·流行病与统计方法·

基于 SHAP 可解释机器学习的心梗患者全因死亡风险预测模型构建

张译予，邢菲，刘宗凤，孟莹莹，吴小娟，丛显铸，黄月，石福艳，王素珍
山东第二医科大学公共卫生学院卫生统计学教研室，山东 潍坊 261053

摘要：目的　探索影响心肌梗死患者全因死亡的关键风险因素并构建可解释的生存预测模型。**方法**　基于英国生物样本库，纳入 5 840 例心肌梗死患者，结合单因素 Cox 回归、最小绝对收缩和选择算子(least absolute shrinkage and selection operator,LASSO)回归和变量重要性测度(variable importance measure, VIMP)法筛选变量。将数据按 7∶3 划分为训练集和测试集，采用 XGBoost 算法构建生存预测模型，并利用该数据库重复评估队列的心肌梗死患者数据作为验证集进行验证；结合沙普利加性解释(SHapley Additive exPlanations,SHAP)方法增强模型可解释性。**结果**　本研究中位随访时间为 14.4 年(95%*CI*：14.36～14.48)，多阶段筛选出胱抑素 C、尿素、糖化血红蛋白、中性粒细胞计数、红细胞分布宽度和年龄 6 个关键风险因素。模型在训练集 3、5 和 7 年 AUC 分别为 0.80(0.76～0.84)、0.82(0.80～0.85)、0.82(0.80～0.84)，测试集为 0.72(0.64～0.79)、0.71(0.66～0.76)、0.70(0.66～0.74)，内部验证集 7 年 AUC 为 0.82(0.73～0.92)。校准曲线显示预测值与实际风险一致性良好。Brier 评分在各数据集中均<0.25，预测误差较低。决策曲线分析显示，在较宽阈值概率范围内净获益均为正值，模型具有良好的临床应用潜力。SHAP 分析证实各特征对风险预测的正向贡献。**结论**　本研究构建的预测模型能有效预测心肌梗死患者全因死亡风险，为个体化预后评估和干预提供依据。

关键词：心肌梗死；机器学习；临床血液指标；全因死亡
中图分类号：R542.22　　**文献标志码**：A　　**文章编号**：1003-8507(2026)19-3477-09
DOI：10.20043/j.cnki.MPM.202606038

Development of a model for predicting all-cause mortality risk in patients with myocardial infarction based on SHAP-based explainable machine learning

ZHANG Yi-yu, XING Fei, LIU Zong-feng, MENG Ying-ying, WU Xiao-juan, CONG Xian-zhu, HUANG Yue, SHI Fu-yan, WANG Su-zhen
Department of Health Statistics, School of Public Health, Shandong Second Medical University, Weifang, Shandong 261053, China

Abstract: Objective To explore key risk factors influencing all-cause mortality in myocardial infarction patients and construct an interpretable survival prediction model. **Methods** Based on the UK Biobank, 5 840 myocardial infarction patients were included. Variables were selected using univariable Cox regression, least absolute shrinkage and selection operator (LASSO) regression, and variable importance measure (VIMP) methods. The dataset was divided into training and test sets at a 7:3 ratio. A survival prediction model was constructed using the XGBoost algorithm and validated using a validation cohort of patients with myocardial infarction from the same database. The shapley additive explanations (SHAP) method was incorporated to enhance model interpretability. **Results** The median follow-up duration in this study was 14.4 years (95% *CI*: 14.36-14.48). Multistage screening identified six core risk factors: cystatin C, urea, glycated hemoglobin, neutrophil count, red cell distribution width, and age. The model achieved AUC values of 0.80 (0.76-0.84), 0.82 (0.80-0.85), and 0.82 (0.80-0.84) on the training set for 3, 5, and 7 years, respectively, while the test set yielded 0.72 (0.64-0.79), 0.71 (0.66-0.76), and 0.70 (0.66-0.74). The internal validation set achieved an AUC of 0.82 (0.73-0.92) for the 7-year prediction. Calibration curves demonstrated good agreement between predicted and observed risks. Brier scores were <0.25 across all datasets, indicating low prediction error. Decision curve analysis showed positive net benefits across a wide range of threshold probabilities. SHAP analysis confirmed the positive contributions of each feature to risk prediction. **Conclusion** The predictive model constructed in this study effectively forecasts all-cause mortality risk in myocardial infarction patients, providing a basis for individualized prognosis assessment and interventions.

Keywords: Myocardial infarction; Machine learning; Clinical blood biomarkers; All-cause mortality

基金项目：山东省科技厅自然科学基金项目（ZR2023MH313）；山东省医疗卫生科技计划项目（202412050529）；潍坊医学院科研创新计划（2023BSQD014）

作者简介：张译予(2000—)，女，硕士在读，研究方向：卫生统计学
通信作者：王素珍，E-mail：wangsz@sdsmu.edu.cn；石福艳，E-mail：shifuyan@sdsmu.edu.cn；王素珍与石福艳为共同通信作者

心肌梗死是一种由冠状动脉粥样硬化所引起的急性缺血性心脏病，发病急、病情危险，易引起恶性心律失常等严重并发症[1]。尽管近年来再灌注治疗及综合药物干预不断优化，心肌梗死患者长期预后仍不理想，全因死亡风险在出院后长期随访中依然较高[2-3]，鉴于此，亟需有效的早期预警和预后评估指标。临床血液指标能够反映炎症、代谢紊乱及心血管损伤等多种病理生理状态，对心肌梗死患者的早期诊断、风险分层和预后评估具有重要价值[4-5]。传统统计模型整合变量维度少，而机器学习能够整合多维度临床数据，解析变量间非线性交互作用，有望建立更精准的预测模型[6]。近年来，机器学习方法逐渐应用于心肌梗死预后研究，并在死亡风险预测方面显示出较传统统计模型更优的性能[7-8]。然而，现有研究构建预测模型所用变量多来源于人口学特征及既往病史，对常规血液生物标志物的系统整合不足；部分机器学习模型（如随机森林）虽提升了预测性能，但“黑箱”特征明显，可解释性不足[6]，限制了其临床推广应用。针对上述不足，本研究系统纳入 58 项常规临床血液指标，采用单因素 Cox、最小绝对收缩和选择算子（least absolute shrinkage and selection operator,LASSO）回归、变量重要性测度（variable importance measure, VIMP）、三阶段互补筛选策略，基于极限梯度提升（extreme gradient boosting,XGBoost）构建心肌梗死全因死亡风险预测模型，并通过沙普利加性解释（SHapley Additive exPlanations,SHAP）分析实现可视化和临床可解释性，同时遵循规范的样本量估算方法，最终构建仅含 6 项常规血液指标的简约模型，在保证良好预测性能的同时降低模型复杂度，提高临床实用性。SHAP 分析进一步量化各变量对风险预测的贡献方向与强度，增强模型透明度和临床可解释性。所有变量均为常规检测指标，成本低、易获取，具有较好的基层推广潜力，可为心肌梗死患者个体化风险评估、长期健康管理及精准干预提供科学依据，并为心血管疾病的智能化风险预警与公共卫生防控提供参考。

1　资料与方法

1.1　研究人群　本研究数据来源于 UK Biobank 数据库，该数据库在 2006—2010 年招募了超过 50 万名 40～69 岁的参与者，是一项大规模的前瞻性队列研究[9]。本研究纳入基线评估时已确诊为心肌梗死的参与者，结局为全因死亡，即随访期间任何原因导致的死亡。随访截至 2023 年 6 月 1 日，随访时间定义为从参与者报告的最早心肌梗死的日期到所有原因的死亡、失访或本次随访结束的时间，中位随访时间为 14.4（14.36～14.48）年。58 项临床血液指标在多组变量的缺失值数量完全一致，缺失率集中在 5%～15% 之间，不满足随机缺失假设。故本研究采用完全病例分析。最终共纳入 5 840 名参与者，筛选流程。

1.2　特征变量选择　本研究纳入 58 项可能与心肌梗死相关的变量，包括年龄、性别及多项临床血液指标[10]。详细的 58 项变量清单见表 1。以生存时间与状态为结局，为避免过拟合风险并提高模型稳定性，采用多阶段筛选策略。先进行单因素 Cox 回归预筛选（$P<0.05$），其目的在于剔除与结局无显著关联的噪声变量，降低后续高维正则化回归的计算负担。尽管单因素 Cox 预筛选可能遗漏部分存在非线性效应或联合作用的变量，但本研究后续采用 LASSO回归和 VIMP[11]排序作为补充筛选环节，VIMP 值按降序排列，以曲线出现明显转折（即“肘点”）的位置作为筛选界值。LASSO的正则化机制可识别变量间的联合作用，VIMP 排序可捕获非线性效应，弥补 LASSO线性假设不足。三阶段互补策略在控制模型复杂度的同时，尽可能保留有预测价值的信息。

表 1　初始纳入的 58 项变量

Table 1　Initially included 58 variables

类别	变量名称
人口学	年龄、性别
血常规	白细胞计数、中性粒细胞计数、中性粒细胞百分比、淋巴细胞计数、淋巴细胞百分比、单核细胞计数、单核细胞百分比、嗜碱性粒细胞计数、嗜碱性粒细胞百分比、嗜酸性粒细胞计数、嗜酸性粒细胞百分比、红细胞计数、血红蛋白浓度、红细胞压积百分比、平均红细胞体积、平均红细胞血红蛋白、平均红细胞血红蛋白浓度、红细胞分布宽度、网织红细胞计数、网织红细胞百分比、高光散射网织红细胞计数、高光散射网织红细胞百分比、未成熟网织红细胞比例、平均网织红细胞体积、平均球状细胞体积、血小板计数、平均血小板体积、血小板分布宽度、血小板压积
血糖	葡萄糖、糖化血红蛋白
肾功能	尿素、肌酐、胱抑素 C
肝功能	丙氨酸氨基转移酶、天门冬氨酸氨基转移酶、碱性磷酸酶、谷氨酰转移酶、白蛋白、总蛋白、总胆红素、直接胆红素、
血脂	胆固醇、甘油三酯、高密度脂蛋白胆固醇、低密度脂蛋白直测、载脂蛋白 A、载脂蛋白 B
炎症指标	C 反应蛋白
电解质与矿物质	钙、磷酸盐
性激素	睾酮、性激素结合球蛋白
其他	胰岛素样生长因子、尿酸盐、维生素 D

1.3　样本量计算　本研究基于四步法[12]估算生存结局预测模型的样本量，根据前期查阅资料结合本研究数据，获得每人年事件发生数为 0.02，总随访时间为 17 年，平均随访时间为 12.71 年，Cox-Snell R^2 为 0.1，经特征变量选择（见方法 1.2），最终确定 6 个预测因子。基于 R 语言 pmsampsize 包实现样本量估算，要求样本量为 510 例，预期发生结局事件 130 例。本研究共纳入 5 840 名参与者，其中结局事件有 1 629 例，远高于阈值，满足预设的样本量要求。

1.4　模型构建与评估　完成特征选择后，将研究对象按 7∶3 比例随机分为训练集和测试集。并从 UK Biobank 2012—2013 年重复评估队列中筛选符合标准的心肌梗死患者作为内部验证队列。本研究采用 XGBoost 算法构建生存预测模型。其内置 survival:cox 目标函数可直接处理右删失数据，自带正则化防过拟合，且与 SHAP 分析兼容，可实现模型可解释性[13]。模型以 cox-nloglik 为评估指标，超参数通过 5 折交叉验证进行调优，在预设参数空间（eta:0.01/0.03/0.05；max_depth:3/5/7；subsample: 0.8/1.0；colsample_bytree: 0.8/1.0；gamma:0/0.5/1.0）中搜索，最终选取交叉验证误差最小的组合：eta=0.03，max_depth=3，subsample=1，colsample_bytree=1，gamma=0.5。最大迭代 1 000 轮，早停 50 轮，随机种子 123。模型在训练集上拟合参数，经测试集和内部验证集评估其预测性能与稳定性。采用时间依赖受试者工作特征（time-dependent receiver operating characteristic，timeROC）曲线、校准曲线、Brier 评分、决策曲线评估模型预测能力，并以 3、5 和 7 年为关键时间节点进行分析。

1.5　模型可解释性　进一步整合 SHAP 分析方法，通过 SHAP 摘要图和依赖图对本研究所构建模型纳入变量的重要性进行度量，同时，描述特征在个体之间的 SHAP 值分布，对每个特征在预后中的作用进行解释[14]。

1.6　统计学分析　采用 R 4.4.2 进行数据分析，正态连续性变量以（$\bar{x} \pm s$）表示，组间比较采用方差分析。非正态分布的计量资料用［$M(P_{25},P_{75})$］表示，组间比较采用秩和检验，计数资料用［n（%）］表示。用 reverse K-M 法计算中位随访时间。检验水准 α=0.05。

2　结　果

2.1　患者基线情况　本研究共纳入 5 840 名心肌梗死患者，其中男性 5 068 人（86.8%），女性 772 人（13.2%），中位年龄为 63 岁。见表 2。

2.2　初步变量筛选　对性别、年龄、56 项临床血液指标进行单因素 Cox 回归，进一步采用 LASSO 回归及交叉优化，选取最优误差内最大的 λ 值（lambda.1se）[15]作为筛选标准，结果见表 3。结果显示，lambda.lse 为 0.017 时，最终保留 24 个变量。除性别、年龄外，涵盖多类生理指标：血常规（嗜酸性粒细胞百分比、淋巴细胞百分比、平均红细胞体积等 11 项）；肝功能指标（白蛋白、谷氨酰转移酶）；血糖指标（葡萄糖、糖化血红蛋白）；肾功能指标（尿素、肌酐、胱抑素 C）；炎症指标（C 反应蛋白）；性激素（性激素结合球蛋白）以及尿酸盐、维生素 D。见图 1。

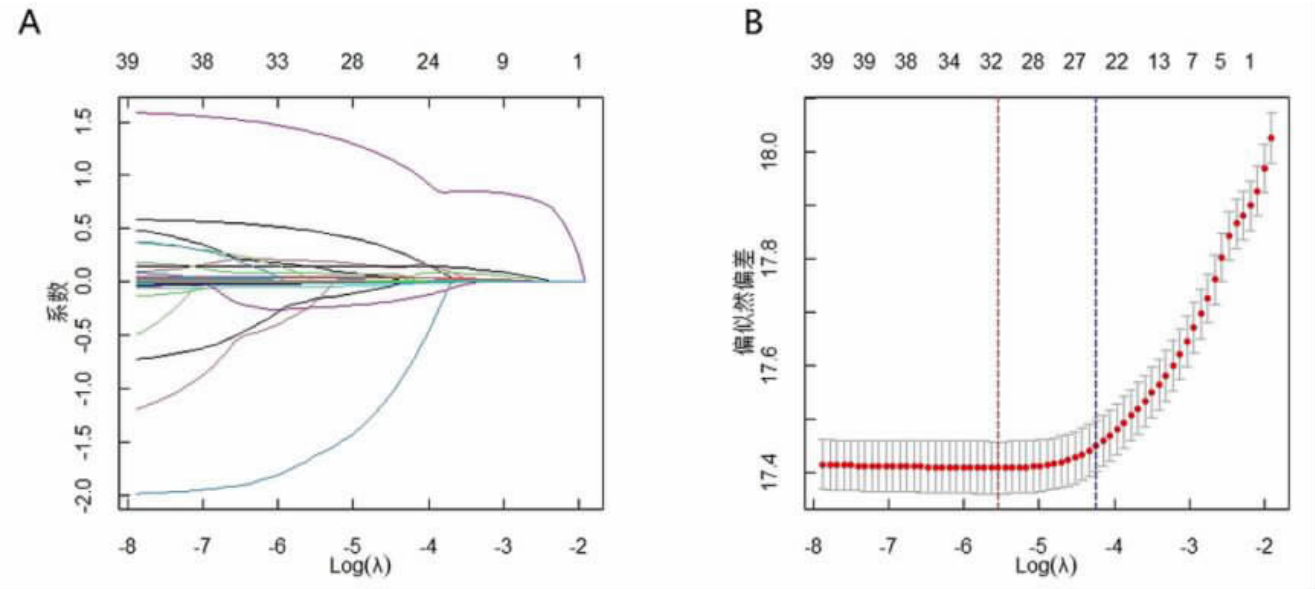
表 2　心肌梗死患者基本信息［n(%)，$M(P_{25},P_{75})$］

Table 1　Baseline characteristics of patients with myocardial infarction［n(%)，$M(P_{25},P_{75})$］

变量		人数(n=5 840)
性别	女	772 (13.22)
	男	5 068 (86.78)
年龄（岁）		63 (59,67)
血常规	白细胞计数	7.27 (6.20,8.53)
	中性粒细胞计数	4.50 (3.67,5.50)
	淋巴细胞百分比	26.40 (21.90,31.40)
	单核细胞计数	0.54 (0.43,0.70)
	嗜酸性粒细胞百分比	2.47 (1.59,3.63)
	平均红细胞体积	92.0 (89.10,94.80)
	红细胞分布宽度	13.50 (13.00,14.10)
	平均网织红细胞体积	107.00 (103.00,112.00)
	平均球状细胞体积	82.60 (79.20,86.20)
	血小板压积	0.21 (0.18,0.24)
	红细胞计数	4.64 (4.37,4.89)
血糖	葡萄糖	5.08 (4.68,5.64)
	糖化血红蛋白	37.90 (35.10,41.80)
肾功能	尿素	5.79 (4.92,6.84)
	肌酐	81.70 (72.00,92.80)
	胱抑素 C	0.99 (0.89,1.11)
肝功能	谷氨酰转移酶	35.30 (24.80,55.30)
	白蛋白	44.90 (43.20,46.60)
炎症指标	C 反应蛋白	1.42 (0.73,2.99)
性激素	性激素结合球蛋白	39.20 (29.60,50.60)
其他	尿酸盐	355.00 (302.00,413.00)
	维生素 D	45.30 (30.90,62.00)

注：本表仅列出经单因素 Cox 回归与 LASSO 回归筛选后保留的 24 个候选预测变量；所有数据基于参与者首次评估时的测量值（见方法部分 1.3 节）。

2.3　关键变量筛选　VIMP 结果如图 2 所示。胱抑素 C 的 VIMP 最高，为 0.096，随后尿素骤降至 0.049，形成第一个明显拐点，但仍处于较高水平。尿素及之后的糖化血红蛋白（0.042）、中性粒细胞计数（0.037）、红细胞分布宽度（0.035）、年龄（0.033）的 VIMP 仍均高于 0.03。自第 7 个变量肌酐（0.026）起，VIMP 降至 0.03 以下，呈现第二个拐点。因此，以 VIMP>0.03 作为入选关键预测因子的界值，最终筛选出前 6 个变量：胱抑素 C、尿素、糖化血红蛋白、中性粒细胞计数、红细胞分布宽度、年龄。



注：图 A 为 LASSO回归系数路径图，表示每个变量所对应的系数值；图 B 为 LASSO交叉验证图，左边虚线 lambda.min 表示在所有的 λ 值中使得均方误差最小的参数，右边虚线 lambda.lse 表示在最小误差一个标准误范围内最大的 λ 值。

图 1　LASSO回归进行特征变量筛选的过程

Figure 1　Feature selection process using LASSO regression

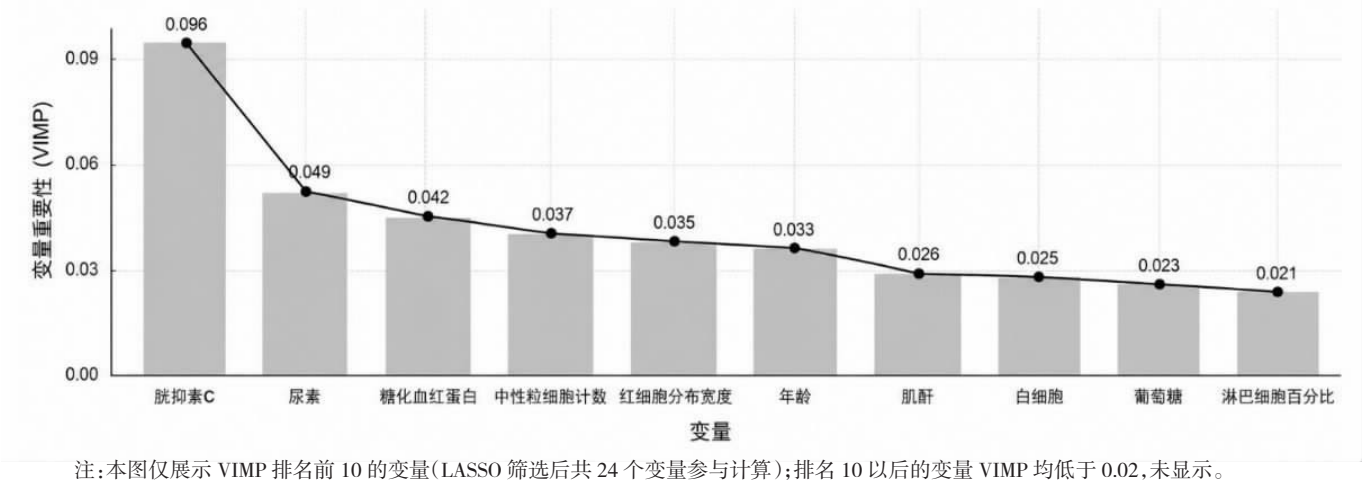
表 3　单因素 Cox 与 LASSO回归筛选后保留的 24 个候选变量的分析结果

Table 3　Results of the analysis of the 24 candidate variables retained after screening using univariate Cox and LASSO regression

变量分组	变量	*HR* (95% *CI*)	*P* 值	LASSO 系数
性别	男 (vs.女)	1.385 (1.180 ~ 1.624)	<0.001	0.176
年龄	年龄	1.068 (1.057 ~ 1.078)	<0.001	0.042
血常规	白细胞计数	1.140 (1.116 ~ 1.165)	<0.001	0.001
	中性粒细胞计数	1.204 (1.173 ~ 1.235)	<0.001	0.089
	淋巴细胞百分比	0.961 (0.955 ~ 0.968)	<0.001	–0.006
	单核细胞计数	1.620 (1.430 ~ 1.836)	<0.001	0.170
	嗜酸性粒细胞百分比	0.957 (0.932 ~ 0.983)	0.001	–0.010
	平均红细胞体积	1.029 (1.019 ~ 1.040)	<0.001	0.010
	红细胞分布宽度	1.276 (1.239 ~ 1.315)	<0.001	0.147
	平均网织红细胞体积	1.029 (1.023 ~ 1.035)	<0.001	0.001
	平均球状细胞体积	1.034 (1.026 ~ 1.043)	<0.001	0.010
	血小板压积	0.358 (0.128 ~ 1.000)	0.05	–0.547
	红细胞计数	0.646 (0.574 ~ 0.726)	<0.001	–0.131
血糖	葡萄糖	1.116 (1.096 ~ 1.137)	<0.001	0.033
	糖化血红蛋白	1.026 (1.022 ~ 1.030)	<0.001	0.013
肾功能	尿素	1.156 (1.135 ~ 1.178)	<0.001	0.001
	肌酐	1.006 (1.005 ~ 1.007)	<0.001	–0.001
	胱抑素 C	2.953 (2.729 ~ 3.195)	<0.001	0.952
肝功能	谷氨酰转移酶	1.002 (1.002 ~ 1.003)	<0.001	0.001
	白蛋白	0.916 (0.899 ~ 0.933)	<0.001	–0.009
炎症指标	C 反应蛋白	1.033 (1.027 ~ 1.039)	<0.001	0.001
性激素	性激素结合球蛋白	1.005 (1.003 ~ 1.008)	<0.001	0.004
其他	尿酸盐	1.002 (1.001 ~ 1.003)	<0.001	<0.001
	维生素 D	0.991 (0.988 ~ 0.993)	<0.001	–0.005

2.4　训练集、测试集和验证集患者临床资料的比较

训练集、测试集和验证集患者在胱抑素 C、尿素、糖化血红蛋白、中性粒细胞计数、红细胞分布宽度、年龄变量中，组间比较差异均无统计学意义，见表 4。



注：本图仅展示 VIMP 排名前 10 的变量（LASSO 筛选后共 24 个变量参与计算）；排名 10 以后的变量 VIMP 均低于 0.02，未显示。

图 2　VIMP 变量重要性排序图（前 10）

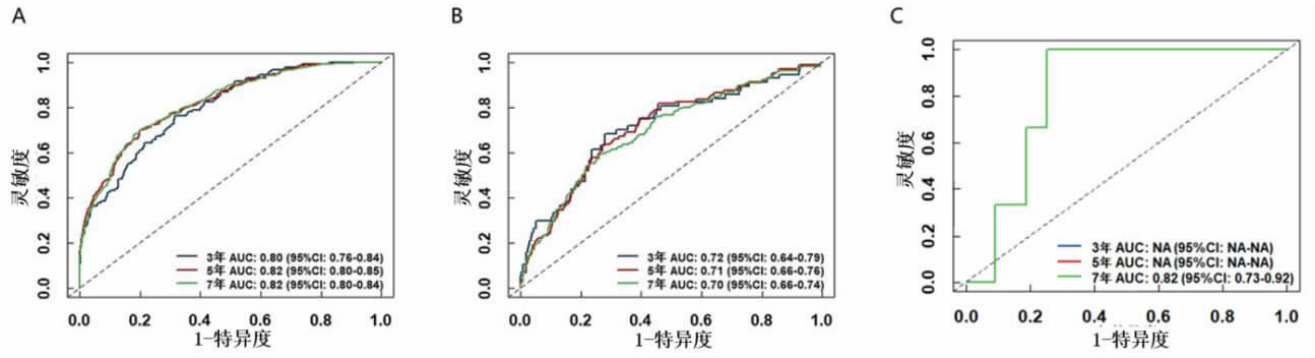
Figure 2　VIMP variable importance ranking (top 10)

表 4　训练集、测试集和验证集患者临床资料的比较［$M(P_{25},P_{75})$］

Table 4　Comparison of clinical characteristics among patients in the training, testing, and validation sets［$M(P_{25}, P_{75})$］

变量	训练集（n=4 088）	测试集（n=1 752）	验证集（n=126）	P 值
脱抑素 C	0.99 (0.89,1.11)	0.99 (0.89,1.11)	0.98 (0.89,1.07)	0.828
尿素	5.76 (4.93,6.82)	5.84 (4.91,6.88)	5.91 (4.99,6.77)	0.575
糖化血红蛋白	37.90 (35.10,41.90)	37.90 (35.10,41.60)	37.45 (35.55,40.68)	0.915
中性粒细胞计数	4.50 (3.67,5.48)	4.50 (3.67,5.50)	4.45 (3.80,5.48)	0.626
红细胞分布宽度	13.51 (13.07,14.07)	13.50 (13.00,14.09)	13.60 (13.10,14.20)	0.242
年龄	63.00 (59.00,67.00)	63.00 (59.00,67.00)	60.00 (63.00,66.75)	0.375

2.5　模型构建与验证评价　基于训练集数据，采用 6 个关键预测因子构建 XGBoost 生存预测模型。在训练集中，模型 3、5、7 的 AUC 值分别为 0.80、0.82 和 0.82，见图 3A；测试集分别为 0.72、0.71 和 0.70，见图 3B；验证集中仅 7 年的 AUC 值为 0.82，模型区分能力良好，见图 3C。各数据集校准曲线均与 45° 线基本吻合，Brier 评分均＜0.25，校准良好，见图 4。决策曲线显示，在较宽阈值概率范围内模型净获益均＞0，模型具有一定临床应用价值，见图 5。

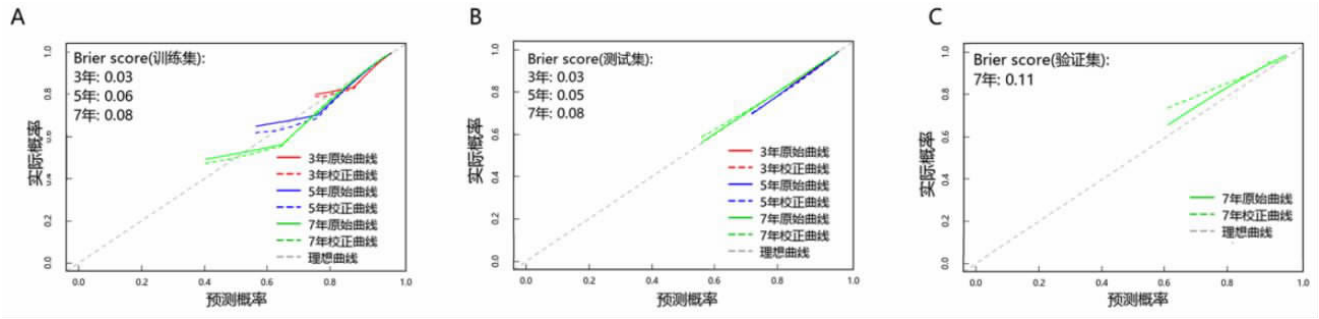


注：A 训练集；B 测试集；C 验证集。

图 3　在训练集、测试集、验证集中模型预测的 timeROC 图

Figure 3　Time-dependent ROC curves of model predictions in the training, testing, and validation sets

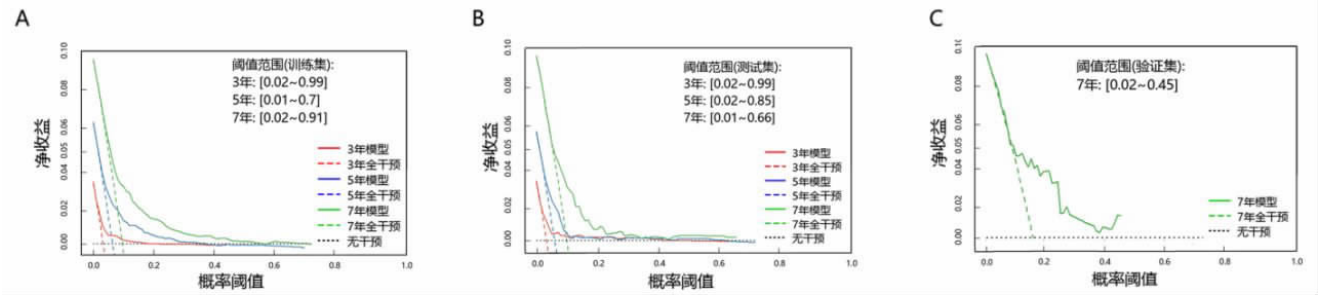
2.6　模型可解释性　SHAP 摘要图显示，脱抑素 C 对模型预测贡献最大，其次为年龄、糖化血红蛋白、红细胞分布宽度、中性粒细胞计数及尿素。上述指标取值升高时对应更高的 SHAP 值，其升高与全因死亡风险增加相关。SHAP 依赖图进一步显示，脱抑素 C、年龄、糖化血红蛋白、红细胞分布宽度、中性粒细胞计数及尿素均与死亡风险呈正相关，且在一定范围内随指标升高风险明显增加。见图 6、7。



注：A 训练集；B 测试集；C 验证集；对角线表示理想校准（预测概率 = 实际概率）；图中标注的数值为对应时间点的 Brier 评分，用于量化预测概率与实际结果之间的平均误差，值越低表示模型校准度越好；Brier 评分与校准曲线从不同维度共同评估模型的校准性能。

图 4　在训练集、测试集、验证集中模型预测的校准曲线

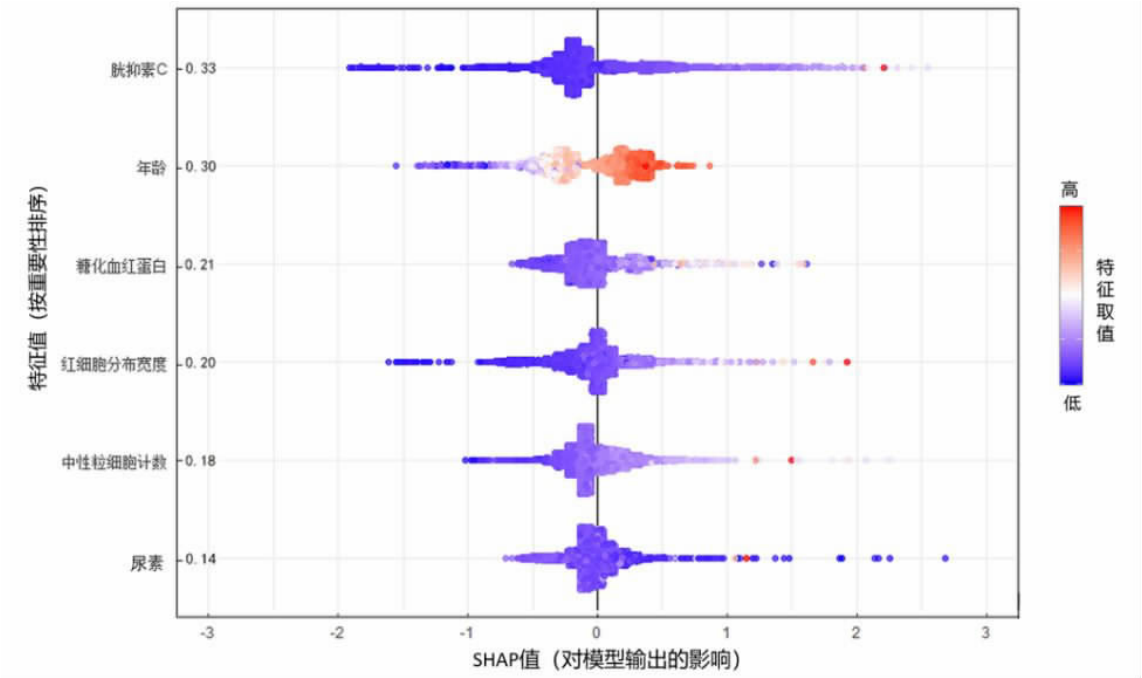
Figure 4　Calibration curves of model predictions in the training, testing, and validation sets



注：A 训练集；B 测试集；C 验证集。

图 5　在训练集、测试集、验证集中模型预测的决策曲线分析曲线

Figure 5　DCA curves of model predictions in the training, testing, and validation sets



注：纵坐标右侧数值表示对应特征的重要性得分（数值越大，重要性越高）。

图 6　XGBoost 在心肌梗死数据集中 6 个特征变量的 SHAP 摘要图

Figure 6　SHAP summary plot for six feature variables in the myocardial infarction dataset using XGBoost

3　讨　论

近年来，多项研究已采用机器学习方法构建心肌梗死患者死亡风险预测模型。但不同研究在变量体系、特征筛选策略及模型解释性方面仍存在差异。Soleimani 等[16]比较了 7 种机器学习算法预测急性心肌梗死患者院内死亡，随机森林表现最优（AUC=0.924），但该研究聚焦于短期院内死亡，未涉及长期预后。Kim 等[17]提出的可靠且可解释的人工智能系统

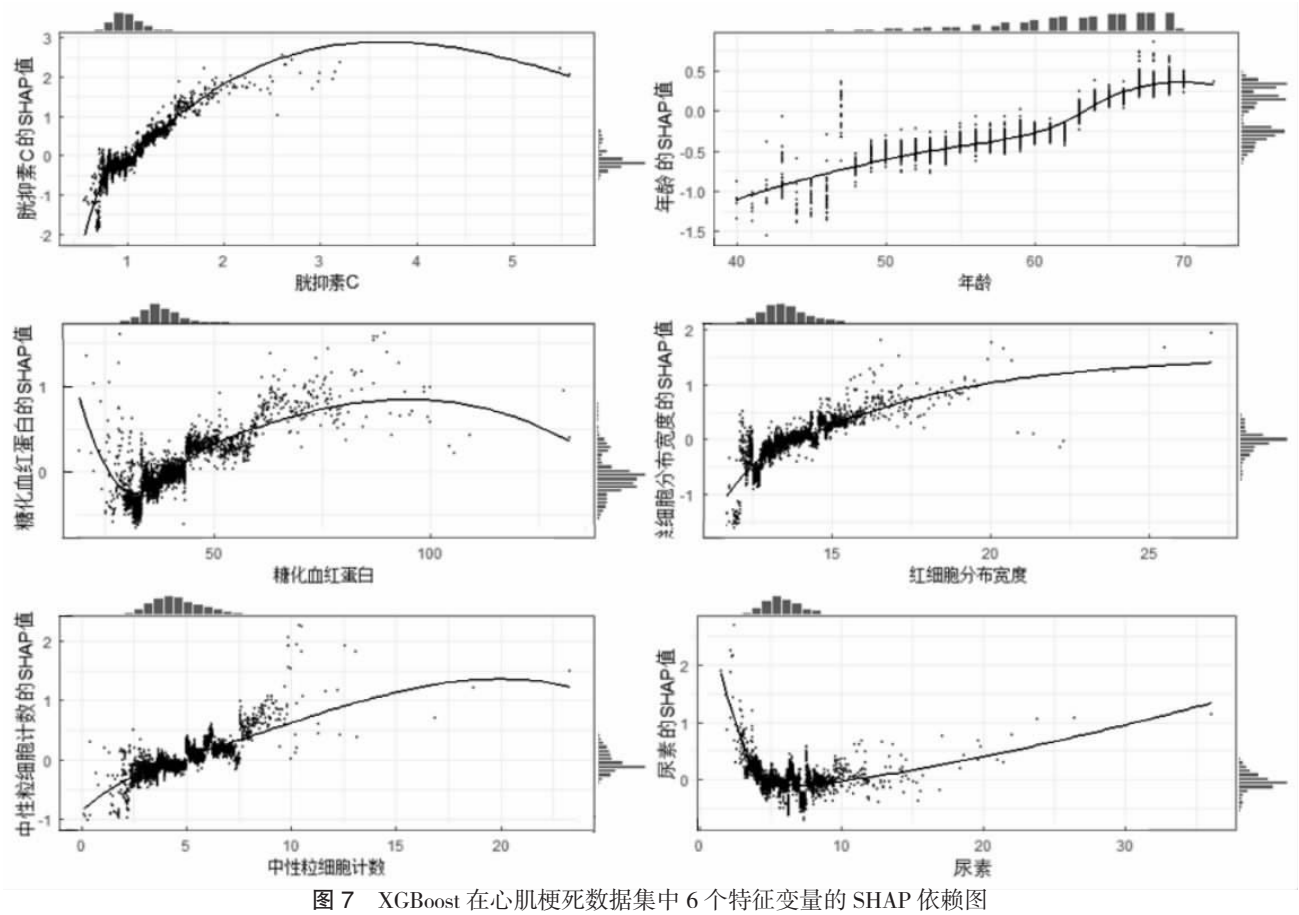


图 7　XGBoost 在心肌梗死数据集中 6 个特征变量的 SHAP 依赖图

Figure 7　SHAP dependence plots of six feature variables in the myocardial infarction dataset using XGBoost

框架虽涵盖短期和长期死亡预测，但其核心在于比较特征标记器变换器、XGBoost 和多层感知机的性能差异，对变量筛选策略的探讨相对有限。Fourkiotis 等[18]系统比较了 12 种模型在心梗后一年死亡预测中的表现，但模型依赖多项临床评估指标，在变量可及性方面存在一定局限。

与上述同类机器学习研究相比，本研究的差异化价值体现在以下方面：在研究定位上，多数现有模型聚焦于院内死亡或短期预后，本研究针对中位随访 14.4 年的长期全因死亡风险；在方法学上，本研究采用 Cox、LASSO、VIMP 三阶段互补筛选策略；在变量选择上，本研究在预测模型上仅采用 6 项常规血液指标，成本低、易获取；在验证设计上，本研究采用训练集、测试集及独立重复评估队列三重评估体系。本研究 XGBoost 模型的预测性能（AUC 约 0.70～0.82）与既往研究报道水平基本一致[19]，并表现出较好的稳定性。本研究虽未观察到更高的 AUC 水平，可能与所纳入变量主要来源于常规临床血液指标、缺乏影像学或遗传信息，以及研究人群整体事件发生率较低有关。校准曲线与决策曲线进一步表明模型具有良好的校准度和临床实用性。与既往研究相比，本研究不仅关注预测性能，还整合了反映肾功能、代谢状态、炎症反应及血液学特征等多维临床指标，从多个生理维度对患者风险进行综合评估。

本研究采用多阶段特征筛选方法，筛选出的关键变量均具有明确的临床依据。胱抑素 C、糖化血红蛋白、尿素、红细胞分布宽度和中性粒细胞计数与肾功能、炎症及红细胞异质性密切相关，已有研究证明其与心血管疾病不良预后相关。胱抑素 C 升高与动脉粥样硬化及全因死亡风险增加相关[20]；糖化血红蛋白反映长期血糖控制及代谢状态异常；尿素水平升高是心血管预后的独立危险因素；红细胞分布宽度上升反映红细胞稳态失调，是心血管疾病不良预后的独立预测因子[21]；较高的中性粒细胞计数与动脉粥样硬化性的发展密切相关[22]。此外，年龄作为重要人口学因素，与心血管疾病风险呈正相关[23]。

心肌梗死患者基数庞大，长期死亡风险及疾病负担较重，建立基于常规临床指标的风险预测模型，有助于在人群层面实现高危人群的早期识别和分层管理。本研究筛选出的关键指标均来源于临床常规检测项目，具有获取便捷、成本较低及基层可推广性较强等特点，可为资源有限地区开展长期随访和慢病管理提供参考。与此同时，机器学习方法能够整合多维临床信息，提高对高危患者的识别能力，有助于推动

心血管疾病防控由被动治疗向主动预警与精准干预转变，从而降低疾病负担并优化医疗资源配置。本研究模型可用于个体化风险评估，未来可基于该模型开发在线风险计算器或嵌入电子病历系统，实现自动化风险分层。例如，在患者入院或随访过程中自动提取相关血液指标并实时输出死亡风险概率，从而识别高风险人群并辅助临床制定个体化干预策略。但在实际应用前仍需关注指标检测标准化、模型外部验证及合理使用边界等问题。

本研究仍存在一些局限性，首先，本研究人群代表性上，样本性别失衡，且仅纳入基线已确诊心梗患者，排除了新发病例及急性期死亡者，可能存在幸存者偏倚。上述因素共同导致模型在女性及新发心梗患者中的外推性受限；同时，数据均来源于 UK Biobank，因人群谱系、环境暴露及临床实践的不同，在中国人群中的适用性有待独立验证。其次，本研究采用 VIMP 排序的“肘点法”筛选变量，该阈值（VIMP＞0.03）的选择具有一定探索性，缺乏正式的统计学检验标准；且仅采用 XGBoost 算法，未与其他生存预测模型进行系统性比较，XGBoost 的相对优势尚不明确。最后，由于内部验证集中 3 年和 5 年内全因死亡事件较少，导致短期 AUC 估计稳定性有限；而 7 年 AUC 达到 0.82，提示模型对远期死亡风险具有较好的识别能力。综上，未来仍需在中国多中心、大样本外部队列中进一步验证模型性能，并开展多种模型的比较研究。

利益冲突声明　本研究不存在任何利益冲突

参考文献

[1] Razavi SR, Szun T, Zaremba AC, et al. 1-Year mortality prediction through artificial intelligence using hemodynamic trace analysis among patients with ST elevation myocardial infarction [J]. Medicina (Kaunas, Lithuania), 2024, 60(4): 558.

[2] Zhang T, Deng W, Deng Y, et al.Mechanisms of ferroptosis regulating oxidative stress and energy metabolism in myocardial ischemia-reperfusion injury and a novel perspective of natural plant active ingredients for its treatment [J]. Biomedecine & Pharmacotherapie, 2023, 165: 114706.

[3] Lee K, Han S, Lee M, et al. Evidence-Based optimal medical therapy and mortality in patients with acute myocardial infarction after percutaneous coronary intervention[J]. Journal of the American Heart Association, 2023, 12(10): e024370.

[4] Liu XL, Wang LL, Wang Y, et al. Myocardial infarction complexity: A multi-omics approach [J]. Clinica Chimica Acta, 2024, 552: 117680.

[5] Frampton J, Ortengren AR, Zeitler EP. Arrhythmias after acute myocardial infarction [J]. The Yale Journal of Biology and Medicine, 2023, 96(1): 83-94.

[6] Boeddinghaus J, Doudesis D, Lopez-Ayala P, et al. Machine learning for myocardial infarction compared with Guideline-Recommended diagnostic pathways[J]. Circulation, 2024, 149(14): 1090-1101.

[7] Gergei I, Pfau T, Kramer BK, et al. Cardiovascular risk prediction-a systems medicine approach [J]. Journal of Hypertension, 2018, 36: e32.

[8] Li XW, Shang CM, Xu CY, et al. Development and comparison of machine learning-based models for predicting heart failure after acute myocardial infarction [J]. BMC Medical Informatics and Decision Making, 2023, 23(1): 165.

[9] 黄夏璇，黄韬，杨瑞，等. UK Biobank 数据的应用介绍[J]. 中国循证医学杂志，2022，22(9)：1099-1107.
Huang XX, Huang T, Yang R, et al. Introduction of UK biobank data application [J]. Chinese Journal of Evidence-Based Medicine, 2022, 22(9): 1099-1107.(In Chinese)

[10] Visioli F, Urbistondo DM, Gkipalis S, et al. Translational biomarkers for integrated cardiovascular disease risk assessment: A multidisciplinary review with applications in precision medicine[J]. Nutrition Metabolism and Cardiovascular Diseases, 2025, 35 (12): 104215.

[11] Zhao B, Nguyen VK, Xu M, et al. Random survival forest for predicting the combined effects of multiple physiological risk factors on all-cause mortality[J]. Scientific Reports, 2024, 14(1): 15566.

[12] Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model [J]. BMJ (Clinical Research ed.), 2020, 368: m441.

[13] Liu XH, Wang XJ, Xie RF, et al. Explainable machine learning for predicting mechanical ventilation in Alzheimer's patients with pneumonia: a SHAP-guided XGBoost nomogram based on MIMIC-IV[J]. Scientific Reports, 2026, 16(1): 17250.

[14] Qi XJ, Wang SJ, Fang CS, et al. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants [J]. Redox Biology, 2025, 79: 103470.

[15] 奚丽婧，郭昭艳，杨雪珂，等. LASSO 及其拓展方法在回归分析变量筛选中的应用 [J]. 中华预防医学杂志，2023，57(1)：107-111.
Xi LJ, Guo ZY, Yang XK, et al. Application of LASSO and its extended method in variable selection of regression analysis [J]. Chinese Journal of Preventive Medicine, 2023, 57 (1): 107-111.(In Chinese)

[16] Soleimani H, Najdaghi S, Davani DN, et al. Predicting in-hospital mortality in patients with acute myocardial infarction: a comparison of machine learning approaches[J]. Clinical Cardiology, 2025, 48(4): e70124.

[17] Kim M, Kang D, Kim MS, et al.Acute myocardial infarction prognosis prediction with reliable and interpretable artificial intelligence system [J]. Journal of the American Medical Informatics Association, 2024, 31(7): 1540-1550.

[18] Fourkiotis KP, Tsadiras A. Machine learning for mortality risk prediction in myocardial infarction: a Clinical-Economic decision support framework[J]. Applied Sciences-Basel, 2025, 15(16): 9192.

[19] Yang JG, Li YX, Li X, et al. A machine learning model for predicting In-Hospital mortality in Chinese patients with ST-Segment elevation myocardial infarction: findings from the China myocardial infarction registry[J]. Journal of Medical Internet Research, 2024, 26: e50067.

[20] Xiao ZB, Riletu AG, Yan XY, et al. Association of serum cystatin C

（下转第 3518 页）

2025, 12: 497–526.

[18] 李欣,戴伟荣,刘文峰,等. 尘肺病患者三年肺功能变化研究[J]. 中华劳动卫生职业病杂志,2020,38(12):891–894.
Li X, Dai WR, Liu WF, et al. Study on changes of pulmonary function in patients with pneumoconiosis in three years [J]. Chinese Journal of Industrial Hygiene and Occupational Diseases, 2020, 38 (12): 891–894.(In Chinese)

[19] Vecchi E, Kilbey T, Sheth M, et al. The 6 minute walk test as a predictor of mortality in idiopathic pulmonary fibrosis: a systematic review[J]. Respiratory Investigation, 2025, 63(3): 334–341.

[20] 葛建文,薛婷,李志敏,等. 基于运动训练的呼吸康复治疗尘肺病患者疗效 meta 分析 [J]. 中国职业医学,2024,51(3):285–291.
Ge JW, Xue T, Li ZM, et al. Meta–analysis on the effects of exercise training–based respiratory rehabilitation therapy in pneumoconiosis patients[J]. China Occupational Medicine, 2024, 51(3): 285–291.(In Chinese)

[21] Pehlivan E, Çetinkaya E, Karaahmetoğlu FS, et al. The impact of different pulmonary rehabilitation approaches on fibrotic interstitial lung disease: a comparative randomized trial [J]. Expert Review of Respiratory Medicine, 2025, 19(9): 1027–1035.

[22] Shen HH, Xu YM, Zhang Y, et al. Efficacy of pulmonary rehabilitation in patients with chronic obstructive pulmonary disease and obstructive sleep apnea; a randomized controlled trial[J]. Journal of Rehabilitation Medicine : Official Journal of the UEMS European Board of Physical and Rehabilitation Medicine, 2024, 56: jrm23757.

[23] Xu JL, Li X, Zeng J, et al. Effect of Baduanjin Qigong on postoperative pulmonary rehabilitation in patients with non–small cell lung cancer: a randomized controlled trial[J]. Supportive Care in Cancer, 2023, 32(1): 73.

[24] Shui LL, Cai JJ, Zhong XQ, et al. Chronic obstructive pulmonary disease patients with high peripheral blood eosinophil counts have better predicted improvement in 6MWD after rehabilitation: a preliminary study[J]. Journal of Cardiopulmonary Rehabilitation and Prevention, 2023, 43(2): 122–128.

[25] Fan YL, Ma RM, Du XQ, et al. Small airway dysfunction in pneumoconiosis: a cross–sectional study [J]. BMC Pulmonary Medicine, 2022, 22(1): 167.

[26] Kakavas S, Kotsiou OS, Perlikos F, et al. Pulmonary function testing in COPD: looking beyond the curtain of FEV1 [J]. NPJ Prim Care Respir Med, 2021, 31(1): 23.

[27] Aubrat P, Albert E, Perreaux M, et al. Pulmonary telerehabilitation in COPD patients:a systematic review to analyse patients'adherence [J]. Healthcare, 2025, 13(15): 1818.

[28] Liu S, Yang A, Yu Y, et al. Exercise prescription training in chronic obstructive pulmonary disease:benefits and mechanisms [J]. International Journal of Chronic Obstructive Pulmonary Disease, 2025, 20: 1071–1082.

[29] Çelik Z, Güzel NA, Allahverdiyeva S, et al. Effects of simultaneous aerobic and inspiratory muscle training on diaphragm function, respiratory muscle strength, endurance, and fatigue index: randomized–controlled trial [J]. European Journal of Applied Physiology, 2025, 125(12): 3769–3783.

[30] Vanka KS, Shukla S, Gomez HM, et al. Understanding the pathogenesis of occupational coal and silica dust–associated lung disease [J]. European Respiratory Review: an Official Journal of the European Respiratory Society, 2022, 31(165): 210250.

[31] Matyga AW, Chelala L, Chung JH. Occupational lung diseases: spectrum of common imaging manifestations [J]. Korean Journal of Radiology : Official Journal of the Korean Radiological Society, 2023, 24(8): 795–806.

[32] Yang BJ, Liu XM, Peng C, et al. Silicosis: from pathogenesis to therapeutics[J]. Frontiers in Pharmacology, 2025, 16: 1516200.

[33] Jenkins I, Takubo T, Lernhardt W, et al. Coal workers' pneumoconiosis in the 21st century: emerging trends, treatment Gaps, and prognostic uncertainties in a resurgent occupational lung disease[J]. Am J Biomed Sci Res, 2026, 31(3): 329–336.

[34] De soomer K, Geubels K, Mendoza H, et al. New reference values for the 6–minute walk distance in a European population across the full adult age range[J]. Respiratory Medicine, 2025, 245: 108205.

[35] Moy ML.Maintenance pulmonary rehabilitation: an update and future directions[J]. Respiratory Care, 2024, 69(6): 724–739.

[36] Al CY, Bailly S, Amroussia I, et al. Trajectories of COPD patients' response to repeated pulmonary rehabilitation programs [J]. Respiratory Medicine, 2021, 190: 106678.

[37] Al CY, Bailly S, Veale D, et al. Predictors of changes in 6–min walking distance following pulmonary rehabilitation in COPD patients: a retrospective cohort analysis [J]. European Journal of Physical and Rehabilitation Medicine, 2022, 58(2): 251–257.

[38] Brunetti G, Malovini A, Maniscalco M, et al.Pulmonary rehabilitation in patients with interstitial lung diseases:Correlates of success [J]. Respiratory Medicine, 2021, 185: 106473.

收稿日期:2026–04–23

(上接第 3484 页)

level and major adverse cardiovascular events in patients with percutaneous coronary intervention[J]. Cardiovascular Diagnosis and Therapy, 2024, 14(4): 621–629.

[21] Vayá A, Rivera L, De LER, et al. Red blood cell distribution width and erythrocyte deformability in patients with acute myocardial infarction [J]. Clinical Hemorheology and Microcirculation, 2015, 59 (2): 107–114.

[22] Luo J, Thomassen JQ, Nordestgaard BG, et al. Neutrophil counts and cardiovascular disease [J]. European Heart Journal, 2023, 44(47): 4953–4964.

[23] Dzubur A, Gacic E, Mekic M. Comparison of patients with acute myocardial infarction according to age [J]. Med Arch, 2019, 73(1): 23–27.

收稿日期:2026–06–03