

## • 诊治分析 •

# 基于机器学习的严重急性肾损伤患者中心静脉导管相关血流感染风险预测模型的构建

郑玲玲, 叶颖颖, 颜彬彬, 温正旺

【关键词】 血液透析; 机器学习模型; 导管相关感染; 急性肾衰竭; 危险因素

doi:10.3969/j.issn.1671-0800.2024.09.032

【中图分类号】 R459.5 【文献标志码】 B 【文章编号】 1671-0800(2024)09-1226-05

急性肾损伤(AKI)是重症监护病房(ICU)危重患者的常见并发症之一,发病率和致死率较高。肾脏替代治疗是主要治疗手段,其中以血液透析最为常见。当前指南推荐动静脉瘘作为血液透析的主要方式<sup>[1]</sup>,然而ICU中患者常因充血性心力衰竭或多发动脉粥样硬化斑块等无法建立人工动静脉瘘<sup>[2]</sup>,此时中心静脉导管成为了血液透析的主要血管通路,但同时带来的中心静脉导管相关感染风险显著增高<sup>[3-4]</sup>,尤其是中心静脉导管相关血流感染(CLABSI)可导致患者经济负担增加、住院及死亡风险加大<sup>[1,5]</sup>,因此明确CLABSI危险因素和构建预测模型有助于临床早期预防、诊疗并提高患者预后。近年来,学者们提出了较多的模型来预测CLABSI<sup>[6-8]</sup>,然而国内外关于机器学习算法在AKI血液透析患者CLABSI风险预测中的优势仍缺乏证据<sup>[9-10]</sup>。本研究回顾性分析温州市中西医结合医院ICU AKI血液透析患者的临床资料,探讨发生CLABSI的独立因素并建立机器学习预测模型,以预测AKI血液透析患者CLABSI的发生风险,现报道如下。

## 1 资料与方法

1.1 一般资料 回顾性选取2015年1月至2023年12月于温州市中西医结合医院ICU接受血液透析治疗的510例严重AKI患者的临床资料。纳入标准:(1)年龄18~80岁;(2)符合改善全球肾脏病预后组织推荐<sup>[11]</sup>的AKI III期(即严重急性肾损伤)诊断标准

及血液透析指征;(3)以临时中心静脉置管接受血液透析;(4)住院临床资料齐全。排除标准:(1)住院时间或导管留置时间<48h;(2)临床资料不全者;(3)合并恶性肿瘤。本研究获温州市中西医结合医院医学伦理委员会批准,豁免/免除知情同意。

## 1.2 方法

1.2.1 临床资料收集 (1)一般资料:年龄、性别、体质指数(BMI)、合并高血压、合并糖尿病、合并冠心病、合并房颤、卒中史、吸烟史及饮酒史;(2)血液透析前实验室资料:血红蛋白水平、血清白蛋白水平、白细胞计数及血小板计数;(3)血液透析相关操作:插管部位(颈内静脉、股静脉)、导管留置时间(<2周,≥2周)、中心静脉插管次数(1次,≥2次)以及3个月内导管相关感染史。

1.2.2 CLABSI诊断及分组 CLABSI的诊断参照2019年肾脏病预后质量倡议血管通路临床实践指南<sup>[12]</sup>,根据患者是否发生CLABSI分为CLABSI组和非CLABSI组。

1.3 自变量筛选及机器学习模型的建立 将研究对象以7:3比例随机分为训练集(358例用于模型建立)和测试集(152例用于模型测试),在训练集中使用单因素、多因素Logistic回归分析纳入指标,确定严重AKI患者接受血液透析治疗后发生CLABSI的独立预测因素,并作为自变量纳入预测模型,分别建立基于逻辑回归(LR)、K-最近邻(KNN)、支持向量机(SVM)、随机森林(RF)、决策树(DT)、反向传播(BP)神经网络6种机器学习算法的CLABSI预测模型。

1.4 统计方法 采用SPSS 26.0统计软件进行处理。正态分布的计量资料表示为均数±标准差,组间比较

基金项目: 温州市基础性公益科研项目(Y2023653)

作者单位: 325000 浙江省温州,温州市中西医结合医院

通信作者: 温正旺, Email: 429317230@qq.com

采用独立样本 *t* 检验；偏态分布的计量资料以中位数(四分位数)表示,组间比较采用非参数检验。计数资料比较采用 $\chi^2$  检验。应用基于 python 3.9 的标准库模块(numpy、pandas、sklearn、matplotlib)进行数据处理、机器学习模型建立及结果可视化输出,绘制受试者工作特征(ROC)曲线,并计算曲线下面积(AUC)、准确性、敏感度、特异度来评估各模型的预测效能。 $P < 0.05$  表示差异有统计学意义。

## 2 结果

**2.1 训练集 AKI 血液透析患者并发 CLABSI 的单因素分析** 510 例患者中发生 CLABSI 111 例(21.8%)。训练集 358 例患者中,非 CLABSI 283 例,CLABSI 75 例;测试集 152 例中,非 CLABSI 116 例,CLABSI 36 例。在训练集中,与非 CLABSI 组相比较,CLABSI 组中年龄 $\geq 60$  岁、合并糖尿病、白细胞计数 $\geq 10 \times 10^9/L$ 、股静脉插管、导管留置时间 $\geq 2$  周、中心静脉插管次数 $\geq 2$  次占比较高(均 $P < 0.05$ ),而吸烟史、血红蛋白水平 $\geq 90$  g/L、血清白蛋白 $\geq 30$  g/L 占比较低(均 $P < 0.05$ ),两组患者其余指标相比较差异无统计学意义(均 $P > 0.05$ ),见表 1。

**2.2 训练集 AKI 血液透析患者并发 CLABSI 的多因素 Logistic 回归分析** 以单因素分析中有统计学意义的 9 个指标作为自变量,以是否发生 CLABSI

为因变量,结果显示年龄( $< 60$  岁=0, $\geq 60$  岁=1)、合并糖尿病(否=0,是=1)、血红蛋白水平( $< 90$  g/L=0, $\geq 90$  g/L=1)、血清白蛋白水平( $< 30$  g/L=0, $\geq 30$  g/L=1)、插管部位(颈内静脉=0,股静脉=1)、导管留置时间( $< 2$  周=0, $\geq 2$  周=1)和中心静脉插管次数(1次=0, $\geq 2$  次=1)是 CLABSI 发生的独立影响因素(均 $P < 0.05$ ),见表 2。

**2.3 机器学习预测模型性能比较** 将基于多因素 Logistic 回归分析筛选出的独立预测因素作为自变量纳入机器学习模型,分别建立 LR、KNN、SVM、RF、DT 和 BP 神经网络模型,ROC 曲线结果见图 1。图 2 为模型算法的准确率比较,SVM 和 DT 模型不仅在中位准确率上表现优异,而且其准确率分布相对较为集中,显示出较高的稳定性和可靠性。6 种模型在训练集和测试集的敏感度、特异度、准确率、AUC(95%CI)见表 3,在测试集和训练集中,6 种机器学习模型的 AUC 值均 $> 0.7$ ,测试集中 DT 模型的准确率(84.4%)、特异度(94.5%)和 AUC 值(0.833)均最高,DT 模型的整体预测能力优于其他 5 种模型,见图 3。

## 3 讨论

CLABSI 作为 AKI 血液透析患者的常见并发症,严重者可致死亡,是影响预后的重要原因之一。

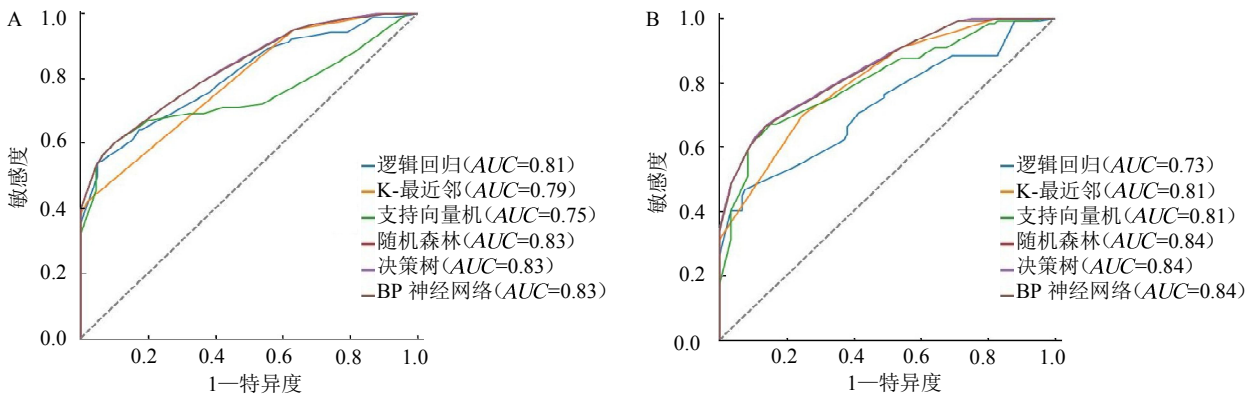
表 1 训练集 AKI 患者并发 CLABSI 的单因素分析

指标	非 CLABSI 组( $n=283$ )	CLABSI 组( $n=75$ )	$\chi^2(Z)$ 值	例(%)	$P$ 值
年龄 $\geq 60$ 岁[例(%)]	130(45.9)	56(65.3)	8.92		$< 0.05$
男性[例(%)]	167(59.0)	19(56.0)	0.69		$> 0.05$
BMI( $kg/m^2$ )	25.1(23.5,27.4)	24.7(22.2,26.6)	(1.36)		$> 0.05$
合并高血压[例(%)]	232(81.6)	64(85.3)	0.56		$> 0.05$
合并糖尿病[例(%)]	101(35.7)	45(60.0)	14.51		$< 0.05$
合并心房颤动[例(%)]	50(17.7)	15(20.0)	0.22		$> 0.05$
卒中史[例(%)]	21(7.4)	6(8.0)	0.30		$> 0.05$
吸烟史[例(%)]	99(35.0)	18(21.3)	5.01		$< 0.05$
饮酒史[例(%)]	93(32.9)	20(26.7)	1.05		$> 0.05$
血红蛋白水平 $\geq 90$ g/L[例(%)]	193(68.2)	28(37.3)	23.91		$< 0.05$
血清白蛋白水平 $\geq 30$ g/L[例(%)]	199(70.3)	32(42.7)	19.80		$< 0.05$
白细胞计数 $\geq 10 \times 10^9/L$ [例(%)]	33(11.7)	18(24.0)	7.39		$< 0.05$
血小板计数( $\geq 180 \times 10^9/L$ )[例(%)]	143(50.5)	35(46.7)	0.35		$> 0.05$
股静脉插管[例(%)]	154(54.4)	57(76.0)	11.41		$< 0.05$
导管留置时间 $\geq 2$ 周[例(%)]	175(61.8)	66(88.0)	18.45		$< 0.05$
中心静脉插管次数 $\geq 2$ 次[例(%)]	22(7.8)	18(24.0)	15.73		$< 0.05$
3 个月内导管相关感染[例(%)]	26(9.2)	6(8.0)	0.10		$> 0.05$

注: BMI 为体质量指数

表2 训练集 AKI 患者并发 CLABSI 的多因素 Logistic 分析

变量	$\beta$ 值	SE 值	Wald $\chi^2$ 值	P 值	OR 值(95%CI)
年龄	0.657	0.321	4.18	< 0.05	1.930(1.028 ~ 3.623)
合并糖尿病	1.246	0.321	15.10	< 0.05	3.476(1.854 ~ 6.516)
吸烟史	-0.187	0.330	0.32	> 0.05	0.830(0.435 ~ 1.583)
血红蛋白水平	-1.318	0.318	17.16	< 0.05	0.268(0.144 ~ 0.499)
血清白蛋白水平	-1.126	0.320	12.41	< 0.05	0.324(0.173 ~ 0.607)
白细胞计数	0.522	0.456	1.31	> 0.05	1.685(0.689 ~ 4.119)
插管部位	1.108	0.352	9.89	< 0.05	3.027(1.157 ~ 6.039)
导管留置时间	1.434	0.413	12.05	< 0.05	4.196(1.867 ~ 9.431)
中心静脉插管次数	1.502	0.430	12.21	< 0.05	4.493(1.934 ~ 10.436)
常量	-3.030	0.615	24.25	< 0.05	—



注:AKI 为急性肾损伤,CLABSI 为中心静脉导管相关感染

图1 训练集(A)与测试集(B)中6种机器学习模型对 AKI 血液透析患者并发 CLABSI 预测的 ROC 曲线

表3 6种机器学习模型在不同数据集中对 AKI 血液透析患者并发 CLABSI 的预测效能

模型	数据集	准确性	敏感度 (%)	特异度 (%)	AUC	95%CI
逻辑回归	训练集	0.847	70.3	88.2	0.81	0.73 ~ 0.90
	测试集	0.784	65.6	88.5	0.73	0.64 ~ 0.85
K-最近邻	训练集	0.773	74.8	86.5	0.79	0.74 ~ 0.87
	测试集	0.619	63.9	70.4	0.81	0.61 ~ 0.90
支持向量机	训练集	0.854	78.8	88.1	0.75	0.65 ~ 0.86
	测试集	0.712	67.5	72.0	0.81	0.65 ~ 0.90
随机森林	训练集	0.728	61.9	74.7	0.83	0.76 ~ 0.92
	测试集	0.625	65.7	78.2	0.84	0.71 ~ 0.95
决策树	训练集	0.888	76.2	88.5	0.83	0.74 ~ 0.93
	测试集	0.844	64.7	94.5	0.84	0.73 ~ 0.95
BP 神经网络	训练集	0.76	63.5	79.2	0.83	0.75 ~ 0.92
	测试集	0.602	54.4	72.5	0.81	0.72 ~ 0.94

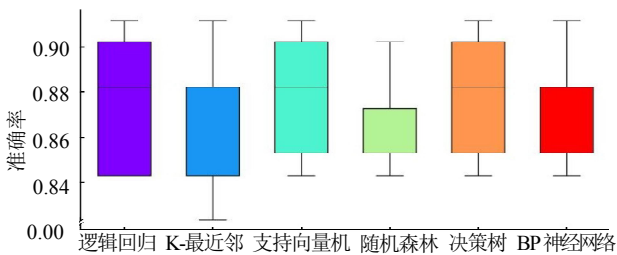


图2 6种机器学习算法预测 AKI 血液透析患者并发 CLABSI 的准确率比较

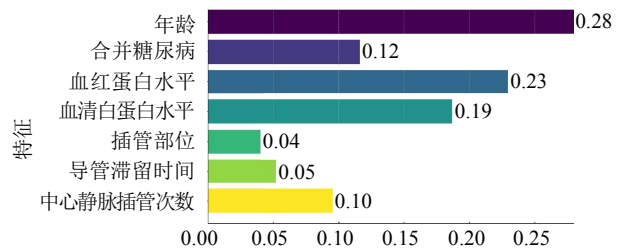


图3 AKI 血液透析患者并发 CLABSI 的独立预测因素在机器学习 DT 模型预测中的特征重要性

本研究中 AKI 血液透析患者后 CLABSI 的发生率为 21.8%，与以往研究结果相似<sup>[2,13]</sup>。目前 CLABSI 的病因及发病机制尚未明确，但普遍认为是多因素综合作用的结果。本研究结果显示，年龄 $\geq 60$ 岁、合并糖尿病、股静脉插管、导管留置时间 $\geq 2$ 周和插管次数 $\geq 2$ 次是 AKI 血液透析患者出现 CLABSI 的危险因素，而血红蛋白水平 $\geq 90$  g/L 和白蛋白水平 $\geq 30$  g/L 为保护性因素。

本研究表明合并糖尿病的 AKI 血液透析患者更易发生 CLABSI，与往研究相符<sup>[6-7]</sup>。这是由于长期高血糖可损伤血管内皮和抑制细胞免疫，且高糖状态有利于病原体生长繁殖<sup>[14]</sup>。年龄作为 CLABSI 的预测因素尚有争议。研究显示，年龄对于 CLABSI 的发生不存在影响<sup>[4,10,15]</sup>。但本研究发现年龄 $\geq 60$ 岁是 AKI 血液透析患者 CLABSI 发生的独立危险因素，与刘亚敏等<sup>[6]</sup>的研究结果相符合。原因可能是，高龄患者基础病多、血管条件差、自体动静脉内瘘建立困难、导管使用率高<sup>[16]</sup>；高龄患者自我卫生条件较差、免疫功能降低、易发生营养不良等，因此增加了导管微生物定植及血流感染的风险<sup>[17]</sup>。

本研究结果显示，股静脉插管、导管留置时间 $\geq 2$ 周、插管次数 $\geq 2$ 次的 AKI 血液透析患者发生 CLABSI 的风险显著增高。一项 Meta 分析<sup>[18]</sup>表明股静脉置管感染风险是颈内静脉的 1.9 倍；张红娟等<sup>[19]</sup>的研究也表明，股静脉置管的感染比例较颈内静脉明显增高，并认为股静脉置管可能无法作为中心静脉留置导管的首选部位。造成这种结果的原因可能与解剖学上股静脉临近会阴部、潮湿多汗及股静脉血流速度相对较慢，更易形成血栓有关。而导管留置时间延长和插管次数增加可能导致病原菌定植风险加大<sup>[5,15]</sup>，感染风险增加，这提示临床在不影响患者治疗的情况，要优先选择颈内静脉置管，同时控制导管留置时间和减少插管次数。

本研究表明血红蛋白水平 $\geq 90$  g/L 和白蛋白水平 $\geq 30$  g/L 是 CLABSI 发生的保护因素。血清白蛋白水平一定程度上反映了患者的营养状态，其浓度降低则免疫球蛋白相应减少，从而导致免疫功能紊乱。同理，当血红蛋白水平较低时，身体各组织器官细胞可能会缺氧，影响其正常功能(包括免疫细胞的正常活动)<sup>[20]</sup>。因此，维持一定的血红蛋白水平和白

蛋白水平有助于减少感染发生。

国内外已广泛报道 Logistic 及 Cox 比例回归风险预测模型在血液透析患者 CLABSI 发生中的应用价值。但至今还没有经过广泛验证的预测工具来识别 ICU 中具有 CLABSI 高风险的 AKA 血液透析患者<sup>[4]</sup>。近些年随着大数据时代的发展，机器学习在临床上的应用逐渐受到重视。Beeler 等<sup>[10]</sup>使用机器学习来预测 CLABSI，模型预测 CLABSI 的 AUC 为 0.87。然而他们的研究依赖于电子病历数据中的医护人员的个人评估，此外模型为实时预测 CLABSI，无法在 CLABSI 发生前预测结果，这限制了这种工具在预防 CLABSI 方面的临床实用性。本研究使用了多个机器学习模型来预测 AKI 血液透析患者 CLABSI 的发生，结果显示在 LR、KNN、SVM、RF、DT 和 BP 神经网络 6 种模型中，DT 模型的整体预测能力更优，可能是因为它对数据样本量要求较小、可适用于非线性数据，多维度输出和特征选择功能较强<sup>[21]</sup>。

综上所述，基于机器学习算法的预测模型能有效预测 AKI 血液透析患者 CLABSI 的发生风险，能够为 CLABSI 的预防和干预提供一定参考。AKI 血液透析患者发生 CLABSI 与年龄、合并糖尿病、血红蛋白水平、白蛋白水平、插管部位、导管留置时间及插管次数等多种因素相关，应引起临床关注。但本研究也存在一定局限性：首先，这是一项单中心回顾性研究，样本量不够充足；其次，研究只纳入了部分生化指标，可能遗漏某些重要的生化数据，未来还需要在多中心、大样本研究中进行验证。

**利益冲突** 所有作者声明无利益冲突

## 参 考 文 献

- [1] 国家慢性肾病临床医学研究中心, 中国医师协会肾脏内科医师分会, 中国急性肾损伤临床实践指南专家组. 中国急性肾损伤临床实践指南[J]. 中华医学杂志, 2023, 103(42): 3332-3366.
- [2] 刘晓天, 叶红坚, 郑勋华, 等. 连续性肾脏替代治疗患者透析导管相关性感染的临床特征和危险因素[J]. 中华肾脏病杂志, 2019, 35(5): 321-328.
- [3] BÖHLKE M, ULIANO G, BARCELLOS F C. Hemodialysis catheter-related infection: Prophylaxis, diagnosis and treatment[J]. J Vasc Access, 2015, 16(5): 347-355.
- [4] RAHMANI K, GARIKIPATI A, BARNES G, et al. Early prediction of central line associated bloodstream infection using machine learning[J]. Am J Infect Control, 2022, 50(4): 440-445.

- [5] 王艳,郑璐.血液透析中心静脉导管相关性血流感染的危险因素分析及风险预测模型的构建与验证[J].全科护理,2023,21(31):4453-4456.
- [6] 刘亚敏.透析导管相关血流感染危险因素分析及风险预测模型[D].郑州:郑州大学,2021.
- [7] 王金红,闫莉,张翠萍.老年急诊血液透析患者中心静脉导管相关性感染现状及危险因素分析[J].老年医学与保健,2023,29(3):565-569.
- [8] 王姗,张志媛,姚佳,等.慢性肾功能衰竭血液透析患者导管相关性血流感染预测模型的构建[J].中华现代护理杂志,2022,28(23):3123-3127.
- [9] 杨可舟,刘宁,王婷,等.机器学习应用于中央导管相关血流感染预测模型的系统评价[J].中华现代护理杂志,2023,29(34):4677-4682.
- [10] BEELER C, DBEIBO L, KELLEY K, et al. Assessing patient risk of central line-associated bacteremia via machine learning[J]. Am J Infect Control, 2018, 46(9): 986-991.
- [11] KELLUM J A, LAMEIRE N, KDIGO AKI Guideline Work Group. Diagnosis, evaluation, and management of acute kidney injury: A KDIGO summary (Part 1)[J]. Crit Care, 2013, 17(1): 204.
- [12] LOK C E, HUBER T S, LEE T, et al. KDOQI clinical practice guideline for vascular access: 2019 update[J]. Am J Kidney Dis, 2020, 75(4 Suppl 2): S1-S164.
- [13] 刘亚敏,赵培翔,王宇飞,等.透析导管相关血流感染的危险因素及风险预测模型[J].中华肾脏病杂志,2022,38(1):23-28.
- [14] BERBUDI A, RAHMADIKA N, TJAHHADI A I, et al. Type 2 diabetes and its impact on the immune system[J]. Curr Diabetes Rev, 2020, 16(5): 442-449.
- [15] 王力红,魏楠,赵霞,等.老年患者中央导管相关血流感染风险预测评分模型构建与验证[J].中国感染控制杂志,2019,18(3):225-231.
- [16] ALSAHOW A, ALHELAL B, ALYOUSEF A, et al. Renal data from the Arab world dialysis in Kuwait: 2013-2019[J]. Saudi J Kidney Dis Transpl, 2020, 31(4): 826-830.
- [17] WANG K, WANG P, LIANG X H, et al. Epidemiology of haemodialysis catheter complications: A survey of 865 dialysis patients from 14 haemodialysis centres in Henan Province in China[J]. BMJ Open, 2015, 5(11): e007136.
- [18] MARIK P E, FLEMMER M, HARRISON W. The risk of catheter-related bloodstream infection with femoral venous catheters as compared to subclavian and internal jugular venous catheters: A systematic review of the literature and meta-analysis[J]. Crit Care Med, 2012, 40(8): 2479-2485.
- [19] 张红娟,赵丽萍,赵素芳.维持性血液透析病人中心静脉留置导管相关性血流感染现状及其危险因素调查研究[J].全科护理,2021,19(4):553-556.
- [20] 陈敏.血液透析患者感染危险因素分析及风险评估模型的建立[D].南昌:南昌大学,2022.
- [21] VAN DER PLOEG T, AUSTIN P C, STEYERBERG E W. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints[J]. BMC Med Res Methodol, 2014, 14: 137.

收稿日期:2024-04-16

(本文编辑:方能)