炎症免疫代谢指标预测直肠癌继发心血管疾病风险

赵啟茗[1],姜文慧[1],郭 有[1,2,3,4]
(1. 赣南医科大学公共卫生学院;2. 赣南医科大学第一附属医院医药大数据中心;
3. 赣州市医药大数据重点实验室;4. 国家基因组科学数据中心医学基因组赣南分中心,江西 赣州 341000)

摘 要:目的:评估炎症-免疫-代谢指标对直肠癌患者继发心血管疾病(Cardiovascular disease,CVD)的预测性能,并基于可解释机器学习构建风险预测模型。**方法:**纳入1 160例首次确诊的直肠癌患者。收集10个候选指标,包括反映系统性炎症的中性粒细胞-淋巴细胞比值(Neutrophil-to-lymphocyte ratio,NLR)、血小板-淋巴细胞比值(Platelet-to-lymphocyte ratio,PLR)、系统性免疫炎症(Systemic immune-inflammation,SII)指数及代谢指标甘油三酯-葡萄糖(Triglyceride-glucose,TyG)指数等。采用多因素Logistic回归评估独立关联。LASSO回归筛选变量后,分步开展变量增量价值评估与算法性能比较。采用沙普利加性解释(SHapley Additive exPlanations,SHAP)进行解释分析,Bootstrap内部验证。**结果:**NLR、PLR、SII与CVD风险独立正相关(P均<0.05),CRP呈边缘显著正相关趋势(P=0.051)。限制性立方样条(Restricted cubic spline,RCS)分析未发现上述指标与CVD风险之间存在显著的非线性关联($P_{nonlinear}$均>0.05)。LASSO筛选出13个变量,含TyG等代谢指标。联合模型(GLM)训练集AUC为0.749(95% CI:0.689~0.811)。Bootstrap校正后AUC为0.699(95% CI:0.657~0.739)。在内部验证集中,该模型的AUC为0.700(95% CI:0.660~0.739),提示模型具备中等区分度,并表现出一定的内部稳定性。亚组分析显示模型在低NLR水平(NLR<2.45,AUC=0.731)、年龄<65岁(AUC=0.729)、女性(AUC=0.724)患者中预测效能相对较优。**结论:**NLR、PLR、SII等炎症-免疫-代谢指标与直肠癌继发CVD风险独立正相关。联合预测模型具有中等区分能力,并表现出一定的临床净获益,可作为直肠癌患者继发心血管疾病风险分层的探索性量化工具。

关键词:直肠肿瘤;心血管疾病;炎症指标;代谢指标;预测模型;可解释机器学习
中图分类号:R735.3;R541 **文献标志码:**A **文章编号:**2097-7174(2026)10-0893-13
DOI:10.3969/j.issn.2097-7174.2026.10.001

Inflammatory-immune-metabolic indicators for predicting the risk of secondary cardiovascular disease in rectal cancer

ZHAO Qiming[1], JIANG Wenhui[1], GUO You[1,2,3,4]
(1. School of Public Health, Gannan Medical University; 2. Medical Big Data Center, The First Affiliated Hospital of Gannan Medical University; 3. Ganzhou Key Laboratory of Medical Big Data; 4. Gannan Subcenter for Medical Genomics, National Genomics Data Center, Ganzhou, Jiangxi 341000)

Abstract: Objective: To evaluate the predictive performance of inflammatory-immune-metabolic indicators for secondary cardiovascular disease (CVD) in patients with rectal cancer, and to construct a risk prediction model based on explainable machine learning. **Methods:** A total of 1 160 patients with newly diagnosed rectal cancer were enrolled. Ten candidate indicators were collected, including the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII) for systemic inflammation, alongside the metabolic triglyceride-glucose (TyG) index. Multivariable logistic regression was used to assess independent associations. After variable selection using LASSO regression, a stepwise framework was adopted to evaluate the incremental predictive value of variable sets and to compare algorithm performance. SHapley Additive exPlanations (SHAP) was used for model interpretation, and bootstrap internal validation was performed. **Results:** NLR, PLR, and SII were independently and positively associated with CVD risk (all

基金项目:国家自然科学基金项目(82060618);江西省重点研发计划项目(20203BBGL73184)
通信作者:郭有,男,博士,副教授,研究方向:肿瘤流行病学。E-mail:gy@gmu.edu.cn

P<0.05), while CRP showed borderline statistical significance (*P*=0.051). Restricted cubic spline (RCS) analysis found no evidence of significant non-linearity for these indicators (all $P_{nonlinear}$>0.05). LASSO selected 13 variables, including the metabolic indicator TyG. The combined model (GLM) achieved an AUC of 0.749 (95%*CI*: 0.689-0.811) in the training set. Bootstrap-corrected AUC was 0.699 (95%*CI*: 0.657-0.739). In the internal validation set, the model yielded an AUC of 0.700 (95%*CI*: 0.660-0.739), indicating moderate discrimination and acceptable performance in internal validation. Subgroup analysis showed relatively better predictive performance in patients with a low NLR level (NLR<2.45, AUC=0.731), age <65 years (AUC=0.729), and the female (AUC=0.724). **Conclusion**: Inflammatory-immune-metabolic indicators, including NLR, PLR, and SII, are independently and positively associated with secondary CVD risk in rectal cancer. The combined prediction model demonstrates moderate discriminative performance and provides favorable clinical net benefit, suggesting its potential as an exploratory quantitative tool for secondary cardiovascular disease risk stratification in rectal cancer patients.

Key words: Rectal cancer; Cardiovascular disease; Inflammatory indicators; Metabolic indicators; Prediction model; Explainable machine learning

结直肠癌是全球范围内发病率和死亡率均位居前列的消化道恶性肿瘤[1]。随着诊疗技术的进步,直肠癌患者的生存期显著延长,但其继发心血管疾病(Cardiovascular disease, CVD)的风险也逐渐引起临床关注[2]。流行病学研究表明,癌症患者死于心血管事件的风险可能不亚于肿瘤本身,这一现象已成为"肿瘤心脏病学"领域的核心议题,反映了肿瘤与心血管系统之间的相互作用与共同的病理机制[3]。对于直肠癌患者而言,手术应激、化疗药物(如氟尿嘧啶类、奥沙利铂)的潜在心脏毒性以及肿瘤相关炎症状态,均可能加剧心血管系统的病理负担[4]。因此,在直肠癌确诊早期识别高CVD风险人群,对于优化个体化治疗策略、改善患者长期预后具有重要意义。

近年来,炎症-免疫-代谢相关外周血指标因其经济、便捷、可重复检测等优势,在肿瘤预后评估中受到广泛关注[5-6]。系统性免疫炎症(Systemic immune-inflammation, SII)指数、中性粒细胞-淋巴细胞比值(Neutrophil-to-lymphocyte ratio, NLR)、血小板-淋巴细胞比值(Platelet-to-lymphocyte ratio, PLR)等复合指标,综合反映了机体炎症激活与免疫抑制之间的动态平衡[7-8]。已有研究证实,NLR升高与结直肠癌患者术后并发症及总生存期缩短相关,SII也被报道为多种实体瘤的独立预后因子[9-10]。此外,C反应蛋白(C-reactive protein, CRP)、纤维蛋白原等传统炎症标志物,以及甘油三酯-葡萄糖指数(Triglyceride-glucose index, TyG)等代谢指标,亦被探索用于评估癌症患者的全身性炎症负荷及代谢紊乱状态[11-12]。然而,上述指标在直肠癌患者中与继发CVD的关联强度、独立预测性能以及最佳组合方式,目前尚缺乏系统性的定量评估。

既往研究存在若干不足:(1)多数研究聚焦于炎症-免疫指标对肿瘤预后(如总生存期、无进展生存期)的预测价值,而专门针对直肠癌患者继发心血管事件风险的预测研究极为稀缺——这恰是肿瘤心脏病学领域的关键研究缺口,提示改善肿瘤患者长期预后的前提之一是防控继发CVD;(2)既往预测模型多采用"黑盒"机器学习算法,虽在一定程度上提升了预测精度,但缺乏对模型决策过程的解释,限制了其在临床实践中的接受度和可推广性;(3)关于上述指标与CVD风险之间是否存在非线性剂量-反应关系,目前尚存争议,不同研究因样本量、人群特征及统计方法的差异而得出了不一致的结论。针对这些不足,本研究旨在评估炎症-免疫-代谢指标与直肠癌患者继发CVD的独立关联及剂量-反应关系;进而整合临床协变量与关键血液指标,构建可解释的预测模型,为直肠癌患者继发心血管疾病的早期风险识别、个体化风险分层及临床管理提供参考依据。

1 材料与方法

1.1 研究对象 基于赣南医科大学第一附属医院电子病历系统,建立了2010年1月至2024年3月首次确诊为直肠癌的患者队列,利用完整的基线临床资料和长期随访记录,评估炎症-免疫-代谢指标对继发CVD的预测性能。纳入标准:(1)首次确诊为直肠癌,诊断依据国际疾病分类(ICD-10)编码确定;(2)年龄≥18岁;(3)有完整的基线血常规、生化指标及临床资料;(4)有明确的随访记录(至少6个月)。排除标准:(1)关键临床数据缺失;(2)既往有其他恶性肿瘤病史;(3)合并严重肝肾功能不全或急性感染性疾病;(4)随访期间失访;(5)直肠癌确诊前或基线期内

已存在本研究定义的心血管疾病。最终纳入1 160例患者。本研究已获赣南医科大学第一附属医院医学伦理委员会批准（2020001），所有患者均签署知情同意书。

1.2　结局定义与随访　结局事件为直肠癌继发CVD，依据ICD-10编码将Ⅰ章中的心血管病分为：冠状动脉疾病、心功能不全、心律失常、肺动脉高压、瓣膜病、心包并发症、周围血管性疾病、卒中和血栓性疾病9种心血管疾病。直肠癌患者合并上述任意一种即认定为直肠癌继发心血管疾病[13]。随访期从直肠癌确诊后3个月开始，至首次心血管事件发生、死亡或随访结束（最长6年）。结局变量为二分类变量。

1.3　变量定义

1.3.1　候选预测指标　基于患者基线（直肠癌首次确诊时）的实验室指标，计算6个复合指标并选取4个单一指标，共计10个候选指标。复合指标包括：系统性免疫炎症指数（SII）=血小板×中性粒细胞/淋巴细胞；中性粒细胞-淋巴细胞比值（NLR）=中性粒细胞/淋巴细胞；血小板-淋巴细胞比值（PLR）=血小板/淋巴细胞；血小板-中性粒细胞比值（Platelet-to-neutrophil ratio，PNR）=血小板/中性粒细胞；中性粒细胞-高密度脂蛋白比值（Neutrophil to high-density lipoprotein cholesterol ratio，NHR）=中性粒细胞/高密度脂蛋白；预后营养指数（Prognostic nutritional index，PNI）=白蛋白+5×淋巴细胞。单一指标包括：C反应蛋白（CRP）、纤维蛋白原、癌胚抗原（Carcinoembryonic antigen，CEA）、甘油三酯-葡萄糖指数（TyG）=ln[甘油三酯（mg/dL）×空腹血糖（mg/dL）/2]。

1.3.2　协变量　收集人口学特征[年龄、性别、体重指数（Body mass index，BMI）]、共病情况[高血压、糖尿病、估算肾小球滤过率（Estimated glomerular filtration rate，eGFR）]、肿瘤特征（营养不良、手术史、化疗史）及生命体征[收缩压（Systolic blood pressure，SBP）、舒张压（Diastolic blood pressure，DBP）]。分类变量为二分类（是/否），连续变量保留原始数值。

1.4　数据预处理　对存在缺失值的变量采用链式方程多重插补（Multivariate imputation by chained equations，MICE）进行填补，设置5次插补，插补模型包含所有分析变量。对偏态分布的CRP、纤维蛋白原、CEA、SII进行自然对数转换，并对所有连续变量（含转换后指标）进行标准化处理（均值为0，标准差为1），以消除量纲影响，使后续分析的比值比（Odds ratio，*OR*）值反映指标每增加一个标准差的效应。

1.5　统计学方法与处理　所有统计分析采用R软件（4.5.1版本）完成。首先按是否发生CVD分组进行基线特征比较：连续变量根据正态性检验结果采用独立样本*t*检验或Mann-Whitney *U*检验，分类变量采用卡方检验。针对10个候选指标，分别构建多因素Logistic回归模型，调整全部11个协变量（年龄、性别、BMI、高血压、糖尿病、eGFR、营养不良、手术史、化疗史、SBP、DBP），计算各指标标准化后的*OR*值及95%置信区间（Confidence interval，*CI*）。对回归分析中显著的指标，采用限制性立方样条（Restricted cubic spline，RCS）分析其与CVD风险的剂量-反应关系，设置3个节点（第10、50、90百分位数），通过似然比检验计算非线性*P*值。

预测模型的构建遵循独立流程。将数据集按7∶3比例分层随机划分为训练集（*n*=813）和验证集（*n*=347）。建模前，计算10个候选血液指标间的Spearman秩相关系数，并在纳入全部候选指标的多因素Logistic回归框架中计算方差膨胀因子（Variance inflation factor，VIF），以评估共线性程度。随后，仅在训练集中，将全部21个候选变量（10个血液指标+11个临床协变量）一次性纳入LASSO回归（L1正则化），采用5折交叉验证，以平均交叉验证二项偏差最小值对应的λ（λ. min）确定惩罚参数，保留非零系数变量。

基于LASSO筛选结果，分两步构建模型。第一步，评估变量集的增量预测性能：统一采用广义线性模型（GLM），依次拟合Model A（仅含临床协变量）、Model B（Model A+NLR；NLR因操作简便、临床可及性高、应用最为广泛，被预先设定为代表性单一炎症指标）和Model C（Model A+LASSO筛选出的血液指标），比较三者的区分度[采用ROC曲线下面积（Area under the receiver operating characteristic curve，AUC）及其95% *CI*评价模型区分度]、校准度（Brier评分）及拟合优度[赤池信息准则（Akaike information criterion，AIC）]。AIC仅用于该GLM框架内的模型间比较；第二步，在确定Model C所纳变量为最优变量集后，比较算法性能：分别构建随机森林（RF）和支持向量机（SVM）模型，与GLM横向比较AUC，以确定最优建模算法。

针对性能最优的Model C（GLM），采用SHAP框架进行可解释性分析：计算特征平均|SHAP值|进行全局重要性排序，绘制蜂群图展示特征值与风险方向

的关系，并选取典型患者绘制瀑布图展示个体预测贡献。基于SHAP重要性排序选取前8个关键变量，重新拟合Logistic回归构建列线图。采用Bootstrap自举重抽样法（1 000次）仅对最终确定的模型进行内部验证，计算校正后AUC。在验证集中按年龄（截断值：65岁）、性别、BMI（截断值：24 kg/m^2，中国成人超重阈值）、NLR水平（截断值：本队列中位数2.45）进行亚组分析，对各亚组AUC及其95% *CI*进行描述性比较，以初步评价模型在不同人群中的预测性能。双侧检验，检验水准α=0.05。

2 结果

2.1 研究对象基线特征 本研究共纳入直肠癌患者1 160例，其中发生CVD者319例（27.5%），未发生者841例（72.5%）。基线特征如表1所示。

在人口学特征方面，发生CVD者年龄中位数为67.0岁，高于未发生CVD者的61.0岁（*P*<0.001）；发生CVD者与未发生CVD者性别构成差异无统计学意义（*P*=0.583）。在共病情况方面，发生CVD者高血压患病率（31.3% *vs* 10.7%）、糖尿病患病率（11.3% *vs* 3.7%）均高于未发生CVD者（*P*<0.001）。实验室指标方面，相比于未发生CVD者，发生CVD者的CRP、NHR、NLR、SII、TyG水平升高，eGFR和PNI水平降低（*P*<0.05），而发生CVD者与未发生CVD者间CEA、PLR、PNR、纤维蛋白原水平差异无统计学意义（*P*>0.05）。

表1 直肠癌患者基线特征比较

变量	未发生CVD者（*n*=841）	发生CVD者（*n*=319）	*P*
年龄[岁，*M*(P_{25},P_{75})]	61.00(53.00,69.00)	67.00(60.00,75.00)	<0.001
性别[*n*(%)]			0.583
男	535(63.6)	209(65.5)	
女	306(36.4)	110(34.5)	
手术史[*n*(%)]	402(47.8)	161(50.5)	0.455
化疗史[*n*(%)]	496(59.0)	189(59.2)	0.987
糖尿病[*n*(%)]	31(3.7)	36(11.3)	<0.001
营养不良[*n*(%)]	86(10.2)	38(11.9)	0.469
高血压[*n*(%)]	90(10.7)	100(31.3)	<0.001
BMI[kg/m^2，*M*(P_{25},P_{75})]	21.83(19.83,24.00)	22.41(20.21,24.46)	0.030
SBP[mmHg，*M*(P_{25},P_{75})]	122.00(111.00,134.00)	125.00(111.00,138.00)	0.030
DBP[mmHg，*M*(P_{25},P_{75})]	77.00(70.00,85.00)	78.00(69.00,85.00)	0.973
CEA[ng/mL，*M*(P_{25},P_{75})]	4.51(2.31,11.90)	4.51(2.30,11.60)	0.964
纤维蛋白原[g/L，*M*(P_{25},P_{75})]	3.24(2.66,3.89)	3.30(2.74,4.01)	0.106
CRP[mg/L，*M*(P_{25},P_{75})]	4.53(1.19,22.02)	6.30(1.75,38.42)	0.006
NHR[*M*(P_{25},P_{75})]	3.48(2.47,5.15)	3.68(2.68,5.50)	0.043
NLR[*M*(P_{25},P_{75})]	2.36(1.69,3.62)	2.53(1.84,4.26)	0.008
PLR[*M*(P_{25},P_{75})]	165.21(123.34,229.31)	166.67(127.01,246.17)	0.146
PNI[*M*(P_{25},P_{75})]	47.55(43.45,51.70)	46.70(41.75,50.62)	0.021
PNR[*M*(P_{25},P_{75})]	65.38(50.27,87.50)	63.76(47.09,83.47)	0.105
SII[*M*(P_{25},P_{75})]	597.59(384.03,943.44)	633.98(430.27,1 052.33)	0.049
TyG[*M*(P_{25},P_{75})]	8.38(8.08,8.73)	8.45(8.11,8.89)	0.015
eGFR[mL/(min·1.73 m^2)，*M*(P_{25},P_{75})]	91.25(77.96,99.58)	83.02(64.85,94.41)	<0.001

注：CVD：心血管疾病；BMI：体重指数；SBP：收缩压；DBP：舒张压；CEA：癌胚抗原；CRP：C反应蛋白；NHR：中性粒细胞-高密度脂蛋白比值；NLR：中性粒细胞-淋巴细胞比值；PLR：血小板-淋巴细胞比值；PNI：预后营养指数；PNR：血小板-中性粒细胞比值；SII：系统性免疫炎症指数；TyG：甘油三酯-葡萄糖指数；eGFR：估算肾小球滤过率。

2.2 候选指标与CVD的独立关联及剂量-反应关系 为评估各候选指标与CVD的独立关联，分别构建10个多因素Logistic回归模型，调整年龄、性别、BMI、高血压、糖尿病、eGFR、营养不良、手术史、化疗史、SBP、DBP。结果显示（表2），NLR（*OR*=1.17，95%*CI*：1.02～1.33，*P*=0.024）、PLR（*OR*=1.15，

95% *CI*：1.00～1.32，*P*=0.042）、SII（*OR*=1.15，95% *CI*：0.99～1.32，*P*=0.049）与CVD风险呈正相关，CRP亦呈边缘显著正向趋势（*OR*=1.14，95% *CI*：0.99～1.32，*P*=0.051），其余指标未显示独立关联。

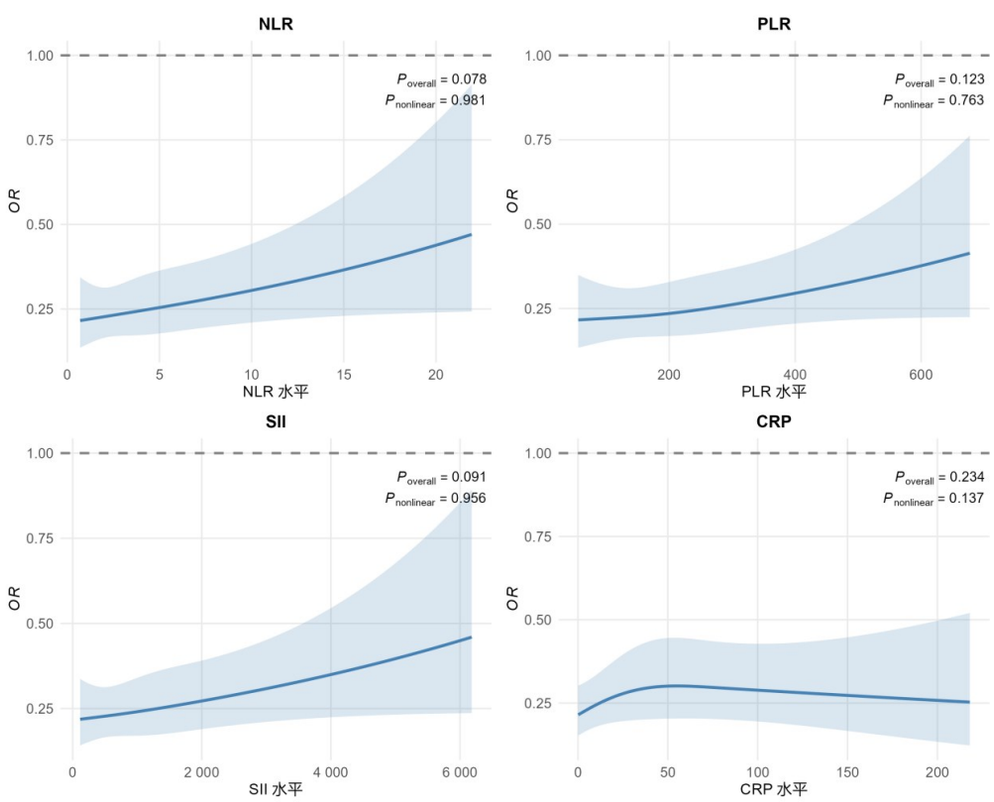
采用RCS进一步探索显著指标的剂量-反应关系。RCS分析结果显示（图1），NLR、PLR、SII与CVD风险的关系呈线性模式（$P_{nonlinear}$>0.05），未发现显著的非线性关系；CRP虽边缘显著（*P*=0.051），但亦未呈现明显非线性特征（$P_{nonlinear}$=0.137），整体表现为剂量-反应的连续递增关系。

2.3　变量筛选与预测模型构建及性能比较　在进行LASSO变量筛选前，对10个候选指标进行系统共线性诊断。Spearman相关系数矩阵（图2）显示，候选指标间存在不同程度的相关性：SII与NLR的相关系数最高（*r*=0.87，*P*<0.001），与PLR亦呈较强相关（*r*=0.78，*P*<0.001）；NLR与PLR之间也存在一定相关性（*r*=0.65，*P*<0.001）。此外，部分候选指标间的相关系数绝对值超过0.40，其余多数指标间的相关性较弱。多因素Logistic回归框架下的VIF诊断（表2）进一步显示，SII的VIF为8.64，在全部候选指标中

表2　10个候选指标与CVD风险的多因素Logistic回归分析及共线性诊断

指标	*OR*（95% *CI*）	*P*	VIF
NLR	1.17（1.02～1.33）	0.024	4.92
PLR	1.15（1.00～1.32）	0.042	4.01
SII	1.15（0.99～1.32）	0.049	8.64
CRP	1.14（0.99～1.32）	0.051	1.19
TyG	1.13（0.98～1.31）	0.103	1.07
纤维蛋白原	1.05（0.91～1.21）	0.528	1.23
NHR	1.03（0.91～1.16）	0.674	1.51
PNI	0.98（0.84～1.14）	0.768	1.61
CEA	0.96（0.83～1.11）	0.543	1.01
PNR	0.94（0.80～1.10）	0.439	1.47

注：*OR*及*P*值为各指标单独调整全部11个协变量（年龄、性别、BMI、高血压、糖尿病、eGFR、营养不良、手术史、化疗史、SBP、DBP）后所得。所有连续变量均经标准化处理（均值=0，标准差=1），偏态分布变量（SII、CRP、纤维蛋白原、CEA）在标准化前先行自然对数转换，*OR*值表示指标每增加1个标准差的效应量。VIF在包含全部10个候选指标的同一Logistic回归框架中计算；VIF≥5提示中等共线性，VIF≥10提示严重共线性。



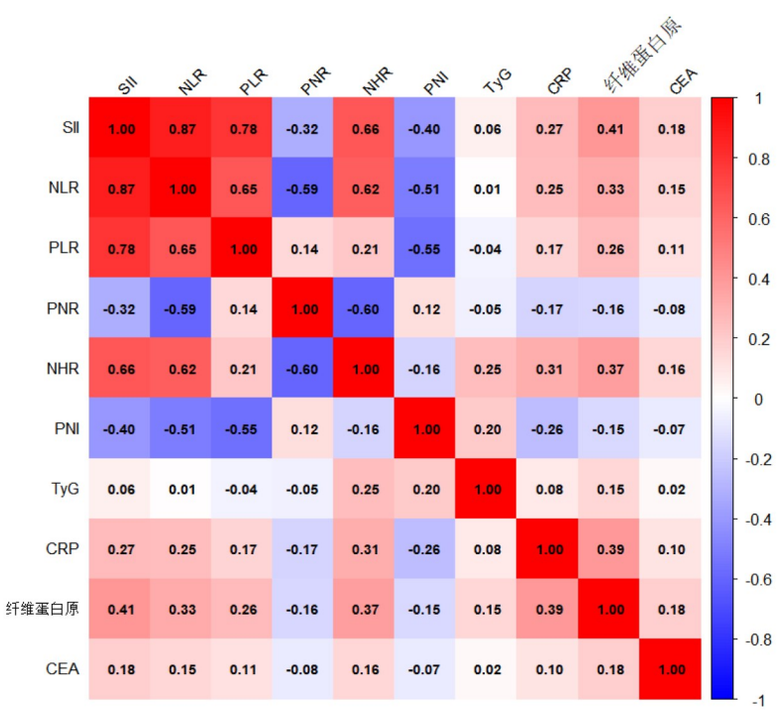
蓝线表示CVD发生风险（比值比）的估计值，阴影区域为95% *CI*。模型调整了年龄、性别、BMI、高血压、糖尿病、eGFR、营养不良、手术史、化疗史、SBP及DBP。节点设置在第10、50、90百分位数。

图1　经多因素校正后NLR、PLR、SII、CRP与CVD风险关联的RCS图

最高，提示存在中等程度的共线性；NLR和PLR的VIF分别为4.92和4.01，其余指标的VIF均低于1.61，未见明显共线性。

LASSO回归采用5折交叉验证，以平均交叉验证二项偏差最小值对应的λ（λ.min）确定惩罚参数，最终保留13个非零系数变量（图3、表3）：血液指标4个（NLR、PLR、TyG、CRP）、临床协变量9个（年龄、BMI、高血压、糖尿病、eGFR、手术史、化疗史、SBP、DBP），其中高血压的系数绝对值最大（0.882），提示其为最强的独立风险因子。值得注意的是，SII虽在多因素回归中差异有统计学意义（P=0.049），但LASSO将其系数压缩为0——这与共线性诊断的结果一致。

基于筛选结果，采用分步框架评估模型性能，其中协变量集经初步筛选后确定为9个（不含性别、营养不良，两者在基线特征比较中组间差异均无统计学意义，见表1）。第一步，评估变量的增量预测性能：在统一采用广义线性模型（GLM）的前提下，比较Model A（9个临床协变量）、Model B（临床协变量+NLR）及Model C（LASSO筛选出的全部13个变量，即Model A的9个临床协变量+NLR、PLR、TyG、CRP）的区分度。结果显示，Model C的AUC为0.749，高于Model A（0.741）与Model B（0.742）（图4）；其准确度（0.764）、敏感性（0.231）及特异性（0.953）亦优于前两者；Brier评分（0.163）与AIC（901.71）均更低，提示校准度及拟合优度均更优（表4）。第二步，确定最佳变量集后比较算法性能：基于Model C纳入的13个变量，分别构建GLM、RF、SVM模型。GLM的AUC为0.749，优于RF（0.718）与SVM（0.704），故选择GLM作为最终建模算法。



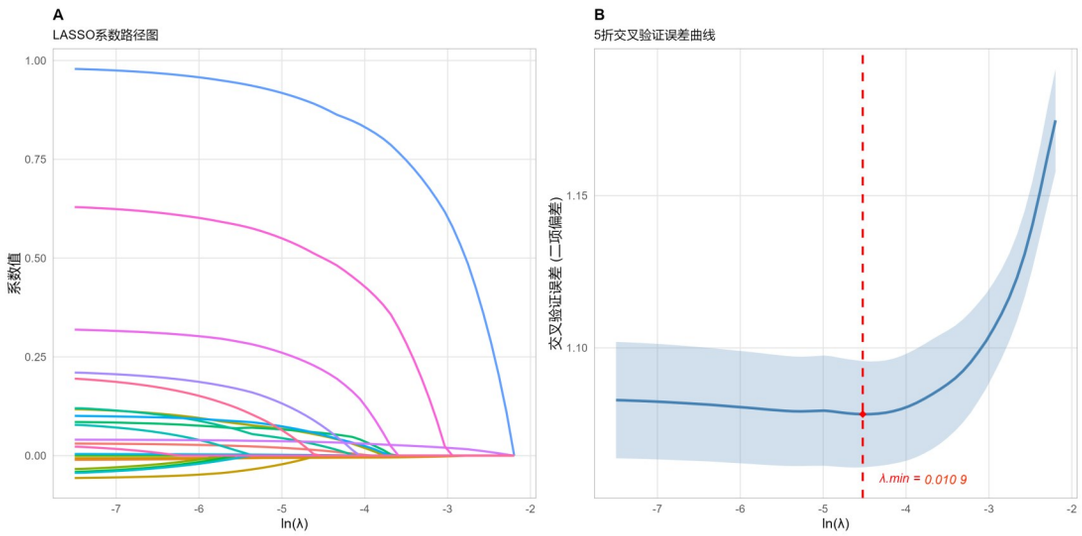
热图颜色及数值表示两两指标间的Spearman秩相关系数；红色代表正相关，蓝色代表负相关，颜色越深、绝对值越大提示相关性越强。SII、CRP、纤维蛋白原、CEA经自然对数转换后纳入分析。

图2 10个候选指标间Spearman秩相关系数矩阵热图

2.4 模型校准度、临床实用性及可解释性分析

针对性能最优的Model C（GLM），进一步评估其校准度与临床实用性。校准曲线显示，预测概率与实际观测概率具有良好的一致性（Brier评分=0.163，Hosmer-Lemeshow检验P=0.180）。决策曲线分析表明，在阈值概率0.10～0.40范围内，模型的净获益均高于“全部干预”与“无干预”两种极端策略，提示具备一定的临床净获益（图5）。

采用SHAP框架对模型进行可解释性分析（图6）。蜂群图揭示了各特征取值与风险方向的关联：高血压、高NLR、高龄、合并糖尿病、高CRP均推高预测风险，而高eGFR则降低预测风险（图6A）。全局特征重要性排序显示，贡献最大的前5个变量依次为年龄、高血压、手术史、TyG、eGFR（图6B）。



A：LASSO 回归系数路径图，不同颜色的曲线表示各变量回归系数随惩罚参数 λ 变化的轨迹；随着 λ 增大，各变量系数逐渐向 0 收缩。B：5 折交叉验证偏差曲线，蓝色实线表示平均交叉验证二项偏差，阴影表示其上下 1 个标准误的范围；红色纵虚线表示平均交叉验证偏差最小时对应的 λ（λ. min＝0. 010 9）。

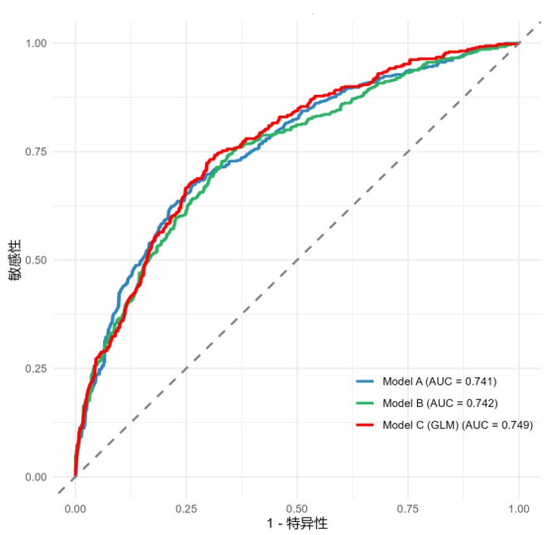
图 3　LASSO 回归变量筛选路径图及交叉验证误差曲线

表 3　LASSO 回归筛选出的非零系数变量

变量	系数
NLR_z	0. 058
CRP_z	0. 051
PLR_z	0. 025
TyG_z	0. 054
年龄	0. 034
BMI	0. 013
高血压	0. 882
糖尿病	0. 502
eGFR	−0. 005
手术史	0. 218
化疗史	0. 077
SBP	0. 001
DBP	−0. 002

注：表中变量名与正文中 NLR、CRP 等对应，_z 后缀表示该变量已经标准化处理。变量筛选基于训练集（*n*=813）进行 5 折交叉验证，并以平均交叉验证二项偏差最小值对应的 λ（λ. min）确定惩罚参数。

个体水平上，瀑布图直观展示了特征的正负向贡献叠加过程：高风险患者的预测主要受高血压、NLR 升高及年龄增加共同贡献（图 6C），低风险患者则因无高血压、NLR 正常及 eGFR 较高等因素，预测概率向低风险方向偏移（图 6D）。基于 SHAP 重要性排序，选取前 8 个关键变量（年龄、高血压、手术史、TyG、eGFR、化疗史、CRP、SBP），重新拟合 Logistic



Model A 为临床协变量模型；Model B 为临床协变量+NLR 模型；Model C 为 LASSO 筛选出的全部 13 个变量模型。所有模型均在训练集（*n*=813）中采用广义线性模型（GLM）训练。

图 4　训练集中 3 种预测模型区分度的 ROC 曲线比较

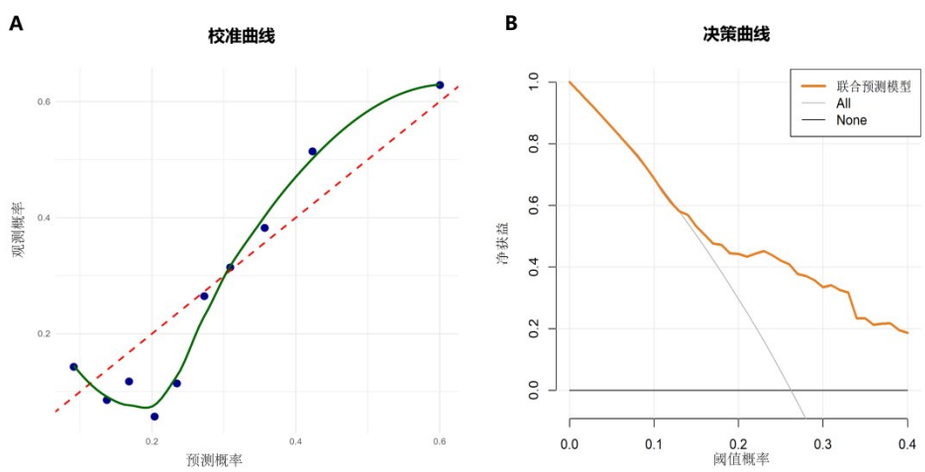
回归，构建可视化列线图（图 7），以便于临床个体化风险评估。

2. 5　模型稳健性验证与亚组分析　采用 Bootstrap 自举重抽样（1 000 次）对训练集模型进行了内部验证。结果显示（表 5），Bootstrap 校正后 AUC 为 0. 699（95% *CI*：0. 657 ~ 0. 739），较训练集原始 AUC0. 749（95% *CI*：0. 689 ~ 0. 811）有所下降，提示模型存在一定程度的乐观偏倚。随后，将最终模型应用于内部验证集（*n*=347）进行验证，其 AUC 为 0. 700（95% *CI*：

表4 不同预测模型在训练集中的性能比较

模型	AUC	准确度	敏感性	特异性	Brier评分	AIC
Model A(临床模型)	0.741	0.762	0.208	0.949	0.166	903.46
Model B(临床+NLR模型)	0.742	0.764	0.219	0.951	0.165	902.98
Model C(GLM)	0.749	0.764	0.231	0.953	0.163	901.71
Model C(RF)	0.718	0.738	0.030	0.985	0.178	–
Model C(SVM)	0.704	0.738	0.129	0.951	0.179	–

注:Model A仅包含临床协变量;Model B为临床协变量+NLR;Model C为LASSO筛选出的全部13个变量(Model A的9个协变量+NLR、PLR、TyG、CRP)。GLM为广义线性模型,RF为随机森林,SVM为支持向量机。AIC仅用于同一GLM框架下Model A/B/C间的比较,不用于跨算法比较。



A:校准曲线。对角线代表理想校准,Brier评分=0.163,提示预测概率与实际观测概率一致性良好;B:决策曲线。在0.10~0.40的阈值概率范围内,模型的净获益高于"全部干预"和"无干预"两种极端策略。

图5 Model C的校准曲线与决策曲线

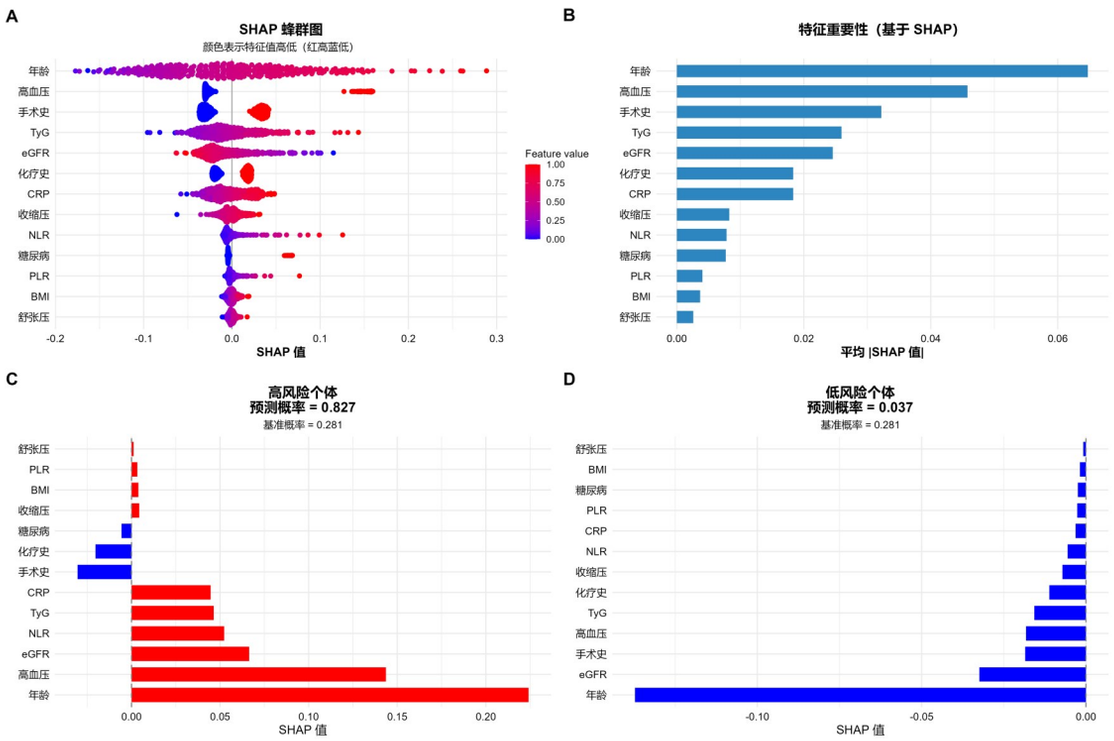
0.660~0.739)。验证集AUC与Bootstrap校正后的AUC接近,提示Bootstrap校正后的性能估计与内部验证集结果基本一致。

亚组分析结果(表6)显示,在BMI分层中,正常组和超重/肥胖组的AUC分别为0.675和0.704;在NLR分层中,低水平组和高水平组的AUC分别为0.731和0.620;在年龄分层中,<65岁组和≥65岁组的AUC分别为0.729和0.619;在性别分层中,女性和男性患者的AUC分别为0.724和0.664。从AUC点估计看,模型在NLR低水平、年龄<65岁及女性患者中的区分度相对较高,而在NLR高水平及年龄≥65岁患者中的区分度相对较低;BMI分层结果未呈现相同方向。鉴于各亚组95% *CI*较宽,且未进行正式的交互作用或亚组间AUC比较检验,相关结果仅作描述性解释。

3 讨论

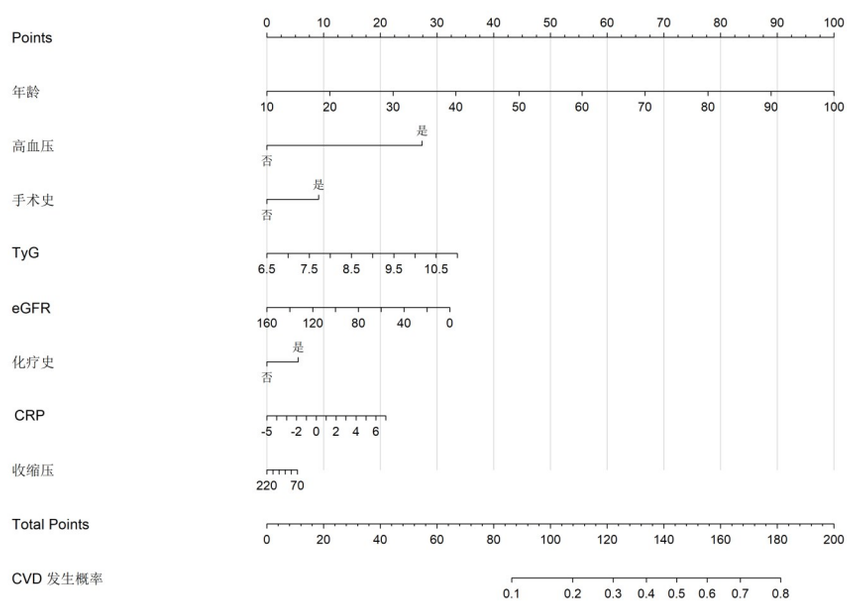
本研究发现,直肠癌患者外周血免疫炎症/代谢相关指标与直肠癌继发CVD风险存在独立关联。多变量分析显示,NLR、PLR及SII与CVD风险呈独立正相关,而CRP呈接近显著的正向趋势。RCS分析未发现上述指标与CVD风险之间存在显著的非线性关联证据($P_{nonlinear}$>0.05),提示其关联形态未表现出明确的非线性阈值。在预测模型层面,Model C(GLM,纳入LASSO筛选变量)在本研究数据中表现出相对较好的区分度和校准表现,并显示出一定的临床净获益。

既往研究[14-15]已提示炎症相关指标与心血管不良结局存在关联,但在肿瘤背景下,炎症信号往往与治疗相关毒性、合并代谢紊乱共同作用,导致风险构成更复杂。本研究在直肠癌患者中进一步补充了"炎症-免疫失衡"与"心血管事件风险"相关联的证据基础:独立关联仍成立,且风险形态更偏向连续谱而非阈值触发。这与NLR、PLR在心血管结局预测中的汇总证据方向一致;同时,SII作为复合免疫炎症指数,其在心血管风险评估中的增量价值



A：蜂群图（各特征值与预测风险方向的关系）；B：特征全局重要性排序（各特征平均|SHAP值|），每个点代表一个样本，颜色从蓝（低值）到红（高值），横轴SHAP值为正表示推高预测风险；C：高风险典型患者个体SHAP瀑布图；D：低风险典型患者个体SHAP瀑布图。SHAP瀑布图中，箭头向右（红色）表示增加预测风险，箭头向左（蓝色）表示降低预测风险。

图6　基于SHAP的联合模型可解释性分析图



变量重要性经SHAP分析确认。TyG、CRP均为标准化后变量。

图7　基于SHAP重要性排序筛选的前8个关键变量构建的CVD风险预测列线图

也得到了印证[16-18]。这一线性递增模式提示，免疫炎症负荷可能以累积式、持续式的方式影响血管病理过程（如内皮损伤、促凝与动脉粥样硬化进展），而非通过某一阈值触发风险的跃升。这一“连续风险”形态也与近年来对连续变量建模重要性的认识相一致：当真实关系更接近线性时，将连续变量

表5 联合预测模型在训练集与内部验证集中的区分度评估

指标	AUC(95% *CI*)
原始AUC(训练集)	0.749(0.689~0.811)
Bootstrap校正后AUC	0.699(0.657~0.739)
验证集AUC	0.700(0.660~0.739)

注:训练集(n=813)采用Bootstrap法(自举重抽样1 000次)进行内部验证以校正乐观偏倚;验证集(n=347)为从原始队列中划出的内部测试集,用于评价模型的内部验证表现。

不加证据地离散化反而可能损失风险信息与可解释性[19]。

需要指出的是,RCS未发现显著非线性拐点,并不意味着炎症指标缺乏临床可操作性,而是提示NLR、PLR及SII更适合作为连续风险标志物加以解读,而非简单的二分类诊断指标。在缺乏明确生物学拐点的情况下,临床阈值设定可采用以下2种互补策略:(1)基于队列分布设定分位数预警值。本研究中NLR中位数为2.45。亚组分析显示,NLR高水平亚组(≥2.45)的模型区分度(AUC=0.620)低于低水平亚组(AUC=0.731),提示在高炎症负荷人群中,NLR对CVD的边际预测贡献相对减弱;结合高水平亚组整体事件率更高的特征,NLR≥2.45可作为识别需加强心血管风险评估人群的初步参考界值。若NLR进一步持续升高,则可视为炎症负荷进一步加重的信号,建议结合年龄、高血压、糖尿病、eGFR及治疗方式综合判断,必要时转诊心内科进行评估;(2)基于模型预测概率的绝对风险分层。结合决策曲线分析,本模型在阈值概率0.10~0.40范围内具有持续正净获益,提示临床上可将模型预测概率≥10%作为加强心血管风险评估与随访的起点,而在预测概率更高时考虑心电图/超声心动图评估及传统危险因素的强化管理。上述2种策略可根据临床实际情境灵活互补:在临床信息不完整、无法获取全部模型变量时,可优先采用NLR分位数进行快速初筛;当完整变量可获取时,建议直接使用模型预测概率进行精细风险分层。需要强调的是,上述阈值均为基于本单中心队列数据的探索性预警参考值,而非具有诊断意义的硬性截断点,其临床适用性仍有待在多中心、前瞻性队列中加以验证。

从生物学机制上看,NLR、PLR与SII均指向炎症-免疫-血栓相关通路的耦联:中性粒细胞相关的炎症放大与淋巴细胞下降所代表的免疫监视减弱,可能共同维持更高水平的系统性炎症;同时,血小板活化与促凝倾向的上升会进一步促进血栓形成与血管重塑。因此,当NLR或PLR升高时,可能通过慢性炎症、内皮功能障碍与促凝血机制协同推动CVD发生概率上移。SII将中性粒细胞、淋巴细胞与血小板的信号整合为复合输出,更贴近“系统性免疫炎症负荷”的整体风险表达,因此在多因素框架中更易保持独立性与稳定贡献[17-18,20]。在肿瘤患者中,上述通路的作用强度还可能被治疗与肿瘤相关炎症进一步放大。肿瘤心脏病学领域的综述与指南普遍强调:癌症与CVD之间存在共享风险因素与互相放大的病理网络(炎症激活、代谢异常、内皮损伤等),这为炎症指标预测直肠癌继发心血管事件提供了更充分的生物学合理性支持[3,21]。

虽然SII在多因素Logistic回归中差异有统计学意义(P=0.049)并被纳入RCS分析,但在LASSO筛选中系数被压缩为0。结合相关系数矩阵(图2)与

表6 联合模型在验证集各亚组中的区分度表现

变量	样本量	事件数	AUC(95% *CI*)
BMI			
正常(<24 kg/m²)	260	67	0.675(0.600~0.749)
超重/肥胖(≥24 kg/m²)	87	24	0.704(0.575~0.834)
NLR			
低水平(<2.45)	181	45	0.731(0.641~0.820)
高水平(≥2.45)	166	46	0.620(0.525~0.716)
年龄			
<65岁	213	53	0.729(0.651~0.807)
≥65岁	134	38	0.619(0.511~0.728)
性别			
女	116	30	0.724(0.615~0.832)
男	231	61	0.664(0.583~0.745)

注:所有AUC及其95% *CI*均在验证集(n=347)中计算。亚组结果为描述性分析,未进行正式的交互作用或亚组间AUC比较检验。

VIF 诊断（表 2）可见，SII 与 NLR 的 Spearman 相关系数高达 0. 87，与 PLR 为 0. 78，VIF 高达 8. 64（全部候选指标中最高），提示存在中等程度的共线性及信息冗余；而 NLR（VIF=4. 92）和 PLR（VIF=4. 01）的独立性相对更强。LASSO 的 L1 正则化在面对高共线变量时自动保留独立信号更强者、剔除冗余者，SII 被剔除正是这一机制的典型表现。综合以上三层证据提示：在已同时获取 NLR 与 PLR 数据的临床场景中，NLR 与 PLR 的组合已能捕捉本模型所需的主要免疫炎症负荷信息，额外纳入 SII 对模型预测能力的边际提升有限。这一发现反映的是本研究模型内部由高共线性导致的统计冗余关系，而非否定 SII 在其他研究背景（如未同时纳入 NLR、PLR 的模型，或以 SII 作为独立复合暴露变量的病理机制研究）中的应用价值。

关于 CRP 仅呈边缘显著（P=0. 051）的现象，可从以下 2 个角度理解。一方面，CRP 属于广谱急性期炎症反应标志物，其生物学意义与 NLR/PLR 等细胞比例指标既有交叉（均反映全身炎症负荷），又有所不同（CRP 主要由肝脏产生，反映上游炎症信号，而 NLR/PLR 更直接反映中性粒细胞与淋巴细胞的动态失衡）；尽管 CRP 的 VIF 仅为 1. 19，共线性程度低，但在多因素框架中，NLR 和 PLR 已从细胞学层面提供了更直接的炎症免疫失衡信息，可能在一定程度上削弱了 CRP 的独立预测贡献。另一方面，尽管 CVD 事件数达到一定规模，但当真实效应量较小且与多个协变量存在相关结构时，CRP 的独立增量仍可能不足以稳定达到阈值显著性。值得特别指出的是，Model C（GLM）的敏感性仅为 0. 231，特异性高达 0. 953，两者的显著不对称界定了本模型的应用边界：凭借极低的误报率，被模型识别为高风险的患者具有较高的 CVD 真实发病置信度，故本模型更适合作为“风险确认”（rule-in）的辅助工具，帮助临床医师精准锁定需重点加强心血管监测的高危患者；而不适合作为面向全体直肠癌患者的广泛筛查工具（rule-out），因约 77% 的真实 CVD 患者会被漏诊，可能错过早期干预时机。这一定位亦得到决策曲线分析的支持：在 0. 10 ~ 0. 60 的宽阈值范围内，模型均显示出优于“全部干预”和“无干预”策略的净获益，提示在中高风险阈值下进行有针对性的干预具有临床价值。同时，本研究在进行 SHAP 全局特征重要性排序时，CRP 虽未进入前 5 位，但依然被成功保留至最终的 8 变量核心列线图中，这提示其虽然在多因素回归中仅呈边缘显著，但在综合模型框架内仍提供了不可忽视的稳健贡献：即在直肠癌患者中，CRP 依然是炎症负荷的关键表征，并被纳入最终 CVD 风险预测模型，提示其仍具有重要的预测价值[22-23]。

在本研究比较的多种算法中，广义线性模型（GLM）获得了最高 AUC（0. 749），同时具有良好的临床可解释性，因此被选为最终预测模型。如前所述，Model C（GLM）在区分度与校准度（以 Brier 评分和拟合优度衡量）方面均优于其他模型，决策曲线亦在 0. 10 ~ 0. 40 阈值概率范围内显示出优于“全部干预/无干预”策略的正净获益。这意味着模型不仅具有较好的排序能力，更可能在真实临床中帮助医师将有限资源用于更可能受益的人群，从而改善随访与管理策略的效率[24-25]。模型训练集经 Bootstrap 校正后 AUC 为 0. 699，较原始 AUC（0. 749）下降约 0. 05，提示存在一定程度的乐观偏倚；内部验证集的 AUC 为 0. 700，与 Bootstrap 校正后的 AUC 接近，提示模型在本研究内部数据中具有一定稳定性。其产生乐观偏倚的潜在成因主要有两方面：一方面，NLR、CRP 等炎症指标本身存在生理性波动，模型在训练过程中可能捕捉到部分样本特异性的噪声信号，导致在重抽样验证时区分度有所稀释；另一方面，本研究为单中心回顾性队列，样本同质性较高，模型容易对该特定人群特征过度拟合，Bootstrap 验证时 AUC 的轻微下降正是这种倾向的客观反映。

在验证集中，亚组分析显示，模型在 NLR 低水平、年龄<65 岁、女性患者等低风险特征人群中区分度更好（AUC：0. 724 ~ 0. 731），在 NLR 高水平、高龄患者中区分度略有下降（AUC：0. 619 ~ 0. 620）。值得注意的是，BMI 分层未呈现与其他变量一致的方向：BMI 正常者 AUC（0. 675）反而低于超重/肥胖者（0. 704）。上述阈值策略在不同患者亚组中的适用性，也可从亚组分析结果的效能差异中得到进一步理解。高危亚组（高龄、高 NLR）中区分度下降，并非预测因子效应方向的改变，而与以下机制密切相关：⑴基线风险饱和效应。当患者已处于高龄或高炎症状态时，年龄和炎症本身已贡献了绝大部分风险信息，模型中其他预测变量的增量区分空间受到压缩，形成类似“天花板效应”；且高危亚组患者的整体风险水平趋于一致，组内风险分层梯度缩窄，模型可用于区分的风险差异空间客观减小，导致 AUC 下降；⑵高危亚组存在更主导的替代致病机

制，削弱了炎症指标的边际预测性能。对于高龄患者，血管老化、动脉僵硬度增加、内皮修复能力下降及多重共病负担（如慢性肾功能减退、多发动脉粥样硬化）可能已成为CVD发生的主要驱动力[14]；对于高NLR患者，炎症水平升高可能同时反映肿瘤负荷增大、潜在感染、严重营养不良或围治疗期应激等多种混杂病理状态，使NLR在该人群中特异性下降，对CVD的增量预测性能受到稀释[15]。此外，直肠癌患者接受手术、放化疗（如奥沙利铂、氟尿嘧啶类）的差异可能通过心肌损伤、内皮毒性及血栓形成等途径影响CVD风险，这些治疗相关心血管毒性在高危亚组中可能较基线炎症指标具有更强的决定性作用[4,14]；(3)竞争性死亡事件。高龄患者同时面临较高的肿瘤相关死亡风险，部分患者在CVD事件发生前已因肿瘤进展死亡，竞争性死亡事件在一定程度上削弱了模型对该亚组CVD风险的判别能力。综上，上述多重机制共同解释了模型在低炎症负荷、年轻及女性患者中区分度较好，而在高龄和高炎症负荷亚组中AUC下降的现象；各亚组AUC的95% *CI*存在重叠，提示上述差异应解释为效能的趋势性异质性，而非具有确定统计学意义的差异。该规律与传统心血管预测模型在高危亚组中区分度普遍低于低危亚组的经验高度一致[26-27]，并非本模型独有的缺陷。

SHAP分析提示年龄、高血压、手术史、TyG与eGFR为全局贡献最大的前5个变量，其中eGFR升高呈保护趋势。该解释与临床既有的肿瘤心脏病学风险框架高度契合：高血压与年龄是传统心血管病理的核心决定因素；手术创伤应激与TyG指数反映的代谢紊乱，进一步叠加了肿瘤患者特有的心血管负担；而在该基础之上，化疗史与CRP等指标的保留，则成功将“治疗相关应激与炎症负荷”引入了风险图谱。这种结构说明，本模型并非单纯依靠某一类指标，而是实质上构建了一个涵盖“传统风险+治疗应激+代谢/炎症负荷”的多维评价体系。这种变量对齐与直观的瀑布图呈现，有助于医师进行透明的临床风险沟通与随访路径制定[28-29]。从方法学角度，近年来TRIPOD+AI等报告指南强调可解释性与透明报告对于临床预测模型转化的重要性[30]。本研究采用LASSO筛选、对比建模并引入SHAP解释，是提高临床可信度的合理路径。

本研究存在以下局限性：(1)受限于回顾性数据中部分CVD事件诊断日期存在区间删失，本研究以二分类结局替代时间-事件建模，采用Logistic回归而非Cox模型，未能在模型中整合个体随访时长的权重。未来前瞻性研究应精确记录事件时间，采用更高效的生存分析模型；(2)所有炎症/免疫/代谢指标均为基线单次测量，未能反映纵向动态变化；(3)缺乏独立外部队列验证，模型外推性尚待检验，当前应视为探索性工具；(4)要求完整基线及随访记录的设计可能引入选择偏倚，纳入样本整体健康状况可能优于实际临床人群；(5)将9种病理机制不尽相同的CVD合并为单一复合结局，尽管已有前期研究支持各亚型在直肠癌背景下共享炎症基础，但这种合并策略仍可能掩盖特定亚型各自的风险决定因素，有待未来更大样本的亚型分层分析加以验证。未来研究应在多中心队列中开展外部验证，并引入指标的纵向轨迹构建动态预测模型，以提升泛化能力与临床转化价值。

所有作者均声明不存在利益冲突关系。

参考文献：

[1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2024, 74(3): 229-263.

[2] Sturgeon K M, Deng L, Bluethmann S M, et al. A population-based study of cardiovascular disease mortality risk in US cancer patients[J]. Eur Heart J, 2019, 40(48): 3889-3897.

[3] Lyon A R, López-Fernández T, Couch L S, et al. 2022 ESC Guidelines on cardio-oncology developed in collaboration with the European Hematology Association (EHA), the European Society for Therapeutic Radiology and Oncology (ESTRO) and the International Cardio-Oncology Society (IC-OS)[J]. Eur Heart J, 2022, 43(41): 4229-4361.

[4] Leong D P, Waliany S, Abdel-Qadir H, et al. Cardiovascular considerations during cancer therapy: gaps in evidence and JACC: CardioOncology expert panel recommendations[J]. JACC CardioOncol, 2024, 6(6): 815-834.

[5] 柳子宁，于滨. 外周血相关指标在直肠癌新辅助治疗中的临床应用价值[J]. 消化肿瘤杂志(电子版)，2023，15(2)：166-170.

[6] Menyhart O, Fekete J T, Győrffy B. Inflammation and colorectal cancer: a meta-analysis of the prognostic significance of the systemic immune-inflammation index (SII) and the systemic inflammation response index (SIRI)[J]. Int J Mol Sci, 2024, 25(15): 8441.

[7] Islam M M, Satici M O, Eroglu S E. Unraveling the clinical significance and prognostic value of the neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, systemic immune-inflammation index, systemic inflammation response index, and delta neutrophil index: an extensive literature review[J]. Turk J Emerg Med, 2024, 24(1): 8-19.

[8] 陈笑，张银洋，陈珂，等. 全身炎症标志物对重症肌无力患者预后影响分析[J]. 神经疾病与精神卫生，2026，26(1)：42-49.

[9] Cho Y, Park S B, Yoon J Y, et al. Neutrophil to lymphocyte ratio can predict overall survival in patients with stage Ⅱ to Ⅲ colorectal cancer[J]. Medicine, 2023, 102(11): e33279.

[10] Terasaki F, Sugiura T, Okamura Y, et al. Systemic immune-inflammation index as a prognostic marker for distal cholangiocarcinoma [J]. Surg Today, 2021, 51(10): 1602-1609.

[11] Burger P M, Pradhan A D, Dorresteijn J A N, et al. C-reactive protein and risk of cardiovascular events and mortality in patients with various cardiovascular disease locations [J]. Am J Cardiol, 2023, 197: 13-23.

[12] 王烁珊，吴敬国，荆小莉，等. 甘油三酯葡萄糖指数与冠状动脉粥样硬化性心脏病关系的研究进展[J]. 中华全科医学，2021，19(6)：1008-1012.

[13] Qiao H, Wang C, Yang C, et al. Association between increased C-reactive protein and cardiovascular disease among patients with rectal cancer [J]. Front Oncol, 2023, 13: 1287619.

[14] Ioffe D, Bhatia-Patel S C, Gandhi S, et al. Cardiovascular concerns, cancer treatment, and biological and chronological aging in cancer: JACC family series [J]. JACC CardioOncol, 2024, 6(2): 143-158.

[15] Zullig L L, Sung A D, Khouri M G, et al. Cardiometabolic comorbidities in cancer survivors JACC: CardioOncology state-of-the-art review [J]. JACC CardioOncology, 2022, 4(2): 149-165.

[16] Zhang X, Wei R, Wang X, et al. The neutrophil-to-lymphocyte ratio is associated with all-cause and cardiovascular mortality among individuals with hypertension[J]. Cardiovasc Diabetol, 2024, 23(1): 117.

[17] Qin P, Ho F K, Celis-Morales C A, et al. Association between systemic inflammation biomarkers and incident cardiovascular disease in 423 701 individuals: evidence from the UK biobank cohort [J]. Cardiovasc Diabetol, 2025, 24(1): 162.

[18] Ye Z, Hu T, Wang J, et al. Systemic immune-inflammation index as a potential biomarker of cardiovascular diseases: a systematic review and meta-analysis [J]. Front Cardiovasc Med, 2022, 9: 933913.

[19] Lopez-Ayala P, Riley R D, Collins G S, et al. Dealing with continuous variables and modelling non-linear associations in healthcare data: practical guide [J]. BMJ, 2025, 390: e082440.

[20] Rawish E, Langer H F. Platelets and the role of P2X receptors in nociception, pain, neuronal toxicity and thromboinflammation[J]. Int J Mol Sci, 2022, 23(12): 6585.

[21] Ananthan K, Lyon A R. The role of biomarkers in cardio-oncology[J]. J Cardiovasc Transl Res, 2020, 13(3): 431-450.

[22] Lee S A, Kwon S O, Park H, et al. Association of serum high-sensitivity C reactive protein with risk of mortality in an Asian population: the Health Examinees cohort [J]. BMJ Open, 2022, 12(7): e052630.

[23] He J, Song C, Zhang R, et al. Discordance between neutrophil to lymphocyte ratio and high sensitivity C-reactive protein to predict clinical events in patients with stable coronary artery disease: a large-scale cohort study [J]. J Inflamm Res, 2023, 16: 5439-5450.

[24] Piovani D, Sokou R, Tsantes A G, et al. Optimizing clinical decision making with decision curve analysis: insights for clinical investigators[J]. Healthcare, 2023, 11(16): 2244.

[25] Riley R D, Archer L, Snell K I E, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study[J]. BMJ, 2024, 384: e074820.

[26] Efthimiou O, Seo M, Chalkou K, et al. Developing clinical prediction models: a step-by-step guide[J]. BMJ, 2024, 386: e078276.

[27] Riley R D, Snell K I E, Archer L, et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study [J]. BMJ, 2024, 384: e074821.

[28] Writing Group for the CKD Prognosis Consortium, Grams M E, Coresh J, et al. Estimated glomerular filtration rate, albuminuria, and adverse outcomes: an individual-participant data meta-analysis [J]. JAMA, 2023, 330(13): 1266-1277.

[29] Ponce-Bobadilla A V, Schmitt V, Maier C S, et al. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development [J]. Clin Transl Sci, 2024, 17(11): e70056.

[30] Collins G S, Moons K G M, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods[J]. BMJ, 2024, 385: e078378.

（收稿：2026-04-07）（修回：2026-07-19）

（责任编辑：眭荣燕）