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.

Platelet-to-Lymphocyte Ratio-Driven Machine Learning Model for Predicting In-Hospital Major Adverse Cardiovascular Events After Percutaneous Coronary Intervention in Acute Myocardial Infarction: A Retrospective Cohort Study

Xiren Meng
Shenyang Medical College
Siyu Pan
Shenyang Medical College
Shuochao Zhao
Shenyang Medical College
Yaxin Wang
Shenyang Medical College
Hao Li
Shenyang Medical College
Zhen Shan
Shenyang Medical College
Yueran Yu
Shenyang Medical College
Yikang Xu
yikangxu@163.com

The Second Affiliated Hospital of Shenyang Medical College

Article

Keywords: Acute myocardial infarction, Percutaneous coronary intervention, Platelet-to-lymphocyte ratio, Machine learning, XGBoost, SHAP, Major adverse cardiovascular events

Posted Date: August 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10458680/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

Accurate risk assessment of major adverse cardiovascular events (MACE) following percutaneous coronary intervention (PCI) in patients with acute myocardial infarction (AMI) is central to optimizing clinical decision-making. Although inflammatory mechanisms play a crucial role in ischemia-reperfusion injury, previous studies have largely been limited to single inflammatory markers, lacking systematic integration and comparative validation of multidimensional inflammatory indices. Moreover, which inflammatory marker possesses the optimal driving efficacy within a machine learning framework remains unclear.

Objective

This study aimed to systematically integrate five commonly used multidimensional inflammatory indices (PLR, MLR, MHR, SII, and SIRI), construct risk prediction models for in-hospital MACE after PCI in AMI patients based on traditional logistic regression and nine machine learning algorithms, establish PLR as a core driving factor, and explore nonlinear relationships between key clinical indicators and MACE.

Methods

A total of 1,163 consecutive AMI patients who underwent PCI at the Second Affiliated Hospital of Shenyang Medical College from 2020 to 2025 were retrospectively enrolled. Patients were randomly divided into a training set (n = 814) and a test set (n = 349) at a 7:3 ratio. Five multidimensional inflammatory indices were calculated: platelet-to-lymphocyte ratio (PLR), monocyte-to-lymphocyte ratio (MLR), monocyte-to-high-density lipoprotein ratio (MHR), systemic immune-inflammation index (SII), and systemic inflammation response index (SIRI). LASSO regression was used for variable selection, and multivariable logistic regression was performed to identify independent risk factors and construct a nomogram. Restricted cubic spline (RCS) analysis was employed to explore nonlinear relationships. The Synthetic Minority Over-sampling Technique (SMOTE) was applied exclusively to the training set to balance class distribution. Based on the identified independent predictors, nine machine learning models (XGBoost, Random Forest, LightGBM, MLP, SVM, AdaBoost, Decision Tree, Gaussian Naive Bayes, and Logistic Regression) were constructed. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), calibration curves, and decision curve analysis (DCA). The SHAP method was used for model interpretability. Robustness was verified through hard endpoint sensitivity analysis and subgroup analyses.

Results

The training and test sets were well-balanced in baseline characteristics. The incidence of MACE was 17.2% (140/814). Univariate analysis showed that PLR, SIRI, SII, and MLR were significantly elevated in the MACE group ($P < 0.05$), while MHR showed no significant difference between groups. LASSO regression ($\lambda_1se = 0.02$) selected 11 candidate variables, with PLR being the only retained

multidimensional inflammatory index. Multivariable logistic regression revealed that age (OR = 1.076, 95% CI: 1.054–1.101, P < 0.001), Gensini score (OR = 1.014, 95% CI: 1.009–1.020, P < 0.001), PLR (OR = 1.010, 95% CI: 1.006–1.014, P < 0.001), and AST (OR = 1.004, 95% CI: 1.002–1.005, P < 0.001) were independent risk factors, whereas EF (OR = 0.965, 95% CI: 0.946–0.983, P < 0.001) and statin use (OR = 0.165, 95% CI: 0.057–0.473, P = 0.001) were independent protective factors. A significant nonlinear relationship was identified between EF and MACE risk (P for non-linearity < 0.001), with risk rising sharply when EF fell below approximately 50%. The training set AUC was 0.842, and the Hosmer-Lemeshow test yielded P = 0.615.

After SMOTE resampling (n = 1,348), comparison of nine machine learning models on the independent test set showed that XGBoost performed best (AUC = 0.937, 95% CI: 0.914–0.959, F1 = 0.859, Brier Score = 0.100), followed by MLP (AUC = 0.906) and Random Forest (AUC = 0.903). DeLong test demonstrated that XGBoost significantly outperformed traditional logistic regression (Z = 5.451, P < 0.001). In the independent test set, XGBoost achieved optimal performance (AUC = 0.937, 95% CI: 0.914–0.959, F1 = 0.859, Brier Score = 0.100), significantly superior to traditional logistic regression (AUC = 0.846, DeLong test: Z = 5.451, P < 0.001), with ΔAUC = 0.091. SHAP analysis revealed that age (39.2%), Gensini score (21.9%), and AST (19.8%) were the most important predictive features, with PLR contributing approximately 10.3%. In hard endpoint sensitivity analysis, XGBoost maintained optimal performance (AUC = 0.965). All five subgroup analyses yielded interaction P-values > 0.05.

Conclusion

Among the five multidimensional inflammatory indices evaluated, PLR was the only one that retained independent predictive value after LASSO regularization and multivariable adjustment, highlighting its potential as a core inflammatory driver within the prediction framework. EF exhibits a nonlinear threshold effect with MACE risk at approximately 50%. The XGBoost model significantly outperforms traditional logistic regression in predictive performance, and when combined with SHAP analysis, enables individualized risk visualization. This PLR-driven machine learning model can serve as a precise and robust quantitative decision-making tool for early risk stratification and individualized intervention in AMI patients after PCI.

1. Introduction

Acute myocardial infarction (AMI) is one of the leading causes of mortality and disability worldwide [1]. Percutaneous coronary intervention (PCI) is currently the preferred revascularization strategy for AMI patients, significantly reducing mortality [2]. However, even after successful PCI, a considerable proportion of patients still experience major adverse cardiovascular events (MACE) during hospitalization, including death, acute heart failure, malignant arrhythmias, myocardial reinfarction, and cardiogenic shock [3]. Therefore, precise early assessment and stratification of MACE risk following PCI in AMI patients is a critical step in optimizing clinical decision-making and improving both short- and long-term prognosis.

Inflammatory responses play a central role in the initiation of atherosclerosis, plaque rupture, and myocardial ischemia-reperfusion injury [4]. In recent years, composite inflammatory indices derived from routine blood cell counts have become a research hotspot in cardiovascular prognostication due to their practicality, low cost, and high reproducibility [5]. Indices such as the platelet-to-lymphocyte ratio (PLR), monocyte-to-lymphocyte ratio (MLR), monocyte-to-high-density lipoprotein ratio (MHR), systemic immune-inflammation index (SII), and systemic inflammation response index (SIRI) have been demonstrated to be associated with adverse outcomes in various cardiovascular diseases [6–9]. However, existing studies have largely focused on single-index analyses, and head-to-head studies systematically integrating and comparing the predictive performance of these five indices within the same population are lacking. Furthermore, within a system of multidimensional inflammatory indices, which specific marker possesses the optimal core driving efficacy in a machine learning framework remains unknown.

In the field of clinical prediction modeling, traditional logistic regression assumes a linear relationship between variables and outcomes. However, the relationship between many continuous clinical variables and MACE risk may be more complex, involving nonlinear associations or threshold effects [10]. Restricted cubic spline (RCS) analysis can effectively identify such nonlinear associations and potential inflection points, providing reliable methodological tools for understanding the dose-response relationships between predictors and prognosis.

Regarding modeling strategies, machine learning algorithms, with their powerful nonlinear fitting capabilities and ability to identify feature interactions, have demonstrated unique advantages in cardiovascular disease risk prediction [11, 12]. However, prediction model studies that systematically compare multidimensional inflammatory indices with traditional logistic regression and multiple machine learning models are currently lacking. Moreover, whether nonlinear relationships and definitive clinical thresholds exist between inflammatory indices and MACE risk has not been adequately validated.

Based on the above research gaps, this study set the following objectives: (1) to systematically integrate and compare the independent predictive performance of five multidimensional inflammatory indices (PLR, MLR, MHR, SII, and SIRI) and identify the inflammatory marker with optimal driving capacity within a machine learning framework; (2) to use RCS to dissect the nonlinear relationships and potential clinical thresholds between key continuous variables (age, left ventricular ejection fraction [LVEF], PLR, and Gensini score) and MACE risk; and (3) based on the selected core predictive variables, to construct traditional logistic regression and nine machine learning models (XGBoost, Random Forest, LightGBM, Multilayer Perceptron, Support Vector Machine, AdaBoost, Decision Tree, Gaussian Naive Bayes, and Logistic Regression), and through comprehensive performance comparison, select the optimal predictive model, thereby providing a quantitative decision-support tool for early risk stratification and individualized intervention for in-hospital MACE in AMI patients after PCI.

2. Materials and Methods

2.1 Study Population

This study was a retrospective cohort study. Consecutive patients diagnosed with AMI who underwent PCI at the Second Affiliated Hospital of Shenyang Medical College from January 2020 to December 2025 were enrolled. For patients with multiple hospitalizations, only the initial PCI hospitalization record was included. This study was approved by the Ethics Committee of the Second Affiliated Hospital of Shenyang Medical College (Approval No.: 2025-SYEY-LL-087_). All patient data were de-identified and the study complied with the ethical principles of the Declaration of Helsinki; the requirement for informed consent was waived.

Inclusion criteria were: (1) diagnosis of AMI according to the Fourth Universal Definition of Myocardial Infarction [1]; (2) confirmation by coronary angiography and treatment with PCI; (3) age ≥ 18 years; and (4) complete clinical data.

Exclusion criteria were: (1) severe valvular heart disease, cardiomyopathy, congenital heart disease, or cor pulmonale; (2) severe hepatic or renal dysfunction, hematological disorders, malignant tumors, or active bleeding.

2.2 Sample Size Estimation

Based on Kendall's sample size estimation principle [13], the sample size should be 5 to 10 times the number of candidate predictor variables. A total of 51 candidate variables were included in this study. Considering 20% incomplete case data, a minimum of 612 patients was required. A total of 1,163 patients were ultimately included, meeting the sample size requirement. In addition, calculated with the number of MACE events (n = 140) and the number of candidate variables (n = 51), the events per variable (EPV) was approximately 2.7:1. Although this is below the conventional EPV ≥ 10 rule of thumb for logistic regression, the subsequent LASSO regularization regression effectively addresses overfitting risk in high-dimensional data through shrinkage estimation [14]. Moreover, the final multivariable model retained only 10 variables (EPV = 14.0), meeting the criterion for model robustness.

2.3 Data Collection

The following data were extracted from the electronic medical record system: (1) sociodemographic data: sex, age, height, weight, BMI, smoking history, alcohol consumption history; (2) disease-related data: history of hypertension, diabetes mellitus, hyperlipidemia, family history of coronary heart disease, chronic kidney disease, atrial fibrillation, peripheral arterial disease, cerebrovascular disease; admission heart rate, systolic blood pressure, diastolic blood pressure; Killip class; type of myocardial infarction (STEMI/NSTEMI); post-PCI TIMI flow grade; number of stents implanted; multivessel disease; left ventricular ejection fraction (LVEF); Gensini score; time from symptom onset to admission; time from symptom onset to PCI; and major medication use (ACEI/ARB/ARNI, β-blockers, statins); (3) laboratory test indices: N-terminal pro-B-type natriuretic peptide (NT-proBNP), cardiac troponin T (cTnT), creatinine,

urea, ALT, AST, potassium, sodium, chloride, fasting blood glucose, total cholesterol, triglycerides, HDL-C, LDL-C, and complete blood count; (4) inflammation-related indices: PLR (platelet count / lymphocyte count), MLR (monocyte count / lymphocyte count), MHR (monocyte count / HDL-C), SII (platelet count × neutrophil count / lymphocyte count), and SIRI (neutrophil count × monocyte count / lymphocyte count) were calculated [15–18]. High-sensitivity C-reactive protein (hs-CRP) was also recorded.

2.4 Outcome Definition

The primary outcome was in-hospital MACE, a composite endpoint including: all-cause death, acute heart failure, malignant arrhythmias (defined as severe sinus bradycardia ≤ 40 beats/min, second-degree or higher atrioventricular block, ventricular tachycardia, ventricular fibrillation, cardiac arrest), myocardial reinfarction, and cardiogenic shock[19–23].The primary outcome was adjudicated according to the Fourth Universal Definition of Myocardial Infarction [1] and relevant clinical guidelines by two independent senior cardiologists; disagreements were resolved through consensus. Hard endpoint sensitivity analysis was conducted using "all-cause death or cardiogenic shock" as the alternative outcome.

2.5 Statistical Analysis

SPSS 27.0 and R 4.3.3 were used for statistical analysis. Continuous variables were first tested for normality (Shapiro-Wilk test). Non-normally distributed variables are presented as median (interquartile range) [M (IQR)], and between-group comparisons were performed using the Mann-Whitney U test. Categorical variables are presented as frequencies (percentages) [n (%)], and comparisons were performed using the χ^2 test or Fisher's exact test (when the theoretical frequency was < 5). All tests were two-sided, and $P < 0.05$ was considered statistically significant.

2.5.1 Variable Selection and Risk Factor Analysis

The entire dataset was randomly divided into a training set (n = 814) and a test set (n = 349) at a 7:3 ratio. Stratified randomization was performed based on MACE status to ensure a comparable proportion of outcome events between the two sets. In the training set, univariate analysis was conducted with in-hospital MACE as the grouping variable. Variables with $P < 0.05$ in univariate analysis were entered into LASSO regression [24]. A 10-fold cross-validation was used to determine the penalty coefficient λ, and variables with non-zero coefficients corresponding to λ_1se were selected. Collinearity analysis was performed on the LASSO-selected variables, and variables with a variance inflation factor (VIF) > 10 were removed. If a variable exhibited an effect direction in multivariable regression contrary to univariate analysis and clinical knowledge, it was excluded based on clinical plausibility. The final variables were entered into multivariable logistic regression [25], with results expressed as odds ratios (OR) and 95% confidence intervals (CI). A nomogram was constructed based on the multivariable regression results, and model calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test and calibration curves [26].

2.5.2 Nonlinear Relationship and Threshold Effect Analysis

For the statistically significant key continuous variables in the final multivariable model (PLR, age, EF, Gensini score), restricted cubic spline (RCS) [27] models were used to fit the dose-response relationships with MACE risk. Based on the Akaike Information Criterion (AIC), four knots were set (at the 25th, 50th, 75th, and 95th percentiles), adjusting for other covariates. Both the overall association P-value and the nonlinear P-value (P for non-linearity) were examined.

2.5.3 Machine Learning Model Construction and Comparison

To avoid data leakage, all data preprocessing and oversampling operations were strictly confined to the training set; the independent test set (n = 349) did not participate in learning any preprocessing parameters. The specific workflow was as follows: First, the Synthetic Minority Over-sampling Technique (SMOTE) [28] was applied to the original training set (n = 814) to balance the class distribution, achieving a 1:1 ratio of MACE-positive to MACE-negative samples (a total of 1,348 samples after balancing, with 674 each for MACE-positive and MACE-negative). Subsequently, this balanced training set was further divided, using a fixed random seed, into an internal training set (n = 943, for model training and hyperparameter tuning) and an internal validation set (n = 405, for early stopping and performance monitoring during tuning). The independent original test set (n = 349) was used only once for final performance evaluation after the models were finalized, ensuring an unbiased estimate of model generalizability.

Based on the six independent predictors ultimately confirmed by multivariable logistic regression (age, LVEF, Gensini score, PLR, AST, statin use), nine models—XGBoost, Random Forest, LightGBM, Decision Tree, Multilayer Perceptron (MLP), Support Vector Machine (SVM), Gaussian Naive Bayes, AdaBoost, and Logistic Regression—were constructed using 10-fold cross-validation combined with grid search for hyperparameter tuning on the SMOTE-balanced internal training set (n = 943) [29–35].

On the independent test set (n = 349), model discrimination and calibration were assessed using the area under the ROC curve (AUC), accuracy, sensitivity, specificity, F1 score, and Brier Score. The DeLong test was used to compare AUC differences between models. Calibration curves and decision curve analysis (DCA) [36] were plotted to evaluate model calibration and net clinical benefit. The SHAP method [37] was used for interpretability analysis of the optimal model, quantifying the contribution of each feature to the model's predictive output.

2.5.4 Sensitivity Analysis

(1) Outcome definition sensitivity: Using "all-cause death or cardiogenic shock" as the hard endpoint, nine machine learning models were reconstructed following the same workflow as the primary model. (2) Subgroup analyses: With PLR as a continuous exposure variable, logistic regression was used to evaluate the association between PLR and MACE risk in different subgroups (age stratified as < 65 / ≥ 65 years, STEMI / NSTEMI, male / female, with / without diabetes mellitus, Killip class I-II / III-IV). The

interaction term "subgroup × PLR" was incorporated into logistic regression models, and the P-value for interaction (P for interaction) was calculated using the likelihood ratio test.

3. Results

3.1 Baseline Characteristics

A total of 1,163 AMI patients who underwent PCI and met the eligibility criteria were enrolled. Among the 814 patients in the training set, 140 (17.2%) experienced in-hospital MACE, and 674 (82.8%) did not. Baseline characteristics of the training and test sets are compared in Table 1. Except for ACEI/ARB/ARNI use (P = 0.007) and cTnT (P = 0.035), there were no statistically significant differences between the training and test sets for all other variables (P > 0.05), indicating good balance achieved by randomization.

Table 1
Comparison of Baseline Characteristics Between the Training and Test Sets

Variable	Training Set(n = 814)	Test Set(n = 349)	t/χ^2/Z	P
Sex, N (%)			2.795	0.095
Male	604(74.201)	275(78.797)		
Female	210(25.799)	74(21.203)		
Smoking history, N (%)			0.47	0.493
No	433(53.194)	178(51.003)		
Yes	381(46.806)	171(48.997)		
Alcohol history, N (%)			0	0.992
No	567(69.656)	243(69.628)		
Yes	247(30.344)	106(30.372)		
Hypertension, N (%)			0.54	0.462
No	331(40.663)	150(42.980)		
Yes	483(59.337)	199(57.020)		
Diabetes mellitus, N (%)			1.443	0.23
No	523(64.251)	237(67.908)		
Yes	291(35.749)	112(32.092)		
Family history of CHD, N (%)			0.505	0.477
No	786(96.560)	334(95.702)		
Yes	28(3.440)	15(4.298)		
Atrial fibrillation, N (%)			1.416	0.234
No	758(93.120)	318(91.117)		
Yes	56(6.880)	31(8.883)		

Note: ACEI/ARB/ARNI, angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker/angiotensin receptor-neprilysin inhibitor; EF, left ventricular ejection fraction; PLR, platelet-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; MHR, monocyte-to-high-density lipoprotein ratio; SII, systemic immune-inflammation index; SIRI, systemic inflammation response index. For complete baseline characteristics, see Supplementary Table S1.

Variable	Training Set(n = 814)	Test Set(n = 349)	t/χ^2/Z	P
Peripheral artery disease, N (%)			1.579	0.209
No	781(95.946)	329(94.269)		
Yes	33(4.054)	20(5.731)		
CKD history, N (%)			1.351	0.245
No	779(95.700)	339(97.135)		
Yes	35(4.300)	10(2.865)		
Cerebrovascular disease, N (%)			0.003	0.959
No	696(85.504)	298(85.387)		
Yes	118(14.496)	51(14.613)		
MI type, N (%)			1.03	0.31
NSTEMI	237(29.115)	112(32.092)		
STEMI	577(70.885)	237(67.908)		
KILLIP class, N (%)			2.194	0.533
I	375(46.069)	171(48.997)		
II	357(43.857)	150(42.980)		
III	28(3.440)	12(3.438)		
IV	54(6.634)	16(4.585)		
Post-PCI TIMI flow, N (%)			1.238	0.744
0	1(0.123)	0(0.000)		
1	1(0.123)	1(0.287)		
2	1(0.123)	0(0.000)		
3	811(99.631)	348(99.713)		

Note: ACEI/ARB/ARNI, angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker/angiotensin receptor-neprilysin inhibitor; EF, left ventricular ejection fraction; PLR, platelet-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; MHR, monocyte-to-high-density lipoprotein ratio; SII, systemic immune-inflammation index; SIRI, systemic inflammation response index. For complete baseline characteristics, see Supplementary Table S1.

Variable	Training Set(n = 814)	Test Set(n = 349)	t/χ^2/Z	P
ACEI/ARB/ARNI, N (%)			7.275	0.007
No	366(44.963)	187(53.582)		
Yes	448(55.037)	162(46.418)		
β-blockers, N (%)			0.637	0.425
No	288(35.381)	115(32.951)		
Yes	526(64.619)	234(67.049)		
Statins, N (%)			0.837	0.36
No	24(2.948)	7(2.006)		
Yes	790(97.052)	342(97.994)		
PLR, median [IQR]	135.115[104.188,184.921]	139.423[107.328,185.106]	-0.566	0.571
MLR, median [IQR]	0.345[0.254,0.489]	0.359[0.259,0.514]	-0.444	0.657
MHR, median [IQR]	0.526[0.381,0.740]	0.546[0.364,0.728]	-0.095	0.924
SII, median [IQR]	871.547[577.711,1381.696]	850.555[547.634,1360.257]	0.064	0.949
SIRI, median [IQR]	0.013[0.008,0.026]	0.012[0.007,0.028]	0.762	0.446
Age, median [IQR]	65.000[56.000,71.100]	64.658[57.000,71.000]	-0.203	0.839
BMI, median [IQR]	25.200[23.000,27.682]	25.000[23.100,27.513]	0.277	0.781
EF, median [IQR]	57.000[50.000,60.000]	57.000[50.000,61.000]	-0.681	0.496
Gensini score, median [IQR]	55.000[30.790,80.000]	56.000[32.000,88.000]	-1.771	0.076
Stents implanted, median [IQR]	1.000[1.000,2.000]	1.000[1.000,2.000]	-1.131	0.222
Multivessel disease, median	1.000[1.000,1.000]	1.000[1.000,1.000]	-0.146	0.837

Note: ACEI/ARB/ARNI, angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker/angiotensin receptor-neprilysin inhibitor; EF, left ventricular ejection fraction; PLR, platelet-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; MHR, monocyte-to-high-density lipoprotein ratio; SII, systemic immune-inflammation index; SIRI, systemic inflammation response index. For complete baseline characteristics, see Supplementary Table S1.

Variable	Training Set(n = 814)	Test Set(n = 349)	t/χ^2/Z	P
[IQR]				
Onset-to-admission, median [IQR]	205.000[90.000,1020.000]	189.000[95.000,880.000]	0.082	0.935
Onset-to-PCI, median [IQR]	365.000[180.000,2405.000]	460.000[193.000,2490.000]	-1.202	0.229
Heart rate, median [IQR]	75.000[66.000,88.000]	73.000[65.000,83.000]	1.873	0.061
SBP, median [IQR]	134.000[120.000,150.000]	135.000[120.000,154.000]	-0.199	0.842
DBP, median [IQR]	80.000[70.000,90.000]	80.000[71.000,90.000]	-0.056	0.955
WBC, median [IQR]	8.510[7.000,11.240]	8.410[6.550,10.990]	1.39	0.165
Neutrophils, median [IQR]	6.520[4.870,8.830]	6.190[4.670,8.890]	0.69	0.49
Monocytes, median [IQR]	0.540[0.430,0.720]	0.550[0.430,0.750]	-0.345	0.73
Lymphocytes, median [IQR]	1.620[1.200,2.120]	1.580[1.180,1.960]	1.48	0.139
Platelets, median [IQR]	215.000[178.000,252.000]	218.000[174.000,249.000]	0.225	0.822
cTnT, median [IQR]	2.020[0.280,6.500]	1.300[0.172,4.799]	2.103	0.035
NT-proBNP, median [IQR]	1338.000[386.000,3300.000]	1157.600[276.000,2874.914]	1.475	0.14
TC, median [IQR]	4.480[3.750,5.200]	4.430[3.730,5.240]	0.433	0.665
TG, median [IQR]	1.520[1.090,2.170]	1.450[1.020,2.190]	0.882	0.378
HDL-C, median [IQR]	1.050[0.892,1.240]	1.070[0.900,1.260]	-0.69	0.49
LDL-C, median [IQR]	2.840[2.240,3.430]	2.770[2.120,3.480]	0.886	0.376

Note: ACEI/ARB/ARNI, angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker/angiotensin receptor-neprilysin inhibitor; EF, left ventricular ejection fraction; PLR, platelet-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; MHR, monocyte-to-high-density lipoprotein ratio; SII, systemic immune-inflammation index; SIRI, systemic inflammation response index. For complete baseline characteristics, see Supplementary Table S1.

Variable	Training Set(n = 814)	Test Set(n = 349)	t/χ^2/Z	P
AST, median [IQR]	35.700[20.900,116.300]	39.700[22.000,126.800]	-1.518	0.129
ALT, median [IQR]	29.400[17.320,54.000]	28.800[17.600,49.400]	0.958	0.338
Creatinine, median [IQR]	73.000[62.000,88.000]	73.000[63.000,87.000]	-0.499	0.618
Urea, median [IQR]	5.310[4.510,7.010]	5.700[4.700,6.940]	-1.05	0.294
K, median [IQR]	4.050[3.810,4.300]	4.050[3.800,4.320]	-0.293	0.769
Na, median [IQR]	140.500[138.600,142.600]	141.000[139.100,142.700]	-1.894	0.058
Cl, median [IQR]	102.700[100.800,104.500]	102.900[100.800,104.900]	-1.041	0.298
Fasting glucose, median [IQR]	6.340[5.430,8.410]	6.340[5.480,8.410]	0.084	0.933
hs-CRP, median [IQR]	15.000[3.500,46.516]	12.500[2.310,45.046]	1.078	0.281

Note: ACEI/ARB/ARNI, angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker/angiotensin receptor-neprilysin inhibitor; EF, left ventricular ejection fraction; PLR, platelet-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; MHR, monocyte-to-high-density lipoprotein ratio; SII, systemic immune-inflammation index; SIRI, systemic inflammation response index. For complete baseline characteristics, see Supplementary Table S1.

3.2 Univariate Analysis

Univariate analysis results for the 814 patients in the training set, grouped by MACE status, are shown in Table 2. Compared to the non-MACE group, patients in the MACE group were older [72.2 (65.2, 79.0) vs. 63.0 (54.0, 70.0), P < 0.001], had lower LVEF [51.7% (41.6, 57.0) vs. 58.0% (52.0, 61.0), P < 0.001], and had higher Gensini scores [70.5 (42.5, 97.1) vs. 52.0 (27.0, 76.0), P < 0.001].

Regarding multidimensional inflammatory indices, PLR [153.3 (123.7, 199.3) vs. 130.0 (101.7, 182.8), P < 0.001], MLR [0.409 (0.284, 0.551) vs. 0.336 (0.250, 0.468), P = 0.003], SII [957.7 (630.4, 1570.6) vs. 839.3 (558.7, 1330.8), P = 0.026], and SIRI [0.021 (0.009, 0.037) vs. 0.012 (0.007, 0.023), P < 0.001] were all significantly elevated in the MACE group. MHR showed no statistically significant difference between groups [0.532 (0.401, 0.808) vs. 0.520 (0.370, 0.727), P = 0.271]. Additionally, the prevalence of atrial fibrillation, Killip class distribution, and the usage rates of β-blockers and statins all differed significantly between groups (P < 0.05).

Table 2
Univariate Analysis Results (Key Variables)

Variable	MACE group(n = 140)	Non-MACE group(n = 674)	t/χ²/Z	P
Sex, N (%)			0.16	0.69
Male	102(72.857)	502(74.481)		
Female	38(27.143)	172(25.519)		
Smoking history, N (%)			0.075	0.784
No	73(52.143)	360(53.412)		
Yes	67(47.857)	314(46.588)		
Alcohol history, N (%)			0.011	0.917
No	97(69.286)	470(69.733)		
Yes	43(30.714)	204(30.267)		
Hypertension, N (%)			0.337	0.561
No	60(42.857)	271(40.208)		
Yes	80(57.143)	403(59.792)		
Diabetes mellitus, N (%)			0.143	0.705
No	88(62.857)	435(64.540)		
Yes	52(37.143)	239(35.460)		
Family history of CHD, N (%)			0.45	0.503
No	137(97.857)	649(96.291)		
Yes	3(2.143)	25(3.709)		
Atrial fibrillation, N (%)			17.404	< 0.001
No	119(85.000)	639(94.807)		
Yes	21(15.000)	35(5.193)		
Peripheral artery disease, N (%)			1.588	0.208
No	137(97.857)	644(95.549)		

Variable	MACE group(n = 140)	Non-MACE group(n = 674)	t/χ²/Z	P
Yes	3(2.143)	30(4.451)		
CKD history, N (%)			1.862	0.172
No	131(93.571)	648(96.142)		
Yes	9(6.429)	26(3.858)		
Cerebrovascular disease, N (%)			2.758	0.097
No	126(90.000)	570(84.570)		
Yes	14(10.000)	104(15.430)		
MI type, N (%)			1.388	0.239
NSTEMI	35(25.000)	202(29.970)		
STEMI	105(75.000)	472(70.030)		
KILLIP class, N (%)			28.887	< 0.001
I	64(45.714)	311(46.142)		
II	47(33.571)	310(45.994)		
III	6(4.286)	22(3.264)		
IV	23(16.429)	31(4.599)		
Post-PCI TIMI flow, N (%)			0.625	0.891
0	0(0.000)	1(0.148)		
1	0(0.000)	1(0.148)		
2	0(0.000)	1(0.148)		
3	140(100.000)	671(99.555)		
Stents implanted, N (%)			3.844	0.698
0	27(19.286)	163(24.184)		
1	71(50.714)	330(48.961)		
2	32(22.857)	144(21.365)		
3	8(5.714)	32(4.748)		

Variable	MACE group(n = 140)	Non-MACE group(n = 674)	t/χ²/Z	P
4	2(1.429)	3(0.445)		
5	0(0.000)	1(0.148)		
6	0(0.000)	1(0.148)		
Multivessel disease, N (%)			0.601	0.438
No	27(19.286)	150(22.255)		
Yes	113(80.714)	524(77.745)		
ACEI/ARB/ARNI, N (%)			2.26	0.133
No	71(50.714)	295(43.769)		
Yes	69(49.286)	379(56.231)		
β-blockers, N (%)			4.961	0.026
No	61(43.571)	227(33.680)		
Yes	79(56.429)	447(66.320)		
Statins, N (%)			16.385	< 0.001
No	12(8.571)	12(1.780)		
Yes	128(91.429)	662(98.220)		
NLR, median [IQR]	4.393[3.035,6.752]	3.951[2.689,5.866]	-2.168	0.03
PLR, median [IQR]	153.338[123.659,199.334]	130.000[101.744,182.759]	-3.686	< 0.001
MLR, median [IQR]	0.409[0.284,0.551]	0.336[0.250,0.468]	-3.016	0.003
MHR, median [IQR]	0.532[0.401,0.808]	0.520[0.370,0.727]	-1.101	0.271
SII, median [IQR]	957.741[630.351,1570.601]	839.304[558.656,1330.790]	-2.221	0.026
SIRI, median [IQR]	0.021[0.009,0.037]	0.012[0.007,0.023]	-4.006	< 0.001
Age, median [IQR]	72.164[65.206,78.961]	63.000[54.000,70.000]	-8.859	< 0.001
BMI, median [IQR]	25.060[22.580,27.500]	25.200[23.030,27.700]	0.943	0.346

Variable	MACE group(n = 140)	Non-MACE group(n = 674)	t/χ²/Z	P
EF, median [IQR]	51.695[41.562,56.957]	58.000[52.000,61.000]	7.571	< 0.001
Gensini score, median [IQR]	70.542[42.541,97.080]	52.000[27.000,76.000]	-5.51	< 0.001
Onset-to-admission, median [IQR]	240.000[100.000,700.000]	197.000[90.000,1080.000]	0.081	0.935
Onset-to-PCI, median [IQR]	340.000[160.000,1580.000]	372.000[182.000,2840.000]	1.721	0.085
Heart rate, median [IQR]	72.000[62.000,85.000]	76.000[67.000,88.000]	2.199	0.028
SBP, median [IQR]	130.000[117.000,150.000]	135.000[120.000,150.000]	1.536	0.124
DBP, median [IQR]	80.000[70.000,86.000]	80.000[72.000,90.000]	2.761	0.006
WBC, median [IQR]	9.170[7.200,11.590]	8.420[6.910,11.020]	-1.854	0.064
Neutrophils, median [IQR]	6.750[4.550,8.760]	6.520[4.910,8.850]	0.29	0.772
Monocytes, median [IQR]	0.560[0.430,0.730]	0.540[0.430,0.720]	-0.395	0.693
Lymphocytes, median [IQR]	1.620[1.230,2.190]	1.620[1.200,2.100]	-0.25	0.803
Platelets, median [IQR]	210.000[172.000,243.000]	216.000[180.000,253.000]	2.117	0.034
cTnT, median [IQR]	4.095[0.355,10.000]	1.813[0.280,5.890]	-2.971	0.003
NT-proBNP, median [IQR]	2013.045[807.000,4445.148]	1130.800[353.000,3004.286]	-3.735	< 0.001
TC, median [IQR]	4.480[3.840,5.080]	4.480[3.720,5.240]	0.227	0.82
TG, median [IQR]	1.400[1.040,1.770]	1.540[1.090,2.250]	2.121	0.034
HDL-C, median [IQR]	1.020[0.880,1.230]	1.050[0.900,1.250]	1.046	0.296
LDL-C, median [IQR]	2.950[2.210,3.500]	2.830[2.240,3.410]	-0.423	0.672
AST, median [IQR]	45.300[21.000,199.300]	35.000[20.900,104.000]	-2.268	0.023

Variable	MACE group(n = 140)	Non-MACE group(n = 674)	t/χ²/Z	P
ALT, median [IQR]	27.300[16.500,55.000]	29.700[17.650,53.600]	0.222	0.824
Creatinine, median [IQR]	76.000[64.400,88.000]	72.100[60.800,87.000]	-1.934	0.053
Urea, median [IQR]	5.700[4.600,7.400]	5.300[4.500,6.930]	-1.519	0.129
K, median [IQR]	4.090[3.800,4.330]	4.050[3.810,4.300]	-0.265	0.791
Na, median [IQR]	140.400[138.500,142.700]	140.500[138.600,142.600]	-0.14	0.889
Cl, median [IQR]	103.000[101.500,104.700]	102.600[100.600,104.500]	-1.893	0.058
Fasting glucose, median [IQR]	6.210[5.500,8.620]	6.370[5.420,8.330]	-0.514	0.608
hs-CRP, median [IQR]	16.539[6.468,55.080]	14.619[3.100,45.452]	-1.685	0.092

3.3 LASSO Regression Variable Selection

Variables with P < 0.05 in univariate analysis were entered into LASSO regression. At λ_1se = 0.02, LASSO regression selected 11 variables with non-zero coefficients: age, EF, Gensini score, heart rate, PLR, NT-proBNP, AST, atrial fibrillation, statins, β-blockers, and NLR. Among the five multidimensional inflammatory indices, SII, SIRI, MLR, and MHR were all compressed to zero; PLR was the only retained multidimensional inflammatory index. Although NLR was retained, its effect direction was reversed in multivariable regression (see Section 3.4), and it was excluded based on clinical plausibility. The LASSO coefficient path plot is shown in Fig. 1.

3.4 Multivariable Logistic Regression Analysis

The LASSO-selected variables (after excluding NLR due to reversed direction) were entered into multivariable logistic regression; results are shown in Table 3.

Table 3
Multivariable Logistic Regression Results

Variable	β	SE	OR (95% CI)	Z-value	P-value
Age	0.073	0.011	1.076 (1.054–1.101)	6.592	< 0.001
Gensini score	0.014	0.003	1.014 (1.009–1.020)	4.862	< 0.001
PLR	0.01	0.002	1.010 (1.006–1.014)	4.729	< 0.001
AST	0.004	0.001	1.004 (1.002–1.005)	4.599	< 0.001
NT-proBNP	< 0.001	< 0.001	1.000 (1.000–1.000)	2.124	0.034
EF	−0.036	0.01	0.965 (0.946–0.983)	−3.698	< 0.001
Statins	−1.800	0.533	0.165 (0.057–0.473)	−3.379	0.001
β-blockers	−0.488	0.24	0.614 (0.383–0.982)	−2.035	0.042
Heart rate	−0.011	0.007	0.989 (0.975–1.002)	−1.591	0.112
Atrial fibrillation	0.528	0.377	1.696 (0.800–3.531)	1.401	0.161

The results showed that age, Gensini score, PLR, AST, and NT-proBNP were independent risk factors for MACE, whereas LVEF, statin use, and β-blocker use were independent protective factors. Heart rate (P = 0.112) and atrial fibrillation (P = 0.161) were not statistically significant after multivariable adjustment. PLR was the only multidimensional inflammatory index that was retained and maintained the correct direction as an independent predictor.

Although NT-proBNP was statistically significant, its OR was approximately 1.000, indicating limited clinical effect size; β-blocker use is a treatment-related variable and carries a risk of treatment bias. Based on these considerations, six variables (age, EF, Gensini score, PLR, AST, statin use) were ultimately included for subsequent machine learning model construction. This traditional logistic regression model achieved an AUC of 0.842 in the training set, with a Hosmer-Lemeshow goodness-of-fit test P = 0.615 and Brier Score = 0.103, indicating good discrimination and acceptable calibration.

3.5 Nonlinear Relationship and Threshold Effect Analysis

RCS analysis was performed on four key continuous variables; results are shown in Table 4.

Table 4
RCS Analysis of Key Continuous Variables and Post-PCI MACE Risk

Variable	Overall P-value	Non-linearity P-value	Conclusion
LVEF	< 0.001	< 0.001	Significant non-linear
Age	< 0.001	0.094	Linear (marginal trend)
PLR	< 0.001	0.267	Linear
Gensini score	< 0.001	0.207	Linear

A significant nonlinear relationship was observed between LVEF and MACE risk (P for non-linearity < 0.001). The RCS curve demonstrated that when LVEF fell below approximately 50%, MACE risk rose sharply with decreasing EF; when LVEF exceeded 50%, the risk curve tended to flatten, suggesting that an EF of approximately 50% represents an important threshold point for MACE risk(Fig. 2).

3.6 Machine Learning Model Performance Comparison

Based on the SMOTE-balanced internal training set (n = 943) and using the six independent predictors (age, LVEF, Gensini score, PLR, AST, statin use), the performance of nine machine learning models on the independent test set (n = 349) is compared in Table 5.

Table 5
Predictive Performance Comparison of Nine Machine Learning Models on the Independent Test Set

Model	AUC (95% CI)	F1 Score	Brier Score
XGBoost	0.937 (0.914–0.959)	0.859	0.100
MLP	0.906 (0.877–0.934)	0.838	0.124
Random Forest	0.903 (0.873–0.932)	0.833	0.131
SVM	0.884 (0.852–0.916)	0.817	0.137
AdaBoost	0.879 (0.845–0.912)	0.822	0.170
LightGBM	0.861 (0.825–0.898)	0.817	0.192
Decision Tree	0.856 (0.818–0.893)	0.796	0.147
Gaussian Naive Bayes	0.850 (0.812–0.887)	0.795	0.175
Logistic Regression	0.846 (0.809–0.883)	0.787	0.160

Note: AUC, area under the receiver operating characteristic curve; CI, confidence interval; Brier Score reflects calibration on the independent test set, with lower values indicating better calibration. XGBoost demonstrated the best overall performance among the nine models.

The ROC curves of the nine machine learning models on the independent test set are displayed in Fig. 3.XGBoost performed best on the independent test set, with an AUC of 0.937 (95% CI: 0.914–0.959), an F1 score of 0.859, and a Brier Score of 0.100. The DeLong test showed that XGBoost significantly outperformed logistic regression (Z = 5.451, P < 0.001), Random Forest (Z = 2.871, P = 0.012), and MLP (Z = 2.451, P = 0.023). Decision curve analysis (DCA) indicated that XGBoost provided net clinical benefit across a wide range of threshold probabilities, from 1% to 98% (Fig. 5).

The calibration curves for all models are shown in Fig. 4. XGBoost achieved a Brier Score of 0.100, indicating good calibration.

3.7 Model Interpretability Analysis

SHAP-based interpretability analysis was performed on the optimal XGBoost model. Feature importance ranking based on the mean absolute SHAP value revealed: age (39.2%), Gensini score (21.9%), and AST (19.8%) as the top three most important predictive features, collectively accounting for 80.9% of model decisions. PLR contributed approximately 10.3% and was the only multidimensional inflammatory index included in the final model, further validating its value as a core driving factor. Statin use contributed the least (< 0.8%). The SHAP feature importance bar plot and beeswarm plot are presented in Fig. 6.

3.8 Sensitivity Analysis

3.8.1 Hard Endpoint Sensitivity Analysis

Using "all-cause death or cardiogenic shock" as the hard endpoint, nine machine learning models were reconstructed following the identical workflow as the primary model (including SMOTE balancing, 10-fold cross-validation, and independent test set evaluation). Results are presented in Table 6.

Table 6
Performance Comparison of Models for Hard Endpoint Sensitivity Analysis

Model	AUC (95%CI)	Accuracy	Sensitivity	Specificity	F1 Score	Brier Score
XGBoost	0.965(0.939–0.991)	0.913	0.955	0.873	0.914	0.073
Random Forest	0.964(0.935–0.992)	0.913	0.94	0.887	0.913	0.098
MLP	0.947(0.908–0.986)	0.906	0.985	0.831	0.91	0.094
SVM	0.911(0.864–0.957)	0.862	0.761	0.958	0.843	0.119
LightGBM	0.888(0.834–0.943)	0.826	0.746	0.901	0.806	0.18
AdaBoost	0.870(0.812–0.929)	0.79	0.776	0.803	0.782	0.178
Decision Tree	0.874(0.818–0.930)	0.804	0.746	0.859	0.787	0.131
Gaussian Naive Bayes	0.790(0.713–0.866)	0.754	0.851	0.662	0.77	0.196
Logistic Regression	0.768(0.687–0.849)	0.768	0.731	0.803	0.754	0.199

Note: Hard endpoint defined as all-cause death or cardiogenic shock. AUC, area under the receiver operating characteristic curve; CI, confidence interval.

In the hard endpoint analysis, XGBoost performed best (AUC = 0.965), with Random Forest closely following (AUC = 0.964); both significantly outperformed traditional logistic regression (AUC = 0.768, DeLong test $P < 0.001$ for both). Consistent with the primary model results, the predictive advantage of machine learning models, particularly ensemble learning methods, was even more pronounced for the hard endpoint, further confirming the robustness of the XGBoost model under different outcome definitions.

3.8.2 Subgroup Analysis

Using PLR as a continuous exposure variable, the association between PLR and MACE risk was assessed across five prespecified subgroups. The interaction P-values for age strata (< 65 / ≥65 years, P for interaction = 0.585), MI type (NSTEMI / STEMI, P for interaction = 0.730), sex (male / female, P for interaction = 0.290), diabetes mellitus (without / with, P for interaction = 0.279), and Killip class (I-II / III-IV, P for interaction = 0.224) were all non-significant (all $P > 0.05$), indicating that the predictive effect of PLR on MACE was robust and consistent across all subgroups.

4. Discussion

4.1 Principal Findings

This study systematically integrated five multidimensional inflammatory indices—PLR, MLR, MHR, SII, and SIRI—and constructed a risk prediction model for in-hospital MACE after PCI in AMI patients. The main findings are summarized as follows: (1) Among the five indices, PLR was the only one that maintained independent predictive performance after LASSO selection and multivariable adjustment (OR = 1.010, 95% CI: 1.006–1.014, P < 0.001), establishing its statistical status as the core driving indicator; (2) A significant J-shaped nonlinear relationship was found between LVEF and MACE risk (P for non-linearity < 0.001), with a threshold of approximately 50%, suggesting that the risk climbs exponentially when LVEF falls below this cut-point; (3) XGBoost performed best on the independent test set (AUC = 0.937), significantly outperforming traditional logistic regression (AUC = 0.846, ΔAUC = 0.091, DeLong test P < 0.001), confirming the unique advantage of ensemble learning methods in capturing complex interaction patterns among clinical features; (4) SHAP interpretability analysis further revealed that age (39.2%), Gensini score (21.9%), and AST (19.8%) were the top three features driving model predictions, with PLR, contributing approximately 10.3%, being the top predictor among inflammatory indices. These findings closely align with the original design objective of this study—to clarify the core driving efficacy of PLR within a machine learning framework.

4.2 Independent Predictive Value of PLR and Systematic Screening of Multidimensional Inflammatory Indices

One of the most critical findings of this study is that among the five multidimensional inflammatory indices, PLR was the only independent factor that maintained a consistent predictive direction across all three levels: univariate analysis, LASSO regularization-based selection, and multivariable logistic regression. Platelets not only participate in hemostasis and thrombosis but also secrete various pro-inflammatory cytokines, promoting endothelial cell activation and leukocyte recruitment [38]. Lymphopenia reflects the massive cortisol release mediated by stress-induced activation of the hypothalamic-pituitary-adrenal axis, which subsequently induces lymphocyte apoptosis and immunosuppression [39]. Therefore, an elevated PLR comprehensively reflects the dual pathophysiological effects of platelet activation and impaired lymphocyte function, serving as a reliable biomarker for adverse prognosis after AMI [40].

Notably, although SII, SIRI, and MLR all showed significant associations with MACE in univariate analysis in this study, their coefficients were reduced to zero after LASSO regularization compression, suggesting that the independent predictive increment of these composite indices is limited after adequately adjusting for strong confounders such as age, LVEF, and the extent of coronary artery disease (Gensini score). This result does not imply that these indices are irrelevant to AMI prognosis; rather, it indicates that the inflammatory information they carry is largely captured by PLR and traditional clinical variables. From a clinical utility perspective, PLR requires only two parameters from a complete blood count—

platelet and lymphocyte counts—making it simple to calculate, inexpensive, and particularly suitable for rapid application in primary care settings and emergency scenarios. Furthermore, all interaction P-values in the subgroup analyses exceeded 0.05, further confirming that the predictive effect of PLR remains robust and consistent across different ages, sexes, MI types, diabetes statuses, and Killip classes.

4.3 Nonlinear Relationship and Threshold Effect of LVEF: From Linear Assumption to Precision Stratification

Another important finding of this study is the significant J-shaped nonlinear relationship between LVEF and MACE risk (P for non-linearity < 0.001), with a threshold of approximately 50%. Traditional logistic regression assumes that each unit decrease in LVEF results in a constant proportional increase in MACE risk. However, the RCS curve clearly demonstrates that when LVEF is above 50%, the risk curve tends to be relatively flat; once LVEF falls below 50%, MACE risk climbs sharply with further decreases in EF. This nonlinear pattern carries important clinical implications—it suggests that among patients with preserved LVEF (> 50%), further improvement in EF may not yield a proportional reduction in MACE risk; conversely, for patients with already impaired EF (< 50%), even small improvements in EF could result in significant clinical benefit.

Clinically, this threshold effect suggests that AMI patients with post-PCI LVEF < 50% should be identified as a high-risk group for in-hospital MACE, warranting closer electrocardiographic and hemodynamic monitoring, earlier cardiac function support (such as vasoactive agents or short-term mechanical circulatory support), and more aggressive optimization and titration of neuroendocrine antagonists (such as ACEI/ARB/ARNI and β-blockers).

4.4 Advantages of Machine Learning Models and the Value of Interpretability

XGBoost performed best on the independent test set (AUC = 0.937, 95% CI: 0.914–0.959), significantly outperforming traditional logistic regression (AUC = 0.846, P < 0.001) and the other seven machine learning models. The Brier Score was 0.100, and the calibration curve indicated good consistency between predicted probabilities and observed event rates. In the hard endpoint sensitivity analysis, the predictive advantage of XGBoost was even more pronounced (AUC = 0.965 vs. 0.768 for logistic regression), suggesting that ensemble learning methods are particularly advantageous when dealing with more severe but sparser outcomes such as "all-cause death or cardiogenic shock." XGBoost, based on the gradient boosting framework, iteratively fits the residuals from the previous round, enabling it to automatically capture nonlinear dependencies and higher-order interactions among features [41]. Its built-in L1/L2 regularization mechanism effectively controls overfitting risk, offering superior generalization capability compared to traditional logistic regression given the sample size (n = 1,163) to feature count (n = 51) ratio in this study.

Another methodological highlight of this study is the deep integration of SHAP interpretability analysis with model construction. Traditional "black-box" machine learning models face barriers to clinical

adoption due to lack of interpretability [42]. SHAP analysis not only confirmed that age, Gensini score, and AST were the top three features driving model predictions—highly consistent with the OR ranking in multivariable logistic regression—but also further revealed the specific direction of impact for each feature's values (Fig. 6B). For instance, the SHAP beeswarm plot indicated that PLR's contribution is not simply linear; rather, it exhibits an accelerating upward trend in the high-value range (above approximately 150), corroborating the distribution characteristics of PLR in univariate analysis (median 153.3 in the MACE group vs. 130.0 in the non-MACE group). This dual-layer interpretative framework of "global importance ranking + local impact direction visualization" helps clinicians understand why the model assigns a high-risk prediction to a given patient, thereby enhancing trust in and willingness to adopt the model's outputs.

4.5 Comparison with Previous Studies

Previous studies on the prognostic value of PLR in AMI patients have largely focused on single-index analyses and traditional regression models [43].

In terms of machine learning applications, recent studies have attempted to apply algorithms such as XGBoost for risk prediction in AMI patients [44].

Compared with previous studies, the differentiating contributions of this study are: (1) a head-to-head comparison of the predictive performance of five multidimensional inflammatory indices within the same population, establishing PLR's optimal driving status; (2) the first systematic exploration of the nonlinear threshold effect of LVEF within an inflammatory marker prediction model; and (3) the integration of traditional logistic regression, nine machine learning algorithms, and SHAP interpretability analysis within a unified framework, realizing a full-chain risk assessment from "prediction" to "explanation."

4.6 Limitations

This study has the following limitations: (1) This was a single-center, retrospective study. Although the sample size (n = 1,163) met statistical requirements and rigorous inclusion/exclusion criteria and independent outcome adjudication were employed, the single-center data source may introduce selection bias and center-specific treatment patterns; external validity requires validation in multicenter prospective cohort studies. (2) The outcome was limited to in-hospital MACE, precluding assessment of long-term prognostic value. (3) Although this study strictly adhered to the principle of "independent test set reserved solely for final evaluation," the inherent absence of temporal sequence in retrospective data means the model cannot simulate the dynamic process of "prospective prediction" in a real clinical setting; future studies based on temporally validated datasets are needed to further evaluate the model's temporal stability. (4) Some laboratory indices (such as hs-CRP) were not included in the final model due to high missing data rates, and their predictive value awaits further evaluation. (5) This study analyzed PLR as a continuous variable and has not yet established an optimal cut-off value for convenient clinical application. Although RCS analysis suggested a linear relationship between PLR and MACE risk (P for

non-linearity = 0.267), determining a clinically actionable threshold still requires further external validation studies.(6) The use of SMOTE oversampling, although strictly confined to the training set, generates synthetic minority-class samples that may not fully capture the heterogeneity of real-world MACE cases; this could lead to an overestimation of model performance when applied to external populations.(7) Certain predictor variables, particularly LVEF and laboratory values, were measured at the time of admission rather than immediately before PCI, and may have been influenced by early therapeutic interventions, thereby limiting their interpretation as true pre-procedural baseline characteristics.(8) High-sensitivity C-reactive protein (hs-CRP) was excluded due to a high proportion of missing data. Given its established role in cardiovascular inflammation, its absence may leave residual confounding. Future studies should prioritize the complete collection of such biomarkers.(9) PLR was analyzed as a continuous variable in this study, and no optimal cut-off value for clinical risk categorization was derived. Determining a clinically actionable threshold remains a task for subsequent validation studies.

5. Conclusion

Through systematic integration of five multidimensional inflammatory indices and comparison of nine machine learning algorithms, this study demonstrates that, among the five inflammatory indices compared, PLR was the only one that independently predicted in-hospital MACE after adjustment (OR = 1.010, 95% CI: 1.006–1.014, P < 0.001). These findings support PLR as a potential core inflammatory driver in this prediction model, though external validation is needed to confirm its generalizability. On this basis, this study is the first to identify a J-shaped nonlinear relationship between LVEF and MACE risk (P for non-linearity < 0.001), with a threshold of approximately 50%, suggesting that AMI patients with EF below this cut-point should be classified as a high-risk group for in-hospital MACE. In terms of model construction, XGBoost performed optimally (independent test set AUC = 0.937, 95% CI: 0.914–0.959), significantly outperforming traditional logistic regression (AUC = 0.846, ΔAUC = 0.091, DeLong test P < 0.001); hard endpoint sensitivity analysis further confirmed its robustness (AUC = 0.965). SHAP interpretability analysis revealed age, Gensini score, and AST as the top three contributing features, with PLR being the leading predictor among inflammatory indices, contributing approximately 10.3%. Nevertheless, given the single-center retrospective design, these findings should be regarded as exploratory and require confirmation in multicenter prospective cohorts.In summary, the PLR-driven XGBoost prediction model, combined with SHAP analysis, can serve as a precise and intuitive quantitative tool for early risk stratification and individualized intervention for in-hospital MACE in AMI patients after PCI; however, its clinical promotion still requires validation in multicenter prospective studies.

Declarations

Data Availability Statement

The datasets generated and analyzed are available from the corresponding author upon reasonable request. Interested researchers may obtain access by contacting the corresponding author via email at:yikangxu@163.com.

Ethics Statement

This study was approved by the Ethics Committee of the Second Affiliated Hospital of Shenyang Medical College (Approval No.: 2025-SYEY-LL-087_). Given the retrospective nature of the research, which utilized exclusively de-identified medical records obtained from routine clinical care, the Ethics Committee granted a waiver of individual informed consent. This study was conducted in strict accordance with the ethical principles of the Declaration of Helsinki. All patient data were fully anonymized prior to analysis and stored in password-protected systems to ensure confidentiality throughout the research process.

Clinical trial number

not applicable.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this manuscript.

Author Contributions

Xiren Meng:Drafted the main manuscript text, Performed statistical analysis and Prepared all figures and tables.

Siyu Pan:Investigation, Data Curation, Manuscript Revision.

Shuochao Zhao and Yaxin Wang : Software, Statistical Analysis, Visualization.

Hao Li and Zhen Shan:Investigation, Resources, Clinical Data Collection.

Yueran Yu:Validation,Quality Control of Laboratory Assays

Yikang Xu(Corresponding Author):Conceived the study, supervised the research and critically revised the manuscript.

All authors have reviewed and approved the final version of the manuscript. Each author has participated sufficiently in the work to take public responsibility for appropriate portions of the content. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding

This study was supported by the Liaoning Provincial Department of Education Scientific Research Platform Construction Project (No. LJ232410164023) and Shenyang Science and Technology Program Project (No. 22-321-33-100).

References

1. Thygesen, K. et al. Executive Group on behalf of the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC)/American Heart Association (AHA)/World Heart Federation (WHF) Task Force for the Universal Definition of Myocardial Infarction. Fourth Universal Definition of Myocardial Infarction Glob Heart. 2018;13(4):305–338. (2018).
2. Qin, Q. et al. A comparison of long-term clinical outcomes between percutaneous coronary intervention (PCI) and medical therapy in patients with chronic total occlusion in noninfarct-related artery after PCI of acute myocardial infarction. *Clin. Cardiol.* **45** (1), 136–144 (2022).
3. Feng, Y. F. et al. Plasma tissue plasminogen activator-inhibitor complex levels in acute myocardial infarction patients: an observational study. *BMC Cardiovasc. Disord*. **24** (1), 722 (2024).
4. Imbesi, A. et al. Targeting Inflammation After Acute Myocardial Infarction. *J. Am. Coll. Cardiol.* **86** (15), 1146–1169 (2025).
5. Yang, Z. et al. The effect of inflammatory markers on mortality in patients with acute myocardial infarction. *Sci. Rep.* **15** (1), 14514 (2025).
6. Chen, Y. et al. Neutrophil-to-Lymphocyte Ratio and Platelet-to-Lymphocyte Ratio are Important Indicators for Predicting in-Hospital Death in Elderly AMI Patients. *J. Inflamm. Res.* **16**, 2051–2061 (2023).
7. Shen, T., Yang, X. & Zhang, Z. Positive Relationship of RDW with NT-proBNP and cTnI in Acute Myocardial Infarction Patients. *Clin. Lab.* ;**68**(6). (2022).
8. Sun, H. et al. Analysis of the Clinical Predictive Value of the Novel Inflammatory Indices SII, SIRI, MHR and NHR in Patients with Acute Myocardial Infarction and Their Extent of Coronary Artery Disease. *J. Inflamm. Res.* **17**, 7325–7338 (2024).
9. Hu, C. et al. The prognostic value of systemic inflammatory response index (SIRI) in acute coronary syndrome patients treated with primary percutaneous coronary intervention: a meta-analysis and systematic review. *Clin. Res. Cardiol.* Mar 16. (2026).
10. Lin, S. et al. Associations of metabolic indicators and inflammation-related indices with adverse cardiovascular events in US adults: NHANES 1999–2018. *Lipids Health Dis.* **25** (1), 57 (2026).
11. Huang, Q. et al. Characterisation of cardiovascular disease (CVD) incidence and machine learning risk prediction in middle-aged and elderly populations: data from the China health and retirement longitudinal study (CHARLS). *BMC Public. Health*. **25** (1), 518 (2025).
12. Gautam, N. et al. Machine Learning in Cardiovascular Risk Prediction and Precision Preventive Approaches. *Curr. Atheroscler Rep.* **25** (12), 1069–1081 (2023).

13. Riley, R. D. et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat. Med.* **38** (7), 1276–1296 (2019).
14. Rafati, S. et al. Comparison of Penalized Cox Regression Methods in Low-Dimensional Data with Few-Events: An Application to Dialysis Patients' Data. *J. Res. Health Sci.* **19** (3), e00452 (2019). PMID: 31586373; PMCID: PMC7183557.
15. Li, Q. et al. Prognostic Impact of Multiple Lymphocyte-Based Inflammatory Indices in Acute Coronary Syndrome Patients. *Front. Cardiovasc. Med.* **9**, 811790 (2022).
16. Villanueva, D. L. E. et al. Monocyte to High-Density Lipoprotein Ratio (MHR) as a predictor of mortality and Major Adverse Cardiovascular Events (MACE) among ST Elevation Myocardial Infarction (STEMI) patients undergoing primary percutaneous coronary intervention: a meta-analysis. *Lipids Health Dis.* **19** (1), 55 (2020).
17. Yang, Y. L. et al. Systemic immune-inflammation index (SII) predicted clinical outcome in patients with coronary artery disease. *Eur. J. Clin. Invest.* **50** (5), e13230 (2020).
18. Becirovic, E. et al. Integration of Hematologic and Metabolic Biomarkers for Outcome Prediction in Acute Coronary Syndromes Without ST Elevation. *Cureus* **17** (6), e86446 (2025).
19. Bosco, E. et al. Major adverse cardiovascular event definitions used in observational analysis of administrative databases: a systematic review. *BMC Med. Res. Methodol.* **21** (1), 241 (2021).
20. Kim, H. Y. et al. Timing of heart failure development and clinical outcomes in patients with acute myocardial infarction. *Front. Cardiovasc. Med.* **10**, 1193973 (2023).
21. Kaczorowski, D. J. et al. 2025 American Association for Thoracic Surgery (AATS) Expert Consensus Document: Surgical management of acute myocardial infarction and associated complications. *J. Thorac. Cardiovasc. Surg.* **170** (5), 1327–1344 (2025).
22. Du, H. et al. Combined prediction value of coronary plaque burden, serum creatinine, MLR and NLR in reinfarction risk after PCI in middle-aged and elderly patients with myocardial infarction. *Eur. J. Med. Res.* **30** (1), 522 (2025).
23. Berg, D. D., Bohula, E. A. & Morrow, D. A. Epidemiology and causes of cardiogenic shock. *Curr. Opin. Crit. Care*. **27** (4), 401–408 (2021).
24. Wu, H. et al. Comparison between traditional logistic regression and machine learning for predicting mortality in adult sepsis patients. *Front. Med. (Lausanne)*. **11**, 1496869 (2025).
25. Huang, B. et al. Development and validation of a nomogram based on Lasso-Logistic regression for predicting splenomegaly secondary to acute pancreatitis. *BMC Gastroenterol.* **24** (1), 281 (2024).
26. Cai, Z. et al. Development and Validation of a Nomogram for Predicting the Risk of Adverse Cardiovascular Events in Patients with Coronary Artery Ectasia. *J. Cardiovasc. Dev. Dis.* **8** (12), 186 (2021).
27. Ni, W. et al. Association between insulin resistance indices and outcomes in patients with heart failure with preserved ejection fraction. *Cardiovasc. Diabetol.* **24** (1), 32 (2025).

28. Jolfayi, A. G. et al. Applied machine learning to predict 1-year major adverse cardiovascular events in elderly patients after percutaneous coronary intervention. *BMC Med. Inf. Decis. Mak.* **25** (1), 401 (2025).
29. Sha, X. et al. Relationship Between Cardiometabolic Index and Post-PCI Coronary Microvascular Dysfunction in Acute STEMI Patients. *J. Multidiscip Healthc.* **18**, 5591–5602 (2025).
30. Van den Eynde, J. et al. Heart team 2.0: A decision tree for minimally invasive and hybrid myocardial revascularization. *Trends Cardiovasc. Med.* **31** (6), 382–391 (2021).
31. Wang, Y. et al. Sex-specific prediction of cardiogenic shock after acute coronary syndromes: the SEX-SHOCK score. *Eur. Heart J.* **45** (43), 4564–4578 (2024).
32. Song, X. et al. Identification of risk genes related to myocardial infarction and the construction of early SVM diagnostic model. *Int. J. Cardiol.* **328**, 182–190 (2021).
33. Zuo, D. et al. Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Med. Inf. Decis. Mak.* **23** (1), 276 (2023).
34. Si, Y. et al. Optimized feature selection and advanced machine learning for stroke risk prediction in revascularized coronary artery disease patients. *BMC Med. Inf. Decis. Mak.* **25** (1), 276 (2025).
35. Ning, Z. et al. Development and validation of a clinical prediction model for in-hospital heart failure risk following PCI in patients with coronary artery disease. *PLoS One*. **20** (6), e0325036 (2025).
36. Wu, Z. F. et al. PTH Predicts the in-Hospital MACE After Primary Percutaneous Coronary Intervention for Acute ST-Segment Elevation Myocardial Infarction. *Ther. Clin. Risk Manag*. **19**, 699–712 (2023).
37. Chen, M. et al. Application of Machine Learning Algorithms in Predicting Major Adverse Cardiovascular Events after Percutaneous Coronary Intervention in Patients with New-Onset ST-Segment Elevation Myocardial Infarction. *Rev. Cardiovasc. Med.* **26** (2), 25758 (2025).
38. Henein, M. Y. et al. The Role of Inflammation in Cardiovascular Disease. *Int. J. Mol. Sci.* **23** (21), 12906 (2022).
39. Shane, D. X. et al. Glucocorticoids impair T lymphopoiesis after myocardial infarction. *Am. J. Physiol. Heart Circ. Physiol.* **327** (2), H533–H544 (2024).
40. Yu, H. et al. Prognostic Value of Platelet to Lymphocyte Ratio for Myocardial Infarction: A Systematic Review and Meta-Analysis. *Clin. Cardiol.* **48** (10), e70215 (2025).
41. Tseng, P. Y. et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit. Care*. **24** (1), 478 (2020).
42. Qi, X. et al. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biol.* **79**, 103470 (2025).
43. Liu, J. et al. The correlation between PLR-NLR and prognosis in acute myocardial infarction. *Am. J. Transl Res.* **13** (5), 4892–4899 (2021). PMID: 34150072; PMCID: PMC8205844.
44. Khera, R. et al. Use of Machine Learning Models to Predict Death After Acute Myocardial Infarction. *JAMA Cardiol.* **6** (6), 633–641 (2021).

Figures

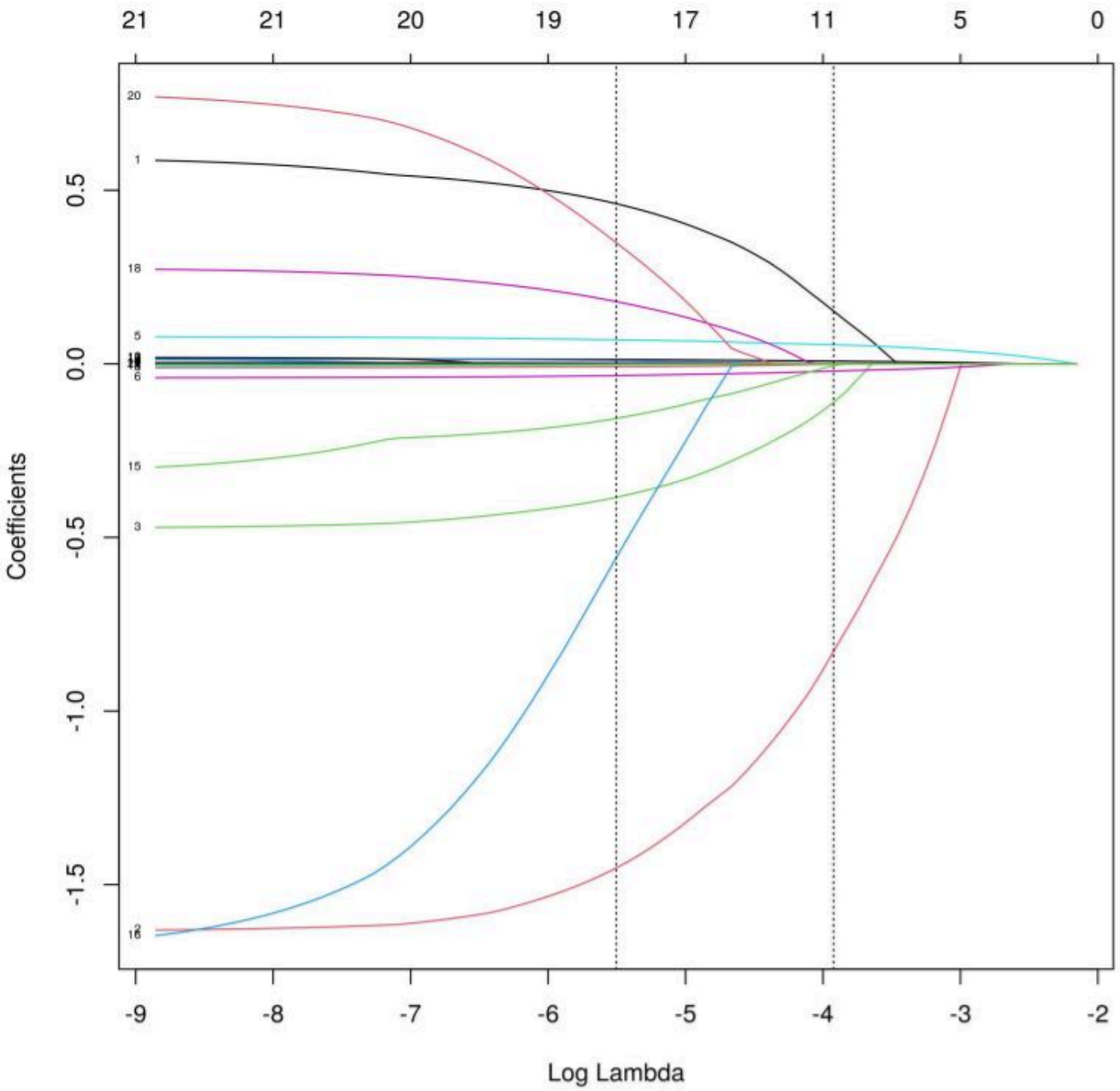


Figure 1

LASSO Coefficient Path Plot. Each curve represents the trajectory of a variable's coefficient as λ changes. The vertical dashed line corresponds to λ_1se = 0.02, at which point 11 variables with non-zero coefficients were selected.

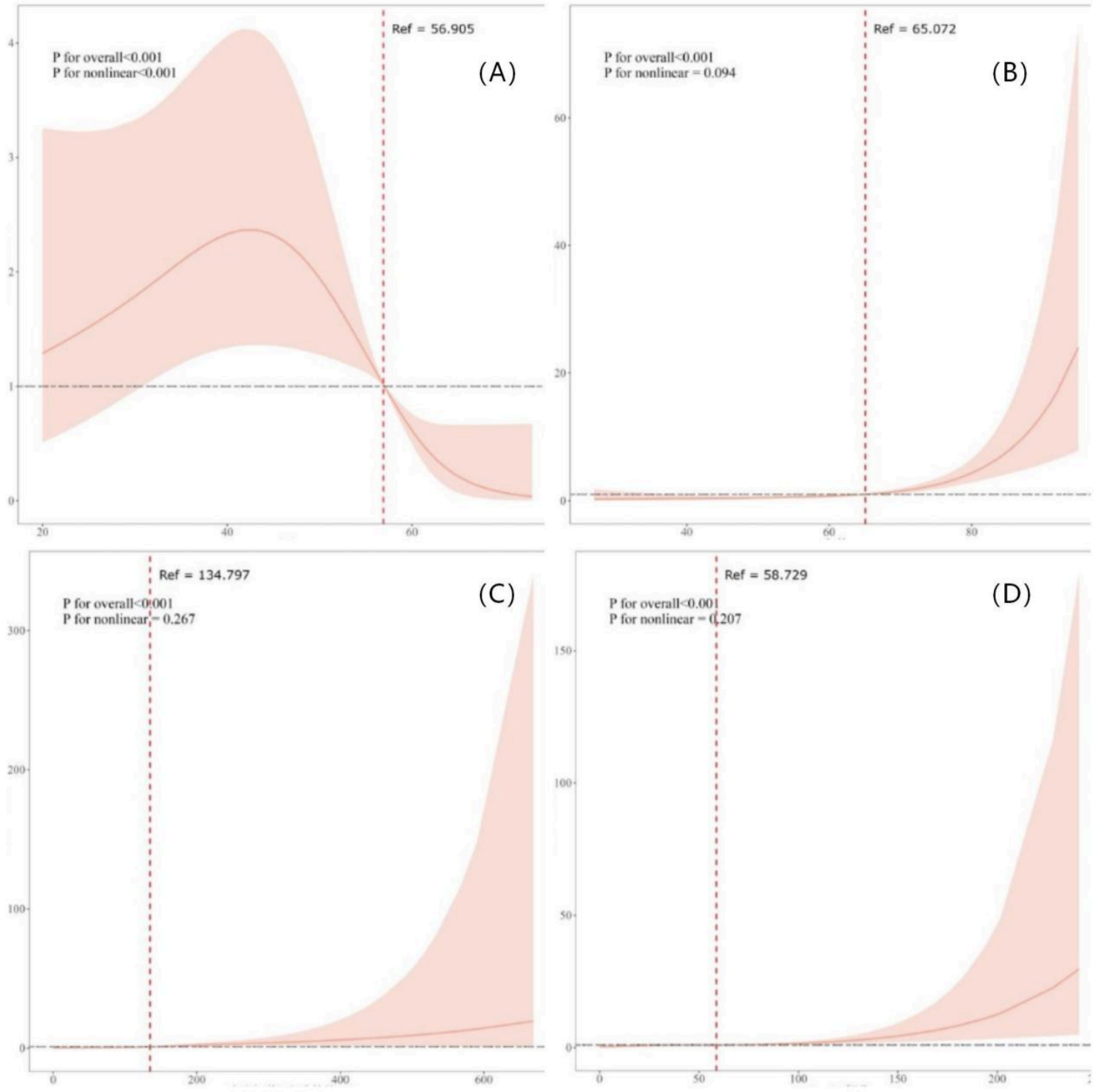


Figure 2

Restricted Cubic Spline Plots for Key Continuous Variables and Post-PCI MACE Risk. (A) LVEF; (B) Age; (C) PLR; (D) Gensini score. Each plot is adjusted for age, LVEF, Gensini score, PLR, and AST (except when the variable itself is the analysis variable). The red solid line represents the OR estimate, and the red shaded area represents the 95% CI. The non-linearity P-value for LVEF was <0.001; for other variables, non-linearity P > 0.05.

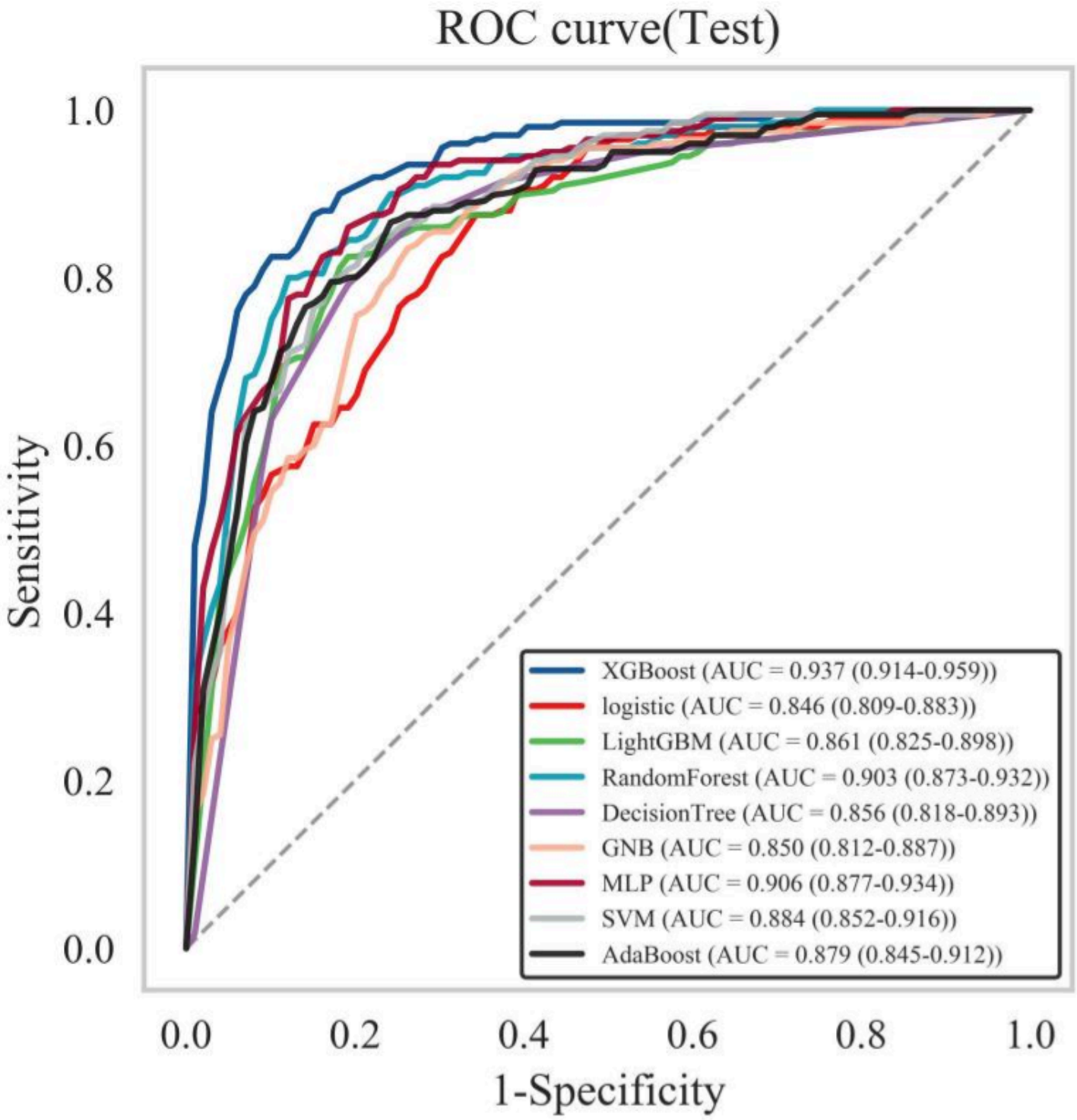


Figure 3

ROC Curves of Nine Machine Learning Models on the Independent Test Set.

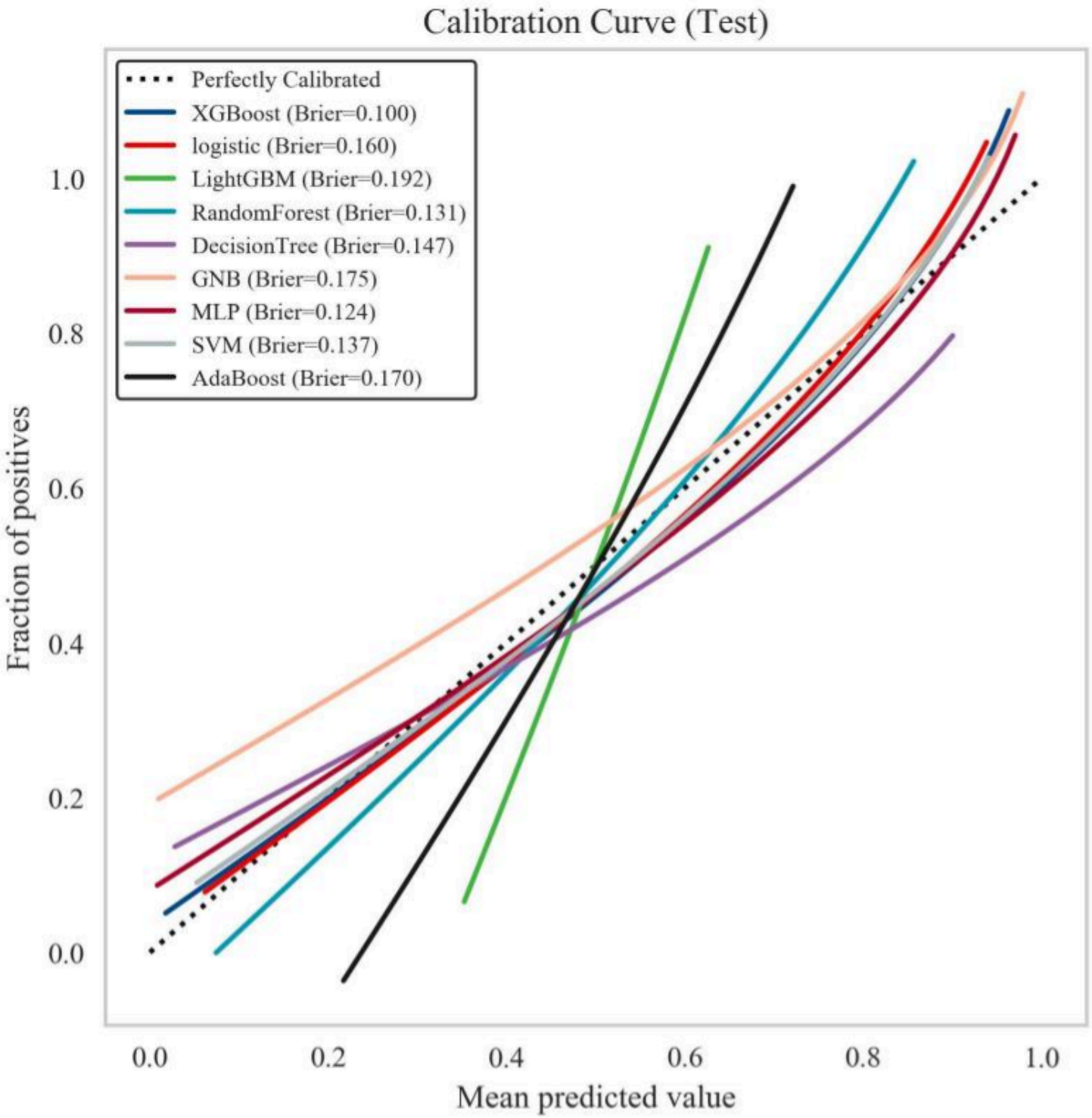


Figure 4

Calibration Curves for Each Model. XGBoost achieved a Brier Score of 0.073, indicating optimal calibration.

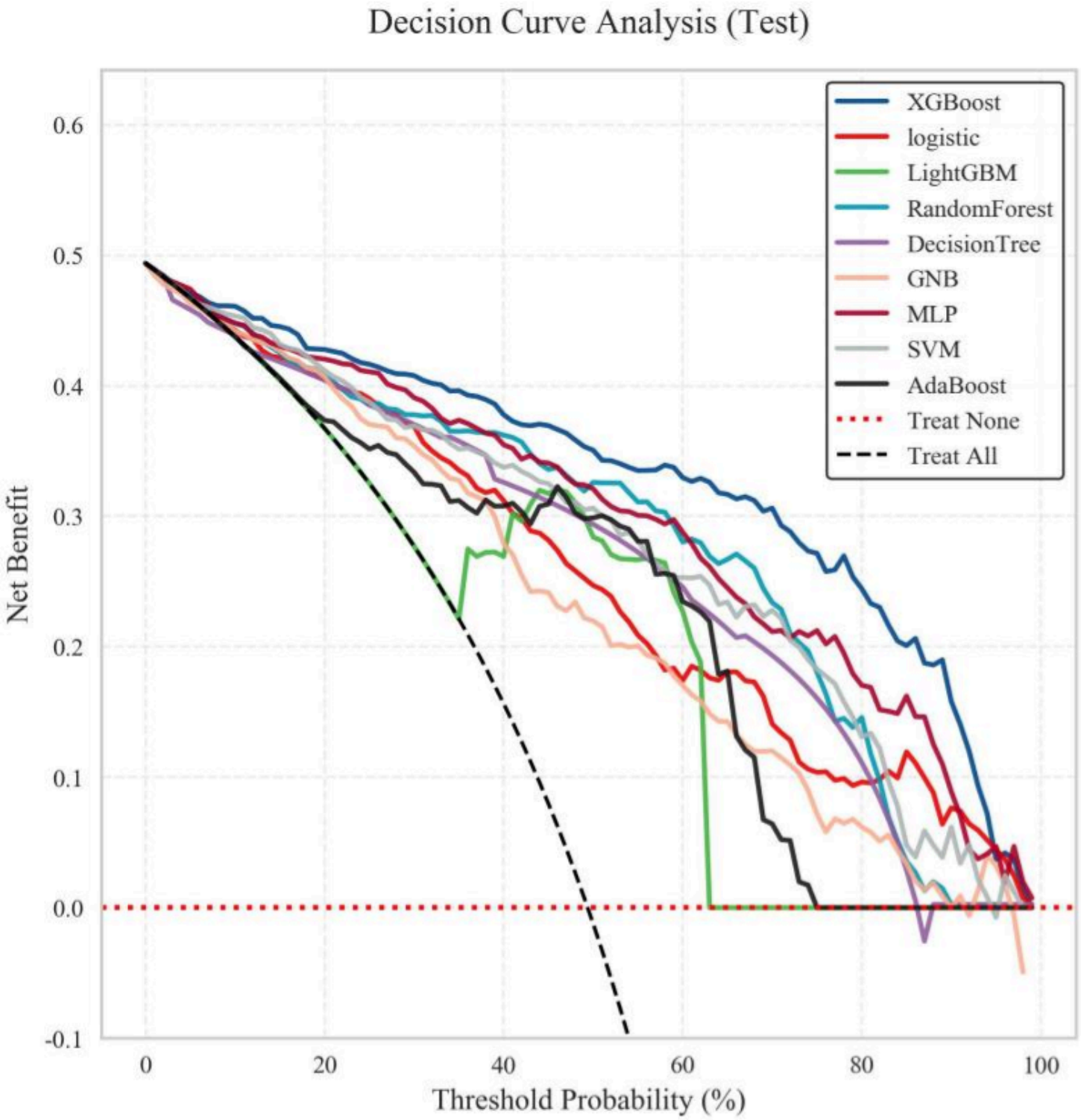


Figure 5

Decision Curve Analysis (DCA) for Each Model. XGBoost demonstrated net clinical benefit within the threshold probability range of 1%–98%.

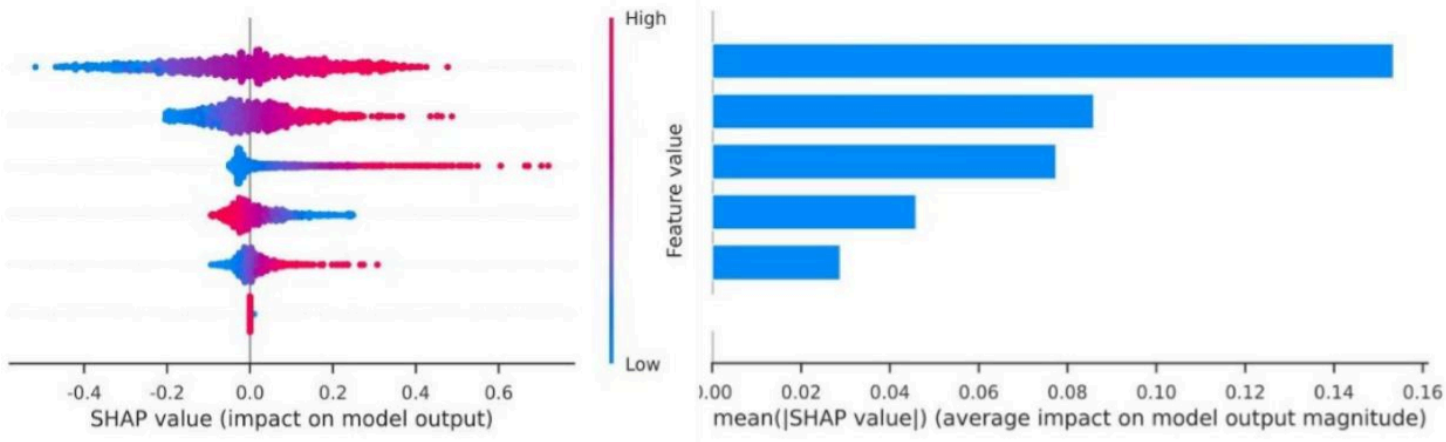


Figure 6

SHAP Feature Importance Plots. (A) Feature importance bar plot based on mean absolute SHAP values; (B) SHAP beeswarm plot illustrating the direction of each feature's impact on model predictions. Red indicates high feature values, blue indicates low feature values. Higher age, higher Gensini score, and higher PLR values correspond to larger SHAP values (increased MACE risk).

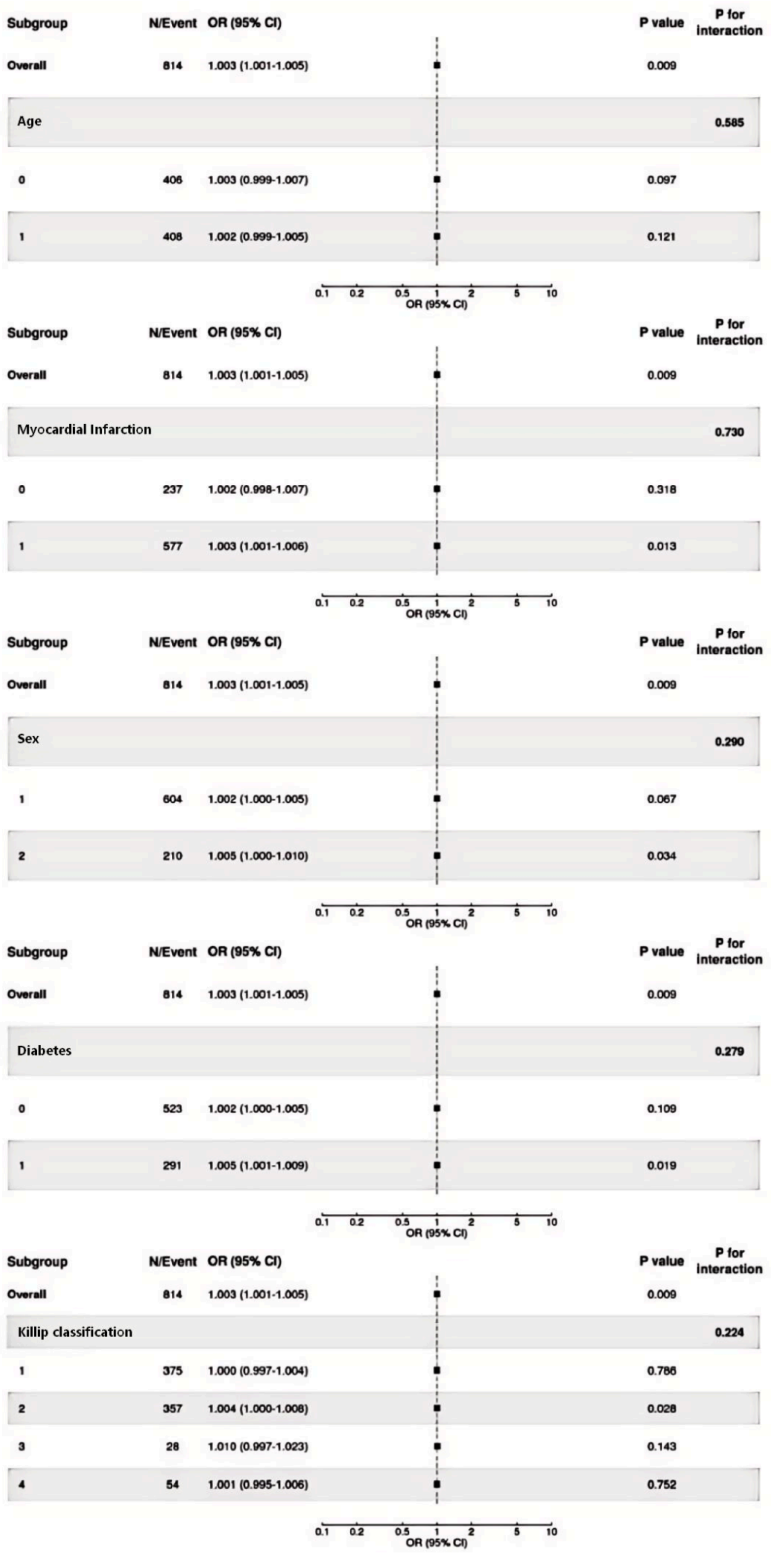


Figure 7

Forest Plot for Subgroup Analysis. Displays the association (OR and 95% CI) between PLR and MACE risk across different subgroups. All interaction P-values were > 0.05.