DOI:10.16016/j.2097-0927.202410090

基于随机生存森林算法的NK/T细胞淋巴瘤患者预后模型的构建与验证

黄 侦[1,2],伍亚舟[1] 陆军军医大学(第三军医大学):[1]军事预防医学系军队卫生统计学教研室,[2]第一附属医院血液病中心,重庆

[摘要] **目的** 探讨NK/T细胞淋巴瘤(natural killer T-cell lymphoma,NKTL)患者的生存预后影响因素。基于随机生存森林(random survival forest,RSF)算法,构建预测NKTL患者总生存期(overall survival,OS)的预后模型。**方法** 从SEER数据库收集2000-2020年的NKTL患者的人口统计学和临床病理资料。按照7:3的比例将患者划分为训练队列(n=471)和验证队列(n=203)。通过Cox回归分析确定影响患者OS的预后因素,并基于分析结果构建列线图模型。同时,使用RSF算法确定影响患者OS的预后因素,并构建RSF模型。采用ROC曲线、校准曲线和决策曲线、净重新分类指数(net reclassification improvement,NRI)和综合判别改善指数(integrated discrimination improvement,IDI)对模型的预测能力进行评估,并比较2种模型的预测效果。通过2种模型计算每位患者的风险得分,根据风险得分的中位数,将患者分为高风险组和低风险组,并绘制生存曲线进行比较。**结果** Ann Arbor分期、年龄、放疗、联合治疗和疾病类型是与生存显著相关的预后变量。在验证队列中,列线图模型的1、3和5年ROC曲线下面积(area under curve, AUC)分别为0.745、0.771和0.748,而RSF模型的AUC为0.764、0.792和0.761。ROC曲线显示2种模型在预测OS方面具有良好的准确性和区分性。校准曲线显示2种模型预测的生存与实际生存之间具有良好的一致性。两种模型均能有效地将患者分为预后差组和预后好组,并且预后差组患者的OS显著低于预后好组(P<0.000 1)。决策曲线显示RSF模型的净获益优于列线图模型。与列线图模型相比,RSF模型的NRI为0.184(95%CI:0.098~0.267,P<0.01),IDI为0.300(95%CI:0.241~0.359,P<0.01),RSF模型的预测能力优于列线图模型。**结论** Ann Arbor分期、年龄、放疗、联合治疗和疾病类型是NKTL患者预后的影响因素,本研究据此建立的RSF模型对NKTL患者预后具有很好的预测能力,可有效评估患者预后。

[关键词] NK/T细胞淋巴瘤;SEER数据库;列线图;机器学习;随机生存森林;生存预测
[中图法分类号] R195.1;R730.7;R733.4 [文献标志码] A

Construction and validation of a prognostic model for NK/T-cell lymphoma based on random survival forest algorithm

HUANG Zhen[1,2], WU Yazhou[1] [1]Department of Health Statistics, Faculty of Military Preventive Medicine, Army Medical University (Third Military Medical University), Chongqing; [2]Center for Hematology, First Affiliated Hospital, Army Medical University (Third Military Medical University), Chongqing, China

[**Abstract**] **Objective** To investigate the prognostic factors affecting survival in patients with natural killer T-cell lymphoma (NKTL), and then develop a prognostic model for predicting their overall survival (OS) based on random survival forest (RSF) algorithm. **Methods** Demographic and clinical pathological data of NKTL patients were collected from the SEER database during 2000 and 2020. The patients were divided into

[基金项目] 国家自然科学基金面上项目(82173621,81872716)
[通信作者] 伍亚舟,E-mail:asiawu@tmmu.edu.cn

2097-0927/Copyright © 2025, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循CC BY许可协议(https://creativecommons.org/licenses/by/4.0/)
引用格式:黄 侦,伍亚舟. 基于随机生存森林算法的NK/T细胞淋巴瘤患者预后模型的构建与验证[J]. 陆军军医大学学报,2025,47(03):275-284. DOI:10.16016/j.2097-0927.202410090.
Citation:HUANG Z, WU Y Z. Construction and validation of a prognostic model for NK/T-cell lymphoma based on random survival forest algorithm[J]. J Army Med Univ,2025,47(03):275-284. DOI:10.16016/j.2097-0927.202410090.

a training cohort (n=471) and a validation cohort (n=203) in a 7∶3 ratio. Cox regression analysis was performed to identify prognostic factors affecting OS, and a nomogram model was constructed based on the obtained factors. Meanwhile, RSF algorithm was used to determine prognostic factors affecting OS to build the RSF model. The models were evaluated using receiver operating characteristic (ROC) curve, calibration curve, decision curve, net reclassification improvement (NRI), and integrated discrimination improvement (IDI), and the predictive performances of the 2 models were compared. Risk scores for each patient were calculated using the 2 models. Then the patients were divided into high- and low-risk groups based on the median risk score, and survival curve was plotted for comparison. **Results** Ann Arbor stage, age, radiotherapy, combined treatment, and type of disease were identified as significant prognostic variables associated with OS. In the validation cohort, the area under the ROC curve (AUC) for the nomogram model at 1, 3, and 5 years was 0.745, 0.771, and 0.748, respectively, while the AUC for the RSF model was 0.764, 0.792, and 0.761 at the same time points. ROC curve analysis indicated that both models demonstrated good accuracy and discrimination in predicting OS. Calibration curve analysis showed a strong consistency between the predicted and actual OS for both models. Both models effectively stratified the patients into poor and favorable prognosis groups, with the OS of patients in the poor prognosis group being significantly shorter than that of the favorable prognosis group (P<0.000 1). Decision curve analysis revealed that the net benefit of the RSF model was superior to that of the nomogram model. Compared to the nomogram model, the NRI for the RSF model was 0.184 (95%*CI*: 0.098~0.267, P<0.01), and the IDI was 0.300 (95%*CI*: 0.241~0.359, P<0.01). Overall, the RSF model demonstrated superior predictive capability than the nomogram model. **Conclusion** Ann Arbor stage, age, radiotherapy, combined treatment, and type of disease are prognostic factors affecting the prognosis of NKTL patients. Our RSF model demonstrates strong predictive capability for the prognosis of NKTL patients and can effectively assess patient outcomes.

[**Key words**] natural killer T-cell lymphoma; SEER database; nomogram; machine learning; random survival forest; survival prediction

Supported by the General Program of National Natural Science Foundation of China (82173621, 81872716). Corresponding author: WU Yazhou, E-mail: asiawu@tmmu.edu.cn

NK/T细胞淋巴瘤(natural killer T-cell lymphoma,NKTL)是一种罕见且具有高度侵袭性的非霍奇金淋巴瘤,与EB病毒感染密切相关[1]。临床多以鼻塞、流涕、鼻出血为首发症状,常伴发热、盗汗、体质量减轻等症状。NKTL多表现为结外侵犯,一般可分为鼻型(主要侵犯鼻腔和上呼吸道)、非鼻型(涉及皮肤、胃肠道、睾丸等器官)以及侵袭性白血病/淋巴瘤亚型(影响骨髓和多个器官)[2]。对于早期局限性NKTL,通常采用放疗或与化疗联合使用,5年总生存率约为70%[3]。然而,对于晚期和复发NKTL,一般采取多学科治疗,包括基于天冬酰胺酶的化疗、放疗和造血干细胞移植巩固治疗等,5年总生存率约为35.1%[4]。总体而言,NKTL预后不良,尤其是在晚期和复发患者中,这突显了对精准生存预测模型的迫切需求,以帮助临床医生及时调整治疗策略,积极应对疾病进展并改善患者预后。

近年来,随着机器学习的快速发展,其应用潜力和实际效果愈发显著。机器学习不仅提高了疾病诊断的准确性和效率,还在治疗方案选择和疾病预后预测等方面展现出卓越能力[5-8]。在众多机器学习算法中,随机生存森林(random survival forest,RSF)逐渐成为生存分析领域的热门工具[9-12]。RSF不仅具备优异的预测性能,还能够提供变量重要性评估,这可以让临床医生直观地识别重要的生存预后影响因素,从而更加精准地进行临床干预。目前,RSF已广泛应用于肺癌、肝癌、乳腺癌等多种疾病的预后评估,然而在NKTL的研究中使用较少。

因此,本研究利用美国监测、流行病学与最终结果(Surveillance, Epidemiology, and End Results,SEER)数据库[13],对NKTL患者的年龄、性别、疾病类型、Ann Arbor分期及治疗方式等一系列预后影响因素进行深入分析。基于这些分析结果,本研究构建了列线图模型和RSF模型,评估了这2种模型的预测价值,并比较了它们的预测效果。本研究旨在为临床医生提供更有效的精准识别工具,从而制定更加个性化和有效的治疗方案,提高患者的疗效和生存率。

1 资料与方法

1.1 研究人群

在本研究中,使用SEER*Stat软件从SEER数据库筛选被诊断为NKTL的患者。纳入标准:①患者必须被正式诊断为NKTL;②确诊时间为2000-2020年。排除标准:①NKTL非原发恶性肿瘤;②确诊年龄<18岁;③生存时间<1个月;④Ann Arbor分期、种族、原发部位、婚姻状况或随访信息等未知。本研究获得SEER数据库的访问权限与使用许可。

1.2 研究变量

本研究从SEER数据库中提取了以下变量:年龄、性别、种族、疾病类型、原发部位、Ann Arbor分期、B症状、放疗、化疗、联合治疗(指治疗方式包括放疗和化疗)、婚姻状态、确诊年份以及随访信息。总生存期(overall survival, OS),定义为从确诊时间至最后一次随访或因任何原因导致死亡的时间间隔。

1.3 统计学分析

本研究采用SPSS 26.0、R语言4.3.2和Python 3.11软件进行统计学分析。计数资料以例(%)表示,组间比较采用χ^2检验或Fisher精确概率法;有序分类变量通过秩和检验进行组间比较。把从SEER数据库筛选的患者按照7:3的比例随机划分为训练队列和验证队列。把所有变量纳入单变量Cox回归分析,P<0.05的单变量被确定用于多变量Cox回归分析,通过多变量Cox回归分析筛选出独立预后变量,并基于这些变量构建列线图模型。同时,采用RSF算法来确定影响患者OS的预后变量,并构建RSF模型。采用ROC曲线及AUC值来评估模型的区分度,使用校准曲线和决策曲线评估模型的校准度和临床适用性,使用风险分组将患者分为高风险组和低风险组,以便更好地识别高危人群。使用净重新分类指数(net reclassification improvement, NRI)和综合判别改善指数(integrated discrimination improvement, IDI)来比较2种模型的预测效果。采用沙普利加性解释(Shapley Additive exPlanations, SHAP)对RSF模型进行可解释性分析。P<0.05被认为差异具有统计学意义。

2 结果

2.1 基线特征

基于纳入和排除标准,本研究从SEER数据库中筛选出674例符合条件的NKTL患者,按照7:3的比例随机划分为训练队列(n=471)和验证队列(n=203),患者的基本人口学和临床病理特征见表1。训练队列中362例(76.9%)诊断为鼻型,418例(88.7%)为结外原发部位,141例(29.9%)处于Ⅲ~Ⅳ期;接受放疗的患者有299例(63.5%),化疗391例(83.0%),联合治疗234例(49.7%);研究结束时,共有290例患者(61.6%)死亡。验证队列中154例(75.9%)为鼻型,177例(87.2%)为结外原发部位,76例(37.4%)为Ⅲ~Ⅳ期;放疗患者120例(59.1%),化疗175例(86.2%),联合治疗97例(47.8%);研究结束时,125例患者(61.6%)死亡。

表1 NK/T细胞淋巴瘤患者的基线特征表[例(%)]

变量	患者总人数(n=674)	训练队列(n=471)	验证队列(n=203)	P
年龄/岁				0.998
≤60	466(69.1)	326(69.2)	140(69.0)	
>60	208(30.9)	145(30.8)	63(31.0)	
性别				0.145
女性	232(34.4)	151(32.1)	81(39.9)	
男性	442(65.6)	320(67.9)	122(60.1)	
种族				0.558
白种人	485(72.0)	348(73.9)	137(67.5)	
黑种人	28(4.15)	19(4.03)	9(4.43)	
其他	161(23.9)	104(22.1)	57(28.1)	
疾病类型				0.962
非鼻型	158(23.4)	109(23.1)	49(24.1)	
鼻型	516(76.6)	362(76.9)	154(75.9)	
原发部位				0.847
结内	79(11.7)	53(11.3)	26(12.8)	
结外	595(88.3)	418(88.7)	177(87.2)	
Ann Arbor分期				0.161
Ⅰ~Ⅱ	457(67.8)	330(70.1)	127(62.6)	
Ⅲ~Ⅳ	217(32.2)	141(29.9)	76(37.4)	
B症状				0.791
否	205(30.4)	144(30.6)	61(30.0)	
是	117(17.4)	76(16.1)	41(20.2)	
未知	352(52.2)	251(53.3)	101(49.8)	
放疗				0.562
否/未知	255(37.8)	172(36.5)	83(40.9)	
是	419(62.2)	299(63.5)	120(59.1)	
化疗				0.584
否/未知	108(16.0)	80(17.0)	28(13.8)	
是	566(84.0)	391(83.0)	175(86.2)	
联合治疗				0.903
否/未知	343(50.9)	237(50.3)	106(52.2)	
是	331(49.1)	234(49.7)	97(47.8)	
婚姻状态				0.866
未婚	256(38.0)	182(38.6)	74(36.5)	
已婚	418(62.0)	289(61.4)	129(63.5)	
确诊年份/年				0.826
2000-2005	153(22.7)	110(23.4)	43(21.2)	
2006-2020	521(77.3)	361(76.6)	160(78.8)	
生存状态				1.000
存活	259(38.4)	181(38.4)	78(38.4)	
死亡	415(61.6)	290(61.6)	125(61.6)	

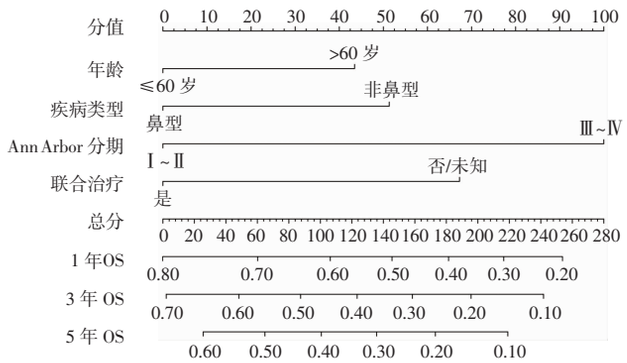


图 1 预测NKTL患者1、3和5年OS的列线图

2.2 列线图模型的构建与验证

在单变量Cox回归分析中,发现年龄、疾病类型、Ann Arbor分期、原发部位、B症状、放疗、联合治疗、确诊年份与NKTL患者的预后相关(P<0.05),将这些变量纳入多变量Cox回归分析,结果显示,年龄>60岁(HR=1.403)、疾病类型为鼻型(HR=0.671)、Ann Arbor分期为Ⅲ~Ⅳ期(HR=2.178)和采取联合治疗(HR=0.592)为NKTL患者OS的独立影响因素(P<0.05,表2)。

根据多变量Cox回归分析确定的独立预后因素,构建了患者1、3和5年的列线图模型(图1)。ROC曲线显示该模型具有良好的准确性和区分性,在训练队列中1、3和5年的AUC分别为0.754、0.723和0.725,而验证队列的AUC为0.745、0.771和0.748(图2)。此外,校准曲线显示该模型预测的生存与实际生存之间具有良好的一致性(图3)。

2.3 RSF模型的构建与验证

机器学习方法中的RSF算法的优势在于能够充分集成多棵树的输出,从而提高模型预测的稳定性和准确性。结果显示,RSF模型在训练到400棵树时性能趋于稳定,并且此时模型获得最优AUC值。最终,构建了一个包含400棵树的RSF模型,以预测NKTL患者的OS。ROC曲线显示该模型在预测OS方面具有良好的准确性和区分性,在训练队列中,1、3和5年的AUC分别为0.791、0.772和0.762,而验证队列的AUC为0.764、0.792和0.761(图4)。此外,校准曲线显示模型预测的生存与实际生存之间具有良好的一致性(图5)。SHAP图显示:Ann Arbor分期、年龄、放疗、联合治疗和疾病类型是5个最显著的预测变量,Ann Arbor分期对预后的影响最大,其次是年龄。具体而言,晚期(Ⅲ~Ⅳ期)和年龄>60岁对模型预测结果产生正贡献,表明具有这些特征的患者死亡概率较高,

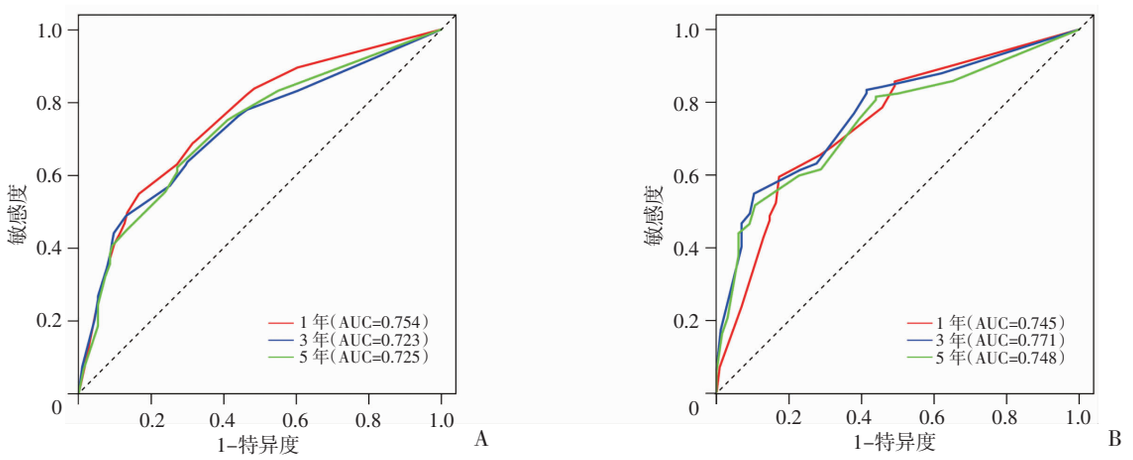
表2 训练队列中NK/T细胞淋巴瘤患者预后的单因素和多因素COX回归分析

变量	人数	单因素分析	多因素分析	
		P	HR(95%CI)	P
年龄/岁		<0.001		0.005
≤60	326		参考	
>60	145		1.403 (1.100~1.789)	
疾病类型		<0.001		0.015
非鼻型	109		参考	
鼻型	362		0.671 (0.507~0.888)	
Ann Arbor分期		<0.001		<0.001
Ⅰ~Ⅱ	330		参考	
Ⅲ~Ⅳ	141		2.178 (1.670~2.841)	
联合治疗		<0.001		<0.001
否/未知	237		参考	
是	234		0.592 (0.457~0.767)	
性别		0.259		
女性	151			
男性	320			
种族		0.774		
白种人	348			
黑种人	19			
其他	104			
原发部位		<0.001		
结内	53			
结外	418			
B症状		0.001		
否	144			
是	76			
未知	251			
放疗		<0.001		
否/未知	172			
是	299			
化疗		0.149		
否/未知	80			
是	391			
婚姻状态		0.966		
未婚	182			
已婚	289			
确诊年份/年		0.030		
2000-2005	110			
2006-2020	361			

预后较差,显著降低OS。相对而言,鼻型、采取放疗和联合治疗则对模型预测结果表现出负贡献,意味着这些患者与较好的预后相关,表明这些治疗措施可以改善患者的OS(图6)。

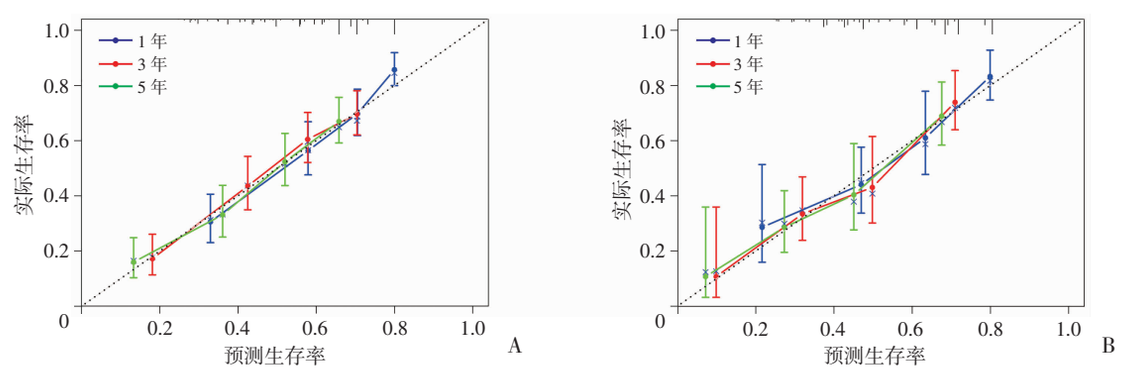
2.4　列线图模型与RSF模型的比较

通过列线图模型和RSF模型计算了训练队列和验证队列中每位患者的风险得分,并根据风险得分的中位数将患者分为高风险组和低风险组,并绘制生存



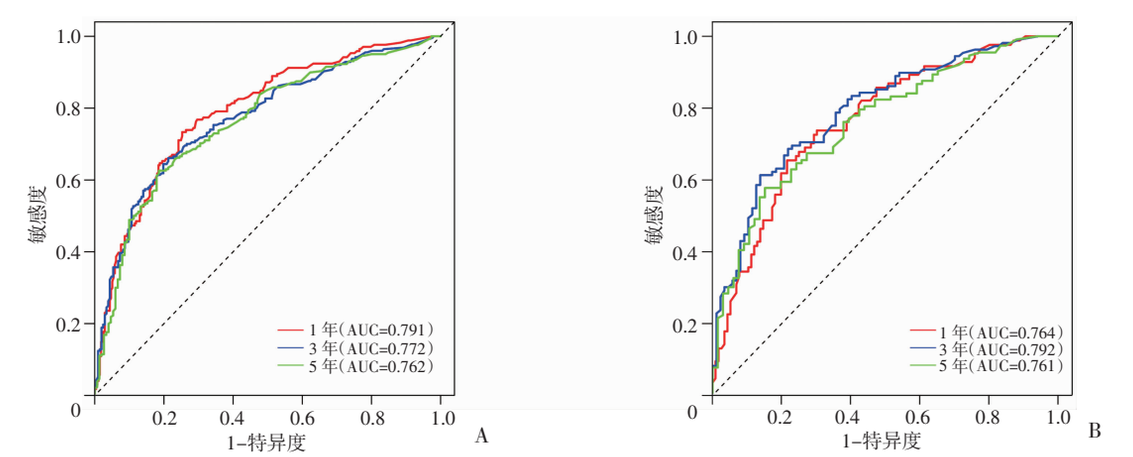
A:训练队列中NKTL患者1、3、5年的ROC曲线及AUC值;B:验证队列中NKTL患者1、3、5年的ROC曲线及AUC值

图2　列线图模型的ROC曲线及AUC值



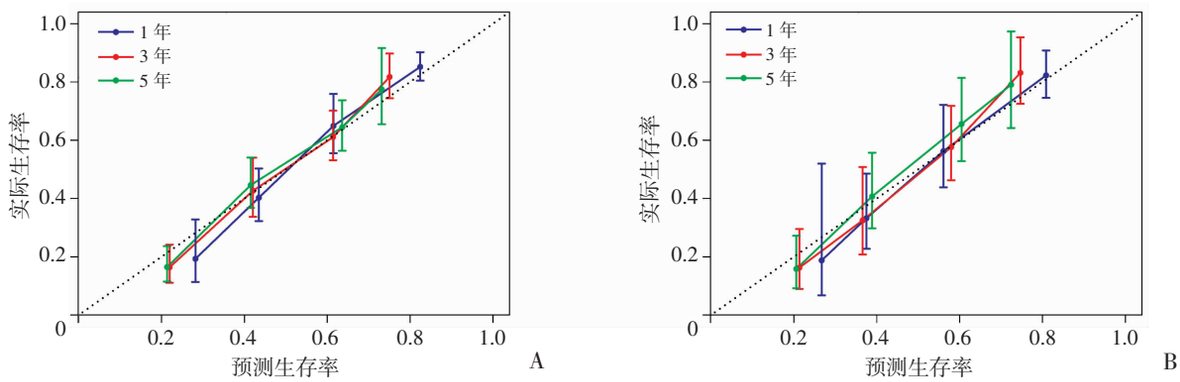
A:训练队列中NKTL患者1、3、5年的校准曲线;B:验证队列中NKTL患者1、3、5年的校准曲线

图3　列线图模型的校准曲线



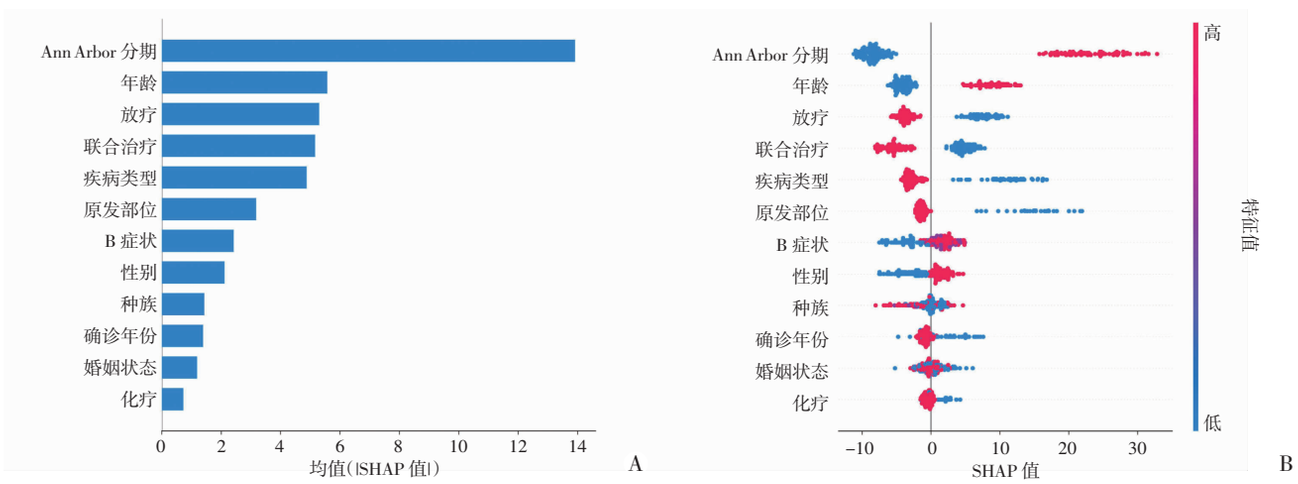
A:训练队列中NKTL患者1、3、5年的ROC曲线及AUC值;B:验证队列中NKTL患者1、3、5年的ROC曲线及AUC值

图4　随机生存森林模型的ROC曲线及AUC值



A:训练队列中NKTL患者1、3、5年的校准曲线;B:验证队列中NKTL患者1、3、5年的校准曲线

图5 随机生存森林模型的校准曲线



A:特征重要性图;B:蜂巢图

图6 随机生存森林模型的SHAP图

曲线。结果显示,2种模型均能有效地将患者区分为预后差组和预后好组,且预后差组的患者OS显著低于预后好组(P<0.000 1,图7)。

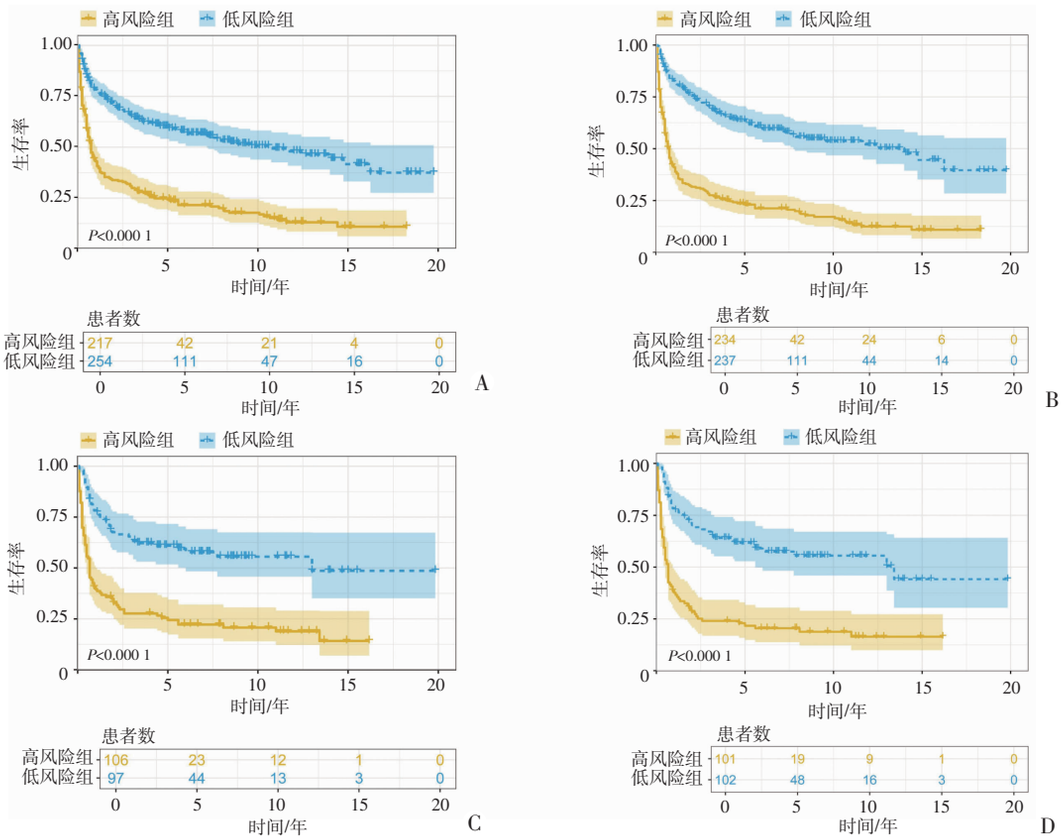
绘制了2种模型的决策曲线。结果显示,在训练队列和验证队列中,列线图模型和RSF模型都具有正的净获益,表明这2种模型在临床应用中均具有良好的潜力。而RSF模型的净获益优于列线图模型(图8)。

计算了RSF模型的NRI和IDI指数。与列线图模型相比,RSF模型的NRI为0.184(95%CI:0.098~0.267,P<0.01),IDI为0.300(95%CI:0.241~0.359,P<0.01),提示RSF模型具有更好的预测能力。

3 讨论

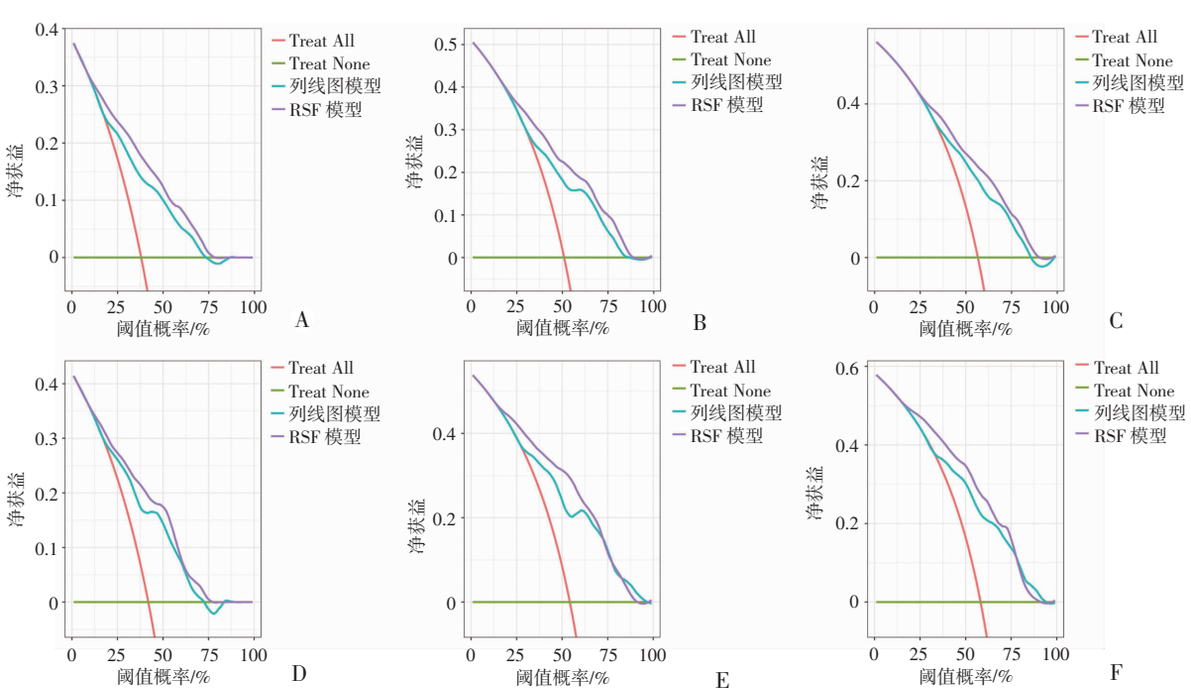
3.1 现有预后模型的比较

NKTL是一种罕见且具有高度侵袭性的非霍奇金淋巴瘤,常表现为鼻腔及鼻窦肿物造成的鼻塞、周围骨质破坏,颌面部肿胀及麻木疼痛感,甚至可能出现鼻中隔和硬腭穿孔等症状[14]。治疗方式通常包括化疗、放疗、造血干细胞移植和新兴的靶向治疗及免疫治疗等。由于该病进展较快,早期干预和准确的预后评估十分重要。在临床实践中,已有一些用于评估NKTL预后的模型或评分工具,包括国际预后指数(International Prognostic Index, IPI)、韩国预后指数(Korean Prognostic Index, KPI)、伴或不伴EBV DNA的NK/T细胞淋巴瘤预后指数(Prognostic Index of Natural Killer Lymphoma with or without EBV DNA, PINK/PINK-E)以及列线图修订风险指数(Nomogram-Revised risk Index, NRI)等。CHEN等[15]报道了NRI模型的性能,验证集中NRI、IPI、KPI和PINK的5年生存预测AUC值分别为0.72、0.61、0.68和0.63。WANG等[16]报道了Nomogram-ES模型,并与IPI、KPI、PINK、PINK-E和NRI进行了比较,其在验证集的AUC值分别为



A、B:训练队列中列线图模型和随机生存森林模型对NKTL患者的风险分组;C、D:验证队列中列线图模型和随机生存森林模型对NKTL患者的风险分组

图 7　列线图模型和随机生存森林模型的风险分组



Treat ALL:示对所有人进行干预;Treat None:示不对任何人进行干预

A~C:训练队列中两模型在1、3和5年的决策曲线;D~F:验证队列中两模型在1、3和5年的决策曲线

图 8　列线图模型和随机生存森林模型的决策曲线

0.76、0.65、0.63、0.65、0.64和0.66。大多数现有模型未能充分考虑患者的个体差异和相关特征,导致预测准确性不足,这些模型在支持个体化治疗方面仍存在一定局限性[15-17]。此外,针对NKTL生存预测的机器学习应用仍显不足。

3.2 RSF模型的分析

RSF是一种机器学习算法,凭借其诸多优势,逐渐成为生存分析和相关领域研究的重要选择[18-20]。RSF无需进行降维或特征选择,可以直接处理高维数据,这意味着研究人员无需手动筛选特征,RSF能够自动识别重要特征,从而降低人为偏差的可能性。其次,RSF基于集成学习的原理,通过构建多棵随机选择样本和特征的决策树,并综合它们的预测结果,提高了模型的准确性和稳定性。这种随机性有效降低了过拟合的风险,并增强了模型对未知数据的泛化能力。此外,RSF表现出良好的鲁棒性,对数据中的噪声和缺失值不敏感。它能够利用其他完整特征的信息来推断缺失数据,从而尽可能保留样本的信息。除了上述优势,RSF的可解释性显著提升了其在实际应用中的价值,常用的解释方法包括SHAP、部分依赖图(partial dependence plot,PDP)和个体条件期望图(individual conditional expectation,ICE)等。SHAP值基于博弈论原理,能够量化每个特征对个体预测结果的贡献,从而提供全局和局部层面的解释。这种方法不仅可以揭示关键特征的重要性,还能直观展示特征对生存时间或风险预测的正负影响。PDP通过绘制特定特征与模型预测之间的关系曲线,清晰地展现该特征对生存概率的平均影响,是分析全局趋势的有力工具。而ICE进一步细化了PDP的分析,能够针对个体绘制特征与预测结果的关系曲线,从而揭示个体层面的异质性。这些方法的结合显著提升了RSF模型预测结果的透明性,为深入理解数据特征和模型行为提供了强有力的支持。本研究对RSF模型的关键参数进行了深入分析,并重点探讨了决策树数量对模型性能的影响。实验结果表明,树的数量在一定范围内对模型性能具有显著影响,但过多的树会显著增加计算成本。通过交叉验证评估,发现当树的数量达到400时,模型性能趋于稳定,进一步增加树的数量对性能的提升并不明显,同时AUC值在此时达到最大。因此,选择400棵树作为模型的参数设置,能够在计算效率、预测性能和过拟合风险之间实现良好的平衡。RSF还能够为个体提供生存函数和风险评估的个性化分析,这在临床决策中具有重要意义,可以帮助医生根据患者的具体情况制定个性化的治疗方案,从而提高临床诊疗的精准性。

本研究构建的RSF模型经SHAP分析显示,Ann Arbor分期、年龄、放疗、联合治疗和疾病类型是5个最显著的预测变量。Ann Arbor分期是最重要的变量,晚期患者(Ⅲ~Ⅳ期)的预后明显较差。既往研究也指出,晚期NKTL患者通常与较低的生存率相关[1]。晚期通常意味着肿瘤已经扩散至其他部位或涉及多个区域,表明疾病的进展程度较重,患者往往面临更大的治疗难度。因此,早发现和早干预对改善预后至关重要。其次,年龄>60岁与较差的预后相关,这可能与老年患者的免疫系统功能下降、合并症增加以及对治疗的耐受性较差有关[21]。此外,发现非鼻型NKTL患者的预后明显差于鼻型患者,这与既往研究结果一致[22]。鼻型NKTL的病理特征较明确、局限性强、早期症状明显,易于早期诊断和治疗,且对标准治疗的响应较好。而非鼻型患者可能表现为更广泛的浸润和复杂临床症状,导致诊断和治疗的延误,进而影响预后。在治疗策略上,放疗对早期NKTL患者有效,联合治疗显著改善患者预后,多种治疗手段的结合能够有效提高患者的生存率,制定个体化的治疗方案对提高NKTL患者预后至关重要[23]。患者的风险分层对于指导患者管理具有重要意义,对列线图模型和RSF模型的风险分层能力进行了评估,结果表明这2种模型均具有良好的风险分层能力。

本研究的结果与既往研究一致,证明了本研究结果的准确性和稳定性。此外,所用变量均为临床常用指标,有助于模型在临床实际中的应用。这不仅为预测NKTL患者生存提供了便捷工具,还能帮助临床医生更有效评估患者预后,优化临床决策,改善患者预后。

3.3 研究的局限

虽然本研究取得了一定的成果,但也存在一些局限性。首先,基于SEER数据库的可用性和数据完整性,纳入的患者时间跨度较大(2000-2020年)。由于缩小时间范围,将导致可用于分析的病例数量显著减少,因此扩大了时间范围。然而由于治疗条件及时代背景的限制,这可能会对结果产生一定的偏倚。其次,ECOG评分、LDH水平、EBV-DNA、STAT3、KMT2D

和TP53等关键预后信息未被收录,这限制了对模型的深入探究,也使无法将构建的RSF模型与现有的临床预测模型进行直接比较。本研究为回顾性队列研究,证据等级相对较低,未来需要通过前瞻性和多中心临床试验进一步验证。尽管如此,SEER数据库仍然是研究罕见疾病的重要资源。

综上所述,分析并确定了影响NKTL患者OS的独立预后因素,构建了新的列线图模型和基于机器学习的RSF模型。模型评估结果显示,RSF模型的预测能力优于列线图模型。RSF模型中,Ann Arbor分期、年龄、放疗、联合治疗和疾病类型被确定为5个最显著的预测变量,这些变量均为临床常用指标,便于在临床实际中应用,具有广泛的实用性。RSF模型可以为临床医生提供重要指导,以更准确地进行风险评估和个体化治疗,从而优化患者管理并改善生存结局。

利益冲突 本文所有作者声明不存在利益冲突

参考文献:

[1] WANG H, FU B B, GALE R P, *et al*. NK-/ T-cell lymphomas[J]. Leukemia, 2021, 35(9): 2460-2468. DOI: 10.1038/s41375-021-01313-2.

[2] TSE E, ZHAO W L, XIONG J, *et al*. How we treat NK/T-cell lymphomas[J]. J Hematol Oncol, 2022, 15(1): 74. DOI: 10.1186/s13045-022-01293-5.

[3] KIM S J, YOON S E, KIM W S. Treatment of localized extranodal NK/T cell lymphoma, nasal type: a systematic review[J]. J Hematol Oncol, 2018, 11(1): 140. DOI: 10.1186/s13045-018-0687-0.

[4] WEI Y C, QI F, ZHENG B M, *et al*. Intensive therapy can improve long-term survival in newly diagnosed, advanced-stage extranodal NK/T-cell lymphoma: a multi-institutional, real-world study[J]. Int J Cancer, 2023, 153(9): 1643-1657. DOI: 10.1002/ijc.34672.

[5] MCINNIS M C, MA J, KARUR G R, *et al*. Chronic lung allograft dysfunction phenotype and prognosis by machine learning CT analysis[J]. Eur Respir J, 2022, 60(1): 2101652. DOI: 10.1183/13993003.01652-2021.

[6] YIN H F, XIE J Y, XING S, *et al*. Machine learning-based analysis identifies and validates serum exosomal proteomic signatures for the diagnosis of colorectal cancer[J]. Cell Rep Med, 2024, 5(8): 101689. DOI: 10.1016/j.xcrm.2024.101689.

[7] YAO F S, ZHANG R Q, LIN Q, *et al*. Plasma immune profiling combined with machine learning contributes to diagnosis and prognosis of active pulmonary tuberculosis[J]. Emerg Microbes Infect, 2024, 13(1): 2370399. DOI: 10.1080/22221751.2024.2370399.

[8] BOEHM K M, AHERNE E A, ELLENSON L, *et al*. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer[J]. Nat Cancer, 2022, 3(6): 723-733. DOI: 10.1038/s43018-022-00388-9.

[9] ZHANG T, AUER P, DONG J, *et al*. Whole-genome sequencing identifies novel predictors for hematopoietic cell transplant outcomes for patients with myelodysplastic syndrome: a CIBMTR study[J]. J Hematol Oncol, 2023, 16(1): 37. DOI: 10.1186/s13045-023-01431-7.

[10] ZHANG Y Y, XU J, HUA J, *et al*. A PD-L2-based immune marker signature helps to predict survival in resected pancreatic ductal adenocarcinoma[J]. J Immunother Cancer, 2019, 7(1): 233. DOI: 10.1186/s40425-019-0703-0.

[11] JAMET B, MORVAN L, NANNI C, *et al*. Random survival forest to predict transplant-eligible newly diagnosed multiple myeloma outcome including FDG-PET radiomics: a combined analysis of two independent prospective European trials[J]. Eur J Nucl Med Mol Imaging, 2021, 48(4): 1005-1015. DOI: 10.1007/s00259-020-05049-6.

[12] ZHANG Z Q, HE G, LUO Z W, *et al*. Individual mortality risk predictive system of patients with acute-on-chronic liver failure based on a random survival forest model[J]. Chin Med J, 2021, 134(14): 1701-1708. DOI: 10.1097/CM9.0000000000001539.

[13] CHE W Q, LI Y J, TSANG C K, *et al*. How to use the Surveillance, Epidemiology, and End Results (SEER) data: research design and methodology[J]. Mil Med Res, 2023, 10(1): 50. DOI: 10.1186/s40779-023-00488-2.

[14] WU R Y, LI Y X, WANG W H, *et al*. Clinical disparity and favorable prognoses for patients with Waldeyer ring extranodal nasal-type NK/T-cell lymphoma and diffuse large B-cell lymphoma[J]. Am J Clin Oncol, 2014, 37(1): 41-46. DOI: 10.1097/COC.0b013e318261084b.

[15] CHEN S Y, YANG Y, QI S N, *et al*. Validation of nomogram-revised risk index and comparison with other models for extranodal nasal-type NK/T-cell lymphoma in the modern chemotherapy era: indication for prognostication and clinical decision-making[J]. Leukemia, 2021, 35(1): 130-142. DOI: 10.1038/s41375-020-0791-3.

[16] WANG H, FU B B, WUXIAO Z J, *et al*. A prognostic survival nomogram for persons with extra-nodal natural

killer-/ T-cell lymphoma[J]. Leukemia, 2022, 36(11): 2724-2728. DOI: 10.1038/s41375-022-01679-x.

[17] CAO J Z, LAN S M, SHEN L H, *et al*. Hemoglobin level, a prognostic factor for nasal extranodal natural killer/T-cell lymphoma patients from stage I to IV: a validated prognostic nomogram[J]. Sci Rep, 2017, 7(1): 10982. DOI: 10.1038/s41598-017-11137-9.

[18] TIAN D, YAN H J, HUANG H, *et al*. Machine learning-based prognostic model for patients after lung transplantation[J]. JAMA Netw Open, 2023, 6(5): e2312022. DOI: 10.1001/jamanetworkopen.2023.12022.

[19] MINAMI T, SATO M, TOYODA H, *et al*. Machine learning for individualized prediction of hepatocellular carcinoma development after the eradication of hepatitis C virus with antivirals[J]. J Hepatol, 2023: S0168-S8278(23)00424-5. DOI: 10.1016/j.jhep.2023.05.042.

[20] HUANG B, SOLLEE J, LUO Y H, *et al*. Prediction of lung malignancy progression and survival with machine learning based on pre-treatment FDG-PET/CT[J]. EBioMedicine, 2022, 82: 104127. DOI: 10.1016/j.ebiom.2022.104127.

[21] WANG Z Y, LI Y X, WANG H, *et al*. Unfavorable prognosis of elderly patients with early-stage extranodal nasal-type NK/T-cell lymphoma[J]. Ann Oncol, 2011, 22(2): 390-396. DOI: 10.1093/annonc/mdq347.

[22] SUZUKI R, SUZUMIYA J, YAMAGUCHI M, *et al*. Prognostic factors for mature natural killer (NK) cell neoplasms: aggressive NK cell leukemia and extranodal NK cell lymphoma, nasal type[J]. Ann Oncol, 2010, 21(5): 1032-1040. DOI: 10.1093/annonc/mdp418.

[23] YANG Y, ZHU Y, CAO J Z, *et al*. Risk-adapted therapy for early-stage extranodal nasal-type NK/T-cell lymphoma: analysis from a multicenter study[J]. Blood, 2015, 126(12): 1424-1432; quiz1517. DOI: 10.1182/blood-2015-04-639336.

(收稿:2024-10-23;修回:2024-12-29)

(编辑 邓强庭)