

基于可解释机器学习的后路腰椎椎体间融合术后慢性疼痛风险预测模型构建

王建民^{1,2}, 李晓峰^{1,2}, 由志涛^{1,2}, 董圣杰^{2,3}, 赵宇驰^{2,3},

李占菊⁴, 邹德鑫^{1,2}, 张剑锋^{1,2}, 孙涛², 杜伟^{1,2}

(滨州医学院附属烟台山医院 1.脊柱外科; 2.烟台市骨与关节修复重建重点实验室;

3.关节外科; 4.信息管理科, 山东 烟台 264003)

摘要: **目的** 基于机器学习构建高精度、可解释性好的后路腰椎椎体间融合术后慢性疼痛 (chronic post-surgical pain, CPSP) 风险预测模型, 为临床早期识别高危人群、实现精准预防提供可靠工具。 **方法** 回顾性分析 2019 年 1 月至 2023 年 12 月在我院行后路腰椎椎体间融合术患者 759 例, 其中男 375 例, 女 384 例, 33~80 (55.28±9.94) 岁。按 7:3 比例分层随机分为训练集 ($n=531$) 和测试集 ($n=228$)。收集所有患者术前、术中、术后共 40 项特征数据。行数据预处理和 LASSO 回归特征筛选后, 构建 7 种机器学习模型, 以受试者工作特征曲线下面积 (area under the curve, AUC)、F1 分数等作为核心指标筛选模型, 使用沙普利加法解释 (Shapley additive explanations, SHAP) 工具对其进行可解释性分析。 **结果** 最终筛选出 CPSP 核心预测特征 10 项, 其中疼痛灾难化量表 (pain catastrophizing scale, PCS)、术前手术部位疼痛、并发症为三大核心驱动因素 (累计贡献 48.21%)。朴素贝叶斯 (AUC=0.914)、逻辑回归 (AUC=0.913) 模型表现出较好的模型效果, 朴素贝叶斯特异度达 0.958, 逻辑回归综合性能更均衡 (F1 分数=0.685)。SHAP 分析明确了各特征对于预测结果的影响作用方向及强度, 揭示 PCS 评分超过 30 分时, CPSP 风险显著升高的阈值效应。 **结论** 利用解释性强的可解释机器学习算法所构建的后路腰椎椎体间融合术后 CP-SP 风险预测模型, 具有良好的判别效能和临床可解释性, 且核心预测因子有助于指导临床个性化防控策略的开展, 推动脊柱外科从“经验医学”向“数据驱动精准预防”转变。

关键词: 后路腰椎椎体间融合术; 术后慢性疼痛; 机器学习; 风险预测模型; 可解释人工智能; 算法比较

中图分类号: R687.3

文献标志码: A

Construction of a chronic post-surgical pain prediction model for posterior lumbar interbody fusion surgery based on interpretable machine learning

WANG Jianmin^{1,2}, LI Xiaofeng^{1,2}, YOU Zhitao^{1,2}, DONG Shengjie^{2,3}, ZHAO Yuchi^{2,3}, LI Zhanju⁴, ZOU Dexin^{1,2}, ZHANG Jianfeng^{1,2}, SUN Tao², DU Wei^{1,2}

(1. Spinal Surgery; 2. Yantai Key Laboratory for Repair and Reconstruction of Bone & Joint; 3. Joint Surgery; 4. Information Management Department, Yantai Shan Hospital Affiliated to Binzhou Medical University, Yantai 264003, Shandong, China)

Abstract: Objective To construct a high-precision and well-interpretable risk prediction model for chronic post-surgical pain (CPSP) after posterior lumbar interbody fusion (PLIF), so as to provide a reliable tool for the early identification of high-risk populations and precise prevention in clinical practice. **Methods** A retrospective study was conducted on 759 patients who underwent PLIF at our hospital from January 2019 to December 2023, including 375 males and 384 females. The patients were 33-80 (55.28±9.94) years old. All cases were stratified and randomly divided into a training

set ($n=531$) and a testing set ($n=228$) at a ratio of 7:3. A total of 40 characteristic variables were collected from all patients before, during, and after surgery. After data pre-processing and LASSO regression feature screening, seven machine learning models were constructed, with area under the curve (AUC), F1 score, and other indicators as the core indicator to select the optimal model, and the Shapley additive explanations (SHAP) tool was used for interpretability analysis. **Results** Ten core predictive features of CPSP were ultimately selected, among which pain catastrophizing scale (PCS) score, preoperative surgical site pain, and complications were the three core driving factors (cumulative contribution: 48.21%). The Naive Bayes (AUC=0.914) and logistic regression (AUC=0.913) models exhibited excellent performance, and logistic regression had a more balanced overall performance (F1 score=0.685). The Naive Bayes specificity was as high as 0.958. SHAP analysis clarified the direction and magnitude of the influence of each feature on the prediction results, revealing the threshold effect of significantly increasing CPSP risk when the PCS score exceeds 30 points. **Conclusion** The CPSP risk prediction model based on machine learning has good discriminative power and clinical interpretability. The core predictive factors provide clear targets for personalized prevention and control strategies in clinical practice, which helps to promote the transformation of spinal surgery from “empirical medicine” to “data-driven precision prevention”.

Key words: Posterior lumbar interbody fusion; Chronic post-surgical pain; Machine learning; Risk prediction model; Explainable artificial intelligence; Algorithm comparison

后路腰椎椎体间融合术是治疗腰椎间盘突出症、腰椎管狭窄症、腰椎滑脱等疾病的常用外科治疗方法,能有效地缓解神经压迫、恢复脊柱稳定性^[1-3]。但是,术后慢性疼痛(chronic post-surgical pain, CPSP)是其常见并发症之一,文献报道发生率为8%~40%^[4-5]。CPSP指术后持续存在或新发、持续时间超过3个月的非神经病理性疼痛,严重影响患者的生活质量,造成功能障碍、心理压力增加及医疗资源消耗增加,已成为脊柱外科领域亟待解决的重点难点^[6-7]。

CPSP发生机制复杂,是术前、术中和术后多种因素综合作用的结果^[8-9]。以往大多数针对CPSP危险因素的研究使用单变量或多变量logistic回归进行探索,无法探究变量间的非线性和相互作用等,预测能力较差^[10-12]。近年来,机器学习技术在医疗预测方面展现出较大的潜力,可以从多维度、复杂的临床数据中学习和挖掘规律,建立起高质量预测模型^[13]。然而,目前基于机器学习构建脊柱术后CPSP预测模型尚存在一定的不足,主要包括:多模态数据融合不足,尚缺乏对心理、社会等关键维度的整合;预测模型难以解释,常被称为“黑箱”,难以转化为临床医生可信赖的决策工具^[14];而且针对后路腰椎椎体间融合术这一特定术式的CPSP,目前尚未系统性地特征工程、多算法比较及可解释性分析的研究。

本研究旨在利用机器学习技术将心理、生理、社会等多维数据融合起来,构建一个可解释的后路腰

椎椎体间融合术术后CPSP风险预测模型,为临床早期识别高危人群提供可靠工具,实现从“经验医学”向“数据驱动精准预防”的跨越。

1 资料与方法

1.1 临床资料

1.1.1 研究对象

回顾性纳入2019年1月至2023年12月于烟台市烟台山医院行后路腰椎椎体间融合术患者759例,其中男375例,女384例,33~80(55.28 ± 9.94)岁。本研究已通过烟台市烟台山医院医学伦理委员会批准(烟台伦准2025209号)。所有患者及其家属均签署知情同意书。

1.1.2 纳入标准与排除标准

纳入标准:①年龄 ≥ 30 岁;②术式为单节段/多节段后路腰椎椎体间融合术;③术后随访 ≥ 6 个月;④临床资料完整(含术前心理评估、术中记录、术后疼痛随访);⑤所有患者均签署知情同意书,自愿参与本研究。排除标准:①因肿瘤、感染、外伤等行手术治疗的患者;②行翻修手术的患者;③合并有神经系统疾病(如脊髓损伤、帕金森病)的患者;④伴有严重认知功能障碍、精神疾病等症状,不能配合完成评分表的患者。

1.2 方法

1.2.1 数据采集

参照电子病历资料及既往研究^[15-18],设计标

准化数据采集表,由2名经过培训的研究人员独立录入数据,并采用双人核对法确保数据准确性,不一致处通过查阅电子病历与手术记录解决。

采集变量涵盖术前、术中、术后3个时间段,共40项。主要包括:疼痛灾难化量表(pain catastrophizing scale, PCS)、医院焦虑抑郁量表(hospital anxiety and depression scale, HADS)、术后活动视觉模拟评分法(visual analogue scale, VAS)、术前手术部位疼痛、并发症、术后不良反应、融合椎体数、吸烟史、糖尿病、高血压等。

主要量表的评分标准:①疼痛灾难化量表总分范围0~52分,分数越高表明疼痛灾难化倾向越严重。本研究将评分<30分定义为“低”分组,≥30分定义为“高”分组。②医院焦虑抑郁量表用于评估焦虑抑郁状态,分为4个等级:无(总分<8分)、轻度(8~10分)、中度(11~14分)、重度(15~21分)。

结局指标:术后疼痛持续≥3个月,且排除感染、螺钉松动等明确病因;术后6个月随访时通过电话或门诊问卷确认CPSP发生情况^[19]。

1.2.2 数据集划分与数据预处理

按7:3比例分层抽样将数据集划分训练集($n=531$)和测试集($n=228$),确保两组CPSP发生率一致,设置随机种子保证结果可复现。采用Python 3.9相关库完成数据预处理。先计算各特征缺失率,把缺失率>20%的特征直接删除,≤20%的特征按类型填补:分类变量采用众数填补,连续变量采用中位数填补,有序分类变量采用线性插值填补。对于连续变量,通过IQR法定义异常值范围,将异常值截断至边界值。对连续变量进行Z-score标准化,分类变量采用One-Hot编码,有序分类变量采用序数编码,最后合并成标准化的数据集。所有数据预处理操作均在训练集上拟合参数,然后将相同的参数应用到测试集。

1.2.3 特征工程

基于方差膨胀因子(variance inflation factor, VIF)分析连续变量和有序分类变量的共线性情况,删除VIF>10的变量。应用最小绝对收缩与选择算子(least absolute shrinkage and selection operator, LASSO)回归进行降维,利用5折交叉验证确定最优正则化参数,筛选非零系数特征,最终保留核心特征用于建模。

1.2.4 类别不平衡处理

训练集存在类别不平衡,采用合成少数类过采样技术(synthetic minority over-sampling technique, SMOTE)算法生成少数类的合成样本,调整样本比例;测试集保持原始分布,避免过拟合。

1.2.5 模型构建与优化

构建支持向量机(support vector machine, SVM)、决策树、随机森林、逻辑回归、朴素贝叶斯、极限梯度提升(extreme gradient boosting, XGBoost)、轻量的梯度提升机(light gradient boosting machine, LightGBM)等7种机器学习模型并设置初始参数。采用网格搜索结合10折分层交叉验证优化各模型参数,以受试者工作特征曲线下面积(area under the curve, AUC)为核心评价指标。在测试集上计算AUC、准确率、灵敏度、特异度、精确度、F1分数等性能指标,绘制受试者工作特征曲线(receiver operating characteristic, ROC),筛选最优模型。采用Bootstrap法(重复抽样1000次)计算关键性能指标的95%置信区间。

1.2.6 模型可解释性分析

采用沙普利加法解释(Shapley additive explanations, SHAP)工具解析最优模型,基于测试集计算SHAP值。根据各特征的SHAP值均值排序特征,明确核心风险因子。选取典型样本(如CPSP高风险、SHAP贡献值构成清晰、代表性强的患者),可视化单个样本中不同特征对CPSP发生风险的具体贡献度。

1.3 统计学处理

采用SPSS 26.0统计学软件和Python Scikit-learn库进行模型评估。计量资料符合正态分布以 $\bar{x}\pm s$ 表示,组间比较采用独立样本 t 检验;计数资料以 $n(\%)$ 表示,组间比较采用 χ^2 检验或Fisher确切概率法。检验水准 $\alpha=0.05$ (双侧)。

2 结果

2.1 临床基本资料与CPSP发生情况

所有患者均完成术后6个月随访,术后发生CPSP 204例,发生率26.88%。训练集和测试集患者在人口学特征、基础疾病、手术相关指标及术后恢复情况等方面均具有较好的均衡性。见表1,图1。

表1 训练集与测试集患者基线特征比较
Table 1 Comparison of baseline characteristics between training and testing sets

项目	训练集($n=531$)	测试集($n=228$)	χ^2/t	P
年龄/岁	55.32±10.00	55.18±9.82	0.181	0.857
体质量/kg	65.30±11.57	65.59±11.19	-0.323	0.747
性别			0.176	0.675
女	266(50.09)	118(51.75)		
男	265(49.91)	110(48.25)		
吸烟			0.001	0.978
无	301(56.69)	129(56.58)		
有	230(43.31)	99(43.42)		
饮酒			0.028	0.867
无	255(48.02)	111(48.68)		
有	276(51.98)	117(51.32)		
就业状态			2.432	0.296
在职	273(51.41)	120(52.63)		
无业/自营	170(32.02)	62(27.19)		
退休	88(16.57)	46(20.18)		
家庭人均月收入			0.972	0.808
<1 000 元	32(6.03)	18(7.89)		
1 000~2 999 元	408(76.84)	172(75.44)		
3 000~4 999 元	69(12.99)	28(12.28)		
≥5 000 元	22(4.14)	10(4.39)		
婚姻状态			0.521	0.771
未婚	76(14.31)	29(12.72)		
已婚	370(69.68)	159(69.74)		
离异/丧偶	85(16.01)	40(17.54)		
文化程度			3.418	0.181
小学及以下	112(21.09)	37(16.23)		
中学	252(47.46)	123(53.95)		
大专及以上	167(31.45)	68(29.82)		
高血压			0.052	0.819
无	305(57.44)	133(58.33)		
有	226(42.56)	95(41.67)		
糖尿病			1.341	0.247
无	338(63.65)	135(59.21)		
有	193(36.35)	93(40.79)		
Frankel 分级			4.478	0.345
A 级	25(4.71)	12(5.26)		
B 级	47(8.85)	28(12.28)		
C 级	66(12.43)	33(14.47)		
D 级	128(24.11)	43(18.86)		
E 级	265(49.91)	112(49.12)		
疼痛灾难化量表			0.728	0.056
低	280(52.73)	138(60.53)		
高	251(47.27)	90(39.47)		
焦虑抑郁评估			1.706	0.636
无	275(51.79)	113(49.56)		
轻度	145(27.31)	71(31.14)		
中度	96(18.08)	36(15.79)		
重度	15(2.82)	8(3.51)		
手术时长/min	160.90±23.16	161.24±23.86	-0.182	0.856

注:变量以 $n(\%)$ 或 $\bar{x}\pm s$ 形式表示。

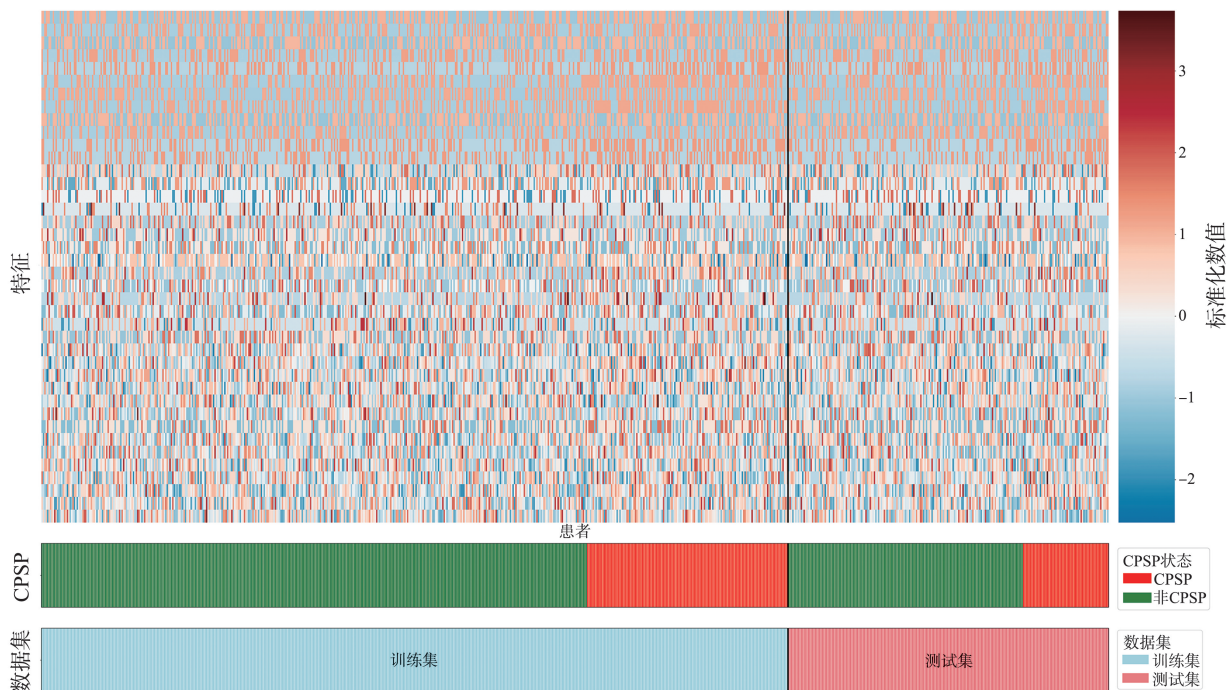


图1 后路腰椎椎体间融合术患者特征
Figure 1 Characteristic of patients undergoing posterior lumbar interbody fusion surgery

2.2 特征工程结果

采用 LASSO 回归进行特征降维,通过交叉验证确定最优正则化参数,绘制 LASSO 正则化路径图显示(图 2,3):随着正则化强度增大,部分特征系数逐渐收缩至零,最终筛选出 10 个非零系数的

核心特征。按 LASSO 系数绝对值从大到小依次为:疼痛灾难化量表、术前手术部位疼痛、并发症、术后不良反应、医院焦虑抑郁量表、融合椎体数、吸烟史、糖尿病、术后活动视觉模拟评分法、高血压(图 4)。

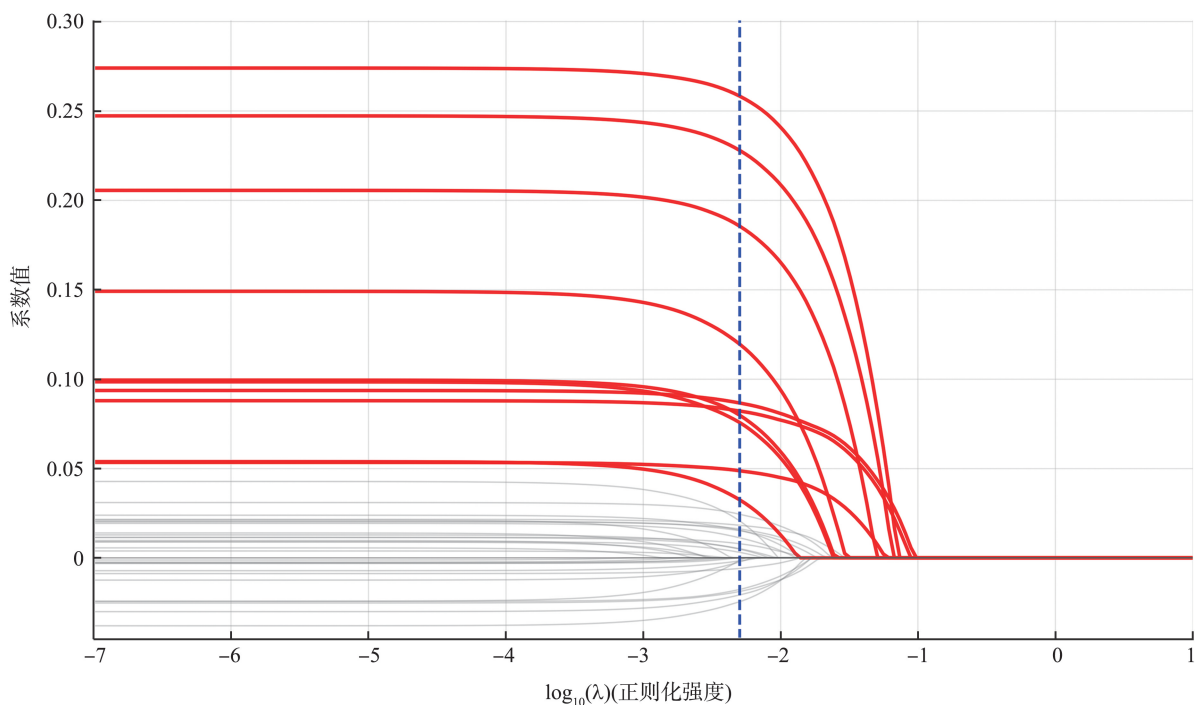


图2 LASSO 回归的系数路径
Figure 2 Coefficient path of LASSO regression

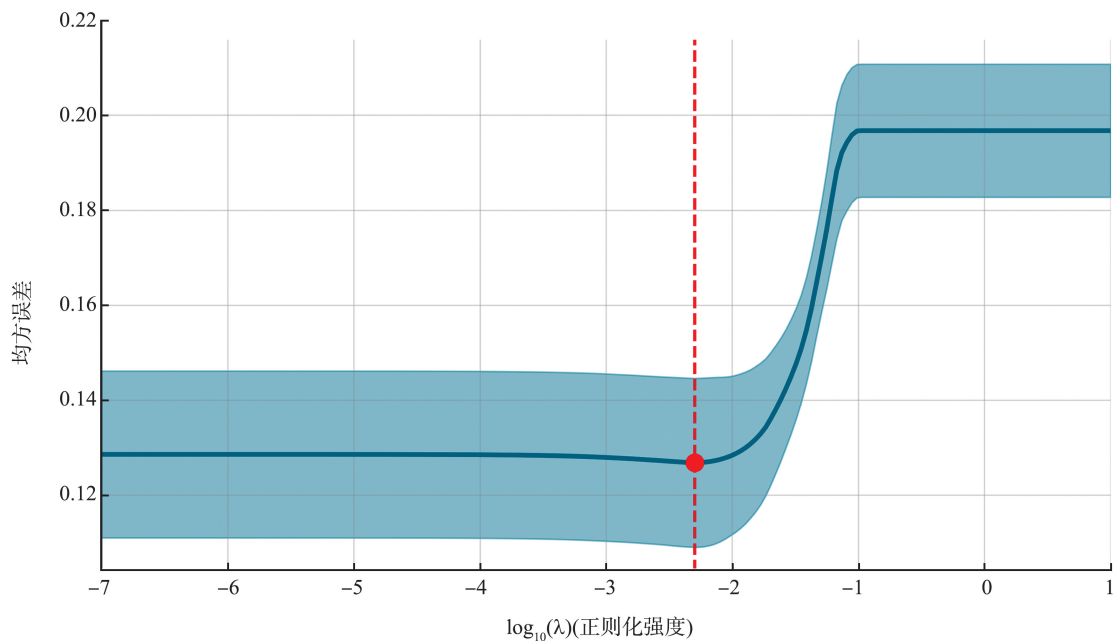


图3 LASSO 回归交叉验证
Figure 3 LASSO regression cross validation

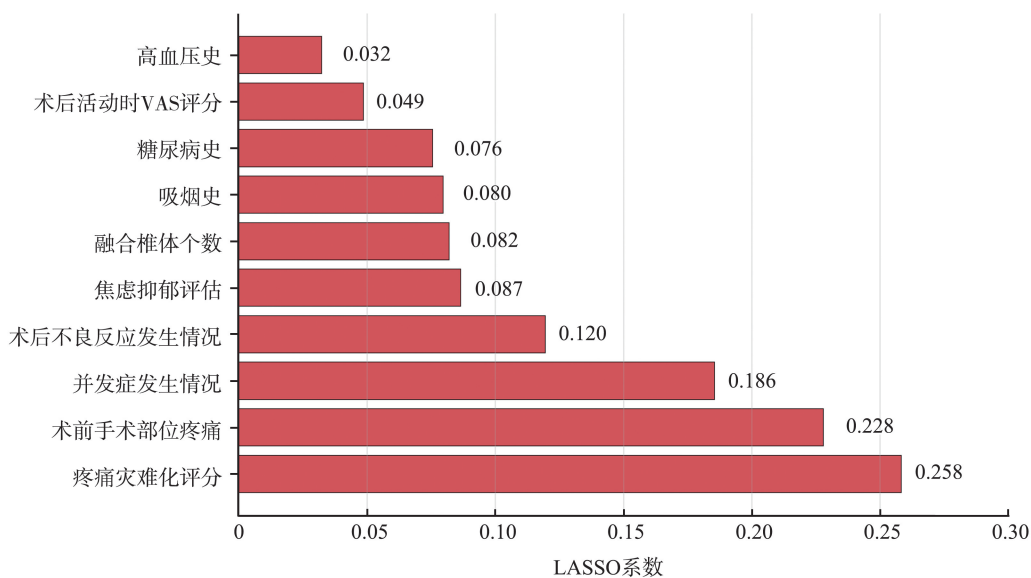


图4 LASSO 回归筛选的 CPSP 预测核心特征及其系数分布
Figure 4 The core features and coefficient distribution of CPSP prediction selected by LASSO regression

2.3 不同模型的预测性能比较

7种模型优化后的测试集性能指标如表2,图5所示。测试集结果表明,朴素贝叶斯表现突出,AUC为0.914(95%CI:0.875~0.949),准确率0.855、特异度0.958,灵敏度0.574偏低。逻辑回归综合性能均衡,AUC为0.913(95%CI:0.873~0.949),准确率0.851、F1分数0.685为模型中最高。随机森林、LightGBM、XGBoost性能中等,决策树AUC为0.772,综合性能较差。支持向量机多项指标异常,推测分类阈值或训练过程存

在问题。经过拟合风险评估,逻辑回归和朴素贝叶斯的泛化能力强。与之相比,支持向量机因为把所有样本都预测成了阴性而没有达到任何效果;决策树模型过拟合严重,测试集AUC只有0.772。为进一步评估预测概率是否可靠,计算每个模型的Brier分数并绘制校准曲线(图6)。7种机器学习模型的Brier分数范围为0.106~0.177,其中逻辑回归(0.106)、朴素贝叶斯(0.109)、支持向量机(0.111)等6种模型的Brier分数<0.150。

表2 机器学习模型优化后的测试集性能比较
Table 2 Performance comparison of machine learning model optimized test set

模型名称	准确率	灵敏度	特异度	精确度	F1 分数	AUC
朴素贝叶斯	0.855 (0.803~0.899)	0.574 (0.446~0.694)	0.958 (0.924~0.988)	0.833 (0.705~0.943)	0.680 (0.563~0.770)	0.914 (0.875~0.949)
逻辑回归	0.851 (0.798~0.895)	0.607 (0.473~0.726)	0.940 (0.904~0.974)	0.787 (0.666~0.900)	0.685 (0.571~0.776)	0.913 (0.873~0.949)
支持向量机	0.732 (0.675~0.785)	0.000 (0.000~0.000)	1.000 (1.000~1.000)	0.000 (0.000~0.000)	0.000 (0.000~0.000)	0.909 (0.869~0.947)
随机森林	0.798 (0.741~0.846)	0.393 (0.260~0.516)	0.946 (0.909~0.977)	0.727 (0.556~0.875)	0.511 (0.368~0.625)	0.880 (0.833~0.921)
LightGBM	0.776 (0.715~0.825)	0.492 (0.359~0.610)	0.880 (0.830~0.926)	0.600 (0.472~0.733)	0.541 (0.415~0.638)	0.860 (0.812~0.907)
XGBoost	0.781 (0.724~0.829)	0.492 (0.359~0.610)	0.886 (0.839~0.932)	0.612 (0.478~0.746)	0.545 (0.420~0.644)	0.858 (0.809~0.904)
决策树	0.732 (0.671~0.789)	0.393 (0.260~0.516)	0.856 (0.802~0.906)	0.500 (0.352~0.640)	0.440 (0.311~0.552)	0.772 (0.704~0.837)

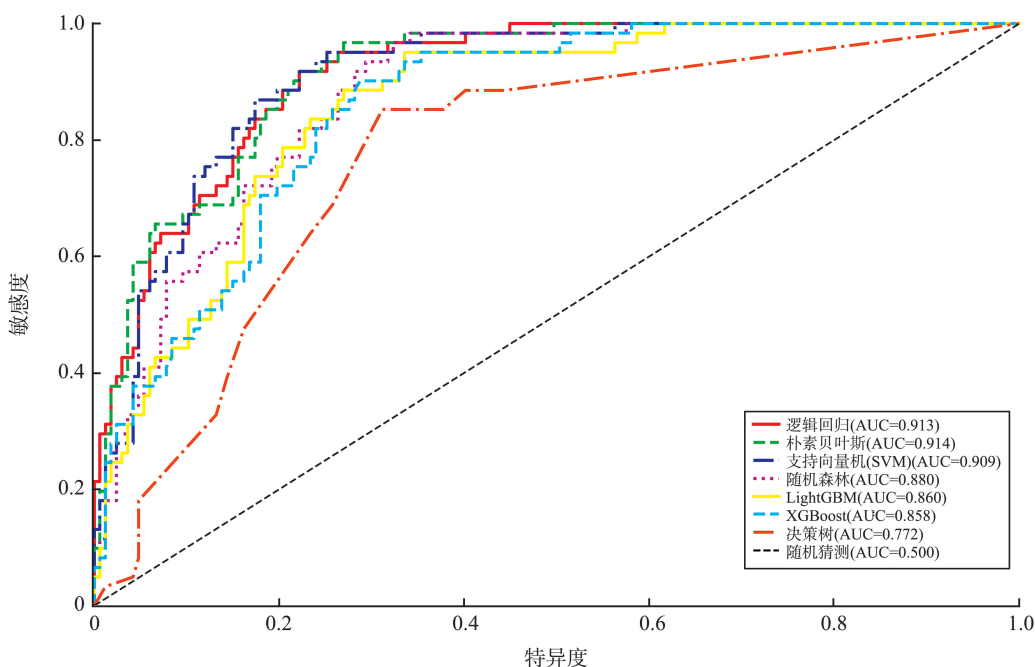


图5 机器学习模型在测试集上的 ROC 曲线对比

Figure 5 Comparison of ROC curves of machine learning models on the test set

2.4 基于 SHAP 的模型可解释性分析

SHAP 摘要图显示,术前疼痛灾难化评分、术前手术部位疼痛和并发症为三大核心特征,累计贡献占比达 48.21%,对预测结果起决定性作用,见图 7。SHAP 依赖图验证了疼痛灾难化评分的临床阈值效应,其标准化值超过 0.58(对应原始量表 30 分)时,

SHAP 值显著上升,提示重度疼痛灾难化患者的 CPSP 风险呈剂量依赖性增加,见图 8。高风险样本(如样本 80)的瀑布图进一步拆解预测过程,融合椎体个数(+1.02)、术前手术部位疼痛(+0.85)和糖尿病史(+0.71)为主要正向风险驱动因素,而疼痛灾难化评分虽未达重度阈值仍贡献-0.59,见图 9。

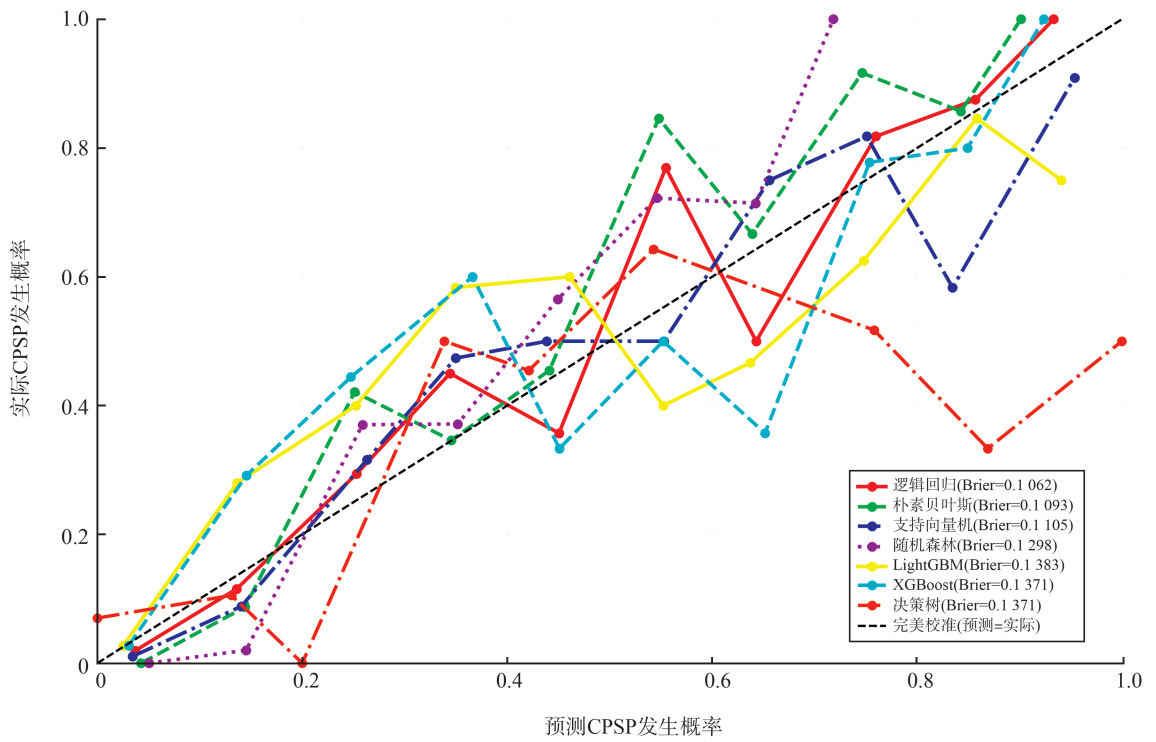


图6 所有模型校准曲线及 Brier 分数对比
Figure 6 Calibration curves and Brier score comparison for all models

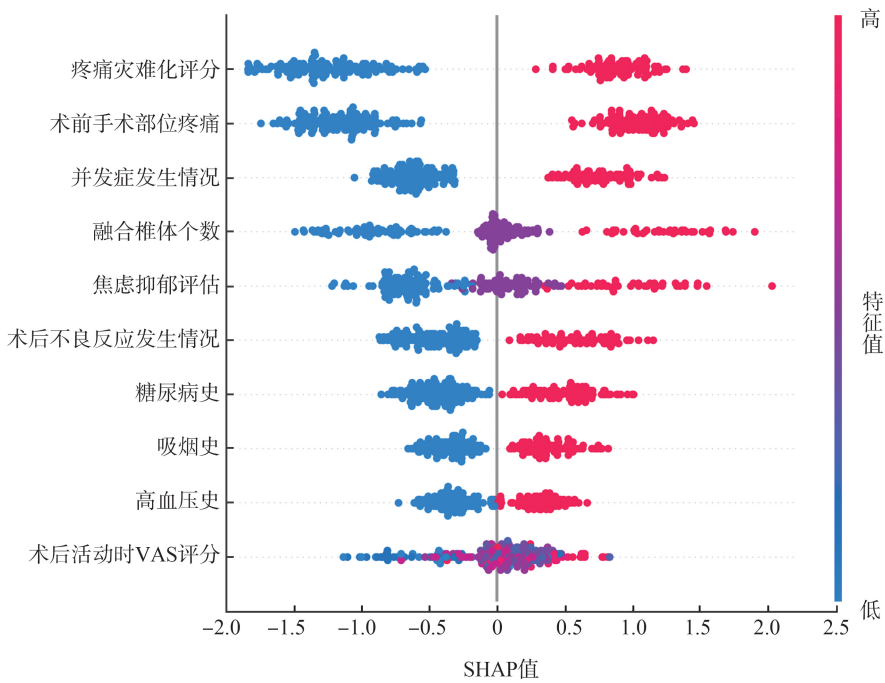


图7 SHAP 摘要图:特征重要性与影响方向
Figure 7 SHAP summary: feature importance and influence direction

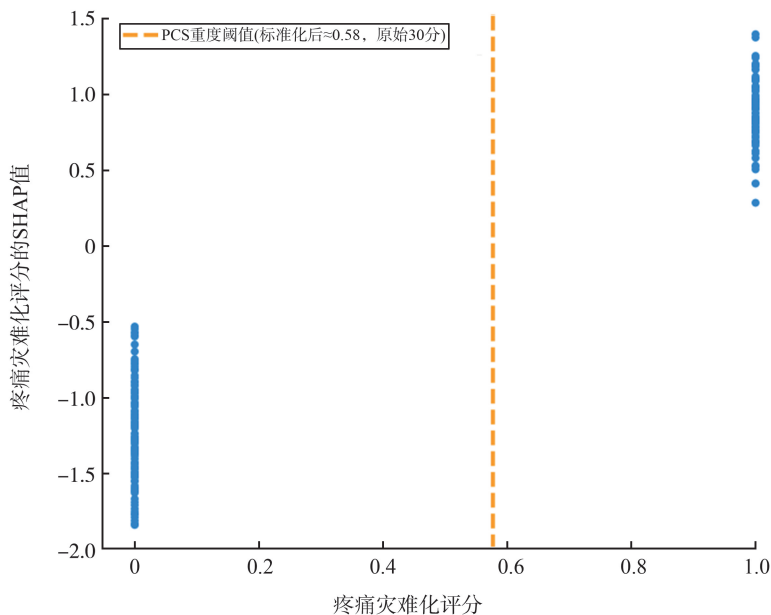


图8 SHAP 依赖图:疼痛灾难化评分与 CPSP 风险的剂量效应关系

Figure 8 SHAP dependency plot: dose-response relationship between PCS-score and CPSP risk

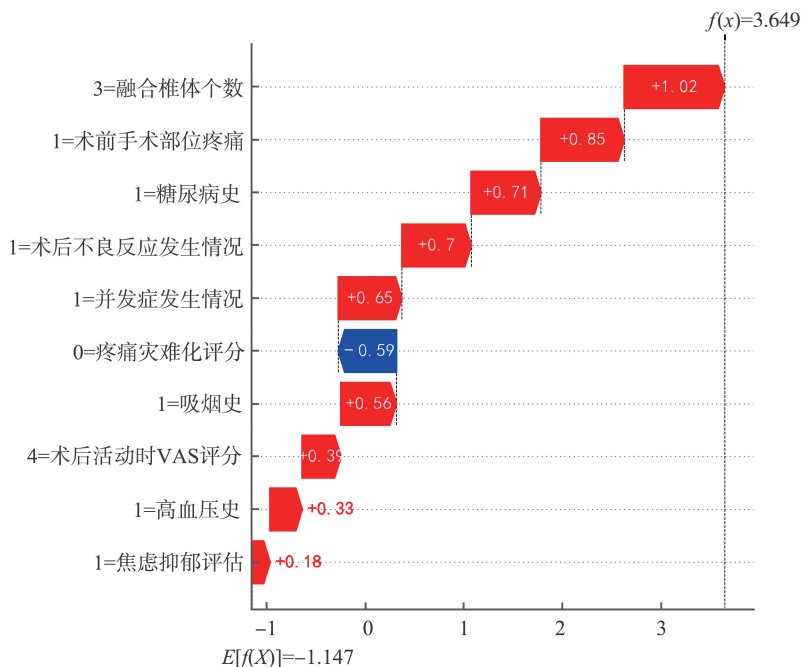


图9 SHAP 瀑布图:样本 80 的 CPSP 预测分解

Figure 9 SHAP waterfall plot: CPSP prediction decomposition of sample 80

3 讨论

后路腰椎椎体间融合术引起的 CPSP 是临床亟待解决的问题之一,现有预测模型多因效能有限或可解释性不足而难以推广^[20]。本研究聚焦该领域痛点,利用机器学习方法构建 CPSP 风险预测模型,不仅通过 LASSO 回归从 40 项初始特征中精准筛选出 10 项核心预测因子,更实现了逻辑回归与朴素贝

叶斯模型 AUC 均大于 0.91 的高性能表现。SHAP 分析结果提示,疼痛灾难化评分、术前手术部位疼痛及并发症是三大主要驱动因素,为临床精准识别高危患者、制定个性化防控策略提供了兼具高性能与可信度的工具。

在核心预测因子的筛选与解析方面,现有国内外研究多依赖单因素分析或简单多因素回归,对预测因子的重要性排序缺乏量化依据^[21-25]。如 Chen 等^[25]强调了心理因素在慢性疼痛中的作用,却并未

对其重要程度及阈值效应给出具体的论述。针对以上问题,本研究采用 LASSO 与 SHAP 结合的方法,在后路腰椎椎体间融合术患者中量化了核心预测因子的贡献度。尤为重要的是,疼痛灾难化评分显示出明确的临床阈值效应:当疼痛灾难化评分为 30 分以上时,CPSP 的风险会显著上升。基于此阈值,在临床实践中,对于疼痛灾难化评分 ≥ 30 分的患者,可考虑提前进行 1~2 次简短心理干预(比如:疼痛认知教育等),并在术后通过电话随访的形式来观察和巩固其认知改善效果,从而降低 CPSP 的发生风险。其次,研究还发现,术前手术部位疼痛与疼痛灾难化评分存在显著的交互效应,表明单纯镇痛或单纯心理干预效果有限,而围术期管理需采取“镇痛-心理”协同的整合模式。

在机器学习模型的构建与应用方面,国内外相关研究普遍存在“单模型依赖”的问题,即仅追求单一模型的性能最优,未考虑不同临床场景的差异化需求^[26-27]。本研究通过 7 种机器学习模型的对比研究发现:逻辑回归与朴素贝叶斯表现最优,其中逻辑回归的 F1 分数最高,综合判别能力最佳,可作为临床常规筛查的基准模型;朴素贝叶斯虽特异度较高,但灵敏度偏低,会漏诊部分高风险患者,更适用于对低危人群进行筛选判断。树模型由于训练难度大且容易过拟合,支持向量机与决策树因分类失效或过拟合均不适用于本研究的临床场景。综上所述,本研究建议采用“多模型优势互补”的临床应用策略:将逻辑回归作为主要预测模型,其均衡的性能与良好的可解释性更符合普适性临床需求;针对特殊情况下(如低风险人群筛查),可结合朴素贝叶斯进行辅助验证,实现敏感性与特异性的最佳平衡。这种多模型融合应用的思路,正是对临床实践中多样化需求的有效响应。

机器学习模型由于存在“黑箱”特性难以获得临床医生认可,且现有机器学习研究虽尝试通过可解释性工具解析 CPSP 预测模型,但仍局限于从宏观角度对重要的特征排序,并没有从“个体样本风险贡献”的微观层面对其解释,难以满足临床“一人一策”的需求^[28-29]。本研究通过 SHAP 框架实现“宏观-微观”双维度解释:从宏观的角度看,SHAP 分析通过计算特征的平均绝对 SHAP 值,明确了 10 个核心特征的贡献占比,证实三大核心驱动因素的累计贡献达 48.21%,为干预优先级提供量化依据;从微观的角度,根据 SHAP 瀑布图对单一样本的预测进行拆解,清晰展示了各特征对该样本风险评估的具体贡献。这种单样本水平的解释,使医生能够明确每位患者的高危原因,并据此制定个性化的干预方案,同时也增强了医生对模型预测结果的信任

度,为模型在临床中的推广应用奠定了基础。

本研究存在以下局限:①数据来源单一,单中心回顾性研究未开展外部验证及时间验证,可能存在一定的泛化能力局限性,未来需开展多中心前瞻性研究以验证模型泛化能力;②未纳入遗传与分子标志物,如 COMT 基因多态性、炎症因子水平等,可能遗漏部分生物学机制相关变量;③未开发临床工具,还未将模型转为可视化计算器或者嵌入至电子病历系统,尚未能应用于临床实践;④虽然 SMOTE 是解决不平衡数据问题常用的过采样方法,但是加入的样本由于是合成的,存在一定偏倚性,后续研究将会加入其他的不平衡学习算法(ADASYN)或者集成学习的方法进行验证。

综上所述,本研究针对后路腰椎椎体间融合术这一特定术式,明确了疼痛灾难化评分、术前手术部位疼痛和并发症等三大核心驱动因素,并量化其贡献。同时,本研究提出“多模型优势互补”策略并通过 SHAP 分析实现深度可解释。模型不仅为临床早期识别高危患者提供可靠工具,更推动脊柱外科 CPSP 管理从“经验判断”向“数据驱动精准预防”转型。

参考文献:

- [1] Wan ZY, Shan H, Liu TF, et al. Emerging issues questioning the current treatment strategies for lumbar disc herniation[J]. *Front Surg*, 2022, 9: 814531. doi:10.3389/fsurg.2022.814531
- [2] Mills ES, Shelby T, Bouz GJ, et al. A decreasing national trend in lumbar disc arthroplasty[J]. *Glob Spine J*, 2023, 13(8): 2271-2277.
- [3] Dantas F, Dantas FLR, Botelho RV. Effect of interbody fusion compared with posterolateral fusion on lumbar degenerative spondylolisthesis: a systematic review and meta-analysis[J]. *Spine J*, 2022, 22(5): 756-768.
- [4] Wu QC, Ford NC, He SQ, et al. Characterizing a new rat model of chronic pain after spine surgery[J]. *Bone Res*, 2025, 13(1): 34. doi:10.1038/s41413-025-00408-1
- [5] Carreon LY, Glassman SD, Howard J. Fusion and non-surgical treatment for symptomatic lumbar degenerative disease: a systematic review of Oswestry Disability Index and MOS Short Form-36 outcomes[J]. *Spine J*, 2008, 8(5): 747-755.
- [6] Wooden S. Chronic postsurgical pain[J]. *Crit Care Nurs Clin N Am*, 2024, 36(4): 585-595.
- [7] Rosenberger DC, Segelcke D, Pogatzki-Zahn EM. Mechanisms inherent in acute-to-chronic pain after surgery - risk, diagnostic, predictive, and prognostic factors[J]. *Curr Opin Support Palliat Care*, 2023, 17(4): 324-337.
- [8] Tassou A, Richebe P, Rivat C. Mechanisms of chronic postsurgical pain[J]. *Reg Anesth Pain Med*, 2025, 50(2): 77-85.

- [9] Pak DJ, Yong RJ, Kaye AD, et al. Chronification of pain; mechanisms, current understanding, and clinical implications[J]. *Curr Pain Headache Rep*, 2018, 22(2): 9. doi: 10.1007/s11916-018-0666-8
- [10] 陈群生, 夏炳江. 脊柱手术患者术后慢性疼痛的危险因素及对策分析[J]. *浙江创伤外科*, 2023, 28(11): 2058-2061.
CHEN Qunsheng, XIA Bingjiang. Risk factors and countermeasures of chronic pain after spinal surgery[J]. *Zhejiang Journal of Traumatic Surgery*, 2023, 28(11): 2058-2061.
- [11] Schober P, Vetter TR. Logistic regression in medical research[J]. *Anesth Analg*, 2021, 132(2): 365-366.
- [12] Bayman EO, Dexter F. Multicollinearity in logistic regression models[J]. *Anesth Analg*, 2021, 133(2): 362-365.
- [13] Wani SUD, Khan NA, Thakur G, et al. Utilization of artificial intelligence in disease prevention; diagnosis, treatment, and implications for the healthcare workforce[J]. *Healthcare (Basel)*, 2022, 10(4): 608. doi:10.3390/healthcare10040608
- [14] Karim MR, Islam T, Shajalal M, et al. Explainable AI for bioinformatics; methods, tools and applications[J]. *Brief Bioinform*, 2023, 24(5): bbad236. doi:10.1093/bib/bbad236
- [15] Rathbone J, Rackham M, Nielsen D, et al. A systematic review of anterior lumbar interbody fusion (ALIF) versus posterior lumbar interbody fusion (PLIF), transforaminal lumbar interbody fusion (TLIF), posterolateral lumbar fusion (PLF)[J]. *Eur Spine J*, 2023, 32(6): 1911-1926.
- [16] Shen N, Zhang YS, Ding ZD, et al. Predicting intraoperative blood loss risk in severe lumbar disc herniation patients undergoing PLIF: a multicenter cohort study using ensemble learning[J]. *Int J Surg*, 2025, 111(9): 5904-5913.
- [17] Mehren C, Ostendorff N, Schmeiser G, et al. Do TLIF and PLIF techniques differ in perioperative complications? -comparison of complications rates of two high volume centers[J]. *Glob Spine J*, 2025, 15(1): 84-93.
- [18] Amaral R, Pokorny G, Marcelino F, et al. Lateral versus posterior approaches to treat degenerative lumbar pathologies-systematic review and meta-analysis of recent literature[J]. *Eur Spine J*, 2023, 32(5): 1655-1677.
- [19] Schug SA, Lavand'homme P, Barke A, et al. The IASP classification of chronic pain for ICD-11: chronic postsurgical or posttraumatic pain[J]. *Pain*, 2019, 160(1): 45-52.
- [20] Martinez V, Lehman T, Lavand'homme P, et al. Chronic postsurgical pain: a european survey[J]. *Eur J Anaesthesiol*, 2024, 41(5): 351-362.
- [21] 许颖, 何龙, 王洁, 等. 脊柱侧弯矫形术患者术后慢性疼痛的危险因素[J]. *临床麻醉学杂志*, 2022, 38(9): 944-948.
XU Ying, HE Long, WANG Jie, et al. Risk factors for chronic post-surgical pain after scoliosis orthopaedics[J]. *J Clin Anesthesiol*, 2022, 38(9): 944-948.
- [22] 宋子扬, 罗浩威, 贺毅, 等. 基于前瞻性队列建立与验证骨科术后慢性疼痛的预测模型[J]. *中国疼痛医学杂志*, 2025, 31(9): 686-695.
SONG Ziyang, LUO Haowei, HE Yi, et al. Development and validation of a prediction model for chronic post-surgical pain in orthopedic surgery based on a prospective cohort study[J]. *Chin J Pain Med*, 2025, 31(9): 686-695.
- [23] Zhang J, Liu TX, Wang WX, et al. Effects of ultrasound-guided erector spinae plane block on postoperative acute pain and chronic post-surgical pain in patients underwent video-assisted thoracoscopic lobectomy: a prospective randomized, controlled trial[J]. *BMC Anesthesiol*, 2023, 23(1): 161. doi:10.1186/s12871-023-02100-5
- [24] 中国康复医学会脊柱脊髓专业委员会腰椎研究学组, 赵杰, 海涌, 等. 后路腰椎椎体间融合技术规范的专家共识[J]. *中国脊柱脊髓杂志*, 2024, 34(12): 1326-1334.
Lumbar Spine Research Group, Spine and Spinal Cord Professional Committee, Chinese Rehabilitation Medical Association, ZHAO Jie, HAI Yong, et al. Expert consensus on technical specifications for posterior lumbar interbody fusion[J]. *Chinese Journal of Spine and Spinal Cord*, 2024, 34(12): 1326-1334.
- [25] Chen D, Yang H, Yang L, et al. Preoperative psychological symptoms and chronic postsurgical pain: analysis of the prospective China Surgery and Anaesthesia Cohort study[J]. *Br J Anaesth*, 2024, 132(2): 359-371.
- [26] 王丽云, 高天勤, 刘雨佳, 等. 基于机器学习产后压力性尿失禁风险预测模型的构建及验证[J]. *山东大学学报(医学版)*, 2025, 63(6): 55-66.
WANG Liyun, GAO Tianqin, LIU Yujia, et al. Development and validation of a postpartum stress urinary incontinence risk prediction model based on machine learning[J]. *Journal of Shandong University (Health Sciences)*, 2025, 63(6): 55-66.
- [27] 贾若曦, 吕丽, 刘涵云, 等. 血细胞计数相关标志物对慢加急性乙型肝炎肝衰竭患者 28 天预后的诊断价值[J]. *山东大学学报(医学版)*, 2025, 63(6): 89-99.
JIA Ruoxi, LYU Li, LIU Hanyun, et al. Diagnostic value of blood cell count-related biomarkers in predicting 28-day mortality in acute-on-chronic hepatitis b liver failure[J]. *Journal of Shandong University (Health Sciences)*, 2025, 63(6): 89-99.
- [28] Setiawan D, Wiranto Y, Girard JM, et al. Individualized machine-learning-based clinical assessment recommendation system[J]. *PLoS Digit Health*, 2025, 4(9): e0001022. doi:10.1371/journal.pdig.0001022
- [29] Azodi CB, Tang JL, Shiu SH. Opening the black box: interpretable machine learning for geneticists[J]. *Trends Genet*, 2020, 36(6): 442-455.