Received: 20 February 2024 | Revised: 26 June 2024 | Accepted: 31 July 2024

DOI: 10.1111/exsy.13708

ORIGINAL ARTICLE

Expert Systems WILEY

CADICA: A new dataset for coronary artery disease detection by using invasive coronary angiography

Ariadna Jiménez-Partinen[1,2] | Miguel A. Molina-Cabello[1,2] | Karl Thurnhofer-Hemsi[1,2,3] | Esteban J. Palomo[1,2] | Jorge Rodríguez-Capitán[2,3,4] | Ana I. Molina-Ramos[2,3,4] | Manuel Jiménez-Navarro[2,3,4,5]

[1]ITIS Software, University of Malaga, Málaga, Spain

[2]Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND, Málaga, Spain

[3]Centro de Investigación Biomédica en Red de Enfermedades Cardiovasculares (CIBERCV), Instituto de Salud Carlos III (ISCIII), Madrid, Spain

[4]Cardiology Deparment, Hospital Universitario Virgen de la Victoria, Málaga, Spain

[5]Facultad de Medicina, University of Malaga, Málaga, Spain

Correspondence
Miguel A. Molina-Cabello and Esteban J. Palomo, ITIS Software, University of Malaga, C/Arquitecto Francisco Peñalosa, 18, Málaga 29071, Spain.
Email: miguelangel@lcc.uma.es; ejpalomo@uma.es

Funding information
Autonomous Government of Andalusia (Spain), Grant/Award Number: UMA20-FEDERJA-108; Ministry of Science and Innovation of Spain, Grant/Award Number: PID2022-136764OA-I00; European Regional Development Fund; University of Malaga (Spain), Grant/Award Numbers: B1-2021_20, B4-2023_13, B1-2022_14, B1-2023_18; Fundación Unicaja, Grant/Award Number: PUNI-003_2023

Abstract

Coronary artery disease (CAD) remains the leading cause of death globally and invasive coronary angiography (ICA) is considered the gold standard of anatomical imaging evaluation when CAD is suspected. However, risk evaluation based on ICA has several limitations, such as visual assessment of stenosis severity, which has significant interobserver variability. This motivates to development of a lesion classification system that can support specialists in their clinical procedures. Although deep learning classification methods are well-developed in other areas of medical imaging, ICA image classification is still at an early stage. One of the most important reasons is the lack of available and high-quality open-access datasets. In this paper, we reported a new annotated ICA images dataset, CADICA, to provide the research community with a comprehensive and rigorous dataset of coronary angiography consisting of a set of acquired patient videos and associated disease-related metadata. This dataset can be used by clinicians to train their skills in angiographic assessment of CAD severity, by computer scientists to create computer-aided diagnostic systems to help in such assessment, and to validate existing methods for CAD detection. In addition, baseline classification methods are proposed and analysed, validating the functionality of CADICA with deep learning-based methods and giving the scientific community a starting point to improve CAD detection.

KEYWORDS
cardiovascular artery disease, classification, deep learning, invasive coronary angiography dataset, medical images

1 | INTRODUCTION

Coronary artery disease (CAD) remains the leading cause of death globally (Murphy et al., 2021; Voudris & Kavinsky, 2019). Clinical presentations of CAD are currently categorised as either acute coronary syndromes or chronic coronary syndromes, and assessing patients with suspected CAD

Abbreviations: CNN, convolutional neural networks; CAD, cardiovascular artery disease; ICA, invasive coronary angiography.

is a significant component of healthcare costs (Saraste et al., 2019). Invasive coronary angiography (ICA) is considered the gold standard of anatomical imaging evaluation when CAD is suspected (Collet et al., 2021; Knuuti et al., 2019). ICA acquisition is based on introducing radiocontrast through a catheter inserted by a percutaneous incision in the femoral or brachial artery, situated in the groin and the arm, respectively. The radiocontrast agent enhances the visibility of coronary arteries; with cardiac angiography equipment, X-ray-based, the state of the arteries is shown, allowing the clinicians to evaluate it and conclude if there is a luminal obstruction. The presence of obstructive CAD in ICA, usually defined as a lesion greater than 70%, has been recognized as an unequivocal sign of a bad cardiovascular prognosis. In contrast, it was initially proposed that non-obstructive CAD (usually defined as a lesion less than 70%) could constitute a condition related to a good cardiovascular prognosis (Kemp et al., 1986), but subsequent evidence has increasingly shown that it confers an adverse prognosis when compared to the prognosis in the absence of CAD (Radico et al., 2018; Rodríguez-Capitán et al., 2021; Wang et al., 2017). Consequently, it is currently accepted that cardiovascular risk increases when the degree of stenosis increases. However, risk evaluation based on ICA has several limitations. There is enough evidence indicating that the visual assessment of stenosis severity alone has significant interobserver variability, so this visual assessment alone does not provide us with enough information upon which to base decisions about revascularization in many patients (Curzen et al., 2014).

In addition to this, angiographic assessment of CAD severity is limited in providing consistent information regarding the physiological significance of coronary lesions. Angiography is especially limited in coronary stenoses of intermediate severity (40%–70% obstruction), where it predicts functional significance in less than 50% of lesions (Tonino et al., 2009). Visual assessment of coronary angiography fails to adequately determine lesion significance because lumen stenosis is only one variable out of many that influence the flow limitation of coronary lesions (Topol & Nissen, 1995). Lesion length, collateral flow, and the amount and health of the myocardial bed supplied are other essential factors that are not readily assessed by coronary angiography (Halon, 2018). In order to overcome the aforementioned limitations, current guidelines recommend the routine assessment of vessel physiology in the form of indices derived from invasive pressure wire, such as fractional flow reserve and the instantaneous wave-free ratio (Collet et al., 2021; Knuuti et al., 2019). Despite these recommendations, the implantation of these functional tests in clinical practice has been especially low (Gabara et al., 2020).

Many medical image datasets have been provided to the research community with the aim of developing an algorithm that can serve as a computer-aided diagnosis system (Codella et al., 2018; Matuszewski & Sintorn, 2021; Spanhol et al., 2015; Tschandl et al., 2018). However, there is a lack of available and high-quality open-access datasets regarding ICA images because most related studies use private image sets (Ovalle-Magallanes, Avina-Cervantes, et al., 2022). Some are provided by an associated medical center and used for image segmentation tasks (Iyer et al., 2021; Nasr-Esfahani et al., 2018), while others are focused on detection and classification (Cong et al., 2019; Zhou et al., 2021). None of them provides access to their data to other researchers, which is necessary to achieve advances in this field. The main contributions of this work can be listed as follows:

- A detailed and rigorous scientific invasive coronary angiography dataset formed by a set of annotated videos for different lesion degrees, as well as metadata related to diseases associated with each patient.
- A set of resources for algorithm testing. An exhaustive revision and expansion process of these tools will be carried out regularly.
- A study of the performance of deep learning architectures using the dataset with the aim of classifying ICA images according to the presence of lesions.

Most of the existing datasets are not publicly available and do not provide a detailed annotation for several CAD lesion degrees. This dataset intends to overcome this gap by providing annotations for <20%, [20%, 50%), [50%, 70%), [70%, 90%), [90%, 98%], 99%, and 100% of lesion degrees. Also, it may serve medical doctors to train their skills in angiographic assessment of CAD severity, and computer scientists to create computer-aided diagnosis systems to help with that kind of evaluation and to validate and improve existing methods for CAD detection and classification by training them on a wide variety of annotated data.

The rest of the paper is structured as follows: Section 2 describes the recent state-of-art works related to ICA. In Section 3, the most important details about the creation and organization of the CADICA dataset are given. The experimental results are shown in Section 4. A discussion is provided in Section 5. Finally, Section 6 is devoted to conclusions.

2 | RELATED WORKS

Deep learning has been thoroughly used for classification, segmentation, and detection tasks in medical imaging diagnosis, including in the area of cardiology, where the most common imaging modalities are MRI (magnetic resonance imaging), and X-ray-based, such as CCTA (Coronary Computed Tomography Angiography) and ICA (Invasive Coronary Angiography) (Song et al., 2022).

In spite of the state-of-the-art devoted to ICA images increasing, it is still at an early stage because of the lack of high-quality open-access available datasets, being the most important downside (Litjens et al., 2019). The proposed learning models are mainly based on convolutional

neural networks (CNNs) architectures, which solve specific tasks using complete images or patch-based processings, where frames are divided into sub-images that are employed to improve the training process (Ovalle-Magallanes, Alvarado-Carrillo, et al., 2022).

Patch-based approaches allow the reduction of the spatial information used as input for the prediction model as well as computational time decreases to process each sample. For instance, Nasr-Esfahani et al. (2018) implemented a segmentation system of ICA images using patches from 44 coronary angiographies, which is composed of two CNNs based on learned kernels. Right coronary artery (RCA) ICA images were used to detect and classify severe lesions, higher than 70%, by a three-phase method based on well-known architectures: YOLONet as a lesion detector, U-Net to segment and simple CNN as a classifier (Au et al., 2018). Also, a three-stage system was proposed by Wu et al. (2020), where whole images and, finally, patches were analysed. First, the U-Net network selected key-frames from 148 ICA videos. In key-frames, bounding boxes for stenosis lesions were located by a deconvolutional single-shot multi-box detector considering the temporal sequence.

However, using whole ICA images could be a first starting point to implement a solution without the necessity of preliminary image processing, where there is no restriction of location or bounding box size, that is, it is independent of lesion position and size, respectively. Additionally, this strategy avoids cutting or missing lesions. Besides, the use of complete images allows models to make use of spatial information, which is lost in patch-based processing and is very useful in determining the existence of lesions. On the contrary, more computational resources are needed, and there might be irrelevant information that affects the model performance. In this context, as first approaches, we can find the classification of healthy ICA images into the left coronary artery (LCA) and right coronary artery (RCA) images using a CNN-based model (Eschen et al., 2022) or the vessel center lines extraction using a U-Net-based model (Zhang et al., 2022). Likewise, Cong et al. (2019) proposed a more complex system based on different Inception architectures to classify ICA images into LCA and RCA images, select key-frames, and classify the present lesions. RCA from ICA images was classified using a system that combines key-frames classification and vessel segmentation by ResNet-18 and U-Net networks, respectively (Zhou et al., 2021). As well, Moon et al. (2021) employed Inception architecture to solve a binary classification between normal and abnormal ($\geq$ 50%) lesions. In order to detect single severe ($\geq$ 70%) lesions, a comparison among eight detector architectures was reported by Danilov et al. (2021), where the architecture based on Faster-RCNN Inception ResNet V2 was the most accurate single-vessel detector. Finally, to solve ICA segmentation tasks, different models based on Efficient-U-Net, Densenet121, ResNet-101, or Inception architectures have demonstrated great performance (Kim et al., 2024; Menezes et al., 2022; Yang et al., 2019; Zhao et al., 2021). Key points of related works reported are summarized in Table 1.

These reported works have some aspects in common with our study, such as the CNN-derived models used, but most of them excluded images with more than one lesion, only classified obstructive lesions, utilized smaller datasets (most of them private datasets), the detail of the artery and projections used is omitted, or a different annotation is implemented.

3 | CADICA DATASET

Next, the most important details about the creation and organization of the CADICA dataset are presented.

3.1 | Patient selection

The dataset proposed in this work consists of 668 invasive coronary angiography videos from 42 patients, acquired at Hospital Universitario Virgen de la Victoria, Málaga, Spain. 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. Prerequisites and data selection have not been performed in order to impose clinical fidelity. Therefore, a wide variability of cases and acquisition configurations were implied, while some cases were laborious to tag. This way, the dataset exhibits a high variety of pathological cases and image quality. Table 2 summarizes the baseline and demographic characteristics of patients included in the dataset.

3.2 | Acquisition protocol

The invasive coronary angiography videos were acquired as Digital Imaging and Communication in Medicine (DICOM) files recorded at 10 frames per second and with different duration (4–8 s) depending on the projection used, but they were converted to PNG images for effortless management. The frame size of each video is 512 $\times$ 512 pixels, while the length of the videos varies from 1 to 151 frames. The cardiac angiography equipment used was Artis Zee (Siemens AG, Muenchen, Germany). The dose of radiation administered in each projection ranges between 5 and 50 mGy. The protocol normally used in each angiography included five projections for the left coronary artery (LCA), such as right anterior oblique (RAO) and left anterior oblique (LAO), both with cranial and caudal angulation, with some additional projections in case of diagnostic difficulties. The projections used for the right coronary artery (RCA) are LAO and RAO, with cranial and caudal angulation.

TABLE 1 Comparison of related works in CAD.

Reference	Projection	Severity	Task	Best results
Nasr-Esfahani et al. (2018)	Unreported	Unreported	Segmentation	81% Dice / 97% Accuracy
Au et al. (2018)	RCA	< 70% lesions	Detection and classification	70% Dice / 82% C-statistic
Wu et al. (2020)	Unreported	< 50% lesions	Detection	87% sensitivity / 79% PPV
Eschen et al. (2022)	LCA and RCA	Unreported	RCA-LCA classification	98% F-measure
Zhang et al. (2022)	Unreported	Unreported	Center-line extraction	85% F-measure / 88% Precision
Cong et al. (2019)	LCA and RCA	All degrees	RCA-LCA classification and $< 25\% / > 25\%$stenosis classification	87% AUC / 60% sensitivity
Zhou et al. (2021)	LCA	< 50% lesions	Segmentation	89% F-measure
Moon et al. (2021)	LCA	< 50% lesions	Normal-abnormal classification	95% AUC / 93% Accuracy
Danilov et al. (2021)	LCA and RCA	< 70% lesions	Detection	83% mAP / 80% F-measure
Menezes et al. (2022)	LCA and RCA	Unreported	Segmentation	92% Dice
Yang et al. (2019)	LCA and RCA	< 30% lesions	Segmentation	91% F-measure
Kim et al. (2024)	LCA and RCA	< 20% lesions	Segmentation	89% sensitivity
Zhao et al. (2021)		< 20% lesions	Segmentation	89% Dice / 86% sensitivity

Figures 1 and 2 show examples of projections for the left coronary artery and the right coronary artery, respectively.

3.3 | Label protocol

A team of cardiologists was involved in the annotation of the dataset, assisted by computer scientists. For each frame, those regions of interest are delimited by a bounding box and classified into categories. This way, for each region of interest, is provided the location of the top left corner of its bounding box, its width, its height, and the label of that region of interest. The possible categories are itemized in Table 3.

To create a benchmark dataset that was as close to reality as possible, all complete recorded videos were provided, but to ease research purposes for each video, a selection of keyframes was carried out. This selection contains the list of frames that exhibit a contrast with enough appearance in order to classify the patient correctly. Videos are organized by patients, where a certain number of videos have been collected for each patient. According to their coronary artery stenosis percentage, patients are grouped into three different categories: < 20% (mild), 20%–50% (moderate), and > 70% (severe). Lesions of 100% imply a total occlusion of the vessel, while a 99% lesion presents a gap where the radiocontrast is imperceptible, but the continuation of the vessel is visible. Those lesions that had a narrowing between 50% and 70% are classified as obstructive in some studies (Finck et al., 2019) and non-obstructive in others (Kang et al., 2016), while lesions with a higher narrowing than 70% in the previous bibliography showed consensus that they should be taken as obstructive and classify as non-obstructive the lesions that are solidly taken as non-obstructive (20%–50%). Figure 3 exhibits some sample images from different patients according to their classification into these categories.

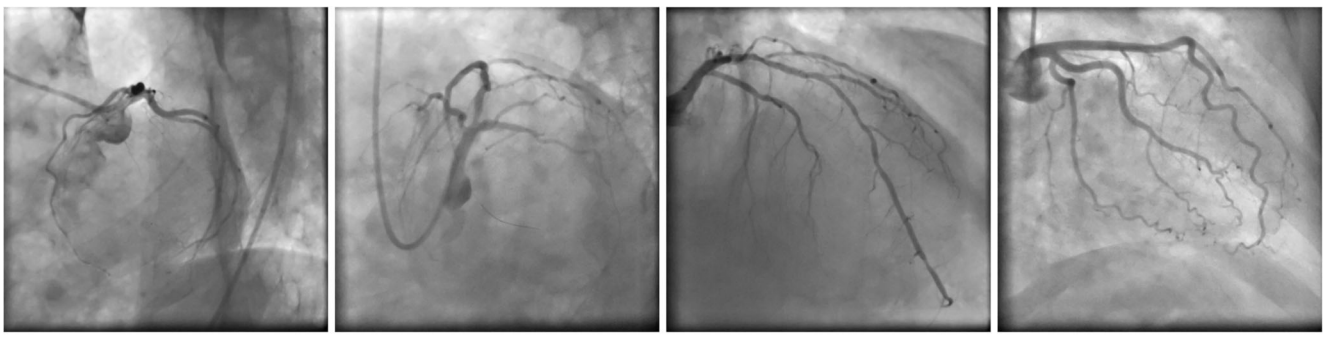
3.4 | Dataset organization

3.4.1 | Video selection

The presented dataset provides a total of 668 ICA videos. However, in some frames of these videos where CAD is present, the lesion is indiscernible, making it difficult to use them for diagnosis but to create a dataset, as close to reality as possible, these videos were also included. In order to

TABLE 2 Demographic and baseline characteristics of the patients. Data are given as percentage or as median (interquartile range).

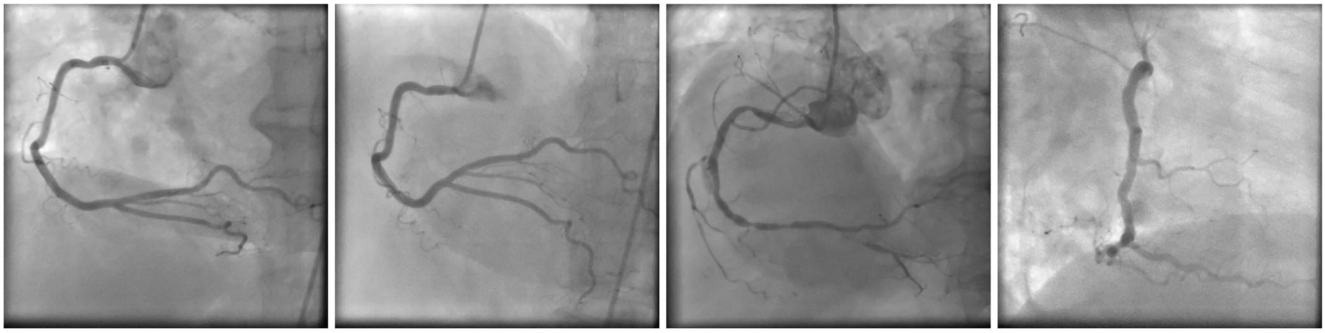
Age (years)	71.5 (58.25–78)
Sex (female–male)	47.6%–52.4%
Diabetes mellitus	40.5%
Dyslipidemia	40.5%
Smoker	45.2%
High blood pressure	61.9%
Kidney failure	14.3%
Heart failure	14.3%
Atrial fibrillation	4.8%
Left ventricular ejection fraction	
Normal (ejection fraction >55%)	68.2%
Mild dysfunction (ejection 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%



(a) LAO 45° CAU 35° (b) LAO 45° CAU 25° (c) RAO 30° CRA 20° (d) RAO 28° CAU 15°

FIGURE 1 Examples for left coronary artery (LCA).

obtain the best videos for CAD classification, a selection of 382 videos was performed. These videos have been chosen by the medical team, where CAD can be visually classified correctly. Thus, videos in which radiocontrast does not perfuse have not been selected for the classification task. The specifications of the CADICA dataset are reported in Table 4, where the number of patients, videos, images from selected videos, and labels from 'lesion' images are itemized. Please note that the number of labels is higher than the number of 'lesion' images since there can be more than one lesion in an image.

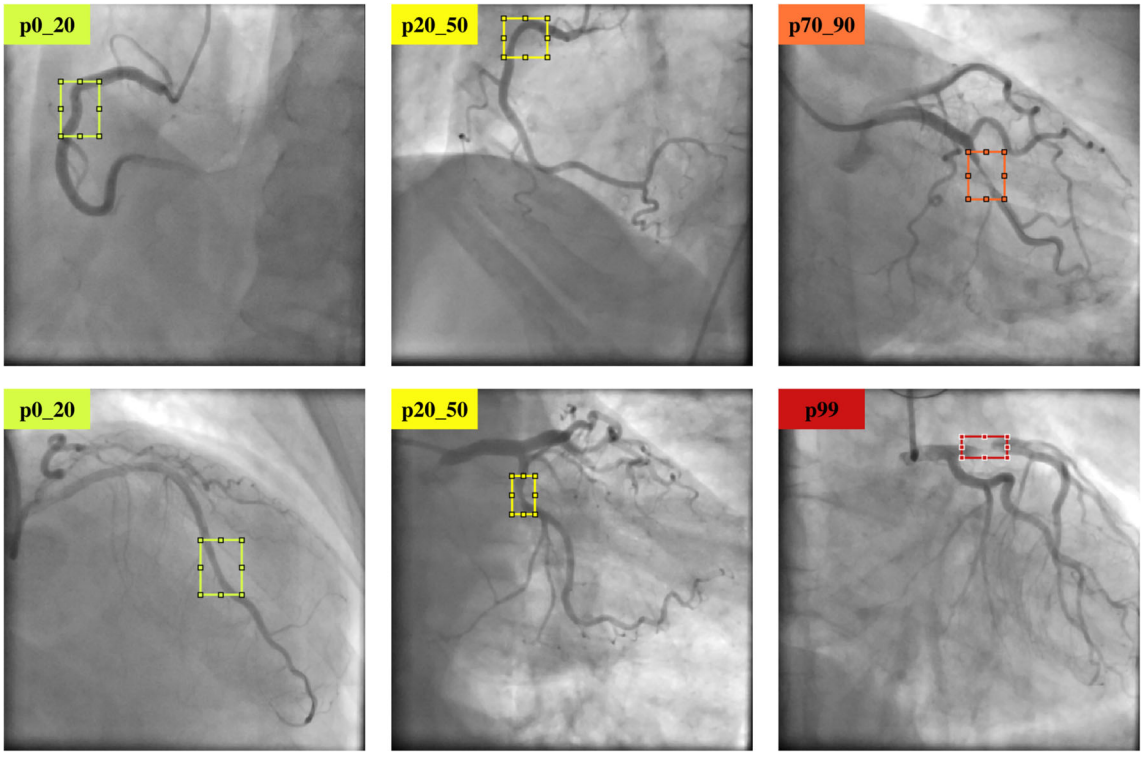


(a) LAO 35° CAU 5° (b) LAO 30° CRA 25° (c) LAO 25° CRA 20° (d) RAO 30°

FIGURE 2 Examples of projections for right coronary artery (RCA).

TABLE 3 Categories into which lesions have been divided.

Label	Lesion range
p0_20	< 20%
p20_50	[20%, 50%)
p50_70	[50%, 70%)
p70_90	[70%, 90%)
p90_98	[90%, 98%]
p99	99%
p100	100%



(a) Mild (b) Moderate (c) Severe

FIGURE 3 Samples of the 3 categories in which lesions are classified and delimited by a bounding box annotated.

TABLE 4 Specifications of CADICA dataset: Number of patients, videos, images from selected videos, and labels from 'lesion' images.

Patients		42	Total
Videos	Non-Selected	286	668
	Selected	382	
Images	Non-Lesion	2130	6126
	Lesion	3996	
Labels	p0_20	1944	6161
	p20_50	1128	
	p50_70	999	
	p70_90	893	
	p90_98	930	
	p99	63	
	p100	204	

3.4.2 | Metadata

The dataset also provides additional clinical data associated with each patient, such as if the patient suffers from diabetes, dyslipidemia, smoking, hypertension, another comorbidity (such as chronic obstructive pulmonary disease), renal insufficiency, heart failure, atrial fibrillation, or left ventricular ejection fraction. Other information such as age, gender, height, weight, and later event (such as non-cardiac death or heart attack) is also reported.

3.4.3 | Structure

The CADICA dataset becomes a directory that contains the *metadata.xlsx* file, which is the file where the clinical data is located, as well as two main folders that differentiate the videos selected by the medical team for each patient: *nonselectedVideos* and *selectedVideos*. Inside each folder, there are several sub-directories with the naming convention *pX* where *X* is the ID of each patient, and *vY*, where *Y* is the ID of the video of that patient.

The folder *pX* contains the following information:

- *vY*: several sub-directories with the videos selected for that patient.
- *lesionVideos.txt*: includes the IDs of the selected videos where appears at least one lesion which is labelled.
- *nonlesionVideos.txt*: contains the IDs of the selected videos where there are no visible lesions.

The folder *vY* contains the following information:

- *input*: a sub-directory containing a separate PNG file for each video frame.
- *pX_vY_selectedFrames.txt*: contains the IDs of the keyframes for the medical team, for all the selected videos.
- *groundtruth*: a sub-directory available only if there are lesions in that selected video.
- The folder *groundtruth* contains the following information:
- *pX_vY_000ZZ.txt*: contains the bounding boxes and their category in each row. There are such files as frames in *pX_vY_selectedFrames.txt*. Bounding boxes are specified in the format $[x,y,w,h]$, where (x,y) are the pixel coordinates of the top left corner, w is the width and h is the height of the bounding box.
- *pX_vY_groundTruthTable.mat*: contains a table with the ground truth information of that video.

4 | EXPERIMENTS

Given the provided dataset in this work, we did an exhaustive performance comparison to classify ICA images according to the presence of lesions, being a binary problem, where the method classifies between 'non-lesion' and 'lesion' (images with any label from Table 3) as validation of the usage of CADICA.

4.1 | Evaluation metrics

In order to measure the performance of a method that classifies coronary angiography images according to their coronary artery stenosis percentage, several well-known metrics have been proposed.

Let us consider the true positives or number of hits (TP), true negatives or correct rejections (TN), false negatives or misses (FN), and false positives or false alarms (FP). The selected metrics and their definitions are as follows:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad Fm = 2 \cdot \frac{PR \cdot RC}{PR + RC} \quad Bal = \frac{RC + SP}{2} \tag{1}$$

$$SP = \frac{TN}{FP + TN} \quad RC = \frac{TP}{TP + FN} \quad PR = \frac{TP}{TP + FP} \tag{2}$$

The most representative measures are the Accuracy (Acc), the F-measure (Fm, also known as F1 score), and the Balanced Accuracy (Bal), which provide a good overall evaluation of the performance of a given method. All these measures represent the percentage of hits of the system by providing values in the interval $[0,1]$, where higher is better.

Meanwhile, other measures are also implicitly considered such as the precision (PR), the recall (RC), and the specificity (SP). In order to analyse these metrics, FN must be considered against FP (lower is better), PR against RC (higher is better).

4.2 | Methods

4.2.1 | Convolutional neural networks

In this study, different Convolutional Neural Networks (CNNs) are used to compare their performance to classify ICA images from CADICA into two classes, 'lesion', which means that appears at least one lesion, and 'non-lesion'.

CNN is a type of deep learning model incorporating at least one convolutional layer, whose purpose is to extract the features from the input image, triggering under a specific condition (J. Wang et al., 2021). CNNs have become successful methods with great versatility of applications in several areas, including medical images, and to solve different problems, such as segmentation, localization, or classification. Also, CNNs are characterized by their transferability of knowledge by applying the transfer learning technique, which is based on employing classification models trained on large datasets, also named pre-trained networks, which are re-trained with a specific dataset to specialize them to the particular problem (Ovalle-Magallanes et al., 2020). In this study, no layer was frozen, so all weights were updated according to the input dataset information. Five known pre-trained CNN architectures are used in this study:

The Residual Networks (ResNets) family (He et al., 2016) introduces the residual connection to the model. These shortcut connections allow skipping some layers in the process. In particular, in this study, ResNet-18 and ResNet-50 networks are used, which are characterized by being composed of 18 and 50 layers deep, respectively.

MobileNet-V2 (Sandler et al., 2018) is a mobile neural network optimized to considerably reduce the number of parameters, compared with other architectures, which decreases the computational load. The MobileNet architecture is based on depthwise separable convolutions, which are a combination of two layers. The first is depthwise convolution, which applies a single filter to the input without extracting features, and the second is named pointwise convolution, which creates a linear combination output with new features (Sae-Lim et al., 2019).

NasNet-Mobile is the smallest version of NasNet models. NasNet models are CNNs based on Neural Architecture Search (NAS), which consist of basic building blocks, called cells, optimized by reinforcement learning method (Reddy et al., 2018).

DenseNet-201 is a deep network based on ensuring the maximum information flow between layers by dense blocks. Dense blocks are blocks of layers, where each layer is connected to all former layers, instead only to the previous one (Huang et al., 2017).

4.2.2 | Data pre-processing

In CADICA, there are 382 selected videos in total, of which two subgroups were done differentiating between views of the left (LCA) and right (RCA) coronary arteries.

The subgroup of LCA views was composed of 216 videos, a total of 3228 images, where 1003 images were labelled as 'non-lesion' and 2225 were labelled as 'lesion'. The subgroup of RCA views consisted of 118 videos, which is 2077 images in total, whose labels were distributed as 617 images labelled as 'non-lesion' and 1460 labelled as 'lesion'.

The input image of the pre-trained architectures selected is an RGB image of size 224 × 224 pixels. Thus, the first processing applied to all images was to resize them and use the colour pre-processing to ensure that images have the number of channels required, in this case, three channels.

To study the binary classification problem 'lesion'/'non-lesion', both sets were divided into training (80%) and test (20%) sets. This division was done by videos, which means that 80% of the 'non-lesion' and 'lesion' videos were used for training and the 20% remaining for testing. This way, frames of the same video of the train set are unavailable for the test set, because frames of a video are very similar between them.

Both sets have unbalanced distributions, which can cause the model to specialize in the majority class, in this case, 'lesion', and be relatively inefficient at classifying the minority class, in this case, 'non-lesion'. To solve this issue a data augmentation strategy had been applied to the training sets. This data augmentation was done by using different random basic operations of the original images, detailed as follows:

- Translations in the *x* and *y* axis of [−25, 25] pixels randomly, Figure 4b.
- Rotation using a random angle between [−25°, 25°], Figure 4c.
- Scaling of the images with a random scale factor in a range of [0.8, 1.7], Figure 4d.

Modifications were applied to the training sets of both classes, 'lesion' and 'non-lesion' images, and to both subsets, LCA and RCA, equalizing and increasing them. Figure 5 shows the original class distributions for LCA and RCA sets, and the final distributions obtained after data augmentation was implemented, obtaining 3640 and 2342 images of each class in the LCA and RCA sets, respectively.

4.2.3 | Experimental setup

Several parameters can be tuned for training convolutional neural networks. The main ones are the following:

- Validation frequency: number of iterations between evaluations of the training process.
- Maximum number of epochs: maximum times that the full dataset is passed to the model to update its weights.
- Optimizer or solver: algorithm applied to update the weights of the network to reduce the loss function.
- Initial learning rate: rate that is going to be used to start the learning procedure to update the network weights.
- Batch size: number of samples, images in this case, that are processed by the model in one iteration.

Due to the many options for possible combinations of training parameters, we started studying the performance behaviour, establishing some values to tune the training parameters. To evaluate the progress in the tuning of the parameters, we used the performance metrics reported in section 4.1, focusing on F-measure, Balanced Accuracy, and Accuracy, because together they provide a global view of the performance. The LCA set was employed for this process because it is the largest and most complex set. Finally, the parameters selected were used to evaluate the RCA set, too.

The first parameter set was validation frequency in 50 iterations to evaluate the training process and the maximum number of epochs set at 10 epochs because increasing it had an unsubstantial improvement compared to time-consuming. For the optimizer, we compared different algorithms: Adam (adaptive moment estimation), SGDM (stochastic gradient descent with momentum), and RMSProp (root mean square propagation). Besides, we proposed several rates for the initial learning rate: 0.01, 0.001, 0.0001, and 0.00001. For the batch size, two values were selected,

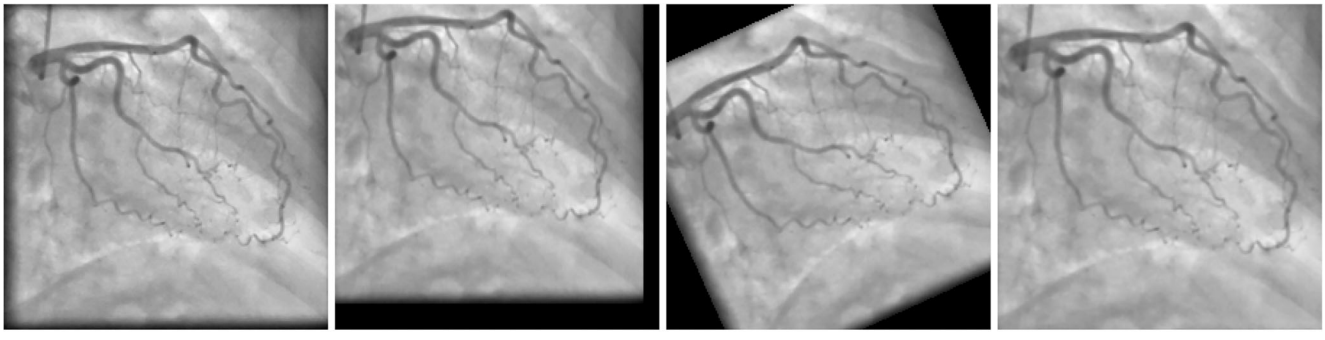


FIGURE 4 Examples of the modifications applied to the training sets to augment data.

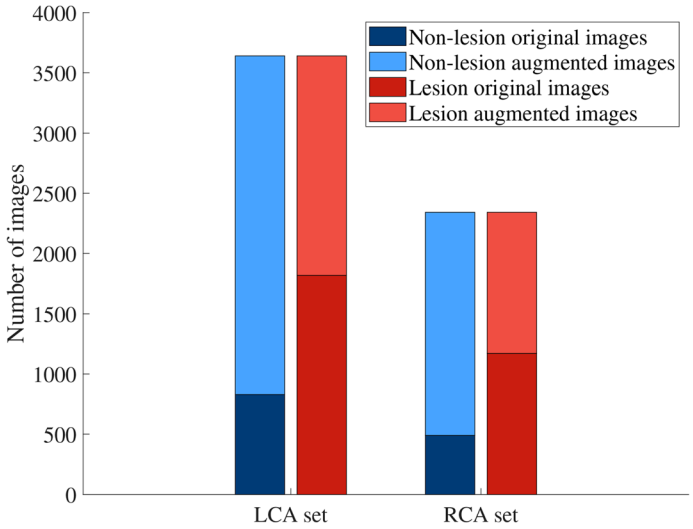


FIGURE 5 Class distribution for Left coronary Artery (LCA) and Right Coronary Artery (RCA) sets.

16 and 64. In total, there are 24 possible combinations. To compare them, we implemented the stratified K-fold cross-validation. This technique is adequate to evaluate the performance in a reliable way because the results are averages from different partitions of the input dataset, so results are independent of the partition employed to validate.

The proposed models were implemented in MATLAB R2022b on a computer system with an Intel Core i9-10, 900X processor, 128 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU card.

First, we divided the training set of the LCA subset into 5 boxes with samples of both classes, 80% for training the models and 20% as the validation set. For each model, the 24 combinations established were executed, except DenseNet-201, which was only trained with batch size 16 because of memory settings.

In order to set the optimal train setting, a grid search for the optimizer algorithm and learning rate was carried out. In Tables 5, 6, and 7, the outcomes obtained on test subsets with 5-fold stratified-cross-validation are reported for F-measure, Balanced Accuracy, and Accuracy, respectively. In these tables, the best results for each model and batch size are shown in bold.

By observing these tables, it can be remarked that the three selected performance measures follow equivalent behaviour. Batch sizes and optimizer outcomes close results independently of their values, but they may be significantly affected by the learning rate chosen, having up to a 10% difference when a low learning rate is chosen. In this case, the most suitable learning rates are 0.0001 and 0.00001.

Focusing on the F-score (Table 5), the lowest score was 0.65 while the best was 0.82, caused mainly by the alteration of the learning rate, as stated before. Balanced Accuracies (Table 6) also yield a similar behaviour, although the scores were significantly lower. For example, there were no differences in using batch sizes 16 or 64, yielding both a value of 0.65. However, in this case, the RMSProp optimizer worked better than SGDM for batch size 16 and vice-versa when using 64, which makes them dependent. A more random pattern was found in the Accuracy results (Table 7), where the Adam optimizer arose for MobileNet-V2 with 72% accuracy. However, in general terms, the other two optimizers were more stable, getting the best results when the 0.00001 learning rate was used. Additionally, it is a remarkable fact that all architectures attain very similar results, failing to clearly stand out one over the other.

Using the three performance metrics presented, the optimizer and the initial learning rate that more times get the highest results were selected. In Table 8, the configuration selected for each model is reported. For instance, the best values for MobileNet-V2 with a batch size of 16 were achieved using the RMSProp solver and an initial learning rate of 0.00001 in the three tables. However, some models present different best configurations depending on the measure taken into account, for example, ResNet-18 with a batch size of 16 obtained the best results using the RMSProp optimizer for F-measure and Balanced Accuracy, whereas the best optimizer was SGDM for Accuracy. In this case, the RMSProp optimizer is selected. Likewise, the chosen initial learning rate was 0.00001 for ResNet-18.

4.3 | Results

The experiments carried out to compare the performance of the different pre-trained models to evaluate the functionality of the implemented dataset based on the selected configurations (Table 8) are summarized here. These configurations were employed to implement a 10-fold stratified cross-validation, where 90% of images were used for training the models and 10% to validate the training process. The experimental results

TABLE 5 F-measure results obtained on the test set for LCA images using 5-fold stratified cross-validation, five convolutional network architectures, and different values for batch size, initial learning rate, and optimizer. The best performances are shown in bold.

		Batch size 16			Batch size 64		
Model	**Learning rate**	***Adam***	***SGDM***	***RMSProp***	***Adam***	***SGDM***	***RMSProp***
MobileNet-V2	0.01	0.758 ± 0.033	0.776 ± 0.028	0.755 ± 0.096	0.767 ± 0.036	0.806 ± 0.013	0.720 ± 0.052
	0.001	0.798 ± 0.035	0.800 ± 0.025	0.791 ± 0.034	0.812 ± 0.024	0.786 ± 0.018	0.807 ± 0.024
	0.0001	0.809 ± 0.030	0.794 ± 0.021	0.818 ± 0.015	0.806 ± 0.016	0.800 ± 0.010	0.808 ± 0.015
	0.00001	0.810 ± 0.014	0.785 ± 0.010	**0.820 ± 0.015**	**0.810 ± 0.009**	0.766 ± 0.011	0.789 ± 0.012
ResNet-18	0.01	0.737 ± 0.054	0.739 ± 0.038	0.730 ± 0.094	0.747 ± 0.047	0.789 ± 0.011	0.731 ± 0.065
	0.001	0.782 ± 0.020	0.798 ± 0.020	0.765 ± 0.025	0.784 ± 0.013	**0.794 ± 0.016**	0.762 ± 0.017
	0.0001	0.784 ± 0.022	0.780 ± 0.022	**0.800 ± 0.021**	0.783 ± 0.029	0.788 ± 0.026	0.775 ± 0.014
	0.00001	0.783 ± 0.017	0.800 ± 0.029	0.789 ± 0.019	0.772 ± 0.016	0.770 ± 0.016	0.782 ± 0.012
ResNet-50	0.01	0.655 ± 0.057	0.777 ± 0.027	0.380 ± 0.418	0.755 ± 0.006	0.793 ± 0.013	0.713 ± 0.097
	0.001	0.773 ± 0.020	0.774 ± 0.012	0.775 ± 0.031	0.772 ± 0.042	0.779 ± 0.021	0.746 ± 0.055
	0.0001	0.785 ± 0.023	0.784 ± 0.025	**0.823 ± 0.017**	0.806 ± 0.024	0.802 ± 0.016	**0.818 ± 0.023**
	0.00001	0.797 ± 0.037	0.796 ± 0.026	0.808 ± 0.025	0.782 ± 0.040	0.783 ± 0.034	0.777 ± 0.024
NasNet-Mobile	0.01	0.743 ± 0.018	0.785 ± 0.017	0.712 ± 0.105	0.723 ± 0.037	**0.798 ± 0.011**	0.717 ± 0.047
	0.001	0.807 ± 0.026	**0.811 ± 0.028**	0.790 ± 0.032	0.790 ± 0.027	0.785 ± 0.019	0.753 ± 0.050
	0.0001	0.794 ± 0.025	0.770 ± 0.036	0.803 ± 0.017	0.779 ± 0.015	0.750 ± 0.021	0.783 ± 0.016
	0.00001	0.778 ± 0.008	0.714 ± 0.011	0.770 ± 0.016	0.773 ± 0.006	0.659 ± 0.049	0.765 ± 0.010
DenseNet-201	0.01	0.694 ± 0.020	0.736 ± 0.032	0.666 ± 0.103			
	0.001	0.761 ± 0.032	0.799 ± 0.019	0.758 ± 0.025			
	0.0001	0.797 ± 0.036	0.774 ± 0.012	0.801 ± 0.034			
	0.00001	0.806 ± 0.025	**0.826 ± 0.017**	0.794 ± 0.014			

using different neural models, batch size, and image subgroups (LCA and RCA) are shown in Table 9, where Balanced accuracy, F-measure, and Accuracy obtained in the test set are reported. In Table 9 are shown in bold the highest values obtained for each measure in both subsets. Note that the best results for Balanced Accuracy and Accuracy were obtained with the MobileNet-V2 model for the LCA subset (0.673 and 0.732, respectively). However, for the RCA set, the NasNet-Mobile model achieved the best Balanced Accuracy and Accuracy (0.658 and 0.744). According to the F-measure, the best results were obtained by the ResNet-50 in both subsets (0.814 and 0.830, respectively).

To study the best suitable model for these inputs, a ranking was implemented to evaluate the results attained considering the three measures. The rankings obtained are reported in Figure 6, scoring the methods by set and batch size. The scores were calculated by sorting the obtained values of a measure in ascending order since a higher value is better according to the considered measures, meaning that the highest value will be in the last position. The position indicates the obtained scores. There are five and four methods for batch sizes 16 and 64, respectively. Therefore, the maximum possible score is 15 points and 12, respectively, indicating that the method attained the highest values in the three measures.

Focusing on the LCA set, Figure 6a,b, the best outcomes are produced by MobileNet-V2, obtaining 15 points and 11 points with a batch size of 16 and 64, respectively. Although comparing the results obtained in Table 9, the best result is produced with a batch size of 64, which reached the maximum balanced accuracy, 0.673, and the second best F-measure and Accuracy, 0.810 and 0.732, respectively.

However, in the case of the RCA set, Figure 6c,d, ResNet-18 and NasNet-Mobile are the architectures that obtain higher performance, being NasNet-Mobile with a batch size of 64 that returns the best Balanced Accuracy and Accuracy values, 0.658 and 0.744, respectively; and the second best F-measure, 0.826, according to the attained results reported in Table 9.

5 | DISCUSSION

In this section, some important aspects to be considered of our proposal are discussed:

- In this work, all degrees of lesions are considered, while other studies only include severe lesions (Au et al., 2018; Shu & Wu, 2021) or exclude those images in which more than one lesion appears (Au et al., 2018). Nevertheless, our results indicate a fair-to-high performance, which means that the dataset is functional, but also that this classification task is complex and challenging.

TABLE 6 Balanced Accuracy results obtained on the test set for LCA images using 5-fold stratified cross-validation, five convolutional network architectures, and different values for batch size, initial learning rate, and optimizer. The best performances are shown in bold.

		Batch size 16			Batch size 64		
Model	**Learning rate**	***Adam***	***SGDM***	***RMSProp***	***Adam***	***SGDM***	***RMSProp***
MobileNet-V2	0.01	0.586 ± 0.059	0.597 ± 0.052	0.514 ± 0.029	0.608 ± 0.029	0.610 ± 0.024	0.578 ± 0.041
	0.001	0.656 ± 0.040	0.644 ± 0.052	0.650 ± 0.044	0.660 ± 0.073	0.630 ± 0.022	0.648 ± 0.052
	0.0001	0.667 ± 0.034	0.643 ± 0.035	0.671 ± 0.011	0.665 ± 0.014	0.638 ± 0.030	0.657 ± 0.018
	0.00001	0.682 ± 0.020	0.618 ± 0.013	**0.688 ± 0.024**	**0.678 ± 0.027**	0.590 ± 0.020	0.662 ± 0.012
ResNet-18	0.01	0.551 ± 0.071	0.559 ± 0.028	0.492 ± 0.034	0.594 ± 0.053	0.625 ± 0.028	0.582 ± 0.120
	0.001	0.609 ± 0.037	0.618 ± 0.023	0.605 ± 0.017	0.636 ± 0.050	**0.656 ± 0.040**	0.583 ± 0.033
	0.0001	0.592 ± 0.014	0.633 ± 0.027	0.623 ± 0.017	0.610 ± 0.037	0.632 ± 0.037	0.583 ± 0.027
	0.00001	0.640 ± 0.009	0.633 ± 0.061	**0.649 ± 0.024**	0.615 ± 0.025	0.618 ± 0.043	0.638 ± 0.014
ResNet-50	0.01	0.558 ± 0.026	0.592 ± 0.029	0.500 ± 0.000	0.578 ± 0.050	0.632 ± 0.027	0.547 ± 0.084
	0.001	0.600 ± 0.022	0.613 ± 0.022	0.564 ± 0.055	0.626 ± 0.014	0.642 ± 0.034	0.577 ± 0.037
	0.0001	0.636 ± 0.031	0.653 ± 0.033	0.645 ± 0.041	0.638 ± 0.057	**0.651 ± 0.022**	0.648 ± 0.030
	0.00001	0.659 ± 0.060	0.623 ± 0.039	**0.679 ± 0.033**	0.650 ± 0.047	0.600 ± 0.066	0.642 ± 0.023
NasNet-Mobile	0.01	0.562 ± 0.020	0.620 ± 0.048	0.531 ± 0.039	0.568 ± 0.043	0.639 ± 0.018	0.597 ± 0.024
	0.001	0.626 ± 0.045	0.686 ± 0.034	0.649 ± 0.053	0.610 ± 0.047	**0.647 ± 0.034**	0.597 ± 0.049
	0.0001	0.625 ± 0.029	0.644 ± 0.048	**0.649 ± 0.025**	0.637 ± 0.033	0.618 ± 0.035	0.629 ± 0.017
	0.00001	0.621 ± 0.010	0.583 ± 0.025	0.629 ± 0.010	0.620 ± 0.013	0.529 ± 0.022	0.607 ± 0.017
DenseNet-201	0.01	0.537 ± 0.084	0.597 ± 0.067	0.512 ± 0.056			
	0.001	0.576 ± 0.041	0.666 ± 0.020	0.611 ± 0.033			
	0.0001	0.623 ± 0.060	0.627 ± 0.022	0.669 ± 0.059			
	0.00001	0.642 ± 0.037	**0.672 ± 0.025**	0.638 ± 0.022			

- This study is only limited to validating the performance on binary classification of the presence or absence of CAD lesions. With the continuous increase of the dataset in the future we expect to test multi-class classification.
- Despite the fact that a Balanced Accuracy between 0.65 and 0.67 was obtained, it represents a good performance since this measure depends on specificity and recall, which quantifies an average of how correctly both classes are classified, and the 'non-lesion' class is worse classified than the 'lesion' class. This is due to the fact that there are fewer 'non-lesion' images than 'lesion' images, although data augmentation was applied to equalize both classes. Since the 'lesion' class includes all degrees of lesion, non-severe lesions can be incorrectly classified as 'non-lesion'.
- Although the DenseNet model usually has a good performance (Chauhan et al., 2021; Guendel et al., 2018; Thurnhofer-Hemsi & Domínguez, 2021), we could not evaluate it with a batch size of 64 due to memory requirements. Nevertheless, training this model with a batch size of 16 yielded good results in terms of F-measure. Therefore, better results will be expected by training with higher batch sizes.
- The original resolution of 512×512 pixels was downsampled to 224×224 pixels because of computational resources; preliminary tests showed an insignificant performance increase compared to the computational cost that the training would require.
- Finally, for LCA images, the selected configurations outcomes similar results independently of the batch size chosen. MobileNet-V2 achieves the highest results for Accuracy and Balanced Accuracy — 0.673 and 0.732, respectively — followed too closely by DenseNet-201 and ResNet-50, which attains the highest F-measure — 0.814. However, the equivalent configurations for RCA images achieve lower results but achieve similar performance with Resnet-50 — 0.830 for F-measure- and NasNet-Mobile — 0.658 and 0.744 for Balanced Accuracy and Accuracy, respectively. Therefore, different models get too similar results in LCA and RCA images being remarkable MobileNet-V2 and NasNet-Mobile in each case.

6 | CONCLUSIONS

The coronary dataset published in this work aims to provide the research community with a conscientious and exhaustive scientific resource. It can serve as a benchmark for both new and existing algorithm implementations, and for medical staff to train their abilities on angiographic assessment of CAD severity. Considering the researchers' feedback, a set of utilities and the already extensive dataset will be regularly revised and expanded.

TABLE 7 Accuracy results obtained on the test set for LCA images using 5-fold stratified cross-validation, five convolutional network architectures, and different values for batch size, initial learning rate, and optimizer. The best performances are shown in bold.

		Batch size 16			Batch size 64		
Model	**Learning rate**	***Adam***	***SGDM***	***RMSProp***	***Adam***	***SGDM***	***RMSProp***
MobileNet-V2	0.01	0.659 ± 0.037	0.678 ± 0.041	0.642 ± 0.074	0.674 ± 0.035	0.710 ± 0.017	0.626 ± 0.045
	0.001	0.716 ± 0.040	0.713 ± 0.039	0.708 ± 0.044	0.730 ± 0.041	0.696 ± 0.018	0.722 ± 0.033
	0.0001	0.729 ± 0.036	0.708 ± 0.029	0.739 ± 0.017	0.726 ± 0.019	0.712 ± 0.018	0.725 ± 0.017
	0.00001	0.733 ± 0.018	0.692 ± 0.012	**0.745 ± 0.021**	**0.732 ± 0.013**	0.667 ± 0.014	0.708 ± 0.012
ResNet-18	0.01	0.629 ± 0.069	0.634 ± 0.037	0.611 ± 0.089	0.653 ± 0.055	0.698 ± 0.014	0.634 ± 0.091
	0.001	0.687 ± 0.021	0.705 ± 0.025	0.671 ± 0.022	0.697 ± 0.023	**0.712 ± 0.025**	0.661 ± 0.017
	0.0001	0.684 ± 0.021	0.692 ± 0.027	0.708 ± 0.023	0.688 ± 0.036	0.699 ± 0.033	0.673 ± 0.019
	0.00001	0.697 ± 0.018	**0.710 ± 0.043**	0.705 ± 0.022	0.679 ± 0.021	0.679 ± 0.026	0.695 ± 0.015
ResNet-50	0.01	0.572 ± 0.030	0.678 ± 0.030	0.473 ± 0.209	0.653 ± 0.017	0.703 ± 0.019	0.617 ± 0.064
	0.001	0.676 ± 0.023	0.681 ± 0.017	0.650 ± 0.027	0.685 ± 0.043	0.694 ± 0.029	0.646 ± 0.570
	0.0001	0.698 ± 0.022	0.702 ± 0.030	**0.737 ± 0.027**	0.717 ± 0.038	0.718 ± 0.019	**0.732 ± 0.029**
	0.00001	0.715 ± 0.050	0.705 ± 0.031	0.731 ± 0.032	0.699 ± 0.049	0.685 ± 0.050	0.692 ± 0.028
NasNet-Mobile	0.01	0.638 ± 0.016	0.692 ± 0.028	0.613 ± 0.074	0.623 ± 0.042	**0.711 ± 0.015**	0.630 ± 0.037
	0.001	0.715 ± 0.037	**0.736 ± 0.035**	0.706 ± 0.040	0.695 ± 0.039	0.701 ± 0.025	0.659 ± 0.056
	0.0001	0.704 ± 0.028	0.687 ± 0.044	0.718 ± 0.021	0.693 ± 0.021	0.661 ± 0.028	0.694 ± 0.018
	0.00001	0.687 ± 0.008	0.621 ± 0.017	0.682 ± 0.016	0.682 ± 0.007	0.562 ± 0.038	0.670 ± 0.014
DenseNet-201	0.01	0.589 ± 0.041	0.643 ± 0.047	0.569 ± 0.081			
	0.001	0.658 ± 0.039	0.719 ± 0.020	0.666 ± 0.030			
	0.0001	0.705 ± 0.049	0.685 ± 0.015	0.723 ± 0.032			
	0.00001	0.719 ± 0.034	**0.746 ± 0.024**	0.706 ± 0.019			

TABLE 8 Selected configurations considering F-measure, Balanced Accuracy and Accuracy obtained with 5-fold stratified cross-validation.

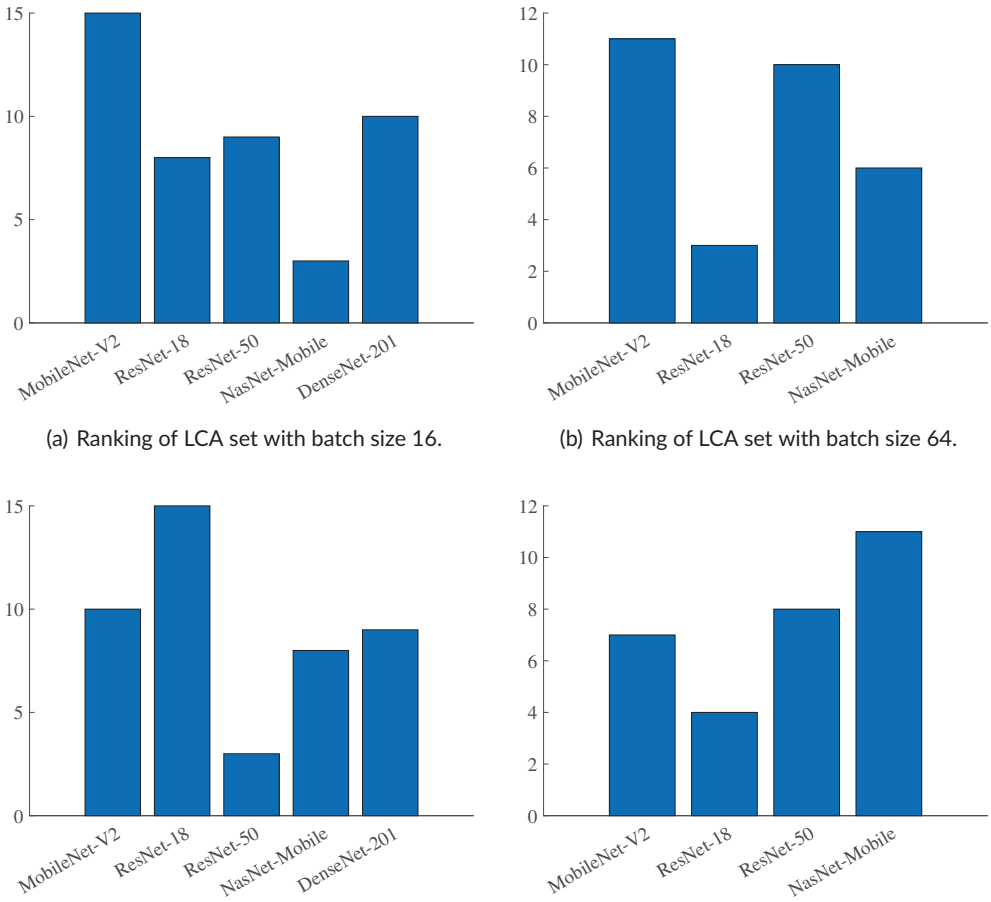
Model	Batch size	Optimizer	Learning rate
MobileNet-V2	16	RMSprop	0.00001
	64	Adam	0.00001
ResNet-18	16	RMSprop	0.00001
	64	SGDM	0.001
ResNet-50	16	RMSprop	0.00001
	64	RMSprop	0.0001
NasNet-Mobile	16	SGDM	0.001
	64	SGDM	0.01
DenseNet-201	16	SGDM	0.00001

Experiments were designed to try the functionality of the dataset, which was divided into LCA and RCA images. Five classification architectures were trained and tested using augmented data to get an overview of the performance classification of the 'lesion' and 'non-lesion' images. Experiments showed that the most suitable models to solve this problem were MobileNet-V2 for LCA images and NasNet-Mobile for RCA images, getting fair-to-high outcomes, around 80% F-measure and Accuracy, and 65% Balanced Accuracy. These results were obtained considering a wide range of lesion levels and support the idea that this classification task is complex, setting up a challenge for physicians and computer-aided diagnosis systems. Therefore, the provided dataset gives the scientific community a starting point to improve CAD detection.

One of the limitations of this work is that, in spite of the wide diversity of categories, the used dataset only contains information about 42 patients from one hospital. Another relevant fact is the image downsampling performed and batch size limitation because of computational resources; further work would focus on higher-resolution images. In future works, we would like to enhance the ICA dataset with more images belonging to patients from different hospitals. These new images would be used to validate proposed models. Given the complexity of the

TABLE 9 Balanced Accuracy, F-measure, and Accuracy results obtained on the test set for LCA and RCA images using 10-fold stratified cross-validation, five convolutional network architectures, and different batch size. The highest values by columns are shown in bold.

		LCA			RCA		
Model	**Batch size**	**Balanced accuracy**	**F-measure**	**Accuracy**	**Balanced accuracy**	**F-measure**	**Accuracy**
MobileNet-V2	16	0.668 ± 0.014	0.805 ± 0.013	0.725 ± 0.015	0.648 ± 0.023	0.811 ± 0.018	0.726 ± 0.023
	64	**0.673 ± 0.026**	0.810 ± 0.020	**0.732 ± 0.025**	0.641 ± 0.041	0.806 ± 0.026	0.719 ± 0.034
ResNet-18	16	0.642 ± 0.026	0.796 ± 0.024	0.710 ± 0.029	0.658 ± 0.043	0.825 ± 0.023	0.743 ± 0.031
	64	0.632 ± 0.013	0.779 ± 0.009	0.691 ± 0.010	0.624 ± 0.045	0.792 ± 0.031	0.702 ± 0.039
ResNet-50	16	0.664 ± 0.035	0.793 ± 0.029	0.713 ± 0.036	0.618 ± 0.022	0.799 ± 0.009	0.705 ± 0.013
	64	0.645 ± 0.044	**0.814 ± 0.021**	0.728 ± 0.029	0.620 ± 0.029	**0.830 ± 0.025**	0.738 ± 0.031
NasNet-Mobile	16	0.634 ± 0.047	0.789 ± 0.023	0.701 ± 0.033	0.651 ± 0.054	0.804 ± 0.058	0.723 ± 0.067
	64	0.637 ± 0.043	0.783 ± 0.020	0.696 ± 0.029	**0.658 ± 0.034**	0.826 ± 0.033	**0.744 ± 0.039**
DenseNet-201	16	0.641 ± 0.024	0.802 ± 0.022	0.715 ± 0.027	0.633 ± 0.037	0.812 ± 0.029	0.723 ± 0.037



(a) Ranking of LCA set with batch size 16.

(b) Ranking of LCA set with batch size 64.

(c) Ranking of RCA set with batch size 16.

(d) Ranking of RCA set with batch size 64.

FIGURE 6 Ranking of methods considering F-measure, Balanced Accuracy and Accuracy obtained for the LCA and RCA subsets using batch sizes 16 and 64.

classification problem posed, other architectures could be tested in order to keep evaluating the performance of different models. Also, future works will focus on multiclass classification, trying to identify, first, between mild, moderate, and severe lesions, but also try to implement a fine-grain classification where each possible range, seven in total, could be identified. Besides, a system to solve the detection task, which involves the classification and location of lesions, could be explored.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGEMENTS

This work is partially supported by the Autonomous Government of Andalusia (Spain) under project UMA20-FEDERJA-108, project name Detection, characterization and prognosis value of the non-obstructive coronary disease with deep learning, and also by the Ministry of Science and Innovation of Spain, grant number PID2022-136764OA-I00, project name Automated Detection of Non Lesional Focal Epilepsy by Probabilistic Diffusion Deep Neural Models. It includes funds from the European Regional Development Fund (ERDF). It is also partially supported by the University of Malaga (Spain) under grants B1-2021_20, project name Detection of coronary stenosis using deep learning applied to coronary angiography; B4-2023_13, project name Intelligent Clinical Decision Support System for Non-Obstructive Coronary Artery Disease in Coronarographies; B1-2022_14, project name Detección de trayectorias anómalas de vehículos en cámaras de tráfico, and B1-2023_18, project name Sistema de videovigilancia basado en cámaras itinerantes robotizadas; and, by the Fundación Unicaja under project PUNI-003_2023, project name Intelligent System to Help the Clinical Diagnosis of Non-Obstructive Coronary Artery Disease in Coronary Angiography. The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga. They also gratefully acknowledge the support of NVIDIA Corporation with the donation of a RTX A6000 GPU with 48 GB. The authors also thankfully acknowledge the grant of the Universidad de Málaga and the Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND. Funding for open access charge: Universidad de Málaga / CBUA.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Mendeley Data at https://data.mendeley.com/datasets/p9bpx9ctcv/2, reference number 10.17632/p9bpx9ctcv.2.

ORCID

Ariadna Jiménez-Partinen https://orcid.org/0000-0002-4668-2946
Miguel A. Molina-Cabello https://orcid.org/0000-0002-8929-6017
Karl Thurnhofer-Hemsi https://orcid.org/0000-0001-6519-1213
Esteban J. Palomo https://orcid.org/0000-0002-8547-9393
Jorge Rodríguez-Capitán https://orcid.org/0000-0001-6650-6264
Ana I. Molina-Ramos https://orcid.org/0000-0002-0735-9417
Manuel Jiménez-Navarro https://orcid.org/0000-0002-6759-0514

REFERENCES

Au, B., Shaham, U., Dhruva, S., Bouras, G., Cristea, E., Al, M. D., Coppi, A., Warner, F., Li, S. X., & Krumholz, H. (2018). Automated characterization of stenosis in invasive coronary angiography images with convolutional neural networks. *arXiv preprint arXiv:1807.10597*.

Chauhan, T., Palivela, H., & Tiwari, S. (2021). Optimization and fine-tuning of densenet model for classification of covid-19 cases in medical imaging. *International Journal of Information Management Data Insights*, *1*(2), 100020.

Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., & Halpern, A. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 168–172). IEEE.

Collet, J. P., Thiele, H., Barbato, E., Barthélémy, O., Bauersachs, J., Bhatt, D. L., Dendale, P., Dorobantu, M., Edvardsen, T., Folliguet, T., & Gale, C. P. (2021). 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *European Heart Journal*, *42*(14), 1289–1367. https://doi.org/10.1093/eurheartj/ehaa575

Cong, C., Kato, Y., Vasconcellos, H. D., Lima, J., & Venkatesh, B. (2019). Automated stenosis detection and classification in x-ray angiography using deep neural network. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1301–1308). IEEE.

Curzen, N., Rana, O., Nicholas, Z., Golledge, P., Zaman, A., Oldroyd, K., Hanratty, C., Banning, A., Wheatcroft, S., Hobson, A., & Chitkara, K. (2014). Does routine pressure wire assessment influence management strategy at coronary angiography for diagnosis of chest pain?: The RIPCORD study. *Circulation. Cardiovascular Interventions*, *7*(2), 248–255. https://doi.org/10.1161/CIRCINTERVENTIONS.113.000978

Danilov, V. V., Klyshnikov, K. Y., Gerget, O. M., Kutikhin, A. G., Ganyukov, V. I., Frangi, A. F., & Ovcharenko, E. A. (2021). Real-time coronary artery stenosis detection based on modern neural networks. *Scientific Reports*, *11*(1), 1–13.

Eschen, C. K., Banasik, K., Christensen, A. H., Chmura, P. J., Pedersen, F., Køber, L., Engstrøm, T., Dahl, A. B., Brunak, S., & Bundgaard, H. (2022). Classification of left and right coronary arteries in coronary angiographies using deep learning. *Electronics*, *11*(13), 2087.

Finck, T., Hardenberg, J., Will, A., Hendrich, E., Haller, B., Martinoff, S., Hausleiter, J., & Hadamitzky, M. (2019). 10-year follow-up after coronary computed tomography angiography in patients with suspected coronary artery disease. *JACC: Cardiovascular Imaging*, *12*(7 Part 2), 1330–1338.

Gabara, L., Hinton, J., Gunn, J., Morris, P. D., & Curzen, N. (2020). Coronary physiology derived from invasive angiography: Will it be a game changer? *Interventional Cardiology*, *15*, e06. https://doi.org/10.15420/icr.2019.25

Guendel, S., Grbic, S., Georgescu, B., Liu, S., Maier, A., & Comaniciu, D. (2018). Learning to recognize abnormalities in chest x-rays with location-aware dense networks. In *Iberoamerican Congress on Pattern Recognition* (pp. 757–765). Springer International Publishing.

Halon, D. A. (2018). Can angiography predict physiology? *International Journal of Cardiology*, *270*, 74–75. https://doi.org/10.1016/j.ijcard.2018.07.029

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). IEEE.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708). IEEE.

Iyer, K., Najarian, C. P., Fattah, A. A., Arthurs, C. J., Soroushmehr, S. R., Subban, V., Sankardas, M. A., Nadakuditi, R. R., Nallamothu, B. K., & Figueroa, C. A. (2021). AngioNet: A convolutional neural network for vessel segmentation in x-ray angiography. *Scientific Reports*, *11*(1), 18066.

Kang, S. H., Park, G. M., Lee, S. W., Yun, S. C., Kim, Y. H., Cho, Y. R., Park, H. W., Suh, J., Yang, D. H., Kang, J. W., & Lim, T. H. (2016). Long-term prognostic value of coronary ct angiography in asymptomatic type 2 diabetes mellitus. *JACC: Cardiovascular Imaging*, *9*(11), 1292–1300.

Kemp, H. G., Kronmal, R. A., Vlietstra, R. E., & Frye, R. L. (1986). Seven year survival of patients with normal or near normal coronary arteriograms: A CASS registry study. *Journal of the American College of Cardiology*, *7*(3), 479–483. https://doi.org/10.1016/s0735-1097(86)80456-9

Kim, Y. I., Roh, J. H., Kweon, J., Kwon, H., Chae, J., Park, K., Lee, J. H., Jeong, J. O., Kang, D. Y., Lee, P. H., & Ahn, J. M. (2024). Artificial intelligence-based quantitative coronary angiography of major vessels using deep-learning. *International Journal of Cardiology*, *405*, 131945.

Knuuti, J., Wijns, W., Saraste, A., Capodanno, D., Barbato, E., Funck-Brentano, C., Prescott, E., Storey, R. F., Deaton, C., Cuisset, T., & Agewall, S. (2019). 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes: The Task Force for the diagnosis and management of chronic coronary syndromes of the European Society of Cardiology (ESC). *European Heart Journal*, *41*(3), 407–477. https://doi.org/10.1093/eurheartj/ehz425

Litjens, G., Ciompi, F., Wolterink, J. M., de Vos, B. D., Leiner, T., Teuwen, J., & Išgum, I. (2019). State-of-the-art deep learning in cardiovascular image analysis. *JACC: Cardiovascular Imaging*, *12*(8 Part 1), 1549–1565.

Matuszewski, D. J., & Sintorn, I.-M. (2021). Tem virus images: Benchmark dataset and deep learning classification. *Computer Methods and Programs in Biomedicine*, *209*, 106318.

Menezes, M. N., Lourenço-Silva, J., Silva, B., Rodrigues, T., Francisco, A. R., Ferreira, P. C., Oliveira, A. L., & Pinto, F. J. (2022). Development of deep learning segmentation models for coronary x-ray angiography: Quality assessment by a new global segmentation score and comparison with human performance. *Revista Portuguesa de Cardiologia*, *41*(12), 1011–1021.

Moon, J. H., Cha, W. C., Chung, M. J., Lee, K. S., Cho, B. H., & Choi, J. H. (2021). Automatic stenosis recognition from coronary angiography using convolutional neural networks. *Computer Methods and Programs in Biomedicine*, *198*, 105819.

Murphy, S. L., Kochanek, K. D., Xu, J., & Arias, E. (2021). Mortality in the United States, 2020. *NCHS Data Brief* (427).

Nasr-Esfahani, E., Karimi, N., Jafari, M. H., Soroushmehr, S. M. R., Samavi, S., Nallamothu, B., & Najarian, K. (2018). Segmentation of vessels in angiograms using convolutional neural networks. *Biomedical Signal Processing and Control*, *40*, 240–251.

Ovalle-Magallanes, E., Alvarado-Carrillo, D. E., Avina-Cervantes, J. G., Cruz-Aceves, I., Ruiz-Pinales, J., & Correa, R. (2022). Deep learning-based coronary stenosis detection in x-ray angiography images: Overview and future trends. Artificial Intelligence and Machine Learning for Healthcare: *Vol. 2: Emerging Methodologies and Trends*, 197–223.

Ovalle-Magallanes, E., Avina-Cervantes, J. G., Cruz-Aceves, I., & Ruiz-Pinales, J. (2020). Transfer learning for stenosis detection in x-ray coronary angiography. *Mathematics*, *8*(9), 1510.

Ovalle-Magallanes, E., Avina-Cervantes, J. G., Cruz-Aceves, I., & Ruiz-Pinales, J. (2022). Improving convolutional neural network learning based on a hierarchical bezier generative model for stenosis detection in x-ray images. *Computer Methods and Programs in Biomedicine*, *219*, 106767.

Radico, F., Zimarino, M., Fulgenzi, F., Ricci, F., Di Nicola, M., Jespersen, L., Chang, S. M., Humphries, K. H., Marzilli, M., & De Caterina, R. (2018). Determinants of long-term clinical outcomes in patients with angina but without obstructive coronary artery disease: A systematic review and meta-analysis. *European Heart Journal*, *39*(23), 2135–2146. https://doi.org/10.1093/eurheartj/ehy185

Reddy, N., Rattani, A., & Derakhshani, R. (2018). Comparison of deep learning models for biometric-based mobile user authentication. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)* (pp. 1–6). IEEE.

Rodríguez-Capitán, J., Sánchez-Pérez, A., Ballesteros-Pradas, S., Millán-Gómez, M., Cardenal-Piris, R., Oneto-Fernández, M., Gutiérrez-Alonso, L., Rivera-López, R., Guisado-Rasco, A., Cano-García, M., & Gutiérrez-Bedmar, M. (2021). Prognostic implication of non-obstructive coronary lesions: A new classification in different settings. *Journal of Clinical Medicine*, *10*(9), 1863. https://doi.org/10.3390/jcm10091863

Sae-Lim, W., Wettayaprasit, W., & Aiyarak, P. (2019). Convolutional neural networks using mobilenet for skin lesion classification. In *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 242–247). IEEE.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Ruecognition* (pp. 4510–4520). IEEE.

Saraste, A., Barbato, E., Capodanno, D., Edvardsen, T., Prescott, E., Achenbach, S., Bax, J. J., Wijns, W., & Knuuti, J. (2019). Imaging in ESC clinical guidelines: Chronic coronary syndromes. *European Heart Journal Cardiovascular Imaging*, *20*(11), 1187–1197. https://doi.org/10.1093/ehjci/jez219

Shu, Y., & Wu, X. (2021). Deep learning based coronary angiography in diagnosis of myocardial ischemia. *Scientific Programming*, *2021*(1), 8491976.

Song, Y., Ren, S., Lu, Y., Fu, X., & Wong, K. K. (2022). Deep learning-based automatic segmentation of images in cardiac radiography: A promising challenge. *Computer Methods and Programs in Biomedicine*, *220*, 106821.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, *63*(7), 1455–1462.

Thurnhofer-Hemsi, K., & Domínguez, E. (2021). A convolutional neural network framework for accurate skin cancer detection. *Neural Processing Letters*, *53*(5), 3073–3093.

Tonino, P. A. L., De Bruyne, B., Pijls, N. H. J., Siebert, U., Ikeno, F., vant Veer, M., Klauss, V., Manoharan, G., Engstrøm, T., Oldroyd, K. G., & Ver Lee, P. N. (2009). Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *The New England Journal of Medicine*, *360*(3), 213–224. https://doi.org/10.1056/NEJMoa0807611

Topol, E. J., & Nissen, S. E. (1995). Our preoccupation with coronary luminology. The dissociation between clinical and angiographic findings in ischemic heart disease. *Circulation*, *92*(8), 2333–2342. https://doi.org/10.1161/01.cir.92.8.2333

Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, *5*(1), 1–9.

Voudris, K. V., & Kavinsky, C. J. (2019). Advances in Management of Stable Coronary Artery Disease: The role of revascularization? *Current Treatment Options in Cardiovascular Medicine*, *21*(3), 15. https://doi.org/10.1007/s11936-019-0720-9

Wang, J., Zhu, H., Wang, S.-H., & Zhang, Y.-D. (2021). A review of deep learning on medical image analysis. *Mobile Networks and Applications*, *26*(1), 351–380.

Wang, Z. J., Zhang, L. L., Elmariah, S., Han, H. Y., & Zhou, Y. J. (2017). Prevalence and prognosis of nonobstructive coronary artery disease in patients undergoing coronary angiography or coronary computed tomography angiography: A meta-analysis. *Mayo Clinic Proceedings*, *92*(3), 329–346. https://doi.org/10.1016/j.mayocp.2016.11.016

Wu, W., Zhang, J., Xie, H., Zhao, Y., Zhang, S., & Gu, L. (2020). Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint. *Computers in Biology and Medicine*, *118*, 103657.

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., & Lee, P. H. (2019). Deep learning segmentation of major vessels in x-ray coronary angiography. *Scientific Reports*, *9*(1), 16897.

Zhang, X., Du, H., Song, G., Bao, F., Zhang, Y., Wu, W., & Liu, P. (2022). X-ray coronary centerline extraction based on c-unet and a multifactor reconnection algorithm. *Computer Methods and Programs in Biomedicine*, *226*, 107114.

Zhao, C., Vij, A., Malhotra, S., Tang, J., Tang, H., Pienta, D., Xu, Z., & Zhou, W. (2021). Automatic extraction and stenosis evaluation of coronary arteries in invasive coronary angiograms. *Computers in Biology and Medicine*, *136*, 104667.

Zhou, C., Dinh, T. V., Kong, H., Yap, J., Yeo, K. K., Lee, H. K., & Liang, K. (2021). Automated deep learning analysis of angiography video sequences for coronary artery disease. *arXiv preprint arXiv:2101.12505*.

AUTHOR BIOGRAPHIES

Ariadna Jiménez-Partinen received her BSc in Health Engineering degree (major in Biomedical Engineering) from the University of Málaga, Spain, in 2019, her MSc in Biomedical Engineering degree from the University of País Vasco, Spain, in 2020, and her MSc in Advanced Biotechnology degree from the University of Málaga and the International University of Andalusia, Spain, in 2021. She is currently a PhD student at the Degree Computer Technologies and a researcher at the Department of Computer Languages and Computer Science, University of Málaga. Her technical interests are in medical image analysis, image processing and classification.

Miguel A. Molina-Cabello received his MSc and PhD degrees in Computer Engineering from the University of Malaga, Spain, in 2015 and 2018. He joined the Department of Computer Languages and Computer Science, University of Malaga, in 2015, where he has a teaching and researching position. He also keeps pursuing research activities in collaboration with other Universities. His technical interests are in visual surveillance, image/video processing, computer vision, neural networks and pattern recognition.

Karl Thurnhofer-Hemsi received his B.Sc. in Computer Engineering and his M.Sc. in Mathematics degrees from the University of Málaga, Spain, in 2014. He joined the Medical and Health Research Center of the University of Málaga in 2015. He is currently a Ph.D. candidate at the Department of Computer Languages and Computer Science, University of Málaga. His technical interests are in medical image analysis, pattern recognition and image processing. He also has interests in evaluation methods and learning systems.

Esteban J. Palomo received his MSc and PhD degrees (with honours) 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, University of Yachay Tech, 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.

Jorge Rodríguez-Capitán acquired his medical degree at University of Malaga (Spain), and obtained his specialization in Cardiology at the Hospital Virgen de la Victoria (Malaga, Spain). He received his PhD degree in Medicine from the University of Malaga (Spain) in 2016. Currently, he develops his career both as a clinical cardiologist and as a clinical researcher. His research concerns include tricuspid valvular disease, heart failure, ischemic heart disease, hypertrophic cardiomyopathy, and recently COVID-19 disease.

Ana I. Molina-Ramos acquired his medical degree at University of Malaga (Spain), and obtained his specialization in Cardiology at the Hospital Virgen de la Victoria (Málaga, Spain). Fellowship in advanced cardiac imaging at Mount Sinai Hospital of New York in 2019. Currently contracted by the Instituto de Salud Carlos III (ISCI III, Madrid, Spain) combining research and clinical activity. Her main area of expertise is cardiac imaging.

Manuel Jiménez-Navarro acquired his medical degree at University of Sevilla (Spain), and obtained his specialization in Cardiology at the Hospital Virgen de la Victoria (Malaga, Spain). He received his PhD degree in Medicine from the University of Malaga (Spain) in 1999. Currently, he works as full professor of Medicine in University of Malaga; as clinical cardiologist in the same hospital and he is the director of the CIBERCV node research group (Carlos III Health Institute, Spain) which is integrated into the Biomedical Research Institute of Malaga (IBIMA).

How to cite this article: Jiménez-Partinen, A., Molina-Cabello, M. A., Thurnhofer-Hemsi, K., Palomo, E. J., Rodríguez-Capitán, J., Molina-Ramos, A. I., & Jiménez-Navarro, M. (2024). CADICA: A new dataset for coronary artery disease detection by using invasive coronary angiography. *Expert Systems*, e13708. https://doi.org/10.1111/exsy.13708