

Anatomy-Informed Multimodal Learning for Myocardial Infarction Prediction

Ivan-Daniel Sievering^{ID}, Ortal Senouf, Thabo Mahendiran^{ID}, David Nanchen^{ID}, Stephane Fournier^{ID}, Olivier Muller^{ID}, Pascal Frossard^{ID}, *Fellow, IEEE*, Emmanuel Abbé, and Dorina Thanou^{ID}, *Senior Member, IEEE*

Abstract—Goal: In patients with coronary artery disease, the prediction of future cardiac events such as myocardial infarction (MI) remains a major challenge. In this work, we propose a novel anatomy-informed multimodal deep learning framework to predict future MI from clinical data and Invasive Coronary Angiography (ICA) images. **Methods:** The images are analyzed by Convolutional Neural Networks (CNNs) guided by anatomical information, and the clinical data by an Artificial Neural Network (ANN). Embeddings from both sources are then merged to provide a patient-level prediction. **Results:** The results of our framework on a clinical study of 445 patients admitted with acute coronary syndromes confirms that multimodal learning increases the predictive power and achieves good performance (AUC: 0.67 ± 0.04 & F1-Score: 0.36 ± 0.12), which outperforms the prediction obtained by each modality independently as well as that of interventional cardiologists (AUC: 0.54 & F1-Score: 0.18). **Conclusions:** To the best of our knowledge, this is the first attempt towards combining multimodal data through a deep learning framework for future MI prediction. Although it demonstrates the superiority of multi-modal approaches over single modality, the results do not yet meet the necessary criteria for practical application.

Index Terms—Coronary artery disease, deep learning, invasive coronary angiography, multimodal data, myocardial infarction.

Impact Statement— Our novel multimodal deep learning framework, integrating clinical data and coronary angiography images, outperforms single-modality methods and expert predictions in predicting myocardial infarction risk.

I. INTRODUCTION

CORONARY artery disease (CAD), a leading cause of death worldwide [1], refers to a disease of the coronary arteries that supply blood to the heart muscle. The disease results, primarily, from the development of plaques of atherosclerosis in the arterial wall, which ultimately lead to narrowings (stenoses) and reduced blood flow. In the acute setting, CAD takes the form of an acute coronary syndrome (ACS), with the most feared manifestation being a myocardial infarction (MI) resulting from the rupture of a plaque of atherosclerosis and the subsequent, abrupt interruption of coronary blood flow. The resulting necrosis of the heart muscle can lead to numerous complications including a reduction in heart function (heart failure), arrhythmia (including cardiac arrest), and death.

However, CAD is a complex pathological process with numerous factors that drive its development, its progression, and its risk of provoking an MI. At a local level, the diameter of the stenosis does correlate with the risk of MI, but it remains insufficient as a predictor of MI, as highlighted by the influence of other local factors such the hemodynamic impact of the stenosis [2]. At a patient level, important drivers of CAD include cardiovascular risk factors (e.g., age, sex, hypertension, diabetes, dyslipidemia), all of which can influence the risk of MI. As a result, in clinical practice, future MI prediction remains a challenge. This is highlighted by the fact that up to 10% of patients with stenoses deemed non-significant (i.e., stenoses < 50% without a significant hemodynamic impact) still present an MI or a need for urgent revascularization (i.e., invasive treatment such as stenting) in the ensuing two years [3]. Even among patients treated for an MI, the risk of short- and long-term adverse outcomes remains significant. Reinfarction represents a significant cause of poor outcomes among MI patients, with rates as high as 4% at one year and 7% at three years [4].

In both acute and chronic settings, invasive coronary angiography (ICA) remains the gold standard investigation for the diagnosis and evaluation of CAD in clinics. ICA involves

Manuscript received 10 September 2023; revised 22 March 2024 and 7 May 2024; accepted 8 May 2024. Date of publication 27 May 2024; date of current version 7 November 2024. This work was supported by the Center for Intelligent Systems, EPFL, Lausanne, Switzerland. The review of this article was arranged by Editor Riccardo Barbieri. (Corresponding author: Ivan-Daniel Sievering.)

Ivan-Daniel Sievering is with the Signal Processing Laboratory 4, EPFL, 1015 Lausanne, Switzerland. He is now with the Swiss Data Science Center, ETH Zurich and EPFL, 1015 Lausanne, Switzerland (e-mail: ivan.sievering@epfl.ch).

Ortal Senouf is with Signal Processing Laboratory 4, EPFL, 1015 Lausanne, Switzerland, and also with the Chair of Mathematical Data Science, EPFL, 1015 Lausanne, Switzerland (e-mail: ortal.senouf@epfl.ch).

Thabo Mahendiran is with the Department of Cardiology, CHUV, 1005 Lausanne, Switzerland, and also with the Chair of Mathematical Data Science EPFL, 1005 Lausanne, Switzerland (e-mail: thabo.mahendiran@epfl.ch).

David Nanchen is with the Unisante, 1011 Lausanne, Switzerland (e-mail: david.nanchen@unisante.ch).

Stephane Fournier and Olivier Muller are with the Department of Cardiology, CHUV, 1005 Lausanne, Switzerland (e-mail: stephane.fournier@chuv.ch; olivier.muller@chuv.ch).

Pascal Frossard is with the Signal Processing Laboratory 4, EPFL, 1015 Lausanne, Switzerland (e-mail: pascal.frossard@epfl.ch).

Emmanuel Abbé is with the Chair of Mathematical Data Science, EPFL, 1015 Lausanne, Switzerland (e-mail: emmanuel.abbe@epfl.ch).

Dorina Thanou is with the Center for Intelligent Systems, EPFL, 1015 Lausanne, Switzerland (e-mail: dorina.thanou@epfl.ch).

Digital Object Identifier 10.1109/OJEMB.2024.3403948

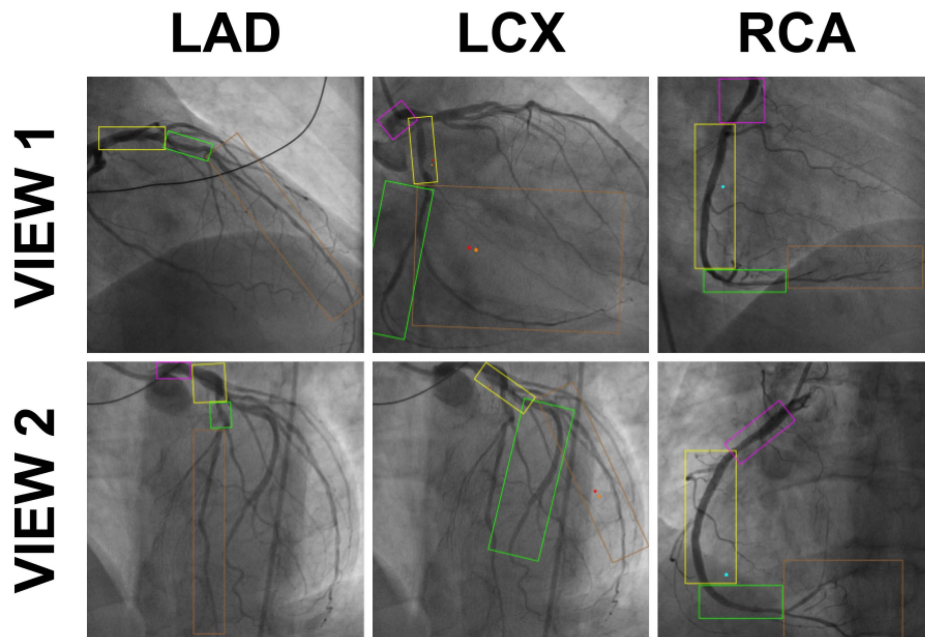


Fig. 1. Annotated ICA images of a patient; two views for each of three different arteries. The boxes indicate different anatomical segments and the red dots that the segment is responsible of future MI.

continuous X-ray (i.e., fluoroscopy) with simultaneous injection of radiopaque contrast into the coronary arteries, thus permitting the identification of coronary stenoses. In current clinical practice, stenosis severity is still often only determined by the physician's estimation of the percentage reduction in arterial diameter. With this approach, a diameter stenosis $\geq 70\%$ is generally considered a strong indicator of a clinically significant lesion, and thus a criterion for treatment (e.g., coronary stenting). New innovative approaches are thus needed to drive progress in this field.

Recent advances in machine learning (ML), in combination with the availability of multimodal data, show promise for capturing and quantifying the complexity of CAD. Some works have already focused on predicting MI in the next months from patients' clinical data using traditional ML algorithms such as Logistic Regression, Random Forest, and Gradient Boosting, (e.g., [5], [6]), with limited success. At the same time, convolutional neural networks (CNNs) have been successfully applied in detecting stenoses and inferring their severity directly from ICA images [7], but without tackling the challenge of future event prediction. First step towards future MI prediction from ICA images was performed in [8] and [9], where a deep learning framework was able to provide significant gains in predicting future culprit lesions, i.e., lesions that lead to future events. This study however was performed at a lesion level and on patients with stable coronary disease. Given the complexity of CAD and the numerous influencers of the risk of future MI, a multimodal approach that extracts knowledge from all available patient data appears crucial for predicting future events.

In this work, we depart from the task of lesion-level MI prediction from ICA images, instead tackling the challenge of patient-level MI prediction using both ICA and clinical information. In particular, we propose an anatomy-informed deep learning framework that combines ICA imaging views

from the three different arteries of the coronary tree (see Fig. 1 for an example of the different arteries), cardiologist guidance on significant anatomical regions, and clinical data in order to predict the occurrence of future MI in patients presenting with an acute coronary syndrome at baseline. The problem is particularly challenging for ML settings not only due to the systemic and biological complexity of the disease but also due to the frequency of the disease. Due to its invasive nature, patients undergo ICA only if there is significant justification. Moreover, out of all these patients, only a small percentage will experience an MI in the future, limiting further the number of MI events. This low and unbalanced data regime is particularly challenging for ML algorithms.

To partially overcome these issues, we exploit our knowledge of the coronary artery tree, and cast the problem as a multi-objective learning framework, with the goal to predict MI both at an artery (auxiliary task) and at a patient level (main task). The MI prediction at an artery level is achieved by learning discriminative features from the corresponding artery views using anatomy-informed CNNs, combined with learned representations of the clinical data, computed by an ANN. These jointly learned artery-specific representations are then concatenated to obtain a prediction at a patient level, which is provided by another ANN. The results obtained in a clinical cohort of 445 patients confirm that considering jointly both image modalities and clinical data (AUC: 0.67 ± 0.04 & F1-Score: 0.36 ± 0.12) provides significantly better predictive power than learning from each modality independently (only using ICA images: AUC: 0.64 ± 0.04 & F1-Score: 0.30 ± 0.11 and only using clinical data: AUC: 0.63 ± 0.04 & F1-Score: 0.28 ± 0.06). The multi-modal learning scheme also outperforms the prediction of interventional cardiologists from the ICA images (AUC: 0.54 ± 0.04 & F1-Score: 0.18 ± 0.04). To the best of our knowledge, this is the first attempt towards providing a global prediction score

for future coronary artery events by exploiting, in a data-driven manner, conjointly the patients' clinical data and their ICA images.

The rest of the paper is organized as follows. First, we provide a description of the clinical study, the different data modalities, and the cardiologist-guided annotation of the anatomical segments of the coronary artery tree. Then, we present our novel multi-objective learning approach for predicting MI from multimodal data. Finally, we illustrate the performance of our framework on the clinical study and compare it with different baselines that rely on learning from a single modality or inferring from interventional cardiologist expertise.

II. MATERIALS AND METHODS

A. Clinical Study

The SPUM-ACS (Special Program University Medicine - Acute Coronary Syndromes) registry is a cohort of consecutive patients admitted with acute coronary syndromes (MI or unstable angina) to four university hospitals in Switzerland between 2009 and 2017. Further details of this registry have been reported previously [10]. For the present study, patients hospitalized with ICA images available for analysis are included. The clinical endpoint considered in this work is MI, i.e., we consider patients who had an MI in the next five years after the acute event (the acute coronary syndrome at baseline). Our approach could be generalized to other datasets that include enough ICA views and similar clinical information about the patient.

1) ICA Images: The dataset consists of clinical data and ICA images of 445 patients, out of which 47 experienced an MI during the follow-up period. Six ICA images were extracted from the baseline ICA of each patient: three arteries (left anterior descending (LAD), left circumflex (LCX), and right coronary artery (RCA)) viewed from two angles with $\approx 30^\circ$ of difference. An illustrative example is shown in Fig. 1. Most of the ICA images are 1524×1524 pixels in size. Those that have different sizes are cropped or interpolated to this size. Only patients with at least one view of each artery are considered. In order to have a fixed number of views per patient, for patients with only one view, we consider that view twice, while for patients with more than two views, we randomly select two views per patient. The initial number of patients in the dataset is 709, but for 264 of them, the dataset does not contain at least one view of each artery, or contain invalid values, and thus only 445 patients are considered.

2) Cardiology-Guided Annotation of Anatomical Segments: Each of the views is annotated by an interventional cardiologist into anatomical segments as defined by the SYNTAX system [11] (colored boxes in Fig. 1, with the color code being consistent across patients). Each segment is then labeled as being responsible for future MI or not (red dots in Fig. 1). The anatomical segments will be used as attention masks in order to guide the learning of the proposed algorithm toward important regions of the coronary artery tree.

3) Clinical Data: The clinical data consists of sex, age, body mass index, diabetes, smoking, hypertension, hypercholesterolemia, previous cardiovascular disease, Killip class, previous

cardiac arrest, and kidney function, which are known to be cardiovascular risk factors. The non-categorical data are normalized (mean subtracted and divided by the standard deviation). Each missing data is replaced by the median, if it belongs to a continuous column (i.e., values of the column can take any value in a given range), or by the most frequent value if it belongs to a categorical column (i.e., values of the column come from a given set of possible values). The distributions of those features are presented per group in Table I.

B. A Multimodal Learning Framework for MI Prediction

1) Proposed Model: We propose a novel, anatomy-informed learning framework that predicts MI by combining coronary artery anatomical information from ICA images with patient-level clinical data, e.g., cardiovascular risk factors. Our approach takes the images of the different ICA views of the artery as input along with the patient's clinical data, permitting the prediction of MI at an artery level. These artery-level predictions are then combined in order to predict whether the patient will have an MI. A summary of the architecture is illustrated in Fig. 2. In what follows, we elaborate on each building block of the architecture, and each data modality.

Imaging data: To extract predictive features from ICA images, we adapt state-of-the-art image-based deep learning architectures to the specificity of the coronary artery images. For each main artery of the coronary tree, we propose a CNN-based architecture, which receives as input the two views of the artery as well as the corresponding attention masks that indicate the bounding boxes drawn by the physicians. These masks help the network focus on the main anatomical segments of the coronary artery tree and not on the background, see Fig. 3. Given that both views represent a snapshot of the same 3D artery from different angles in the 2D space, we use this anatomy-informed fact to process them jointly through two siamese networks as we expect that similar features are present in both pictures. The backbone is a ResNet-18 network [12], which is a well-known state-of-the-art model. We then achieve a global representation of the ICA images by concatenating the representations of both views, followed by average pooling, flattening, and dropout, in order to reduce the dimension of the embedded space and to diminish overfitting. Finally, this global representation of the views is given as input to a classification layer, which is responsible for predicting MI in each artery only from imaging data. This global representation will be used for the multimodal prediction, jointly with the embedding of the clinical data.

Clinical data: In parallel, clinical data are processed through two hidden layers of 50 and 10 neurons (Fully connected (FC) layer, batch normalization, ReLU activation, dropout), inspired by [13]. The feature representation extracted from this pipeline defines the embedding of the clinical data. This embedding is given as an input to a classification layer (FC, Sigmoid), which is responsible for predicting patient-level MI only from clinical data, which provides another auxiliary loss. This embedding will also be used for the multimodal prediction, jointly with the images' embedding.

TABLE I
DEMOGRAPHIC TABLE: DISTRIBUTION OF THE CLINICAL FEATURES W.R.T. GROUPS

	All		MI		Non-MI	
	Value	Miss.	Value	Miss.	Value	Miss.
Age	63.5±12.6	0.0%	65.8±12.7	0.0%	63.3±12.5	0.0%
BMI	26.9±4.3	0.3%	27.6±3.8	3.8%	26.9±4.3	0.3%
Kidney function	81.3±23.7	3.2%	72.3±22.9	39.7%	82.1±25.8	3.4%
Sex (% of females)	23.1	0.0%	15.4	0.0%	23.7	0.0%
Diabetes	19.5	0.0%	37.2	0.0%	18.0	0.0%
Hypertension	55.1	0.0%	66.7	0.0%	54.1	0.0%
Hypercholesterolemia	61.8	0.1%	70.5	1.3%	61.0	0.1%
Previous CVD	29.8	0.0%	57.7	0.0%	27.4	0.0%
Previous cardiac arrest	3.4	0.3%	2.6	3.8%	3.4	0.3%
Smoking (% of >0)	58.3	0.0%	61.5	0.0%	58.1	0.0%
Killip class (% of >0)	13.4	1.0%	17.9	12.8%	13.0	1.1%

Mean±standard deviation for the age, BMI, and kidney function. % of positive for the other features. The 'Miss.' columns are the % of missing values.

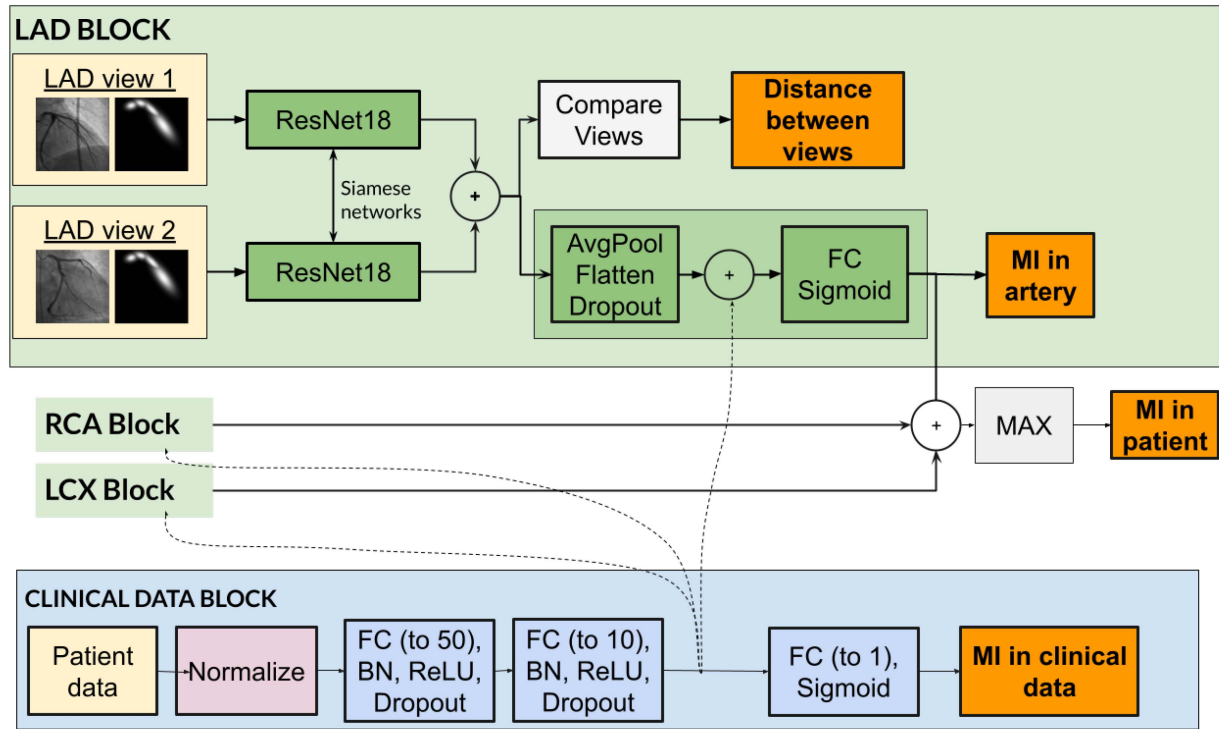


Fig. 2. Anatomy-informed multimodal framework for MI prediction. The patient data is processed by the clinical data block and the two views of each artery are processed by artery blocks. A prediction is provided for each artery and the prediction at the patient level is the maximum of these three predictions. Note that the decision at the artery level is influenced by the clinical data.

Patient-level prediction (multimodal model): Once ICA and clinical data have been processed independently, their representations are combined in order to achieve a multimodal embedding per artery. We follow a strategy similar to [14] and [15], by concatenating the representations of the ICA images and the clinical data, and further analyzing them through an FC layer activated by a Sigmoid function. This output is used to compute the probability of an MI for a given artery. Finally, the prediction at the patient level is defined as the worst-case prediction of

the three arteries, i.e., a single MI prediction at an artery level is enough to predict MI at a patient level. The independent prediction for each artery is motivated by the anatomy-informed fact that MI is often a local event, i.e., information from an artery does not necessarily improve the prediction in another one.

2) Training Procedure: Our multimodal network is trained by minimizing a loss function ($\mathcal{L}$) that takes into account each of the building blocks mentioned above. In particular, we aim to improve the overall MI predictions by

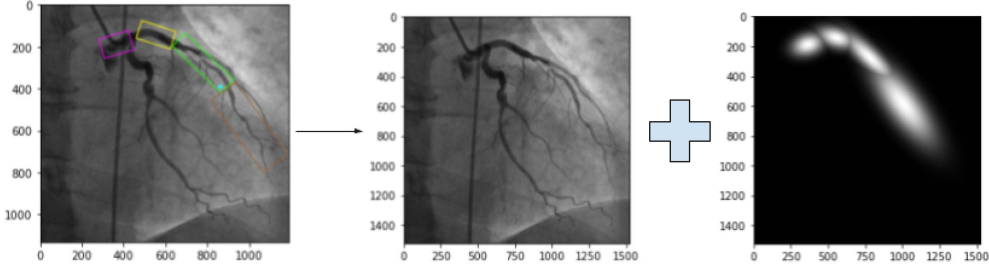


Fig. 3. Annotated ICA images (left) are converted to a raw image (center) and a mask (right) that indicates the different anatomical segments. The mask is created by generating Gaussian functions centered on the sections' rectangle and using the same width and height.

optimizing the artery level predictions from multimodal embeddings ($l_{pred(LAD/LCX/RCA)}$), the patient level predictions from clinical data ($l_{clinical}$), and the overall patient-level prediction from both clinical and artery level embeddings ($l_{patient}$). In addition, for each artery view, we define an auxiliary loss function that enforces similarity between views in the embedding space ($l_{simi(LAD/LCX/RCA)}$). The loss will be zero if the views share the same information (this loss helps the model to learn the features to detect as we expect that both views mostly contain similar information). This function is set to be the mean Euclidean distance between the features extracted from the two pairs of views of the same artery.

Thus, the complete loss that consists of the above-mentioned terms, can be defined as presented in (1) and (2). The weight of the main loss (patient level prediction) is always fixed to one as it is our main target and the others ($w = [w_1, w_2, w_3]$) are considered as hyperparameters (ranging from 0 to 1).

$$l = l_{patient} + l_{aux}, \quad (1)$$

where

$$l_{aux} = w_1 * l_{clinical} + \sum_{n=artery}^{[LAD, LCX, RCA]} w_2 * l_{pred(n)} + w_3 * l_{simi(n)}. \quad (2)$$

Given the imbalanced nature of our dataset (47 MI out of 445), we employ the AUC Margin loss function as classification loss [16] along with a PESG optimizer [17]. This loss function serves as a margin-based surrogate for optimizing the AUC score. It has shown (along with the PESG optimizer) good performances on the classification of large-scale medical image datasets [16]. Additionally, to tackle the imbalance, the MI samples of the dataset are oversampled (the minority samples are duplicated until parity is reached). Moreover, different image-based augmentation methods are applied to ICA images in the training set (no data augmentation is applied to the clinical data). Each data augmentation technique has a given probability of happening, and is applied sequentially to the images in the following order:

- 1) Random cropping (probability of happening: 20%): a subimage is cropped on the image, which can have a ratio between 75% and 125% of the original image and

- between 80% and 99% of the size of the image. The image is then resized to the input size (bi-linear interpolation);
- 2) Random rotation (probability of happening: 20%): the image is rotated between -30° and $+30^\circ$;
- 3) Color (probability of happening: 20%): brightness is altered between 80% and 120%, contrast between 80% and 120%, saturation between 80% and 120% and hue between -20% and 20% .

For the cropping and rotation, the same values are applied to the raw image and its mask (otherwise, the mask does not provide the correct insight). Quantitative analysis regarding the impact of these augmentation methods have not been conducted; however, their implementation was deemed necessary to prevent overfitting the duplicated inputs. Thus, the parameters' values have been fixed heuristically.

The weights of the ResNet-18 backbones are initialized with a pre-trained network on ImageNet [18] (it was found that pre-trained models reach better performance, even though our images are not “natural” images). The models are trained using 5-fold cross-validation on the training set, with a similar number of MI patients in each fold. The multimodal network is trained during 20 epochs with PESG (gamma: 595, margin: 0.99) and a batch size of 4. The weight decay and the dropout are fixed at 0.003, and 31% respectively. The learning rate is initially set to 0.077, and a scheduler divides by 10 the learning at each plateau (3 consecutive epochs without metric improvement). The weight of artery loss, distance loss, and clinical data loss are set to 0.06, 0.007 and 0.006 respectively. All these hyperparameters have been selected using naive grid search (iterating over all the parameter combinations in the selected range).

III. RESULTS

In this section, we evaluate our multimodal frameworks for MI prediction on the clinical study presented in Section II-A. First, we introduce our baseline methods that are based on predicting from each modality independently, as well as a comparison with the prediction score achieved from the visual inspections of interventional cardiologists. Next, the obtained results are discussed.

A. Baselines

- a) **Single modality: ICA:** We truncate the multimodal model so that it uses only the ICA images, i.e., without the

TABLE II
DIFFERENT MODALITIES: VALIDATION SCORES (MEAN $\pm$ STD) OBTAINED FROM A 5-FOLD CROSS-VALIDATION ON THE TRAINING SET

Predictor	F1-Score	AUC-ROC	Precision	Recall
Naive	0.10	0.50	0.10	0.10
Interventional Cardiologist 1	0.14	0.49	0.10	0.25
Interventional Cardiologist 2	0.22	0.58	0.15	0.47
Cardiologists' mean	0.18	0.54	0.13	0.36
Single modality: Clinical data	0.28 $\pm$ 0.06	0.63 $\pm$ 0.04	0.18 $\pm$ 0.04	0.72$\pm$0.17
Single modality: ICA images	0.30 $\pm$ 0.11	0.64 $\pm$ 0.04	0.21 $\pm$ 0.13	0.69 $\pm$ 0.11
Multimodal framework	0.36$\pm$0.12	0.67$\pm$0.04	0.36$\pm$0.18	0.44 $\pm$ 0.10

“Clinical Data Block” in Fig. 2. We compute predictions per artery and define the patient-level prediction as the worst-case scenario from each artery. The model is trained using the AUC-loss (as for the multimodal).

The model is trained for 20 epochs with a PESG optimizer (gamma: 411, margin: 0.81) and a batch size of 4. The weight decay is set to 0.007 and the dropout to 0.6%. The starting learning rate is 0.03, divided by 10 after 3 epochs without improvement. The weight of the artery loss is 0.002, and the weight of the distance loss is 0.0008.

b) Single modality: Clinical data: We use only the “Clinical Data Block” from the multimodal model presented in Fig. 2 and thus predict MI only based on clinical data.

The model is trained during 300 epochs with PESG (gamma: 470, margin: 0.92) and a batch size of 32. A Kaiming Normal [19] is used for initialization. The weight decay is 0.0048, and the dropout is 49.76%. The starting learning rate is 0.004, divided by 10 after 25 epochs without improvement.

c) Visual inspection from interventional cardiologists: The exact same set of patients is evaluated by two blinded interventional cardiologists, who analyzed the ICA views of each patient and provided their patient-level prediction of MI.

d) Naive predictor: A naive strategy is applied: it classifies the sample as positive 10.6% of the time, and otherwise as negative. This probability corresponds to the distribution of MI in the dataset (47 MI among 445 patients (10.6%)). Thus, this benchmark shows the performance of random guesses.

B. Evaluation Metrics

First, the dataset is split into a training set and a testing set (while conserving the same ratio of positive and negative in each). Then, the performances are computed by applying a 5-fold stratified procedure on the training set only; the mean and standard variation across the validation folds are computed. Finally, a model is trained on the whole training set and tested on the testing set, which was not used in the previous step.

Four evaluation metrics [20] are considered:

- AUC-ROC measures the balance between the True Positive Rate and the False Positive Rate;
- Precision measures the percentage of True Positive among the samples classified as positive;
- Recall measures the percentage of positive correctly classified;

- F1-Score is the harmonic mean of precision and recall.

C. Performance Analysis

In Table II, we compare the MI predictive performance of the proposed multimodal framework, as well as the baselines mentioned above during the 5-fold cross-validation procedure. We observe that the performance of our multimodal framework is better than learning from each modality independently, highlighting the benefit of extracting predictive features from both. The performances obtained only from clinical data and only from ICA are similar, highlighting the difficulty of learning directly from ICA. This is most likely attributed to several reasons with the main ones being the limited number of data points, and in particular MI patients, and the important quantity of uninformative background in ICA images. At the same time, the multimodal approach outperforms learning from each of the single modalities independently. These results confirm our intuition that each modality contains different information that, if combined properly, can provide a better predictive score at a patient level. The low predictive performance achieved from expert clinicians indicates the complexity of the task and the importance of building data-driven tools that could assist clinical decision-making.

Table III shows the performances but by training on the whole training set and testing on the testing dataset. These results have to be considered with great care, the main reason being that the testing dataset contains very few positive cases (5 MI over 89 patients). Moreover, the results are outside of the confidence intervals computed from the validation dataset. However, these results lead to similar conclusions to the ones presented previously.

D. Alternative Models

a) Alternative single modality: ICA: Different variations of the single modality ICA model are considered. First, two losses are used for training, i.e., the AUC-loss (as for the multimodal) and the Binary Cross Entropy (BCE) loss (optimized by Stochastic Gradient Descent (SGD)). Second, two different ways for providing patient predictions are compared: (i) *Max analysis*: We compute predictions per artery and define the patient level prediction as the worst-case scenario from each artery. (ii) *Common*: We concatenate the output of the Siamese

TABLE III
PREDICTION PERFORMANCE ON A TEST SET

Predictor	F1-Score	AUC-ROC	Precision	Recall	Specificity
Interventional Cardiologist 1	0.095	0.484	0.054	0.400	0.568
Interventional Cardiologist 2	0.125	0.562	0.070	0.600	0.524
Cardiologists' mean	0.110	0.523	0.062	0.500	0.546
Multimodal framework	0.167	0.627	0.100	0.600	0.654

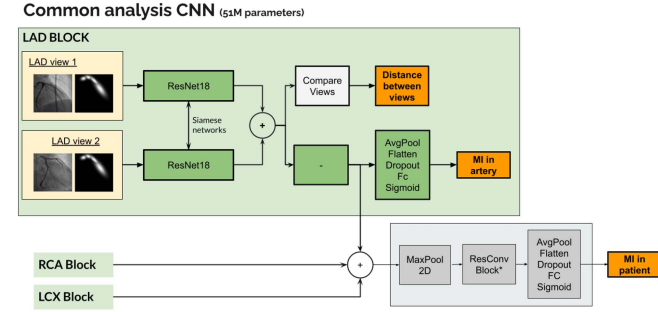


Fig. 4. Single modality ICA framework for MI prediction. The two views of each artery are processed separately before being concatenated together into a bigger feature map. This new feature map is further processed through convolutional layers and poolings to finally provide a patient-level prediction. The *ResConvBlock* is the set of convolutional blocks connected with skip connections presented in the section III-A.

TABLE IV

DIFFERENT IMPLEMENTATIONS OF SINGLE MODALITY: VALIDATION SCORES (MEAN $\pm$ STD) OBTAINED FROM A 5-FOLD CROSS-VALIDATION ON THE TRAINING SET

Modality	Implementation	F1-Score	AUC-ROC
ICA	Max + AUC	0.30$\pm$0.11	0.67$\pm$0.04
ICA	Max + BCE	0.28 $\pm$ 0.09	0.63 $\pm$ 0.06
ICA	Common + AUC	0.30 $\pm$ 0.12	0.66 $\pm$ 0.07
ICA	Common + BCE	0.31 $\pm$ 0.07	0.65 $\pm$ 0.03
Clinical data	AUC	0.28 $\pm$ 0.06	-
Clinical data	BCE	0.25 $\pm$ 0.05	-

networks for all arteries and process the entire feature representation to predict the MI at the patient level. It is represented in Fig. 4. To improve the discriminative power, we add some additional layers in the architecture. Those layers are similar to the ones used in ResNet [12]: max pooling (to reduce the size of the embedding space), a residual convolutional block (two convolutional layers with batch normalization and ReLU activation, connected by a residual connection), average pooling, flattening, dropout, and finally a classification layer activated by Sigmoid. We note that these additional layers increase the total number of parameters of the second model: the *Common* approach has five times more parameters than the *Max* one.

In Table IV, the different implementations of the single modality models are compared. We notice that the *Max* approaches for the ICA modality (taking the maximum of the prediction of each artery) reaches the same performance as the *Common* one (analyzing all the arteries together) while having much fewer parameters. Overall, using the AUC-loss is better or similar to the

BCE loss. For that reason, our solution uses a *Max* architecture trained with AUC-loss. The hyperparameters of those models can be found in section VI-A.

b) Alternative single modality: clinical data: Different traditional ML classifiers have been considered to predict MI from clinical data. Their performance is reported in Table V. We observe that overall all methods obtain comparable performance. They all obtained a higher recall than precision, which is good in the current application, as missing an MI could be fatal for the patient. ANN slightly outperforms the others and is more easily implementable in a Neural Network framework but uses significantly more parameters. The ANN is trained with AUC and BCE, and the obtained results are compared in Table IV. The hyperparameters of those ANNs can be found in section VI-B. Overall, those results are close to the ones obtained in [5] and [6]. The similarity between the reported performance across different works that predict future outcomes from clinical data, despite using different datasets and models, could be a strong indicator that we may have reached the limits of the predictive capacity of the clinical data, making the use of ICA images on top of clinical data necessary for better predictions.

IV. DISCUSSION

To the best of our knowledge, there are no previous studies that aim to predict a patient's risk of future MI by combining both image and patient data via a deep learning framework. Thus this pilot study represents the first of its kind and demonstrates the efficacy (albeit modest) of such an approach.

From the cardiology perspective, the importance of this specific prediction task cannot be understated. Cardiovascular disease remains the leading cause of death worldwide [1], despite significant advances in the prevention and treatment of cardiovascular disease in recent years. Even with the application of the best available predictors of cardiovascular risk (e.g. degree of coronary stenosis, hemodynamic impact of a stenosis, risk factors such as diabetes and hypertension), a significant number of patients still go on to experience an MI. This can likely be explained by the complex pathophysiology of coronary artery disease. To tackle this complexity, we propose an approach that integrates pertinent information from different sources.

With regards to the performance of our approach, some average quantitative comparisons can be performed with other works that predict future MI in different settings. Compared to [8], where a deep learning framework was able to predict future culprit lesions from ICA with an F1-score of 0.57, our performance is overall lower. This study was performed at a lesion level and on patients with stable coronary disease. On the

TABLE V

DIFFERENT ML ALGORITHM FOR MI PREDICTION FROM CLINICAL DATA: VALIDATION SCORES (MEAN $\pm$ STD) OBTAINED FROM A 5-FOLD CROSS-VALIDATION ON THE TRAINING SET

Model	F1-Score	Accuracy (%)	Precision (%)	Recall (%)
Decision Tree	0.23 $\pm$ 0.08	70.18 $\pm$ 0.08	15.13 $\pm$ 6.40	57.22 $\pm$ 20.18
Balanced Random Forest	0.21 $\pm$ 0.11	66.23 $\pm$ 19.97	13.33 $\pm$ 9.43	56.93 $\pm$ 30.61
Logistic Regression	0.24 $\pm$ 0.10	68.37 $\pm$ 5.03	14.97 $\pm$ 5.48	64.80 $\pm$ 16.90
Naive Bayes	0.24 $\pm$ 0.09	65.63 $\pm$ 5.03	15.09 $\pm$ 5.16	58.52 $\pm$ 8.65
Gradient Boosting	0.24 $\pm$ 0.06	71.36 $\pm$ 2.95	17.39 $\pm$ 4.16	63.12 $\pm$ 11.65
Artificial Neural Network	0.28$\pm$0.06	81.12$\pm$4.64	17.79$\pm$0.04	72.00$\pm$17.20

one hand, working at a lesion level is a simplified problem for deep learning, as it provides as input the exact lesion, as opposed to the whole artery. At the same time, the clinical cohort of [8] consists of a completely different population, with stable coronary disease, all of which eventually had an MI. In our setup, the cohort consists of patients with acute coronary syndromes, with only 10% suffering from MI in the follow-up period. Thus our scenario is more representative of real clinical practice where not all patients will have an MI during follow-up, and the challenge is identifying the ones that will. Compared to algorithms that predict only from clinical data, our results are close to the ones obtained by [5] and [6]. However, while we consider the risk of MI within a follow-up period of 5 years using a relatively small dataset ($\approx$ 500 patients) and only 11 clinical features, these works present models developed with significantly larger datasets with many more variables and predict the risk of MI within different time intervals (shorter ones). More specifically, the work in [5] reaches an F1-Score of 0.101 while predicting MI within the next six months for 2 million patients with 8'000 features. The work of [6] reaches an AUC of 0.72 while predicting MI within 12 months based on 7'000 patients and 192 features. These comparisons suggest that the framework proposed in the current study builds significantly on previous work and has the potential to improve the accuracy of ML-driven predictions of future MI significantly.

Importantly, we recognize that the generalisability of these results to other clinical cohorts needs to be demonstrated due to differing patient populations, as well as variance with respect to the quality and nature of the clinical and imaging data. As such, this study needs to be extended in future work.

V. CONCLUSION

In this work, we proposed a multimodal framework based on deep learning, which exploits the knowledge of the main arteries of the coronary tree, and the ICA images corresponding to each of those, as well as clinical patient data in order to predict future MI in patients with acute coronary syndromes. The ICA images are processed by anatomy-guided CNNs, and the clinical patient data is analysed by an ANN. Embeddings from both modalities are then combined to finally provide a patient-level prediction. Experimental results confirmed the superior performance of our method in comparison to learning from each modality separately, but also in comparison to human-based predictions from experienced interventional cardiologists. Although the numerical results of this study should be considered with caution due to the small number of MI patients, the performance obtained with

the current data is non-trivial. It indicates that the integration, via a well-designed learning framework, of imaging, clinical variables, and clinical expertise (including knowledge of the coronary anatomy), has the potential to improve on current approaches to the highly complex and challenging MI prediction task.

However, at its current state, this solution does not exhibit sufficiently high-performance levels for practical usage. A potential advancement for enhancing performance involves exploring attention mechanisms for integrating features from different sources [21]. Alternatively, one could leverage the arterial structure to organize patches centered around the artery (and thus get rid of the non-informative background) and then feed those patches to transformers or Graph Neural Networks, which have already shown promise in detecting severe stenosis [22] and coronary plaques [23], respectively.

AUTHOR CONTRIBUTIONS

IS, OS, DT and TM designed and implemented the method, performed data analysis and wrote the manuscript. EA, PF, OM and DT were responsible for the conception of the study. DN, SF, OM and TM provided support on the clinical aspects and they also provide the data used in this study. All authors critically reviewed the manuscript and approved the final version for submission.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

- [1] I. Abubakar, T. Tillmann, and A. Banerjee, "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," *Lancet*, vol. 385, pp. 117-171, 2015.
- [2] G. Ciccarelli et al., "Angiography versus hemodynamics to predict the natural history of coronary stenoses: Fractional flow reserve versus angiography in multivessel evaluation 2 substudy," *Circulation*, vol. 137, pp. 1475-1485, 2018.
- [3] P. Xaplanteris et al., "Five-year outcomes with PCI guided by fractional flow reserve," *New England J. Med.*, vol. 379, pp. 250-259, 2018.
- [4] G. Stone et al., "Heparin plus a glycoprotein IIb/IIIa inhibitor versus bivalirudin monotherapy and paclitaxel-eluting stents versus bare-metal stents in acute myocardial infarction (HORIZONS-AMI): Final 3-year results from a multicentre, randomised controlled trial," *Lancet*, vol. 377, pp. 2193-2204, 2011.
- [5] D. Mandair, P. Tiwari, S. Simon, K. Colborn, and M. Rosenberg, "Prediction of incident myocardial infarction using machine learning applied to harmonized electronic health record data," *BMC Med. Inform. And Decis. Mak.*, vol. 20, pp. 1-10, 2020.

- [6] S. Gupta et al., "Evaluation of machine learning algorithms for predicting readmission after acute myocardial infarction using routinely collected clinical data," *Can. J. Cardiol.*, vol. 36, 878–885, 2020.
- [7] E. Ovalle-Magallanes, J. Avina-Cervantes, I. Cruz-Aceves, and J. Ruiz-Pinales, "Transfer learning for stenosis detection in X-ray coronary angiography," *Mathematics*, vol. 8, 2020, Art. no. 1510.
- [8] T. Mahendiran et al., "Deep learning-based prediction of future myocardial infarction using invasive coronary angiography: A feasibility study," *Open Heart*, vol. 10, no. 1, 2023, Art. no. e002237.
- [9] D. Thanou et al., "Predicting future myocardial infarction from angiographies with deep learning," in *Proc. Med. Imag. Meets NeurIPS Workshop*, 2021, pp. 1–4.
- [10] R. Auer et al., "Quality of care after acute coronary syndromes in a prospective cohort with reasons," *PLoS One*, vol. 9, no. 3, 2014, Art. no. e93147.
- [11] F. Neumann et al., "ESC/EACTS Guidelines on myocardial revascularization," *Eur. Heart J.*, vol. 40, pp. 87–165, 2019.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, 770–778.
- [13] J. Kojuri et al., "Prediction of acute myocardial infarction with artificial neural networks in patients with nondiagnostic electrocardiogram," *J. Cardiovasc. Dis. Res.*, vol. 6, no. 2, pp. 51–59, 2015.
- [14] S. Thomas, "Combining image features and patient metadata to enhance transfer learning," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. & Biol. Soc.*, 2021, pp. 2660–2663.
- [15] J. Ellen, C. Graff, and M. Ohman, "Improving plankton image classification using context metadata," *Limnol. And Oceanogr.: Methods*, vol. 17, pp. 439–461, 2019.
- [16] Z. Yuan, Y. Yan, M. Sonka, and T. Yang, "Large-scale robust deep auc maximization: A new surrogate loss and empirical studies on medical image classification," in *Proc. IEEE/CVF Int. Conf. Computer Vis.*, 2021, pp. 3040–3049.
- [17] Z. Guo, Y. Yan, Z. Yuan, and T. Yang, "Fast Objective & duality gap convergence for non-convex strongly-concave min-max problems with PL condition," *J. Mach. Learn. Res.*, vol. 24, pp. 1–63, 2023.
- [18] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, 248–255.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Computer Vis.*, 2015, pp. 1026–1034.
- [20] C. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [21] A. Pacheco and R. Krohling, "An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification," *IEEE J. Biomed. Health Informat.*, vol. 25, pp. 3554–3563, Sep., 2021.
- [22] X. Ma, G. Luo, W. Wang, and K. Wang, "Transformer network for significant stenosis detection in CCTA of coronary arteries," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2021, pp. 516–525.
- [23] M. Viti, H. Talbot, and N. Gogin, "Transformer graph network for coronary plaque localization in CCTA," in *Proc. IEEE 19th Int. Symp. Biomed. Imag.*, 2022, pp. 1–5.