Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cross-modal image-text collaborative deep-learning model for predicting severe coronary stenosis from CCTA: a retrospective single-center study

Li Ma
The Fifth People's Hospital of Huai'an
Jiangpeng Wang
The Fifth People's Hospital of Huai'an
Dongdong Tang
The Fifth People's Hospital of Huai'an
Huan Tang
The Fifth People's Hospital of Huai'an
Haibao Yuan
The Fifth People's Hospital of Huai'an
Linghui Zhao
xnkzlh@163.com

The Fifth People's Hospital of Huai'an

Research Article

Keywords: Coronary computed tomography angiography, Deep learning, Cross-modal learning, Radiomics, Coronary stenosis, Artificial intelligence

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11431659/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Coronary computed tomography angiography (CCTA) is a cornerstone non-invasive modality for coronary artery evaluation. Nevertheless, conventional manual reading suffers from subjective bias and limited reproducibility. We proposed a cross-modal deep-learning framework integrating image and clinical text information to improve detection of severe coronary stenosis.

Objective

To develop and validate a cross-modal radiomics deep-learning model based on CCTA data and further evaluate its clinical performance for identifying severe coronary artery stenosis.

Methods

In this single-center retrospective study, 96 patients receiving both CCTA and invasive coronary angiography (ICA) were enrolled, yielding 288 raw coronary vessel-level samples; two vessels were excluded for vessel-specific poor image quality, resulting in 286 evaluable coronary vessel-level (unit of analysis) samples. A dual-branch cross-modal model combining ResNet-18 image encoder and lightweight Transformer text encoder was constructed. Class-differentiated augmentation and multi-loss joint optimization were applied to mitigate class imbalance. Dataset was split into training/test set (8:2). Using ICA as reference gold-standard, ROC-based diagnostic metrics (AUC, sensitivity, specificity, accuracy, PPV, NPV) were calculated for model and manual interpretation.

Results

Our cross-modal deep-learning model achieved AUC = 0.83 (95%CI 0.71-0.94), sensitivity 64.0%, specificity 93.0%, accuracy 79.0%, PPV 89.0%, NPV 74.0%. It outperformed manual CCTA reading (AUC = 0.676; sensitivity 46.38%, specificity 88.7%, accuracy 68.5%).

Conclusion

The proposed cross-modal image-text collaborative deep-learning model improves diagnostic performance for severe coronary stenosis. It could serve as a promising intelligent auxiliary tool for CCTA-based non-invasive assessment of coronary artery disease.

Introduction

Coronary heart disease (CHD) remains the leading global cause of mortality and permanent disability, with its pathological basis rooted in lipid-driven chronic immune-mediated inflammation and fibroproliferative atherosclerosis (1). Epidemiological data show that the absolute number of global all-cause deaths attributable to CHD increased by 41.5% between 2000 and 2023. This trend is largely attributable to population aging and the high global prevalence of modifiable metabolic risk factors including hypertension, dyslipidaemia, type 2 diabetes mellitus, and long-term tobacco exposure(2). In contrast to rising absolute mortality burden, the age-standardized CHD mortality rate decreased by 28.9%, and age-standardized years of life lost (YLL) secondary to premature CHD death decreased by 27.4% over the same timeframe (2). This paradoxical epidemiological pattern suggests that population-wide cardiovascular management strategies, including community-based CHD screening, guideline-directed lipid-lowering and antihypertensive therapy, acute coronary syndrome emergency care, and tiered preventive interventions, have delivered measurable survival benefits at the population level.

Within the full diagnostic and therapeutic spectrum of CHD, timely, quantitative and reproducible non-invasive cardiovascular imaging serves as an irreplaceable cornerstone for early disease screening, cardiovascular risk stratification and personalized clinical decision-making. Coronary computed tomography angiography (CCTA) has been established as the first-line non-invasive anatomical imaging modality for patients presenting with stable chest pain or suspected obstructive CHD, enabling comprehensive volumetric visualization of the left anterior descending artery (LAD), left circumflex artery (LCX), right coronary artery (RCA) and all their secondary branches(3). Beyond identifying luminal obstructive stenosis, CCTA supports comprehensive qualitative phenotyping and volumetric quantification of atherosclerotic plaques, allowing discrimination of high-risk lesions such as calcified plaques, non-calcified plaques and low-attenuation vulnerable plaques, which secures its central role in contemporary CHD management workflows (4).

Despite widespread clinical adoption of CCTA, conventional manual image interpretation carries inherent drawbacks that impair diagnostic reproducibility and sensitivity, thereby creating an urgent clinical demand for deep learning-powered intelligent computer-aided diagnosis. As reported by Shrivastava et al.(5), traditional semi-automated plaque quantification relies heavily on labor-intensive manual slice-by-slice annotation by radiologists; complete image analysis for a single patient takes more than 25 minutes, which substantially increases radiologists' workload and prolongs diagnostic waiting time for both outpatient and emergency patients. Furthermore, subjective visual assessment by readers and inter-individual anatomical variations inevitably introduce prominent inter-observer discrepancies in CCTA interpretation. When evaluating heavily calcified coronary segments, junior radiologists exhibit markedly lower detection rates of high-grade obstructive stenosis, with up to 40% disagreement in stenosis grading between trainee readers and senior attending physicians. Such inconsistency may lead to missed haemodynamically significant lesions and delayed referral for invasive revascularization(6).

Recent advances in deep learning architectures, including convolutional neural networks (CNNs), recurrent convolutional neural networks and lightweight Transformer encoders, enable end-to-end fully automatic processing of raw CCTA volumetric datasets. A unified algorithmic pipeline can sequentially

complete coronary centerline extraction, vessel lumen segmentation, volumetric quantification of total and component-specific plaques, and categorical stenosis grading(5, 7–9). Head-to-head validation trials against intravascular ultrasound(IVUS)—the invasive gold standard for luminal and plaque quantification—have verified excellent concordance between deep learning-derived plaque and stenosis metrics and IVUS reference values. Meanwhile, automated deep learning analysis shortens CCTA interpretation from dozens of minutes to merely seconds, delivering diagnostic performance comparable to or even superior to that of senior cardiovascular radiologists (6, 10).

Owing to their outstanding capacity for automated hierarchical radiomic feature extraction and lesion pattern recognition without handcrafted imaging descriptors, deep learning techniques have been widely adopted across all subfields of medical imaging over the past decade. For CCTA-specific clinical scenarios, deep learning algorithms are capable of overcoming core limitations of human visual reading, such as blooming artifacts induced by severe calcification, detection of subtle eccentric plaques, and inconsistent stenosis grading thresholds among different readers. Accordingly, deep learning models possess promising translational potential for standardized, high-performance intelligent quantitative evaluation of coronary artery lesions in routine clinical practice.

Materials and Methods

1.1 Study Population

We retrospectively enrolled consecutive patients who underwent both CCTA and invasive coronary angiography (ICA) at The Fifth People's Hospital of Huai'an from 1 January 2024 to 30 December 2024. The unit of analysis for diagnostic performance was coronary vessel-level rather than patient-level. A total of 96 patients fulfilled patient-level eligibility criteria, yielding 288 raw coronary vessel samples (LAD, LCX, RCA). Two vessel samples were further excluded due to vessel-specific poor image quality that precluded feature extraction, leaving 286 evaluable vessel-level CCTA samples covering the left anterior descending artery (LAD), left circumflex artery (LCX), and right coronary artery (RCA). According to the Coronary Artery Disease Reporting and Data System (CAD-RADS) criteria, coronary stenosis was dichotomized into non-severe stenosis (< 70%) and severe stenosis (≥ 70%). The study protocol received ethical approval from the Institutional Review Board of The Fifth People's Hospital of Huai'an (Ethics reference number: KY-P-2026-51). Given the retrospective nature of this study using de-identified clinical data, the requirement for written informed consent was waived. Inclusion criteria were as follows: (I) age between 18 and 80 years; (II) complete CCTA acquisition with diagnostic-quality images sufficient for coronary segmentation, plaque analysis, and stenosis quantification; and (III) invasive CAG performed within 1 week after CCTA scanning. Exclusion criteria were as follows: (I) severe renal insufficiency or iodinated contrast allergy precluding standard contrast-enhanced CCTA; (II) prior cardiac bypass grafting, percutaneous coronary stent implantation, severe congenital heart disease, or cardiomyopathy; and (III) suboptimal CCTA image quality prohibiting coronary segmentation and feature extraction. A STARD-compliant flowchart depicting patient-level screening followed by subsequent vessel-level exclusions is shown in Fig. 1.

1.2 CCTA Image Acquisition

All CCTA examinations were performed using a Siemens SOMATOM Force dual-source CT scanner. According to cardiovascular CT clinical guidelines, pre-scan beta-blockers, sublingual nitroglycerin, and intravenous iodinated contrast medium were administered when clinically indicated. The scan range extended from the inferior margin of the tracheal carina to the cardiophrenic sulcus. Scan parameters included tube voltage of 100–120 kV with automatic tube-current modulation. Contrast medium was injected via an antecubital vein using a high-pressure injector; scan triggering was synchronized to contrast bolus arrival within the aortic root. Raw CT datasets were reconstructed with a slice thickness of 0.625 mm and transferred to a dedicated post-processing workstation. Two radiologists with ≥ 5 years of cardiovascular-imaging experience independently assessed stenosis severity for each coronary vessel. Consensus reading was performed to resolve diagnostic discrepancies. Interventional cardiologists performing reference-standard ICA were blinded to CCTA manual readings and deep-learning model outputs. Invasive ICA performed by interventional cardiologists following standardized angiographic projection protocols served as the diagnostic gold standard for severe coronary stenosis.

1.3 Statistical Analysis

Statistical analyses were performed using SPSS version 25.0. A two-tailed $P < 0.05$ was defined as the threshold for statistical significance. Continuous variables were presented as mean ± standard deviation; categorical variables were summarized as counts and percentages [*n*(%)]. The chi-square test was applied for inter-group comparison of categorical variables. ROC curve analysis was implemented to evaluate model predictive performance, and the AUC, accuracy, sensitivity, and specificity were computed. Calibration curves were plotted to assess agreement between model-predicted risk probabilities and actual clinical outcomes, and the calibration result was reported in the Results section. All statistical and diagnostic-accuracy reporting complies with STARD 2015 recommendations.

1.4 Image Processing and Computational Analysis

1.4.1 Construction of Cross-Modal Image-Text Collaborative Classification Model

This study developed a vision-text dual-branch cross-modal deep learning model that jointly learns features from CCTA coronary images and clinical stenosis descriptive text to perform binary classification of severe versus non-severe coronary stenosis. The overall training pipeline is illustrated in Fig. 2. The model architecture integrates three core modules: class-differentiated data augmentation, dual-branch feature extraction, and cross-modal contrastive collaborative regularization, designed to address key challenges including limited coronary dataset size, scarcity of severe stenosis samples, heterogeneous pathological morphology driven by coronary calcification, and disconnection between discriminative visual features and clinical diagnostic logic inherent to image-only classification models. All deep-learning experiments were implemented with PyTorch 1.13, running on NVIDIA GPU hardware.

1.4.2 Feature extractor

The image branch employs an ImageNet pre-trained ResNet-18 backbone, with only high-layer network parameters fine-tuned in a layer-wise manner to adapt the model to CCTA data distribution while retaining general fundamental visual representations. Augmented local coronary vessel images are fed into residual blocks to extract 512-dimensional high-level visual features, which are simultaneously delivered to a classification head and a projection head. The classification head outputs binary stenosis predictions, whereas the projection head maps visual features into a shared 256-dimensional joint embedding space followed by L2 normalization. The text branch utilizes a lightweight Transformer encoder, accepting standardized radiologist stenosis diagnostic descriptions as textual input. Word embedding, positional encoding and multi-head attention modules extract clinical semantic features, which are projected via a dedicated adapter into the identical 256-dimensional joint embedding space as visual features, enabling spatial alignment between imaging phenotypes and clinical diagnostic semantics. Image feature extraction is illustrated in Fig. 3.

1.4.3 Dataset Preprocessing and Class-Aware Data Augmentation

The annotated CCTA dataset was limited in overall volume, with severe coronary stenosis constituting a minority class. Direct model training without mitigation would induce classifier bias toward the majority non-severe stenosis cohort and elevate false-negative rates for severe lesions. A dual strategy combining class-differentiated image augmentation and weighted stochastic sampling was proposed to alleviate class imbalance.

Class-differentiated augmentation pipeline: Distinct transformation intensities were applied to the two lesion subgroups. Non-severe stenosis samples only underwent low-magnitude geometric transformations including minor random horizontal flipping, slight rotation within ± 10°, and narrow-range scaling perturbation to preserve native morphological characteristics of normal vessel walls and minimal plaque burden. For severe stenosis samples, fundamental geometric augmentations were supplemented with extensive rotation, multi-scale resizing, histogram equalization and adaptive contrast adjustment, simulating diverse pathological imaging phenotypes such as heavy calcified plaques, irregular vessel wall thickening, and distal perfusion defects to expand morphological diversity of the minority severe stenosis class.

Weighted stochastic sampling: Sampling weights for each training batch were inversely proportional to the total sample count of the corresponding class, ensuring balanced representation of severe and non-severe stenosis cases per training iteration to prevent majority-class dominance of gradient updates and stabilize classification decision boundaries.

Model Comparison and Performance Metrics

Conventional manual CCTA reading served as the comparative baseline for multi-dimensional performance assessment of the proposed deep learning model. Using invasive CAG results as the reference gold standard, diagnostic metrics including sensitivity, specificity, accuracy, positive predictive value (PPV) and negative predictive value (NPV) were computed for both modalities. ROC curves were plotted, with AUC values calculated to quantify overall discriminative power.

Results

2.1 Baseline Clinical Characteristics

Baseline demographic and clinical data of the 96 enrolled patients are summarized in Table 1. The cohort consisted of 44 males and 52 females, aged 40–80 years with a mean age of (66.32 ± 8.15) years. Twenty-four patients reported a history of tobacco smoking, 64 had hypertension, 27 suffered from type 2 diabetes mellitus, and elevated low-density lipoprotein cholesterol (LDL-C) was documented in 67 individuals.

Table 1
Baseline Clinical Characteristics of 96 Enrolled Patients

Clinical Characteristic	Value
Sex (Male/Female)	44/52
Age (years)	66.32 ± 8.15
Body mass index (BMI)	25.12 ± 3.26
Smoking history, n (%)	24(25.9%)
Hypertension, n (%)	64(66.7%)
Diabetes mellitus, n (%)	27(28.0%)
Elevated LDL-C, n (%)	67(69.7%)

2.2 Deep Learning vs. Conventional Interpretation for CCTA Diagnosis

A total of 286 coronary vessels from the 96 patients were analyzed, with each vessel classified based on its maximally stenotic segment. Conventional manual CCTA reading identified severe stenosis in 81 vessels, yielding a sensitivity of 46.38%, specificity of 88.7%, accuracy of 68.5%, NPV of 64.4% and PPV of 79.0%. The overall discriminative capacity of manual CCTA for detecting severe coronary stenosis was suboptimal, with an AUC of 0.676 (95%CI: 0.613–0.739) and moderate inter-modality consistency (Kappa = 0.357). Statistically significant diagnostic discordance was observed between manual CCTA interpretation and invasive CAG ($P < 0.001$). The proposed deep learning model achieved a sensitivity of 64%, specificity of 93%, accuracy of 79%, NPV of 74%, PPV of 89%, and an AUC of 0.83 (95%CI: 0.71–

0.94). Calibration analysis demonstrated acceptable agreement between model-predicted probabilities of severe coronary stenosis and reference outcomes verified by invasive coronary angiography. Detailed comparative metrics are listed in Table 2, and the ROC curve of the deep learning classifier is presented in Fig. 4.

Table 2
Diagnostic Metrics: Manual CCTA Reading vs. Deep Learning Model

Metric	**Manual CCTA**	**Our Model**
Sensitivity	46.38%	**64%**
Specificity	88.7%	**93%**
Accuracy	68.5%	**79%**
Negative Predictive Value (NPV)	64.4%	**74%**
Positive Predictive Value (PPV)	79.0%	**89%**
AUC (95% CI)	0.67 (95% CI: 0.61–0.74)	**0.83 (95%CI: 0.71–0.94)**

Discussion

CCTA serves as the first-line non-invasive imaging modality for patients with chest pain. It enables comprehensive visualization of coronary anatomy, luminal stenosis grading, and quantitative characterization of atherosclerotic plaque morphology as well as plaque burden(3, 11). Nevertheless, conventional manual interpretation has notable clinical drawbacks, including limited inter-reader measurement reproducibility, time-consuming reading workflow, and suboptimal performance in identifying in-stent luminal borders and tiny lipid-rich high-risk plaques. Artificial intelligence (AI) centered on deep learning empowers CCTA diagnostic workflows from multiple dimensions, covering image reconstruction pre-processing, fully automatic coronary segmentation, vulnerable plaque detection, and non-invasive quantitative evaluation of hemodynamic ischemia. It generates objective and reproducible quantitative imaging biomarkers to support population-level screening and individualized risk stratification for coronary artery disease (CAD)(5, 12). In the present study, we developed a cross-modal image-text collaborative deep learning model integrating a ResNet-18 image encoder and a lightweight Transformer text encoder. The proposed model achieved an AUC of 0.83 for predicting severe coronary stenosis, outperforming conventional manual reading (AUC = 0.676). This finding is consistent with accumulating evidence that deep learning improves objectivity and accuracy in CCTA interpretation. Furthermore, our cross-modal architecture mitigates inherent limitations of vision-only models when processing complex pathological CCTA images.

CCTA images reconstructed by conventional iterative algorithms frequently suffer from elevated background noise, prominent blooming artifacts induced by metallic stents, and blurred contours of small coronary branches and in-stent lumens, which impair diagnostic accuracy for luminal stenosis and in-stent restenosis(13, 14). Super-resolution deep learning reconstruction (SR-DLR) simultaneously

suppresses background noise and enhances vessel-wall boundary contrast. Compared with conventional model-based iterative reconstruction, SR-DLR substantially reduces stent-related artifacts and increases the proportion of evaluable in-stent vessel segments(13). Even under low-radiation-dose scanning protocols, SR-DLR preserves diagnostically acceptable image quality(15). Multiple prospective studies have demonstrated that CAD-RADS classifications derived from SR-DLR-reconstructed images exhibit higher diagnostic agreement with invasive coronary angiography. SR-DLR improves inter-observer consistency for stenosis grading and reduces the risk of missed or misdiagnosed coronary lesions(14, 15). These advances in image reconstruction provide high-quality input data and lay a solid foundation for guaranteeing the diagnostic performance of downstream deep-learning models.

End-to-end transfer-learning deep-learning frameworks built on curved multi-planar reformatted images eliminate manual region-of-interest delineation and pre-segmentation steps, simplifying intelligent CCTA analysis. However, such pixel-level models have well-recognized shortcomings: they demonstrate high sensitivity for calcified plaques yet limited performance in detecting low-density lipid-rich soft plaques(6, 10, 16). A systematic review and meta-analysis conducted by Tu et al. confirmed that deep-learning approaches achieve overall favorable diagnostic accuracy for coronary stenosis, whereas model performance is heavily affected by input image quality and lesion phenotypes(16). These observations highlight the promise of image-only deep learning while revealing its bottlenecks, implying that incorporating multi-modal information such as clinical textual descriptions to assist feature extraction may boost detection rates of occult and atypical lesions. This constitutes the core design rationale for our cross-modal image-text collaborative model.

Luminal stenosis severity alone cannot fully reflect the hemodynamic significance of coronary lesions. Lipid-enriched high-risk plaques may trigger microvascular dysfunction via maladaptive vascular remodeling and endothelial inflammatory injury(6). The extreme gradient boosting (XGBoost) algorithm captures complex non-linear interactions among plaque morphology, luminal features, and vascular quantitative metrics. Without offline computational fluid dynamics (CFD) calculations, XGBoost can simultaneously predict myocardial ischemia validated by invasive fractional flow reserve (FFR) and hyperemic myocardial perfusion impairment detected by positron emission tomography (PET)(6, 10). It improves detection efficiency of hemodynamically significant stenosis and reduces unnecessary referrals for invasive FFR measurements. The prospective REVEALPLAQUE study further reported excellent agreement between AI-assisted CCTA plaque quantification and intravascular ultrasound (IVUS), supporting deep-learning-enhanced CCTA as a non-invasive alternative to invasive imaging modalities(6).

CAD-RADS represents the standardized reporting framework for contemporary CCTA examinations. Cascaded gradient-boosting classifiers can be constructed by extracting first-order statistics and wavelet-based texture radiomic features from coronary multi-planar reformatted images with multiple feature-integration strategies to realize automatic six-category CAD-RADS classification(17). Majority-voting schemes MV_P and MV_C yield superior performance compared with statistical aggregation based on mean or extreme values. The MV_C strategy assigns patient-level grade according

to the most severe lesion within individual vessels, which aligns well with real-world clinical decision-making and achieves 100% classification accuracy in heterogeneous subgroups with inconsistent stenosis severity across coronary arteries. SHapley Additive exPlanations (SHAP) interpretability analysis identifies texture metrics such as gray-level size-zone matrix features as key predictors corresponding to tissue heterogeneity caused by plaque lipid and calcified components. Compared with black-box deep-learning models, radiomic features possess clear imaging-pathological correlations and superior clinical interpretability. Combined application of super-resolution reconstruction and coronary radiomics narrows predictive gaps between high-risk and low-risk plaques and further improves CAD-RADS grading consistency(17). Post-hoc analyses from the ISCHEMIA trial demonstrated that CCTA-derived atherosclerotic plaque burden provides incremental prognostic value for the composite endpoint of cardiovascular death or myocardial infarction after adjustment for conventional risk factors, ejection fraction, prior CAD, and number of diseased vessels, underscoring the clinical importance of standardized plaque quantification(11).

Artificial neural networks trained solely on demographic, biochemical, and epidemiological baseline variables achieve a maximum validation accuracy of 96.25% for long-term CAD risk prediction after SMOTE resampling for class imbalance and Adam optimizer tuning. Nevertheless, these models do not incorporate CCTA anatomical information or plaque quantitative metrics; they are only suitable for general-population screening and cannot accomplish vessel-specific ischemic risk assessment, serving merely as complementary tools for imaging-based AI models(12). Standardized AI quantitative pipelines enable automatic co-registration of serial follow-up CCTA datasets and segment-wise quantification of dynamic changes in plaque volume, stenosis severity, and high-risk plaque burden to accurately identify plaque progression, stability, or regression. When combined with full AI reconstruction and plaque-quantification workflows, photon-counting ultra-low-dose CT fulfills quantitative follow-up requirements while lowering contrast medium consumption and radiation exposure. It provides continuous objective quantitative endpoints for evaluating lipid-lowering and anti-atherosclerotic therapies and overcomes limitations of conventional qualitative follow-up assessments(12). Cho et al. validated the feasibility of AI-assisted serial CCTA analysis for tracking CAD progression, enabling longitudinal clinical monitoring of plaque dynamics(12).

This study has several notable limitations. First, this is a single-centre retrospective investigation with limited sample size, and external-cohort validation is absent. Analyses were performed at vessel-level rather than patient-level, which may overestimate clinical performance; selection bias cannot be fully excluded given retrospective enrolment. Larger multi-centre datasets are required to further verify generalizability of the proposed cross-modal image-text collaborative model. More broadly, multiple obstacles hamper clinical translation of coronary-AI tools. Most training datasets originate from single centres using one CT scanner vendor. Prospective validations of AI algorithms optimised for photon-counting CT remain insufficient. Variations in scanning parameters and iterative-reconstruction protocols alter image texture and plaque-quantification outputs, leading to shifts in model performance. Pixel-wise deep-learning models tend to miss soft plaques; XGBoost-based pipelines depend on semi-automatic plaque-segmentation pre-processing and cannot generate instant reports immediately

after scanning. Myocardial deep-learning models offer prominent benefits for patients with intermediate-grade stenosis yet show limited value for evaluating total coronary occlusion. The vast majority of existing studies use FFR or CAD-RADS as surrogate endpoints. Large-sample multi-centre long-term follow-up evidence built upon hard cardiovascular endpoints (myocardial infarction, revascularisation) remains scarce. Uniform industry standards for CCTA post-processing, plaque quantification, and radiomic-feature extraction have not been established across medical institutions. Measurement discrepancies across different algorithms hinder multi-centre data integration and large-scale clinical deployment.

Future translational research on coronary-AI can be prioritised across five directions: construction of integrated multi-modal-AI pipelines; optimisation of lightweight networks for low-dose and novel-generation CT scanners; prospective-cohort validations grounded on long-term hard cardiovascular endpoints; deployment of standardised interpretable-AI modules; and hierarchical application for diverse clinical scenarios. The integrated pipeline combines super-resolution reconstruction, unsupervised vascular segmentation, end-to-end stenosis recognition, XGBoost-derived ischaemia scoring, myocardial-function deep-learning, and coronary-radiomics modules. It achieves one-stop post-processing after CCTA acquisition covering image optimisation, anatomical identification, functional prediction, standardised CAD-RADS grading, and longitudinal quantitative follow-up. Reconstruction and plaque-recognition algorithms should be optimised for photon-counting CT to expand eligibility of non-invasive coronary screening for elderly patients and subjects with renal insufficiency. Multi-centre long-term cohorts should be established to validate predictive performance of AI tools for major adverse cardiovascular events. Industry-wide standards for plaque quantification and radiomic-feature extraction should be unified; SHAP analysis and heat-map visualisation should be embedded within commercial CCTA workstations. A hierarchical diagnostic framework should be built: artificial neural networks for general-population CAD screening, multi-modal CCTA-AI for precise risk stratification for chest-pain patients, serial quantitative-AI for long-term plaque surveillance. Ultimately, a full-process standardised AI-assisted diagnosis-treatment system covering CAD screening, individualised risk stratification, and long-term disease management can be realised to improve long-term cardiovascular outcomes for CAD patients.

Conclusion

In this study, we constructed a cross-modal image-text collaborative deep-learning model built upon ResNet-18 image encoder and lightweight Transformer text encoder. Combined with class-differentiated data augmentation, weighted stochastic sampling, and multi-loss joint-optimisation strategies, our cross-modal deep-learning model achieved significantly improved predictive performance for severe coronary stenosis compared with conventional manual CCTA interpretation. Through cross-modal contrastive learning, the model fuses coronary imaging features and semantic information extracted from clinical diagnostic texts and enables automatic detection of subtle pathological signatures including atherosclerotic plaques and luminal stenosis. This framework effectively alleviates dataset class imbalance and substantially reduces missed-diagnosis rates for occult severe coronary stenosis.

In summary, our CCTA-based cross-modal image-text collaborative deep-learning model supports efficient, accurate non-invasive evaluation of coronary-stenosis severity and overcomes major drawbacks of manual reading including high subjectivity and insufficient sensitivity. It may serve as a reliable intelligent auxiliary tool for non-invasive coronary-artery-disease screening, myocardial-ischaemia risk stratification, and individualised clinical decision-making.

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the Institutional Review Board of The Fifth People's Hospital of Huai'an. Due to retrospective design with de-identified clinical imaging data, written informed consent requirement was waived by the institutional review board (Ethics reference number: KY-P-2026-51). All procedures complied with the principles of the Declaration of Helsinki.

Consent for publication

Not applicable. No individual identifiable patient data are presented in this manuscript.

The raw CCTA clinical imaging dataset cannot be publicly released due to hospital patient-privacy regulations. De-identified intermediate feature data and model code are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Science-Education Integration Clinical Research Project of Jiangsu College of Nursing (Grant No. SH202510250312). The funder played no role in study design, data collection, analysis, manuscript writing or decision to submit.

Author Contribution

[+]These authors contributed equally to this work and share first authorship.

Li Ma[+]: Conceptualization, Methodology, Formal analysis, Writing-original draft;

Jiangpeng Wang[+]: Conceptualization, Resources, Data curation, Writing-original draft;

Dongdong Tang: Resources, Data curation;

Huan Tang: Data curation;

Haibao Yuan: Data curation;

Linghui Zhao*: supervised this study. All authors contributed to manuscript drafting and revision.

All authors contributed to manuscript drafting and revision. All authors read and approved the final manuscript.

Data Availability

The raw CCTA image dataset contains sensitive patient clinical information and cannot be publicly released to protect patient privacy under hospital regulations. De-identified intermediate feature data and the model code that support the findings of this study are available from the corresponding author upon reasonable written request.

References

1. Shaya GE, Leucker TM, Jones SR, Martin SS, Toth PP. Coronary heart disease risk: Low-density lipoprotein and beyond. Trends Cardiovasc Med. 2021;32(4):181–94.
2. Global burden. of 292 causes of death in 204 countries and territories and 660 subnational locations, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. Lancet. 2025;406(10513):1811–72.
3. Honigberg MC, Lander BS, Baliyan V, Jones-O'Connor M, Healy EW, Scholtz J-E, et al. Preventive Management of Nonobstructive CAD After Coronary CT Angiography in the Emergency Department. JACC Cardiovasc Imaging. 2019;13(2 Pt 1):437–48.
4. Cury RC, Leipsic J, Abbara S, Achenbach S, Berman D, Bittencourt M, et al. Computed Tomography (SCCT), the American College of Cardiology (ACC), the American College of Radiology (ACR), and the North America Society of Cardiovascular Imaging (NASCI). JACC Cardiovasc Imaging. 2022;15(11):1974–2001. CAD-RADS™ 2.0–2022 Coronary Artery Disease-Reporting and Data System: An Expert Consensus Document of the Society of Cardiovascular.
5. Shrivastava P, Kashikar S, Parihar PH, Kasat P, Bhangale P, Shrivastava P. A systematic review on deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction. Eur J Radiol Open. 2025;14:100652.
6. Narula J, Stuckey TD, Nakazawa G, Ahmadi A, Matsumura M, Petersen K, et al. Prospective deep learning-based quantitative assessment of coronary plaque by computed tomography angiography compared with intravascular ultrasound: the REVEALPLAQUE study. Eur Heart J Cardiovasc Imaging. 2024;25(9):1287–95.

7. Wang L, Zhang X, Tian C, Chen S, Deng Y, Liao X, et al. PlaqueNet: deep learning enabled coronary artery plaque segmentation from coronary computed tomography angiography. Vis Comput Ind Biomed Art. 2024;7(1):6.
8. Adolf R, Nano N, Chami A, von Schacky CE, Will A, Hendrich E, et al. Convolutional neural networks on risk stratification of patients with suspected coronary artery disease undergoing coronary computed tomography angiography. Int J Cardiovasc Imaging. 2023;39(6):1209–16.
9. Hamana T, Nishimori M, Shibata S, Kawamori H, Toba T, Hiromasa T, et al. Deep-learning-driven optical coherence tomography analysis for cardiovascular outcome prediction in patients with acute coronary syndrome. Eur Heart J Digit Health. 2024;5(6):692–701.
10. Ihdayhid AR, Sehly A, He A, Joyner J, Flack J, Konstantopoulos J, et al. Coronary Artery Stenosis and High-Risk Plaque Assessed With an Unsupervised Fully Automated Deep Learning Technique. JACC Adv. 2024;3(9):100861.
11. Nurmohamed NS, Min JK, Anthopolos R, Reynolds HR, Earls JP, Crabtree T, et al. Atherosclerosis quantification and cardiovascular risk: the ISCHEMIA trial. Eur Heart J. 2024;45(36):3735–47.
12. Cho GW, Anderson L, Quesada CG, Jennings RS, Min JK, Earls JP, et al. Serial analysis of coronary artery disease progression by artificial intelligence assisted coronary computed tomography angiography: early clinical experience. BMC Cardiovasc Disord. 2022;22(1):506.
13. Orii M, Sone M, Osaki T, Ueyama Y, Chiba T, Sasaki T, et al. Super-resolution deep learning reconstruction at coronary computed tomography angiography to evaluate the coronary arteries and in-stent lumen: an initial experience. BMC Med Imaging. 2023;23(1):171.
14. Zou L, Xu C, Liu X, Wang L, Zhou T, Yang L, et al. Super-Resolution Deep Learning Reconstruction for Coronary CT Angiography: Coronary Stenosis Assessment and CAD-RADS Reclassification. Radiology. 2026;318(2):e252163.
15. Zou L-M, Xu C, Xu M, Xu K-T, Zhao Z-C, Wang M, et al. Ultra-low-dose coronary CT angiography via super-resolution deep learning reconstruction: impact on image quality, coronary plaque, and stenosis analysis. Eur Radiol. 2025;35(8):4674–84.
16. Tu L, Deng Y, Chen Y, Luo Y. Accuracy of deep learning in the differential diagnosis of coronary artery stenosis: a systematic review and meta-analysis. BMC Med Imaging. 2024;24(1):243.
17. Corti A, Lo Iacono F, Ronchetti F, Chiesa M, Pontone G, Colombo GI, et al. Patient-level CAD-RADS scoring from coronary radiomic features. Sci Rep. 2026;16(1):4330.

Figures

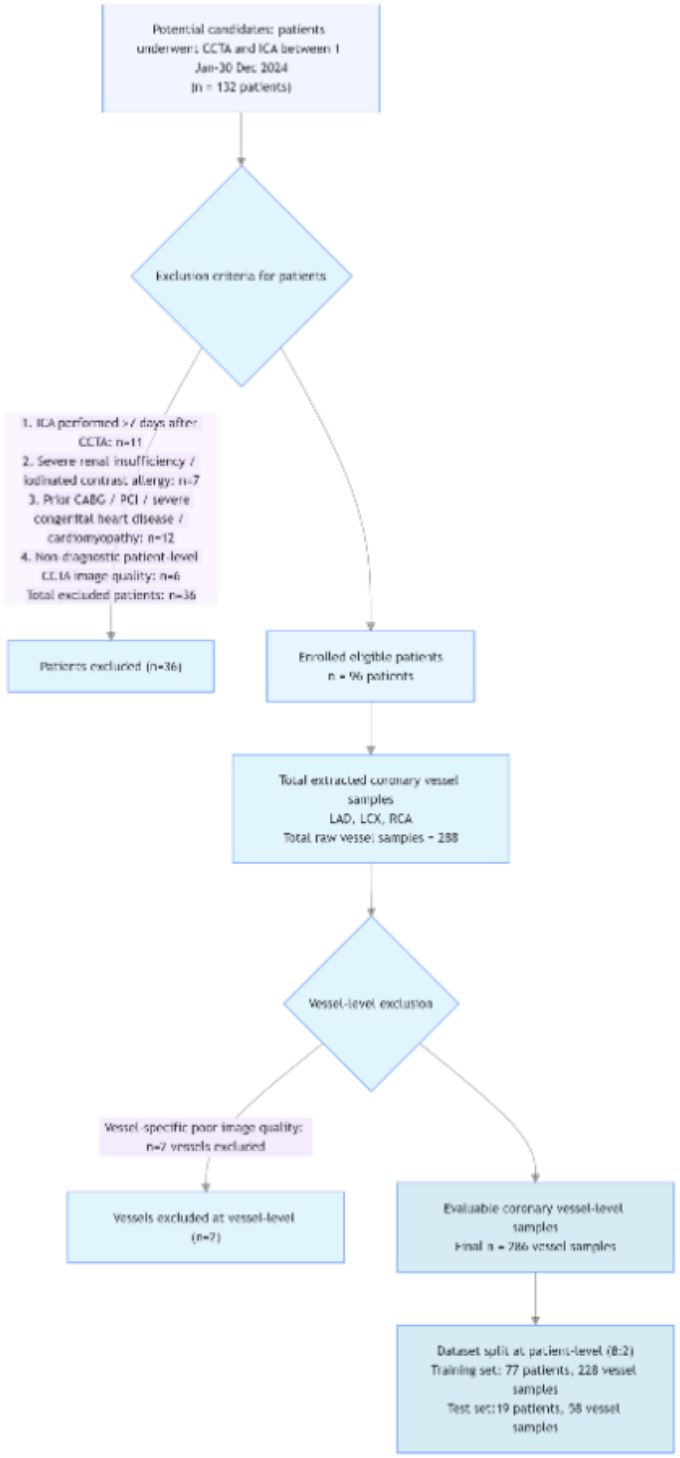


Figure 1

STARD-compliant patient-and-vessel-level participant selection flowchart

Notes: Analysis unit for diagnostic performance was coronary vessel-level, not patient-level. Two out of 288 raw vessel samples were excluded due to vessel-specific poor image quality. Splitting was performed on patient-level rather than vessel-level to prevent data leakage: vessels originating from one

patient were entirely assigned either to training or test set and were never distributed across both subsets. CABG=coronary artery bypass grafting; PCI=percutaneous coronary intervention; CCTA=coronary computed tomography angiography; ICA=invasive coronary angiography.

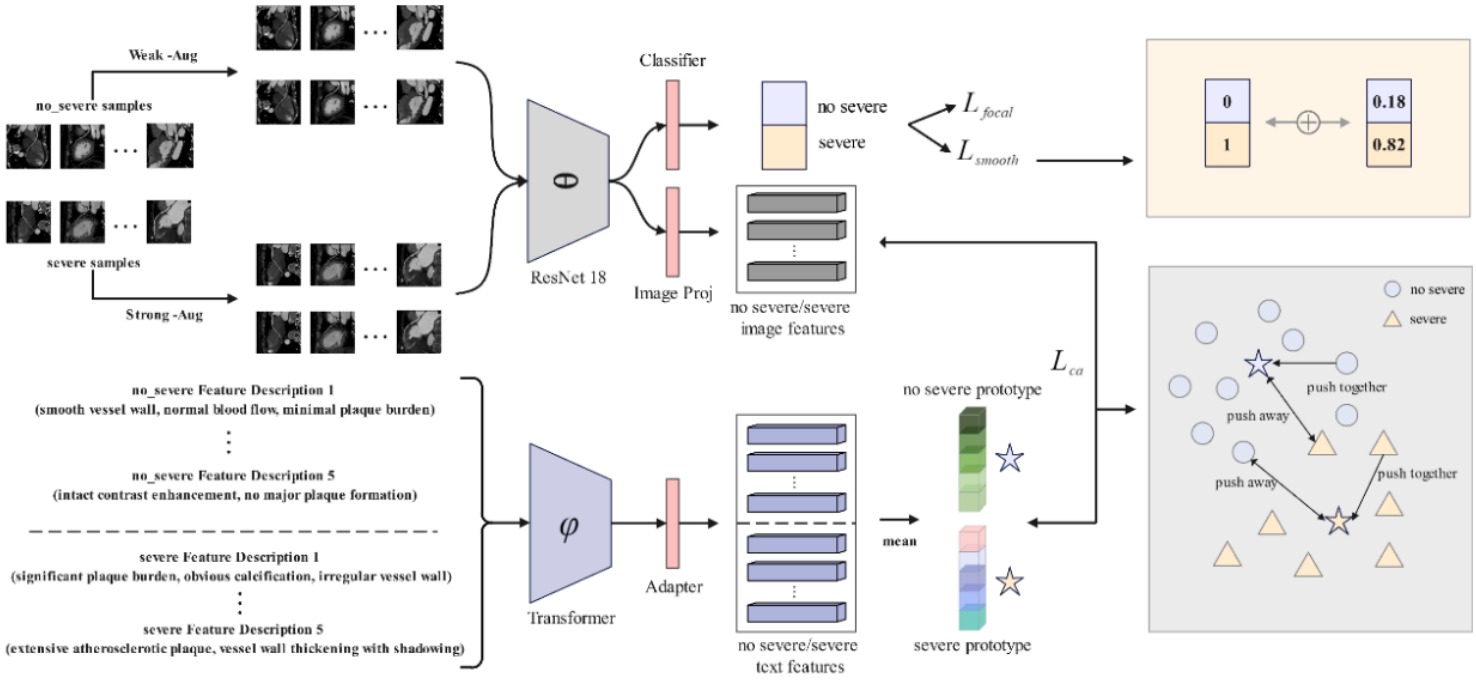


Figure 2

Overall architecture diagram of model training

Notes: Weak-Aug = low-intensity data augmentation for non-severe stenosis samples; Strong-Aug = high-intensity data augmentation for severe stenosis samples. The image encoder adopts ResNet-18, while a lightweight Transformer functions as the text encoder. Multi-loss joint optimization was implemented during model training: Focal loss: Down-weights easily classified normal vessel samples, prioritizing hard-to-distinguish cases with calcified and eccentric plaques to mitigate class imbalance; Label smoothing loss: Suppresses model overfitting and enhances generalization capacity; Cross-modal contrastive loss: Leverages class semantic prototypes as anchor points to narrow feature distances between matched image-text pairs of identical categories while widening inter-category feature gaps. Clinical diagnostic text guides the model to capture vascular features consistent with cardiovascular diagnostic reasoning, eliminating misclassification induced by irrelevant imaging artifacts such as calcification and vessel wall grayscale heterogeneity.

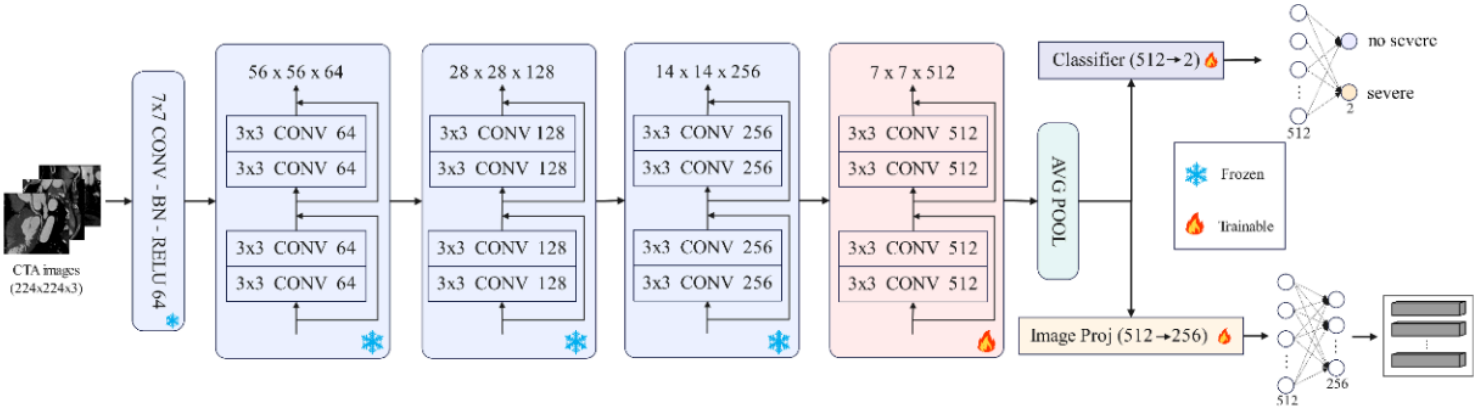


Figure 3

Image feature extractor

Note: The backbone adopts a pre-trained ResNet-18 initialized on ImageNet. Frozen layers (blue) preserve original parameters during training, while the top layer (red) is selectively fine-tuned. Input augmented CCTA images pass through sequential residual blocks, yielding a 512-dimensional feature vector. This representation branches into two downstream heads: a classifier for binary stenosis outcome and a projector for cross-modal alignment.

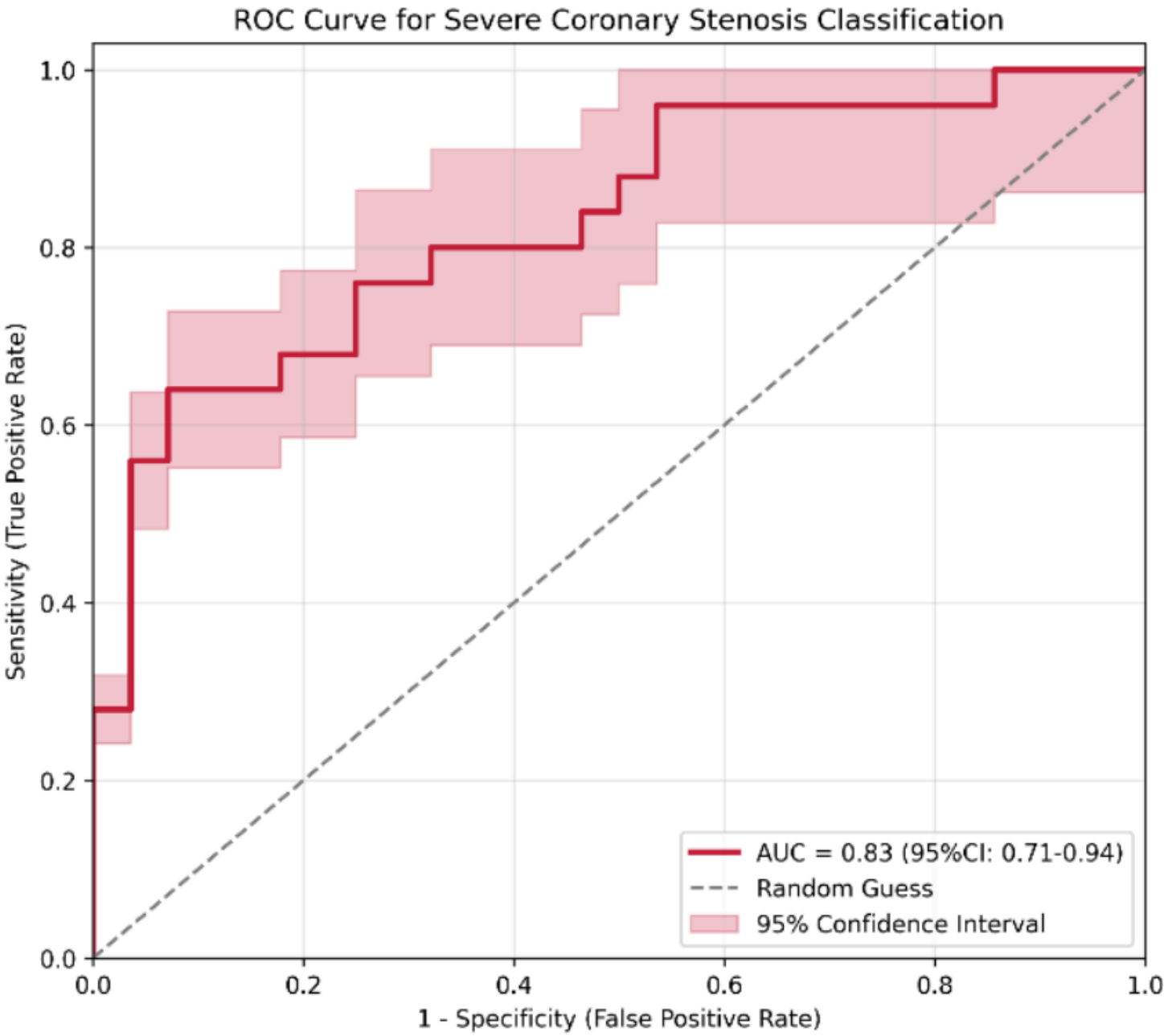


Figure 4

ROC curve for severe coronary stenosis classification.

Note: AUC = 0.83 (95% CI: 0.71–0.94). The red solid line represents model performance, the dashed diagonal line indicates random guess, and the shaded area shows the 95% confidence interval.