Classification of Coronary Calcifications using Deep Learning Approach: A Feasibility Study with the orCaScore Database

Jessica Cristina Santos do Nascimento[1], **Italo Francyles Santos da Silva**[1], **Aristófanes Corrêa Silva**[1], **Anselmo Cardoso de Paiva**[1]

[1]Applied Computing Group, Federal University of Maranhão (UFMA), São Luís - MA, Brazil

{jess, francyles, ari, paiva}@nca.ufma.br

Abstract. *This document presents a deep learning method for the automatic detection and classification of calcified lesions in coronary arteries using Convolutional Neural Networks (CNN) based on the AlexNet architecture. The demonstration of this proposal uses orCaScore, a standardized and labeled database of low-dose radiation computed tomography (CT) images of the heart. The methodology division was designed starting with the region of interest (ROI) extraction for consecutively utilizing a patch-based approach. We tested this approach in 7,386 patches and achieved an accuracy of 67%, sensitivity of 100%, precision of 67% and specificity of 75%. Our technique aims to reinforce the detection and quantification of coronary calcified lesions, enabling accurate diagnosis and treatment of cardiovascular diseases.*

1. Introduction

World Health Organization (WHO) data show that cardiovascular diseases caused 17.9 million deaths in 2019 which represented 32% of all deaths in the world in that year [(WHO) et al. 2020]. Calcificated plaque deposits on the wall of coronary arteries is a strong and independent predictor of cardiovascular events. Though analysis of Computed Tomography (CT) images is a well established diagnostic practice, it is still characterized by manually identifying coronary calcifications. This is an approach that is prohibitive in large-scale clinical trials and impractical in daily exercise and is a well-known problem in current clinical practice [Wolterink et al. 2016].

This problem reflects the need for an automated and accurate solution for detecting coronary calcifications more efficiently and at scale. In addition, it is important to highlight how manual detection may be limited by the subjectivity of the observer and therefore accuracy may be compromised. This inaccuracy can lead to diagnostic errors and inadequate treatment. The automatic detection of calcified lesions in coronary arteries through the use of algorithms is a promising solution, as it offers precision, scalability and objectivity in detecting coronary calcifications.

To enhance automatic detection and quantification of calcified lesions in coronary arteries, thus contributing to better diagnosis and treatment of cardiovascular diseases, we propose a method for classifying coronary calcifications based on CT using Convolutional Neural Networks (CNN). The feasibility of this strategy is studied through the CNN architecture AlexNet [Krizhevsky et al. 2017] and using the orCaScore[1] database.

[1]Available on: https://orcascore.grand-challenge.org

The organization of this work was divided as follows: Section 2 shows the steps of planning and organizing the method used; Section 3, the main results achieved with the realization of this initiative; and Section 4 some final considerations about the work developed, discussing the knowledge obtained and the challenges to be overcome for enabling our strategy to be replicated within the area of computing applied to healthcare.

2. Materials and Method

In this Section, we present the methodological process of this work with the help of the dataset and the proposed method for coronary calcification classification. Its visual representation can be seen in Figure 1.

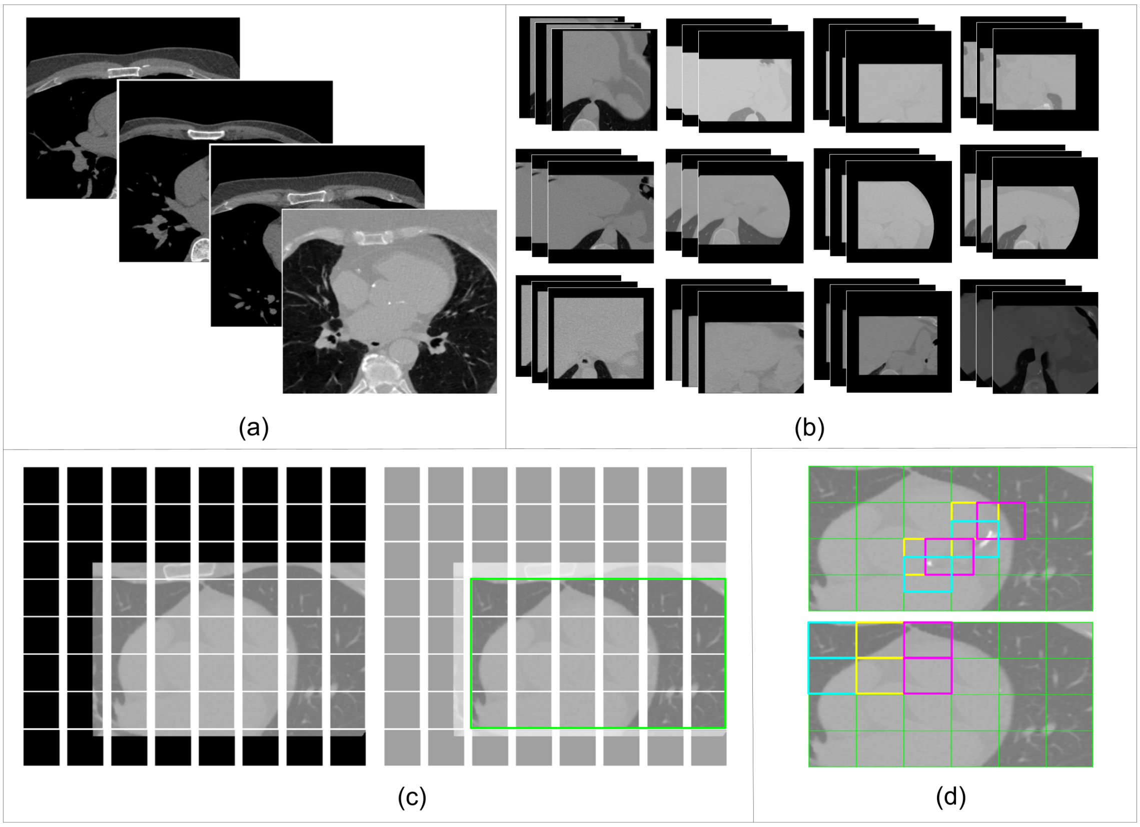

Figure 1. Samples of: (a) original database slices; (b) ROI extraction; (c) process of discarding patches with black pixels to optimize the amount; and (d) demonstration of overlapping positive patches and non-overlapping negatives.

2.1. Data Description

The orCaScore [Wolterink et al. 2016] is a standardized and labeled dataset of low-dose radiation CT images of the heart. The images were reviewed by cardiovascular radiology experts to identify coronary calcifications. Each image was labeled as positive or negative for the presence of calcifications by experienced cardiovascular radiologists. The purpose of this dataset is to assist in the development of machine learning algorithms for early detection of coronary artery disease and treatment of cardiovascular disease. The dataset composition used in this study to train and evaluate our model consists of a set of ECG-triggered CT scans containing one contrast-enhanced and one non-contrast scan from a total of 32 patients. These images were acquired from 4 different scanners from 4 different providers in 4 different hospitals.

Once this dataset includes images from multiple scanners and hospitals, providing a diverse and representative sample for training and evaluation, we used it to train and evaluate our model through a patch-based approach for precise localization and classification of lesions in this process. For the preparation of the data, the following steps were performed:

- **ROI Extraction** - To restrict the identification of all the potential candidate lesions to the heart's region of interest, the ROIs were extracted from the images by trained experts. They carefully delineated the heart region, excluding non-heart structures from further analysis. While this extraction was done manually, the authors suggest it is better to develop an automatic method for this step in future work. This approach aimed to reduce the number of irrelevant structures in the dataset and improve the classification performance.
- **Patch Extraction** - After ROI extraction, the images were divided into smaller windows, known as patches. Their extraction consisted of generating new images by subdividing the previously segmented images into smaller windows sized 52 x 52 mm (52 x 52 pixels). The objective was to classify them into patches belonging to or not belonging to coronary calcification. This step highlighted the number of patches containing calcification becoming significantly less than the ones without it. Images that did not have calcification were subjected to a side-by-side approach for the patch extraction process, where the objective is to reduce the set size. The remaining images, all containing calcification, were part of an extracting process involving an overlap of patches based on the work of Junior et al. (2021) . We discarded the calcification-free patches from these remaining images to reduce the general quantity of preprocessed images.
- **Data Distribution** - The distribution was based on the results of the earlier-mentioned extractions, which generated a dataset of 37,610 images. They were divided into training, validation and testing sets. The training set consisted of 22,156 images (60% of the total), the validation set consisted of 7,386 images (20% of the total) and the testing set consisted of 7,386 images (20% of the total). The dataset included 921 images containing calcification and 36,689 images without it.

2.2. Proposed Method

In this study, we propose the use of a CNN based on the AlexNet architecture to classify coronary calcifications. The AlexNet architecture was chosen due to its simplicity and effectiveness in image classification tasks. Additionally, it has been widely used in medical image analysis, making it a suitable choice for this preliminary study.

The CNN was implemented using Keras[2] library version 2.11.0 and TensorFlow[3] library version 2.12.0. Its architecture was defined using the Sequential class. The model

[2]Available on: https://keras.io/
[3]Available on: https://www.tensorflow.org/

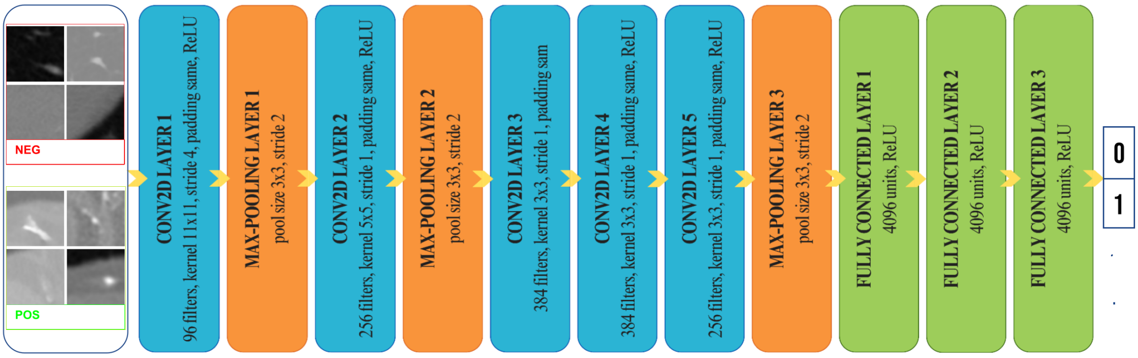

Figure 2. Representation of AlexNet architecture.

consisted of 13 layers including 5 convolutional layers, 3 max-pooling layers and 3 fully connected layers (Figure 2). The output layer had 2 units that corresponded to the two classes of the dataset. To optimize the model, we used the Adam optimizer and categorical cross-entropy as the loss function. The training process was performed over 50 epochs with a batch size of 32.

It is worth noting that the choice of CNN architecture, optimizer, loss function and training parameters can significantly affect the performance of the model. Therefore, we provide a detailed description of these aspects to ensure the reproducibility of our work. The image preprocessing and preparation steps were based on the work of Santini et al. (2018) and included image normalization and resizing to a fixed size of 227 x 227 pixels to standardize the images and ensure compatibility with the AlexNet architecture.

In the next Section, we describe the evaluation metrics used to assess the performance of the proposed method.

3. Results

In this Section, we present the main results achieved with the implementation of the proposed method. Primarily, the training process involved a convolutional neural network based on the AlexNet architecture, consisting of 13 layers including 5 convolutional layers, 3 max-pooling layers and 3 fully connected layers. The model was trained using the Adam optimizer and categorical cross-entropy as the loss function, over 50 epochs with a batch size of 32. The classification model achieved an accuracy of 0.67 on the testing set, indicating that 67% of the predictions were correct. The sensitivity score of 1.00 demonstrated that the model correctly identified all positive cases while the specificity score of 0.75 indicated that the model correctly identified 75% of negative cases. The precision score of 0.67 represented that when the model predicted a positive result, it was correct 67% of the time.

However, while the sensitivity score indicates that the model correctly identified all positive cases, the specificity score suggests that the model had some difficulty in identifying negative cases. Therefore, additional optimization of the model is necessary to increase its overall performance. Further improvements will be made to the proposed method, such as exploring different deep learning architectures, increasing the size of the dataset through the use of data augmentation techniques in the training set and including more features in the analysis, since the most significant limitation is notably when it comes to the imbalance in the dataset, with only 921 images containing calcification and 36,689 images without it. This can have a prejudicial impact on the ability of the model to accurately classify images and may result in biased results.

In summary, the results of this study demonstrate — according to data from the systematic analysis of various evaluation metrics addressed in the work of Sokolova et al. (2009) — that the developed model has deficiencies in accurately predicting the outcome of the target variable. Although the model achieved a sensitivity of 1.00, its overall accuracy was only 0.67. This indicates a considerable number of false positives and false negatives. The precision of the model was also only 0.67, suggesting that it may not be reliable in identifying true positive cases. These limitations must be addressed and further optimization and validation are necessary before it can be used in clinical practice. Nevertheless, the proposed method demonstrated good first steps for the classification of coronary calcifications using CNNs.

4. Conclusion

In this section, we delve into the proposed approach by initially examining its placement within the existing literature. We also shed light on the limitations of the method and present conclusive findings.

Given the variations in study design, data distribution and quantitative assessment methods, it is noteworthy that direct comparisons between the results are challenging. This is principally the case when observing the low-value metrics obtained in this current work, highlighting that the method still does not have a prominent place in the literature. However, observing the deficiencies and the advances proposed in the automatic deep learning approach for coronary artery calcium segmentation developed by Santini et al. (2018), we draw inspiration from their techniques for ROI and patch extraction to achieve validation metrics as promising as those presented by them. As a result of that inspiration, our approach differs from the works analyzed by Litjens et al. (2019) as we focus on ROI extraction and a patch-based approach for classification, rather than performing a complete exam evaluation. By focusing on these extractions, we aspire to improve the efficiency of classification, particularly in scenarios where detailed analysis of specific regions is of utmost importance. Additionally, the methods evaluated by Wolterink et al. (2016) using orCaScore served as the foundation for our usage of this same database. By leveraging the insights from these works and through the adoption of these inspirations, we aim to achieve validation metrics comparable to the state-of-the-art methods in the field.

Further improvements are necessary to enhance the performance of the model and reduce its limitations. One possible solution to refine our technique is to automate the extraction process of the heart's ROI. Currently, the ROI is manually extracted which can be time-consuming and potentially subject to human error since it is prone to inter-observer variability. An automatic tactic for ROI extraction could potentially improve the accuracy and efficiency of the proposed method. Another approach is to explore different CNN architectures or to use transfer learning techniques to leverage pre-trained models. It may also be beneficial to incorporate additional features or imaging modalities to enhance the predictive power of the model. The authors will continue their work by applying data augmentation techniques to the training set to balance the image database.

We concede that our current approach is still a work in progress with room for refinement and advancement. We hope to continue our work to achieve a future state where our approach can make significant contributions to improve the efficiency and accuracy of detecting and quantifying calcified lesions within coronary arteries. In due course, this will bestow upon the medical community a tool for the accurate diagnosis and treatment of cardiovascular diseases. The challenges faced during the implementation of the proposed method have provided valuable learning experiences that can be applied to future research in this field including the extension of this approach to other areas of healthcare technology.

Acknowledgments

The authors acknowledge the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and Fundação de Amparo à Pesquisa Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) (Brazil), Empresa Brasileira de Serviços Hospitalares (Ebserh) Brazil (Grant number 409593/2021-4) for the financial support.

References

Júnior, R. J. F. A., da Silva, I. F. S., Silva, A. C., and de Paiva, A. C. (2021). Detecçao de nódulos da tireoide em exames de termografia utilizando redes neurais convolucionais em cascata. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 269–280. SBC.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. volume 60, pages 84–90. AcM New York, NY, USA.

Litjens, G., Ciompi, F., Wolterink, J. M., de Vos, B. D., Leiner, T., Teuwen, J., and Išgum, I. (2019). State-of-the-art deep learning in cardiovascular image analysis. volume 12, pages 1549–1565. American College of Cardiology Foundation Washington, DC.

Santini, G., Latta, D. D., Martini, N., Valvano, G., Gori, A., Ripoli, A., Susini, C. L., Landini, L., and Chiappino, D. (2018). An automatic deep learning approach for coronary artery calcium segmentation. In *EMBEC & NBC 2017: Joint Conference of the European Medical and Biological Engineering Conference (EMBEC) and the Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC), Tampere, Finland, June 2017*, pages 374–377. Springer.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. volume 45, pages 427–437. Elsevier.

(WHO), W. H. O. et al. (2020). Global health estimates 2020: Deaths by cause, age, sex, by country and by region, 2000-2019: Who; 2020.

Wolterink, J. M., Leiner, T., De Vos, B. D., Coatrieux, J.-L., Kelm, B. M., Kondo, S., Salgado, R. A., Shahzad, R., Shu, H., Snoeren, M., et al. (2016). An evaluation of automatic coronary artery calcium scoring methods with cardiac ct using the orcascore framework. volume 43, pages 2361–2373. Wiley Online Library.