Interdisciplinary Research in Medical Sciences Specialty
ISSN: 3062-4401
2026, Volume 6, Issue 1, Page No: 194-206

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

Explainable and Reproducible Machine Learning Framework for Assessment of Coronary Calcification and Segmental Stenosis on Computed Tomography Angiography

Santiago Morales[1]*, Laura Rojas[2], Camila Vega[1], Diego Molina[3]

[1]Department of Cardiovascular Imaging and Machine Learning, Faculty of Medicine, University of Buenos Aires, Buenos Aires, Argentina.
[2]Department of Coronary Calcification Assessment, Faculty of Medicine, Pontifical Catholic University of Chile, Santiago, Chile.
[3]Department of CT Angiography and Stenosis Prediction, Faculty of Medicine, National University of La Plata, La Plata, Argentina.

***E-mail** ✉ santiago.morales@med.uba.ar

Received: 12 February 2026; **Revised:** 02 May 2026; **Accepted:** 05 May 2026

ABSTRACT

CCTA has become a standard noninvasive imaging technique for the initial diagnosis and clinical management of coronary artery disease (CAD). Advanced machine learning techniques applied to CCTA data demonstrate considerable potential for precise quantitative evaluation of CAD. In a post hoc study involving 909 subjects from the SCOT-HEART trial (with a median follow-up period of 5.8 years), we initially examined the variability of CCTA-based imaging parameters in a reference group of 221 individuals characterized by zero coronary artery calcium scores, stenoses under 10%, and absence of any CAD findings on imaging. This evaluation spanned 21 distinct image processing configurations. We subsequently constructed and validated interpretable ML models designed to measure the extent of coronary calcification and degree of narrowing in the primary coronary artery branches (LMA, LCX, LAD, pRCA, mRCA). Comprehensive evaluation was performed on calcified plaque burden, stenosis severity, and myocardial infarction outcomes. Across the different processing approaches, a set of 549 robust and stable imaging features was determined. Performance of six machine learning classifiers (SVM, KNN, MLP, Naïve Bayes, gradient boosting, and LightGBM) was assessed for the prediction of coronary calcification and stenoses. The leading model delivered 84.2% accuracy along with an AUC of 0.973. Segment-level stenosis classification accuracy remained above 84.8% for all major vessels. Model performance showed only minor variations (less than 0.05) between those utilizing the complete feature set and those limited to stable features alone. SHAP interpretability analysis highlighted differing levels of importance for imaging characteristics and traditional clinical risk variables. The identified stable imaging features establish a reliable foundation for future development of ML-driven quantitative coronary assessments. The proposed interpretable machine learning models exhibited encouraging results in the quantification of coronary calcification and location-specific stenoses.

Keywords: Computed tomography angiography, Stable feature, Calcification quantification, Major coronary segments, Stenoses assessment, Interpretable machine learning model

How to Cite This Article: Morales S, Rojas L, Vega C, Molina D. Explainable and Reproducible Machine Learning Framework for Assessment of Coronary Calcification and Segmental Stenosis on Computed Tomography Angiography. Interdiscip Res Med Sci Spec. 2026;6(1):194-206. https://doi.org/10.51847/vlyNZ5PaMM

Introduction

Cardiovascular disease, particularly coronary artery disease (CAD), remains the foremost cause of death globally and a major source of long-term disability [1, 2]. Atherosclerosis constitutes the main underlying mechanism responsible for most cases of CAD [3]. Coronary computed tomography angiography (CCTA) offers an excellent noninvasive approach for evaluating atherosclerotic changes, delivering precise information on both plaque

composition and vessel narrowing [4]. Prior research has demonstrated that quantitative plaque assessment and pericoronary fat attenuation index can enhance risk stratification beyond conventional cardiovascular risk factors [5, 6]. Nevertheless, these methods fail to extract the complete range of detailed information available in CCTA datasets. Radiomic analysis of images, when integrated with machine learning techniques, shows strong potential for more comprehensive biomarker quantification, including detailed plaque characterization. Radiomics has already been applied to identify high-risk plaque features [7, 8] and to analyze pericoronary adipose tissue for distinguishing inflammatory versus fibrotic alterations, as well as for predicting major adverse cardiovascular events [9].

Current machine learning investigations into CAD exhibit two notable patterns. First, the majority of automated tools for detecting coronary stenosis or calcification employ deep learning architectures that handle feature extraction, reduction, and prediction in a single end-to-end pipeline, resulting in models that are largely opaque to clinicians. This has prompted concerns within the medical community about feature reliability and model transparency, issues that are especially relevant in cardiovascular imaging applications of AI [10]. Second, most studies evaluate stenosis on a per-patient basis, with relatively few addressing segment-specific analysis. Yet, evaluating stenosis at the individual segment level better supports precision medicine and tailored patient care. Clinical decision-making in CAD relies heavily on segment-level details such as the number of diseased vessels, exact lesion sites and severity, lesion length, and vessel tortuosity.

Achieving standardization is essential for improving the reproducibility of machine learning models in medical imaging and advancing the integration of imaging into personalized medicine [11]. Radiomic studies face several practical difficulties, however. The same radiomic feature can produce inconsistent values depending on variations in image acquisition protocols or preprocessing steps, which undermines feature stability. Additionally, radiomic pipelines often generate thousands of features with substantial redundancy, necessitating feature selection or dimensionality reduction. Various strategies have been explored, including interobserver variability filtering, ROI perturbation-based variance reduction, or correlation-driven selection [12]. The comparative effectiveness of these methods and their influence on downstream model performance remain insufficiently investigated, and outcomes can differ based on specific parameter choices. Model development further introduces variability due to subjective parameter tuning by developers. Collectively, these issues complicate efforts to standardize radiomics-based ML workflows for clinical use, even as initiatives like the Image Biomarker Standardization Initiative (IBSI) continue to provide guidance [13], alongside research on preprocessing effects in carotid CTA [14] and voxel intensity standardization in oncologic CT imaging [15]. Most prior work has also concentrated on diseased or lesion-specific images, with few studies leveraging truly healthy reference populations.

To address these limitations, the present study seeks to establish an interpretable and reproducible framework for quantitative CCTA analysis. This approach integrates radiomics and machine learning using data from the SCOT-HEART trial cohort and identifies a subset of stable coronary radiomic features based on their consistency in individuals without CAD. We evaluate the comparative performance of models using only stable features versus the full feature set for both patient-level classification (calcified plaque burden) and segment-level classification (stenoses). Finally, we apply Shapley analysis to quantify the marginal contributions of key features across six ML models, thereby improving model transparency and interpretability.

Materials and Methods

Data processing and study design

This post hoc investigation draws on the multicentre randomized SCOT-HEART trial. The study received approval from the relevant ethics committee, and all participants gave written informed consent. The trial enrolled patients presenting with symptoms suggestive of stable angina due to suspected coronary artery disease. Main results from the SCOT-HEART trial have been reported previously [16-20].

Computed tomography acquisition and analysis

All participants underwent coronary artery calcium scoring and CCTA on one 320-slice scanner (Aquilion ONE, Toshiba Medical Systems, Tochigi, Japan) or two 64-slice systems (Brilliance 64, Philips Healthcare, Netherlands; Biograph mCT, Siemens, Germany). Scans were acquired after sublingual GTN administration, with beta-blockers given when heart rate exceeded 60 beats per minute. Gating was performed prospectively or

retrospectively depending on heart rate and rhythm. Contrast-enhanced CCTA used 50–70 mL of iodinated contrast injected at 5.5–6.5 mL/s. Complete CCTA data were available for all participants in the imaging arm, with no missing imaging datasets. Clinical variables extracted from electronic health records via NHS Scotland's eDRIS service had missing continuous data (e.g., age, BMI) imputed using mean values; participants with missing categorical data or sparse missing entries were excluded.

Coronary calcium scores were computed from non-contrast scans using the Agatston method [21]. Scores were grouped into categories: 0–10, 11–100, 101–399, and >400 AU. Stenosis severity was graded segmentally as no stenosis (<10%), mild (10%–49%), moderate (50%–70%), or severe (>70%). The current analysis focused on five major coronary segments of highest prognostic relevance: left main artery (LMA), left circumflex artery (LCX), left anterior descending artery (LAD), proximal right coronary artery (pRCA), and mid-right coronary artery (mRCA).

Segmentation

A semi-automatic vessel segmentation pipeline incorporating the nnU-Net algorithm [22, 23] was employed to outline the coronary lumen. An experienced radiologist (J. R. W. M., with over 10 years of CCTA expertise) performed manual corrections to ensure accuracy. The periluminal region of interest (ROI) was defined as a 2-pixel radial outward extension from the outer vessel contour, encompassing the immediate perivascular area around the coronary wall to capture adjacent plaque and fat tissue while minimizing contouring variability. A total of 1834 radiomic features across seven categories were extracted from this ROI using PyRadiomics (version 3.1.0) [24]. For patient-level calcification assessment, the five major segments were combined into a single ROI. For segment-level stenosis analysis, features were extracted independently from each of the five segments (LMA, LCX, LAD, pRCA, mRCA). Image intensities were restricted to the –170 to 200 HU range to focus on relevant fat and plaque tissues while excluding lumen contamination. All computations were carried out using PyTorch on an NVIDIA RTX 3080 GPU and Intel Core i9-12900 CPU.

Stable feature test

Stable radiomic features resistant to variations in noise, reconstruction, and ROI definition were identified by testing 221 participants with zero calcium score (CACS = 0), stenosis <10%, and no CAD. Feature stability was assessed across 21 imaging configurations involving different resampling voxel sizes and interpolation methods. Images were resampled using both isotropic and anisotropic voxel dimensions covering common clinical resolutions, including up- and down-sampling scenarios. The 21 settings consisted of:

1. Original images without resampling.
2. Isotropic voxel sizes: (0.50, 0.50, 0.50) mm, (0.75, 0.75, 0.75) mm, (1.0, 1.0, 1.0) mm, (1.25, 1.25, 1.25) mm, and (1.5, 1.5, 1.5) mm.
3. For each resampling, four interpolation algorithms were tested: linear, B-spline, Welch windowed sinc, and nearest-neighbour.

These stable features served as a pre-selection strategy compared against using the complete radiomic feature set. A feature was defined as stable when the average variance across the 21 configurations in the 221 subjects was below 25% ($\Delta r < 0.25$), and highly stable when below 10% ($\Delta r < 0.1$).

Feature selection

Dimensionality reduction was carried out through a data-driven strategy involving Pearson correlation analysis followed by LASSO regression. In the Pearson correlation step, when features showed a correlation coefficient exceeding 0.9, only one representative feature was kept. The LASSO-CV approach integrated least absolute shrinkage and selection operator (LASSO) with cross-validation to boost both the robustness and computational efficiency of the selection process.

Comparison of approaches

To evaluate the impact of various data processing strategies, five distinct modeling configurations were tested: (1) the complete set of radiomic features with no selection applied, (2) only the stable radiomic features without additional data-driven reduction, (3) all radiomic features after data-driven selection, (4) stable radiomic features combined with data-driven selection, and (5) clinical variables integrated with the fourth configuration.

Building of machine learning models

Six different algorithms were utilized to construct and test models for quantifying coronary calcification and evaluating stenosis severity at the segmental level using CTA data. The algorithms included support vector machine (SVM), Gradient Boosting Classifier (GBC), Light Gradient Boosting Machine (LightGBM), K-nearest neighbours (KNN), Naive Bayes, and multilayer perceptron (MLP). SVM seeks the optimal separating hyperplane for class differentiation. GBC works by sequentially improving weak learners and focusing on previously misclassified instances through weight adjustments. KNN classifies new instances based on the majority label among its nearest neighbors. MLP is a feedforward neural network that captures intricate patterns via multiple hidden layers with nonlinear activation functions. LightGBM employs a gradient boosting framework with leaf-wise tree growth for greater speed and performance on large datasets. Naive Bayes is a fast probabilistic classifier based on the assumption of feature independence and Gaussian distribution modeling. Employing multiple algorithms allowed not only accurate prediction of calcification and branch-specific stenosis but also assessment of the generalizability of periluminal CCTA-derived features across varied machine learning architectures.
To mitigate overfitting and enhance generalizability, fivefold cross-validation was implemented. In each fold, 80% of the data was randomly allocated for training (with 20% of the training portion reserved as an internal test set), while the remaining 20% served as the validation set. Model performance was evaluated using receiver operating characteristic (ROC) curves and area under the curve (AUC) metrics.

Shapley additive explanations analysis
SHAP (SHapley Additive exPlanations), introduced by Lundberg and Lee [25], provides a unified framework for interpreting individual model predictions. It calculates the contribution of each feature to a specific prediction by applying concepts from cooperative game theory to assign fair Shapley values. In this context, SHAP explains how each input feature influences the model's output. For medical applications, inputs may consist of electronic health record variables or imaging-derived features. In the current study, radiomic features extracted from coronary CT scans were analyzed with SHAP to determine feature importance and prediction contributions across eight machine learning models.

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \quad (1)$$

In this framework, represents the explanation model, M is the number of input features, and z indicates the presence (1 or 0) of corresponding features. Here, the presence pertains to contexts such as images and textual data (for instance, in text, after one-hot encoding a word, not all words appear in every sentence); denotes the attribution value (Shapley value) for each j feature, and is a constant. Since tree models use structured data as input, all features are present for a sample ; thus, the formula can be written as follows:

$$g(x') = \phi_0 + \sum_{j=1}^{M} \phi_j \quad (2)$$

Cox proportional hazards regression was used to examine relationships between radiomic features, clinical risk factors, and the outcome of myocardial infarction. Univariate survival analysis was first conducted to identify candidate predictors, followed by multivariate Cox regression to evaluate their combined effects while adjusting for confounders. Bootstrap resampling was applied to evaluate the stability and variability of performance metrics, including the time-dependent concordance index.

Results and Discussion

Baseline characteristics of the study populations
The overall study population consisted of 909 participants, split into two groups.

1. a) Six hundred eighty-one participants (mean age 58.27 ± 9.11 years, 392 males [57.56%]) were allocated for model development and internal validation. Within this group, 193 had no stenosis, 175 had mild stenosis, 124 had moderate stenosis, and 189 had severe stenosis. Individuals with more advanced stenosis were more frequently male and had higher rates of hypertension and prior coronary artery disease. Fourteen participants

experienced the primary endpoint of fatal or nonfatal myocardial infarction. Complete baseline details are presented in **Table 1**.

2. b) An external validation cohort comprised 228 CCTA examinations performed on the two 64-detector row scanners.

Table 1. Characters baseline of participants' population

Characteristic	No significant CAD *N* = 193	Participants *N* = 681	Severe CAD *N* = 189	Moderate CAD *N* = 124	Mild CAD *N* = 175	*p*-value
Age	53.04 ± 9.71	58.27 ± 9.11	61.20 ± 7.36	58.83 ± 8.64	60.67 ± 7.93	< 0.001
Gender						< 0.001
Female — *n* (%)	118 (61.14)	289 (42.44)	46 (24.34)	46 (37.10)	79 (45.14)	
Male — *n* (%)	75 (38.86)	392 (57.56)	143 (75.66)	78 (62.90)	96 (54.86)	
Total cholesterol (mmol/L)	4.88 ± 1.95	4.96 ± 1.96	5.0 ± 1.99	5.14 ± 1.72	4.84 ± 2.14	0.451
HDL cholesterol	0.87 ± 0.76	0.87 ± 0.70	0.84 ± 0.65	0.98 ± 0.65	0.83 ± 0.70	0.255
Body mass index (kg/m^2)	29.09 ± 5.72	29.23 ± 5.01	29.25 ± 4.18	29.61 ± 5.28	29.06 ± 4.84	0.779
Hypertension — *n* (%)	49 (25.39)	234 (34.36)	82 (43.39)	46 (37.10)	57 (32.57)	< 0.001
CHD prior history — *n* (%)	4 (2.07)	62 (9.10)	30 (15.87)	18 (14.52)	10 (5.71)	< 0.001
Smoking history — *n* (%)	95 (49.22)	372 (54.63)	104 (55.03)	84 (67.74)	89 (50.86)	0.0062
CHD family history — *n* (%)	75 (38.86)	275 (40.38)	81 (42.86)	46 (37.10)	73 (41.71)	0.875
Diabetes — *n* (%)	20 (10.36)	83 (12.19)	25 (13.23)	16 (12.90)	22 (12.57)	0.832
Cerebrovascular disease — *n* (%)	5 (2.59)	32 (4.70)	7 (3.7)	11 (8.87)	9 (5.14)	0.083
CAC score (AU)	2 ± 13	286 ± 706	737 ± 1070	343 ± 696	70 ± 78	< 0.001

1. Mean (standard deviation); number (percent); *CAC* coronary artery calcium, *CHD* coronary heart disease, *BMI* body mass index

Stable features in participants without CAD

We evaluated the distribution of radiomic features derived from CTA scans in a group of 221 individuals who had zero coronary artery calcium scores (CACS = 0), coronary stenoses below 10%, and no evidence of CAD. These assessments were performed across 21 varied imaging configurations to determine feature consistency. The goal was to pinpoint reliable features that could improve the reproducibility of downstream machine learning models. We also examined how these CTA-derived imaging characteristics were influenced by differences in voxel dimensions and interpolation techniques. Out of the initial 1834 radiomic features, 549 proved stable under all 21 imaging settings ($\Delta r < 0.25$); **(Figure 1)**. For any given imaging configuration among the 21 tested, roughly 66 percent to 77 percent of features demonstrated stability, while approximately 20% to 33% qualified as highly stable **(Figure 2)**.

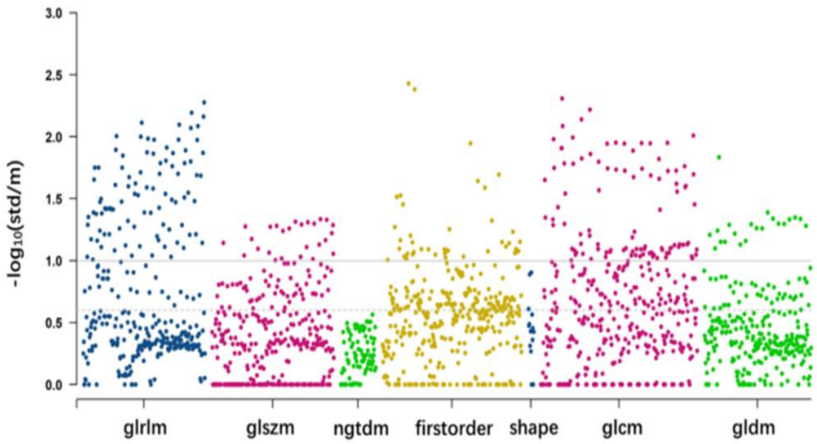
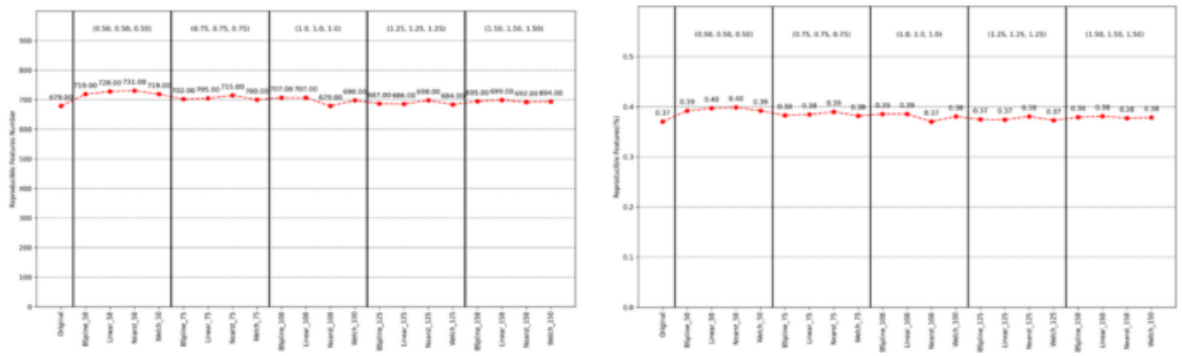


Figure 1. Manhattan plot illustrating the stability distribution of the 1834 radiomic features across the study population without CAD.

The x-axis displays the seven radiomics feature categories. Each individual point corresponds to one radiomic feature. The dotted horizontal line indicates the threshold for stable features, while the solid line marks the threshold for highly stable features.



a) b)

Figure 2. Stability Statistics and Percentage Plot of Features in the Population without CAD

Clinical and biological relevance of stable radiomic features

We evaluated the practical clinical importance of the stable radiomic features through Cox proportional hazards models, testing their link to myocardial infarction (MI) events across the 5-year follow-up window. Key outcomes were expressed as hazard ratios with accompanying 95% confidence intervals and illustrated via forest plots. A number of these features displayed meaningful prognostic associations, demonstrating their capacity to help distinguish individuals facing higher MI likelihood.

To further establish their biological grounding, we assessed how these stable features related to coronary artery calcium score (CACS), a recognized indicator of overall atherosclerotic plaque amount. Findings, revealed several strong statistically significant correlations. Notably, squareroot_ngtdm_Strength correlated positively at Spearman $\rho = 0.849$, and gradient_ngtdm_Strength at $\rho = 0.803$, implying that stronger myocardial texture patterns align with greater calcification levels. In comparison, log_sigma_2_0_mm_3D_glcm_InverseVariance showed a clear negative link ($\rho = -0.722$), which could indicate lower tissue uniformity in cases with heavier plaque accumulation.

These combined evaluations confirm that the selected stable features possess strong statistical reliability along with clear clinical predictive power and biological meaning, strengthening their value for improved cardiovascular risk profiling and patient grouping.Comparison of Feature Selection Pipelines for Predicting Coronary Artery Calcification Score Category

Participants showed an overall average coronary artery calcium score of 286 ± 706 AU. The scores were distributed as follows: 201 people in the 0–10 range, 165 in the 11–100 range, 125 in the 101–399 range, and 109 above 400. The effectiveness of the five distinct feature selection strategies was benchmarked using the six standard machine learning algorithms **(Table 2)**.

Table 2. Performance of Machine Learning Models for Coronary Calcification Assessment

Machine Learning Models	All Features Based Test (95% CI)	All Features Based Train (95% CI)	Unstable Features Based Test (95% CI)	Unstable Features Based Train (95% CI)	Stable Features Based Test (95% CI)	Stable Features Based Train (95% CI)	Stable + Clinical Features Based Test (95% CI)	Stable + Clinical Features Based Train (95% CI)
SVM	0.842 (0.775, 0.897)	0.885 (0.855, 0.909)	0.733 (0.654–0.802)	0.754 (0.715–0.793)	0.800 (0.72, 0.86)	0.883 (0.853, 0.907)	0.808 (0.735, 0.866)	0.908 (0.881, 0.930)
KNN	0.750 (0.671, 0.815)	0.827 (0.794, 0.857)	0.708 (0.627–0.779)	0.723 (0.683–0.763)	0.742 (0.663, 0.809)	0.800 (0.764, 0.831)	0.708 (0.624, 0.776)	0.812 (0.778, 0.843)

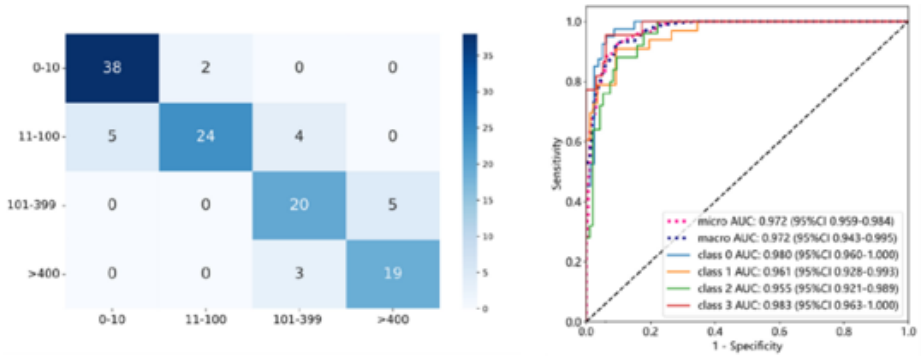
LightGBM	0.825 (0.751, 0.878)	0.931 (0.906, 0.949)	0.735 (0.656–0.794)	0.848 (0.816–0.880)	0.752 (0.671, 0.815)	0.908 (0.881, 0.930)	0.752 (0.671, 0.815)	0.919 (0.893, 0.939)
Gradient Boosting	0.750 (0.671, 0.815)	0.896 (0.867, 0.918)	0.703 (0.612–0.775)	0.817 (0.782–0.852)	0.743 (0.663, 0.809)	0.892 (0.863, 0.915)	0.733 (0.655, 0.802)	0.885 (0.855, 0.909)
MLP	0.833 (0.759, 0.885)	0.883 (0.853, 0.907)	0.750 (0.673–0.807)	0.783 (0.746–0.819)	0.783 (0.703, 0.841)	0.871 (0.841, 0.897)	0.800 (0.727, 0.860)	0.869 (0.839, 0.895)
Naïve Bayes	0.785 (0.711, 0.847)	0.800 (0.764, 0.831)	0.708 (0.627–0.789)	0.723 (0.683–0.763)	0.767 (0.687, 0.828)	0.783 (0.747, 0.816)	0.783 (0.703, 0.841)	0.796 (0.760, 0.828)

Among the five modeling strategies tested, the approach relying on the complete set of radiomic features without any reduction performed the worst overall. This configuration displayed clear overfitting, reaching perfect training accuracy of 1.0 while experiencing a substantial decline in performance on the test set — the lowest accuracy observed among all five pipelines across the six machine learning algorithms.

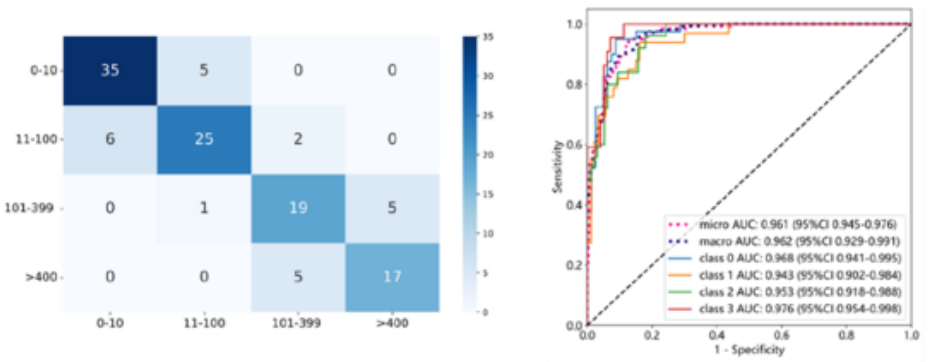
The model built exclusively on stable radiomic features (without additional data-driven reduction) also exhibited overfitting, with training accuracies ranging from 0.98 to 1.0. Although it showed a noticeable drop in test set performance, the decline was slightly less severe compared to using the full radiomic feature set.

When comparing the two pipelines that incorporated data-driven feature selection (Pearson correlation combined with LASSO), the version starting from all radiomic features achieved superior accuracy in both training and testing phases, along with a smaller performance gap between them. On the test set, the accuracy advantage of this pipeline over the one using pre-filtered stable features ranged from 0.007 to 0.073 across the six algorithms. This improvement was similar in scale to the difference observed between the strongest and weakest performing models overall (0.092).

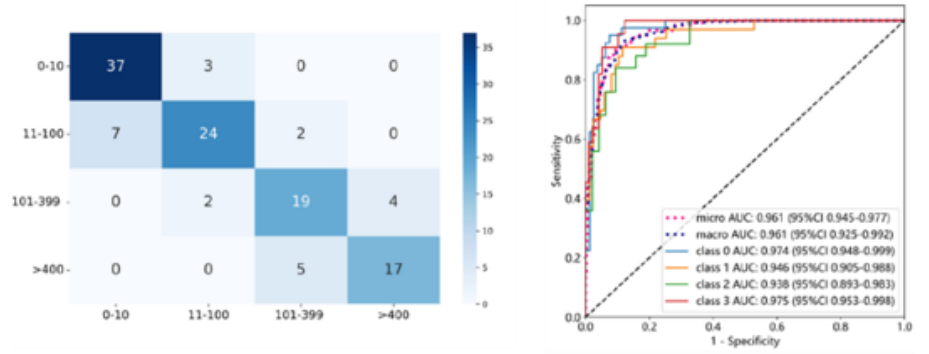
Across the six algorithms, SVM, MLP, and LightGBM delivered accuracies above 80 percent, whereas KNN, Naïve Bayes, and gradient boosting reached between 75 percent and 80 percent. The SVM model stood out as the top performer, attaining an overall four-class classification accuracy of 84.2 percent for predicting coronary artery calcium score categories. Its per-class AUC values were: class_0 (CACS 0–10): 0.980; class_1 (CACS 11–100): 0.961; class_2 (CACS 101–399): 0.955; and class_3 (CACS ≥ 400): 0.983. Both the micro-average and macro-average AUCs reached 0.972 **(Figure 3)**.



a)



b)



c)

Figure 3. Performance of Coronary Calcification Quantification in Machine Learning Models.

Top left: Heatmap displaying prediction outcomes from the all-feature model on the test set. Top right: AUC curves for the all-feature model in the test set. Middle: Performance metrics for the model utilizing stable features. Bottom: Results for the model combining stable radiomic features with clinical risk factors.

The heatmaps present the classification outcomes across the four calcium score categories, with each cell indicating the frequency of correct predictions versus actual labels. This visualization highlights categories the model identifies most reliably and areas prone to errors. ROC curves demonstrate the model's capacity to separate the different classes, while AUC values quantify overall discriminative ability (ranging from 0.5 for random guessing to 1.0 for perfect separation). AUC scores above 0.9 signify excellent performance. The macro-AUC provides an unweighted average across all classes, whereas the micro-AUC evaluates global performance by accounting for all samples collectively, placing greater emphasis on more prevalent categories. These metrics together deliver a thorough assessment of the model's classification reliability.

When contrasting models built on stable imaging features alone against those integrating both clinical risk factors and stable features, the addition of clinical variables produced only modest gains in calcium score severity prediction. Performance still fell short of models that began with the full radiomic feature set. For the SVM algorithm specifically, incorporating clinical factors improved training accuracy by 0.025 and test accuracy by only 0.008 relative to the stable-features-only model. The top-performing SVM model maintained strong results across datasets, achieving 0.811 accuracy and 0.939 AUC.

Comparison of feature selection for predicting coronary artery stenosis

This analysis encompassed 3083 coronary artery segments in total. Of these, 2038 segments showed stenosis below 10%, 760 segments had 10–49% stenosis, and 285 segments had stenosis exceeding 50%. The prediction models attained high accuracy when evaluating stenosis severity across the five primary coronary artery branches. In the test set, accuracy surpassed 80% for all territories **(Table 3 and Figure 4)**.

Table 3. Performance for major coronary segments' stenosis classification in SVM model

Segments	Total	Stenosis ≥50%	Stenosis 10–49%	Stenosis <0%	Test accuracy (95% *CI*)	Train accuracy (95% *CI*)
All segments	3083	285	760	2038	86.7% (83.8–89.2%)	87.3% (85.9–88.6%)

pRCA	614	36	132	446	84.4% (77.1–89.9%)	86.7% (83.5–89.5%)
mRCA	658	83	130	445	84.2% (76.9–89.4%)	86.9% (83.7–89.5%)
LMA	518	14	89	415	92.7% (85.6–96.1%)	93.9% (91.2–95.9%)
LCX	643	44	149	450	89.2% (82.6–93.4%)	92.9% (90.5–94.9%)
LAD	650	108	260	282	84.5% (77.4–89.8%)	85.8% (82.5–88.5%)

1. *pRCA* right coronary artery-proximal segment, *mRCA* right coronary artery-mid segment, *LMA* left coronary artery, *LCX* left circumflex artery, *LAD* left anterior descending artery

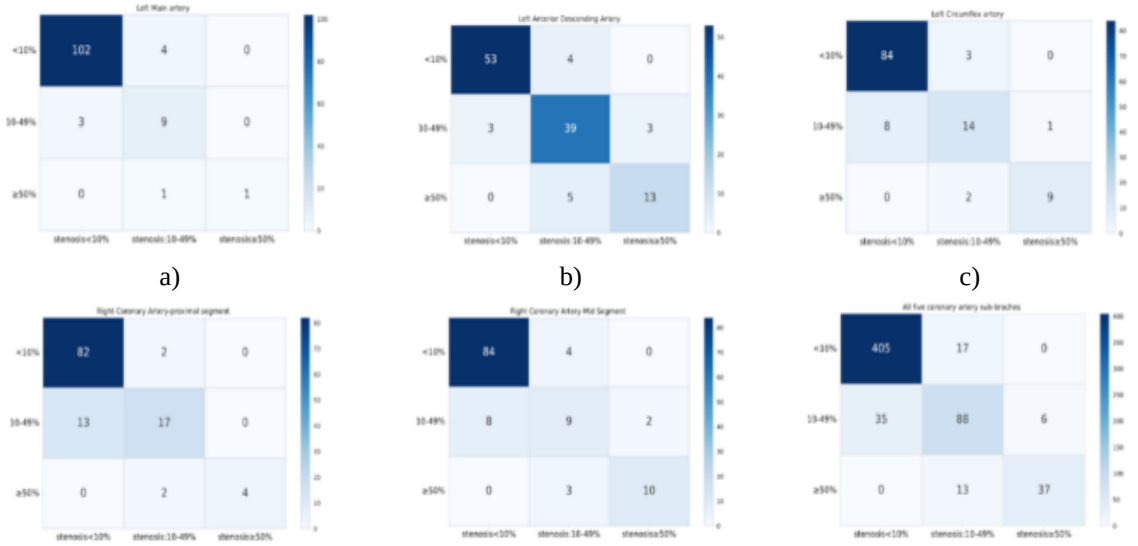
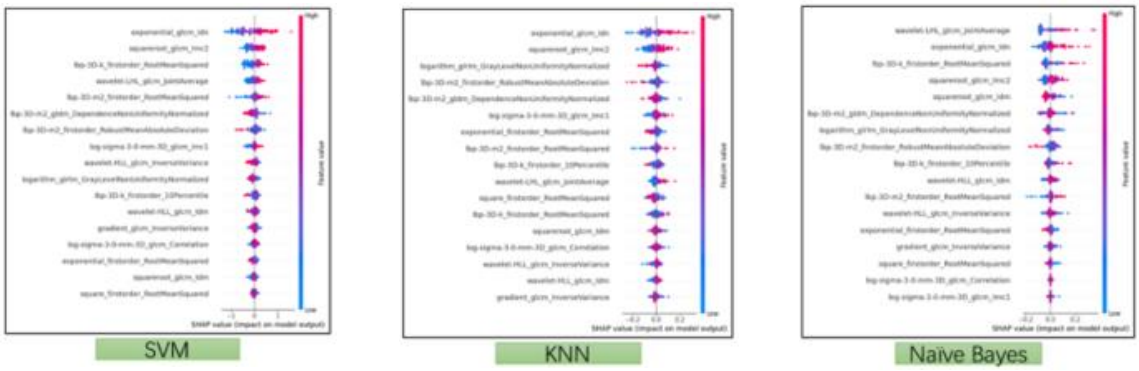
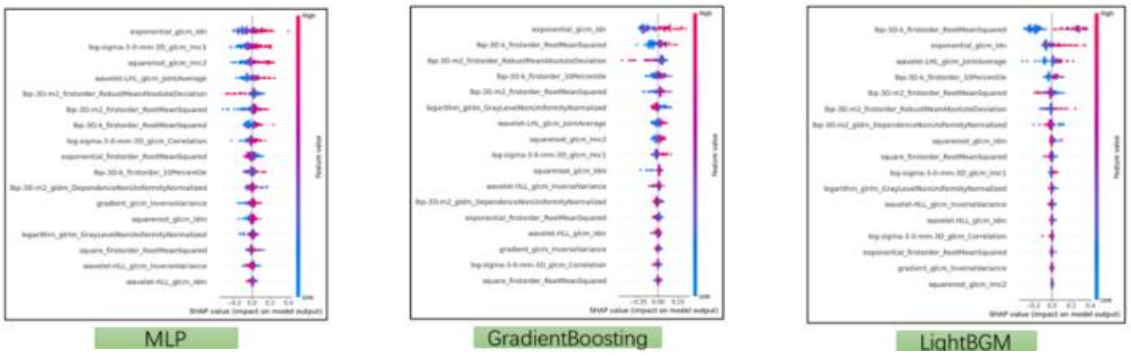


Figure 4. Performance Evaluation of Machine Learning Models for Segment-Level Coronary Stenosis

Six heatmaps display the classification performance of the models regarding stenosis severity in each of the five key coronary artery segments (LMA, LCX, LAD, pRCA, and mRCA). The panel in the bottom right presents the combined prediction outcomes for all five segments together.

Comparison between the model using the full collection of radiomic features and the model limited to stable radiomic features revealed highly comparable results for stenosis evaluation in the five major coronary branches. Accuracy differences stayed under 0.03 in both the training and testing datasets. This level of consistency was maintained uniformly across the six machine learning algorithms included in the analysis. Shapley value analysis further indicated that the influence of specific imaging features on model predictions showed considerable variation depending on the particular algorithm employed **(Figures 5 and 6)**.





d) e) f)

Figure 5. Shapley summary plot showing radiomic feature importance.

Each point corresponds to an individual participant. The plot indicates the direction and magnitude of each feature's contribution to the model output, where red dots represent higher feature values and blue dots represent lower feature values.

a) b) c)

d) e) f)

Figure 6. Local Model Explanation by the Shapley Method

In the present work, we thoroughly assessed the consistency of coronary and perivascular radiomic features obtained from CCTA scans under 21 distinct image preprocessing conditions in subjects free of CAD. We also explored the impact of feature stability on the effectiveness of machine learning models for evaluating coronary calcification and stenosis. Notably, selecting features based on their stability resulted in only minimal reductions in performance during both training and testing phases, irrespective of the specific algorithm employed. These outcomes indicate that emphasizing feature stability can strengthen model reproducibility and applicability across settings while preserving overall diagnostic capability.

To our knowledge, this represents the first in-depth examination of the distribution and robustness of periluminal coronary radiomic features extracted from CCTA in a population without CAD. Out of 1834 initial features, 549 were determined to be stable across the various preprocessing setups. The identified stable features offer a useful benchmark for subsequent quantitative CCTA imaging research. Earlier reproducibility studies have primarily addressed oncologic imaging [14] or carotid vessels [15], with comparatively little attention given to cardiovascular contexts. Our investigation characterizes CCTA feature behavior in CAD-free individuals, selects stable candidates, and validates them using multiple machine learning techniques in affected patients. This may help advance standardization efforts in cardiovascular radiomics.

Models constructed with these stable features delivered solid results in coronary calcification quantification, reaching accuracies above 80% in the four-class calcium scoring task. Their performance closely matched that of models using the entire feature set, confirming that well-chosen stable features preserve essential discriminatory information for accurate calcium assessment. Several features also correlated strongly with CACS, an established

measure of atherosclerotic plaque load. These observations underscore the value of stable radiomic features for more reliable automated calcium evaluation and sustained cardiovascular risk monitoring. While prior deep learning studies have reported strong results in automated calcium scoring [26-28], they have often been constrained by modest sample sizes and limited interpretability. Our study benefits from a large randomized controlled trial dataset acquired across multiple scanner types and imaging conditions, thereby improving the robustness and generalizability of the findings.

Beyond calcification, we applied the approach to segment-level coronary stenosis classification. Analyzing over 3000 coronary segments from the SCOT-HEART trial, our interpretable ML models achieved high accuracy in stenosis grading across the major branches (LMA, LCX, LAD, pRCA, and mRCA). This segment-specific evaluation offers advantages over most earlier ML investigations, which typically focused on patient-level stenosis using invasive angiography or CT data [29].

To promote transparency, we utilized Shapley value analysis to determine the individual contribution of each imaging feature to model outputs. This revealed that feature importance differed substantially among the six algorithms, consistent with their unique underlying learning strategies. Such interpretability aids in clarifying how imaging-derived metrics drive predictions and can foster greater confidence in AI-supported cardiovascular diagnostics. Shapley methods have been previously used in studies of kidney injury [30] and liver inflammation [31]; our application extends this technique to CCTA imaging.

Overall, our results emphasize the value of selecting features that are both statistically stable and clinically meaningful. Features that demonstrate consistency despite variations in imaging protocols are more likely to transfer well across different centers, populations, and scanners. This strategy helps minimize overfitting to particular acquisition conditions, thereby improving reproducibility and clinical trust. From a practical standpoint, such reproducible and interpretable ML systems could facilitate more uniform quantitative CCTA analyses in varied clinical environments.

Several limitations of the study should be noted. The sample size and variety of scanners were restricted. Larger, more diverse multicentre studies incorporating broader imaging protocols and patient demographics will be important for further validation. Although we employed one established feature reduction method (Pearson correlation combined with LASSO) for clarity, future research could test additional techniques such as mutual information, recursive feature elimination, or tree-based selection to strengthen model performance. Prospective comparisons between the AI models and experienced radiologists will also be valuable for evaluating interpretability, reader variability, and real-world consistency. Lastly, validation in more heterogeneous groups, including patients with both stable and unstable plaques, will be necessary to confirm the robustness of these features across the full range of coronary artery disease.

Conclusion

We identified a set of stable coronary and periluminal radiomic features from CCTA scans in individuals without CAD. Machine learning models built using these stable features achieved strong performance in quantifying coronary calcification and in assessing stenosis within the major coronary segments, with accuracy levels comparable to models using unrestricted feature sets. These results underscore the promise of reproducible and interpretable approaches for enhancing CCTA-based research and clinical evaluation of coronary artery disease.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. 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. 2018;392(10159):1736-88.

2. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. J Am Coll Cardiol. 2020;76(25):2982-3021.
3. Jensen RV, Hjortbak MV, Bøtker HE. Ischemic heart disease: an update. Semin Nucl Med. 2020;50(3):195-207.
4. Suhny A, Leslee JS. Past, present, and future of CTA. Circulation. 2024;150(2):87-90.
5. Greenland P, Blaha MJ, Budoff MJ, Erbel R, Watson KE. Coronary calcium score and cardiovascular risk. J Am Coll Cardiol. 2018;72(4):434-47.
6. Sandoval Y, Bielinski S, Daniels L. Atherosclerotic cardiovascular disease risk stratification based on measurements of troponin and coronary artery calcium. J Am Coll Cardiol. 2020;76(4):357-70.
7. Kolossváry M, Karády J, Szilveszter B, Kitslaar P, Hoffmann U, Merkely B, et al. Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign. Circ Cardiovasc Imaging. 2017;10(12):e006843.
8. Dey D, Commandeur F. Radiomics to identify high-risk atherosclerotic plaque from computed tomography: the power of quantification. Circ Cardiovasc Imaging. 2017;10(12):e006843.
9. Oikonomou EK, Williams MC, Kotanidis CP, Desai MY, Marwan M, Antonopoulos AS, et al. A novel machine learning-derived radiotranscriptomic signature of perivascular fat improves cardiac risk prediction using coronary CT angiography. Eur Heart J. 2019;40(43):3542-54.
10. Hunter DJ, Holmes C. Where medical statistics meets artificial intelligence. N Engl J Med. 2023;389(13):1211-9.
11. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-62.
12. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology. 2018;288(1):203-11.
13. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology. 2020;295(2):328-38.
14. Le EPV, Rundo L, Tarkin JM, Evans NR, Gopalan D, Sala E, et al. Assessing robustness of carotid artery CT angiography radiomics in the identification of culprit lesions in cerebrovascular events. Sci Rep. 2021;11(1):3499.
15. Escudero Sanchez L, Rundo L, Gill AB, Hoare M, Mendes Serrao E, Sala E, et al. Robustness of radiomic features in CT images with different slice thickness, comparing liver tumour and muscle. Sci Rep. 2021;11(1):8262.
16. SCOT-HEART Investigators. CT coronary angiography in patients with suspected angina due to coronary heart disease (SCOT-HEART): an open-label, parallel-group, multicentre trial. Lancet. 2015;385(9985):2383-91.
17. SCOT-HEART Investigators. Coronary CT angiography and 5-year risk of myocardial infarction. N Engl J Med. 2018;379(10):924-33.
18. Williams MC, Kwiecinski J, Doris M, McElhinney P, D'Souza MS, Cadet S, et al. Low-attenuation noncalcified plaque on coronary computed tomography angiography predicts myocardial infarction. Circulation. 2020;141(18):1452-62.
19. Williams MC, Kwiecinski J, Doris M, McElhinney P, D'Souza MS, Cadet S, et al. Sex-specific computed tomography coronary plaque characterization and risk of myocardial infarction. JACC Cardiovasc Imaging. 2021;14(9):1804-14.
20. Tzolos E, Williams MC, McElhinney P, Lin A, Grodecki K, Cadet S, et al. Pericoronary adipose tissue attenuation, low-attenuation plaque burden, and 5-year risk of myocardial infarction. JACC Cardiovasc Imaging. 2022;15(6):1078-88.
21. Woodward M, Brindle P, Tunstall-Pedoe H. Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). Heart. 2007;93(2):172-6.
22. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203-11.

23. Shi Y. The effect of subbranch for the quantification of local hemodynamic environment in the coronary artery: a computed tomography angiography–based computational fluid dynamic analysis. Emerg Crit Care Med. 2022;2(4):181-90.
24. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 2017;77(21):e104-7.
25. Lundberg S, Lee S-I. Consistent feature attribution for tree ensembles. Nat Commun. 2017;8:3031.
26. Eng D, Chute C, Khandwala N, Rajpurkar P, Long J, Shleifer S, et al. Automated coronary calcium scoring using deep learning with multicenter external validation. NPJ Digit Med. 2021;4(1):88.
27. Mu D, Junjie B, Wenping C. Calcium scoring at coronary CT angiography using deep learning. Radiology. 2022;302(2):309-16.
28. Biavati F, Saba L, Boussoussou M. Coronary artery calcium score predicts major adverse cardiovascular events in stable chest pain. Radiology. 2024;310(1):e231557.
29. Zreik M. A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography. IEEE Trans Med Imaging. 2019;38(7):1588-98.
30. Hu J, Xu J, Li M. Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. EClinicalMedicine. 2024;68:102409.
31. Rui F, Yeo YH, Xu L, Huang R. Development of a machine learning-based model to predict hepatic inflammation in chronic hepatitis B patients with concurrent hepatic steatosis: a cohort study. EClinicalMedicine. 2024;68:102419.