



OPEN

DATA DESCRIPTOR

# Coronary Dominance: Angiogram dataset for coronary dominance classification

Ivan Kruzhilov<sup>1,2,5</sup> , Gleb Mazanov<sup>3,5</sup>, Alexander Ponomarchuk<sup>2</sup>, Galina Zubkova<sup>2</sup>, Artem Shadrin<sup>4</sup>, Ruslan Utegenov<sup>4</sup>, Pavel Blinov<sup>2</sup> & Ivan Bessonov<sup>4</sup>

We release a new dataset containing invasive coronary angiograms for the coronary dominance classification task, an essential aspect of SYNTAX score estimation for assessing the severity of coronary artery disease. The dataset contains 1,574 angiographic studies, a set of X-ray multi-view videos classified as right or left dominant. The dataset has three parts: main, real distribution, and domain shift. Each study in the main dataset falls into five categories: normal, poor quality, artifact, high uncertainty, or right coronary artery (RCA) occlusion. These tags help classify the dominance more accurately and allow for uncertainty estimation and outlier detection. In the real distribution part, there are 400 randomly selected studies. In contrast the domain shift part includes studies from an interventional angiography system that differs from that used in the main and real distribution parts.

## Background & Summary

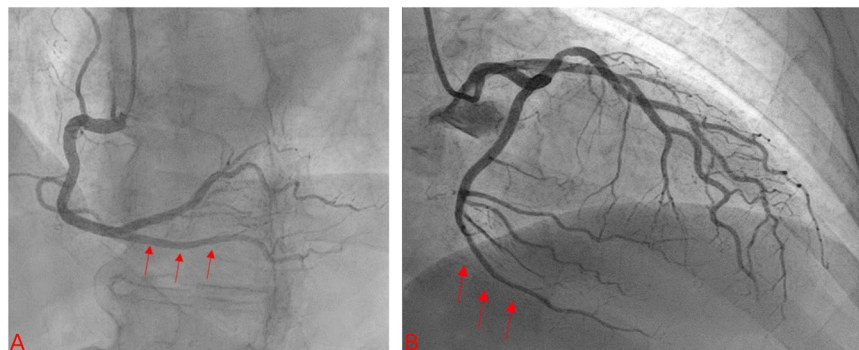
Coronary dominance classification is an essential step in the SYNTAX (Synergy Between Percutaneous Coronary Intervention with Taxus and Cardiac Surgery) score estimation, which has become a crucial tool for assessing the severity of coronary artery disease<sup>1</sup>. This classification is also necessary when determining the severity of coronary lesions using the Gensini score<sup>2</sup>.

Coronary dominance is determined by the coronary artery branch (Fig. 1) that supplies the posterior descending artery (PDA). There are three main categories: left dominance, right dominance, and co-dominance<sup>3</sup>. Approximately 70–80% of the population has right coronary dominance, while approximately 5–10% have left coronary dominance<sup>4</sup>. In some regions<sup>5</sup>, the proportion of individuals with left coronary dominance may be as high as 20%. Co-dominance occurs in approximately 10–20% of cases, where the PDA receives blood from the left and right coronary arteries<sup>6</sup>.

It is important to note that co-dominance is *not* in the SYNTAX scoring system. Therefore, when dealing with complex cases involving arteries supplying the PDA almost equally (PDA arises from both the circumflex artery and the right coronary artery), clinicians have to choose ‘right dominance’ when calculating SYNTAX scores. However, this approach can be challenging when the right coronary artery is small and supplies a limited myocardial blood supply zone, creating ambiguity in the classification. This issue likely contributes to discrepancies observed between local site assessments and angiographic core lab evaluations. For instance, in the EXCEL trial<sup>7</sup>, the right dominance type was identified significantly more often in the core lab compared to site operators (95.3% vs. 89.7%,  $p < 0.001$ ). These differences highlight the subjectivity inherent in classifying borderline cases of co-dominance and the potential for variability between evaluators. To address these challenges, we introduced a special tag ‘RCA small diameter’ into the dataset for this complex case.

The use of advanced computational approaches, such as neural network algorithms, could enhance the consistency of coronary dominance assessments by reducing inter-observer variability. By providing an objective and reproducible framework, such tools could mitigate the limitations of current classification methods and improve the reliability of dominance determination in clinical and research settings.

<sup>1</sup>Applied mathematics and artificial intelligence, Moscow Power Engineering Institute, 14, Krasnokazarmennaya, Moscow, 111250, Russia. <sup>2</sup>Sber AI Lab, 32A, Kutuzovskiy, Moscow, 121170, Russia. <sup>3</sup>Skoltech, 30, Bolshoy Boulevard, Moscow, 121205, Russia. <sup>4</sup>Tyumen Cardiology Research Center, Tomsk National Research Medical Center of Russian Academy of Science, 111, Melnikaita, Tyumen, 625026, Tyumen oblast, Russia. <sup>5</sup>These authors contributed equally: Ivan Kruzhilov, Gleb Mazanov. ✉e-mail: [kruzhilovis@mpei.ru](mailto:kruzhilovis@mpei.ru)



**Fig. 1** The determination of heart dominance is based on the origin of the PDA (red arrows). The PDA can originate from either the right coronary artery (A); the left coronary artery (B), or both.

Interventional cardiologists use X-ray video, known as angiographic view, to estimate the SYNTAX score and classify coronary dominance. To get these videos, interventional cardiologists inject a contrast agent into the left (LCA) and right (RCA) coronary arteries. By capturing heart pulses from different angles, an angiography study provides valuable information about the cardiovascular system. Figure 2 provides examples of typical angiographic frames from the dataset.

The study<sup>8</sup> found that the following factors may complicate the classification of coronary dominance: a total occlusion, a small RCA diameter, a poor-quality angiogram, and artifacts. A RCA with a small diameter means there is significant disagreement among experts about how to classify a patient. Cases with a poor-quality angiogram, which can lead to incorrect diagnoses, typically have low-contrast medium filling. Studies that involve artifacts include those with pacing electrodes or sternal wires. The Fig. 3 shows examples of these cases. Our dataset includes all these categories and allows us to measure the impact of each factor on classification metrics.

The domain shift is an important problem when working with medical data. A change in a medical device's parameters or settings could significantly impact the neural network's prediction accuracy. To allow ML researchers to experience this phenomenon, we present angiograms from two different cardiovascular imaging systems in our dataset – Phillips Allura Clarity and Philips Azurion.

The presented dataset is the first complete angiogram dataset for classifying coronary dominance. However there are several other datasets of angiography frames (not full angiographic studies) with segmented vessels or stenosis bounding boxes. One such study<sup>9</sup> published a dataset containing 134 X-ray images of the coronary arteries with segmented vessels. Another study<sup>10</sup> provided a dataset containing 8,325 angiography frames with labeled stenoses collected from 100 patients using Siemens and General Electric devices. Most recently<sup>11</sup>, introduced an ARCADE dataset containing 1,500 images for coronary vessel segmentation, classification, and stenotic lesion detection. The study<sup>12</sup> presented an annotated Invasive Coronary Angiography dataset of 42 patients including manually labeled lesion bounding boxes and selected clinical features.

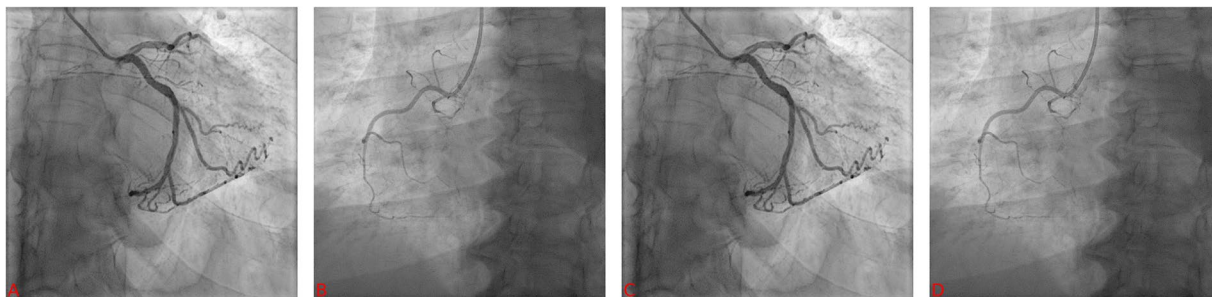
## Methods

Invasive coronary angiography is a widely utilized diagnostic procedure for suspected coronary artery disease patients. Our dataset contains the study of patients who underwent invasive coronary angiography at the Tyumen Cardiology Research Center (Branch of Tomsk National Research Medical Center), Russian Academy of Science, between January 2016 and November 2023. The study strictly adheres to the ethical principles outlined in the Declaration of Helsinki and has been approved by the ethics committee of Tyumen Cardiology Research Center (Protocol № 184, dated 12/02/2024). The waiver for informed consent was granted by ethics committee of Tyumen Cardiology Research Center as the study would not be feasible or practicable without the modification of informed consent, it has important social value, and it poses no more than minimal risk to participants. Furthermore, we have fully anonymized all data to protect patient confidentiality.

This study contains 1,574 patients who underwent invasive coronary angiography procedures. Interventional cardiologists used the latest guidelines<sup>13</sup> and utilized the state-of-the-art interventional angiography systems Phillips Allura Clarity (Philips Healthcare, Best, The Netherlands) and Philips Azurion (Philips Healthcare, Best, The Netherlands). The contrast agent is iopromide. The frame rate is 15 frames per second.

Among the 1,574 patients, 567 were diagnosed with acute coronary syndrome (ACS), while 1,007 had stable coronary artery disease (CAD). For those with stable CAD, coronary angiography was performed based on the presence of chronic coronary syndrome or abnormalities revealed by noninvasive testing. The average age of the patients was  $58 \pm 9.9$  years. Of the total cohort, 1,228 patients (78%) were men, with an average age of  $57 \pm 9.7$  years, while 346 patients (22%) were women, with an average age of  $63 \pm 9.9$  years. The standard coronary angiographic views of the RCA have the following distribution: 1.4% RAO-CAU, 16% RAO-CRA, 38% LAO-CAU and 44% LAO-CRA. The standard coronary angiographic views of the LCA have the following distribution: 21.9% RAO-CAU, 24.3% RAO-CRA, 25.3% LAO-CAU and 28.5% LAO-CRA, where RAO/LAO is right/left anterior oblique, CAU and CRA – are caudal/cranial views.

Using X-ray technology, coronary angiography captures images on 2D planes and creates video images of the internal organs and blood vessels. These views are not synchronized in time. The study consists of the LCA and RCA views, which typically contain between 20 and 70 frames,  $512 \times 512$  each. The pixels values are in a range



**Fig. 2** Typical examples of angiographic frames (A) LCA, Left dominance, Azurion; (B) RCA, Left dominance, Allura Clarity; (C) LCA, Right dominance, Allura Clarity; (D) RCA, Right dominance, Azurion.



**Fig. 3** Examples of cases complicating dominance classification from the main data set (A) bad quality angiogram; (B) artifacts; (C) RCA small diameter; (D) RCA occlusion.

from 0 to 255. The angiograms from Phillips Azurion have a frame around the image which we deleted and resized the images to  $512 \times 512$ . The original angiography studies were in dicom format. We converted them to Numpy's compressed array format (npz) to reduce the dataset size and make it easier to use. We also normalized the data, removed artifacts, and automatically classified the RCA and LCA arteries using a neural network.

The data for the dataset were collected during standard medical studies under the supervision of Dr. Ivan Bessonov, the Head of Interventional Cardiology Laboratory at the Tyumen Cardiology Research Center, who is a co-author of this article. Three interventional cardiologists, with experience ranging from three to fifteen years, provided ground-truth information on coronary dominance. Dr. Ivan Bessonov has more than fifteen years of experience, and we considered his labels to be the gold standard. He acted as a moderator to verify the labeling. He also labeled studies with additional tags, such as occlusions and artifacts. The disagreement rate between the experts was 2.8%. After excluding complex categories such as poor quality, artifacts, and small RCA diameters, the disagreement rate on the main data set dropped to 1.5%.

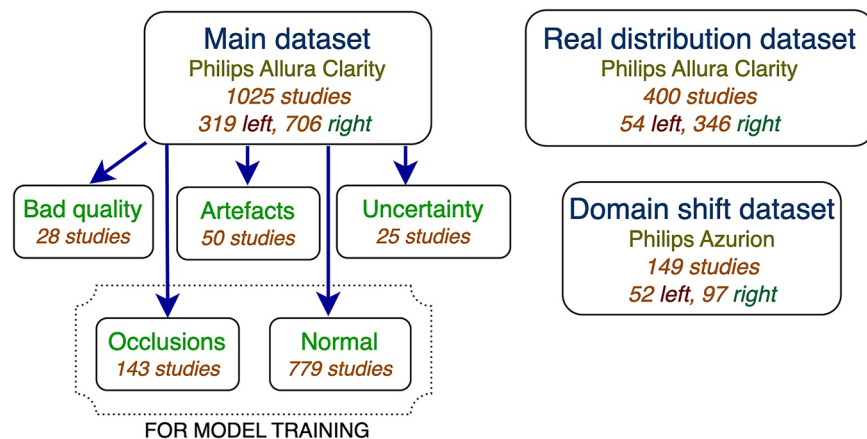
### Data Records

The data are available at the CoronaryDominance<sup>14</sup> Hugging Face repository <https://huggingface.co/datasets/BearSubj13/CoronaryDominance>.

The data we present here includes three parts (Fig. 4) - the main dataset (Table 1), the real distribution dataset (Table 2), and the domain shift dataset (Table 3). The main dataset with 1,025 studies (319 of whom have left dominance, and 706 of whom have right dominance) from the Philips Allura Clarity cardiovascular imaging system is for model training and hyper-parameters fine-tuning. For the main dataset, each study falls into one of five categories: bad quality, artifact, RCA small diameter (uncertainty), occlusion, or normal (Table 4). "Normal" is the largest category. We enhanced the main data set with additional left dominant cases and occlusions in order to address the problem of imbalance.

The real distribution dataset consists of 400 randomly selected studies (54 left dominant, and 346 right dominant) for the Philips Allura Clarity cardiovascular imaging system. Due to this, unlike the main dataset, the real distribution dataset follows the same distribution as all the studies in the clinic. The real distribution dataset does not have group subdivision, so it includes studies with poor quality, artifacts, and high uncertainty. The domain shift dataset comprises 149 studies (52 left and 97 right dominant) from the Philips Azurion cardiovascular imaging system. This dataset also does not have group subdivision. However, we have enhanced it with an additional left dominant studies.

The data structure is as follows: the root directory contains two subdirectories: "Left\_Dominance" and "Right\_Dominance", which contain angiographic studies of left and right dominant patients, respectively. Each study consists of several angiographic views - gray-scale videos saved in Numpy's compressed array format (.npz). The study directory also contains two subdirectories named "RCA" (for the Right Coronary Artery) and "LCA" (for the Left Coronary Artery), respectively. Figure 5 shows the file structure of the dataset.



**Fig. 4** Dataset structure.

Number of	LCA	RCA	Total
Patients	-N/A-	-N/A-	1,025
Angiographic views	4,965	1,777	6,742
Frames	219,691	84,122	303,803

**Table 1.** Number of patients, angiographic views and frames in the main data set.

Number of	LCA	RCA	Total
Patients	-N/A-	-N/A-	400
Angiographic views	2,033	857	2,890
Frames	101,424	44,403	145,827

**Table 2.** Number of patients, angiographic views and frames in the real distribution data set.

Number of	LCA	RCA	Total
Patients	-N/A-	-N/A-	149
Angiographic views	668	243	911
Frames	31,756	12,720	44,476

**Table 3.** Number of patients, angiographic views and frames in the domain shift data set.

Number of	bad quality	artefacts	uncertainty	occlusion	normal	Total
Left dominant	13	15	9	31	251	319
Right dominant	15	35	16	112	528	706
Total	28	50	25	143	779	1,025

**Table 4.** Number of patients by categories. Main part of the dataset.

The repository<sup>14</sup> contains the following archive files: domain\_shift\_dataset.7z (7.8 Gb) - with the domain shift dataset, real\_distribution.7z (25.9 Gb) - with the real distribution dataset, and main\_normal.7z (40.7 Gb), and main\_others.7z (14.2 Gb) are for the main dataset. We split the main dataset into two archives due to its large size; the archive main\_normal.7z is for normal data of main dataset (779 angiography studies) and main\_others.7z are for other categories such as bad quality, artefacts, uncertainty/small RCA diameter, and occlusions - 246 angiography studies in total.

The study folder (Fig. 5) follows the structure “Study0xxxx\_study\_id”, where “0xxxx” is the sequence number of the study and “study\_id” is the study ID. The name of the .npz file containing the angiographic view video matches the series ID. The study ID and the series ID are unique within the dataset and follow the same format as the corresponding ID in DICOM (.dcm) files. Our data are fully anonymized; therefore the IDs in the dataset do not match those in the original .dcm source files. The sequence number is unique within a specific dataset. We have introduced this for the convenience of referencing, such as “main dataset study 580” instead of “1.1.11.11111.11.903846392438863822081872923423339322”.

## Left\_Dominance

Study00579\_1.1.11.11111.11.869085947840901039288407951315343760 ← Study

## LCA

1.1.11.11111.11.16182961545013696635109168881626416.1.1.npz  
 1.1.11.11111.11.29293839621542019225230002264659525.1.1.npz  
 1.1.11.11111.11.55838673388937743275110424745479320.1.1.npz  
 1.1.11.11111.11.70261562781233949411185287142806880.1.1.npz ← View  
 1.1.11.11111.11.71034023915188661310312686828867066.1.1.npz  
 1.1.11.11111.11.71644854664347303576173739914223886.1.1.npz  
 1.1.11.11111.11.83887744120255052354019603483659580.1.1.npz  
 1.1.11.11111.11.99315245548855271023720958106112817.1.1.npz

## RCA

1.1.11.11111.11.29591834230959834488576624009266388.1.1.npz ← View

Study00580\_1.1.11.11111.11.903846392438863822081872923423339322

Study00581\_1.1.11.11111.11.403000224476940931128247336780644522

...

## Right\_Dominance

Study00051\_1.1.11.11111.11.697247659798095106724889393224196074

Study00052\_1.1.11.11111.11.386654325874967122036316957888226295 ← Study

Study00053\_1.1.11.11111.11.795830296846926976021952631863000280

...

Study number

Study ID

Fig. 5 Data structure in repository.

Name	Type	range	Description
pixel_array	Float array	frames × 512 × 512	angiographic 3D view
seriesid	string	consists of numbers and “.”	unique ID of an angiographic view
studyid	string	consists of numbers and “.”	unique ID of an angiographic study
series_number	integer	0 - 1025	ID of an angiographic study
is_collaterals	Boolean	{True, False} None	Collaterals in LCA
primary_angle	Float	-180 - 180 None	positioner angle, degree
secondary_angle	Float	-90 - 90 None	positioner angle, degree
is_occlusion	Boolean	{True, False}	Occlusion in RCA
is_undefined_type	Boolean	{True, False}, None	Studies with high uncertainty
is_artifact	Boolean	{True, False}, None	The presence of artifacts
artery_type	String	{LCA, RCA}	Coronary artery type

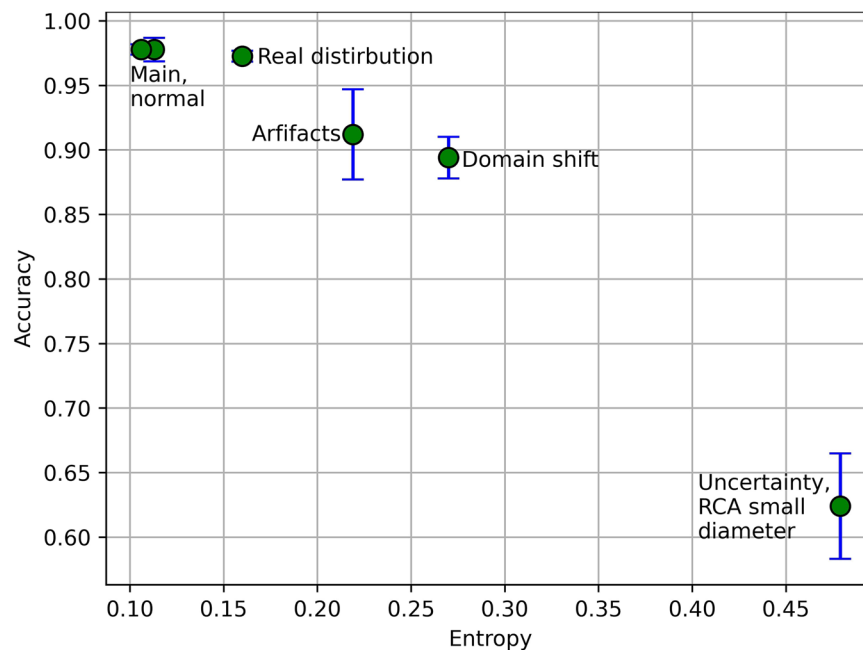
Table 5. Fields of a dictionary in .npz file.

Table 5 shows the data structure in .npz files. We extracted the primary and secondary angles from the fields (0018,1510) and (0018,1511) in the original .dcm file. You can consult the DICOM Standard Browser <https://dicom.innolitics.com/ciods/digital-x-ray-image/dx-positioning/00181511> for more information on these angles. In most cases, the *is\_collaterals* tag is *True* when the *is\_occlusion* tag is also *True*.

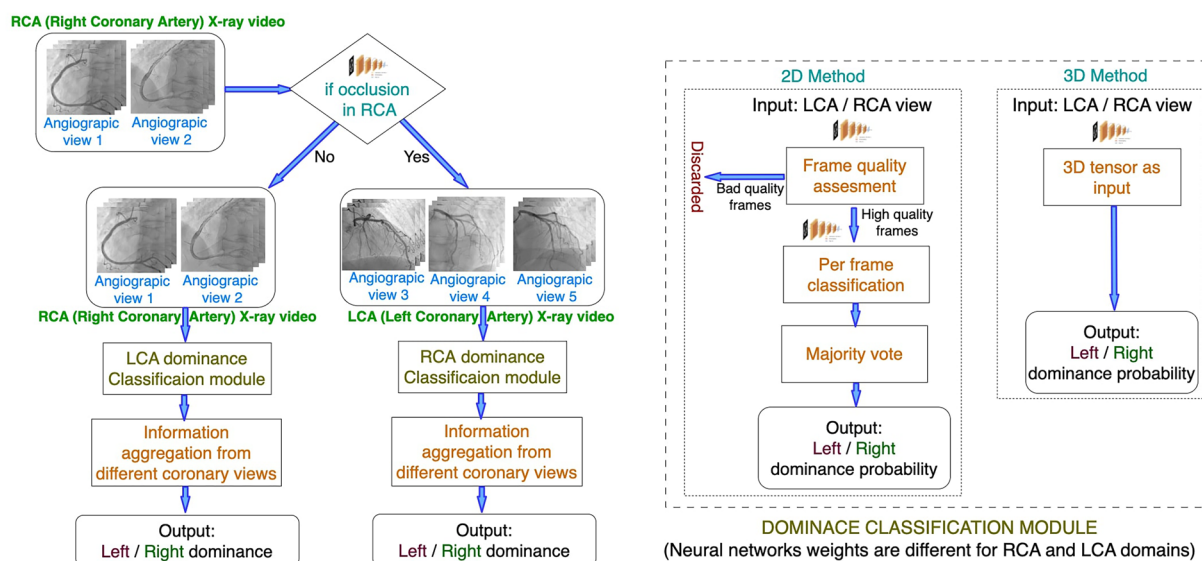
## Technical Validation

To guarantee the labeling quality, a PhD interventional cardiologist with more than 15 years of experience verified the labels of other experts. In cases where there was disagreement between the experts, he added comments explaining why the case caused a controversy. Most often, the disagreements were in studies on RCA with a small diameter, which we specially labeled. In the majority of cases, our “gold standard” interventional cardiologist changed the left-dominant label to right-dominance, which is a common mistake in the cardiological community<sup>7</sup>. It is worth noting that the predictions of our model trained on data with no “gold standard corrections” coincide in most cases with the gold standard labeling. This fact indicates the correctness of our labeling, as machine learning models have the ability to generalize and disregard some false labels. Furthermore, Fig. 6 clearly shows the influence of our special tags, such as “RCA small diameter”, “domain shift” and “artifacts”, on the model’s accuracy and uncertainty, which in turn demonstrates the relevance of our labels.

We developed a baseline model for coronary dominance classification and used exactly the same data for it as we present in the dataset. We trained a neural network on normal and occlusion categories from the main dataset. The model achieved about 97% accuracy when tested on 400 studies of the real distribution dataset. The accuracy dropped to 89% for the domain shift dataset. The recall of occlusion classification was about 99% and precision about 85%. The interventional cardiologists from our team analyzed all cases where our baseline model failed regularly. In most cases, there is a clear explanation for why the study was difficult for the model to analyze. This also confirms the correctness of our baseline model and data we used to train them. The more information about our baseline model is in the following subsection.



**Fig. 6** Dominance classification accuracy dependence on the entropy of the model output probabilities.



**Fig. 7** Proposed coronary dominance classification pipeline. The dominance classification module allows choosing between 2D and 3D neural networks.

**Baseline model for coronary dominance classification.** We provide a benchmark for the coronary dominance classification for future researchers to compare their results with. Figure 7 shows the scheme of our classification algorithm. Our approach for coronary dominance classification works as follows: first, we classify occlusion in RCA using a neural network. If there is no occlusion in the RCA, we use the RCA views for classification, otherwise, we use LCA views. Both 2D and 3D neural networks can be used for coronary dominance classification. For classification of RCA occlusion, we used ConvNeXt with a frame quality check. To analyze angiogram projections, we analyzed them frame-wise with 2D networks, and only informative frames were included in the decision process. We used a specifically trained neural network to determine if a frame was informative or not, and then made the final decision for each angiogram view based on a majority vote of all relevant analyzed frames. The majority vote is an unweighted sum of the probabilities from all informative frames. The use of 3D neural networks may allow for omitting the majority vote step. Finally, our approach uses either four 2D or one 2D and two 3D networks for coronary dominance classification. The first approach uses the following 2D networks for: 1) RCA occlusion; 2) frame quality check; 3) RCA dominance classification; 4) and LCA dominance

Metric/Dataset	5-fold cross-valid	Hold out	Real distribution	Domain shift	Artifact	Uncertainty / RCA small diameter
dataset size	822	100	400	149	50	25
Recall right ↑	97.9 $\pm$ 1.0%	97.3 $\pm$ 0.0%	98.1 $\pm$ 0.6%	93.6 $\pm$ 1.5%	95.4 $\pm$ 3.8%	53.8 $\pm$ 6.4%
Recall left ↑	97.6 $\pm$ 2.0%	99.2 $\pm$ 1.7%	91.2 $\pm$ 1.0%	81.5 $\pm$ 4.5%	81.3 $\pm$ 5.0%	77.8 $\pm$ 0.0%
Precision right ↑	97.9 $\pm$ 1.0%	99.7 $\pm$ 0.5%	98.7 $\pm$ 0.1%	90.5 $\pm$ 2.1%	92.3 $\pm$ 2.0%	81.0 $\pm$ 1.9%
Precision left ↑	95.4 $\pm$ 1.9%	92.3 $\pm$ 1.2%	87.5 $\pm$ 3.3%	87.3 $\pm$ 2.6%	89.1 $\pm$ 8.4%	48.9 $\pm$ 3.4%
Recall <sub>macro</sub> ↑	97.73 $\pm$ 1.1%	98.3 $\pm$ 0.8%	94.6 $\pm$ 0.3%	87.6 $\pm$ 2.2%	88.4 $\pm$ 3.6%	65.8 $\pm$ 3.2%
Accuracy ↑	97.8 $\pm$ 0.9%	97.8 $\pm$ 0.4%	97.3 $\pm$ 0.4%	89.4 $\pm$ 1.6%	91.2 $\pm$ 3.5%	62.4 $\pm$ 4.1%
F1 <sub>macro</sub> ↑	97.4 $\pm$ 1.1%	97.1 $\pm$ 0.6%	93.9 $\pm$ 0.8%	88.1 $\pm$ 1.9%	89.3 $\pm$ 4.1%	62.2 $\pm$ 3.9%
MCC ↑	94.9 $\pm$ 2.1%	94.2 $\pm$ 1.2%	87.8 $\pm$ 1.6%	76.5 $\pm$ 3.6%	79.0 $\pm$ 8.2%	30.7 $\pm$ 5.9%
inter-model entropy ↓	0.113	0.106	0.160	0.270	0.219	0.479
ensemble entropy ↓	-NA-	0.025	0.039	0.077	0.063	0.052

**Table 6.** Coronary dominance classification metrics for ConvNeXt model using different datasets. We used a 5-fold cross-validation approach on 822 studies to train the models and fine-tune hyper-parameters. The remaining data was used for testing purposes. We utilized 5 models trained during 5-fold cross-validation to obtain an average value with  $\pm$  STD and estimate the entropy in bits.

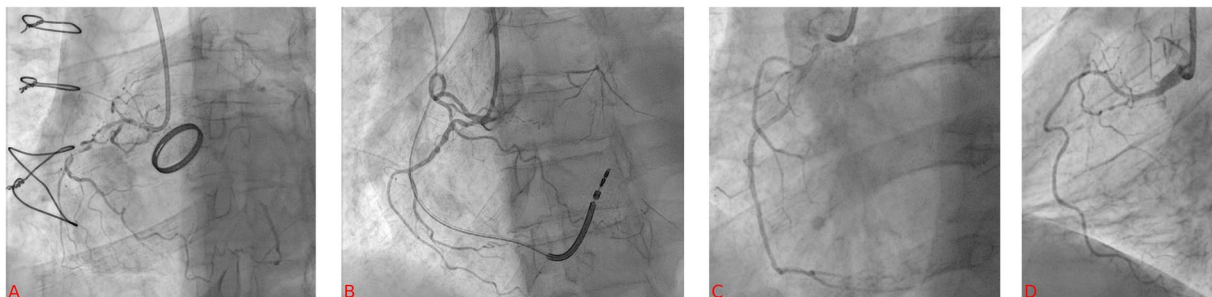
Nº	True dominance	Models failed out of 5	entropy	Artifact	RCA small diameter	Comment
1	Left	4	0.437	✓	–	Easy-to-classify case, pacing electrodes
2	Left	3	0.330	✓	–	Difficult for classification cardiac structure, sternal wires, heart valve. Fig. 7A
3	Left / ?Right	5	0.801	–	✓	Difficult to classify the case. The posterior descending artery arises from the LCA, but it may also arise from the RCA. The second RCA projection has poor quality. Our “gold standard” expert labelled the study as being left-dominant, but the majority of experts considered it to be right-dominant
4	Left	4	0.752	✓	–	Difficult for classification case, pacing electrode. Fig. 7B
5	Right	4	0.754	–	✓	Fig. 7C
6	Right	5	0.708	✓	✓	Sternal wires, heart valve artifact
7	Right	4	0.837	–	–	RCA fluoroscopy instead of angiography. For this reason the model used LCA for classification
8	Right	4	0.672	–	✓	–
9	Right	4	0.312	–	–	The model failed to classify an RCA occlusion and used RCA for classification. The occlusion misclassification was on a narrow margin. Fig. 7D.

**Table 7.** Studies where our approach regularly failed - 3 or more out of 5 models trained during 5-fold cross-validation were unable to correctly predict coronary dominance. The output probability entropy of the model  $>$  0.5 bits, indicating a high level of uncertainty.

classification. An alternative 3D approach uses a 2D network for occlusion detection, and two additional 3D nets for RCA and LCA classification.

For model training, validation, and testing we utilized 922 “normal” and “occlusion” studies from the main dataset. We also used other categories such as “artifact” and “uncertainty” as well as the real distribution and domain shift datasets for testing only. We have split 100 hold out studies from 922 “normal” and “occlusion” categories to reserve them for the testing. The rest 822 studies we exploited for model training and hyper-parameter fine-tuning using a 5-fold cross-validation strategy. After training 5 models during 5 folds of cross-validation, we applied them to the test data, which allowed us to present STD along with the mean value for all our metrics.

Table 6 presents the performance of the 2D ConvNeXt model for coronary dominance classification, as it achieved the best results among all the models. The Table 6 includes recall, precision, and macro metrics such as accuracy, recall macro, F1 macro, and Matthew correlation (MCC<sup>15</sup>) for hold-out, artifacts, uncertainty, real-distribution, and domain-shift data. The occlusion prediction recall for RCA is about 99% with a precision of 90%. For artifacts and domain-shift, the model performed poorly in left dominant studies. There are three basic approaches<sup>16</sup> to uncertainty estimation: model sampling, single-network methods, and data augmentation. The Table 6 presents the uncertainty estimation of coronary dominance classification using model sampling and single-network methods. For the model sampling approach, we used an ensemble approach that utilizes 5



**Fig. 8** RCA of studies (Table 7) where our approach regularly failed (a) case 2; (b) case 4; (c) case 5 (d); case 9.

models from a cross-validation fold (inter-model). For the single-network method (intra-model), we calculated the Shannon entropy between the class probability (Equation (1)) predictions (mean value of 5 models).

$$entropy = -p_{right} \log_2 p_{right} - p_{left} \log_2 p_{left} \quad (1)$$

Analyzing hard cases where the *model regularly failed* can provide valuable insight into potential areas for further research into better solutions. The summary of such cases is in Table 7 and the examples are in Fig. 8. The main reasons for the false classification of studies with left-dominance are artifacts and for studies with right-dominance the cause is RCA small diameter.

The Fig. 6 shows the dependence of the model's prediction accuracy on the intra-model entropy. The data from another cardiovascular imaging system or angiography with artifacts have significantly higher entropy due to a domain shift, since these cases were not included in the training dataset. The uncertainty cases have the highest entropy and the lowest accuracy, because they lie near the decision boundary.

### Usage Notes

Our dataset of angiography studies for coronary dominance classification could be a valuable benchmark for 3D multi-view methods. Since the data come from different cardiovascular imaging systems, one can test the effect of domain shift. The dataset is useful for estimating uncertainty and detecting outliers. If one trains a neural network using data without artifacts and then tests the network on a subset with artifacts, one can estimate the algorithm's stability concerning outliers. Studies involving RCA with a small diameter are particularly interesting for uncertainty estimation since there is significant disagreement among experts regarding dominance classification in these studies.

We recommend using normal and occlusion studies from the main dataset for model training and the other studies for testing only. For normal studies, one could use both the RCA and LCA views for coronary dominance classification. However, in our opinion, the RCA views are more suitable for this purpose. For studies with occlusion, only LCA views provide essential information for the dominance classification.

The utilization of the presented data set for SYNTAX score prediction extends beyond its intended purpose, as it offers a vast amount of full angiograms that can be utilized for pretraining in other tasks such as artery segmentation or stenosis detection. One potential approach for pretraining is the prediction of primary and secondary positioning angle, or the classification of angiograms based on their positioning (such as LAO or RAO). This approach has been successfully implemented in a previous study<sup>17</sup> using the *CathAI* neural network.

### Code availability

The tags provided by experts was stored in .xlsx file and original angiographic studies was in dicom format. We used python 3.8 and pydicom package 2.4.4 to normalize data and convert .dcm files with angiographic views into .npz dictionaries. The code is available at [https://github.com/ivankru/dataset\\_dominance](https://github.com/ivankru/dataset_dominance). The code for the baseline coronary dominance estimation model is available at [https://github.com/ivankru/dominance\\_classification](https://github.com/ivankru/dominance_classification).

Received: 17 June 2024; Accepted: 19 February 2025;

Published online: 26 February 2025

### References

1. Sianos, G. *et al.* The SYNTAX score: an angiographic tool grading the complexity of coronary artery disease. *EuroIntervention* **1**, 219–227 (2005).
2. Rampidis, G. P., Benetos, G., Benz, D. C., Giannopoulos, A. A. & Buechel, R. R. A guide for Gensini score calculation. *Atherosclerosis* **287**, 181–183 (2019).
3. Thiene, G., Frescura, C., Padalino, M., Basso, C. & Rizzo, S. Coronary arteries: Normal anatomy with historical notes and embryology of main stems. *Frontiers in cardiovascular medicine* **8**, 649855 (2021).
4. Wang, L. *et al.* Association between coronary dominance and acute inferior myocardial infarction: a matched, case-control study. *BMC cardiovascular disorders* **19**, 1–7 (2019).
5. Das, H., Das, G., Das, D. C. & Talukdar, K. A study of coronary dominance in the population of Assam. *Journal of Anatomical Society of India* **59**, 187–191 (2010).
6. Shriki, J. E. *et al.* Identifying, characterizing, and classifying congenital anomalies of the coronary arteries. *Radiographics* **32**, 453–468 (2012).

7. Genereux, P. *et al.* Outcomes of PCI versus CABG in left main disease according to SYNTAX score by site versus angiographic core laboratory assessment: Insights from the excel trial. *Journal of the American College of Cardiology* **69**, 972–972 (2017).
8. Kruzhilov, I. *et al.* Neural network-based coronary dominance classification of RCA angiograms. *arXiv preprint arXiv:2309.06958* (2023).
9. Cervantes-Sanchez, F., Cruz-Aceves, I., Hernandez-Aguirre, A., Hernandez-Gonzalez, M. A. & Solorio-Meza, S. E. Automatic segmentation of coronary arteries in X-ray angiograms using multiscale analysis and artificial neural networks. *Applied Sciences* **9**, 5507 (2019).
10. Danilov, V. V. *et al.* Real-time coronary artery stenosis detection based on modern neural networks. *Scientific reports* **11**, 7582 (2021).
11. Popov, M. *et al.* Dataset for automatic region-based coronary artery disease diagnostics using X-ray angiography images. *Scientific Data* **11**, 20 (2024).
12. Jiménez-Partinen, A. *et al.* Cadica: a new dataset for coronary artery disease detection by using invasive coronary angiography. *Expert Systems* **41**(12), e13708, <https://doi.org/10.1111/exsy.13708> (2024).
13. Lawton, J. S. *et al.* 2021 ACC/AHA/SCAI guideline for coronary artery revascularization: executive summary: a report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. *Circulation* **145**, e4–e17 (2022).
14. Ivan Kruzhilov, Gleb Mazanov, Alexander Ponomarchuk, etc. CoronaryDominance (Revision 5729d4a). Hugging Face. <https://doi.org/10.57967/hf/4360> (2025).
15. Chicco, D. & Jurman, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics* **21**, 1–13 (2020).
16. Kurz, A. *et al.* Uncertainty estimation in medical image classification: systematic review. *JMIR Medical Informatics* **10**, e36427 (2022).
17. Avram, R. *et al.* CathAI: fully automated coronary angiography interpretation and stenosis estimation. *npj Digital Medicine* **6**, 142 (2023).

### Author contributions

Ivan Kruzhilov – dataset preparation and article writing, Gleb Mazanov – dataset preparation and baseline model, Alexander Ponomarchuk – dataset preparation, Galina Zubkova – data acquisition, Artem Shadrin – data labelling, Ruslan Utegenov – data labelling, Pavel Blinov – project coordination, and Ivan Bessonov – data labelling and quality control.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04676-8>.

**Correspondence** and requests for materials should be addressed to I.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025