IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 April 2024, accepted 8 May 2024, date of publication 15 May 2024, date of current version 22 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3401465

RESEARCH ARTICLE

Coronary Artery Disease Classification With Different Lesion Degree Ranges Based on Deep Learning

ARIADNA JIMÉNEZ-PARTINEN[1,3], KARL THURNHOFER-HEMSI[1,3,4], (Member, IEEE), JORGE RODRÍGUEZ-CAPITÁN[2,3,4], ANA I. MOLINA-RAMOS[2,3,4], AND ESTEBAN J. PALOMO[1,3], (Member, IEEE)

[1]ITIS Software, University of Malaga, 29071 Málaga, Spain
[2]Department of Cardiology, Hospital Universitario Virgen de la Victoria, 29010 Málaga, Spain
[3]Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina (IBIMA Plataforma BIONAND), Campanillas, 29590 Málaga, Spain
[4]Centro de Investigación Biomédica en Red de Enfermedades Cardiovasculares (CIBERCV), Instituto de Salud Carlos III (ISCIII), 28029 Madrid, Spain

Corresponding author: Ariadna Jiménez-Partinen (ariadna@uma.es)

This work was supported in part by the Autonomous Government of Andalusia, Spain (Project Name: Detection, characterization and prognosis value of the non-obstructive coronary disease with deep learning) under Project UMA20-FEDERJA-108; in part by the Ministry of Science and Innovation of Spain (Project Name: Automated Detection of Non-Lesional Focal Epilepsy by Probabilistic Diffusion Deep Neural Models) under Grant ID2022-136764OA-I00; in part by European Regional Development Fund (ERDF); in part by the University of Málaga, Spain (Project Name: Detection of Coronary Stenosis Using Deep Learning Applied to Coronary Angiography) under Grant B1-2021_20; in part by the University of Málaga, Spain (Project Name: Intelligent Clinical Decision Support System for Non-Obstructive Coronary Artery Disease in Coronarographies) under Grant B4-2023_13; in part by the University of Málaga, Spain (Project Name: Detección de Trayectorias Anómalas de Vehículos en Cámaras de Tráfico) under Grant B1-2022_14; in part by the University of Málaga, Spain (Project Name: Sistema de videovigilancia basado en cámaras itinerantes robotizadas) under Grant B1-2023_18; and in part by the Fundación Unicaja (Project Name: Intelligent System to Help the Clinical Diagnosis of Non-Obstructive Coronary Artery Disease in Coronary Angiography) under Project PUNI-003_2023.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Comité de Ética de la Investigación (CEI) Provincial de Málaga.

ABSTRACT Invasive Coronary Angiography (ICA) images are considered the gold standard for assessing the state of the coronary arteries. Deep learning classification methods are widely used and well-developed in different areas where medical imaging evaluation has an essential impact due to the development of computer-aided diagnosis systems that can support physicians in their clinical procedures. In this paper, a new performance analysis of deep learning methods for binary ICA classification with different lesion degrees is reported. To reach this goal, an annotated dataset of ICA images containing the ground truth, the location of lesions, and seven possible severity degrees ranging between 0% and 100% was employed. The ICA images were divided into ''lesion'' or ''non-lesion'' patches. We aim to study how binary classification performance is affected by the different lesion degrees considered in the positive class. Therefore, five Convolutional Neural Network architectures – DenseNet-201, MobileNet-V2, NasNet-Mobile, ResNet-18, and ResNet-50 – were trained with different input images where different lesion degree ranges were gradually incorporated until considering the seven lesion degrees. Besides, four types of experiments with and without data augmentation were designed, whose F-measure and Area Under Curve (AUC) were computed. Reported results achieved an F-measure and AUC of 92.7% and 98.1%, respectively. However, lesion classification is highly affected by the degree of the lesion intended to be classified, with 15% less accuracy when <99% lesion patches are present.

INDEX TERMS Invasive coronary angiography, medical images, classification, deep learning.

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

I. INTRODUCTION

Invasive Coronary Angiography (ICA) images are one of the methods for anatomical imaging evaluation. Although the use of other non-invasive methods of assessment for Coronary Artery Disease (CAD) is increasing, it remains the gold standard method for evaluating the coronary artery state, confirming CAD, and guiding for interventions through X-ray imaging technology [1], [2], [3]. During an ICA procedure, a catheter is inserted by a percutaneous incision in the radial or femoral artery to introduce the radiocontrast agent [4].

The assessment of the stenosis severity is done visually and has a crucial subjective part that depends on the experience of the expert, having a substantial interobserver variability [5], [6]. Computer-aided diagnosis could improve the efficiency of diagnosis, supporting clinician decisions. This fact motivates the scientific community to develop and analyze different approaches to solve stenosis classification and detection tasks in ICA images. Nowadays, Convolutional Neural Networks (CNN) use the power of GPU-accelerated algorithms to recognize objects successfully and have been widely used for decision support systems and image classification, more specifically in medical images [7], [8].

In this context, only a few methods are proposed, being this research field is in an early stage because of the need for available open-access datasets [9]. To relieve this problem, Ovalle-Magallanes et al. [10] proposed a Bezier-based Generative Model, which generates synthetic image patches as a data augmentation technique.

ICA image state-of-the-art has been addressed using different techniques and strategies. Some works had been dedicated to ICA feature extraction and Principal Component Analysis (PCA) to use Machine Learning methods as classifiers [11], [12], [13]. Gil-Rios et al. [12] used the Univariate Marginal Distribution Algorithm and statistical comparison between five metaheuristics to explore the search space in order to develop an automatic feature selection of ICA images. As a result of the PCA study, a subset of 20 features was established to correctly classify ICA images, with an accuracy of 89%.

Deep Learning-based solutions are the main techniques to solve classification, segmentation, and detection tasks in ICA imaging. In order to detect single severe lesions ($\geq$ 70% of narrowing) in ICA images, a comparison among eight detector architectures considering both detection metrics and real-time data processing was presented by Danilov et al. [14], where the architecture based on Faster-RCNN Inception ResNet V2 was the most accurate single-vessel detector.

Pang et al. [15] designed a two-stage network as an object detector based on ResNet-50 structure that was developed using sequence image information from single projection ICA images. Firstly, a feature map was extracted and candidate boxes were generated and classified into stenosis or non-stenosis in the second stage. A method based on keyframes selection and classification into normal (<50% narrowing) and abnormal ($\geq$ 50% of narrowing) images using a GoogleNet Inception-V3 as based architecture was proposed by Moon et al. [16]. The location of the stenosis was also provided. Zhou et al. [17] used a three-stage method for extracting keyframes using ResNet-18 structure, vessel segmentation with U-Net model, and stenosis measurement from segmentation masks to classify Right Coronary Artery (RCA) images according to the lesion degree presented was proposed.

The main contribution of this work is to evaluate how the binary classification performance of ICA images is affected by the different lesion degrees considered in the ''lesion'' class, whose effects have been unreported before in the literature. In addition, a comparison between well-known deep neural network models is analyzed to determine the most effective model for the different lesion degree ranges. This is an exhaustive study to increase understanding of shortcomings, requirements, and potential improvements for deep learning solutions in invasive coronary angiography, approaching solutions for clinical settings, having the potential to alleviate pressure on healthcare services in general and to improve the catheterization laboratory diagnoses, treatment, and logistics in particular, as described below.

Firstly, improving understanding of the shortcomings of coronary stenoses can facilitate operators to identify lesions that might have otherwise been unnoticed, which would have a beneficial impact on patient outcomes. Secondly, recording and reporting the results of procedures (such as lesion location, severity, and whether stents have been placed) through automated ICA interpretation shortens the duration and increases the efficiency of ICA. This leads to higher amortization of the catheterization laboratory. Thirdly, this study and further work are focused on finding and applying models that might also be used to guide real-time Percutaneous Coronary Intervention (PCI) procedures. Peri-procedural analysis of ICA images, including automated functional assessment, could optimize PCI outcomes by providing a lesion-specific recommendation on a revascularization strategy, eventually with advice on stent size, length, location, and preferred strategy. Even after stenting, automated measurements on the proportion of stent under expansion and hemodynamic function may inform the operator and patient about the expected short- and long-term outcomes. Finally, we suggest that the comprehensive requirements of deep learning solutions of ICA images could be potential tools that could streamline the calculation of scales to guide clinical decision-making in complex CAD (e.g. the calculation of the SYNTAX score).

The organization of the rest of the paper is structured as follows: Section II specifies the details of the dataset used and training models. The experimental setup and results are presented in Section III. Finally, Section IV is devoted to conclusions.

TABLE 1. Demographic and baseline characteristics of patients. Data are given as % or as median (interquartile range).

Main characteristics	
Age (years)/	71.5 (58.25-78)
Sex (female-male)	47.6% - 52.4%
Diabetes mellitus	40.5%
Dyslipidemia	40.5%
Smoker	45.2%
High blood pressure	61.9%
Kidney failure	14.3%
Heart failure	14.3%
Atrial fibrillation	4.8%
Left ventricular ejection fraction	
Normal (ejection fraction >55%)	68.2%
Mild dysfunction (ejection fraction 45%-55%)	9.8%
Moderate dysfunction (ejection fraction 45%-35%)	0%
Severe dysfunction (ejection fraction <35%)	22%
Clinical indication for angiography	
Chronic coronary syndrome	4.9%
Non-ST segment elevation acute coronary syndrome	65.9%
ST segment elevation acute coronary syndrome	29.3%
Number of vessels affected	
0	23.8%
1	50%
2	14.3%
3	11.9%
Maximum degree of the coronary artery involvement	
<20%	14.3%
20-50%	66.7%
>70%	19%

II. METHODOLOGY

A. SOURCE DATA

CADICA dataset is composed of videos from 42 anonymized patients acquired at the Hospital Universitario Virgen de la Victoria in Málaga (Spain) with Artis Zee (Siemens AG, Muenchen, Germany) as cardiac angiography equipment. They have been included within the regulation set by the local ethical committee of the hospital and patient consent was waived because this is a retrospective study with anonymized data. The dataset includes different projections for the left and right coronary arteries, such as the right and left anterior obliques, with cranial and caudal angulation. Table 1 describes the demographic characteristics of patients included in the dataset.

In conjunction with a team of cardiologists, a selection of frames, where the radiocontrast had been perfusing correctly or the lesion was discernible, was done for each video. Furthermore, these frames were annotated, delimiting the region of interest by bounding boxes and organizing them into categories. The possible clinical categories were established according to seven possible lesion degree ranges depending on the narrowing of the vessel, in ascending order: $< 20\%$, $[20\%, 49\%]$, $[50\%, 69\%]$, $[70\%, 89\%]$, $[90\%, 98\%]$, 99% and 100%. The 99% and 100% lesion categories have a particular morphology. A 100% lesion is a total occlusion of the vessel, from which the continuation of the vessel is imperceptible. The 99% lesions present a gap, the radiocontrast is imperceptible in the narrow, but the continuation of the vessel is visible. The rest of the categories are assessed depending on the grade of narrowing with respect to the lumen on the vessel. In total, there are 3,900 images with at least one lesion and 1,943 images with no visible lesions.

B. DATA PREPROCESSING

The present work focuses on the classification of patches, i.e. equal subdivisions, of ICA images. The raw images have size 512×512 pixels, which were divided into a 4×4 grid, and then resized to 32×32 pixels. This way, we want to preserve spatial information near the lesion, and the downscale was done for training performance. The use of bigger patches outcomes worse results and requires more computational resources. These patches were labeled with the corresponding lesion degree if the centroid of the lesion bounding box fell into it. However, it is possible to encounter different types of lesions on a single patch. The severest category was chosen if more than one bounding box centroid was found. Otherwise, it was assessed as ''non-lesion'' patches. Fig. 1 shows representative samples of patches for each possible category.

This procedure implied a vast increase in the ''non-lesion'' class, the negative class. To relieve this imbalance between patches with and without lesions, the latter set was reduced before the training. First, the background ''non-lesion'' patches were removed: a basic mask of the ICA images was extracted to segment main vessels to carry out this reduction. The process of generating masks was based on applying a sequence of morphological operations, as follows: subtraction of the result of histogram equalization and average filter, thresholding, removal of little objects, dilatation, erosion, and Gaussian filter.

The masks were split into patches and those patches where the mask had less than 2% of vessel pixels were discarded. However, both sets still are unbalanced, so to prevent this issue, a random reduction of the ''non-lesion'' class was applied, equalizing both classes to have the same number of elements.

Once both classes had been equalized, data augmentation was employed by applying different random basic spatial operations of the original patches. Particularly, these basic operations were:

- Translations in the X and Y axis in a random range of $[-4, 4]$ pixels, Fig. 2(a).
- Scaling of the images randomly with a scale factor in a range of $[0.9, 1.2]$, Fig. 2(c).
- Flip horizontally and vertically, Fig. 2(d).

C. CONVOLUTIONAL NEURAL NETWORKS

In this study, different well-known Convolutional Neural Network (CNN) architectures were employed to analyze their performance concerning the positive class assigned to binary classify ICA images into ''lesion'' and ''non-lesion'' classes. CNNs are based on convolutional layers, where former layers extract basic features, latter layers extract more specific features, and pooling layers are used to subsample features maps, and fully connected layers as final classifier [18], [19].

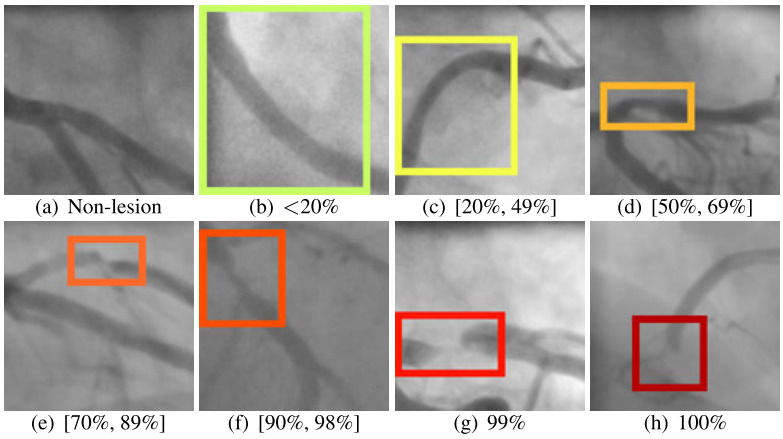


FIGURE 1. Samples of the eight lesion ranges in which patches are categorized. A colored bounding box indicates the location and degree of the lesion.

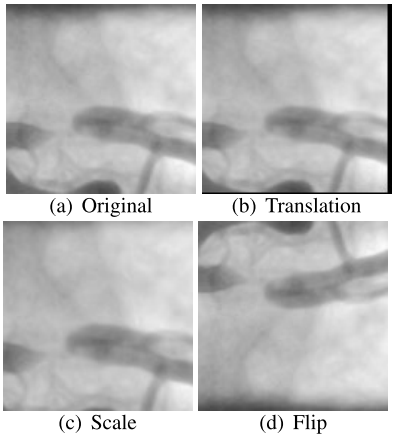


FIGURE 2. Examples of the modifications applied to the training sets to augment data.

Five widely used in the literature pre-trained architectures were selected:

- DenseNet-201, characterized by implementing dense blocks connecting their layers to all former layers [20].
- MobileNet-V2 is a mobile neural network based on the combination of depthwise convolution, which applies to the input a single filter without new features, and pointwise convolution, which produces a linear combination output extracting features [21], [22].
- NasNet-Mobile is the smallest model of NasNet versions, whose architectures are designed by Neural Architecture Search (NAS) which finds the best cells or basic blocks using the reinforcement learning technique [23].
- ResNet-18 and ResNet-50 are Residual Networks (ResNets) that introduce the concept of residual connections, implementing shortcut connections where certain convolutional layers can be skipped at one time [19], [24].

D. EVALUATION METRICS

In order to quantify the performance of different methods to classify ICA images as a binary classification task, the main four representative parameters are used: True Positive (*TP*), True Negative (*TN*), False Positive (*FP*) and False Negative (*FN*) [25]. F-measure is one of the related metrics that provides good overall performance because it integrates Precision and Recall measures under the concept of harmonic mean [26]. Precision indicates the rate of correctly positive samples over total positive predicted samples, while Recall is the proportion of correctly positive samples overall actual positive samples. Another measure involving two measures is the area under a ROC curve (AUC), which is also calculated. AUC corresponds to the integral of an ROC curve which shows the Recall versus the Specificity for different thresholds of classification scores. Specificity is the proportion of correctly classified negative samples out of the total actual negative samples.

The mentioned measures range in $[0, 1]$ (the higher is better), and are defined as follows:

$$F\text{-}measure = 2 \cdot \frac{Pre \cdot Rec}{Pre + Rec} \qquad Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \qquad Specificity = \frac{TN}{FP + TN} \tag{2}$$

III. EXPERIMENTAL RESULTS

A. TRAINING AND EXPERIMENTS DESCRIPTION

This work aims to analyze the impact on the performance of the binary "lesion"/"non-lesion" classification when different degrees of lesions are considered into the positive class. For each experiment, the positive class, i.e., the "lesion" class, was set up by lesion degrees, including all higher degrees. For example, $\geq$ 90% positive class includes 100%, 99%, and [98%, 90%] categories. Besides, the "non-lesion" class was randomly reduced to equalize the number of patches of the lesion class, as above mentioned. Both classes

TABLE 2. Training/test patches used for each established strategy and lesion range. The number of samples of each class, lesion and non-lesion, is one-half.

Lesion Range	With 100% and 99% lesions	With 100% and 99% lesions + Data Augmentation	W/o 100% and 99% lesions	W/o 100% and 99% lesions + Data Augmentation
100%	242 / 62	484 / 62	-	-
$\geq$ 99%	258 / 90	516 / 90	-	-
$\geq$ 90%	874 / 342	1748 / 342	706 / 206	1412 / 206
$\geq$ 70%	1442 / 586	2884 / 586	1386 / 338	2772 / 338
$\geq$ 50%	3170 / 664	6340 / 664	2820 / 710	5640 / 710
$\geq$ 20%	5064 / 1354	10128 / 1354	4894 / 1220	9788 /1220
$>$ 0%	8626 / 2344	17252 / 2344	8230 / 2436	16460 /2436

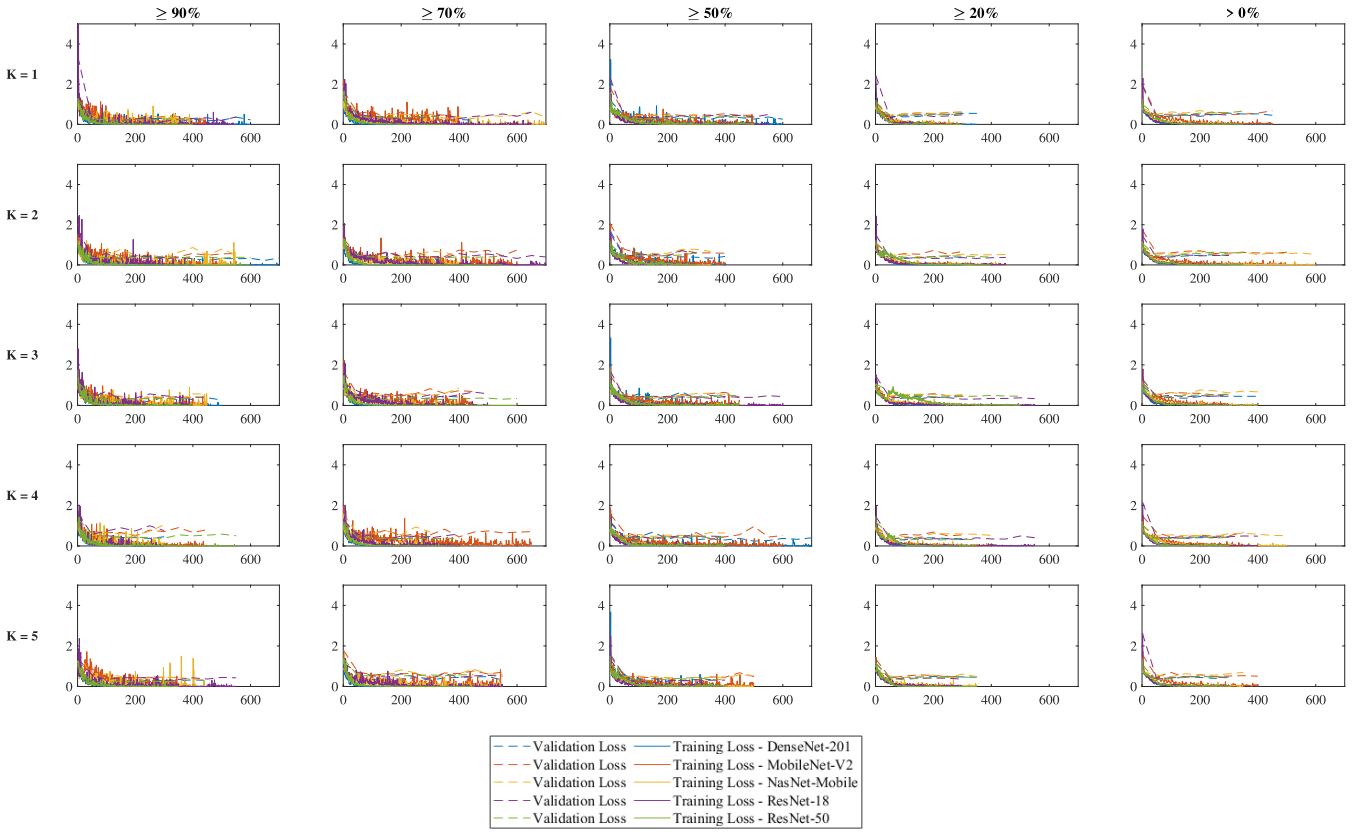


FIGURE 3. Training and validation loss for the "*W/o 100% and 99% lesions*" strategy and each proposed CNN architecture. Columns represent the lesion range considered as positive class and rows the five folds applied for stratified cross-validation.

were divided into training (80%) and test (20%) sets by videos, i.e., frames of the same video in the train set are unavailable for the test set because frames of the same sequence are very similar. This way allows for estimating a fairer performance evaluation.

Due to the clear morphological difference between lesions of 100% and 99% severity compared to the remaining degrees, four training strategies were established, including or not these high levels of severity and the use of data augmentation. The first experiment, named "*With 100% and 99% lesions*", contemplates seven categories, i. e. all possible degrees, therefore seven positive classes are established, in each of which the higher former degrees are included. On the contrary, another strategy, "*W/o 100% and 99% lesions*", considers only five categories, excluding lesions of 100% and 99% severity, so five positive classes are determined. Finally, two extra strategies were considered by applying data augmentation to the training set (80% of each class), doubling the amount of data. The number of patches used in the training process is reported in Table 2, where the number of each class is one-half. It must be considered that the sizes of train sets of adjacent categories may mismatch because of the different number of frames in video sequences.

Regarding CNNs training, we set some hyperparameters: validation frequency = 50, validation patience = 5, and maximum epochs = 50, while the batch size was set according to the number of training patches to keep the rate of iterations in all training processes. The binary cross entropy function was used as the loss function to train the models. In contrast, we tuned up the optimizer and the initial learning rate. Three different algorithms were compared: Adam (adaptive moment estimation), SGDM (Stochastic

TABLE 3. F-measure obtained on the test set using 5-fold stratified cross-validation for the four strategies. The highest values by rows are shown in bold.

Strategy	Lesion Range	Convolutional Neural Networks Models				
		DenseNet-201	MobileNet-V2	NasNet-Mobile	ResNet-18	ResNet-50
With 100% and 99% lesions	100%	0.897 ± 0.019	0.866 ± 0.091	0.884 ± 0.056	0.853 ± 0.089	$\mathbf{0.920 \pm 0.031}$
	$\geq 99\%$	$\mathbf{0.915 \pm 0.026}$	0.915 ± 0.040	0.908 ± 0.059	0.850 ± 0.110	0.893 ± 0.093
	$\geq 90\%$	0.592 ± 0.082	0.575 ± 0.031	0.574 ± 0.060	$\mathbf{0.651 \pm 0.051}$	0.582 ± 0.047
	$\geq 70\%$	$\mathbf{0.721 \pm 0.049}$	0.635 ± 0.057	0.655 ± 0.035	0.677 ± 0.040	0.654 ± 0.019
	$\geq 50\%$	$\mathbf{0.660 \pm 0.027}$	0.647 ± 0.035	0.632 ± 0.045	0.631 ± 0.057	0.624 ± 0.023
	$\geq 20\%$	$\mathbf{0.674 \pm 0.025}$	0.669 ± 0.046	0.655 ± 0.051	0.651 ± 0.028	0.633 ± 0.036
	$> 0\%$	0.690 ± 0.019	$\mathbf{0.704 \pm 0.012}$	0.673 ± 0.030	0.679 ± 0.029	0.675 ± 0.019
With 100% and 99% lesions + Data Augmentation	100%	0.894 ± 0.034	0.893 ± 0.098	0.803 ± 0.081	$\mathbf{0.927 \pm 0.032}$	0.914 ± 0.060
	$\geq 99\%$	$\mathbf{0.905 \pm 0.023}$	0.860 ± 0.040	0.894 ± 0.038	0.868 ± 0.045	0.884 ± 0.017
	$\geq 90\%$	0.520 ± 0.062	0.522 ± 0.118	0.479 ± 0.069	0.582 ± 0.101	$\mathbf{0.602 \pm 0.033}$
	$\geq 70\%$	0.676 ± 0.035	0.683 ± 0.024	$\mathbf{0.684 \pm 0.024}$	0.652 ± 0.081	0.637 ± 0.055
	$\geq 50\%$	0.600 ± 0.056	0.636 ± 0.062	$\mathbf{0.642 \pm 0.053}$	0.619 ± 0.079	0.540 ± 0.029
	$\geq 20\%$	0.645 ± 0.019	$\mathbf{0.646 \pm 0.009}$	0.633 ± 0.029	0.644 ± 0.017	0.615 ± 0.024
	$> 0\%$	0.689 ± 0.064	0.672 ± 0.030	$\mathbf{0.696 \pm 0.020}$	0.693 ± 0.009	0.678 ± 0.026
W/o 100% and 99% lesions	$\geq 90\%$	0.596 ± 0.148	0.497 ± 0.119	$\mathbf{0.606 \pm 0.106}$	0.561 ± 0.123	0.420 ± 0.070
	$\geq 70\%$	$\mathbf{0.715 \pm 0.053}$	0.701 ± 0.082	0.666 ± 0.016	0.659 ± 0.124	0.598 ± 0.069
	$\geq 50\%$	0.711 ± 0.047	$\mathbf{0.719 \pm 0.034}$	0.694 ± 0.015	0.700 ± 0.067	0.673 ± 0.079
	$\geq 20\%$	0.659 ± 0.024	0.641 ± 0.022	0.657 ± 0.028	$\mathbf{0.661 \pm 0.030}$	0.605 ± 0.069
	$> 0\%$	$\mathbf{0.694 \pm 0.031}$	0.666 ± 0.015	0.669 ± 0.019	0.658 ± 0.031	0.656 ± 0.032
W/o 100% and 99% lesions + Data Augmentation	$\geq 90\%$	0.494 ± 0.057	$\mathbf{0.614 \pm 0.118}$	0.558 ± 0.103	0.450 ± 0.055	0.446 ± 0.115
	$\geq 70\%$	$\mathbf{0.738 \pm 0.052}$	0.707 ± 0.061	0.721 ± 0.044	0.671 ± 0.065	0.567 ± 0.025
	$\geq 50\%$	$\mathbf{0.750 \pm 0.052}$	0.740 ± 0.034	0.729 ± 0.007	0.695 ± 0.026	0.675 ± 0.032
	$\geq 20\%$	0.651 ± 0.032	0.655 ± 0.031	$\mathbf{0.673 \pm 0.022}$	0.660 ± 0.020	0.664 ± 0.018
	$> 0\%$	0.686 ± 0.029	0.676 ± 0.012	$\mathbf{0.693 \pm 0.008}$	0.663 ± 0.028	0.676 ± 0.024

TABLE 4. AUC obtained on the test set using 5-fold stratified cross-validation, for the four strategies. The highest values by rows are shown in bold.

Strategy	Lesion Range	Convolutional Neural Networks Models				
		DenseNet-201	MobileNet-V2	NasNet-Mobile	ResNet-18	ResNet-50
With 100% and 99% lesions	100%	$\mathbf{0.971 \pm 0.023}$	0.950 ± 0.035	0.943 ± 0.039	0.969 ± 0.020	0.958 ± 0.013
	$\geq 99\%$	$\mathbf{0.981 \pm 0.013}$	0.967 ± 0.027	0.952 ± 0.075	0.954 ± 0.028	0.967 ± 0.045
	$\geq 90\%$	0.800 ± 0.024	0.746 ± 0.029	0.800 ± 0.063	$\mathbf{0.839 \pm 0.026}$	0.825 ± 0.030
	$\geq 70\%$	$\mathbf{0.863 \pm 0.024}$	0.845 ± 0.028	0.823 ± 0.014	0.837 ± 0.021	0.823 ± 0.015
	$\geq 50\%$	$\mathbf{0.840 \pm 0.011}$	0.791 ± 0.019	0.828 ± 0.028	0.815 ± 0.038	0.786 ± 0.046
	$\geq 20\%$	$\mathbf{0.785 \pm 0.009}$	0.775 ± 0.010	0.779 ± 0.021	0.783 ± 0.011	0.767 ± 0.012
	$> 0\%$	0.807 ± 0.009	0.813 ± 0.004	$\mathbf{0.814 \pm 0.012}$	0.806 ± 0.007	0.807 ± 0.013
With 100% and 99% lesions + Data Augmentation	100%	0.949 ± 0.022	0.924 ± 0.073	0.914 ± 0.052	$\mathbf{0.973 \pm 0.017}$	0.973 ± 0.026
	$\geq 99\%$	$\mathbf{0.968 \pm 0.017}$	0.930 ± 0.025	0.954 ± 0.027	0.944 ± 0.005	0.952 ± 0.029
	$\geq 90\%$	0.732 ± 0.046	0.761 ± 0.035	0.749 ± 0.035	0.749 ± 0.041	$\mathbf{0.800 \pm 0.017}$
	$\geq 70\%$	$\mathbf{0.828 \pm 0.022}$	0.809 ± 0.030	0.825 ± 0.016	0.820 ± 0.041	0.783 ± 0.031
	$\geq 50\%$	0.799 ± 0.037	0.782 ± 0.032	$\mathbf{0.812 \pm 0.022}$	0.795 ± 0.020	0.743 ± 0.016
	$\geq 20\%$	$\mathbf{0.767 \pm 0.022}$	0.753 ± 0.016	0.751 ± 0.021	0.749 ± 0.021	0.739 ± 0.019
	$> 0\%$	$\mathbf{0.807 \pm 0.042}$	0.797 ± 0.013	0.796 ± 0.024	0.807 ± 0.005	0.785 ± 0.012
W/o 100% and 99% lesions	$\geq 90\%$	$\mathbf{0.846 \pm 0.040}$	0.771 ± 0.051	0.800 ± 0.047	0.828 ± 0.047	0.772 ± 0.071
	$\geq 70\%$	$\mathbf{0.854 \pm 0.010}$	0.818 ± 0.061	0.838 ± 0.013	0.823 ± 0.060	0.800 ± 0.021
	$\geq 50\%$	$\mathbf{0.858 \pm 0.015}$	0.853 ± 0.020	0.820 ± 0.012	0.844 ± 0.026	0.829 ± 0.029
	$\geq 20\%$	$\mathbf{0.805 \pm 0.014}$	0.785 ± 0.018	0.772 ± 0.014	0.794 ± 0.027	0.765 ± 0.054
	$> 0\%$	$\mathbf{0.817 \pm 0.014}$	0.789 ± 0.012	0.794 ± 0.011	0.801 ± 0.016	0.789 ± 0.026
W/o 100% and 99% lesions + Data Augmentation	$\geq 90\%$	0.777 ± 0.041	0.794 ± 0.061	$\mathbf{0.803 \pm 0.035}$	0.729 ± 0.052	0.727 ± 0.066
	$\geq 70\%$	$\mathbf{0.858 \pm 0.013}$	0.790 ± 0.031	0.828 ± 0.025	0.805 ± 0.024	0.760 ± 0.011
	$\geq 50\%$	$\mathbf{0.861 \pm 0.026}$	0.830 ± 0.018	0.838 ± 0.014	0.830 ± 0.023	0.825 ± 0.020
	$\geq 20\%$	0.780 ± 0.021	0.772 ± 0.025	0.778 ± 0.008	$\mathbf{0.786 \pm 0.014}$	0.771 ± 0.028
	$> 0\%$	$\mathbf{0.805 \pm 0.014}$	0.776 ± 0.011	0.788 ± 0.015	0.780 ± 0.014	0.777 ± 0.012

Gradient Descent with Momentum), and RMSProp (Root Mean Square Propagation). Four initial learning rates were tested: 0.01, 0.001, 0.0001, and 0.00001. In total, there are 12 possible hyperparameter combinations for each threshold, delimiting the positive class in each strategy and network. 5-fold stratified cross-validation was implemented to

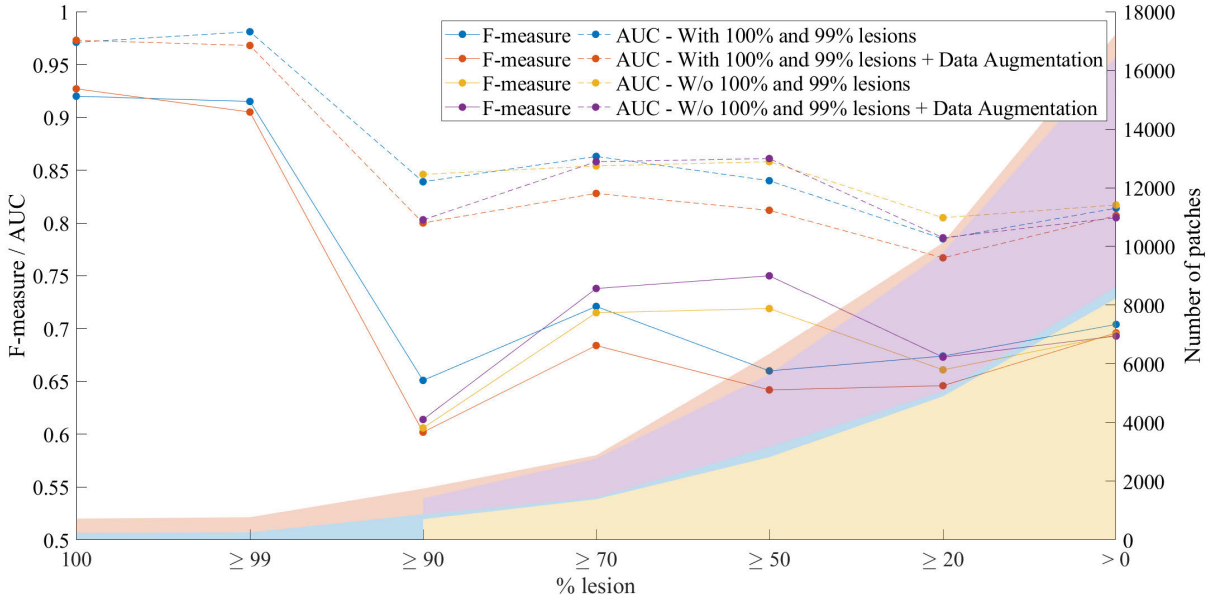


FIGURE 4. Highest F-measure and AUC obtained in the test set with 5-fold cross-validation. The number of patches in train sets are shown in the right Y-axis and are represented using shaded areas, where each color represents a strategy whose lesion degrees are included and indicated in the X-axis.

compare all these possibilities reliably. The learning rate and optimizer were selected for each CNN based on the average validation accuracy among the 5 folds and these optimal settings were reported in the Supplementary Material. In Fig. 3 training and validation loss per fold and lesion range for each trained model are depicted for "*W/o 100% and 99% lesions*" strategy. Note that there was no overfitting during the training process.

The proposed models were implemented in MATLAB R2022b on a computer system with an Intel Core i9-10900X processor, 128 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU card. Furthermore, no layer of chosen pre-trained methods was frozen, so all weights were updated during the training process according to the input class information.

B. RESULTS

Next, we present the result outcomes of the experiments above, based on applying the optimized model, and evaluating the validation accuracy obtained in the training process, to the test set. In Table 3 and Table 4, the F-measure and AUC of the chosen model are reported, respectively. The highest F-measure and AUC values by positive class and strategy established are shown in bold, standing out one architecture among the 5 CNNs for each case. These highest values are plotted in Fig. 4, where the numbers of patches for the training process are depicted too.

By analyzing Table 3, the first remarkable aspect is that all lesion ranges established as the positive class have the same tendencies independently of the strategy or methods employed. Note that the best results are attained with high-severe lesions; for 100% positive class, ResNet-18 with data augmentation and ResNet-50 stand out with the highest values, 0.927 and 0.920, respectively. However, for $\geq$ 99%, the DenseNet-201 model obtained the best outcomes for both with and without data augmentation. DenseNet-201 achieved good results even considering lower degrees ($\geq$ 70%, $\geq$ 50% and $\geq$ 20%) but its performance decreased sightly when data augmentation is applied, being NasNet-Mobile more robust. If 100% and 99% lesions are excluded, no model stands out above the rest. For instance, DenseNet-201 and MobileNet-V2 with data augmentation have a fair-to-high performance for $\geq$ 70% and $\geq$ 50% positive class, around 0.7, being the rest of the models under this. Finally, ResNet-50 for $\geq$ 90% positive class yielded a poor result, below 50% of F-measure.

Regarding Table 4, AUC values are reported. The AUC measure considers the specificity and the recall, which helps to check how well each class is classified. The first outstanding fact is that any value is above 0.7, the lowest value is 0.727 for $\geq$ 90% and using ResNet-50 with the "W/o 100% and 99% lesions + Data augmentation" strategy. It is a fair-to-high value but the corresponding F-measure from Table 3 is a poor value, 0.446. This fact points out that AUC needs to be supported by another performance metric. As F-measure, all architectures obtain similar results along all positive classes and strategies. The highest values are obtained with high-severe lesion ranges: 0.981 with DenseNet-201 for $\geq$ 99% without data augmentation, 0.973 with ResNet-18 and data augmentation, and 0.971 with DenseNet-201, both for 100% positive class. In this case, the outcomes obtained decrease slightly when moderate and mild lesion degrees are considered into the positive class. This fact makes sense because lesions are more complex to discern from healthy vessels. For positive classes which include lower lesion degrees, AUC decreases under 0.9. The best AUC results

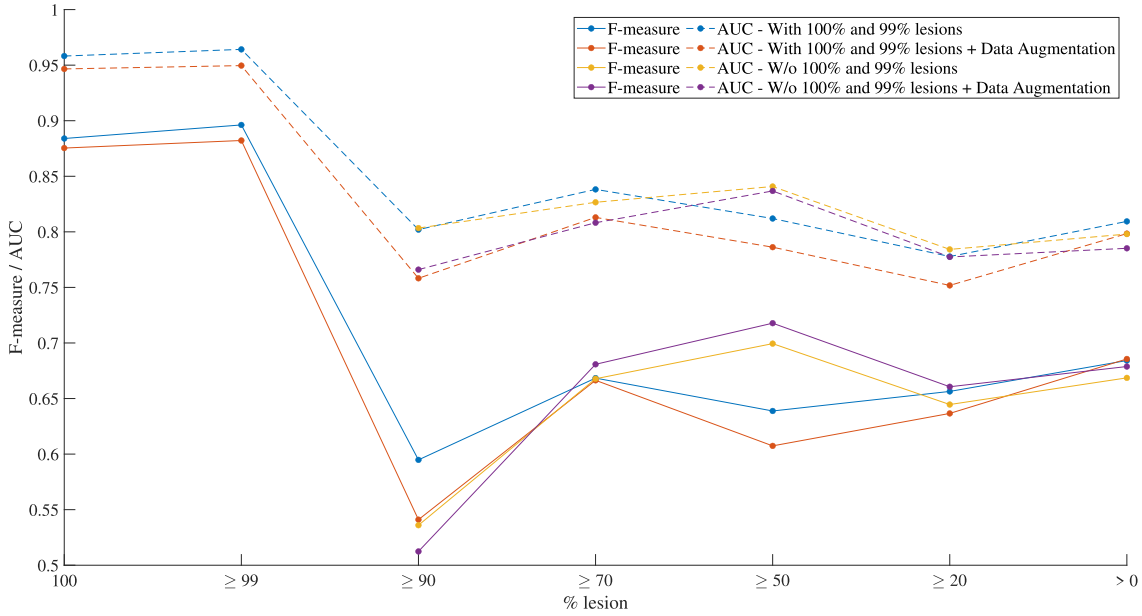


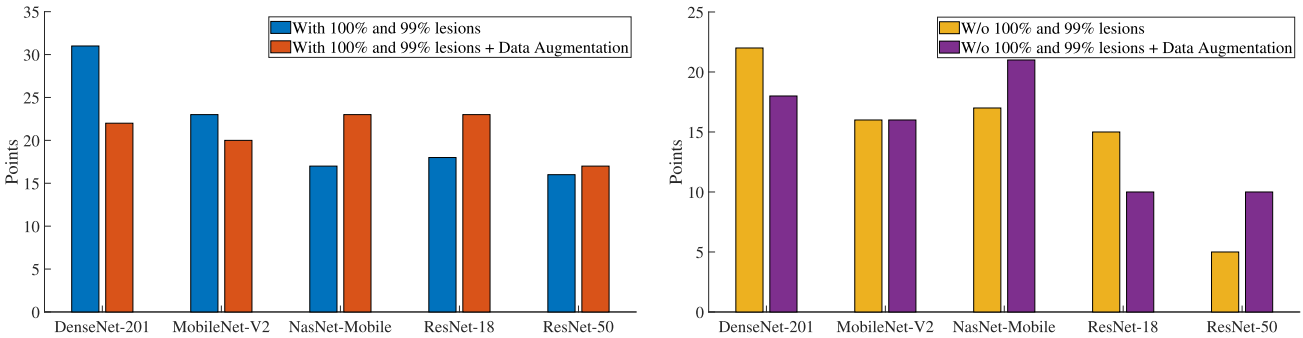
FIGURE 5. Average of F-measure and AUC obtained across all CNNs in the test set with 5-fold cross-validation to compare results in a model-independent way.

for lesion degrees below 99% are 0.863 for $\geq$ 70% using "With 100% and 99% lesions" strategy, and 0.861 for $\geq$ 50% using 'W/o 100% and 99% lesions + Data augmentation" strategy, both for the DenseNet-201 model, which is clearly the architecture that stands out over the rest, achieving the highest AUC results in most cases.

Highest F-measure and AUC obtained results in the test set with 5-fold cross-validation are given in Fig. 4. The Y-axis represents the measure (line plots, left) and the number of patches in train sets (shaded areas, right), respectively. Besides, this graph shows the tendencies for the four implemented strategies. The behavior of the performance of the strategies is similar along all "lesion" classes independently of the strategy followed and network architecture used, as can be seen in Fig. 5, where an average of the five architectures considered are computed to evaluate performance behavior in a model-independent way.

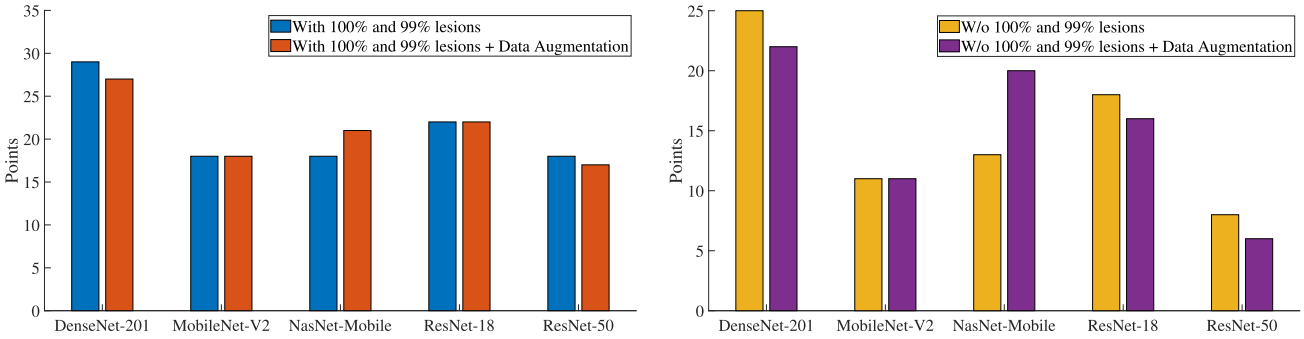
It can be seen clearly how the positive classes 100% and $\geq$ 99% are very well classified, either considering F-measure or AUC, in spite of the small number of patches used. Whereas when the $[98\%, 90\%]$ range is included the performance drops significantly, around 25% and 15% for F-measure and AUC, respectively. The lower outcomes attained with the $\geq$ 90% category are similar along all strategies. Then, the performance increases for $\geq$ 70% and $\geq$ 50% positive classes, decreasing slightly again when the lowest range degrees are included. This tendency is followed by both measures, F-measure and AUC. It could be interpreted considering the kind of lesions and the number of patches. Despite the growth of the number of "lesion" patches, the results do not improve because the classification task becomes more complex, as lower lesion degrees are more difficult to discern from "non-lesion" patches. Therefore, categories of 100% and $\geq$ 99% lesions are well classified, achieving excellent results because of their clear morphological difference, despite the small number of patches used. Furthermore, $\geq$ 70% and $\geq$ 50% ranges have good results because a larger number of patches are employed, and the lesions considered remain clearly distinguishable. Also, it could stand out that the augmentation data implies null improvement when 100% and 99% lesion degrees are considered and a slight improvement when they are excluded. This fact supports the idea that these fine-grain categories represent a highly complex problem since, despite the data augmentation applied, the methods still have difficulties in improving their performance. Additionally, there is a great difference between AUC and F-measure values, both measures involve Recall, which measures how well is classified the positive class ("lesion" class in this case), but AUC takes into account Specificity, the rate of how well classifies is the negative class, and F-measure considers Precision, which rates the positive samples correctly classified. Considering the difference obtained between measures, it could be interpreted as better Specificity than Precision, which means that the negative class, i.e., the "non-lesion" class, in some cases is slightly better classified than the positive class.

In addition, a ranking among the proposed architectures was computed in Fig. 6 and 7, where obtained points are divided by strategies. The scores were set by sorting the corresponding performance metric, F-measure or AUC, obtained by positive class in ascending order, considering better a higher value. The position indicates the points obtained. The points obtained for each lesion range were accumulated for each model. There are 7 and 5 lesion ranges, resp., including or excluding 100% and 99%, and



(a) Strategies including 100% and 99% lesion degrees. (b) Strategies excluding 100% and 99% lesion degrees.

FIGURE 6. Ranking of methods considering F-measure obtained for each positive class established.



(a) Strategies including 100% and 99% lesion degrees. (b) Strategies excluding 100% and 99% lesion degrees.

FIGURE 7. Ranking of methods considering AUC obtained for each positive class established.

5 methods, so the maximum possible score is 35 and 25 points, respectively. Focusing on strategies without data augmentation, DenseNet-201 got the highest points for the four cases, while when data augmentation is applied, points are more spread out, standing out DenseNet-201 and NasNet-Mobile. Achieving the highest points means the model is suitable to solve most of the binary classification problems analyzed here with good performance. Then, despite the type of strategy followed, DenseNet-201 stands out as the best classification network for CAD lesions.

The test sets have different sizes since positive classes are cumulative, i.e., they grow. Therefore, a more restrictive test was carried out, where all established problems use the same test size. To do it, test sets were randomly reduced to the lowest number of patches of the test sets, in this case, 62 patches in the 100% lesion category. The results obtained are provided in the Supplementary Material. In Table 5 are reported the computed F-measure and AUC obtained with DenseNet-201, one of the most suitable architectures according to the results obtained above for F-measure, AUC, and ranking evaluation. Concerning Table 5, the assumptions made previously are also corroborated in this case, despite the reduction of the test set. 100% and $\geq$ 99%, with and without data augmentation, achieve the highest values ($>$ 85% F-measure, $>$ 95% AUC), suffering a tough decrease when milder categories are included, falling to unacceptable results lower than 50% of F-score, although

TABLE 5. F-measure and AUC obtained on the test set using 5-fold stratified cross-validation with DenseNet-201 for the four strategies and the same size for the test sets.

Strategy	Lesion Range	F-measure	AUC
With 100% and 99% lesions	100%	0.897 ± 0.019	0.971 ± 0.023
	$\geq$ 99%	$\mathbf{0.900 \pm 0.033}$	$\mathbf{0.983 \pm 0.011}$
	$\geq$ 90%	0.599 ± 0.058	0.738 ± 0.044
	$\geq$ 70%	0.661 ± 0.110	0.834 ± 0.088
	$\geq$ 50%	0.534 ± 0.126	0.796 ± 0.058
	$\geq$ 20%	0.706 ± 0.031	0.820 ± 0.045
	$>$ 0%	0.669 ± 0.088	0.805 ± 0.052
With 100% and 99% lesions + Data Augmentation	100%	$\mathbf{0.894 \pm 0.034}$	$\mathbf{0.949 \pm 0.022}$
	$\geq$ 99%	0.851 ± 0.157	0.933 ± 0.092
	$\geq$ 90%	0.597 ± 0.066	0.709 ± 0.034
	$\geq$ 70%	0.699 ± 0.024	0.829 ± 0.020
	$\geq$ 50%	0.533 ± 0.103	0.796 ± 0.033
	$\geq$ 20%	0.671 ± 0.025	0.791 ± 0.030
	$>$ 0%	0.651 ± 0.098	0.777 ± 0.056
W/o 100% and 99% lesions	$\geq$ 90%	0.627 ± 0.171	0.811 ± 0.034
	$\geq$ 70%	$\mathbf{0.781 \pm 0.068}$	0.816 ± 0.058
	$\geq$ 50%	0.667 ± 0.050	$\mathbf{0.843 \pm 0.035}$
	$\geq$ 20%	0.634 ± 0.055	0.784 ± 0.023
	$>$ 0%	0.583 ± 0.096	0.778 ± 0.059
W/o 100% and 99% lesions + Data Augmentation	$\geq$ 90%	0.462 ± 0.088	0.713 ± 0.071
	$\geq$ 70%	0.650 ± 0.109	0.773 ± 0.077
	$\geq$ 50%	$\mathbf{0.769 \pm 0.072}$	$\mathbf{0.863 \pm 0.023}$
	$\geq$ 20%	0.608 ± 0.049	0.772 ± 0.049
	$>$ 0%	0.593 ± 0.042	0.767 ± 0.036

AUC remains a fair-to-high value ($>$ 70%). In all cases, AUC attains higher values than the F-measure, this fact remarks that the positive class is worse classified than the negative class, despite the class balance applied. Additionally, models trained with data augmentation neither increase

their performance substantially, reinforcing the necessity of increasing the lesion data and equalizing categories to avoid specialization on ''non-lesion'' patches.

IV. CONCLUSION

This work presents a classification methodology for coronary artery disease using invasive coronary angiography images. A total of 5 state-of-the-art deep neural models were used to distinguish between lesion and non-lesion images, varying the threshold of lesion degree to consider into the ''lesion'' class. The dataset was divided into non-overlapping patches and four types of experiments were carried out, including data augmentation and removing high-severe classes.

Results showed that the 99% and 100% categories are easy to classify as lesions (>90% F-measure, >95% AUC) even with little data, while when a lower degree is included in the positive class, the performance drops significantly (65% F-measure, 80% AUC). If those extreme cases are discarded, the networks reach 75% of F-measure and 85% AUC when data augmentation is applied when $\geq$ 70% and $\geq$ 50% severity is intended to be detected. Besides, DenseNet-201 and NasNet-Mobile demonstrated their effectiveness in solving most of the binary classification problems raised.

One of the limitations of this work is that the used dataset only contains information about 42 patients from one hospital. As future work, we would like to enhance the performed experiments with ICA images from more patients from different hospitals. Another limitation is that the different lesion degree ranges are not balanced, as can be seen in Table 2. Although they do not represent different classes, the lack of patches in each lesion degree range may affect the experimental results since the used models will learn better those lesion degree ranges with more patches available.

Further work will be focused on improving the overall classification performance. On one side, classifying each severity degree separately and including more sophisticated preprocessing steps could bring more homogeneity and, therefore, produce better results. Another approach would be the training of custom deep networks from scratch, using structures that focus one local spatial features.

The work presented aims to improve the understanding of the shortcomings, requirements, and potential improvements for deep learning solutions of ICA images, approaching solutions to clinical settings. Finally, we suggest that the requirement comprehension of deep learning solutions of ICA images could be potential tools that could streamline the automated interpretation or the calculation of scales to guide clinical decision-making in complex CAD (e.g., the calculation of the SYNTAX score), thus shorting the duration and increasing the efficiency of the CAD diagnosis and treatment procedures.

SUPPLEMENTARY INFORMATION

In addition, a Supplementary Material is provided with additional information: a table with optimal settings established for reported results is provided, and the complete results regarding the experiments with balanced test sets carried out and summarized in Table 5.

ACKNOWLEDGMENT

The authors would like to thank the computer resources, technical expertise, and assistance provided by the Supercomputing and Bioinformatics (SCBI) Center, University of Málaga, also would like to thank the support of NVIDIA Corporation with the donation of an RTX A6000 GPU with 48Gb, and also would like to thank the Universidad de Málaga and the Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina (IBIMA Plataforma BIONAND).

DATA AVAILABILITY

CADICA dataset is open-access available at the Mendeley Data repository with the data identification number: 10.17632/p9bpx9ctcv.2, and direct URL to data: https://data.mendeley.com/datasets/p9bpx9ctcv/2.

REFERENCES

[1] J.-P. Collet et al., ''2020 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation,'' *European Heart J.*, vol. 42, no. 14, pp. 1289–1367, Apr. 2021. [Online]. Available: https://academic.oup.com/eurheartj/article/42/14/1289/5898842

[2] J. Knuuti, W. Wijns, A. Saraste, D. Capodanno, E. Barbato, C. Funck-Brentano, E. Prescott, R. F. Storey, C. Deaton, T. Cuisset, and S. Agewall, ''2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes,'' *Eur. Heart J.*, vol. 41, no. 3, pp. 407–477, Jan. 2020, doi: 10.1093/eurheartj/ehz425.

[3] Y. Zhou, H. Guo, J. Song, Y. Chen, and J. Wang, ''Review of vessel segmentation and stenosis classification in X-ray coronary angiography,'' in *Proc. 13th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2021, pp. 1–5.

[4] G. Rigatelli, F. Gianese, and M. Zuin, ''Modern atlas of invasive coronary angiography views: A practical approach for fellows and young interventionalists,'' *Int. J. Cardiovascular Imag.*, vol. 38, no. 5, pp. 919–926, May 2022.

[5] L. L. Leape, R. E. Park, T. M. Bashore, J. K. Harrison, C. J. Davidson, and R. H. Brook, ''Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures,'' *ACC Current J. Rev.*, vol. 9, no. 3, pp. 95–97, May 2000.

[6] L. M. Zir, S. W. Miller, R. E. Dinsmore, J. Gilbert, and J. Harthorne, ''Interobserver variability in coronary angiography,'' *Circulation*, vol. 53, no. 4, pp. 627–632, 1976.

[7] L. Cai, J. Gao, and D. Zhao, ''A review of the application of deep learning in medical image classification and segmentation,'' *Ann. Transl. Med.*, vol. 8, no. 11, p. 713, Jun. 2020.

[8] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, ''A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises,'' *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, May 2021.

[9] G. Litjens, F. Ciompi, J. M. Wolterink, B. D. de Vos, T. Leiner, J. Teuwen, and I. Isgum, ''State-of-the-art deep learning in cardiovascular image analysis,'' *JACC, Cardiovascular Imag.*, vol. 12, no. 8, pp. 1549–1565, Aug. 2019.

[10] E. Ovalle-Magallanes, J. G. Avina-Cervantes, I. Cruz-Aceves, and J. Ruiz-Pinales, ''Improving convolutional neural network learning based on a hierarchical Bezier generative model for stenosis detection in X-ray images,'' *Comput. Methods Programs Biomed.*, vol. 219, Jun. 2022, Art. no. 106767.

[11] B. Qin, M. Jin, and S. Ding, ''Extracting heterogeneous vessels in X-ray coronary angiography via machine learning,'' in *Cardiovascular and Coronary Artery Imaging*. Amsterdam, The Netherlands: Elsevier, 2022, pp. 89–127.

[12] M.-A. Gil-Rios, I. V. Guryev, I. Cruz-Aceves, J. G. Avina-Cervantes, M. A. Hernandez-Gonzalez, S. E. Solorio-Meza, and J. M. Lopez-Hernandez, "Automatic feature selection for stenosis detection in X-ray coronary angiograms," *Mathematics*, vol. 9, no. 19, p. 2471, Oct. 2021.
[13] E. O. Rodrigues, L. O. Rodrigues, J. J. Lima, D. Casanova, F. Favarim, E. R. Dosciatti, V. Pegorini, L. S. N. Oliveira, and F. F. C. Morais, "X-ray cardiac angiographic vessel segmentation based on pixel classification using machine learning and region growing," *Biomed. Phys. Eng. Exp.*, vol. 7, no. 5, Sep. 2021, Art. no. 055026.
[14] V. V. Danilov, K. Y. Klyshnikov, O. M. Gerget, A. G. Kutikhin, V. I. Ganyukov, A. F. Frangi, and E. A. Ovcharenko, "Real-time coronary artery stenosis detection based on modern neural networks," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, Apr. 2021.
[15] K. Pang, D. Ai, H. Fang, J. Fan, H. Song, and J. Yang, "Stenosis-DetNet: Sequence consistency-based stenosis detection for X-ray coronary angiography," *Computerized Med. Imag. Graph.*, vol. 89, Apr. 2021, Art. no. 101900.
[16] J. H. Moon, D. Y. Lee, W. C. Cha, M. J. Chung, K.-S. Lee, B. H. Cho, and J. H. Choi, "Automatic stenosis recognition from coronary angiography using convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 198, Jan. 2021, Art. no. 105819.
[17] C. Zhou, T. V. Dinh, H. Kong, J. Yap, K. K. Yeo, H. K. Lee, and K. Liang, "Automated deep learning analysis of angiography video sequences for coronary artery disease," 2021, *arXiv:2101.12505*.
[18] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017.
[19] J. Wang, H. Zhu, S.-H. Wang, and Y.-D. Zhang, "A review of deep learning on medical image analysis," *Mobile Netw. Appl.*, vol. 26, pp. 351–380, 2021.
[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[22] W. Sae-Lim, W. Wettayaprasit, and P. Aiyarak, "Convolutional neural networks using MobileNet for skin lesion classification," in *Proc. 16th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Jul. 2019, pp. 242–247.
[23] N. Reddy, A. Rattani, and R. Derakhshani, "Comparison of deep learning models for biometric-based mobile user authentication," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–6.
[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[25] E. Ovalle-Magallanes, J. G. Avina-Cervantes, I. Cruz-Aceves, and J. Ruiz-Pinales, "Hybrid classical–quantum convolutional neural network for stenosis detection in X-ray coronary angiography," *Exp. Syst. Appl.*, vol. 189, Mar. 2022, Art. no. 116112.
[26] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: An overview," 2020, *arXiv:2008.05756*.

ARIADNA JIMÉNEZ-PARTINEN received the B.Sc. degree in health engineering (major in biomedical engineering) from the University of Málaga, Spain, in 2019, the M.Sc. degree in biomedical engineering from the University of País Vasco, Spain, in 2020, and the joint M.Sc. degree in advanced biotechnology from the University of Málaga and the International University of Andalusia, Spain, in 2021. She is currently pursuing the Ph.D. degree in computer technologies with the Department of Computer Languages and Computer Science, University of Málaga. She is also a Researcher with the Department of Computer Languages and Computer Science, University of Málaga. Her technical and research interests include medical image analysis, image processing, and classification.

KARL THURNHOFER-HEMSI (Member, IEEE) received the B.Sc. degree in computer engineering and the M.Sc. degree in mathematics from the University of Málaga, Spain, in 2014, and the Ph.D. degree (Hons.) in computer technologies from the University of Málaga, in 2021. In 2023, he joined the Department of Computer Languages and Computer Science, University of Málaga, where he is currently an Assistant Professor. His technical and research interests include medical image analysis, pattern recognition, image processing, evaluation methods, and learning systems.

JORGE RODRÍGUEZ-CAPITÁN received the Medical degree from the University of Malaga, Spain, the degree in cardiology from the Hospital Universitario Virgen de la Victoria, Málaga, Spain, and the Ph.D. degree in medicine from the University of Malaga, Spain, in 2016. Currently, he is developing his career both as a Clinical Cardiologist and as a Clinical Researcher. His research interests include tricuspid valvular disease, heart failure, ischemic heart disease, hypertrophic cardiomyopathy, and recently COVID-19 disease.

ANA I. MOLINA-RAMOS received the Medical degree from the University of Malaga, Spain, and the degree in cardiology from the Hospital Universitario Virgen de la Victoria, Málaga, Spain. Fellowship in advanced cardiac imaging with The Mount Sinai Hospital, New York City, in 2019. Currently, she contracted by the Instituto de Salud Carlos III (ISCI III), Madrid, Spain, combining research and clinical activity. Her research interest includes cardiac imaging.

ESTEBAN J. PALOMO (Member, IEEE) received the M.Sc. and Ph.D. degrees (Hons.) in computer engineering from the University of Málaga, Spain, in 2008 and 2013, respectively. In 2015, he joined the School of Mathematics and Computer Sciences, Yachay Tech University, Ecuador. In 2017, he joined the Department of Computer Languages and Computer Science, University of Málaga, where he is currently an Associate Professor. His current research interests include unsupervised learning, self-organization, data mining, image/video processing, and deep learning.