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.

Development and Validation of a Machine Learning Model Integrating SPECT MPI Multiparametric Features for Risk Prediction in Coronary Artery Disease with Preserved Ejection Fraction

Yuanyuan Li
First Hospital of Shanxi Medical University
Jiaxin Cao
First Hospital of Shanxi Medical University
Yuting Zhao
First Hospital of Shanxi Medical University
Ping Wu
First Hospital of Shanxi Medical University
Ruonan Wang
First Hospital of Shanxi Medical University
Zhifang Wu
First Hospital of Shanxi Medical University
Haiyan Liu
First Hospital of Shanxi Medical University
Sijin Li
lisjnm123@163.com
First Hospital of Shanxi Medical University

Research Article

Keywords: Machine Learning, Interpretable, Predictive Model, Myocardial Perfusion Imaging, Peak Filling Rate, Total Perfusion Deficit, Entropy.

Posted Date: August 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10383224/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

Objective

Early identification of high-risk patients with coronary artery disease (CAD) and preserved left ventricular ejection fraction (LVEF) is essential for optimizing management. This study aimed to develop a machine learning model integrating clinical and myocardial perfusion imaging (MPI)–derived parameters to predict major adverse cardiac events (MACE).

Methods

We retrospectively included 1,675 patients with suspected or confirmed CAD and preserved LVEF who underwent resting MPI. Demographic, clinical, and imaging variables were collected, and missing data were handled via multiple imputation. Ten machine learning algorithms were developed and compared. Model performance was assessed by the area under the receiver operating characteristic curve (AUC) and additional metrics. Shapley additive explanations analysis (SHAP) was applied to interpret feature contributions.

Results

The CatBoost model showed the best performance, with cross-validated AUC of 0.693, accuracy 0.632, precision 0.551, F1 score 0.451, and Brier score 0.234. SHAP analysis identified nine key predictors: prior cardio-cerebrovascular disease, age, peak filling rate, low-density lipoprotein cholesterol, total perfusion deficit, eccentricity index (no-gated), entropy, coronary stenosis ≥ 50%, and New York Heart Association functional class.

Conclusions

We developed an interpretable CatBoost model integrating clinical and MPI features to predict MACE risk in CAD patients with preserved LVEF. The model demonstrated a balanced performance in terms of stability, generalizability, and clinical applicability, and may serve as a valuable tool for risk stratification and personalized management in this patient population.

Trial registration

Not applicable. This study is a retrospective analysis.

1. Introduction

Cardiovascular disease (CVD) remains the leading cause of mortality worldwide, accounting for approximately one third of all global deaths. Among all forms of CVD, CAD represents the most lethal subtype and constitutes the greatest contributor to the overall disease burden[1]. With progressive population aging, the rising prevalence of metabolic disorders, and continuous advances in therapeutic strategies, the clinical phenotypes and prognostic trajectories of patients with CAD have become increasingly heterogeneous. This heterogeneity is particularly pronounced across different categories of left ventricular ejection fraction. Accumulating evidence indicates that patients with CAD remain at substantial risk of adverse cardiovascular outcomes even in the presence of preserved left ventricular ejection fraction (LVEF). This subgroup is frequently associated with heart failure with preserved ejection fraction (HFpEF), left ventricular diastolic dysfunction, and coronary microvascular dysfunction[2, 3]. Consequently, accurate risk stratification and prognostic assessment in patients with CAD and preserved LVEF have emerged as critical challenges in contemporary cardiovascular research and clinical practice.

A wide range of diagnostic and risk assessment modalities are currently available for CAD, including transthoracic echocardiography, myocardial perfusion imaging (MPI), coronary computed tomography angiography, cardiac magnetic resonance imaging, and invasive coronary angiography, each offering complementary information[4]. Among these techniques, single-photon emission computed tomography (SPECT) MPI is one of the earliest developed and most extensively validated noninvasive imaging modalities in cardiology[5]. Owing to its favorable safety profile, high reproducibility, and broad applicability, SPECT MPI plays an indispensable role in the diagnosis of CAD, evaluation of myocardial ischemia, and prognostic risk assessment. Beyond perfusion assessment, SPECT MPI enables the simultaneous quantification of left ventricular morphology, systolic and diastolic function, and mechanical synchrony, parameters whose clinical relevance in cardiovascular risk stratification has been well established[6, 7].

With the expanding clinical use of SPECT MPI, a variety of derived quantitative parameters—including perfusion indices such as total perfusion deficit (TPD), functional parameters such as peak filling rate (PFR) and LVEF, structural and geometric indices such as eccentricity index (EI) and shape index (SI), as well as measures of left ventricular systolic mechanical synchrony such as entropy—have demonstrated considerable potential for predicting adverse cardiovascular events[8–11]. However, conventional risk assessment strategies typically rely on a limited number of individual variables and are therefore unable to fully integrate the multidimensional information provided by advanced imaging. This limitation restricts their predictive performance in complex and heterogeneous disease populations. In recent years, advances in big data analytics, artificial intelligence, and quantitative imaging have led to growing interest in machine learning–based approaches for disease prediction and risk stratification. Machine learning models offer distinct advantages in handling high-dimensional data, capturing nonlinear relationships, and modeling complex interactions among multiple parameters, thereby providing a powerful framework for integrative prognostic assessment.

Despite increasing evidence supporting the utility of machine learning in cardiovascular prognostication, studies focusing specifically on patients with CAD and preserved LVEF—particularly those integrating comprehensive SPECT MPI–derived parameters—remain limited[12, 13]. Therefore, in the present retrospective study, we analyzed clinical and SPECT MPI data from CAD patients treated at our institution over the past five years. We systematically developed and validated multiple machine learning models to evaluate their performance in risk stratification and prognostic assessment among patients with CAD and preserved LVEF, with the goal of providing novel insights to support individualized clinical decision-making.

2. Materials and Methods

2.1 Study Population

Clinical data were retrospectively collected from patients with suspected or established CAD who underwent resting SPECT MPI between January 2016 and March 2021. CAD diagnoses were defined according to the International Classification of Diseases and included angina pectoris, acute myocardial infarction, subsequent myocardial infarction, early complications following acute myocardial infarction, and other acute or chronic ischemic heart diseases. The study protocol was approved by the Institutional Ethics Committee of the First Hospital of Shanxi Medical University (approval number: 2022-K-128) and was conducted in accordance with the principles of the Declaration of Helsinki.

The inclusion criteria were as follows:(1) patients with clinically suspected or confirmed CAD who underwent resting SPECT MPI at the Department of Nuclear Medicine of the First Hospital of Shanxi Medical University;(2) availability of transthoracic echocardiography performed within three months of SPECT MPI, demonstrating a LVEF ≥ 50%; and (3) absence of major comorbid conditions that could substantially affect LVEF or life expectancy, including severe congenital heart disease, cardiomyopathy, significant valvular heart disease, myocarditis, malignancy, hematologic disorders, respiratory failure, severe infection, or thyroid dysfunction.

The exclusion criteria were as follows:(1) occurrence of significant arrhythmias during gated resting SPECT MPI acquisition or other technical issues that could compromise image quality;(2) incomplete clinical data and/or missing SPECT MPI–derived parameters; and (3) lack of follow-up information. Eligible patients were randomly assigned in a 7:3 ratio to a training cohort and an internal validation cohort **(**Fig. 1**).**

2.2 Clinical Characteristics

Variables assessed in this study were extracted from the inpatient electronic medical records, including demographic characteristics, clinical information, and SPECT MPI–derived parameters. Collected variables included sex, age, body mass index (BMI; calculated as weight divided by height squared), New York Heart Association (NYHA) functional class, medical history (including prior myocardial infarction, ischemic stroke, peripheral vascular disease, diabetes mellitus, hypertension, and previous revascularization), family history of cardiovascular disease, smoking status, and medication use. Medications included β-blockers, angiotensin-converting enzyme inhibitors or angiotensin receptor blockers (ACEI/ARBs), calcium channel blockers, diuretics, lipid-lowering agents, antiplatelet and/or anticoagulant agents, and glucose-lowering medications.

2.3 SPECT MPI Image Acquisition and Processing

2.3.1 Imaging System and Radiotracer

All SPECT MPI examinations were performed using a dual-head IQ-SPECT/CT system (Symbia T16, Siemens Healthineers, Germany) equipped with SMART-ZOOM collimators. Image acquisition was conducted with a photopeak of 140 keV, a 20% energy window, a 128 × 128 matrix, and a zoom factor of 1.0. The radiotracer used was ^{99m}Tc-methoxyisobutyl isonitrile (^{99m}Tc-MIBI). Radiolabeling was performed in the Department of Nuclear Medicine of the First Hospital of Shanxi Medical University, with a radiochemical purity > 95%. ^{99m}Tc was obtained from a ^{99}Mo/^{99m}Tc generator supplied by Beijing Atom High-Tech Co., Ltd., and MIBI was provided by the Jiangsu Institute of Atomic Medicine.

2.3.2 Imaging Protocol

After exclusion of contraindications, patients fasted for at least 3 hours before intravenous injection of 740–925 MBq of ^{99m}Tc-MIBI under resting conditions. A fatty meal was administered 30 minutes after injection, and gated SPECT acquisition was initiated 60 minutes post-injection. Imaging was performed in the supine position with both arms raised and fixed above the head.

2.3.3 Acquisition Parameters and Reconstruction

The two detectors were positioned close to the chest wall with a 76° angular configuration. Data were acquired using a step-and-shoot protocol from 38° right anterior oblique to 66° left posterior oblique, with 25 seconds per frame, 6° per step, resulting in 34 projections over 208° in approximately 8 minutes. Images were reconstructed using an ordered subsets expectation maximization algorithm to generate short-axis, horizontal long-axis, and vertical long-axis images of the left ventricle.

2.3.4 Image Analysis

Quantitative parameters were automatically derived using Quantitative Perfusion SPECT (QPS) and Quantitative Gated SPECT (QGS) software. The extracted parameters included: Perfusion parameters: TPD. Functional parameters: Systolic parameters: LVEF, peak ejection rate (PER); Diastolic parameters: PFR, mean filling rate over the first third of diastole (MFR/3), and time to peak filling (TTPF); Regional wall motion and wall thickening scores. Morphological parameters: SI (no gated), end-diastolic shape index (EDSI), end-systolic shape index (ESSI), EI and EI (no gated), end-diastolic volume (EDV), and end-systolic volume (ESV). Machanical synchrony parameters: entropy, phase histogram bandwidth, and phase standard deviation.

2.4 Follow-Up and Outcomes

All enrolled patients were followed up through outpatient visits, hospital records, and telephone interviews. Major adverse cardiac events (MACE) were recorded, including all-cause mortality, nonfatal myocardial infarction or ischemic stroke, coronary revascularization, and rehospitalization due to cardiovascular or cerebrovascular causes. Based on follow-up outcomes, patients were classified into groups with favorable or unfavorable prognosis.

2.5 Data Preprocessing

Before analysis, continuous variables were examined for outliers. Extreme values were adjusted toward the data distribution using boxplot-based methods (**Supplementary Figure S1**). All laboratory variables were retained as continuous variables, and categorical predictors were predefined before model construction. Variables with missing data in the training and test cohorts are summarized in **Supplementary Figure S2**. Variables with more than 20% missing values were excluded. For the remaining variables, missing values were imputed using multiple imputation with the "mice" package in R. To assess the robustness of model performance across different imputation strategies, six missing data imputation methods were compared: Bayesian linear regression, classification and regression trees (CART), k-nearest neighbors (kNN), mean/mode imputation, predictive mean matching with logistic and polynomial regression (pmm_logreg_polyreg), and random forest imputation. The results of the sensitivity analysis are presented in **Supplementary Figure S3 and Table S1**.

2.6 Feature Selection

After multiple imputation, least absolute shrinkage and selection operator (LASSO) regression was applied for feature selection to reduce model complexity and enhance predictive stability. LASSO is a widely used feature selection technique in machine learning that removes less informative variables and identifies an optimal subset of predictors by shrinking regression coefficients toward zero through L1 regularization. Feature selection was performed using 10 repetitions of 10-fold cross-validation to ensure robustness of the selected feature set.

2.7 Model Development and Validation

Based on the features selected by LASSO, multiple machine learning models were developed and evaluated in a Python environment, including logistic regression, k-nearest neighbors, support vector machines, naive Bayes, decision trees, random forest, multilayer perceptron (MLP), and gradient boosting–based models (XGBoost, LightGBM, and CatBoost). Continuous variables were standardized using z-score normalization, and categorical variables were encoded using label encoding. The study population was randomly stratified by outcome into a training cohort (70%) and an independent test cohort (30%).

2.8 Model Performance Evaluation

Model performance was assessed using multiple commonly accepted metrics, including the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, F1 score, and Brier score. The Hosmer–Lemeshow test was used to evaluate the agreement between predicted probabilities and observed outcomes. Logarithmic loss (log-loss) was calculated to quantify the discrepancy between predicted probabilities and true labels. Calibration curves were constructed to assess the concordance between predicted and observed risks. The DeLong test was applied to compare differences in AUC values between models. Decision curve analysis (DCA) was performed to evaluate the net clinical benefit across a range of threshold probabilities. The optimal predictive model was selected based on overall performance in both the training and test cohorts. Shapley additive explanations (SHAP) were used to interpret model outputs and assess feature importance.

2.9 Statistical Analysis

Continuous variables are presented as median (interquartile range) and were compared using the Mann–Whitney U test. Categorical variables are presented as counts (percentages) and were compared using the chi-squared test or Fisher's exact test, as appropriate. Statistical significance was defined as a two-sided P value < 0.05. All statistical analyses were performed using R software version 4.4.3. Machine learning model development and SHAP analyses were conducted using Python version 3.12 with relevant libraries, including scikit-learn, XGBoost, LightGBM, CatBoost, and SHAP.

3. Results

3.1 Baseline Clinical Characteristics

The median follow-up duration for the study cohort was 3.1 years (interquartile range, 1.7–4.9 years; range, 0.1–6.6 years). During follow-up, a total of 662 patients experienced MACE (39.5%), including 71 deaths (4.2%). The test set included 1,173 patients, among whom 461 experienced MACE (39.3%), whereas the training set comprised 502 patients, of whom 201 experienced MACE (40.0%). Baseline clinical and imaging characteristics of all patients are summarized in Table 1. No significant differences were observed between the training and test cohorts across all variables, indicating balanced patient distribution and comparability between the datasets.

Table 1
Patient demographics and baseline characteristics.

Characteristic, n (%), Median (Q1, Q3)	Overall N = 1,675	Training Cohort N = 1,173	Internal Test Cohort N = 502	p-value[1]
Sex (man)	1,109 (66.2%)	781 (66.6%)	328 (65.3%)	0.622
Age	62 (53, 71)	62 (53, 71)	61 (52, 70)	0.351
Body Mass Index	24.3(22.2, 26.7)	24.2 (22.2, 26.6)	24.6 (22.4, 26.8)	0.290
Coronary stenosis ≥ 50%	503 (30.0%)	354 (30.2%)	149 (29.7%)	0.839
Previous history of Cardio-cerebrovascular disease	856 (51.1%)	609 (51.9%)	247 (49.2%)	0.309
NYHA Functional Class				0.408
I	107 (6.4)	58 (4.9)	50 (10.0)	
II	603 (36.0)	446 (38.0)	156 (31.0)	
III	677 (40.4)	463 (39.5)	216 (43.0)	
IV	288 (17.2)	206 (17.6)	80 (16.0)	
Hyperlipidemia	428 (25.6%)	293 (25.0%)	135 (26.9%)	0.411
Hyperhomocysteinemia	188 (11.2%)	135 (11.5%)	53 (10.6%)	0.572
Hyperuricemia	102 (6.1%)	74 (6.3%)	28 (5.6%)	0.567
hypertension	957 (57.1%)	663 (56.5%)	294 (58.6%)	0.439
Smoking	729 (43.5%)	517 (44.1%)	212 (42.2%)	0.486
Diabetes	437 (26.1%)	310 (26.4%)	127 (25.3%)	0.630
White Blood Cell Count	6.30 (5.20, 7.50)	6.30 (5.20, 7.50)	6.20 (5.10, 7.70)	0.895
Neutrophil count	3.70 (2.90, 4.80)	3.70 (2.92, 4.80)	3.70 (2.90, 4.80)	0.641
Lymphocyte count	1.80 (1.40, 2.20)	1.79 (1.40, 2.20)	1.80 (1.40, 2.20)	0.309
Red Blood Cell Count	4.49 (4.14, 4.83)	4.48 (4.14, 4.83)	4.54 (4.18, 4.85)	0.372
Hemoglobin	142 (130, 152)	141 (130, 152)	142 (132, 152)	0.253
Creatinine	71 (60, 81)	71 (61, 82)	70 (60, 81)	0.515
Blood urea nitrogen	5.20 (4.24, 6.25)	5.23 (4.26, 6.26)	5.06 (4.18, 6.24)	0.119
Uric acid	331 (276, 400)	333 (278, 402)	330 (271, 394)	0.158
Fasting blood glucose	5.20 (4.71, 6.28)	5.24 (4.72, 6.32)	5.15 (4.67, 6.15)	0.235
Triglyceride	1.45 (1.06, 1.96)	1.44 (1.05, 1.99)	1.47 (1.10, 1.92)	0.702
Total Cholesterol	3.99 (3.23, 4.77)	3.98 (3.23, 4.76)	4.00 (3.21, 4.79)	0.888
High-density lipoprotein cholesterol	0.99 (0.85, 1.15)	0.99 (0.85, 1.16)	1.00 (0.85, 1.14)	0.937
Low-density lipoprotein cholesterol	2.38 (1.81, 3.01)	2.37 (1.82, 3.01)	2.41 (1.80, 3.01)	0.689
Total perfusion deficit	8 (4, 16)	8 (4, 17)	7 (4, 14)	0.109
Peak filling rate	1.63 (1.27, 2.07)	1.61 (1.27, 2.06)	1.69 (1.28, 2.09)	0.216
The first third of the end-systole to end-diastole phase	0.60 (0.48, 0.73)	0.59 (0.48, 0.73)	0.61 (0.50, 0.74)	0.174
Time to peak filling from end-systole	255 (225, 283)	256 (223, 284)	254 (227, 281)	0.924
Summed Motion Score	13 (6, 26)	14 (6, 27)	13 (6, 25)	0.357
Summed Thickening Score	8 (2, 16)	8 (2, 17)	7 (1, 16)	0.262
Left ventricular ejection fraction	47 (36, 54)	46 (36, 54)	47 (37, 55)	0.268
End-diastolic volume	84 (64, 112)	85 (64, 113)	84 (64, 108)	0.530
End-systolic volume	45 (30, 69)	45 (30, 70)	44 (29, 67)	0.379
Shape index (no-gated)	0.60 (0.55, 0.65)	0.60 (0.55, 0.65)	0.60 (0.55, 0.65)	0.948

[1]Continuous variables are presented as median (interquartile range) and were compared using the Wilcoxon rank-sum test. Categorical variables are presented as counts (percentages) and were compared using Pearson's chi-squared test or Fisher's exact test, as appropriate.

Characteristic, n (%), Median (Q1, Q3)	Overall N = 1,675	Training Cohort N = 1,173	Internal Test Cohort N = 502	p-value[1]
Eccentricity index (no-gated)	0.82 (0.79, 0.85)	0.82 (0.79, 0.85)	0.82 (0.79, 0.85)	0.921
End diastolic shape index	0.63 (0.59, 0.68)	0.63 (0.59, 0.68)	0.63 (0.59, 0.68)	0.742
End systolic shape index	0.51 (0.46, 0.58)	0.51 (0.46, 0.58)	0.51 (0.47, 0.58)	0.908
Eccentricity index	0.84 (0.80, 0.87)	0.84 (0.80, 0.87)	0.84 (0.80, 0.87)	0.965
Histogram bandwidth	54 (36, 84)	54 (36, 84)	54 (36, 84)	0.473
Mean	130 (119, 142)	130 (119, 142)	131 (119, 142)	0.833
Phase standard deviation	19 (12, 27)	19 (12, 27)	19 (12, 27)	0.677
Entropy	47 (40, 54)	47 (40, 54)	46 (39, 54)	0.247

[1]Continuous variables are presented as median (interquartile range) and were compared using the Wilcoxon rank-sum test. Categorical variables are presented as counts (percentages) and were compared using Pearson's chi-squared test or Fisher's exact test, as appropriate.

Furthermore, patients with favorable versus adverse outcomes demonstrated significant differences in multiple clinical and imaging parameters (**Supplementary Table S2**). Compared with patients without MACE, those who experienced MACE were more likely to be male, older, and have a history of prior cardiovascular or cerebrovascular disease, as well as worse functional status. Several imaging-derived parameters-including PFR, TPD, and entropy-also differed significantly between groups, providing a rationale for subsequent multivariable modeling.

3.2 Feature Selection

To ensure robust feature selection, a recursive feature elimination (RFE) framework was implemented in conjunction with six distinct missing-data imputation methods. Post-imputation, the distributions of all variables were compared with pre-imputation distributions, revealing high concordance (**Supplementary Tables S3 and S4)**, indicating that multiple imputation did not introduce substantial bias.

Correlation analyses of continuous variables revealed significant inter-variable associations, such as between histogram bandwidth and phase entropy (Fig. 2a). To mitigate multicollinearity, LASSO regression was applied for feature selection (Figs. 2b and 2c). Based on the LASSO results and current literature, nine variables were ultimately retained: history of prior cardiovascular or cerebrovascular disease, age, PFR, low-density lipoprotein cholesterol, TPD, EI (no-gated), entropy, ≥ 50% coronary stenosis on angiography, and NYHA functional class.

To further evaluate the stability and consistency of these features across different machine learning algorithms, feature importance was systematically compared within the RFE framework for multiple models (**Supplementary Figure S4**).

3.3 Model Development and Performance Comparison

To identify the optimal model for predicting MACE, ten machine learning algorithms were evaluated using the training dataset with ten-fold cross-validation repeated ten times. Cross-validation results (Table 2) demonstrated that the CatBoost model exhibited the highest stability and generalizability, achieving the highest mean cross-validated area under the receiver operating characteristic curve (cross-validated area under the receiver operating characteristic curve, CV_AUC_Mean = 0.693) with the lowest standard deviation (CV_AUC_S = 0.012). Other top-performing models included logistic regression (CV_AUC_Mean = 0.690) and random forest (CV_AUC_Mean = 0.687).

Table 2
Performance parameters of the ten machine learning prediction models in the training set.

Model	CV_AUC_Mean	CV_AUC_Std	Train_AUC	Test_AUC	Accuracy	Precision	Recall	F1	Sensitivity	Specificity	Brier
CatBoost	0.693	0.012	0.871	0.629	0.632	0.551	0.382	0.451	0.382	0.796	0.234
Logistic	0.690	0.021	0.700	0.634	0.626	0.537	0.397	0.457	0.397	0.776	0.230
RF	0.687	0.015	0.935	0.627	0.628	0.542	0.392	0.455	0.392	0.783	0.233
SVM	0.686	0.025	0.776	0.598	0.612	0.513	0.392	0.444	0.392	0.757	0.239
XGBoost	0.679	0.017	0.879	0.600	0.584	0.468	0.367	0.411	0.367	0.727	0.244
Naïve Bayes	0.677	0.024	0.683	0.627	0.618	0.520	0.457	0.487	0.457	0.724	0.294
LightGBM	0.665	0.014	0.981	0.583	0.594	0.485	0.397	0.436	0.397	0.724	0.261
Decision Tree	0.638	0.034	0.810	0.562	0.581	0.466	0.412	0.437	0.412	0.691	0.292
KNN	0.633	0.023	0.806	0.584	0.586	0.475	0.437	0.455	0.437	0.684	0.272
MLP	0.600	0.034	0.997	0.538	0.545	0.424	0.422	0.423	0.422	0.625	0.392

RF, Random Forest; SVM, Support Vector Machine; KNN, K-Nearest Neighbors; MLP, Multilayer Perceptron; CV_AUC, Cross Validated_Area Under the Receiver Operating Characteristic Curve.

3.4 Model Discrimination, Calibration, and Clinical Utility

In the training cohort, all models demonstrated generally high discriminative performance as measured by the AUC (Fig. 3a). Notably, the MLP, LightGBM, and random forest models achieved training AUCs of 0.997, 0.981 and 0.935, respectively, suggesting potential overfitting. In contrast, the CatBoost and logistic regression models exhibited more moderate training performance, with AUCs of 0.871 and 0.700, respectively, indicating appropriate model fitting.

When evaluated in the independent test cohort (Fig. 3b), the generalization performance of the models diverged. logistic regression achieved the highest test AUC (0.634), followed closely by CatBoost (AUC = 0.629) and random forest (AUC = 0.627). More complex models such as XGBoost and LightGBM exhibited pronounced declines in test AUC, indicating relatively weaker generalizability. **Supplementary Figure S5** presents the confusion matrices and corresponding accuracy for all ten machine learning models in the test cohort. Although some models demonstrated comparable overall AUC, differences in sensitivity-specificity trade-offs were observed. Among them, CatBoost maintained a relatively balanced performance in identifying both event and non-event patients, suggesting more robust and clinically practical predictive capacity.

Calibration curves (Fig. 4) indicated that in both the training (Fig. 4a) and test cohorts (Fig. 4b), the predicted probabilities of CatBoost and logistic regression models closely aligned with the observed event rates, approximating the ideal diagonal line. This was further confirmed by Brier scores, with CatBoost (0.234) and logistic regression (0.230) achieving the lowest scores, reflecting superior probability calibration.

The DCA was used to evaluate the net clinical benefit of each model across a range of decision thresholds. As shown in Fig. 5, in both the training (Fig. 5a) and test cohorts (Fig. 5b), CatBoost provided the highest standardized net benefit across the majority of threshold probabilities (approximately 0.1–0.8), indicating excellent clinical utility. To further assess the robustness of clinical benefit, **Supplementary Figure S6** presents bootstrap-corrected DCA curves (200 iterations; 95% confidence interval), confirming the stability of CatBoost's superior performance.

Although logistic regression achieved similar performance on certain metrics (e.g., test AUC), its cross-validation stability (CV_AUC_SD = 0.0209) was inferior to CatBoost. Conversely, models such as random forest and XGBoost exhibited clear overfitting, as evidenced by large discrepancies between training and test AUCs.

To further assess the stability of key predictive features across different machine learning models, **Supplementary Figure S7** displays a normalized feature importance heatmap for all models in the test cohort. Despite differences in modeling algorithms, several core variables consistently demonstrated high relative importance, indicating robust associations with major adverse cardiac events. This finding supports the rationality of the feature selection strategy employed and lays the groundwork for subsequent SHAP-based interpretability analyses. Based on these comprehensive evaluations, CatBoost was selected as the optimal model for predicting distant adverse outcomes and is recommended for subsequent clinical risk assessment and decision support.

3.4 Model Interpretation

SHAP analysis was applied to enhance the interpretability of the predictive model. Figure 6a displays the summary bar plot of SHAP values, ranking features in descending order of importance. Figure 6b visualizes the top nine predictors for MACE, with the most influential features being: history of prior cardiovascular or cerebrovascular disease, age, PFR, low-density lipoprotein cholesterol, TPD, EI (no-gated), entropy, ≥ 50% coronary stenosis, and NYHA functional class. In the plot, red indicates higher feature values associated with increased risk, while blue indicates lower values associated with reduced risk.

Figure 6c presents partial dependence plots for the four most important variables, illustrating their relationships and interactions. Figure 6d shows example predictions for three individual cases, where red and green arrows indicate positive and negative contributions of each feature to the predicted risk, and the

arrow length represents the magnitude of influence. This visualization provides a transparent and intuitive understanding of the model's decision logic at the individual patient level.

4. Discussion

To our knowledge, this is the first large-scale, real-world cohort study integrating clinical data with SPECT MPI parameters to systematically compare ten machine learning models for predicting adverse clinical outcomes in patients with CAD and preserved ejection fraction. We identified a set of predictive risk factors and developed a risk prediction model combining clinical, laboratory, and imaging data, tailored for CAD patients with preserved left ventricular systolic function.

Previous studies have reported risk prediction for MACE in patients with preserved ejection fraction, but most relied solely on clinical data[14], potentially overlooking unmeasured residual risk factors. SPECT MPI, widely accessible in clinical practice, enables simultaneous assessment of myocardial perfusion, morphology, function, and mechanical synchrony, making it a valuable noninvasive tool. Integration of electronic health record data, SPECT MPI parameters, and advanced machine learning algorithms facilitates the development of robust, clinically relevant predictive models.

Currently, there are no consensus guidelines on the optimal number of features for predictive models. In this study, 45 candidate variables were initially considered. To reduce multicollinearity and enhance model stability, we employed LASSO regression for feature selection. While LASSO retained 24 variables at λ_min, this was excessive for clinical interpretability. At λ_1se, only 7 variables were retained, which excluded key SPECT MPI parameters well-documented in the literature. Balancing statistical results with clinical relevance, we ultimately selected 9 core features for model construction, ensuring both physiological interpretability and practical complexity suitable for clinical application.

Among the ten machine learning models, CatBoost consistently outperformed others across multiple metrics. Although the test-set AUC of CatBoost was approximately 0.63, this should be interpreted in the context of real-world data. MACE is a multifactorial, long-term outcome influenced not only by imaging and clinical variables but also by treatment adherence, medication adjustments, and unmeasured confounders, making high AUCs challenging. Previous studies using SPECT MPI or clinical data for external validation reported AUCs ranging from 0.60–0.70, indicating the comparability of our results[9, 15, 16]. Furthermore, our strict training–test separation, low event rate in the test set, and independent cohort design more realistically reflect generalizability, whereas some complex models showed overfitting, with markedly decreased test performance. Importantly, CatBoost demonstrated stable calibration (low Brier score) and superior clinical net benefit across a wide range of thresholds in decision curve analysis, supporting its utility for risk stratification and clinical decision-making even with moderate discriminative power. The final CatBoost model incorporated nine features, combining clinical, biochemical, and SPECT imaging parameters, enabling scalable individualized risk prediction. History of prior cardiovascular or cerebrovascular disease emerged as the most important predictor, reflecting the impact of systemic atherosclerotic burden. Age remained a robust risk factor, consistent with vascular aging, structural remodeling, and cumulative cardiovascular risk[17]. Coronary stenosis ≥ 50% and NYHA functional class reflected anatomical and functional severity, representing traditional clinical risk markers[18].

Several SPECT MPI-derived imaging parameters ranked highly in feature importance. PFR reflects early left ventricular diastolic function; reduced PFR indicates impaired myocardial relaxation and decreased compliance, often associated with ischemia, microvascular dysfunction, and early diastolic abnormalities[10]. TPD quantifies the extent and severity of ischemia, correlating with coronary stenosis, ischemic burden, and myocardial viability[19]. EI (no-gated) reflects left ventricular geometric remodeling, with higher values indicating progression toward spherical morphology, often associated with chronic ischemic cardiomyopathy[20]. Entropy measures mechanical synchrony, with higher values indicating dyssynchrony from multifocal ischemia, microvascular dysfunction, or myocardial fibrosis[21].

The consistent high-ranking of these imaging features reveals the continuum of cardiovascular pathology from anatomical stenosis to functional impairment and structural remodeling. SHAP analysis enhanced model interpretability, allowing clinicians to visualize feature contributions at both the population and individual patient levels. This transparency facilitates trust and supports the use of the model in personalized risk assessment, guiding tailored therapy and follow-up strategies.

Overall, these findings highlight the independent prognostic value of diastolic function, myocardial perfusion, ventricular geometry, and mechanical synchrony derived from SPECT MPI, and demonstrate the utility of machine learning models in integrating multidimensional imaging and clinical data for precise cardiovascular risk prediction.

This study has several limitations. First, it was a single-center, retrospective study, which may limit the generalizability of the findings. Although rigorous cross-validation and an independent test cohort were employed, further validation in multicenter, prospective studies is warranted. Second, the sample size and number of events were relatively limited, potentially affecting the stability of more complex models. Third, due to clinical workflow constraints and patient conditions, some participants did not undergo stress SPECT imaging, preventing the assessment of myocardial perfusion and function under stress conditions. This limitation may reduce the ability to detect microvascular dysfunction or impaired diastolic response during stress. Nevertheless, resting SPECT remains valuable for evaluating myocardial morphology, wall thickness, chamber size, and diastolic function. In this study, resting measures such as PFR, TPD, entropy, and EI were still strongly associated with clinical outcomes, suggesting that resting imaging retains predictive utility when stress data are unavailable. Additionally, the present model predicts fixed-time risk, and future work could incorporate time-to-event machine learning approaches to enhance prognostic information. Although SHAP analysis improves interpretability, it remains an associative method and cannot establish causality.

5. Conclusion

In summary, we successfully developed an interpretable machine learning model integrating multimodal clinical and imaging data to predict the risk of MACE in CAD patients with preserved ejection fraction. The CatBoost model outperformed other models in predicting adverse outcomes. Feature importance analyses further highlighted the potential pathophysiological relevance of key predictors, providing additional insights for understanding disease mechanisms and supporting clinical decision-making. Future studies should validate the model in larger, prospective, multicenter cohorts and explore its deployment in real-world clinical settings.

Abbreviations

coronary artery disease
CAD
left ventricular ejection fraction
LVEF
myocardial perfusion imaging
MPI
major adverse cardiac events
MACE
the area under the receiver operating characteristic curve
AUC
Shapley additive explanations
SHAP
New York Heart Association
NYHA
heart failure with preserved ejection fraction
HFpEF
single-photon emission computed tomography
SPECT
total perfusion deficit
TPD
peak filling rate
PFR
eccentricity index
EI
Random Forest
RF
Support Vector Machine
SVM
K-Nearest Neighbors
KNN
Multilayer Perceptron
MLP
classification and regression trees
CART
positive predictive value
PPV
negative predictive value
NPV
Decision curve analysis
DCA
least absolute shrinkage and selection operator
LASSO

Declarations

Ethics approval and consent to participate:

The study protocol received ethical clearance from the Ethics Committee of the First Hospital of Shanxi Medical University (No. 2022-K-128). All research activities involving human subjects complied with the Declaration of Helsinki, and written informed consent was obtained from all participants.

Consent for publication:

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This study was financially supported by the National Natural Science Foundation of China (U22A6008, 82027804), Shanxi Province Higher Education "Billion Project" Science and Technology Guidance Project, Research Project Supported by Shanxi Scholarship Council of China (No.2024 – 147).

Conflict of Interest:

I have nothing to declare.

Author Contribution

YL drafted the original manuscript, performed investigation, conducted formal analysis, and contributed to conceptualization. JC reviewed and edited the manuscript, and developed the methodology. YZ, PW, and RW curated the data. ZW, HL, and SL reviewed and edited the manuscript, supervised the study, and acquired funding. All authors read and approved the final manuscript.

Acknowledgements:

Not applicable.

Availability of data and materials

All the data in this study are shown in the manuscript, and Additional file materials, and unprocessed data are available from the corresponding author upon reasonable request.

References

1. 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;76:2982–3021.
2. Fayyaz AU, Eltony M, Prokop LJ, Koepp KE, Borlaug BA, Dasari S, et al. Pathophysiological insights into HFpEF from studies of human cardiac tissue. Nat reviews Cardiol. 2025;22:90–104.
3. Capone F, Sotomayor-Flores C, Bode D, Wang R, Rodolico D, Strocchi S, et al. Cardiac metabolism in HFpEF: from fuel to signalling. Cardiovascular Res. 2023;118:3556–75.
4. Sirajuddin A, Mirmomen SM, Kligerman SJ, Groves DW, Burke AP, Kureshi F, et al. Ischemic Heart Disease: Noninvasive Imaging Techniques and Findings. Radiographics: Rev publication Radiological Soc North Am Inc. 2021;41:990–1021.
5. George RT, Mehra VC, Chen MY, Kitagawa K, Arbab-Zadeh A, Miller JM, et al. Myocardial CT Perfusion Imaging and SPECT for the Diagnosis of Coronary Artery Disease: A Head-to-Head Comparison from the CORE320 Multicenter Diagnostic Performance Study. Radiology. 2015;274:626.
6. Garcia EV, Slomka P, Moody JB, Germano G, Ficaro EP. Quantitative Clinical Nuclear Cardiology, Part 1: Established Applications. J nuclear cardiology: official publication Am Soc Nuclear Cardiol. 2020;27:189–201.
7. Slomka PJ, Moody JB, Miller RJH, Renaud JM, Ficaro EP, Garcia EV. Quantitative clinical nuclear cardiology, part 2: Evolving/emerging applications. J nuclear cardiology: official publication Am Soc Nuclear Cardiol. 2021;28:115–27.
8. Fukumoto K, Yoda S, Tanaka Y, Hatta T, Suzuki Y, Matsumoto N, et al. Prognostic value of stress phase entropy for predicting major cardiac events in patients with known or suspected stable coronary artery disease. Heart Vessels. 2025;40:677–87.
9. Otaki Y, Singh A, Kavanagh P, Miller RJH, Parekh T, Tamarappoo BK, et al. Clinical Deployment of Explainable Artificial Intelligence of SPECT for Diagnosis of Coronary Artery Disease. JACC Cardiovasc imaging. 2022;15:1091–102.
10. Xu B, Liu L, Abdu FA, Yin G, Mohammed AQ, Xu S, et al. Prognostic Value of Diastolic Dysfunction Derived From D-SPECT in Coronary Artery Disease Patients With Normal Ejection Fraction. Front Cardiovasc Med. 2021;8:700027.
11. Zhao Y, Hu Y, Li Y, Wang Y, Xiao Y, Xu L, et al. Spherization indices measured by resting SPECT improve risk stratification in patients with ischemia with non-obstructive coronary artery disease (INOCA). EJNMMI Res. 2024;14:16.
12. Wang J, Fan X, Qin S, Shi K, Zhang H, Yu F. Exploration of the efficacy of radiomics applied to left ventricular tomograms obtained from D-SPECT MPI for the auxiliary diagnosis of myocardial ischemia in CAD. Int J Cardiovasc Imaging. 2022;38:465–72.
13. Forrest IS, Petrazzini BO, Duffy Á, Park JK, Marquez-Luna C, Jordan DM, et al. Machine learning-based marker for coronary artery disease: derivation and validation in two longitudinal cohorts. Lancet (London England). 2023;401:215–25.

14. Si J, Ding Z, Hu Y, Zhang X, Zhang Y, Cao H, et al. Predictors and prognostic implications of left ventricular ejection fraction trajectory improvement in the spectrum of heart failure with reduced and mildly reduced ejection fraction. J Cardiol. 2024;83:250–7.
15. Miller RJH, Kuronuma K, Singh A, Otaki Y, Hayes S, Chareonthaitawee P, et al. Explainable Deep Learning Improves Physician Interpretation of Myocardial Perfusion Imaging. Journal of nuclear medicine: official publication. Soc Nuclear Med. 2022;63:1768–74.
16. Singh A, Miller RJH, Otaki Y, Kavanagh P, Hauser MT, Tzolos E, et al. Direct Risk Assessment From Myocardial Perfusion Imaging Using Explainable Deep Learning. JACC Cardiovasc imaging. 2023;16:209–20.
17. Jansen SA, Huiskens B, Trompet S, Jukema J, Mooijaart SP, van Willems K, et al. Classical risk factors for primary coronary artery disease from an aging perspective through Mendelian Randomization. GeroScience. 2022;44:1703–13.
18. Wang C, Yu W, Yang X, Sun F, Tian Y, Yue Y, et al. Association of epicardial fat volume index with coronary artery disease phenotypes based on anatomical and functional imaging: a linkage with the severity of coronary artery disease. Cardiovasc Diabetol. 2025;24:387.
19. Azadani PN, Miller RJH, Sharir T, Diniz MA, Hu LH, Otaki Y, et al. Impact of Early Revascularization on Major Adverse Cardiovascular Events in Relation to Automatically Quantified Ischemia. JACC Cardiovasc imaging. 2021;14:644–53.
20. Zhao Y, Liu Y, Zheng Y, Xu X, Wang Y, Zhang X. Left ventricular sphericity indexes by gated single photon emission computed tomography myocardial perfusion imaging: potential opportunities and challenges for cardiac remodeling and clinical prognosis-a narrative review. Quant imaging Med Surg. 2025;15:12857–70.
21. Kuronuma K, Miller RJH, Otaki Y, Van Kriekinge SD, Diniz MA, Sharir T, et al. Prognostic Value of Phase Analysis for Predicting Adverse Cardiac Events Beyond Conventional Single-Photon Emission Computed Tomography Variables: Results From the REFINE SPECT Registry. Circulation Cardiovasc imaging. 2021;14:e012386.

Figures

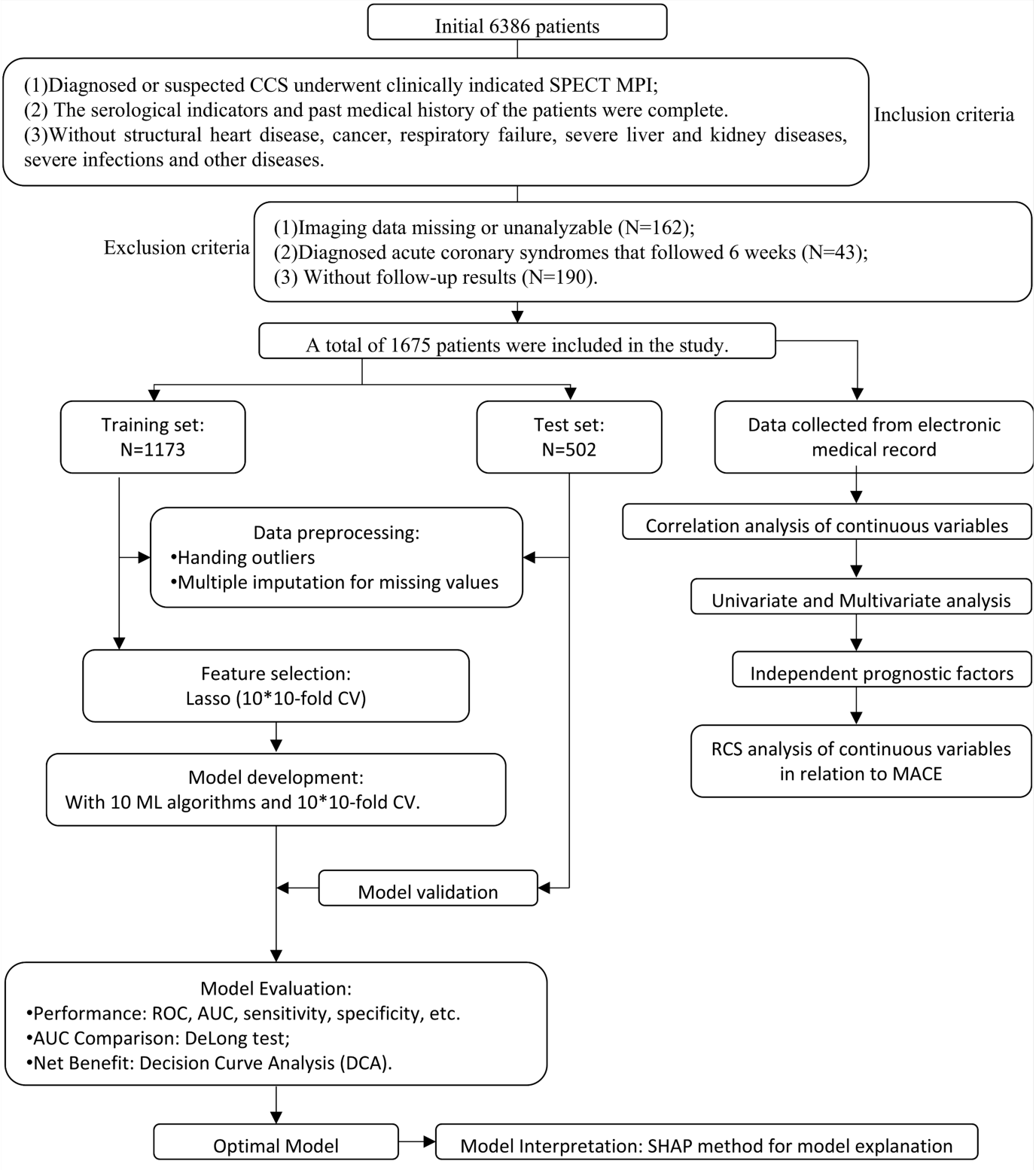


Figure 1

The flowchart of the study.

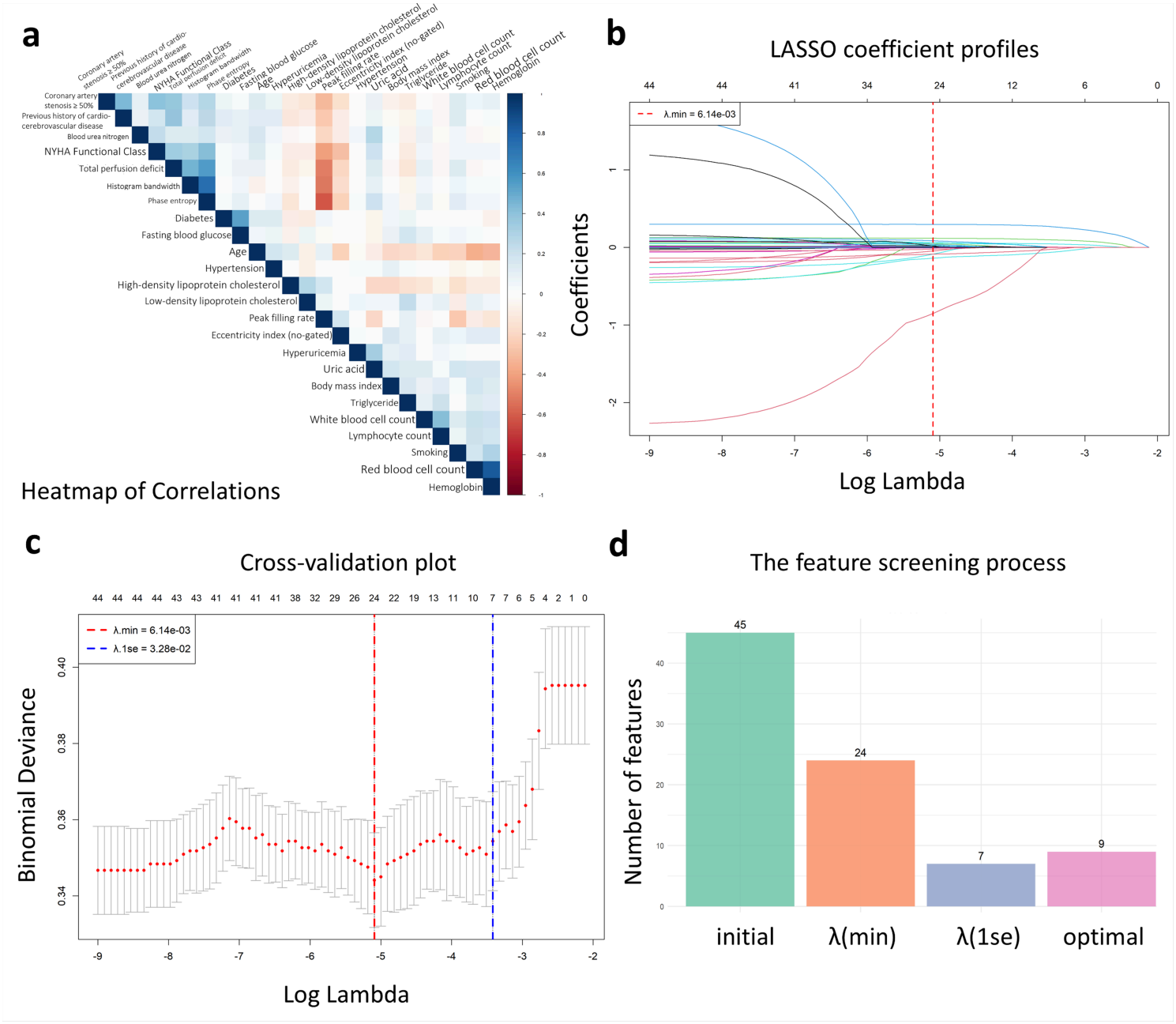


Figure 2

Process of Selecting Variables through LASSO Regression.

(a) heatmap depicting pairwise correlations among variables, (b) the LASSO coefficient spectrum of 45 candidate variables across log(λ), illustrating coefficient shrinkage trends, (c) the binomial deviance curve with optimal λ indicated by a vertical dashed line, and (d) the integration of LASSO results with literature-based evidence to select the final nine predictors for model development.

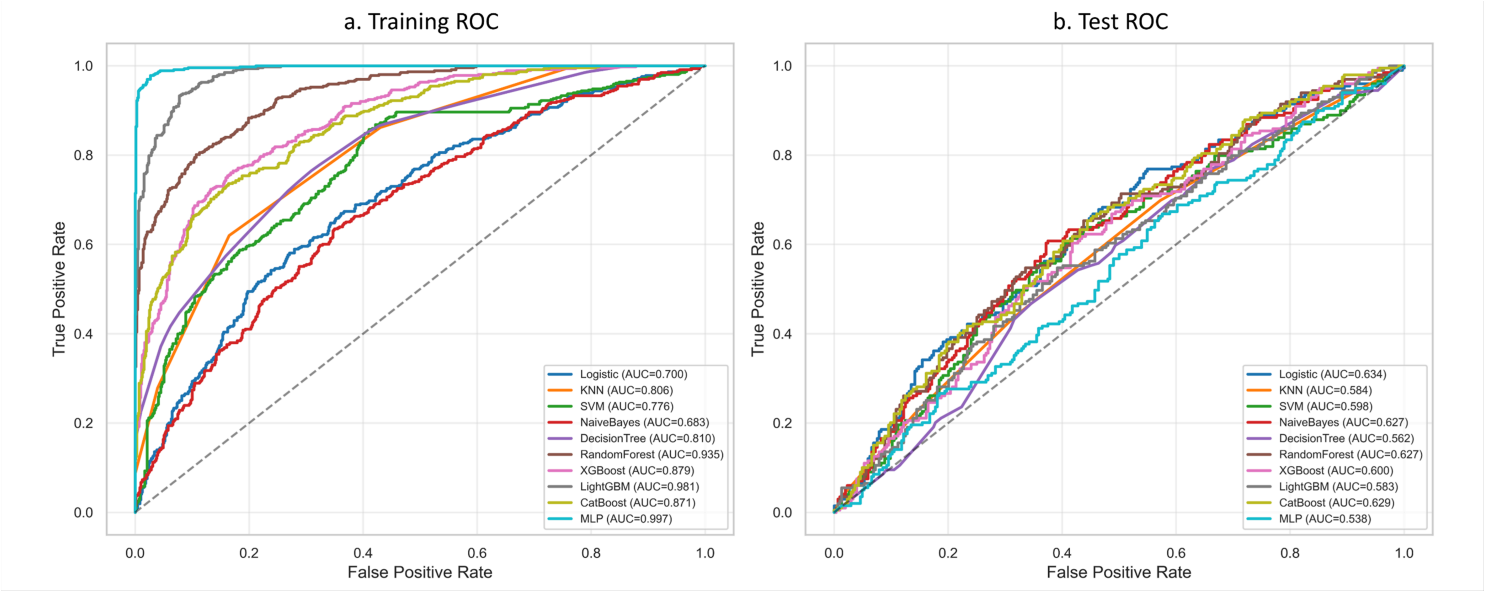


Figure 3

ROC curve analysis.

(a) ROC curves in the training cohort. The MLP model demonstrated near-perfect discriminative performance with an area under the curve (AUC) of 0.997. (b) ROC curves in the independent test cohort. The logistic model maintained the highest generalizability among all models, with an AUC of 0.634.

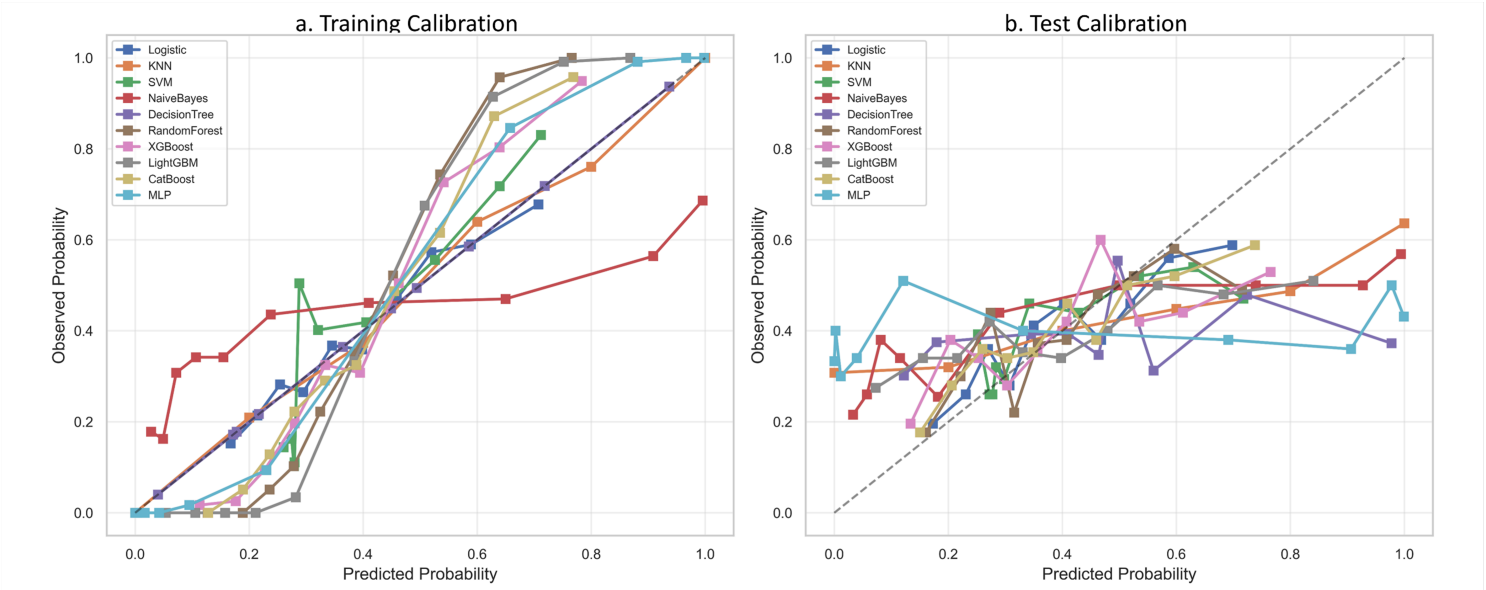


Figure 4

Calibration curve analysis.

(a) Calibration in the training cohort. The RF model's calibration curve closely aligned with the ideal 45° line, indicating near-perfect calibration (calibration score = 0.004). (b) Calibration in the test cohort. The RF model achieved a post-calibration score of 0.057, and the Hosmer–Lemeshow test confirmed good fit (P > 0.05). The 45° dashed line represents perfect calibration, where predicted probabilities equal observed event rates.

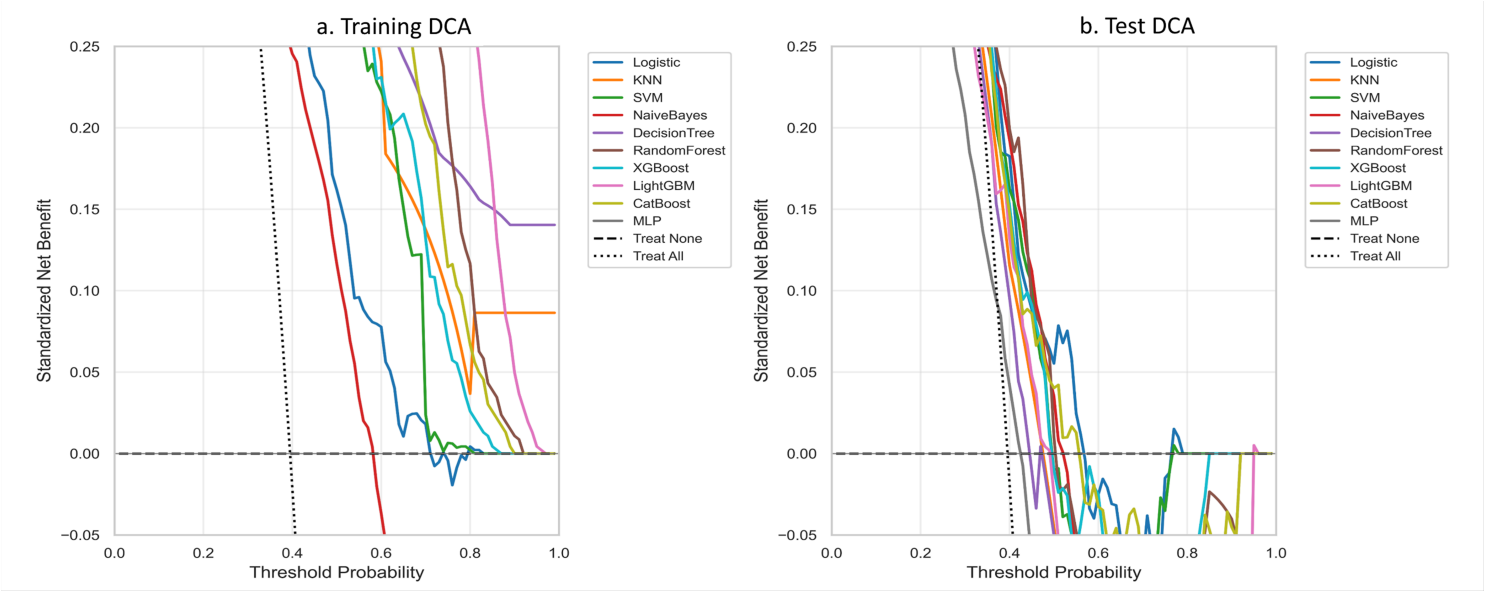


Figure 5

DCA curves for each model.

(a) DCA in the training cohort. (b) DCA in the test cohort. Across the full range of threshold probabilities (training 0–1.0; test 0–0.8), the CatBoost model provided the highest standardized net benefit over the majority of thresholds (~0.1–0.8).

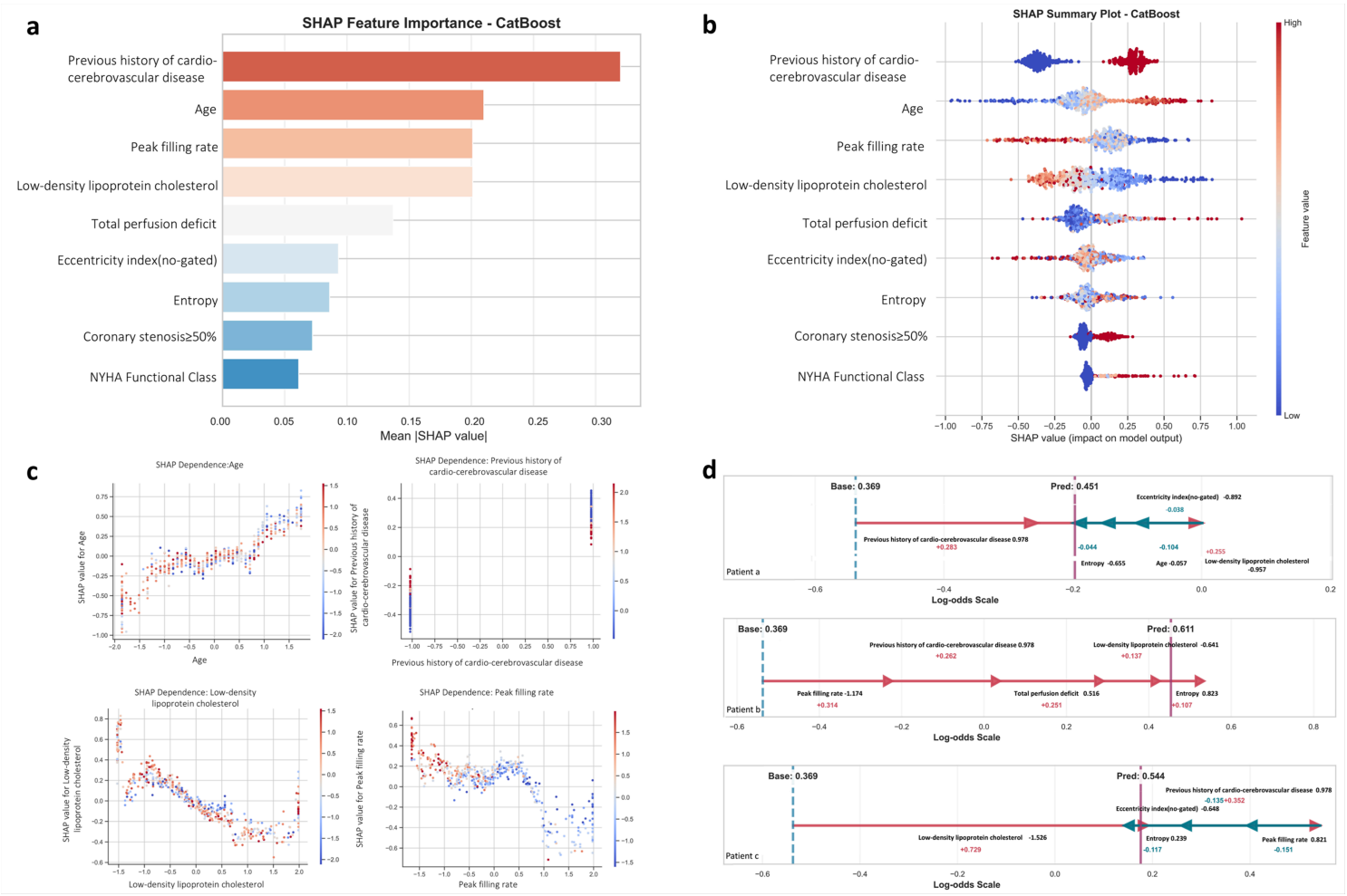


Figure 6

SHAP interpretation of the CatBoost model.

(a) Summary bar plot of SHAP values, ranking features by mean absolute SHAP value and highlighting the top nine predictors. (b) Summary dot plot of SHAP values for each feature; each row represents a feature, the x-axis represents SHAP value, and color indicates feature magnitude (red = high, blue = low), illustrating the relationship between feature values and impact on risk. (c) Partial dependence plots for the four most important features, showing their effect on predicted MACE risk. (d) SHAP plots for three representative cases; red and green arrows indicate positive and negative contributions of each feature to the prediction, with arrow length reflecting the magnitude of impact, demonstrating the model's logic for individual predictions.

Supplementary Files

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

- SupplementaryMaterial.docx