

网络出版时间: 2025-06-30 13:45:19 网络出版地址: <https://link.cnki.net/urlid/34.1065.R.20250630.0857.015>

◇ 临床医学研究 ◇

基于多因素 Logistic 回归构建剖宫产术后子宫 疤痕憩室形成风险预测模型

张梦媛¹, 何 晔¹, 吴媛媛², 王 晶¹

(安徽医科大学第一附属医院¹ 妇产科、² 超声科, 合肥 230022)

摘要 目的 筛选剖宫产术后发生子宫疤痕憩室(CSD)的危险因素并构建风险预测模型。方法 调取行剖宫产术的491例产妇作为研究对象,并收集数据库中术后12个月对产妇进行随访的阴超数据,将数据集按照7:3随机分为训练集和测试集;使用最小绝对值收敛和选择算子(LASSO)进行变量筛选得到特征变量,对特征变量进行多因素Logistic回归分析,进一步使用R软件构建列线图联合预测模型,同时对模型进行内部评价及验证。结果 共纳入491例样本数据,其中训练集344例,测试集147例;通过LASSO交叉验证算法筛选特征变量,最终筛选出5个变量:剖宫产次数、2次剖宫产间隔年数、24 h出血量、手术时间和子宫位置($P < 0.05$)。利用其构建的CSD风险预测模型内部评价和内部验证的决策分析曲线准确度较高;训练集及测试集中诊断模型的AUC(95% CI)分别为0.75(0.71~0.80)和0.79(0.71~0.87)。结论 采用LASSO交叉验证算法建立的风险预测模型对产后CSD发生具有良好的预测价值,值得临床关注。

关键词 子宫疤痕憩室;机器学习;LASSO交叉验证;风险预测;剖宫产;列线图

中图分类号 R 719.82

文献标志码 A **文章编号** 1000-1492(2025)07-1297-08

doi: 10.19405/j.cnki.issn1000-1492.2025.07.019

剖宫产术后子宫疤痕憩室(cesarean scar diverticula, CSD)是指剖宫产手术后,子宫前壁切口处愈合不良而形成的子宫壁囊性膨出^[1-2]。近年来,全球剖宫产率持续上升,我国部分地区已高达50.0%^[3]。剖宫产率的上升导致CSD发生率也随之激增,可导致阴道异常出血、痛经及不孕等问题^[4]。合并CSD再次妊娠可增加子宫破裂、胎盘植入及凶险性前置胎盘等严重并发症的风险^[5]。目前对CSD的形成原因已有较多阐述,但对于这一并发症发生的预测研究较少。近年来,通过机器学习建立预测模型是临床领域的研究热点^[6-7],最小绝对值收敛和选择算子(least absolute shrink-age and selection operator, LASSO)回归是机器学习算法的一种,可提升预测能力并增强模型解释性,适用于大量特征的数据集分析^[8]。通过交叉验证减少过拟合的情况,从而提升模型的泛化能力与稳定性,增强模型解释性^[9]。因此,该研究拟运用LASSO回归交叉

验证的算法从临床数据中筛选指标,结合术后随访憩室的超声形态学数据,构建CSD风险预测模型并验证该模型,为临床上预测及有效预防CSD的发生提供科学依据。

1 材料与方法

1.1 病例资料 Excel随机数字表法抽取2013年12月—2021年8月于安徽医科大学第一附属医院妇产科行剖宫产手术产妇共491例作为研究对象。术后12个月进行电话随访并随诊阴道超声。根据阴道超声评估参数筛选子宫切口是否发生疤痕憩室分为憩室形成组($n=193$)与无憩室形成组($n=298$)。

1.2 纳入与排除标准 纳入标准:年龄20~40岁;无慢性传染病史者;于安徽医科大学第一附属医院行剖宫产术者;阴道超声提示CSD诊断明确;签署知情同意书。排除标准:检查时各项指标不正常者;子宫结构先天发育异常;前次剖宫产有严重并发症;未签署知情同意书;既往接受子宫肌瘤剔除等妇科疾病子宫手术史;1年内再次受孕现为妊娠期。

1.3 研究方法 从安徽医科大学第一附属医院病案管理系统收集临床资料。①基本信息:CSD形成(疤痕处缺损深度 >2 mm)、剖宫产时年龄、孕周、2次剖宫产间隔年数、既往有阴道分娩史、子宫位置

2025-03-25 接收

基金项目:安徽省自然科学基金(编号:2208085QH234)

作者简介:张梦媛,女,硕士研究生;

王 晶,女,主任医师,硕士生导师,通信作者,E-mail:
wangjing@ustc.edu.cn

(将子宫位置按照子宫体纵轴与宫颈纵轴的前夹角来划分^[10]: 前夹角 $< 150^\circ$ 定义为前位, $150^\circ \sim < 180^\circ$ 为平位, $180^\circ \sim < 210^\circ$ 为平后位, $\geq 210^\circ$ 定义为后位); ② 手术相关因素: 术中出血量、24 h 出血量、术前后血色素、B-Lynch 缝合、急诊手术、手术时间; ③ 感染相关因素: 羊水污染程度、宫口扩张(宫口开大 ≥ 1 cm)、胎膜早破、阴道检查; ④ 切口张力相关因素: 孕前体质指数 (body mass index, BMI)、孕期 BMI、巨大儿、是否计划分娩、使用缩宫素时间; ⑤ 切口愈合相关因素: 剖宫产次数、妊娠期合并症及并发症 (包括妊娠期高血压、糖尿病、肝内胆淤积症)、辅助生殖技术助孕、孕期使用激素保胎治疗情况。

1.4 统计学处理 采用 R 4.2.2 和 SPSS25.0 软件进行数据分析。按 7 : 3 的比例随机分为训练集和测试集, 并对变量进行比较, 正态数据中计量资料用 $\bar{x} \pm s$ 表示, 计数资料用 $n(\%)$ 表示, 非正态数据以 $M(P_{25} P_{75})$ 表示。在单变量分析中, 对分类变量比较采用 χ^2 检验或 Fisher 确切概率检验, 对连续变量比较采用 Student's t 检验或秩和检验。在训练集中, 采用 LASSO-Logistic 回归分析进行多因素分析, 筛选独立危险因素, 构建 CSD 的预测列线图。采用受试者工作特征曲线 (receiver operating characteristic, ROC) 和校准曲线评估该模型的性能, 还进行了决策曲线分析 (decision curve analysis, DCA) 以确定预测的净效益阈值。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 基线特征描述 按照 7 : 3 的比例将患者随机分为训练集 ($n = 344$) 及测试集 ($n = 147$)。对训练集和测试集的临床特征进行比较, 结果显示两组数据整体具有可比性, 见表 1。按有无疤痕憩室分组对训练集的临床特征进行差异性分析, 结果显示在训练集中, 剖宫产时年龄、2 次剖宫产间隔年数、剖宫产次数、孕周、术中出血、24 h 出血量、羊水情况、手术时间、子宫位置在两组间的差异有统计学意义 ($P < 0.05$), 见表 2。

2.2 LASSO 回归筛选特征 在训练集中利用 LASSO 回归模型筛选特征, 采用十折交叉验证。最终模型包括 5 个变量: 剖宫产次数、2 次剖宫产间隔年数、24 h 出血量、手术时间和子宫位置, 交叉验证误差在最小的一个标准误差内, 见图 1、图 2。

2.3 Logistic 回归模型分析 将 LASSO 回归筛选出的变量纳入 Logistic 回归模型中, 单因素 Logistic

回归中所有变量均存在统计学差异。最终发现手术时间 OR (95% CI) 为 2.04 (1.04 ~ 4.00), 子宫位置与前位比较, 后位 OR (95% CI) 为 5.57 (3.43 ~ 9.04), 这 2 个特征均是 CSD 的独立影响因素, 见表 3。进一步基于多因素 Logistic 回归结果构建联合预测模型。

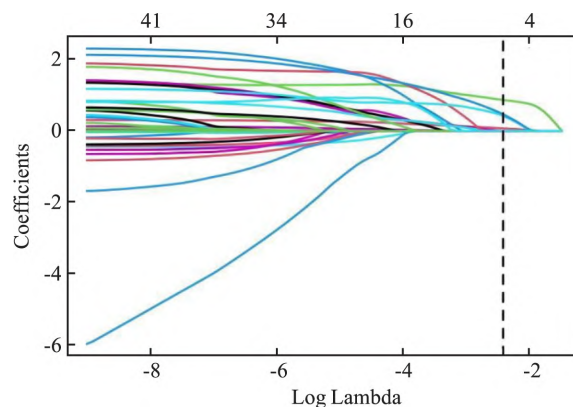


图 1 LASSO 回归变量选择路径

Fig. 1 Path of LASSO regression variable selection

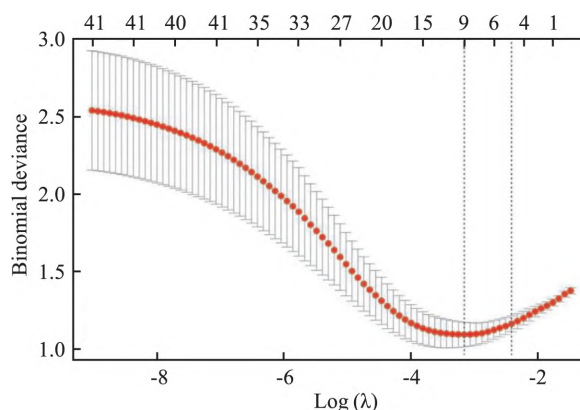


图 2 LASSO 回归的交叉验证图

Fig. 2 Cross-validation plot of LASSO regression

2.4 模型构建 进一步对筛选出的独立预测因素构建联合预测模型。每一个预测变量均可向评分轴垂直投射一个分值, 然后将每个预测变量的分值相加得到一个总评, 在总评分轴上找到相应位置, 向风险轴上垂直投射的数值即为该研究对象的 CSD 发生风险, 见图 3。

2.5 模型验证 对构建的模型进行评价, 如图 4 所示, 该模型在训练集和测试集上的 AUC (95% CI) 分别为 0.75 (0.71 ~ 0.80) 和 0.79 (0.71 ~ 0.87), 相比较于单个变量, 该联合预测模型的稳定性更好, 预测能力更强, 见图 5。训练集和测试集的列线图预测模型校准图如图 6 所示, 该模型的校准曲线与理

表 1 训练集和测试集临床特征比较 [$M(P_{25} P_{75})$ $n(\%)$ $\bar{x} \pm s$]Tab.1 Comparison of clinical characteristics between training and test sets [$M(P_{25} P_{75})$ $n(\%)$ $\bar{x} \pm s$]

Clinical characteristics	Research population		P value
	Training set ($n = 344$)	Test set ($n = 147$)	
Age at cesarean section (year)	29.0 (27.0 , 32.0)	30.0 (28.0 , 33.0)	0.065
Interval years (year)	0 (0 , 0)	0 (0 , 0)	0.506
Number of cesarean sections	0 (0 , 0)	0 (0 , 0)	0.688
Number of vaginal births	0 (0 , 0)	0 (0 , 0)	0.246
Assisted reproduction			0.223
No	295 (85.8)	132 (89.8)	
Yes	49 (14.2)	15 (10.2)	
Week of pregnancy (weeks)	39.0 (38.0 , 40.0)	39.0 (38.0 , 40.0)	0.036
Use of Hormones			0.575
No	279 (81.1)	116 (78.9)	
Yes	65 (18.9)	31 (21.1)	
Emergency cesarean section			0.805
No	268 (77.9)	116 (78.9)	
Yes	76 (22.1)	31 (21.1)	
Preoperative hematocrit (g/L)	114.0 $\pm$ 12.0	115.0 $\pm$ 11.0	0.286
Postoperative hematocrit (g/L)	105.0 $\pm$ 14.0	106.0 $\pm$ 12.0	0.298
Intraoperative bleeding (mL)	300.0 (200.0 , 400.0)	250.0 (200.0 , 300.0)	0.058
24-hour bleeding volume (mL)	420.0 (340.0 , 510.0)	383.0 (311.0 , 498.0)	0.050
Fetal weight (g)	3 500.0 (3 150.0 , 3 800.0)	3 400.0 (3 200.0 , 3 700.0)	0.343
Labor with precursor sign			0.110
No	311 (90.4)	125 (85.0)	
Yes	33 (9.6)	22 (15.0)	
Uterine opening (cm)	0 (0 , 0)	0 (0 , 0)	0.920
Premature rupture of membranes			0.149
No	284 (82.6)	129 (87.8)	
Yes	60 (17.4)	18 (12.2)	
Number of vaginal examinations	0 (0 , 0)	0 (0 , 1.0)	0.513
Duration of use of oxytocin (h)	0 (0 , 0)	0 (0 , 0)	0.146
Planned delivery			0.383
No	320 (93.0)	135 (91.8)	
Yes	24 (7.0)	12 (8.2)	
Complication			0.754
No	227 (65.9)	99 (67.3)	
Yes	117 (34.1)	48 (32.7)	
Amniotic fluid properties			0.920
Clear	287 (83.4)	123 (83.6)	
I - II	29 (8.4)	12 (8.2)	
III	27 (7.9)	12 (8.2)	
Bloody	1 (0.3)	0 (0)	
Operative time (h)	1.0 (0.8 , 1.0)	0.8 (0.8 , 1.0)	0.075
Uterine binding style			0.551
No	321 (93.3)	141 (95.9)	
Bly	8 (2.3)	2 (1.4)	
Improved	15 (4.4)	4 (2.7)	
Uterine position			0.620
Anteflexion	253 (73.5)	101 (68.7)	
Retroflexion	77 (22.4)	38 (25.9)	
Little retroflexion	5 (1.5)	3 (2.0)	
Little anteflexion	9 (2.6)	5 (3.4)	
BMI during pregnancy (kg/m ²)	27.8 (25.7 , 29.5)	27.3 (25.7 , 28.9)	0.293
Non-pregnancy BMI (kg/m ²)	22.0 (20.2 , 23.9)	21.5 (20.1 , 23.4)	0.278

表2 训练集临床特征组间比较 [$M(P_{25} P_{75})$ $n(\%)$ $\bar{x} \pm s$]Tab.2 Intergroup comparison of clinical characteristics of the training set [$M(P_{25} P_{75})$ $n(\%)$ $\bar{x} \pm s$]

Clinical characteristics	Training set		P value
	NCSD ($n = 215$)	CSD ($n = 129$)	
Age at cesarean section (year)	29.0 (27.0 , 32.0)	31.0 (28.0 , 33.0)	0.011
Interval years (year)	0 (0 , 0)	0 (0 , 5.0)	<0.001
Number of cesarean sections	0 (0 , 0)	0 (0 , 1.0)	<0.001
Number of vaginal births	0 (0 , 0)	0 (0 , 0)	0.216
BMI during pregnancy (kg/m^2)	27.8 (25.7 , 29.6)	27.7 (25.8 , 29.4)	0.984
Assisted reproduction			0.605
No	186 (86.5)	109 (84.5)	
Yes	29 (13.5)	20 (15.5)	
Week of pregnancy (weeks)	39.0 (38.0 , 40.0)	39.00 (38.0 , 39.0)	0.025
Use of Hormones			0.696
No	173 (80.5)	106 (82.2)	
Yes	42 (19.5)	23 (17.8)	
Emergency cesarean section			0.502
No	170 (79.1)	98 (76.0)	
Yes	45 (20.9)	31 (24.0)	
Preoperative hematocrit (g/L)	114.0 $\pm$ 12.0	113.0 $\pm$ 11.0	0.516
Postoperative hematocrit (g/L)	105.0 $\pm$ 14.0	105.0 $\pm$ 15.0	0.856
Intraoperative bleeding (mL)	300.0 (200.0 , 300.0)	300.0 (200.0 , 400.0)	0.006
24-hour bleeding volume (mL)	400.0 (330.0 , 488.0)	450.0 (350.0 , 540.0)	0.002
Fetal weight (g)	3 500.0 (3 150.0 , 3 800.0)	3 450.0 (3 150.0 , 3 800.0)	0.594
Labor with precursor sign			0.646
No	176 (91.2)	115 (89.1)	
Yes	19 (8.8)	14 (10.9)	
Uterine opening (cm)	0 (0 , 0)	0 (0 , 0)	0.278
Premature rupture of membranes			0.106
No	183 (85.1)	101 (78.3)	
Yes	32 (14.9)	28 (21.7)	
Number of vaginal examinations	0 (0 , 1.0)	0 (0 , 0)	0.245
Duration of use of oxytocin (h)	0 (0 , 0)	0 (0 , 0)	0.343
Planned delivery			0.197
No	203 (94.4)	117 (90.7)	
Yes	12 (5.6)	12 (9.3)	
Complication			0.100
No	148 (68.8)	79 (61.2)	
Yes	67 (31.2)	50 (38.8)	
Amniotic fluid properties			0.016
Clear	190 (88.4)	97 (75.2)	
I - II	9 (4.2)	20 (15.5)	
III	15 (7.0)	12 (9.3)	
Bloody	1 (0.4)	0 (0)	
Operative time (h)	0.8 (0.8 , 1.0)	1.0 (0.8 , 1.0)	0.035
Uterine binding style			>0.999
No	201 (93.5)	120 (93.0)	
Bly	5 (2.3)	3 (2.3)	
Improved	9 (4.2)	6 (4.7)	
Non-pregnancy BMI (kg/m^2)	22.0 (20.3 , 24.0)	21.6 (20.0 , 23.6)	0.205
Uterine position			<0.001
Anteflexion	181 (84.2)	72 (55.8)	
Retroflexion	26 (12.1)	51 (39.6)	
Little retroflexion	2 (0.9)	3 (2.3)	
Little anteflexion	6 (2.8)	3 (2.3)	

表 3 单因素和多因素 Logistic 回归分析结果
Tab.3 Univariate and multivariate Logistic regression analysis

Clinical characteristics	Univariate analysis					Multivariate analysis				
	β	S. E	Z value	P value	OR (95% CI)	β	S. E	Z value	P value	OR (95% CI)
Number of cesarean sections	1.36	0.23	5.96	<0.001	3.90 (2.49 – 6.10)	0.71	0.52	1.38	0.168	2.04 (0.74 – 5.62)
Interval years	0.22	0.04	5.77	<0.001	1.24 (1.15 – 1.34)	0.10	0.08	1.24	0.215	1.11 (0.94 – 1.30)
Uterine position										
Anteflexion					1.00 (Reference)					1.00 (Reference)
Retroflexion	1.65	0.23	7.12	<0.001	5.23 (3.32 – 8.25)	1.72	0.25	6.94	<0.001	5.57 (3.43 – 9.04)
Little retroflexion	0.87	0.72	1.21	0.225	2.38 (0.59 – 9.72)	0.81	0.77	1.05	0.294	2.25 (0.49 – 10.25)
Little anteflexion	0.28	0.57	0.49	0.622	1.32 (0.43 – 4.05)	0.55	0.58	0.95	0.340	1.74 (0.56 – 5.41)
24-hour bleeding volume	0.01	0	2.24	0.025	1.01 (1.01 – 1.01)	0	0	0.47	0.637	1.00 (1.00 – 1.00)
Operative time	0.95	0.30	3.21	0.001	2.59 (1.45 – 4.62)	0.71	0.34	2.08	0.038	2.04 (1.04 – 4.00)

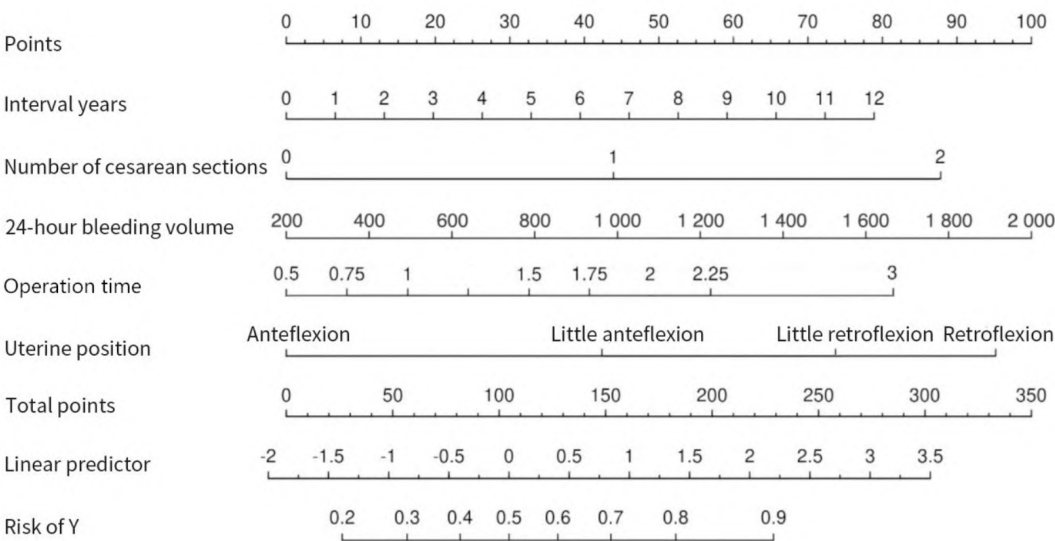


图 3 剖宫产术后 CSD 形成的预测模型列线图
Fig.3 Nomogram prediction model for CSD

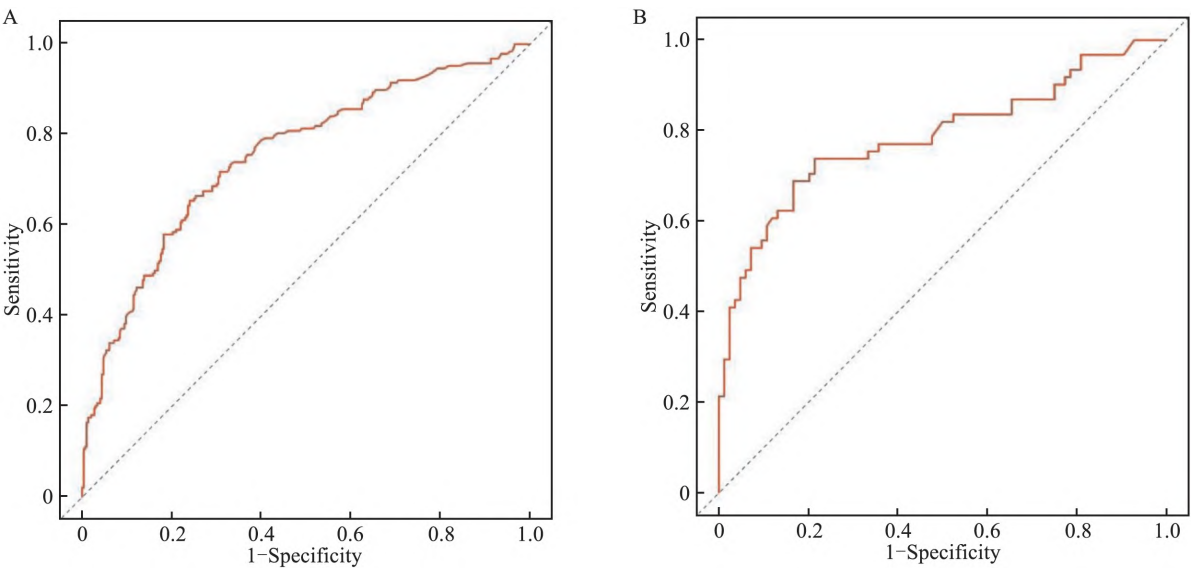


图 4 训练集和测试集 ROC 曲线
Fig.4 Training set and Test set ROC curves
A: Training set; B: Test set.

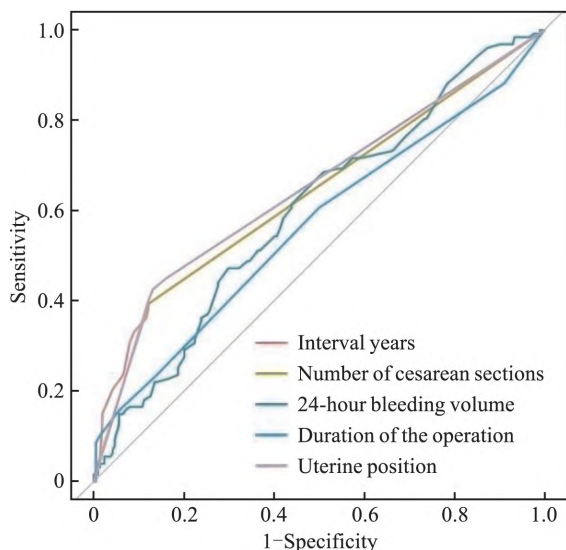


图5 单个变量预测模型 ROC 曲线

Fig. 5 Single variable prediction model ROC curve

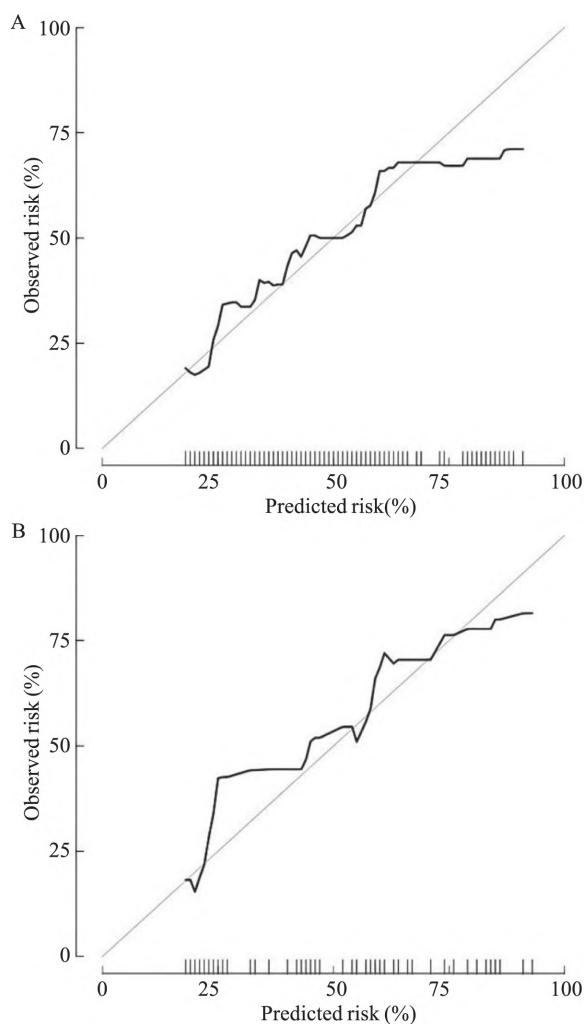


图6 训练集和测试集校准曲线

Fig. 6 Training and test set calibration curves

A: Training set; B: Test set.

想曲线相对接近,表明预测结果与实际发现一致。训练集和测试集的列线图预测模型 DCA 曲线如图 7 所示,该曲线揭示了模型在不同风险阈值下提供的净收益,与对所有患者进行干预或不对任何患者进行干预的极端策略进行了比较。模型在风险阈值为 0.4 附近提供较大的净收益,这强调了模型在临床决策中的实用性。在测试集中,DCA 曲线表现出与训练集相似的趋势,同时说明该模型泛化能力较好。

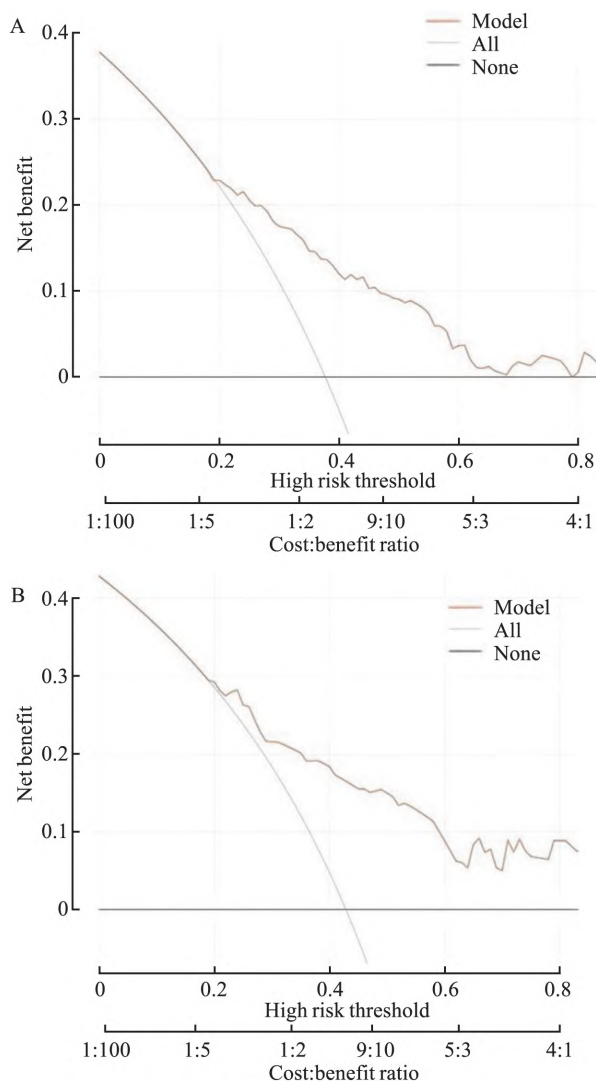


图7 训练集和测试集 DCA 曲线

Fig. 7 DCA curves for training and test sets

A: Training set; B: Test set.

3 讨论

CSD 作为剖宫产术的远期并发症,发生率为 19.4% ~ 88.0% 本研究结果显示,491 例足月产妇

经 12 个月随访,有 193 例发生 CSD,发生率为 39.31%。目前对于 CSD 检测的方式包括磁共振成像、宫腔镜检查以及经腹或阴道超声等^[2,11],但成本较高或为侵入性^[12]。因此,本研究结合临床因素对 CSD 的发生筛选出独立危险因素并构建个性化风险预测模型,为临床判断 CSD 的发生提供一种无创且更为简易方便的途径。

本研究使用机器学习设计 CSD 预测模型,利用训练集数据构建风险预测模型,利用测试集进行模型验证,该模型可靠度优异且具有良好的判别效度和受益。相较于一般算法,在变量过多或变量质量参差不齐的情况下,直接构建传统 Logistic 回归模型可能会降低预测性能,出现特征重要性不合理的情况。故本研究采用 LASSO 回归交叉验证的方法更科学全面的筛选出 5 个特征危险因素,包括剖宫产次数、2 次剖宫产间隔年数、24 h 出血量、手术时间和子宫位置。

本研究发现剖宫产间隔年数及剖宫产次数的增加是 CSD 发生的危险因素,其原因可能与切口纤维化修复及机化相关^[13],切口的纤维化及机化导致子宫肌层缺乏弹性及血管分布减少,影响切口愈合。此外,术中出血量已有相关研究认为是 CSD 发生的危险因素^[13],但本研究认为 24 h 出血量与 CSD 发生的相关性更大,即使术中出血量相等,但因术后出血增多导致总 24 h 出血量增大的产妇术后并发 CSD 的风险更高,其原因可能与子宫肌层收缩不佳、局部肌层血液灌流减少导致切口愈合不良有关。一项对 567 例剖宫产后子宫切口愈合情况的多因素回归分析发现,手术时间减少是 CSD 发生的保护因素^[14],原因可能在于切口暴露时间较短,出血量相应减少。另外,本研究证实后位子宫是 CSD 发生的危险因素。子宫过度后倾后屈引起下段张力增大,会导致切口缺血并延缓愈合,血液滞留在子宫内,容易引发反复感染及愈合不良,加上宫腔压力增加,最终导致 CSD 的形成^[15]。

综上所述,本研究所开发的风险预测模型具有可视化、直观以及易于获取研究指标等优点,便于在临床应用推广。同时,较大的样本量有助于增强研究结果的稳健性,并提高了预测模型的可靠性。然而,本研究也存在一些局限性。首先,该模型属于单中心研究,数据来源较为单一,缺乏多中心的资料。其次,本研究仅为机器学习在临床数据挖掘与整理中的初步探讨,仍需进一步的前瞻性研究进行验证。

参考文献

- [1] Jordans I P M, De Leeuw R A, Stegwee S I, et al. Sonographic examination of uterine niche in non-pregnant women: a modified Delphi procedure[J]. *Ultrasound Obstet Gynecol*, 2019, 53(1): 107–15. doi: 10.1002/uog.19049.
- [2] 赵琪锦,李 楚,杨云萍,等. 子宫瘢痕憩室的研究进展[J]. *中国计划生育学杂志*, 2022, 30(9): 2165–70. doi: 10.3969/j.issn.1004-8189.2022.09.050.
- [2] Zhao Q J, Li C, Yang Y P, et al. Research progress of uterus scar diverticulum[J]. *Chin J Fam Plan*, 2022, 30(9): 2165–70. doi: 10.3969/j.issn.1004-8189.2022.09.050.
- [3] Xu C, Fu Q, Tao H B, et al. Effect of cesarean section on the severity of postpartum hemorrhage in Chinese women: the Shanxi study[J]. *Curr Med Sci*, 2018, 38(4): 618–25. doi: 10.1007/s11596-018-1922-1.
- [4] Klein Meuleman S J M, Murji A, Van Den Bosch T, et al. Definition and criteria for diagnosing cesarean scar disorder[J]. *JAMA Netw Open*, 2023, 6(3): e235321. doi: 10.1001/jamanet-workopen.2023.5321.
- [5] Hochler H, Wainstock T, Lipschuetz M, et al. Grandmultiparity, maternal age, and the risk for uterine rupture—a multicenter cohort study[J]. *Acta Obstet Gynecol Scand*, 2020, 99(2): 267–73. doi: 10.1111/aogs.13725.
- [6] Qadri A M, Hashmi M S A, Raza A, et al. Heart failure survival prediction using novel transfer learning based probabilistic features[J]. *Peer J Comput Sci*, 2024, 10: e1894. doi: 10.7717/peerj-cs.1894.
- [7] Chen A T, Zhang Y, Zhang J. Explainable machine learning and online calculators to predict heart failure mortality in intensive care units[J]. *ESC Heart Fail*, 2025, 12(1): 353–68. doi: 10.1002/ehf2.15062.
- [8] Yu J, Mao W, Sun S, et al. Identification of an m6A-related lncRNA signature for predicting the prognosis in patients with kidney renal clear cell carcinoma[J]. *Front Oncol*, 2021, 11: 663263. doi: 10.3389/fonc.2021.663263.
- [9] Yu J, Mao W, Xu B, et al. Construction and validation of an autophagy-related long noncoding RNA signature for prognosis prediction in kidney renal clear cell carcinoma patients[J]. *Cancer Med*, 2021, 10(7): 2359–69. doi: 10.1002/cam4.3820.
- [10] Xholli A, Londero A P, Scovazzi U, et al. Elasticity of the cervix in relation to uterus position[J]. *J Clin Med*, 2024, 13(9): 2572. doi: 10.3390/jcm13092572.
- [11] Daido S, Kido A, Kataoka M, et al. MR imaging of uterine morphology and dynamic changes during lactation[J]. *J Magn Reson Imaging*, 2017, 45(2): 617–23. doi: 10.1002/jmri.25375.
- [12] Allameh Z, Rouholamin S, Rasti S, et al. A transvaginal ultrasound-based diagnostic calculator for uterus post-cesarean scar defect[J]. *BMC Womens Health*, 2023, 23(1): 558. doi: 10.1186/s12905-023-02715-3.
- [13] Tang X, Wang J, Du Y, et al. Cesarean scar defect: risk factors and comparison of evaluation efficacy between transvaginal sonog-

- raphy and magnetic resonance imaging[J]. *Eur J Obstet Gynecol Reprod Biol*, 2019, 242: 1 – 6. doi: 10.1016/j.ejogrb.2019.09.001.
- [14] Vikhareva O, Valentin L. Risk factors for incomplete healing of the uterine incision after Cesarean section[J]. *BJOG Int J Obstet Gynaecol*, 2010, 117(9): 1119 – 26. doi: 10.1111/j.1471-0528.2010.02631.x.
- [15] Wang J, Pang Q, Wei W, et al. Definition of large niche after Cesarean section based on prediction of postmenstrual spotting: Chinese cohort study in non-pregnant women[J]. *Ultrasound Obstet Gynecol*, 2022, 59(4): 450 – 6. doi: 10.1002/uog.24817.

Construction of a predictive model for the risk of cesarean scar diverticula formation after cesarean delivery based on multifactorial Logistic regression

Zhang Mengyuan¹, He Ye¹, Wu Yuanyuan², Wang Jing¹

(¹Dept of Obstetrics and Gynecology, ²Dept of Ultrasound, The First Affiliated Hospital of Anhui Medical University, Hefei 230022)

Abstract **Objective** To screen the risk factors of cesarean scar diverticula (CSD) after cesarean section and to construct a risk prediction model. **Methods** 491 cases of mothers who underwent cesarean section were recruited as the study subjects, and the data from the database of negative ultrasound of mothers who returned to the hospital 12 months after operation were collected, and the dataset was randomly divided into the training set and the test group according to 7 : 3; the variables were screened to obtain the risk factors of CSD and the risk prediction model was constructed by the use of least absolute shrinkage and selection operator (LASSO); the variables were screened using the LASSO to obtain the characteristic variables, and the characteristic variables were analyzed by multifactorial logistic regression analysis, and the nomogram prediction model was constructed by using the R software. **Results** A total of 491 cases of sample data were included, including 344 cases in the training set and 147 cases in the test set; feature variables were screened by LASSO, and ten-fold cross-validation was used. Five variables were finally screened: number of cesarean deliveries, number of years between two cesarean deliveries, 24-hour hemorrhage, operation time and uterine position ($P < 0.05$). The accuracy of the decision analysis curves for internal evaluation and internal validation of the CSD risk prediction model constructed using it was high; the AUC (95% CI) of the diagnostic model in the training set and the test set were 0.75 (0.71 – 0.80) and 0.79 (0.71 – 0.87), respectively. **Conclusion** The risk prediction model established using the LASSO cross-validation algorithm has good predictive value for the occurrence of postpartum scar diverticula, which deserves clinical attention. **Key words** cesarean scar diverticula; machine learning; LASSO cross-validation; risk prediction; cesarean section; nomogram

Fund program Natural Science Foundation of Anhui Province (No. 2208085QH234)

Corresponding author Wang Jing, E-mail: wangjing@ustc.edu.cn