Deep Learning for Coronary Artery Disease Severity Classification

Ariadna Jiménez-Partinen[1], Karl Thurnhofer-Hemsi[1,3,4], Esteban J. Palomo[1,3] and Ana I. Molina-Ramos[2,3,4]

[1] Department of Computer Languages and Computer Science, University of Málaga, Bulevar Louis Pasteur, 35, 29071, Malaga, Spain

[2] Cardiology Deparment, Hospital Universitario Virgen de la Victoria, 29010, Malaga, Spain

[3] Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND. C/ Severo Ochoa, 35, Málaga TechPark, Campanillas, 29590, Malaga, Spain

[4] Centro de Investigación Biomédica en Red de Enfermedades Cardiovasculares (CIBERCV), Instituto de Salud Carlos III (ISCIII). Avenida Monforte de Lemos, 3-5. Pabellón 11. Planta 0, 28029, Madrid, Spain

Tel.: +34 952133392

E-mail: ariadnapartinen@uma.es

Summary: Medical imaging evaluations are one of the fields where computed-aid diagnosis could improve the efficiency of diagnosis supporting physician decisions. Cardiovascular Artery Disease (CAD) is diagnosed using the gold standard, Invasive Coronary Angiography (ICA). In this work, performance analysis for binary classification of ICA images considering the severity ranges separately is reported, evaluating how performance is affected depending on the degree of lesions considered. For this purpose, an annotated dataset of ICA images was employed, which contains the ground truth, the location and the category of lesions into seven possible ranges: <20%, [20%, 49%], [50%, 69%], [70%, 89%], [90%, 98%], 99%, and 100%. The ICA images were pre-processed, divided into patches and balanced by downsampling and data augmentation. In this study, four known pre-trained CNN architectures were trained using different categories of lesion degree as input, whose F-measures are computed. Results report that the F-measures showed a behavior dependent on the narrow presents of the image, being lesions with more than 50% severity were better classified, achieving an F-measure of 75%.

Keywords: Invasive Coronary Angiography; Deep Learning; Classification; Healthcare

1. Introduction

The use of Machine Learning, and more concrete Deep Learning, techniques are widely extended in different image areas. Medical imaging evaluations are one of the fields where computed-aid diagnosis could improve the efficiency of diagnosis supporting physician decisions. Cardiovascular Artery Disease (CAD) is usually diagnosed using Invasive Coronary Angiography (ICA), which remains the gold standard [1]. ICA imaging is an X-ray-based evaluation image, where a radiocontrast is injected into the arteries of the myocardium through a catheter which is inserted by a percutaneous incision in the radial or femoral artery [2].

The assessment of the degree of the narrowing is usually done visually, involving a substantial impact of subjectivity because it depends on the experience of the clinician, generating interobserver variability [3]. It has been proven the successful use of Convolutional Neural Networks (CNN) classifying objects in the image classification system [4,5,6].

The aim of this work is to evaluate how the performance of the binary classification problem of ICA images is affected depending on which category is considered as the positive class, being affected by the morphology and grade of narrowing of the lesion.

2. Methodology

The ICA images used in the present study are from a dataset composed of 42 anonymized videos acquired at Hospital Universitario Virgen de la Victoria (Málaga, Spain) by the cardiac angiography equipment Artis Zee (Siemens AG, Muenchen, Germany) under the corresponding regulations and permissions of the local ethical committee of the hospital. The original Digital Imaging and Communication in Medicine (DICOM) files were converted to PNG files for ease of use, obtaining images with a size of 512×512 pixels. The database is composed of different projections of the right coronary artery (RCA) and left coronary artery (LCA), such as right and left anterior obliques (RAO, LAO), with cranial and caudal angulations.

The frames with enough radiocontrast to visualize the arteries correctly were annotated (bounding box and category) by both the informatic and cardiology team. The possible severity categories of a lesion are <20%, [20%, 49%], [50%, 69%], [70%, 89%], [90%, 98%], 99%, and 100%. In Fig. 1, a sample of each category is reported. In total, there are 3,900 images with at least one lesion and 1,943 images without visible lesions.

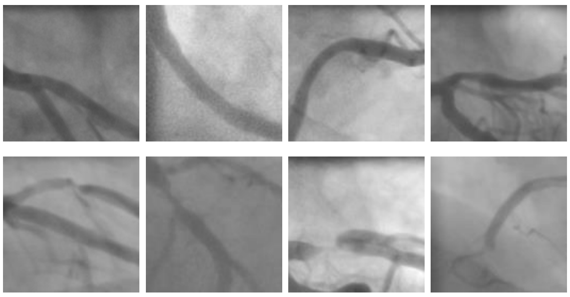

Fig. 1. Samples of the eight lesion categories. From left to right and top to bottom: non-lesion, <20%, [20%, 49%], [50%, 69%], [70%, 89%], [90%, 98%], 99%, and 100%.

We followed a patch classification methodology, that is, the raw images were divided into a 4×4 grid. These patches were resized to 32×32 pixels and assigned to the category corresponding if the centroid of the bounding box of a lesion falls into it, and the rest were considered "non-lesion" patches. This procedure implied a huge increase in the "non-lesion" class, unbalancing the distribution between both classes. To lessen it, the "non-lesion" class was filtered, removing background patches using a basic mask of the ICA images, which were obtaining applying morphological operations segmenting the vessels. Those patches whose mask had less than 2% of vessel pixels were discarded. Additionally, the negative class was randomly reduced to equalize the positive class. Then, data augmentation was applied using the following basic operations:

- Translations in the X and Y axis in a random range of [−4, 4] pixels
- Scaling of the images randomly with a scale factor in a range of [0.9, 1.2].
- Flip horizontally and vertically.

Finally, for the training process, the patches obtained for each input established were: 5904, 3536, 2992, 2620, 2268, 84, and 668, for <20%, [20%, 49%], [50%, 69%], [70%, 89%], [90%, 98%], 99%, and 100%, respectively.

Four pre-trained deep network architectures widely used were chosen: MobileNet-V2, ResNet-18, ResNet-50, and DenseNet-201.

The performance of the different models was evaluated and compared using the F-measure, which provides an overall performance, where the precision (P) and recall (R) measures under the concept of the harmonic mean are included. Considering the main four representative parameters: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [7].

$$F-measure = 2 \cdot \frac{P \cdot R}{P + R} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

3. Experimental results

3.1. Experiments description

The present work studies how the performance is affected depending on the lesion category considered the "lesion" class. Both classes were divided into training (80%) and test (20%) sets by videos, i.e., frames of the same videos in the test set are unavailable for the train set because frames of a sequence could be very similar, so a fairer performance evaluation is carrying out. To compare in a reliable way the results attained in test sets, K-fold cross-validation was implemented, with K=5.

For the training, we set: validation frequency = 50, validation patience = 5, maximum epochs = 70, initial learning rate = 0.0001, and SGDM solver (Stochastic Gradient Descent with Momentum). The batch size was set according to the number of training patches to keep the rate of iterations in all training processes.

The methods were implemented in MATLAB R2023a on a computer system with an Intel Core i9-10900X processor, 128 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU card. Moreover, no layer of chosen pre-trained models was frozen, updating all weights during training.

3.2. Results

In Table 1, the average and standard deviation of F-measure values obtained in the test sets are reported, showing in bold the highest values for each positive class. The first fact to stand out is that the four architectures employed have the same tendency in each positive class, without large differences between them.

Another remarkable aspect is the poor performance obtained in the classification task of 100% lesions. However, in the rest of the ranges, performance increased slightly, although it decreased again with low lesion degrees ([20%, 49%] and <20%). It could be explained because of the morphology of the lesion, although there are more patches to train, these lesions are more difficult to discern between non-lesion vessels being a more complex classification task. Whereas lesions higher than 50%, i.e., 99%, [90%, 98%], [70%, 90%], and [50%, 69%], are clearly distinguishable from non-lesion vessels getting higher outcomes, 0.698, 0.676, 0.685, and 0.752, respectively.

Table 1. F-measure obtained on the test set using 5-fold cross-validation for each category.

Severity	*MobileNet*	*ResNet-18*	*ResNet-50*	*DenseNet-201*
100%	0.328 ± 0.211	0.254 ± 0.161	**0.360 ± 0.234**	0.211 ± 0.089
99%	**0.698 ± 0.047**	0.577 ± 0.128	0.585 ± 0.242	0.632 ± 0.163
[90%, 98%]	0.619 ± 0.087	0.569 ± 0.053	0.555 ± 0.061	**0.676 ± 0.069**
[70%, 89%]	0.594 ± 0.058	**0.685 ± 0.015**	0.581 ± 0.046	0.623 ± 0.033
[50%, 69%]	**0.752 ± 0.015**	0.745 ± 0.035	0.695 ± 0.032	0.742 ± 0.027
[20%, 49%]	0.562 ± 0.031	**0.571 ± 0.028**	0.553 ± 0.034	0.511 ± 0.038
<20%	**0.602 ± 0.022**	0.559 ± 0.029	0.585 ± 0.020	0.600 ± 0.017

4. Conclusions

This study reports the evaluation of how the performance of the binary classification is affected depending on the severity degree of ICA images. Four CNN architectures were employed for the binary classification task, using lesion and non-lesion patches. The lesion class, positive class, was established considering the seven possible ranges of lesion degree separately. Results reported that severe lesions, higher than 50% (75% F-measure), are better classified than 100% (36% F-measure) and lower than 50% (60% F-measure) lesions.

Acknowledgments

This work is partially supported by the Autonomous Government of Andalusia (Spain) under grant UMA20-FEDERJA-108, and by the University of Málaga (Spain) under grant B1-2021_20 and grant UMA of Pla, and by the Instituto de Investigación Biomédica de Málaga - IBIMA Plataforma BIONAND (Spain), under grant PLAN_PROPIO (Área Ibima Tech) 2021. The authors thankfully the funding from the University of Malaga (Spain).

References

[1]. J. Knuuti, W. Wijns, A. Saraste, D. Capodanno, E. Barbato, C. Funck-Brentano, ... , J. J. Bax. 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*, Vol. 41, Issue 3, 2020, pp. 407-477.

[2]. G. Rigatelli, F. Gianese, M. Zuin. Modern atlas of invasive coronary angiography views: a practical approach for fellows and young interventionalists, *The International Journal of Cardiovascular Imaging*, Vol. 38, Issue 5, 2022, pp. 919-926.

[3]. L. M. Zir, S. W. Miller, R. E. Dinsmore, J. P. Gilbert, J. W. Harthorne. Interobserver variability in coronary angiography, *Circulation*, Vol. 53, Issue 4, 1976, pp. 627-632.

[4]. G. Litjens, F. Ciompi, J. M. Wolterink, B. D. de Vos, T. Leiner, J. Teuwen, I. Išgum. State-of-the-art deep learning in cardiovascular image analysis, *JACC: Cardiovascular imaging*, Vol. 12, Issue 8 Part 1, 2019, pp. 1549-1565.

[5]. J. H. Moon, W. C. Cha, M. J. Chung, K. S. Lee, B. H.Cho, J. H. Choi. Automatic stenosis recognition from coronary angiography using convolutional neural networks, *Computer methods and programs in biomedicine*, Vol. 198, 2021, pp. 105819.

[6]. K. Pang, D. Ai, H. Fang, J. Fan, H. Song, J. Yang. Stenosis-DetNet: Sequence consistency-based stenosis detection for X-ray coronary angiography, *Computerized Medical Imaging and Graphics*, Vol. 89, 2021, pp. 101900.

[7]. E. Ovalle-Magallanes, J.G. Avina-Cervantes, I. Cruz-Aceves, J. Ruiz-Pinales. Hybrid classical–quantum convolutional neural network for stenosis detection in xray coronary angiography, *Expert Systems with Applications*, Vol. 189, 2022, pp. 116112.