Evaluation of Stenoses Using AI Video Models Applied to Coronary Angiographies

Robert Avram (robert.avram.md@gmail.com)
Montreal Heart Institute https://orcid.org/0000-0002-8490-0270

Elodie Labrecque-Langlais
Polytechnique Montreal

Denis Corbin
Montreal Heart Institute

Olivier Tastet
Montreal Heart Institute

Ahmad Hayek
Montreal Heart Institute

Gemina Doolub
Montreal Heart Institute

Sebastián Mrad
Montreal Heart Institute

Jean-Claude Tardif
https://orcid.org/0000-0002-8200-8983

Jean-Francois Tanguay
Montreal Heart Institute, Université de Montréal

Guillaume Marquis-Gravel
Montreal Heart Institute

Geoffrey Tison
University of California, San Francisco https://orcid.org/0000-0002-0310-3326

Samuel Kadoury
Polytechnique Montreal

William Le
Polytechnique Montreal

Richard Gallo
Montreal Heart Institute

Frederic Lesage
Polytechnique Montreal

Article

Keywords: Machine learning, artificial intelligence, coronary angiography, angiogram, automated interpretation, coronary artery disease, stenosis, deep neural network

Posted Date: November 23rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3610879/v1

Additional Declarations: (Not answered)

Evaluation of Stenoses Using AI Video Models Applied to Coronary Angiographies

Élodie Labrecque Langlais[1,2], Denis Corbin[2,3], Olivier Tastet[2,3], Ahmad Hayek[3], Gemina Doolub[3], Sebastián Mrad[3], Jean-Claude Tardif[3], Jean-François Tanguay[3], Guillaume Marquis-Gravel[3], Geoffrey Tison[4], Samuel Kadoury[5], William Le[5], Richard Gallo[3], Frederic Lesage[1,3*], Robert Avram[2,3*]

1. Department of Electrical Engineering, Polytechnique Montréal, Montreal, Canada.

2. Heartwise (heartwise.ai), Montreal Heart Institute, Montreal, Canada.

3. Department of Medicine, Montreal Heart Institute, Université de Montréal, Montreal, Canada.

4. Department of Medicine, University of California, San Francisco, United States of America

5. Department of Computer Engineering, Polytechnique Montréal, Montreal, Canada.

*Contributed equally

Keywords: Machine learning, artificial intelligence, coronary angiography, angiogram, automated interpretation, coronary artery disease, stenosis, deep neural network

Corresponding Author:
Robert Avram, MD, MSc
Division of Cardiology, Department of Medicine
Montreal Heart Institute
Montreal, Québec, H1T 1C8
robert.avram.md@gmail.com

Word Count: 6,643

Author Conflict of Interest Disclosures: Dr. Avram and Dr. Tison are co-inventors in the patent pending 63/208,406 (Method and System for Automated Analysis of Coronary Angiograms). There are no disclosures for the remaining authors. The other authors declare no conflict of interest. Dr. Tison has received research grants from Myokardia, General Electric and Janssen Pharmaceuticals.

Author Contribution: ELL wrote the majority of the manuscript as well as conceived and conducted the experiments. All other co-authors reviewed and improved the manuscript. FL and RA supervised the project conducted by ELL. DC and OT assisted with the experiments. GT supervised the CathAI project. SK and WL offered their advice for the conception of the project. AH, SM and GD annotated Dataset B. All authors read and approved the final manuscript.

Data Availability: The data used for training and validation in this study are derived from clinical care and thus are not made publicly available due to data privacy concerns. Algorithm 4 and 6 weights are available at https://huggingface.co/heartwise/DeepCoro

Code Availability: The underlying code for DeepCoro's inference of Algorithm 3 to 6 for this study is available in the DeepCoro GitHub repository and can be accessible via this link at https://github.com/HeartWise-AI/DeepCoro

ABBREVIATIONS

AI: Artificial Intelligence

ARCADE: Automatic Region-based Coronary Artery Disease diagnostics using X-ray angiography imagEs

AUPRC: Area Under the Precision-Recall Curve

AUROC: Area Under the Receiver Operating Curve

CABG: Coronary Artery Bypass Grafting

CAD: Coronary Artery Disease

CAG: Coronary Angiography

CI: Confidence Interval

DICOM: Digital Imaging and Communications in Medicine

LAD: Left Anterior Descending Artery

LCA: Left Coronary Artery

LCX: Left Circumflex Artery

MAE: Mean Absolute Error

MHI: Montreal Heart Institute

PCI: Percutaneous Coronary Intervention

PPV: Positive Predictive Value (Precision)

QCA: Quantitative Coronary Angiography

r: Pearson's correlation coefficient

RCA: Right Coronary Artery

SD: Standard Deviation

UCSF: University of California, San Francisco

ABSTRACT

The coronary angiogram is the gold standard for evaluating the severity of coronary artery disease stenoses. Presently, the assessment is conducted visually by cardiologists, a method that lacks standardization. This study introduces DeepCoro, a ground-breaking AI-driven pipeline that integrates advanced vessel tracking and a video-based Swin3D model that was trained and validated on a large dataset comprised of 182,418 coronary angiographies spanning 4 years. DeepCoro achieved a notable precision of 71.89% in identifying coronary artery segments and demonstrated a mean absolute error of 20.15% (95% CI: 19.88-20.40) and a classification AUROC of 0.8294 (95% CI: 0.8215-0.8373) in stenosis percentage prediction compared to traditional cardiologist assessments. When compared to two expert interventional cardiologists, DeepCoro achieved lower variability than the clinical reports (19.09%; 95% CI: 18.55-19.58 vs 21.00%; 95% CI: 20.20-21.76, respectively). In addition, DeepCoro can be fine-tuned to a different modality type. When fine-tuned on quantitative coronary angiography assessments, DeepCoro attained an even lower mean absolute error of 7.75% (95% CI: 7.37-8.07), underscoring the reduced variability inherent to this method. This study establishes DeepCoro as the first video-based, adaptable tool in coronary artery disease analysis, significantly enhancing the precision and reliability of stenosis assessment.

INTRODUCTION

Cardiovascular diseases account for roughly 17.9 million annual deaths, making them the leading global cause of mortality [1]. A significant contributor is atherosclerotic coronary artery disease (CAD), where stenoses (i.e. obstructions caused by atherosclerotic plaque) can lead to myocardial infarction if untreated [2-5]. Reliable and accurate identification of the extent and severity of CAD directly impacts the decision to pursue an invasive revascularization procedure (i.e. percutaneous coronary intervention (PCI), or coronary artery bypass grafting (CABG)), generally conducted when stenoses are severe [2,5]. In addition, this stenosis assessment is essential to provide necessary treatment and prevent unnecessary revascularization [6]. These stenoses are usually identified through visual interpretation of coronary angiography (CAG) videos, a minimally invasive procedure involving iodine dye and X-ray imaging [4,7-9]. Despite the routine use of visual estimation of stenosis in CAG, this approach lacks standardization and shows high intra-observer and inter-observer variability, generally ranging from 6.9% to 26.4% between observers [10]. Yet, this visual assessment remains the clinical standard for assessing CAD severity [4,11,12] and is also endorsed by clinical guidelines [13,14]. Quantitative coronary angiography (QCA) offers more reproducible results but requires clinician input for image selection and is mainly used in research [6,15-17]. Hence, there is a need for an efficient and objective tool to assess coronary artery stenoses in routine clinical cardiology practice.

Artificial intelligence (AI) algorithms offer the potential for more standardized assessments of diagnostic tests, such as CAG, often performing as well or better than medical experts in various tasks [1]. However, existing AI methods for the interpretation of CAG face several challenges that hinder their clinical implementation: they often were trained on small datasets [3,8,18], have extensive exclusion criteria [3,15,19], rely on classifying vessels in CAG images as normal or abnormal instead of providing a continuous percentage of severity for every stenosis [8,19], and require human inputs to assist with the interpretation [2,3,15,18]. These limitations make them less representative of real-world clinical data. For example, some focus only on specific projection angles [3] or the simpler structure of the right coronary artery (RCA), avoiding the more complex left coronary artery (LCA) [3,15,19]. A recent method, CathAI [6], automates the assessment of stenosis severity from CAG images. However, its algorithm for identifying coronary artery segments could be improved due to suboptimal performance. Additionally, like other aforementioned methods, CathAI's use of static images rather than dynamic videos may overlook critical information that clinicians often obtain from video analysis for evaluating stenosis severity.[6]

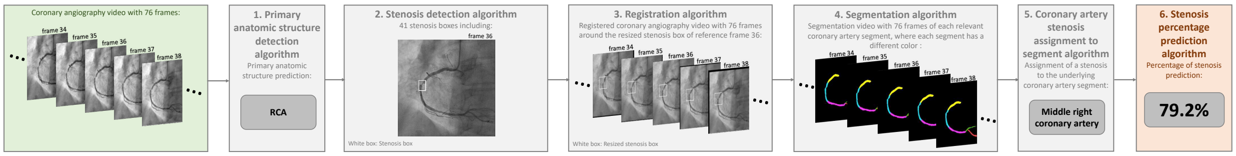
In this work, as a primary objective, we aimed to develop a video-based algorithmic pipeline called DeepCoro, which goal is to automatically localize stenosis and assess their severity in CAG videos, rather than images, of both the LCA and RCA, and evaluate its performance using visual assessments made by cardiologists on a large real-world CAG dataset spanning 5 years from the Montreal Heart Institute (MHI). DeepCoro is the first pipeline that leverages videos instead of static images for the automatic evaluation of CAGs. Distinguished by its innovative coronary artery segment recognition and stenosis percentage prediction algorithms, DeepCoro aims to enhance diagnostic accuracy by using the temporal dimensions present in CAG videos, mimicking the comprehensive assessment performed by cardiologists. The pipeline consists of six integral algorithms tackling various aspects from anatomic structure detection to stenosis percentage prediction. As secondary objectives, we aimed to benchmark the effectiveness of DeepCoro against an existing state-of-the-art image-based pipeline, CathAI [6], which was re-trained on the same dataset. Additionally, we aimed to evaluate the performance of DeepCoro against existing CAG evaluation methods, concentrating on its consistency relative to human evaluators and its correlation with QCA labels.

METHODS

DeepCoro Algorithm Development

DeepCoro employs a unique multi-step pipeline to detect and analyze stenosis in CAG videos stored in Digital Imaging and Communications in Medicine (DICOM) format. It leverages six specialized algorithms, where the data flows from one algorithm to the other. It has been trained and validated on an extensive dataset from MHI. DeepCoro's architecture builds on the foundational work of CathAI [6], integrating its essential algorithms for detecting primary anatomical structures and identifying stenosis (Algorithms 1 and 2; Figure 1). These algorithms were adopted without further training on our dataset. Our novel contribution is detailed in Algorithms 3-6, representing our advancements in this field [6]. DeepCoro initiates with the detection of primary anatomical structures (Algorithm 1), selectively focusing on videos pertaining to the RCA and LCA. It then employs RetinaNet [20] models (Algorithm 2) to locate stenoses within these coronary segments. A registration algorithm (Algorithm 3) follows, aligning frames to account for heart and respiratory movement, creating a stable video in reference to a stenotic coronary segment. Our sophisticated multi-class segmentation algorithm (Algorithm 4) plays a pivotal role in the process by categorizing the coronary artery into proximal, middle, and distal segments. Algorithm 5, in each frame, evaluates the content of the resized stenosis box, focusing on pixels within the central region of the reference area, which are matched to the underlying coronary artery segment as predicted by Algorithm 4. This method is designed around the typical central placement of stenosis within annotations, thus concentrating on this area with the assumption that it houses the relevant segment. Lastly, a stenosis severity prediction algorithm (Algorithm 6), using a modified Swin3D [21] transformer model, quantifies stenosis severity from the aligned video to predict a continuous percentage in targeted artery segments. This integrated approach facilitates automatic interpretation of CAG videos without any human input. More details are available in Supplementary Methods.

Figure 1. DeepCoro Pipeline Overview



Legend. Overview of DeepCoro's algorithmic pipeline and example of outputs from each algorithm of a 76 frames coronary angiography video. In practice, steps 3 through 6 must be performed for all stenosis boxes detected at step 2, but the figure shows an example using frame 36 as a reference frame. **White box**: Stenosis localisation box. **Green background box**: DeepCoro's input representing videos of left or right coronary angiograms. **Grey background box**: DeepCoro's intermediary output. **Orange background box**: DeepCoro's final output representing a continuous stenosis percentage as well as the underlying coronary artery segment with the stenosis. **Abbreviations**: RCA: Right Coronary Artery.

Algorithm 1: Primary Anatomic Structure Detection Algorithm

The first algorithm used an Xception [22] image-based model previously trained on over 14,000 CAG images which were annotated by cardiologists at the University of California, San Francisco (UCSF). This model was trained as part of CathAI's pipeline and results were previously published [6]. The algorithm distinguished the primary anatomic structures, like the RCA, LCA, aorta, radial artery, left ventricle, catheter, and femoral artery, present in the most frames of the video (Supplementary Table 3) and was used to exclude videos not mostly containing the RCA or LCA from further analyses.

Algorithm 2: Stenosis Detection Algorithm

The second algorithm uses the RetinaNet [20] architecture, a state-of-the-art model for object detection, to pinpoint the locations of coronary artery segments and stenoses. It achieves this by drawing bounding boxes around these areas, thereby defining their precise coordinates. RetinaNet models were previously trained on UCSF data as part of the previously published CathAI's pipeline to detect specific anatomical structures, procedural instruments associated to a PCI and stenoses (Supplementary Table 3), and were previously published [6]. The RetinaNet models were applied to all frames of LCA and RCA videos, as identified by Algorithm 1.

Each detected stenosis overlapping with a coronary artery's bounding box was preliminarily assigned to the respective artery segment, in a specific frame (Supplementary Figure 1a). To optimize dataset handling and computational efficiency, especially for Algorithm 3, we limited inclusion to just one stenosis box per artery segment, in each video. We selected the central instance which is typically near the video's midpoint, characterized by peak dye intensity, enhancing vessel visibility and thus, the accuracy of subsequent algorithm predictions in the DeepCoro pipeline. This selection criterion was crucial to prevent an overwhelming dataset size that would result from including multiple instances of the same coronary artery segment stenosis. By focusing on the most centrally located stenosis in the temporal dimension, particularly for videos with multiple stenoses linked to the same artery segment, we efficiently managed computational resources. This approach of assigning stenoses to coronary segments is the one used as part of CathAI's pipeline [6] and it was compared to the method proposed under Algorithm 5 [6].

Algorithm 3: Registration Algorithm

Given the inherent motion of cardiac structures during systole or breathing, this algorithm aims to align the stenosis bounding box derived by Algorithm 2, resized to measure the closest to 17.5 mm by 17.5 mm. To achieve this, spatial translations were used for aligning the stenosis box in a reference frame to previous and subsequent frames, guided by a Discriminative Correlation Filter from the OpenCV Python library (Supplementary Figure 2) [23]. A registered video was generated for each stenosis box obtained after Algorithm 2.

Algorithm 4: Segmentation Algorithm

Algorithm 4 segments full videos by applying an ensemble of seven segmentation algorithms (Supplementary Table 10) frame by frame, to generate registered multi-class segmented videos depicting 11 epicardial coronary artery segments (5 for the RCA – i.e. proximal RCA, middle RCA, distal RCA, posterolateral branch from the RCA and posterior descending artery – and 6 for the LCA – i.e. left main artery, proximal left anterior descending artery (LAD), middle LAD, distal LAD, proximal left circumflex artery (LCX) and distal LCX), based on SYNTAX score definition (Supplementary Table 3) [24]. These algorithms were refined through extensive testing of nine advanced segmentation models over 200 epochs, combined with five different loss functions, and optimized through a random search of batch sizes and learning rates, validated through 5-fold cross-validation.

Algorithm 5: Coronary Artery Stenosis Assignment to Segment Algorithm

This algorithm pinpointed the coronary artery segment affected by stenosis. It examined the central pixels of the resized stenosis-indicative region of the segmented registered video in each frame (Supplementary Figure 1b). These central pixels were matched with the coronary segments identified by Algorithm 4. The final video prediction was the segment that was most consistently predicted across frames. When the algorithm is unable to predict any artery segments in a video, no further predictions are made in that video.

For assessing stenosis severity at the video-level, a targeted approach is employed. When multiple videos from the same DICOM file are related to an identical artery segment, only the central instance is retained for the final calculation of stenosis percentage. The remaining videos, which are effectively duplicates for the same coronary segment, are removed. While most of these duplicates are initially filtered out by Algorithm 2, there are instances where Algorithm 5 might reassign videos to the same segment. In such cases, only one video per segment is kept. This strategy ensures a thorough and accurate evaluation of stenosis severity.

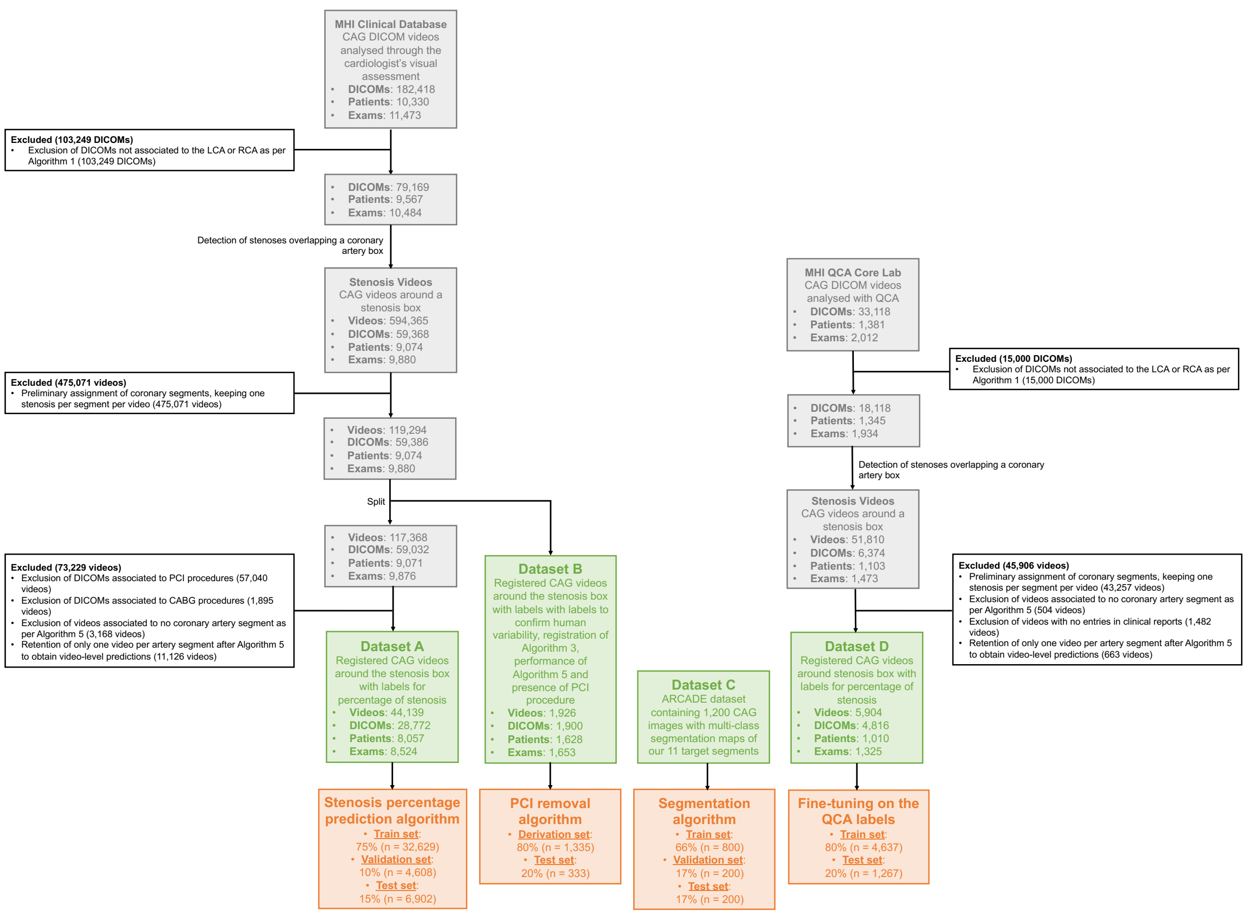
Algorithm 6: Stenosis Percentage Prediction Algorithm

For the final step, DeepCoro uses a modified version of the Swin3D [21] architecture, a state-of-the-art video classification transformer model adapted for regression tasks to output stenoses percentages ranging between 0 to 100%. Through extensive testing, we have established that our final Algorithm 6 stands as the most optimal among the options considered (Supplementary Table 11).

Description of datasets

Our algorithms were trained and validated using four distinct datasets, detailed below, and illustrated in Figure 2 (Supplementary Figure 5 shows a detailed form of this figure). Comprehensive dataset characteristics and patient demographics are available in Supplementary Table 1 and 2. We eliminated videos that did not feature the LCA or RCA, as identified by Algorithm 1. We further reduced the dataset due to computational constraints as outlined in the methodology associated with Algorithm 2, keeping only one video per DICOM per coronary artery segment preliminary predictions. Videos depicting PCI and CABG procedures were also removed. Any videos that did not display recognizable coronary artery segments as determined by Algorithm 5 were excluded. To facilitate video-level predictions, we maintained a single video for each artery segment in a DICOM after the application of Algorithm 5.

Figure 2. Datasets and patients used to train and validate DeepCoro.



Legend. Datasets change in size when our algorithms are applied to our datasets. ARCADE is a public dataset described in [25]. **White box**: Exclusion details. **Grey box**: Intermediate datasets. **Green box**: Final datasets. **Orange box**: Dataset split for the development of an algorithm. **Abbreviations**: ARCADE: Automatic Region-based Coronary Artery Disease diagnostics using X-ray angiography imagEs, CABG: Coronary Artery Bypass Grafting, CAG: Coronary Angiography, DICOM: Digital Imaging and Communications in Medicine, MHI: Montreal Heart Institute, PCI: Percutaneous Coronary Intervention, QCA: Quantitative Coronary Angiography.

Dataset A was derived from the MHI's comprehensive CAG videos database, which contains 182,418 videos of LCA or RCA in DICOM format, captured at 15 frames per second, and includes patients aged 18 years and older, spanning the period from January 1, 2017, to December 31, 2021. The dataset was used for the training, validation, and testing of DeepCoro. Clinical reports detail the percentage of coronary artery segment stenosis, ranging from 0 to 100%, associated to these videos as described by the interventional cardiologist after visual assessment. Segments with ≥70% stenosis were classified as severe, while those with less were labeled as non-severe. After applying Algorithms 1 to 5 to this database, we obtained a final dataset of 44,139 videos from 8,057 patients with associated stenosis percentages from clinical reports. This dataset was partitioned into mutually exclusive training (75%), validation (10%), and test (15%) sets, ensuring that each patient could only belong to one of these subsets, for the training and evaluation of DeepCoro's Algorithm 6.

Dataset B consists of 1,926 videos of the LCA and RCA from 1,628 patients, randomly selected from the MHI database. This selection occurred after the initial application of Algorithms 1 to 3. The purpose of this specific curation was to ensure that the dataset included only LCAs and RCAs with stenoses, facilitating the testing of our pipeline's performance against a human-annotated dataset. All videos were simultaneously annotated by two cardiologists (interface shown in Supplementary Figure 3) with over 10 years of experience in reading CAG videos to describe stenosis percentage (1,926 labels), correctness of the artery segment registration (Algorithm 3; 1,926 labels), identification of the coronary artery segment with a stenosis (1,926 labels) and presence of PCI in the video (i.e. presence of a guidewire, balloon, or stent; 1,668 labels). The purposes of this dataset were to assess inter-observer variability, determine the performance of Algorithm 3, 4 and 6, and develop an automatic algorithm to remove videos associated to PCI from our dataset. For the latter, we randomly divided the dataset into an 80% derivation set (to identify the optimal threshold for excluding PCI procedures) and a 20% hold-out test set where we tested this threshold to exclude PCI procedures. A video was deemed correctly registered when the targeted coronary segment remained in the predefined stenosis box from the onset of iodine dye visibility to its disappearance. For evaluating Algorithm 4's accuracy, we excluded videos that did not correspond to the 11 pre-specified coronary segments as determined by both the algorithm and a cardiologist, removing 119 videos and reducing the dataset from 1,926 to 1,807 videos. To assess the performance of Algorithm 4, the Dataset B was cleaned from PCI and CABG procedures with our algorithm; in total, 965 videos remained in Dataset B.

Dataset C is sourced from the Automatic Region-based Coronary Artery Disease diagnostics using X-ray angiography imagEs (ARCADE) [25] public dataset and includes 1,200 X-ray CAG images, with 25 different multi-class coronary artery segmentations. We only considered the 11 coronary artery segments pertinent to our approach. Segments with very few annotated examples were excluded. We randomly split 800 images (66% of the dataset) for Algorithm 4 training and 200 images each (17% of the dataset each) for validation and testing.

Dataset D was extracted from the MHI QCA Core Laboratory's, which is a separate dataset from the MHI clinical database that features CAG videos from randomized controlled trials focused on lipid-lowering therapies [26]. This dataset comprised a unique patient population distinct from our primary clinical dataset, as it mainly contained mild-to-moderate coronary stenoses, with an average QCA stenosis severity of 33.7%±11.7%. These angiograms were recorded at 15 frames per second. Each angiogram underwent QCA analysis by trained technicians and was supervised by an expert physician. For this dataset, stenosis was categorized as severe with QCA stenosis percentage of ≥50% [27,28]. After applying Algorithms 1, 2, 3, and 5, the dataset was narrowed down from 33,118 CAG video to 5,904 videos cropped around stenoses. This dataset provided an ideal setting to evaluate DeepCoro adaptability to different clinical contexts, using QCA stenosis percentage labels for retraining. Patients' videos were randomly broken down into an 80%-20% split, resulting into a training set of 4,637 videos and a test set of 1,267 videos for the external validation of DeepCoro.

Removal of PCI procedures

The study aimed to ensure accurate labeling of CAG videos for diagnostic evaluation, prior to any plaque modification during a PCI. We excluded videos taken during or following a PCI because the stenosis observed could be modified during the intervention. This modification could lead to discrepancies between the video and the initial diagnostic stenosis labels. For example, a significant stenosis visualized during balloon angioplasty type PCI might appear less severe in the imaging captured during the procedure. Such intra-procedural changes in stenosis are usually not recorded; only the diagnostic stenosis percentage before PCI and the result after PCI are typically reported. However, we kept the initial diagnostic CAG videos of patients who subsequently underwent PCI treatments. To distinguish diagnostic CAG videos from those recording PCI, we developed a dual-method approach for identifying videos linked to PCI procedures:

1. **Method 1:** We used Algorithm 2 to identify frames containing procedural instruments indicative of a PCI (i.e. stent, balloon, guidewire). A metric *m1* was calculated by summing the number of frames containing the instruments according to RetinaNet [20] predictions, then dividing by the total frame count. For a video, if *m1* was greater or equal to a pre-determined threshold, the video was identified as a PCI-related. This threshold was optimized using Youden's index, applied to 100 equally spaced thresholds between 0 and 1 in a derivation set and then compared to the human annotations.
2. **Method 2:** We also leveraged clinical reports that specify the start and end times of each PCI procedure for both RCA and LCA views. Videos within these timeframes were flagged as PCI-related. Due to possible time discrepancies between clinical report timestamps and the video recording, an offset value *m2* was subtracted from each stenting event's start time to account for inconsistencies. The optimal offset was determined by maximizing Youden's index a derivation set, considering offsets of 0, 5, 10, 15, 20, 25, and 30 minutes.

The optimal thresholds for both Method 1 and Method 2 were initially determined using the derivation set of Dataset B. These thresholds were then validated on a separate hold-out test set of Dataset B. The combined approach associates a video to a PCI if one is detected with either of the two methods, for subsequent removal. By combining both methods, we systematically identified and excluded videos associated with PCI procedures, as well as any subsequent videos of a patient, from Dataset A. This ensures that DeepCoro is only trained on diagnostic CAG videos to make sure the labeling of coronary artery stenoses is the most accurate.

Removal of patients with CABG

We also excluded videos connected to CABG procedures from Dataset A. This is due to the significant anatomical differences in patients who have undergone CABG surgery, which interferes with our objective of creating a universally applicable analysis pipeline for native coronary vessels. Clinical reports indicate whether a patient has undergone a CABG operation. Therefore, patients featuring bypassed vessels were excluded.

Algorithm Evaluation and Statistical Analysis

Primary Objective: Assessment of the DeepCoro Algorithm Suite

Algorithm 1 and 2 performances were previously published [6]. Our current objective was to evaluate the novel Algorithm 3 to 6 within the DeepCoro pipeline. For Algorithm 3, video registration performance was defined by the proportion of correctly registered videos, determined using human annotated data from Dataset B. Algorithm 4's segmentation quality was validated on Dataset C's test set by employing the Dice Score, positive predictive value (PPV), and sensitivity for the delineation of the 11 targeted coronary artery segments. We also described sensitivity, PPV, and the F1-score for each of the 11 coronary artery segments of interest against ground truth labels from Dataset B. Algorithm 6's efficacy in stenosis percentage prediction was gauged using mean absolute error (MAE) and Pearson's correlation coefficient (r) on Dataset A's test set. Its capability to classify stenosis severity (non-severe versus severe) was assessed using the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, PPV, and the area under the precision-recall curve (AUPRC). Clinician-reported stenosis percentages, which were dichotomized for severity classification at a 70% threshold, served as the gold standard for Dataset A, whereas a 50% threshold [27,28] was used in Dataset D (evaluated through QCA). Due to important class imbalance favoring the non-severe stenosis class [29], AUPRC and AUROC best represent DeepCoro's performance for a comprehensive performance overview. Binarization of predictions used the optimal threshold derived from maximizing Youden's index among 100 equally spaced thresholds between 0 and 1 on the validation set. Our primary performance was assessed at the artery-level which was derived by computing the mean stenosis severity for each coronary segment, using videos from the same patient taken on a single day. This approach provided an aggregated measure of stenosis for each artery segment, leveraging multiple angiographic views to enhance the robustness of the assessment.

We calculated 95% confidence intervals (CIs) using the bootstrapping method, which involves randomly selecting 80% of the test set data to recompute the performance metrics. This process was repeated across 1,000 iterations, and the CIs were established from the resulting distributions of these metrics.

Comparison with the CathAI pipeline

CathAI represents a notable advancement in AI-driven automated interpretation of CAGs [6]. However, it has limitations, such as suboptimal coronary artery segment prediction and reliance on still images, contrasting with the dynamic video analysis typically employed by cardiologists. Our goal was to compare our video-based DeepCoro pipeline with CathAI’s image-based one to assess if it could address these issues and enhance overall performance.

We first compared coronary artery segment predictions between CathAI and DeepCoro on Dataset B, using human annotations as the standard. CathAI's approach, using RetinaNet [20] models for stenosis identification, was evaluated against DeepCoro's segmentation-based method (Algorithm 5). Performance metrics PPV, F1-Score, and sensitivity were computed for both systems.

Further, we analyzed stenosis severity on Dataset A, contrasting CathAI's image-based model with DeepCoro's video-based approach. To ensure a fair comparison for this task, we integrated DeepCoro's coronary artery segment predictions into CathAI's framework rather than using the coronary segment prediction from RetinaNet. CathAI's stenoses, delineated with bounding boxes, were processed, and compared with DeepCoro's Algorithm 6. Training parameters for CathAI, including epochs, learning rate, optimizer, and batch size, were aligned with DeepCoro's settings (Supplementary Methods). The comparison was based on uniform metrics, with model superiority determined by non-overlapping 95% CIs. We applied the DeLong test [30] to evaluate statistical differences of AUROC and AUPRC between CathAI and DeepCoro for stenosis severity categorization, with a p-value below 0.05 indicating a significant difference.

Assessment of Inter-Observer Variability

To evaluate inter-observer variability, we conducted a comparative analysis between the performance of DeepCoro's Algorithm 6 and the clinical reports annotations, against the average annotations from two expert interventional cardiologists with over ten years of experience provided in Dataset B. These annotations served as the ground truth, against which we compared the performance of Algorithm 6 as well as the accuracy of clinical reports. Clinical report predictions were obtained by retrieving the percentage of obstructions of the DICOM in clinical reports associated to the artery segment identified by the cardiologist in Dataset B, and were binarized with a 70% threshold. Key metrics such as the AUROC, AUPRC, sensitivity, specificity, MAE, and r were computed to quantitatively benchmark the performance of Algorithm 6 and clinical reports against the human annotated stenosis percentage.

External validation against QCA

Fine-tuning is pivotal for adapting DeepCoro's Algorithm 6 to new tasks, such as predicting stenosis percentages that match the inherently more consistent measurements from QCA. To test its capability for this precision-demanding task, we fine-tuned the algorithm using the train set of Dataset D—CAG videos annotated with QCA by the MHI core lab as the gold standard. The fine-tuning involved keeping the model's parameters fixed except the last two linear layers which are adjusted over 100 epochs. We varied learning rates from 1e-2 to 1e-7 and determined the optimal rate to be 1e-3, based on the lowest loss achieved. The model, optimized for QCA annotations, was then rigorously evaluated on the test set to confirm its enhanced regression accuracy.

Human Subjects Research

This study was reviewed and approved by the MHI Institutional Review Board. The need for individual informed consent was waived.

RESULTS

Performance of the DeepCoro pipeline

To assess the efficacy of the DeepCoro pipeline, we evaluated the performance of its component algorithms (Algorithms 3-6) as well as our PCI removal algorithm. Dataset A was a central dataset for training and evaluating DeepCoro. We had 182,418 CAG videos in DICOM format of the LCA and RCA in our MHI clinical database which, after applying Algorithm 1 to 5, resulted in 44,138 cropped videos of stenosed coronary artery segments that formed Dataset A. Average age of patients was 67.6±11.0 years old, and 12,917 (29%) were identified as female and 30,067 (68%), as male. The average stenosis percentage was 20.6±30.2% and we had 16% severe stenoses. The relatively low mean and wide standard deviation (SD) reflect a wide spectrum of stenosis percentages, including a substantial proportion of segments with 0% stenosis as identified by cardiologists, while our stenosis detection algorithm detected one.

A significant challenge in analyzing CAG videos is the movement of vessels within the videos, which can complicate the analysis of specific areas of interest. Algorithm 3 successfully registered 96.63% of the videos in Dataset B, as verified by expert annotators. The videos that were not correctly registered mostly either contained obstructive background elements, such as pacemaker leads or sternotomy wires, or suffered from poor contrast injections.

For the multi-vessel segmentation algorithm (Algorithm 4), our best performance was observed using an ensemble of seven models, averaging the predictions across the different models. We obtained a Dice score, PPV and sensitivity of respectively 73.93%, 75.96% and 70.12% using a weighted average across segments in the test set of Dataset C (Supplementary Table 4). This suggests that there is a strong agreement and significant overlap between the predicted segmentations and ground truths. For coronary artery segment prediction, we obtained similar performances against human annotators on Dataset B for Algorithm 5 with a PPV, sensitivity and F1-score across all 11 coronary segments of 71.89%, 70.72% and 70.71% respectively (Supplementary Table 5). Owing to the pipeline's high accuracy for upstream algorithms, it was subsequently applied to our dataset for the training and inference stages of Algorithms 6.

We optimized DeepCoro's PCI removal algorithm for high sensitivity to ensure accurate identification and exclusion of PCI-related videos, as PCIs can greatly increase error rates in the stenoses labelling, due to plaque modification. Using Youden's index, we determined the optimal cut-offs to be 0.16 for Method 1 and a 25-minute offset for Method 2. Individually, Method 1 showed a sensitivity of 91.89%, and Method 2, 79.73% and, when combined, the sensitivity increased to 95.27%, indicating that the integrated approach effectively identified most video recordings of PCI procedures for subsequent removal. After determining that the registration, multi-vessel segmentation and PCI removal algorithm were working properly, we excluded all videos that were associated to PCI or post-PCI (n=57,040 videos) and patients that underwent a CABG (n=1,895 videos) from the MHI clinical database.

Next, Algorithm 6 demonstrated high classification performance and moderate correlations for both vessels on Dataset A using the Swin3D approach compared to other video-classification models (Supplementary Table 11). Specifically, the RCA exhibited higher stenosis severity classification performance (AUROC=0.8643; 95% CI: 0.8537-0.8745), sensitivity (76.20%; 95% CI: 73.98-78.60), and precision-recall balance (AUPRC=0.5578: 95% CI: 0.5242-0.5890), along with a stronger correlation coefficient (r=0.6200; 95% CI: 0.6018-0.6372) for stenosis percentage regression task (Table 1). These results suggest that the model is more adept at identifying and quantifying severe stenoses in RCA exams than LCA. Stratified results across age groups and sex are shown in Supplementary Table 8. DeepCoro's predictions plotted against ground truth for Dataset A are shown in Supplementary Figure 4a.

Comparison with the CathAI pipeline

In a head-to-head comparison of CAG interpretation tools, DeepCoro significantly outperformed CathAI [6]. The evaluation focused on stenosis assignment to coronary artery segments and stenosis percentage prediction performances. DeepCoro's segmentation method demonstrated a more accurate stenosis assignment, achieving a higher overall average PPV (71.89% versus 59.10%), sensitivity (70.72% versus 56.50%), and F1-score (70.71% versus 56.50%) than CathAI's bounding box approach. These results were consistently robust across individual segments in Dataset B, indicating DeepCoro's reliable performance enhancement without significant variance, as detailed in Supplementary Table 5.

Secondly, comparing stenosis classification in combined LCA and RCA, DeepCoro's video-based approach demonstrated significant improvements over CathAI with AUROC values at the artery-level of 0.8294 (95% CI: 0.8215-0.8373) versus 0.7953 (95% CI: 0.7875-0.8038) ($p<0.01$, as determined by DeLong's test) and AUPRC values of 0.5239 (95% CI: 0.5041-0.5421) versus 0.4670 (95% CI: 0.4497-0.4849) ($p<0.01$, as determined by DeLong's test). In addition, Algorithm 6 (DeepCoro's video-based model) had lower variability with the report stenosis, with a MAE of 20.15% (95% CI: 19.88-20.40) versus 21.61% (95% CI: 21.35-21.87), and higher correlation, with a r of 0.5497 (95% CI: 0.5360-0.5630) versus 0.4571 (95% CI: 0.4430-0.4711). We present a comparative performance table at the artery-level and video-level for both approaches in Supplementary Table 6 and 7 respectively which demonstrates superior performance for DeepCoro's video-based model over the CathAI's image-based model.

Inter-Observer Variability of Visual Assessment Methods

In the expert-annotated Dataset B, DeepCoro demonstrated an AUROC of 0.8699 (95% CI: 0.8564-0.8853) and an AUPRC of 0.7042 (95% CI: 0.6717-0.7339) for combined LCA and RCA in stenosis severity classification, surpassing the clinical reports which showed an AUROC of 0.7533 (95% CI: 0.7328-0.7744) and an AUPRC of 0.4737 (95% CI: 0.4382-0.5086) (Table 2). For regression tasks, DeepCoro's MAE was 19.09% (95% CI: 18.55-19.58) with a r of 0.6792 (95% CI: 0.6598-0.7004), in contrast to clinical reports with an MAE of 21.00% (95% CI: 20.20-21.76) and a r of 0.5000 (0.4702-0.5302; Table 2; Supplementary Figure 4b). Overall, DeepCoro's performance was closer to two experts annotating the CAG videos, rather than the clinical report, demonstrating DeepCoro's potential as a standardized approach to assess CAG videos. Ultimately, DeepCoro's Algorithm 6 accuracy more closely mirrored the assessments of two expert cardiologists in annotating CAG videos and was superior to the conventional clinical report.

Re-training and performance of DeepCoro on the MHI QCA Dataset – Dataset D

To verify if the performance of our model mirrors the lower variability of QCA, we fine-tuned and tested the regression performance of our pipeline on Dataset D, a dataset of CAG videos which uses QCA assessment as ground truth. The average percentage annotated in this dataset is 33.7 ± 11.7 %. Overall, the fine-tuned model had a MAE against QCA of 7.75% (95% CI: 7.37-8.07; Supplementary Table 9; Supplementary Figure 4c).

DISCUSSION

In the current study, we introduced DeepCoro, the first video-based pipeline for the interpretation of CAG videos. Our approach describes a novel ensemble segmentation approach, artery tracking algorithm and stenosis percentage prediction algorithm which mark a significant advancement in the automated analysis of CAG videos. We demonstrated good classification and regression performance for stenosis severity assessment on a large real-world dataset, spanning 5 years and over 40,000 CAG videos. Second, when benchmarked against the existing image-based CathAI pipeline [6], DeepCoro demonstrated superior performance, setting a new standard for state-of-the-art automatic CAG interpretation. Third, DeepCoro's performance not only aligned more closely with the expert evaluations from seasoned cardiologists but also exhibited lower variability when compared to clinical reports, enhancing the reliability of CAG video assessments. Fourth, our findings underscore DeepCoro's versatility, showing its ability to be fine-tuned for diverse applications, such as QCA assessments, where it displayed acceptable classification accuracy. Finally, our model weights were made publicly available which will accelerate the research in this field by enabling researchers and cardiologists to fine-tune them to their own dataset and develop novel applications.

During the evaluation of DeepCoro's Algorithm 6 for stenosis prediction, we noted a MAE of 20.15% (95% CI: 19.88-20.40), which falls within the typical variability range (6.9 to 26.5% [31]) noted among practitioners as reported in the literature. By aligning Algorithm 6 with the consensus annotations from two veteran interventional cardiologists, we reduced its variability to 19.09% (18.55-19.58), outperforming the inter-observer variability of 21.00% (20.20-21.76) found within this dataset. The precision of Algorithm 6 was further enhanced through calibration with QCA assessments. This fine-tuning highlights DeepCoro's capability to diminish variability and improve the accuracy of stenosis evaluations, with results reflecting the nature of the training or fine-tuning dataset. DeepCoro's benefits extend beyond its capacity to reduce variability. The algorithm's design also allows for scalability across different datasets and adaptability to new diagnostic criteria, potentially setting a new standard for reproducibility in CAG interpretation. Its application could lead to more consistent and reliable stenosis assessments, by acting as an independent observer in the interpretation of CAG, which could lead to better-informed clinical decisions and potentially improving patient outcomes by ensuring a higher degree of diagnostic accuracy.

DeepCoro demonstrated an AUROC of 0.8294 (0.8215-0.8373) on a comprehensive real-world dataset, aligning with prior stenosis classification research but significantly expanding on the scope with data spanning five years. This contrasts with earlier studies that used under 500 annotated angiogram frames, but obtained a higher AUC of 0.97 [19]. Au *et al.* achieved an AUROC of 0.825 for their stenosis severity classification algorithm alone applied solely on RCA videos, but their performance declined when they applied their pipeline end-to-end to automatically interpret angiograms [15]. Notably, our end-to-end pipeline maintained a comparable performance for both RCA and LCA videos, matching the AUROC scores of algorithms tested solely on less complex RCA images. [15] Zhao *et al.* achieved a sensitivity of 55.56% and PPV of 50.00% for severe stenosis classification, which is comparable to our pipeline, which displayed a sensitivity of 67.64% (66.09-69.31) but PPV of 40.25% (38.97-41.55) [18]. Crucially, our pipeline's validity is bolstered by a dataset vastly larger than the 99-patient cohort used by Zhao *et al.*, enhancing the clinical applicability of our findings [18]. Zhou *et al.* achieved a MAE of 15.9±13.3% in a study focusing on RCA stenoses among 102 patients [3]. This highlights DeepCoro's wider clinical applicability and robustness. Our model's accuracy, combined with a diverse and extensive real-world dataset, offers superior generalizability compared to earlier studies that were restricted by smaller patient numbers and limited scope in terms of views analyzed.

Perhaps the most comprehensive work involved an algorithmic pipeline called "CathAI" that was trained on 13,843 studies spanning 4 years of data. They achieved an AUC of 0.862 (95% CI: 0.843-0.880) [6]. CathAI, while a state-of-the-art pipeline, had several limitations, notably its reliance on static images rather than video data. This approach may overlook crucial temporal information, which is essential for accurate diagnosis. Video-based models, by leveraging the inherent variability in multiple cardiac cycles, enhance diagnostic precision and effectively address the fluctuations in cardiac function that occur from one heartbeat to the next. [32] For instance, EchoNet-Dynamic, a video-based deep learning algorithm, exemplifies this advancement by outperforming human experts and image-based models in key diagnostic tasks like left ventricle segmentation and ejection fraction estimation from echocardiographic videos, showcasing the potential for video models to improve reproducibility and precision in cardiovascular disease diagnosis. [32] DeepCoro addressed these challenges by using a video-based analysis framework, effectively capturing the dynamic nature of cardiac cycles and thereby enabling accurate stenosis estimation. In our study, DeepCoro consistently surpassed CathAI [6] in regression metrics across all coronary arteries and matched or outperformed it in most classification metrics. The system proved especially effective in assessing the LCA, demonstrating the superior capability of video-models in analyzing complex anatomical structures. DeepCoro excelled in accurately assigning stenoses to coronary segments, significantly reducing data mislabeling. This contrasts with CathAI's RetinaNet approach, which sometimes incorrectly interpreted the coronary artery tree's structure by including irrelevant background details in its predictions. In contrast, DeepCoro's segmentation method effectively recognized the interconnectedness of artery segments. This approach is more aligned with the methods cardiologists use, assessing artery segments based on their positions within the coronary artery tree, thereby enhancing stenosis prediction precision and the overall robustness of the algorithm.

Direct comparisons between DeepCoro and other models in existing literature present challenges, primarily due to the unique and extensive dataset that our approach used, a dataset not commonly used for testing other methods. We rigorously applied DeepCoro to a comprehensive collection of videos, covering both the RCA and LCA from all projection angles. This approach was undertaken without imposing extensive exclusion criteria, ensuring a broad and representative dataset that accurately reflects real-world clinical scenarios. In contrast, other models in the field often rely on partial automation or restrictive selection criteria, which may not capture the full spectrum of clinical scenarios, making DeepCoro a more comprehensive and clinically pertinent tool.

We also demonstrated that DeepCoro can be applied to new tasks. Upon fine-tuning with video annotations derived from QCA, DeepCoro achieved a reduced MAE of 7.75% (95% CI: 7.37-8.07%), aligning with the reduced variability typically associated with QCA compared to visual assessment. This MAE is notably less than the 10 to 17% variability range reported in literature when comparing QCA annotations with visual assessments [28]. Such approach could be undertaken in the future to fine-tune DeepCoro for calcium estimation, identifying the vulnerable plaque [33] or predicting the physiological impact of stenoses [34].

Results stratified across age groups and sexes demonstrate no significant bias on Dataset A’s test set (Supplementary Table 8), although the best performance was observed at extremes of age and those aged 60 to 75 had slightly reduced performance. Some differences in performance may be due to the varying proportion of severe stenoses across groups. Moreover, as individuals age, their vessels typically become stiffer due to increased calcification and changes in the vessel walls [35]. This age-related transformation in the coronary arteries' structure can influence the diagnostic process and the accuracy of interpretations derived from CAG videos.

Finally, we are the first to make our model weights publicly available, a step that promises to expedite progress in this research domain. By providing access to these weights, cardiologists may fine-tune the models to their specific datasets and foster the development of new applications for interpretation of CAGs.

Understanding the limitations of DeepCoro is crucial for a comprehensive evaluation of its capabilities. A primary limitation is that the stenosis percentage used for training and testing is based on clinician interpretation from CAG videos, which may not always align with the actual stenosis values. This discrepancy underscores the potential benefit of employing large-scale datasets analyzed through objective measures like QCA to improve reproducibility. Despite this, our approach, when compared to the clinical report in Dataset B, had lower variability, suggesting that DeepCoro could be used to reduce variability in CAG interpretation. Of note, DeepCoro tends to underestimate stenosis severity reflecting the underlying stenosis distribution, likely due to the high prevalence of non-severe stenoses in the training dataset. Also, the current version of DeepCoro is optimized for detecting stenoses in 11 specific coronary segments on patients that did not receive a CABG surgery, thus the performance of our approach on bypass grafts or on non-included segments is unknown. Furthermore, the pipeline's reliance on multiple algorithms, each introduced a compounding degree of error, potentially affecting the final output's quality. For example, our registration algorithm did not perfectly align all videos, with only 96.63% correctly registered, suggesting a need for further refinement to achieve consistent success across all cases. However, our approach remains the most comprehensive one, it underscores the different steps that must be taken to automatically interpret CAGs and these performances can be improved by annotating more data.

In conclusion, DeepCoro marks a substantial leap in CAG video interpretation. This multi-step video analysis pipeline adeptly mirrors the dynamic analysis conducted by cardiologists, offering a more standardized approach to CAG evaluation. It shows great promise in reducing variability in clinical assessments and is versatile enough for tasks like interpreting QCA labels. Looking to the future, DeepCoro has the potential for further development to identify stenoses in more anatomically complex coronary segments such as the marginal or diagonal arteries. Additionally, it could be enhanced to evaluate other critical features in CAG videos, like calcification severity, which plays a vital role in plaque stability assessment and treatment planning. A crucial next step is to assess DeepCoro's impact on clinical decision-making, particularly regarding revascularization strategies. Conducting a randomized controlled trial to compare revascularization decisions based on AI-assisted CAG interpretations versus traditional methods will be key in understanding DeepCoro's effect on clinical outcomes.

AKNOWLEDGEMENT

This study was funded by the Fonds de la recherche en santé du Québec (Grant 312758), the Montreal Heart Institute Research Centre, the Montreal Heart Institute Foundation, the Des Groseillers-Bérard Interventional Cardiology Research Chair, Natural Sciences and Engineering Research Council of Canada (NSERC), the Institute for Data Valorization (IVADO) and the Fonds de recherche du Québec — Nature et technologies (FRQNT). The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Table 1. Artery-Level Performance of DeepCoro in the Test Set of Dataset A

Task	Metric	Coronary Artery		
		LCA	*RCA*	*LCA + RCA*
Number of exams		2,568	2,259	4,827
Number of severe stenoses, ⩾ 70% (n (%))		536 (21%)	345 (15%)	881 (18%)
Number of non-severe stenoses, <70% (n (%))		2,032 (81%)	1,914 (85%)	3,946 (82%)
Number of healthy vessels, 0% stenoses (n (%))		1,253 (49%)	1,075 (48%)	2,328 (48%)
Classification*	AUROC	0.8017 (0.7919 - 0.8124)	0.8643 (0.8537 - 0.8745)	0.8294 (0.8215 - 0.8373)
	AUPRC	0.5092 (0.4868 - 0.5329)	0.5578 (0.5242 - 0.5890)	0.5239 (0.5041 - 0.5421)
	Sensitivity (%)	70.70 (68.75 - 72.73)	76.20 (73.98 - 78.60)	72.86 (71.24 - 74.47)
	Specificity (%)	74.51 (73.56 - 75.43)	79.03 (78.10 - 80.04)	76.71 (76.05 - 77.36)
	PPV (%)	41.06 (39.48 - 42.70)	37.08 (35.11 - 39.00)	39.42 (38.15 - 40.68)
	F1-score (%)	51.95 (50.32 - 53.58)	49.88 (47.86 - 51.78)	51.15 (49.81 - 52.39)
Regression	MAE (%)	22.19 (21.82 - 22.52)	17.82 (17.48 - 18.16)	20.15 (19.88 - 20.40)
	r	0.4890 (0.4704 - 0.5087)	0.6200 (0.6018 - 0.6372)	0.5497 (0.5360 - 0.5630)

Legend. The performance at the artery-level of the DeepCoro's pipeline on the test set of Dataset A. The range in parentheses is the 95% confidence interval generated by bootstrapping. *DeepCoro predictions were binarized with a threshold of 0.23, as determined on the validation set. **Abbreviations**. AUPRC: Area Under the Precision-Recall Curve, AUROC: Area Under the Receiver Operating Curve, LCA: Left Coronary Artery, MAE: Mean Absolute Error, PPV: Positive Predictive Value, *r*: Pearson's correlation coefficient, RCA: Right Coronary Artery.

Table 2. Video-Level Performance of Coronary Angiography Clinical Reports and DeepCoro against Expert Annotated Videos Present in Dataset B

<table>
<tr><th rowspan="3">Task</th><th rowspan="3">Metric</th><th colspan="6">Coronary artery</th></tr>
<tr><th colspan="2">LCA</th><th colspan="2">RCA</th><th colspan="2">RCA + LCA</th></tr>
<tr><th>Clinical reports</th><th>DeepCoro</th><th>Clinical reports</th><th>DeepCoro</th><th>Clinical reports</th><th>DeepCoro</th></tr>
<tr><td colspan="2">Number of videos</td><td colspan="2">475</td><td colspan="2">490</td><td colspan="2">965</td></tr>
<tr><td colspan="2">Number of severe stenoses, ⩾ 70% (n (%))</td><td colspan="2">105 (22%)</td><td colspan="2">92 (19%)</td><td colspan="2">197 (20%)</td></tr>
<tr><td colspan="2">Number of non-severe stenoses, <70% (n (%))</td><td colspan="2">370 (78%)</td><td colspan="2">398 (81%)</td><td colspan="2">768 (80%)</td></tr>
<tr><td colspan="2">Number of healthy vessels, 0% stenoses (n (%))</td><td colspan="2">335 (71%)</td><td colspan="2">322 (66%)</td><td colspan="2">657 (68%)</td></tr>
<tr><td rowspan="4">Classification</td><td>AUROC</td><td>0.7430 (0.7136 - 0.7735)</td><td>0.8781 (0.8608 - 0.8965)</td><td>0.7649 (0.7359 - 0.7959)</td><td>0.8599 (0.8386 - 0.8867)</td><td>0.7533 (0.7328 - 0.7744)</td><td>0.8699 (0.8564 - 0.8853)</td></tr>
<tr><td>AUPRC</td><td>0.4908 (0.4400 - 0.5369)</td><td>0.7066 (0.6659 - 0.7494)</td><td>0.4684 (0.4168 - 0.5213)</td><td>0.7040 (0.6592 - 0.7483)</td><td>0.4737 (0.4382 - 0.5086)</td><td>0.7042 (0.6717 - 0.7339)</td></tr>
<tr><td>Sensitivity (%)</td><td>46.63 (41.57 - 51.76)</td><td>79.91 (76.54 - 83.95)</td><td>47.88 (42.31 - 52.78)</td><td>77.20 (72.86 - 81.94)</td><td>47.30 (43.95 - 50.64)</td><td>78.74 (76.05 - 81.70)</td></tr>
<tr><td>Specificity (%)</td><td>91.87 (90.57 - 93.27)</td><td>77.05 (75.00 - 79.12)</td><td>93.73 (92.63 - 94.94)</td><td>77.40 (75.39 - 79.50)</td><td>92.81 (91.95 - 93.77)</td><td>77.20 (75.78 - 78.60)</td></tr>
<tr><td rowspan="2">Regression</td><td>MAE (%)</td><td>22.28 (21.18 - 23.39)</td><td>19.76 (19.02 - 20.53)</td><td>19.79 (18.74 - 20.84)</td><td>18.45 (17.75 - 19.15)</td><td>21.00 (20.20 - 21.76)</td><td>19.09 (18.55 - 19.58)</td></tr>
<tr><td>r</td><td>0.4661 (0.4221 - 0.5084)</td><td>0.6653 (0.6406 - 0.6916)</td><td>0.5332 (0.4929 - 0.5757)</td><td>0.6914 (0.6624 - 0.7232)</td><td>0.5000 (0.4702 - 0.5302)</td><td>0.6792 (0.6598 - 0.7004)</td></tr>
</table>

Legend. Video-level performance of DeepCoro and clinical reports on Dataset B. The statistically significant metrics where the confidence intervals don't overlap are shown in bold. DeepCoro predictions were binarized with a threshold of 0.23, as determined on the validation set, and clinical used a threshold of 0.7 for sever stenosis classification. The range in parentheses is the 95% confidence interval generated by bootstrapping. **Abbreviations**. AUPRC: Area Under the Precision-Recall Curve, AUROC: Area Under the Receiver Operating Curve, MAE: Mean Absolute Error, *r*: Pearson's correlation coefficient, RCA: Right Coronary Artery, LCA: Left Coronary Artery.

REFERENCES

1 Langlais-Labrecque, É. *et al.* Novel artificial intelligence applications in cardiology: current landscape, limitations, and the road to real-world applications. *Journal of Cardiovascular Translational Research*, 1-13 (2022).

2 Jungiewicz, M. *et al.* Vision Transformer in stenosis detection of coronary arteries. *Expert Systems with Applications* **228**, 120234 (2023).

3 Zhou, C. *et al.* Automated deep learning analysis of angiography video sequences for coronary artery disease. *arXiv preprint arXiv:2101.12505* (2021).

4 Grech, E. Pathophysiology and investigation of coronary artery disease. *ABC of Interventional Cardiology. 2nd edn. BMJ Books, Oxford* (2011).

5 Singh, A. & Prakash, N. A REVIEW OF AI MODELS FOR PREDICTION AND DETECTING HEART DISEASES FOR IMPROVED WELLBEING.

6 Avram, R. *et al.* CathAI: fully automated coronary angiography interpretation and stenosis estimation. *NPJ Digital Medicine* **6**, 142 (2023).

7 Scanlon, P. J. *et al.* ACC/AHA guidelines for coronary angiography: executive summary and recommendations: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Coronary Angiography) developed in collaboration with the Society for Cardiac Angiography and Interventions. *Circulation* **99**, 2345-2357 (1999).

8 Algarni, M., Al-Rezqi, A., Saeed, F., Alsaeedi, A. & Ghabban, F. Multi-constraints based deep learning model for automated segmentation and diagnosis of coronary artery disease in X-ray angiographic images. *PeerJ Computer Science* **8**, e993 (2022).

9 Çimen, S., Gooya, A., Grass, M. & Frangi, A. F. Reconstruction of coronary arteries from X-ray angiography: A review. *Med Image Anal* **32**, 46-68 (2016).

10 Leape, L. L. *et al.* Effect of variability in the interpretation of coronary angiograms on the appropriateness of use of coronary revascularization procedures. *American Heart Journal* **139**, 106-113 (2000).

11 Alizadehsani, R. *et al.* Coronary artery disease detection using artificial intelligence techniques: A survey of trends, geographical differences and diagnostic features 1991–2020. *Computers in Biology and Medicine* **128**, 104095 (2021).

12 Nallamothu, B. K. *et al.* Comparison of clinical interpretation with visual assessment and quantitative coronary angiography in patients undergoing percutaneous coronary intervention in contemporary practice: the Assessing Angiography (A2) project. *Circulation* **127**, 1793-1800 (2013).

13 Patel, M. R. *et al.* ACC/AATS/AHA/ASE/ASNC/SCAI/SCCT/STS 2017 appropriate use criteria for coronary revascularization in patients with stable ischemic heart disease: a report of the American College of Cardiology appropriate use criteria task force, American Association for Thoracic Surgery, American Heart Association, American Society of Echocardiography, American Society of Nuclear Cardiology, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and Society of Thoracic Surgeons. *J Am Coll Cardiol* **69**, 2212-2241 (2017).

14 GN, L. 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention: Executive summary: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. *Circulation* **124**, 2574-2609 (2011).

15 Au, B. *et al.* Automated characterization of stenosis in invasive coronary angiography images with convolutional neural networks. *arXiv preprint arXiv:1807.10597* (2018).

16 Anderson, R. D. & Pepine, C. J. Vol. 127 1760-1762 (Am Heart Assoc, 2013).

17 Avram, R. *et al.* CathAI: Fully Automated Interpretation of Coronary Angiograms Using Neural Networks. (2021).

18 Zhao, C. *et al.* Automatic extraction and stenosis evaluation of coronary arteries in invasive coronary angiograms. *Computers in biology and medicine* **136**, 104667 (2021).

19 Moon, J. H. *et al.* Automatic stenosis recognition from coronary angiography using convolutional neural networks. *Computer methods and programs in biomedicine* **198**, 105819 (2021).

20 Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. in *Proceedings of the IEEE international conference on computer vision.* 2980-2988.

21 Liu, Z. *et al.* in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 3202-3211.

22 Chollet, F. in *Proceedings of the IEEE conference on computer vision and pattern recognition.* 1251-1258.

23 OpenCV. *cv::Tracker Class Reference*, <https://docs.opencv.org/3.4/d0/d0a/classcv_1_1Tracker.html> (

24 Serruys, P. W. *et al.* Assessment of the SYNTAX score in the Syntax study. *EuroIntervention* **5**, 50-56 (2009).

25 Maxim Popov, A. A., Nuren Zhaksylyk, Alsabir Alkanov, Adilbek Saniyazbekov, Temirgali Aimyshev, Eldar Ismailov, Ablay Bulegenov, Alexey Kolesnikov, Aizhan Kulanbayeva, Arystan Kuzhukeyev, Orazbek Sakhov, Almat Kalzhanov, Nurzhan Temenov, & Siamac Fazli1. (ed Zenodo) (2023).

26 Tardif, J.-C. *et al.* Effects of reconstituted high-density lipoprotein infusions on coronary atherosclerosis: a randomized controlled trial. *Jama* **297**, 1675-1682 (2007).

27 Garrone, P. *et al.* Quantitative coronary angiography in the current era: principles and applications. *Journal of interventional cardiology* **22**, 527-536 (2009).

28 Shah, R. *et al.* Comparison of visual assessment of coronary stenosis with independent quantitative coronary angiography: Findings from the Prospective Multicenter Imaging Study for Evaluation of Chest Pain (PROMISE) trial. *American heart journal* **184**, 1-9 (2017).

29 Avram, R., Olgin, J. E. & Tison, G. H. The rise of open-sourced machine learning in small and imbalanced datasets: Predicting in-stent restenosis. *Canadian Journal of Cardiology* **36**, 1574-1576 (2020).

30 DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837-845 (1988).

31 Zhang, H. *et al.* Comparison of physician visual assessment with quantitative coronary angiography in assessment of stenosis severity in China. *JAMA internal medicine* **178**, 239-247 (2018).

32 Ouyang, D. *et al.* Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252-256 (2020).

33 Jun, T. J. *et al.* Automated detection of vulnerable plaque in intravascular ultrasound images. *Medical & Biological Engineering & Computing* **57**, 863-876 (2019).

34 Tu, S. *et al.* Diagnostic accuracy of quantitative flow ratio for assessment of coronary stenosis significance from a single angiographic view: A novel method based on bifurcation fractal law. *Catheterization and Cardiovascular Interventions* **97**, 1040-1047 (2021).

35 Piccirillo, F. *et al.* Changes of the coronary arteries and cardiac microvasculature with aging: Implications for translational research and clinical practice. *Mechanisms of ageing and development* **184**, 111161 (2019).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.pdf