Πανεπιστήμιο Θεσσαλίας
Σχολή Θετικών Επιστημών
Τμήμα Πληροφορικής με Εφαρμογές στη Βιοιατρική

Κατηγοριοποίηση εικόνων και βίντεο αγγειογραφίας

Χαρίκλεια Γκραντούνη

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων
Δελήμπασης Κωνσταντίνος

Λαμία 2025

University of Thessaly
School of Science
Department of Computer Science and Biomedical Informatics

Automated image and video analysis for Coronary Angiography

Charikleia Gkrantouni

THESIS PROJECT

Supervisor
Delibasis Konstantinos

Lamia 2025

Contents

Ευχαριστίες

Θα ήθελα να εκφράσω την ειλικρινή μου ευγνωμοσύνη στον επιβλέποντά μου, Καθηγητή Κ. Δελημπάση, για την ανεκτίμητη καθοδήγηση, τις εποικοδομητικές παρατηρήσεις και τη συνεχή υποστήριξή του καθ' όλη τη διάρκεια της παρούσας έρευνας. Η επιστημονική του κατάρτιση και η ενθάρρυνσή του υπήρξαν καθοριστικές για την ολοκλήρωση αυτής της διατριβής.

Θα ήθελα επίσης να ευχαριστήσω τον υποψήφιο διδάκτορα Γ. Νούσια για την πολύτιμη συμβολή του στην πρόοδο της παρούσας έρευνας, του οποίου οι διορατικές συζητήσεις και η βοήθειά του υπήρξαν πηγή έμπνευσης και ουσιαστικής υποστήριξης. Επιπλέον, εκφράζω την ευγνωμοσύνη μου στον υποψήφιο διδάκτορα Σ. Κάμνη για τις εποικοδομητικές του συζητήσεις, ιδιαίτερα σε σχέση με τις τεχνικές πτυχές του έργου.

Τέλος, θέλω να εκφράσω τις θερμές ευχαριστίες μου στον κ. Γεώργιο Παπανάγνου, Συντονιστή Διευθυντή της Μονάδας Στεφανιαίων Νόσων και Αιμοδυναμικού Εργαστηρίου του Γενικού Νοσοκομείου Λαμίας, ο οποίος, στο πλαίσιο της Μεταδιδακτορικής του ερευνητικής συνεργασίας με το τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική, παρείχε όλα τα κλινικά δεδομένα που χρησιμοποιήσαμε, αλλά και για την καθοδήγηση και συμβουλές του σχετικά με την ερμηνεία τους. Χωρίς την συνεισφορά του, η εργασία αυτή δε θα ήταν δυνατό να επιτελεστεί.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Professor K. Delibasis, for his invaluable guidance, constructive feedback, and continuous support throughout the course of this research. His expertise and encouragement have been instrumental in the completion of this thesis.

I also wish to thank PhD candidate G. Nousias for his valuable contribution to the progress of this research, whose insightful discussions and assistance have been both inspiring and helpful. In addition I am grateful to PhD candidate S.Kamnis for his constructive discussions, especially in the technical aspects of the project.

Finally, I would like to express my thanks to Dr George Papanagnou, Director of the Coronary Disease Unit and the Hemodynamic Laboratory of the General Hospital of Lamia, who, in the framework of a Post-doctoral cooperation between the Hospital and the Department of Computer Science and Biomedical Informatics, provided us with all the clinical data that we utilized in this work, as well as their invaluable interpretation. This work would not have been possible without his contribution.

Abstract

Coronary artery disease (CAD) remains the leading cause of mortality worldwide, with invasive coronary angiography (ICA) being the standard method for diagnosis and intervention. However, manual interpretation of angiographic videos is time-consuming, prone to observer variability, and limited in real-time applications. In this thesis, we present a fully automated deep learning pipeline for ICA video analysis, integrating frame selection, anatomical classification, procedural detection, and stenosis identification.

The dataset consisted of 11,929 anonymized ICA videos from 767 patients. Pre-processing included anonymization of DICOM files, and binary coding of diagnosis that where given by the General Hospital of Lamia. The pipeline of the research includes automatic frame selection using vesselness filtering, a Frame Score metric and structural similarity index (SSIM). Anatomical classification between left (LCA) and right (RCA) coronary arteries was performed using a fine-tuned AlexNet model, achieving excellent performance (AUC = 0.991). Procedural frame detection, specifically balloon inflation was addressed using a Vision Transformer (ViT-Base) model, giving very strong classification results. Finally, stenosis detection and localization were implemented with AlexNet, achieving AUCs of 0.963–0.984 across proximal, middle, and distal RCA segments, with similarly high accuracy for LCMA, LAD, and LCx segments. During testing, predictions were made frame-by-frame across each video, and a majority voting scheme was applied to assign the final video-level label.

Compared to previous approaches, which typically relied on static CCTA images or isolated ICA frames and addressed single tasks, our work introduces the first end-to-end automated pipeline for full ICA videos, encompassing both anatomical and clinical event recognition.

This study presents an end-to-end automated pipeline for ICA video analysis, integrating anatomical classification, procedural event detection, and stenosis identification with automatic frame selection. The system achieved state-of-the-art results on a large multi-patient dataset and demonstrates significant potential to assist clinicians in diagnosis, interventional guidance, and treatment planning.

1 Introduction

Coronary angiography is an invasive imaging procedure used to provide visualization information about the coronary arteries and support the diagnosis and therapeutic intervention of coronary artery disease. The procedure produces real time X-ray video sequences, in which the vessels are highlighted by contrast medium (iodine-based radio-opaque dye), allowing the assessment of their anatomy and blood flow. Although it remains a very important procedure, the interpretation of angiography is challenging due to factors like contrast flow, overlapping vessels, image quality, real time performance and many pathological features. Until recently, the manual selection of frames and the classification by experts has been time consuming and error-prone. These challenges motivate the development of automated computational approaches that can assist cardiologists by improving efficiency, reproducibility, and diagnostic accuracy.

1.1 Anatomy of the Heart and Imaging of Coronary Vessels

The standard anatomy of the coronary artery, Figure 1 includes the left main coronary artery (LMCA), which bifurcates into the left anterior descending artery (LAD) and the left cirmumflex artery (LCx). The LAD descends consistently along the anterior interventricular groove, while the LCx follows the atrioventricular groove that supplies the lateral and posterior walls of the left ventricle. The LAD and LCx part of the LCA divides into 3 segments:

- The proximal segment, which extends from its origin at the bifurcation to the first major septal branch;
- The mid segment which includes the first and second septal;
- The distal segment which continues beyond the second septal branch toward the apex.

The right coronary artery (RCA) exhibits variability in curvature and length transverse to the right atrioventricular sulcus. It is also divided into three segments: the proximal segment, from its origin at the right coronary sinus to the first acute marginal branch; the mid segment from the acute marginal branch to the crux of the heart; and the distal segment which goes beyond the crux to give rise to the posterior descending artery (PDA) .

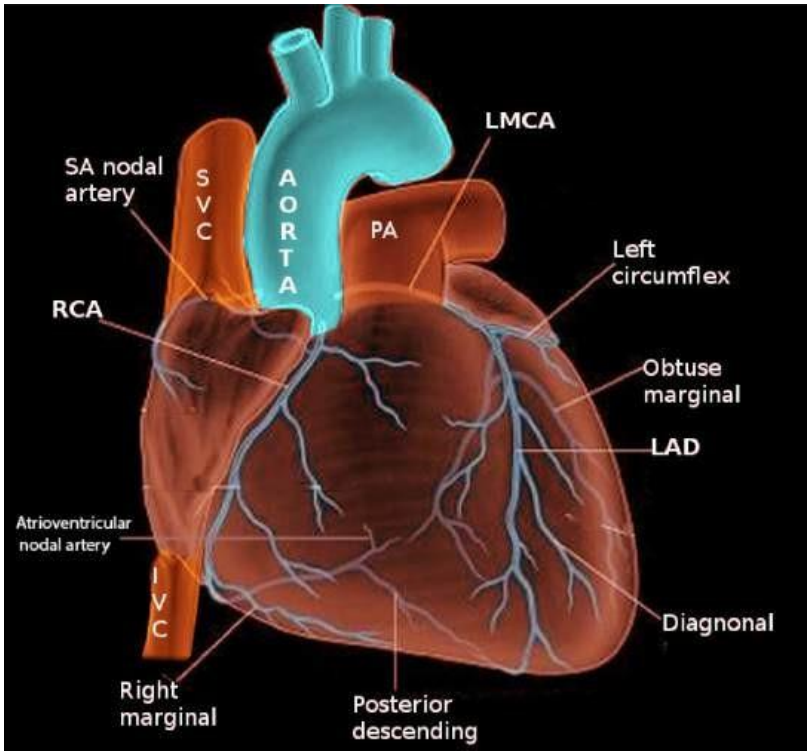


Figure 1: A representation of Heart anatomy [1]

The coronary angiography procedure is performed primarily to obtain detailed imaging of the heart's vascular anatomy, with a particular focus on the coronary arteries and associated structures. The main goal is to visualize the coronary vessels through the use of intravenous contrast medium injection, therefore facilitating an accurate assessment of cardiac morphology and function. Specific diagnostic aims include the identification of the main coronary arteries in relation to the aortic root, visualization of the left coronary main artery (LCMA), the left anterior descending (LAD), the left circumflex artery (LCx) and the right coronary artery (RCA), and the measurement of the aortic diameter. Through this procedure, clinicians can detect abnormalities in the heart, such as vessel stenosis, atheromatic plaques, or the status of previous interventions, such as stent placement.

[1]Image adapted from [1].

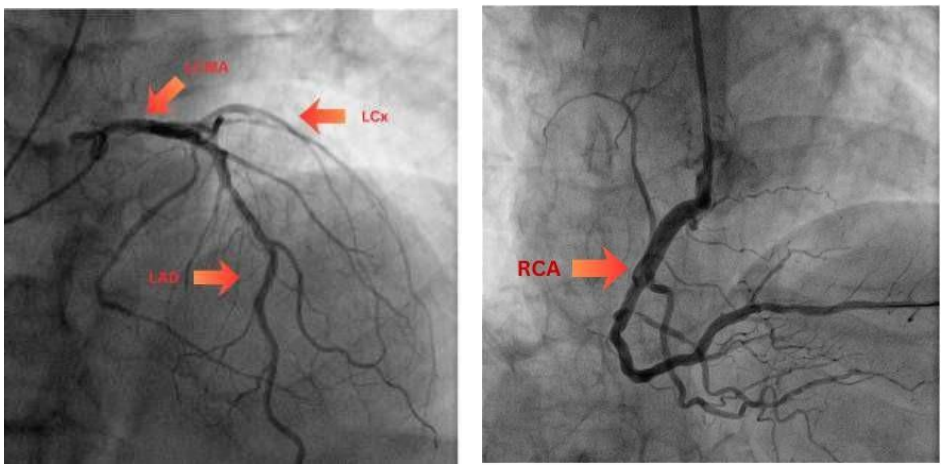


Figure 2: LCA and RCA example

The imaging system used to acquire the videos is often called ”C-arm”. The X-ray source is positioned below the patient, while the detector is placed above the thoracic cavity. The X-beam travels in an upward direction through the chest to the direction of the detector, producing an image based on different attenuation of the X-rays. A typical C-arm is shown in Figure 3.

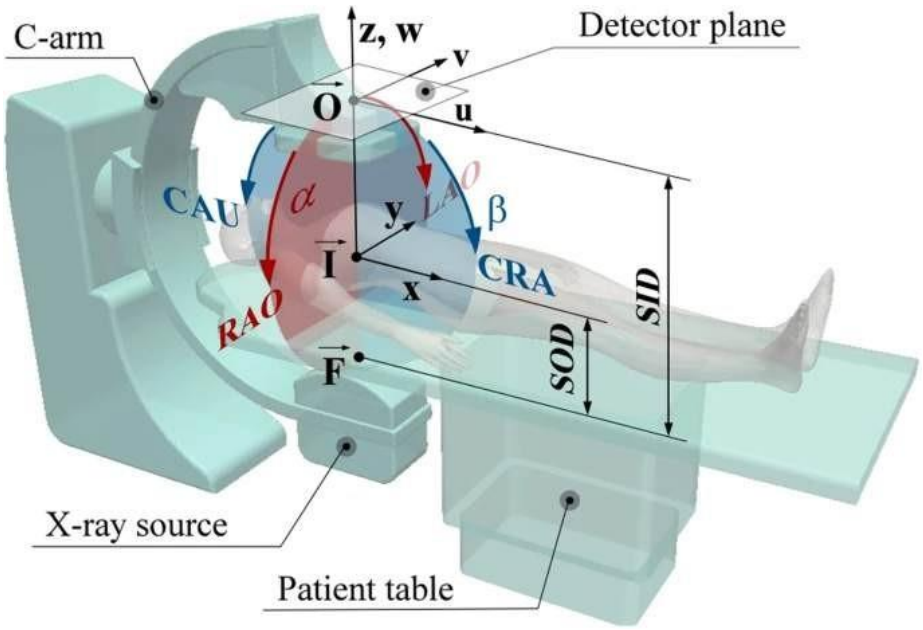


Figure 3: Three-dimensional reconstruction of coronary arteries [2]

[2]Image adapted from [2].

The image geometry includes antero-posterior view (AP), cranial view (CRA), caudal view (CAU), and the LCA, RCA, AP. In cranial view (CRA) the X-beam is angled upward from the feet to the head of the patient, its particularly useful for visualizing the LAD and its diagonal branches. In caudal views (CAU) the X-ray beam is angled downward from the head to the feet of the patient, visualizing the LCx and obtuse marginal branches. The antero-posterior (AP) refers to a straight frontal projection, without cranial or caudal tilt; it is used to assess the overall outline of the coronary tree. These projections enable an accurate assessment of coronary morphology and support clinical decision-making during diagnosis or procedures.

1.2 Related work

In [3] a real time radiation dose monitoring system was developed using DICOM data and leveraging the Modality Performed Procedure Step (MMPS) to automatically capture dose information during coronary angiography. This pioneering real-time DRL framework not only enhances dose tracking but also provides procedural feedback.

While in [4] an automatic method for selecting the most appropriate frame from coronary angiographic videos was proposed. The approach applies classical image processing filters based on the eigen-analysis of the Hessian matrix, such as Canny[5], Sato [6], Frangi [7], Mejering [8] and selects the frame with the highest contrast, achieving a 85.7% accuracy with the Sato filter. The authors provided quantified evaluation of their method based on frames selected by cardiology experts, but they do not identify specific placement of the vessels.

The authors of [9] present a Deep AngioKey model, which automates the extraction of keyframes from coronary angiography videos by identifying the frame where vessel structures are most visible. The approach is based on a U-net architecture and achieved an accuracy of 98.1% with an average frame distance of 1,63 from expert selected keyframes. However, the method evaluates frames independently and requires further validation on diverse datasets.

A systematic approach, often used by clinicians is the Syntax Score [10], an angiographic classification designed to quantify the complexity of coronary artery disease in patients with left main and three-vessel CAD. In [11] an approach is proposed that did not use deep learning networks and classification procedures, while utilizing the SYNTAX Score. For each lesion, the specialists classified its anatomical characteristics: Location (proximal, mid, distal, ostial), lesion length, whether it was bifurcation/trifurcation, tortuosity, calcification, thrombus, diffuse disease. All these parameters were then scored and summed giving the SYNTAX Score for the patient.

In [12] the paper introduces a feature-based model for coronary vessel segmentation, without labeling them. The proposed approach fuses classical filter-based features with deep neural network representations. These are fed into ensemble classifiers Gradient Boosting Decision Tree (GBDT) and Deep Forest, which achieved AUROC 95%, and F1 score 87%, with notably high specificity 99% on a test set of 130 images. The ground truth vessel masks were created manually by experts and were used as the reference standard for training and evaluating the segmentation models.

A number of approaches were focused on U-net architectures. More specifically, in [13] a DR-LCT-UNet was introduced in order to improve coronary segmentation in Coronary Computed Tomography Angiography (CCTA) scans. The model extended the 3D U-Net by incorporating Dense Residual and Local Contextual Transformer modules, achieving a high Dice Similarity Coefficient (DSC) of 85.8%. The ground truth annotations and the appropriate keyframe extraction were decided by medical specialists. However, the study reported limited comparisons with other methods and lacked detailed training settings.

Similarly in [14] the paper uses a SE-RegUNet, which integrates RegNet encoders with squeeze-and-excitation (SE) blocks to improve feature extraction *for coronary vessel segmentation* from X-ray angiography images. A two-step preprocessing pipeline was applied to improve image con- trast and vessel delineation. The ground truth was established with the assistance of cardiologists, who selected a key frame with optimal vessel opacification for each patient and generated consensus- based coronary masks. Using the open-access segmentation tool MedSeg, the vessels were manu- ally traced to produce the final ground truth masks. The best-performing variant—SE-RegUNet 4GF—achieved a Dice score of 0.72 and accuracy of 0.97 on internal test data. The external vali- dation on the DCA1 dataset slightly improved with a Dice score of 0.76, supporting the robustness of the method.

In [15] the paper uses a TVS-Net(Temporal Vessel Segmentation Network), a 2D+time neural network that combines a Dense 3D encoder with 2D decoder, to include temporal information from angiography frames. The network generates a binary segmentation of the coronary vessels, without any localization of the vessels. Results are presented using frames, manually annotated by specialists. The authors report a DCS of 83.4% on the internal test set, while external validation had slightly lower performance.

Stenosis classification is another important objective in coronary angiography. In [16] the authors evaluated eight state-of-the-art object detection networks for real-time coronary artery stenosis detection in invasive angiography, without any classification of the vessel placement in the heart. The Faster R-CNN Inception ResNet V2 achieved the highest accuracy (mAP 0.95, F1-score 0.96), while SSD MobileNetV2 delivered the fastest inference (38 fps, mAP 0.83, F1-score 0.80) and The R-FCN ResNet-101V2 model provided the best balance (mAP 0.94, F1-score 0.96). Keyframe ex- traction happened manually from specialists which labeled the frames with bounding boxes around stenotic regions, which served as the ground truth.

Another approach by [17] implements an integrated deep learning framework combining segmentation and classification on coronary angiography images. Specifically, they performed a binary classification for the existence of stenosis and tested multiple networks for vessel segmentation. They compared segmentation models (U-Net, UNet++, ResUNet-a), with U-Net achieving the highest accuracy DCS 84%. In terms of classification, DenseNet201 outperformed other pretrained models, reaching an accuracy of 90%. The angiography frames were selected by cardiology experts during routine diagnostic procedures then manually delineated the coronary vessels to produce segmentation masks.

The CathAI system was introduced in [18], as a multi-stage deep learning pipeline designed to fully automate the interpretation of coronary angiograms. The pipeline consists of four neural networks, each tailored for a specific task: projection angle classification, anatomical structure identification, object localization and stenosis severity estimation. This system identifies the right and Left coronary arteries (RCA and LCA respectively) and predicts the stenosis classified into two severities (above or below 70%). Frame selection is automated using a very simple method based on SSIM. While the modular approach and real-time capability seem to have great results, the low accuracy in bounding-box localization (48% Mean Average Precision) and variability across external datasets suggest further improvement.

The work of [19] presents an AI-based Quantitative Coronary Angiography (AI-QCA) system to quantify vessel stenosis. The algorithm identifies the minimum lumen diameter (MLD) location and computes stenosis metrics on the major coronary arteries. In this approach, the appropriate frame is manually selected and three different neural models are applied to generate the vessel segmentation. Stenosis is then quantified by measuring the vessel width perpendicularly to its

medial axis.

In [20] the authors proposed an end-to-end deep learning model, built on pre-existing Curved Multiplanar Reformats (cMPR) of coronary CCTA scans, where the arteries were already known to contain stenosis, to detect moderate (>50%) and severe (>70%) stenosis in the LAD, RCA, and LCX arteries, the ground truth was identified by specialists. They authors applied a transfer learning approach, including an EfficientNet, ResNet, DenseNet, and Inception-ResNet, with class-weighted strategies. Activation maps as heatmaps were included for understanding and the model achieved AUROC values of 0.95/0.94 (LAD), 0.92 (RCA), and 0.88 (LCX) with the use of EfficientNetB3/B0. The authors mentioned limited external validation. The following Table summarizes the details of the aforementioned works.

Study	Year	Modality	Detect Stenosis	Stenosis Localization	Vessel Segmentation	Vessel Labelling	Frame Selection	Other Processing	Dataset Size	Method
[20]	2025	CCTA	YES (>50, >70)	LAD, RCA, LCx	NO	NO	MANUAL	cMPRs	6,293 CT scans	cMPRs
[14]	2024	ICA	NO	NO	YES	NO	MANUAL	CLAHE	1000 videos	SE-RegUNet
[16]	2021	ICA	Binary	Continuous	NO	NO	MANUAL	Data augmentation	8,325 frames	Faster R-CNN, SSD, R-FCN
[17]	2023	ICA	Binary	NO	YES	NO	MANUAL	Augmentation, Syngo QCA	170 images	U-Net, DenseNet201
[12]	2022	ICA	NO	NO	YES	NO	MANUAL	Filter features, DL features	130 images	GBDT, Deep Forest
[19]	2024	CCTA	YES (severity)	YES (MLD location)	YES	LAD, LCx, RCA	MANUAL	Contrast normalization	7 658 images	AI-QCA
[4]	2023	ICA	NO	NO	NO	NO	AUTOMATIC	Sato, Frangi, Canny filters	37,209 images	Filter keyframe extraction
[13]	2023	CCTA	NO	NO	YES	NO	MANUAL	Intensity normalization	81 cases	DR-LCT-UNet
[18]	2023	ICA	YES (>70%)	Bounding boxes	NO	YES (LCA/RCA)	MANUAL	Frame preprocessing	13 843 studies	CathAI
[9]	2023	ICA	NO	NO	YES	NO	AUTOMATIC	Vessel visibility scoring	2 949 frames	Deep AngioKey
[15]	2025	ICA	NO	NO	YES	NO	MANUAL	Normalization, augmentation	323 sequences	TVS-Net
[3]	2015	ICA	NO	NO	NO	NO	N/A	Extracted DICOM dose data	4 571 procedures (CAG+PCI)	DRLs
[11]	2005	ICA	YES (>50%)	YES (manual vessel/branch)	NO	YES (all branches)	MANUAL	Pure manual scoring	1 800 patients	SYNTAX Score

Table 1: Comparison of methods for coronary imaging analysis across different studies.

1.3 Problem definition and aims of this thesis

The problem this thesis addresses is the automatic classification of coronary angiography videos. In particular, the work focuses on three main tasks: automatic selection of the most relevant frames using deep learning–based methods, classification of left and right coronary arteries through neural networks, and detection of interventional devices (balloons) and stenosis with advanced deep learning technique. Coronary angiography remains a highly challenging domain due to the complex nature of videos, the variability in vessel structures, and the clinical relevance of the features. These challenges make manual interpretation time-consuming, subjective and prone to variability among experts, thus the need for automated and reliable computational solutions is needed. While previous works have focused on some of these tasks, few have addressed the combined challenge of frame selection, artery classification, balloon detection and stenosis detection in a pipeline.

2 Methodology

2.1 Overall description of the proposed system

The proposed analysis pipeline initiates with the collection of patient angiographic videos, which are chronologically ordered according to their acquisition time. Each video is subsequently evaluated in sequence until no further recordings remain, at which point the workflow terminates with the final diagnostic outcome. For each available video, the system first determines whether the sequence corresponds to the right coronary artery (RCA) or the left coronary artery (LCA). Following artery classification, the algorithm verifies whether a balloon intervention has already been identified in prior sequences of the same artery. If a balloon has been previously detected, subsequent diagnostic frame selection for that artery is discarded, and the analysis proceeds to the next available video. Otherwise, in the absence of a previously detected balloon, the system performs a dedicated balloon detection step. If a balloon is identified in the current sequence, the remaining diagnostic steps are discarded, otherwise, the pipeline advances to the selection of diagnostic frames, which is performed using a combined strategy of vesselness filtering to enhance vascular structures and structural similarity index (SSIM) analysis to retain only the most informative frames.

All the selected frames from the current patient are then processed within the stenosis detection module, executed separately for RCA and LCA sequences, to identify the presence of significant vascular narrowing. This iterative process continues across all available patient videos until completion, ensuring a structured and systematic pathway from raw acquisition to final stenosis diagnosis.

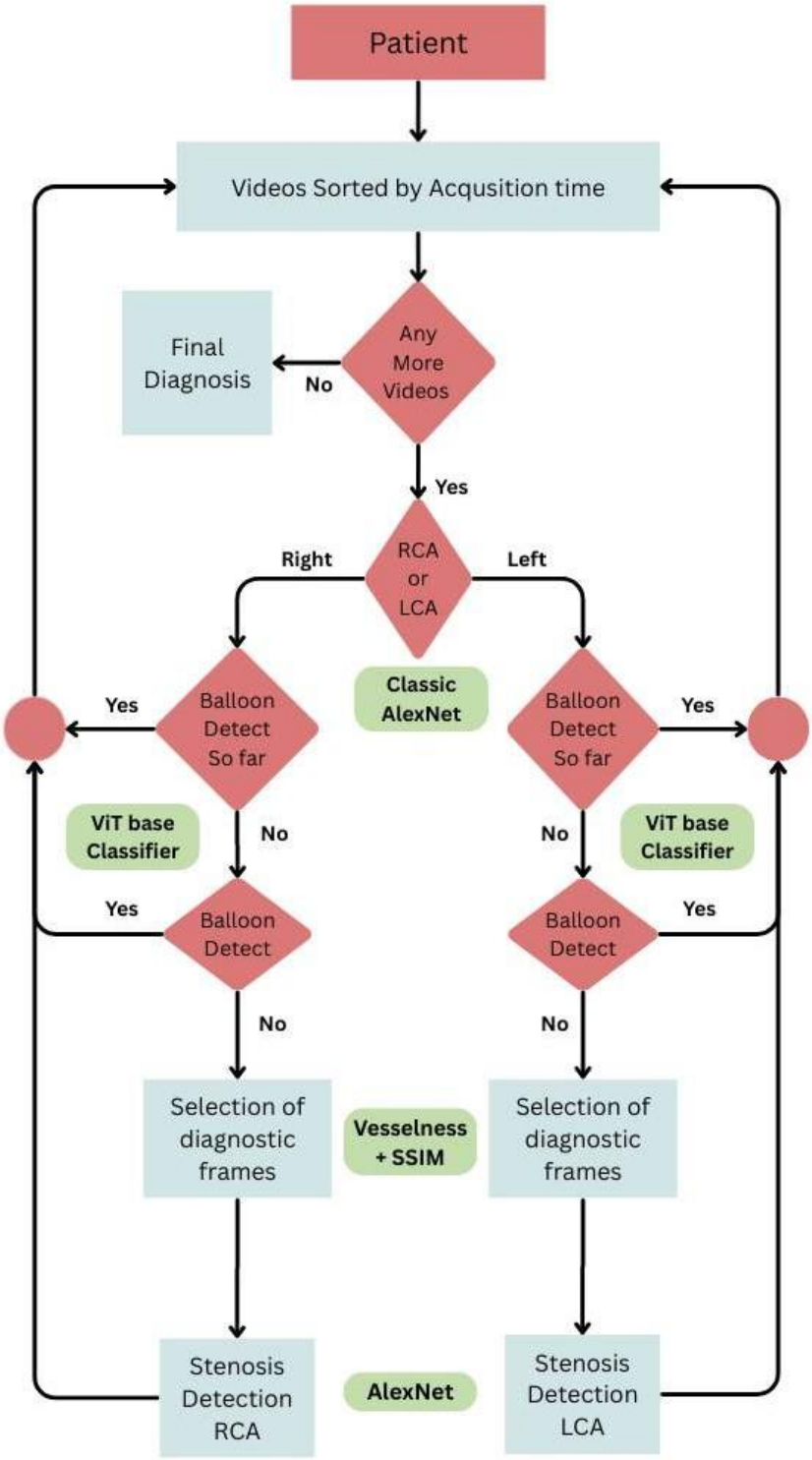


Figure 4: Overall description of the pipeline

2.2 Data handling

2.2.1 Anonymization procedure

It was essential to prevent any risk of data leakage during processing of medical imaging data, as medical data fall under a special category requiring strict safeguards. To ensure compliance with the General Data Protection Regulation (GDPR), an anonymization procedure was applied to the DICOM files (Digital Imaging and Communications in Medicine). This process involved exhaustive overwriting sensitive header fields by setting them equal to null values (e.g.,PatientName=”Null), effectively removing all personally identifiable information (PII). This includes the patient’s full name, date of birth, series number, and study ID number, as shown in Figure 5, in accordance with GDPR requirements.

To allow for traceability without compromising confidentiality, we retained two identifying fields: the internal patient ID and the year the imaging procedure took place. This allowed for the creation of a unique reference, while ensuring that no directly identifying information was preserved. This method aligns with the principle of data minimization under GDPR and follows the DICOM PS3.15 confidentiality guidelines for identification. The structural and diagnostic integrity of the imaging data was preserved throughout the process, ensuring the data set remained suitable for use in academic and clinical research without compromising confidentiality.

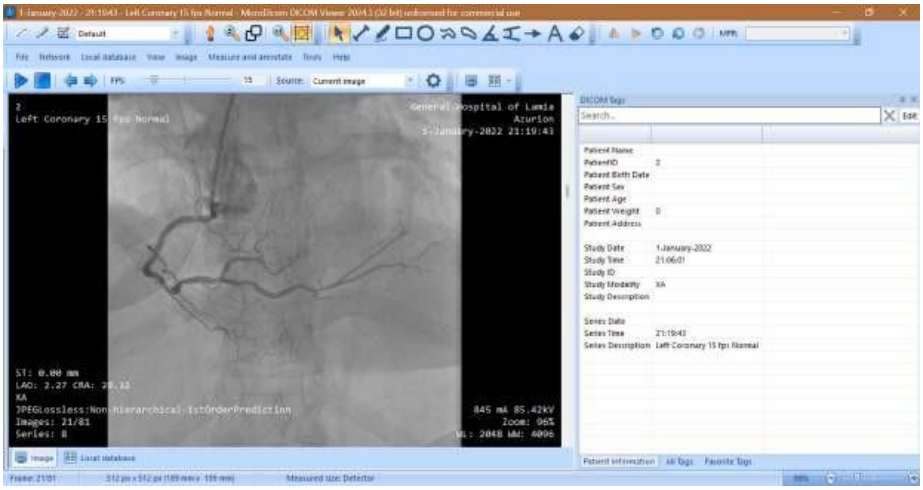


Figure 5: An anonymized DICOM image with all personal identifiers removed.

2.2.2 Coronary disease Diagnosis encoding as Multi-label Binary classification

In addition to DICOM file anonymization, a different process was followed to handle patient diagnosis data obtained from physical medical records, that were handed by the hospital. To ensure compliance with GDPR and prevent exposure of sensitive health information, all patient names were removed from the paper records. Each case was assigned a consistent internal patient code (matching the code used in anonymized DICOM files), provided by the hospital.

The full text of the diagnoses were then manually transcribed and entered in an Excel spreadsheet. This spreadsheet included binary-encoded columns for each coronary segment, ”Left Main

Coronary Artery (LMCA)", "Left Anterior Descending Artery(LAD)", "Left Circumflex Artery (LCx)", "Right coronary artery"). Each artery was subdivided into proximal, middle segment, and distal segment, with the value of 1 and 0 indicating the presence and absence of stenosis respec- tively. In total, 8 binary columns encode the presence and location of stenosis in the left coronary arteries and 3 binary columns encode the presence and location of stenosis in the right coronary artery. Figure 6 shows an example of the proposed multi-label encoding of disease classification.

Στέλεχος: Χωρίς στενώσεις. Πρόσθιος κατιών: Μη αιμοδυναμικά σημαντική αθηρωμάτωση μετά την έκφυση του 1ου διαγωνίου. Περισπωμένη: Χωρίς στενώσεις. Δεξιά στεφανιαία:
Στέλεχος: Χωρίς στενώσεις. Πρόσθιος κατιών: Ολική απόφραξη στην έκφυση. Παράπλευρη σκιαγράφηση της περιφέρειας από διαφραγματικούς κλάδους. Διάμεσος: Μεγάλος κλάδο
Στέλεχος: Χωρίς στενώσεις. Πρόσθιος κατιών: Χωρίς στενώσεις. Περισπωμένη: Χωρίς στενώσεις. Δεξιά στεφανιαία: Επικρατούσα, χωρίς στενώσεις.
Στέλεχος: Χωρίς στενώσεις. Πρόσθιος κατιών: Ολική απόφραξη στο 1ο τριτημόριο. Περισπωμένη: Χωρίς στενώσεις. Δεξιά στεφανιαία: Επικρατούσα. Διάχυτα αθηρωματική, χωρίς σ

	A	B	C	D	E	F	G	H	I	J	K	L	M
	YEAR	CODE	ΣΤΕΛΕΧΟΣ		ΠΡΟΣΘΙΟΣ ΚΑΤΙΩΝ			ΠΕΡΙΣΠΩΜΕΝΗ			ΔΕΞΙΑ ΣΤΕΦΑΝΙΑΙΑ		
			ΕΚΦΥΣΗ	ΜΕΣΟΤΗΤΑ	ΕΚΦΥΣΗ	ΜΕΣΟΤΗΤΑ	ΠΕΡΙΦΕΡΕΙΑ	ΕΚΦΥΣΗ	ΜΕΣΟΤΗΤΑ	ΠΕΡΙΦΕΡΕΙΑ	ΕΚΦΥΣΗ	ΜΕΣΟΤΗΤΑ	ΠΕΡΙΦΕΡΕΙΑ
	2021	001	0	0	1	0	0	0	0	0	0	1	1
	2021	010	0	0	1	0	0	0	0	0	0	1	0
	2021	100	0	0	0	0	0	0	0	0	0	0	0
	2021	101	0	0	0	1	0	0	0	0	0	0	0

Figure 6: Diagnostic and Binary coded samples of patients.

In addition, a free text column was included to capture the diagnosis from the original records, ensuring that all information relevant to each patient's condition is preserved in both structured and unstructured formats. Manual entries were reviewed to minimize errors and prevent any data leakage.

2.3 Fluoroscopic image acquisition

The geometry of video acquisition depends mainly on two angles that describe the orientation of the X-ray bean relative to the patient:

- Primary angle: DICOM tag (0018,1510)- represents the rotation of the C-arm in the transverse plane. The primary angle ranges from -90 (right side of the patient) to +90 (left side), as depicted in Figure 7 from the documentation of DICOM.
- Secondary angle: DICOM tag (0018, 1511)- represents cranial or caudal tilt. ranges from -90 degrees (caudal/feet direction) to +90 degrees (cranial/head direction), as shown in Figure 8.

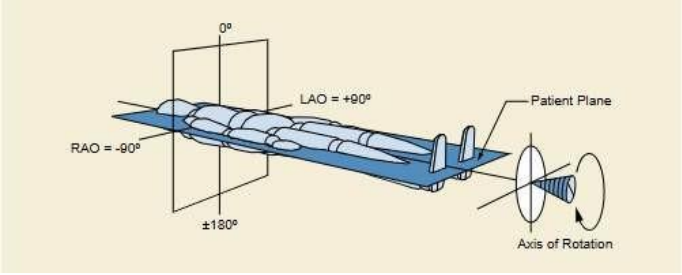


Figure 7: Positioner Primary Angle.[3]

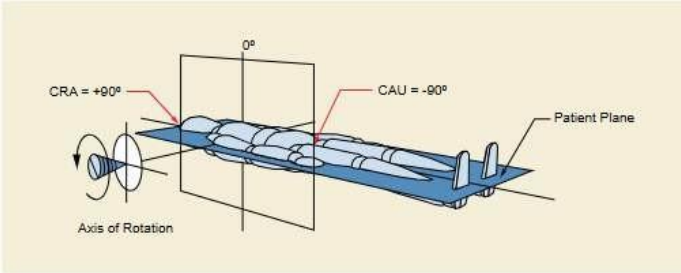


Figure 8: Positioner Secondary Angle. [4]

We compiled the angular data, stored in the DICOM metadata, we applied the K-means clus- tering algorithm (K = 9) to the data and grouped them into 9 angle-based clusters, The centers of the clusters are superimposed in Figure 9 . The angular team of each video of all patients were stored into a structured Excel file. These data are visualized in Figure 10.

9:

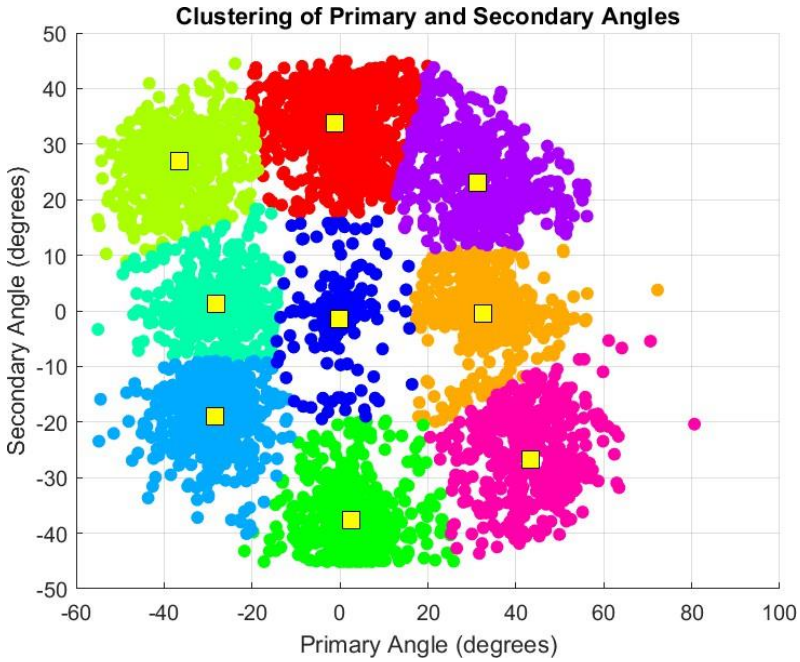


Figure 9: K-means clustering results. Each dot in the 2D angle acquisition space, represents a video

[4]Image adapted from [21].

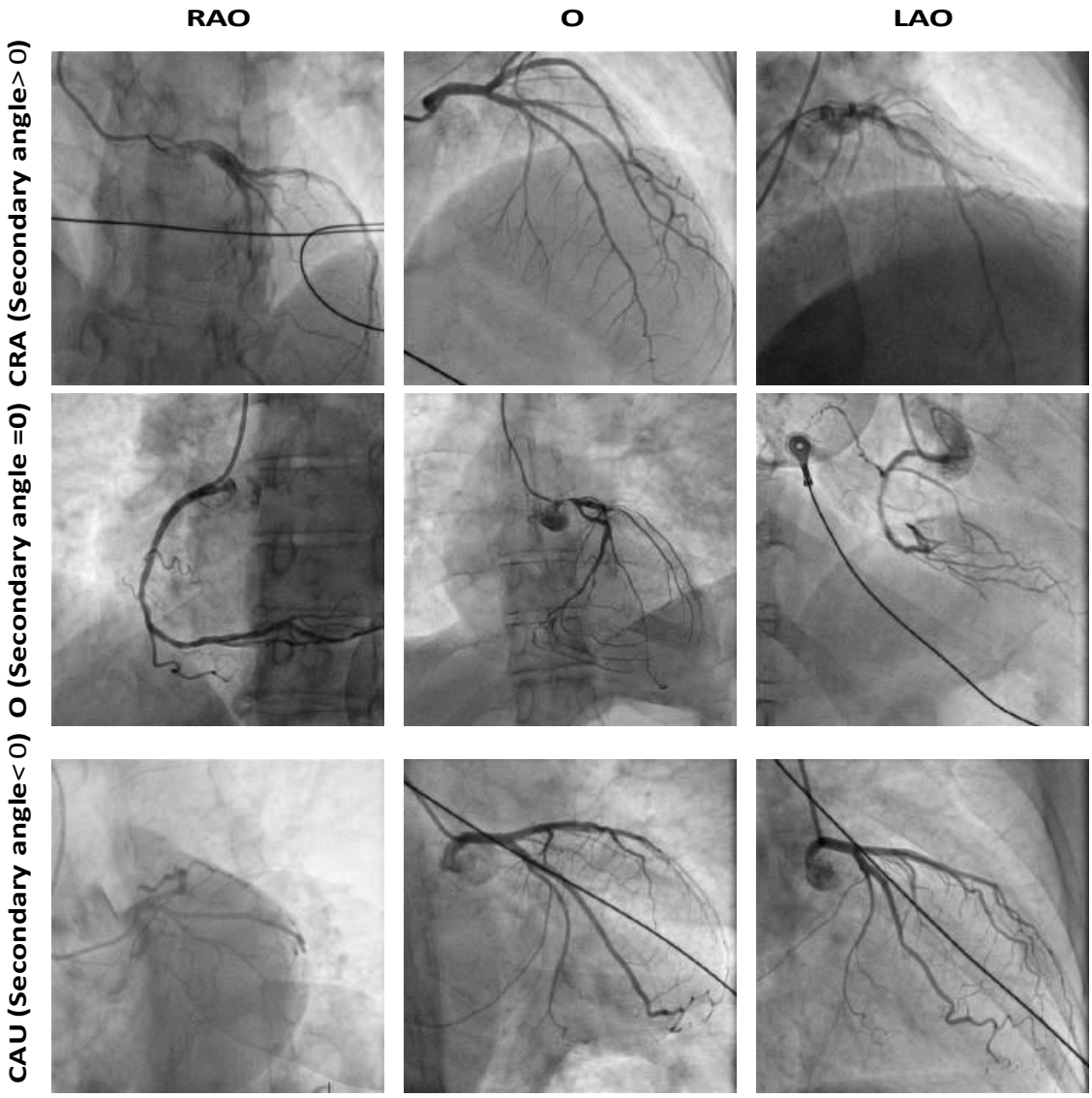


Figure 10: Exemplary images from each of the 9 combinations of the primary - secondary angle of Fig. 8

2.4 Left-Right Coronary Artery Frame and Video classification

The goal is to classify the entire video as LCA or RCA. To this end, we first train a deep learning model to classify *individual frames* and then utilize the trained deep learning model to inference the class of each frame of the video. The class of the *whole video* is determined by the majority of frame classes.

An early step of our proposed methodology involves the classification of the left coronary artery (LCA) and the right coronary artery (RCA) from coronary angiographic data. Anatomically, the human heart includes two primary coronary arteries: LCA and RCA, each of which branches into several sub-arterial structures 1. These arteries exhibit unique characteristics that can be important for diagnostic cardiology. Typically, the RCA and LCA are imaged from distinct, specific angles. Specifically,

- Primary angle α near -10° and below were predominantly associated with RCA
- Primary angle values near +10° and above were typically associated with LCA

However, this is not a rule strictly adhered to in clinical practice. Therefore, we did not resort into determining the LCA or RCA from the viewing angles alone.

Deep learning and more specifically convolutional neural networks (CNNs) have been used with remarkable success during the last 10 years or image classification tasks in almost any domain, including the medical images [22] and videos [23], [24]. To facilitate this frame classification, we adopted a scheme based on a Deep learning model. To train this model, we extracted specific frames from DICOM coronary angiography videos that clearly include the arterial structure. Specifically, we selected frames where the coronary arteries are visible, to ensure that we can observe the morphology in full detail, and we also selected frames, in which the tools used (e.g. guide wires) differ depending on whether the LCA or RCA is being examined, thus capturing context-specific features.

In total, we extracted 6,556 LCA images, and 2,372 RCA images, which were then labeled and split into two categories. Typical examples are shown in Figure 11 and 12 respectively. Although many approaches exist for balancing asymmetric data [25], we did not resort to any of them, since the performance of the system proved very high. In order to construct the ground truth, we utilized the primary angle, as following. We defined a threshold value that distinguished between left and right coronary arteries (above and below the threshold respectively). Subsequently, the resulting classification was reviewed by an expert to correct possible misclassifications.

For the deep learning model we selected AlexNet [26] as the base architecture due to its proven performance in image classification tasks and it had low computational cost compared to models like ResNet [27], VGG16 and VGG19 [28]. Our dataset was organized by unique patient code and year of examination to prevent data leakage, and was split at the patient level into 70% training, 15% validation and 15% testing. Thus, patient-wise splitting was ensured.

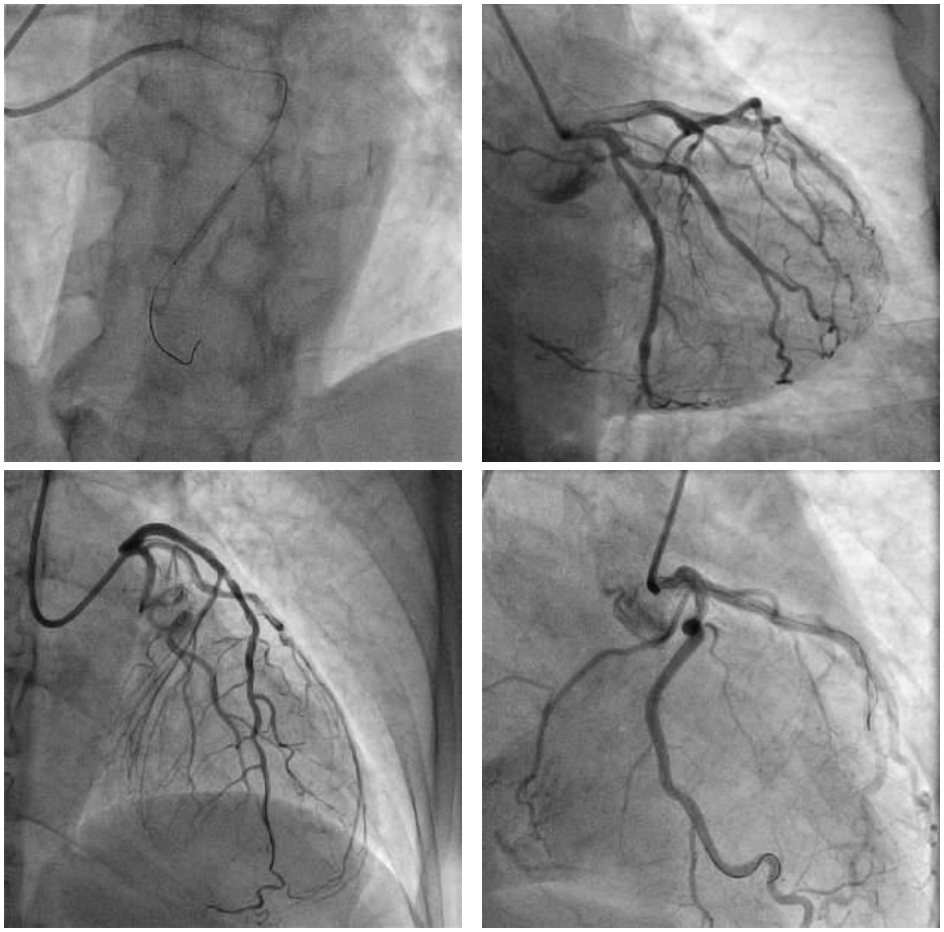


Figure 11: Examples of frames from the LCA dataset.

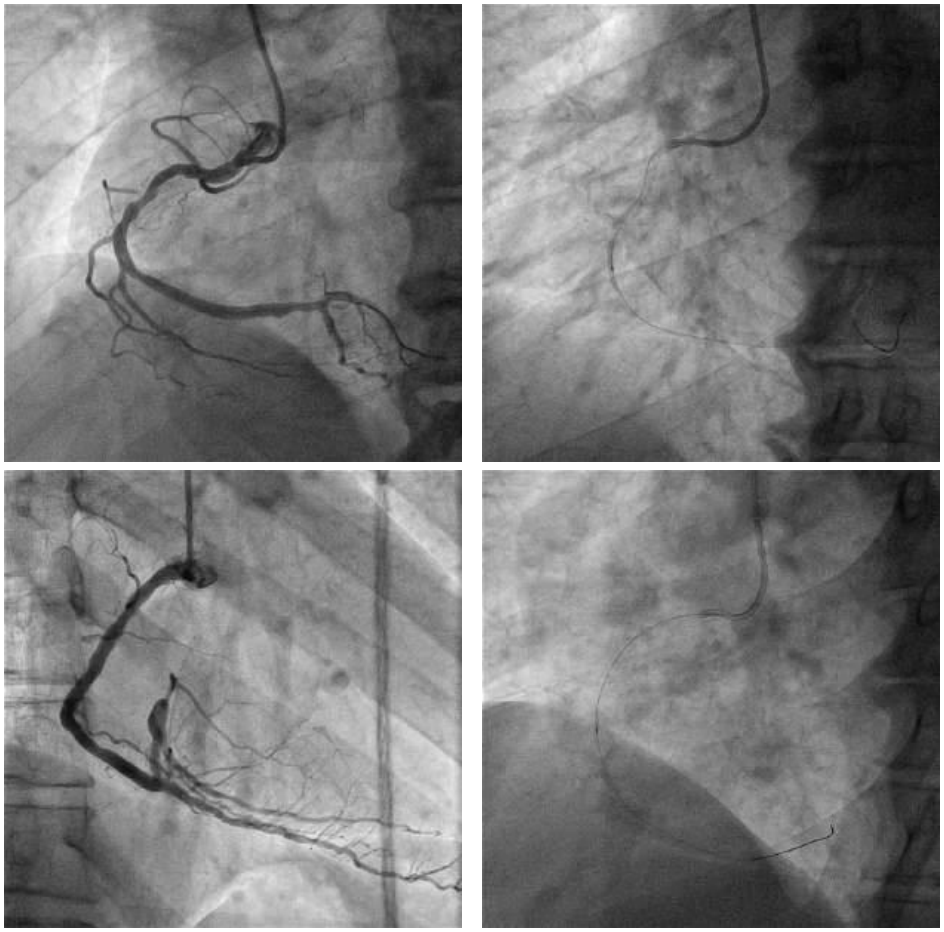


Figure 12: Examples of frames from the RCA dataset.

2.4.1 Pretrained AlexNet finetuned for LCA-RCA classification

Initially, we used the AlexNet convolutional neural network, pretrained on ImageNet [29] to classify coronary angiographic images into two categories Left Coronary Artery (LCA) and Right Coronary Artery (RCA). The pretrained Alexnet was further finetuned using our dataset. Images were resized to 227×227 to match the original AlexNet input, while grayscale frames were converted to three channels to meet the RGB format that was expected. Data augmentation included scaling (0.8–1.2), translation within 64 pixels range and small rotation not more than $\pm 10^{\circ}$, so variability could be enhanced.

The pretrained AlexNet was modified by replacing the final fully connected layer with a new layer that corresponded in the number of output we needed (2 outputs) while for the output layer a classification layer was used. Earlier convolutional layers were frozen and retained their pre-trained weights from ImageNet. The higher fully connected layers were fine-tuned with adjusted learning factors. For the training of the model stohastic gradient with momentum (SGDM), using a mini

batch size of 32, an initial learning rate of $1 \cdot 10^{-4}$ and a maximum number of 50 epochs. To prevent overfitting, early stopping was applied with a validation patience of 10, so, on average, training stopped after 18.7 epochs.

2.4.2 Customized AlexNet, finetuned for LCA-RCA classification

On a second attempt, we designed a custom hybrid network that retained the core structure of AlexNet, but integrated a custom sub-network. This addition was made to incorporate metadata, specifically the primary angle recorded during angiography, according to the DICOM documenta- tion. Each image was paired with its primary angle, allowing better representation of the location of the artery. Each image was resized at 227×227 and converted into three channels for RGB to fit the AlexNet input.

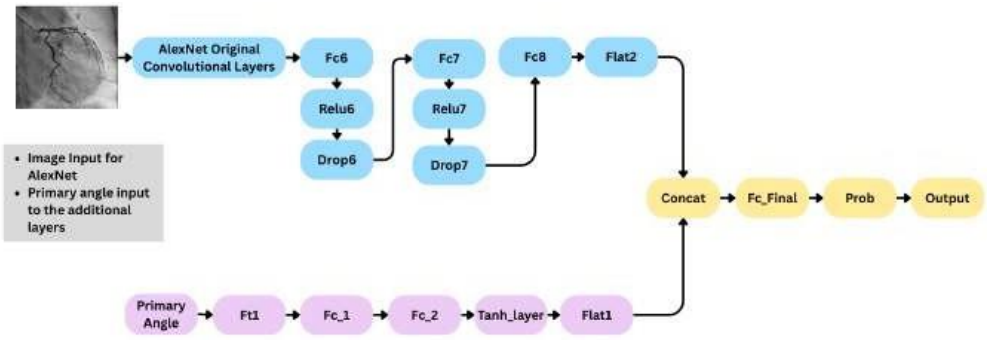


Figure 13: Customized AlexNet Architecture.

The angles were normalized and passed through a small MLP (FC(512) → FC(1) → tanh), flattened, and concatenated with the flattened AlexNet feature vector. The output of the 2 networks were inserted to a final fully connected layer with two outputs and a softmax classification layer (cross-entropy loss).

Convolutional layers of AlexNet were initialized from ImageNet and frozen (zero learn-rate factors), while higher fully connected layers were fine-tuned.Similar to the previous model SGDM was used for training with a mini batch size of 32, learning rate of $1 \cdot 10^{-4}$ and maximum epochs of 50. The training of the model stopped at 11 epochs since it met the restrictions of early stopping to prevent overfitting during training.

2.5 Balloon vs Non-Balloon Frame and Video classification

The goal is to classify the entire video as containing ballon or non-ballon. To this end, we first train a deep learning model to classify *individual frames*. Subsequently we utilize the trained deep learning model to inference the class of each frame of the video and propose an empirical rule to determine the class of the *whole video*.

In order to classify individual frames of video DICOM files as ”Balloon” or ”non-Balloon”, we extracted specific frames from DICOM coronary angiography videos, including frames with interventional tools such as balloon catheters clearly visible, and diagnostic frames where the coronary artery anatomy is distinctly represented. In total, we extracted manually 7,475 artery images, Fig- ure 14, and 7,451 balloon procedure images (Figure 15), which were then labeled and split into two categories. The ground truth was constructed by manual inspection of the frames.

To develop a robust deep learning-based classifier we evaluated three architectures. Firstly we employed AlexNet, due to its proven performance in image classification tasks. Secondly we implemented a Vision-Transformer (ViT-Base) model to leverage the self-attention mechanism and capture global dependencies across the image and lastly we tested a lightweight ViT-Tiny model to access the performance-efficiency trade-off in a lower complexity setting. Each approach was trained and evaluated under consistent data splitting and augmentation protocols to ensure fair comparison.

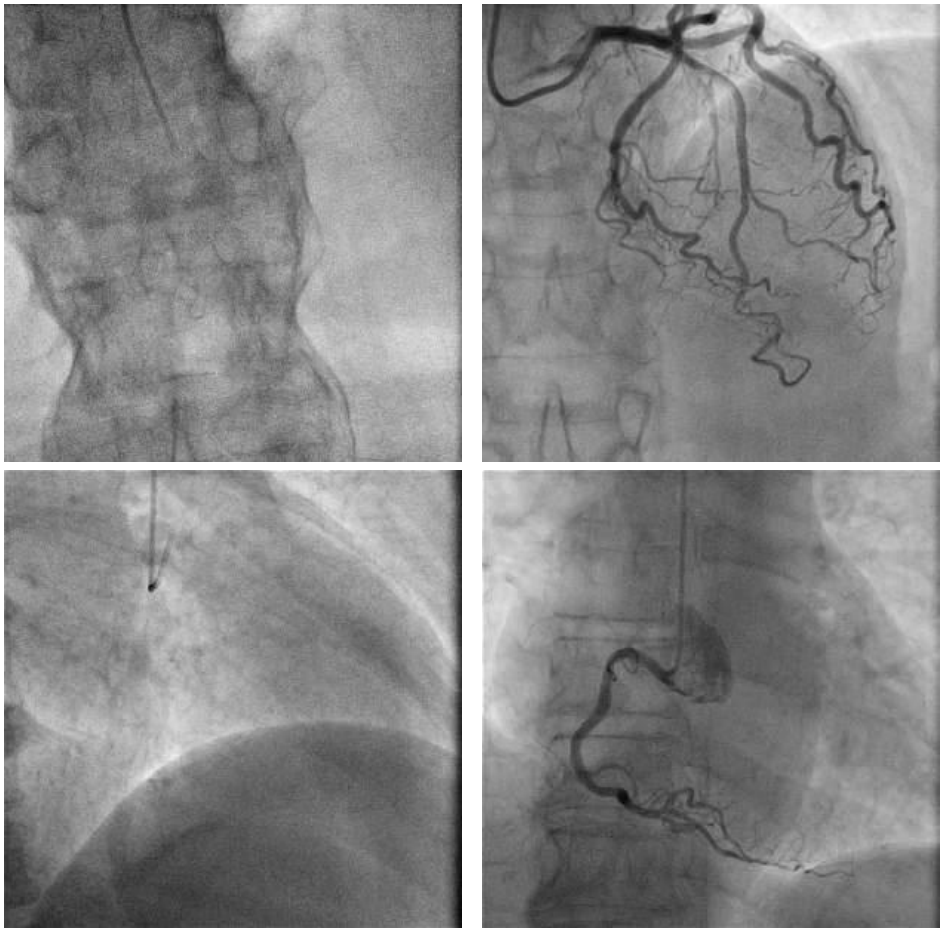


Figure 14: Example of Non-Balloon dataset

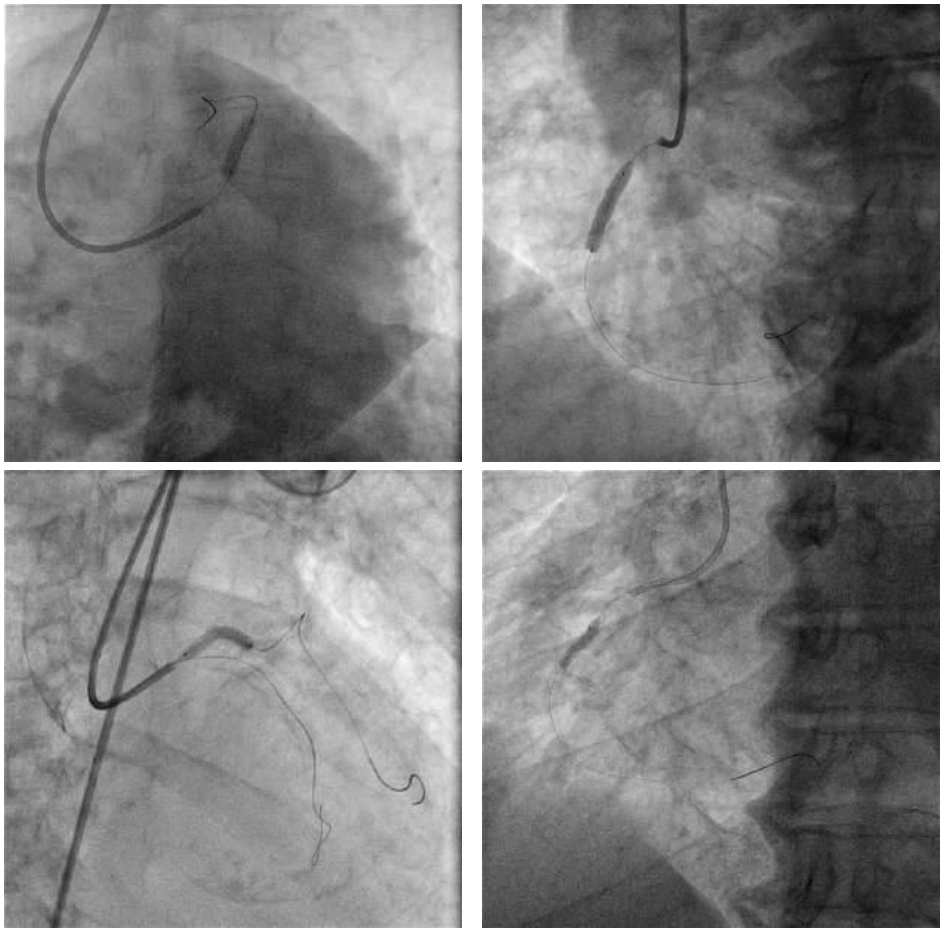


Figure 15: Example of Balloon dataset

2.5.1 Classic AlexNet approach

Firstly a baseline classification model was developed, using transfer learning with a pre-trained AlexNet architecture, fine-tuned on our dataset that contained images with categories Balloon or non-Balloon. To prevent data leakage, unique patient identifiers were again extracted (year of examination and patient ID) and used for patient-wise splitting into 70% training, 15% validation, and 15% test sets. Images were again resized to fit the AlexNet input size and each subset was processed with data augmentation that included random scaling (0.8-1.2), rotation (0–15°) horizontal/vertical translations of up to ± 64 pixels to improve model generalization.

During fine-tuning, the pretrained weights in the earlier convolutional layers were preserved and higher learning rates were assigned to the final layers to facilitate better task-specific adaptation. The network was trained using SGDM with a batch size of 32 and maximum of 50 epochs. To prevent overfitting early stopping was applied and finally the network was trained for 8 epochs.

2.5.2 Vit-Base Model Transformer

For a more advanced feature representation, a Vit-Base model [30] pre-trained on AlexNet was selected. The model was adapted for the binary classification task by freezing the core self-attention layers and replacing the final classification layers with a new fully connected layer of two output neurons. The same data pipeline was maintained as with the AlexNet model, including patient-level data split and augmentation. The model was trained using the Adam optimizer with a mini batch size of 8 and maximum epochs of 10. A warm-up learning rate schedule is applied, during which the learning rate is increasing. It is then followed by a step-wise decay (drop factor 0.1 every 3 epochs) was used. Early stopping was also applied here, so overall training was limited to 9 epochs. Performance was evaluated on the held-out test set using overall accuracy and confusion matrix analysis.

2.5.3 Vit-Tiny Model Transformer

In addition to the ViT-Base model, a Vision Transformer [31] of tiny configuration pre-trained on the ImageNet was investigated. This model processes images as a sequence of patches is capable of capturing long-range interactions with significantly fewer parameters than ViT-Base. Similar to the base variant, the self attention layers were frozen to preserve their general feature representations, and a new classification head with a fully connected layer specific to our binary classification task was added. The same splitting and augmentation were applied while early stopping was also included. The network was trained with a mini batch size of 8 and the maximum epochs were 10. Finally, the model met the validation criteria and stopped after 9 epochs. The resulting model was saved and evaluated using standard accuracy metrics and a confusion matrix to assess its predictive power.

2.5.4 Video classification as Ballon - Non Ballon

Let N be the total number of frames in the video sequence and n_b be the number of frames in the video sequence labeled as Balloon. The label of the whole video is decided based on the following empirical rule:

$$\text{Video Label} = \begin{cases} \text{Balloon}, & \text{if } (n_b > T_b \wedge \frac{n_b}{N} > T_R) \\ \text{Non ballon}, & \text{otherwise} \end{cases}, \quad (1)$$

where T_b and T_R are the thresholds for the number and ratio of ”balloon”-labeled frames, respectively.

This empirical rule ensures that the final label for a sequence is not based on single-frame predictions alone, but on the consistency and proportion of balloon detections across the video. Specifically, if no balloon frames are detected, the sequence is automatically classified as Artery. Otherwise, the ratio of detected balloon frames to the total number of frames is computed. A sequence is classified as Balloon only if the values n_b and $\frac{n_b}{N}$ are above the threshold values T_b and T_R, if these conditions are not met, the sequence defaults to Non-Balloon. After multiple trials on the Train dataset and Validation Dataset the threshold values were finalized i T_b=1 and T_R=0.8, this combination ensured the higher accuracy of the model and reduced the misclassified videos.

2.6 Frame Extraction

The purpose of this module is to extract few frames from each DICOM videos, which are appropriate for coronary stenosis detection. To this end the general approach was to apply classical image processing algorithms to each of the video frames and subsequently select the most appropriate frame based on some metric. The final solution integrates image enhancement with a selection metric and similarity assessment to identify the most informative frames. Multiple image processing algorithms were explored and combined in a hybrid method that uses vessel enhancement techniques and structural similarity measures.

2.6.1 Bottom-Hat transformation approach

The bottom-hat morphological transform was applied to each frame using a line-shaped structuring element, selected to align with the anatomical structure of coronary arteries. The bottom-hat method is a morphological image processing technique designed to enhance dark objects on a bright background. It is effective for highlighting features like cracks, veins or other features that are smaller than the structuring element. Mathematically, it is defined as the difference between the morphological closing of an image and the original image:

$$BH_{theta} = BottomHat_{\vartheta}(I) = Closing(I) - I = I \cdot s_{\vartheta} = (I \oplus s_{\vartheta}) \ominus s_{\vartheta}, \quad (2)$$

where I is the input image in grayscale, $\cdot$ is the image closing operator, $\oplus$, $\ominus$ the dilation and erosion morphological operators and s_{ϑ} is the structuring element used in the morphological operations. Since we are looking for linear structures in any orientation, we defined the structuring element as a binary linear segment of length 30 pixels with variable orientation $\vartheta = 0, 10, \cdots, 170^0$. Let us define the bottom hat result for each angle as $BottomHat_{\vartheta}$. The pixel-wise maximum along all different BH_{ϑ} was obtained.

$$I_{BH}(i,j) = \max_{\vartheta}(BH_{\vartheta}(i,j), \vartheta = 0, 10, \cdots, 170^0) \quad (3)$$

The closing operation suppresses small dark regions and smooths contours while preserving overall image brightness. After subtracting the original image from the closed image the result is an image that emphasizes the dark regions that were previously diminished, thereby enhancing vessel-like structures. Thus I_{BH} has large values in pixels that belong to dark linear structures of any orientation. This image was subsequently normalized to the range [0,255] to standardize intensity values. The next step is the application of hysteresis thresholding (HT) with large and low thresholds, $t_H = 200$ and $t_L = 90$ (these limits were decided after multiple trials) to generate the final binary image B

$$B = HT(I_{BH}, t_H, t_L). \quad (4)$$

The resulting binary image was then passed through a Frame Score metric that is analyzed bellow and was used for the final results.

2.6.2 Vesselness filtering approach

In this approach, the vesselness filter [7], [32], was used to enhance vascular structures across the image frames. The vesselness filtering is an image enhancement technique that highlights tubular or curvilinear structures, based on they second-order intensity profile implemented in [33]. This method utilizes the Frangi filter which is based on the Hessian matrix to assess vessel-like

characteristics. Initially, the image is convolved with a Gaussian kernel at different scales σ to handle structures of various sizes:

$$I(x, y; \sigma) = I(x, y)G(x, y; \sigma) \tag{5}$$

At each pixel, the Hessian matrix is computed from the second order derivatives of the smoothed image, the eigenvalues λ_1 and λ_2 are calculated. For tubular structures one eigenvalue is small and the other is large and negative. For blob like structures both eigenvalues are large and negative and for flat regions both eigenvalues are near zero. The Vesselness response is computed from a vessel like profile:

$$V = \begin{cases} 0, & \lambda_2 > 0 \\ \exp\left(-\frac{R_B^2}{2\beta^2}\right) \cdot \left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right), & \text{otherwise} \end{cases} \tag{6}$$

Where:

- $R_B = \frac{\lambda_1}{\lambda_2}$ — a measure of deviation from a line-like (tubular) shape.
- $S = \sqrt{\lambda_1^2 + \lambda_2^2}$ the second-order structureness, capturing local contrast.
- β and c — parameters that control sensitivity to blob-like structures and noise, respectively.

This function ensures that high vesselness scores are assigned to pixels within elongated, low-contrast structures The vesselness measure is computed over range of scales σ and the final response is the the maximum response at each pixel:

$$V_{\text{final}}(x, y) = \max_{\sigma} V(x, y; \sigma) \tag{7}$$

The resulting Vesselness probability map was normalized to the intensity range [0,255] to standardize values for subsequent processing. A hysteresis thresholding operation was then applied with a high threshold of 230 and a low threshold of 140, ensuring robust detection of continuous vessels segments while suppression weak responses due to noise. The binarized Vesselness map was then subjected a Frame Score metric analysis to determine the number of distinct vascular regions in the binary image.

2.6.3 Steger line detection approach

Steger's line detection method [34] was employed to extract the center lines of elongated structures in images. It is based on differential geometry principles and relies on the second-order structure of image intensity, combined with subpixel interpolation methods. Specifically, it is designed to detect the medial axis of line structures, such as bright or dark lines of finite width on contrasting background. This method uses the Hessian matrix to identify locations where the gradient orthogonal to the line vanishes and where the second derivative across the line is maximal. Initially the input image is convolved with a Gaussian filter at a σ scale :

$$I_\sigma(x, y) = I(x, y) * G(x, y; \sigma), \tag{8}$$

where: σ controls the scale at which lines are detected (small *sigma* for fine lines and large *sigma* for thicker lines. Following this, the Hessian matrix is computed, at each pixel, the one eigenvalue

is negative (for bright lines) or positive (for dark lines) indicating a high curvature perpendicular to the line and the eigenvector points orthogonal to the line.

$$H = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}$$

A pixel is considered a line point if the first derivative of intensity in the direction orthogonal to the line is zero and if the second derivative in that direction is maximal. Mathematically this is equivalent to calculating the following quantity t:

$$t = -\frac{r_x n_x + r_y n_y}{r_{xx} n_x^2 + 2 r_{xy} n_x n_y + r_{yy} n_y^2} \tag{9}$$

and selecting the pixels for which $|t| < 0.5$. Subpixel accuracy is achieved by interpolating the intensity profile along the direction of the eigenvector and locating the precise extremum; this interpolation allows a better accuracy than simple pixel resolution and a smooth representation of the line. Unlike the previous methods, the resulting vessel image is binary.

Example output of the three tested pre-processing methods applied to the same angiographic frame. The frames that were selected by Bottom-Hat and Steger, provided less information compared to Vesselness and missed the vessel structure while in some case they failed to select the most diagnostically relevant frame.

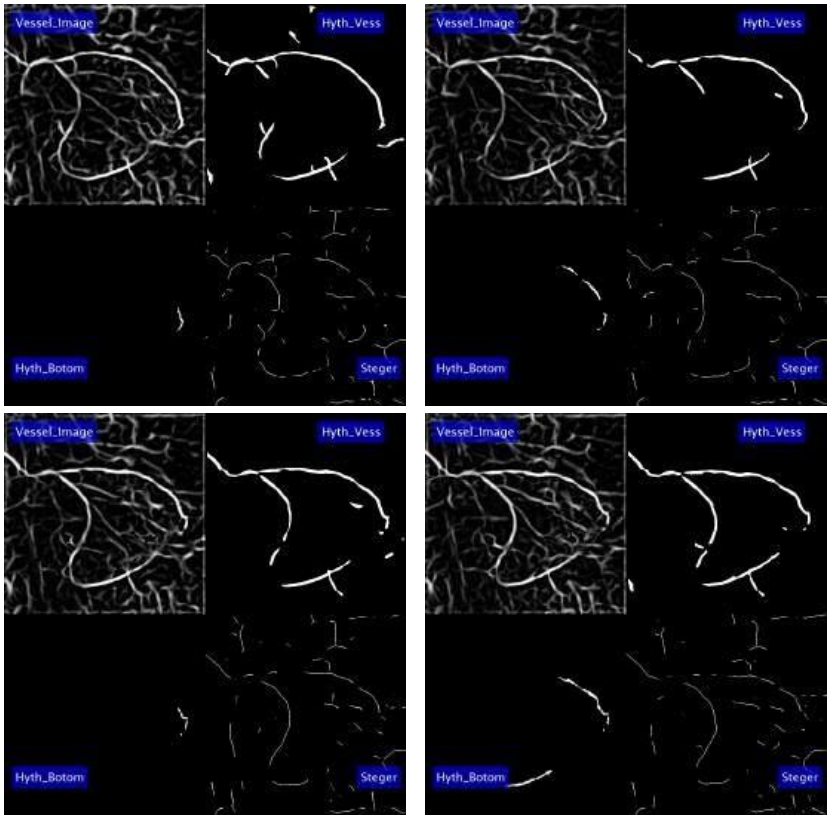


Figure 16: Multiple frames extracted from each method. Original image: Upper left, Vesselness with hysteresis thresholding

2.7 Frame-based statistics

This subsection is applied after each one of the aforementioned vessel enhancing methods (i.e. bottom hat, Vesselness, and Frangi's method). As already mentioned, the binarized vessel image is subjected to connected component analysis using 8-connectivity to determine the number of distinct vascular regions in the binary image. For any frame let us denote by k the number of connected components and by n_i the area (number of pixels of the ith component.

The extracted measures included the area of all components defined as the number of non-zero pixels $\sum_{i=1}^{k} n_i$ and the area of components larger than T_p = 400 pixels $\sum_{n_i>T_p} n_i$. The Frame Score was calculated by dividing the area of larger components and the number of connected components:

$$FS = \frac{\text{Number of non-zero Pixels of large components}}{\text{Number of Connected Components}} = \frac{1}{k} \sum_{n_i>T_p} n_i \quad (10)$$

Components smaller than T_p pixels were excluded, since empirical trials showed that values below 400 introduced mid to excessive noise and did not contribute to meaningful structural in- formation, while values above 400 eliminated important vascular segments. The Frame score ratio was chosen because a higher number of separate connected structures to the overall area indicates important structural complexity. We define as local maxima in Frame score, frames q that have higher frame score than the previous and the next frame (irrespectively of the absolute value of the score). More formally

$$q : FS_q > FS_{q-1} \; AND \; FS_q > FS_{q+1} \quad (11)$$

This metric provided a normalized measure of spatial activity in each frame. By choosing frames that corresponded to the local maxima, a clear vessel information was ensured, which is essential for diagnostic interpretation and deep learning training.

These local maxima frames q were compared to the first frame of the sequence using the Structural Similarity Index (SSIM) [35], since we assume that no radiopaque contrast has been injected in frame 1. $SSIM(i, j)$ is a perceptual metric that evaluates image similarity between frames i and j, based on luminance, contrast, and structural information.

The final selected frame q_0 is defined as the local maxima frame q with the lowest SSIM value relative to the first frame, i,e. the most dissimilar to the 1st frame

$$q_0 = \min_q SSIM(1, q) \forall q \quad (12)$$

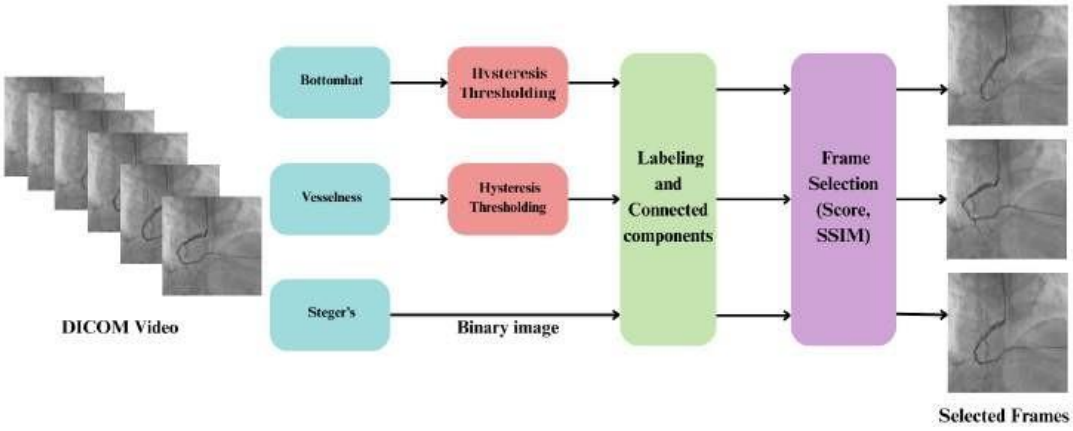


Figure 17: Alternative pipelines for frame selection from DICOM video.

The histogram analysis of Figure 18 as well as the expert's opinion indicated that the Vesselness method shows superior performance compared to the Bottom-hat and Steger approaches. It appears to exhibit higher and more consistent pixel counts across frames, with the vessels appearing in a much clearer form. In contrast, the Bottom-hat and Steger methods show greater variability and lower overall pixel intensities, and they also miss important points of the vessel, which implies reduced robustness in capturing fine vascular details.

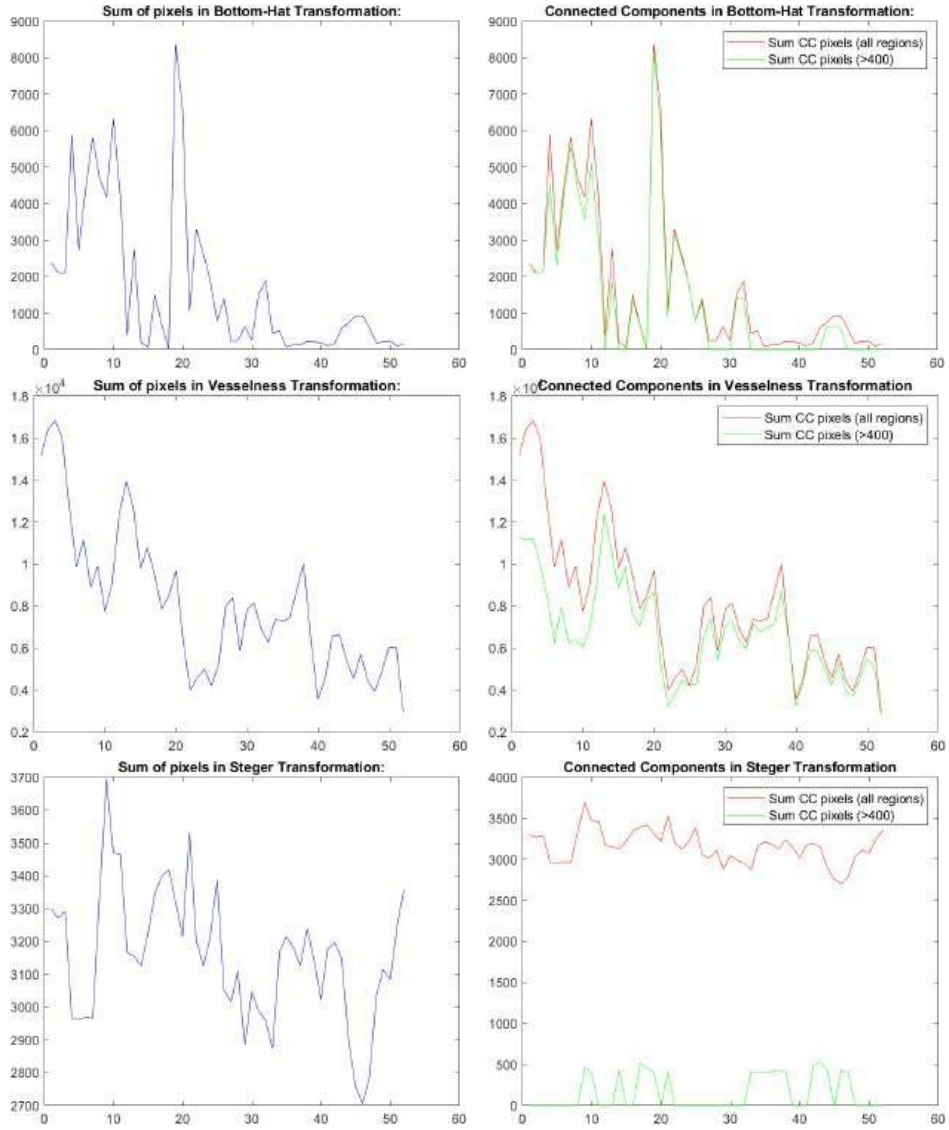


Figure 18: Number of Non-Zero pixels for each method (left) and number of CC before thresholding (red) and number of CC after thresholding (green) (right)

The outcomes of the applied transformations for each method are illustrated in Figure 19. In the case of Bottom-Hat, the frame score exhibits multiple prominent peaks. These peaks are compared with the first frame, and the minimum SSIM value is identified at the 11th local maximum. For Vesselness approach, several local maxima are highlighted, however, the SSIM method selects the frame which corresponds to the 7th peak of the computed Frame Score as the most representative. Finally, in the case of Steger method, the selected peak is observed to be closely with the one obtained by Bottom-Hat transformation, since it selected the 10th peak of frame score.

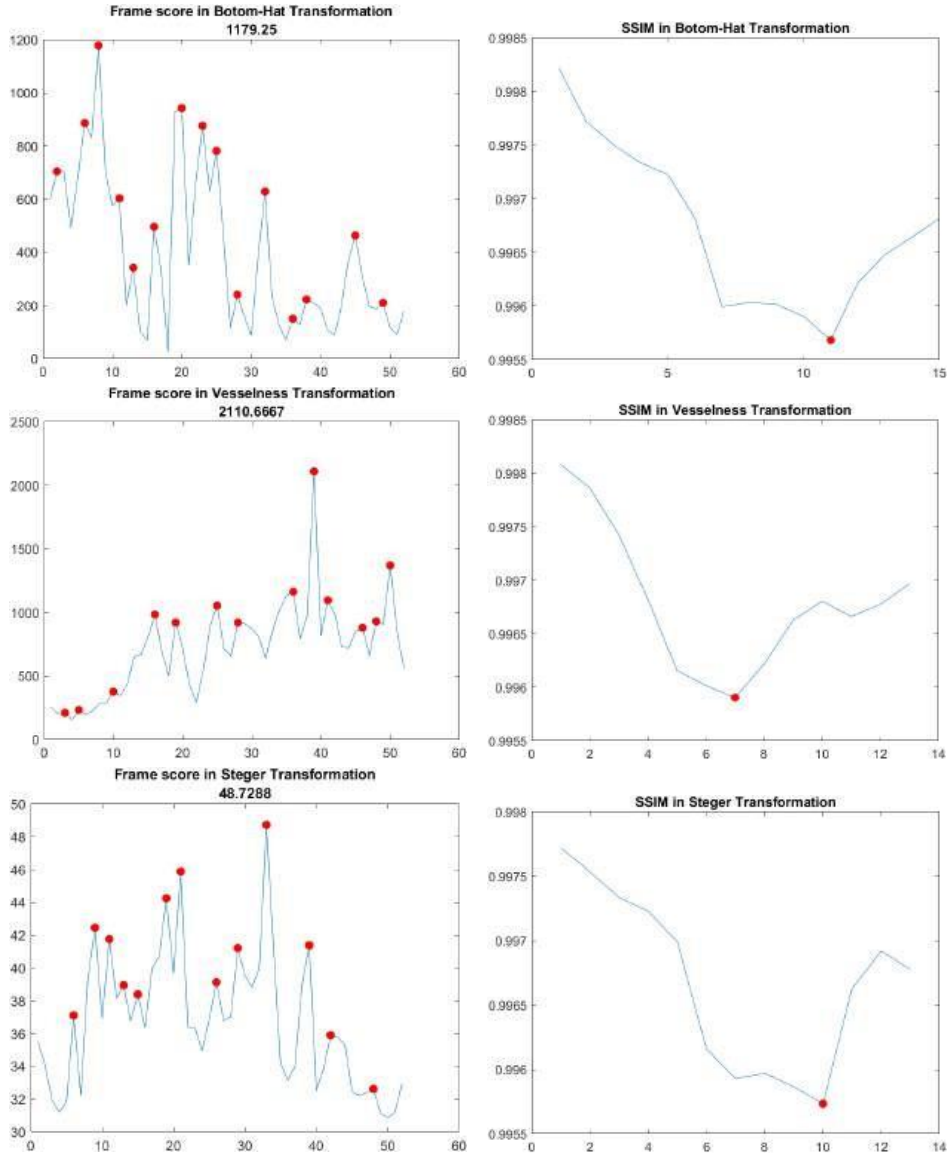


Figure 19: Plot of Frame score and SSIM for each method

In Figure 20, the representative frames selected by each vessel enhancement method are highlighted with colored bounding boxes. The red box corresponds to the Bottom-hat method, the blue box to the Steger approach, and the green box to the Vesselness method. These annotations allow a direct visual comparison of the structural clarity achieved by each technique. It can be observed that the vesselness-based frame (green) preserves a more continuous and contrasted vessel tree, while the bottom-hat (red) and Steger (blue) selections display partial loss of vascular detail or reduced robustness.

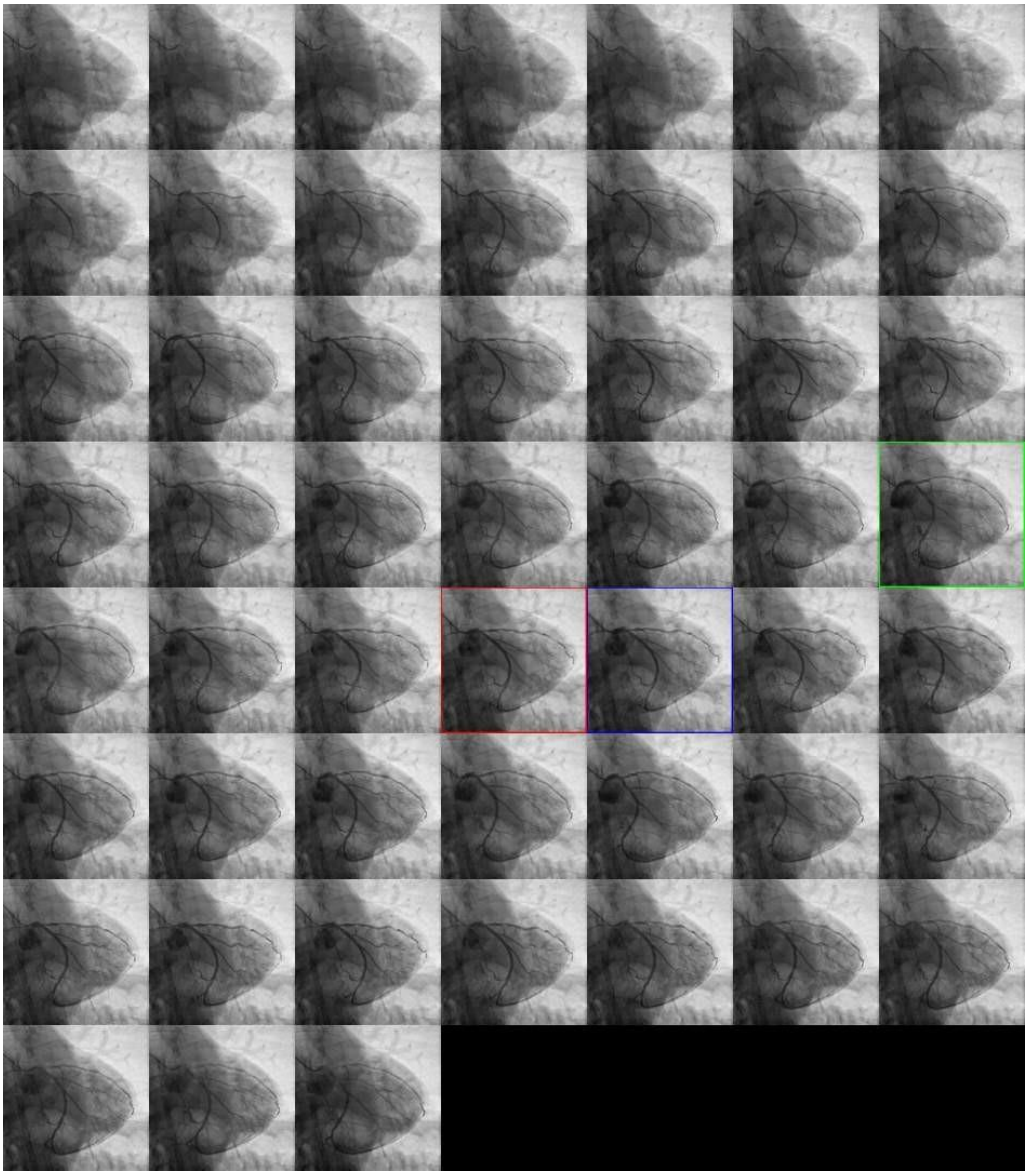
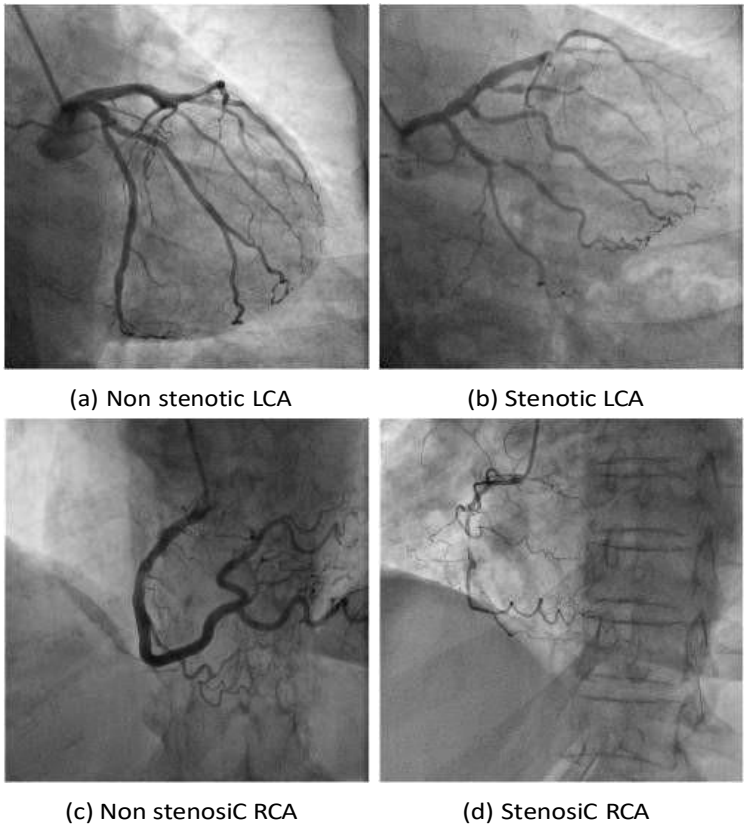


Figure 20: Frame selection for each method

2.8 Stenosis-No Stenosis Frame Classification

To classify coronary angiographic images for the presence of stenosis we used a convolutional neural network based on AlexNet, adapted for a multi-label setting. As a preprocessing step, stenosis detection was applied only on frames that had already passed through the preceding stages of the pipeline, coronary artery side classification (LCA vs. RCA), and balloon/non-balloon detection

and frame selection. This ensured that only diagnostically relevant frames corresponding to the correct artery and intervention type were considered in the stenosis classification task. Overall the dataset consisted of 3,347 images, with 2,549 and 798 images displaying LCA and RCA vessels, respectively. In Figure 21 an example of one stenotic vessel and one non stenotic vessel is displayed for both LCA and RCA artery.



(a) Non stenotic LCA (b) Stenotic LCA

(c) Non stenosiC RCA (d) StenosiC RCA

Figure 21: Examples of the Stenosis Dataset.

The classification was performed separately for the right and the left coronary artery, due to their anatomical differences. Each image was resized to $227 \times 227 \times 3$ to match the AlexNet input. Ground truth labels were derived from the original medical diagnosis, on which we applied multi binary labeling that describe each segment of the coronary vessels separately. One (1) indicates the existence of stenosis (positive) and zero (0) the absence of stenosis (negative). In our custom dataset, each image was defined by a unique patient code and the year of the examination, to enforce patient-wise splitting into 70% training, 15% validation, and 15% test sets, with a fixed random seed to ensure reproducibility. Three different classification models were developed:

1. RCA-model, trained specifically on RCA specific images, with 3 output neurons corresponding to the three anatomical RCA segments as shown below.

Proximal	Middle	Distal
Neuron_1	Neuron_2	Neuron_3

Table 2: ”Neurons corresponding to 3 anatomical RCA segments”

2. LCA-model (3 neurons) trained on LCA images, with 3 output neurons, that represent combined sections that derived from the original eight-segment classification.

LCMA	LAD	LCx
Neuron_1	Neuron_2	Neuron_3

Table 3: ”Neurons corresponding to 3 anatomical LCA segments”

3. LCA-model (8 neurons) trained on LCA images, with 8 output neurons each corresponding to one of the original LCA.

	Proximal	Middle	Distal
LCMA	Neuron_1	Neuron_2	N/A
LAD	Neuron_3	Neuron_4	Neuron_5
LCx	Neuron_6	Neuron_7	Neuron_8

Table 4: ”Neurons corresponding to 8 anatomical LCA segments”

In Figure 22 illustrates how the detailed coronary tree segments that was defined in [36] were consolidated into the anatomical classes used for the stenosis classification task in our work. For the left coronary artery, the eight original segments were grouped into three main territories: the left main coronary artery (LCMA), the left anterior descending artery (LAD), and the left circumflex artery (LCx), each further subdivided into proximal, mid, and distal regions. Specifically, segments 5, 6, 7, 8, 9, 9a, 10, 10a, 11, 12, 12a, 12b, 13, 14, 14a, 14b, and 15 were mapped into these categories according to their anatomical position and functional role. Similarly, for the right coronary artery, the original subdivisions were merged into three classes: proximal (segment 1), mid (segment 2), and distal (segments 3, 4, 16, 16a–c), thereby covering the full course of the vessel including the posterior descending and posterolateral branches. These schematic representations ensured that the neural network models were trained with anatomically consistent labels while reducing unnecessary granularity from side branches.

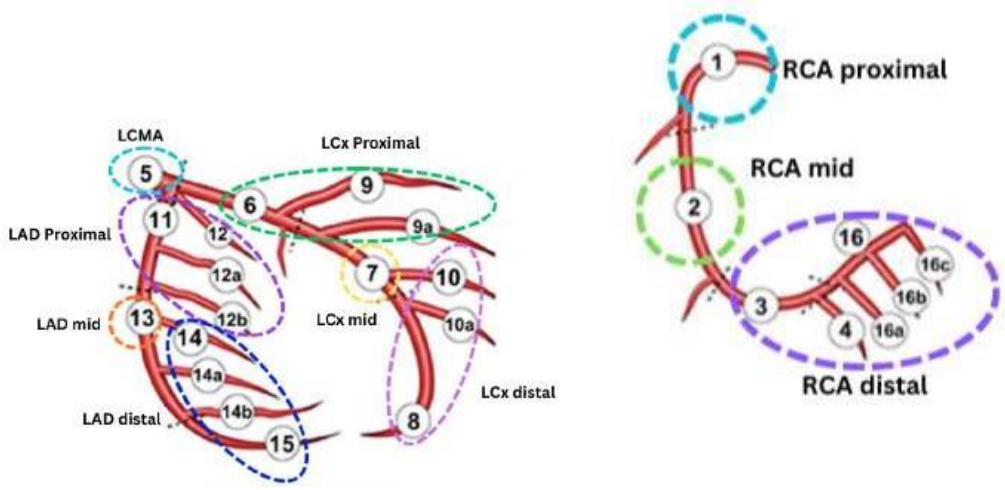


Figure 22: Segmentation of anatomy of the heart [5]

For the LCA, two models were explored: one with three output neurons corresponding to clinically groups of the eight original segments, and one with eight neurons corresponding to the full segmental scheme. For the RCA, a model with three neurons was trained, each representing one anatomical segment.

In each case the AlexNet was finetuned by replacing the final layer with the number of the required output neurons, depending on the model, followed by a sigmoid activation layer. The rest of the convolutional layers retained their pre-trained weights, while higher layers were trained with an increased learning rate to adapt to the domain-specific classification task. Training was performed with the Adam optimizer, a mini-batch size of 32, a maximum of 50 epochs, and binary cross-entropy loss, with early stopping (validation patience of 10) applied to prevent overfitting.The LCA (3 neurons) was overall trained for 14 epochs while the LCA (8 neurons) was trained for 34 epochs. The RCA model was stopped after 20 epochs. The network was trained with binary cross-entropy loss, reflecting the multi-label nature of the problem.

[5]Image adapted from [36].

3 Results

For the evaluation of classification results we employed multiple metrics. Frame selection performance was assessed using distance-based measures across the three methods, while the classification tasks (LCA vs RCA, Balloon detection, Stenosis Detection) were evaluated through confusion matrices and general metrics.

In a binary classification setting, **precision** is the number of correclty predicted positives (true positives) among all sample that were predicted positives (true positives and false positives). **Recall** it is the correctly predicted positives samples among all actual positive samples (true positives and false negatives). Finally the **F1-Score** is the mean of precision and recall.

$$Precision = \frac{\text{Relevant Retrieved Instances}}{\text{All Retrieved Instances}} = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{\text{Relevant Retrieved Instances}}{\text{All Relevant Instances}} = \frac{TP}{TP + FN} \tag{14}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{15}$$

3.1 Left-Right Coronary detection

The detection of whether an angiographic frame corresponds to the LCA) or RCA is an important and initial step in automated analysis, since subsequent classification tasks (e.g., stenosis localiza- tion or balloon detection) rely on the correct identification of the arterial vessel. The following subsections present the results of both models that were explained previously in Section 2.4 with performance reported in terms of confusion matrices, precision, recall, F1-scores, and overall accu- racy.

3.1.1 Classic AlexNet approach for LCA and RCA Frame classification

In the initial model, we fine tuned a pretrained AlexNet for the binary classification of LCA and RCA images. The resulting confusion matrix is shown in the Figure 23. The model achieved 823 true positives for the Left class, while misclassifying 32 LCA as RCA, at the same time it correctly predicted 314 true negatives for the Right class, with 17 incorrectly predicted samples as LCA.

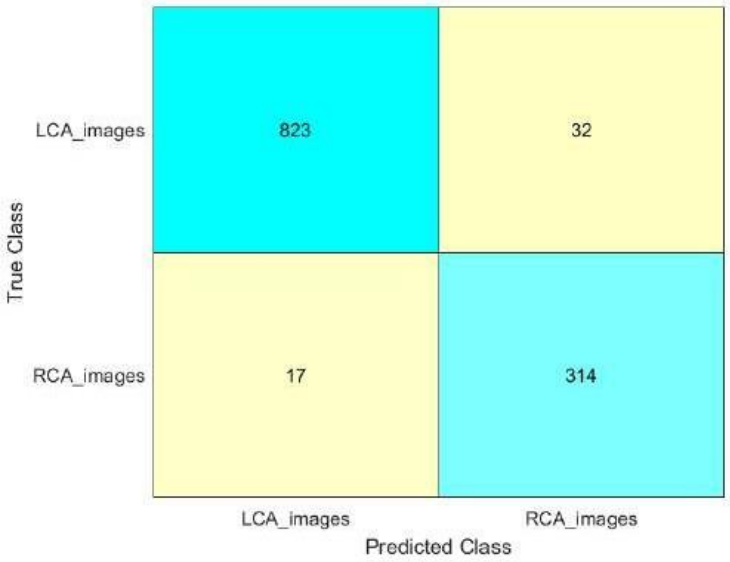


Figure 23: Performance metrics for LCA vs. RCA classification using the classic AlexNet model.

As summarized in the Table 5 this corresponds to a precision values 97.9 % and recall value of 96.2% for the Left class and 90.7% precision and 94.8% for the Right class. These metrics indicate high sensitivity and overall balanced performance across classes. The model reached an overall accuracy of 95.87% and a general F1-Score of 94.8%.

Class	Precision%	Recall%	F1-score%
LCA	97.9	96.2	97.0
RCA	90.7	94.8	92.7
Overall Accuracy			95.87%

Table 5: Performance metrics for LCA vs. RCA frame classification using the classic AlexNet.

The ROC curve rises and aligns with the upper boundary, indicating near-perfect separability between right- and left-coronary images, even if the datasets are unbalanced. High sensitivity is observed at a low false positive rate close to 0.05 allowing a high score of specificity while correctly identifying almost all RCA frames. The straight ROC curve at the end suggests stable performance across a broad threshold range.

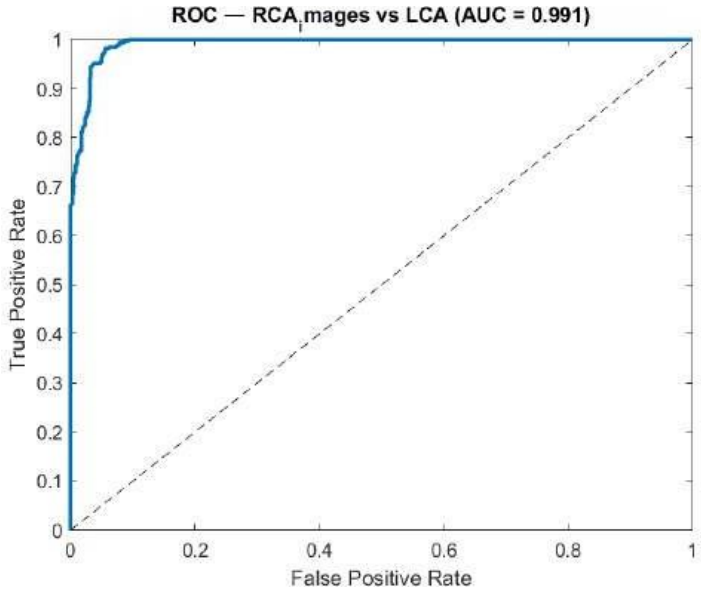


Figure 24: ROC curve for frame-based LCA/RCA AlexNet.

3.1.2 Custom AlexNet approach for LCA and RCA classification

The second approach was based on a custom network that combined a fine tuned version of AlexNet and some additional input features, incorporating the angiographic primary angle metadata. As shown in the Figure 25, the model achieved 822 true positives for the Left class and 300 true negatives for the Right class. Misclassifications included 33 Left-class images incorrectly predicted as Right (false negatives for the Left class) and 31 Right-class images incorrectly predicted as Left (false positives for the Left class). According to the Table 6 this corresponds to a precision values 96.3 % for the Left class and 90.0% for the Right class, with recall values of 96.1 % and 90.6% respectively, highlighting high sensitivity and overall balanced performance across classes. The model showed a general accuracy of 94.8% and a general F1-Score of 93.3%.

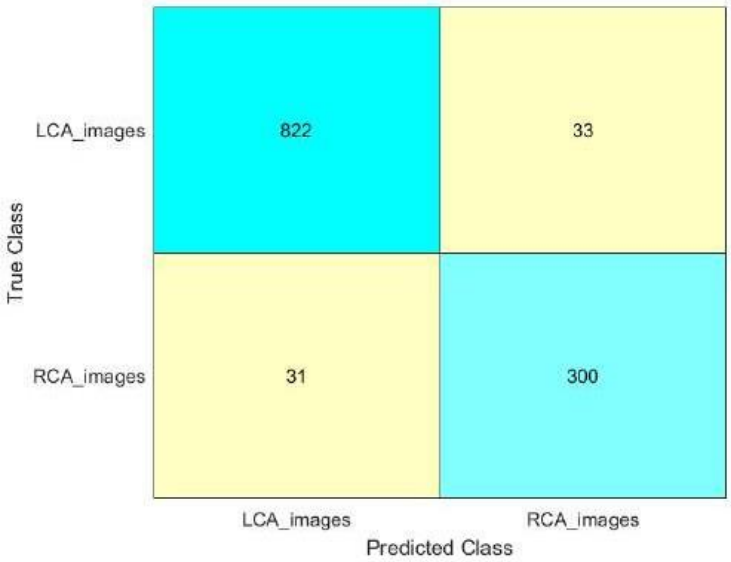


Figure 25: Frame-based confusion matrix for Left-Right Coronary artery with Angles as input.

Class	Precision%	Recall%	F1-score%
LCA	96.3	96.1	96.2
RCA	90.0	90.6	90.3
Overall Accuracy			94.8%

Table 6: Frame-based performance metrics for LCA vs. RCA classification using the custom AlexNet with angle input.

While the inclusion of angular metadata slightly showed a more balanced metrics for both classes, the overall performance was marginally lower than the classic approach. This may indicate that the dataset contained some faulty or inconsistent labels, specifically cases where the recorded angle did not represent the correct artery, likely due to procedural alterations during angiography. The metrics for discrimination between RCA images and LCA views still remain very high. The ROC curve remains close to the top-left corner, achieving high true-positive rates with minimal false alarms. The slight reduction relative to the RCA comparison of the Classic AlexNet approach may be due to class imbalance, but overall the classifier exhibits robust and threshold-insensitive behavior for this task.

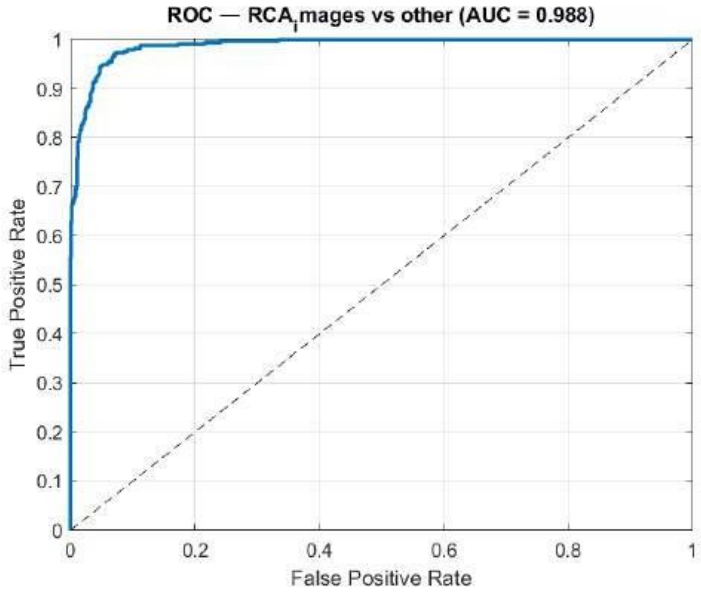


Figure 26: ROC curve LCA/RCA angles

3.1.3 RCA-LCA Video Classification

The confusion matrix in Figure 27 illustrates the performance of the model in classifying videos between left coronary artery (LCA) and right coronary artery (RCA) cases, achieving accuracy of 97.3%. Out of the total LCA cases of test dataset, 806 were correctly classified as LCA, while only 26 were misclassified as RCA. Similarly, among RCA cases, 347 were correctly predicted, and 6 were misclassified as LCA. These results indicate a very high overall accuracy, with the misclassification rate being minimal even in the video classification, suggesting that the model can reliably differentiate between LCA and RCA.

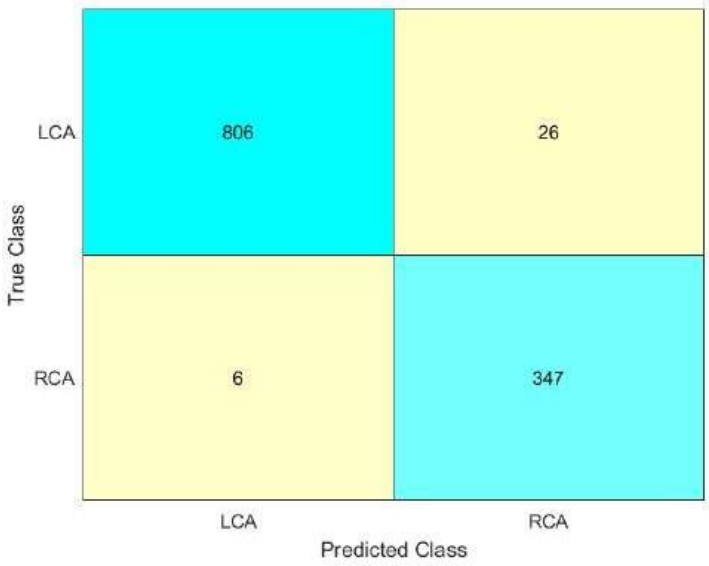


Figure 27: Video-based Classification of RCA and LCA.

3.2 Balloon Detection

The next step that we evaluated was balloon detection, distinguishing between balloon-based and non-balloon coronary interventions. We reviewed the performance of each of the three models: a fine-tuned AlexNet baseline, a Vision Transformer (ViT-Base), and a Vision Transformer in a Tiny configuration (ViT-Tiny). For each model, confusion matrices were generated and standard classification metrics were computed.

3.2.1 Frame classification with Classic AlexNet Model

The first model was based on AlexNet [26], as shown in the Figure 28, the model achieved 1136 true positives for the Non-Balloon class and 1191 true negative for the Balloon class. The misclassifications cases included 118 false positives and 289 false negatives.

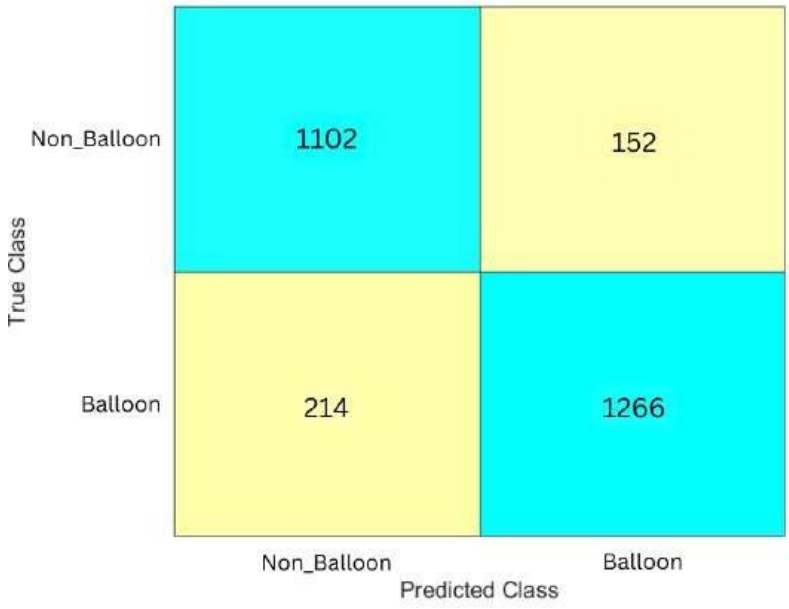


Figure 28: Confusion matrix of Balloon vs. Non-Balloon classification using the AlexNet model.

This resulted in an overall accuracy of 86.6%. According to Table 7, the model reached a precision of 83.7% and recall of 87.8% for the Non-Balloon class and achieved a precision of 89.2% and recall of 85.54% for the Balloon class. The results indicate that although the model was able to detect balloons

Class	Precision%	Recall%	F1-score%
Arteries_other	83.7	87.8	85.7
Balloon	89.2	85.5	87.3
Overall Accuracy			86.61%

Table 7: Performance metrics for Balloon vs. Non-Balloon classification using the AlexNet model.

3.2.2 Frame classification with Vision Transformer (ViT), base model

A base visual transformer was used for the classification that produced the following results with a confusion matrix. As shown in the Figure 29, the model achieved 1269 true positives for the Non-Balloon class and 1486 true negative for the Balloon class. The misclassifications cases included 201 false negatives and 4 false positives. That shows a bigger sensitivity in misclassified non-balloon for balloon.

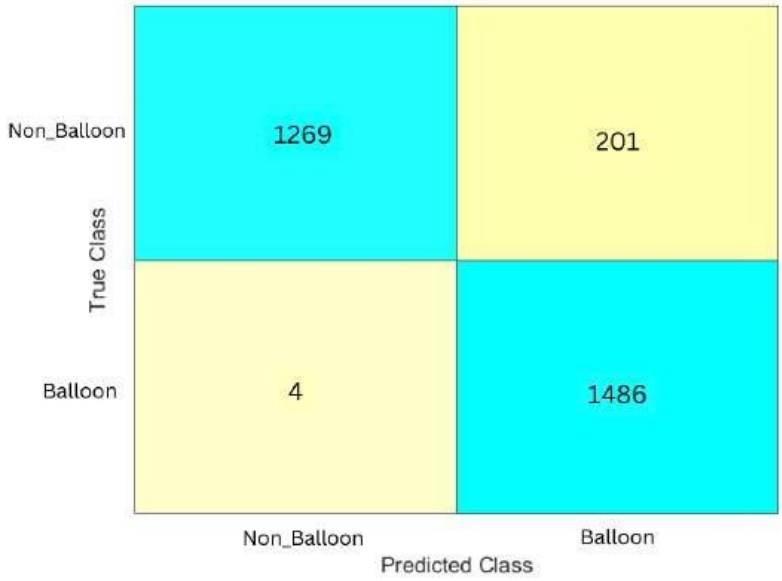


Figure 29: Confusion matrix of Balloon vs. Non-Balloon classification using the ViT-Base model.

As summarized in Table 8, this corresponds to an accuracy of 93.07%. The Non-Balloon class achieved a very high precision of 99.0%, although the recall was lower 86.0%, suggesting the model is missing certain types of images for Balloon. Conversely, the Balloon class achieved a precision of 88.0% and a recall of 99.0% , this indicates that the model is more sensitive in identifying Balloon cases so the misclassifying cases of balloons as non-balloons are significantly less

Class	Precision%	Recall%	F1-score%
Arteries_other	99.0	86.0	92.0
Balloon	88.0	99.0	93.0
Overall Accuracy			93.07%

Table 8: Performance metrics for Balloon vs. Non-Balloon classification using the ViT-Base model

3.2.3 Frame classification with Vision Transformer, tiny model

For the second approach of the transformer a tiny visual transformer was used. As shown in the Figure 30, the model achieved 1280 true positives for the Non-Balloon class and 1444 true negative for the Balloon class. The misclassifications cases included 46 false positives and 190 false negatives. That shows a bigger sensitivity in misclassified non-balloon for balloon.

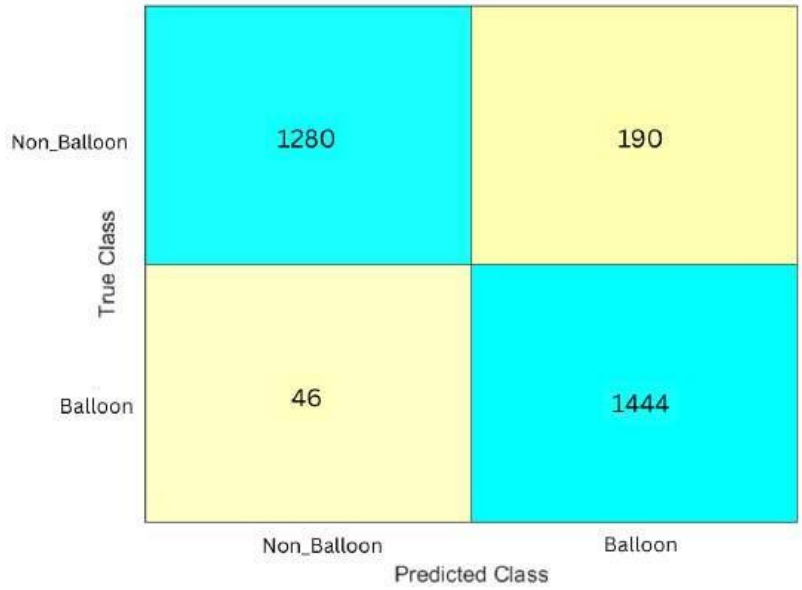


Figure 30: Confusion matrix of Balloon vs. Non-Balloon classification using the ViT-Tiny model.

This yielded an accuracy of 92.06%. As shown in Table **??**, the Non-Balloon class reached a precision of 96.5% and recall of 87.0%, while the Balloon class achieved 88.0% precision and 96.9% recall. Performance was slightly below that of ViT-Base, but still considerably higher than the AlexNet baseline, demonstrating the advantage of transformer-based architectures even in smaller configurations.

Class	Precision%	Recall%	F1-score%
Arteries_other	96.5	87.0	91.5
Balloon	88.0	96.9	92.4
Overall Accuracy			92.06%

Table 9: Performance metrics for Balloon vs. Non-Balloon classification using the ViT-Tiny model.

3.2.4 Balloon- Non-Balloon Video Classification

The confusion matrix in Figure **??** presents the *video* classification results for balloon inflation or non-balloon (artery) sequences, achieving accuracy equal to 96%.

The model correctly identified 547 non-balloon frames and 407 balloon frames, while it misclassified 23 non-balloon cases as balloon and 13 balloon cases as non-balloon. Although the overall performance is strong, the model shows some tendency to confuse non-balloon frames as balloon, which could impact the precision of balloon detection. However the balanced recognition of both classes supports the model's usefulness in filtering interventional phases for subsequent stenosis analysis.

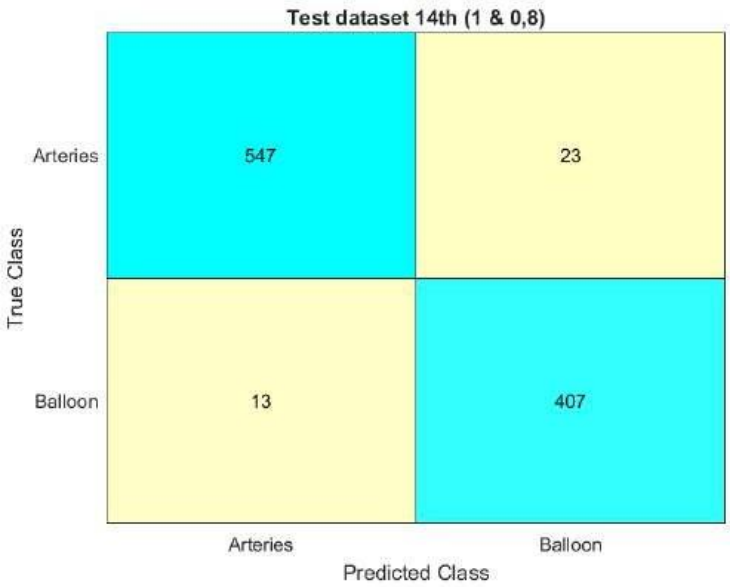


Figure 31: Video Base Classification of Balloon and Non-Balloon

3.3 Results of Classification RCA/LCA with AlexNet and Balloon/Non-Balloon with Base transformer

To extend the frame-level classification into a clinically meaningful video-level prediction we adopted a majority-voting strategy across all frames of each angiographic sequence. Specifically as shown in Figures 32, 33, 34 each frame was independently classified with the predicted class labels for the coronary artery type (LCA or RCA) using the classic AlexNet approach and the balloon status (Balloon or Non-Balloon) by using the Base Transformer. The annotation appears in white when the prediction is correct, while incorrect predictions are highlighted in red. This visualization enables a frame-by-frame inspection of the classifier's performance, making it possible to identify both consistent classification regions and specific frames where errors occur. By analyzing the spatial and temporal distribution of these misclassifications, the robustness of the model can be assessed, as well as potential challenges arising from overlapping anatomical structures, low contrast, or motion artifacts.

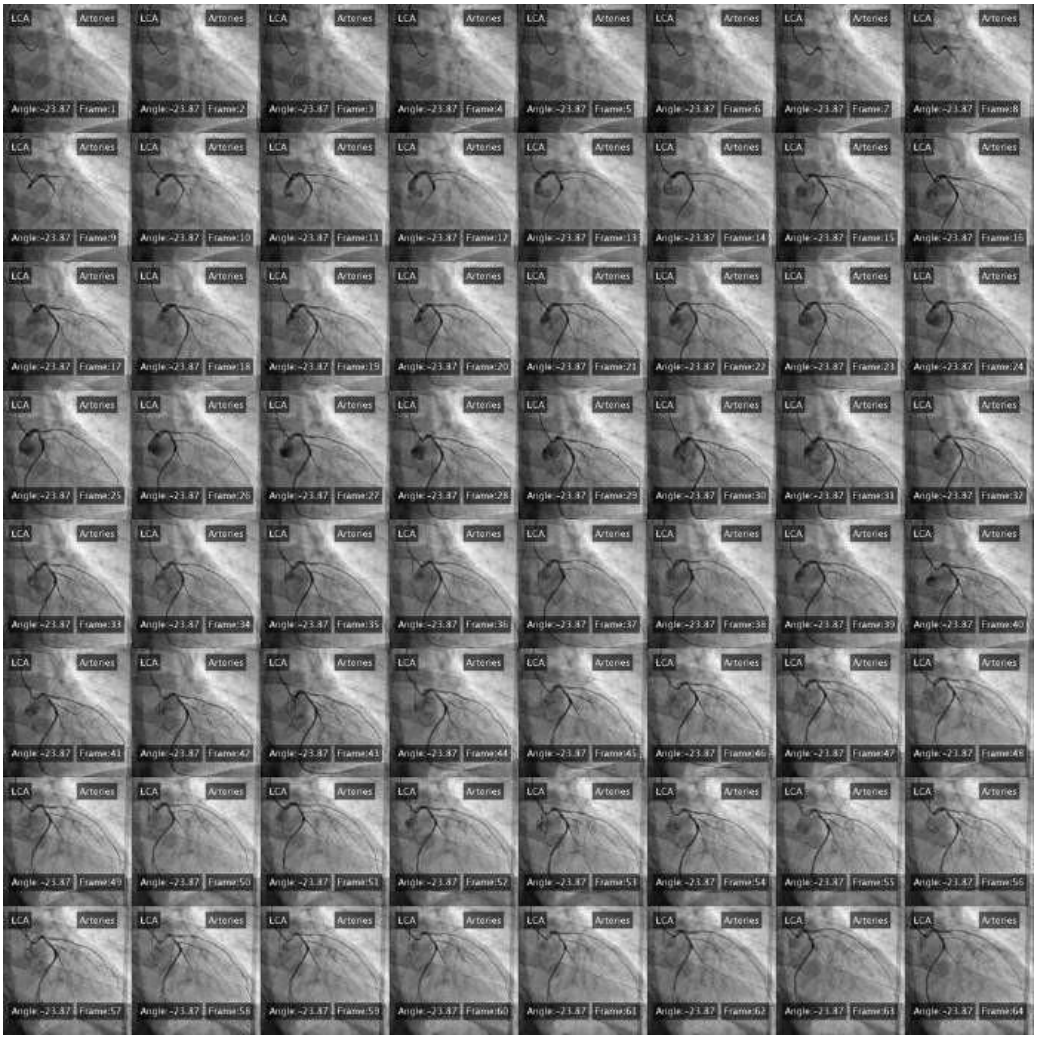


Figure 32: LCA Artery

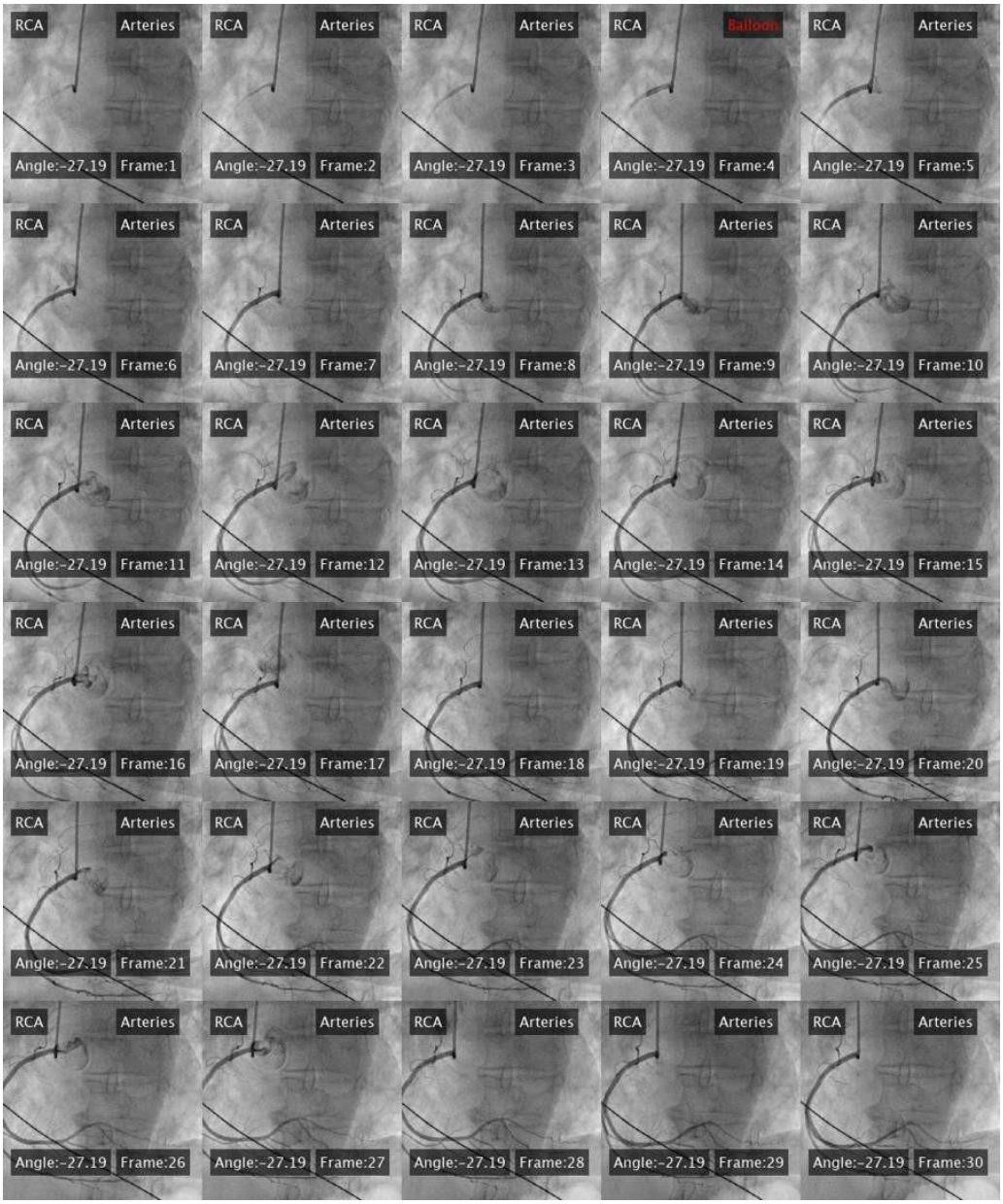


Figure 33: RCA Artery

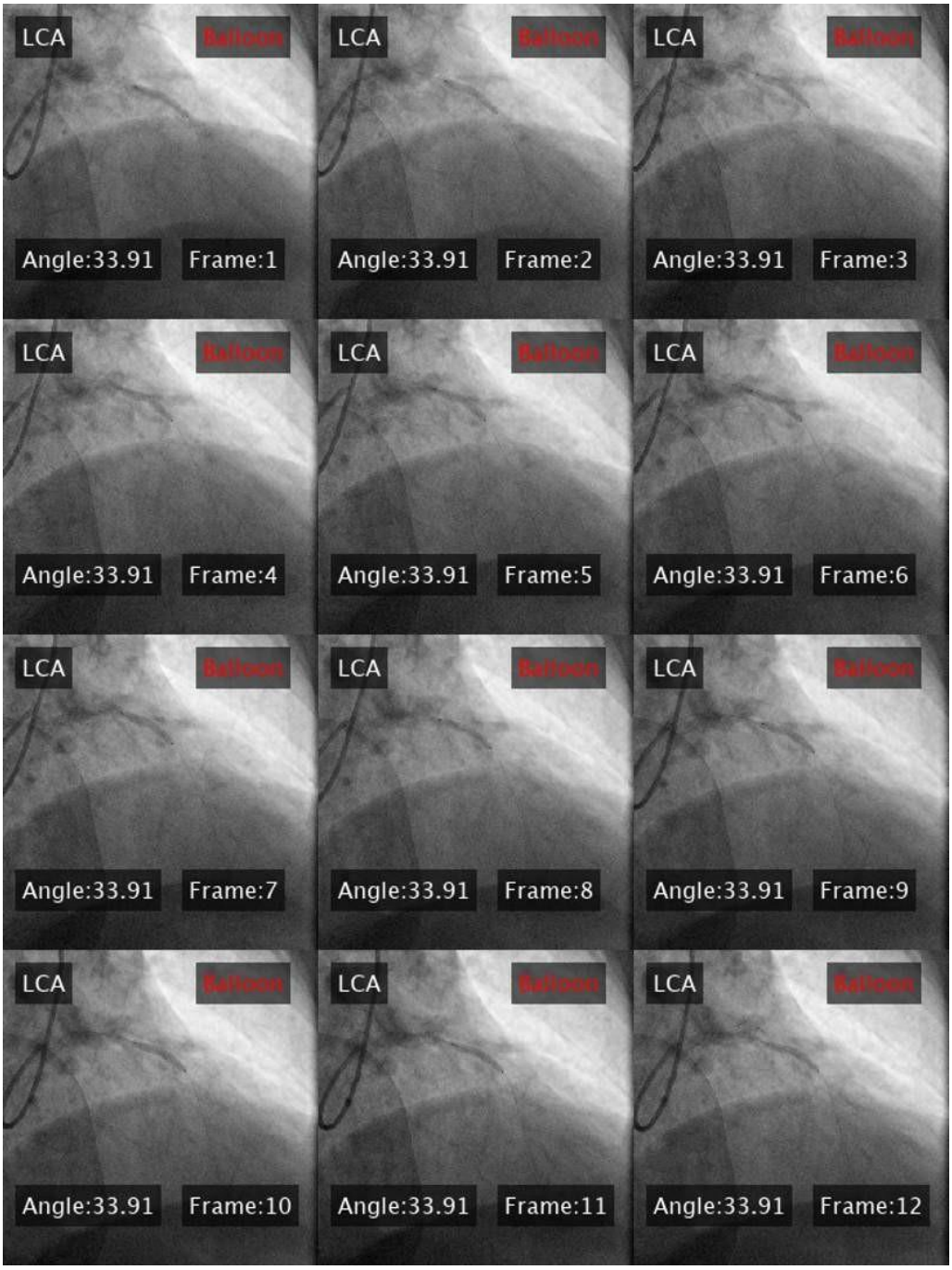
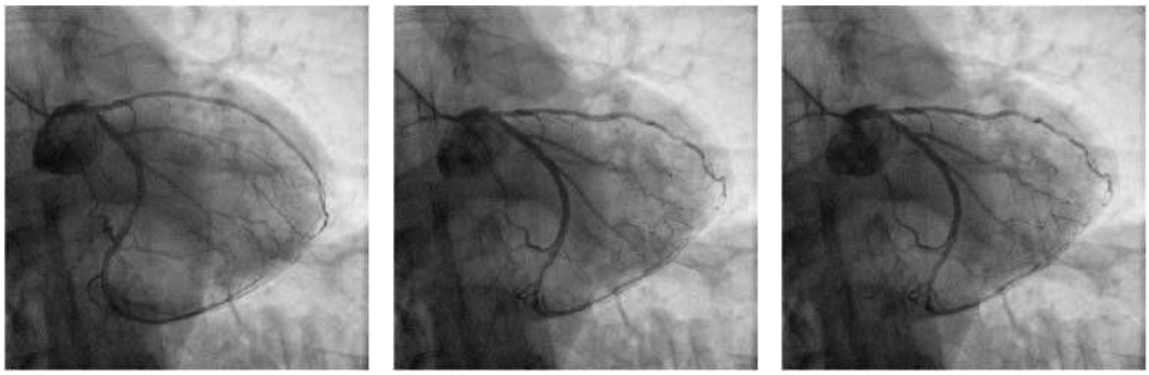


Figure 34: Balloon

3.4 Frame selection

After the system determines wether the sequence corresponds to RCA or LCA, the algorithm verifies wether a Balloon intervention has already been identified in the current or prior sequences (according to their acquisition time). In this case, the video of this sequence is discarded. Otherwise all three frame selection methods are applied to the current video, returning one frame per video (per method).

Even though the three frame selection methods resulted in very small temporal distances between their selected frames, as shown in Figure 35, some quality differences were observed in the representative frames.



(a) Vesselness (b) Bottom-hat (c) Steger lines

Figure 35: Selected frames for each method.

3.5 LCA Stenosis detection in the 3 main coronary vessels

In this subsection, stenosis is detected in the 3 main coronary vessels of LCA (LCMA, LAD, LCx), as a 3-class multi-label problem. Consequently, the output layer of the CNN consists of 3 neurons.

Results are presented for the test subset, which has been split per patient. In the corresponding confusion matrices, the class labels are defined as follows: 0 indicates the absence of stenosis, while 1 indicates the presence of stenosis. This binary scheme provides an intuitive representation of the classifier's predictions, where correct classifications appear along the diagonal of the matrix (true negatives for label 0 and true positives for label 1), and misclassifications are reflected in the off-diagonal entries (false positives and false negatives). Such a representation allows direct assessment of the model's ability to correctly discriminate between normal and stenotic segments of the left coronary artery.

3.5.1 LCMA Stenosis Detection

The first confusion matrix in Figure 36 represents the classification results of the networks first output neuron, which refers to the Left Main Coronat Artery (LMCA) identification of stenosis. All of the 1,486 non-stenosis frames were correctly identified (true positives) with the model achieving perfect sensitivity. For the stenosis frames, 362 out of 365 were correctly detected (true negatives), with 3 misclassifications.

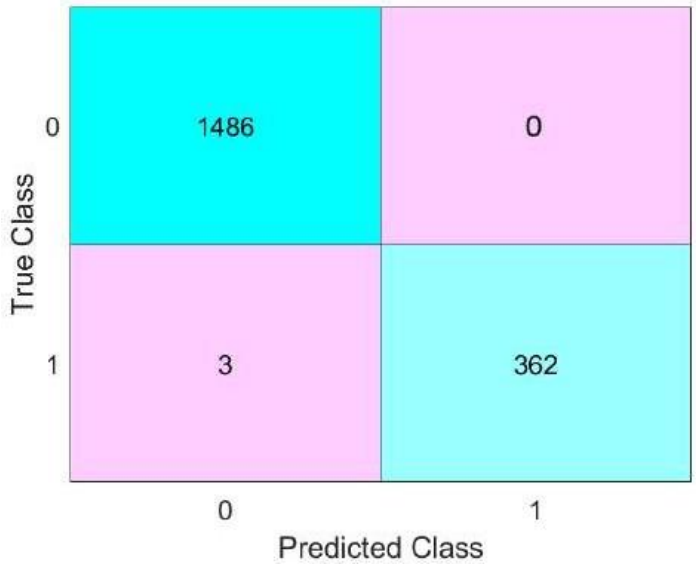


Figure 36: Confusion matrix for stenosis detection in the proximal LCA segment.

The left main coronary artery (LMCA) ROC aligns with the upper boundary, yielding near- perfect performance across thresholds. The model sustains > 0.99 true-positive rate at close to zero false-positive rate, indicating highly stable discrimination and flexible threshold selection without appreciable loss of accuracy.

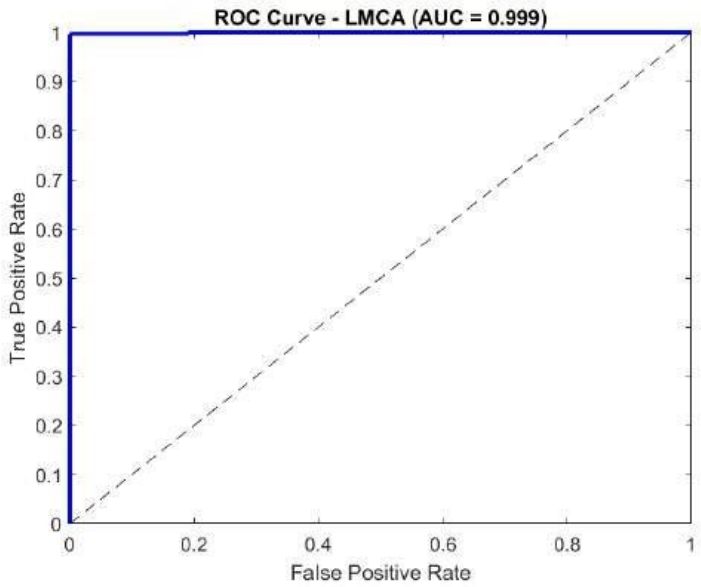


Figure 37: ROC curve for stenosis detection in the proximal LCA segment.

3.5.2 LAD Stenosis Detection

The second confusion matrix refers to the Left Anterior Oblique (LAD) vessel of the LCA, which is represented by the second output neuron. In Figure 38 for the 519 non-stenosis frames the model did not appear to have any false negatives. For stenosis cases, the model correctly detected 1,329 out of 1,332, with 3 false positives. The true positives and true negatives suggest the capability of the model to detect stenosis while the higher rate of false positives means that a small number of mid segment stenosis were missed, possibly due to more subtle or overlapping vessel appearances.

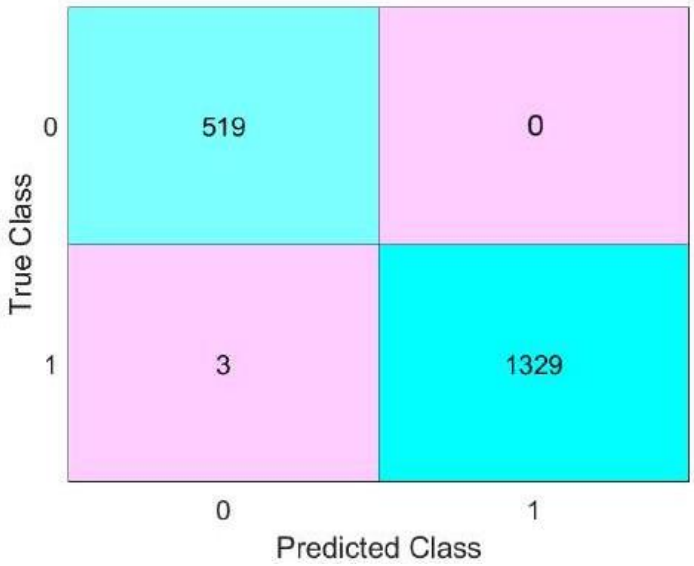


Figure 38: Confusion matrix for stenosis detection in the mid LCA segment.

The curve aligns with the top border, implying perfect ranking of positives above negatives for the left anterior descending (LAD) segment. This reflects separability across thresholds, consistent with an effectively error-free classifier within the evaluated data.

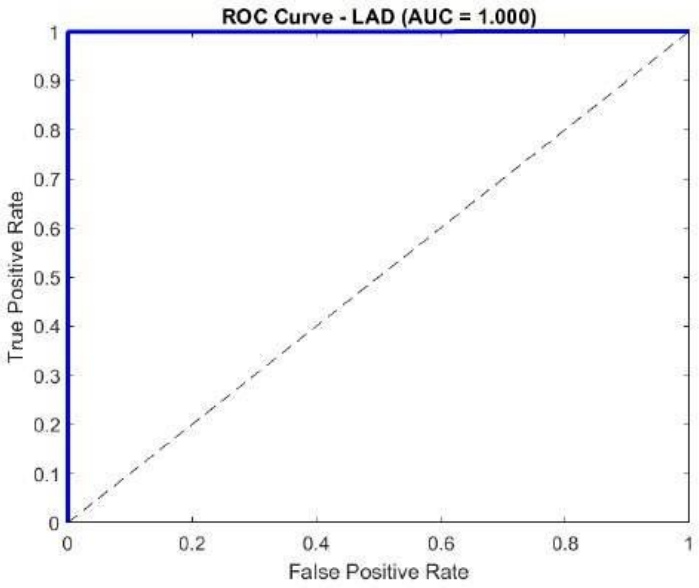


Figure 39: ROC curve for stenosis detection in the middle LCA segment.

3.5.3 LCx Stenosis Detection

The third confusion matrix corresponds to the Left Circumflex vessel (LCx) of the LCA, so the Figure 40 represents the classification results of the networks third output neuron. Out of 740 non-stenosis frames, 737 were correctly classified, and 3 were false negatives. For stenosis cases, 1,107 out of 1111 were correctly detected, with 4 false positives. Performance here remains very high, though the slightly increased number of misclassifications compared to the proximal region suggests that distal stenoses, potentially smaller or with more subtle angiographic features.

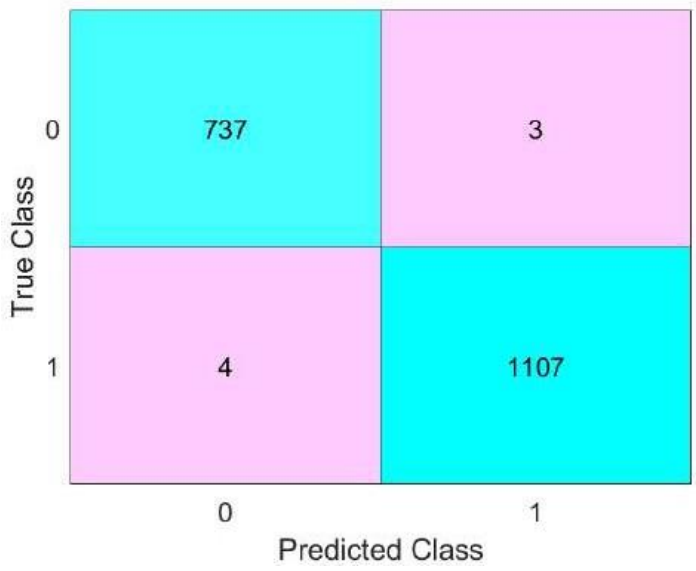


Figure 40: Confusion matrix for stenosis detection in the distal LCA segment.

The ROC curve 41 is pinned to the y-axis and upper boundary for almost the entire range of thresholds, indicating near-perfect discrimination for left circumflex (LCx) stenosis. A very high sensitivity is achieved at false positive rates.

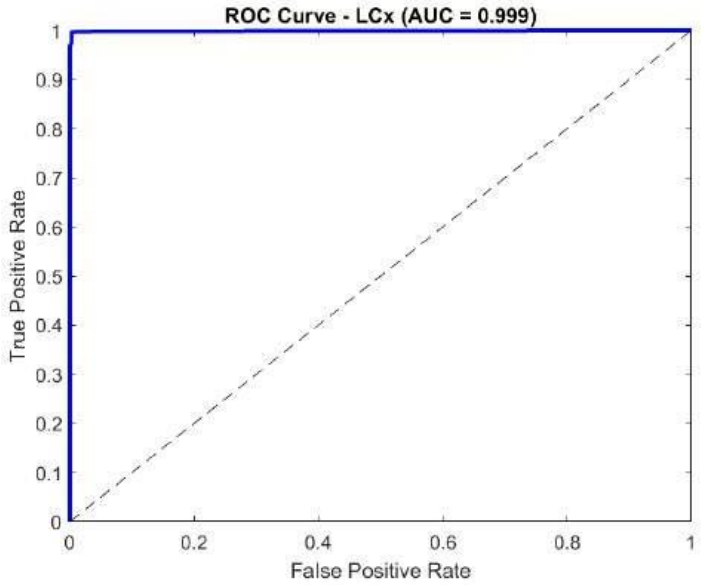


Figure 41: ROC curve for stenosis detection in the distal LCA segment.

3.5.4 Overall results for the 3-class LCA stenosis detection

Table 10 summarizes the stenosis detection metrics for each one of the main 3 LCA vessels. It can be observed that the model achieved excellent performance in detecting stenosis across all three LCA vessels. Precision, recall, and F1-scores were consistently above 0.99, with an average precision of 99.9%, recall of 99.5%, and F1-score of 99.7%. These results demonstrated the robustness of the classifier in detecting stenosis in the three different regions of the left artery.

Coronary Vessel	Neuron	Precision%	Recall%	F1-score%
LCMA	1	100	99.2	99.6
LAD	2	100	99.8	99.9
LCx	3	99.7	99.5	99.7
Average		99.9%	99.5%	99.7%

Table 10: Performance metrics (precision, recall, and F1-score) for the three LCA output neurons.

Figure 42 presents typical examples of correct and incorrect stenosis detections. The rows correspond to anatomical segments LMCA, LAD, and LCx and columns depict the four outcome categories (TP, FN, FP, TN). True-positive examples display stenosis in the vessels that align with the predictions. False-negative frames contain a lesion that is missed, typically under sub-optimal imaging conditions (e.g., foreshortening, overlap with guidewires, low contrast, or motion). Specif- ically, for LMCA and LAD no false positives were observed ("Zero missed"), consistent with their near-perfect ROC performance. The LCx row includes an isolated false positive, where contrast patterns might mimic a focal stenosis. Overall the LMCA and LAD achieve essentially error-free

discrimination, while LCx errors are rare and predominantly arise from anatomical/contrast arti- facts rather than systematic misclassification.

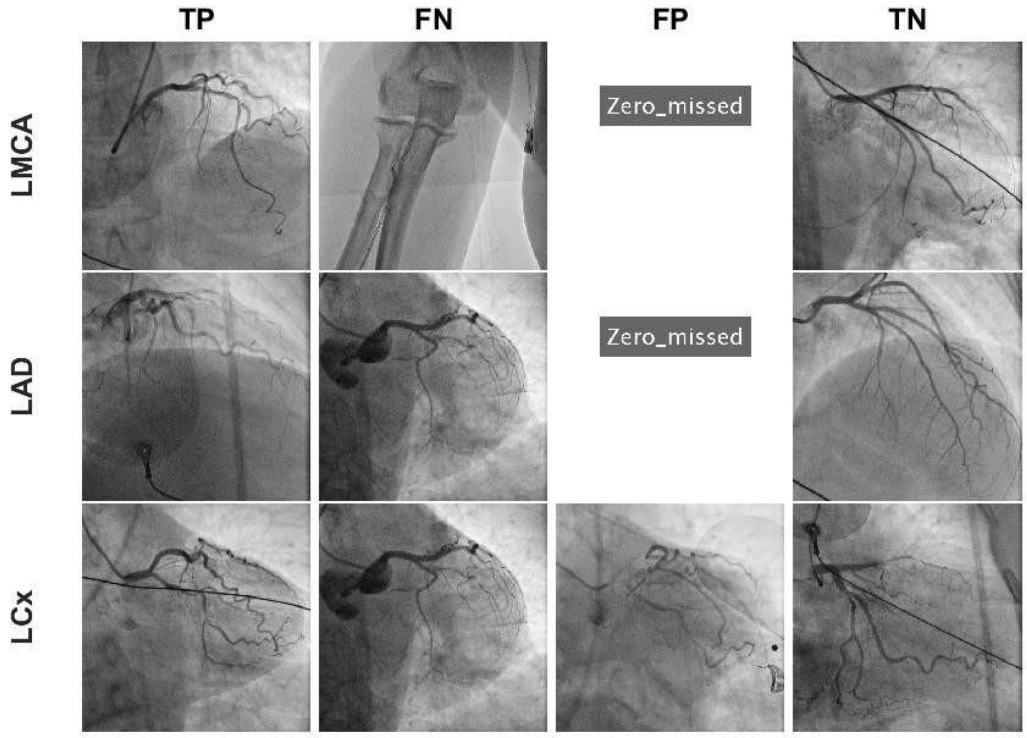


Figure 42: Typical results of correct and wrong classifications (TP, FN, FP, TN) for each binary class i.e. anatomical vessel.

3.6 RCA stenosis detection

In this subsection, stenosis is detected in the 3 main anatomical regions of the Right Coronary Artery (RCA Proximal, RCA Middle, RCA Distal), as a 3-class multi-label problem. Consequently, the output layer of the CNN consists of 3 neurons.

Results are presented for the test subset, which has been split per patient. In the confusion matrices of the following subsections, the class labels follow the same binary definition: 0 denotes the absence of stenosis (positive) and 1 denotes the presence of stenosis (negative). Correct predictions are aligned along the diagonal (true negatives for class 0 and true positives for class 1), while errors are captured in the off-diagonal entries (false positives and false negatives).

3.6.1 RCA Proximal Segment stenosis detection

The confusion matrix in Figure 43 refers to the proximal segment of RCA, which corresponds to the first output neuron. Out of 217 non-stenosis cases, 213 were correctly identified and only 4 were

misclassified as stenosis. For stenosis cases, the model achieved correct classifications with 68 out of 78. These results indicate a very high sensitivity and precision in detecting stenosis.

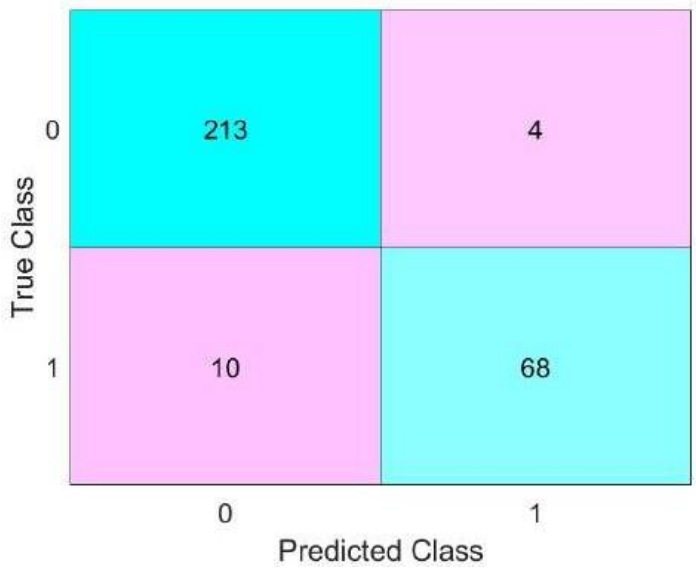


Figure 43: Confusion matrix for stenosis detection in the proximal RCA segment.

The ROC exhibits a steep initial ascent and remains well above the chance diagonal, indicating excellent discriminative ability. Sensitivity above 0.90 is attainable at low false-positive rates (0.05–0.10), suggesting reliable detection of proximal RCA lesions without a large increase in false alarms.

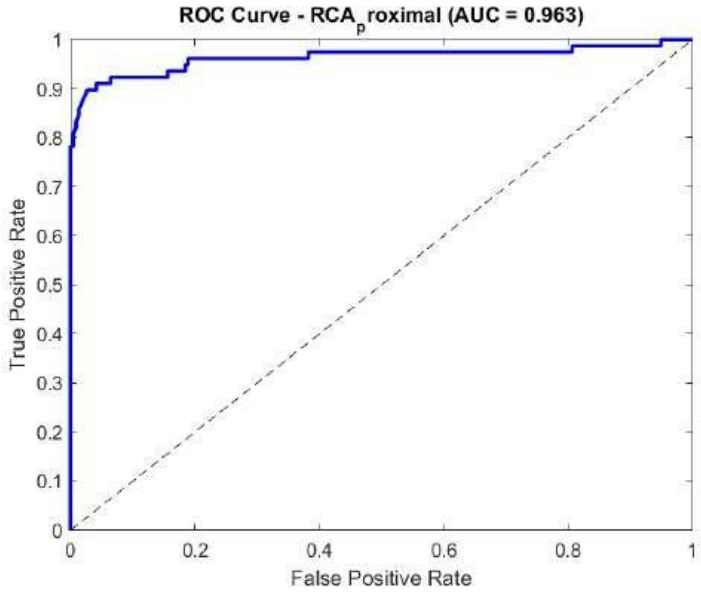


Figure 44: ROC curve for stenosis detection in the proximal RCA segment.

3.6.2 RCA Mid Segment stenosis detection

The second confusion matrix Figure 45 represents the classification results of the networks second output neuron, which refers to the mid segment of RCA and the identification of stenosis. For non-stenosis cases, the model correctly detected 213 of non-stenosis (true positives), while the number of mis-classified of non-stenotic frames for stenosis was only 7. The stenosis frames are predicted correctly in 70 out of 75 frames. The number of false predictions is very low, resulting in increased precision and recall.

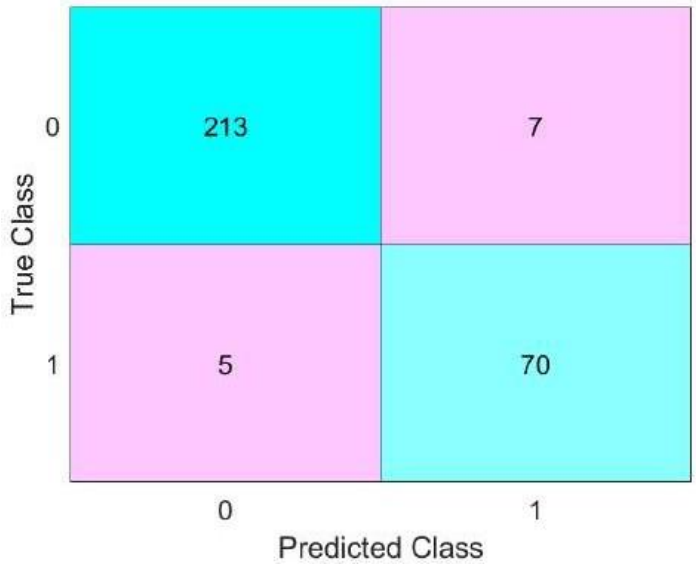


Figure 45: Confusion matrix for stenosis detection in the mid RCA segment.

This curve aligns with the top border for most thresholds, reflecting near-ceiling performance and the best separability among the three segments. Very high sensitivities are sustained while keeping the false-positive rate minimal, indicating a robust and stable classifier for mid-segment disease

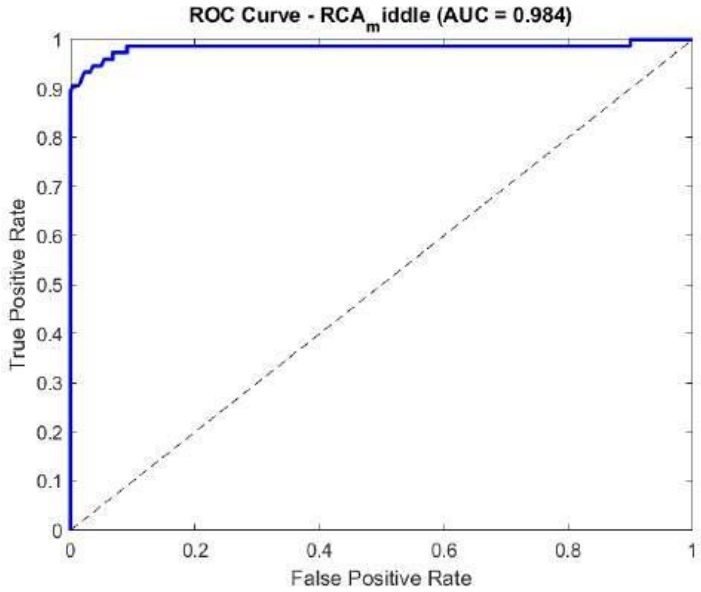


Figure 46: ROC curve for stenosis detection in the middle RCA segment.

3.6.3 RCA Distal Segment stenosis detection

The third output neuron corresponds to the distal RCA segment. Figure 47 represents the classification results of the networks third output neuron. The results showed that 130 of non-stenosis cases and 141 stenosis cases were correctly classified. The wrongly classified frames included 12 false negative and 12 false positives. These results are balanced but the misclassifications are relatively increased compared to the middle segment, which indicates that the distal segment stenosis of RCA is harder to predict, probably due to the anatomy of the heart.

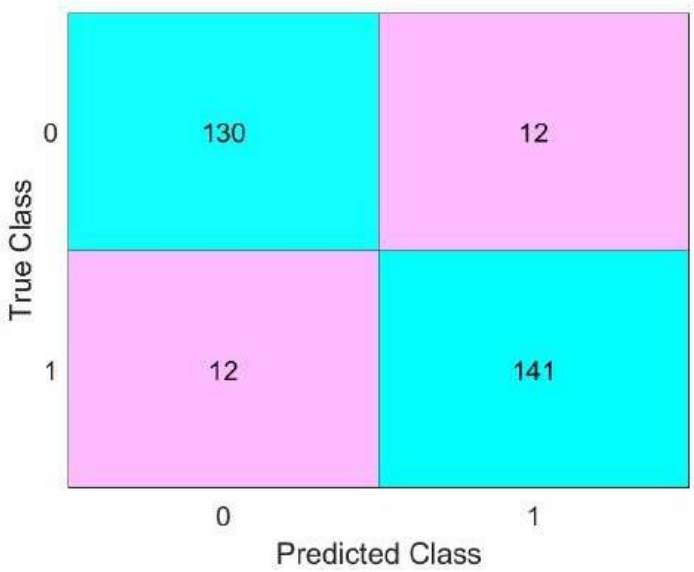


Figure 47: Confusion matrix for stenosis detection in the distal RCA segment.

The ROC curve exhibits a small early elbow, due to a stricter-threshold trade-off (some initial false negatives), but the curve rapidly approaches the upper boundary, achieving near-perfect sensitivity with only modest growth in false positives.

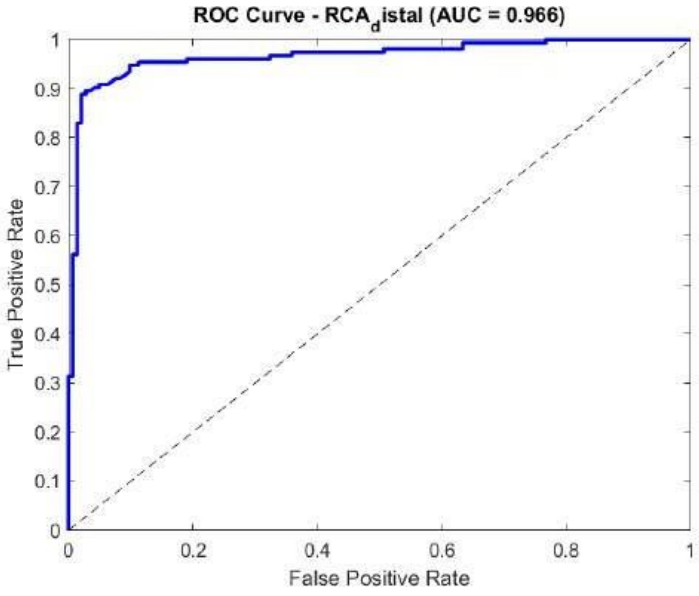


Figure 48: ROC curve for stenosis detection in the distal RCA segment.

3.6.4 Overall results for the 3-class RCA stenosis detection

Table 11 summarizes the stenosis detection metrics for each of the 3 anatomic regions of the RCA. According to the results, the three output neuron model for the right coronary artery (RCA) achieved a very high performance across all segments. The model achieved a precision of 92.5%, recall 90.9% and a F1-Score of 91.6% indicating a strong ability to recognize stenosis cases.

Table 11: Performance metrics (precision, recall, and F1-score) for the three RCA output neurons.

Coronary Vessel	Neuron	Precision%	Recall%	F1-score%
RCA proximal	1	94.4	87.2	90.7
RCA mid	2	90.9	93.3	92.1
RCA distal	3	92.2	92.2	92.2
Average		92.5%	90.9%	91.6%

The montage in Figure 49 presents representative qualitative outcomes of the right coronary artery (RCA) classifier across anatomical segments and decision outcomes. Rows correspond to the three RCA segments (proximal, middle, distal), while columns correspond to the four confusion- matrix categories (true positive, false negative, false positive, true negative). True-positive examples show clear narrowing within the annotated segment, with preserved background contrast and limited vessel overlap. False-negative cases contain a lesion that is not detected maybe because of guidewire overlap, reduced contrast, or motion blur. False-positive frames lack clinically relevant focal stenosis but may mimic one (vessel crossings). True-negative examples display vessels that stenosis was not detected.

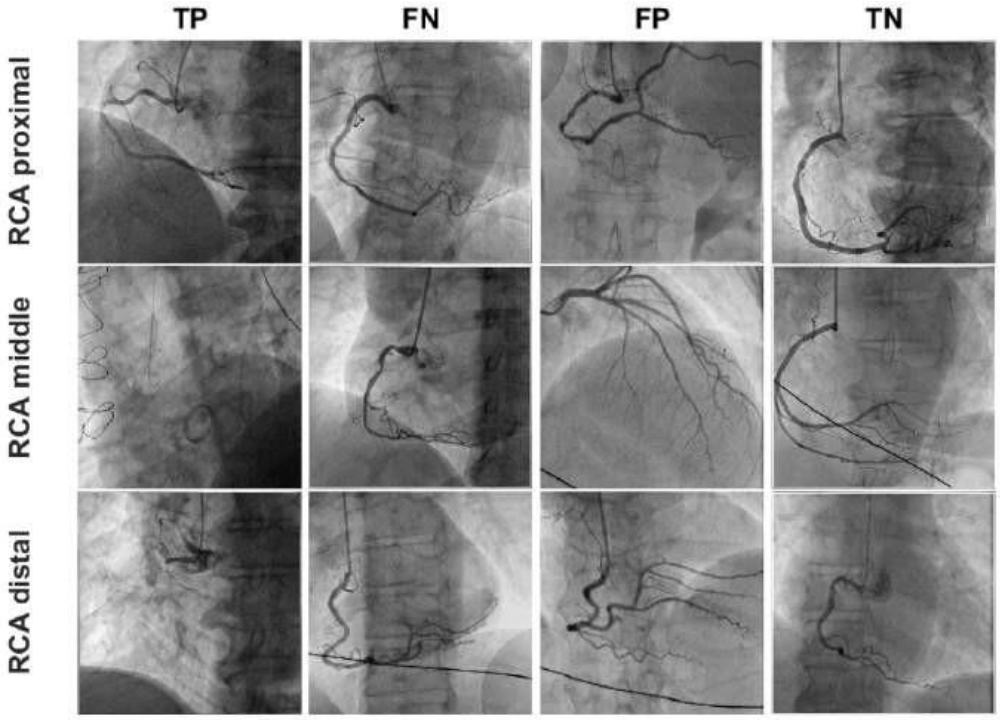


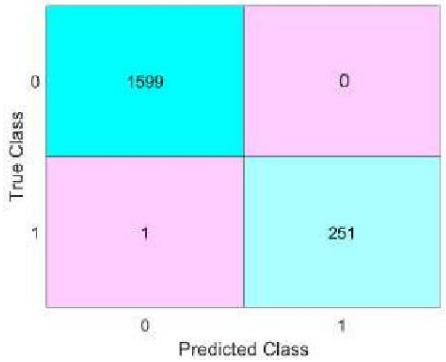
Figure 49: Typical results of correct and wrong classifications (TP, FN, FP, TN) for each binary class i.e. anatomical vessel.

3.7 LCA Stenosis detection in 8 anatomical regions

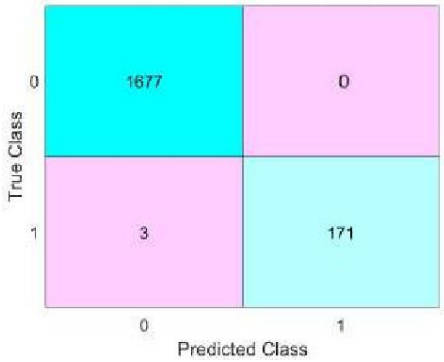
In this subsection, we refine stenosis detection in LCA considering 8 anatomical regions, namely LCMA proximal, LCMA middle, LAD proximal, LAD middle, LAD distal, LCx proximal, LCx middle, LCx distal). The problem is formulated as an 8-class multi-label problem. Consequently, the output layer of the CNN consists of 8 neurons.

Results are presented for the test subset, which has been split per patient. Figure 50 shows the confusion matrices for each of the 8 binary classes, with 0 representing the absence of stenosis (negative), and 1 representing the presence of stenosis (positive) in the respective segment. Each neuron corresponds to a specific anatomical subdivision of the LCA, and the confusion matrices allow evaluation of the model's performance for each region independently.

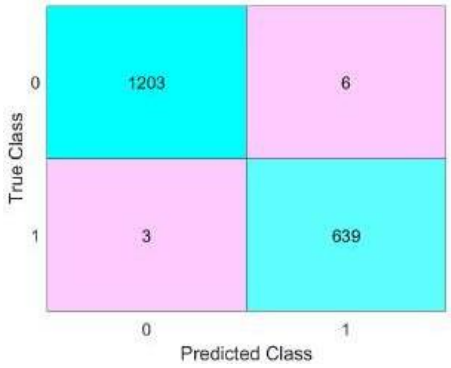
Figure 50 demonstrates consistently strong performance of the stenosis detection model. In nearly all segments, the classifier achieves perfect separation between stenotic and non-stenotic cases, with almost no false positives and no false negatives.These results indicate that the model achieves extremely high sensitivity and specificity across all segments of the LCA, ensuring reliable identification of stenotic lesions while minimizing the risk of misdiagnosis.



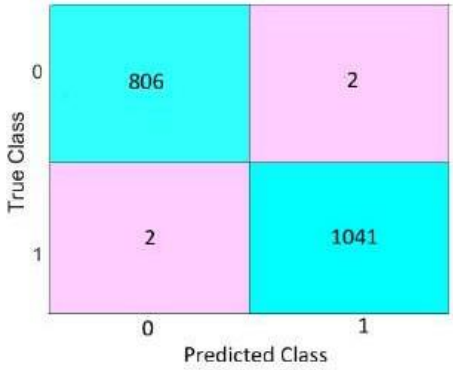
(a) LCMA proximal



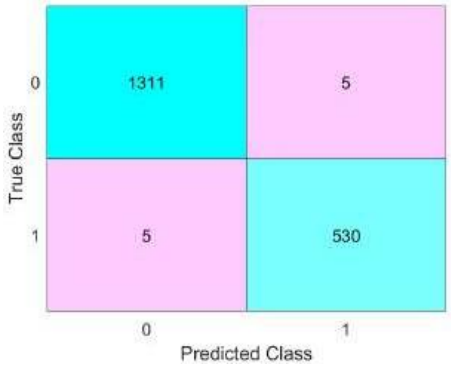
(b) LCMA middle



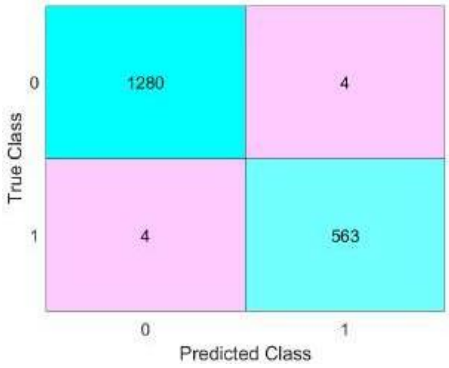
(c) LAD proximal



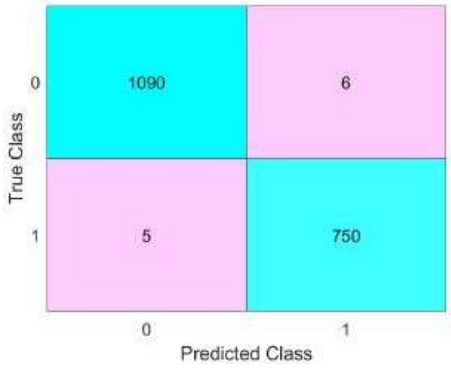
(d) LAD middle



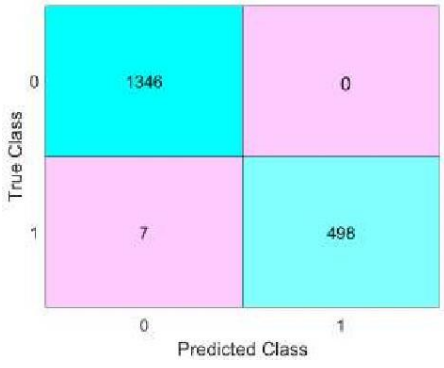
(e) LAD distal



(f) LCx proximal



(g) LCx middle



(h) LCx distal

Figure 50. Confusion matrices of LCA stenosis (1) vs non-stenosis (0) with 8 neurons.

The classification performance across all eight classes is very high, with most classes achieving perfect precision, recall and F1-scores according to the table 12. Neurons 1,2 and 8 reached 1.000 in recall indicating that the model identified non-stenosis cases without missing any while the stenosis cases where nearly perfect. Labels 3,4,5,6 and 7 also achieve near perfect results, with only a marginal difference in precision or recall, suggesting minimal misclassifications. But nonetheless their F1-score remained very strong which indicates that the balance between precision and recall is very high. The overall accuracy indicates that the model maintains consistently high accuracy across all classes without a bias to any category.

Coronary Vessels	Neuron	Precision%	Recall%	F1-score%
LCMA proximal	1	99.6	100	99.8
LCMA mid	2	98.2	100	99.09
LAD proximal	3	99.5	99.06	99.2
LAD mid	4	99.8	99.8	99.8
LAD distal	5	99.06	99.06	99.06
LCx proximal	6	99.2	99.2	99.3
LCx mid	7	99.3	99.2	99.2
LCx distal	8	98.6	100	99.2
Macro avg		**99.1**	**99.5**	**99.3**

Table 12: Per-class precision, recall and F1-score (one-vs- rest).

Comparing it with the LCA model of 3 neurons the model performance remained equally high despite finer segmentation, demonstrating robustness to granularity of annotation.

The panel in Figure 51 summarizes representative frames for each left-coronary segment—LMCA (proximal, middle), LAD (proximal, middle, distal), and LCx (proximal, middle, distal) arranged by prediction outcome (TP, FN, FP, TN). True-positive examples exhibit a clear stenosis in vessels within the annotated segment. False-negative cases contain a lesion that is missed, typically under unclear viewing conditions (foreshortening, guidewire overlap, low contrast). False positives are less common in LCMA and LCx Distal segments (”Zero missed”) while in the rest anatomic parts image quality and other observations might mimic a stenosis for the remaining segments. True-negative frames represent vessels without any stenosis which the model classified correctly.

Some of the representative results are as follows.

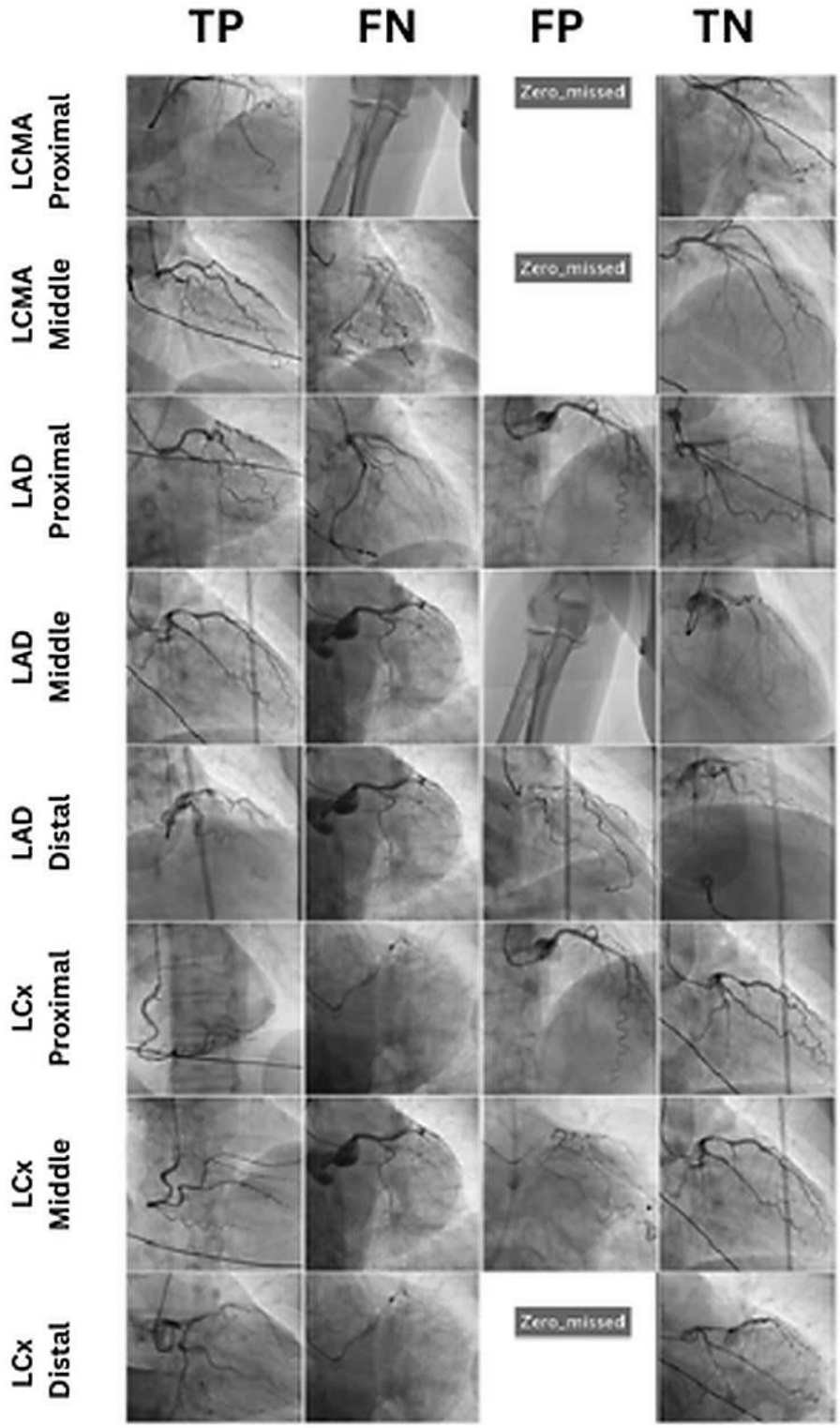


Figure 51: Typical results of correct and wrong classifications (TP, FN, FP, TN) for each binary class i.e. anatomical vessel.

4 Conclusions and Discussion

This study presents a fully automated pipeline for the analysis of coronary angiography videos, evaluated on 11,929 videos from 767 patients, addressing multiple clinically relevant tasks: automatic frame selection, anatomical classification of the left and right coronary arteries, detection of interventional events, such as balloon inflation and stenosis detection and localization.

The methodologies and deep learning models that were used, demonstrated high performance across the tasks. The AlexNet-based artery classification model achieved an AUC of 99.1% for distinguishing between LCA and RCA. Balloon detection with ViT-Base appeared to have strong accuracy, confirming the suitability of transformer-based approaches for identifying procedural events. Stenosis detection achieved AUCs between 96.3% and 98.%4 across RCA segments, while in LCA segments had similarly very high values reaching 99.1% macro precision despite its challenging anatomy. Importantly, we trained models at the frame level and applied video-level majority vot- ing during testing, which reduced frame-wise misclassifications and provided clinically meaningful video-level predictions.

In prior research, individual tasks of coronary angiography interpretation have been explored in isolation. For frame selection, [4] proposed classical filter-based methods and later introduced Deep AngioKey [9], achieving up to 98.1% accuracy. In contrast, our frame selection method combines vesselness filtering, structural similarity, and a complexity-based Frame Score, ensuring that only diagnostically relevant frames are retained before classification and detection. Segmentation approaches such as SE-RegUNet [14], DR-LCT-UNet [13], and TVS-Net [15] achieved Dice scores ranging from 72% to 86% on coronary vessel masks. They introduced state of the art methods but they were based on manual annotations and often operated on static frames and CCTA data. In this thesis we focus on providing higher-level diagnostic outputs that are directly applicable in real clinical workflows.

Anatomical classification of LCA vs. RCA was included in CathAI [18], but bounding-box localization had limited accuracy (48% mAP). Our AlexNet-based model achieved stronger discrimination, with near-perfect AUC.

Balloon detection is an important contribution to these studies. To our knowledge, no prior studies explicitly addressed procedural event detection in ICA frames and videos. Our results with ViT-Base demonstrate that transformer-based architectures can reliably capture changes associated with balloon inflation, opening avenues for real-time interventional support.

Finally, stenosis detection has been addressed in studies such as [16], who applied object detection networks achieving mAP up to 95%, and [17], that reported 90% accuracy with DenseNet201. However, these relied on manually curated frames with bounding-box annotations. Similarly, AI-QCA [19] and [20] introduced methods for stenosis quantification in CCTA, requiring manual frame or vessel selection. Our pipeline performs stenosis detection and localization (only the category of the vessel), achieving comparable AUCs.

Key strengths of this study include the use of a large and diverse dataset of ICA videos and end-to-end automation without relying on manual frame selection. By focusing on multiple frame analysis, our approach aligns closely with real-world clinical practice.

Our models achieved strong performance across tasks, with AlexNet reaching an AUC of 99.1% for LCA/RCA classification, ViT-Base balloon recognition with high accuracy and stenosis detection achieving AUCs of 96.3%–98.4% across coronary segments.

Compared to previous studies, this work is among the first to provide a multi-task, video-based pipeline for ICA. The results highlight its potential for real-time decision support, reducing inter-

pretation time and variability in the catheterization laboratory. Overall, this study demonstrates the promise of automated ICA video analysis as a valuable tool for supporting clinicians in the diagnosis and treatment of coronary artery disease.

This work also has limitations. First, the dataset originates from a single center so more data are needed to ensure generalizability. Second, while balloon detection was demonstrated successfully, further validation in larger procedural datasets is required. Third, the ground truth annotations for stenosis were based on angiographic interpretation and clinical reports given by the hospital and binary codes we generated, which may not fully align with other methods of diagnosis. Finally, most of our models operate on frames and use majority voting for video-level classifications.

Future research should extend validation to multi-center datasets. The methodologies and deep learning models could be adapted to train and inference on video sequences, instead of single frames, which would alleviate the need for frame selection. Deploying a software that would offer the whole, or part of the pipeline as a service could enable diagnostic and interventional support, erasing the gap between computational methods and practical clinical application. Finally, the proposed pipeline and its architecture could be modified for real-time application during the interventional clinical procedure.

Abbreviations

Abbreviation	Definition
AP	Anterior Posterior
AI-QCA	AI-based Quantitative Coronary Angiography
CAD	Coronary Artery Disease
CMBR	Curved Multiplanar Reformats
CCTA	Coronary Computed Tomography Angiogram
CRA	Cranial View
CAU	Caudal View
DICOM	Digital Imaging and Communications in Medicine
DSC	Dice Similarity Coefficient
GBDT	Gradient Boosting Decision Tree
GDPR	General Data Protection Regulation
ICA	Invasive Coronary Angiography
LCA	Left Coronary Artery
LCMA	Left Coronary Main Artery
LAD	Left Anterior Descending
LCx	Left Circumflex
MMPs	Modality Performed Procedure Step
MLD	Minimum Lumen Diameter
RCA	Right Coronary Artery
PDA	Posterior Descending Artery
PII	Personally Identifiable Information
SE-RegUNet	Squeeze-and-Excitation RegUNet
SSIM	Structural Similarity Index
SGDM	Stochastic Gradient with Momentum
TVS-Net	Temporal Vessel Segmentation Network

Table 13: Glossary of abbreviations used in coronary imaging analysis.

References

[1] StatPearls Publishing. *Anatomy, Thorax, Coronary Arteries*. Available from https://www.ncbi.nlm.nih.gov/books/NBK534790/. Treasure Island (FL): StatPearls Publishing, 2024.

[2] Arso M. Vukicevic et al. "Three-dimensional reconstruction and NURBS-based structured meshing of coronary arteries from the conventional X-ray angiography projection images". In: *Scientific Reports* 8.1 (2018), p. 1711. DOI: 10.1038/s41598-018-19440-9.

[3] Jungsu Kim et al. "Development of diagnostic reference levels using a real-time radiation dose monitoring system at a cardiovascular center in Korea". In: *Journal of Digital Imaging* (2015). DOI: 10.1007/s10278-015-9773-9.

[4] Hounaida Moalla et al. "Automatic Coronary Angiogram Keyframe Extraction". In: *Proceed- ings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*. SciTePress, 2023, pp. 582–589. DOI: 10.5220/0011850700003411.

[5] John Canny. "A Computational Approach to Edge Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (1986), pp. 679–698. DOI: 10 .1109 / TPAMI.1986.4767851.

[6] Yoshinobu Sato et al. "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images". In: *Medical Image Analysis* 2.2 (1998), pp. 143–168. DOI: 10.1016/S1361-8415(98)80009-1.

[7] Alejandro F. Frangi et al. "Multiscale vessel enhancement filtering". In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*. Vol. 1496. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 1998, pp. 130–137. DOI: 10.1007/BFb0056195.

[8] Erik H. W. Meijering, Wiro J. Niessen, and Max A. Viergever. "Design and validation of a tubularity measure for enhanced skeletonization of vascular structures in 3D". In: *IEEE Transactions on Medical Imaging* 20.12 (2001), pp. 1221–1230. DOI: 10.1109/42.974918.

[9] Houna
¨ıda Moalla et al. *Deep AngioKey: A Novel Approach for Objective Keyframe Extraction in Coronary Angiography Analysis*. Preprint, Research Square. Version Version 1. Posted on Research Square, not peer-reviewed. Oct. 2023. DOI: 10.21203/rs.3.rs-3498564/v1.

[10] SYNTAX Score Consortium. *SYNTAX Score Tutorial: Definitions*. https://syntaxscore.org/index.php/tutorial/definitions. Accessed: 2025-09-15. 2025.

[11] G. Sianos et al. "The SYNTAX Score: an angiographic tool grading the complexity of coronary artery disease". In: *EuroIntervention* 1.2 (Aug. 2005), pp. 219–227.

[12] Zijun Gao et al. "Vessel segmentation for X-ray coronary angiography using ensemble methods with deep learning and filter-based features". In: *BMC Medical Imaging* 22.1 (Jan. 2022), p. 10. DOI: 10.1186/s12880-022-00734-4.

[13] Qianjin Wang et al. "Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer". In: *Frontiers in Physiology* 14 (Aug. 2023), p. 1138257. DOI: 10.3389/fphys.2023.1138257.

[14] Shih-Sheng Chang et al. "Optimizing ensemble U-Net architectures for robust coronary vessel segmentation in angiographic images". In: *Scientific Reports* 14 (Mar. 2024), Article 6640. DOI: 10.1038/s41598-024-57198-5.

[15] H. He et al. “Deep learning based coronary vessels segmentation in X-ray angiography using temporal information”. In: *Medical Image Analysis* 102 (Feb. 2025). DOI: 10.1016/j.media.2025.103496.

[16] Viacheslav V. Danilov et al. “Real-time coronary artery stenosis detection based on modern neural networks”. In: *Scientific Reports* 11 (Apr. 2021), Article 7582. DOI: 10.1038/s41598-021-87174-2.

[17] Şerife Kaba et al. “The Application of Deep Learning for the Segmentation and Classification of Coronary Arteries”. In: *Diagnostics* 13.13 (July 2023), p. 2274. DOI: 10.3390/diagnostics13132274.

[18] Robert Avram et al. “CathAI: fully automated coronary angiography interpretation and stenosis estimation”. In: *npj Digital Medicine* 6 (Aug. 2023). Published 11 August 2023, Article 142. DOI: 10.1038/s41746-023-00880-1.

[19] Young In Kim et al. “Artificial intelligence-based quantitative coronary angiography of major vessels using deep-learning”. In: *International Journal of Cardiology* 405 (June 2024), p. 131945. DOI: 10.1016/j.ijcard.2024.131945.

[20] Vibha Gupta et al. “End-to-end deep-learning model for the detection of coronary artery stenosis on coronary CT images”. In: *Open Heart* 12.1 (Jan. 2025), e002998. DOI: 10.1136/openhrt-2024-002998.

[21] Innolitics, LLC. *DICOM Standard Browser — X-Ray Angiographic Image IOD*. 2025. URL: https://dicom.innolitics.com/ciods/x- ray- angiographic- image/xa- positioner (visited on 09/03/2025).

[22] Spiros V Georgakopoulos et al. “Improving the performance of convolutional neural network for skin image classification using the response of image analysis filters”. In: *Neural Computing and Applications* 31.6 (2019), pp. 1805–1822.

[23] George Nousias et al. “Video-based eye blink identification and classification”. In: *IEEE Journal of Biomedical and Health Informatics* 26.7 (2022), pp. 3284–3293.

[24] George Nousias, Konstantinos K Delibasis, and Georgios Labiris. “Blink Detection Using 3D Convolutional Neural Architectures and Analysis of Accumulated Frame Predictions”. In: *Journal of Imaging* 11.1 (2025), p. 27.

[25] Maria Katranzopoulou, Konstantinos Delibasis, and Ilias Maglogiannis. “Class Imbalance and Data Splitting Approaches for Face Image Pain Classification: Survey and Experimental Comparisons”. In: *Data Science in Applications: Towards AI-Driven Approaches*. Springer, 2025, pp. 165–217.

[26] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc., 2012, pp. 1097–1105.

[27] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[28] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *International Conference on Learning Representations (ICLR)* (2015). arXiv: 1409.1556 [cs.CV].

[29] Jia Deng et al. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[30] Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *International Conference on Learning Representations (ICLR)* (May 2021). arXiv:2010.11929. DOI: 10.48550/arXiv.2010.11929.

[31] Kan Wu et al. “TinyViT: Fast Pretraining Distillation for Small Vision Transformers”. In: *arXiv preprint* (July 2022). arXiv:2207.10666. DOI: 10.48550/arXiv.2207.10666.

[32] I. Maglogiannis and K. K. Delibasis. “Enhancing classification accuracy utilizing globules and dots features in digital dermoscopy”. In: *Computer Methods and Programs in Biomedicine* 118.2 (2015), pp. 124–133. DOI: 10.1016/j.cmpb.2014.11.010.

[33] Tim Jerman. *Jerman Enhancement Filter: Enhancement of Vessel/Tubular and Blob/Spherical Structures in 2D/3D Images Using Hessian Eigenvalues*. https:// www. mathworks. com/ matlabcentral/ fileexchange/ 63171 - jerman- enhancement- filter. MATLAB Central File Exchange. Retrieved September 2025. 2017.

[34] Carsten Steger. “An unbiased detector of curvilinear structures”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.2 (1998), pp. 113–125. DOI: 10 . 1109 / 34 . 659930.

[35] Zhou Wang et al. “Image quality assessment: From error visibility to structural similarity”. In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612. DOI: 10.1109/TIP.2003.819861.

[36] Mayank Yadav et al. “X-ray Coronary Angiogram images and SYNTAX score to develop Machine-Learning algorithms for CHD Diagnosis”. In: *Scientific Data* 12.1 (2025), p. 104. DOI: 10.1038/s41597-025-04727-0.

Automated image and video analysis for Coronary Angiography

Charikleia Gkrantouni

Judge committee

1 Delimpasis Konstantinos
2 Tasoulis Sotirios
3 Plagianakos Vasilios

Κατηγοροποίση εικόνων και βίντεο αγγειογραφίας

Γκραντούνη Χαρίκλεια

Τριμελής επιτροπή

1 Δελήμπασης Κωνσταντίνος
2 Τασουλής Σωτήριος
3 Πλαγιανάκος Βασίλειος