

食管鳞癌淋巴结转移的影像组学与预后的关联

陈小琪¹, 徐华霞¹, 王小玥¹, 李锐进², 邱明铤³, 邱模良⁴, 吴凯明⁵, 胡志坚⁶

摘要: **目的** 探讨使用卷积神经网络(CNN)提取的食管鳞癌(ESCC)淋巴结转移相关的影像组学特征与 ESCC 预后的关联。**方法** 回顾性分析 308 例 ESCC 患者的临床资料,其中 154 例有淋巴结转移、154 例无淋巴结转移,按 2:1 比例将其分为训练集和验证集。通过 MRICroGL 软件标记 CT 影像中的淋巴结,使用 CNN 分割提取 ESCC 淋巴结影像特征,通过 LASSO 回归和随机森林筛选与 ESCC 淋巴结转移相关的影像组学特征并构建预测模型,采用 Cox 回归进行特征选择,建立影像组学标签,分析其与 ESCC 预后的关联,进而构建列线图评价模型预测能力。**结果** 使用 CNN 自动提取到 999 个影像组学特征值;使用 LASSO 回归筛选出的 19 个特征,在全模型、训练集和验证集中构建模型的曲线下面积(AUC) (95% CI)分别为 0.747(0.694,0.801)、0.751(0.686,0.817)和 0.766(0.672,0.860);使用随机森林筛选出 9 个特征的预测模型中 AUC (95% CI)分别为 0.692(0.633,0.751)、0.683(0.610,0.755)和 0.723(0.624,0.822)。多因素 Cox 回归分析显示,高风险的影像组学标签与 ESCC 不良的预后相关($P<0.05$)。列线图结果显示,影像组学标签能很好地预测 ESCC 预后,总生存率 C 指数(95% CI)为 0.710(0.670,0.749),无病进展生存率 C 指数(95% CI)为 0.775(0.746,0.804)。**结论** 基于 CNN 和 CT 影像组学构建的影像组学标签对预测 ESCC 预后具有很高的价值。

关键词: 食管鳞癌;影像组学;淋巴结转移;卷积神经网络

文献标志码: A **文章编号:** 1672-4194(2023)02-0103-07

我国是食管癌(esophageal carcinoma, EC)的高发地区,发病率为 11.130/100 000,居我国肿瘤发病率第六位;其死亡率为 8.280/100 000,居我国肿瘤死亡率第五位^[1]。EC 的病理类型包括鳞状细胞癌和腺癌,我国约 90% EC 为食管鳞癌(esophageal squamous cell carcinoma, ESCC)^[2]。淋巴结转移(lymph node metastasis, LNM)是 ESCC 的主要转移方式,是 TNM 分期中重要的病理特征之一,也是影响 ESCC 患者总生存期的关键性预后指标^[3]。由于独特的淋巴引流方式,EC 极易在食管黏膜及黏膜下层丰富的淋巴管网内发生广泛或跳跃性转移^[4],具有较高的 LNM 发生率^[5],导致患者预后不佳。因此,识别 ESCC 患者是否发生 LNM 并评估患者预后显得尤为重要。

影像组学作为一种非侵入性方法,能够从医学影像中挖掘出定量特征,并结合统计学方法筛选出最有价值的影像特征,用于疾病的诊断、疗效评估及

预后预测,包括乳腺癌、肺腺癌、结直肠癌等恶性肿瘤的诊断及预后评估^[5-6],特别是用于识别原发性肿瘤是否伴有 LNM^[7]。已有研究^[8-9]证实,影像组学在预测 ESCC LNM 和预后方面的潜在价值。然而,这些研究仅仅关注原发性肿瘤的影像特征,未探讨淋巴结影像特征与 LNM 的关联。同时,这些研究大多使用人工手动法对影像特征进行分割与提取^[10-12]。人工手动分割不仅耗时,且受经验与主观因素影响比较大,难以适用于大规模的病例研究^[12]。与人工手动分割相比,用卷积神经网络(convolutional neural networks, CNN)获取到的特征可以让复杂的内部信息得到更好的描述,能够发现更深层次的特征^[13],适用于大数据研究。本研究基于 CNN 分割提取到的淋巴结影像组学特征,探讨 LNM 相关的影像组学特征与 ESCC LNM 及预后的关联,为 ESCC 淋巴结的识别及预后判断提供理论依据。

1 对象与方法

1.1 对象 收集 2014 年 6 月—2021 年 9 月福建医科大学附属第一医院 308 例新发 ESCC 患者,其中淋巴结未转移者 154 例,LNM 者 154 例。将研究对象按照 2:1 随机分为训练集($n=206$)和验证集($n=102$)。纳入标准:(1)经手术病理组织学确诊的原发性 ESCC;(2)术前未接受放、化疗;(3)在福建省居住 10 a 以上。排除标准:(1)经病理诊断确诊

收稿日期:2022-06-07

资助项目:福州市级科技计划项目(2019-S-72);福建医科大学大学生创新创业训练计划项目(202110392001)

作者单位:1. 福建医科大学 公共卫生学院,福州 350122;

2. 福建医科大学 基础医学院,福州 350122;

3. 福建医科大学 附属第一医院胸外科,福州 350004;

4. 福建医科大学 附属福州市第一医院影像科,福州 350009;

5. 福建医科大学 附属第一医院影像科,福州 350004;

6. 福建医科大学 公共卫生学院流行病与卫生统计学系,福州 350122

作者简介:陈小琪,女,福建医科大学 2019 级预防医学专业本科生

通信作者:胡志坚. Email: huzhijian@fjmu.edu.cn

为非原发性 ESCC、复发性 ESCC 患者;(2)伴有肝肾功能不全或急慢性感染;(3)病情危重不能清晰回答问题者。本研究通过福建医科大学伦理委员会批准(审批号:201495),患者及家属均知情同意。

1.2 影像资料收集 影像资料来源于福建医科大学附属第一医院胸外科。采用 Toshiba Aquilion One 320 排 CT 机,从胸廓入口到肺底,进行胸部螺旋扫描。扫描参数:根据医院统一标准执行。所有患者数据格式均保持一致,采用 digital imaging and communications in medicine (DICOM)格式。

1.3 影像组学特征提取与筛选

1.3.1 影像资料标记 根据影像指南统一标记标准,由 2 位具有 5 a 以上工作经验的影像科医生使用 MRICroGL 软件标记 ESCC 患者 CT 影像资料中的淋巴结。

1.3.2 图像分割 为确保分割的准确性及节省人力物力,本研究使用 CNN 对标记的淋巴结区域进行自动分割,进而提取 ESCC 淋巴结的影像组学特征。

1.3.3 影像资料提取 为减少各种 CT 扫描仪的影响,统一将 DICOM 图像转换为灰度值图像,并对每个患者自动提取感兴趣区(region of interest, ROI)的影像特征进行进一步分析,使用 CNN 自动提取 ESCC 的 CT 淋巴结特征,共提取到 999 个影像组学特征值用于后续研究。

1.4 统计学处理 采用 SPSS 24.0 软件及 R(4.10)软件进行统计分析。使用 χ^2 检验比较 2 组患者临床特征的分布情况。本研究使用 CNN 对标记的淋巴结区域进行自动分割,并自动提取 ESCC 淋巴结的影像组学特征;进而分别使用 LASSO 回归和随机森林筛选影像组学特征构建 ESCC LNM 的预测模型。根据最佳 λ 纳入 LASSO 回归筛选的影像组学特征;根据十折交叉验证拟合的模型误差及 increase in mean squared error(%IncMSE)纳入随机森林筛选的影像组学特征。以曲线下面积(area under the curve, AUC)作为判断模型的效能;并使用 Delong 法比较,选择最佳模型进行后续研究;根据 Cox 回归(向后法)算法所得回归系数构建与筛选出 ESCC 预后相关的影像组学标签(radiomics signature),并使用 Cox 回归探讨影像组学标签与 ESCC 预后的关联;以 C 指数评价由多因素 Cox 回归分析变量所建列线图的预测能力。

2 结果

2.1 临床特征比较 训练集和验证集在性别、年龄、肿瘤部位、肿瘤直径、分化程度和 T 分期分布均衡可比($P>0.05$,表 1)。

表 1 训练集和验证集患者的临床特征比较
Tab. 1 The comparison of clinical characteristics of training set and validation set

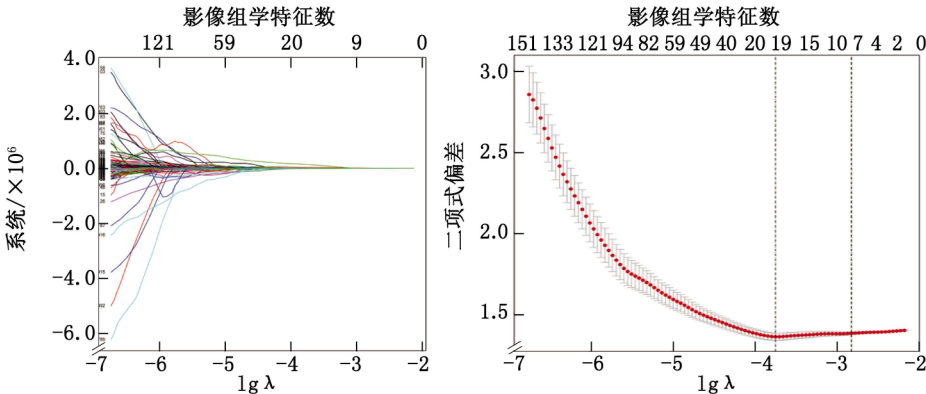
变 量	训练集 ($n=206$)	验证集 ($n=102$)	χ^2	P
性别			0.205	0.651
女	45(21.80)	20(19.60)		
男	161(78.20)	82(80.40)		
年龄/岁			<0.001	0.994
≤ 60	93(45.10)	46(45.10)		
>60	113(54.90)	56(54.90)		
肿瘤部位			1.438	0.487
上段	21(10.20)	14(13.70)		
中段	112(54.40)	56(54.90)		
下段	57(27.70)	23(22.50)		
$d_{\text{肿瘤}}/\text{cm}$			0.998	0.318
≤ 4	122(59.20)	53(52.00)		
>4	77(37.40)	43(42.20)		
分化程度			1.958	0.376
高分化	19(9.20)	7(6.90)		
中分化	122(59.20)	56(54.90)		
低分化	50(24.30)	32(31.40)		
T 分期			2.002	0.157
I / II	66(32.00)	41(40.20)		
III	140(68.00)	61(59.80)		

2.2 筛选与 ESCC LNM 相关影像组学特征构建 ESCC LNM 的预测模型

2.2.1 使用 LASSO 回归筛选影像组学特征构建 ESCC LNM 的预测模型 使用 LASSO 回归筛选影像特征,最佳 λ 为 0.024(图 1)。使用 LASSO 回归,共筛选出 19 个影像特征,分别为:173、417、457、473、504、521、546、583、585、611、678、695、761、812、884、905、932、974 和 987。

使用上述 19 个影像特征构建预测模型,在全模型、训练集和验证集中,其 AUC (95% CI)分别为:0.747(0.694, 0.801)、0.751(0.686, 0.817)和 0.766(0.672, 0.860)(图 2)。

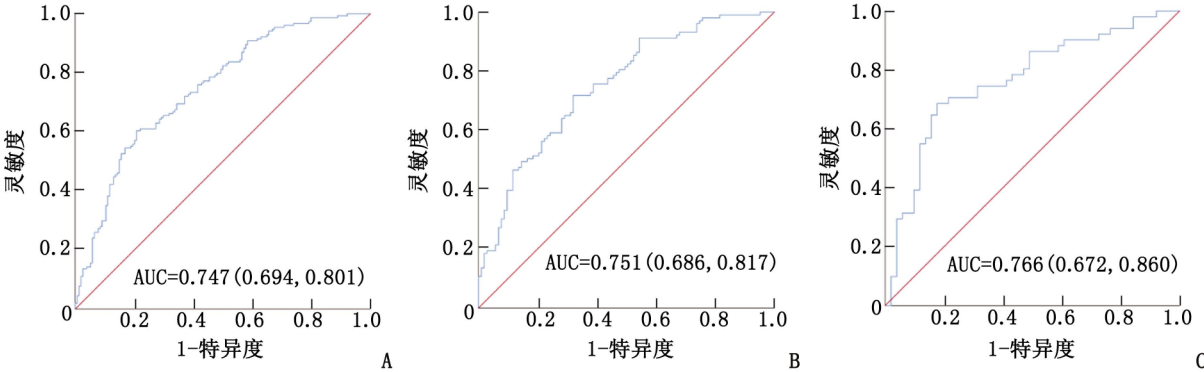
2.2.2 使用随机森林筛选影像组学特征构建 ESCC LNM 的预测模型 根据随机森林计算结果(图 3)。以十折交叉验证曲线中模型误差的稳定性及 %IncMSE 作为筛选影像特征的依据,选择排序为前 9 位的影像特征用于构建 ESCC LNM 预测模型,9 个影像特征分别为 987、306、505、438、111、546、799、566 和 850。



CNN:卷积神经网络。

图 1 使用 LASSO 回归筛选 CNN 提取的影像组学特征

Fig. 1 LASSO was used to screen the radiomics features extracted by CNN



ESCC:食管鳞癌;LNM:淋巴结转移;AUC:曲线下面积。A:全模型;B:训练集;C:验证集。

图 2 使用 LASSO 回归筛选的影像组学特征构建 ESCC LNM 预测模型

Fig. 2 The radiomics features screened by LASSO were used to construct the prediction model of ESCC LNM



%IncMSE:增长的错误率平方均值;CNN:卷积神经网络。

图 3 使用随机森林筛选 CNN 提取的影像组学特征

Fig. 3 The random forest was used to screen the radiomics features extracted by CNN

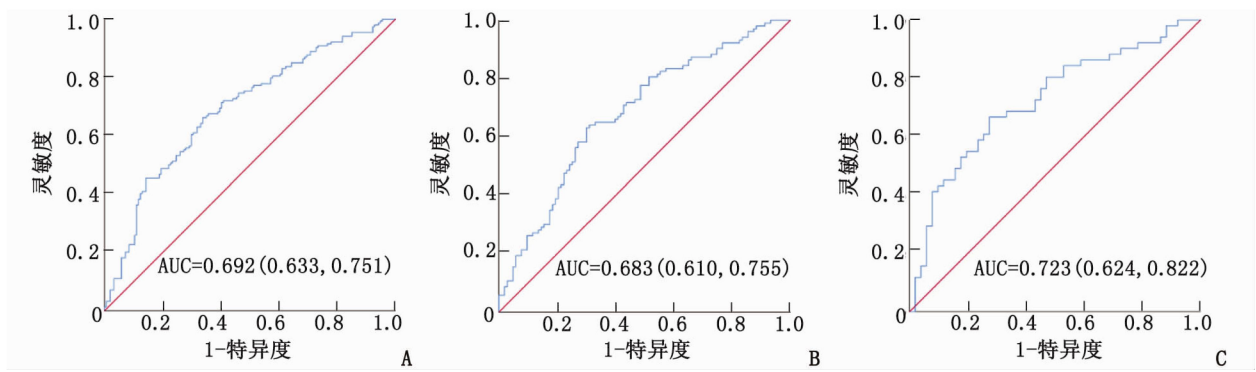
使用上述 9 个影像特征构建预测模型,在全模型、训练集和验证集中,其 AUC (95% CI)分别为:0.692 (0.633, 0.751)、0.683 (0.610, 0.755) 和 0.723 (0.624, 0.822)(图 4)。

2.2.3 筛选最佳预测模型 使用 Delong 法比较 LASSO 回归与随机森林筛选的影像组学特征构建的 ESCC LNM 预测模型。分析发现,使用 LASSO 回归所构建模型的 AUC 优于使用随机森林所构建的模型($Z=-3.955, P<0.001$)。因此,本研究选择使用 LASSO 回归筛选的变量进行后续的分析。

2.3 使用 ESCC LNM 相关的影像组学特征构建 ESCC 预后相关的影像组学标签 通过 Cox 回归(向后法)算法进一步筛选出与 ESCC 预后相关的影像组学特征,根据回归系数构建影像组学标签,具体计算方式如下:

$$\text{影像组学标签} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n \quad (1)$$

式中 X_n 表示影像组学特征变量, β_n 表示所对应的回归系数^[14]。



ESCC:食管鳞癌;LNM:淋巴结转移。AUC:曲线下面积。A:全模型;B:训练集;C:验证集。

图4 使用随机森林筛选的影像组学特征构建 ESCC LNM 预测模型

Fig. 4 The radiomics features screened by random forest were used to construct the prediction model of ESCC LNM

2.4 影像组学标签与 ESCC 预后的关联分析 本研究对象的生存时间中位数为 17.430 个月,1、3 及 5 a 生存率分别为 77% (95% CI: 0.720, 0.830)、52% (95% CI: 0.450, 0.590) 和 42% (95% CI: 0.310, 0.520)。根据影像组学标签的中位数,将其分

为低风险影像组学标签组和高风险影像组学标签组。

调整性别、年龄、肿瘤直径、分化程度、T 分期、术后放化疗和影像组学标签,多因素 Cox 回归分析发现,高风险的影像组学标签与 ESCC 不良的预后相关($P < 0.05$, 表 2)。

表 2 多因素 LASSO Cox 回归分析

Tab. 2 Multivariate LASSO Cox regression analysis

变 量	全模型	训练集	验证集
OS			
低风险指数	1.000	1.000	1.000
高风险指数	1.880(1.220,2.900)	1.420(0.850,2.360)	3.990(1.730,9.190)
DFS			
低风险指数	1.000	1.000	1.000
高风险指数	1.780(1.210,2.620)	1.920(1.190,3.120)	2.040(1.080,3.850)

OS: 总生存率;DFS: 无病进展生存率。

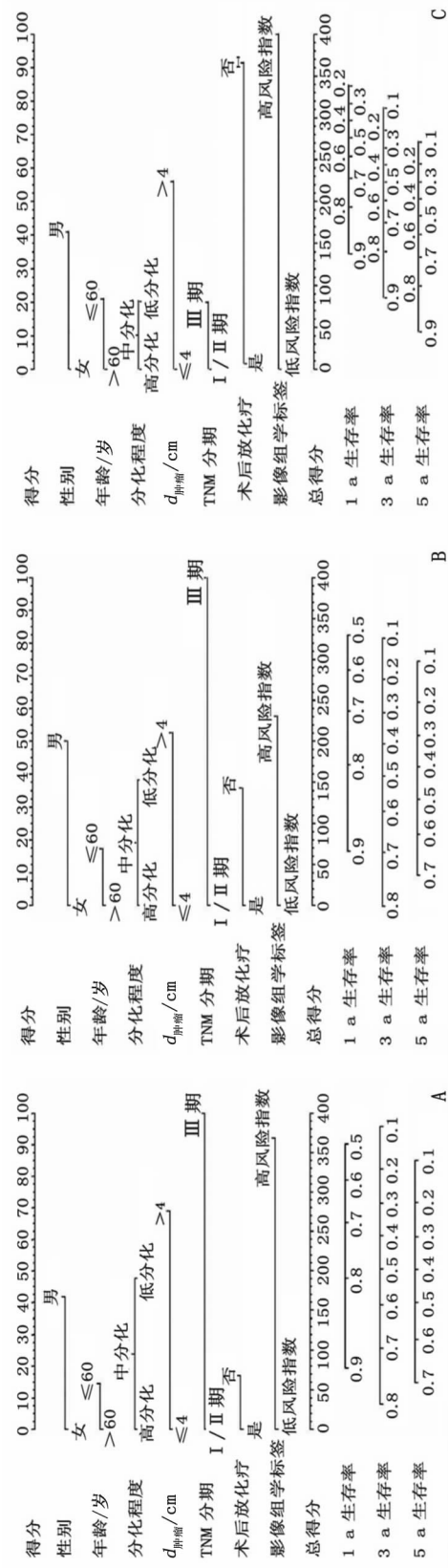
2.5 列线图 使用多因素 Cox 回归分析的变量构建列线图(图 5,图 6),影像组学标签能很好地预测食管癌的预后,总生存率(overall survival, OS) C 指数(95% CI)在全模型、训练集和验证集中分别为 0.710(0.670, 0.749)、0.765(0.742, 0.788)和 0.652(0.568, 0.736),无病进展生存率(disease free survival, DFS) C 指数(95% CI)在全模型、训练集和验证集中分别为 0.775(0.746, 0.804)、0.754(0.726, 0.781)和 0.774(0.760, 0.789)。

3 讨 论

影像组学在医疗领域应用越来越广泛,其在疾病的辅助诊断、疗效评估和预后预测方面已有初步效果^[8-9]。本研究通过 CNN 提取影像组学特征,并使用 LASSO 回归筛选与 ESCC LNM 相关的影像组学特征,构建 ESCC LNM 的预测模型,发现该模型识别 LNM 的效果较好(AUC=0.750, 95% CI:

0.690, 0.800)。同时,多因素 Cox 回归分析发现,高风险影像组学标签与 ESCC 不良预后有关。此外,列线图结果显示,影像组学标签在预测 ESCC 预后方面具有较高的价值。

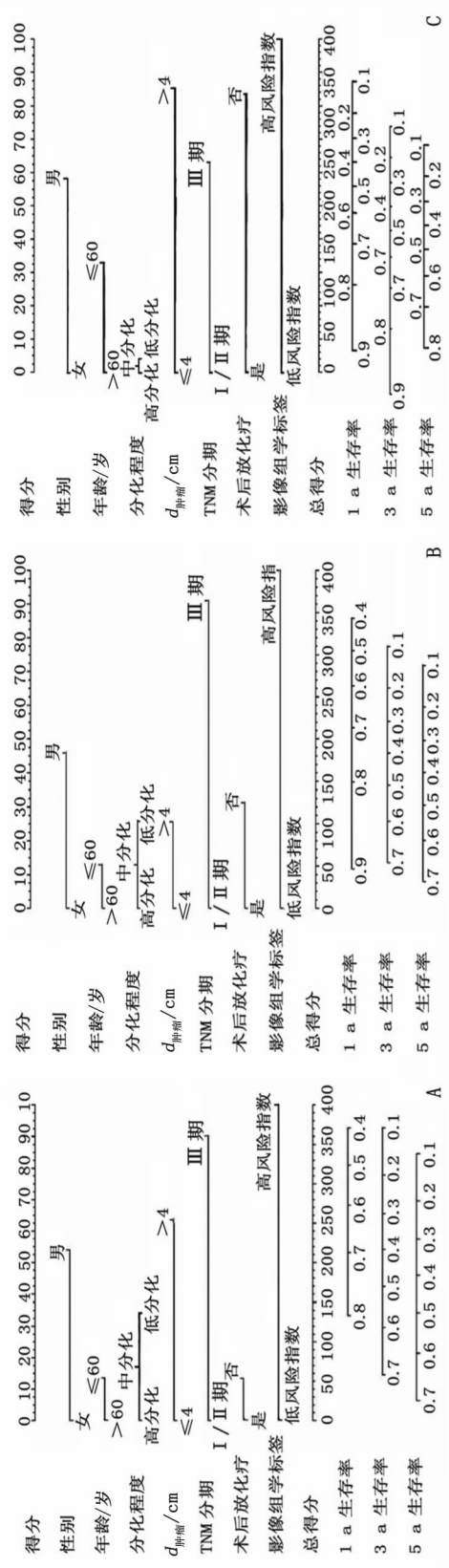
近年来,多项研究^[5,15-17]证实影像组学在 LNM 的预测价值。如 ZHAO 等^[18]基于 ESCC 患者临床特征及计算机断层扫描提取到的影像组学特征,构建 LNM 预测模型,训练集和验证集的 AUC 分别为 0.870 和 0.840。赵博^[19]的研究发现,基于 CT 影像组学可以较好地预测 ESCC LNM,且对“隐匿”LNM 的预测也有一定的潜力,训练组和验证组的 AUC 分别为 0.870 和 0.840。然而,上述研究大多采用手动分割对 CT 影像资料进行图像分割和特征提取,耗时又耗力,且手动分割存在分割不准确、对边界判断具有主观性等问题^[10-12]。本研究通过 CNN 来分割并提取 ESCC LNM 的影像组学特征,其分割精准,且能节省时间,并择优选择 LASSO 回



A: All model; B: Training set; C: Validation set. OS: Overall survival.

图 5 预测 OS 的临床预后列线图

Fig. 5 Nomogram for predicting the clinical prognosis of OS



A: All model; B: Training set; C: Validation set. DFS: Disease-free survival.

图 6 预测 DFS 的临床预后列线图

Fig. 6 Nomogram for predicting the clinical prognosis of DFS

归对提取到的 999 个特征值进行筛选,共筛选出 19 个影像特征并构建预测 LNM 的预测模型,其 AUC (95% CI)在全模型、训练集和验证集中分别为 0.747 (0.694, 0.801)、0.751 (0.686, 0.817) 和 0.766 (0.672, 0.860)。

使用 Cox 回归筛选与预后相关的影像组学特征,并根据回归系数构建影像组学标签,根据中位数将其分为低风险和高风险影像组学标签,结果显示,与低风险影像组学标签比较,高风险影像组学标签与不良预后相关,说明影像组学标签与 ESCC 预后存在关联,可用于 ESCC 预后的预测,与文献 [8-9] 结论一致。此外,基于影像组学标签,构建 ESCC 预后预测的列线图,分析发现总 OS C 指数 (95% CI) 在全模型、训练集和验证集中分别为 0.710 (0.670, 0.749)、0.765 (0.742, 0.788) 和 0.652 (0.568, 0.736), DFS C 指数 (95% CI) 分别为 0.775 (0.746, 0.804)、0.754 (0.726, 0.781) 和 0.774 (0.760, 0.789),说明影像组学标签能较好地预测 ESCC 的预后。但本研究为小样本、单中心研究,模型的外推使用及稳定性有待进一步进行多中心、大样本的验证。

参考文献:

- [1] ZHENG R, ZHANG S, ZENG H, et al. Cancer incidence and mortality in China, 2016[J]. JNCC, 2022, 2(1): 1-9.
- [2] RICE T W, ISHWARAN H, HOFSTETTER W, et al. Esophageal cancer: Associations with (pN+) lymph node metastases[J]. Annals of Surgery, 2017, 265(1): 122-129.
- [3] CHEN L, YU S, JIANG X, et al. Supraclavicular lymph node metastasis in elderly patients undergoing esophageal squamous cell carcinoma radical surgery: Construction of risk and prognostic predictive nomograms[J]. J Thorac Dis, 2021, 13(1): 18-30.
- [4] ZHONG H, XU Y, WANG J, et al. Overexpression of microRNA-19a-3p promotes lymph node metastasis of esophageal squamous cell carcinoma via the RAC1/CDC42-PAK1 pathway[J]. Transl Cancer Res, 2021, 10(6): 2694-2706.
- [5] ZHU Y, YANG L, SHEN H. Value of the application of CE-MRI radiomics and machine learning in preoperative prediction of sentinel lymph node metastasis in breast cancer[J]. Front Oncol, 2021, 11: 757111.
- [6] RAN J, CAO R, CAI J, et al. Development and validation of a nomogram for preoperative prediction of lymph node metastasis in lung adenocarcinoma based on radiomics signature and deep learning signature[J]. Front Oncol, 2021, 11: 585942.
- [7] LIU J, SUN D, CHEN L, et al. Radiomics analysis of dynamic contrast-enhanced magnetic resonance imaging for the prediction of sentinel lymph node metastasis in breast cancer[J]. Front Oncol, 2019, 9: 980.
- [8] LI X, LIU Q, HU B, et al. A computed tomography-based clinical-radiomics model for prediction of lymph node metastasis in esophageal carcinoma[J]. J Cancer Res Ther, 2021, 17(7): 1665-1671.
- [9] 姜泽坤. 影像组学技术研究及临床肿瘤诊疗应用[D]. 济南: 山东师范大学, 2021.
- [10] 谭显政. 基于 CT 影像组学精准预测可切除性食管鳞癌淋巴结转移和无病生存期[D]. 广州: 南方医科大学, 2019.
- [11] 董璐. 基于肿瘤-淋巴双区域影像组学的食管癌生存期预测[D]. 深圳: 中国科学院大学(中国科学院深圳先进技术研究院), 2021.
- [12] TRULLO R, PETITJEAN C, NIE D, et al. Fully automated esophagus segmentation with a hierarchical deep learning approach[J]. Conf Proc IEEE Int Conf Signal Image Process Appl, 2017, 2017: 503-506.
- [13] WANG L, CHANG C, LIU Z, et al. A medical image fusion method based on SIFT and deep convolutional neural network in the SIST domain[J]. J Healthc Eng, 2021, 2021: 9958017.
- [14] 王卿宇. 食管癌 CT 影像组学标签构建及预后模型研究[D]. 太原: 太原理工大学, 2020.
- [15] 罗何三. 基于治疗前 CT 影像组学预测食管鳞癌放疗疗效和预后的临床研究[D]. 广州: 南方医科大学, 2020.
- [16] YU Y, HE Z, OUYANG J, et al. Magnetic resonance imaging radiomics predicts preoperative axillary lymph node metastasis to support surgical decisions and is associated with tumor microenvironment in invasive breast cancer: A machine learning, multicenter study[J]. EBioMedicine, 2021, 69: 103460.
- [17] BI L, LIU Y, XU J, et al. A CT-based radiomics nomogram for preoperative prediction of lymph node metastasis in perampullary carcinomas[J]. Front Oncol, 2021, 11: 632176.
- [18] ZHAO B, ZHU H, LI X, et al. Predicting lymph node metastasis using computed tomography radiomics analysis in patients with resectable esophageal squamous cell carcinoma [J]. J Comput Assist Tomogr, 2021, 45(2): 323-329.
- [19] 赵博. CT 影像组学在预测食管癌淋巴结转移、化疗疗效及预后中的应用[D]. 北京: 北京大学医学部, 2021.

Association between Radiomics Related to Lymph Node Metastasis and Prognosis of Esophageal Squamous Cell Carcinoma

CHEN Xiaoqi¹, XU Huaxia¹, WANG Xiaoyue¹, LI Ruijin², QIU Minglian³,
QIU Moliang⁴, WU Kaiming⁵, HU Zhijian⁶

1. Public Health School of Fujian Medical University, Fuzhou 350122, China;

2. School of Basic Medical Sciences Fujian Medical University, Fuzhou 350122, China;

3. Department of Thoracic Surgery, The First Affiliated Hospital of Fujian Medical University, Fuzhou 350004, China;

4. Department of Imaging, Fuzhou First Hospital Affiliated to Fujian Medical University, Fuzhou 350009, China;

5. Department of Imaging, The First Affiliated Hospital of Fujian Medical University, Fuzhou 350004, China;

6. Epidemiology and Health Statistics, Public Health School of Fujian Medical University, Fuzhou 350122, China

ABSTRACT: **Objective** To explore the relationship between the radiomics features associated with lymph node metastasis (LNM) of esophageal squamous cell carcinoma(ESCC) and the prognosis of ESCC by using convolutional neural networks (CNN) screening. **Methods** This study enrolled 308 patients with pathologically confirmed advanced ESCC, including 154 with LNM and 154 without LNM. All enrolled patients were randomly divided into a training cohort and a testing cohort at a ratio of 2 : 1. The lymph nodes in CT images are marked by MRIcroGL software, then radiomics features of ESCC lymph nodes were segmented and extracted by CNN. Radiomics features associated with ESCC LNM were screened by LASSO method and the random forest, these features were used to construct a prediction model. Using Cox regression to feature selection, meanwhile establishing radiomics signature based on those selected features. And analyzing the association between radiomics signature and ESCC prognosis, and then constructing the nomogram to evaluate the predictive ability of the model. **Results** By using CNN, 999 radiomics signature feature values are automatically extracted. 19 features were screened out by LASSO, using these features to build predictive models of ESCC LNM, with 0.747(95% CI: 0.694, 0.801) in the whole cohort, 0.751(95% CI: 0.686, 0.817) in the training cohort and 0.766(95% CI: 0.672, 0.860) in the testing cohort. 9 features were screened out by the random forest and the prediction model of ESCC LNM was constructed, which were 0.692(95% CI: 0.633, 0.751), 0.683(95% CI: 0.610, 0.755) and 0.723(95% CI: 0.624, 0.822) in the whole cohort, training cohort and testing cohort, respectively. Multivariate Cox regression analysis showed that high-risk radiomics signature was associated with poor prognosis of ESCC ($P < 0.05$). The Nomogram results showed that radiomics signature could more precisely predict the prognosis of ESCC. The C-index of the nomogram for overall survival was 0.710 (95% CI: 0.670, 0.749) and disease free survival was 0.775 (95% CI: 0.746, 0.804). **Conclusion** The radiomics signature constructed based on the computed tomography radiomics features related to LNM of ESCC screened by CNN has a high value in predicting the prognosis of ESCC.

KEY WORDS: esophageal squamous cell carcinoma; radiomics; lymphatic metastasis; convolutional neural networks

(编辑:张慧茹)