Attention-based learning of views fusion applied to myocardial infarction diagnosis from x-ray CT

Jakub Gwizdala
EPFL, Switzerland
jakub.gwizdala@epfl.ch

Ortal Senouf
EPFL, Switzerland
ortal.senouf@epfl.ch

Denise Auberson
CHUV, Switzerland
denise.auberson@chuv.ch

David Meier
CHUV, Switzerland
David.Meier@chuv.ch

David Rotzinger
CHUV, Switzerland
David.Rotzinger@chuv.ch

Stephane Fournier
CHUV, Switzerland
Stephane.Fournier@chuv.ch

Salah Qanadli
CHUV, Switzerland
Salah.Qanadli@chuv.ch

Olivier Muller
CHUV, Switzerland
Olivier.Muller@chuv.ch

Pascal Frossard
EPFL, Switzerland
pascal.frossard@epfl.ch

Emmanuel Abbé
EPFL, Switzerland
emmanuel.abbe@epfl.ch

Dorina Thanou
EPFL, Switzerland
dorina.thanou@epfl.ch

Abstract

Despite being a non-invasive imaging modality, coronary computed tomography angiography (CCTA) is still not the clinical gold-standard modality for the diagnosis and evaluation of Coronary Artery Diseases (CAD), which is typically performed with an invasive coronary angiography (ICA). In this work, we aim at bringing CCTA diagnosis performance closer to the level of the ICA. We propose a deep attention learning framework that takes as an input non-invasive CCTA images and is able to predict a clinical decision, such as revascularization, that is typically based on invasive modalities such as ICA. We represent the CCTA volumetric imaging by two cross-sectional views that follow the curvature of the coronary artery, and we use an attention mechanism that learns a fused representation for better diagnosis. Experimental results on a clinical study of 80 patients indicate that the learned fused model achieves a significant gain in the performance (F1-score: 0.53 ± 0.11) with respect to the CT fractional-flow-reserve (FFR_{CT}), a clinical baseline estimating the drop of flow from CCTA (F1-score: 0.46 ± 0.09). These preliminary results confirm that a data-driven approach can boost the diagnosis power of CCTA and eventually contribute towards the wider adoption of this non-invasive imaging modality in clinical settings.

1 Introduction

Myocardial infarction (MI), commonly referred to as a heart attack, is attributed to an impairment in blood flood by atherosclerotic plaque causing a narrowing, or stenosis, in the coronary artery. In clinics, Invasive coronary angiography (ICA) is considered the gold standard for the assessment of the hemodynamic significance of stenosis. However, it has been identified that only around 1%-13% of such invasive examinations target patients with obstructive coronary artery disease (Agewall et al., 2017). Avoiding unnecessary invasive procedures and replacing them with a non-invasive alternative,

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

such as the X-ray coronary computed tomography angiography (CCTA), is a desirable solution (Gorenoi et al., 2012). CCTA provides a 3D anatomical assessment of the coronary arteries, plaque composition, morphology and stenosis quntification. Fractional-flow-reserve derived from CCTA (FFR-CT) provides a non-invasive assessment of the hemodynamic impact of a stenosis (Nørgaard et al., 2014). Nevertheless, CCTA and FFR-CT has not been sufficiently validated in higher risk settings like acute myocardial infarction (Collet et al., 2021).

In the last years, there have been attempts at using machine learning and deep learning methods to alleviate some of the challenges that hinder CCTA from becoming the gold standard in the diagnosis of flow-limiting coronary artery disease. Along these lines, various works try to estimate predictive measures concerning stenoses (Griffin et al., 2022; Hong et al., 2019; Wolterink et al., 2016), some involving coronary artery segmentation (Lin et al., 2022). However, these works are limited to extracting handcrafted predictive measures rather than discovering new ones. Finally, the works by Denzinger et al. (2019) and Zou et al. (2020) approach the task of directly predicting the cardiologist decision of administering revascularization treatment procedure with promising results. They represent the 3D content of the CT using a sequence of axial views following the curvature (center line) of the artery, (see Figure 1, left). In our work, we tackle a similar task of predicting the cardiologist's decision directly from CCTA. However, we use two reconstructed cross-sectional views to represent a coronary segment (see Figure 1, right). This choice, commonly referred to in the literature as 2.5D representation, is motivated by the need to reduce the overall computational effort Denzinger et al. (2020, 2022). We investigate approaches for fusing information from these two views within a weak labeling setting, i.e, the labels are given in the image level without the annotation of the specific region of interest (lesion) in the image. To this end, we propose an attention learning framework that is able to fuse the views effectively. The framework is validated on a clinical study of 80 patients, with promising results.

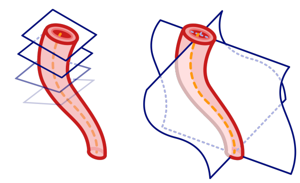

Figure 1: Left: axial views along the artery centerline. Right: two cross-sectional views used in our work. Illustration based on Kanitsar et al. (2002).

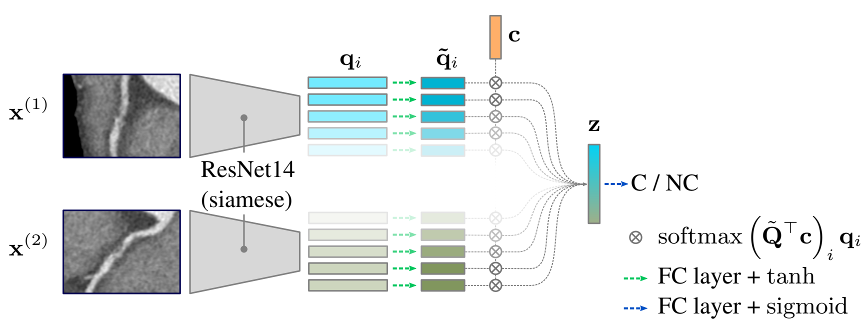

Figure 2: Architecture of the culprit lesion classification model. Both views $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are passed through two siamese ResNet14 networks to generate feature maps – sets of $\mathbf{q}_i$ vectors (Pang et al., 2021). After projecting to a lower dimensional space and stacking into a $\tilde{\mathbf{Q}}$ matrix, they are combined through the content attention module to generate an artery segment representation $\mathbf{z}$ (Ilse et al., 2018). Finally, the classification of the sample as culprit (C) or non-culprit (NC) is performed.

2 Methodology

2.1 Clinical study

The imaging data used in our work comes from patients recruited in an observational trial. The patients, who were in the emergency department with a suspected MI, had a CCTA prior getting an ICA. From each CCTA scan, based on a coronary artery centerline extraction, curved planar reformation (CPR) views were reconstructed for segments of the three major artery branches: right coronary artery (RCA), segments 1, 2 and 3; left anterior descending coronary artery (LAD), segments

1, 2 and 3; and the first segment of left circumflex artery (LCX). Each segment was visualized with two CPR views. The primary view was selected automatically by the hospital software (GE AW Advanced Visualisation) used to generate the reconstructed views. The secondary view was reconstructed in the orthogonal plane to the primary view, adjusted by the medical expert by up to 20°.

Each artery segment was separately labeled as "culprit", i.e., containing a culprit lesion, or "non-culprit", i.e., containing a lesion that was not identified as the one responsible for the heart attack or in which no lesions were present. This classification of segments was performed by the medical expert based on the visual assessment of ICA and FFR measurement, if available. The resulting dataset included 514 coronary artery segments from 80 patients among which 63 (12.3%) segments were labeled as culprit.

2.2 Methods

Our goal is to design a learning framework that takes as an input the two orthogonal views of each artery segment obtained from CCTA scans and infers if a segment is culprit or not. We pose the problem as a binary classification problem, and build a framework that takes into consideration two main challenges of our setup: (i) weak labeling: the labels are given in the image level, while only specific region in the image correspond with its label; (ii) the representation of the 3D segments from the two axial views.

Weak labeling treatment. Following similar reasoning as the one in Ilse et al. (2018); Pang et al. (2021), we treat this problem within the multiple instance learning (MIL) framework, where an image is represented by forming a bag of instances associated with one label. In our work, however, the instances are not represented by patches sampled from the image, but by learned features that aggregate information from different regions of the image. This reduces the problem to learning a feature attention layer to highlight information from areas in the image that are relevant for the classification task at hand.

Learned two-views fusion. In order to extract predictive features from each view, we pass them separately through a siamese ResNet (He et al., 2016) backbone to obtain two feature maps. After computing the embedding for each view, we fuse them by learning an attention mechanism inspired by the work of Pang et al. (2021). We use content attention and apply it to the set of feature vectors from both views together, as visualized in Figure 2. This generates a learned weighted average representation for the artery segment that can be ultimately classified as culprit or non-culprit by a fully-connected (FC) layer. Multi-view fusion in the context of coronary artery disease analysis has been applied in Zhang et al. (2020) in a task of stenosis quantification in ICA, employing a fixed fusion, rather than a learned one as proposed in our work. The features extraction backbone, together with the views fusion attention layer and the final classifier are trained end-to-end. The objective function is chosen to be the focal-loss (Lin et al., 2017) to account for the imbalance between the culprit and non-culprit classes.

3 Results and discussion

Due to the scarcity of the available dataset, we split the data into five folds and use cross-validation to train and validate the performance of the proposed pipeline. Samples were randomly split into folds patient-wise in a stratified manner, i.e., the ratio between patients with and without culprit segments is similar in each fold. Table 3 summarizes the results of our proposed approach (*learned fusion*) and compares them to five baselines: (1) *Naive*, a label is assigned randomly based on the frequency of each class in the train set; (2) *Views as channels*, the two views are stacked in the input's channels dimension; (3) *Learned attention*, application of the method by Pang et al. (2021), views still stacked as channels; (4) *Feature concatenation*, employing the method by Zhang et al. (2020) concatenating max-pooled feature representations of the two views; (5) FFR_{CT}, FFR values estimated from the CCTA by an external laboratory (HeartFlow ®, Redwood City, CA 94063, USA). The results suggest that learning the views fusion while taking into account the imbalance between culprit-related areas and non-culprit or background areas significantly contributes to the performance of the MI culprit prediction task. We believe that this performance can be further improved by introducing more than two cross-sectional views, similarly to what has been shown in (Denzinger et al., 2022) for a different task and by using radiomics for the representation of the views (Denzinger et al., 2019).

Method	F1	Precision	Recall	Specificity	AUC	Accuracy
Naive	0.12	0.12	0.12	0.88	–	0.78
Views as channels	0.36 ± 0.12	0.39 ± 0.17	0.37 ± 0.14	0.91 ± 0.04	0.35 ± 0.14	0.84 ± 0.03
Learned attention	0.40 ± 0.10	0.39 ± 0.10	0.45 ± 0.15	0.90 ± 0.04	0.40 ± 0.08	0.84 ± 0.03
Feature concatenation	0.46 ± 0.13	0.53 ± 0.18	0.46 ± 0.18	**0.93 ± 0.05**	0.45 ± 0.13	0.87 ± 0.04
Learned fusion	**0.53 ± 0.11**	**0.55 ± 0.18**	0.55 ± 0.14	**0.93 ± 0.05**	**0.52 ± 0.17**	**0.88 ± 0.04**
FFR_{CT}	0.46 ± 0.09	0.33 ± 0.06	**0.77 ± 0.18**	0.78 ± 0.04	–	0.78 ± 0.04

References

Agewall, S., Beltrame, J. F., Reynolds, H. R., Niessner, A., Rosano, G., Caforio, A. L., De Caterina, R., Zimarino, M., Roffi, M., Kjeldsen, K., et al. (2017). Esc working group position paper on myocardial infarction with non-obstructive coronary arteries. *European heart journal*, 38(3):143–153.

Collet, J.-P., Thiele, H., Barbato, E., Barthélémy, O., Bauersachs, J., Bhatt, D. L., Dendale, P., Dorobantu, M., Edvardsen, T., Folliguet, T., et al. (2021). 2020 esc guidelines for the management of acute coronary syndromes in patients presenting without persistent st-segment elevation: the task force for the management of acute coronary syndromes in patients presenting without persistent st-segment elevation of the european society of cardiology (esc). *European heart journal*, 42(14):1289–1367.

Denzinger, F., Wels, M., Breininger, K., Gülsün, M. A., Schöbinger, M., André, F., Buß, S., Görich, J., Sühling, M., and Maier, A. (2020). Automatic CAD-RADS Scoring Using Deep Learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 45–54. Springer.

Denzinger, F., Wels, M., Ravikumar, N., Breininger, K., Reidelshöfer, A., Eckert, J., Sühling, M., Schmermund, A., and Maier, A. (2019). Coronary Artery Plaque Characterization from CCTA Scans using Deep Learning and Radiomics. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 593–601. Springer.

Denzinger, F., Wels, M., Taubmann, O., Gülsün, M. A., Schöbinger, M., André, F., Buss, S. J., Görich, J., Sühling, M., Maier, A., et al. (2022). CAD-RADS Scoring using Deep Learning and Task-Specific Centerline Labeling. *arXiv preprint arXiv:2202.03671*.

Gorenoi, V., Schönermark, M. P., and Hagen, A. (2012). CT coronary angiography vs. invasive coronary angiography in CHD. *GMS Health Technology Assessment*, 8.

Griffin, W. F., Choi, A. D., Riess, J. S., Marques, H., Chang, H.-J., Choi, J. H., Doh, J.-H., Her, A.-Y., Koo, B.-K., Nam, C.-W., Park, H.-B., Shin, S.-H., Cole, J., Gimelli, A., Khan, M. A., Lu, B., Gao, Y., Nabi, F., Nakazato, R., Schoepf, U. J., Driessen, R. S., Bom, M. J., Thompson, R., Jang, J. J., Ridner, M., Rowan, C., Avelar, E., Généreux, P., Knaapen, P., de Waard, G. A., Pontone, G., Andreini, D., and Earls, J. P. (2022). AI evaluation of stenosis on coronary CT angiography, comparison with quantitative coronary angiography and fractional flow reserve. *JACC: Cardiovascular Imaging*.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

Hong, Y., Commandeur, F., Cadet, S., Goeller, M., Doris, M. K., Chen, X., Kwiecinski, J., Berman, D. S., Slomka, P. J., Chang, H.-J., and Dey, D. (2019). Deep learning-based stenosis quantification from coronary CT Angiography. In *Medical Imaging 2019: Image Processing*, volume 10949, pages 643–651. SPIE.

Ilse, M., Tomczak, J. M., and Welling, M. (2018). Attention-based Deep Multiple Instance Learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 2127–2136.

Kanitsar, A., Fleischmann, D., Wegenkittl, R., Felkel, P., and Groller, E. (2002). CPR – curved planar reformation. In *IEEE Visualization, 2002. VIS 2002.*, pages 37–44.

Lin, A., Manral, N., McElhinney, P., Killekar, A., Matsumoto, H., Kwiecinski, J., Pieszko, K., Razipour, A., Grodecki, K., Park, C., Otaki, Y., Doris, M., Kwan, A. C., Han, D., Kuronuma, K., Tomasino, G. F., Tzolos, E., Shanbhag, A., Goeller, M., Marwan, M., Gransar, H., Tamarappoo, B. K., Cadet, S., Achenbach, S., Nicholls, S. J., Wong, D. T., Berman, D. S., Dweck, M., Newby, D. E., Williams, M. C., Slomka, P. J., and Dey, D. (2022). Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. *The Lancet Digital Health*, 4(4):e256–e265.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal Loss for Dense Object Detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988.

Nørgaard, B. L., Leipsic, J., Gaur, S., Seneviratne, S., Ko, B. S., Ito, H., Jensen, J. M., Mauri, L., De Bruyne, B., Bezerra, H., Osawa, K., Marwan, M., Naber, C., Erglis, A., Park, S.-J., Christiansen, E. H., Kaltoft, A., Lassen, J. F., Bøtker, H. E., and Achenbach, S. (2014). Diagnostic Performance of Noninvasive Fractional Flow Reserve Derived From Coronary Computed Tomography Angiography in Suspected Coronary Artery Disease: The NXT Trial (Analysis of Coronary Blood Flow Using CT Angiography: Next Steps). *Journal of the American College of Cardiology*, 63(12):1145–1155.

Pang, G., Ding, C., Shen, C., and Hengel, A. V. (2021). Explainable Deep Few-shot Anomaly Detection with Deviation Networks. *arXiv preprint arXiv:2108.00462*.

Wolterink, J. M., Leiner, T., de Vos, B. D., van Hamersvelt, R. W., Viergever, M. A., and Išgum, I. (2016). Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Medical Image Analysis*, 34:123–136.

Zhang, D., Yang, G., Zhao, S., Zhang, Y., Ghista, D., Zhang, H., and Li, S. (2020). Direct Quantification of Coronary Artery Stenosis Through Hierarchical Attentive Multi-View Learning. *IEEE Transactions on Medical Imaging*, 39(12):4322–4334.

Zou, Z.-Z., Xie, K., Zhao, Y.-F., Wan, J., Lan, L., and Wen, C. (2020). Intelligent Assessment of Percutaneous Coronary Intervention Based on GAN and LSTM Models. *IEEE Access*, 8:90640–90651.