Ecole Polytechnique Fédérale de Lausanne

Master project, 2021-2022, Mathematics section

Myocardial infarction prediction from angiographies based on deep learning

Pierre Vuillecard

Supervisors: Dorina Thanou, Ortal Senouf

Professor: Pascal Frossard

LTS4, EPFL

Lausanne, January 21, 2022

Abstract

Stenosis is a coronary artery disease that can lead to heart attack or myocardial infarction (MI). It is one of the main cause of death in the past two decades. Predicting whether a stenosis will lead to a heart attack or not is crucial for modern society. The signal of the vessel surrounding the stenosis might give valuable information on the stenosis condition. Computer-aided diagnosis systems gained in popularity since they could help and assist experts in critical decision-making. Previous researches developed several methods to detect significant stenosis in X-ray coronary angiography image. They demonstrated excellent performances based on classification from extracted vessels patches or based on detection from full X-ray images, via object detection models. In this thesis, a special effort is put to evaluate the prediction of future MI based on extracted vessel segment from X-ray coronary angiograms. Doctors have annotated angiography images for this work. It brings many challenges that need to be addressed as the limited amount of data, the class imbalance, the annotation specificities, and the qualities of the image. In order to understand if there is some predictive signal in the angiograms, we approach the problem from different angles. First, we predict MI from segment extracted from the annotated angiograms. Then, we try to simultaneously detect the segment and predict if it will lead to an MI. Next, we try to classify segments but taking into account not only the segment information but the entire image, with an attention map on the segment level. Finally, we simplify the problem by exploiting the exact position of the stenosis on the vessel, and predict MI from patches extracted around the stenosis. The different techniques demonstrated poor predictions of vessels leading to a MI. Based on a clinical study of 469 patients, out of which only 58 had MI, we present evidence that the signal of vessels responsible for heart attacks might not be sufficient to make a clear distinction between a severe and benign stenosis inside the vessels.

Contents

1 Introduction

1.1 Objectives and challenges

Coronary artery disease is the most common cardiovascular disease noa (2015), and the leading cause of death in the world during the past two decades Mendis et al. (2011). Heart attack or myocardial infarction (MI) is a consequence of this disorder: it happens in case of hypoxia in the heart tissue. Oxygen can't reach the heart muscle properly via the bloodstream because the latter is blocked. A specific type of coronary artery disease might cause this pathology, called stenosis. In fact, stenosis is an unnatural narrowing of coronary arteries over time, which can partially or entirely block out the bloodstream Lampros et al. (2017). Therefore, the diagnosis and prevention of stenosis are crucial for modern society. Coronary angiography image is the standard method to estimate the severity of coronary artery stenosis. However, the variability of diagnoses allows the automatic computer-aided diagnosis (CAD) systems to play a vital role in cardiology to detect stenosis. Thus, the objective of this work is to use modern computer vision to classify at the vessel level, whether stenosis will lead to an MI or not.

Currently, X-ray coronary angiography (XCA) remains the gold-standard imaging technique for the medical diagnosis of stenosis and other related conditions. During the process, a liquid dye, such as fluorescein, is injected into the blood vessels of the heart. As the dye diffuse in the cardiovascular system, the structure of the system is revealed with X-rays. From the images, interventional cardiologists are able to detect narrowed or blocked areas through coronary arteries as shown in figure 1. The goal of this work is to classify whether a stenosis detected in XCA leads to an MI or not using modern machine learning techniques, based on deep learning.

For this reason, annotated XCA images were collected with the collaboration of the CHUV, Lausanne. In total this study counts 469 patient, with only 58 that had MI and with a total of 2694 images with 105 that contain MI vessel segment. MI patients have already been treated resulting in revacularized vessel segment in the XCA images and MI vessels segments are detected because it contain severe stenosis that might lead to an MI. Doctors annotated XCA images indicating, for specific parts of the vessels, whether it led to an MI or not. An example of an annotated XCA is presented in figure 1. In the figure, two annotations can be seen: the boxes delimiting the vessel segment location and the dots corresponding to the box label.

In this work, the input will be the boxes since we are interested in the stenosis consequence at the vessel level. In fact, a hypothesis is that the vessel segment around the stenosis could bring valuable information for the model to detect whether it will lead to an MI or not. The target in this experiment is the condition of the vessel, MI or non MI. Thus, the objective is slightly different from standard stenosis detection as Ovalle-Magallanes et al. (2020).

The above mentioned dataset poses many challenges from the machine learning point of view. Indeed, as many medical studies Shin et al. (2016), the dataset is imbalanced as the number of MI vessels segments are less common. This is a challenge for a supervised learning approach since models tend to overfit the over represented classes Garg et al. (2020).

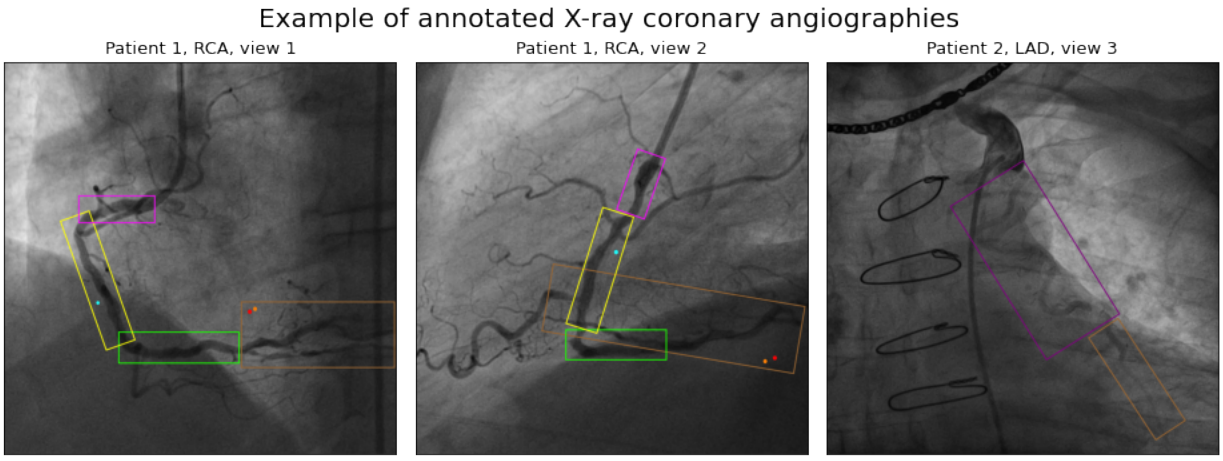


Figure 1: Example of annotated X-ray coronary angiographies for different patients. Two annotated images are shown for *Patient 1*. Both refer to the same heart segment *RCA* with different views. As the brown boxes contain a red dot, it means that this segment contains stenosis that leads to an MI. *Patient 2* exhibits bad quality images. It also shows medical tools that might be inside the patient body.

Moreover, the amount of available data are limited and it is known that deep learning model needs an important number of data for good performances as it could easily suffer from overfitting since the number of parameters is large. Furthermore, the configuration of the problem and the anatomies of vessel segment involve various annotated boxes sizes and aspect ratios. The heterogeneity in the box dimension makes the input format difficult to work with at a vessel segment level.The data quality could be different among the XCA images, as shown figure 1. The contrast of the vessel could be different and some images could be blurry as demonstrated *Patient 1* and *Patient 2* in figure 1. Also, some patients have already been treated, thus the images might contain medical tools like stents, shown in *Patient 2*. The patients could have some revascularisation vessel segment that could affect the vessel quality. All of these reasons could add noise to this data and reduce the data quality.
To sum up, the objective is to classify at the box level whether it will lead to an MI or not. Our hypothesis is that at the vessel level, valuable information could be used in order to predict if, the stenosis contained inside the vessel segment, will lead to an MI. The annotated XCA images bring a few challenges:

- CH1: Class imbalance
- CH2: Limited amount of data
- CH3: Boxes sizes heterogeneity
- CH4: Data quality

Challenges CH1, CH2, and CH4 are known in medical imagery studies as data acquisition is difficult, and the quality of annotations is costly Shin et al. (2016). However, challenge

CH3 is understudied and brings an important issue for our task. This work develops and experiments ideas trying to solve or reduce the effect of those challenges.

1.2 Related work

In this section, an overview of previous studies on deep learning methods applied to angiograms is presented. In order to reduce the variability of diagnoses, CAD systems have emerged as a solution. It started to interest the scientific communities and they tried different approaches like in Vidya et al. (2015) where they use statistical analysis on 2D echocardiography images.
In the literature, several methods for detecting coronary stenosis in XCA images were developed. The first studies were focused on classical image processing techniques. The aim was to measure the width of the vessel: abnormal large vessels were classified as stenosis. Then, with the development of convolutional neural network (CNN), CNN became a gold standard in images classification in many applications including medical imagery Shin et al. (2016). CNN is a powerful deep learning method in the image processing field, which can automatically learn the hierarchical feature representations from data. Regarding stenosis classification a study Ovalle-Magallanes et al. (2020) showed that they could achieved accurate scores on extracted patches from XCA images. They had 32×32 patches from stenosis and other non stenosis patches. They used a combination of 10,000 artificial artery images produced by a generative model and 250 real patches. The classes were balanced and they split in half for train and test. They used a pre-trained ResNet50 CNN He et al. (2015) on ImageNet and achieved an F1-score of 0.91 on real data and 0.92 using artificial artery images. One could notice that the real stenosis showed in the report were really visible, the arteries had clear narrowing and there is no ambiguity concerning the presence of a stenosis. Thus, this study showed that stenosis classification is feasible and even demonstrated decent performances. They also exposed that transfer learning using ImagNet on artery patches improved the accuracy by 12%. Additionally, it exhibited that we can use a deep learning model on limited amount of data which is CH2 in section 1.1.
The same idea was used in the work of the previous semester project Thanou et al. (2021). The task was slightly distinct: the objective was to classify whether a stenosis led to an MI or not from an extracted patch centered around the stenosis. The stufy consisted respectively of 522, 142, and 75 training, validation, and test set patches, extracted from 83 patients. The patches were imbalanced, the MI patches represented 28%. The authors used a ResNet18 architecture for the CNN. They experimented different techniques to deal with the imbalanced data. The best configuration was to balance the dataset during training, they reached an F1-score of at most 0.4 using a balance dataloader and 0.7 by augmenting the MI class using transformation (rotation, flipping, blurring). Note that 0.7 was tested on a balanced dataset. Various other ideas were tested without major improvement including Frangi filter and self-contrastive learning approach. In this work, they had to deal with CH1, CH2, and CH4. It seems that in order to deal with class imbalance CH1, balance dataloader and augmenting the less represented class are the best options. Here, all patches contained stenosis, it seems that a signal exists, that might distinguish an MI stenosis from a non MI

stenosis since the model reaches a relatively good performance. Recall that our hypothesis is that at the vessel level, valuable information could be used in order to predict if, the stenosis contained inside the vessel segment, will lead to an MI. Thus, we hope that our task on vessel segment could improve those results.

The previous works mainly focused on the analysis at the stenosis level using patches to extract the stenosis information from the XCA images. Other works, used the full XCA images to detect stenosis. In fact, giving as input the full images to a model could bring valuable information to detect the stenosis. This could also solve the challenge CH3 of boxes sizes heterogeneity. Three studies caught our attention, Moon et al. (2021), Du et al. (2020), and Danilov et al. (2021).

The first one Moon et al. (2021) used a CNN architecture from the original XCA images to predict if the heart segment contained a stenosis. The task is very similar to the first paper Ovalle-Magallanes et al. (2020) but at the image level and not using patches. 452 coronary artery angiography movie clips were labeled as significant stenosis (abnormal) and non significant stenosis (normal). The dataset was balanced and they used the five key frames to train the model to classify the images, generating around 2,260 images. Using a five random cross validation, they reached an accuracy of 0.93 and using an external dataset, they obtained an accuracy of 0.887. The results are comparable to the one in Ovalle-Magallanes et al. (2020), thus working at the original image level doesn't seem to be an issue. Additionally, this study also showed that challenge CH2 is not an issue for their data, they also exhibited that augmentation could improve the performance and thus augmentation can be a solution for challenge CH2.

Then, Du et al. (2020) tried to detect lesion types and locations using object detection techniques. Lesion types included significant stenosis detection. They had 7,239 annotated lesions with 1,315 stenosis (location surrounded by bounding box and lesion classes). They used a model architecture similar to a U-net Ronneberger et al. (2015) followed by a region proposal network (RPN) to finally predict the class and the location of the bounding box. They selected an intersection over union (IoU) of 0.5. They obtained an F1-score of 0.82, a recall of 0.91, and a precision of 0.77 for the stenosis lesion. With the same idea, Danilov et al. (2021) explored the detection of significant stenosis using object detection. They have 8,325 XCA images annotated with the location of the stenosis surrounded by a bounding box. They tried different standard object detection architectures including faster RCNN with a ResNet50 as backbone CNN. They reached an F1-score of 0.88 for the ResNet RCNN and a mean average precision for an IoU of 0.5 (mAP@0.5) of 0.92. Both studies showed promising performances. Using object detection techniques seems to be an option for our task. In the setting of object detection, we would have two classes: one for non MI bounding box and one for MI bounding box. It could partially solve challenge CH3 because it works at the original image level but the box sizes to detect are still very heterogeneous which might be an issue for the box coordinate regression. CH1 is still a major problem in this case because in both studies they had around 8,000 images.

To conclude, all these studies showed promising performances. Our task is different from those studies except in Thanou et al. (2021), as they aimed at classifying stenosis and not

predicting MI. Note that in all of these studies, except in Thanou et al. (2021), the data quality is satisfying. As a matter of fact, all the stenosis were significant and clearly visible, thus they do not find solutions for challenge CH4. Two approaches were established: classify from extracted patches as in Thanou et al. (2021) and Ovalle-Magallanes et al. (2020) studies or from the original images as Moon et al. (2021), Du et al. (2020), and Danilov et al. (2021). The second approach seems more promising regarding challenge CH3. Otherwise, CH2 doesn't seem to be an issue in all of these studies but augmentation showed improvement for limited data. Then, for CH1, only Thanou et al. (2021) work deals with it, balance dataloader and augmenting the MI class seem to be great solutions in order to make the data more balanced. However, our dataset is more challenging because MI stenosis are labeled from doctors assumption about the actual condition of the stenosis and not from clinical history of the patient as in Thanou et al. (2021).

1.3 Contribution

The objective of this work is to predict if a vessel segment leads to an MI in the next five years. However, we saw in section 1.1 that this dataset comes with many different challenges. This work mainly focuses on the 58 MI patients and develops three main approaches in order to overcome these challenges. The first approach is focusing on patches classification where patches are extracted from the annotated boxes in the XCA images. The second approach is about detecting the annotated boxes locations and classes at the XCA images level. Then, the third approach is trying to focus the model attention on the boxes to classify, at the XCA images level, whether the highlighted vessel segment leads to an MI or not. Finally, an MI stenosis classification from extracted patches centered around stenosis in our annotated data is performed to estimate the qualities of the stenosis. All the approaches demonstrate the complexities and challenges that come with the data. The last approach suggests that the signal from MI stenosis might not be sufficient to make a distinction between benign and sever stenosis, adding another annotation level would be required to confirm this hypothesis.

2 Clinical study

In this section, data exploration is performed to better understand the input images and the challenges that they bring. For each patient involved in the study, the doctors annotated three main coronary arteries (RCA, LDA, and LCX), each included from one to four different views. They put boxes and dots on the XCA images. Each box color corresponds to a specific vessel segment, there are in total six vessels segments. Then, each dot color corresponds to the condition of the vessel, there are three colors, seen in figure 1. Orange and blue are vessel segments that had a revascularization. Red corresponds to a stenosis in the vessel segment that might lead to a future MI. The target in this work is defined as follows: a box containing a red dot corresponds to MI class and a box without it, to the non MI class. It is important to note here that there might exist a vessel segment inside a box that contains a stenosis but is classified as non MI since doctors judged that the stenosis was not bad

enough to lead to an MI.
In table 1, we can see that the number of input is 10,901 for all patients and 1,192 for MI patients. In the same table, a statistical summary of the annotated images shows the class imbalance between non MI and MI. In fact, only 12.4% of the annotated patients had an MI. By looking at the target, the ratio of MI box in all patients is 1.3% and 12% in MI patients. In this study, we only focus on the MI patient since the class imbalance is too important with all the patients. Thus, we can see that the class imbalance challenge (CH1) and the limited amount of data (CH2) are two challenges that need to be addressed. Additional

Annotation	Patients	Images	Boxes	Ratio MI box
All	469	2694 (105)	10901 (143)	1.3 %
Only MI	58	307 (105)	1192 (143)	12 %

Table 1: Summary of the annotation at different levels. *All* concerns all the annotated patients and *Only MI* are the patients that had an MI. Numbers in parenthesis () correspond to the number of MI (a red dot in the annotation).

challenge is the heterogeneity of the boxes size and aspect ratio (CH3). Figure 2 visually exhibits the distribution of the boxes dimension for the six colors. The average box size for each color varies significantly, magenta boxes are on average much smaller than brown boxes. Even per color, there is a large variance, like in the case of green and brown boxes. One can also notice that the number of boxes is equally distributed for magenta, yellow, green, and brown boxes, but blue and magenta dark boxes are less represented in the data. Additionally, the MI boxes are not equally distributed among the colors. Indeed, magenta and brown boxes are less likely to lead to an MI than blue boxes which have almost 50% chance to lead to an MI. Moreover, the dimension of the images varies from 1524 to 1133 pixels but the majority of the images have a size of 1524 pixels. Image quality is another challenge we have encountered throughout this study (CH4). It is a very common problem when working on real clinical imaging data. As mentioned in section 1.1, the XCA images included in this data are of poor quality in terms of vessel contrast and blurriness. This is also affected by artifacts caused by medical tools, like stents.

3 Experimental settings

The experimental setup described in the following subsection is unified and shared among all experiments.

3.1 Data splitting and evaluation metrics

In all of the experiments, we defined a common training approach for the purpose of comparing them. Seven patients were excluded from the train set to be used as an out-of-sample test set. The different trained models are tuned and evaluated with a patient-level five-fold cross-validation protocol (CV-5). It consists of splitting the MI patients samples into five

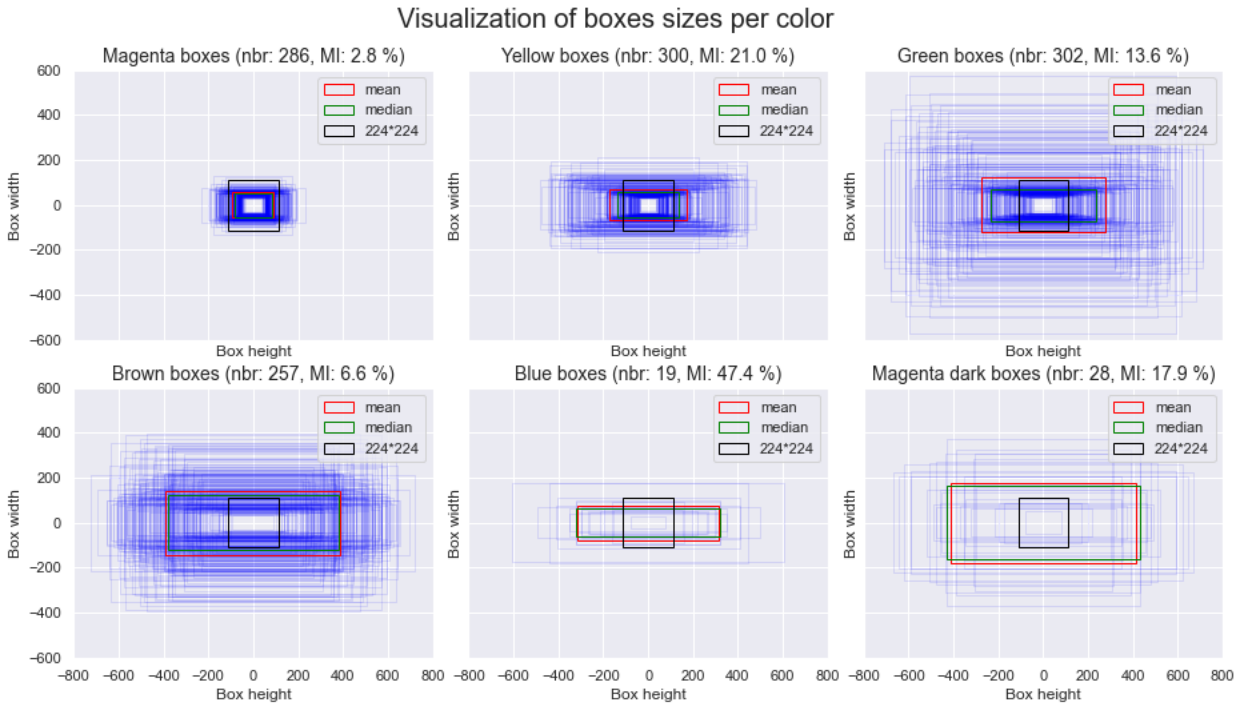


Figure 2: Visualization of boxes sizes for each color. *nbr* refers to the number of boxes in the XCA images annotation. *MI* is the number of MI boxes among the specific color. In this plot only the boxes of MI patients are shown. As a reference a box of size 224×224 (ImageNet input size Russakovsky et al. (2015)) is shown to compare with the current inputs. The red and green boxes show respectively the average and median dimension of boxes per color.

folds. Since each patient contains approximately the same number of MI vessels segments, each fold contains the same number of MI vessel. Then, by iterating over the folds, one fold is selected as the validation set per iteration, the other folds are used as a train set. The model is trained and validated for the five iterations. At the end, the average performance metrics are computed over the folds to get a statistically significant performance analysis. Cross validation can also estimate the variance of the performance, which is interesting especially when dealing with large number of model parameters and small training data. In this work, a cross validation at patient-level is performed because images from the same patient present in both training and validation sets could add a bias to the performance analysis Shin et al. (2016). Note that sometimes a patient-level three-fold cross-validation protocol is used in case the model requires a large computational time.
In this work, we chose multiple metrics to evaluate the performance of the models. As the data is unbalanced (section 1.1), the accuracy is not a valid metric to use Garg et al. (2020). Thus, the F1 score was chosen to estimate the performance since it is an harmonic mean between precision and recall. Garg et al. (2020). We calculated additional metrics in order to have a clear understanding of the performances and biases of each model:

- Precision: the fraction of relevant instances among the retrieved instances, therefore the number of true MI among the predicted MI.
- Recall: the fraction of relevant instances that were retrieved, therefore the total number of MI retrieve by the model.
- Area under the curve (AUC): the tradeoff between precision and recall for different decisions thresholds. This is a visual summary of both precision and recall. The AUC is the area under the precision-recall curve, a high area under the curve represents both a high recall and a high precision.
- Specificity: the number of non MI retrieved by the model.

Both precision and recall are important but in the context of MI prediction, a high recall means that the model does not miss a stenosis that could lead to an MI, causing the death of a patient.
In the object detection experiments, the metric used is slightly different since the location of the predicted bounding box plays an important role. The usual metric used in this field is the intersection over union (IoU), a measure between 0 and 1 that computes the overlap between two boxes where 1 indicates a perfect overlap. In the context of object detection, it measures the overlap between the predicted and the ground truth bounding boxes. The threshold to differentiate correct and wrong prediction is arbitrary. Then, we can define the mean average precision (mAP) based on a fixed IoU threshold. The mAP is the average of the AUC for the precision and recall curve for each class. Then, mAP@0.5 refers to the average of AUC for the precision and recall curves for each class regarding boxes that overlap from a least 0.5. Finally, mAP@0.25:0.05:0.75 is the average mAP for different IoU thresholds from 0.25 to 0.75. Similarly, mAR is the average recall.

3.2 Training strategies

In this section, the model selection is presented along with three different training strategies in order to deal with challenge CH3 of class imbalance defined in section 1.1.
In all the different experiments elaborated in the following sections, hyperparameters tuning is conducted through patient-level five-fold cross validation. The selected final configuration is the one that yields the best mean F1-score. Then the model is trained on the MI patients train set with the selected hyperparameters.
The first approach used to tackle the class imbalance challenge CH3 is a weighted cross entropy loss Phan and Yamamoto (2020) defined as the following :

$$Loss(x, class) = -\frac{\sum_{i=1}^{N} w_{class_i} log\left(\frac{exp(x_{i,class_i})}{\sum_{j=1}^{C} exp(x_{i,j})}\right)}{\sum_{i=1}^{N} w_{class_i}}$$

where $x \in \mathbb{R}^{NC}$, $class \in \{1, ..., C\}^N$, and $w \in \mathbb{R}^C$. N is the number of elements in a batch and C is the number of classes, which is 2 in this work. It is a weighted average over the

cross entropy per element in the batch. With this method, it is possible to give more importance to a specific class. It seems promising for unbalanced data since one can favor the less represented class. We set the weight to be inversely proportional to the class representation $w = [N/C_1, N/C_2]$ where here N is the number of data and C_i is the number of data from class i.
Another approach concerns the way the data is sampled every batch. In fact, because the data are not balanced, we can over-sample randomly the less represented class at every batch in order to make the batch balanced. With this method, some elements of the less represented class are seen multiple times during an epoch. This method is referred to as balance dataloader throughout this report.
The last approach consists of augmenting the less represented class to make the data balanced. To augment the data, we use different transformations that do not change the overall aspect of the images. We use horizontal and vertical flip as well as a 90 degree rotation. We use augmentation on the fly, which means that every time an image of the less represented class is loaded, a transformation can be applied with a probability of 0.5. Per batch, we over-sampled the less represented class and randomly applied a transformation to it. Thus, within each epoch, an image from the less represented class should be only seen once because it might be transformed at random. Augmenting the MI class also helps with the challenge of limited data (CH2) since we add new versions of the images that the model has never seen before.

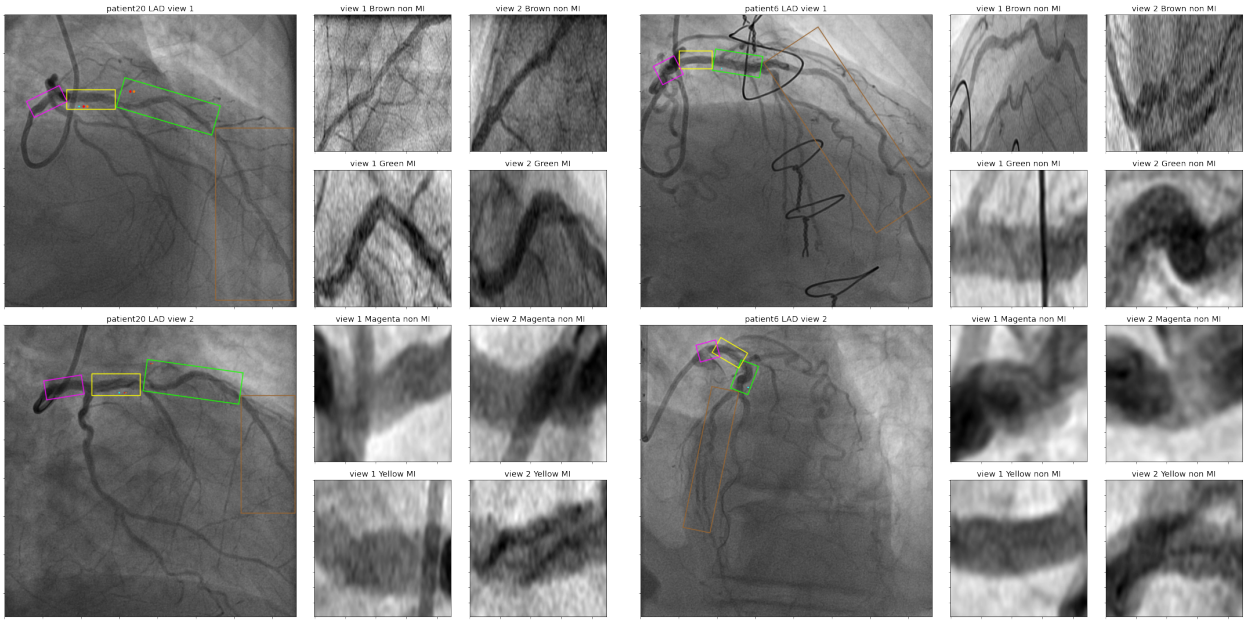
4 MI prediction: a segment classification approach

In this section, similarly to Ovalle-Magallanes et al. (2020) and Thanou et al. (2021), this approach deals with patches but patches are extracted from the anatomical segments defined by the boxes in the XCA image and then are classify as MI or non MI.

4.1 Data configuration

The doctors annotated the original images by boxes for the vessel location and dots for the vessel label. For patches classification, the boxes need to be extracted from the original annotated images along with the associated label to define the non MI and MI classes. An algorithm that detects the boxes and dots coordinates were developed. Then, using the coordinates of the box and if it contains a red dot or not in the annotated images, patches are extracted with the associated class. It results in 1,192 MI patients with only 143 patches that contain an MI stenosis. This approach does not solve CH1 and CH2 (section 1.1), as the class imbalance is maintained across the number of patches in each class.
Then, the extracted patches have a very heterogeneous size and aspect ratio as it was presented in section 1.1, figure 2. It refers to the challenge CH3. A CNN architecture could work with various input size, but to train it to solve a particular task, it is better if the input dimension is fixed. Common preprocessing methods are resizing or zero-padding Hashemi (2019). However, the difference in size between the larger and the smaller patches is very

large thus, resizing seems to be the best way to format the input dimension to a fixed size. Thus, we decided to resize the patches to a fixed size of 224×224 which is a standard input size for CNN, as in the ImageNet dataset. Figure 3 shows an example of the procedure of patches extraction and resizing, with examples of patients with two different views from the same heart segment (LAD).
Additionally, challenge CH4, referring to data quality, is still an issue. We mentioned, in section 1.1, that some medical tool artifacts could be seen in some XCA images and can tamper images quality. Therefore, for each patient, we choose to give as input of the learning algorithm two different views of the same heart segment. Indeed, each vessel segment has different views and medical tool artifacts might not be present in all of the views. For example in figure 3b, one can see the patches extraction for a patient with a medical tool artifact. The artifact is not present in both views, see the brown and green patches for an example. Therefore, with this approach, we hope to limit the impact of the medical tool artifacts present in the images.



(a) Patient with MI stenosis.

(b) Patient with medical tool artifacts

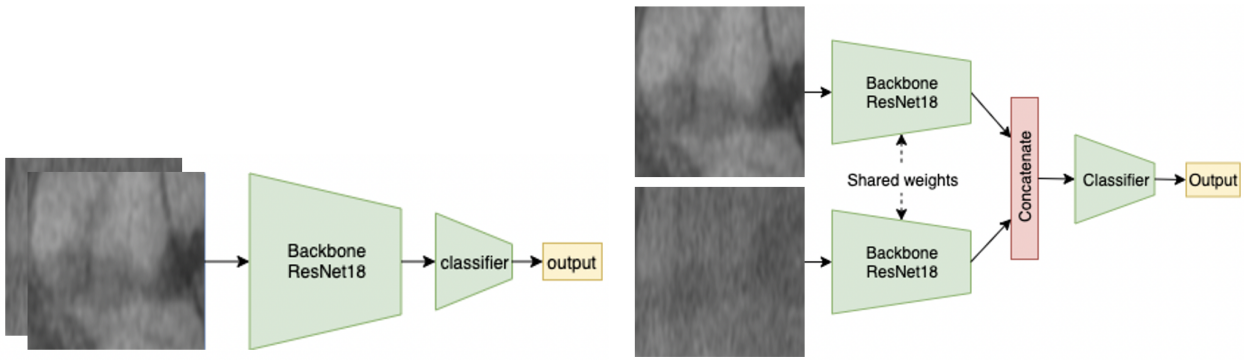
Figure 3: Visualization of patches extraction per view. For the same patient and same heart segment but the different view, we extract the boxes from the original images and then resize it to 224×224 image.

Nevertheless, figure 3 exhibits the limitation of resizing the patches. Indeed, as the patches have very different scales, resizing small images results in blurry images. Additionally, when patches have a large width over height ratio, resizing the patches distorts the images and affects the quality of the images.
To sum up, our first approach is to learn a model where the inputs are pairs of patches from different views of the same patient and heart segment, and the target output is MI or

non-MI prediction. Patches are resized to a fixed 224×224 size. Patches are split between a train and a test set. The seven patients selected for the test set are extracted from the MI patients. The test set contains 72 non MI pairs of patches and 11 MI pairs of patches. The train set contains 515 non MI and 70 MI pairs of patches. The train set is selected to run the patient-level five-fold cross-validation.

4.2 Model and training

In this section, the models are introduced. For this application, a CNN architecture seems to be the most natural choice because it can automatically learn the hierarchical feature representations from data He et al. (2015). As stated in section 1.2, ResNet is the model that has been often used in all previous studies. In fact, the depth of representations is of central importance for many visual recognition tasks and ResNet allows deep architecture thanks to identity pass-through connection to keep a structured gradient and distributed representation He et al. (2015). Here two slightly different architectures are proposed: the first architecture refers to the standard CNN architecture, as seen in figure 4a. The two patches are given as one input with two channels to the ResNet18. Usually, ResNet18 for Imagnet classification works with three inputs channels for RGB input images. Here, we set the input channel to two. At the end of the ResNet18, an average pool layer transforms the feature map into a flat tensor, before being fed to a linear classifier. In the second architecture, each patch is treated separately through two ResNet18 with shared weights, as shown in figure 4b. This architecture is called a Siamese neural network and already demonstrates its strength in computer vision Koch et al.. In the end, similarly to the previous architecture, an average pool layer flattens the feature map before the results of the two parallel branches are concatenated together. Finally, a linear classifier is used at the end. In this model, we duplicate it three times to be in the RGB format for simplicity.



(a) Standard CNN (b) Siamese CNN

Figure 4: Two different architectures that deal with two patches as input.

Regarding the initialization and optimization, three different strategies are tested. The first one is to optimize the model parameters from scratch using the default PyTorch parameter initialization. The second one is to initialize the weights of the ResNet18 from a pre-trained

ResNet18 on ImageNet. Then, the final strategy is similar but uses the weights of the best Resnet18 pre-trained on stenosis patches from Thanou et al. (2021) work. Indeed, Shin et al. (2016) showed the potential of transfer learning from pre-trained ImageNet via fine tuning model on medical images, which improves the model's performances. Model pre-trained on stenosis patches should yield even better results because it has been trained on similar data. Thus, it should mitigate the challenge CH4. Concerning the optimizer, we chose an SGD optimizer. When we fine tuned the model with pre-trained weights, the learning rate for the base parameter (backbone parameter) is decreased by a factor of 10, with respect to the transfer learning strategy defined in Shin et al. (2016). In this classification task, a cross entropy loss is used. Furthermore, four different experiments during the training are tested as presented in section 3.2 that help for challenges CH1 and CH2.
To sum up, we presented two slightly different models for MI prediction from segments. Three promising parameter initialization and four different experiments will be used in order to deal with the imbalanced data. There are in total 24 configurations to train and test with five fold cross validation.

4.3 Results

The models and training strategies are designed to mitigate the effect of data imbalance and partially the limited amount of data CH1, CH2.
The first architecture tested is the standard CNN. The baseline experiment presented in figure 5 demonstrate poor performance. Indeed, the testing evaluation shows clear overfitting because the validation accuracy does not increase and is lower than 0.88, which is the proportion of non MI in the data. However, the training shows good performance that demonstrate a low bias but high variance. The model overfits the non MI class as demonstrated by the specificity, which is almost 1 every epoch, and the recall that is almost zero. The variance that we can observe in the accuracy validation by the large band can be explained by the low number of data and a large number of parameters in the model. It is also worth noting that the proportion of MI class in each validation fold is not perfectly equal due to the patient splits increasing a bit the variance. Then, the second experiment try to give more importance on the MI class to attenuate the overfitting using the weighted loss approach explained in section 3.2, the results are almost similar to the baseline experiment. The model still overfits the non MI class, as exhibited in figure 6. Different weights were tested giving more importance toward the MI class but the model started to overfit the MI class. Therefore, there is no evidence that a weighted loss helps the model to learn a significant separable representation. The next experiment uses the balance dataloader approach presented in section 3.2. This experiment should also reduces the overfitting toward the non MI class because it makes the data balance in the training. In figure 7, the summary of the performances is presented. During the training, the model achieves an accuracy of 1. However, the validation accuracy increases and reaches a plateau of around 0.83. The model is still overfitting the non MI class because the specificity increases straight to 0.9 with a 0.1 confidence interval and the recall decreases 0.15 with a 0.15 confidence interval. Thus, the model is wrong in around 10% of non MI classes and finds around 10% of MI classes.

However, the F1 score is decreasing with a large variance.

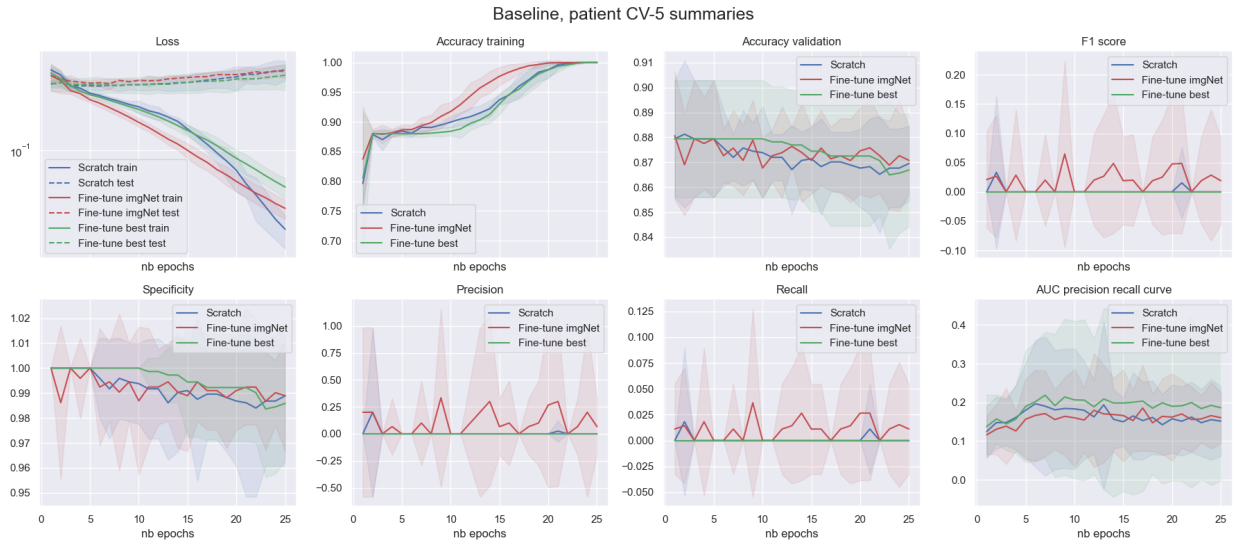


Figure 5: Baseline experiment for the standard CNN architecture.

In the experiment where the MI class is augmented at random by some transformations, as

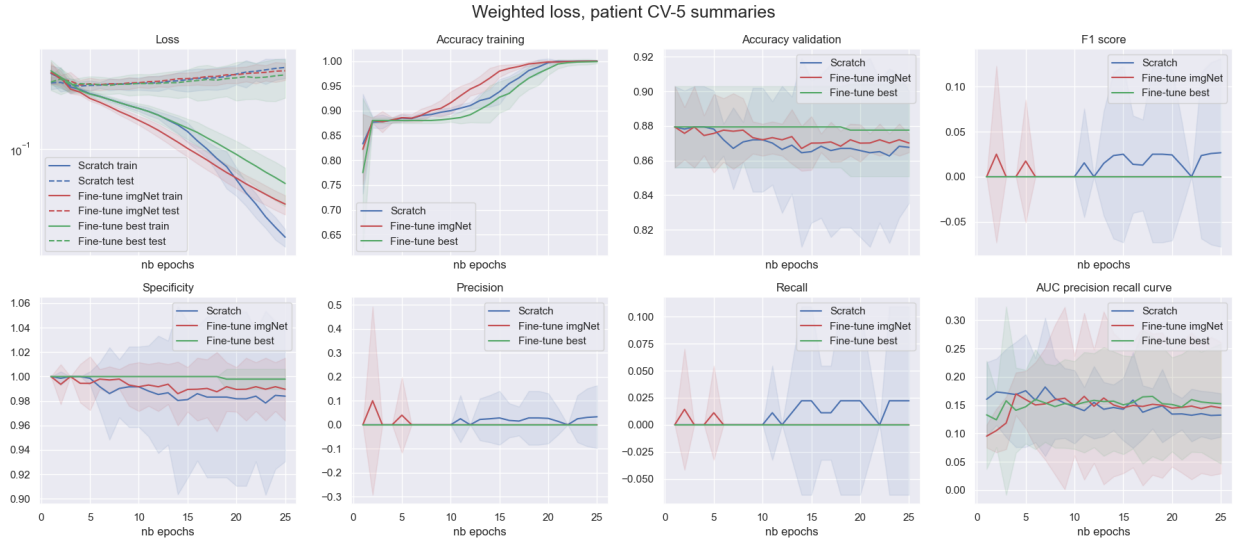


Figure 6: Weighted loss experiment for the standard CNN architecture.

explained in section 3.2, the results are slightly similar as before, except that in the training accuracy, the model does not reach an accuracy of 1. Even if the variation is large, the model has a recall different from 0 and a specificity different from 1. Thus, the model misses some non MI patches and predicts well some MI patches. In all the experiments, the accuracy does not exceed the 0.88 thresholds, which would be necessary for the model to predict well

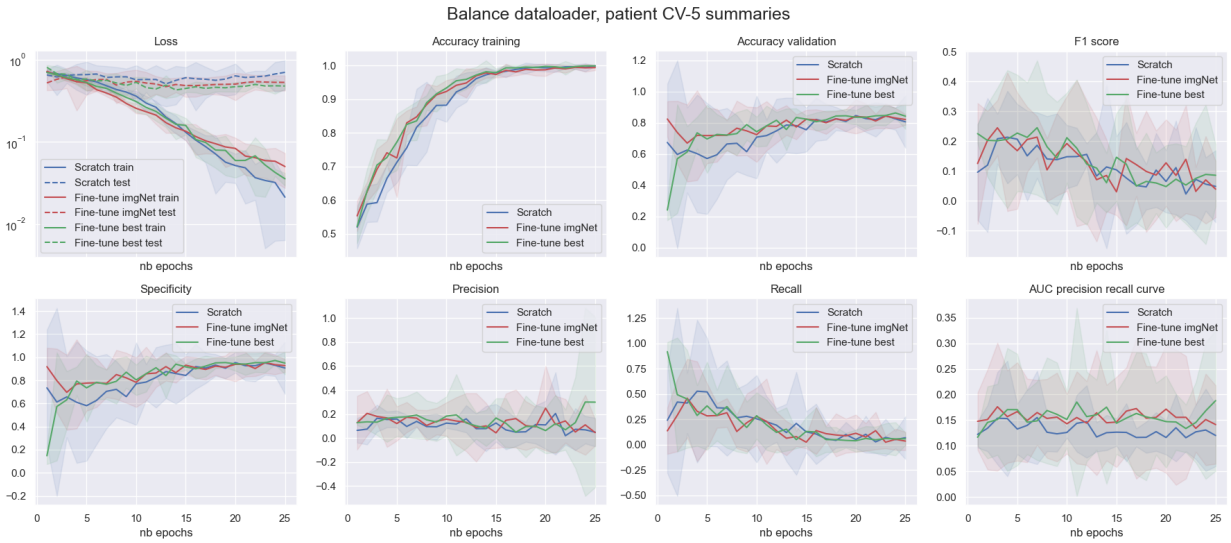


Figure 7: Balance dataloader experiment for the standard CNN architecture.

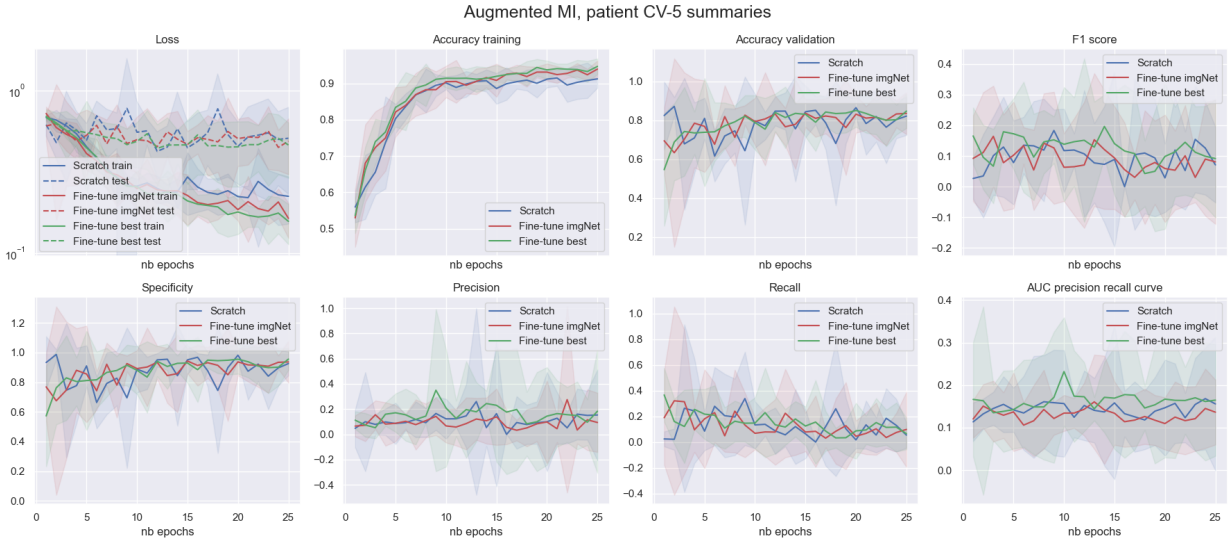


Figure 8: Augmented MI experiment for the standard CNN architecture.

all the non MI and also predict well some MI patches.
From those four experiments, it is difficult to tell which configuration is the best since they all seem to overfit the non MI class. However, balancing the data in the training using either the augmentation or the balance dataloader results in a non zero F1 score and recall, thus reducing the extreme overfitting of predicting only one class.
The experiments from the Siamese architecture network show a similar pattern. The baseline and weighted loss experiments show overfitting towards the non MI class. From all possible configurations, the ImageNet initialization seems to be slower than the others to learn on the

train set. For the balance dataloader, all the methods tend to overfit the non MI class but the ImageNet initialization stays uncertain because it leads to 0.5 recall and 0.6 specificity as shown in figure 9. For the augmented MI experiment, the results are comparable to the one obtained with the standard CNN architecture. Since they are comparable the results are not presented here.

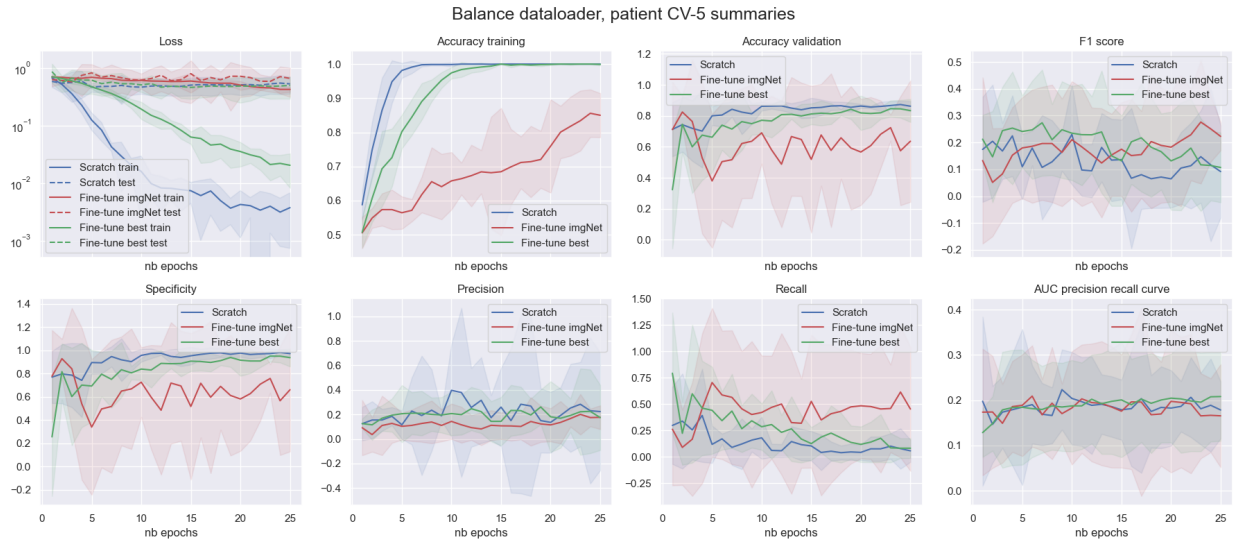


Figure 9: Balance dataloader experiment for siamese CNN architecture.

Finally, the least worse model seems to be the augmented MI experiment with the standard CNN architecture. Before testing the model on the test patients, an hyperparameter tuning is performed to find the best parameters combination. A standard grid search is performed over the learning rate $\{0.0005, 0.001, 0.003, 0.005\}$, the l2 penalized weights $\{0, 0.01, 0.1, 0.2\}$ and a dropout of 0.5 is added or not before the linear classifier. It leads to 32 experiments to run with a CV-5 for each of them. The best parameters configuration is based on the F1 score since it is the fairest metric in an imbalance scenario. The results are not very promising because most of the best F1 configuration is reached at early epochs (less than 5 epochs), then it decreases. Thus, an acceptable configuration is a learning rate of 0.003 with no penalized weight (0) and no dropout. The best model is trained one final time on the entire dataset excluding the seven test patients. It was tested for five runs in order to have statistically significant results. The results are shown in table 2. The results are selected accordingly to the best F1 score epochs. The model is very uncertain as the confidence interval is large and you can see that the best F1 score is reached at the first epochs with the weights initialization from Thanou et al. (2021) which shows that the model does not generalize at all.

There are various reasons to explain why the model predictions are so poor. The first one can be the annotation or the data quality that is not sufficient for the model to learn something about the vessel segment that leads to an MI. Another reason would be that the patches extracted and resized bring important distortion and blurriness to the input, which alters

Initialization	F1	Recall	Precision	Specificity	AUC	Epoch
Scratch	0.17 ±0.20	0.11 ±0.14	0.47 ±0.7	0.97 ±0.04	0.18 ±0.07	27
ImageNet	0.18 ±0.14	0.23 ±0.35	0.19 ±0.07	0.83 ±0.27	0.18 ±0.09	16
Best	0.18 ±0.20	0.16 ±0.21	0.26 ±0.31	0.91 ±0.07	0.17 ±0.11	1

Table 2: Results of five runs of train and test on the set of test patients for the three different initialization using the standard CNN architecture and the augmented approach.

the data quality and add too much noise to the data. The volume of input data can also lead to overfitting, the model learns the noise inside the data and not the MI signal. Finally, it is possible that the model itself is not appropriate to learn a linearly separable representation of the vessel segment.

5 MI prediction: a segment detection approach

In the previous experiment, the classification was done at the patches level by resizing the extracted patches to a fixed size. However, due to the wide variation of the dimension across patches, the input quality was poor. In this section, we tried to detect and classify vessel segment between MI and non MI boxes at the image levels using object detection. We saw in section 1.2 that Du et al. (2020) and Danilov et al. (2021) demonstrate excellent performance on stenosis detection.

5.1 Data configuration

The goal behind object detection is to predict the object location and the class of the detected object. The annotated RCA images contain boxes that delimit the anatomic portion of the vessel, and each of the boxes is classified as MI or not MI. Thus, object detection seems to be a natural choice in this setting. Note that the annotated data boxes follow the orientation of the vessel, as demonstrated in figure 1. Indeed, the boxes are not horizontally oriented as most standard object detection models would require, an example is COCO Lin et al. (2015).
The first idea could be to define a bigger horizontal box that contains an oriented annotated box. However, the horizontal box could be too big and might include another stenosis of the potentially wrong class. It would increase the noise in our data, which is already not great. The solution chosen in this experiment is to work with a model that accepts rotated boxes and keeps the original annotated box as the ground truth bounding boxes as in Follmann and König (2020). Regarding the data preparation, we need to create an annotation similar to the one used in the COCO dataset. For each image, we need to specify the location and the class of each box. With the rotated box approach, the location of the box is defined by the location of the box center, the width, the height, and finally the rotation angle of the box. Also, we created two dataset: one with both class MI, non MI and one with only the MI class similarly to Danilov et al. (2021). The first dataset contains 225 images in the train

set and 59 images in the validation set. Note that in total there are 1,192 boxes including 143 MI boxes. The second dataset contains 82 images in the training set and 39 images in the validation set. In this validation set, there are 20 images that contain an MI and 19 that do not. Note that the number of data is really low, especially for the second dataset, with only the MI boxes compared to the 8,000 and 7,000 images in Danilov et al. (2021) and Du et al. (2020) respectively. All images were resized such that the maximum size side was 1,524 pixels and the minimum side was 1,464 pixels.

5.2 Model and training

In this approach of object detection, we chose to work with fast R-CNN. This model already showed very good performances on various datasets Ren et al. (2016). Danilov et al. (2021) used a Faster R-CNN with a ResNet CNN as backbone and achieved very good accuracy on stenosis detection. Thus, for this approach, a ResNet50 pre-trained on ImageNet is selected as backbone of the model. The architecture of the model is exhibited in figure 10. The

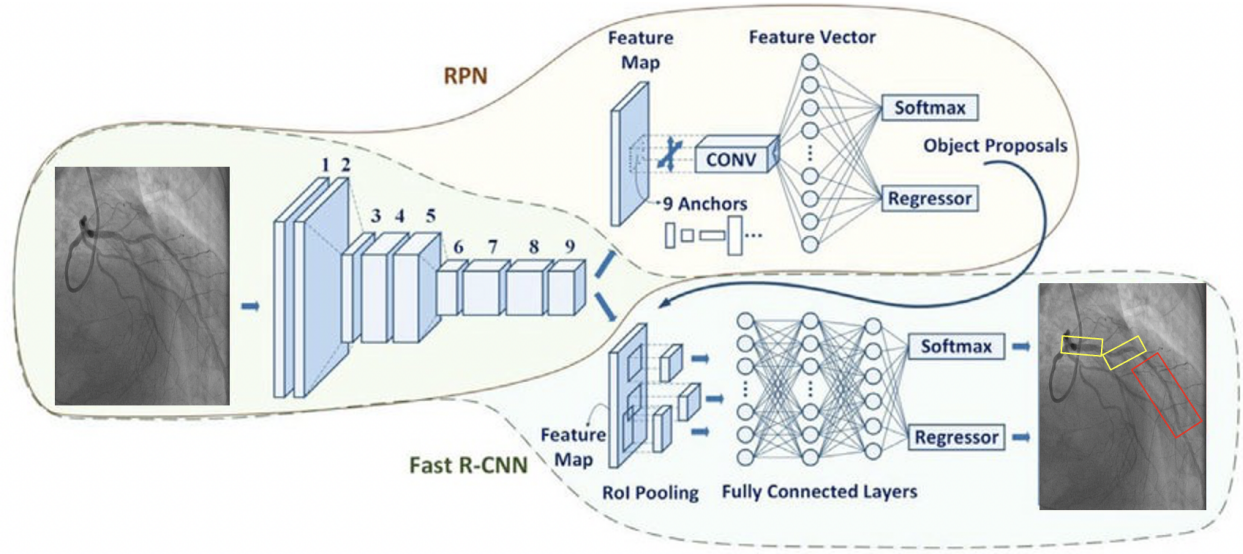


Figure 10: Scheme of a faster R-CNN architecture.

model is composed of three different parts: the first part is a CNN, more precisely a ResNet in our case, that automatically learns the hierarchical feature representations from the data and outputs a feature map; then, the feature map is sent to a region proposal network called RPN, where for different locations of the feature map, a small neural network predicts whether there is an object or not as well as the bounding box of those objects; in the end, the best region proposals are extracted and resized from the feature map by a region of interest (RoI) pooling, so that they can be the input for another fully connected neural network which predicts the object class (classification) and the bounding boxes (regression).
In this model, many parameters can be tuned. The implementation is based on Detectron by Facebook, Girshick et al. (2022). They explained how to choose many of the parameters and

we also referred to Follmann and König (2020) and Du et al. (2020) implementation. Other parameters were tweaked by hand to give the best results such as the anchors box proposal (scale, aspect ratio and angle) allowing to have good box proposal and the non-maximum suppression (NMS) threshold limiting the number of predictions for the same bounding box. Then, two slightly different architectures were tested for the RPN network. The anchors are applied at different scales in order to detect object of different size in the model. Another architecture called feature pyramidal network (FPN) can be used similarly to Du et al. (2020), where, instead of scaling the anchors, the anchors are applied at different levels of the feature map from the ResNet architecture.
Additionally, augmentation was applied similarly to He et al. (2018) in the Cityscapes dataset application, where they randomly resized the minimum size of the image in order to reduce the overfitting. They also had a reduced number of training images but still higher than ours (2,975 which is much larger than our 225 or 82 images). We randomly resized the smallest side of the images between $[1400, 1464]$, all the inferences were made on 1464. A random horizontal flip was used in the training.
Finally, regarding the training setup, SGD is used to optimize the model. The initial learning rate is 0.001 with 2,500 iterations, as advised by the Detectron Readme Girshick et al. (2022). A warm up of 500 iterations is used with a warm up factor of 0.33. The learning decreases by a factor of 10 after 1,500 iterations. In this setting, only one image is used per batch with one GPU.

5.3 Results

In this approach, we tried to solve the challenge CH3 by working at the images level, so that we do not have to resize the patches. In this experiment, two main datasets are used: the first one consists of both MI and non MI classes whereas the second is only using MI class. In this part, only the results from the two different architectures with and without FPN are presented.
Using the first configuration, where there are two classes to predict, the model achieves poor performances. Results are not presented since no MI boxes are predicted and model predicts not accurately non MI boxes. With all the different parameters tested, the model still overfits the non MI class. Thus, the first configuration, with both classes, is removed from the experiment.
Using as input only the MI boxes, the model becomes naturally better to predict the MI class since now there is only one class to predict, besides the fact that the model also predicts whether the object is from the background. This is the same approach used by Danilov et al. (2021), where they detected only the stenosis class. With the first architecture without FPN, the model predictions are rather messy and sparse. In figure 11a and 11c, the five highest score predictions are shown in red, we can see that the bounding boxes are large and not accurate, especially in figure 11c. None of the predictions has an IoU bigger than 0.5. The ground truth corresponds to the green box. Moreover, the detection achieves poor performances in terms of mAP and mAR, as shown in table 3. Regarding the five highest scores, the model reaches 0.003 and 0.037 respectively for the mAP and mAR. Using the

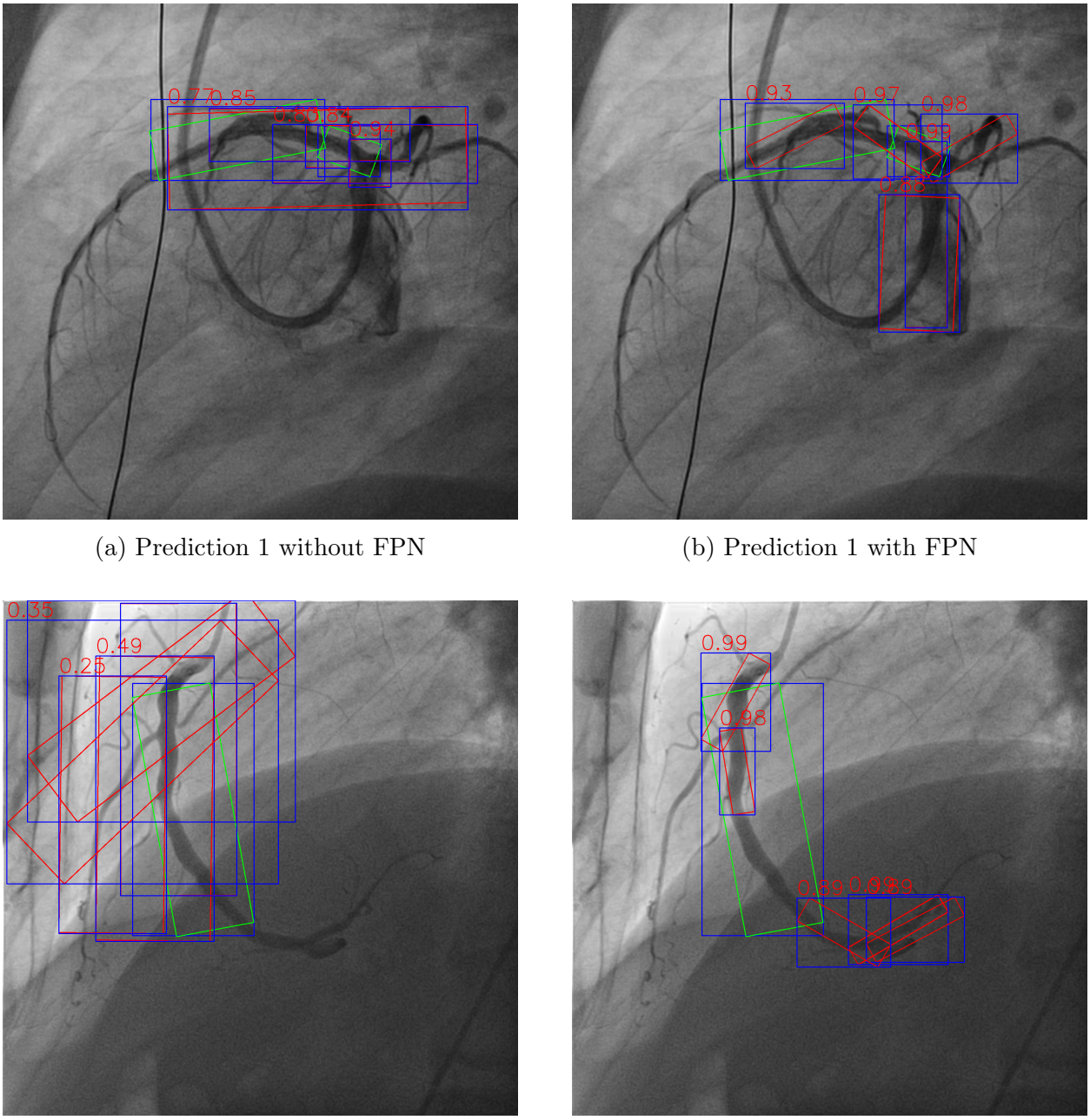
FPN architecture, the model shows better prediction as demonstrated in figure 11. The predictions are much smaller and accurate for vessel detection. The scores of the predictions are also more confident. However, even if the predicted boxes seem better, none of them has an IoU bigger than 0.5. The resulting mAP and mAR are better than before but far from excellent. In both experiments, the training loss for the box coordinates regression doesn't decrease much. It can explain why the predicted bounding boxes didn't match well the ground truth. Since the model doesn't reach reasonable performance, the model is not tested on the test set.
We already saw in section 1.2 that Du et al. (2020) and Danilov et al. (2021) demonstrate excellent performances on stenosis detection. The prediction of MI at the vessel level is similar since it works on the same types of data but is slightly different at the annotation levels. Thus, the first hypothesis to understand why it doesn't work is that the annotation is not sufficient for an object detection task. Indeed, we saw that the model with FPN could detect the vessel, but the mAP score is still low. Figure 11b and 11d show that even if the predicted bounding box is inside the ground truth, with a possible focus on the stenosis inside, the IoU is small because the predicted bounding box doesn't match the ground truth bounding box. Thus, the model has issues on the box coordinates regression that could come from the annotation quality or the fact that bounding box dimensions are really heterogeneous. Another explication could be the volume of data that is not sufficient compared to the other studies.

IoU 0.25:0.05:0.75	mAP			mAR		
	max_det = 5	max_det = 10	max_det = 100	max_det = 5	max_det = 10	max_det = 100
without FPN	0.003	0.004	0.004	0.037	0.062	0.12
with FPN	0.013	0.014	0.02	0.11	0.15	0.48

Table 3: Results of the faster R-CNN architecture with and without FPN on the validation dataset containing only the MI class.mAP and mAR are metrics computed for IoU from 0.25 to 0.75, see section 3.1. *max_det* is the maximum of detection, a *max_det* of 5 means that it looks only at the 5 highest score detected boxes.

6 Custom attention classification

In the previous approaches, we tried to predict if the vessel leads to an MI at the patches levels. We assumed that it didn't work well because of the boxes size heterogeneity. We tried object detection as a solution to this issue since it uses full images. However, results were poor. We suspected that the cause could be the annotation quality as well as the small volume of input data, since it only dealt with MI images. Thus, in the following approach we decided to still work at the image level, but focus the model on a specific vessel segment, with the idea of Moon et al. (2021). We used custom attention to highlight vessels segments to then classify whether it leads to an MI or not. With this approach, the heterogeneous boxes dimension is not an issue, because the model is built at the image level. The class



(a) Prediction 1 without FPN

(b) Prediction 1 with FPN

(c) Prediction 2 without FPN

(d) Prediction 2 with FPN

Figure 11: Comparison of two predictions made with faster RCNN, with/without FPN. Green boxes are the true bounding boxes and Red are the predictions. The blue boxes are only here to associate the prediction score with the right prediction. In this plot, only the five highest score bounding box predictions are presented.

imbalance CH1, and the limited amount of data CH2 could still be challenging, but methods used in the patches classification in section 4 are used again to attenuate this problem.

6.1 Data configuration

In this approach, the model receives as input the original images, with also the information about the segment of interest in the vessel. For that, we created a custom attention input highlighting the location of the vessel in the original images. The attention is a Gaussian distribution centered at the middle of the box that locates the vessel segment, with the same scale as the box such that 95% of the distribution is inside the box. Thus, the Gaussian $\mathcal{N}(\mu, \Sigma)$ is defined by the following mean and covariance matrix :

$$\mu = \begin{pmatrix} C_x \\ C_y \end{pmatrix}, \text{ and } \Sigma = M \begin{pmatrix} (\lambda \frac{w}{2})^2 & 0 \\ 0 & (\lambda \frac{h}{2})^2 \end{pmatrix} M^T$$

where $M \in \mathbb{R}^{2\times 2}$ is a rotation matrix defined by the angle of the box, $C \in \mathbb{N}^2$ is the coordinate of the center of the box, $\lambda \in \mathbb{R}$ is a constant that scales the Gaussian and finally $w \in \mathbb{R}$, $h \in \mathbb{R}$ are respectively the width and height of the box. Note that we additionally rescaled the Gaussian distribution to be in $[0, 1]$.
There are two methods in order to give the attention location to the model input. The first one is to give the Gaussian attention as an additional input channel. As figure 12 demonstrates in the second image, the Gaussian attention corresponding to the green box is added as a channel. It is also possible to apply the Gaussian attention directly to the original image, as presented in the third image in figure 12. In the second method, the model only focuses on one vessel segment, since the remaining image contains no signal, while in the first method, the model could use other information around the vessel segment of interest.

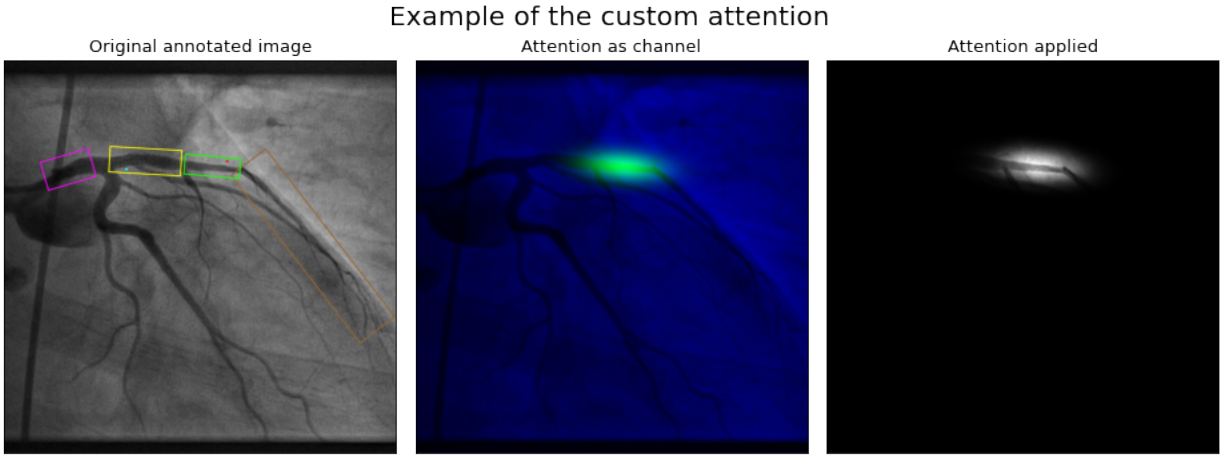


Figure 12: Example of the two different methods to incorporate the attention location of the *green* vessel segment in the input. The first image shows the original annotation. Then, the second image presents the attention as an additional input channel. Finally, the last image exhibits the Gaussian applied to the original image.

In this custom attention approach, the input is the original image but there is as much input as there are boxes in an image since the Gaussian attention changes for each box. For the

five-fold cross validation, there are 925 non MI and 124 MI images with the custom attention. In the test set, there are 124 non MI and 19 MI images from the same patient, used in the patches classification section 4. The input images have a large size of 1500×1500.

6.2 Model and training

The model used in this section is very similar to the one used for the patches classification, in section 4.2. In fact, CNN architecture seems to be the most natural architecture for image classification. Moreover, the same backbone CNN architecture is used from the previous patches classification approaches, since the objective is also to compare the approaches and not the model architectures. Thus, a ResNet18 is chosen for this experiment as well. The only difference is that the input images are much larger than in the patches extraction task. Therefore, an additional block of layers is added before the backbone ResNet18 to reduce the input size, as shown in figure 13. The reduction block is composed of a convolution with a kernel size of 7 and a stride of 2, which reduces the input size by two. It is followed by a batchnorm and a ReLU activation layer before being fed to a max pool layer, with a kernel size of 3 and a stride of 2 again reducing the size by two. Thus the reduction block reduces the input size by 4. We didn't use the ResNet18 directly after the input since the output feature map would be too large, too much information would be lost if the final average pooling layer was applied before the tensor was flattened, for the classifier block.

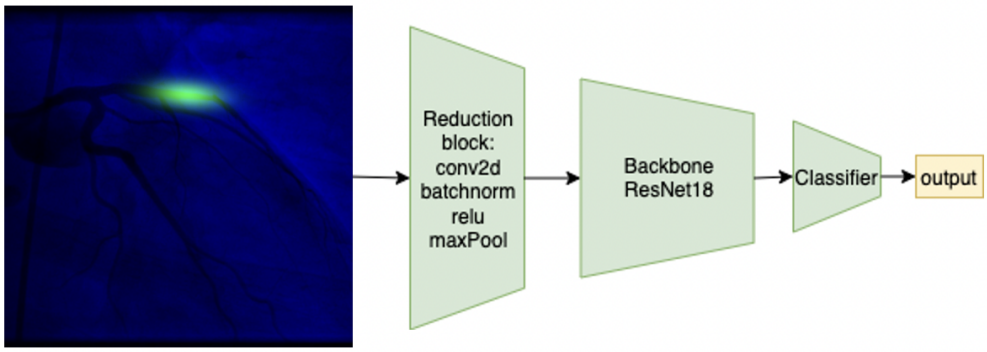


Figure 13: Architecture used for the custom attention input. An additional block is added before the ResNet backbone to reduce the input size.

Regarding the training, three backbone initialization are tested similarly to those presented in section 4.2 for the patches classification: the default initialization (learn from scratch), the pre-trained ImagNet ResNet18, and finally the pre-trained stenosis MI classification from Thanou et al. (2021) study. This transfer method could help with the limited amount of data available, CH2.
Then, four different experiments are tested, similarly to the patches classification presented in section 4.2: the baseline, the weighted cross entropy loss, the balance dataloader, and finally the augmented MI class. In this approach, the augmentation is done similarly to the

one in the patches classification, but additionally, when the Gaussian custom attention is created, it is done for different scales and slightly different locations.

6.3 Results

The two different ways of incorporating the custom attention are compared along with the four experiments and the three different initialization. The model selected for this approach is then tested on the same test patients as in patches classification.
First, in both ways of including the custom attention, the baseline, and the weighted loss experiment overfit towards the non MI class, similarly to the experiment in the patches classification (see section 4 for a reminder) you can find the resulting plot in figure 24, 25, 28, and 29. Then, for the balance dataloader experiment, both attention inputs performed nearly as badly as the one in the patches classification. Here the initialization with ImageNet yields the best learning on the train set. Results for the apply custom attention are similar to the channel input attention except for the scratch initialization, which seems to overfit the non MI class for early epochs, plot shown in figure 26 and 30. The setup where we augmented the samples from the MI class demonstrates slightly worse performance, as presented in 27 and **??**. Thus the hyperparameter focus only on the balance dataloader experiment since it seems the most promising experiment. The ImageNet initialization demonstrates more promising performance. The results of the four-fold cross validation are presented in figure 14. In the plot, the best parameter configuration for the ImageNet initialization that yields the best F1-score is shown. The channel method seems to be slightly better than the applied method of incorporating the attention. It is demonstrated by the AUC and F1-score that are higher. The overall results are not great since for an F1-score of 0.3 at the 14 epochs, the recall reaches 0.27, precision 0.4, and specificity 0.87. It still overfits the non MI class.
The best configuration is the channel attention with the ImageNet initialization. This configuration was trained on the data, and then tested on the seven test patients. In table 4, the results show an F1 score of 0.22, recall 0.22, precision 0.25 specificity 0.88, and AUC 0.19. The confidence interval is also quite small compared to the performance in the patches classification. The overall performance is slightly better than the patches classification on the test set but still not sufficient for a CAD systems.
This approach is similar to the patches classification since we classify vessel segments but here the inputs are the full images with particular attention on the vessel segment. In this configuration, the heterogeneous boxes dimension might no longer be an issue. The reason why it doesn't perform well could be that attention used in this approach is not a suitable solution for this problem. The annotation quality or image quality might be still one of the reasons. Finally, it could be because the signal of MI vessels is not sufficient for the model to make a clear distinction between non MI and MI vessels.

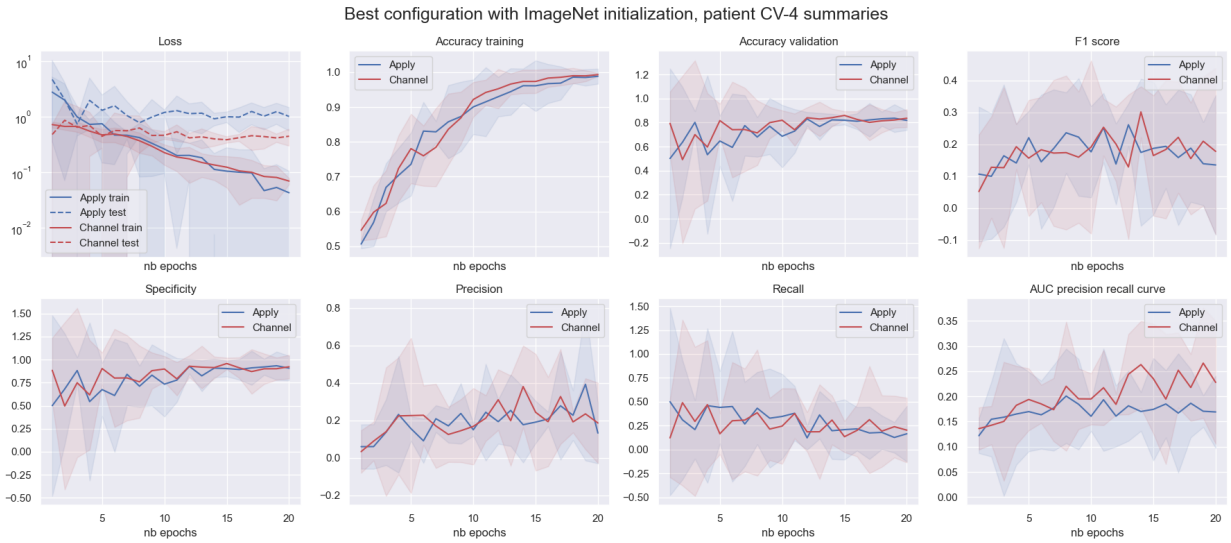


Figure 14: Best parameter configuration for the ImageNet initialization after running a four-fold cross validation using a balance dataloader. Both *Apply* and *Channel* ways of incorporating the attention are presented.

Initialization	F1	Recall	Precision	Specificity	AUC	Epoch
ImageNet	0.22 ±0.04	0.22 ±0.10	0.25 ±0.13	0.88 ±0.09	0.19 ±0.04	17

Table 4: Performance of the custom attention with the channel attention on the seven test patients. The ImageNet initialization is chosen since it yields the best performance in the cross validation.

7 MI stenosis classification

In the three previous approaches, the input demonstrates many challenges: the dimension of the boxes in the patches classification approach, the annotation quality, and the limited amount of data for the object detection approach. In the last approach, using custom attention, the prior knowledge of the vessel segment location shows no benefit in the generalization performance. To understand why all of the approaches can not well succeed, we tried to evaluate the predictive capacity on MI stenosis which is a simpler task. Indeed, classify MI stenosis using fixed size patches center around the stenosis is simpler than classifying MI vessel that have heterogeneous sizes. The task is still slightly different than classical stenosis detection Ovalle-Magallanes et al. (2020) because, in the annotation, only MI stenosis are spotted and the non stenosis patches might contain normal stenosis.

7.1 Data configuration

In stenosis classification, the objective is to classify patches extracted from the vessels whether it contains stenosis or not. In this experiment, we needed a new annotation level of

the data. Indeed, since the objective is to classify the stenosis, the location of the stenosis is needed. Thus, the doctors took care of annotating the stenosis for the MI patients as in figure 15 In total, 167 stenosis were found. The next step is to create a dataset from these annotations. Patches centered around the stenosis are extracted. The patches size is 250×250 pixels so that we can either randomly crop a 224×224 image inside of it to augment the data or simply crop a smaller 224×224 image with the same center. Regarding the patches without stenosis, patches of 250×250 are extracted from the center of boxes of non MI patients. Therefore, 167 stenosis and 167 non stenosis patches of size 250×250 are extracted. Note that the amount of patches is slightly bigger than the data set from Ovalle-Magallanes et al. (2020). Thus, the volume of data shouldn't be an issue if the quality is sufficient. In figure 15, an example of patches extraction for an MI patient is shown. This particular example shows a significant stenosis inside the *Green* box. The lesion of the stenosis is clearly visible, the vessel is narrowing at the location of the blue dot. However, the stenosis contained in the *Yellow* and *Magenta* box is less obvious. The data are split such that a fair comparison with the other methods can be made. The test set contains 48 patches with 24 stenosis patches from the seven test patients defined in section 3.1. The remaining patches are used for the patient-level five-fold cross validation. Therefore, in each iteration, the model is trained on 250 patches and validated on 62 patches, equally distributed between stenosis and non stenosis.

7.2 Model and training

The model used in this section is similar to the one used in the patches classification, section 4.2, and the one used for the custom attention approach, section 6.2. A CNN architecture is used and, similarly to the previous approaches, a ResNet18 model is used as a backbone to extract valuable features from the image. The final classification task is performed by a linear classifier.

Regarding the training, the three backbone initialization mentioned in the patches classification and custom attention section 4.2 and 6.2 respectively are used again. In this stenosis detection, the box dimension CH3 and class imbalance CH1 are not challenges anymore because the patches are balanced and each patch now has a fixed size that brings no distortion to the patch quality. Thus, the different experiments to run in the stenosis classification are slightly different than the ones used in patches classification and custom attention. Here only two experiments are performed. The first one is the baseline and the second one is using augmentation. In the augmentation, transformation is performed online during training for both MI and non MI classes, see section 3.2. It includes random crop, horizontal and vertical flipping. In this approach, since there are only two experiments, hyperparameter tuning is done for each experiment. A standard grid search is performed over the learning rate $\{0.004, 0.007, 0.01, 0.03\}$, the l2 penalized weights $\{0, 0.01, 0.055, 0.1, 0.3\}$. This leads to 20 experiments to run with a 5-CV for each of them. The best parameter configuration is based on the F1 score, in order to be consistent with the previous experiment.

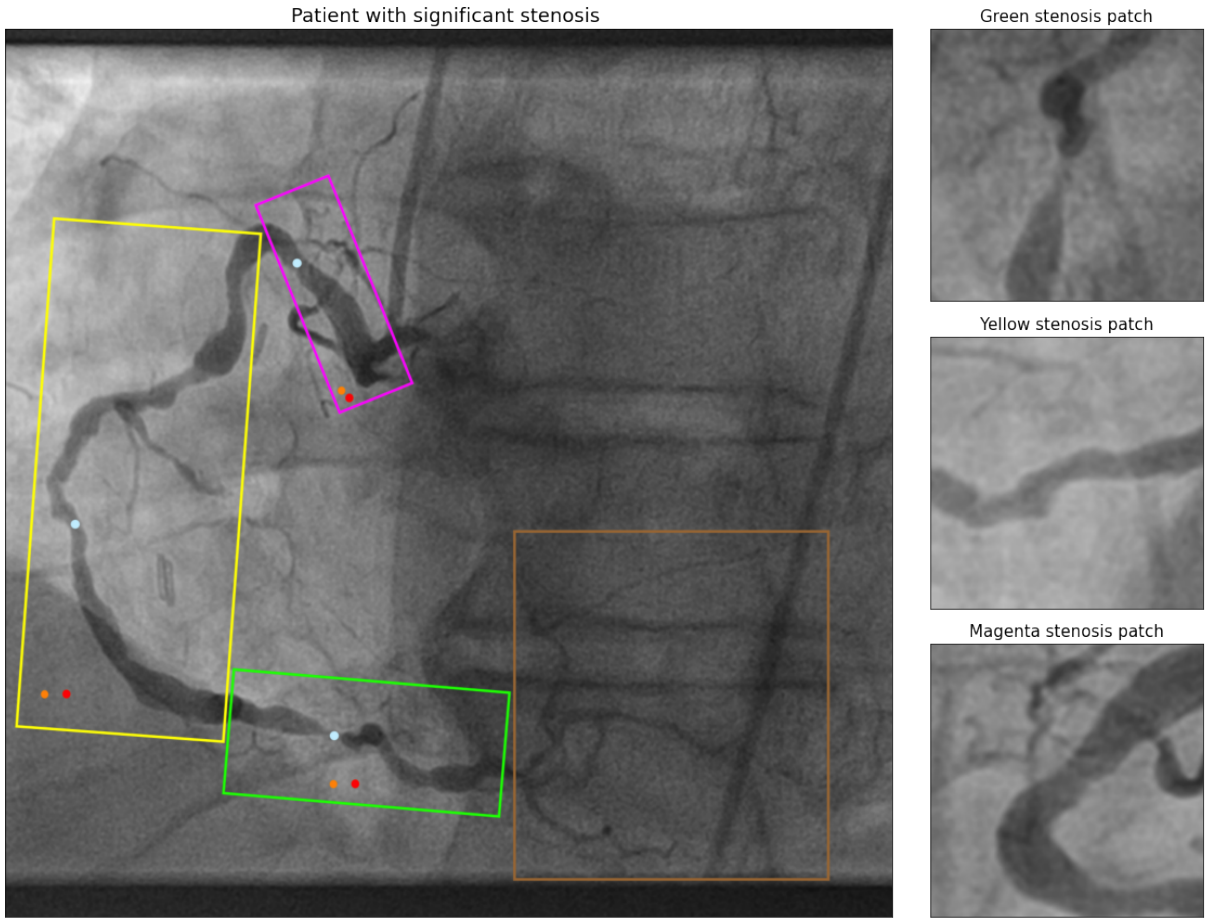


Figure 15: Example of the patches stenosis extraction for an MI patient. Doctors annotated the stenosis using sky blue color. The patches have a size of 250×250 pixels.

7.3 Results

The baseline and the augmented experiment are presented in this section. For each experiment and each initialization, the parameter configuration is selected to have the highest F1 score.

The baseline experiment demonstrates interesting performance as you can see in figure 17. ImageNet initialization shows the best performance. It shows a recall bigger than 0.75 and the specificity above 0.40, with an average F1 score of 0.66. Note that the model is still uncertain as demonstrated by the confidence interval on the figure. The augmentation experiment presents less promising results. No initialization exceed the 0.6 F1 score threshold on average. The model seems to predict well some of the non stenosis images because the specificity is above 0.5 while keeping a constant predictive accuracy on the stenosis patches since it oscillate around 0.5.

The baseline experiment with the ImageNet initialization demonstrates the best performance. Thus to validate the performance the model is trained on all the data excluding the

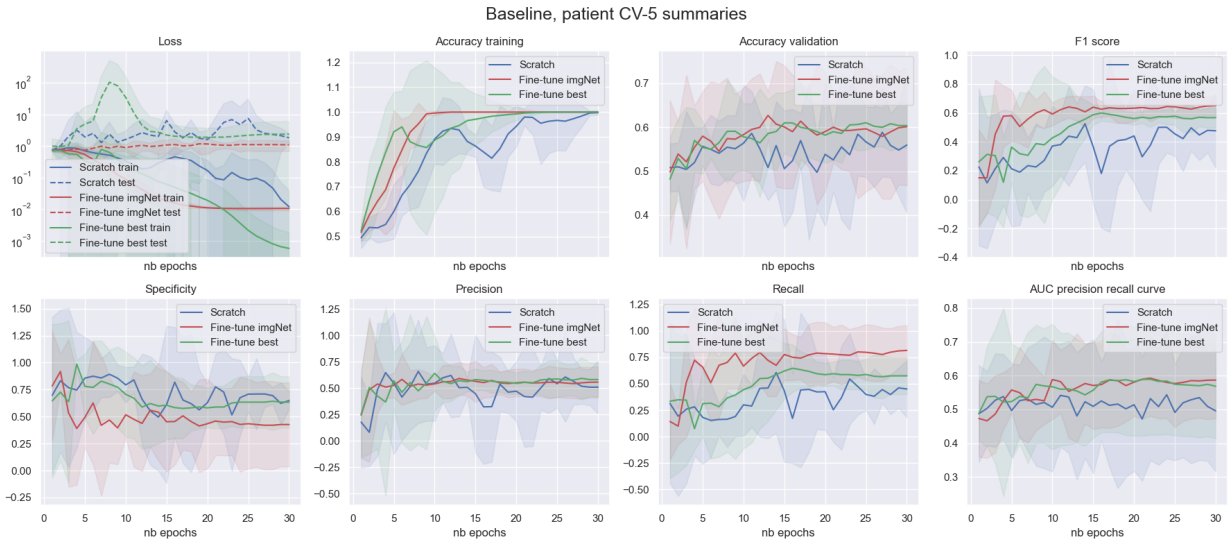


Figure 16: Baseline experiment with five cross validation for the MI stenosis classification.

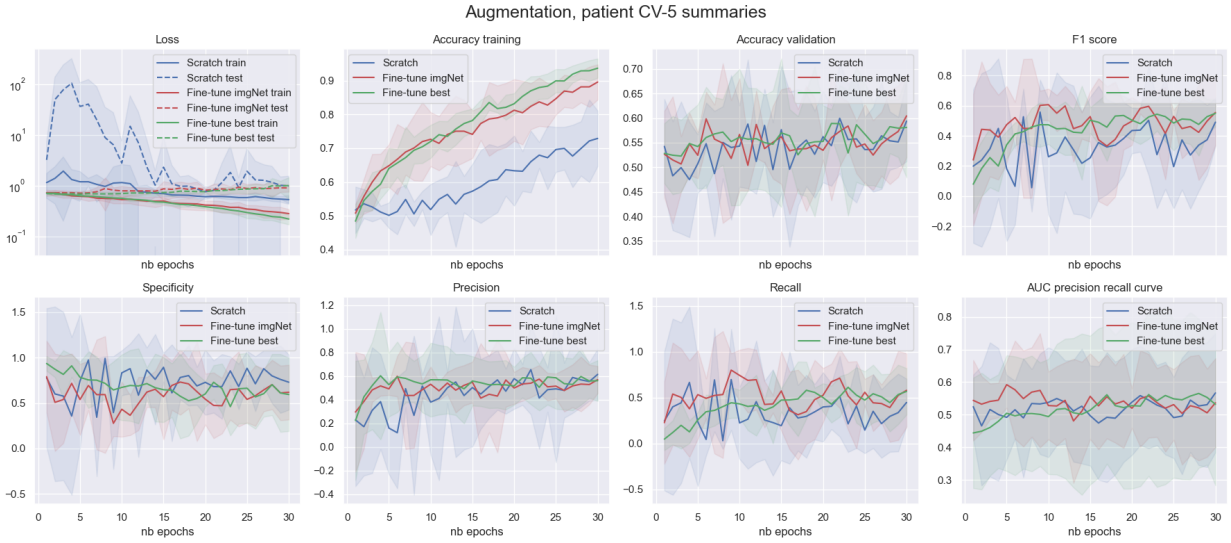


Figure 17: Augmentation experiment with five cross validation for the MI stenosis classification.

seven test patients and tested on the seven patients. The results are presented in table 5. ImageNet shows as expected the best performance. It reaches an F1 score of 0.79, a recall of 0.8, a precision of 0.78, a specificity of 0.77. The results on the test set are better than the one on the cross validation. It could be because the qualities of the stenosis in the test set are much better than the one in the cross validation. The number of training data is also higher when trained on all data excluding the seven patients than in the cross validation. On the same test set, the results are not better with the augmentation experiment as expected

from the cross validation.

Initialization	F1	Recall	Precision	Specificity	AUC	Epoch
Scratch	0.67 ±0.007	0.99 ±0.33	0.50 ±0.018	0.025 ±0.10	0.53 ±0.084	1
ImageNet	0.79 ±0.11	0.80 ±0.14	0.78 ±0.14	0.77 ±0.16	0.85 ±0.07	28
Best	0.66 ±0.056	0.49 ±0.06	1.0 ±0.0	1.0 ±0.0	0.94 ±0.026	16

Table 5: Baseline experiment performance on the seven test patients. Results correspond to the best F1 score and 5 runs are performed to get statistically significant results with the 95% confidence interval.

Since this task is slightly different from a stenosis classification, we can not compare it with different studies like Ovalle-Magallanes et al. (2020) or Moon et al. (2021). But from the test evaluation, it seems possible from a patches stenosis annotation to find some of the MI stenosis.

8 Discussion and conclusion

In this thesis, we studied the very challenging problem of prediction MI from angiograms, using deep learning techniques. In particular, we focused on a clinical study of 58 patient that have MI vessel segment. This dataset presents many different challenges as the class imbalance CH1, the small amount of data CH2, the annotation quality CH3, CH4, and the qualities of the image CH4. In the present study, a special effort was put to find solutions to tackle these challenges.

Our first approach consisted of classifying patches, extracted from the annotated boxes. The patches were resized to a fixed size in order to alleviate the heterogeneous patches dimensions CH3. Patches were used as pairs of different views of the same vessel segment in order to attenuate the medical tool artifact that might be present in some views and improve the data qualities CH4. Different instances of the models that dealt differently with the input images, the experiments that tried to reduce the class imbalance CH1, the transfer learning with different initialization that tried to overcome the limited amount of data CH1, and the hyperparameter tuning to get the best models configuration, unfortunately all yield poor performance on the patient-level five-fold cross validation. One major issue with this approach turned out to be the data quality of the resized extracted patches. Indeed, we resized heterogeneous patches size and aspect ratio, that turned out to affect the quality by adding distortion, blurriness, and various scales vessels representations, as shown in figure 18.

Therefore, to avoid resizing the patches, we decided to work with the full image as input. This approach was expected to increase the data quality CH4 and also reduce the effect of boxes dimension CH3. We tried an approach that simultaneously detects the segment, and predict the probability of being an MI segment. The object detection approach has the advantage to detect the MI boxes with the full image as input. Thus, this approach is chosen to detect only the MI boxes in the annotated images. The model seemed to detect

the vessel quite accurately but showed poor precision to predict bounding boxes that match ground truth MI bounding boxes. The reason could be that the annotated bounding boxes do not accurately delimit the target MI vessel or that the MI vessels do have not a sufficiently different signal than normal vessel segments and thus could not be classified as MI bounding boxes accurately.

Since we know the location of the vessel segment, in our next modeling approach, we tried to add that information as prior knowledge for the prediction. Thus, a third approach that we developed still works with the full images as input but contains an attention that highlights the vessel location before classifying the image as MI or non MI. As before, this might increase the data quality CH4 and also reduce the effect of boxes dimension CH3 but here a prior knowledge is added. The model developed, the experiments that tried to reduce the class imbalance CH1, the transfer learning with different initialization that tries to overcome the limited amount of data CH1, and the hyperparameter tuning to get the best models configuration yielded slightly better performance on the patient-level five-fold cross validation than the first approach. However, the predictions were not sufficient for a CAD system to work in practice. With all of those approaches, and a careful study of the obtained results, we suspect that the data quality or the signal itself is not sufficient for a model to predict whether a vessel segment will lead to an MI in the future.

In our last attempt to investigate if angiograms contains a predictive signal for MI, we decided to simplify the problem even further. In particular, our last experiment tried to see if the MI stenosis has a significant signal to be well predicted. The doctors annotated the precise location of the MI stenosis so that we could extract patches center around them. We also added non MI stenosis patches from the non MI patients. The model seemed to predict well a part of the MI stenosis and demonstrated a reasonable performance on the test set. This might be promising however, a bottleneck of this approach is that we do not know whether the non MI stenosis patches contain stenosis or not. This might be an issue because, if we suppose that non MI stenosis patches don't contain stenosis then the classification task is a stenosis classification and we have no insight into the MI stenosis signal quality. Whereas, if non MI stenosis patches contain normal stenosis then we could say that MI stenosis has a different signal and could be predicted by the model. Therefore, we need an additional annotation level in order to evaluate the MI stenosis signal quality. In figure 19 examples of input stenosis patches are presented

Finally, this work shows no evidence that the vessel segment information is relevant to predict MI stenosis. However, in this study, the data are challenging and bring a lot of additional perturbation. In terms of data qualities, adding normal and MI stenosis locations could be beneficial. Also, the annotation quality could be improved by collecting annotations from different doctors and then aggregating them. One could also add an estimation of the quality of the image by using stratified noisy cross-validation from Hsu et al. (2020). In terms of approach and experiment, supervised contrastive learning could be a promising approach like the one mentioned in Khosla et al. (2021). Indeed, in figure 2, we saw that specific anatomic vessels segments are more likely to lead to an MI. Thus, we could group together the vessel segment by color and MI in the latent space representation before linearly classifying the

vessel segment. It should incorporate the anatomic vessel types as prior knowledge.

Code is available at : https://github.com/Vuillecard/CVD_predicition.

Pierre Vuillecard, Lausanne, 21.01.2022 :

References

Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, 385(9963):117–171, January 2015. ISSN 0140-6736, 1474-547X. doi: 10.1016/S0140-6736(14)61682-2. Publisher: Elsevier.

Viacheslav V. Danilov, Kirill Yu Klyshnikov, Olga M. Gerget, Anton G. Kutikhin, Vladimir I. Ganyukov, Alejandro F. Frangi, and Evgeny A. Ovcharenko. Real-time coronary artery stenosis detection based on modern neural networks. *Scientific Reports*, 11(1): 7582, April 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-87174-2. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Applied mathematics;Computational science;Computer science;Interventional cardiology;Outcomes research Subject_term_id: applied-mathematics;computational-science;computer-science;interventional-cardiology;outcomes-research.

Tianming Du, Lihua Xie, Honggang Zhang, Xuqing Liu, Xiaofei Wang, Donghao Chen, Yang Xu, Zhongwei Sun, Wenhui Zhou, Lei Song, Changdong Guan, Alexandra Lansky, and Bo Xu. Automatic and multimodal analysis for coronary angiography: training and validation of a deep learning architecture. *EuroIntervention : journal of EuroPCR in collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology*, 17, August 2020. doi: 10.4244/EIJ-D-20-00570.

Patrick Follmann and Rebecca König. Oriented Boxes for Accurate Instance Segmentation. *arXiv:1911.07732 [cs]*, March 2020. arXiv: 1911.07732 version: 2.

Armaan Garg, Vishali Aggarwal, and Neeti Taneja. Classification of Imbalanced Data: Addressing Data Intrinsic Characteristics. In Pradeep Kumar Singh, Sanjay Sood, Yugal Kumar, Marcin Paprzycki, Anton Pljonkin, and Wei-Chiang Hong, editors, *Futuristic Trends in Networks and Computing Technologies*, volume 1206, pages 264–277. Springer Singapore,

Singapore, 2020. ISBN 9789811544507 9789811544514. doi: 10.1007/978-981-15-4451-4_21. Series Title: Communications in Computer and Information Science.

Ross Girshick, Radosavovic Ilija, Gkioxari Georgia, and He Kaiming. Detectron, January 2022. original-date: 2017-10-05T17:32:00Z.

Mahdi Hashemi. Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. *Journal of Big Data*, 6(1): 98, November 2019. ISSN 2196-1115. doi: 10.1186/s40537-019-0263-7.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*, December 2015. arXiv: 1512.03385.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. *arXiv:1703.06870 [cs]*, January 2018. arXiv: 1703.06870.

Joy Hsu, Sonia Phene, Akinori Mitani, Jieying Luo, Naama Hammel, Jonathan Krause, and Rory Sayres. Improving Medical Annotation Quality to Decrease Labeling Burden Using Stratified Noisy Cross-Validation. *arXiv:2009.10858 [cs, eess]*, September 2020. arXiv: 2009.10858.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised Contrastive Learning. *arXiv:2004.11362 [cs, stat]*, March 2021. arXiv: 2004.11362.

Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese Neural Networks for One-shot Image Recognition. page 8.

Michalis Lampros, Dimitrios Fotiadis, and Athanasiou Lambros. Atherosclerotic Plaque Characterization Methods Based on Coronary Imaging - 1st Edition, May 2017.

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common Objects in Context. *arXiv:1405.0312 [cs]*, February 2015. arXiv: 1405.0312.

Shanthi Mendis, Pekka Puska, Bo Norrving, and World Health Organization. *Global atlas on cardiovascular disease prevention and control*. World Health Organization, 2011.

Jong Hak Moon, Da Young Lee, Won Chul Cha, Myung Jin Chung, Kyu-Sung Lee, Baek Hwan Cho, and Jin Ho Choi. Automatic stenosis recognition from coronary angiography using convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 198:105819, January 2021. ISSN 0169-2607. doi: 10.1016/j.cmpb.2020.105819.

Emmanuel Ovalle-Magallanes, Juan Gabriel Avina-Cervantes, Ivan Cruz-Aceves, and Jose Ruiz-Pinales. Transfer Learning for Stenosis Detection in X-ray Coronary Angiography. *Mathematics*, 8(9):1510, September 2020. doi: 10.3390/math8091510. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Trong Huy Phan and Kazuma Yamamoto. Resolving Class Imbalance in Object Detection with Weighted Cross Entropy Losses. *arXiv:2006.01413 [cs]*, June 2020. arXiv: 2006.01413.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:1506.01497 [cs]*, January 2016. arXiv: 1506.01497.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*, May 2015. arXiv: 1505.04597.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *arXiv:1409.0575 [cs]*, January 2015. arXiv: 1409.0575.

Hoo-Chang Shin, H. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, D. Mollura, and R. Summers. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Char-

acteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 2016. doi: 10.1109/TMI.2016.2528162.

Dorina Thanou, Ortal Senouf, Omar Raita, Emmanuel Abbé, and Pascal Frossard. Predicting future myocardial infarction from angiographies with deep learning. page 4, 2021.

K. Sudarshan Vidya, E. Y. K Ng, U. Rajendra Acharya, Siaw Meng Chou, Ru San Tan, and Dhanjoo N. Ghista. Computer-aided diagnosis of Myocardial Infarction using ultrasound images with DWT, GLCM and HOS methods: A comparative study. *Computers in Biology and Medicine*, 62: 86–93, 2015. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2015.03.033.

A Data quality inspection

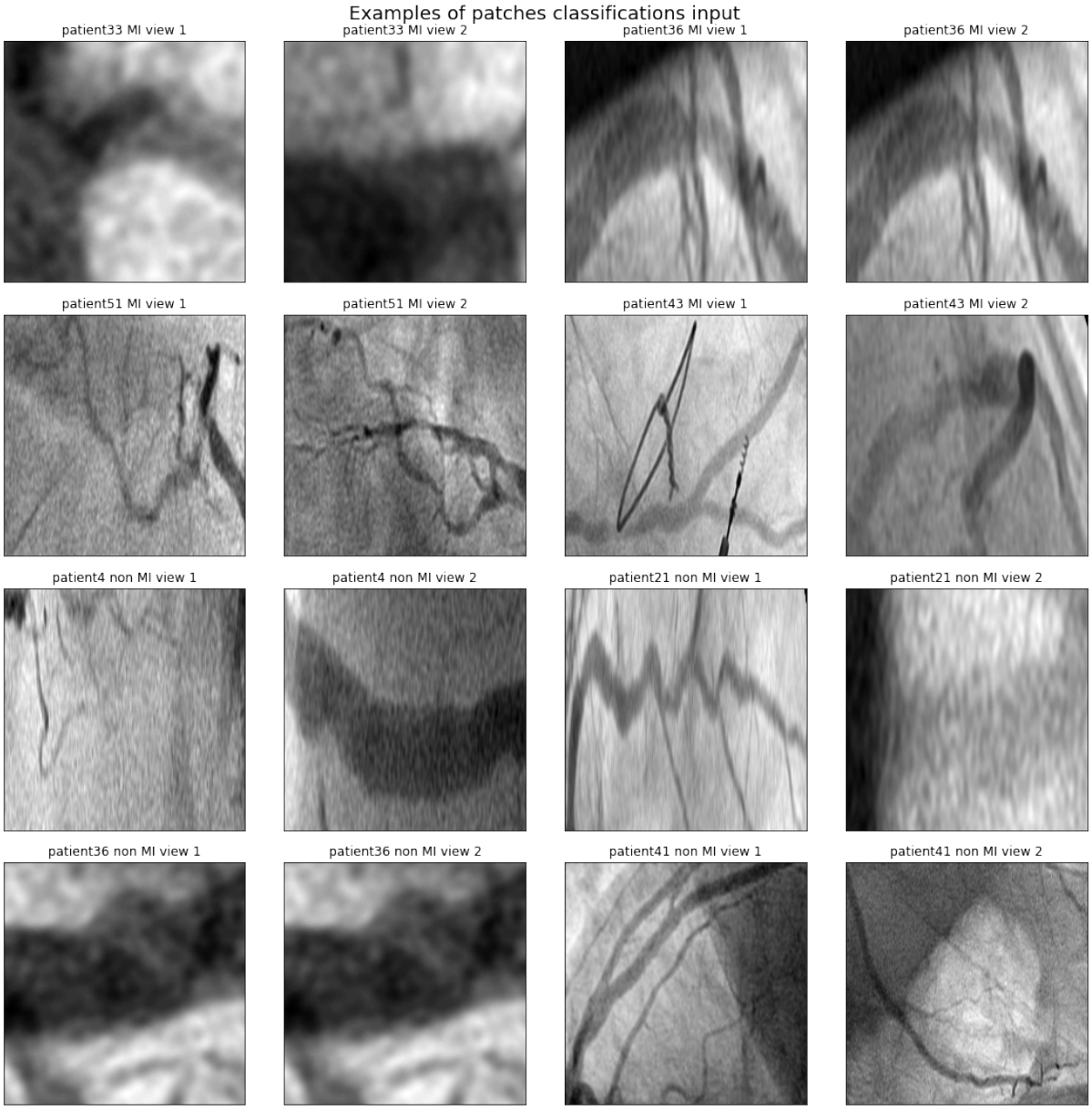


Figure 18: Examples inputs for the patches classification approach, two views are given as input to the model, thus *view 1* and *view 2* refer to the two inputs. Therefore, 8 inputs are presented 4 MI top and 4 non MI bottom.

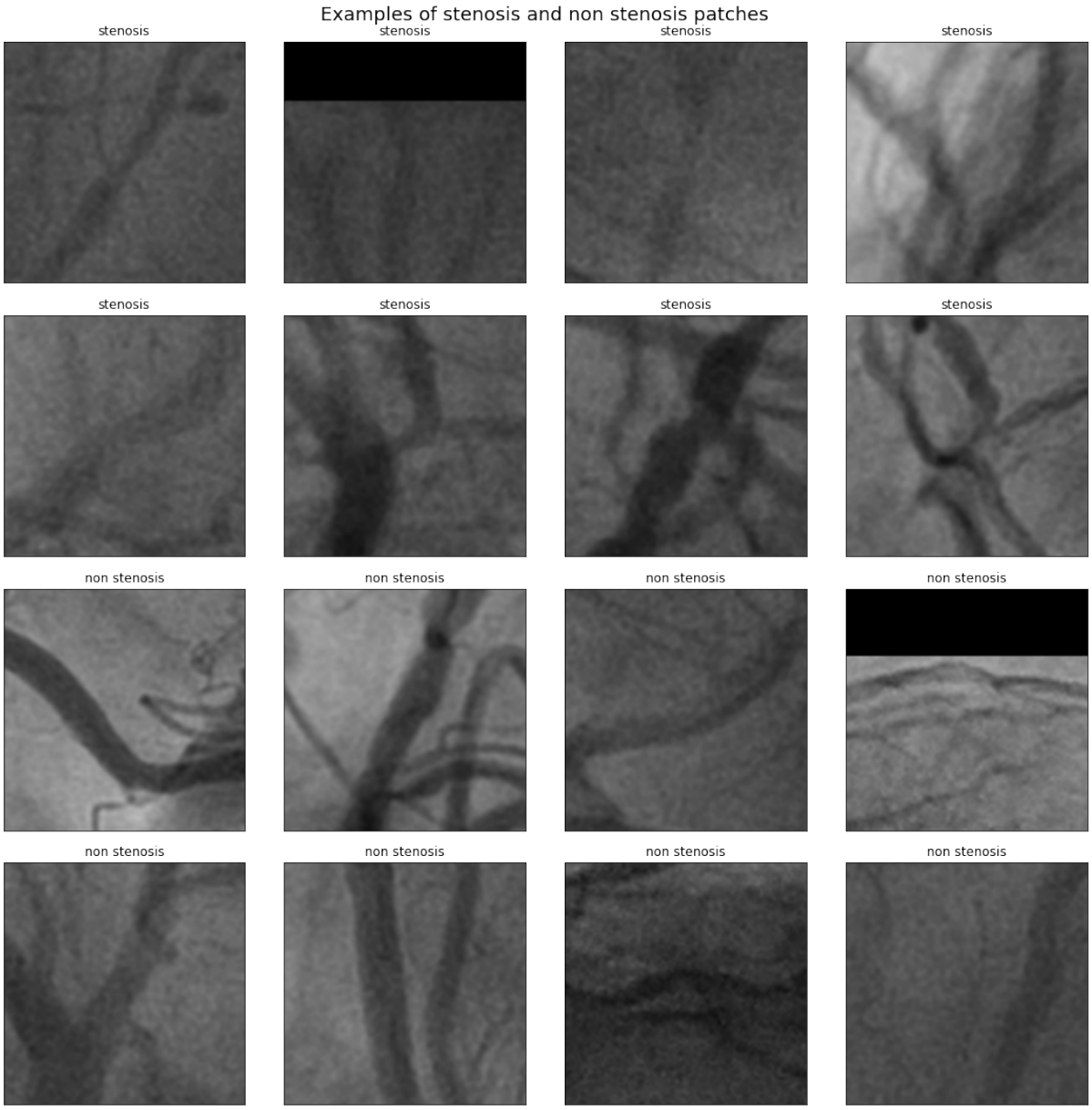


Figure 19: Examples of inputs for the MI stenosis classification approach. 16 inputs are presented 8 MI stenosis top and 4 non MI stenosis bottom.

B Patches classification

In this section, the results of the different experiment is shown as an information complement.

B.1 Results Siamese model

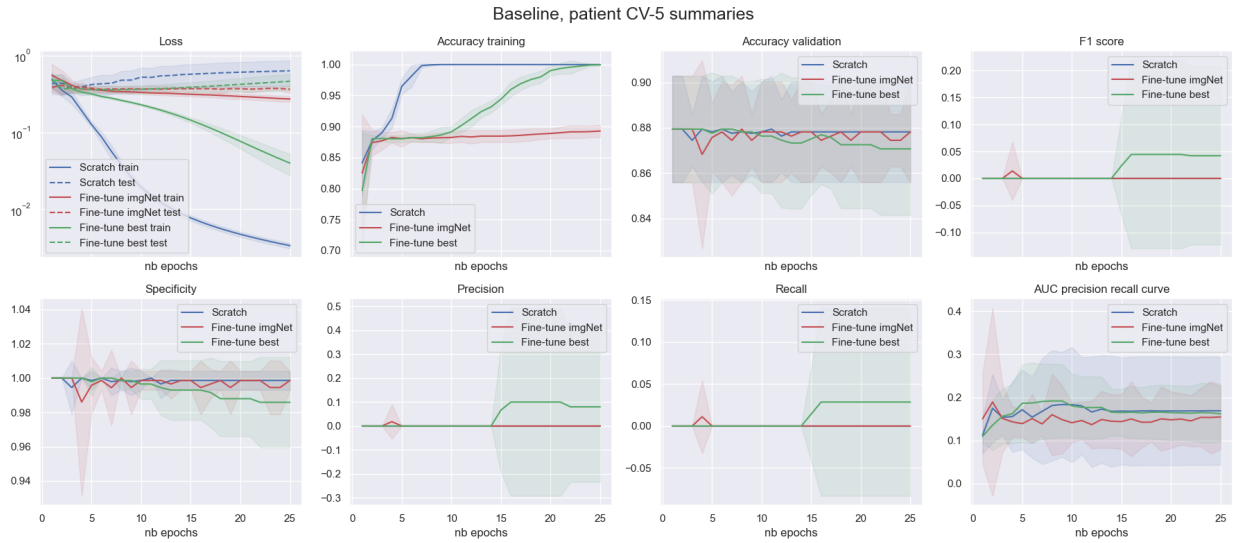


Figure 20: Baseline experiment for Siamese model on patches classification.

Figure 21: Weighted loss experiment for Siamese model on patches classification.

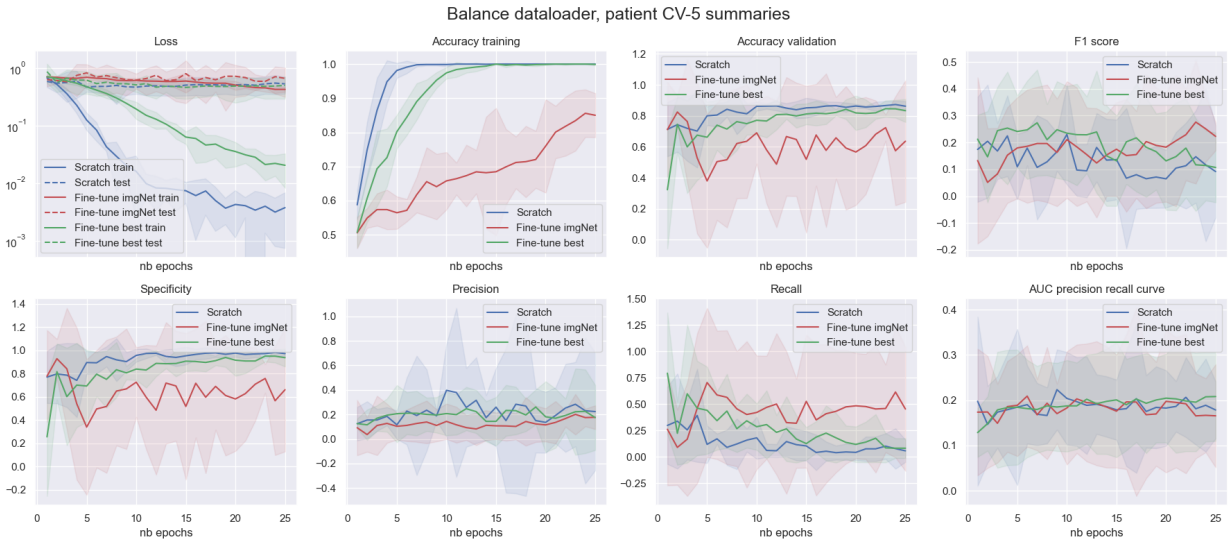


Figure 22: Balance dataloader experiment for Siamese model on patches classification.

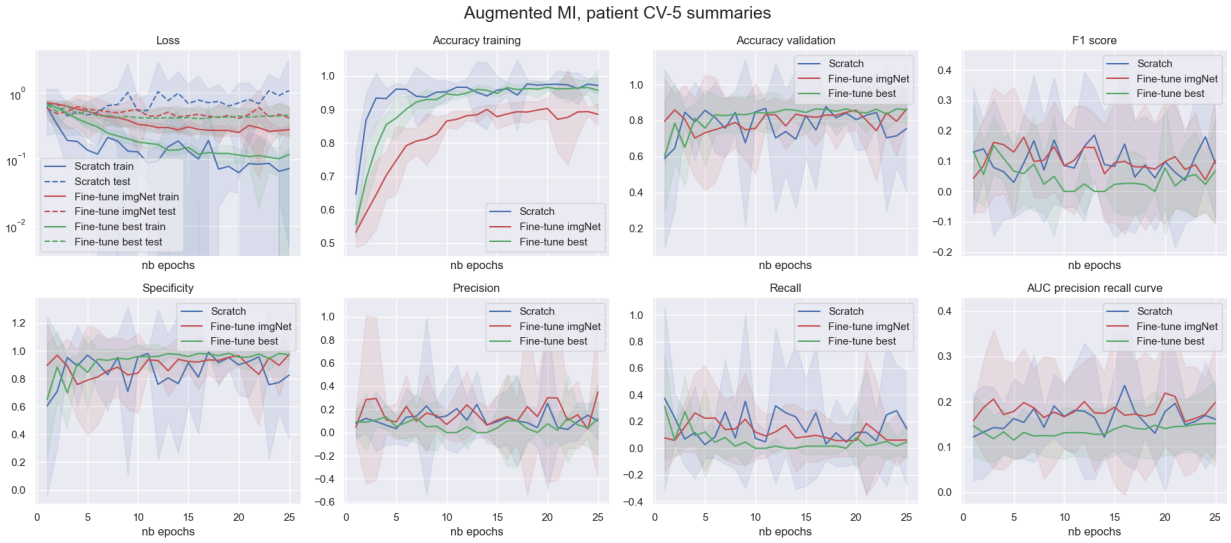


Figure 23: Augmented MI experiment for Siamese model on patches classification.

C Custom attention

C.1 Results attention as channel

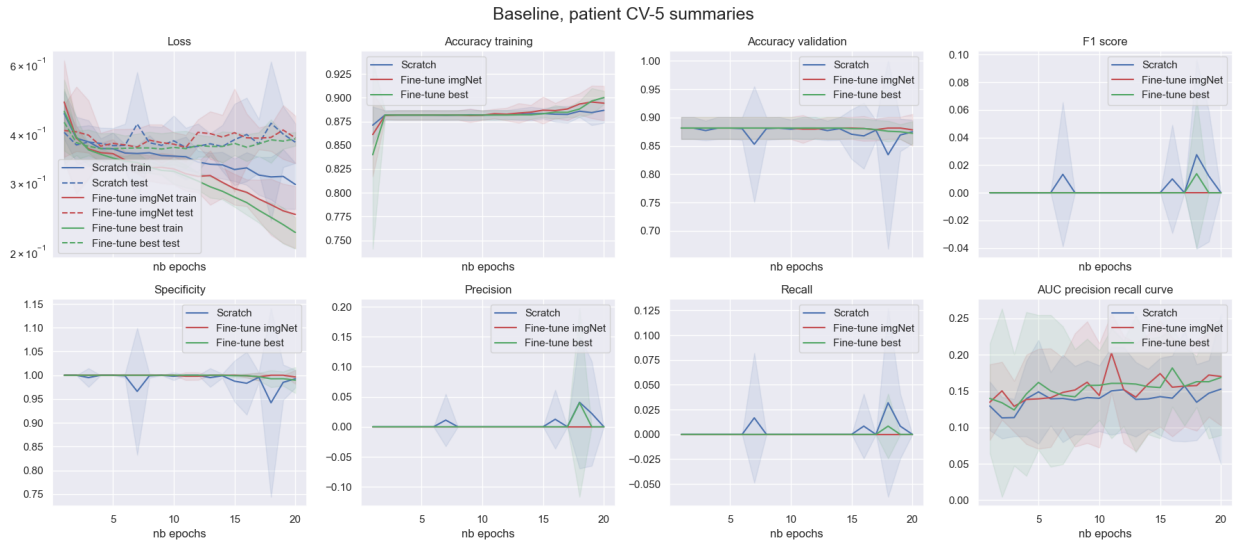


Figure 24: Baseline experiment for channel input on custom attention.

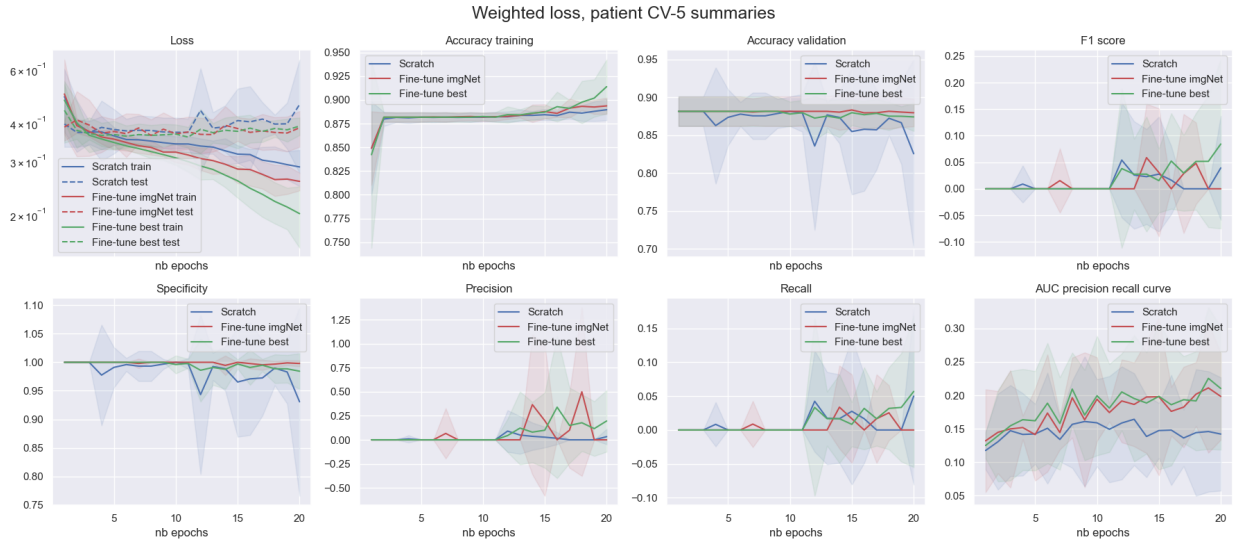


Figure 25: Weighted loss experiment for channel input on custom attention.

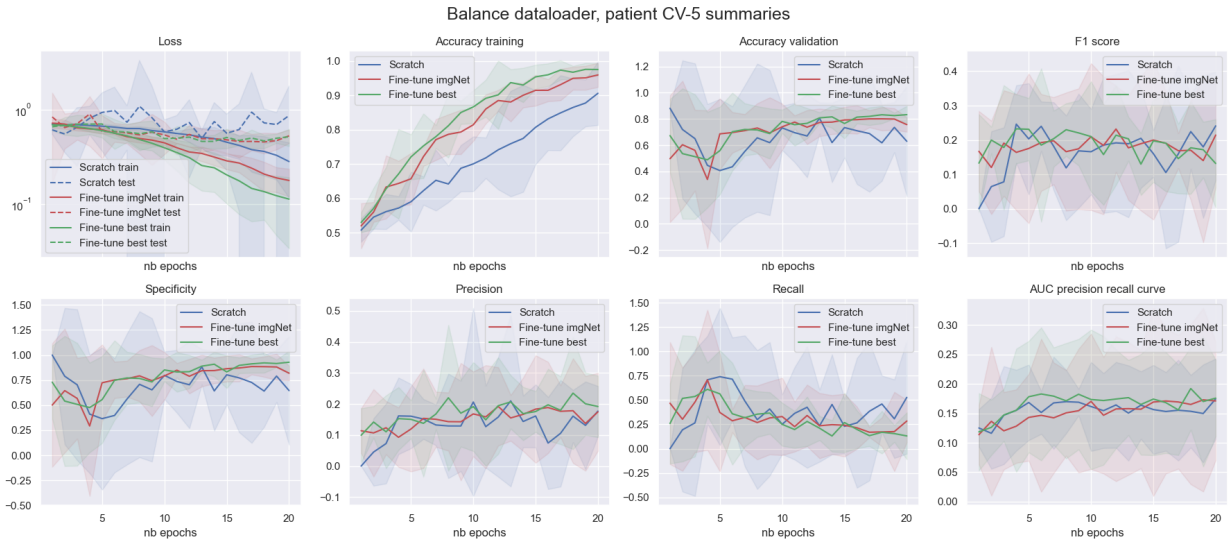


Figure 26: Balance dataloader experiment for channel input on custom attention.

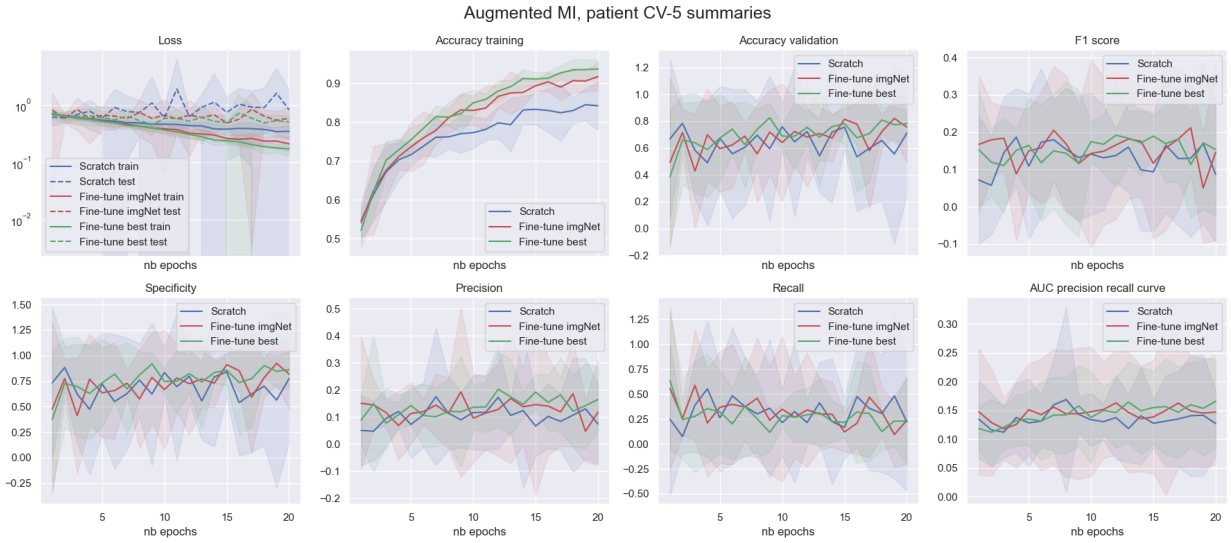


Figure 27: Augmented MI experiment for channel input on custom attention.

C.2 Results attention apply

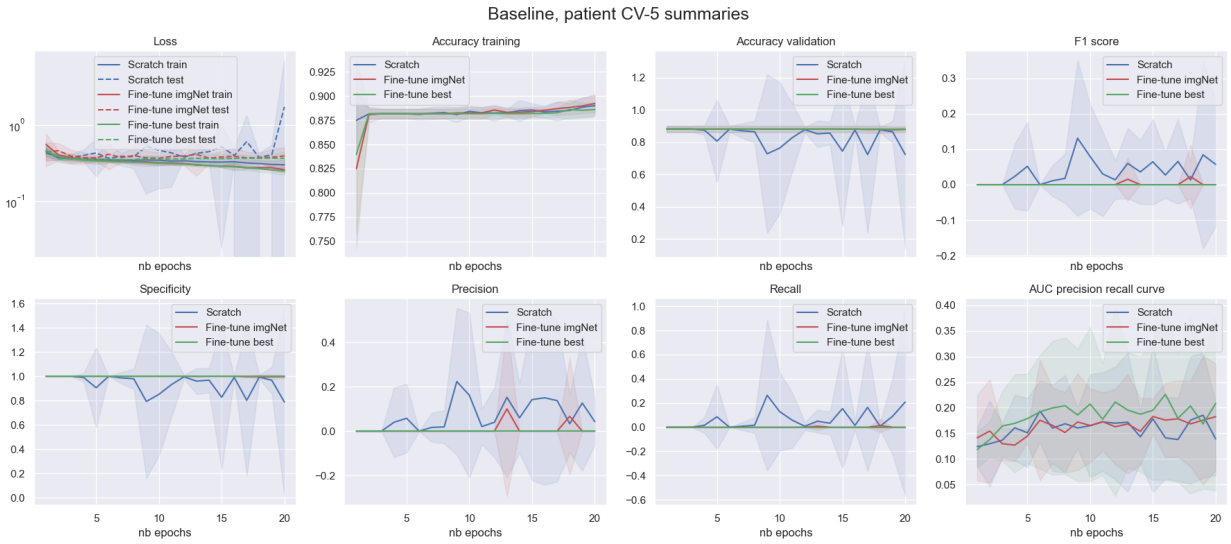


Figure 28: Baseline experiment for applying attention input.

Figure 29: Weighted loss experiment for applying attention input.

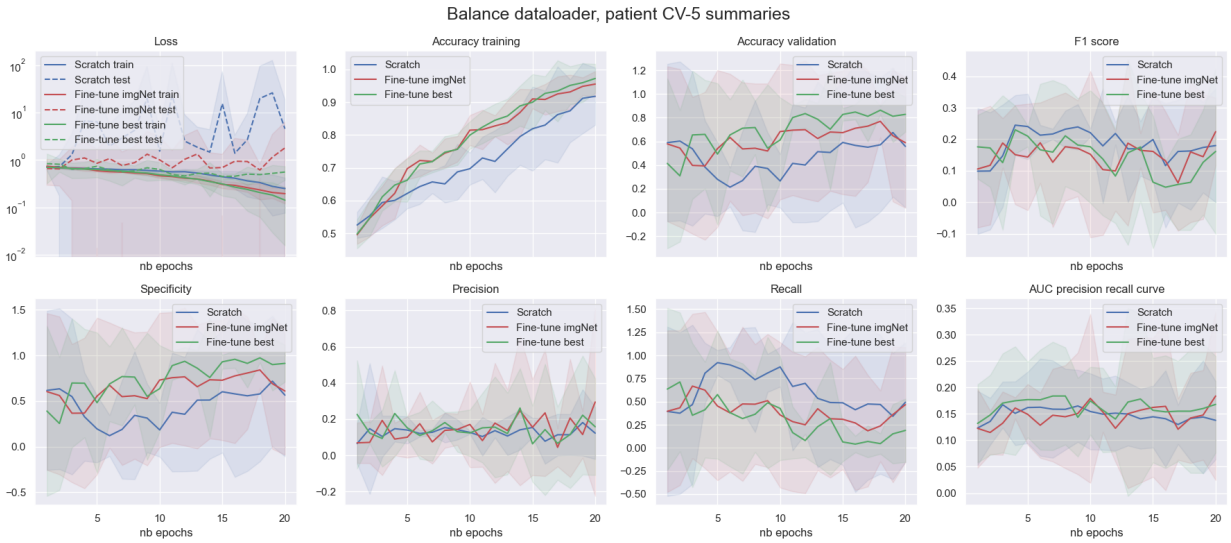


Figure 30: Balance dataloader experiment for applying attention input.

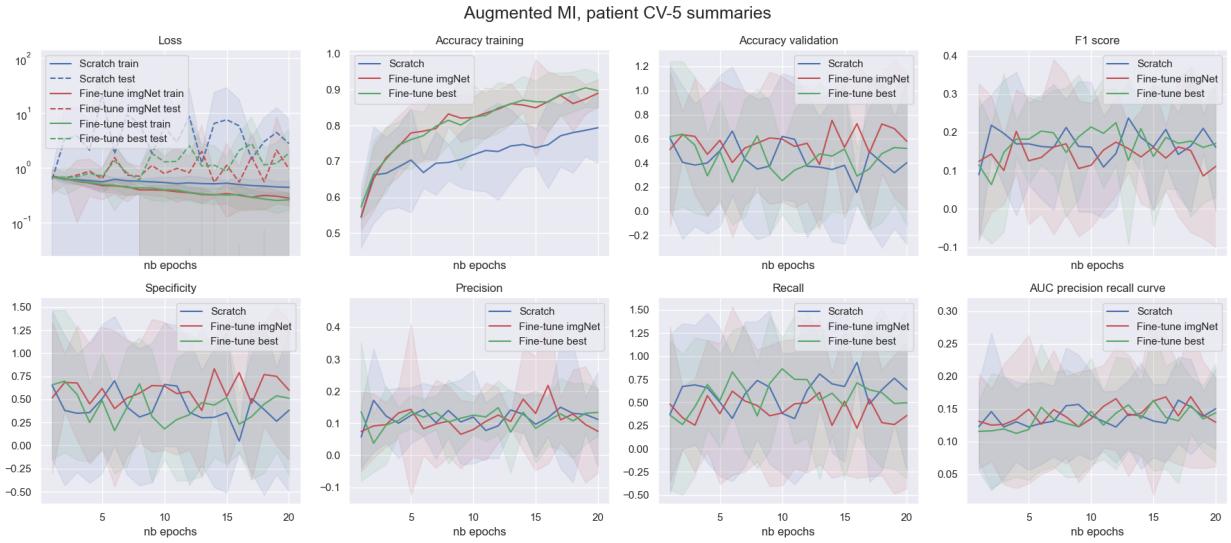


Figure 31: Augmented MI experiment for applying attention input.