

Article

The Application of Deep Learning for the Segmentation and Classification of Coronary Arteries

Şerife Kaba ¹, Huseyin Haci ^{2,*}, Ali Isin ³, Ahmet Ilhan ⁴ and Cenk Conkbayir ⁵

¹ Department of Biomedical Engineering, Near East University, TRNC Mersin 10, Nicosia 99138, Turkey; serife.kaba@neu.edu.tr

² Department of Electrical-Electronic Engineering, Near East University, TRNC Mersin 10, Nicosia 99138, Turkey

³ Department of Biomedical Engineering, Cyprus International University, TRNC Mersin 10, Nicosia 99138, Turkey; aisin@ciu.edu.tr

⁴ Department of Computer Engineering, Near East University, TRNC Mersin 10, Nicosia 99138, Turkey; ahmet.ilhan@neu.edu.tr

⁵ Department of Cardiology, Near East University, TRNC Mersin 10, Nicosia 99138, Turkey; cenkconk@hotmail.com

* Correspondence: huseyin.haci@neu.edu.tr

Abstract: In recent years, the prevalence of coronary artery disease (CAD) has become one of the leading causes of death around the world. Accurate stenosis detection of coronary arteries is crucial for timely treatment. Cardiologists use visual estimations when reading coronary angiography images to diagnose stenosis. As a result, they face various challenges which include high workloads, long processing times and human error. Computer-aided segmentation and classification of coronary arteries, as to whether stenosis is present or not, significantly reduces the workload of cardiologists and human errors caused by manual processes. Moreover, deep learning techniques have been shown to aid medical experts in diagnosing diseases using biomedical imaging. Thus, this study proposes the use of automatic segmentation of coronary arteries using U-Net, ResUNet-a, UNet++, models and classification using DenseNet201, EfficientNet-B0, Mobilenet-v2, ResNet101 and Xception models. In the case of segmentation, the comparative analysis of the three models has shown that U-Net achieved the highest score with a 0.8467 Dice score and 0.7454 Jaccard Index in comparison with UNet++ and ResUnet-a. Evaluation of the classification model's performances has shown that DenseNet201 performed better than other pretrained models with 0.9000 accuracy, 0.9833 specificity, 0.9556 PPV, 0.7746 Cohen's Kappa and 0.9694 Area Under the Curve (AUC).

Keywords: coronary artery disease (CAD); coronary arteries; angiography; U-Net; pretrained models



Citation: Kaba, Ş.; Haci, H.; Isin, A.; Ilhan, A.; Conkbayir, C. The Application of Deep Learning for the Segmentation and Classification of Coronary Arteries. *Diagnostics* **2023**, *13*, 2274. <https://doi.org/10.3390/diagnostics13132274>

Academic Editor: Andreas Kjaer

Received: 16 June 2023

Revised: 30 June 2023

Accepted: 3 July 2023

Published: 5 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Coronary Artery Disease (CAD) is one of the most common forms of cardiovascular disease affecting the human population globally. CAD is the leading cause of death in both developed and developing countries [1]. About 30% of all deaths worldwide are due to cardiovascular disease. The disease is also considered one of the most frequent causes of death in Europe on an annual basis [2].

The coronary arteries supply blood to the heart muscles. This is an important function for the heart muscle and thus for the body. The Left Main Coronary Artery (LMCA) and the Right Coronary Artery (RCA) are the primary types of coronary arteries. The LMCA carries blood to the left part of the heart, whereas the RCA is mainly responsible for carrying blood to the right side of the heart. LMCA is composed of two major branches. The Left Anterior Descending artery (LAD) is one of the major branches of the LMCA that transports blood to the left anterior side of the heart. Another branch of the LMCA is the Circumflex Artery (Cx) which surrounds the heart muscle (myocardium). Through this artery, the outside and the back part of the heart are supplied with blood [3,4].

Most heart diseases are caused by atherosclerosis, which causes the arteries to narrow or be blocked due to the build-up of plaque. CAD occurs when plaque (fatty deposits) builds up, narrowing (stenosis) the passageway and interfering with the heart's supply of blood, oxygen, and other vital nutrients. Plaque build-up can block the coronary arteries and prevent nutrients and oxygen from reaching the heart muscle, which can lead to myocardial infarction (heart attacks) and sometimes death. When plaque is deposited on the wall of a coronary artery, the lumen of that artery is affected. Therefore, it is noteworthy that the diagnosis of stenosis can be made by measuring the lumen diameter with an angiography imaging system [2,4]. The differences between normal and blocked arteries are illustrated in Figure 1.

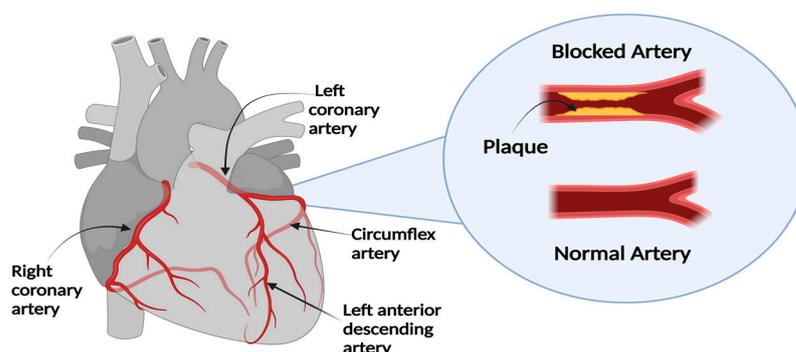


Figure 1. Normal versus blocked Artery.

1.1. Diagnosis and Treatment of CAD

Early diagnosis of CAD significantly reduces mortality and morbidity rates [5]. In fact, for decades, CAD has been diagnosed and treated using interventional radiology procedures or surgical operations [6].

One of these procedures is Invasive Coronary Angiography (ICA) uses a catheter, and X-ray imaging and remains the “gold standard” for the diagnosis of coronary artery stenosis [7,8]. The primary function of the catheters used in this imaging technique is to monitor blood pressure in real-time. Blockage (occlusion) or stenosis in human arteries can be detected with the help of X-rays in which a contrast material is injected through the catheter [2]. ICA provides the cardiologist with information about the severity of luminal stenosis and is also used to diagnose atherosclerotic disease. For example, angiography can reveal calcifications, plaque ruptures severe lesions and luminal thrombosis [4].

In addition, ICA is a decision-making aid for revascularization treatments such as stent placement and coronary artery bypass surgery [9].

1.2. Challenges

Cardiologists view frames in angiographic videos with their naked eyes during angiography procedures in order to evaluate various problems affecting blood vessels, such as stenosis, plaque and blockage. Based on their experience, cardiologists can diagnose the stenosis of coronary arteries and decide medical treatment options for CAD (i.e., stent placement). However, this direct method, which requires more accuracy, objectivity, and consistency, is heavily influenced by human factors [6].

On the other hand, automated segmentation of the cardiovascular system has been reported to reduce the error rate in the diagnosis of stenosis. In fact, an image segmentation method has already been proposed to extract blood vessels from images so that cardiologists can quickly diagnose plaque and stenosis.

Recently, with the development of Deep Learning (DL), various Convolutional Neural Network (CNN) architectures have been proposed and used for image segmentation [10]. CNNs are also typically used for feature extraction, image recognition, image formation, and image-based rendering. These capabilities are essential elements of the technological processes of clinical data acquisition and processing. In addition, image enhancement,

image classification, and visual extraction techniques, including pattern classification, have also become very popular. The accuracy and robustness of image enhancement can also be measured using DL techniques [11]. The primary goal of using DL networks in medical image analysis is to classify images into different classes or categories (such as positive or negative, abnormal or normal, etc.). This involves analyzing the input images and classifying the output in order to determine whether or not a particular disease is present [12].

1.3. Contribution

The main research question of this study revolves around how segmentation and classification models can help increase stenosis classification performance and thus relieve the workload of cardiologists in Northern Cyprus hospitals. Consequently, the objective of this study is tailored toward applying these models to the classification of patient images.

It is essential to mention that in this study, no artificial data were generated by data augmentation. On the contrary, all images used were clinical images and none were synthesized. Instead of developing a network from scratch, which requires a substantial amount of data, pretrained models (based on transfer learning) are used only for classification tasks, where learned parameters are transferred to solve a problem in another task. [13].

It is noteworthy to emphasize that the system proposed in this study is intended to aid cardiologists in conducting an automated and accuracy-increased diagnosis of stenosis. In the healthcare field, it is vital to minimize diagnostic errors. For this reason, decision accuracy can be significantly improved in the case that the experience and instinct of the doctors are complemented with novel automated technologies, such as Artificial Intelligence (AI). An automated diagnosis system can provide a second (additional) opinion/verification. It should be noted that using such systems on their own can result in misdiagnosis in problematic/rare cases since they do not have human-like instincts [6,11].

Thus, this study suggests automated diagnosis technologies be used in conjunction with health professionals.

In this study, based on the U-Net model and its variants (ResUNet-a and UNet++), we applied the DL approach to segment the major coronary arteries, including the RCA, LMCA, LAD, and Cx. Finally, five different pretrained models are used to automatically classify angiographic images with or without stenosis of the major coronary arteries. The study's main contributions are highlighted below:

- The application of UNet, Unet++ and ResUNet-a for the automatic segmentation of coronary angiograms.
- Comparison of model's performances for both segmentation and classification tasks.
- One of the major contributions of this study is the application of DL-based transferred learning using several pretrained models (EfficientNet-B0, DenseNet201, Mobilenet-v2, ResNet101 and Xception) for the classification of coronary angiograms.
- Comparison between the performance of models trained on the coronary artery small datasets can aid cardiologists in the selection of the best-performing model and also aid them in making appropriate decision-making.
- Another contribution of this study is the use of raw/unaltered data that are obtained from real patients by the cardiology department of Near East University Hospital, instead of using a dataset curated from an online repository system. We have noticed that a significant number of images available from online repository systems have been altered (i.e., cropped, rotated, and enlarged) to aid the segmentation and classification performance of coronary arteries. However, this is not the case in clinical applications.
- Performance evaluation of models based on accuracy, sensitivity, specificity, precision, Dice Score (F1 Score), Jaccard Index and Matthews correlation coefficient (MCC), negative predictive value (NPV), Cohen's kappa, Area Under Curve (AUC) and Receiver Operating Characteristic (ROC) curve.

1.4. Related Work

This section briefly reviews previous methods relevant to our work, including conventional medical image segmentation of coronary arteries, DL-based segmentation, and image classification and diagnosis of coronary stenosis.

Various methods have been proposed for the automatic segmentation of blood vessels based on traditional techniques. The following methods are commonly used in image processing: Filtering, thresholding, tracking-based and region growing. To illustrate, typical filtering methods utilize Gaussian and Gabor filters to extract information about edges iteratively [14]. Moreover, in tracking-based segmentation methods, seed points are selected at vessel edges, and then tracking is conducted under the guidance of constraints derived from the image. Tracking-based methods often create deformable models to capture the shape variations of coronary vessels in angiogram sequences [15].

In addition, Kerkeni et al. [16] presented a multiscale region-growing method for coronary artery segmentation. In line with this, a region-growing rule was developed that is known to uniquely integrate vessel and directional information into the region-growing approach. An iterative multiscale search is then performed based on this criterion. It is important to note that the points selected in each step serve as seed cells for the next step.

Although these traditional image segmentation methods have improved the segmentation of vessel structures in image analysis, they could be less efficient in clinical practice due to the high computational complexity of pixel-by-pixel analysis. DL techniques using CNNs have recently gained more attention in medical imaging due to their promising results in vessel segmentation [17–21].

There are various CNN architectures that have been proposed for image segmentation. The most commonly used architectures are U-Net and SegNet models. SegNet is a DL semantic segmentation model that was originally developed for modeling scene segmentors and pixel-wise classification layers. It is an efficient architecture, especially for road and building image segmentation [22,23].

Zhao et al. [24] proposed a semantic segmentation algorithm to extract coronary arteries and classify them into LCA, LAD and other types of arterial segments based on the Support vector machine (SVM) algorithm using 225 angiography images. The result of the study achieved the mean accuracy of multi-class artery classification was 70.33% and the mean intersection over union for semantic segmentation of arteries was 0.6868 [24].

The U-Net model is considered the most successful architecture for medical image segmentation competitions [25]. The U-Net model is a CNN that enables efficient segmentation of biomedical data. Despite its relatively simple architecture and high accuracy in binary segmentation problems, it is also used in many other areas [26]. In addition, CNNs are used to detect specific features and to distinguish elements in the image according to their context in the original image. Several applications of these algorithms have shown promising results, including medical image processing, image enhancement, classification, and segmentation. Among the most common approaches with regard to CNNs are pretrained neural networks. This technique is conducted in order to increase the accuracy rate while significantly reducing the training time, which also increases the effectiveness of the CNN. The result of this process is what is known as transfer learning [11,27]. Among the known CNN models, GoogleLeNet is becoming increasingly popular due to its efficiency in visual object classification and recognition. The AlexNet model became the most popular after winning the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) competition in 2012, and future networks are typically compared to its achievements [28].

Ever since, several models have been developed and tested using ImageNet datasets. Some of the models that achieved high accuracy scores include VGGNet, Inception, ResNet, DenseNet, etc. In comparison with traditional machine learning (ML) models, DL models can automatically learn features and patterns, and classify different types of images with higher generalization. DL models have been shown to aid healthcare experts in terms of classifying medical images of diseases such as different kinds of cancers, tuberculosis and pneumonia [29].

Furthermore, data augmentation is often used to expand the range of data available to CNNs. For example, Antczak and Liberadzki [30] proposed that CNNs be used to detect stenosis. The study utilized several data augmentation techniques in order to increase the training sets using images acquired from online repositories. The developed CNN model was capable of generating random (positive and negative) stenosis patches with a Gaussian kernel. The performance evaluation of the model resulted in an overall accuracy of 90% [30]. Gil-Rios, Miguel-Angel, and colleagues proposed the use of a Support Vector Machine (SVM) for the detection of coronary stenosis from the Antczak and Liberadzki image dataset as well as the dataset of the Mexican Social Security Institute. The model performance yielded remarkable results [31]. Ovalle-Magallanes, Emmanuel, and colleagues presented a method to automatically detect coronary artery stenosis based on X-ray coronary angiograms. The images were trained using a pretrained CNN on a large amount of data, including synthetic and clinical data. The results obtained demonstrated the ability of pretrained networks in classifying coronary artery stenosis from coronary angiography images [32].

The study conducted by Lee et al. [33] applied deep learning for the quantification of Coronary Artery Calcium (CAC). The approach was designed according to ECG-gated coronary CT angiography (CCTA) with reference to coronary calcium scoring CT (CSCT). The data used in this retrospective study were acquired from 315 patients who underwent both CCTA and CSCT. The patients were divided into two groups, 200 in the internal and 115 in the external validation tests. The outcome of the study has shown that the automated algorithms were able to extract CACs in less than 300 s mean time with a 1.3% failure rate. The model's volume and Agatston scores have displayed high agreement with those from CSCT with concordance correlation coefficients of 0.76–0.94 for external sets and 0.90–0.97 for external sets. In terms of the model classification performance, the study was able to record 0.94 weighted kappa and 92% accuracy for the internal sets and 0.91 weighted kappa and 86% accuracy for the external sets.

Jamil and Roy [34] proposed the application of DL models for the detection of Valvular Heart Disease (VHD). The study employed a relatively simpler DL model in terms of network structure and performance. In order to train and validate the model performance, the study employed three different frameworks which include 1D and 2D Phonocardiography (PCG) raw signals. The frameworks revolve around the use of linear prediction cepstral coefficients (LPCC) and Mel frequency cepstral coefficients (MFCC) features for 1D PCG signal while the second framework revolves around the use several deep convolutional neural networks (D-CNNs) features for 2D PCG signals. The third framework revolves around the application of nature/bio-inspired algorithms (NIA/BIA) which include two algorithms namely, genetic algorithm (GA) and particle swarm optimization (PSO) for efficient and automatic selection of raw PCG signals. In addition, a vision transformer (ViT) was implemented in order to improve the performance of the classifiers. The evaluation of DL-based algorithms and classifiers has shown that ViT achieved the best result with an average accuracy of 99.90% and 99.95% F1-score.

Considering the fact that the manual approach for quantitatively assessing the stenosis severity of coronary arteries is very tedious and is subject to the experience of cardiologists; therefore, developing automatic quantitative coronary analysis (QCA) is crucial. In order to address this concern, Liu et al. [35] proposed the application of artificial-intelligent-based QCA also known as AI-QCA for the accurate and fast quantitative assessment of the severity of stenosis. The framework is designed according to three main units which include the boundary-aware segmentation on the coronary angiogram (CAG) images followed by the construction of the coronary artery tree which is enabled by AI and lastly the diameter fitting and detection of stenosis. The evaluation of the AI-QCA based on segmentation resulted in a precision value of 0.866, recall of 0.897 and F1-score of 0.879 using 1322 CAGs. A comparison of the diameter assessment of 249 CAGs between AI-QCA and senior experts using Root Mean Square Error (RMSE) resulted in 0.064 and 0.765 Pearson coefficients.

Moreover, the use of AI-QCA has been shown to reduce the operation time from tens of minutes to a few seconds. The summary of the related work is presented in Table 1.

Table 1. Literature review table.

Reference	Image Acquisition	No. of Images	Method	Results
[14]	Cardiology Department of the Mexican Social Security Institute	130 coronary angiograms	Multiscale versions of the Gaussian filter and Gabor filter	ACC: 0.9698, Sens: 0.6364, Spec: 0.9880, PPV: 0.7434, Dice coefficient: 0.6857
[16]	Department of The University Hospital Fattouma Bourguiba, Monastir, Tunisia	50 coronary angiograms	Sato filter, Vessel enhancing diffusion filter, and Frangi filter Multiscale region growing	Mean precision: 82%, First dataset of Dice coefficient: $80 \pm 5\%$, second dataset of Dice coefficient: $70 \pm 5\%$
[18]	Asan Medical Center (Internal) Chungnam National University Hospital (External)	Internal:3302 coronary angiograms, External:181 coronary angiograms	U-Net, ResNet101, DenseNet121, InceptionResNet-v2	Average F1 score: 0.917, and 93.7% of the image
[19]	University of Michigan Hospital	462 coronary angiograms	Convolutional neural network, AngioNet	Dice score: 0.864, pixel accuracy (PA):0.983, Sens: 0.918, Spec:0.987
[20]	Fuwai Central China Cardiovascular Hospital	109 patient's Coronary angiogram	Threshold segmentation, Region-based segmentation, PSPNet with TL	ACC: 0.957 Sens: 0.865 Spec: 0.949
[21]	Public Database	134 Coronary angiograms	CLAHE, Multiresolution strategy and multiscale strategy with U-Net	ACC: 0.9765 Sens: 0.7978 Spec: 0.9885 PPV: 0.8137 Dice coefficient: 0.7905
[31]	Mexican Social Security Institute (First dataset), Antczak and Liberadzki dataset (second dataset)	First dataset: 180 coronary angiograms (500 patches) Second dataset: 250 coronary angiograms	SVM-based classifier, UDMA	First database ACC: 0.89 and Jaccard Index: 0.80, and the second database, ACC: 0.88 Jaccard Index. 0.79
[32]	Antczak and Liberadzki dataset (public)	10,000 synthetic images and 250 real coronary angiograms images	Pretrained (VGG16, ResNet50, and Inception-v3 with Transfer Learning	ACC: 0.95, Precis: 0.93, Sens: 0.98, Spec: 0.92, and F1 score: 0.95

ACC: Accuracy, Sens: Sensitivity, Spec: Specificity, PPV: Positive predictive value, Precis: Precision, F1 score: harmonic mean of PPV (precision) and sensitivity (recall), Contrast Limited Adaptive Histogram Equalization (CLAHE), Pyramid Scene Parsing Network (PSPNet), Transfer Learning (TL), Univariate Marginal Distribution Algorithm (UMDA).

2. Methodology

The proposed system of this study contains two main stages. The first stage consists of image pre-processing and segmentation and the second stage is comprised of image classification. The dataset is evaluated based on K-fold cross-validation set. In the first stage, U-Net, ResUNet-a, and UNet++, models are used to segment the coronary arteries. In the second stage, the coronary arteries are classified using five different pretrained models according to the presence of stenosis based on the original images of the patients. These pretrained models are evaluated using five-fold cross-validation. DenseNet201, EfficientNet-B0, Mobilenet-v2, and ResNet101 models adjust the input size of the RGB images to $224 \times 224 \times 3$, whereas for the Xception model, the input images are resized to $229 \times 229 \times 3$. The block diagram of the proposed system is presented in Figure 2.

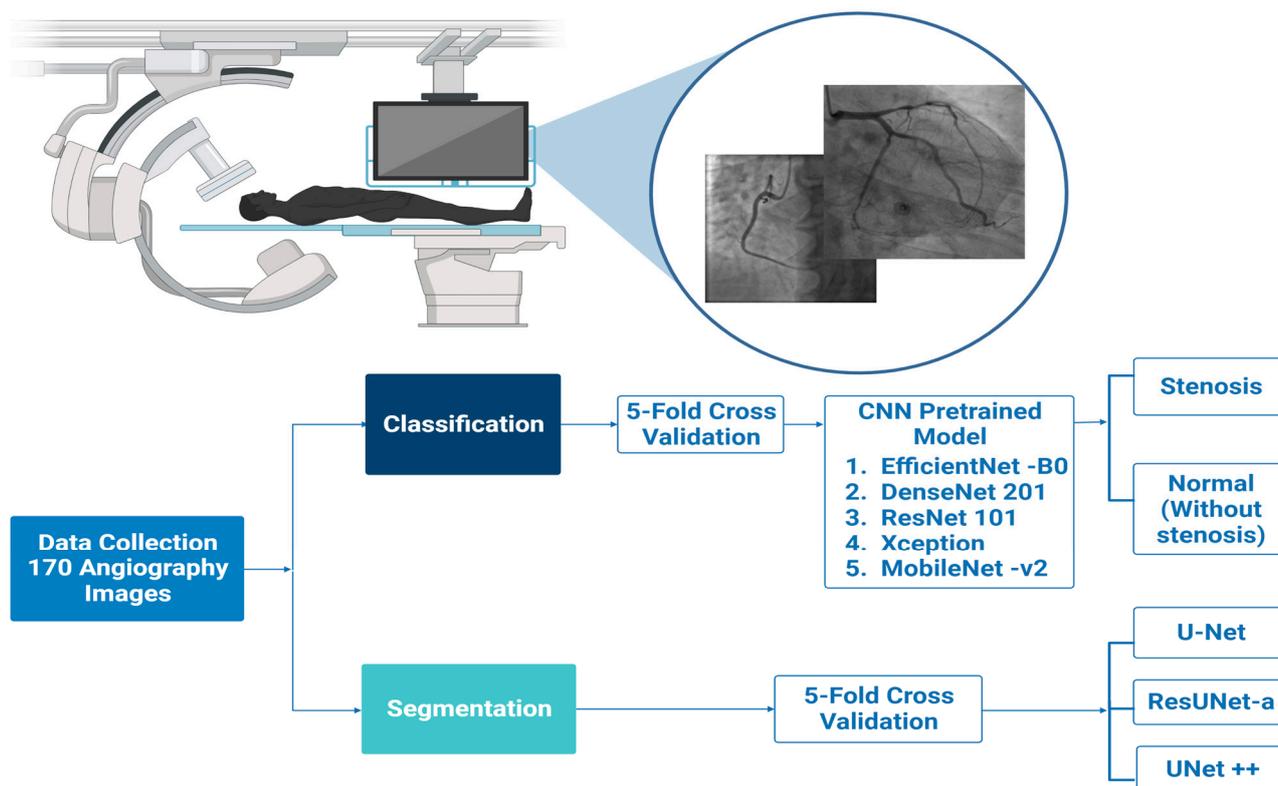


Figure 2. The Block Diagram of the Proposed System.

2.1. Dataset Description

The dataset used in this study consisted of 170 coronary angiograms from 22 patients, including 68 narrowed (with stenosis) and 102 normal (without stenosis), acquired using the GE (General Electric) Innova 3100 Cath/Angio System (Angiography machine) at the Department of Cardiology, Near East University Hospital, Nicosia, TRNC. Image examples of coronary arteries with and without stenosis are shown in Figure 3. Patient data were collected with the approval of the hospital management ethics committee. Data collection took place between October 2020 and February 2020. Data collection was conducted according to the university's ethical guidelines. The images captured by the medical imaging device used in this study were in Digital Imaging and Communications in Medicine (DICOM) format with a size of 512×512 in an RGB color model. In addition, the medical images were converted to PNG format. First, the labels for all images were created manually for segmentation by a board-certified cardiologist. Then, the catheter was marked with the major coronary arteries in the images.

In this study, a cardiologist and an angiography technician from the Near East University Hospital analyzed the images using Syngo QCA clinical software. Catheter calibration was used as the reference standard to analyze the QCA (Quantitative coronary angiography). It should be mentioned that the QCA is one of the most commonly used tools in clinical practice to assess coronary artery stenosis. The QCA was developed to objectively measure the lumen diameter of the coronary artery [36].

The scale currently proposed by the Society of Cardiovascular Computed Tomography for grading the severity of stenosis is used to rate the grade according to a standardized reporting system for patients with coronary Computed Tomography Angiography (CTA) images. The percentage of narrowing of the luminal diameter of the coronary arteries can be classified according to specific grades. According to the Coronary Artery Disease Reporting and Data System (CAD-RADS) [37], if there is "no visible stenosis," the stenosis of the coronary arteries is classified as 0%. If the percentage of narrowing of the coronary artery is 1–24%, it is called minimal stenosis or plaque with no stenosis. It is called mild

stenosis if the rate is 25–49%. If the percentage of stenosis is 50–69%, it is called “moderate”; if there is “severe stenosis,” the narrowing of the coronary arteries is categorized as 70–99%, and if the degree of narrowing (stenosis) is 100%, it is called “occluded”. The dataset used in this study was classified as coronary arteries without stenosis between 0% and 24%. Coronary arteries, with more than 24%, were classified as stenosis. Figure 4 shows an example of a coronary artery with stenosis and without stenosis from a QCA analysis performed at Near East University Hospital.

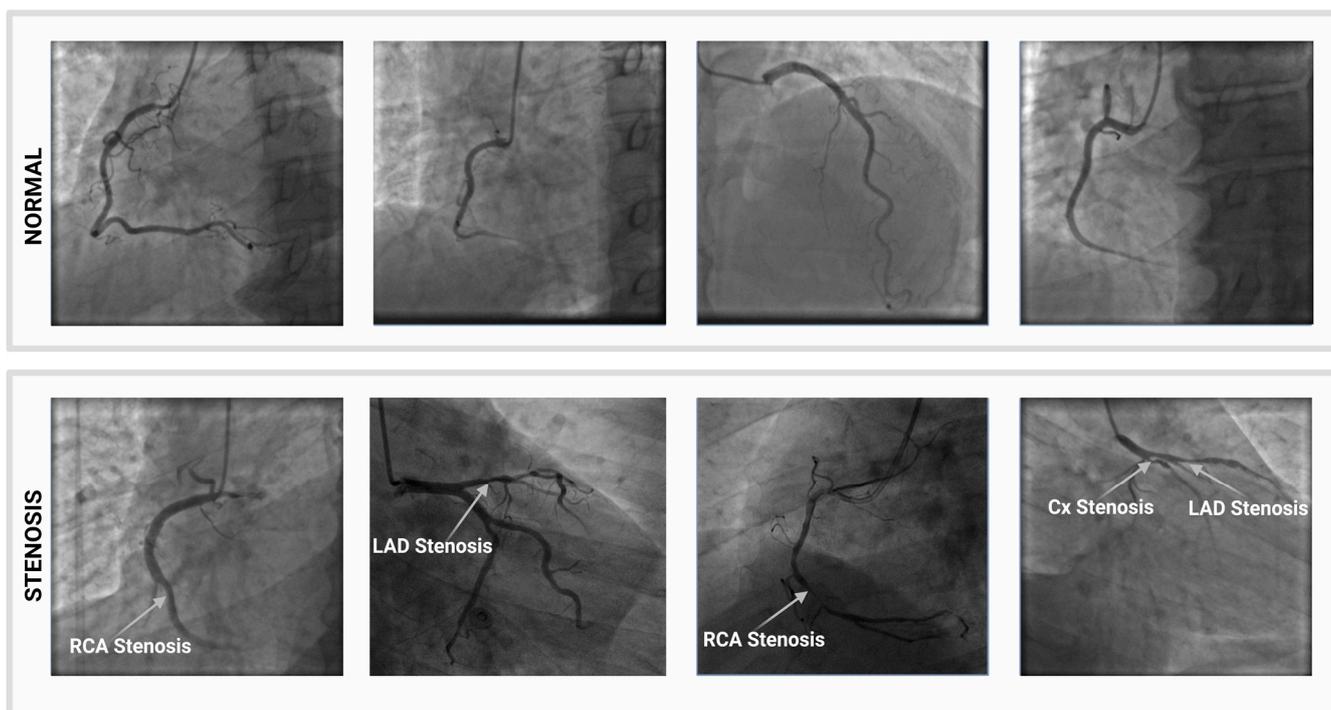


Figure 3. An example of images with and without stenosis in the dataset.

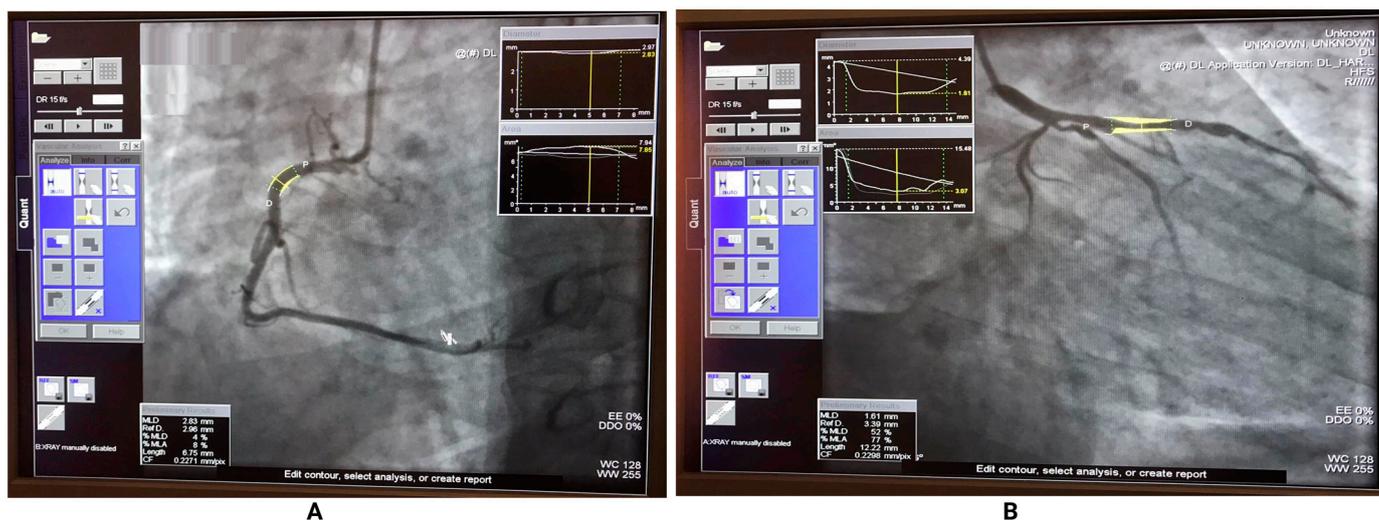


Figure 4. QCA analysis from our database, (A) An example of an arterial RCA without stenosis from our database. (B) An example of a LAD arterial stenosis from our database.

2.2. Image Pre-Processing

Image pre-processing was used in this study to prepare images before coronary artery segmentation. First, the images in the dataset were converted from RGB to grayscale. In

addition, the images were resized to 256×256 to meet memory requirements by reducing training time.

2.3. Cross Validation

Cross-validation is a popular technique used in AI to select parameters and evaluate learning parameters and predictive performance. In this technique, the original sample is randomly divided into groups of equal size in K-fold cross-validation. Then, the training of the dataset is repeated several times for K (e.g., $n = K$). Finally, the average performance of the training and testing datasets is calculated as an evaluation index for the models. This approach is very effective, especially when dealing with a limited number of images, as it uses the entire dataset [38,39]. In this study, we used a five-fold cross-validation to determine training and testing sets for both segmentation and classification tasks.

2.4. Segmentation

2.4.1. U-Net

Ronneberger et al. [40] proposed U-Net as a CNN for biomedical image segmentation. U-Net is a neural network architecture that focuses on image segmentation. There are two main paths in the U-Net architecture. The first path is considered the contracting path, also known as the encoder path or analysis path. Classification information is provided by the convolutional network, which is similar to a standard convolutional network. The second path is the expansion path, also called the decoder or synthesis path, which involves the upward convolution and concatenation of features from the contraction path. Thus, the network learns localized classification information. Furthermore, the extension path increases the output resolution, which can then be fed into a final convolutional layer to obtain a fully segmented image [41]. A convolutional network is used to create the contraction path. It consists of two 3×3 convolutions (unpadded convolutions) applied simultaneously, traced by a rectified linear unit (ReLU), a 2×2 max-pooling operation, and a stride-2 downsampling operation for each. In each downsampling stage, the number of feature channels is increased by two times their value. Each step on the expansion path is preceded by an upsampling of the feature map. The convolution path involves two 3×3 convolutions, followed by ReLUs, a 2×2 convolution (upconvolution) that halves the number of feature channels, and the concatenation of the contraction path through the corresponding feature map. Since the edge pixels are lost with each convolution, cropping is required. Convolutions of 1×1 are used in the last layer to divide the 64-component feature vectors into classes. There are 23 convolution layers in the network. The resulting network has a U-shaped topology and is nearly symmetric [40].

2.4.2. Deep Residual U-Net (ResUNet)

The modification of U-Net gives rise to ResUNet and is categorized under the semantic segmentation neural networks. ResUNet which combines the characteristics of U-Net and the residual networks [42]. The name “ResUNet-a” is established due to the fact that the architecture is composed of residual building blocks with multiple atrous convolutions and a U-Net backbone design. ResUNet-a utilizes a U-Net encoder and decoder backbone coupled with residual connections, atrous convolutions, multi-tasking inference and pyramid scene parsing pooling which is placed in both the middle and the end of the network. The inferred tasks are applied inside the network prior to the formation of the final segmentation mask. ResUNet-a determines the boundary of the objects, the distance transform of the segmentation mask, the segmentation mask and a colored reconstruction of the input sequentially. Based on the ResUNet-a model there are two basic types of architecture, d6 and d7, which differ in depth such as the total number of layers. The encoder part of ResUNet-a d6 includes six ResBlock-a building blocks, followed by the PSPooling layer. The encoder part of ResUNet-a d7 there are seven ResBlock-a building blocks. There are three different output options for each model (d6 and d7), which are referred to as single

task semantic segmentation layer, multi-task layer (mtsk), and conditioned multi-task output layer (cmtsk) [43].

2.4.3. UNet++

U-Net++ is proposed by Zhou et al. which applied the model for the segmentation of medical images. U-Net++ is another form of the U-net model derived from the DenseNet model, using a dense network of skip connections between the expansive and the contracting paths. The architecture is known for its encoder and decoder sub-networks which are linked via a series of nested, dense skip pathways. As an updated or modified version of U-Net, U-Net++ is designed with skip pathways that link the decoder and the encoder network. Another variation between the two is that the feature map in the encoder network is straightly transmitted to the decoder network in U-Net while the feature map of the encoder network is mapped to the decoder network via dense convolution blocks through the redesigned skip pathways in U-Net++. In essence, the feature graph semantic level in the encoder is proximate to the feature graph semantic level in the decoder via dense convolution blocks [44].

2.5. Classification

The pretrained CNN model is already trained using millions of images (ImageNet) of different objects such as pencils, coffee mugs, cats, dogs, keyboards, etc. Most pretrained CNN models are trained on a subset of the ImageNet database, used in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). The architecture is designed to classify up to 1000 categories [45,46].

In this study, five different pretrained CNN models were used to classify coronary angiography images. DenseNet201, EfficientNet-B0, Mobilenet-v2, ResNet101 and Xception models were used for the classification system. This section briefly summarizes these pretrained models.

2.5.1. DenseNet

Increasing the model's depth through blocks and layers has been shown to increase efficiency. DenseNet is one of these models along with ResNet with higher depths. Densely Connected Convolutional Networks (DenseNet) shared a lot of similarities with ResNet, one of the main differences between the 2 is that ResNet utilized an additive approach in which the model takes previous output as an input for future layers while DenseNet takes in all the entire previous output as an input for a future layer. DenseNet is stacked with narrow layers (i.e., 12 filters), which add a lesser set of new feature maps. Currently, there are several variants of DenseNet with the latest one as DenseNet264 followed by DenseNet201, DenseNet169 and the oldest as DenseNet121. The computational cost of this architecture is lower compared to ResNet which owes to the fact that each layer in the DenseNet model has uninterrupted access to the original input as well as a gradient from the loss function. This specific characteristic makes DenseNet one of the superior DL architectures for image classification tasks [47,48]. For example, in DenseNet201, the direct connections of the hidden layers are replaced with dense connections which enable the re-utilization of network features as well as optimize data transmission between the model's layers. The model is designed to take in input with a size of 224×224 followed by a series of convolution and pooling operations. The four dense blocks of the model are interspersed with ternary transition blocks which result in an output based on 14×14 feature maps [49].

2.5.2. EfficientNet

EfficientNet is a DL model developed by Mingxing Tan and Quoc V. Le [50]. The model provides an exceptional approach for scaling neural networks by enhancing precision, depths and widths. The CNN model's scaling method revolves around the application of compounded coefficients that evenly scale up the resolution dimensions, width, and

depth [50]. Moreover, EfficientNet is a type of DL model that is developed from baseline models established via neural network search [51]. In terms of architecture, EfficientNet-B0's main building block is a mobile inverted bottleneck convolution (MBConv) that is slightly modified as a result of the addition of special blocks known as squeeze-and-excitation optimization blocks. Thus, each of these MBConv blocks depends on shortcut connections and depthwise convolutions between the blocks [52].

2.5.3. MobileNet

MobileNet is CNN developed by a team in Google with the aim of minimizing the amount of memory used for computing and simultaneously producing high performance. The CNN architecture was trained using the large ImageNet dataset which contains thousands of categories. As the name implies, the model is designed for mobile phones or devices with lower computational power. The layers of MobileNet-v1 are divided into 2; the first layer is a depthwise convolutional layer designed for lightweight filtering (which utilized one convolutional filter for each input channel) while the second layer is a pointwise (1×1 convolution) layer designed for building new features through computing linear combinations of input channels. The MobileNet-v2 model is comprised of two residual blocks, the first block is designed with a stride of 1 and the second block with a stride of 2 where their main function is for downsizing. Each of these residual blocks is equipped with three main layers which include a convolutional layer (1×1) with a rectified linear unit (ReLU), followed by a depthwise convolution and a third of convolutional layer (1×1) without non-linearity. The MobileNet-v2 architecture comprises a fully convolutional layer with 32 filters followed by 19 residual bottleneck blocks [53,54].

2.5.4. ResNet

ResNet is a kind of DL based on residual learning. The ResNet model was crowned the best-performing model in the 2015 ILSVRC competition. The model was developed by a team working at Microsoft. Currently, there are a handful of ResNet variants designed with a different number of layers. Some of these variants include ResNet1202, ResNet164, ResNet152, ResNet110, ResNet101, ResNet50, ResNet34 and ResNet18. Among these variants, ResNet101 is utilized in this study. There are three main layers in the ResNet101 model: convolutional layers, pooling layers, and fully connected layers. The model is designed with 101 layers with 33 residual blocks. The main function of the residual blocks of ResNet101 is to enable shortcuts or skip connections to other layers [55–57].

2.5.5. Inception

In order to improve the training efficiency of neural networks, Google's team first developed the Inception model back in 2014. The introduction of 1×1 convolutional kernels after 3×3 convolutional kernels, 5×5 convolutional kernels and 3×3 pooling kernels significantly minimizes the number of parameters and simultaneously increases the width of the model and optimizes the model's adaptability to scale. The modification of the Inception model gives rise to Xception (with better performance) due to the introduction of both depthwise separable convolution and residual structure. Thus, the Xception model makes more efficient use of its parameters compared to the Inception model despite having the same number of parameters. The architecture of Xception CNN consists of 36 convolutional layers which form the feature extraction base of the CNN. In this study, the evaluation method will solely investigate image classification of CAD and thus, the convolutional base will be coupled with a logistic regression layer. Another option is to add fully connected layers before adding the logistic regression layer [58,59].

2.6. Experimental Design

All our experiments were conducted on a computer with 32 GB RAM, an NVIDIA GeForce RTX 2080 Ti graphics card, and an Intel i7-8th Generation CPU. We implemented all experiments with DL networks in the MATLAB 2021b programming language. For the

training process, the Adam optimizer and five-fold cross-validation are employed to train and test the dataset. Coronary artery segmentation is performed using 0.0001, 100, and 8 parameters for learning rate, epoch, and batch size, respectively. The training parameters for classifying coronary arteries with and without stenosis are set to 0.0001, 100, and 32.

2.7. Performance Metrics and Confusion Matrix

The efficiency of the image segmentation and classification system is evaluated using a variety of performance metrics [60,61]. In this study, various performance metrics including accuracy, sensitivity, specificity, precision, Dice Score (F1 score), Jaccard Index and Matthews correlation coefficient (MCC) were used to evaluate the segmentation system.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (1)$$

Sensitivity measures the ability of the system to segment coronary artery pixels, as seen in Equation (2).

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (2)$$

In contrast to sensitivity, *specificity* represents the segments of the background pixels and is described in Equation (3).

$$Specificity = \frac{TN}{(TN + FP)} \quad (3)$$

Precision (Positive Predictive Value) represents the ratio of correctly classified coronary artery pixels to all pixels classified as coronary arteries by the system. This led the researchers to analyze the system's success in generating true positives compared to all pixels classified as arteries, as defined in Equation (4).

$$Precision (PPV) = \frac{TP}{(TP + FP)} \quad (4)$$

Image segmentation algorithms are often evaluated using the *Dice score*, which ignores the correct classification of negative samples, such as background pixels. The equation of the Dice score is given in Equation (5) [62].

Dice score measures the pixel similarity between the segmented image and the ground truth that ranges from 0 (no similarity) to 1 (identical) [63].

$$Dice \text{ Score (F1 score)} = \frac{2TP}{(2TP + FP + FN)} \quad (5)$$

The *Jaccard Index* is calculated as the overlap ratio between the predicted and ground truth segmentation. This is the union area between the predicted and ground truth segmentation, as given in Equation (6) [64].

$$Jaccard \text{ Index} = \frac{TP}{(TP + FP + FN)} \quad (6)$$

Mathews Correlation Coefficient (*MCC*). This metric was first introduced in 1975 by a biochemist known as Brian W. Matthews who was working on the prediction of enzyme structure. As a marginal measure unaffected by the unbalanced datasets issue, the *MCC* is a contingency matrix approach to calculating the Pearson product-moment correlation coefficient between predicted and actual values. *MCC* is represented mathematically in the equation below:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (7)$$

The interpretation of *MCC* is based on -1 as the worst value and $+1$ as the best value. Moreover, the metric is the only binary classification rate that produces a higher value only if the binary predictor was able to accurately predict the majority of positive and negative data cases [65].

A confusion matrix is a popular approach for estimating the performance of a model for segmentation and classification based on the number of true positives (*TP*), true negatives (*TN*), false positives (*FP*), and false negatives (*FN*) [66,67]. *TP* correctly identifies segmented coronary arteries, while *FP* identifies segmented background pixels as arteries. *TN* indicates background pixels classified as background, but *FN* incorrectly identifies arterial pixels segmented as background.

This study used accuracy, sensitivity, specificity, positive predictive value (*PPV*), and negative predictive value (*NPV*), *F1 score*, *MCC*, Cohen's kappa, Area Under Curve (*AUC*) and Receiver Operating Characteristic (*ROC*) curve to evaluate image classification. Accuracy indicates the overall success rate of the system in classifying images with and without coronary artery stenosis. The sensitivity metric is the ability of an algorithm to detect coronary artery stenosis. Specificity refers to the power of a classification technique to see images with non-stenotic coronary arteries. The *PPV* or precision defines the probability of coronary stenosis in the images when the stenosis is present. The *NPV* indicates the probability that the result is negative when stenosis is not present. The equation of the *NPV* is given in Equation (8).

$$NPV = \frac{TN}{(TN + FN)} \quad (8)$$

The model correctly identifies *TPs* as positive cases on images with coronary artery stenosis. *TNs* are correctly identified by the model as negative cases on images without stenosis. *Fps* occur when the model identifies coronary artery stenosis on images without stenosis according to the ground truth classification. *Fns* describe the case when the model does not identify stenosis on the image where stenosis is present according to the ground truth classification.

The *F1 score* is another evaluation metric that determined how precisely an ML model operates by differentiating the actual true positives from the expected ones [68]. Thus, the *F1 score* is regarded as the harmonic mean of recall and precision which depicts that it penalizes extreme values of both. However, one of the limitations of the *F1 score* is that it is not symmetric between classes as it depends solely on which class is labeled as positive or negative [69].

The kappa coefficient is another performance metric approach the is used to summarize the agreement between 2 nominal classifications according to similar categories. The evaluation metric is popularly utilized in the field of social and behavioral sciences as well as the medical field as a measure of agreement between two raters. The concept was first introduced in 1960 by Jacob Cohen which is an alternative evaluation to accuracy. The kappa coefficient is mostly used as a performance measure or sample statistic. For instance, when calculating kappa for a sample of subjects is one step in a series of experimental steps, or when kappa is used for analyzing a binary classification [65].

Cohen's kappa can be represented mathematically [70]:

$$\kappa = \frac{pA - pE}{1 - pE} \quad (9)$$

where *pA* is the observed relative agreement between two annotators

pE is the hypothetical probability of agreement by chance (with data labels randomly assigned).

$\kappa = 1$ corresponds to the case of perfect agreement

$\kappa = 0$ indicates no agreement other than what would be expected by chance

AUC is regarded as one of the most significant performance metrics for evaluating the performance of ML models, especially classification tasks. *AUC*-score summarizes the Receiver Operating Characteristic (*ROC*) curve or is a metric that is used to represent the

area under the ROC curve. The higher the score, the better the performance of the model and its ability to discriminate between positive and negative cases.

Mathematically, *AUC* is represented as: [68]

$$AUC = \frac{TPR - FPR + 1}{2} \quad (10)$$

AUC score is showcased as a random classification score which is represented by a diagonal ROC curve in the unit square. The value can range between 0.5 to 1, where 1 represents a perfect classification score. Nevertheless, a score range of 0.7–0.8 is acceptable while any score below 0.7 is considered as poor performance [71]. The ROC curve is graphically represented by the false-positive rate (*FPR*) and the true-positive rate (*TPR*) [51].

3. Experimental Results

3.1. Segmentation Results

The results of our segmentation study and the steps that led to these results are shown in Figure 5. The images provided include the original input image, the patient's images marked by the cardiologist (ground truth), and the output images (segmented images) resulting from the segmentation models performed on them. The visual output results of the U-Net model are closer to ground truth images than ResUNet-a and UNet++ as seen in Figure 5.

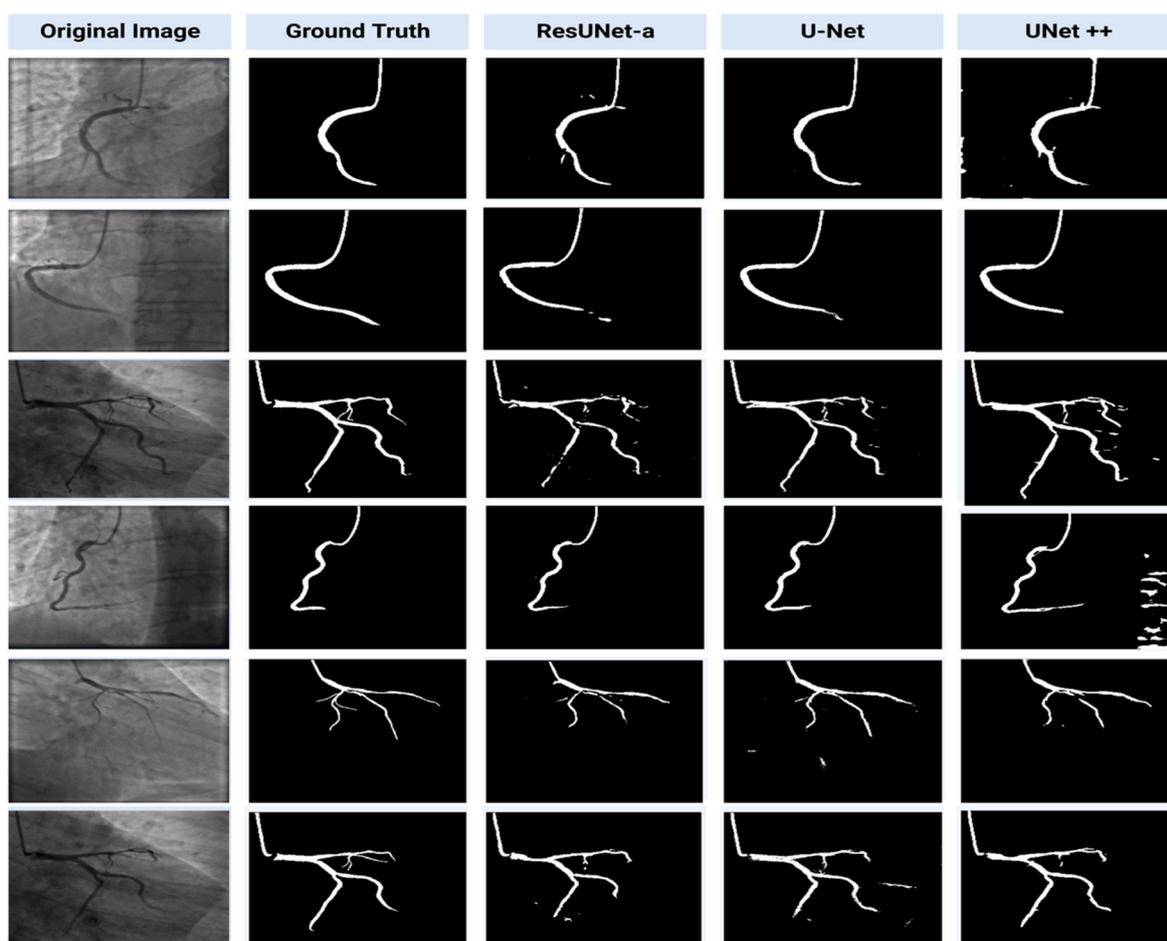


Figure 5. Sample segmentation results from the dataset.

We performed K-fold cross-validation to evaluate coronary artery segmentation and classify non-stenotic and stenotic coronary arteries using DL models. The parameter K was set to 5 in our experiments. In our dataset, for each fold, 136 images were used for the

training set and 34 for the test set. All the images (total = 170) in our dataset were tested. Tables 2–4, respectively, show the K-fold cross-validation and segmentation of coronary arteries with the U-Net, ResUNet-a and UNet++ models.

Table 2. The segmentation results of the K-fold cross-validation of the U-Net model (k = 5).

Folds	Accuracy	Sensitivity	Specificity	Precision	Dice Score	Jaccard Index	MCC
Fold1	0.9929	0.8477	0.9970	0.8809	0.8611	0.7628	0.8591
Fold2	0.9905	0.8355	0.9960	0.8628	0.8429	0.7390	0.8413
Fold3	0.9917	0.8946	0.9942	0.8006	0.8416	0.7446	0.8405
Fold4	0.9926	0.8397	0.9968	0.8697	0.8521	0.7542	0.8497
Fold5	0.9911	0.8822	0.9942	0.7976	0.8360	0.7266	0.8335
Average	0.9918	0.8599	0.9957	0.8423	0.8467	0.7454	0.8448

Table 3. The segmentation results of the K-fold cross-validation of the ResUNet-a model (k = 5).

Folds	Accuracy	Sensitivity	Specificity	Precision	Dice Score	Jaccard Index	MCC
Fold1	0.9926	0.8224	0.9975	0.8925	0.8534	0.7508	0.8518
Fold2	0.9892	0.7728	0.9966	0.8760	0.8129	0.6964	0.8134
Fold3	0.9914	0.8133	0.9960	0.8407	0.8238	0.7178	0.8211
Fold4	0.9923	0.8394	0.9966	0.8615	0.8462	0.7452	0.8445
Fold5	0.9921	0.8029	0.9974	0.8872	0.8379	0.7324	0.8376
Average	0.9915	0.8101	0.9968	0.8716	0.8348	0.7285	0.8337

Table 4. The segmentation results of the K-fold cross-validation of the UNet++ model (k = 5).

Folds	Accuracy	Sensitivity	Specificity	Precision	Dice Score	Jaccard Index	MCC
Fold1	0.9878	0.9128	0.9900	0.7092	0.7945	0.6673	0.7969
Fold2	0.9872	0.8673	0.9915	0.7456	0.7932	0.6637	0.7935
Fold3	0.9879	0.8937	0.9904	0.7291	0.7919	0.6779	0.7959
Fold4	0.9907	0.9046	0.9932	0.7709	0.8292	0.7176	0.8288
Fold5	0.9867	0.9086	0.9890	0.6825	0.7742	0.6435	0.7786
Average	0.9881	0.8974	0.9908	0.7274	0.7966	0.6740	0.7988

Training and evaluation of U-Net yield an average accuracy of 0.9918, average Sensitivity of 0.8599, average Specificity of 0.9957, precision of 0.8423, average Dice score of 0.8467, average Jaccard Index of 0.7454 and average MCC of 0.8448 as shown in Table 2.

Training and evaluation of ResUNet-a yield an average accuracy of 0.9915, average Sensitivity of 0.8101, average Specificity of 0.9968, precision of 0.8716, average Dice score of 0.8348, average Jaccard Index of 0.7285 and average MCC of 0.8337 as shown in Table 3.

Training and evaluation of UNet++ yield an average accuracy of 0.9881, average Sensitivity of 0.8974, average Specificity of 0.9908, precision of 0.7274, average Dice score of 0.7966, average Jaccard Index of 0.6740 and average MCC of 0.7988 as shown in Table 4.

3.2. Classification Results

The performance measures for classifying coronary arteries with and without stenosis using five different pretrained classification models with 5-fold cross-validation are presented in Tables 5–9. For each fold, the performance measures accuracy, sensitivity,

specificity, PPV, and NPV, F1 score, MCC and Cohen's Kappa were calculated for the system evaluation. The result of this study validated that machine learning algorithms can automatically classify coronary arteries with high accuracy. Figure 6 presents the confusion matrix of five pretrained CNN models for classifying coronary artery images as normal (without stenosis) and stenosis.

Table 5. The classification results of the K-fold cross-validation of the pretrained Densenet201 model (k = 5).

Folds	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score	MCC	Cohen's Kappa
Fold1	0.9412	0.8889	1	1	0.8889	0.9412	0.8889	0.8826
Fold2	0.8824	0.6923	1	1	0.8400	0.8182	0.7626	0.7354
Fold3	0.8529	0.7000	0.9167	0.7778	0.8800	0.7368	0.6369	0.6352
Fold4	0.9412	0.8182	1	1	0.9200	0.9000	0.8676	0.8589
Fold5	0.8824	0.7500	1	1	0.8182	0.8571	0.7834	0.7606
Average	0.9000	0.7699	0.9833	0.9556	0.8694	0.8507	0.7879	0.7746

Table 6. The classification results of the K-fold cross-validation of the pretrained EfficientNet-B0 model (k = 5).

Folds	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score	MCC	Cohen's Kappa
Fold1	1	1	1	1	1	1	1	1
Fold2	0.8529	0.6154	1	1	0.8077	0.7619	0.7050	0.6640
Fold3	0.8824	0.8000	0.9167	0.8000	0.9167	0.8000	0.7167	0.7167
Fold4	0.7941	0.8182	0.7826	0.6429	0.9000	0.7200	0.5711	0.5609
Fold5	0.9118	0.8125	1	1	0.8571	0.8966	0.8345	0.8211
Average	0.8882	0.8092	0.9399	0.8889	0.8963	0.8357	0.7655	0.7525

Table 7. The classification results of the K-fold cross-validation of the pretrained MobileNet-v2 model (k = 5).

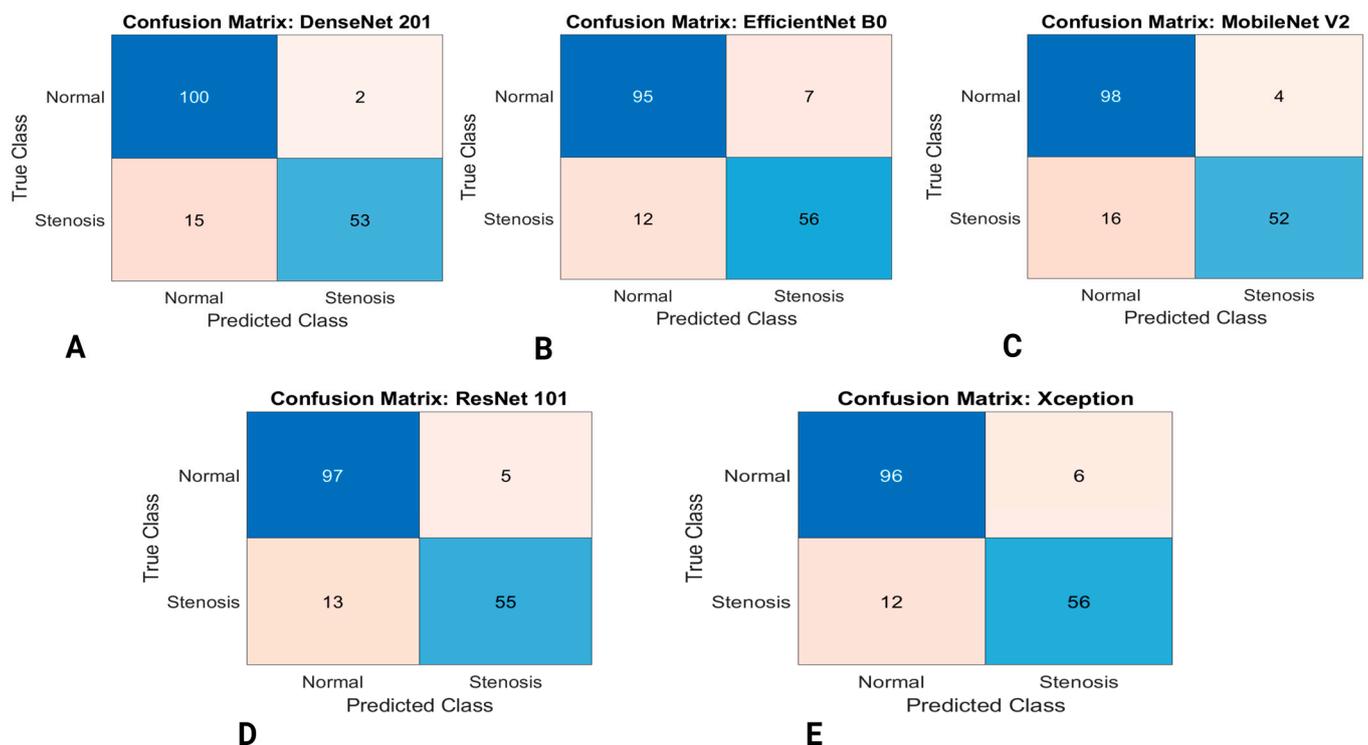
Folds	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score	MCC	Cohen's Kappa
Fold1	0.9118	0.8333	1	1	0.8421	0.9091	0.8377	0.8247
Fold2	0.8824	0.6923	1	1	0.8400	0.8182	0.7626	0.7354
Fold3	0.8824	0.8000	0.9167	0.8000	0.9167	0.8000	0.7167	0.7167
Fold4	0.9118	0.8182	0.9565	0.9000	0.9167	0.8571	0.7954	0.7935
Fold5	0.8235	0.6875	0.9444	0.9167	0.7727	0.7857	0.6600	0.6408
Average	0.8824	0.7663	0.9635	0.9233	0.8576	0.8340	0.7545	0.7422

Table 8. The classification results of the K-fold cross-validation of the pretrained ResNet101 model (k = 5).

Folds	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score	MCC	Cohen's Kappa
Fold1	0.9706	0.9444	1	1	0.9412	0.9714	0.9428	0.9412
Fold2	0.8824	0.7692	0.9524	0.9091	0.8696	0.8333	0.7496	0.7434
Fold3	0.8529	0.7000	0.9167	0.7778	0.88000	0.7368	0.6369	0.6352
Fold4	0.9412	0.9091	0.9565	0.9091	0.9565	0.9091	0.8656	0.8656
Fold5	0.8235	0.6875	0.9444	0.9167	0.7727	0.7857	0.6600	0.6408
Average	0.8941	0.8021	0.9540	0.9025	0.8840	0.8473	0.7710	0.7652

Table 9. The classification results of the K-fold cross-validation of the pretrained Xception model (k = 5).

Folds	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score	MCC	Cohen's Kappa
Fold1	0.9412	0.8889	1	1	0.8889	0.9412	0.8889	0.8828
Fold2	0.8824	0.6923	1	1	0.8400	0.8182	0.7626	0.7354
Fold3	0.7941	0.8000	0.7917	0.6154	0.9048	0.6957	0.5548	0.5441
Fold4	0.9706	0.9091	1	1	0.9583	0.9524	0.9334	0.9312
Fold5	0.8824	0.8125	0.9444	0.9286	0.8500	0.8667	0.7677	0.7622
Average	0.8941	0.8206	0.9472	0.9088	0.8884	0.8548	0.7815	0.7711

**Figure 6.** Confusion matrix for the coronary artery classification (A) DenseNet201 (B) EfficientNet-B0 (C) Mobilenet-v2 (D) ResNet101 (E) Xception.

The AUC/ROC for the classification of coronary arteries is presented in Figure 7. AUC/ROC are regarded as two of the most significant performance metrics for the eval-

uation of classification models. It can be seen in Figure 7. The DenseNet201 model outperformed other classification models with an AUC score of 0.9694.

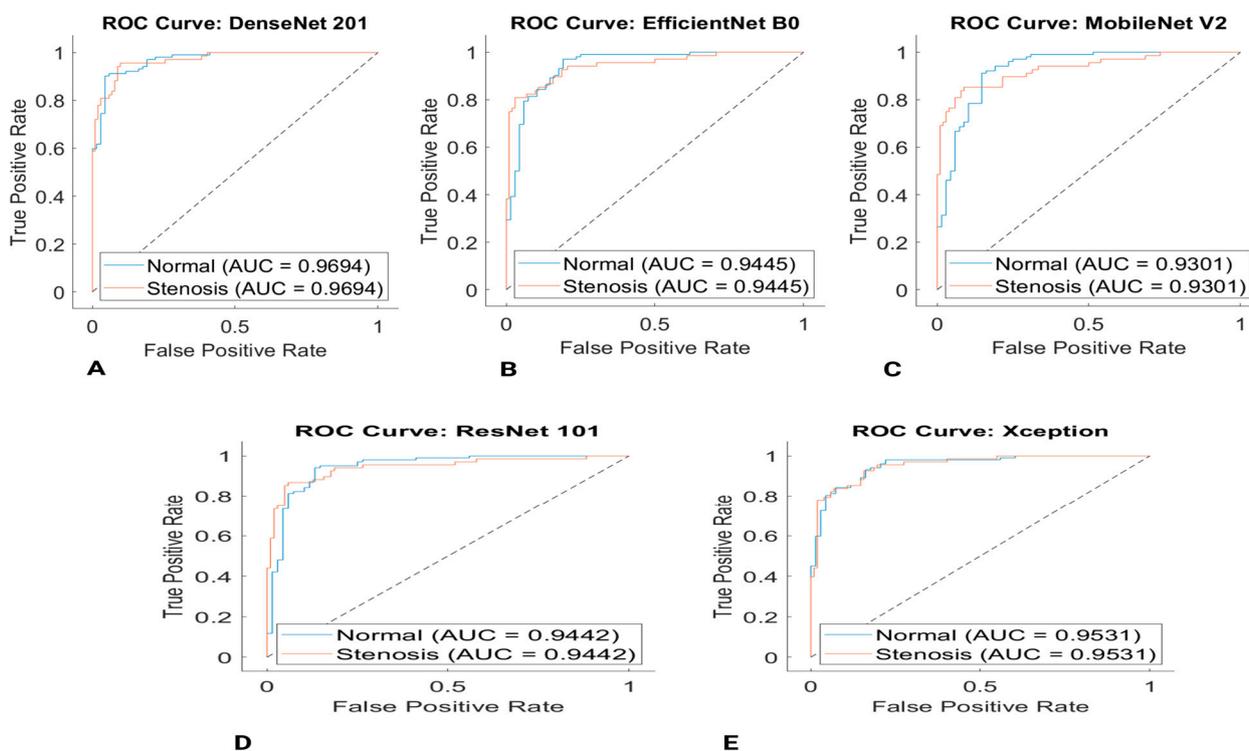


Figure 7. ROC curve for the coronary artery classification (A) DenseNet201 (B) EfficientNet-B0 (C) Mobilenet-v2 (D) ResNet101 (E) Xception.

Training and evaluation of the DenseNet201 model yield an average accuracy of 0.9000, average Sensitivity of 0.7699, average Specificity of 0.9833, average PPV of 0.9556, average NPV of 0.8694, average F1 score of 0.8507, average MCC of 0.7879 and average Cohen's Kappa of 0.7746 as shown in Table 5.

Training and evaluation of the EfficientNet-B0 model yield an average accuracy of 0.8882, average Sensitivity of 0.8092, average Specificity of 0.9399, average PPV of 0.8889, average NPV of 0.8963, average F1 score of 0.8357, average MCC of 0.7655 and average Cohen's Kappa of 0.7525 as shown in Table 6.

Training and evaluation of the MobileNet-v2 model yield an average accuracy of 0.8824, average Sensitivity of 0.7663, average Specificity of 0.9635, average PPV of 0.9233, average NPV of 0.8576, average F1 score of 0.8340, average MCC of 0.7545 and average Cohen's Kappa of 0.7422 as shown in Table 7.

Training and evaluation of the ResNet101 model yield an average accuracy of 0.8941, average Sensitivity of 0.8021, average Specificity of 0.9540, average PPV of 0.9025, average NPV of 0.8840, average F1 score of 0.8473, average MCC of 0.7710 and average Cohen's Kappa of 0.7652 as shown in Table 8.

Training and evaluation of the Xception model yield an average accuracy of 0.8941, average Sensitivity of 0.8206, average Specificity of 0.9472, average PPV of 0.9088, average NPV of 0.8884, average F1 score of 0.8548, average MCC of 0.7815 and average Cohen's Kappa of 0.7711 as shown in Table 9.

4. Discussion

Previous studies have shown that early detection of CAD is crucial for the timely treatment and for prolonging patients' lives. Since heart disease is the leading cause of death around the world, cardiologists are entrusted with diagnosing and treating many patients. In addition, most healthcare facilities rely on invasive coronary angiography

procedures. However, despite its high accuracy and efficiency, reviewing these angiographic images and interpreting many images can be tedious for cardiologists and can also lead to misdiagnosis [2–4,6]. To overcome these obstacles, several studies have proposed the DL approach, in which scientists can use algorithms to classify and segment medical images [11,12,17–21].

In line with existing studies, this study used pretrained models to classify angiographic images. However, unlike previous studies that made use of datasets from various online repositories and many clinical datasets, this study used datasets generated by cardiologists at Near East University Hospital in North Cyprus. In addition, the classification of angiographic images (i.e., 170 images) was achieved with five pretrained CNN models as shown in Tables 5–9. Several studies have already used pretrained models such as DenseNet, EfficientNet and Xception to classify medical images efficiently which is in line with the result achieved [48,72,73].

4.1. Limitations and Clinical Implications

The growing field of AI is changing the landscape of several fields ranging from computer and electronics, medicine, agriculture, banking and finance, business, and advertisements, etc. Over the last decade, scientists have developed several platforms that integrated artificial intelligence in healthcare. Computer-aided systems are currently in use for modeling biological components (such as patient's organs), assisted surgery (using robots), prediction of diseases outbreak, gene profiling, drug invention and repurposing and computer-aided diagnosis [74]. The use of DL models such as Artificial neural networks (ANN) and CNN are transforming the field of medical images into autonomic systems that allow real-time classification of medical data with high efficiency. This study aimed to use DL techniques for the automatic segmentation and classification of coronary arteries based on angiography images. The result achieved in this study is in line with other studies regarding the role of Computer-aided systems which receive medical data as input and produce output based on a series of algorithms. Similar approaches have been integrated into several hospitals and are currently in use to help cardiologists. Despite the fact that DL models produce excellent results, they are hindered by several challenges. For example, the lack of a limited amount of datasets can limit performance. Therefore, training DL models using a substantial amount of data is crucial for high performance. To counter this problem, scientists proposed the use of pretrained models (also known as Transfer Learning (TL) models). TL models enable scientists to extract weights and features from trained models and repurpose them on new tasks with insufficient datasets. Thus, the use of pretrained models has been shown to outperform models developed from scratch [74,75].

Another method proposed by scientists that can be used to resolve the issue of the limited amount of data is the use of data augmentation techniques which include zooming, rotation, cropping, mirroring, flipping, etc. Lastly, the last five years have seen the rise of hybrid and ensemble models. Extracting features and training using fused models have been shown to result in high performance. Moreover, the use of classifiers such as Support Vector Machine (SVM), K-nearest Neighbor (KNN) Decision Tree (DT) and Random Forest (RF) instead of SoftMax is another dimension integrated by scientists in evaluating models with higher performances [74].

4.2. Comparison of Segmentation Models Performance

Based on the comparison segmentation models with the U-Net model and its variants presented in Table 10. Dice score and Jaccard index are considered more reliable and significant performance metrics for the image segmentation system [76]. Thus, a comparative assessment of the segmentation models has shown that U-Net outperformed both UNet++ and ResUnet-a in both the Dice score (0.8467) and Jaccard Index (0.7454).

Table 10. Comparison of Segmentation Models Performance of the 5-Fold Cross-Validation with the average score.

Models	Accuracy	Sensitivity	Specificity	Precision	Dice Score	Jaccard Index	MCC
U-Net	0.9918	0.8599	0.9957	0.8423	0.8467	0.7454	0.8448
ResUnet-a	0.9915	0.8101	0.9968	0.8716	0.8348	0.7285	0.8337
UNet++	0.9881	0.8974	0.9908	0.7274	0.7966	0.6740	0.7988

4.3. Comparison of Classification Models Performance

Based on the comparison between different pretrained classification models as presented in Table 11, it can be seen that in terms of accuracy, the DenseNet201 model achieved the highest accuracy with 0.9000. In terms of sensitivity, Xception achieved the best result with 0.8206. While in the case of specificity, DenseNet201 achieved the best result with 0.9833. Comparative evaluation of both PPV and NPV has shown that DenseNet201 and EfficientNet-B0 achieved the best result with 0.9556 PPV and 0.8963 NPV, respectively. In the case of the F1 score, the model that achieved the highest score is Xception with 0.8548. Moreover, in the case of MCC, DenseNet201 achieved the best result with 0.7879. The comparison of the model's Cohen's Kappa result has shown that DenseNet201 achieved the best score with 0.7746. Therefore, in the overall assessment of performance metrics, DenseNet201 is considered the best-performing model that can automatically classify coronary arteries.

Table 11. Comparison of Classification Pretrained Models Performance of the 5-Fold Cross-Validation with the average score.

Task	Models	Accuracy	Sensitivity	Specificity	PPV	NPV	F1 Score	MCC	Cohen's Kappa
Normal and Stenosis	Densenet201	0.9000	0.7699	0.9833	0.9556	0.8694	0.8507	0.7879	0.7746
	EfficientNet-B0	0.8882	0.8092	0.9399	0.8889	0.8963	0.8357	0.7655	0.7525
	MobileNet-v2	0.8824	0.7663	0.9635	0.9233	0.8576	0.8340	0.7545	0.7422
	ResNet101	0.8941	0.8021	0.9540	0.9025	0.8840	0.8473	0.7710	0.7652
	Xception	0.8941	0.8206	0.9472	0.9088	0.8884	0.8548	0.7815	0.7711

5. Conclusions

The need to develop accurate diagnosis approaches is crucial for minimizing diagnostic errors and increasing treatment efficiency. In line with this statement, we proposed a system that includes U-Net and its variants and pretrained CNN models for the segmentation and classification of coronary angiograms. Another contribution of this research was the use of raw data acquired from real patients instead of using a dataset collected from online repositories. This is associated with the claim that a significant number of images available in online repositories have been altered (e.g., cropped, rotated, and enlarged), to facilitate coronary artery segmentation and classification. However, this is not the case for clinical applications. This work employed U-Net, ResUNet-a, and UNet++ models for the segmentation of coronary angiography images. Consequently, we used five different pretrained models (EfficientNet-B0, DenseNet201, Mobilenet-v2, ResNet101 and Xception) to classify coronary arteries into binary classes, e.g., with or without stenosis. The main goal of this research was to implement an automatic segmentation and classification system for coronary arteries. We believe that this will help doctors make more accurate diagnoses, reduce workload, and minimize the risk of error.

It should be mentioned that one of the study's limitations is using a small number of datasets for training and validation. In addition, the study is limited to binary classification. Therefore, future studies should focus on improving diversity in multiple coronary artery stenosis classifications according to their specific grades. Moreover, future studies will also

focus on the precise measurement of coronary artery stenosis diameter using angiography images based on AI and image processing techniques. Furthermore, future studies will focus on a large number of datasets curated from various hospitals in Northern Cyprus. It is also suggested to develop an IoT/AI-based system that will allow cardiologists to upload images via online systems and obtain results in real-time.

Author Contributions: Methodology, Ş.K., H.H., A.I. (Ahmet İlhan) and A.I. (Ali Isin); Software, A.I. (Ahmet İlhan), Ş.K. and A.I. (Ali Isin); Dataset preparation, C.C. and Ş.K. processed coronary angiographic images for labeling and classification and provided the dataset with patient images. Writing original draft preparation, Ş.K.; writing—review and editing, Ş.K., H.H., A.I. (Ali Isin), A.I. (Ahmet İlhan) and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Institutional Review Board Statement: The Near East University Scientific Research Ethics Committee approved the study under the Declaration of Helsinki with protocol code NEU/2019/74/929 and date 21 November 2019.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Institutional Review Board has not granted permission to share the patient data generated and analyzed in the current study. However, they are available on reasonable request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Malakar, A.K.; Choudhury, D.; Halder, B.; Paul, P.; Uddin, A.; Chakraborty, S. A Review on Coronary Artery Disease, Its Risk Factors, and Therapeutics. *J. Cell. Physiol.* **2019**, *234*, 16812–16823. [[CrossRef](#)] [[PubMed](#)]
2. Pandey, R.; Kumar, M.; Majdoubi, J.; Rahimi-Gorji, M.; Srivastav, V.K. A Review Study on Blood in Human Coronary Artery: Numerical Approach. *Comput. Methods Programs Biomed.* **2020**, *187*, 105243. [[CrossRef](#)] [[PubMed](#)]
3. Bit, A.; Suri, J.S.; Ranjani, A. Anatomy and physiology of blood vessels. In *Flow Dynamics and Tissue Engineering of Blood Vessels*; IOP Publishing Ltd: Bristol, UK, 2020. [[CrossRef](#)]
4. Taki, A.; Kermani, A.; Ranjarnavazi, S.M.; Pourmodheji, A. Overview of different medical imaging techniques for the identification of coronary atherosclerotic plaques. In *Computing and Visualization for Intravascular Imaging and Computer-Assisted Stenting*; Academic Press: Cambridge, MA, USA, 2017; pp. 79–106. [[CrossRef](#)]
5. Sayadi, M.; Varadarajan, V.; Sadoughi, F.; Chopannejad, S.; Langarizadeh, M. A Machine Learning Model for Detection of Coronary Artery Disease Using Noninvasive Clinical Parameters. *Life* **2022**, *12*, 1933. [[CrossRef](#)]
6. Wang, L.; Liang, D.; Yin, X.; Qiu, J.; Yang, Z.; Xing, J.; Dong, J.; Ma, Z. Coronary Artery Segmentation in Angiographic Videos Utilizing Spatial-Temporal Information. *BMC Med. Imaging* **2020**, *20*, 110. [[CrossRef](#)] [[PubMed](#)]
7. Bertolone, D.T.; Gallinoro, E.; Esposito, G.; Paolisso, P.; Bermpeis, K.; De Colle, C.; Fabbriatore, D.; Mileva, N.; Valeriano, C.; Munhoz, D.; et al. Contemporary Management of Stable Coronary Artery Disease. *High Blood Press. Cardiovasc. Prev.* **2022**, *29*, 207–219. [[CrossRef](#)]
8. Liu, Y.; Zhang, X.; Wan, W.; Liu, S.; Liu, Y.; Liu, H.; Zeng, X.; Zhang, Q. Two New Stenosis Detection Methods of Coronary Angiograms. *IJCARS* **2022**, *17*, 521–530. [[CrossRef](#)]
9. Hashemi, J.; Rai, S.; Ghafghazi, S.; Berson, R.E. Blood Residence Time to Assess Significance of Coronary Artery Stenosis. *Sci. Rep.* **2020**, *10*, 11658. [[CrossRef](#)]
10. Barriada, R.G.; Masip, D. An Overview of Deep-Learning-Based Methods for Cardiovascular Risk Assessment with Retinal Images. *Diagnostics* **2022**, *13*, 68. [[CrossRef](#)]
11. Moorthy, J.; Gandhi, U.D. A Survey on Medical Image Segmentation Based on Deep Learning Techniques. *BDCC* **2022**, *6*, 117. [[CrossRef](#)]
12. Balaji, K.; Lavanya, K. Medical image analysis with deep neural networks. In *Deep Learning and Parallel Computing Environment for Bioengineering Systems*; Academic Press: Cambridge, MA, USA, 2019; pp. 75–97. [[CrossRef](#)]
13. Han, X.; Zhang, Z.; Ding, N.; Gu, Y.; Liu, X.; Huo, Y.; Qiu, J.; Yao, Y.; Zhang, A.; Zhang, L.; et al. Pre-Trained Models: Past, Present and Future. *AI Open* **2021**, *2*, 225–250. [[CrossRef](#)]
14. Cervantes-Sanchez, F.; Cruz-Aceves, I.; Hernandez-Aguirre, A.; Hernandez-Gonzalez, M.A.; Solorio-Meza, S.E. Automatic Segmentation of Coronary Arteries in X-Ray Angiograms Using Multiscale Analysis and Artificial Neural Networks. *Appl. Sci.* **2019**, *9*, 5507. [[CrossRef](#)]
15. Nwogu, I.; Lorigo, L. Fast temporal tracking and 3D reconstruction of a single coronary vessel. In Proceedings of the 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 16 September–19 October 2007; Volume 5, p. V-537.

16. Kerkeni, A.; Benabdallah, A.; Manzanera, A.; Bedoui, M.H. A Coronary Artery Segmentation Method Based on Multiscale Analysis and Region Growing. *CMIG* **2016**, *48*, 49–61. [[CrossRef](#)]
17. Jiang, Y.; Zhang, H.; Tan, N.; Chen, L. Automatic Retinal Blood Vessel Segmentation Based on Fully Convolutional Neural Networks. *Symmetry* **2019**, *11*, 1112. [[CrossRef](#)]
18. Yang, S.; Kweon, J.; Roh, J.-H.; Lee, J.-H.; Kang, H.; Park, L.-J.; Kim, D.J.; Yang, H.; Hur, J.; Kang, D.-Y.; et al. Deep Learning Segmentation of Major Vessels in X-Ray Coronary Angiography. *Sci. Rep.* **2019**, *9*, 16897. [[CrossRef](#)] [[PubMed](#)]
19. Iyer, K.; Najarian, C.P.; Fattah, A.A.; Arthurs, C.J.; Soroushmehr, S.M.R.; Subban, V.; Sankardas, M.A.; Nadakuditi, R.R.; Nallamotheu, B.K.; Figueroa, C.A. AngioNet: A Convolutional Neural Network for Vessel Segmentation in X-Ray Angiography. *Sci. Rep.* **2021**, *11*, 18066. [[CrossRef](#)]
20. Zhu, X.; Cheng, Z.; Wang, S.; Chen, X.; Lu, G. Coronary Angiography Image Segmentation Based on PSPNet. *Comput. Methods Programs Biomed.* **2021**, *200*, 105897. [[CrossRef](#)]
21. Jiang, Z.; Ou, C.; Qian, Y.; Rehan, R.; Yong, A. Coronary Vessel Segmentation Using Multiresolution and Multiscale Deep Learning. *IMU* **2021**, *24*, 100602. [[CrossRef](#)]
22. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
23. Saood, A.; Hatem, I. COVID-19 Lung CT Image Segmentation Using Deep Learning Methods: U-Net versus SegNet. *BMC Med. Imaging* **2021**, *21*, 19. [[CrossRef](#)]
24. Zhao, C.; Bober, R.; Tang, H.; Tang, J.; Dong, M.; Zhang, C.; He, Z.; Esposito, M.; Xu, Z.; Zhou, W. Semantic Segmentation to Extract Coronary Arteries in Invasive Coronary Angiograms. *J. Adv. Appl. Comput. Math.* **2022**, *9*, 76–85. [[CrossRef](#)]
25. Tran, S.-T.; Cheng, C.-H.; Nguyen, T.-T.; Le, M.-H.; Liu, D.-G. TMD-Unet: Triple-Unet with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation. *Healthcare* **2021**, *9*, 54. [[CrossRef](#)]
26. Zunair, H.; Ben Hamza, A. Sharp U-Net: Depthwise Convolutional Network for Biomedical Image Segmentation. *Comput. Biol. Med.* **2021**, *136*, 104699. [[CrossRef](#)] [[PubMed](#)]
27. Cai, L.; Gao, J.; Zhao, D. A Review of the Application of Deep Learning in Medical Image Classification and Segmentation. *Ann. Transl. Med.* **2020**, *8*, 713. [[CrossRef](#)] [[PubMed](#)]
28. Semoglou, A.-A.; Spiliotis, E.; Assimakopoulos, V. Image-Based Time Series Forecasting: A Deep Convolutional Neural Network Approach. *Neural Netw.* **2023**, *157*, 39–53. [[CrossRef](#)] [[PubMed](#)]
29. Wang, J.; Liu, Q.; Xie, H.; Yang, Z.; Zhou, H. Boosted EfficientNet: Detection of Lymph Node Metastases in Breast Cancer Using Convolutional Neural Networks. *Cancers* **2021**, *13*, 661. [[CrossRef](#)]
30. Antczak, K.; Liberadzki, Ł. Stenosis Detection with Deep Convolutional Neural Networks. *MATEC Web Conf.* **2018**, *210*, 04001. [[CrossRef](#)]
31. Gil-Rios, M.-A.; Guryev, I.V.; Cruz-Aceves, I.; Avina-Cervantes, J.G.; Hernandez-Gonzalez, M.A.; Solorio-Meza, S.E.; Lopez-Hernandez, J.M. Automatic Feature Selection for Stenosis Detection in X-Ray Coronary Angiograms. *Mathematics* **2021**, *9*, 2471. [[CrossRef](#)]
32. Ovalle-Magallanes, E.; Avina-Cervantes, J.G.; Cruz-Aceves, I.; Ruiz-Pinales, J. Transfer Learning for Stenosis Detection in X-Ray Coronary Angiography. *Mathematics* **2020**, *8*, 1510. [[CrossRef](#)]
33. Lee, J.O.; Park, E.-A.; Park, D.; Lee, W. Deep Learning-Based Automated Quantification of Coronary Artery Calcification for Contrast-Enhanced Coronary Computed Tomographic Angiography. *J. Cardiovasc. Dev. Dis.* **2023**, *10*, 143. [[CrossRef](#)]
34. Jamil, S.; Roy, A.M. An Efficient and Robust Phonocardiography (PCG)-Based Valvular Heart Diseases (VHD) Detection Framework Using Vision Transformer (ViT). *Comput. Biol. Med.* **2023**, *158*, 106734. [[CrossRef](#)]
35. Liu, X.; Wang, X.; Chen, D.; Zhang, H. Automatic Quantitative Coronary Analysis Based on Deep Learning. *Appl. Sci.* **2023**, *13*, 2975. [[CrossRef](#)]
36. Suzuki, N.; Asano, T.; Nakazawa, G.; Aoki, J.; Tanabe, K.; Hibi, K.; Ikari, Y.; Kozuma, K. Clinical Expert Consensus Document on Quantitative Coronary Angiography from the Japanese Association of Cardiovascular Intervention and Therapeutics. *Cardiovasc. Interv. Ther.* **2020**, *35*, 105–116. [[CrossRef](#)]
37. Cury, R.C.; Leipsic, J.; Abbara, S.; Achenbach, S.; Berman, D.; Bittencourt, M.; Budoff, M.; Chinnaiyan, K.; Choi, A.D.; Ghoshhajra, B.; et al. CAD-RADS™ 2.0—2022 Coronary Artery Disease-Reporting and Data System. *JACC Cardiovasc. Imaging* **2022**, *15*, 1974–2001. [[CrossRef](#)]
38. Fan, C.; Hauser, H. Fast and Accurate CNN-Based Brushing in Scatterplots. *Comput. Graph. Forum* **2018**, *37*, 111–120. [[CrossRef](#)]
39. Anguita, D.; Ghelardoni, L.; Ghio, A.; Oneto, L.; Ridella, S. The 'K' in K-fold Cross Validation. In Proceedings of the 31th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 25–27 April 2017; pp. 441–446.
40. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science, Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241. [[CrossRef](#)]
41. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access* **2021**, *9*, 82031–82057. [[CrossRef](#)]
42. Ashraf, H.; Waris, A.; Ghafoor, M.F.; Gilani, S.O.; Niazi, I.K. Melanoma Segmentation Using Deep Learning with Test-Time Augmentations and Conditional Random Fields. *Sci. Rep.* **2022**, *12*, 3948. [[CrossRef](#)] [[PubMed](#)]

43. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [[CrossRef](#)]
44. Zhou, Z.; Siddiquee, M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2018; pp. 3–11.
45. ImageNet. Available online: <https://www.image-net.org/update-mar-11-2021.php> (accessed on 15 April 2023).
46. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
47. Yang, Y.; Zhang, L.; Du, M.; Bo, J.; Liu, H.; Ren, L.; Li, X.; Deen, M.J. A Comparative Analysis of Eleven Neural Networks Architectures for Small Datasets of Lung Images of COVID-19 Patients toward Improved Clinical Decisions. *Comput. Biol. Med.* **2021**, *139*, 104887. [[CrossRef](#)]
48. Rahman, T.; Chowdhury, M.E.H.; Khandakar, A.; Islam, K.R.; Islam, K.F.; Mahbub, Z.B.; Kadir, M.A.; Kashem, S. Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-Ray. *Appl. Sci.* **2020**, *10*, 3233. [[CrossRef](#)]
49. Almuayqil, S.N.; Abd El-Ghany, S.; Elmogy, M. Computer-Aided Diagnosis for Early Signs of Skin Diseases Using Multi Types Feature Fusion Based on a Hybrid Deep Learning Model. *Electronics* **2022**, *11*, 4009. [[CrossRef](#)]
50. Tan, M.; Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.
51. Putra, T.A.; Rufaida, S.I.; Leu, J.-S. Enhanced Skin Condition Prediction Through Machine Learning Using Dynamic Training and Testing Augmentation. *IEEE Access* **2020**, *8*, 40536–40546. [[CrossRef](#)]
52. Gorji, H.T.; Shahabi, S.M.; Sharma, A.; Tande, L.Q.; Husarik, K.; Qin, J.; Chan, D.E.; Baek, I.; Kim, M.S.; MacKinnon, N.; et al. Combining Deep Learning and Fluorescence Imaging to Automatically Identify Fecal Contamination on Meat Carcasses. *Sci. Rep.* **2022**, *12*, 2392. [[CrossRef](#)]
53. Seidaliyeva, U.; Akhmetov, D.; Ilipbayeva, L.; Matson, E.T. Real-Time and Accurate Drone Detection in a Video with a Static Background. *Sensors* **2020**, *20*, 3856. [[CrossRef](#)] [[PubMed](#)]
54. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
55. Ichim, L.; Popescu, D. Melanoma Detection Using an Objective System Based on Multiple Connected Neural Networks. *IEEE Access* **2020**, *8*, 179189–179202. [[CrossRef](#)]
56. Ardakani, A.A.; Kanafi, A.R.; Acharya, U.R.; Khadem, N.; Mohammadi, A. Application of Deep Learning Technique to Manage COVID-19 in Routine Clinical Practice Using CT Images: Results of 10 Convolutional Neural Networks. *Comput. Biol. Med.* **2020**, *121*, 103795. [[CrossRef](#)] [[PubMed](#)]
57. Li, F.; Yan, L.; Wang, Y.; Shi, J.; Chen, H.; Zhang, X.; Jiang, M.; Wu, Z.; Zhou, K. Deep Learning-Based Automated Detection of Glaucomatous Optic Neuropathy on Color Fundus Photographs. *Graefe's Arch. Clin. Exp. Ophthalmol.* **2020**, *258*, 851–867. [[CrossRef](#)]
58. Liu, Y.; Zhang, L.; Hao, Z.; Yang, Z.; Wang, S.; Zhou, X.; Chang, Q. An Xception Model Based on Residual Attention Mechanism for the Classification of Benign and Malignant Gastric Ulcers. *Sci. Rep.* **2022**, *12*, 15365. [[CrossRef](#)]
59. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
60. Al-antari, M.A.; Al-masni, M.A.; Choi, M.-T.; Han, S.-M.; Kim, T.-S. A Fully Integrated Computer-Aided Diagnosis System for Digital X-Ray Mammograms via Deep Learning Detection, Segmentation, and Classification. *Int. J. Med. Inform.* **2018**, *117*, 44–54. [[CrossRef](#)] [[PubMed](#)]
61. Nai, Y.-H.; Teo, B.W.; Tan, N.L.; O'Doherty, S.; Stephenson, M.C.; Thian, Y.L.; Chiong, E.; Reilhac, A. Comparison of Metrics for the Evaluation of Medical Segmentations Using Prostate MRI Dataset. *Comput. Biol. Med.* **2021**, *134*, 104497. [[CrossRef](#)] [[PubMed](#)]
62. Tohka, J.; van Gils, M. Evaluation of Machine Learning Algorithms for Health and Wellness Applications: A Tutorial. *Comput. Biol. Med.* **2021**, *132*, 104324. [[CrossRef](#)] [[PubMed](#)]
63. Baskaran, L.; Al'Aref, S.J.; Maliakal, G.; Lee, B.C.; Xu, Z.; Choi, J.W.; Lee, S.-E.; Sung, J.M.; Lin, F.Y.; Dunham, S.; et al. Automatic Segmentation of Multiple Cardiovascular Structures from Cardiac Computed Tomography Angiography Images Using Deep Learning. *PLoS ONE* **2020**, *15*, e0232573. [[CrossRef](#)] [[PubMed](#)]
64. Rizwan, I.; Haque, I.; Neubert, J. Deep Learning Approaches to Biomedical Image Segmentation. *IMU* **2020**, *18*, 100297. [[CrossRef](#)]
65. Chicco, D.; Warrens, M.J.; Jurman, G. The Matthews Correlation Coefficient (MCC) Is More Informative Than Cohen's Kappa and Brier Score in Binary Classification Assessment. *IEEE Access* **2021**, *9*, 78368–78381. [[CrossRef](#)]
66. Sammut, C.; Webb, G.I. *Encyclopedia of Machine Learning and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2017. [[CrossRef](#)]
67. Zhu, W.; Zeng, N.; Wang, N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. In Proceedings of the NESUG Proceedings: Health Care and Life Sciences, Baltimore, MD, USA, 14–17 November 2010; Volume 19, p. 67.
68. Rashed, B.M.; Popescu, N. Performance Investigation for Medical Image Evaluation and Diagnosis Using Machine-Learning and Deep-Learning Techniques. *Computation* **2023**, *11*, 63. [[CrossRef](#)]

69. Hicks, S.A.; Strümke, I.; Thambawita, V.; Hammou, M.; Riegler, M.A.; Halvorsen, P.; Parasa, S. On Evaluation Metrics for Medical Applications of Artificial Intelligence. *Sci. Rep.* **2022**, *12*, 5979. [[CrossRef](#)]
70. Wang, J.; Yang, Y.; Xia, B. A Simplified Cohen's Kappa for Use in Binary Classification Data Annotation Tasks. *IEEE Access* **2019**, *7*, 164386–164397. [[CrossRef](#)]
71. Rodríguez-Hernández, M.M.; Pruneda, R.E.; Rodríguez-Díaz, J.M. Statistical Analysis of the Evolutive Effects of Language Development in the Resolution of Mathematical Problems in Primary School Education. *Mathematics* **2021**, *9*, 1081. [[CrossRef](#)]
72. Shamila Ebenezer, A.; Deepa Kanmani, S.; Sivakumar, M.; Jeba Priya, S. Effect of Image Transformation on EfficientNet Model for COVID-19 CT Image Classification. *Mater. Today Proc.* **2022**, *51*, 2512–2519. [[CrossRef](#)]
73. Fan, Y.-J.; Tzeng, I.-S.; Huang, Y.-S.; Hsu, Y.-Y.; Wei, B.-C.; Hung, S.-T.; Cheng, Y.-L. Machine Learning: Using Xception, a Deep Convolutional Neural Network Architecture, to Implement Pectus Excavatum Diagnostic Tool from Frontal-View Chest X-Rays. *Biomedicines* **2023**, *11*, 760. [[CrossRef](#)] [[PubMed](#)]
74. Irkham, I.; Ibrahim, A.U.; Nwekwo, C.W.; Al-Turjman, F.; Hartati, Y.W. Current Technologies for Detection of COVID-19: Biosensors, Artificial Intelligence and Internet of Medical Things (IoMT). *Sensors* **2022**, *23*, 426. [[CrossRef](#)] [[PubMed](#)]
75. Kora, P.; Ooi, C.P.; Faust, O.; Raghavendra, U.; Gudigar, A.; Chan, W.Y.; Meenakshi, K.; Swaraja, K.; Plawiak, P.; Acharya, U.R. Transfer learning techniques for medical image analysis: A review. *Biocybern. Biomed. Eng.* **2021**, *42*, 79–107. [[CrossRef](#)]
76. Bertels, J.; Eelbode, T.; Berman, M.; Vandermeulen, D.; Maes, F.; Bisschops, R.; Blaschko, M.B. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. In *Lecture Notes in Computer Science, Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2019 Shenzhen, China, 13–17 October 2019*; Springer: Cham, Switzerland, 2019; pp. 92–100. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.