

• 临床研究 • doi:10.3969/j.issn.1671-8348.2024.24.018

网络首发 [https://link.cnki.net/urlid/50.1097.R.20241030.1336.004\(2024-10-30\)](https://link.cnki.net/urlid/50.1097.R.20241030.1336.004(2024-10-30))

基于机器学习的消化道手术患者中心静脉导管相关性血栓风险模型的构建^{*}

范连娣¹, 王 宁^{2△}, 郭振江¹, 刘防震¹, 崔朝勃²

(衡水市人民医院:1. 胃肠外科;2. 呼吸与危重症医学科, 河北衡水 053000)

[摘要] **目的** 探讨消化道手术患者中心静脉导管相关性血栓形成的危险因素,并基于机器学习算法构建风险预测模型。**方法** 选取 2018 年 5 月至 2024 年 3 月该院接受消化道手术并留置中心静脉导管的患者 385 例为研究对象,根据是否形成导管相关性血栓分为血栓组($n=62$)和非血栓组($n=323$),收集患者年龄、BMI、合并症、现患肿瘤、中性粒细胞/淋巴细胞比值(NLR)、手术时间、置管静脉、系统性免疫炎症指数(SII)、D-二聚体及导管留置时间,比较两组间基线资料的差异。将研究对象按 7:3 比例随机分为训练集和测试集,基于训练集建立 logistic 回归模型及随机森林、支持向量机、决策树和朴素贝叶斯风险预测模型,比较不同模型在测试集中预测导管相关性血栓时受试者操作特征曲线下面积(AUC)、准确度、灵敏度、特异度及 F1 值,对最佳预测模型中预测因素的重要性进行可视化排序。**结果** 两组基线资料中现患肿瘤患者所占比例、NLR、手术时间及 D-二聚体水平比较,差异有统计学意义($P<0.05$)。5 种风险预测模型的 AUC 值从大到小依次为随机森林(0.773)、logistic 回归模型(0.734)、支持向量机(0.680)、朴素贝叶斯(0.666)和决策树(0.650),其中随机森林模型的准确度(0.853)、灵敏度(0.599)、特异度(0.877)和 F1 值(0.414)均为最高。D-二聚体、手术时间、现患肿瘤、NLR 是随机森林模型中前 4 重要预测因素。**结论** 所建立消化道手术患者中心静脉导管相关性血栓的随机森林模型显示出良好性能,D-二聚体、手术时间、现患肿瘤、NLR 是主要的预测因素。

[关键词] 消化道手术;中心静脉导管;静脉血栓形成;机器学习;预测模型**[中图法分类号]** R656.6 **[文献标识码]** A **[文章编号]** 1671-8348(2024)24-3779-05

Construction of risk model for central venous catheter-related thrombosis based machine learning in patients undergoing gastrointestinal tract surgery^{*}

FAN Liandi¹, WANG Ning^{2△}, GUO Zhenjiang¹, LIU Fangzhen¹, CUI Chaobo²

(1. Department of Gastrointestinal Surgery; 2. Department of Respiratory and Critical Care Medicine, Hengshui Municipal People's Hospital, Hengshui, Hebei 053000, China)

[Abstract] **Objective** To explore the risk factors of central venous catheter related thrombosis in the patients undergoing gastrointestinal tract surgery, and to construct a risk prediction model based on machine learning algorithms. **Methods** A total of 385 patients receiving gastrointestinal tract surgery and central venous catheter indwelling in this hospital from May 2018 to March 2024 were selected as the study subjects and divided into the thrombus group ($n=62$) and non-thrombus group ($n=323$) based on whether or not the catheter-related thrombosis forming. The age, body mass index (BMI), comorbidities, current tumors, neutrophil/lymphocyte ratio (NLR), surgery time, catheterization vein, systemic immune inflammation index (SII), D-dimer and catheter indwelling time of the patients were collected, and the differences in baseline data were compared between the two groups. The research subjects were randomly divided into the training set and testing set by a 7:3 ratio. Based on the training set, the logistics regression model, random forest, support vector machine, decision tree and naive Bayes risk prediction models were established. The area under the operating characteristic curve (AUC), accuracy, sensitivity, specificity and F1 value in predicting catheter-related thrombosis were compared among different models in the testing set. The importance of the predictive factors in the best prediction model conducted the visualized ranking. **Results** There were statistically significant differences in the proportion of tumor patients, NLR, surgical time and D-dimer level in the baseline data between the two groups (all $P<0.05$). The AUC values of the five risk prediction models from great to small were the random

^{*} 基金项目:河北省医学科学研究课题计划(20230262)。 [△] 通信作者, E-mail: wcc19871107@163.com。

forest (0.773),logistics regression model (0.734),support vector machine (0.680),naive Bayes (0.666) and decision tree (0.650). Among them,the accuracy (0.853),sensitivity (0.599),specificity (0.877) and F1 value (0.414) of the random forest model were the highest. D-dimer,surgery time,current tumor and NLR were the top four important predictive factors in the random forest model. **Conclusion** The constructed random forest model for central venous catheter-related thrombosis in the patients undergoing gastrointestinal tract surgery demonstrates good performance,and the D-dimer,surgery time,current tumor and NLR are the main predictive factors.

[Key words] gastrointestinal tract surgery;central venous catheter;venous thrombosis formation;machine learning;prediction model

中心静脉导管(CVC)可为消化道手术患者提供中长期静脉通路,满足患者快速静脉补液、全肠外营养及监测中心静脉压力的临床需求^[1-2]。然而 CVC 的应用可能出现导管相关性血栓(CRT),导致导管功能障碍^[3],影响原发病的治疗,增加护理成本^[4],甚至出现肺栓塞等严重后果^[5]。及时发现和处理 CRT 对减轻不良后果至关重要^[6-7]。多数 CRT 是无症状的,仅能通过影像学技术确诊,有的仅表现为 CVC 导管堵塞,只有少数患者出现症状,表现为血栓性静脉炎或静脉回流受阻的症状^[8-9]。对所有患者进行实时的血栓发展动态监测难度较大。近年来,机器学习算法与医学临床研究的结合改变了风险预测和决策过程。机器学习算法能够分析复杂的数据集,增加模型泛化性能,从而提高预测准确性^[10-12]。现有的 CRT 风险预测模型通常依赖于传统的统计方法,这可能会忽略预测变量之间的非线性关系和复杂的交互作用,包括随机森林、支持向量机和决策树等多种机器学习算法在处理高维数据和变量之间的复杂关系方面表现更好的预测性能。

本研究通过收集各种人口统计学、临床和手术因素等临床数据,建立消化道手术患者传统的 logistic 回归和应用机器学习算法开发 CRT 风险预测模型,比较不同模型的预测性能,为 CRT 患者早期实施有针对性的预防及干预措施提供依据^[13-14]。

1 资料与方法

1.1 一般资料

选择 2018 年 5 月至 2024 年 3 月在本院实施消化道手术并留置 CVC 的 385 例患者为研究对象。纳入标准:(1)接受消化道手术的患者;(2)围手术期置入 CVC;(3)年龄≥18 岁;(4)CVC 使用时间≥3 d。排除标准:(1)存在 CVC 禁忌证;(2)既往有上肢深静脉血栓病史;(3)急诊手术。本研究取得本院伦理委员会批准(审批号:2018-1-007),患者及家属均签署知情同意书。

1.2 方法

1.2.1 CVC 置入

采用标准方法完成 CVC(美国泰利福公司,ARROW CS-27702-E)置入,即由麻醉医师于右侧颈内静脉或锁骨下静脉完成 CVC 置入。

1.2.2 预测变量

收集的变量包括年龄、BMI、合并症(高血压、糖尿病、高脂血症或其他慢性病)、现患肿瘤、置管静脉、中性粒细胞/淋巴细胞比值(NLR)、手术时间、系统性免疫炎症指数(SII)、D-二聚体、导管留置时间。其中手术时间为从麻醉开始至麻醉结束时间。NLR=中性粒细胞计数/淋巴细胞计数;SII=血小板计数×中性粒细胞计数/淋巴细胞计数。血小板计数、中性粒细胞、淋巴细胞以及 D-二聚体分别从入院 24 h 内血常规及凝血功能中获取相应的数值。

1.2.3 CRT 诊断标准

对置入 CVC 的患者常规超声筛查,CRT 形成的诊断标准为:(1)探头加压时静脉管腔不能压瘪;(2)CVC 内、导管壁或静脉管腔内可探及实性回声;(3)彩色多普勒未检测到血流信号。超声诊断为 CRT 患者纳入血栓组,否则纳入非血栓组。

1.2.4 CRT 治疗策略

CRT 目前尚无标准治疗。本中心根据指南推荐,对 CVC 位置正常、功能良好且未发生导管相关感染的患者,继续留置导管的同时给予 3 个月以上的治疗量抗凝,抗凝药物选用低分子肝素类或利伐沙班。若临床不再需要应用 CVC,则抗凝 1~2 周后拔除导管^[15]。

1.2.5 风险预测模型建立

将研究对象按 7:3 比例随机分为训练集和测试集,使用最值归一化法对训练集中连续变量进行预处理,以减少特征规模差异对模型的影响。基于训练集建立传统的 logistic 回归模型、随机森林、支持向量机、决策树和朴素贝叶斯 5 种风险预测模型。

1.2.6 不同模型的比较

比较不同预测模型在测试集中的受试者特征工作特征曲线下面积(AUC)、准确度、灵敏度、特异度及 F1 值。在实际应用中,F1 值被广泛用于机器学习模型的评估,尤其在处理正负样本比例严重不平衡时,本研究中血栓组和非血栓组样本存在不平衡性,因此 F1 值能够更准确地反映模型在识别正例方面的性能。F1 值越高,说明模型的性能越好^[16]。对最佳预测模型中预测因素的重要性进行可视化排序。

1.3 样本量计算

参考 Kendall 样本量估算方法,选取样本量是研究因子 10~20 倍,本研究共 11 个自变量,考虑增加 20%样本流失率,最低样本量 $n=132$ 例,本研究最终纳入 385 例患者。

1.4 统计学处理

应用 R4.2.3 软件进行统计学分析。计数资料以例数或百分数表示,组间比较采用 χ^2 检验;非正态分布的计量资料采用 $M(Q_1, Q_3)$ 表示,组间比较采用 Mann-Whitney U 非参数检验。应用 rsample 包进行数据拆分, tidymodels 包建立 5 种风险预测模型, pPOC 包计算模型的 AUC。以 $P<0.05$ 为差异有统

计学意义。

2 结 果

2.1 CRT 发生率

本研究共纳入了 385 例患者,其中男 223 例,女性 162 例;年龄 22~83 岁,中位年龄 65 岁。根据是否形成 CRT 分为血栓组($n=62$)和非血栓组($n=323$),CRT 的发生率为 16.1%(62/385)。

2.2 两组间影响因素的比较

两组间现患肿瘤所占比例、NLR、手术时间及 D-二聚体水平比较差异均有统计学意义($P<0.05$),见表 1。

表 1 血栓组和非血栓组患者影响因素的比较

变量	血栓组($n=62$)	非血栓组($n=323$)	χ^2/U	P
年龄[$n(\%)$]			0.126	0.723
<65 岁	34(54.8)	185(57.3)		
≥ 65 岁	28(45.2)	138(42.7)		
性别[$n(\%)$]			1.207	0.272
男	32(51.6)	191(59.1)		
女	30(48.4)	132(40.9)		
BMI[$n(\%)$]			0.715	0.398
<25 kg/m ²	29(46.8)	170(52.6)		
≥ 25 kg/m ²	33(53.2)	153(47.4)		
合并症[$n(\%)$]			0.507	0.776
1 种	20(32.3)	114(35.3)		
2 种	23(37.1)	105(32.5)		
≥ 3 种	19(30.6)	104(32.2)		
现患肿瘤[$n(\%)$]			5.462	0.019
是	31(50.0)	212(65.6)		
否	31(50.0)	111(34.4)		
置管静脉[$n(\%)$]			0.132	0.716
锁骨下静脉	28(45.2)	154(47.7)		
颈内静脉	34(54.8)	169(52.3)		
NLR[$M(Q_1, Q_3)$]	2.55(2.25, 2.85)	2.25(1.88, 2.70)	-3.879	<0.001
手术时间[$M(Q_1, Q_3)$, 分]	140.00(130.00, 150.00)	160.00(140.00, 180.00)	-5.353	<0.001
SHI[$M(Q_1, Q_3)$]	1 033.50(416.75, 1 783.25)	667.00(267.00, 1 400.00)	-1.939	0.053
D-二聚体[$M(Q_1, Q_3)$, mg/L]	0.84(0.64, 0.94)	0.57(0.44, 0.76)	-6.054	<0.001
导管留置时间[$M(Q_1, Q_3)$, d]	18.00(10.00, 26.25)	16.00(10.00, 24.00)	-0.291	0.771

2.3 风险模型预测性能比较

不同风险预测模型的 AUC 值从大到小依次为随机森林(0.773)、logistic 回归模型(0.734)、支持向量机(0.680)、朴素贝叶斯(0.666)、决策树(0.650)。随机森林模型的准确度(0.853)、灵敏度(0.599)、特异度(0.877)和 F1 值(0.414)均为最高,见表 2。

2.4 特征重要性排列

进一步对随机森林模型中的预测变量重要性进行了可视化排序,从上到下变量的重要性依次减小。结果显示,D-二聚体、手术时间、现患肿瘤、NLR 是位

于前 4 位的重要预测因素,见图 1。

表 2 不同风险模型预测性能比较

模型	准确度	灵敏度	特异度	F1 值
随机森林	0.853	0.599	0.877	0.414
logistic 回归模型	0.819	0.428	0.873	0.364
决策树	0.836	0.500	0.868	0.345
支持向量机	0.845	0.571	0.862	0.308
朴素贝叶斯	0.802	0.333	0.856	0.258

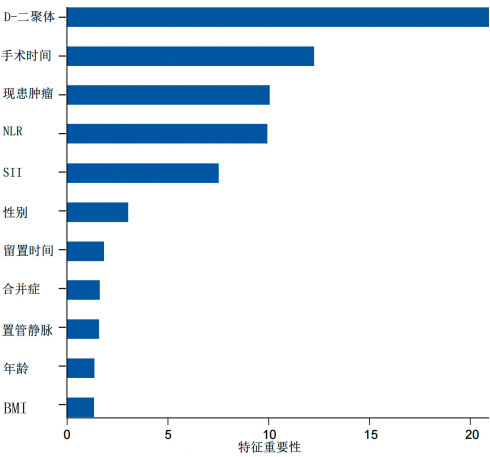


图 1 随机森林预测模型变量重要性排序

3 讨 论

CVC 是消化道手术患者常用的静脉通路,在临床实践中发挥着重要的作用。然而,CVC 的置入与静脉血栓形成风险的增加有关。血栓形成可能会影响疾病的治疗,甚至导致肺栓塞等严重后果^[17]。早期识别发生该并发症的高危患者至关重要^[18]。然而,大多数 CRT 患者是无症状的。即便是有症状发现也比较晚,往往导致不良临床结果^[19]。因此,在胃肠外科置入 CVC 的患者中,一种简单、可靠、准确的评估 CRT 个体风险的方法至关重要^[20]。

本研究中单因素分析及随机森林模型中变量重要性中排名为前 4 位的 CRT 影响因素均为现患肿瘤、NLR、手术时间和 D-二聚体水平。同样有研究表明,肿瘤患者住院期间发生了静脉血栓栓塞的风险显著高于非肿瘤患者^[21],CVC 的置入进一步加剧了静脉血栓形成的风险。此外,手术后的患者更容易出现静脉血栓,手术会直接损伤血管壁,出现血流减慢,手术损伤后组织因子入血激活了体内的凝血系统,使血液凝固性增高。2022 年发布的《普通外科患者静脉血栓栓塞症风险评估与预防护理专家共识》^[22]中指出,对于普通外科手术而言,手术时间越长,患者静脉血栓栓塞症的发生风险越高。本研究也显示,手术时间越长 CRT 发生的风险越高。D-二聚体来源于纤溶酶溶解的交联纤维蛋白凝块,主要反映纤维蛋白溶解功能。D-二聚体的临床检测主要应用在静脉血栓栓塞、深静脉血栓形成和肺栓塞的诊断。D-二聚体升高提示更高的静脉血栓风险,同样本研究中基线 D-二聚体水平越高,提示 CRT 的风险越高^[23-24]。炎症不仅参与动脉血栓形成的病理生理过程,而且对静脉血栓形成过程中的血栓前状态有影响。有研究表明,炎症反应在静脉血栓的发病机制中起着重要作用,其代表性的炎症标志物包括 NLR、PLR 和 SII。中性粒细胞在静脉血栓栓塞中的作用已在许多研究中得到证实^[25]。NLR 是一种非特异性炎症标记物,代表炎症中中性粒细胞和淋巴细胞之间的关系,可从常规实验室研究中获得,并提供有关全身炎症状态的重要信息。NLR 越高,说明中性粒细胞计数越高,淋巴细胞计数越低。

中性粒细胞数量的增加表示全身炎症过程,而淋巴细胞数量的减少则表示疾病造成的持续应激^[26],因此,较高的 NLR 可能提示静脉血栓栓塞。本研究显示了机器学习算法(尤其是随机森林)在预测 CRT 风险方面的临床实用性。虽然传统的 logistic 回归模型仍是风险预测建模的重要工具,但机器学习技术在处理预测变量之间复杂的非线性关系方面具有更多优势。本研究通过对不同模型的比较,随机森林模型的 AUC 值及 F1 值均优于其他模型,展现出更好的预测性能,可成为消化道手术患者在留置 CVC 期间 CRT 风险分层的重要工具。本研究存在一些局限性。首先,本研究为回顾性研究,尚需前瞻性研究验证其预测性能,以进一步提高其准确性和可推广性。其次,该研究是单中心研究,需开展涉及不同患者群体的多中心研究及检测模型的外部有效性。最后,未来的模型应考虑纳入更多的生物标志物和影像学参数。

综上所述,本研究建立的随机森林风险预测模型,可早期识别消化道手术 CRT 高危患者,从而实施有针对性的血栓预防策略,以降低 CRT 风险并改善患者预后。

参考文献

[1] SRIDHAR C D, ABOU-LSMAIL M Y, AHUJA S P. Central venous catheter-related thrombosis in children and adults[J]. Thromb Res, 2020,187:103-112.

[2] LASAGNI D, NOSADINI M, MOLINARI A C, et al. Systemic catheter-related venous thromboembolism in children: data from the italian registry of pediatric thrombosis[J]. Front Pediatr, 2022,10:843643.

[3] LEUNG A, HEAL C, PERERA M, et al. A systematic review of patient-related risk factors for catheter-related thrombosis[J]. J Thromb Thrombolysis, 2015,40(3):363-373.

[4] CHEN Y, CHEN H, YANG J, et al. Patterns and risk factors of peripherally inserted central venous catheter-related symptomatic thrombosis events in patients with malignant tumors receiving chemotherapy[J]. J Vasc Surg Venous Lymphat Disord, 2020,8(6):919-929.

[5] 王宁, 郭振江, 张媛媛, 等. 定期超声检查在中心静脉置入设备相关深静脉血栓诊治中的应用价值[J]. 山东大学学报(医学版), 2021,59(7):63-67.

[6] ABDEL-RAZEQ H, AL-JAGHBEER M J. Primary thromboprophylaxis for the prevention of venous thromboembolism in cancer patients with central venous catheters: a literature review[J]. J Clin Med, 2024,13(6):1660.

- [7] ELIAS A, DEBOURDEAU P, ESPITIA O, et al. Central venous catheter associated upper extremity deep vein thrombosis in cancer patients: diagnosis and therapeutic management [J]. Arch Cardiovasc Dis, 2024, 117(1): 72-83.
- [8] GOLDENBERG N A, DONADINI M P, KAHN S R, et al. Post-thrombotic syndrome in children: a systematic review of frequency of occurrence, validity of outcome measures, and prognostic factors [J]. Haematologica, 2010, 95(11): 1952-1959.
- [9] GIORDANO P, SARACCO P, GRASSI M, et al. Recommendations for the use of long-term central venous catheter (CVC) in children with hemato-oncological disorders: management of CVC-related occlusion and CVC-related thrombosis. On behalf of the coagulation defects working group and the supportive therapy working group of the Italian Association of Pediatric Hematology and Oncology (AIEOP) [J]. Ann Hematol, 2015, 94(11): 1765-1776.
- [10] JANDY K, WEICHBROTH P. A machine learning approach to classifying New York Heart Association (NYHA) heart failure [J]. Sci Rep, 2024, 14(1): 11496.
- [11] HASSAN A, GULZAR AHMAD S, ULLAH MUNIR E, et al. Predictive modelling and identification of key risk factors for stroke using machine learning [J]. Sci Rep, 2024, 14(1): 11498.
- [12] 李惠萍, 胡安民. 机器学习算法对危重患者死亡率预测效能的比较 [J]. 广东医学, 2020, 41(16): 1642-1646.
- [13] 郭振江, 王宁, 张媛媛, 等. 利伐沙班预防经外周静脉穿刺中心静脉置管相关上肢静脉血栓疗效探讨 [J]. 天津医药, 2021, 49(9): 963-967.
- [14] 王宁, 郭振江, 张媛媛, 等. 利伐沙班与低分子肝素治疗中心静脉置入设备相关性上肢深静脉血栓的疗效比较 [J]. 天津医药, 2021, 49(4): 432-435.
- [15] MARIN A, BULL L, KINZIE M, et al. Central catheter-associated deep vein thrombosis in cancer: clinical course, prophylaxis, treatment [J]. BMJ Support Palliat Care, 2021, 11(4): 371-380.
- [16] WANG J, SOURLOS N, HEUVELMANS M, et al. Explainable machine learning model based on clinical factors for predicting the disappearance of indeterminate pulmonary nodules [J]. Comput Biol Med, 2024, 169: 107871.
- [17] FALLOUH N, MCGUIRK H M, FLANDERS S A, et al. Peripherally inserted central catheter-associated deep vein thrombosis: a narrative review [J]. Am J Med, 2015, 128(7): 722-738.
- [18] PARKER R I. Catheter-associated deep venous thrombosis prevention: which path to choose? [J]. Crit Care Med, 2021, 49(3): 537-540.
- [19] MARIN A, BULL L, KINZIE M, et al. Central catheter-associated deep vein thrombosis in cancer: clinical course, prophylaxis, treatment [J]. BMJ Support Palliat Care, 2021, 11(4): 371-380.
- [20] LIN S, ZHU N, ZHANG Y, et al. Development and validation of a prediction model of catheter-related thrombosis in patients with cancer undergoing chemotherapy based on ultrasonography results and clinical information [J]. J Thromb Thrombolysis, 2022, 54(3): 480-491.
- [21] KHORANA A A, FRANCIS C W, CULAKOVA E, et al. Frequency, risk factors, and trends for venous thromboembolism among hospitalized cancer patients [J]. Cancer, 2007, 110(10): 2339-2346.
- [22] 中华护理学会外科护理专业委员会, 中华医学会外科学分会护理学组. 普通外科患者静脉血栓栓塞症风险评估与预防护理专家共识 [J]. 中华护理杂志, 2022, 57(4): 444-449.
- [23] LOFFREDO L, VIDILI G, SCIACQUA A, et al. Asymptomatic and symptomatic deep venous thrombosis in hospitalized acutely ill medical patients: risk factors and therapeutic implications [J]. Thromb J, 2022, 20(1): 72.
- [24] 陈颖, 秦贤, 孙乔, 等. 危重患者下肢深静脉血栓风险预测模型的构建及评价 [J]. 护理学杂志, 2021, 36(6): 35-38.
- [25] HU J, CAI Z, ZHOU Y. The association of neutrophil-lymphocyte ratio with venous thromboembolism: a systematic review and meta-analysis [J]. Clin Appl Thromb Hemost, 2022, 28: 10760296221130061.
- [26] KREMERS B, WÜBBEKE L, MEES B, et al. Plasma biomarkers to predict cardiovascular outcome in patients with peripheral artery disease: a systematic review and meta-analysis [J]. Arterioscler Thromb Vasc Biol, 2020, 40(9): 2018-2032.