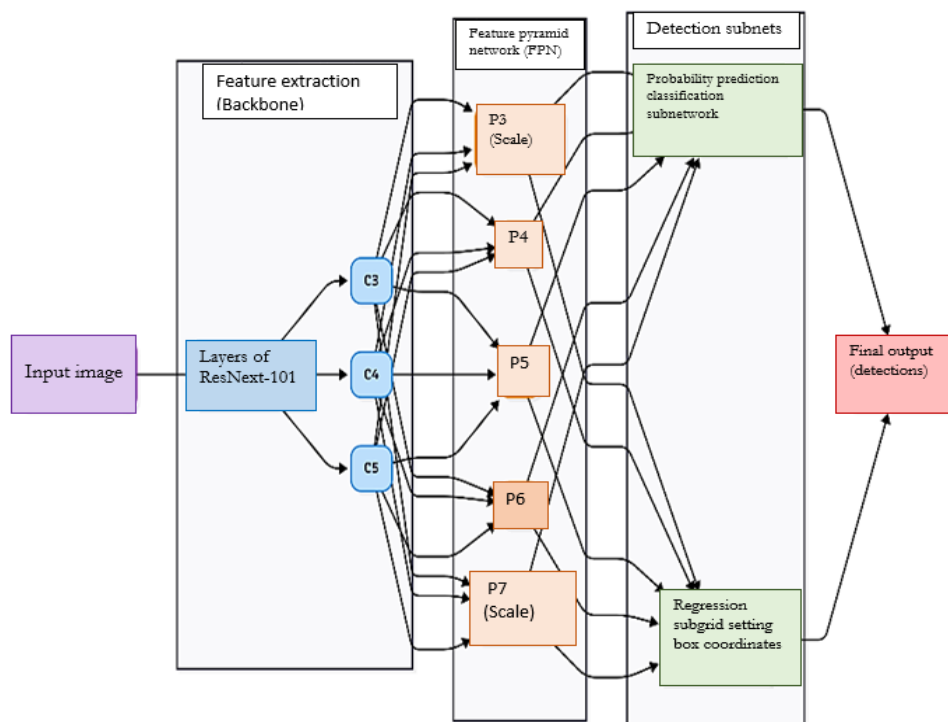


Fig. 2. Architecture of the retinaNet network.

FPN: in conventional deep neural networks, deeper layers contain strong semantic information but have low spatial resolution, which makes them unsuitable for detecting small objects such as mild stenotic regions. RetinaNet addresses this limitation by employing a FPN built on top of a ResNeXt-101 backbone. The FPN structure consists of a bottom-up pathway and a top-down pathway with lateral connections. This architecture enables the integration of strong semantic features from higher layers with fine-grained spatial details from lower layers. As a result, the model can accurately localize coronary stenosis across multiple scales.

Focal Loss Function: one of the major challenges in object detection training is class imbalance, where most regions in an image correspond to the background (healthy tissue), while only a small portion contains stenotic lesions. Standard loss functions, such as cross-entropy loss, tend to be dominated by a large number of easy negative samples. RetinaNet addresses this issue by introducing the focal loss *Function 1*. In this formulation, the parameter γ (set to 2 in this study) significantly down-weights easy examples, forcing the model during training to focus more on hard-to-classify samples, particularly stenotic regions.

$$FL(p_t) = -\alpha t(1 - p_t)^\gamma \log(p_t). \quad (1)$$

3.2 | EfficientDet-D3 Model

This model has been developed with the aim of bridging the gap between high detection accuracy and computational efficiency, making it an ideal choice for clinical environments with limited hardware resources. The architecture of EfficientDet-D3 is illustrated in *Fig. 3*. The key innovations of this model include:

Bi-directional Feature Pyramid Network (BiFPN): in contrast to the conventional FPN, where information flows in a single direction, EfficientDet employs a more advanced structure known as BiFPN. In this design, nodes that contribute minimally to feature extraction are removed, and additional edges are introduced to facilitate efficient feature transfer across different levels. A key characteristic of BiFPN is its ability to enable bidirectional information flow—both top-down and bottom-up—as well as repeated

stacking of these fusion blocks. This results in richer feature representations and more effective integration of multi-level semantic information.

Compound scaling: most object detection models improve accuracy by increasing only one dimension of the network, such as depth or width. In contrast, EfficientDet utilizes a compound scaling strategy in which all dimensions of the network—including input resolution, depth, and width—are scaled simultaneously and in a coordinated manner using a single compound coefficient (ϕ). In this study, the EfficientDet-D3 variant is employed, which is optimized for high-resolution images (896×896 pixels). This choice enables the model to capture subtle stenotic patterns in large angiographic images without incurring excessive computational overhead.

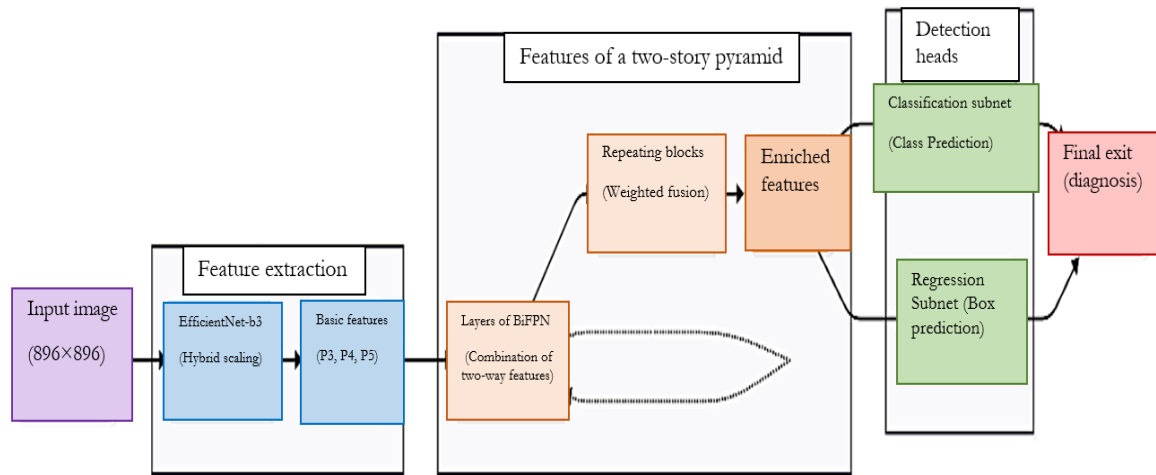


Fig. 3. Architecture of the EfficientDet-D3 network.

4 | Implementation and Experimental Results

4.1 | Dataset

In this study, a dataset comprising 8,325 angiographic images from 100 patients with a confirmed diagnosis of single-vessel CAD was used. The data were collected from patients who underwent CAG using Siemens Coroscop and GE Innova systems at the institute for complex cardiovascular disease research in Kemerovo, Russia. All participants provided written informed consent, and the images were acquired in accordance with ethical protocols. The raw images had variable spatial resolutions ranging from 512×512 to 1000×1000 pixels. Data preparation was performed within an object detection framework, where clinical experts annotated stenotic regions using bounding boxes. A notable characteristic of this dataset is that stenotic lesions typically occupy only a small portion of the vessel, while the majority of the artery remains healthy. This property enables the model to learn not only discriminative features of stenotic regions but also contextual information from healthy vascular tissue within the same frame. For unbiased evaluation, the dataset was split into training (80%), validation (10%), and test (10%) sets. A patient-level, non-random splitting strategy was adopted, ensuring that all images belonging to a single patient were included exclusively in one subset. This approach prevents data leakage between training and testing phases and guarantees that the model is evaluated on entirely unseen patients, thereby providing a realistic assessment of generalization performance. The demographic and clinical characteristics of the dataset are presented in *Table 1*.

Table 1. Demographic and clinical characteristics of the dataset.

Clinical / Demographic Feature	Value / Distribution
Total number of patients	100 patients
Total number of images	8,325 frames
Mean age (years)	60.3±13.8
Gender	Male: 68% / Female: 32%
Body Mass Index (BMI)	21.6±5.1 (kg/m ²)
Coronary artery stenosis severity	70% in all patients
Comorbidities	Hypertension: 53% Diabetes mellitus: 14% Chronic heart failure: 36%
Final diagnosis	CAD

4.2 | Evaluation Metrics

In this study, the mAP is used as the standard quantitative metric to evaluate model performance and the quality of lesion localization. This widely adopted metric in object detection tasks assesses the model's ability to accurately determine the spatial location of coronary stenosis. To compute mAP, two key components—precision and recall—are first defined as follows:

- I. Precision (P): indicates the proportion of positive predictions made by the model that are actually correct, reflecting the accuracy of the predictions.
- II. Recall (R): indicates the proportion of true stenotic regions correctly identified by the model, reflecting its sensitivity.

The final evaluation is based on the Intersection over Union (IoU) criterion. In this study, the IoU threshold is set to 0.5. This means that a predicted bounding box is considered a correct detection if its overlap with the ground truth bounding box exceeds 50%. The mAP value is computed as the mean of Average Precision (AP) across all detections, where higher values indicate superior model performance in both discrimination and localization of stenotic regions.

4.3 | Data Preprocessing and Preparation

In this study, a carefully designed preprocessing pipeline was implemented to enhance the quality of input data and ensure compatibility with the requirements of deep neural networks. Given the noisy nature of angiographic images and the variability in raw image dimensions, the first step involved image standardization. For the RetinaNet-101 model, all images were resized to a fixed resolution of 640 × 640 pixels to match the network input requirements. In contrast, for the EfficientDet-D3 model, which requires finer vascular detail extraction, images were resized to a higher resolution of 896 × 896 pixels. This increase in resolution enables the model to more accurately capture subtle and clinically significant stenotic patterns that may be lost at lower resolutions.

To improve model generalization and prevent overfitting, targeted data augmentation techniques were applied. Random horizontal flipping was employed to enable the model to recognize stenosis regardless of vessel orientation. In addition, random cropping was used to extract smaller image regions, thereby encouraging the network to focus on local features and small-scale stenotic structures. Specifically for the EfficientDet model, a more advanced augmentation strategy involving random scaling, cropping, and padding-to-square transformation was applied to enhance robustness under varying spatial scales and configurations, reflecting real-world clinical variability.

In the final stage of data preparation, the dataset format was optimized for computational efficiency. Raw images and their corresponding XML annotation files were converted into the binary TFRecord format. TFRecord files store data as serialized binary sequences, significantly reducing Input/Output (I/O)

bottlenecks during training. The use of this unified format eliminated the overhead associated with reading large numbers of individual files from directories and maximized data throughput to the Graphics Processing Unit (GPU) during model training.

4.4 | Implementation Settings and Hardware Configuration

In this study, the training and evaluation process was implemented using the TensorFlow Object Detection Application Programming Interface (API) (version 2.13) and the Python programming language (version 3.9). To accelerate convergence and leverage previously learned visual representations, transfer learning was employed. Accordingly, the initial weights of the models were pre-trained on the Common Objects in Context (COCO) dataset and subsequently fine-tuned using the angiographic dataset of this study.

Due to differences in architectural complexity and computational requirements, two separate hardware environments were utilized. The RetinaNet-101 model, which operates on standard 640×640 pixel inputs, was trained on a local workstation equipped with an NVIDIA GeForce RTX 3080 Ti GPU (12 GB VRAM) and 16 GB of RAM. This configuration was sufficient to handle the computational demands of the FPN and ResNet backbone.

In contrast, the EfficientDet-D3 model, which processes high-resolution inputs (896×896 pixels) and employs the more complex BiFPN architecture, required significantly higher memory resources. Therefore, training was conducted on the Google Colab cloud platform using an NVIDIA A100 GPU (40 GB VRAM) and 83 GB of system memory. This hardware separation ensured that each model was trained under optimal conditions without memory bottlenecks, thereby maximizing computational efficiency. A summary of hyperparameter settings and hardware specifications is provided in *Table 2*.

Table 2. Summary of hyperparameter settings and hardware specifications.

Parameter/Setting	RetinaNet-101	EfficientDet D3
Input image size	640×640 pixels	896×896 pixels
Batch size	4	4
Number of training steps	65000	25000
Base learning rate	0.0025	0.0025
Optimizer	Momentum (0.9)	Momentum (0.9)
Focal Loss parameters	$\gamma=2.0, \alpha=0.25$	$\gamma=1.5, \alpha=0.25$
Computational hardware (GPU)	NVIDIA RTX 3080 Ti (12GB)	NVIDIA A100 SXM4 (40GB)
Execution environment	Local system	Cloud platform (Google Colab)

4.5 | Experimental Results

In this section, the performance of the proposed models (RetinaNet and EfficientDet-D3) is evaluated in terms of accuracy and speed and compared with state-of-the-art methods. The primary metric used for assessing localization accuracy is mAP with an IoU threshold of 0.5.

Table 3 presents the results of evaluating the proposed models in comparison with other recent studies. As shown, the EfficientDet-D3 model achieves the highest accuracy among all evaluated methods, with a mAP of 96.6%. This model outperforms strong existing approaches, including the method proposed by Danilov et al. [9] with an mAP of 95.4% and the complex quantum-based framework of Li et al. [10] with a score of 92.39%. This superior performance can be attributed to the BiFPN architecture and the use of higher-resolution input images (896×896), which enable more precise delineation of stenotic boundaries. The RetinaNet-101 model also demonstrates highly competitive performance, achieving an mAP of 93.2%, indicating its robustness in handling the challenges of angiographic imaging.

In addition to accuracy, processing speed is of critical importance for real-time clinical applications. *Table 4* reports the inference time of both models on the test set. RetinaNet achieves a faster average inference time of 57 ms per frame compared to EfficientDet-D3, which requires 95 ms. Although EfficientDet-D3 is

slower, its significantly higher accuracy (96.6%) makes it a more suitable option for high-stakes diagnostic scenarios where accuracy is the primary priority. Importantly, both models operate within an acceptable time range for intraoperative and clinical use.

Table 3. Comparison of the proposed methods with previous studies.

Study	Task Type	Evaluation Metric	Result (Percent)
Ovalle-Magallanes et al. [13]	Classification	Accuracy	93
Antczak and Liberadzki [12]	Classification	Precision	90
Moon et al. [11]	Classification	Precision	94.3
Pang et al. [8]	Object detection	Accuracy	94.87
Danilov et al. [9]	Object detection	mAP	95.4
Li et al. [10]	Object detection	F1-score	92.39
Proposed method (RetinaNet)	Object detection	mAP	93.2
Proposed method (EfficientDet-D3)	Object detection	mAP	96.6

Table 4. Comparison of speed and accuracy of the proposed models.

Model	Input Resolution	Processing Speed (ms)	mAP on COCO	Final Accuracy
SSD RetinaNet101	640×640	57	35.6	93.2%
EfficientDet D3	896×896	95	45.4	96.6%

A visual inspection of the model outputs demonstrates that the algorithm has a strong capability in distinguishing complex background tissues (such as bone structures and imaging noise) from coronary vessels. The models were able to detect and localize coronary stenoses with high confidence levels (e.g., 53% and 75% as shown in *Fig. 4*). These visual results confirm that the proposed model is reliable not only in quantitative metrics but also in practical, real-world scenarios.

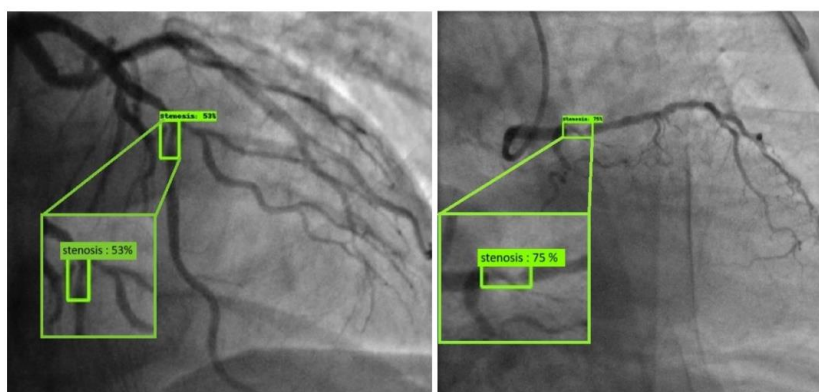


Fig. 3. Visualization of model outputs.

The findings of this study demonstrate that the use of lightweight object detection architectures such as EfficientDet-D3 and RetinaNet can provide an effective solution for bridging the gap between high accuracy and high processing speed in coronary artery stenosis detection. The achieved accuracy of 96.6% in the EfficientDet-D3 model confirms that, by leveraging mechanisms such as BiFPN, it is possible to deliver performance superior to that of more complex and computationally expensive models, even in noisy angiographic images. However, this study is subject to certain limitations: 1) although the dataset used is sufficiently large, it is single-center in nature, meaning that all images were acquired from a specific hospital using particular imaging devices. This may limit the generalizability of the model when applied to data from other clinical centers with different imaging protocols, and 2) the quality of input images remains a persistent challenge; the presence of interfering anatomical structures such as vertebrae or catheters in some frames may introduce noise and reduce detection accuracy. Furthermore, the current model is focused

exclusively on single-vessel disease, and its performance in patients with multi-vessel involvement has not yet been fully evaluated. Future research is recommended to address these limitations by incorporating multi-center datasets and integrating advanced denoising techniques.

5 | Conclusion

This study introduced an automated deep learning-based diagnostic framework and demonstrated the effectiveness of two architectures, EfficientDet-D3 and RetinaNet-101, in the precise localization of coronary artery stenosis. The experimental results showed that EfficientDet-D3 achieved a mAP of 96.6%, while RetinaNet achieved 93.2%, both outperforming previous state-of-the-art methods and providing an effective balance between accuracy and speed for real-time applications.

The implementation of this approach has significant potential to support clinical decision-making by reducing reliance on manual interpretation and minimizing human error, thereby accelerating diagnostic and therapeutic processes. Future research should focus on evaluating the proposed models on multi-center datasets with greater device variability and developing advanced noise reduction techniques (e.g., suppression of anatomical artifacts such as vertebrae and catheters), in order to facilitate the deployment of this system as a reliable clinical decision-support tool.

Authors' Contributions

All aspects of the research and manuscript preparation were carried out by the author. The author has read and approved the final version of the manuscript.

Funding

This study did not receive any specific funding from public, commercial, or non-profit funding agencies.

Data Availability

All data are included in the text.

Conflict of Interest

The author declares that he does not have any conflict of interest.

Consent for Publication

The author has given consent for the publication of this manuscript.

Ethics Approval and Consent to Participate

This study does not involve any research conducted on human participants or animals.

References

- [1] Huang, B., Luo, Y., Wei, G., He, S., Shao, Y., Zeng, X., & Zhang, Q. (2025). Deep learning model for coronary artery segmentation and quantitative stenosis detection in angiographic images. *Medical physics*, 52(7), e17970. <https://doi.org/10.1002/mp.17970>
- [2] Myrmel, G. M. S., Wereski, R., Karaji, I., Saeed, N., Aakre, K. M., Mills, N. L., & Pedersen, E. R. (2025). Cardiac biomarkers and CT coronary angiography for the assessment of coronary heart disease. *Clinical biochemistry*, 135, 110857. <https://doi.org/10.1016/j.clinbiochem.2024.110857>
- [3] Sakamoto, M., Yoshimura, T., & Sugimori, H. (2025). Automated coronary artery identification in CT angiography: A deep learning approach using bounding boxes. *Applied sciences*, 15(6), 3113. <https://doi.org/10.3390/app15063113>

- [4] Kim, Y. I., Roh, J. H., Kweon, J., Kwon, H., Chae, J., Park, K., ... & Kim, Y. H. (2024). Artificial intelligence-based quantitative coronary angiography of major vessels using deep-learning. *International journal of cardiology*, 405, 131945. <https://doi.org/10.1016/j.ijcard.2024.131945>
- [5] Zeng, X., Huang, B., Luo, Y., Wei, G., He, S., & Shao, Y. (2025). SAM-VMNet: Deep neural networks for coronary angiography vessel segmentation. <https://doi.org/10.48550/arXiv.2406.00492>
- [6] Li, S., & Fan, Y. (2024). Coronary artery segmentation in X-ray angiography based on deep learning approach. *2024 43rd Chinese control conference (CCC)* (pp. 7345-7350). IEEE. <https://doi.org/10.23919/CCC63176.2024.10662007>
- [7] Chen, Z., Chen, Y., Si, J., Xiao, C., Liu, X., Wang, C., ... & Guo, Y. (2025). An innovative model for diagnosing lesions in coronary angiography imagery using an improved YOLOv4 model. *Bioengineering*, 12(11), 1241. <https://doi.org/10.3390/bioengineering12111241>
- [8] Pang, K., Ai, D., Fang, H., Fan, J., Song, H., & Yang, J. (2021). Stenosis-DetNet: Sequence consistency-based stenosis detection for X-ray coronary angiography. *Computerized medical imaging and graphics*, 89, 101900. <https://doi.org/10.1016/j.compmedimag.2021.101900>
- [9] Danilov, V. V., Klyshnikov, K. Y., Gerget, O. M., Kutikhin, A. G., Ganyukov, V. I., Frangi, A. F., & Ovcharenko, E. A. (2021). Real-time coronary artery stenosis detection based on modern neural networks. *Scientific reports*, 11(1), 7582. <https://doi.org/10.1038/s41598-021-87174-2>
- [10] Li, X., Ai, D., Song, H., Fan, J., Fu, T., Xiao, D., ... & Yang, J. (2024). STQD-Det: Spatio-temporal quantum diffusion model for real-time coronary stenosis detection in X-ray angiography. *IEEE transactions on pattern analysis and machine intelligence*, 46(12), 9908-9920. <https://doi.org/10.1109/TPAMI.2024.3430839>
- [11] Moon, J. H., Cha, W. C., Chung, M. J., Lee, K. S., Cho, B. H., & Choi, J. H. (2021). Automatic stenosis recognition from coronary angiography using convolutional neural networks. *Computer methods and programs in biomedicine*, 198, 105819. <https://doi.org/10.1016/j.cmpb.2020.105819>
- [12] Antczak, K., & Liberadzki, Ł. (2018). Stenosis detection with deep convolutional neural networks. *MATEC web of conferences* (p. 04001). EDP Sciences. <https://doi.org/10.1051/mateconf/201821004001>
- [13] Ovalle-Magallanes, E., Avina-Cervantes, J. G., Cruz-Aceves, I., & Ruiz-Pinales, J. (2020). Transfer learning for stenosis detection in X-ray coronary angiography. *Mathematics*, 8(9), 1510. <https://doi.org/10.3390/math8091510>