Coronary Artery Stenosis Assessment in X-Ray Angiography Through Spatio-Temporal Attention for Non-Invasive FFR and iFR Estimation

Raffaele Mineo¹*^a, Federica Proietto Salanitri¹*, Giovanni Bellitto¹, Ovidio De Filippo²,

Fabrizio D'Ascenzo², Simone Palazzo¹ and Concetto Spampinato¹

¹PeRCeiVe Lab, University of Catania, Catania, Italy ²Department of Medical Sciences, University of Turin, Turin, Italy

Keywords: Attention Methods, Coronary Angiography, Medical Imaging Analysis.

Abstract: Determining the degree of stenosis in coronary arteries through X-ray angiography imaging is a multifaceted task, given their appearance variability, the overlapping of vessels, and their small size. Traditional automated approaches utilize 2D deep models processing multiple angiography views as well as key frames. In this research, we propose a new deep learning model to non-invasively evaluate the fractional flow reserve (FFR) and instantaneous wave-free ratio (iFR) of moderate coronary stenosis from angiographic videos to better analyze spatial and temporal correlation without manual preprocessing. Our strategy harnesses 3D Convolutional Neural Networks (CNNs) to learn local spatio-temporal features and integrates self-attention layers to understand broad correlations within the feature set. At training time, both FFR and iFR values are employed for supervision, with missing targets suitably handled through multi-branch outputs. The resulting model can be employed to predict the presence of a clinically-significant coronary artery stenosis and to directly determine the FFR and iFR values. We also include an explainability strategy to show which parts of a video the model focuses on in the assessment of FFR and iFR values. Our proposed model demonstrates superior results than competitors on a dataset of 778 angiography exams from 389 patients. Importantly, our model doesn't require key frames, thus reducing the efforts required by clinicians.

1 INTRODUCTION

Invasive evaluation of coronary conditions utilizing Fractional Flow Reserve (FFR) and/or Instantaneous Free wave Ratio (iFR) serves as an essential guide for Percutaneous Coronary Revascularization (PCI) of intermediate grade coronary lesions (Neumann et al., 2018; Knuuti and Revenco, 2020). Despite its proven reduction of subsequent revascularization procedures and associated prognostic benefits, its real-world application remains modest. This can be attributed to factors such as the extensive setup and measurement time, the considerable cost of the diagnostic probe, and the invasive nature of the procedure that may present a low, but not negligible, risk of complications. Furthermore, these evaluations can be subject to significant inter-observer variations. Nevertheless, clinicians are not interested to the absolute values of FFR/iFR, rather, if these values are under or over threshold, which is set to 0.80 for FFR and 0.89 for iFR (Neumann et al., 2018; Tonino et al., 2009; De Bruyne et al., 2012). In light of

^aEqual contribution by R. Mineo and F. Proietto Salanitri these challenges, Artificial Intelligence (AI) and Machine Learning (ML), with the aid of convolutional neural networks (CNNs) and more recently vision transformers (Dosovitskiy et al., 2020), have shown immense potential (Proietto Salanitri et al., 2021; Tomar et al., 2022; Salanitri et al., 2022; Valanarasu et al., 2021). They can relax these constraints by enhancing risk assessment and cardiovascular imaging analysis and automating artery stenosis quantification from coronary angiography. Despite the advancements, existing strategies require key frame selection alongside the incorporation of multiple angiography views (Zhang et al., 2020; Zhang et al., 2019) (see Fig. 1, increasing the burden on both the patients and cardiologists.

To address the above limitations, in this paper we propose a deep network that employs two views for each exam, but it does not require any key frame selection, thus balancing the need for comprehensive information. Our approach specifically seeks to evaluate stenosis severity through both direct and indirect estimation of FFR/iFR values from angiography videos. It harnesses the capabilities of both Convolutional Neural Networks (CNNs) and attention mecha-

Mineo, R., Proietto Salanitri, F., Bellitto, G., De Filippo, O., D'Ascenzo, F., Palazzo, S. and Spampinato, C.

DOI: 10.5220/0012449200003657

Paper published under CC license (CC BY-NC-ND 4.0)

ISBN: 978-989-758-688-0; ISSN: 2184-4305

Coronary Artery Stenosis Assessment in X-Ray Angiography Through Spatio-Temporal Attention for Non-Invasive FFR and iFR Estimation

In Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 1, pages 305-312

Proceedings Copyright © 2024 by SCITEPRESS - Science and Technology Publications, Lda



Figure 1: Two sample angiography views for the same patient. Red bounding boxes show the major stenosis.

nism to draw out meaningful spatio-temporal features and capture long-range dependencies within the input video. The CNN architecture excels at extracting meaningful spatio-temporal features from the input video, while the attention mechanism is adept at capturing long-range dependencies within the video. This combination allows our model to focus on the most relevant features for the task at hand, enhancing its predictive capabilities. Our approach is unique in that it assesses stenosis severity from two perspectives: classification and regression. The classification perspective allows us to categorize the severity of stenosis, while the regression perspective enables us to predict the precise FFR and iFR values. This dual-faceted approach provides a more nuanced understanding of the patient's condition, offering significant support to clinicians in their decision-making process.

We validated the feasibility and accuracy of our approach using a dataset collected from multiple Italian hospitals, consisting of 778 angiographic exams from 389 patients. Our approach demonstrated superior performance compared to conventional methods, underscoring its potential as a robust, adaptable, and effective solution for stenosis assessment based on coronary angiography. Moreover, we delved into the interpretability of our model to provide a more comprehensive understanding of its functionality.

Thus, the contributions of our paper can be summarised as follows:

- We put forward a novel convolutional model specifically designed to process and analyze X-Ray angiography videos, thereby addressing a significant gap in the current literature.
- We pioneer a multi-branch architecture that allows for diverse assessment modalities, including

both classification and regression. This innovative design not only promotes robust feature learning but also facilitates the training of heterogeneous datasets, thereby enhancing the model's versatility and applicability.

• We conduct an exhaustive experimental analysis to validate the efficacy of our proposed method. The results clearly demonstrate the superior performance of our model, outperforming existing solutions in terms of accuracy and robustness.

2 RELATED WORK

Coronary stenosis is a leading cause of heart failure due to impaired blood flow resulting from vessel narrowing. The severity of the condition may indicate its possible treatment, either through pharmaceutical methods or surgery (Neumann et al., 2018). Over the past decade, deep learning has been extensively utilized for diagnosing the severity of stenosis, its detection, and FFR quantification from imaging data. In particular, two main categories of methods exist: 2D approaches that analyze individual frames from angiography videos, and 3D models that directly extract spatio-temporal features from the entire video. Most 2D methods classify stenosis by severity levels or identify hemodynamically-significant stenosis by thresholding FFR/iFR values. Key frames are typically identified through CNN architectures (Moon et al., 2021; Rodrigues et al., 2021) or through a combination of convolutional and recurrent networks (Cong et al., 2019; Ma et al., 2017; Ovalle-Magallanes et al., 2022). A subset of these techniques limit the analysis to blood vessels by incorporating a pre-processing segmentation step (Wu et al., 2020; Au et al., 2018). Stenosis detection on individual frames is also prevalent and generally involves key frame identification and object detection models for stenosis location. A comprehensive benchmark of state-of-the-art object detection models for coronary stenoses is presented in (Danilov et al., 2020). Another set of 2D methods analyze the form and visual appearance of blood vessels on the key frame to locate stenoses (Zhao et al., 2021b; Zhao et al., 2021a). Additionally, interpretability approaches on framebased stenosis classification models generate activation maps to assist in stenosis detection (Moon et al., 2021; Cong et al., 2019). Recently, a few 3D models have emerged that operate on the entire angiography videos for quantitative coronary analysis and stenosis detection (Zhang et al., 2019; Zhang et al., 2020; Xue et al., 2018; Han et al., 2023). (Zhang et al., 2019; Zhang et al., 2020) are particularly relevant to our work as they conduct a quantitative coronary analysis (QCA) of stenoses. In detail, these methods carry out regression of several clinical indices, such as minimum lumen diameter, proximal and distal reference vessel diameters, among others, utilizing a primary angiography view alongside an additional side view and a manually chosen key frame. These methods are based on a 3D convolutional backbone, shared between the two angiography views, whose features are further processed by an attention layer in (Zhang et al., 2020). They also employ 2D dilated residual convolutions to extract features from the key frame. These two feature sets are then processed by a hierarchical self-attention mechanism for the final QCA regression.

Our proposed approach contrasts with existing ones in that it does not require a manually-selected key frame, thereby reducing the load on physicians. We employ a 3D CNN model combined with a global attention mechanism in conjunction with a multi-task formulation of stenosis severity assessment, which encourages the learning of more generic features and supports supervision via both discrete class labels and continuous FFR/iFR scores.

3 METHOD

The proposed model, as depicted in Fig. 2, is a deep learning architecture that combines a sequence of 3D convolutional kernels, inspired by the 3D ResNet-18 (Tran et al., 2018), with attention modules. The model is designed to process two views of angiography exams, which serve as its input. These inputs are processed by a shared 3D convolutional network,

which is trained to extract spatio-temporal features from the angiography data. More specifically, the feature extractor is a ResNet3D model (Tran et al., 2018), pre-trained on the Kinetics-400 dataset (Carreira and Zisserman, 2017) for video action recognition. To adapt this model from RGB to the X-ray inputs, the first-layer convolutional kernels are averaged over the channel dimensions, allowing the model to effectively process the angiography data. This feature extractor is shared among the two input views and for each view it produces a spatio-temporal tensor [W, H, T, C] where W, H, T and C are, respectively, weight, height, time and channels (feature maps). This tensor is then serialized into a $[W \times H \times T, C]$ tensor and processed by a multi-head spatio-temporal self attention module that performs the following operation:

 $MHA(Q, K, V) = \|\text{head}_1, \text{head}_2, \dots, \text{head}_h\| \times W_O$ (1)

where:

- Q, K, and V are the input queries, keys, and values, respectively.

- head_i = SelfAttention $(Q \times W_{Qi}, K \times W_{Ki}, V \times W_{Vi})$ represents the self-attention mechanism applied to each head.
- *h* is the number of attention heads.
- $\|\text{head}_1, \text{head}_2, \dots, \text{head}_h\|$ denotes the concatenation of the output of each head.
- W_O is the output transformation weight matrix.

The SelfAttention function is defined as:

SelfAttention
$$(Q, K, V) =$$
Softmax $\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V$
(2)

where d_k is the dimension of the key vectors.

This self-attention mechanism allows the model to focus on the most relevant features for the task at hand, enhancing its predictive capabilities. The self-attended features are then simultaneously fed to a three-branch layer. This layer is responsible for performing binary classification and quantification, through regression, of either FFR (for the data for which FFR is provided) or iFR (for the data for which iFR is provided). This multi-task approach allows the model to provide a comprehensive analysis of the angiography data. The classifier predicts a binary class on the hemodynamical significance of a stenosis, using established iFR and FFR thresholds of 0.89 and 0.80, respectively, as reported in (Neumann et al., 2018; Tonino et al., 2009; De Bruyne et al., 2012; Baumann et al., 2018).



Figure 2: Architecture overview. A 3D CNN, common to all views, extracts spatio-temporal features which are subsequently refined through a multi-head spatio-temporal self-attention. The resulting features are then channeled to a three-output layer which carries out binary classification (over the threshold), and quantification of iFR and FFR.

The model is trained using a combination of binary-cross entropy loss (3) for the classification task and L1 loss (4) for the regression task on the continuous values of iFR and FFR.

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3)$$
$$L_{L1} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (4)$$

where y is the true label and \hat{y} is the predicted label. The total loss function, combining both the binarycross entropy loss and the L1 loss, can be represented as:

$$L_{\text{total}} = \alpha L_{BCE} + \beta L_{L1} \tag{5}$$

where α and β are hyperparameters that control the importance of the two loss terms. This dual loss function (5) approach allows the model to effectively learn and predict both binary and continuous outcomes, enhancing its versatility and predictive power.

4 EXPERIMENTAL RESULTS

4.1 Dataset

The proposed method was trained and evaluated using a private dataset of 778 coronary angiographies from 389 patients, gathered between January 2020 and January 2022. The patient group consisted of 303 males and 86 females, with an average age of 67.9 ± 9.61 years. IRB protocol number is 0092163.

The study encompassed patients diagnosed with either chronic coronary syndrome (CCS) or acute coronary syndrome (ACS). For each patient, two Xray angiographies, each from a different view, were available. These angiographies were assessed by two expert cardiologists who conducted invasive physiological evaluations of intermediate coronary stenosis using iFR, FFR, or both.

Specifically, FFR values were available for 251 patients (64.5%), iFR values for 228 patients (58.6%), and both values were provided for a subset of 90 patients (23.1%). For each exam, the major stenosis was identified by radiologists and labeled as hemodynamically significant if the FFR value was less than 0.80 (Neumann et al., 2018; Johnson et al., 2015; Tonino et al., 2009; De Bruyne et al., 2012) or if the iFR value was less than 0.89 (Neumann et al., 2018; Baumann et al., 2018; Davies et al., 2017). Consequently, 93 patients (23.9%) were labeled as positive, resulting in a significant class imbalance.

The coronary angiography and physiological measurements were conducted following standardized clinical practice, and key frames were annotated by two expert cardiologists. The angiographies were collected using different machines and practices, resulting in variations in spatial sizes (ranging from 512×512 to 1024×1024) and frame rates (ranging from 15 fps to 30 fps). To standardize the data, all samples were resized to 256×256 pixels and adjusted to 15 fps, with all collected videos cut to a length of 60 frames, equivalent to 4 seconds.

4.2 Training and Evaluation

We conducted a 5-fold nested cross-validation to estimate the accuracy of the proposed approach and the comparative methods. In each split, we allocated 60%

Methods	Accuracy	AUC	Sensitivity	Specificity
S3D (Xie et al., 2018)	79.5±5.54	0.93 ± 0.03	65.4±14.17	93.6±4.81
MVCNN (Su et al., 2015)	84.5±6.11	0.85 ± 0.07	76.8 ± 10.20	92.2 ± 2.18
GVCNN (Feng et al., 2018)	78.4±4.52	$0.87 {\pm} 0.05$	71.0±7.29	85.8±4.36
DMQCA (Zhang et al., 2019)	81.7±3.69	$0.82{\pm}0.04$	$70.2{\pm}8.00$	$93.2{\pm}1.87$
HEAL (Zhang et al., 2020)	$79.5{\pm}5.18$	$0.84{\pm}0.05$	67.9 ± 9.32	$91.4 {\pm} 3.26$
DMTRL (Xue et al., 2018)	79.1±3.79	$0.85{\pm}0.05$	$67.0{\pm}6.56$	$91.2{\pm}3.48$
Ours	$87.3{\pm}6.14$	$0.93{\pm}0.03$	$82.4{\pm}12.24$	$92.2{\pm}1.79$

Table 1: Comparison of our method with state-of-the-art general deep learning and clinic AI methods.

of the data for training, 20% for validation, and the remaining 20% for testing, maintaining the original label proportion. The input X-ray angiographies were normalized to a range between 0 and 1 and standardized to have a mean of 0 and a variance of 1. Data augmentation was implemented through random horizontal and vertical flipping and random 90-degree rotation, applied identically to all frames. The training process involved minimizing the combination of binary cross-entropy loss for the classification branch, and the L1 losses for the two iFR/FFR regression branches. We used the AdamW optimizer with a learning rate of 1e-5 and a batch size of 8, over a total of 300 epochs. During training, not all samples have both FFR and iFR values, so the two regression branches are not always activated. A specific branch is activated/trained only when the corresponding label is available for a given sample. The experiments were conducted on two NVIDIA Tesla T4 GPUs using automatic mixed precision (amp) training. The proposed approach was implemented using the Py-Torch and MONAI frameworks.

We assess the performance of our method comparing it with both general state-of-the-art deep architectures (S3D (Xie et al., 2018), MVCNN (Su et al., 2015), GVCNN (Feng et al., 2018)) and clinic AI techniques (DMQCA (Zhang et al., 2019), HEAL (Zhang et al., 2020), DMTRL (Xue et al., 2018)) designed for angiography videos, that have been adapted to perform classification. We employed the balanced accuracy to address class imbalance, the area under the Receiver Operating Characteristic (ROC) curve (AUC), as well as sensitivity and specificity. We also assessed the model's performance in terms of regression, i.e., the model's ability to quantify iFR and FFR as continuous values. The metrics used for the regression task were Mean Square Error (MSE) (6) and Mean Absolute Error (MAE) (7)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
(6)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$
(7)

where y_i is the actual value and \hat{y}_i is the predicted value.

4.3 Results

As reported in Table 1, our model shows satisfactory accuracy in determining the hemodynamic significance of coronary stenoses, outperforming both general deep learning models and clinic AI techniques, with an average accuracy score of 87.3, significantly higher than the others: the closest competitor (Zhang et al., 2019) has a notably lower accuracy of 81.7.

In terms of Area Under the Curve (AUC), as shown in Fig. 3 our method yields 0.93, indicating a higher true positive rate for the same false positive rate, which is a desirable characteristic in medical applications. Our method also shows superior performance in terms of sensitivity, with a score of 82.4. This means that our method is better at correctly identifying positive cases. The specificity of our method is 92.2, which is slightly lower than (Zhang et al., 2019) but higher than the other two methods. This indicates that our method is quite good at correctly identifying negative cases.

Additionally, as reported in Table 2 the performance of our proposed model was also satisfactory when FFR and iFR values were treated as continuous variables rather than dichotomous ones.

Table 2: Performance (in terms of MSE and MAE) when regressing FFR and iFR values.

Measure	MSE	MAE	
FFR	$0.060 {\pm} 0.005$	$0.037 {\pm} 0.002$	
iFR	$0.045 {\pm} 0.005$	$0.026{\pm}0.003$	

Fig. 3 reports the comparison in terms of ROC and precision-recall curves between our approach and



Figure 3: ROC (left) and Precision-Recall (right) curves comparison between our approach and state-of-the-art clinical AI methods.



Figure 4: **Impact of attention strategies**. Interpretability maps computed through M3D-cam (Chattopadhay et al., 2018), when using attention and not. For each image, we also report, with an arrow, the major stenosis as identified by clinicians. In each part, the yellow parts are the most activated ones, while the purple areas are the least activated ones. In our model we can see that attention is focused on stenoses or arteries, while the model without attention also targets the morphology of the bones and tissues visible at X-rays.

the state-of-the-art methods specifically designed for stenosis quantification.

In addition, in Table 3 we investigate whether the use of a keyframe for each view, inserting an input branch with a ResNet-50 with late fusion strategy on our model, would lead to performance improvements. Our findings reveal a substantial equality in performance, showing that our approach is autonomously able to identify key information in the video, without the need for manual human interaction.

Finally, we investigate the impact of the employed spatio-temporal attention mechanism. In particular, we evaluate the performance of our model when using a) *global attention*, i.e., each location in the fea-

ture volume attends to all other locations in space and time; b) no attention applied to the CNN extracted features. The comparison is carried out using different attention strategies using interpretability maps, computed through M3D-cam (Chattopadhay et al., 2018). Fig. 4 shows that our spatial and temporal attention is an effective strategy to make the model focus on the stenosis for FFR quantification, demonstrating the importance of both spatial and temporal information in coronary angiographies. When no attention is used the model fails to focus on the major stenosis, thus leading to incorrect and highly uncertain predictions.

Overall, these results suggest that our method pro-

Coronary Artery Stenosis Assessment in X-Ray Angiography Through Spatio-Temporal Attention for Non-Invasive FFR and iFR Estimation

Table 3: Effect of keyframe integration.

Methods	Accuracy	AUC	Sensitivity	Specificity
Ours	87.3±6.14	$0.93{\pm}0.03$	82.4±12.24	92.2±1.79
Ours + keyframe	$87.1 {\pm} 7.08$	$0.94{\pm}0.05$	$82.3 {\pm} 10.94$	$91.9{\pm}1.91$

vides a more accurate and reliable performance compared to the other state-of-the-art methods.

5 CONCLUSIONS

In this work, we presented an approach for the noninvasive evaluation of Fractional Flow Reserve (FFR) and instantaneous wave-free ratio (iFR) from standard coronary angiography, based on a combination of 3D Convolutional Neural Networks and self-attention layers, without the requirement of manual intervention in the identification of a keyframe or vessel region in the video. Our approach provides a reliable evaluation of coronary stenoses without the need for hyperemic flow induction, eliminating risks associated with intracoronary wire passage and reducing additional equipment, training, and procedural costs.

Our model has demonstrated exceptional accuracy and specificity across diverse cases, showcasing its robust performance in hemodynamic evaluation and its potential to enhance both operator and patient access to physiologically guided decision-making, which could have a consequential impact on clinical outcomes and costs.

Future research directions include the employment of specific synthetic data generation techniques, as exemplified by (Pennisi et al., 2023), both for data augmentation purposes and to facilitate secure data sharing while preserving privacy. Moreover, given that a substantial portion of patients in the dataset underwent invasive FFR/iFR for clinical reasons, a potential selection bias towards a relatively high burden of angiographic and functional coronary disease cannot be entirely dismissed. To address this concern, ongoing efforts involve expanding the dataset and refining the model to ensure its robustness across a broader spectrum of patient profiles and clinical scenarios.

ACKNOWLEDGEMENTS

This research was supported by MUR, PRIN 2020, project: "LEGO.AI: LEarning the Geometry of knOwledge in AI systems", n. 2020TA3K9N, CUP: E63C20011250001, and by Piano della Ricerca di Ateneo 2020/2022, Linea 2D, Università di Catania.

Raffaele Mineo is a PhD student enrolled in the National PhD in Artificial Intelligence, XXXVII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma.

REFERENCES

- Au, B., Shaham, U., Dhruva, S., Bouras, G., Cristea, E., Coppi, A., Warner, F., Li, S.-X., and Krumholz, H. (2018). Automated characterization of stenosis in invasive coronary angiography images with convolutional neural networks. *arXiv preprint arXiv:1807.10597*.
- Baumann, S., Chandra, L., Skarga, E., Renker, M., Borggrefe, M., Akin, I., and Lossnitzer, D. (2018). Instantaneous wave-free ratio (ifr®) to determine hemodynamically significant coronary stenosis: a comprehensive review. World Journal of Cardiology.
- Carreira, J. and Zisserman, A. (2017). Quo vadis, action recognition? a new model and the kinetics dataset. In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6299–6308.
- Chattopadhay, A., Sarkar, A., Howlader, P., and Balasubramanian, V. N. (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE.
- Cong, C., Kato, Y., Vasconcellos, H. D., Lima, J., and Venkatesh, B. (2019). Automated stenosis detection and classification in x-ray angiography using deep neural network. In *IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE.
- Danilov, V., Gerget, O., Klyshnikov, K., Ovcharenko, E., and Frangi, A. (2020). Comparative study of deep learning models for automatic coronary stenosis detection in x-ray angiography. In *Proceedings of the* 30th International Conference on Computer Graphics and Machine Vision.
- Davies, J. E., Sen, S., Dehbi, H.-M., Al-Lamee, R., Petraco, R., Nijjer, S. S., Bhindi, R., Lehman, S. J., Walters, D., Sapontis, J., et al. (2017). Use of the instantaneous wave-free ratio or fractional flow reserve in pci. *New England Journal of Medicine*, 376(19):1824–1834.
- De Bruyne, B., Pijls, N. H., Kalesan, B., Barbato, E., Tonino, P. A., Piroth, Z., Jagic, N., Möbius-Winkler, S., Rioufol, G., Witt, N., et al. (2012). Fractional flow reserve–guided pci versus medical therapy in stable coronary disease. *New England Journal of Medicine*.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is

worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

- Feng, Y., Zhang, Z., Zhao, X., Ji, R., and Gao, Y. (2018). Gvcnn: Group-view convolutional neural networks for 3d shape recognition. In *CVPR*.
- Han, T., Ai, D., Li, X., Fan, J., Song, H., Wang, Y., and Yang, J. (2023). Coronary artery stenosis detection via proposal-shifted spatial-temporal transformer in x-ray angiography. *Computers in Biology and Medicine*.
- Johnson, N. P., Johnson, D. T., Kirkeeide, R. L., Berry, C., De Bruyne, B., Fearon, W. F., Oldroyd, K. G., Pijls, N. H., and Gould, K. L. (2015). Repeatability of fractional flow reserve despite variations in systemic and coronary hemodynamics. *JACC: Cardiovascular Interventions*, 8(8):1018–1027.
- Knuuti, J. and Revenco, V. (2020). 2019 esc guidelines for the diagnosis and management of chronic coronary syndromes. *European heart journal*, 41(5):407–477.
- Ma, H., Ambrosini, P., and van Walsum, T. (2017). Fast prospective detection of contrast inflow in x-ray angiograms with convolutional neural network and recurrent neural network. In *MICCAI*.
- Moon, J. H., Cha, W. C., Chung, M. J., Lee, K.-S., Cho, B. H., Choi, J. H., et al. (2021). Automatic stenosis recognition from coronary angiography using convolutional neural networks. *Computer methods and programs in biomedicine*, 198:105819.
- Neumann, F.-J., Sousa-Uva, M., Ahlsson, A., Alfonso, F., Banning, A. P., Benedetto, U., Byrne, R. A., Collet, J.-P., Falk, V., Head, S. J., Jüni, P., Kastrati, A., Koller, A., Kristensen, S. D., Niebauer, J., Richter, D. J., Seferović, P. M., Sibbing, D., Stefanini, G. G., Windecker, S., Yadav, R., Zembala, M. O., and Group, E. S. D. (2018). 2018 ESC/EACTS Guidelines on my
 - ocardial revascularization. European Heart Journal.
- Ovalle-Magallanes, E., Avina-Cervantes, J. G., Cruz-Aceves, I., and Ruiz-Pinales, J. (2022). Hybrid classical-quantum convolutional neural network for stenosis detection in x-ray coronary angiography. *Expert Systems with Applications*, 189:116112.
- Pennisi, M., Salanitri, F. P., Bellitto, G., Palazzo, S., Bagci, U., and Spampinato, C. (2023). A privacy-preserving walk in the latent space of generative models for medical applications. In *MICCAI*.
- Proietto Salanitri, F., Bellitto, G., Irmakci, I., Palazzo, S., Bagci, U., and Spampinato, C. (2021). Hierarchical 3d feature learning forpancreas segmentation. In *MLMI* (*MICCAI workshop*).
- Rodrigues, D. L., Menezes, M. N., Pinto, F. J., and Oliveira, A. L. (2021). Automated detection of coronary artery stenosis in x-ray angiography using deep neural networks. arXiv preprint arXiv:2103.02969.
- Salanitri, F. P., Bellitto, G., Palazzo, S., Irmakci, I., Wallace, M., Bolan, C., Engels, M., Hoogenboom, S., Aldinucci, M., Bagci, U., et al. (2022). Neural transformers for intraductal papillary mucosal neoplasms (ipmn) classification in mri images. In *EMBC*.
- Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E. (2015). Multi-view convolutional neural networks for 3d shape recognition. In *ICCV*.

- Tomar, N. K., Jha, D., Bagci, U., and Ali, S. (2022). Tganet: Text-guided attention for improved polyp segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 151–160. Springer.
- Tonino, P. A., De Bruyne, B., Pijls, N. H., Siebert, U., Ikeno, F., vant Veer, M., Klauss, V., Manoharan, G., Engstrøm, T., Oldroyd, K. G., et al. (2009). Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *New England Journal of Medicine*, 360(3):213–224.
- Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., and Paluri, M. (2018). A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of* the IEEE conference on Computer Vision and Pattern Recognition, pages 6450–6459.
- Valanarasu, J. M. J., Oza, P., Hacihaliloglu, I., and Patel, V. M. (2021). Medical transformer: Gated axial-attention for medical image segmentation. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24, pages 36–46. Springer.
- Wu, W., Zhang, J., Xie, H., Zhao, Y., Zhang, S., and Gu, L. (2020). Automatic detection of coronary artery stenosis by convolutional neural network with temporal constraint. *Computers in biology and medicine*, 118:103657.
- Xie, S., Sun, C., Huang, J., Tu, Z., and Murphy, K. (2018). Rethinking spatiotemporal feature learning: Speedaccuracy trade-offs in video classification. In *Proceedings of the European conference on computer vision (ECCV)*, pages 305–321.
- Xue, W., Brahm, G., Pandey, S., Leung, S., and Li, S. (2018). Full left ventricle quantification via deep multitask relationships learning. *Medical image analysis*.
- 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*.
- Zhang, D., Yang, G., Zhao, S., Zhang, Y., Zhang, H., and Li, S. (2019). Direct quantification for coronary artery stenosis using multiview learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 449–457. Springer.
- Zhao, C., Tang, H., McGonigle, D., He, Z., Zhang, C., Wang, Y.-P., Deng, H.-W., Bober, R., and Zhou, W. (2021a). A new approach to extracting coronary arteries and detecting stenosis in invasive coronary angiograms. arXiv preprint arXiv:2101.09848.
- Zhao, C., Vij, A., Malhotra, S., Tang, J., Tang, H., Pienta, D., Xu, Z., and Zhou, W. (2021b). Automatic extraction and stenosis evaluation of coronary arteries in invasive coronary angiograms. *Computers in Biology* and Medicine, 136:104667.