

# 随机森林和决策树模型在脓毒症患儿预后不良风险预测中的效能比较

叶爱兰, 姚桂爱, 杜晓贞

【关键词】 脓毒症; 预后不良; 随机森林模型; 决策树模型

doi:10.3969/j.issn.1671-0800.2024.11.021

【中图分类号】 R725.9 【文献标志码】 A 【文章编号】 1671-0800(2024)11-1475-05

脓毒症是由感染引起的全身性炎症反应综合征, 并常伴随器官功能障碍, 严重时危及生命<sup>[1]</sup>。资料显示, 脓毒症是儿童重病监护病房中的主要疾病之一, 病死率高达 5%~25%, 全球每年有数百万儿童死于脓毒症<sup>[2]</sup>。目前临床治疗脓毒症主要依靠抗生素、血管活性药物及液体复苏等方式, 但临床效果存在不确定性, 患儿预后也存在较大差异, 预测脓毒症患儿的预后风险已成临床医生关注的重要问题。传统的临床评分系统和生物标志物检测缺乏足够的敏感性和特异性, 预测准确性有限。随机森林和决策树模型作为两种常见的机器学习算法, 在多个领域显示出强大的预测能力, 但不同模型在具体应用中的效能可能存在差异。因此, 本研究通过建立脓毒症患儿预后不良风险预测的随机森林和决策树模型, 比较两种模型的预测效能, 现报道如下。

## 1 资料与方法

1.1 一般资料 收集 2019 年 1 月至 2023 年 12 月丽水市中心医院收治的 530 例脓毒症患儿的临床资料, 经计算机生成随机数字表法, 根据 7:3 的比例分为建模集、验证集, 分别为 371 例、159 例。纳入标准: (1) 年龄 28 d 至 14 岁; (2) 符合儿童脓毒症诊断标准<sup>[3]</sup>, 即存在感染证据, 并伴有全身炎症反应综合征 (SIRS) 的表现; (3) 患儿临床资料及随访资料完整。排除标准: (1) 合并非脓毒症所致的严重血液系

统及心、肝、肾功能障碍; (2) 合并恶性肿瘤; (3) 合并先天性疾病、多发畸形、染色体病; (4) 合并免疫缺陷或正在接受免疫抑制治疗者; (5) 入院前 1 个月有大量失血或输血史; (6) 未接受标准化的脓毒症治疗方案。本研究获得丽水市中心医院伦理委员会科研伦理小组批准, 豁免/免除知情同意。

## 1.2 方法

1.2.1 一般资料收集 通过查询病案、询问家属等方式收集患儿一般资料, 包括性别、年龄、脓毒症家族史、脓毒症严重程度、原发感染部位、合并器官损伤、小儿危重病例评分 (PCIS)、氧合指数、白蛋白 (ALB)、白细胞计数 (WBC)、降钙素原 (PCT)、红细胞分布宽度与血小板计数比值 (RPR)、C 反应蛋白 (CRP)、血尿素氮 (BUN)、肌酐 (Scr)、D-二聚体 (D-D)、乳酸脱氢酶 (LDH) 及机械通气情况。

1.2.2 预后情况随访 患儿入院后 28 d, 统计其生存情况, 死亡患儿纳入不良组, 存活患儿纳入良好组。

1.3 统计方法 采用 SPSS 21.0 软件分析数据, 计数资料采用  $\chi^2$  检验; 符合正态分布的计量资料以均数±标准差表示, 不符合的以  $M(Q_1, Q_3)$  表示, 分别采用独立样本  $t$  检验、Mann-Whitney  $U$  检验; 用 R Studio 软件中的 Random Forest 程序包建立随机森林模型, 对变量的重要性进行排序, 采用 LASSO 分析进行变量选择; 采用 Logistic 回归模型分析脓毒症患儿预后不良的影响因素, 基于决策树 C5.0 算法构建对应的风险预测模型; 采用受试者工作特征曲线 (ROC) 及曲线下面积 (AUC) 验证模型效能。  $P < 0.05$  表示差异有统计学意义。

基金项目: 浙江省医药卫生科技计划项目 (2020KY379)

作者单位: 323000 浙江省丽水, 丽水市中心医院

通信作者: 叶爱兰, Email: 18368094528@163.com

## 2 结果

2.1 建模集脓毒症患儿预后不良的单因素分析 建模集患儿入院后 28 d 死亡 61 例 (16.44%)，验证集死亡 26 例 (16.35%)。将建模组死亡患儿纳入不良组，其余纳入良好组。两组年龄、脓毒症严重程度、合并器官损伤、机械通气、PCT、RPR 及 PCIS 评分差异均有统计学意义 (均  $P < 0.05$ )，见表 1。

2.2 基于随机森林算法的 FCR 森林图构建 经梯度提升回归树算法对随机森林模型的参数进行调整，从上到下 3 条线分别为不良组、袋处数据 (即内

置的支叉验证变量)、良好组患儿的错误率随决策树数目变化而发生变化的情况，从决策树数量为 100 棵开始，随机森林算法误差逐渐趋于平缓，泛化能力逐渐增强。通过随机森林算法对初步纳入的 18 个变量的重要性进行排序，从大到小前 7 位 (贡献率  $> 5%$ ) 依次为 PCIS 评分 (18.47%)、脓毒症严重程度 (17.39%)、PCT (10.25%)、合并器官损伤 (9.35%)、RPR (7.21%)、机械通气 (7.04%) 及年龄 (6.03%)，见图 1。

2.3 建模集脓毒症患儿预后不良的多因素分析 以预后不良 (否=0，是=1) 为因变量，年龄、PCIS 评分、

表 1 建模集脓毒症患儿预后不良的单因素分析

指标	不良组 (n=61)	良好组 (n=310)	$\chi^2(t)[Z]$ 值	P 值
性别[例(%)]			0.54	> 0.05
男	35(57.38)	162(52.26)		
女	26(42.62)	148(47.74)		
年龄(岁)	4.6±1.3	3.2±0.9	(10.64)	< 0.05
脓毒症家族史[例(%)]			3.34	> 0.05
有	6(9.84)	13(4.49)		
无	55(90.16)	297(95.81)		
脓毒症严重程度[例(%)]			7.52	< 0.05
脓毒症	13(21.31)	110(35.48)		
严重脓毒症	38(62.30)	176(56.77)		
脓毒症休克	10(16.39)	24(7.74)		
原发感染部位[例(%)]			7.65	> 0.05
胰腺炎	15(24.59)	62(20.00)		
肺部感染	10(16.39)	79(25.48)		
肠道感染	12(19.67)	53(17.10)		
泌尿系统感染	15(24.59)	96(30.97)		
颅内感染	9(14.75)	20(6.45)		
合并器官损伤[例(%)]			9.87	< 0.05
有	10(16.39)	16(5.16)		
无	51(83.61)	294(94.84)		
PCIS 评分(分)	72.89±14.30	77.88±13.74	(2.58)	< 0.05
氧合指数(mmHg)	275.48±29.16	280.63±29.15	(1.26)	> 0.05
ALB(g/L)	35.21±6.13	36.98±7.24	(1.79)	> 0.05
WBC( $\times 10^9/L$ )	15.86±4.06	14.79±4.42	(1.75)	> 0.05
PCT(ng/L)	1.60±0.47	1.28±0.32	(6.55)	< 0.05
RPR	0.085±0.012	0.056±0.009	(21.68)	< 0.05
CRP(mg/L)	15.42(12.39, 18.43)	16.09(13.09, 18.15)	[1.96]	> 0.05
BUN(mg/dL)	5.23±0.97	4.99±0.89	(1.90)	> 0.05
Scr(mg/dL)	1.45±0.36	1.34±0.42	(1.91)	> 0.05
D-D( $\mu g/L$ )	1.25±0.32	1.19±0.34	(1.27)	> 0.05
LDH(U/L)	226.65±39.98	218.46±37.41	(0.98)	> 0.05
机械通气[例(%)]			6.22	< 0.05
是	16(26.23)	42(13.55)		
否	45(73.77)	268(86.45)		

注: PCIS 为小儿危重病例评分, ALB 为白蛋白, WBC 为白细胞计数, PCT 为降钙素原, RPR 为红细胞分布宽度与血小板计数比值, CRP 为 C 反应蛋白, BUN 为血尿素氮, Scr 为肌酐, D-D 为 D-二聚体, LDH 为乳酸脱氢酶。1 mmHg $\approx$ 0.133 kpa

PCT、RPR(实际值)、脓毒症严重程度(脓毒症、严重脓毒症=0, 脓毒症休克=1)、合并器官损伤(否=0, 是=1)及机械通气(否=0, 是=1)为因变量。当方差膨胀因子(VIF)值 > 10 排除共线性变量。PCIS 评分与脓毒症严重程度表现为共线性, 排除脓毒症严重程度, 保留 PCIS 评分。Logistic 回归模型分析显示, 年龄、PCIS 评分是脓毒症患儿预后的保护性因素(均  $P < 0.05$ ), 合并器官损伤、PCT、RPR 及机械通气是脓毒症患儿预后的危险因素(均  $P < 0.05$ ), 见表 2。

**2.4 脓毒症患儿预后不良风险预测的决策树模型构建** 将 Logistic 回归模型筛选出的 6 个变量(年龄、合并器官损伤、PCIS 评分、PCT、RPR 及机械通气)采用 C5.0 算法进行筛选, 构建决策树模型。决策树模型纳入的因素为年龄、PCT、RPR, 见图 2。

**2.5 随机森林和决策树模型在脓毒症患儿预后不良**

风险预测中的效能验证及比较 ROC 曲线显示, 随机森林模型预测建模集脓毒症患儿预后不良的曲线下面积(AUC)为 0.960(95%CI: 0.935 ~ 0.978,  $P < 0.05$ ), 灵敏度为 88.52%, 特异度为 94.52%; 预测验证集的 AUC 为 0.928 (95%CI: 0.876 ~ 0.963,  $P < 0.05$ ), 灵敏度为 84.62%, 特异度为 90.98%, 见图 3 ~ 4。决策树模型预测建模集脓毒症患儿预后不良的 AUC 为 0.932(95%CI: 0.901 ~ 0.955,  $P < 0.05$ ), 灵敏度为 96.72%, 特异度为 83.55%; 预测验证集脓毒症患儿预后不良的 AUC 为 0.900 (95%CI: 0.842 ~ 0.942,  $P < 0.05$ ), 灵敏度为 76.92%, 特异度为 85.71%, 见图 5 ~ 6。

### 3 讨论

脓毒症发病机制复杂, 涉及宿主免疫反应、病原体特性和环境因素的多重交互作用<sup>[4]</sup>。脓毒症早期

表 2 建模集脓毒症患儿预后不良影响因素的多因素分析

自变量	$\beta$ 值	SE 值	Wald $\chi^2$ 值	P 值	OR 值	95% CI
年龄	-0.587	0.217	7.32	< 0.05	0.556	0.363 ~ 0.851
合并器官损伤	0.672	0.248	7.31	< 0.05	1.957	1.203 ~ 3.186
PCIS 评分	-0.654	0.319	4.20	< 0.05	0.520	0.278 ~ 0.972
PCT	0.500	0.241	4.32	< 0.05	1.648	1.029 ~ 2.641
RPR	0.778	0.254	9.40	< 0.05	2.176	1.324 ~ 3.578
机械通气	0.868	0.310	7.87	< 0.05	2.383	1.299 ~ 4.372
常数项	-14.379	4.073	12.46	< 0.05		

注: PCIS 为小儿危重病例评分, PCT 为降钙素原, RPR 为红细胞分布宽度与血小板计数比值

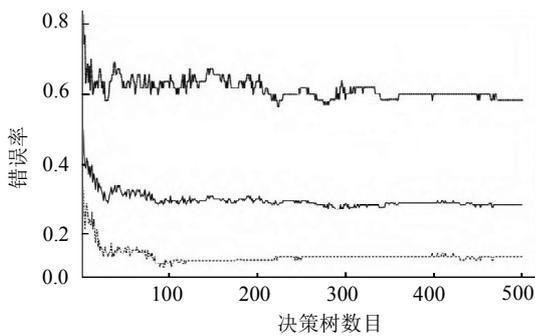


图 1 决策树数目与错误率的关系

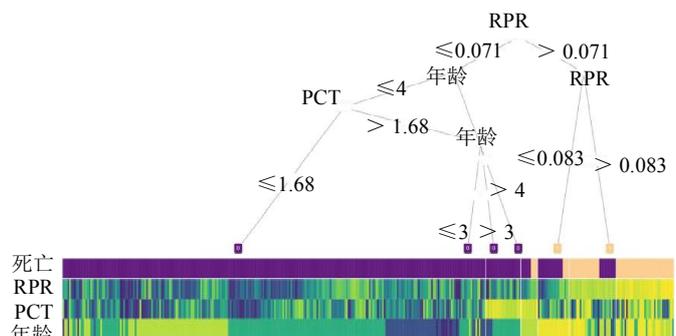


图 2 脓毒症患儿预后不良风险预测的决策树模型

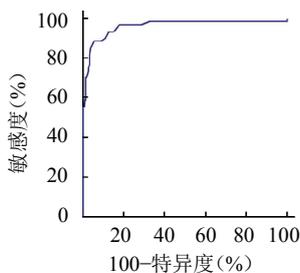


图 3 随机森林模型预测建模集脓毒症患儿预后不良的 ROC 曲线

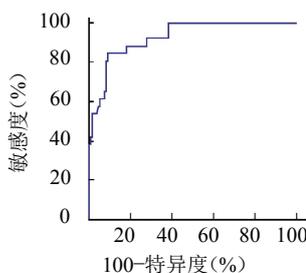


图 4 随机森林模型预测验证集脓毒症患儿预后不良的 ROC 曲线

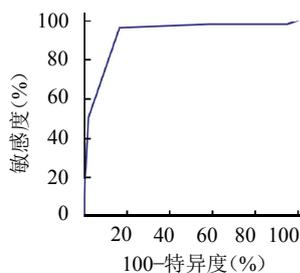


图 5 决策树模型预测建模集脓毒症患儿预后不良的 ROC 曲线

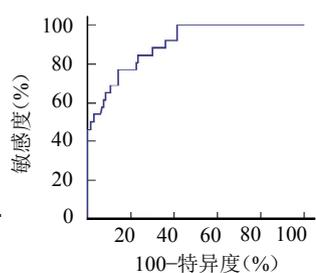


图 6 决策树模型预测验证集脓毒症患儿预后不良的 ROC 曲线

症状缺乏特异性,容易与其它感染性疾病混淆,诊断困难,加之其病理过程迅速且不可预测,病情可能在短时间内迅速恶化,发展为脓毒性休克和多器官功能衰竭。本研究结果显示建模集患儿死亡率为16.44%,验证集为16.35%,与郑聪颖等<sup>[5]</sup>的16.95%接近,但低于徐涛等<sup>[6]</sup>的30.00%及霍风文等<sup>[7]</sup>的25.56%,可能与不同医疗机构资源、患儿遗传易感性存在差异有关。目前临床对脓毒症患儿风险预测方法存在显著局限性,传统的临床评分系统主要基于生理和实验室指标,通过加权评分来评估病情严重程度和预后风险,在成人脓毒症中有一定的应用价值,但儿童的生理特点和病理反应与成人存在明显差异,简单将成人评分系统应用于儿童可能导致预测准确性和可靠性下降<sup>[8]</sup>。因此探索和应用更为先进的预测模型尤为必要。

随着大数据和机器学习技术的发展,数据驱动的预测模型在医学领域展现出巨大潜力,其可在短时间内处理大量复杂的临床数据,捕捉传统统计方法难以发现的非线性关系,为疾病的预测和诊断提供新的视角。本研究经梯度提升回归树算法调整随机森林模型的参数,对变量的重要性进行排序,从大到小前7位依次为PCIS评分、脓毒症严重程度、PCT、合并器官损伤、RPR、机械通气及年龄。Logistic回归模型分析显示,年龄、PCIS评分是脓毒症患儿预后的保护性因素,合并器官损伤、PCT、RPR及机械通气是脓毒症患儿预后不良的危险因素。分析原因为:(1)随着年龄的增长,儿童的免疫系统逐渐发育成熟,对病原体的识别和反应能力增强,更能有效对抗感染,降低脓毒症导致严重并发症的风险。此外,较大年龄的儿童各项器官功能和代谢能力逐渐完善,可更好地应对脓毒症引起的全身性炎症和代谢紊乱。(2)PCIS评分低的脓毒症患儿通常存在显著的生理失衡,如低血压、低氧血症、代谢性酸中毒和意识障碍等,表明患儿的器官和系统功能严重受损,难以维持正常的生命活动和恢复能力,其需要更密集的监护和治疗,包括高级生命支持和持续的多学科医疗团队协作,预后相对更差。宋林等<sup>[9]</sup>研究发现,PCIS评分增加是脓毒症患儿死亡的独立保护因素,提示其与脓毒症患儿预后密切相关。(3)脓毒症中一个器官的损伤可能影响其它器官的功能,形成

“器官-器官交互作用”,增加感染性休克、呼吸衰竭、急性肾损伤等并发症风险,且合并器官损伤的脓毒症患儿需要更复杂的治疗策略,包括血流动力学支持、器官功能支持和抗感染治疗,治疗的复杂性可能影响疗效。(4)高PCT水平通常伴随着严重细菌感染,提示体内存在较高的感染负荷,病情进展迅速,感染控制困难,恶化风险高。Li等<sup>[10]</sup>研究发现,血清PCT在脓毒症患者中表达异常,与疾病严重程度密切相关,且对患者康复具有一定的预测价值。(5)RPR的升高与全身炎症反应、骨髓功能障碍和血液动力学变化有关,反映体内严重的炎症和凝血紊乱,预示病情加重。高RPR值也提示机体存在显著微循环障碍,红细胞变异性增加和血小板减少可能导致组织供氧不足,加重器官损伤和功能衰竭。一项基础动物研究显示<sup>[11]</sup>,脓毒症小马驹模型的RPR显著高于健康小马驹,诊断脓毒症的最佳临界值为0.09,且与预后有密切关联。(6)有机械通气的需求表明患儿已发生严重的呼吸衰竭,肺功能严重受损,而呼吸衰竭是脓毒症进展为重症和死亡的重要原因,且机械通气可能带来呼吸机相关性肺炎、气胸、肺泡损伤等一系列并发症,导致病情恶化。

本研究将Logistic回归模型筛选出的6个变量采用C5.0算法进行筛选,构建决策树模型,纳入的因素为年龄、PCT、RPR。经ROC曲线发现,随机森林模型预测建模集、验证集脓毒症患儿预后不良的AUC均高于决策树模型,表明随机森林的预测效能优于决策树模型。本研究中随机森林模型在预测脓毒症患儿预后不良风险方面的优势可能与其集成学习的特性有关。随机森林通过构建多个决策树,可以减少单一决策树的预测误差,并提高模型的泛化能力。尽管随机森林在多个评估指标上表现优异,但其模型复杂度较高,训练和预测时间相对较长,在临床实时应用中可能会受到一定限制。而决策树模型由于其结构简单、易于解释,仍然在临床实践中具有重要的应用价值,尤其是在需要快速、直观决策的情况下。决策树模型可以通过可视化的树状结构帮助临床医生理解预测过程和结果,提供有力的决策支持。因此,随机森林和决策树模型各有所长,选择哪一种模型或是否结合使用,应根据具体情况和需求来决定。

综上所述,在脓毒症患儿预后不良风险预测中,随机森林的预测效能优于决策树模型。本研究的创新性在于首次尝试应用随机森林算法于特定人群(脓毒症患儿)的风险预测,将临床数据与机器学习技术结合起来,为脓毒症患儿的预后评估提供了新的视角和方法。不足之处在于数据集来源于单一中心,限制了模型的泛化能力。未来研究方向应进行多中心、前瞻性研究,以增强模型的泛化能力和临床适用性,并其他机器学习技术在脓毒症患儿预后评估中的应用。

**利益冲突** 所有作者声明无利益冲突

### 参 考 文 献

- [1] WENG L, XU Y, YIN P, et al. National incidence and mortality of hospitalized sepsis in China[J]. *Crit Care*, 2023, 27(1): 84.
- [2] REHN M, CHEW M S, OLKKOLA K T, et al. Surviving sepsis campaign: International guidelines for management of sepsis and septic shock in adults 2021-endorsement by the Scandinavian society of anaesthesiology and intensive care medicine[J]. *Acta Anaesthesiol Scand*, 2022, 66(5): 634-635.
- [3] SCHLAPBACH L J, WATSON R S, SORCE L R, et al. International consensus criteria for pediatric sepsis and septic shock[J]. *JAMA*, 2024, 331(8): 665-674.
- [4] YANG K, CHEN J Q, WANG T, et al. Pathogenesis of sepsis-associated encephalopathy: More than blood-brain barrier dysfunction[J]. *Mol Biol Rep*, 2022, 49(10): 10091-10099.
- [5] 郑聪颖,林小飞,李江书,等.血小板聚集率和血小板黏附率评价脓毒症患儿病情及预后的价值[J]. *河北医科大学学报*, 2023, 44(9): 1053-1057.
- [6] 徐涛,李瑶,梁君峰,等.超声心动图及血清肿瘤坏死因子- $\alpha$ 、白细胞介素-1 $\beta$ 对脓毒症患儿的诊断及预后评估[J]. *临床超声医学杂志*, 2022, 24(6): 440-444.
- [7] 霍风文,张俊艳,张德重,等.血清 APN、PGRN 联合检测对脓毒症患儿疾病的预测价值及与其预后的相关性分析[J]. *现代生物医学进展*, 2022, 22(8): 1596-1600.
- [8] DRONAMRAJU S, AGRAWAL S, KUMAR S, et al. Comparison of piro, APACHE iv, and sofa scores in predicting outcome in patients with sepsis admitted to intensive care unit: A two-year cross-sectional study at rural teaching hospital[J]. *Indian J Crit Care Med*, 2022, 26(10): 1099-1105.
- [9] 宋林,邹惠,曾玲,等.脓毒症患儿预后的影响因素分析及 pSOFA 评分、PCIS 评分及早期血乳酸测定的预测价值探讨[J]. *现代生物医学进展*, 2023, 23(3): 494-499.
- [10] LI Q, YAN W, LIU S, et al. Study on the correlation and clinical significance of T-lymphocyte Subsets, IL-6 and PCT in the severity of patients with sepsis[J]. *Pak J Med Sci*, 2023, 39(1): 227-231.
- [11] SCALCO R, DE OLIVEIRA G N, DA ROSA CURCIO B, et al. Red blood cell distribution width to platelet ratio in neonatal foals with sepsis[J]. *J Vet Intern Med*, 2023, 37(4): 1552-1560.

收稿日期:2024-07-16

(本文编辑:吴迪汉)