ORIGINAL ARTICLE

Facilitating Timely Decision-Making in Healthcare: An Object Detection Approach for Automated Coronary Artery Stenosis Detection

Hadis Keshavarz [1], Hossein Sadr [2,3*], Mojdeh Nazari [1,4], Arsalan Salari [1]

[1] *Cardiovascular Diseases Research Center, Department of Cardiology, Heshmat Hospital, School of Medicine, Guilan University of Medical Sciences, Rasht, Iran*

[2] *Guilan Road Trauma Research Center, Trauma Institute, Guilan University of Medical Sciences, Rasht, Iran*

[3] *Department of Artificial Intelligence in Medicine, Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran*

[4] *Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

*Corresponding Author: Hossein Sadr
Received: 21 May 2024 / Accepted: 09 April 2025
Email: sadr@qiau.ac.ir

Abstract

Purpose: Coronary Artery Disease (CAD), characterized by coronary artery stenosis—the narrowing of arteries supplying blood to the heart—is a leading global cause of morbidity and mortality. Timely detection and management of stenosis are crucial to preventing severe outcomes such as myocardial infarction and heart failure. Despite advancements in medical imaging, current diagnostic methods rely heavily on the manual interpretation of coronary angiograms, which is time-consuming, subjective, and prone to variability. To address these limitations, this study proposes an automated object detection-based framework for identifying coronary artery stenosis in medical imaging.

Materials and Methods: The study employs two state-of-the-art deep learning models, RetinaNet and EfficientDet D3, to detect stenotic regions in X-ray angiography images. A dataset of 8,325 annotated images from 100 patients with single-vessel CAD, sourced from the Research Institute for Complex Issues of Cardiovascular Diseases in Kemerovo, Russia, was used for training and evaluation. To enhance model performance, a comprehensive preprocessing pipeline was applied, including image resizing, data augmentation, and intensity normalization. These steps ensured robustness and generalizability across diverse imaging conditions.

Results: Both models demonstrated high accuracy in stenosis detection. RetinaNet achieved a mean Average Precision (mAP) of 93.2%, while EfficientDet D3 outperformed with an mAP of 96.6%. These results highlight the models' ability to accurately identify stenosis, even in noisy and variable angiographic images. The superior performance of EfficientDet D3 underscores its potential for clinical integration, offering precise and reliable stenosis localization.

Conclusion: This study presents a robust and efficient deep learning framework for the automated detection of coronary artery stenosis. By reducing reliance on manual interpretation and enhancing diagnostic accuracy, the proposed approach supports timely and informed clinical decision-making. This innovation has the potential to streamline diagnostic workflows, improve patient outcomes, and advance the application of artificial intelligence in cardiovascular healthcare.

Keywords: Object Detection; Deep Learning; Medical Image; Coronary Angiography; Digital Medicine.

DOI:

1. Introduction

Coronary Artery Disease (CAD) has emerged as a prevalent condition impacting individuals worldwide, particularly in developed nations. According to global health organization data, cardiovascular diseases account for more than 31% of the global mortality rate [1, 2]. Coronary artery disease results from atherosclerosis, where plaque, consisting of fats, cholesterol, calcium, and other blood substances, accumulates in arteries due to elevated LDL cholesterol. This plaque hardens over time, narrowing arteries and restricting the flow of oxygen-rich blood [3, 4].

Data has emerged as a vital resource for organizations across various sectors, including healthcare, in an era characterized by technological advancement and digital transformation. Presently, the majority of hospitals maintain electronic health records that encompass a comprehensive array of patient information, such as clinical histories, medical imaging, symptoms, diagnoses, and treatment regimens [5, 6]. The volume of medical data generated and collected on a daily basis is on the rise. However, the full potential of this data often remains untapped due to limitations in effective analytics tools and a shortage of adequately trained personnel. Leveraging this data to develop predictive screening and diagnostic models has the potential to alleviate workload pressures on healthcare professionals while simultaneously enhancing the healthcare system by facilitating early detection and timely intervention for patients [7, 8].

Current clinical practice for assessing the presence and extent of CAD relies on medical images obtained by various diagnostic procedures. In clinical practice, X-ray coronary angiography (CABG) is the main imaging technique for diagnosing coronary diseases [5, 9, 10]. Detailed X-ray images help the cardiologist identify blockages. However, this process is time-consuming, and the limited number of experts necessitates computer-aided diagnosis systems. However, these systems play an important role in cardiology in the detection of arterial anomalies, as this process is very time-consuming and there is a limited number of clinical experts available. Today, imaging research has improved by using deep learning to support medical decision-making processes [11, 12].

Our study presents a detailed analysis of the available neural network architectures and their potential in terms of accuracy in single-vessel disease detection. This approach aims to use efficient CNN architectures, namely RetinaNet and EfficientDet D3, for the object detection method and further explore the possibilities of their modification and optimization to provide superior real-time classification potential for the detection of multi-vessel coronary artery stenosis. It is worth mentioning that although previous studies have demonstrated significant advancements in stenosis detection using deep learning, many approaches rely on computationally expensive models or frameworks that lack real-time applicability. For instance, architectures such as Faster R-CNN or complex hybrid methods often require extensive resources and processing time, limiting their practicality in clinical settings. In contrast, we leverage the lightweight architectures of RetinaNet and EfficientDet D3 in this study, which strike a balance between computational efficiency and detection accuracy. These models are optimized to handle noisy and high-resolution angiographic images while maintaining fast inference times, making them highly suitable for real-time deployment in resource-constrained environments.

A freely available dataset, including data from about a hundred patients with confirmed one-vessel coronary artery disease who underwent coronary angiography at the Research Institute for Complex Problems of Cardiovascular Diseases in Kemerovo, Russia, was utilized in our implementation. Based on the empirical results, our proposed methods were able to achieve a good level of accuracy for Coronary Artery Stenosis Detection. It can be stated that our proposed methods can also be used to develop software that optimizes and facilitates invasive angiography.

The remainder of this paper is organized as follows. A review of the literature is provided in Section 1.1. Section 2 includes the details of our proposed model. The details of the experiments and the obtained results are summarized in Section 3. Conclusion and future research direction are also mentioned in Section 4.

1.1. Related Work

Stenosis detection based on traditional machine learning methods can be summarized in three steps: Feature extraction, selection, and classification. However, the Deep Learning paradigm can perform these three steps consistently during the optimization process. In particular, Convolutional Neural Networks (CNNs) have provided excellent performance improvements in the field of medical images. When working with X-ray video sequences, it is necessary to filter the entire video and keep only the images with a visible arterial tree to detect the stenosis. The classification process is performed on the candidate images using an object-based framework or an image classification task. In image classification, the probabilities of the class labels for an image are calculated. In contrast, object localization and classification draw a bounding box around the objects of interest in an image and assign them a class label [13].

1.2. Object-Based Classification

In the field of medical images, object-based classification refers to the identification of the location of lesions and their classification. Accordingly, the candidate boxes are created around the object detection suggestions. Created around the object detection suggestions. The boxes that are close to each other are combined into a single cluster. Then, only clusters that contain the most candidate boxes remain. Finally, a single bounding box is created for each object to assign a label. Various approaches have been suggested to address the challenge of stenosis detection, involving the automatic feature extraction of the image through a CNN [13].

Pang *et al.* introduced the Stenosis-DetNet, an end-to-end network that selects candidate frames from raw X-ray angiography videos. The network then detects candidate object bounding boxes containing stenotic regions, optimizing them using prior coronary artery displacement information and image features. A Residual Network (ResNet) is utilized as the backbone model for feature extraction [14].

Danilov *et al.* explored eight object detection CNN architectures to locate a single stenotic lesion. The configurations include Single Shot multi-box Detector (SSD), Faster-RCNN, and Region-based Fully Convolutional Networks (R-FCN), with backbone networks such as MobileNet-v2, ResNet50, RFCN ResNet101, Inception-v4, and NASNet. The RFCN ResNet101 configuration demonstrates an optimal trade-off between accuracy and speed [15].

In recent advancements, Li *et al.* [16] introduced the STQD-Det framework, achieving significant results in real-time coronary stenosis detection with an F1 score of 92.39% and a processing speed of 25.08 frames per second. Their model's innovative use of spatio-temporal feature sharing and quantum diffusion demonstrated notable improvements over several state-of-the-art methods. However, the complexity of their framework and the need for specialized quantum modules may pose challenges for implementation in resource-constrained environments.

1.3. Image-Based Classification

Labeling medical image data requires considerable expertise. Rather than recognizing and classifying each object within an image individually, it may be more useful to categorize the entire image as a single class. For a more sophisticated classification, a patch-based method can be used where each patch in the image is assigned a class. In this patch-based strategy, each patch is treated as a tagged object [13].

Moon *et al.* proposed a two-step approach for automated stenosis recognition in coronary angiography. Initially, a keyframe detection mechanism is employed to select images highlighting the most opacified coronary artery. Subsequently, a deep learning-based stenosis classification is conducted on each key frame using a pre-trained Inception-v3 model from the ImageNet dataset. The model incorporates single-channel and spatial-wise attention modules from CBAM arranged in parallel at the model's base to enhance attention. In cases where there is limited availability of labeled images, the authors recommend a patch-based approach for stenosis detection. This approach involves performing the classification at the patch level instead of categorizing the entire image. Consequently, a full-size image generates multiple patches of a predetermined size (e.g., 32 × 32 pixels), thereby increasing the number of training images. Refer to

Figure 8.14 for a detailed depiction of this process [17].

In this particular scenario, Antczak and Liberadzki released a dataset specifically designed for patch-based stenosis detection. The study utilized a shallow patch-based CNN, consisting of only five layers. To enhance the network's detection rates, a pre-training step was implemented. This involved creating a synthetic dataset through a generative model, assuming that a vessel could be represented as a Bezier curve [18].

Ovalle-Magallanes *et al.* introduced a network-cut approach to minimize the number of layers requiring transfer and fine-tuning from a source model pre-trained on a large dataset, employing transfer learning. Through an exhaustive search, the optimal cut and trainable layers were selected for three different state-of-the-art architectures: VGG16, ResNet50, and Inception-v3. The Inception-v3 model demonstrated the most effective detection results, retaining and fine-tuning only the first three inception blocks [19].

While prior research has illustrated substantial progress in the detection of stenosis through deep learning techniques, many of these methodologies depend on computationally intensive models or frameworks that do not facilitate real-time application. For example, architectures such as Faster R-CNN and various complex hybrid approaches typically necessitate considerable computational resources and processing durations, thereby reducing their feasibility in clinical environments. In contrast, the method presented in this study utilizes the lightweight architectures of RetinaNet and EfficientDet D3, effectively achieving a balance between computational efficiency and detection accuracy. These models are specifically optimized to process noisy, high-resolution angiographic images while ensuring rapid inference times, thus rendering them highly appropriate for real-time implementation in resource-limited settings. By integrating scalability with efficiency, our approach offers a robust and practical solution for the automated detection of coronary artery stenosis.

2. Materials and Methods

The proposed methodology aims to develop an effective system for the detection of coronary artery stenosis through object detection techniques using deep learning models (Figure 1). This process involves identifying and evaluating the location of each observable stenosis within X-ray angiography images, framing the task as an object detection problem where stenosis is treated as the object of interest. The process of stenosis detection is approached as an object detection challenge, supported by annotated bounding boxes that mark the stenosis within the optimal detection range. Modern object detection techniques, primarily based on deep learning models, can be categorized into two primary types: Single-shot Detectors (SSD) and Two-stage Detectors. This study focuses on leveraging advancements from Convolutional Neural Networks (CNNs) to facilitate the recognition of object classes effectively [20].

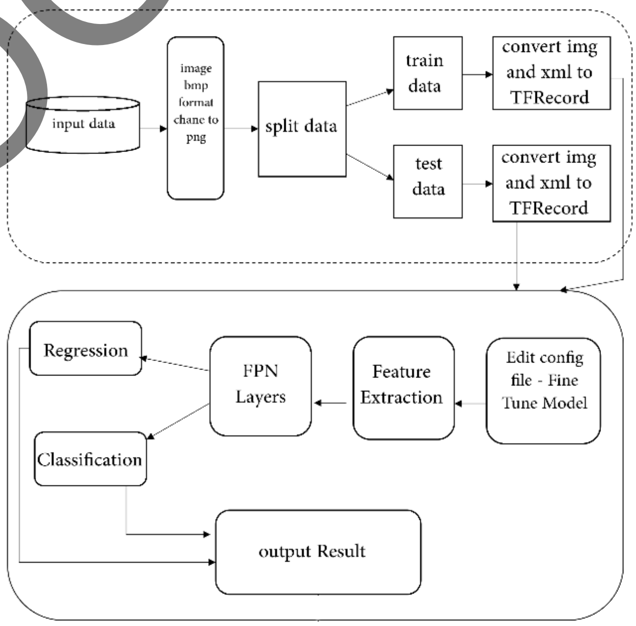

Figure 1. The process of the proposed model

RetinaNet was chosen for its innovative focal loss function, designed to alleviate class imbalance—a common issue in medical datasets characterized by sparse annotations. This feature, along with its Feature Pyramid Network (FPN), enhances multi-scale feature extraction, making RetinaNet particularly adept at detecting small and intricately detailed stenosis features with high precision.

EfficientDet D3 provides a scalable and lightweight architecture for real-time applications. Its Bidirectional Feature Pyramid Network (BiFPN) allows for the efficient aggregation of features across multiple scales, while a compound scaling approach ensures a balance between accuracy and computational efficiency. This model excels in processing noisy medical images without overfitting, rendering it an optimal choice for high-resolution angiographic datasets.

It is worth mentioning that alternative approaches, such as Faster R-CNN, while achieving high accuracy, suffer from slower inference speeds due to their two-stage architecture, which limits real-time applicability. Alternatively, models like YOLO (You Only Look Once), although fast, often encounter challenges in precisely detecting small anomalies within high-resolution images. By leveraging the strengths of RetinaNet and EfficientDet D3, our methodology aims to strike an optimal balance between precision and efficiency for real-time coronary artery stenosis detection [21, 22].

2.1. RetinaNet

RetinaNet is one of the most powerful one-stage object detection models, used especially for small object detection [23]. The structure incorporates ResNext-101 as its foundational backbone along with feature pyramid networks (FPN), which combines a more intricate backbone feature extractor comprising two supplementary sub-networks (Figure 2). These sub-networks play a crucial role in accurately classifying a bounding box and regressing the estimated coordinates [24, 25]. A key advantage of using Feature Pyramid Networks (FPN) over excessively deep feature maps that are too deep is twofold. First, deep feature maps encounter challenges in object localization, as small shifts in the deep feature map lead to significant localization errors when mapped back to the input image. FPN mitigates this issue by incorporating multiple levels of feature pyramids to provide more accurate object localization.

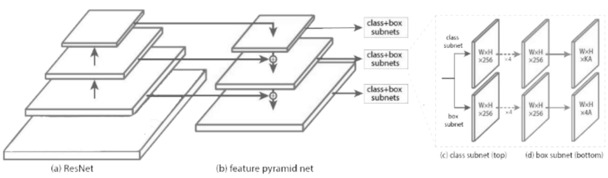

Figure 2. RetinaNet FPN architecture

Second, deep feature maps prove to be disadvantageous for small objects because spatial resolution diminishes, resulting in a loss of information related to these objects. FPN addresses this drawback by utilizing early layers to predict small objects. Despite the challenge posed by a potential lack of semantic information, FPN's strategy of integrating features from different scales improves its ability to accurately classify object categories, making it a superior choice compared to conventional deep feature maps [24]. Also, RetinaNet introduces focal-loss, which seeks to address the imbalance between positive and negative samples, as well as easy and hard samples, improving the quality of bounding boxes largely. In handling the abundance of generated boxes and addressing the class imbalance between background and stenosis assignment, we adopt the values of α=0.25 and γ=2 for the focal loss, which is experimented found to have the best performance in [25] (Equation 1).

$$FL(p_t)=\alpha_t(1-p_t)^{\gamma}\, log(p_t) \qquad (1)$$

2.2. EfficientDet D3

EfficientDet is one of the latest approaches to object detection, and the reason for its emergence is that the existing, albeit very accurate, networks have become costly in terms of resources and time. In the case of this network, the goal is to develop a model that takes into account both accuracy and efficiency to achieve results that can truly be used in real time. The starting point of this model is the feature pyramid networks, also known as FPN (feature pyramid network), as well as the Retinanet. This network represents a general solution for the construction of feature pyramids in convolutional networks. The construction of the pyramid includes a top-down and a bottom-up approach. In the bottom-up method, a hierarchy of features is calculated, which consists of a series of feature maps obtained in different proportions with a scaling step of 2. To create the pyramid, a layer is defined for each level, and the output of the last layer of each level is used as a reference set for the next one. In particular, 1 x 1 convolutions are used to minimize the size of bottom-up feature maps. Conversely, the higher resolution characteristics in the top-down approach are more semantically grounded but more geographically distant. Each side connection integrates feature maps from both top-down and bottom-up approaches that have the same spatial size. The architecture of the EfficientDet model is based on the EfficientDet network (Figure 3). This convolutional neural network is a method by which all dimensions are uniformly scaled using a compound coefficient. These factors are usually arbitrarily scaled, but the EfficientDet method uniformly scales the width, depth, and resolution of the network with a fixed set of scaling coefficients. The composite scaling method is justified by the idea that if the input image is larger, then the network needs more layers to increase the receptive field and more channels to capture finer patterns on the larger image. The network of architectural features is BiFPN (Bidirectional Feature Pyramid Network), one of the innovations of the model, which aims to aggregate features at different resolutions. With a list of features of different sizes, the goal is to find a transformation that can effectively assemble these features and provide a new list of features. The characteristics obtained at this level are taken over by a network that outputs the object class together with the rectangle surrounding the object [26-28].

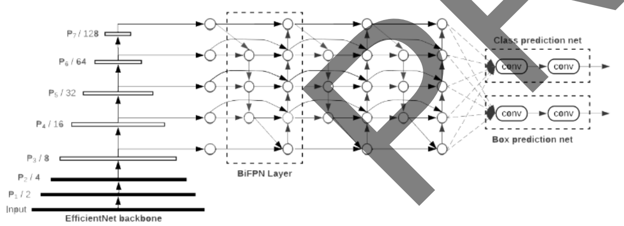

Figure 3. EfficientDet BiFPN architecture

3. Results

3.1. Dataset

In this retrospective study, a hundred patients with confirmed one-vessel coronary artery disease underwent coronary angiography using Siemens Coroscop and GE Healthcare's Innova at the Research Institute for Complex Problems of Cardiovascular Diseases in Kemerovo, Russia [15]. The patients provided written informed consent, and angiographic images were retrospectively collected and processed.

A total of 8325 grayscale images, ranging from 512 × 512 to 1000 × 1000 pixels, were selected, with 80% used for training, 10% for validation, and 10% for testing. The non-randomized data split ensured independent subsets for validation and testing, preventing bias in performance metrics and enabling joint data labeling by multiple specialists. In the context of our study, it is important to note that coronary artery stenosis typically affects only a small portion of the vessel, while the majority of the artery remains unaffected. As a result, even within images labeled as containing stenosis, there are extensive regions of the vessel that are free of any narrowing (Figure 4). These non-stenotic regions inherently provide the model with examples analogous to frames without stenosis. Consequently, the model is implicitly exposed to features of non-stenotic areas during training, enabling it to distinguish between stenotic and non-stenotic regions within a single frame.

To assess the source dataset, the stenotic region's size is estimated by computing bounding box areas. Similar to the COCO dataset, objects are categorized by area into small, medium, and large. Table 1 presents the clinical and demographic characteristics

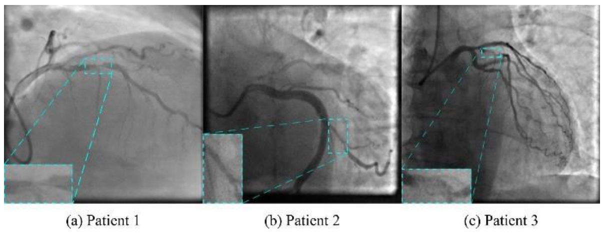

Figure 4. Data labeling of the source images

Table 1. Clinical and demographic data of the study population

Parameter	Value
Total number of patients	100
Mean age ± SD, years	60.3 ± 13.8
Men, n (%)	68 (68%)
Women, n (%)	32 (32%)
Body mass index (kg/m2)	21.6 ± 5.1
Diagnosis	CAD
Class I NYHA	5 (5%)
Class II NYHA	84 (84%)
Class III NYHA	11 (11%)
Comorbidities	
Arterial hypertension	53 (53%)
Diabetes mellitus	14 (14%)
Chronic heart failure, classes 1–2	36 (36%)
Coronary artery stenosis > 70% (n, %)	(100%)

of the study participants, providing comprehensive information about the population under investigation.

3.2. Pre-Processing

In this study, a series of preprocessing steps was applied to the input images to ensure they were optimally prepared for training with RetinaNet-101 and EfficientDet-D3. These steps were designed to standardize the data, enhance the models' robustness to variations in the images, and address the specific challenges posed by the noisy and high-resolution nature of X-ray angiography images.

For RetinaNet-101, the preprocessing began with resizing all input images to a fixed dimension of 640 × 640 pixels. This resizing ensured uniformity across the dataset and compatibility with the model's input layer. Data augmentation techniques were also employed to improve the diversity and generalizability of the training data. A random horizontal flip was applied to introduce variability in object orientation, enabling the model to handle directional changes effectively. Additionally, random cropping was performed to extract smaller regions of the input images. This augmentation strategy helped the model learn to detect features of varying sizes and positions, particularly small and localized features such as stenosis.

For EfficientDet-D3, preprocessing involved scaling the images to a larger dimension of 896 × 896 pixels. This allowed the model to capture finer details that are critical for detecting subtle features in the angiographic images. The resizing process was complemented by a more complex augmentation strategy known as "random scale, crop, and pad to square." This method involved scaling the image to a random size, cropping a square patch, and padding as necessary to achieve the target input dimensions. This approach not only standardized the input size but also improved the model's ability to handle images of varying scales and compositions. Like RetinaNet-101, random horizontal flipping was used to further augment the data and introduce variability.

These preprocessing techniques were carefully selected to address the specific challenges of the dataset while maximizing the models' performance. By standardizing image dimensions and applying targeted augmentations, the models were better equipped to generalize across noisy and high-resolution medical images, improving their robustness and accuracy in detecting coronary artery stenosis.

3.3. Evaluation Metrics

In this work, mAP has been used as a metric to check the model's performance. It is a common metric used to evaluate object detection models and tells how good the localization of an object is within an image. This work focuses on X-ray images; hence, the mAP score comments on how effective the localization of opacity is within the chest. The expression of Average Precision (AP) calculates the average of AP, giving mAP. Precision is a metric that measures how accurate the predictions are, i.e., the percentage of predictions that are correct, whereas recall measures the percentage of actual positives how many are identified correctly (Equation 2):

$$AP = \frac{1}{n} * \sum_{i=o}^{n} P * Ri \tag{2}$$

Where P is Precision, Ri are the values of recalls from 0 to n, where n is an integer value. The better the mAP score of predictions, the better the model [29].

3.4. Hyperparameters and GPU Devices

In object detection, hyperparameter tuning refers to the process of selecting the optimal values for the various parameters and settings used in the training of an object detection model. These parameters and settings, which are often referred to as hyperparameters, can have a significant impact on the performance of the model, including factors such as accuracy, precision, and recall. Hyperparameters and GPU models in object detection can include settings related to the architecture of the model, such as the number and size of the convolutional layers, as well as parameters related to the training process, such as the learning rate, the batch size, and the number of training epochs [30, 31]. A step refers to one operation for the optimizer to update the weights of the model. The relationships between the number of steps to epochs and batch size are (Equation 3):

$$num\ of\ steps\ = (epochs * num\_of\_training\ imgs) / batch\_size \tag{3}$$

Table 2 furnishes a summary of the hyperparameters employed in training the proposed model.

The models used in this research were executed on two different systems. RetinaNet was run on a local system with a GeForce RTX 3080 Ti GPU (12 GB VRAM) and 16 GB of RAM. EfficientDet3, on the other hand, was executed on Google Colab with an A100 GPU and 83 GB of RAM. Both systems utilized Python version 3.9 and TensorFlow version 2.13.

Table 2. Hyperparameters

RetinaNet 101	
Parameter	Value
Num classes	1
Height * width	640 * 640
regularizer	l2 regularizer
gamma	2.0
alpha	0.25
Batch size	4
Data augmentation options	Random Horizontal Flip, Random Crop Image
Momentum optimizer value	0.9
Learning rate base	0.0025
Total steps	65000
GPU (Local system)	Geforce RTX 3080 Ti 12GB
Ram	16 GB
EfficientDet D3	
Parameter	Value
Num classes	1
Min dimension max dimension	896 * 896
regularizer	l2 regularizer
gamma	1.5
alpha	0.25
Batch size	4
Data augmentation options	Random Horizontal Flip, random scale crop and pad to square
Momentum optimizer value	0.9
Learning rate base	0.0025
Total steps	25000
GPU (google colab)	NVIDIA A100 SXM4 40 GB
Ram	83 B

3.5. Implementation Details

The object detection accuracy depends on the model used. To improve the accuracy, it is possible to train the model. TensorFlow allows using custom models by training custom datasets. The object detection library has the tools needed to export a new model or an updated model. Initially, we convert BMP files to PNG format, and the corresponding XML files are converted to TFRecord files.

The TensorFlow training data-set format is TFRecord. A TFRecord file stores the data as a sequence of binary strings, this format enabled us to encode all the information needed into a single file. With the help of TFRecord, we never needed to read files from different directories and worry about inconsistent file format problems [32, 33].

TensorFlow supports a wide range of tensor models. The use of a more complex model will yield higher accuracy results, while training the model will improve the accuracy of trained objects. This means that a basic model can achieve high-accuracy results when focused on certain objects [33]. In this work, we used the Table 3 pre-trained SSD models based on the TensorFlow 2 Detection Model Zoo.

Then, the default parameters were modified to suit our specific model by utilizing a configuration file based on Table 2 and specifying the paths to the TFRecord files. The model config block is about the configuration of a model. These config files are used to configure parameters for the initial setting of some of the computer codes.

The fine-tuned model was fed to TensorFlow Object Detection API. After training models on data separately, we tested the trained model on eval data. For mAP, a predefined threshold value for Intersection over Union equal to 0.5 was used.

Table 3. Fine-tuned Models

Model	**Speed (ms)**	**COCO mAP**
EfficientDet D3 896x896	95	45.4
SSD ResNet101 V1 FPN 640x640 (RetinaNet101)	57	35.6

4. Discussion

In this section, we present the findings of our study on the performance of advanced deep learning models in detecting coronary artery stenosis from medical images. As mentioned, EfficientDet D3 and SSD RetinaNet101 architectures, which are known for their state-of-the-art capabilities in object detection tasks, were used to assess their effectiveness specifically in the context of coronary artery disease.

We obtained the following results (Table 4) based on Detection Boxes Precision/mAP@.50IOU. The sample of obtained results is also illustrated in Figure 5.

Table 4. Results comparison

Model	Computational Performance	Task
Ovalle-Magallanes *et al.* [19]	Precision: 93	Classification
Antczak and Liberadzki [14]	Accuracy: 90	Classification
Moon *et al.* [17]	Accuracy: 94.3	Classification
Pang *et al.* [14]	Precision:94.87	Detection
Danilov *et al.* [15].	mAP: 95.4	Detection
Li *et al.* [16]	F1 score: 92.39	Detection
EfficientDet D3	**mAP: 96.6**	**Detection**
SSD RetinaNet101	**mAP: 93.2**	**Detection**

Based on the empirical results, our study contributes significantly to the field of medical image analysis by demonstrating the exceptional capabilities of advanced deep learning models, particularly EfficientDet D3 and SSD RetinaNet101, in the detection of coronary artery stenosis. Our EfficientDet D3 model achieved a mean Average Precision (mAP) of 96.6%, outperforming existing methodologies, including Danilov *et al.* [15] with an mAP of 95.4% and Li *et al.* [16] with an F1 score of 92.39. This marked improvement is critical for clinical applications where high precision and accuracy are paramount for timely diagnosis and intervention. Moreover, the SSD RetinaNet101 model also showcased impressive performance with an mAP of 93.2%, highlighting the robustness of our approach against other contemporary models.

By leveraging these advanced architectures, our research not only provides a comprehensive comparison against state-of-the-art detection methods but also emphasizes their versatility and effectiveness in handling complex medical imaging scenarios. These results illustrate the potential for implementing such models in real-time clinical settings, where detecting small and localized features of stenosis can significantly impact patient outcomes. Furthermore, our findings lay a substantial foundation for future investigations aimed at optimizing deep learning frameworks for medical applications, encouraging further exploration of automated diagnostic tools that can enhance cardiovascular health management.

4.1. Limitations

One limitation of this study is the variability in X-ray angiography images acquired from different devices, each producing images with distinct resolutions and noise levels. Although our method was tested on a diverse dataset, it primarily relied on single-center data, which may limit the generalizability of the results to other clinical environments. Furthermore, while the proposed models strike a balance between speed and accuracy,

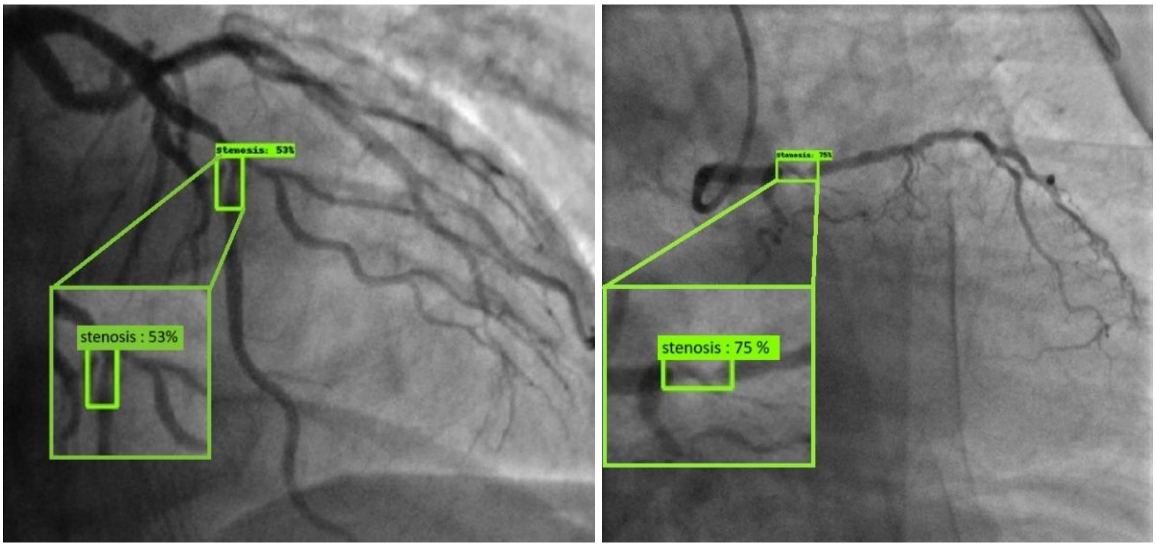

Figure 5. Samples of obtained results

there is still room for improvement in handling extreme cases of image quality degradation or atypical anatomical structure.

To further enhance the model's generalization and diagnostic accuracy, we propose augmenting the dataset by including additional entirely non-stenotic frames. While the current dataset already offers a blend of stenotic and non-stenotic regions, incorporating more purely non-stenotic frames would give the model a broader understanding of the vessel characteristics, ultimately improving its performance in distinguishing between stenotic and non-stenotic regions in real-world applications. This extension of the dataset would help reduce false positives and ensure more reliable predictions, especially in cases where no stenosis is present.

5. Conclusion

This study demonstrates the effectiveness of advanced deep learning models, specifically EfficientDet D3 and SSD ResNet101, in accurately detecting stenosis in coronary angiograms with high precision. The EfficientDet D3 model achieved an impressive mean average precision (mAP) of 96.6%, while the SSD ResNet101 model achieved an mAP of 93.2%, both surpassing the detection capabilities of prior methods. These results underline the potential of these models in identifying stenotic segments with exceptional reliability. The findings highlight several key contributions of this research:

- Automated Detection for Enhanced Accuracy: By leveraging pre-trained models, this study addresses the challenge of accurately detecting stenosis locations and reduces the reliance on operator expertise. The models consistently demonstrated high confidence in identifying stenotic segments with a 50% overlap threshold, thereby improving diagnostic precision.

- Benchmarking Against Existing Methods: The study provides a comprehensive comparison with existing models, showcasing the superior performance of EfficientDet D3 and SSD ResNet101 in detection tasks. With mAP values of 96.6% and 93.2%, respectively, these models set a new standard for detection accuracy in the domain of coronary angiogram analysis.

- Practical Implications for Clinical Settings: Given that this project is ultimately intended to be implemented as a hospital-grade application, both speed and accuracy are of paramount importance. The proposed algorithm has demonstrated its capability to meet these requirements effectively, showcasing its potential to streamline diagnostic processes and enhance decision-making efficiency in clinical environments.

In conclusion, this study successfully demonstrates the applicability and robustness of deep learning models in addressing the complexities of stenosis detection in coronary angiograms. The proposed algorithm not only achieves state-of-the-art detection accuracy but also paves the way for future advancements in automated medical imaging analysis. Further research and validation in diverse clinical settings will ensure the scalability and generalisability of this approach, ultimately contributing to improved patient outcomes and streamlined healthcare workflows.

Future work should focus on incorporating multi-center datasets that capture a broader range of imaging devices and patient populations. Additionally, the integration of domain adaptation techniques could further improve the model's performance across varying imaging conditions. Considering that X-ray angiography images are highly noisy and the presence of bone structures, spinal vertebrae, catheters, and diaphragm curtains exacerbates this issue, future research will also focus on these aspects to generalize the proposed models to images of other patients. This research aims to assist in developing a tailored program for a clinical decision support system, which would help in providing assistance in the interpretation of images from various patients and enhance their overall quality.

Acknowledgment

The authors received no financial support for the research, authorship, and/or publication of this article.

References

1- Haidong Wang *et al.*, "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015." *The lancet,* Vol. 388 (No. 10053), pp. 1459-544, (2016).

2- Mojdeh Nazari, Hassan Emami, Reza Rabiei, Azamossadat Hosseini, and Shahabedin Rahmatizadeh, "Detection of Cardiovascular Diseases Using Data Mining Approaches: Application of an Ensemble-Based Model." *Cognitive Computation,* pp. 1-15, (2024).

3- JG Myers, JA Moore, M Ojha, KW Johnston, and CR Ethier, "Factors influencing blood flow patterns in the human right coronary artery." *Annals of biomedical engineering,* Vol. 29pp. 109-20, (2001).

4- Erling Falk, "Pathogenesis of atherosclerosis." *Journal of the American College of cardiology,* Vol. 47 (No. 8S), pp. C7-C12, (2006).

5- Hossein Sadr, Arsalan Salari, Mohammad Taghi Ashoobi, and Mojdeh Nazari, "Cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models." *European Journal of Medical Research,* Vol. 29 (No. 1), p. 455, (2024).

6- Zahra Alkan Saberi, Hossein Sadr, and Mohammad Reza Yamaghani, "An Intelligent Diagnosis System for Predicting Coronary Heart Disease." in *2024 10th International Conference on Artificial Intelligence and Robotics (QICAR)*, (2024): *IEEE*, pp. 131-37.

7- Patrick Kierkegaard, "Electronic health record: Wiring Europe's healthcare." *Computer law & security review,* Vol. 27 (No. 5), pp. 503-15, (2011).

8- Amanda H Gonsalves, Fadi Thabtah, Rami Mustafa A Mohammad, and Gurpreet Singh, "Prediction of coronary heart disease using machine learning: an experimental analysis." in *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies*, (2019), pp. 51-56.

9- Zeinab Khodaverdian, Hossein Sadr, and Seyed Ahmad Edalatpanah, "A shallow deep neural network for selection of migration candidate virtual machines to reduce energy consumption." in *2021 7th International conference on web research (ICWR)*, (2021): *IEEE*, pp. 191-96.

10- Mojdeh Nazari, Shadi Moayed Rezaie, Fereshteh Yaseri, Hossein Sadr, and Elham Nazari, "Design and analysis of a telemonitoring system for high-risk pregnant women in need of special care or attention." *BMC Pregnancy and Childbirth,* Vol. 24 (No. 1), p. 817, 2024/12/18 (2024).

11- Emmanuel Ovalle-Magallanes, Juan Gabriel Avina-Cervantes, Ivan Cruz-Aceves, and Jose Ruiz-Pinales, "Hybrid classical–quantum Convolutional Neural Network for stenosis detection in X-ray coronary angiography." *Expert Systems with Applications,* Vol. 189p. 116112, (2022).

12- Shirin Dehghan, Reza Rabiei, Hamid Choobineh, Keivan Maghooli, Mozhdeh Nazari, and Mojtaba Vahidi-Asl, "Comparative study of machine learning approaches integrated with genetic algorithm for IVF success prediction." *Plos one,* Vol. 19 (No. 10), p. e0310829, (2024).

13- Emmanuel Ovalle-Magallanes, Dora E Alvarado-Carrillo, Juan Gabriel Avina-Cervantes, Ivan Cruz-Aceves, Jose Ruiz-Pinales, and Rodrigo Correa, "Deep Learning-based Coronary Stenosis Detection in X-ray Angiography Images: Overview and Future Trends." *Artificial Intelligence and Machine Learning for Healthcare: Vol. 2: Emerging Methodologies and Trends,* pp. 197-223, (2022).

14- Kun Pang, Danni Ai, Huihui Fang, Jingfan Fan, Hong Song, and Jian Yang, "Stenosis-DetNet: Sequence consistency-based stenosis detection for X-ray coronary angiography." *Computerized Medical Imaging and Graphics,* Vol. 89p. 101900, (2021).

15- Viacheslav V Danilov *et al.*, "Real-time coronary artery stenosis detection based on modern neural networks." *Scientific reports,* Vol. 11 (No. 1), p. 7582, (2021).

16- X. Li *et al.*, "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,* Vol. 46 (No. 12), pp. 9908-20, (2024).

17- Jong Hak Moon, Won Chul Cha, Myung Jin Chung, Kyu-Sung Lee, Baek Hwan Cho, and Jin Ho Choi, "Automatic stenosis recognition from coronary angiography using convolutional neural networks." *Computer methods and programs in biomedicine,* Vol. 198p. 105819, (2021).

18- Karol Antczak and Łukasz Liberadzki, "Stenosis detection with deep convolutional neural networks." in *MATEC web of conferences*, (2018), Vol. 210: *EDP Sciences*, p. 04001.

19- Emmanuel Ovalle-Magallanes, Juan Gabriel Avina-Cervantes, Ivan Cruz-Aceves, and Jose Ruiz-Pinales, "Transfer learning for stenosis detection in X-ray coronary angiography." *Mathematics,* Vol. 8 (No. 9), p. 1510, (2020).

20- Hafsa Ouchra and Abdessamad Belangour, "Object detection approaches in images: a survey." in *Thirteenth International Conference on Digital Image Processing (ICDIP 2021)*, (2021), Vol. 11878: *SPIE*, pp. 132-41.

21- N Bilous, V Malko, M Frohme, and A Nechyporenko, "Comparison of CNN-Based Architectures for Detection of Different Object Classes. AI 2024, 5, 2300–2320." ed, (2024).

22- Samain Abid, Muhammad Haris, Mansoor Iqbal, Nadia Khan, Khalid Munir, and Hamza Yousaf, "Comparative Analysis of Machine Learning Pre-Trained Object Detection Models: Performance, Efficiency, and Application Suitabilit." in *2024 International Conference* *on Electrical, Communication and Computer Engineering (ICECCE)*, (2024): *IEEE*, pp. 1-6.

23- Zihao Zhou, Yang Li, Chenglei Peng, Hanrong Wang, and Sidan Du, "Image processing: Facilitating retinanet for detecting small objects." in *Journal of Physics: Conference Series*, (2021), Vol. 1815 (No. 1): *IOP Publishing*, p. 012016.

24- Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature pyramid networks for object detection." in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2017), pp. 2117-25.

25- Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection." in *Proceedings of the IEEE international conference on computer vision*, (2017), pp. 2980-88.

26- Mingxing Tan, Ruoming Pang, and Quoc V Le, "Efficientdet: Scalable and efficient object detection." in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (2020), pp. 10781-90.

27- Dan Munteanu, Diana Moina, Cristina Gabriela Zamfir, Ştefan Mihai Petrea, Dragos Sebastian Cristea, and Nicoleta Munteanu, "Sea mine detection framework using YOLO, SSD and EfficientDet deep learning models." *Sensors,* Vol. 22 (No. 23), p. 9536, (2022).

28- Brett Koonce and Brett Koonce, "EfficientNet." *Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization,* pp. 109-23, (2021).

29- Vinayak Tiwari, Amit Singhal, and Nischay Dhankhar, "Detecting COVID-19 Opacity in X-ray Images Using YOLO and RetinaNet Ensemble." in *2022 IEEE Delhi Section Conference (DELCON)*, (2022): *IEEE*, pp. 1-5.

30- Carina Isabel Andrade Albuquerque, "Object Detection in medical imaging." (2023).

31- Ming Zhou, Bo Li, and Jue Wang, "Optimization of Hyperparameters in Object Detection Models Based on Fractal Loss Function." *Fractal and Fractional,* Vol. 6 (No. 12), p. 706, (2022).

32- Fengkai Wan, "Deep Learning Method Used in Skin Lesions Segmentation and Classification." ed, (2018).

33- Mohammed Noman, Vladimir Stankovic, and Ayman Tawfik, "Object detection techniques: Overview and performance comparison." in *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, (2019): *IEEE*, pp. 1-5.