scientific reports

Check for updates

OPEN

# Patient-level CAD-RADS scoring from coronary radiomic features

Anna Corti[1]✉, Francesca Lo Iacono[1], Francesca Ronchetti[2], Mattia Chiesa[2,3], Gianluca Pontone[2,5], Gualtiero I. Colombo[4] & Valentina D. A. Corino[1,2]

**Synthesizing coronary radiomic data to obtain a single patient-wise Coronary Artery Disease-Reporting and Data System (CAD-RADS) score remains challenging. This work proposes four strategies for summarizing radiomic features extracted from 2779 multiplanar reconstruction images derived from coronary computed tomography angiography of 238 patients. A cascade pipeline was developed to train gradient boosting classifiers for CAD-RADS scoring through consecutive tasks, considering 80%-20% training/test split with five-fold cross-validation on the training set. Two statistical-based and two majority voting approaches were implemented to obtain patient-level classification. The former consisted in computing features average, minimum, maximum and standard deviation, across the coronary images, leading to intermediate coronary classification, followed by patient classification according to the worst coronary class. The latter consisted in single image predictions and the application of majority voting either to all the images, to obtain patient classification (MV_P), or to the images of single coronary arteries, followed by patient classification according to the worst coronary class (MV_C). Majority-voting approaches outperformed statistical-based ones, with MV_P achieving an AUC of CAD-RADS_0 = 0.94, CAD-RADS_1 = 0.92, CAD-RADS_2 = 0.97, CAD-RADS_3 = 0.77, CAD-RADS_4 = 0.88, CAD-RADS_5 = 0.85, and MV_C of CAD-RADS_0 = 0.82, CAD-RADS_1 = 0.78, CAD-RADS_2 = 0.84, CAD-RADS_3 = 0.96, CAD-RADS_4 = 0.98 and CAD-RADS_5 = 0.85. This study represents a significant advancement toward robust and reproducible coronary radiomics tools for automated CAD-RADS scoring.**

**Keywords** Coronary artery disease (CAD), Stenosis, Coronary computed tomography angiography (CCTA), Radiomics, CAD-RADS, Machine learning, Atherosclerotic plaque

With about 8.9 million annual deaths, coronary artery disease (CAD) represents the major cause of morbidity and mortality worldwide[1]. CAD is usually caused by atherosclerosis, which consists of the formation of fibro-fatty plaques in the arterial wall of coronary arteries, leading to coronary lumen stenosis with subsequent impaired blood supply to the myocardium[2]. Recently, guidelines from the European Society of Cardiology and the National Institute for Health and Care Excellence have recommended coronary computed tomography angiography (CCTA) as the first-line imaging modality for CAD diagnosis in patients for whom CAD cannot be ruled out by clinical assessment alone[3]. To guide patient management and treatment decisions following CCTA, the Coronary Artery Disease-Reporting and Data System (CAD-RADS) was proposed, representing a standard approach for scoring CAD from CCTA[4]. CAD-RADS grades the severity of coronary stenosis into six categories, namely, class 0 (no stenosis: 0%), class 1 (minimal nonobstructive stenosis: 1–24%), class 2 (mild nonobstructive stenosis: 25–49%), class 3 (moderate stenosis: 50–69%), class 4 (severe stenosis: 70–99%), and class 5 (total occlusion: 100%). Today, CAD-RADS scoring is performed manually by expert radiologists. In this context, automated tools for stenosis severity assessment and CAD-RADS scoring can be useful in supporting clinicians' practice and addressing the challenges of interobserver variability and time consumption. To this end, several image-based deep-learning models have been developed[5–11]. However, deep learning models often operate as "black boxes", and while various explainability techniques have been developed to shed light on their internal decision-making processes, the underlying criteria are frequently based on abstract latent representations that are not readily linked to clinically meaningful characteristics. This can limit the trust and adoption of such models in clinical settings, where transparent and explainable outcomes are essential. The same challenge applies to workflows involving deep feature extraction followed by machine learning classifiers, as the so-called "deep

[1]Department of Electronics, Information and Bioengineering, Politecnico Di Milano, Via Ponzio 34/5, 20133 Milan, Italy. [2]Department of Perioperative Cardiology and Cardiovascular Imaging, Centro Cardiologico Monzino IRCCS, Milan, Italy. [3]Bioinformatics and Artificial Intelligence Facility, Centro Cardiologico Monzino IRCCS, Milan, Italy. [4]Unit of Immunology and Functional Genomics, Centro Cardiologico Monzino IRCCS, Milan, Italy. [5]Department of Biomedical, Surgical and Dental Sciences, University of Milan, Milan, Italy. ✉email: anna.corti@polimi.it

features" are still abstract and lack direct interpretability. In contrast, radiomic features are explicitly defined and quantifiable image descriptors that reflect various aspects of tissue characteristics, such as heterogeneity, texture, shape, and intensity. This makes radiomics, namely the process of radiomic feature extraction and mining, inherently more interpretable, offering not only predictive value but also potential insights into the underlying biological mechanisms of disease. Such interpretability is particularly advantageous in clinical settings, where understanding the rationale behind a model's output is critical for informed decision-making. Radiomics has recently been applied to the cardiovascular field for noninvasive image biomarker identification[12–15]. In the context of CAD, radiomics has shown potential in identifying vulnerable plaques, detecting plaque composition, grading stenosis, or predicting major adverse cardiovascular events[16–19]. However, being a recent field of application, robust methodologies for coronary plaque radiomic analyses still need to be defined.

An open challenge is the translation of radiomic features to patient-level outcomes. Specifically, radiomic analyses usually consider a region of interest (ROI) on which features are extracted, which mostly consists of the coronary artery wall or the atherosclerotic plaque. However, patient-level information (e.g., occurrence of adverse cardiovascular events, need for revascularization) rather than coronary plaque-level information is more relevant in clinical practice. This requires the definition of methodologies to integrate plaque-level information for a patient-level prediction. Accordingly, here we developed and tested different approaches of radiomic feature summarization to predict a patient's disease severity, as indexed by CAD-RADS scoring, from CCTA images of the three main coronary arteries. To the best of the authors' knowledge, this represents the first radiomics study applied to coronary arteries dealing with the problem of integrating plaque information for a patient-level prediction. We developed and tested four methods to establish a robust and reproducible approach for implementing a radiomics-based machine learning model for predicting CAD severity.

## Methods

### Patient enrollment and image acquisition

*Patient dataset*

Two hundred and thirty-eight patients undergoing CCTA for CAD evaluation at Centro Cardiologico Monzino IRCCS (Milan, Italy) between 2016 and 2018 were included in the study[11]. CAD-RADS scoring was performed by two expert readers, certified with Level III of the European Association of Cardiovascular Imaging certification in CT, among a pool of 10 (with 5 to 10 years of experience) by following the Society of Cardiovascular Computed Tomography (SCCT) guidelines[20–22]. In case of disagreement a third reader was involved to achieve a consensus. The average time required to assign CAD-RADS score by each reader was of 16 ± 7 min. Table 1 details the dataset composition across the 6 CAD-RADS classes and patient clinical characteristics. This retrospective study was approved by the Ethics Committee of the IRCCS European Institute of Oncology and Monzino Cardiology Center (registration number: R1061/19-CCM 11 25) under the ethical standards of the Declaration of Helsinki. Data acquisition followed the General Data Protection Regulation of the EU. All patients signed the informed consent.

*CCTA scan protocol*

CCTA scans were acquired using Revolution CT (GE Healthcare, Milwaukee, IL), with acquisition and imaging protocols following the SCCT guidelines[23]: slice configuration 256 × 0.625, mm, gantry rotation time 270 ms, tube voltage 100 KVp, effective tube current 500 mAs. A 40% adaptive iterative reconstruction algorithm was used for imaging reconstruction (ASIR, GE Healthcare, Milwaukee, Wisconsin).

The analysis encompassed the three primary coronary arteries: left anterior descending (LAD), left circumflex (LCX), and right coronary artery (RCA), with a total of 373 coronary arteries. Table 2 shows the number of patients and coronary arteries included in the study. It is worth noting that, in patients with multiple arteries affected by stenosis, the arteries exhibited the same degree of severity. There were no cases in which different coronary arteries within the same patient fell into different CAD-RADS categories. For patients with CAD-RADS = 0 the images of all the arteries were available, except for two patients presenting with two rather than three coronary arteries. Differently, in patients with CAD-RADS > 0, only the coronary arteries with plaques were available.

Post-processing was applied to enhance contrast and obtain straightened vessels. Moreover, up to 8 multiplanar reconstruction (MPR) 2D images were obtained for each coronary artery by rotating 45 degrees along the vessel's longitudinal axis, leading to a dataset of 2779 images. The dedicated software (AW Server3.2

| Patients | All | Class 0 | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 |
|---|---|---|---|---|---|---|---|
| N | 238 | 45 | 41 | 40 | 42 | 46 | 24 |
| Age (y), mean ± SD | 60.6 ± 12.3 | 50.3 ± 13.3 | 57.4 ± 9.0 | 62.0 ± 12.5 | 64.8 ± 10.1 | 65.4 ± 10.0 | 66.8 ± 8.0 |
| Male, n (%) | 165 (70) | 20 (44) | 26 (63) | 32 (80) | 30 (71) | 35 (76) | 22 (92) |
| Hypertension, n (%) | 84 (35) | 8 (17) | 11 (27) | 17 (43) | 13 (31) | 22 (48) | 13 (54) |
| Hyperlipidemia, n (%) | 76 (32) | 7 (16) | 8 (19) | 9 (23) | 17 (40) | 21 (46) | 14 (58) |
| Diabetes, n (%) | 21 (9) | 0 (0) | 0 (0) | 2 (5) | 3 (7) | 4 (9) | 12 (50) |
| Smoker, n (%) | 45 (19) | 5 (11) | 5 (12) | 6 (15) | 10 (24) | 10 (22) | 9 (38) |
| Family history, n (%) | 86 (36) | 16 (36) | 17 (41) | 15 (38) | 17 (40) | 17 (37) | 4 (17) |

**Table 1**. Patient and image characteristics of the study population.

| Images | All | Class 0 | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 |
|---|---|---|---|---|---|---|---|
| Coronary arteries, n | 373 | 133 | 54 | 48 | 53 | 57 | 28 |
| Images, n | 2779 | 1164 | 271 | 350 | 414 | 456 | 224 |
| Images per patient, n | [2–24] | [16–24] | [2–16] | [3–19] | [6–16] | [8–24] | [8–16] |

**Table 2**. Image dataset.

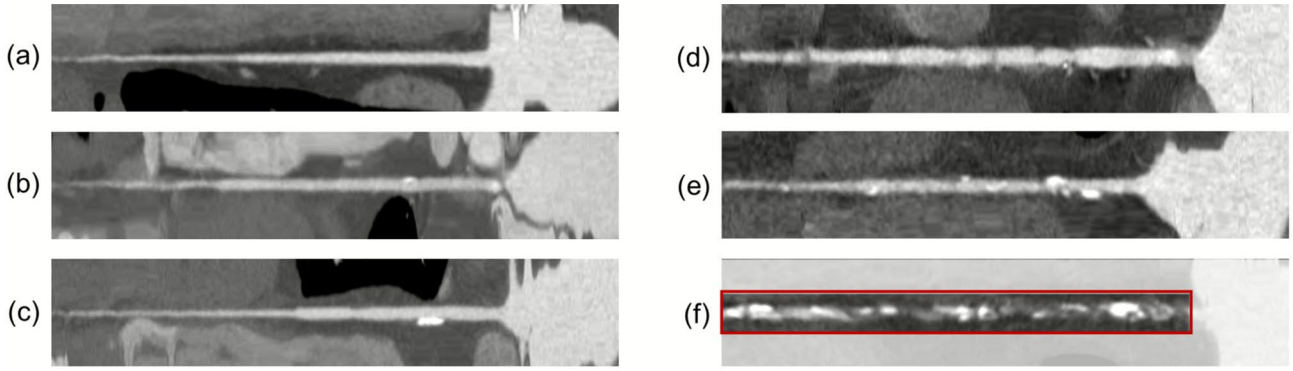


**Fig. 1**. Example of straightened coronary artery CCTA image belonging to CAD-RADS class 0 (**a**), class 1 (**b**), class 2 (**c**), class 3 (**d**), class 4 (**e**), class 5 (**f**) and example of the region of interest (ROI) for radiomic features extraction.

Ext 4.8, GE Healthcare, Milwaukee, Wisconsin) was used for MPR generation. Image quality was assessed for each coronary segment and categorized into four levels: *excellent* (no artifacts, fully interpretable), *good* (minor artifacts, but sufficient for a confident diagnosis), *adequate* (moderate artifacts, still acceptable for routine clinical evaluation), and *poor/non-evaluable* (severe artifacts preventing accurate interpretation). Segments rated as poor/non-evaluable were excluded from the analysis, regardless of plaque presence. Thus, only the views in which the plaque was clearly visible were saved by radiologists and available for analysis. Figure 1 shows an example of a MPR image for each CAD-RADS class.

*Proof of concept case study*

Although the primary dataset included patients whose coronary arteries shared the same CAD-RADS score, a preliminary proof-of-concept case study was conducted to assess the method's ability to classify coronary arteries with varying degrees of stenosis within the same patient. To this end, patients exhibiting multiple stenoses of different severity levels across different coronary arteries were included. Only five patients presented this characteristic and were analyzed. The patients underwent the same clinical and image acquisition protocol, as described in Sect. "CCTA scan protocol". In these cases, the CAD-RADS score was clinically assigned considering the highest degree of stenosis observed. Four patients presented with one coronary artery with severe stenosis (≥ 70%) (8 images in class 4) and one with moderate stenosis (50–69%) (8 images in class 3). One patient presented with one occluded coronary artery (8 images in class 5) and one coronary artery with severe stenosis (≥ 70%) (8 images in class 4).

## Radiomic feature extraction

The ROI was semi-automatically generated from MPR images, by defining a rectangle, 40-pixel wide, centered along the straightened coronary centerline, with the lower boundary aligned with the image inferior margin, and the upper boundary to exclude the ventricular muscle (only this last boundary was manually identified). For each coronary vessel, the ROI was defined on a single view and then applied to all the remaining images.

Following ROI delineation, default Pyradiomics 3.0 settings were considered for extracting 2D radiomic features, namely B-spline interpolation, and fixed-bin histogram discretization, with 25 bins[24,25]. Features were extracted from the original image, its exponential transformation and its Wavelet transformation (four wavelet decomposition images as HH, LL, LH, HL, where 'H' stands for a high-pass filter and 'L' for a low-pass filter). Thus, 18 first-order statistics and 75 textural features (24 gray level co-occurrence matrix, 16 gray level run length matrix, 16 gray level size zone matrix, 5 neighboring gray-tone difference matrix, and 14 gray level dependence matrix) from the original image, the exponential transform and the 4 wavelet transforms were computed, resulting in a total of 558 radiomic features.

## Model development

*Feature summarization approaches*

Radiomic features were extracted from each MPR image of the patient's coronary arteries, and, as reported in Table 1, each patient may have up to 24 images (from 8 MPR images and 3 coronary arteries). Consequently,

different approaches were implemented to predict patient CAD-RADS score from radiomic features extracted from all images (Fig. 2).

The first approach (Av_C) consists of averaging the radiomic features across all images of each patient's coronary artery to predict an artery-based class: thus, a score prediction is obtained for each coronary artery and the patient CAD-RADS score is assigned to the most critical artery-based score. The second approach (Stat_C), reproduces the first approach but including additional statistical metrics besides the average, namely the minimum, maximum and standard deviation of the radiomic features. The third approach (MV_P) consists of predicting a score for each image (from the features extracted) and then performing a majority voting to predict the patient's class: thus, the patient's CAD-RADS score will be the most recurrent score among the image-based predictions. Finally, the last approach (MV_C) consists of predicting a score for each image and then performing a majority voting at the coronary level to predict an artery-based class: the artery-based score will be the most recurrent score among image-based predictions referring to that coronary artery and, then, the patient's CAD-RADS score is assigned to the most critical artery-based score. Thus, the first two approaches (Av_C, Stat_C) summarize the radiomic features of different images by some statistical metrics (average, or maximum, minimum, average and standard deviation) while the other two approaches (MV_P, MV_C) consider the features of individual images to make the predictions, and then "summarize" the predictions through a majority voting. While the MV_P approach directly predict the patient's class from all the images, the Av_C, Stat_C and MV_C approaches introduce a coronary artery-based scoring, with the most severe class among arteries as the final patient class.

*Machine learning pipeline*

For each of the four approaches a cascade pipeline, consisting of sequential sub-classification tasks, was implemented (Fig. 3). Specifically, four gradient boosting classifiers[26] were trained, the first one distinguishing class 0 vs. 1–2-3–4 vs. 5, the second one distinguishing 1–2 vs. 3–4, the third one 1 vs. 2, and the fourth one 3 vs. 4. This hierarchical design was adopted to simplify the original six-class classification problem by progressively breaking it down into more manageable binary or ternary tasks. The first classifier targets the most clinically distinct categories, healthy (CAD-RADS 0), intermediate (CAD-RADS 1–4), and total occlusion (CAD-RADS 5), which are generally easier to separate based on imaging features. Subsequent models focus on resolving

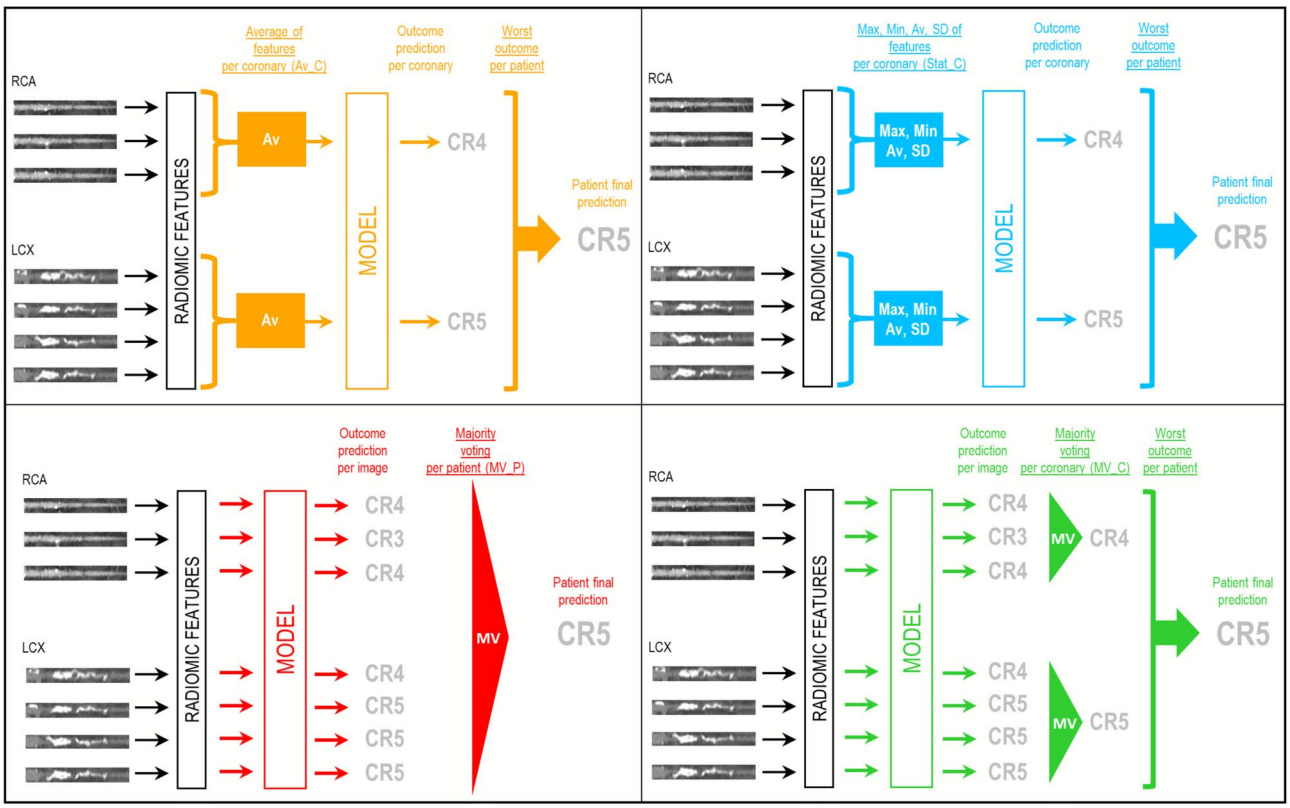


**Fig. 2**. Feature summarization approaches. Top-left: average-based approach with coronary artery-based prediction followed by patient prediction (Av_C). Top-right: statistical-based approach (including maximum, minimum, average and standard deviation) with coronary artery-based prediction followed by patient prediction (Stat_C). Bottom-left: majority voting approach with direct patient prediction (MV_P). Bottom-right: majority voting approach with coronary artery-based prediction followed by patient prediction (MV_C). CR 0: CAD-RADS class 0; CR 1: CAD-RADS class 1; CR 2: CAD-RADS class 2; CR 3: CAD-RADS class 3; CR 4: CAD-RADS class 4; CR 5: CAD-RADS class 5.

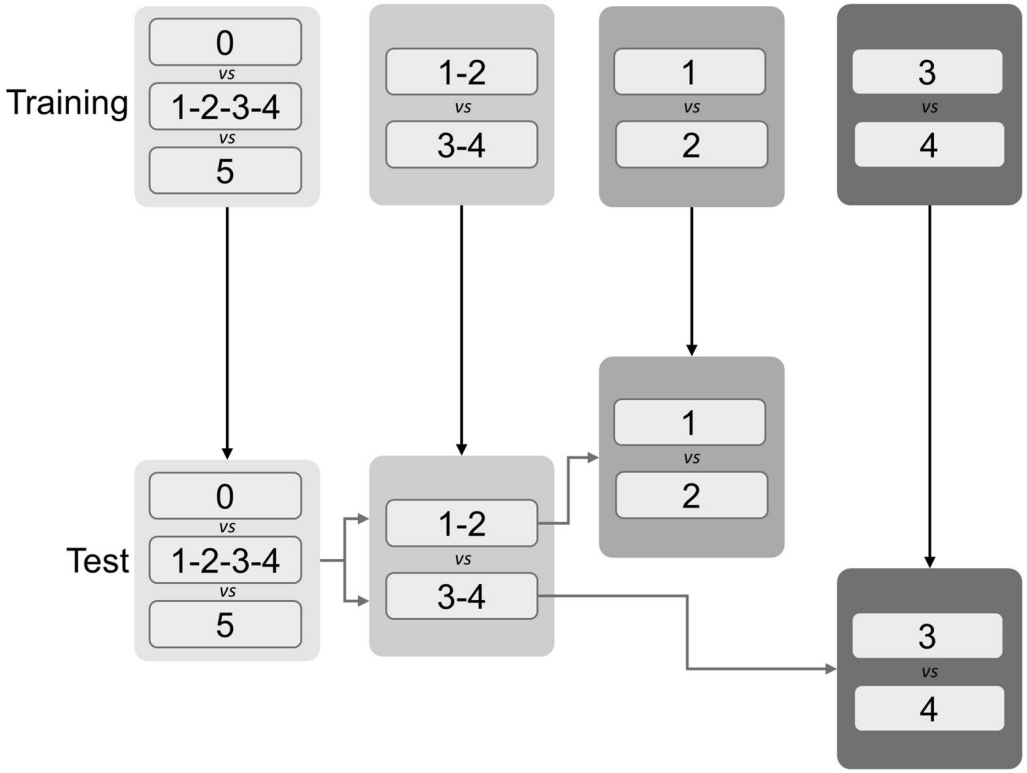


**Fig. 3**. Cascade pipeline. Different gray rectangles represent different models.

| Images | All | Class 0 | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 |
|---|---|---|---|---|---|---|---|
| Patient Training set | 198 | 36 | 36 | 36 | 36 | 36 | 18 |
| Image Training set | 2259 | 848 | 245 | 312 | 350 | 336 | 168 |
| Patient Test set | 40 | 9 | 5 | 4 | 6 | 10 | 6 |
| Test set | 520 | 316 | 26 | 38 | 64 | 120 | 56 |

**Table 3**. Training and test sets composition.

finer-grained distinctions among the intermediate classes, where decision boundaries are subtler. Such a cascade approach can enhance overall performance by allowing each classifier to specialize in a narrower classification task, reducing the complexity at each stage and improving the ability to learn class-specific features. To develop the machine learning models, about 80% of the image dataset was considered for training and the remaining 20% for testing (with a patient-stratified partitioning), as shown in Table 3. For the test analysis, for the sake of simplicity, at each task, only the subset of test samples correctly classified by the preceding step was passed on for further classification. Importantly, this approach does not exclude misclassified samples from the final evaluation. Misclassified instances at any stage are counted as incorrect in the final performance metrics. However, once a sample is misclassified at a given stage, it is not processed further in the cascade. For instance, if a sample truly belonging to class 0 is incorrectly classified at the first stage as belonging to classes 1–4, it is considered misclassified overall, and is not further evaluated to determine whether it would have been classified as 1, 2, 3, or 4. Thus, even if the cascade can be applied up to the final step (as in real inference scenarios), for the purpose of the test set evaluation, it is irrelevant. This design reflects the hierarchical decision structure of the cascade and ensures that the reported performance reflects both the accuracy of individual tasks and the compounded effect of errors across the cascade.

Within the training phase, at each sub-classification task, feature scaling and selection, and data balancing were performed. Specifically, features were standardized (z-score) and the mean and standard deviation values of each feature were used to subsequently normalize the test set. Moreover, to reduce the dimensionality of the radiomic dataset and avoid overfitting, a two-step feature selection was applied. First, significantly discriminant features were identified through a statistical test, and features with p-values $< 0.01$ were retained. Mann–Whitney U-test was used for binary tasks, one-way ANOVA for multiclass tasks with normal distribution, and the Kruskal–Wallis test for multiclass tasks with non-normal distribution (with normality assumption determined through Shapiro test for groups with less than 50 instances or the Kolmogorov–Smirnov test). Second, the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm was applied, by selecting the penalty

strength parameter in LASSO regression α. Finally, following data balancing through the Synthetic Minority Over-sampling TEchnique (SMOTE) algorithm [27], the gradient boosting classifier was trained on the selected K features.

To identify the optimal value of penalty strength parameter α, a five-fold cross-validation was conducted during the training phase. The process involved varying α and selecting the value that maximized the performance of the validation metrics.

Only the best performing feature aggregation approaches (Av_C, Stat_C, MV_P, MV_C) across the five-fold cross-validation were applied to the test set. Moreover, for said approaches a direct six-class classification approach was implemented to evaluate the potential benefit of a cascade pipeline and.

*Statistical and explainability analysis*

The predictive performance of the model was evaluated through the balanced accuracy, sensitivity, specificity, f1-score, and Area Under the Curve of the Receiver Operating Characteristic curve (AUC-ROC). For multiclass tasks, the macro averaging version of f1-score, AUC-ROC, sensitivity, and specificity were used. Bootstrap with 100 iterations was performed to compute the 95% confidence interval (CI). In each iteration, a new test set was generated by sampling with replacement from the original test set, preserving the original sample size. Performance metrics were computed for each resampled set, and the median and 95% CI were derived from the obtained distribution of metric values across all the iterations. Statistical difference among the ROC curves was assessed through the DeLong test. T-test was performed to compare the performance metrics between the models. Statistical significance was assumed for p-value $< 0.05$.

SHAP (SHapley Additive exPlanations) analysis was performed on the trained models to gain insights into the decision-making processes and the most relevant selected features[28].

*Code availability*

The developed pipeline is available online at https://doi.org/10.5281/zenodo.17792272.

## Results

Figure 4 provides the results of five-fold cross-validation for the different classification tasks for each of the feature summarization approaches. Overall, the task 0 vs. 1–2-3–4 vs. 5 shows higher performance than the following tasks, for all the approaches. Moreover, approaches based on majority voting (MV_P and MV_C) outperformed those based on feature statistical metrics (Av_C and Stat_C). Additionally, to evaluate the potential benefit of a cascade pipeline, the two best-performing approaches (MV_P and MV_C) were implemented also for direct six-class classification. However, during cross-validation, this direct classification strategy yielded poor performance across both feature integration approaches (Supplementary Table S1), and was therefore not considered for application to the test set.

Consequently, the majority voting approaches (MV_P and MV_C) with cascade pipeline classification were applied to the test set. Figure 5 shows the ROC curves and the confusion matrix of the MV_P and MV_C approaches in the test set, for the full six-class CAD-RADS classification. In both cases excellent performances were obtained. Although no significant differences emerged between the ROC curves of the two approaches (DeLong $p > 0.05$), minor stratification differences can be appreciated from the confusion matrix. Particularly, MV_P identified all the CR 2 patients and misclassified only 1 patient in CR 0 (over 9), in CR 1 (over 5) and in CR 5 (over 6), outperforming MV_C for said classes. Differently, MV_C outperformed MV_P for CR 3, identifying 5 patients over 6 and CR 4 identifying 9 patients over 10, and equally performed for CR 5, identifying 5 patients over 6. The MV_P approach provided higher AUC than the MV_C approach for class 0 ($AUC_{MV\_P} = 0.94$; $AUC_{MV\_C} = 0.82$), class 1 ($AUC_{MV\_P} = 0.92$; $AUC_{MV\_C} = 0.78$), class 2 ($AUC_{MV\_P} = 0.97$; $AUC_{MV\_C} = 0.84$) and lower in class 3 ($AUC_{MV\_P} = 0.77$; $AUC_{MV\_C} = 0.96$) and class 4 ($AUC_{MV\_P} = 0.88$; $AUC_{MV\_C} = 0.98$), but without statistical significance (DeLong $p > 0.05$), and the same AUC for class 5 ($AUC_{MV\_P} = 0.85$; $AUC_{MV\_C} = 0.85$). On average, the MV_P yielded an AUC = 0.89 and the MV_C an AUC = 0.87 (average over the 6 classes). Both approaches misclassified 9 patients over 40, five of which were in common, as detailed in the Supplementary Table S2. The 40 patients within the test set presented 10 calcific, 12 mixed and 20 fibrolipidic plaques and, across the misclassified patients, all but one involved either mixed or fibrolipidic plaques, which are known to be more challenging to distinguish, particularly when compared to calcified plaques. Moreover, in the majority of misclassified cases, the model assigned the patient to an adjacent class. Tables 4 and 5 show the macro-average performance metrics of the MV_P and MV_C approaches in the test set for the overall six-class classification, and for each of the four tasks, respectively, obtained across the 100 bootstrap iterations. As regards the overall six-class classification, no significant differences were identified between the performance metrics of the MV_P and the MV_C approaches. However, significant differences between the MV_P and MV_C approaches emerged within the tasks, as reported in Table 5. Analyzing the tasks, differently from the cross-validation results, a clear trend of a task outperforming the others for both approaches did not emerge. MV_P provided similar higher performances in the tasks 0 vs. 1–2-3–4 vs. 5, 1–2 vs. 3–4 and 1 vs. 2, with respect to the task 3 vs. 4 associated with the lowest performance; differently MV_C provided similar higher performances in the tasks 0 vs. 1–2-3–4 vs. 5, 1–2 vs. 3–4 and 3 vs. 4, with respect to the task 1 vs. 2 associated with the lowest performance. As a preliminary proof-of-concept case study, five patients with stenoses of different degrees in multiple coronary arteries were analyzed, as detailed in Sect. 2.1.3. By applying the MV_P and MV_C approaches to these five additional patients who had multiple coronary arteries with different severities of stenosis, the MV_C approach correctly identified the CAD-RADS score of all 5 patients, while the MV_P only of 4 out of 5. For all the four approaches, the model processing time for each patient, from ROI generation to final prediction was of about 2 min.

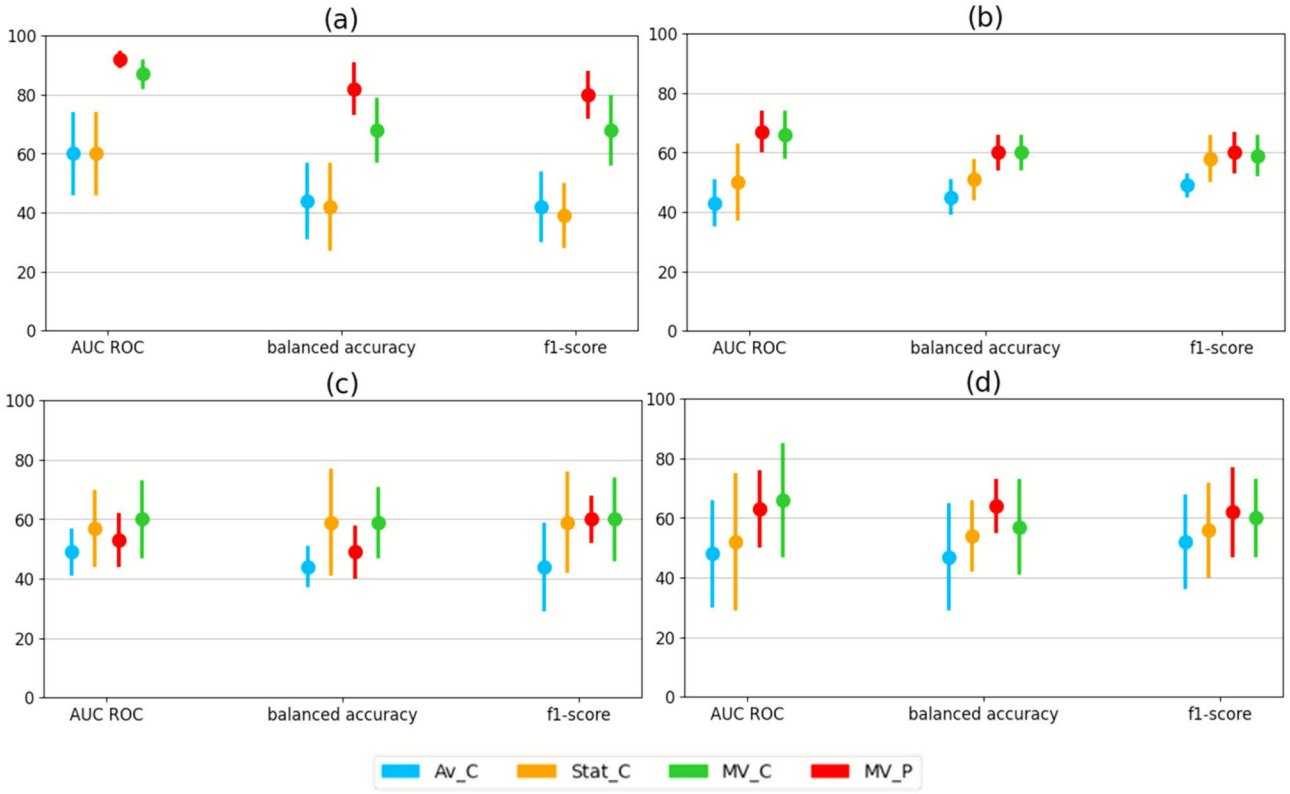


**Fig. 4**. Performance metrics of the 4 feature summarization approaches in the cross-validation sets for the 4 classification tasks. **a** First classification task: 0 vs. 1–2-3–4 vs. 5; **b** second classification task 1–2 vs. 3–4; **c** third classification task 1 vs 2; **d** fourth classification task 3 vs. 4. Av_C: average-based approach with coronary artery-based prediction followed by patient prediction. Stat_C: statistical-based approach (maximum, minimum, average and standard deviation) with coronary artery-based prediction followed by patient prediction. MV_P: majority voting approach with direct patient prediction. MV_C: majority voting approach with coronary artery-based prediction followed by patient prediction.

Supplementary Table S3 reports all the selected radiomic features of the MV_P and MV_C models. Notably, most of the selected features derived from the wavelet transforms of the images and belonged to the textural class, suggesting that the texture, accounting for arterial tissue composition, contributes to the prediction of the stenosis grading. Results of the SHAP analysis on the best performing model, namely the MV_P, are reported in the Supplementary Materials. Figure S1 reports the model features importance, while Figure S2 details the role of the most important features in determining the prediction of a patient correctly classified in the CAD-RADS 4 class, representing an explanatory case. SHAP analysis revealed that the majority of the top 10 most influential features belong to the textural class. Specifically, for the MV_P approach (Supplementary Figure S1), seven texture-based features were consistently selected across the sub-classification tasks. Gray Level Size Zone Matrix (GLSZM) features were the most frequently chosen (eight occurrences), capturing characteristics such as the size of homogeneous gray-level zones (e.g., *Small Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis*), gray-level heterogeneity (*Gray Level Non-Uniformity Normalized*), and size heterogeneity (*Size-Zone Non-Uniformity*). The Neighbouring Gray Tone Difference Matrix (NGTDM) *Coarseness* feature, which reflects the average difference between a pixel's gray level and that of its neighbors, appeared six times. Gray Level Run Length Matrix (GLRLM) features were selected four times, with *Run Variance* and *Run Length Non-Uniformity Normalized* indicating variability in the lengths of continuous gray-level runs. Lastly, Gray Level Dependence Matrix (GLCM) *Inverse Difference Moment Normalized (IDMN)*, selected three times, measures local homogeneity within the image. Overall, these features offer a comprehensive characterization of gray-level spatial distribution, capturing intensity variation, neighborhood contrast, uniformity, and texture structure. Such descriptors are potentially indicative of tissue composition, highlighting patterns associated with high-intensity regions (e.g., fibrous tissue, calcifications) and low-intensity areas (e.g., lipid-rich or soft tissue zones).

## Discussion

The present study addresses the challenge of translating coronary plaque radiomics into clinically actionable, patient-level assessments. Given that radiomic features are usually extracted at the plaque or coronary level, while clinical endpoints are commonly defined at the patient level, effective feature summarization strategies are essential. To this end, we developed and compared four different approaches for deriving a patient-level CAD-RADS score from CCTA-based MPR images of the main coronary arteries. Two methods were based on

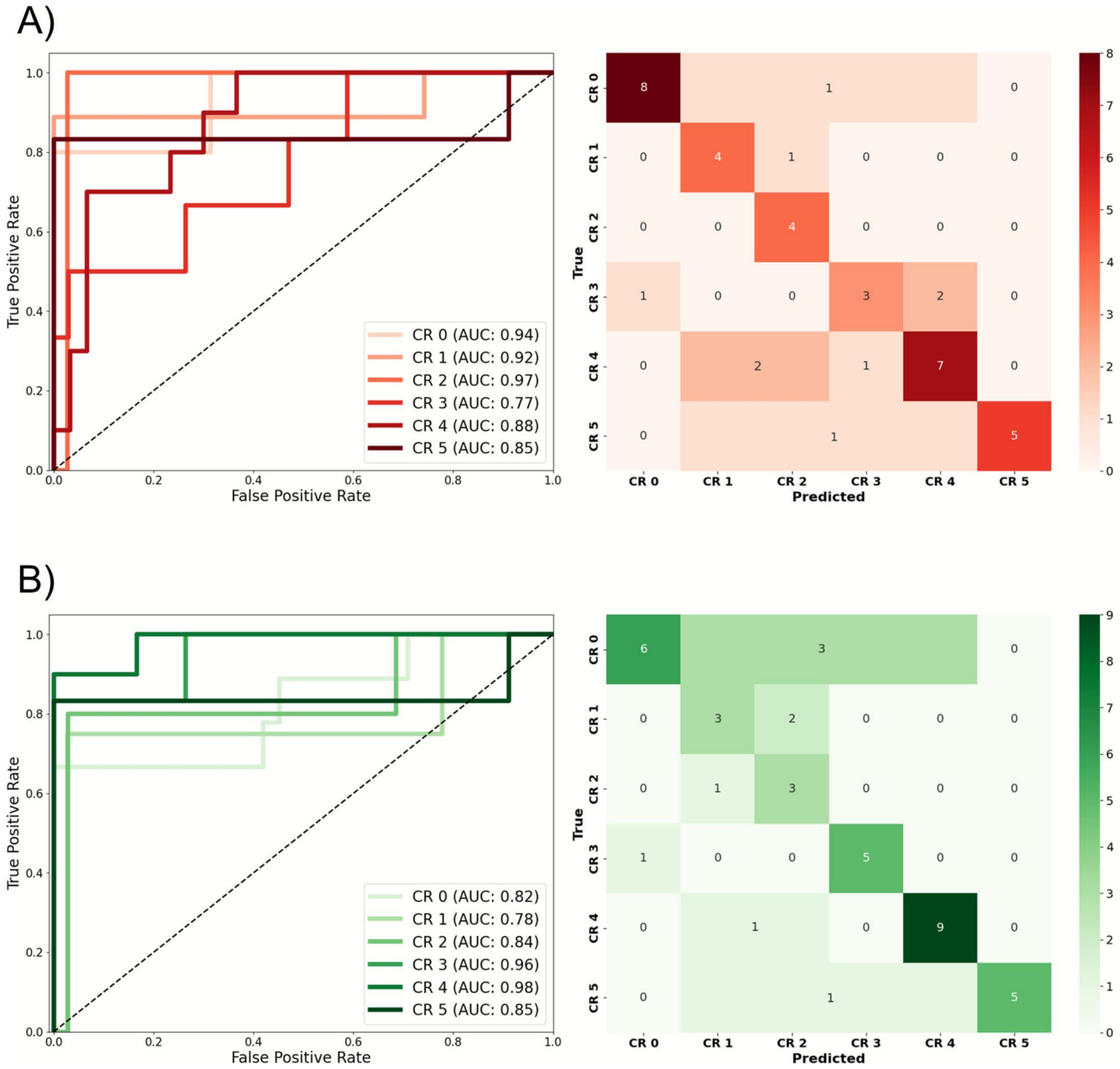


**Fig. 5**. Receiver operating characteristic (ROC) curves for each CAD-RADS class and confusion matrix for the two majority voting approaches in the test set. **A** Majority voting approach with direct patient prediction (MV_P). **B** majority voting approach with coronary artery-based prediction followed by patient prediction (MV_C). CR 0: CAD-RADS class 0; CR 1: CAD-RADS class 1; CR 2: CAD-RADS class 2; CR 3: CAD-RADS class 3; CR 4: CAD-RADS class 4; CR 5: CAD-RADS class 5. DeLong $p > 0.05$ for all the pairwise comparisons at same CR between MV_P and MV_C.

feature statistical metrics computation across the images, and two involved image-based prediction followed by a majority voting criterion. Moreover, both a direct patient-level classification, and an intermediate coronary-level classification, were evaluated. These strategies serve to bridge the gap between detailed image-based data and patient-level clinical decision-making in a personalized medicine context.

On the validation set, the majority voting approaches (MV_P and MV_C) outperformed the statistical-based approaches (Av_C and Stat_C). This suggests that considering only the average, minimum, maximum and standard deviation of the radiomic features from multiple images of coronary arteries results in a loss of information on the most severe characteristics, which are those that drive the clinical assignment of CAD-RADS. Differently, an image-based prediction with majority voting is intrinsically more conservative of the most severe disease if this is captured by at least half of the views.

On the test set, both MV_P and MV_C provided high stratification performance with a mean macro-averaged AUC of 0.89 and 0.87, respectively, and a mean balanced accuracy of 0.77 across the six-class scoring. Additionally, within the individual tasks, no consistent trend emerged indicating that a specific task outperformed

| | AUC | Balanced accuracy | F1 score | Sensitivity | Specificity |
|---|---|---|---|---|---|
| MV_P | CR0: 0.94 (0.92–0.95) | 0.77 (0.75–0.78) | 0.74 (0.73–0.76) | 0.77 (0.75–0.78) | 0.95 (0.95–0.96) |
| | CR1: 0.93 (0.91–0.94) | | | | |
| | CR2: 0.97 (0.96–0.98) | | | | |
| | CR3: 0.76 (0.73–0.77) | | | | |
| | CR4: 0.87 (0.86–0.88) | | | | |
| | CR5: 0.87 (0.81–0.88) | | | | |
| MV_C | CR0: 0.85 (0.81–0.85) | 0.77 (0.74–0.77) | 0.75 (0.72–0.75) | 0.77 (0.74–0.77) | 0.96 (0.95–0.96) |
| | CR1: 0.79 (0.72–0.80) | | | | |
| | CR2: 0.84 (0.80–0.86) | | | | |
| | CR3: 0.97 (0.95–0.97) | | | | |
| | CR4: 0.99 (0.98–0.99) | | | | |
| | CR5: 0.87 (0.81–0.88) | | | | |

**Table 4**. Macro-average performance metrics of the MV_P and MV_C approaches in the test set. Data are expressed as the median and 95% confidence interval.

| | | AUC_M | Balanced accuracy | F1 score | Sensitivity | Specificity |
|---|---|---|---|---|---|---|
| MV_P | 0–1234-5 | 0.96 (0.95–0.96) | 0.91 (0.88–0.91) | 0.92 (0.89–0.92) | 0.91 (0.88–0.91) | 0.95 (0.94–0.95) |
| | 12–34 | 0.99 (0.99–0.99)* | 0.94 (0.93–0.94) | 0.93 (0.92–0.94) | 0.88 (0.86–0.89) | 1.00 (1.00–1.00)* |
| | 1–2 | 0.80 (0.75–0.83) | 0.90 (0.87–0.92) | 0.90 (0.84–0.90) | 1.00 (1.00–1.00) | 0.80 (0.75–0.83) |
| | 3–4 | 0.63 (0.59–0.67) | 0.71 (0.69–0.75) | 0.82 (0.79–0.84) | 0.90 (0.86–0.91) | 0.50 (0.50–0.60) |
| MV_C | 0–1234-5 | 0.92 (0.91–0.93) | 0.83 (0.81–0.84) | 0.85 (0.83–0.86) | 0.83 (0.81–0.84) | 0.91 (0.89–0.91) |
| | 12–34 | 1.00 (0.99–1.00)* | 0.97 (0.96–0.97) | 0.96 (0.96–0.97) | 0.93 (0.92–0.95) | 1.00 (1.00–1.00)* |
| | 1–2 | 0.56 (0.53–0.62) | 0.69 (0.63–0.70) | 0.67 (0.59–0.67) | 0.75 (0.67–0.77) | 0.62 (0.57–0.66) |
| | 3–4 | 1.00 (1.00–1.00) | 1.00 (1.00–1.00) | 1.00 (1.00–1.00) | 1.00 (1.00–1.00) | 1.00 (1.00–1.00) |

**Table 5**. Macro-average performance metrics of the MV_P and MV_C approaches in the test set, for each task. Data are expressed as the median and 95% confidence interval. All pairwise comparisons between MV_P and MV_C were statistically significant but one (identified by *).

the others across both approaches. In both cases, the first two tasks provided excellent performances; then, while MV_P demonstrated high performance for the task 3 vs. 4, but not for the task 1 vs. 2, MV_C showed the opposite. The performance decline observed with increasing task granularity is expected, as distinguishing between adjacent classes is inherently more difficult than separating more distinct categories. However, due to the limited size of the test set, definitive conclusions cannot be drawn at this stage. In future, expanding the number of samples in classes 0, 1, 2, 3, and 4 within the training set would likely enhance the model's ability to learn the subtle distinctions between neighboring classes. From a clinical standpoint, the high accuracy in detecting extreme categories such as CAD-RADS 0 and 5, but also intermediate categories such as 1–2 vs. 3–4, is particularly encouraging, as these cases often require the most definitive clinical decisions, ranging from patient treatment exclusion to immediate intervention. Nonetheless, enhancing the model's ability to differentiate both 1 vs. 2 and 3 vs. 4 classes remains essential to support more refined and patient-specific decision-making, and represents an important focus for future development.

When the two majority voting approaches (MV_P and MV_C) were applied to the test set, no significant differences in the classification performances were obtained, considering the final six-class classification. However, it must be noted that in the dataset used for this analysis, each patient presented with only images related to arteries that had the same degree of stenosis. This dataset composition inherently reduces the potential advantage of the MV_C approach over MV_P because both methods operate on arteries that are largely homogeneous in disease severity for each patient. Consequently, both summarization strategies are methodologically valid and yielded comparable results in this context. However, to the authors' opinion, the MV_C approach is conceptually more generalizable and better aligned with clinical practice, particularly in scenarios where imaging data includes all major coronary arteries with varying degrees of stenosis. In such cases, the MV_P approach may be disproportionately influenced by the majority of images from healthy or less diseased arteries, potentially underestimating the severity of the most critical lesion. In contrast, MV_C mimics the clinical decision-making process by assigning the patient's CAD-RADS score according to the artery with the highest severity, independent of the number or proportion of images from healthy or less diseased arteries. To explore this hypothesis, we performed a preliminary proof-of-concept analysis on five patients whose imaging included arteries with heterogeneous stenosis degrees, some with lower severity than the artery defining their CAD-RADS class. In this limited sample, The MV_P approach misclassified 1 out of 5 patients, while the MV_C approach correctly classified all the patients. This explanatory analysis was included as preliminary proof-of-concept case study of how the two approaches potentially work in a more general setting. However, the analysis

was limited to five patients, as only a small number of patients met this specific inclusion criterium within our dataset. As such, although these preliminary findings suggest that MV_C may better capture clinically relevant severity when heterogeneous arterial involvement is present, the small sample size limits the strength of any conclusions. Future studies with expanded cohorts including heterogeneous artery-level disease are necessary to fully validate this hypothesis and clarify the comparative performance of these summarization strategies.

Overall, both approaches provided high performances in predicting six-class CAD-RADS scoring, demonstrating the potentialities of the developed radiomic feature summarization approaches, combined with the implemented cascade pipeline for automated CAD-RADS grading from MPR CCTA images of coronary arteries. In particular, the benefit of the cascade pipeline compared to a direct six-class scoring was demonstrated for the two best performing approaches (MV_P and MV_C). Performing a direct classification provided low stratification performance. Differently, by decomposing the problem into a sequence of simpler binary classification tasks, the cascaded pipeline allowed the model to focus on more manageable decisions at each stage (e.g., distinguishing low- from high-severity cases), determining high stratification performance.

To the best of the authors' knowledge, this study represents the first methodological analysis exploring different approaches to how artery-level radiomic features can be summarized to obtain a patient-level prediction. Indeed, most of the studies analyzing coronary plaque with radiomics were based on a plaque-level analysis, aimed at predicting plaque composition[29,30], plaque stenosis grading[13,30,31], and plaque vulnerability[16,17,19]. Three studies used plaque radiomics to predict major adverse cardiac events[13,16,32], thus using plaque-level information to predict a patient-level outcome. The first study considered one plaque for each patient[13] (namely the culprit plaque for patients undergoing the event and one matched plaque for patients without events), thus simplifying the factual reality. The second study performed statistics of the features of different plaques, considering the minimum, 1$^{st}$ quartile, mean, 3$^{rd}$ quartile, maximum, and standard deviation values as "summarized" features[32], and the last one did not provide details on how the plaque features were integrated to obtain a patient-level radiomic signature[16].

Given the recent application of plaque and coronary artery radiomics and the paucity of related studies, it is of utmost importance to propose and develop potential methods for the definition of a common approach that can be used to standardize plaque/artery radiomic analyses. The feature summarization approaches based on a majority voting approach presented here address this need and, thanks to their versatility, can be applied to predict patient outcomes (e.g., major adverse cardiac events) from multiple plaques. Accordingly, the present study aims to pave the way toward the development of a robust and reproducible coronary artery radiomic-based machine learning approach that can be used for the prediction of patient stenosis severity.

In terms of prediction of stenosis grading, the developed majority voting approaches, provided classification performance on the test set in line with the multiclass radiomic classification model for stenosis grading proposed by Jin et al.[30] (AUC of 0.91). To the best of the authors' knowledge, besides the study by Jin et al.[30], this work represents the first multiclass classification analysis for predicting stenosis grading and CAD-RADS with a radiomic approach. Differently from Jin et al.[30], herein, the radiomic analysis was performed on MPR images of the whole coronary arteries, rather than 3D plaque volumes, thus providing a more computationally efficient approach. Our radiomic approach also outperformed several deep learning studies aiming at CAD-RADS grading from CCTA[6,11] and provided similar classification performance as the six-class model by Denzinger et al.[8,9], that achieved an average accuracy of 0.86 on 759 patients[9], and as the 3-class model by Gerbasi et al.[10], that achieved an AUC = 0.93 on a test set of 50 patients. Thus, the present study, besides providing novel feature integration approaches, demonstrates the potentiality of radiomics in predicting CAD-RADS scores.

The SHAP analysis results further underscore the potential of radiomics for plaque characterization. Unlike deep learning models, which rely on abstract latent representations often lacking direct clinical interpretability, radiomic features are explicit, quantifiable image descriptors that capture key aspects of tissue properties, such as texture, heterogeneity, shape, and intensity. In this study, the top 10 most influential features were predominantly textural, reflecting structural heterogeneity, intensity distributions and spatial relationships. These features may correspond to biologically relevant plaque markers, such as the presence of large lipid-rich cores, or necrotic regions or small, spotty calcifications. Such a level of interpretability facilitates the translation of radiomics into the clinical workflow. The potential for clinical translation is further strengthened by the processing time: with a processing time of approximately 2 min, significantly faster than the average 16 ± 7 min required for manual scoring, the automated system enables a quicker assessment of CAD severity, which can potentially enhance the efficiency in clinical practice.

The study presents some limitations. First, a simplified bounding-box ROI generation was applied, which on one side offers advantages in terms of computational efficiency and time savings, but on the other side might affect the model performance. To the best of our knowledge, there is currently no publicly available automated coronary artery segmentation approach that includes segmentation of the arterial wall. Therefore, considering the significant effort involved in developing fully automated or deep learning-based solutions, any potential performance improvements must be carefully weighed against the increased methodological complexity such approaches would add. Second, in the dataset used, each patient had coronary artery images showing only stenoses of the same severity, corresponding to the lesion that determined the patient's overall CAD-RADS score. This dataset peculiarity likely limited the performance difference between the MV_P and MV_C. To obtain a more comprehensive dataset, it is recommended to include coronary arteries exhibiting lower degrees of stenosis than those defining a patient's higher CAD-RADS category, thereby encompassing all three main coronary arteries for each patient. Future studies with larger, more diverse cohorts featuring heterogeneous artery-level disease are needed to more accurately compare the performance of the MV_P and MV_C approaches. Finally, our study was conducted using a single-center, single-vendor dataset comprising 238 patients, which inherently limits the generalizability of our findings. A known challenge in radiomics is the variability of feature stability across different scanner models and imaging protocols. Scanner-specific factors can affect radiomic

feature reproducibility, potentially impacting the robustness of predictive models when applied to external datasets acquired with different equipment. Therefore, to strengthen the validity and clinical applicability of our methodology, future work should prioritize testing and validating the model on larger, multicentric datasets collected from multiple vendors and scanner types. Such efforts will be essential to enhance the reliability and generalizability of our approach. Despite the aforementioned limitations, the present study achieved a score of 73% (classified as "good") according to the METhodological RadiomICs Score (METRICS), which is a quality scoring tool to assess research quality of radiomic studies[33].

## Conclusions

In the present study, a novel CCTA-based coronary artery radiomics analysis was proposed to provide a step forward toward the definition of a radiomics-based tool for the patient-level stenosis severity assessment. In particular, we addressed the open challenge of how to summarize multiple radiomic information at different coronary artery levels to obtain a unique patient-level prediction that can be used in the clinical setting. Four feature integration approaches were proposed and developed, and their performance was evaluated within a cross-validation scheme: the majority voting approaches (MV_P, MV_C) outperformed the statistical-based ones (Av_C, Stat_C), and thus further applied on the test set. Overall, the two majority voting approaches provided similar performance on the test set (with a mean macro-averaged AUC of 0.89 and 0.87, respectively). However, the MV_C approach, which assigns the patient class to the worst coronary artery-based score, better aligns with current CAD-RADS clinical scoring and is more generalizable, as demonstrated by the better performance in the case study scenario. If confirmed in larger studies, the proposed algorithm has the potential to provide automated CAD-RADS scoring, thus providing an objective and repeatable coronary artery stenosis assessment.

## Data availability

**All relevant data and materials have been included in the article. Further inquiries can be directed to the corresponding authors.**

Received: 11 January 2025; Accepted: 9 December 2025
Published online: 29 January 2026

## References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England)* **392**, 1736–1788 (2018).
2. Bentzon, J. F., Otsuka, F., Virmani, R. & Falk, E. Mechanisms of Plaque Formation and Rupture. *Circ. Res.* **114**, 1852–1866 (2014).
3. Knuuti, J. et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur. Heart J.* **41**, 407–477 (2020).
4. Cury, R. C. *et al.* CAD-RADS$^{TM}$ 2.0 - 2022 Coronary Artery Disease-Reporting and Data System: An Expert Consensus Document of the Society of Cardiovascular Computed Tomography (SCCT), the American College of Cardiology (ACC), the American College of Radiology (ACR), and the N. *J. Cardiovasc. Comput. Tomogr.* **16**, 536–557 (2022).
5. Paul, J.-F., Rohnean, A., Giroussens, H., Pressat-Laffouilhere, T. & Wong, T. Evaluation of a deep learning model on coronary CT angiography for automatic stenosis detection. *Diagn. Interv. Imaging* **103**, 316–323 (2022).
6. Penso, M. et al. A token-mixer architecture for CAD-RADS classification of coronary stenosis on multiplanar reconstruction CT images. *Comput. Biol. Med.* **153**, 106484 (2023).
7. Huang, Z. et al. Clinical evaluation of the automatic coronary artery disease reporting and data system (CAD-RADS) in coronary computed tomography angiography using convolutional neural networks. *Acad. Radiol.* **30**, 698–706 (2023).
8. Denzinger, F. *et al.* Automatic CAD-RADS Scoring Using Deep Learning. In *Medical Image Computing and Computer Assisted Intervention -- MICCAI 2020* (eds. Martel, A. L. et al.) 45–54 (Springer International Publishing, 2020).
9. Denzinger, F. *et al.* CAD-RADS Scoring using Deep Learning and Task-Specific Centerline Labeling. In *Proceedings of the 5th international conference on medical imaging with deep learning* (eds. Konukoglu, E. et al.) vol. 172 315–324 (PMLR, 2022).
10. Gerbasi, A. et al. CAD-RADS scoring of coronary CT angiography with multi-axis vision transformer: A clinically-inspired deep learning pipeline. *Comput. Methods Progr. Biomed.* **244**, 107989 (2024).
11. Muscogiuri, G. et al. Performance of a deep learning algorithm for the evaluation of CAD-RADS classification with CCTA. *Atherosclerosis* **294**, 25–32 (2020).
12. Lo Iacono, F. et al. Identification of subclinical cardiac amyloidosis in aortic stenosis patients undergoing transaortic valve replacement using radiomic analysis of computed tomography myocardial texture. *J Cardiovasc. Comput. Tomogr.* https://doi.org/10.1016/j.jcct.2023.04.002 (2023).
13. Hu, W. et al. Novel radiomics features from CCTA images for the functional evaluation of significant ischaemic lesions based on the coronary fractional flow reserve score. *Int. J. Cardiovasc. Imaging* **36**, 2039–2050 (2020).
14. Shang, J. et al. Prediction of acute coronary syndrome within 3 years using radiomics signature of pericoronary adipose tissue based on coronary computed tomography angiography. *Eur. Radiol.* **32**, 1256–1266 (2022).
15. Izquierdo, C. et al. Radiomics-based classification of left ventricular non-compaction, hypertrophic cardiomyopathy, and dilated cardiomyopathy in cardiovascular magnetic resonance. *Front. Cardiovasc. Med.* **8**, 764312 (2021).
16. Chen, Q. et al. A coronary CT angiography radiomics model to identify vulnerable plaque and predict cardiovascular events. *Radiology* **307**, e221693 (2023).
17. Chen, Q. et al. CT texture analysis of vulnerable plaques on optical coherence tomography. *Eur. J. Radiol.* **136**, 109551 (2021).
18. Li, X.-N. et al. Identification of pathology-confirmed vulnerable atherosclerotic lesions by coronary computed tomography angiography using radiomics analysis. *Eur. Radiol.* **32**, 4003–4013 (2022).
19. Lin, A. et al. Radiomics-based precision phenotyping identifies unstable coronary plaques from computed tomography angiography. *JACC. Cardiovasc. Imaging* **15**, 859–871 (2022).
20. Cury, R. C. et al. Coronary artery disease - reporting and data system (CAD-RADS): An expert consensus document of SCCT, ACR and NASCI: endorsed by the ACC. *JACC. Cardiovasc. Imaging* **9**, 1099–1113 (2016).
21. Pontone, G. et al. Clinical applications of cardiac computed tomography: a consensus paper of the European Association of Cardiovascular Imaging-part II. *Eur. Hear. J. Cardiovasc. Imaging* **23**, e136–e161 (2022).

22. Pontone, G. et al. Clinical applications of cardiac computed tomography: a consensus paper of the european association of cardiovascular imaging-part I. *Eur. Hear. J. Cardiovasc. Imaging* **23**, 299–314 (2022).
23. Abbara, S. et al. SCCT guidelines for the performance and acquisition of coronary computed tomographic angiography: A report of the society of Cardiovascular Computed Tomography Guidelines Committee: Endorsed by the North American Society for Cardiovascular Imaging (NASCI. *J. Cardiovasc. Comput. Tomogr.* **10**, 435–449 (2016).
24. Pyradiomics features description. *Pyradiomics features description* (2018).
25. van Griethuysen, J. J. M. et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
26. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
27. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority over-Sampling Technique. *J. Artif. Int. Res.* **16**, 321–357 (2002).
28. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. in *Advances in Neural Information Processing Systems* (eds. Guyon, I. et al.) vol. 30 (Curran Associates, Inc., 2017).
29. Yunus, M. M. et al. Automated classification of atherosclerotic radiomics features in coronary computed tomography angiography (CCTA). *Diagnostics (Basel, Switzerland)* **12**, (2022)
30. Jin, X. et al. Automatic coronary plaque detection, classification, and stenosis grading using deep learning and radiomics on computed tomography angiography images: a multi-center multi-vendor study. *Eur. Radiol.* **32**, 5276–5286 (2022).
31. Li, L. et al. Radiomic features of plaques derived from coronary CT angiography to identify hemodynamically significant coronary stenosis, using invasive FFR as the reference standard. *Eur. J. Radiol.* **140**, 109769 (2021).
32. Eslami, P. et al. Radiomics of coronary artery calcium in the framingham heart study. *Radiol. Cardiothorac. imaging* **2**, e190119 (2020).
33. Kocak, B. et al. METhodological RadiomICs Score (METRICS): a quality scoring tool for radiomics research endorsed by EuSoMII. *Insights Imaging* **15**, 8 (2024).


## Acknowledgements

None.

## Author contributions

AC: conceptualization, formal analysis, investigation, methodology, validation, visualization, writing – original draft, writing – review and editing. FLI: formal analysis, investigation, methodology, visualization, writing – review and editing. FR: formal analysis, software, visualization, writing – review and editing. MC: visualization, writing – original draft, writing – review and editing. GP: data curation, funding acquisition, writing – review and editing. GIC: data curation, writing – original draft, writing – review and editing. VDAC: conceptualization, funding acquisition, investigation, project administration, resources, supervision, visualization, writing – original draft, writing – review and editing.

## Declarations

## Competing interests

This work has been partially funded by Fondazione Regionale per la Ricerca Biomedica (Regione Lombardia), project ID 3432721. AC is funded by the National Plan for NRRP Complementary Investments (PNC, established with the decree-law 6 May 2021, n. 59, converted by law n. 101 of 2021) in the call for the funding of research initiatives for technologies and innovative trajectories in the health and care sectors (Directorial Decree n. 931 of 06–06-2022)—project n. PNC0000003—AdvaNced Technologies for Human-centrEd Medicine (project acronym: ANTHEM). This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them. Gianluca Pontone receives a Speaker honorarium and/or research grant from GE Healthcare and BRACCO, Heartflow, Boheringer. All other authors do not report any conflicts.

## Ethics approval

The protocols were approved by the Ethics Committee of the IRCCS European Institute of Oncology and Monzino Cardiology Center (registration number: R1061/19-CCM 11 25) of the participating center, under the ethical standards of the Declaration of Helsinki. Data acquisition followed the General Data Protection Regulation of the EU. All patients signed the informed consent.

## Consent for publication

All authors have read and approved the manuscript and agree with submission to Scientific Reports.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-32352-9.

**Correspondence** and requests for materials should be addressed to A.C.

**Reprints and permissions information** is available at www.nature.com/reprints.

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