

进展期胃癌生存预测:基于增强 CT 深度学习模型的构建

张文娟,张利文,邓娟,任铁柱,徐敏,周俊林

【摘要】 目的:探讨基于术前增强 CT 构建的深度学习(DL)模型对进展期胃癌(AGC)1、2、3 年生存概率的预测价值。**方法:**回顾性分析 2013 年 1 月—2015 年 12 月在本院经病理证实为 AGC 的 337 例患者的临床和 CT 资料。按照 7:3 的比例将患者随机分为训练集($n=237$)和验证集($n=100$)。采用数据增强技术增加训练集的数据量,随后基于术前 CT 增强静脉期图像构建残差卷积神经网络结构的 DL 模型,预测 AGC 患者 1、2、3 年的生存概率。经 Cox 单因素及多因素分析构建临床模型,然后联合 DL 模型和临床模型构建综合模型并绘制其诺莫图。计算各模型的 Harrel 一致性指数(C-index)和风险比(HR),并应用 Kaplan-Meier 曲线、校准曲线及临床决策曲线比较 3 种模型对 OS 的预测效能。**结果:**在训练集和验证集中,临床模型、DL 模型和综合模型的 C-index 值分别为 0.70(95%CI:0.65~0.75)、0.72(95%CI:0.67~0.76)、0.74(95%CI:0.69~0.78)和 0.64(95%CI:0.56~0.71)、0.66(95%CI:0.58~0.73)、0.67(95%CI:0.59~0.74),表明综合模型具有最优的生存期预测能力;三个模型的 HR 分别为 2.72(95%CI:2.06~4.02)、2.88(95%CI:1.89~4.39)、2.72(95%CI:2.13~3.49)和 2.11(95%CI:1.43~3.11)、4.32(95%CI:1.66~11.24)、1.89(95%CI:1.36~2.60),均以 DL 模型的 HR 最高,表明 DL 模型预测的高危人群具有更高的死亡风险。校准曲线分析显示基于综合模型的诺莫图预测 AGC 患者 1、2、3 年生存概率与实际的预后随访结果具有较高的一致性。临床决策曲线显示综合模型的净收益优于其它 2 种模型。**结论:**基于 CT 增强静脉期图像利用残差卷积神经网络构建的 DL 模型是一种良好的 AGC 患者生存风险评估模型,对 AGC 患者生存期的早期预判具有较高的临床应用价值。

【关键词】 进展期胃癌;体层摄影术,X 线计算机;残差卷积神经网络;深度学习;预后

【中图分类号】 R814.42;R735.2 **【文献标志码】** A **【文章编号】** 1000-0313(2024)04-0488-08

DOI:10.13609/j.cnki.1000-0313.2024.04.010

开放科学(资源服务)标识码(OSID):



The value of deep learning in the construction of survival prediction models for advanced gastric cancer based on enhanced CT ZHANG Wen-juan,ZHANG Li-wen,DENG Juan,et al.Department of Radiology,Lanzhou University Second Hospital,Lanzhou 730030,China

【Abstract】 Objective: To explore the value of deep learning (DL) model based on preoperative enhanced CT to predict the 1-, 2- and 3-year survival probability in patients with advanced gastric cancer (AGC). **Methods:** From January 2013 to December 2015, the clinical and CT data of 337 patients with AGC confirmed by pathology in our hospital were retrospectively analyzed. All subjects were divided into a training set ($n=237$) and an external validation set ($n=100$) according to a ratio of 7:3. A DL model of residual convolutional neural network was constructed based on preoperative contrast-enhanced venous phase CT images to predict the 1-, 2- and 3-year survival probability of AGC patients. Data enhancement technology was used to increase the data amount of training set. Univariate and multivariate Cox regression analysis methods were used to construct clinical models, and then DL model and clinical model were integrated to construct comprehensive prediction models and a corresponding nomogram was created. Harrel agreement index (C-index) and risk ratio (HR) were calculated, and Kaplan-Meier curve, calibration curve and clinical decision curve were used to compare the prediction performance of the three models for OS. **Results:** In the training set, the C-index values of

作者单位:730030 甘肃兰州:兰州大学第二医院放射科/甘肃省医学影像重点实验室/医学影像人工智能甘肃省国际科技合作基地(张文娟、邓娟、任铁柱、徐敏、周俊林);100190 北京,中国科学院自动化所分子影像重点实验室(张利文)

作者简介:张文娟(1989—),女,甘肃武威人,博士,主治医师,主要从事腹部影像诊断工作。

通讯作者:周俊林,E-mail:ery_zhoujl@lzu.edu.cn

基金项目:国家自然科学基金资助项目(82102151);兰州市青年科技人才创新项目(2023-2-44);甘肃省省级人才项目青年个人项目(甘组通字【2024】4号)

comprehensive model, DL model and clinical model were 0.74 (95% CI: 0.69~0.78), 0.72 (95% CI: 0.67~0.76) and 0.70 (95% CI: 0.65~0.75), respectively; and in the validation set, they were 0.67 (95% CI: 0.59~0.74), 0.66 (95% CI: 0.58~0.73) and 0.64 (95% CI: 0.56~0.71), respectively. The results indicated that the comprehensive model had the best survival prediction ability. In the training set, the HRs of the three models were 2.72 (95% CI: 2.06~4.02), 2.88 (95% CI: 1.89~4.39) and 2.72 (95% CI: 2.13~3.49); in the validation set, the HRs of the three models were 2.11 (95% CI: 1.43~3.11), 4.32 (95% CI: 1.66~11.24) and 1.89 (95% CI: 1.36~2.60). The results showed that the DL model had the highest HR in both training and validation set, which indicated that the patients in high-risk group predicted by the DL model had a higher risk of death. Calibration curve analysis showed that the probability of 1-, 2- and 3-years survival probability predicted by nomogram based on the comprehensive prediction model was in good agreement with the actual prognostic follow-up results. The clinical decision curve also proved that the net benefit of the comprehensive model was better than that of the other two models. **Conclusion:** The DL model constructed based on residual convolutional neural network in this study is a good survival risk assessment model, which has good application value for realizing early prediction of survival probability in AGC patients.

【Key words】 Advanced gastric cancer; Tomography, X-ray computed; Residual convolutional neural network; Deep learning; Prognosis

胃癌是全球最常见的消化道恶性肿瘤,在我国发现的胃癌病例中,大部分为进展期胃癌(advanced gastric cancer, AGC),早期胃癌仅占约 11.5%^[1]。尽管近年来早期诊断和多学科治疗策略均有所改进,但由于复发率高,接受了根治性手术的 AGC 患者的生存率仍然很低,5 年生存率低于 20%^[2]。准确地预后评估对个体化治疗方案的选择和疗效评价至关重要,但由于肿瘤异质性,部分相同 TNM 分期的 AGC 患者存在明显的疗效差异,预后结果不一^[3-4]。因此,如果能提出比现有的 TNM 分期更加准确的预后评估方法,将患者进行风险分层,提供更个性化的概率预测,将更有助于临床决策。增强 CT 是临床上最常用的胃癌诊断及术前分期方法,随着大数据及人工智能技术的发展,基于增强 CT 的影像组学模型在多种肿瘤中均有较好的预后预测价值,为肿瘤的预后预测提供了强有力的依据^[5-8]。但影像组学分析中需要精准分割病灶来保证提取特征的稳定性,而深度学习(deep learning, DL)技术的最显著的特点就是可以通过多隐层神经网络进行自主学习,通过深层非线性网络结构无限逼近复杂函数,直接获得与数据相关联的特征,实现让机器自主学习并提取数据特征^[9]。因此,相比人工方法,基于 DL 技术提取的影像组学特征(例如形状、大小和纹理)与临床结果的相关性更大^[10]。基于以上研究背景,本研究中采用一种已经被广泛认可的残差卷积神经网络(residual convolutional neural network, RCNN)算法并基于术前三期增强 CT 图像来构建并验证 AGC 生存风险预测 DL 模型,旨在进一步提高对 AGC 患者的术前影像诊断水平,改善患者的预后。

材料与方法

1. 病例资料

回顾性搜集 2013 年 1 月—2015 年 12 月在本院经病理证实且进行了术后随访的 AGC 患者的临床和影像学资料。纳入标准:①经病理证实为胃腺癌;②经国际抗癌联盟(Union for International Cancer Control, UICC)及美国肿瘤联合会(American Joint Committee on Cancer, AJCC)第 8 版 TNM 分期中临床 TNM 分期(clinical TNM, cTNM)明确诊断为进展期胃癌且无远处转移($\geq T2 + M0$);③行根治性切除术及 D2 淋巴结(第一站及第二站区域淋巴结)清扫术,且为 R0 切除(切缘无肿瘤残余);④送检淋巴结数 ≥ 15 ;⑤手术前 2 周内行腹部 CT 平扫和三期增强扫描;⑥有完整的基线临床资料和随访资料;⑦术前行胃镜活检。排除标准:①术前接受放、化疗;②CT 图像有伪影;③胃充盈不良。

患者的基线临床资料及随访信息应包括以下内容。①一般资料:年龄、性别;②胃镜检查结果,包括肿瘤部位、Borrmann 分型、WHO 组织学病理分级;③血清学肿瘤标志物:癌胚抗原、CA125;④病理结果:pTNM 分期、WHO 组织学病理分类;⑤影像评估:影像医师分析肿瘤的 CT 征象并进行 TNM 分期(cTNM)。⑥随访信息:记录每例患者自接受手术至死亡或最后随访日期的时间间隔。所有患者在胃癌根治术后至少随访 1 年。患者在术后的前 2 年内每 3~6 个月复查一次,在随后的 3~5 年中每 6~12 个月复查一次,随访内容包括临床病史、体格检查、血液学检查、功能状

态评分(performance status, PS)、体重监测、胃镜、腹部彩超和 CT 等。本研究中患者的随访时间为 3~74 个月,中位随访时间为 32 个月。

最终共纳入符合研究要求的连续 337 例 AGC 患者,将患者按照 7:3 的比例随机分为训练集(training set)和验证集(external validation set),训练集有 237 例,验证集有 100 例。

2. CT 扫描方法

患者检查前禁食 6~12 h,检查前 20 min 饮用温开水 1000~2000 mL 使胃腔充盈,并肌肉注射 20 mg 山莨菪碱,以抑制胃肠道蠕动。

使用 Discovery CT750 HD 宝石能谱 CT 机,扫描参数:120 kVp,160 mA,0.4~0.6 s/r,探测器准直宽度 64×0.625 mm,矩阵 512×512 ,视野 350 mm $\times$ 350 mm,扫描层厚 5 mm,螺距 0.992,重建层厚 1.25 mm。经肘静脉团注等渗对比剂碘佛醇(320 mg I/mL),注射速率 3.0 mL/s,行动脉期、静脉期和延迟期三期增强扫描,延迟时间分别为 30~40、60~80 和 120 s。

3. DL 模型的构建和验证

在注射对比剂后 30~35 s(动脉期)在 CT 图像上即可见累及黏膜层的肿瘤(T1,早癌)出现强化,而侵犯肌层的肿瘤($\geq$ T2,进展期胃癌)的强化高峰时间一般在 60~70 s 之后(静脉期)出现,强化时间明显晚于黏膜层且持续时间较长,因此本研究中选择静脉期图像进行肿瘤 ROI 的勾画和分割。由 2 位经验丰富的放射科医师在每例患者的增强 CT 静脉期图像中选择肿瘤最大层面及相邻的 2 个层面,分别独立勾画矩形框将肿瘤区域(包括肿瘤病灶及其邻近的组织结构)包含进去。使用 ITK-SNAP 3.6.1 软件(www.itk-snap.org)进行图像分割。

org)进行图像分割。

残差网络(residual network, RN)可以增加深度学习网络的深度,同时不会引起训练集中误差的增多^[11],残差学习结构可以通过前向神经网络结合捷径连接实现,整个网络仍然可以通过端到端的反向传播进行训练。本研究中基于残差网络的原理,构建了一个 18 层的端到端的 RCNN(图 1),大小为 224×224 。模型由 8 个残差块组成,这些残差块可以通过捷径连接高效传输梯度,并加速网络收敛,输入端为原始 CT 图像。我们还定义了专门的损失函数(loss)来训练模型进行风险预测,其计算公式:

$$Loss = -\frac{1}{C \sum_{s=1}^C} \sum_{LS_i=1}^C \{D_{\beta}(\chi_i) - \log \sum_{j \in A(T_j)} e^{D_{\beta}(\chi_i)}\} \quad (1)$$

$C_{S=1}$ 为观察状态的患者人数, C 是事件总数,指患者的终点事件被观测到。在本研究中,终点事件指在随访时间终点患者的生存结局; i 表示患者集合序号, j 是在随访时间大于第 i 个患者患者集合中的患者序号。 $Si=1$ 表示这个集合里面的患者随访终点事件是被观测到的。 $A(T_j)$ 是一个患者集合,在该集合中,要求所包含的患者随访时间都比第 i 个患者随访时间长。 T_j 是随访期间的生存时间, β :表示深度学习模型的参数, $D_{\beta}(\chi_i)$ 是 DL 模型的输出,代表患者的生存风险概率。

同时采用数据增强、dropout 和微调等技术对模型进行训练以避免数据过拟合。本研究中使用 Imageaug 库来扩大数据集。数据增强技术包括翻转、转换、旋转、缩放、加性高斯白噪声和剪切等步骤。此技术仅适用于训练集,在将增强数据输入 DL 模型之前,将所有 CT 图像的大小调整为 224×224 ,并将每个被

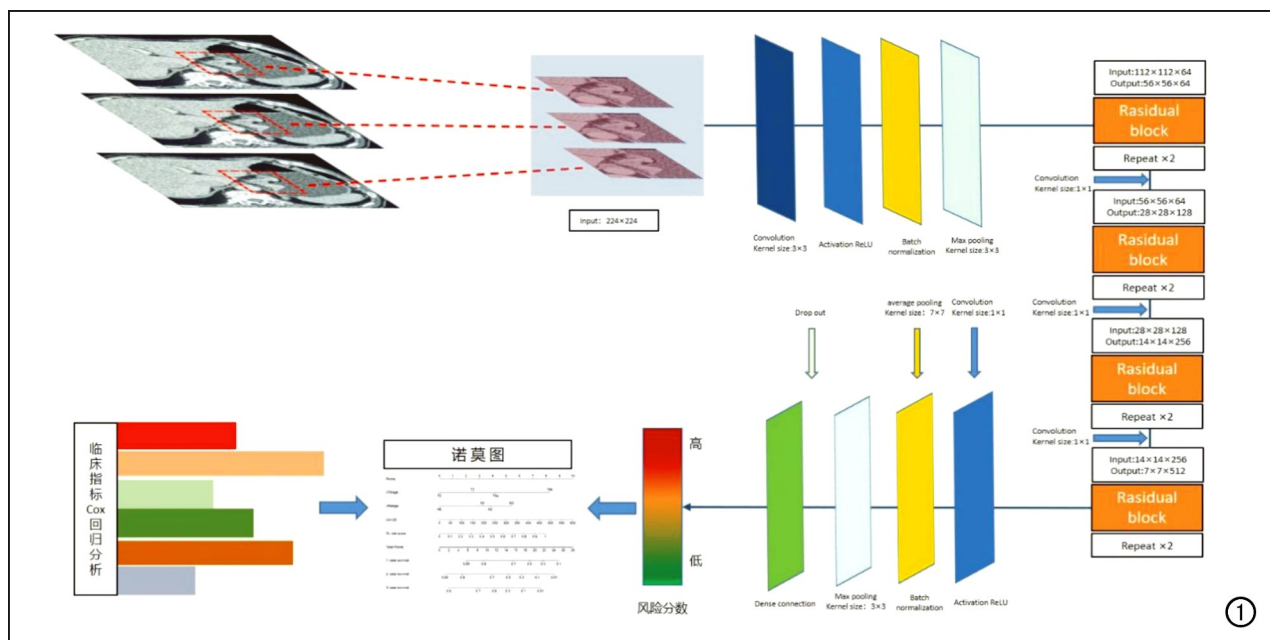


图 1 基于 RCNN 的深度学习模型及其诺莫图构建的主要步骤和方法。

调整大小的切片堆叠为三通道图像。利用数据增强技术,可将包含 237 例患者的 711 帧 CT 的训练集数据放大 8 倍,最终共有 5688 帧图像输入 DL 模型进行生存风险预测模型的训练。然后,使用验证集中 100 例患者的 300 帧图像(未应用数据增强技术)来验证模型。本研究中构建的 RCNN 的输入端为患者的 CT 图像,输出端为每个患者的 1、2、3 年生存概率,与 Cox 模型中 log-risk 函数的估计相同。以每例患者所有层面 CT 图像输入 DL 模型获得的预测概率的平均值作为此例患者的生存风险概率。通过定义 flatten 层之后的 dropout 层来细化 ResNet18,将全连接层(fully-connected layer)定义为 sigmoid 激活函数的输出。

本研究中采用随机初始化的方法加载图像并构建模型。所有实验均在一个计算集群中进行:4 个服务器 NVIDIA TiTAN Xp(12TB 帧缓存),基于深度学习的框架软件 Tensorflow 1.10.0(www.tensorflow.org)和程序包 Keras 2.2.4(www.tensorflow.org/guide/keras)。使用公式(1)作为损失函数来训练模型。Batch_size 设置为 24,学习率设置为 0.001,迭代的 epoch 设置为 1000,并在 Keras 库中使用 Adam 优化器进行默认设置,使用 earllystup 和 reduceLronplatea 的回调和一些默认设置来有效地训练模型。

DL 模型的验证:本研究中采用 Kaplan-Meier (KM)曲线研究 DL 模型与生存期之间的潜在联系。以训练集中得到的中位风险评分作为截断值,得分低于截断值的患者被归为低风险组,其它患者被归为高风险组。采用校准曲线(calibration curve)评价基于 DL 预测模型获得的预测风险结果与实际结果的一致性。计算 Harre 一致性指数(Harrell's concordance index, C-index)和风险比(hazard ratio,HR)比较不同模型的预后价值。通过决策曲线分析(decision curve analysis,DCA)观察在一定的阈值概率下不同模型的净收益。

4. 统计学分析

使用 R3.6.1 软件(<http://www.R-project.org>)进行统计分析。组学特征和临床变量的比较采用 Mann-Whitney U 检验和 χ^2 检验;采用单因素及多因素 Cox 回归分析筛选与预后高度相关的特征;对模型 KM 曲线的比较采用 Log-Rank 检验;采用 G-rho Rank 检验计算 HR。以 $P<0.05$ 为差异有统计学意义。

结 果

1. 临床变量分析

训练集和验证集中临床特征的比较结果详见表 1。2 个数据集之间临床特征的差异均无统计学意义($P>0.05$)。

单因素和多因素 Cox 回归分析结果见表 2。结果显示,cT 分期、cN 分期及 CA125 是患者生存期的独立临床风险预测因子。

表 1 训练集及验证集中患者临床特征的比较

临床特征	训练集 N=237	验证集 N=100	Z 值	P 值
年龄/年	56(54.95±9.06)	54(55.02±10.00)	0.054	0.816
性别/例			2.135	0.138
男	173(73.0%)	81(81.0%)		
女	64(27.0%)	19(19.0%)		
肿瘤部位/例			4.287	0.039
胃窦	113(47.7%)	43(43.0%)		
胃体	58(24.5%)	38(38.0%)		
贲门	66(27.8%)	19(19.0%)		
分化程度/例			0.897	0.329
低/未分化	120(50.6%)	47(47.0%)		
中分化	104(43.9%)	44(44.0%)		
高分化	13(5.5%)	9(9.0%)		
Borrmann 分型/例			0.152	0.618
I 型	5(2.1%)	5(5.0%)		
II 型	51(21.5%)	20(20.0%)		
III 型	156(65.8%)	64(64.0%)		
IV 型	25(10.5%)	11(11.0%)		
CA125/例			3.856	0.054
正常	222(93.7%)	87(87.0%)		
异常升高	15(6.3%)	13(13.0%)		
CEA/例			2.816	0.128
正常	165(69.6%)	61(61.0%)		
升高	72(30.4%)	39(39.0%)		
生存期/个月	36	17	2.763	0.129
cN 分期/例			6.152	0.016
N0	59(24.9%)	23(23.0%)		
N1	59(24.9%)	12(12.0%)		
N2	51(21.5%)	19(19.0%)		
N3a/N3b	68(28.7%)	46(46.0%)		
cT 分期/例			4.971	0.028
T2	43(18.1%)	14(14.0%)		
T3	135(57.0%)	48(48.0%)		
T4a	58(24.5%)	37(37.0%)		
T4b	1(0.4%)	1(1.0%)		

2. 模型的构建

我们在研究过程中发现,模型中的残差块和恒等映射不仅可以提高学习能力,而且可以解决退化问题,尤其是在数据集有限的情况下。因此,我们构建了基于 RCNN 的端到端 DL 模型,以显示风险预测模型的学习能力。通过数据增强技术重新生成 5688 帧图像输入到 DL 模型中,最后输出每个患者的风险预测概率。

经 Cox 单因素及多因素分析,cT 分期、cN 分期及 CA125 是预后预测的独立影响因子,构建包含这 3 个预后因子的临床模型。同时,基于 DL 模型获得的风险预测概率和临床模型中的变量,采用风险比例线性回归方法建立临床-DL 综合模型并绘制其诺莫图(图 2)。诺莫图可实现个体化预测胃癌患者 1、2、3 年生存期的生存概率。

3. 模型预测效能的评估

基于 DL 模型的输出结果,可以将训练集和验证集中的患者分别分为两个不同的风险亚组:低风险组和高风险组。在训练集中,高风险组和低风险组中患

表 2 AGC 患者生存期的独立危险因素

临床变量	单因素分析		多因素分析	
	HR(95%CI)	P 值	HR(95%CI)	P 值
年龄	0.99(0.99~1.01)	0.48	1.01(0.99~1.03)	0.49
性别				
男	1.00(reference)	—	1.00(reference)	—
女	1.02(0.70~1.49)	0.91	1.17(0.74~1.85)	0.98
肿瘤位置				
胃窦	1.00(reference)	—	1.00(reference)	—
胃体	0.82(0.53~1.26)	0.36	0.83(0.49~1.43)	0.52
贲门	1.04(0.70~1.54)	0.85	1.27(0.68~2.39)	0.46
分化程度				
低/未分化	1.00(reference)	—	1.00(reference)	—
中分化	0.58(0.41~0.82)	<0.01	0.61(0.34~1.10)	0.10
高分化	0.23(0.07~0.74)	0.02	0.54(0.15~2.04)	0.37
Borrmann 分型				
I 型	1.00(reference)	—	1.00(reference)	—
II 型	2.19(0.30~16.23)	0.44	0.92(0.16~7.31)	0.94
III 型	3.38(0.47~24.27)	0.23	1.42(0.18~11.01)	0.73
IV 型	10.95(1.48~81.23)	0.02	3.68(0.44~30.57)	0.23
CA125				
正常	1.00(reference)	—	1.00(reference)	—
升高	2.73(1.54~4.86)	<0.01	1.98(1.32~2.97)	<0.01
癌胚抗原				
正常	1.00(reference)	—	1.00(reference)	—
升高	2.32(1.65~3.27)	<0.01	1.76(0.92~3.37)	0.09
cN 分期				
N0	1.00(reference)	—	1.00(reference)	—
N1	2.13(1.21~3.74)	<0.01	0.89(0.43~1.56)	0.89
N2	2.90(1.65~5.11)	<0.01	1.67(1.10~2.84)	<0.01
N3a/N3b	3.95(2.33~6.70)	<0.01	2.86(1.71~4.36)	<0.01
cT 分期				
T2	1.00(reference)	—	1.00(reference)	—
T3	2.74(1.49~5.03)	<0.01	2.77(1.36~27.34)	<0.01
T4a	5.72 (3.03~10.81)	<0.01	5.23(1.19~53.96)	<0.01
T4b	9.52(1.23~74.04)	0.03	17.4(0.36~846.16)	0.15

表 3 三种预测模型的 C-index 和 HR 值

模型类型	C-index(95%CI)		HR(95%CI)	
	训练集	验证集	训练集	验证集
DL 模型	0.70(0.65~0.75)	0.64(0.56~0.71)	2.88(1.89~4.39)	4.32(1.66~11.24)
临床模型	0.72(0.67~0.76)	0.66(0.58~0.73)	2.72(2.06~4.02)	2.11(1.43~3.11)
综合模型	0.74(0.69~0.78)	0.67(0.59~0.74)	2.72(2.12~3.49)	1.89(1.38~2.60)

者的中位生存时间分别为 21 和 57 个月;在验证集中,则分别为 14 和 28 个月。

为了评价和比较 DL 模型和综合模型的预测效能,采用 KM 曲线分析基于模型的输出结果对患者进行危险度分层的效能,结果见图 3~4。在训练集和验证集中,DL 模型和临床模型都可以成功地将患者分为高风险组和低风险组,说明两个模型均可将患者进行危险度分层,具有较好的预后预测能力。此外,综合模型在训练集和验证集中均显示出最佳的预测效能,其 C-index 值最高,DL 模型在训练集和内部验证集中的 HR 最高,表明 DL 模型预测的高危人群较其它模型预测的高危人群有更高的死亡危险(表 3)。

综合模型的校准曲线显示其具有良好的风险预测效能(图 5)。校准曲线显示综合模型的诺莫图与完美

拟合曲线(Perfect-fit)之间的对比,显示出不论是在训练集还是验证集中,综合模型预测的生存风险概率和实际随访观察到的 1、2、3 年生存概率之间具有很好的 consistency。DCA 结果表明综合模型较 DL 模型和临床模型为患者提供了更大的临床净收益(图 6)。

讨 论

胃癌组织的高度异质性以及复杂的病因使得预测患者的预后具有很大的挑战性。随着胃癌综合治疗方法的发展及多学科诊疗模式(multiple disciplinary team,MDT)的推广,我们旨在寻求通过更加多元化的生存预测工具来满足胃癌个体化诊疗的需求和期望。DL 模型具有以下优点:①通过层次神经网络结构提取与临床问题直接相关的多层次抽象特征;②不需要

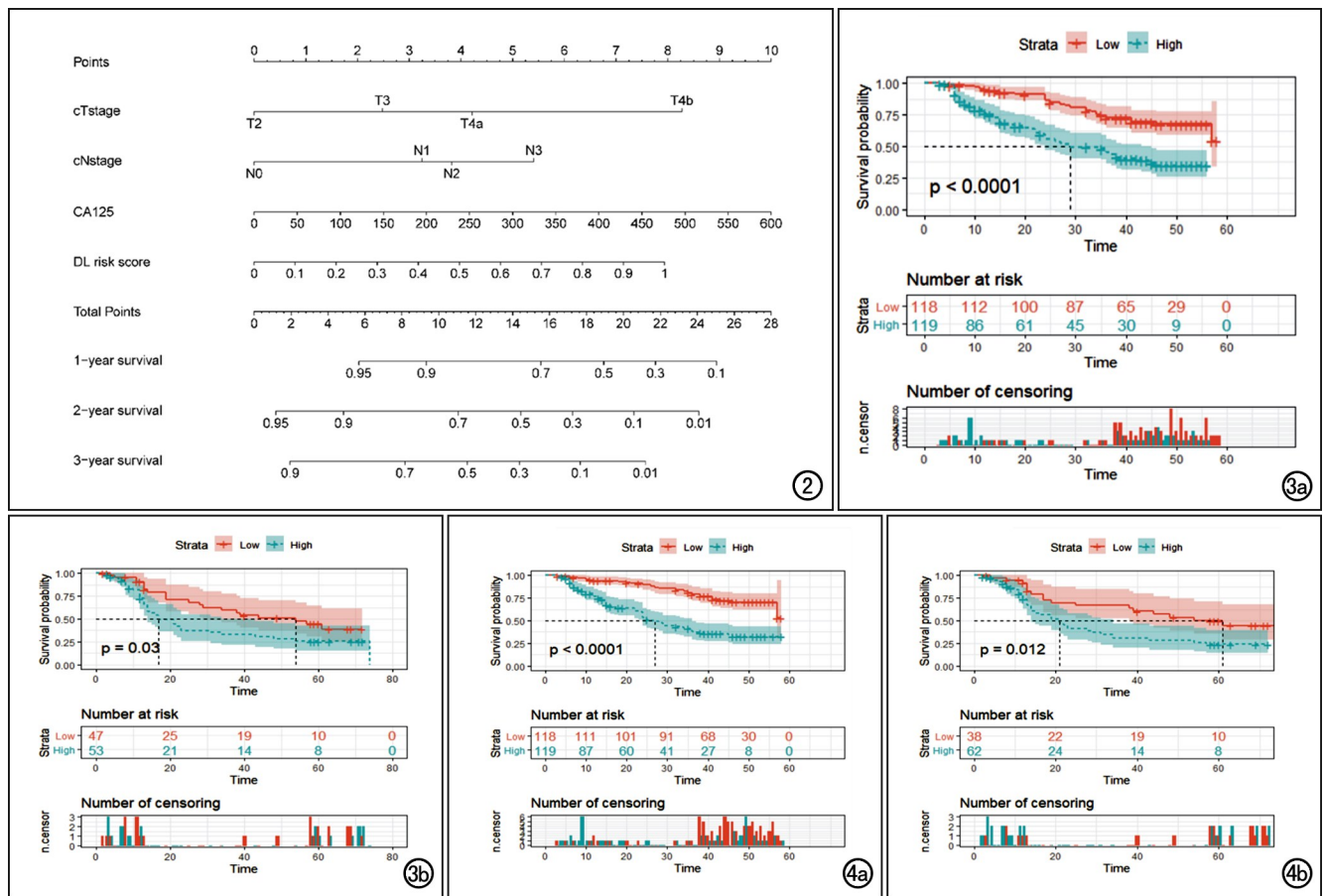


图 2 基于综合模型的诺莫图。图 3 DL 模型的 KM 曲线,纵坐标为生存概率,横坐标为时间。可见随着时间的增加,低风险组和高风险组患者的生存概率存在差异,说明 DL 模型可将患者分为不同的风险亚组。a)训练集;b)验证集。图 4 临床模型的 KM 曲线,纵坐标为生存概率,横坐标为时间。可见随着时间的增加,低风险组和高风险组患者的生存概率存在差异,说明临床模型可将患者分为不同的风险亚组。a)训练集;b)验证集。

对肿瘤边界进行耗时的勾勒标注,并考虑到了肿瘤周边微环境以及肿瘤与附着组织的关系;③相应的软件等工具快速且使用方便,其端对端的设计,只需操作者将原始 CT 图像进行输入,无需进一步的人工干预即

可直接获得临床问题的预测结果。

本研究中开发并验证了一个基于残差网络的端到端的 DL 生存预测模型,旨在通过输入术前 CT 图像即可获得胃癌患者在接受根治性切除术后的生存期预测结果。DL 模型在训练集和验证集中将预后有差异的胃癌患者分为高、低危不同分组的能力显示出良好的结果;而且,将 DL 模型与临床模型融合构建的综合模型较 DL 模型具有更好的预测性能。同时,针对 DL 模型常见的过拟合问题及为了更好的提高模型的预测性能,本研究中采用损失函数、Dropout 和微调技术等一系列方法来训练模型。He 等^[11]的研究结果表明残差块和恒等映射可以提高模型的学习能力,解决退化问题。我们的研究结果亦显示,在某些情况下残差网络也可以解决 CT 图像分析的退化问题。

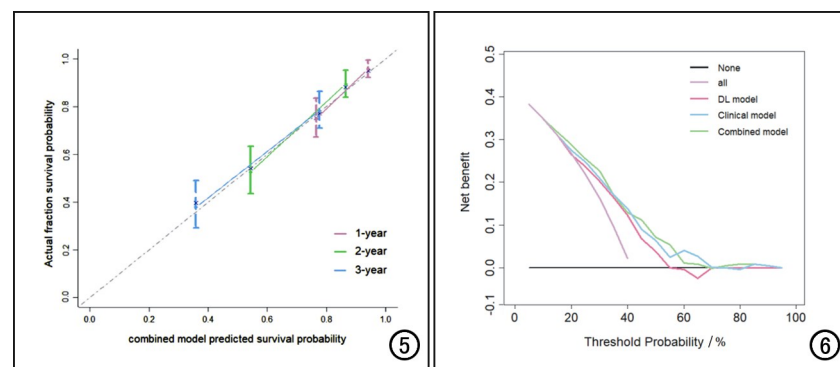


图 5 综合模型诺莫图在训练集中的校准曲线,综合模型的预测结果和完美曲线之间拟合良好,表明其预测效能较高。图 6 DL 模型、临床模型及综合模型的临床决策曲线,显示综合模型的净收益优于其它两个模型。

迄今为止,大多数的 DL 应用聚焦于疾病诊断过程中的检测和分类^[12-13]。在预后研究方面,Cox 比例风险(cox proportional hazard,CPH)模型是目前最常用的癌症患者生存分析方法,但由于该方法的假设结果是协变量的线性组合,其预测结果可能过于简单^[14-16]。研究者们最近转向利用 AI 算法从医学图像中提取信息建立模型来进行肿瘤患者生存期的预测,这些方法目前在提高肿瘤复发和生存预测的准确性方面发挥着重要作用。如 Chi 等^[17]将人工神经网络(artificial neural networks,ANN)应用于生存分析,因为 ANN 可以很容易地考虑变量间的相互作用,建立非线性预测模型,从而提供比传统方法更灵活的生存时间预测。他们的研究中比较了两个不同的乳腺癌数据集的人工神经网络的结果,显示神经网络能成功预测复发概率并将预后分层。Kim 等^[16]采用基于 DL 的生存预测算法 DeepSurv 对口腔癌患者进行生存预测,并与随机生存森林模型(random survival forest,RSF)和 CPH 模型进行比较。结果显示 3 种模型中,DeepSurv 的表现最好,在训练集和验证集中的 C-index 值分别达 0.810 和 0.781,优于 RSF 的 0.770、0.764 和 CPH 的 0.756、0.694。Matsuo 等^[16]比较了 DL 神经网络模型和 CPH 模型在预测宫颈癌患者生存中的作用,与 CPH 模型相比,DL 模型在预测无进展生存期(progression free survival,PFS)方面显示出更好的性能;而且研究中发现当加入更多特征时,DL 模型的性能进一步提高。Wang 等^[18]从 245 例高级别浆液性卵巢癌患者的 8917 帧 CT 图像中训练新型深度学习网络,提取预后生物标志物,并开发了一个结合 DL 特征和 CPH 的 DL-CPH 模型,对患者的个体复发风险和 3 年复发概率显示出良好的预测性能。本研究中提出的 DL 模型同样也表现出了较高的预测预后性能,优于 TNM 分期以及基于传统的临床病理风险因素的模型。

在本研究中,Cox 回归分析结果显示 cT 分期、cN 分期以及 CA125 是生存风险预测的独立预测因子。TNM 分期是预测 AGC 患者预后最常用的评估系统^[19-20]。既往的研究结果显示累及浆膜的肿瘤(T3/T4)和 N2/N3 分期的胃癌复发率高^[21],T2/T3 和 T4 期胃癌的腹膜转移、血行转移和淋巴结转移的 5 年累积复发率分别为 8.0% 和 53.6%、13.0% 和 33.1%、7.7% 和 31.2%^[22-23]。淋巴结是否发生转移也是胃癌的重要预后指标,同样是 T3/T4a 的患者,淋巴结转移阴性(N0)和淋巴结转移阳性(N+)的胃癌患者,其 1、3、5 年生存率分别为 68.9% 和 66.4%、41.4% 和 33.1%、33.1% 和 25.9%。我们前期的研究也表明,cT4a/T4b 和/或 cN2/N3 患者与 cT2/T3 和/或 cN0/

N1 患者相比,具有更大的早期复发风险^[24]。CA125 是常见的血清学肿瘤标志物,术前血清 CA125 水平与胃癌的侵袭性和生存期有一定的相关性^[25];而且是评估腹膜转移的敏感标志物,如崔等^[26]的分析了术前血清 CA125 及 CEA、CA199、CA242 和 AFP 等实验室指标对胃癌腹膜转移的临床诊断价值,结果显示 CA125 对胃癌腹膜转移的诊断价值最高,AUC 为 0.964。在本研究中,上述提到的 cT 分期、cN 分期以及 CA125 均与预后有关,被纳入 DL 诺莫图,根据不同指标的权重,计算出每个患者的风险预测概率,再根据截断值进行危险度分层。

尽管本研究中基于 CT 增强静脉期图像采用 RC-CN 方法构建的 DL 模型的预测性能较好,但也存在一些局限性:①本研究是单中心回顾性研究,下一步有必要搜集多中心、前瞻性的数据以获得 DL 模型更泛化及稳健的预测性能;②本模型仅基于术前 CT 图像构建 DL 模型,若能结合病理图像及其它类型的图像可能会有更重要的发现。

综上所述,基于残差卷积神经网络构建的 DL 模型对 AGC 患者生存期具有良好的预测价值,相对于 TNM 分期,DL 生存风险模型的诺莫图可提供更具体、更个性化的预测概率。

参考文献:

- [1] 王玮,孙哲,邓靖宇,等.基于多中心大样本数据库的胃癌外科治疗相关数据的整合与分析[J].中华胃肠外科杂志,2016,19(2):179-185.
- [2] Tegels JJ, de Maat MF, Hulsewe KW, et al. Improving the outcomes in gastric cancer surgery[J]. World J Gastroenterol, 2014, 20(38):13692-13704.
- [3] Yachida S, Jones S, Bozic I, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer[J]. Nature, 2010, 467(7319):1114-1117.
- [4] Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing[J]. N Engl J Med, 2012, 366(10):883-892.
- [5] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach[J/OL]. Nat Commun, 2014, 5:e4644[2014 Jun 3]. DOI: 10.1038/ncomms5006.
- [6] Huang YQ, Liang CH, He L, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer[J]. J Clin Oncol, 2016, 34(18):2157-2164.
- [7] Zhou Y, He L, Huang Y, et al. CT-based radiomics signature: a potential biomarker for preoperative prediction of early recurrence in hepatocellular carcinoma[J]. Abdom Radiol, 2017, 42(6):1695-1704.
- [8] 黄玲玲,余鑫,赵泉,等.基于增强 CT 影像组学模型预测食管癌患者放疗完全缓解后肿瘤复发的价值[J].中国临床医学影像杂志,2023,34(2):91-96.

- [9] Le Cun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553):436-444.
- [10] Liu Z, Wang S, Dong D, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges[J]. *Theranostics*, 2019, 9(5):1303-1322.
- [11] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. *The IEEE conference on computer vision and pattern recognition (CVPR)*, 2016:770-778.
- [12] 安超, 张晨, 郑广平, 等. 基于深度学习的胸部 X 线肺结核检测研究及多中心临床验证[J]. *放射学实践*, 2022, 36(6):704-709.
- [13] 陈美容, 周涛, 陈炫幸, 等. 基于 CT 图像的深度学习模型鉴别良恶性肾肿瘤[J]. *放射学实践*, 2023, 37(4):468-473.
- [14] Kim DW, Lee S, Kwon S, et al. Deep learning-based survival prediction of oral cancer patients[J/OL]. *Sci Rep*, 2019, 9(1):e6994 [2019 May 6]. DOI:10.1038/s41598-019-43372-7.
- [15] Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network[J/OL]. *BMC Med Res Methodol*, 2018, 18(1):e24 [2018 Feb 26]. DOI:10.1186/s12874-018-0482-1.
- [16] Matsuo K, Purushotham S, Jiang B, et al. Survival outcome prediction in cervical cancer: Cox models vs. deep-learning model[J/OL]. *Am J Obstet Gynecol*, 2019, 220(4):e381 [2018 Dec 21]. DOI:10.1016/j.ajog.2018.12.030.
- [17] Chi CL, Street WN, Wolberg WH. Application of artificial neural network-based survival analysis on two breast cancer datasets[J/OL]. *AMIA Annu Symp Proc*, 2007: e130-e134 [2007 Oct 11]. PMID:18693812.
- [18] Wang S, Liu Z, Rong Y, et al. Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer[J/OL]. *Radiother Oncol*, 2019, 132: e171-e177 [2023 Feb 16]. DOI:10.1016/j.radonc.2018.10.019. Epub 2018 Nov 1.
- [19] Bando E, Makuuchi R, Tokunaga M, et al. Impact of clinical tumor-node-metastasis staging on survival in gastric carcinoma patients receiving surgery[J]. *Gastric Cancer*, 2017, 20(3):448-456.
- [20] Bando E, Makuuchi R, Irino T, et al. Validation of the prognostic impact of the new tumor-node-metastasis clinical staging in patients with gastric cancer[J]. *Gastric Cancer*, 2019, 22(1):123-129.
- [21] Lee SR, Kim HO, Son BH, et al. Prognostic significance of the metastatic lymph node ratio in patients with gastric cancer[J]. *World Journal of Surgery*, 2012, 36(5):1096-1101.
- [22] Nakanishi Y, Ohara M, Domen H, et al. Differences in risk factors between patterns of recurrence in patients after curative resection for advanced gastric carcinoma[J/OL]. *World J Surg Oncol*, 2013, 11:e98 [17 May 2013]. DOI:10.1186/1477-7819-11-98.
- [23] Kim DH, Oh CA, Oh SJ, et al. Validation of seventh edition AJCC gastric cancer staging modifications[J]. *J Surg Oncol*, 2012, 105(1):26-30.
- [24] Zhang W, Fang M, Dong D, et al. Development and validation of a CT-based radiomic nomogram for preoperative prediction of early recurrence in advanced gastric cancer[J/OL]. *Radiother Oncol*, 2019, 145: e13-e20 [2019 Dec 21]. DOI:10.1016/j.radonc.2019.11.023. Epub 2019 Dec 21.
- [25] Shimada H, Noie T, Ohashi M, et al. Clinical significance of serum tumor markers for gastric cancer: a systematic review of literature by the Task Force of the Japanese Gastric Cancer Association[J]. *Gastric Cancer*, 2014, 17(1):26-33.
- [26] 崔丹丹. 血清 CA125 对胃癌腹膜转移的临床诊断价值[D/OL]. 吉林大学, 2022. DOI:10.27162/d.cnki.gjlin.2021.004259.

(收稿日期:2023-09-06 修回日期:2023-11-27)