

基于深度学习算法开发和验证的肝细胞癌预后预测模型:一项大样本队列和外部验证研究

韦英婷 覃家盟 樊金莲 梁敏 周海华 覃东 李婷

作者单位:537000 贵港 贵港市人民医院质控科

【摘要】目的 基于深度学习算法开发和验证可评估肝细胞癌(hepatocellular carcinoma, HCC)患者预后的预测模型,并评估其价值。**方法** 选择 2011 年 1 月—2015 年 12 月美国国立癌症研究所的监测、流行病学和最终结果(Surveillance, Epidemiology and Results, SEER)数据库中经病理确诊的 HCC 患者为训练队列用于构建模型,选择同期在本院经病理确诊的 HCC 患者为外部验证队列用于评估模型。主要预测结局为 1、3、5 年总生存率。分别使用深度学习算法 DeepSurv、随机生存森林(RFS)、Cox 比例风险回归开发模型,使用 C-index 评价模型的区分度,一致性曲线评估模型的校准度,log-rank 检验评估危险分层能力。**结果** 训练队列 9 730 例患者和外部验证队列 405 例患者被纳入分析。在训练队列中,DeepSurv 算法 1、3、5 年的 C-index 分别为 0.85 (95%CI:0.80~0.90)、0.82 (95%CI:0.77~0.89)、0.80 (95%CI:0.73~0.87),在外部验证队列中分别为 0.83 (95%CI:0.78~0.87)、0.79 (95%CI:0.74~0.83)、0.72 (95%CI:0.67~0.77)。无论在训练队列还是外部验证队列中,DeepSurv 算法的 C-index 和校准度均优于 RFS、Cox 回归和 TNM 分期(均 $P < 0.05$);log-rank 检验显示,DeepSurv 算法具有良好的危险分层能力($P < 0.001$)。**结论** 基于深度学习算法 DeepSurv 开发的预测模型可有效预测 HCC 患者死亡率,且优于常规的算法和预后评价指标。

【关键词】 肝细胞癌;预测模型;深度学习算法;机器学习;随机生存森林

【中图分类号】 R735.7

【文献标识码】 A

【文章编号】 1674-5671(2021)03-0294-07

DOI: 10.3969/j.issn.1674-5671.2021.03.13

Deep learning - based development and validation of a predictive model for evaluating the prognosis of hepatocellular carcinoma: A study of a large-sample cohort and external validation

WEI Yingting, QIN Jiameng, FAN Jinlian, LIANG Min, ZHOU Haihua, HUI Dong, LI Ting (Department of Quality Control, Guigang City People's Hospital, Guigang 537000, China)

Corresponding author: FAN Jinlian, E-mail: 1536111561@qq.com

【Abstract】 Objective To develop and validate a predictive model for evaluating the prognosis of patients with hepatocellular carcinoma (HCC) based on deep-learning algorithms, and to evaluate its value. **Methods** The SEER (Surveillance, Epidemiology and Results) database of the pathologically diagnosed HCC patients in the National Cancer Institute (USA) from January 2011 to December 2015 were selected as the training cohort to construct predictive models, and the HCC patients who were also pathologically diagnosed in the

【通信作者】 樊金莲, E-mail: 1536111561@qq.com

IVL): a multicentre, single-arm, phase 2 trial [J]. Lancet Oncol, 2020, 21(4): 593-602.

[28] 徐卫, 梁金花. 弥漫大 B 细胞淋巴瘤新基因分型及分子靶向的治疗进展[J]. 临床血液学杂志, 2020, 33(9): 594-598.

[29] WRIGHT G W, HUANG D W, PHELAN J D, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications[J]. Cancer Cell, 2020, 37(4): 551-568.

[30] PROVASI M, BAGNOLI F, FANONI D, et al. CD30 expression in a case of intravascular large B-cell lymphoma, cutaneous variant [J]. J Cutan Pathol, 2019, 46(6): 447-451.

[31] GUPTA G K, JAFFE E S, PITTALUGA S. A study of PD-L1 expression in intravascular large B cell lymphoma: correlation with clinical and pathological features [J]. Histopathology, 2019, 75(2): 282-286.

[收稿 2021-01-25][修回 2021-05-02][编辑 罗惠予]

People's Hospital of Guigang City during the same period were selected as the external verification cohort to evaluate the model. The main predictions were 1-, 3-, and 5-year overall survival rates. The deep-learning algorithm DeepSurv, random survival forest (RFS), Cox proportional hazard regression were used to develop the models. C-index was used to evaluate the discrimination, the calibration curve was used to evaluate the calibration, and the log-rank test was used to evaluate the ability of risk stratification. **Results** A total of 9,730 patients in the training cohort and 405 patients in the external verification cohort were finally included in the study. In the training cohort, the C-index of DeepSurv algorithm in 1-, 3-, and 5-year were 0.85 (95%CI: 0.80–0.90), 0.82 (95%CI: 0.77–0.89) and 0.80 (95%CI: 0.73–0.87), respectively; in the external validation cohort, they were 0.83 (95%CI: 0.78–0.87), 0.79 (95%CI: 0.74–0.83), 0.72 (95%CI: 0.67–0.77), respectively. The C-index and calibration of DeepSurv algorithm were better than those of RFS, Cox regression and TNM staging in both the training and external validation cohort (all $P < 0.05$). The log-rank test showed that the DeepSurv algorithm had a good capability of risk stratification ($P < 0.001$). **Conclusions** The predictive model developed based on the deep learning algorithms DeepSurv can effectively predict the mortality of HCC patients, and is superior to conventional algorithms and prognostic evaluation indicators.

【Key words】 Hepatocellular carcinoma; Prediction model; Deep learning algorithm; Machine learning; Random forest

肝细胞癌 (hepatocellular carcinoma, HCC) 是常见的消化道恶性肿瘤, 肝切除术是主要的治疗方法之一, 但 5 年复发率高达 70%^[1-2], 5 年总生存率仅为 37%^[3]。准确评估 HCC 患者的生存情况对选择治疗策略有重要意义。但是肝癌异质性大, 患者的预期寿命仍难以准确评估。临床预测模型的开发有助于提高临床医师评估患者预后的能力^[4-5]。目前基于 HCC 预后影响因素已经开发了一系列预测模型, 但外推性、准确度仍不足^[6-8]。近年来, 随着庞大数据库出现和计算能力增强, 一些新颖的预测分析方法不断涌现^[9]。其中深度学习是人工智能的一个子学科, 有助于研究人员分析大量数据, 通过预测解决问题^[10]。在过去的十余年中, 深度学习在医疗保健中的使用越来越多^[11-12]。在临床研究中, 深度学习主要用于提供更准确的预测结果以评估肿瘤患者的预后^[13-14]。随着患者病理学和遗传信息的可用性增加, 深度学习可能成为预测生存的新工具^[15-16]。本研究收集美国国立癌症研究所的监测、流行病学和最终结果 (Surveillance, Epidemiology and Results, SEER) 数据库中的数据, 采用深度学习算法 DeepSurv 开发了一种可预测 HCC 患者 1、3、5 年生存率的模型, 并选择本院 HCC 患者验证模型的可推广性, 以期通过个性化预测模型更好地帮助临床医师做出治疗决策或临床试验设计。

1 资料与方法

1.1 研究设计与对象

本研究为回顾性队列研究。选择 2011 年 1 月—2015 年 12 月 SEER 数据库中经病理确诊的 HCC 患者为训练队列用于构建模型, 随访截至 2019 年 12 月; 选

择同期于本院经病理确诊的 HCC 患者为外部验证队列用于评估模型, 随访截至 2019 年 6 月。遵循 TRIPOD 报告指南^[17]开发和验证模型。纳入标准: (1) 病理明确诊断为 HCC; (2) 年龄 ≥ 18 岁; (3) HCC 为第一原发性癌。排除标准: (1) 尸检获取或只有死亡报告上的病例; (2) 非 HCC 相关性死亡者; (3) 合并其他肿瘤者; (4) 随访信息不完整者。本研究经本院理论委员会审核批准。由于本研究为回顾性研究且数据匿名分析, 因此不需要患者知情同意。

1.2 数据提取及定义

训练队列数据采用 SEER*stat 8.3.5 软件 (<http://www.seer.cancer.gov/seerstat>) 在 SEER 数据库中提取。该数据库于 1973 年建立, 记录了美国部分州县 40 余年来肿瘤患者的相关信息, 是目前美国最具代表性的大型肿瘤登记注册数据库之一^[18]。外部验证队列数据从本院电子病历系统中提取。收集患者的性别、年龄、种族、组织学分级、TNM 分期、肿瘤大小、转移情况、手术情况、生存时间、生存状态等资料。组织学分级以 ICD-O-3 为标准分为 I 级、II 级、III 级、IV 级; TNM 分期以 AJCC 第 7 版为标准。

1.3 预测结局

本研究将 HCC 患者的相关性死亡作为研究终点, 主要预测结局为 1、3、5 年总生存率。总生存期 (overall survival, OS) 定义为患者确诊到死亡或随访截止的时间。

1.4 DeepSurv 算法

DeepSurv 算法是将深度学习技术应用于非线性 Cox 回归的风险网络, 其网络是一种深度前馈神经网络, 通过网络学习权重, 预测患者协变量对其危险率的影响^[19]。其中构建 DeepSurv 模型网络参数设置是

关键。首先,需要设置网络层数和各层节点数,其中输入层节点数为预后影响因素的个数,不需要用户输入;输出层只有一个节点,即对数风险函数。其次,需要调节相关参数,关键参数有学习率、学习率衰减、激活函数选择、正则化及优化方法等。

1.5 RFS算法

RFS算法是一种用于右删失生存数据进行分析的随机森林方法,引入了生存森林的事件保留原则,可用于定义总体死亡率,是一种简单的可解释的死亡率度量,可以用作预测结果^[20]。构建RFS的一般流程:(1)模型通过“自助法”(Bootstrap)将原始数据以有放回的形式随机抽取样本,建立样本子集,并将每个样本中37%的数据作为袋外数据排除在外。(2)每个样本随机选择特征构建其对应的生存树。(3)利用Nelson-Aalen法估计随机生存森林模型的总累积风险。在R语言环境下随机森林的建模过程主要包含两个重要参数:ntree(树的数目)和mtry(随机选择特征的数目)。

1.6 统计学方法

计量资料以均数±标准差($\bar{x} \pm s$)表示;分类数据以 $n(\%)$ 表示,组间比较采用 χ^2 检验;采用Kaplan-Meier法计算生存率,组间比较采用log-rank检验。在训练队列中采用Cox比例风险回归分析影响OS的因素,计算风险比(HR)及其对应的95%可信区间(CI),采用Schoenfeld残差法检验等比例假设。由于本研究的预测变量较少,因此采用影响OS的独立因素构建DeepSurv和RFS模型。为防止过度拟合,选择合适的模型参数,模型训练均采用5折交叉验证。分别在训练队列和外部验证队列中对DeepSurv、RFS和Cox模型进行评价。采用C-index和校准曲线评价模型的区分度和校准度,计算方法采用Bootstrap法,重抽样次数为1 000次。同时采用时间依赖的C-index评估模型区分度的持续性。为评估模型的危险分层能力,首先采用X-tile软件计算风险概率的最佳临界值;其次根据最佳临界值将患者分为低风险组、中风险组和高风险组;最后采用log-rank检验比较各组生存曲线的差异。采用R 3.4.3软件(<https://www.r-project.org/>)中的survival程序包构建Cox比例风险回归模型,采用randomForestSRC程序包构建RFS模型;采用Python 3.0中的TFDeepSurv深度生存分析包构建DeepSurv

(<https://github.com/jaredleekatzman/DeepSurv>)模型;其余数据均采用R 3.4.3软件分析。本研究以双侧 $P < 0.05$ 为差异有统计学意义。

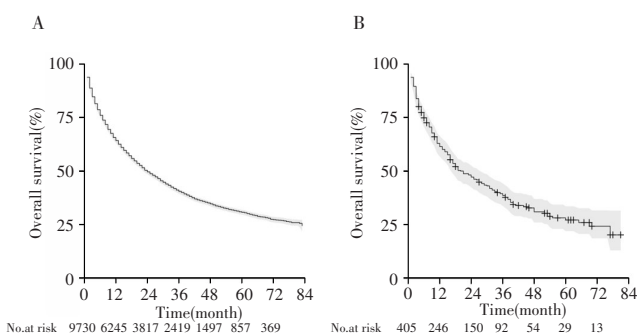
2 结果

2.1 一般资料及生存情况

在SEER数据库中筛选了78 456例HCC患者,最终纳入符合标准9 730例构成训练队列;在本院电子病历系统中共筛选了731例患者,最终纳入符合标准405例构成外部验证队列。在训练队列中,中位随访44个月时死亡5 836例(60.0%),中位OS为24.0个月(95%CI: 23.0~25.0个月),1、3、5年生存率分别为64.2%、40.5%、30.7%,见图1A;在外部验证队列中,中位随访41个月时死亡236例(61.5%),中位OS为21.0个月(95%CI: 17~28.0个月),1、3、5年生存率分别为61.4%、38.6%、27.1%,见图1B。两个队列中,诊断时的平均年龄分别为(65.0±10.9)岁和(65.1±11.3)岁,均以男性为主(训练队列男性占75.8%,外部验证队列男性占76.3%),其余资料比较见表1。

2.2 训练队列中影响OS的独立因素

在训练队列中采用Cox比例风险回归模型分析影响HCC患者OS的因素。单因素Cox回归分析显示,性别、种族、年龄、组织学分级、T分期、N分期、M分期、肿瘤大小、转移、手术与OS有关(均 $P < 0.05$)。多因素Cox回归分析显示,年龄、组织学分级、T分期、N分期、M分期、肿瘤大小、转移、手术是影响OS的独立因素(均 $P < 0.05$),见表2。



A: Training cohort (n=9 730); B: External validation cohort (n=405)

图1 HCC患者的Kaplan-Meier生存曲线

Fig.1 Kaplan-Meier curves of HCC patients

表 1 训练队列和外部验证队列中 HCC 患者的基线资料

Tab.1 Characteristics of HCC patients in the training cohort and external validation cohort

Characteristic	Training cohort (n=9 730)	External validation cohort (n=405)	χ^2	P^a
Age at diagnosis (year)			0.183	0.669
Mean±SD	65.0±10.9	65.1±11.3		
<60	3 037 (31.2%)	131 (32.3%)		
≥60	6 693 (68.8%)	274 (67.7%)		
Gender			0.023	0.879
Female	2 351 (24.2%)	96 (23.7%)		
Male	7 379 (75.8%)	309 (76.3%)		
Race			0.035	0.983
Black	1 315 (13.5%)	—		
White	6 622 (68.1%)	—		
Other	1 793 (18.4%)	405 (100.0%) ^b		
Histological grade			3.441	0.328
I	3 030 (31.1%)	128 (31.6%)		
II	4 713 (48.4%)	186 (45.9%)		
III	1 863 (19.1%)	82 (20.2%)		
IV	124 (1.3%)	9 (2.2%)		
T stage			2.227	0.694
T1	4 747 (48.8%)	200 (49.4%)		
T2	2 250 (23.1%)	92 (22.7%)		
T3	2 197 (22.6%)	91 (22.5%)		
T4	328 (3.4%)	10 (2.5%)		
Tx	208 (2.1%)	12 (3.0%)		
N stage			1.534	0.464
N0	8 781 (90.2%)	358 (88.4%)		
N1	534 (5.5%)	27 (6.7%)		
Nx	415 (4.3%)	20 (4.9%)		
M stage			0.112	0.737
M0	8 853 (91.0%)	366 (90.4%)		
M1	877 (9.0%)	39 (9.6%)		
Surgery			1.820	0.177
No	4 868 (50.0%)	217 (53.6%)		
Yes	4 862 (50.0%)	188 (46.4%)		
Tumor size (cm)			0.013	0.994
Mean±SD	60.5±46.6	61.6±55.7		
<3	2 465 (25.3%)	103 (25.4%)		
3-5	2 841 (29.2%)	119 (29.4%)		
>5	4 424 (45.5%)	183 (45.2%)		
Metastasis			0.065	0.798
No	8 653 (88.9%)	358 (88.4%)		
Yes	1 077 (11.1%)	47 (11.6%)		

Data are n(%), unless otherwise specified. ^a P values are calculated by Pearson's chi-squared test. ^b All patients are of yellow race in the external validation cohort

表 2 影响 HCC 患者 OS 的 Cox 回归分析*

Tab.2 Cox regression analysis for OS in HCC patients*

Variable	No. of death	Univariable		Multivariable	
		HR (95% CI)	P	HR (95% CI)	P
Age at diagnosis (year)					
<60	1 662	1.00		1.00	
≥60	4 174	1.29 (1.22–1.36)	<0.001	1.14 (1.07–1.20)	<0.001
Gender					
Female	1 332	1.00			
Male	4 504	1.13 (1.06–1.20)	<0.001		
Race					
Black	854	1.00			
White	4 053	0.88 (0.81–0.94)	<0.001		
Other	929	0.69 (0.63–0.75)	<0.001		
Histological grade					
I	1 747	1.00		1.00	
II	2 641	0.99 (0.93–1.05)	0.772	1.16 (1.09–1.24)	<0.001
III	1 352	1.70 (1.59–1.83)	<0.001	1.70 (1.58–1.83)	<0.001
IV	96	1.90 (1.55–2.33)	<0.001	1.74 (1.41–2.14)	<0.001
T stage					
T1	2 346	1.00		1.00	
T2	1 180	1.07 (0.99–1.14)	0.072	1.18 (1.10–1.27)	<0.001
T3	1 847	3.12 (2.93–3.32)	<0.001	1.58 (1.47–1.70)	<0.001
T4	2 86	3.36 (2.97–3.80)	<0.001	1.72 (1.51–1.96)	<0.001
Tx	1 77	3.24 (2.78–3.77)	<0.001	1.42 (1.20–1.68)	<0.001
N stage					
N0	5 020	1.00		1.00	
N1	487	3.50 (3.19–3.85)	<0.001	1.31 (1.19–1.45)	<0.001
Nx	329	2.15 (1.92–2.40)	<0.001	1.13 (0.99–1.29)	0.065
M stage					
M0	5 015	1.00		1.00	
M1	821	4.19 (3.88–4.52)	<0.001	1.37 (1.13–1.65)	0.001
Surgery					
No	4 027	1.00		1.00	
Yes	1 809	0.22 (0.21–0.23)	<0.001	0.28 (0.26–0.30)	<0.001
Tumor size (cm)					
<3	956	1.00		1.00	
3–5	1 513	1.60 (1.47–1.73)	<0.001	1.27 (1.17–1.37)	<0.001
>5	3 367	3.27 (3.04–3.51)	<0.001	1.71 (1.57–1.86)	<0.001
Metastasis					
No	4 852	1.00		1.00	
Yes	984	3.73 (3.48–4.01)	<0.001	1.32 (1.11–1.58)	0.002

HR: Hazard ratio; CI: Confidence interval. *No variables violate the assumption of proportionality, calculated by the schoenfeld residual test

2.3 预测模型的开发

选择影响 OS 的独立因素年龄、组织学分级、T 分期、N 分期、M 分期、肿瘤大小、转移、手术, 基于 5 折交叉验证构建 DeepSurv、RFS 和 Cox 模型。DeepSurv 模型最优网络设置: 输入层有 9 个节点, 隐藏层有两层 (第一隐藏层有 7 个节点, 第二隐藏层有 4 个节点), 输出层有 1 个节点; 学习率为 0.05, 学习率衰减为 1.0, 激活函数为 rule, 正则化方法为 sgd。RFS 模型最优参数设置: ntree 为 500, mtry 为 4, 其余参数均为默认设置或自动学习。Cox 模型方程: 风险概率 = 年龄 ≥ 60 岁 $\times 0.129$ + 组织学分级 II $\times 0.151$ + 组织学分级 III $\times 0.532$ + 组织学分级 IV $\times 0.552$ + T2 期 $\times 0.167$ + T3 期 $\times 0.458$ + T4 期 $\times 0.542$ + Tx 期 $\times 0.354$ + N1 期 $\times 0.273$ + Nx 期 $\times 0.124$ + M1 期 $\times 0.312$ - 手术 $\times 1.267$ + 肿瘤大小 3~5 cm $\times 0.237$ + 肿瘤大小

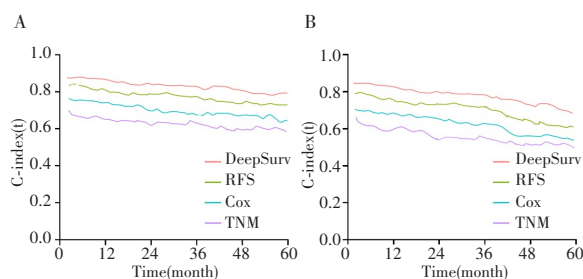
> 5 cm $\times 0.536$ + 转移 $\times 0.281$ 。TNM 分期方程: 风险概率 = T2 期 $\times 0.051$ + T3 期 $\times 0.970$ + T4 期 $\times 0.881$ + Tx 期 $\times 0.834$ + N1 期 $\times 0.545$ + Nx 期 $\times 0.384$ + M1 期 $\times 0.922$ 。

2.4 预测模型的评估与比较

在训练队列和外部验证队列中, 深度学习算法 DeepSurv 的 C-index 均高于 RFS、Cox 回归以及 TNM 分期 (均 $P < 0.05$), RFS、Cox 回归的 C-index 高于 TNM 分期 (均 $P < 0.05$), 见表 3。比较每种模型预测性能的持续性趋势, 发现无论在训练队列还是外部验证队列, DeepSurv 在整个计算期间 (第 1~60 个月), C-index 均高于 RFS、Cox 回归以及 TNM 分期, 见图 2。一致性曲线亦发现, 无论在训练队列还是外部验证队列中, DeepSurv 的 1、3、5 年的校准曲线亦显示出良好的校准度, 见图 3。

表 3 预测模型的区分度评价
Tab.3 Discrimination evaluation of predictive model

Model	Training cohort			External validation cohort		
	1-year	3-year	5-year	1-year	3-year	5-year
DeepSurv	0.85(0.80-0.90)	0.82(0.77-0.89)	0.80(0.73-0.87)	0.83(0.78-0.87)	0.79(0.74-0.83)	0.72(0.67-0.77)
RFS	0.80(0.75-0.84)	0.78(0.70-0.86)	0.75(0.68-0.87)	0.77(0.72-0.82)	0.72(0.66-0.78)	0.62(0.57-0.67)
Cox	0.72(0.68-0.76)	0.70(0.64-0.77)	0.66(0.60-0.77)	0.68(0.63-0.72)	0.61(0.56-0.66)	0.58(0.53-0.62)
TNM	0.66(0.60-0.72)	0.65(0.59-0.71)	0.60(0.55-0.65)	0.61(0.56-0.67)	0.57(0.51-0.63)	0.56(0.50-0.62)



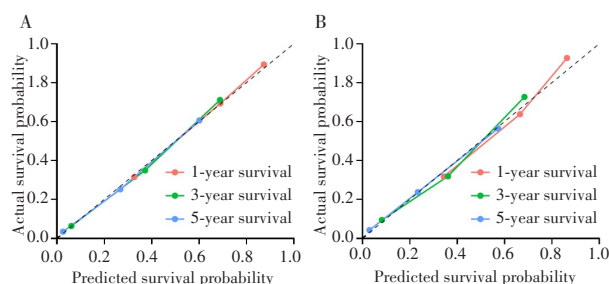
A: Training cohort; B: External validation cohort

图 2 时间依赖的 ROC 曲线评估模型区分度的持续性

Fig.2 Time-dependent ROC curves evaluated the persistence of model discrimination

2.5 DeepSurv 模型的危险分层能力

在训练队列中, 以 DeepSurv 模型预测患者的风险概率, X-tile 软件计算风险概率的最佳临界值, 根据最佳临界值 0.87、2.93 将患者分为低风险组 (< 0.87)、中风险组 (0.87~2.93) 和高风险组 (> 2.93), 任意两组之

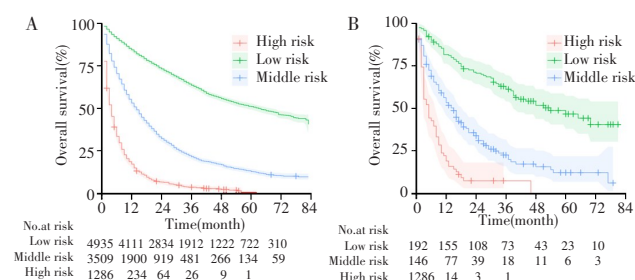


A: Training cohort; B: External validation cohort. The 45° gray line was the reference line that indicated where a perfect calibration would lie. The predicted line overlapped well with the reference line, demonstrating the good performance of the DeepSurv model

图 3 校准曲线评估预测 1、3、5 年总生存率的校准度

Fig.3 Calibration curves estimated the calibration for predicting the survival rate at 1, 3, and 5 years

间的生存曲线差异均有统计学意义 (均 $P < 0.001$), 见图 4A; 在外部验证队列中亦获得相同的结果, 见图 4B。表明 DeepSurv 预测模型具有良好的危险分层能力。



A: Training cohort; B: External validation cohort. The X-tile software was used to calculate the cut-off of risk probability and divided it into low risk group (<0.87), middle risk group ($0.87-2.93$), and high risk group (>2.93)

图4 Kaplan-Meier 曲线评估 DeepSurv 模型的危险分层能力

Fig.4 Kaplan-Meier curves evaluated the risk stratification ability of DeepSurv model

3 讨论

本研究基于大样本数据库 SEER 和深度学习算法 DeepSurv 构建了可预测 HCC 患者 1、3、5 年总生存率的模型,并将该模型与机器学习算法 RFS、Cox 回归和 TNM 分期构建的模型进行比较,发现 DeepSurv 模型的预测性能优于 RFS、Cox 回归和 TNM 分期。本研究同时证明了深度学习可提供一种新颖的解决方案,通过个性化预测模型能更好地帮助临床医师做出治疗决策或进行临床试验设计。

鉴于简单性和相对准确性,TNM 分期已成为临床实践中预测 HCC 患者预后的常用指标之一,但不能精确评估预后。目前主要通过结合临床和病理变量改善 TNM 分期的预测能力。本研究中 Cox 比例风险模型将性别、年龄、组织学分级、肿瘤大小、转移、手术等与 TNM 分期相结合,发现其 C-index 无论在训练队列还是外部验证队列中均高于 TNM 分期。Cox 比例风险模型是一个半参数模型,可用于计算观察到的协变量对事件发生风险的影响^[21]。医学研究人员常使用该模型评估预后协变量在死亡或癌症复发等事件中的重要性,并随后告知患者其治疗选择^[22]。Cox 比例风险模型假设患者的死亡风险是其协变量的线性组合,该假设称为“等比例假设”。然而在真实世界的数据集,等比例假设往往不能满足^[23]。因此,需要更丰富的生存模型,以更好地将生存数据与非线性风险函数拟合。

在拥有大量患者健康数据的电子时代,可以将深度学习或将机器学习整合到电子健康记录中,从而为临床医师提供有价值的预后信息^[24-25]。本研究分别

采用深度学习算法 DeepSurv 和机器学习算法 RFS 构建预测 HCC 患者 1、3、5 年总生存率的模型,在训练队列和外部验证队列中,DeepSurv 和 RFS 的 C-index 均大于 Cox 回归和 TNM 分期,且 DeepSurv 的 C-index 最大,表明 DeepSurv 具有最高的模型区分度,校准曲线亦表明 DeepSurv 有较好的校准度,危险分层系统也发现 DeepSurv 模型可将 HCC 患者区分为低、中、高风险 3 组。DeepSurv 算法是将深度学习技术应用于非线性 Cox 回归的风险网络,其网络是一种深度前馈神经网络,通过网络学习权重,预测患者协变量对其危险率的影响^[19]。医学数据具有高可变性,非线性相互作用和异构分布^[26]。在这种情况下,深度学习技术可能是一种更合适的预测方法,因为其已被证明可处理具有复杂、非线性、异构分布的大数据集^[26-27]。此外,深度学习的独特之处在于可以应用布尔逻辑、绝对条件、条件概率和其他非常规策略对数据建模。但深度学习也存在一定缺陷:一是错误或分类不当的数据将影响模型的质量^[15];二是分析缺乏透明度。虽然深度学习涉及多个分析层,可能做出有意义的预测^[9],但这些分层通常无法被有意义地解释^[28]。如本研究中的 Cox 回归和 TNM 分期结果可用具体的数学公式呈现,但 DeepSurv 和 RFS 算法难以做到。

本研究采用深度学习算法构建预测模型,并进行了外部验证,证明所构建的模型具有一定外推性;样本量大,有足够的随访时间;模型的构建采用了 5 折交叉验证,避免了过度拟合或欠拟合,以上保证了本研究结果的可靠性。但本研究也存在以下局限性:(1)本研究为回顾性研究,无法避免回顾性偏倚及随访依从性等造成的偏倚。(2)训练队列和外部研究队列来自不同国家,虽然 DeepSurv 模型在内部验证和外部验证下均表现良好,但不同种族人群可能会限制模型的适用性,即使本研究多因素 Cox 回归表明种族并不是影响 HCC 预后的危险因素,但仍需进一步研究。(3)本研究虽然构建了预测模型,但未对其进行部署。临床环境中使用深度学习算法的方式有很多种,可以在网站上发布机器学习模型的界面,如 Azure Machine Learning Studio 支持此功能,临床医师可以将患者数据输入基于 Web 的表单中以预测 HCC 患者 1、3、5 年总生存率。此外可将训练好的深度学习模型直接集成到电子病历系统中,这是最有效但也是最难实现的方法^[29]。

综上所述,本研究采用深度学习算法 DeepSurv 构建可预测 HCC 患者生存的模型,该模型的预测变量为年龄、组织学分级、TNM 分期、肿瘤大小、转移、手术,较 TNM 分期预测模型可以更好地预测 HCC 患者的生存期。在大数据时代,深度学习和人工智能通过提高对患者风险估计的量化能力而在评估预后方面起重要作用,为开发更准确预测预后方法提供了新的方向。

参 考 文 献

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021. DOI: 10.3322/caac.21660.
- [2] ZHONG J H, KE Y, GONG W F, et al. Hepatic resection associated with good survival for selected patients with intermediate and advanced-stage hepatocellular carcinoma[J]. Ann Surg, 2014, 260(2): 329-340.
- [3] ZHONG J H, KE Y, WANG Y Y, et al. Liver resection for patients with hepatocellular carcinoma and macrovascular invasion, multiple tumours, or portal hypertension[J]. Gut, 2015, 64(3): 520-521.
- [4] STEYERBERG E W, VERGOUWE Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation[J]. Eur Heart J, 2014, 35(29): 1925-1931.
- [5] MOONS K G, ALTMAN D G, VERGOUWE Y, et al. Prognosis and prognostic research: application and impact of prognostic models in clinical practice[J]. BMJ, 2009, 338: b606.
- [6] HUO R R, LIU X, CUI J, et al. Development and validation of a nomogram for assessing survival in patients with hepatocellular carcinoma after hepatectomy[J]. Biosci Rep, 2020, 40(6): 1-10.
- [7] NG K K C, CHENG N M Y, HUANG J, et al. Development and validation of a novel nomogram predicting 10-year actual survival after curative hepatectomy for hepatocellular carcinoma[J]. Surgeon, 2021. DOI: 10.1016/j-surge.2020.11.013.
- [8] LUO X Y, DU C Y, WEI X F, et al. The study of nomogram based on Ishak inflammation score for recurrence of hepatocellular carcinoma after curative resection[J]. Zhonghua Wai Ke Za Zhi, 2018, 56(2): 124-129.
- [9] KOUROU K, EXARCHOS T P, EXARCHOS K P, et al. Machine learning applications in cancer prognosis and prediction[J]. Comput Struct Biotechnol J, 2015, 13: 8-17.
- [10] BUR A M, SHEW M, NEW J. Artificial Intelligence for the Otolaryngologist: A State of the Art Review[J]. Otolaryngol Head Neck Surg, 2019, 160(4): 603-611.
- [11] CRUZ J A, WISHART D S. Applications of machine learning in cancer prediction and prognosis[J]. Cancer Inform, 2007, 2: 59-77.
- [12] THRALL J H, LI X, LI Q, et al. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success[J]. J Am Coll Radiol, 2018, 15(3 Pt B): 504-508.
- [13] OBERMEYER Z, EMANUEL E J. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine[J]. N Engl J Med, 2016, 375(13): 1216-1219.
- [14] SHEW M, NEW J, BUR A M. Machine Learning to Predict Delays in Adjuvant Radiation following Surgery for Head and Neck Cancer[J]. Otolaryngol Head Neck Surg, 2019, 160(6): 1058-1064.
- [15] SHAH N D, STEYERBERG E W, KENT D M. Big Data and Predictive Analytics: Recalibrating Expectations[J]. JAMA, 2018, 320(1): 27-28.
- [16] HINTON G. Deep Learning - A Technology With the Potential to Transform Health Care[J]. JAMA, 2018, 320(11): 1101-1102.
- [17] COLLINS G S, REITSMA J B, ALTMAN D G, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement[J]. BMJ, 2015, 350: g7594.
- [18] MULLER C, FAHS M C. Technology assessment and accountable health services for women[J]. Am J Public Health, 1999, 89(12): 1895.
- [19] KATZMAN J, SHAHAM U, CLONINGER A, et al. Deep Survival: A Deep Cox Proportional Hazards Network[J]. Arxiv, 2016. DOI: 10.1186/S12874-018-0482-1.
- [20] ISHWARAN H, KOGALUR U B, BLACKSTONE E H, et al. Random survival forests[J]. The Annals of Applied Statistics, 2008, 2(3): 841-860.
- [21] ZUCKER D M, ZHOU X, LIAO X, et al. A modified partial likelihood score method for Cox regression with covariate error under the internal validation design[J]. Biometrics, 2019, 75(2): 414-427.
- [22] HUO R R, LIU H T, DENG Z J, et al. Dose-Response Between Serum Prealbumin and All-Cause Mortality After Hepatectomy in Patients With Hepatocellular Carcinoma[J]. Front Oncol, 2020, 10: 596691.
- [23] ZIEGLER A, LANGE S, BENDER R. Survival analysis: Cox regression[J]. Dtsch Med Wochenschr, 2007, 132 (Suppl 1): e42-44.
- [24] WANG H, JIANG Y, LI B, et al. Single-Cell Spatial Analysis of Tumor and Immune Microenvironment on Whole-Slide Image Reveals Hepatocellular Carcinoma Subtypes[J]. Cancers (Basel), 2020, 12(12): 1-10.
- [25] SHI J Y, WANG X, DING G Y, et al. Exploring prognostic indicators in the pathological images of hepatocellular carcinoma based on deep learning[J]. Gut, 2020. DOI: 10.1136/gutjnl-2020-320930.
- [26] WANG G, LAM K M, DENG Z, et al. Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques[J]. Comput Biol Med, 2015, 63: 124-132.
- [27] MCCARTHY J F, MARX K A, HOFFMAN P E, et al. Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management[J]. Ann N Y Acad Sci, 2004, 1020: 239-262.
- [28] YU M K, MA J, FISHER J, et al. Visible Machine Learning for Biomedicine[J]. Cell, 2018, 173(7): 1562-1565.
- [29] BUR A M, HOLCOMB A, GOODWIN S, et al. Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma[J]. Oral Oncol, 2019, 92: 20-25.

[收稿 2021-03-15] [修回 2021-05-10] [编辑 霍荣瑞]