

基于增强CT可解释深度学习与临床特征 构建腹膜后软组织肉瘤病理分级及 预后预测列线图

夏 珺¹, 庞 婧¹, 刘媛媛¹, 邱子尧¹, 郭 莉², 任辰晨², 刘 源², 丁伟伟³, 王鹤翔^{1*}

¹青岛大学附属医院放射科, 山东 青岛

²青岛大学计算机科学技术学院, 山东 青岛

³宁夏医科大学总医院放射科, 宁夏 银川

收稿日期: 2026年8月11日; 录用日期: 2026年9月5日; 发布日期: 2026年9月11日

摘 要

目的: 探讨基于三期增强CT构建的可解释性深度学习列线图(Deep Learning Nomogram, DLN)模型在术前无创预测腹膜后软组织肉瘤(retroperitoneal sarcoma, RPS)病理分级及评估患者预后中的价值。方法: 回顾性分析2010年1月至2023年12月两家三甲医院的193例初诊原发性RPS患者数据(训练集116例, 外部验证集77例)。提取三期增强CT的影像组学与ResNet34深度学习特征, 通过残差多头注意力机制融合后结合临床预测因子构建DLN模型。采用ROC曲线、校准曲线及DCA评估模型诊断效能与临床获益; 利用Grad-CAM和SHAP实现像素级解释; 结合Kaplan-Meier分析与多因素Cox回归评估其预后分层及独立预测价值。结果: 在外部验证集中, DLN模型预测高级别RPS的AUC为0.882 (95% CI: 0.792~0.971), 优于单纯临床模型(AUC: 0.530, $P < 0.001$)及深度学习模型(AUC: 0.878), 且DCA显示最高临床净获益。可解释性分析证实高级别病灶激活区集中于肿瘤异质性核心。以0.69为截断值, DLN高风险组患者的PFS显著劣于低风险组(训练集HR = 2.555, $P = 0.013$; 验证集HR = 7.382, $P = 0.002$), 多因素Cox回归证实DLN是PFS的独立危险因素。结论: 基于三期增强CT的DLN模型能够在术前非侵入性评估RPS的病理分级和预后风险, 为临床制定个体化治疗方案和优化决策提供有力工具。

关键词

腹膜后软组织肉瘤, 深度学习, 病理分级, 风险分层, 预后

Contrast-Enhanced CT-Based Interpretable Deep Learning and Clinical Nomogram for Predicting Pathologic Grade and Prognosis in Retroperitoneal Sarcoma

*通讯作者。

文章引用: 夏珺, 庞婧, 刘媛媛, 邱子尧, 郭莉, 任辰晨, 刘源, 丁伟伟, 王鹤翔. 基于增强CT可解释深度学习与临床特征构建腹膜后软组织肉瘤病理分级及预后预测列线图[J]. 临床医学进展, 2026, 16(9): 1427-1438.

DOI: 10.12677/acm.2026.1693180

Jun Xia¹, Jing Pang¹, Yuanyuan Liu¹, Ziyao Qiu¹, Li Guo², Chenchen Ren², Yuan Liu², Weiwei Ding³, Hexiang Wang^{1*}

¹Department of Radiology, The Affiliated Hospital of Qingdao University, Qingdao Shandong

²College of Computer Science and Technology, Qingdao University, Qingdao Shandong

³Department of Radiology, General Hospital of Ningxia Medical University, Yinchuan Ningxia

Received: August 11, 2026; accepted: September 5, 2026; published: September 11, 2026

Abstract

Objective: To establish and validate an interpretable deep learning nomogram (DLN) model based on contrast-enhanced CT (CECT) for noninvasively predicting pathologic grade and prognosis in retroperitoneal sarcoma (RPS). **Methods:** Retrospective data of 193 patients with primary RPS were collected from two centers between January 2010 and December 2023 (training cohort: $n = 116$; external validation cohort: $n = 77$). Radiomic and ResNet34-derived deep learning features from triphasic CECT were fused via a residual multi-head attention mechanism and integrated with independent clinical predictors to establish the DLN. Model performance was evaluated using ROC curves, calibration plots, and DCA. Model interpretability was achieved using Grad-CAM and SHAP. Kaplan-Meier analysis and multivariate Cox regression were applied to evaluate its prognostic accuracy for PFS. **Results:** The DLN model achieved an AUC of 0.882 (95% CI: 0.792~0.971) for high-grade RPS prediction in the external validation cohort, outperforming the clinical model (AUC: 0.530, $P < 0.001$) and deep learning signature (AUC: 0.878). DCA demonstrated maximum clinical net benefit for DLN. Interpretability analysis revealed high model focus concentrated within the heterogeneous tumor core of high-grade lesions. High-risk patients defined by DLN had significantly worse PFS in both training (HR = 2.555, $P = 0.013$) and validation cohorts (HR = 7.382, $P = 0.002$), with DLN identified as an independent prognostic factor. **Conclusions:** The externally validated DLN model offers a non-invasive, accurate tool for preoperative grading and prediction of progression-free survival in patients with RPS, enabling personalized therapy and improving clinical decision making.

Keywords

Retroperitoneal Sarcoma, Deep Learning, Pathological Grading, Risk Stratification, Prognosis

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

腹膜后软组织肉瘤(retroperitoneal sarcoma, RPS)是一类体积庞大且解剖关系复杂的罕见肿瘤, 约占所有软组织肉瘤的 12%~15%, 其预后显著差于肢体肉瘤[1][2]。由于 RPS 解剖位置深在、确诊时肿瘤体积往往较大且易侵犯邻近脏器, 给临床根治性治疗带来了极大挑战[3]。目前, 外科手术仍是 RPS 唯一可能实现治愈的治疗手段[4]。然而, 多中心研究表明, 术前放疗仅能使法国总抗癌中心联盟(FNCLCC) 1 级(低级别) RPS 患者获益, 对 2 级或 3 级(高级别)患者并无显著疗效[5]。因此, 术前预测 RPS 的病理分级对于制定个体化治疗方案具有重要临床意义。

临床上, 对软组织肉瘤患者进行穿刺活检以获取术前病理分级是常规流程[6]。但由于 RPS 位置深且

具局部侵袭性, 穿刺取样不仅操作难度大, 且因肿瘤高度异质性极易产生取样偏倚[7]。因此, 亟需开发一种无创、准确且为临床提供有参考价值的术前评估方法。

多期增强 CT (CECT) 在 RPS 的诊断、分期及术后随访中发挥着不可替代的作用[8] [9]。近年来, 人工智能及深度学习(DL)等先进图像分析技术在肢体及躯干软组织肉瘤的分期与预后评估中展现出巨大潜力[10] [11], 但其在 RPS 中的应用仍处于探索阶段。现有研究虽然尝试利用 CT 影像组学预测 RPS 常见亚型的组织学分级[8], 但往往受限于特定病理类型, 缺乏在全谱系 RPS 中的泛化能力。近年来, Transformer 网络因其核心的自注意力机制突破了传统卷积操作的局部感受野限制, 在捕获长距离依赖关系方面展现出优异性能, 并在医学图像分析中取得了诸多突破[12]-[14]。然而, 深度学习模型的“黑盒”特性极大地限制了其在实际临床中的推广。基于梯度加权类激活映射(class activation maps, CAMs)与沙普利加性解释(Shapley additive explanations, SHAP)等可解释性技术, 可以提供像素级别的决策可视化, 从而提升临床医师对模型预测结果的合理性评估与信任度[15] [16]。

鉴于此, 本研究旨在构建并验证一种基于三期增强 CT 的可解释性深度学习列线图(DLN)模型, 以期在术前非侵入性预测原发性 RPS 患者的病理分级, 并评估其在无进展生存期(PFS)风险分层中的预测价值。

2. 资料与方法

2.1. 患者资料

本研究为回顾性队列研究。回顾性收集 2010 年 1 月至 2023 年 12 月青岛大学附属医院与宁夏医科大学总医院经病理证实为原发性 RPS 的患者。纳入标准:(1) 经术后病理确诊为 RPS 且临床病理资料完整;(2) 术前具备完整的动脉期、静脉期及延迟期三期腹部 CECT 图像。排除标准:(1) 术前接受过新辅助化疗、放疗等相关治疗;(2) 临床或随访资料缺失;(3) CT 图像存在严重伪影影响肿瘤观察评估。共纳入 193 例患者, 其中青岛大学附属医院 116 例作为训练集, 宁夏医科大学总医院 77 例作为外部验证集[17]。本研究遵循《个体预后或诊断多因素预测模型透明报告声明》(TRIPOD)标准。

2.2. 临床资料收集及影像学分析

收集患者基本信息(年龄、性别、Ki67 指数等)和第 8 版美国癌症联合委员会分级系统(T 分期、N 分期、M 分期)。肿瘤病理分级依据法国癌症中心联合会(French Fédération Nationale des Centres de Lutte Contre le Cancer, FNCLCC)分级系统进行评定[18]。其中, FNCLCC 1 级定义为低级别肿瘤, FNCLCC 2 级和 3 级定义为高级别肿瘤。

由 2 名具有 6 年以上 RPS 影像诊断经验的放射科医师分别评估 CECT 图像, 评估前不告知临床和病理学结果, 意见不一致时通过讨论达成一致。多发病灶 CAM 选取体积最大的病灶进行分析。

2.3. 感兴趣区(ROI)分割

采用 ITK-SNAP 开源软件(v.3.8.0)对三期增强 CT 图像进行感兴趣区(ROI)勾画。勾画范围沿肿瘤边缘紧贴进行, 包含肿瘤内部的囊变坏死及出血区域, 并尽可能排除邻近正常组织与粗大血管。为兼顾肿瘤整体形态评估与特征提取的一致性, 首先由医师定位肿瘤体素范围, 随后鉴于不同中心 CT 扫描层厚(3~5 mm)存在差异, 为消除三维重采样可能引入的层间内插伪影, 最终确定并提取肿瘤最大异质性截面(最大单层截面)的 ROI 用于后续特征提取与深度模型构建。所有图像的 ROI 勾画由 1 名具有 6 年放射诊断经验的医师完成勾画, 并由 1 名具有 10 年以上经验的高级放射科医师进行审核; 若存在分歧, 通过两人共同讨论协商达成一致。

2.4. 图像预处理、数据增强与不平衡调优

针对三期增强 CT 图像, 首先应用软组织窗(窗宽 400 HU, 窗位 40 HU)标准化像素强度, 并线性缩放至[0, 255]范围, 以优化肿瘤显示的对比度。以最大截面层面审核通过的 ROI 为基准提取肿瘤外接边界框, 并在高度和宽度方向上均按 1.05 倍系数进行等比例外扩, 用以捕获肿瘤边缘及周围微环境的隐匿征象。将裁剪后的各期图像作为独立通道, 按动脉期、静脉期和延迟期进行通道堆叠, 构建 3 通道输入张量, 从而使深度学习模型能够同时提取多期动态强化模式。最后, 将像素值归一化至[0, 1]范围, 并采用双三次插值法将所有通道统一调整分辨率至 256×256 像素。

为防止模型过拟合, 训练集独立应用了涵盖几何形变(弹性形变、网格畸变)、颜色与对比度调节(CLAHE、锐化)及噪声模拟(高斯噪声、模糊)的数据增强算法(应用概率 20%~25%)。此外, 针对复发(标签 1)与未复发(标签 0)样本的类别不平衡及临床漏诊风险, 引入加权交叉熵损失函数, 将两类损失权重分别设为 0.8 与 2.6, 通过加大对少数派误判的惩罚力度以消除预测偏置。

2.5. 深度学习融合网络与模型构建

2.5.1. 影像组学特征提取与筛选

使用 PyRadiomics 从三期 CECT 提取组学特征(每期 1220 个)并进行 Z-score 标准化。基于训练集, 依次通过 Mann-Whitney U 检验($P < 0.05$)和基于最优 $\lambda_{\min}$ 的 10 折交叉验证 LASSO 回归进行特征降维, 各期独立筛选出 16 个特征并拼接为 48 维组学特征向量。

2.5.2. 多模态特征融合与模型构建

将堆叠的三通道图像张量输入 ImageNet 预训练的 ResNet34 网络提取深度特征, 通过 6 头残差多头注意力模块(Dropout = 0.2)自适应重加权融合深度特征与组学特征向量, 构建深度学习(DL)模型。采用单因素与多因素 Logistic 回归筛选独立临床预测因子。最后将 DL 评分与独立临床因子联合构建 DLN 模型。

2.6. 生存分析

患者随访采用三期 CECT, 术后 2 年内每 3~6 个月复查 1 次, 2 年后每 6 个月复查 1 次。无进展生存期(PFS)定义为从手术切除之日起至疾病进展(局部复发或远处转移)或末次随访的时间。根据训练集中 DLN 计算出的最佳截断值(0.69), 将患者分为高风险组与低风险组。

2.7. 统计学分析

统计学分析使用 R 软件(v.4.2.2)和 SPSS26.0。Kolmogorov-Smirnov 检验用于连续变量的正态性检验, Mann-Whitney U 检验用于连续变量的组间比较, χ^2 检验或 Fisher 确切概率法用于分类变量的组间比较。采用一致性指数、ROC 曲线下面积(area under the ROC curve, AUC)评价模型在训练集和外部验证集中的模型效能。采用 DeLong 检验比较各模型间 AUC 的差异。采用校准曲线评价模型的拟合优度, 决策曲线用于评价临床净获益。采用 Kaplan-Meier 生存曲线和 log-rank 检验评价基于 DLN 的 PFS 风险分层表现。采用单因素和多因素 Cox 比例风险模型评估独立预后因子。 $P < 0.05$ 为差异有统计学意义。

3. 结果

3.1. 临床资料

193 例 RPS 病理学类型见表 1。训练集(116 例)与外部验证集(77 例)RPS 患者的临床及 CT 形态学特征比较见表 2。训练集中, 高级别组男性比例(67.65% vs. 47.56%, $P = 0.048$)及分叶状形态比例(35.29% vs. 14.63%, $P = 0.028$)显著高于低级别组。外部验证集中, 低级别与高级别组间所有基线变量差异均无统计

学意义($P > 0.05$)。

Table 1. Pathologic type distribution of retroperitoneal sarcoma in training and external validation cohorts

表 1. 训练集和外部验证集中腹膜后软组织肉瘤病理类型分布(例)

病理类型	训练队列 (n = 116)	外部验证队列 (n = 77)
多形性肉瘤，未分化型	3	3
恶性周围神经鞘瘤	0	5
神经母细胞瘤	0	2
尤文肉瘤	2	0
横纹肌肉瘤	1	1
间皮瘤	0	2
神经内分泌肿瘤	0	3
平滑肌肉瘤	22	7
上皮样肉瘤	0	2
原始神经外胚层肿瘤	1	1
未分化肉瘤	1	0
黏液纤维肉瘤	1	1
脂肪肉瘤	74	40
其他	11	10

Table 2. Comparison of clinical characteristics and CT features between low- and high-grade RPS in training and external validation cohorts

表 2. 训练集和外部验证集中低级别与高级别腹膜后软组织肉瘤患者临床资料及 CT 特征比较

变量	训练队列(n = 116)			验证队列(n = 77)		
	低级别 (n = 82)	高级别 (n = 34)	<i>P</i>	低级别 (n = 54)	高级别 (n = 23)	<i>P</i>
Ki67 (%), 中位数(IQR)	20.00 (5.00, 30.00)	15.00 (5.00, 40.00)	0.654	10.00 (5.00, 30.00)	25.00 (9.00, 30.00)	0.109
性别(例)			0.048			0.861
男性	39 (47.56)	23 (67.65)		27 (50.00)	12 (52.17)	
女性	43 (52.44)	11 (32.35)		27 (50.00)	11 (47.83)	
AJCC (例)			0.120			0.521
I	25 (30.49)	4 (11.76)		17 (31.48)	4 (17.39)	
II	3 (3.66)	1 (2.94)		1 (1.85)	1 (4.35)	
III	50 (60.98)	26 (76.47)		32 (59.26)	17 (73.91)	
IV	4 (4.88)	3 (8.82)		4 (7.41)	1 (4.35)	
形态(例)			0.028			0.927
类圆形	17 (20.73)	9 (26.47)		15 (27.78)	5 (21.74)	
分叶状	12 (14.63)	12 (35.29)		10 (18.52)	4 (17.39)	
哑铃状	1 (1.22)	0 (0.00)		2 (2.7)	1 (4.35)	
不规则形	52 (63.41)	13 (38.24)		27 (50.00)	13 (56.52)	

续表

囊变坏死(例)			0.082			0.332
无	41 (50.00)	11 (32.35)		30 (55.56)	10 (43.48)	
有	41 (50.00)	23 (67.65)		24 (44.44)	13 (56.52)	
PFS (月), 中位数(IQR)	9.00 (3.25, 14.00)	8.50 (5.25, 13.75)	0.792	20.50 (8.00, 39.00)	20.00 (9.50, 27.50)	0.705

注: IQR = 四分位距; AJCC = 美国癌症联合委员会; PFS = 无进展生存期。

3.2. 临床模型的诊断效能

多因素 Logistic 回归结果显示, 性别与 AJCC 分期是预测 RPS 高级别病理分级的独立预测因素(见表 3), 并进一步用于构建 DLN 模型。临床模型在训练集和外部验证集中的 AUC 分别为 0.923 (95% CI: 0.879~0.968)和 0.530 (95% CI: 0.402~0.659)。

Table 3. Univariable and multivariable Logistic regression analysis of clinical and CT features associated with RPS pathologic grade

表 3. RPS 病理分级相关临床及 CT 特征的单因素与多因素 Logistic 回归分析

变量	单因素			多因素		
	OR	95% CI	P	OR	95% CI	P
Ki67	1.015	0.994~1.035	0.157	0.987	0.951~1.024	0.489
性别	0.484	0.213~1.103	0.084	0.224	0.053~0.938	0.041
AJCC	38.515	5.585~265.605	<0.001	110.369	8.698~1400.501	<0.001
形态	0.79	0.582~1.072	0.130	0.573	0.304~1.082	0.086
囊变坏死	1.87	0.821~4.257	0.136	2.249	0.529~9.572	0.273

注: OR = 比值比; CI = 置信区间。

3.3. DL 模型及 DLN 模型的诊断效能评价

在外部验证队列中, DL 模型的预测效能优于临床模型, AUC 分别为 0.878 (95% CI: 0.788~0.967)和 0.530 (95% CI: 0.402~0.659), 差异具有统计学意义(DeLong 检验, $P < 0.001$)。进一步融合临床特征后构建的 DLN 模型在外部验证队列中的 AUC 为 0.882 (95% CI: 0.792~0.971), 较单独 DL 模型略有提高, 但两者差异无统计学意义(DeLong 检验, $P = 0.157$)。

各模型在训练集和外部验证集中评估 RPS 患者病理分级的效能见表 4, 图 1。DLN 得分构成见图 2(A), 在融合列线图中, 深度学习特征比传统临床变量展现出更高的权重贡献度。在训练集和外部验证集中, 高级别与低级别 RPS 患者的 DLN 评分差异均具有统计学意义(见图 2(B)、图 2(C))。校准曲线与决策曲线分析证明, DLN 模型预测概率与实际观察一致性良好, 且能获得最高临床净获益(见图 3)。

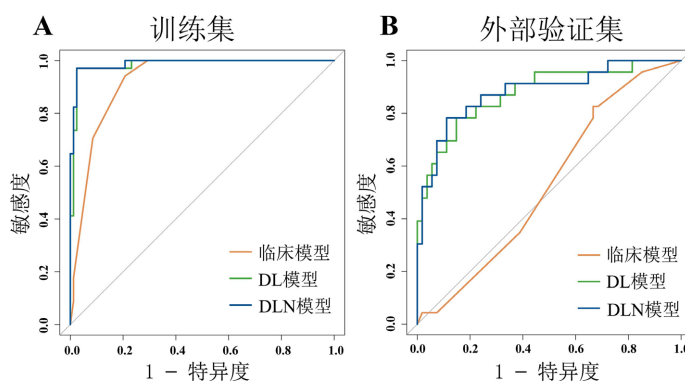
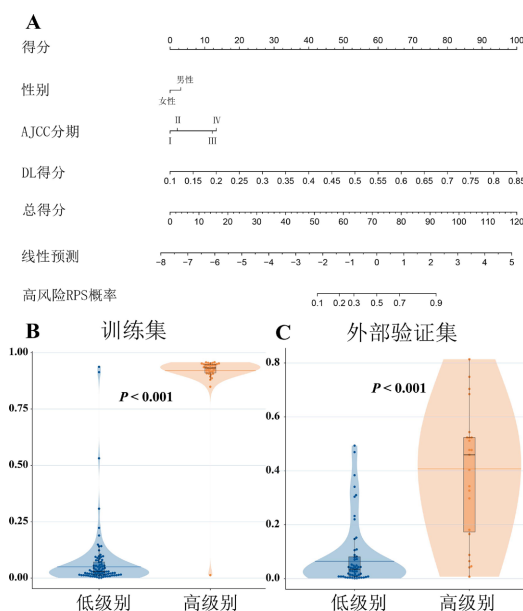
3.4. DLN 模型的可解释性分析

图 4(A)展示了训练集中 DLN 风险评分分布。为进一步提高 DLN 模型的可解释性, 本研究采用 CAM 和 SHAP 分析方法对模型预测依据进行可视化解释[19]。结合 CAM、SHAP 及个体化列线图, 本研究对不同风险梯度患者的决策模式进行了深度可视化(见图 4(B))。CAM 可反映模型预测过程中不同像素区域

Table 4. Performance of different models in evaluating pathologic grade of RPS patients**表 4.** 模型评估腹膜后软组织肉瘤患者病理分级的效能

队列	模型	AUC	95% CI	准确率	灵敏度	特异度
训练队列(n = 116)	临床模型	0.923	0.879~0.968	0.836	0.941	0.793
	DL 模型	0.984	0.963~1.000	0.974	0.971	0.976
	DLN 模型	0.988	0.973~1.000	0.974	0.971	0.976
外部验证队列(n = 77)	临床模型	0.530	0.402~0.659	0.481	0.826	0.333
	DL 模型	0.878	0.788~0.967	0.831	0.783	0.852
	DLN 模型	0.882	0.792~0.971	0.857	0.783	0.889

注: DL 模型 = 深度学习模型; DLN 模型 = 深度学习列线图模型; CI = 置信区间; AUC = 受试者工作特征曲线下面积。

**Figure 1.** Receiver operating characteristic (ROC) curves of the clinical model, deep learning (DL) model, and DL nomogram (DLN) for predicting pathologic grade in the training and external validation cohorts**图 1.** 训练集和外部验证集中临床模型、DL 模型与 DLN 模型预测病理分级的 ROC 曲线**Figure 2.** Construction and predictive performance of the DLN model. (A) Nomogram visualization of the DLN model. (B) (C) Associations between DLN score and pathologic grade in the training and external validation cohorts**图 2.** DLN 模型的构建及预测效能。(A) DLN 模型的列线图可视化。(B) (C) 训练集与外部验证集中 DLN 评分与病理分级的相关性分析

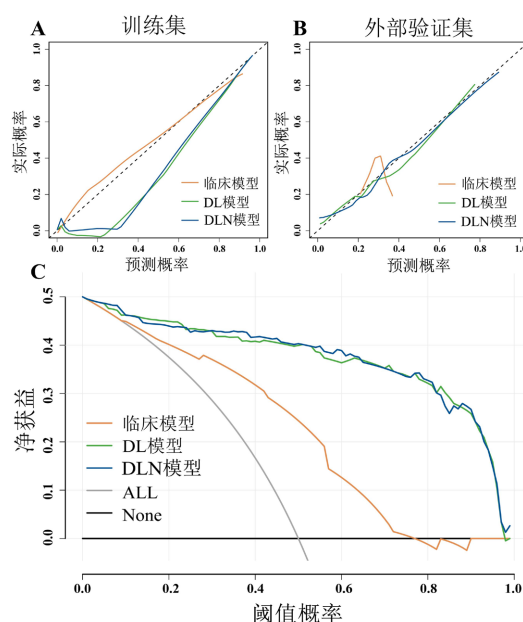


Figure 3. Diagnostic performance and clinical utility of different models. (A) (B) Calibration curves of the clinical model, DL model, and DLN model in the training and external validation cohorts. (C) Decision curve analysis for all cohorts. DL, deep learning; DLN, deep learning nomogram

图 3. 不同模型的诊断效能与临床实用性评估。(A) (B) 临床模型、DL 模型及 DLN 模型在训练集与外部验证集中的校准曲线。(C) 所有数据集的决策曲线分析。DL: 深度学习; DLN: 深度学习列线图

的激活强度,用于定位模型重点关注区域;SHAP 图则进一步量化单个像素特征对病理分级预测结果的贡献。在预测正确病例中,高级别肉瘤(患者 A, DLN 评分 = 0.965)在 CAM 图上表现为肿瘤核心强化区的高度红热激活,而低级别肉瘤(患者 C, DLN 评分 = 0.016)则呈弥散的蓝色低响应状态。进一步分析误判病例发现,假阳性患者 B (低级别但 DLN 评分 = 0.893)因病灶内存在局灶性富血管强化区(CAM 呈现局灶红热),且伴有较高的临床分期(T3N1M0),导致模型综合评分偏高;假阴性患者 D (高级别但 DLN 评分 = 0.013)则因病灶体积较小且内部密度相对均匀,缺乏典型的空间异质性特征,导致深度学习模块提取的风险响应不足。

3.5. PFS 风险分层

训练集和外部验证集的中位 PFS 分别为 9.00 个月(IQR: 4.00~14.00)和 20.00 个月(IQR: 8.00~38.00)。以 0.69 为阈值,DLN 可成功将患者分为高风险组和低风险组。Kaplan-Meier 生存曲线表明,在训练集(HR = 2.555, $P=0.013$)和外部验证集(HR = 7.382, $P=0.002$)中,高风险组患者的 PFS 均显著差于低风险组(见图 5)。

为全面评估模型的预后预测价值并避免二分法带来的阈值选择偏倚,本研究将 DLN 评分作为连续变量直接纳入 Cox 比例风险回归模型。单因素及多因素分析结果显示,连续变量 DLN 评分在训练集与外部验证集中均是患者 PFS 的独立预后危险因素(见表 5)。

4. 讨论

本研究构建并外部验证了基于三期 CECT 的 DLN 模型,实现了术前无创评估 RPS 病理分级与 PFS 预后分层。研究发现,单纯临床模型在外部验证集中效能衰减严重(AUC 0.530),表明依赖宏观离散临床指标难以克服跨中心入组偏倚及捕捉深层生物学异质性。而 DLN 通过融合三期动态强化高维特征与注意力机制,成功打破了跨中心泛化瓶颈(验证集 AUC 0.882),验证了无创深度学习特征对传统临床评估的强

大增益价值。

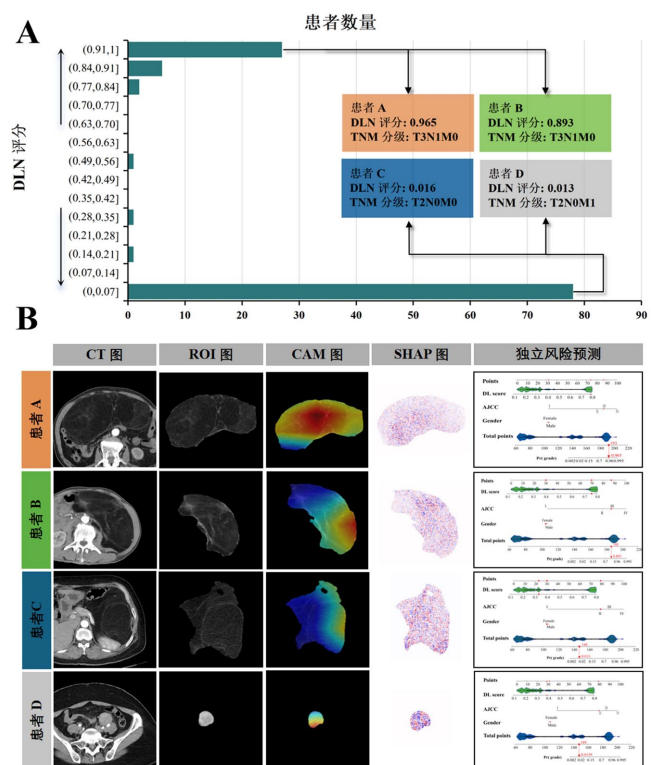


Figure 4. Distribution of DLN scores and visualization of representative cases. (A) Distribution of DLN scores in the training cohort and the location of four representative cases (Patients A~D) selected based on the score histogram, along with their clinical TNM stages. (B) Original CT images, region of interest (ROI) maps, CAM maps, SHAP maps, and individualized nomogram prediction score displays for the four representative cases. Tumor staging was defined according to the eighth edition of the American Joint Committee on Cancer (AJCC) TNM classification system. DLN, deep learning nomogram

图 4. DLN 评分的分布特征及典型病例可视化。(A) 基于训练集 DLN 评分分布直方图挑选的 4 例代表性病例(患者 A~D)的分布位置及临床 TNM 分期。(B) 4 例代表性病例的原始 CT 图、感兴趣区图、CAM 图、SHAP 图及个体化预测列线图评分展示。肿瘤分期依据美国癌症联合委员会(AJCC)第 8 版 TNM 分期系统定义。DLN: 深度学习列线图

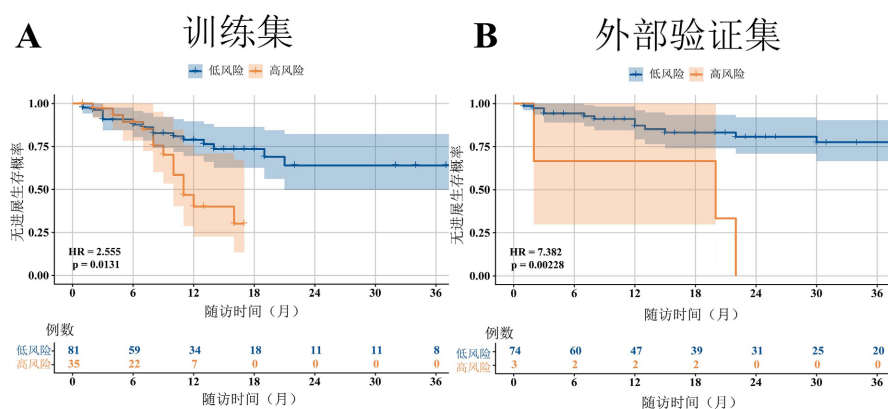


Figure 5. Kaplan-Meier survival curves for progression-free survival (PFS) stratified by DLN risk categories. Survival differences between the high- and low-risk groups defined by DLN are presented in the training and external validation cohorts, respectively. PFS, progression-free survival; DLN, deep learning nomogram

图 5. 基于 DLN 高低风险分组的无进展生存期(PFS) Kaplan-Meier 生存曲线。分别展示了训练集与外部验证集中由 DLN 定义的高低风险组的生存差异。PFS: 无进展生存期; DLN: 深度学习列线图

Table 5. Univariable and multivariable Cox regression analyses for progression-free survival in patients with RPS
表 5. RPS 无进展生存期的单因素与多因素 Cox 回归分析

队列		FNCLCC	AJCC 分期	临床模型	DLN 评分
训练队列 (n = 116)	单因素	HR (95% CI)	1.970 (1.187~3.270)	1.440 (0.996~2.082)	23.255 (7.310~73.978)
					2.870 (1.260~6.541)
	多因素	p	0.009	0.052	<0.001
					0.012
外部验证队列 (n = 77)	单因素	HR (95% CI)	1.142 (0.628~2.077)	23.376 (6.622~82.514)	2.545 (1.077~6.015)
	多因素	p	0.662	<0.001	0.033
外部验证队列 (n = 77)	单因素	HR (95% CI)	5.109 (2.210~11.811)	1.671 (0.951~2.936)	2.267 (0.499~10.306)
					136.387 (18.499~1005.515)
	多因素	p	<0.001	0.074	0.289
					<0.001
外部验证队列 (n = 77)	单因素	HR (95% CI)	4.908 (1.871~12.870)		64.468 (9.522~436.488)
	多因素	p	0.001		<0.001

注：HR = 风险比；CI = 置信区间；FNCLCC = 法国癌症中心联合会；AJCC = 美国癌症联合委员会；DLN 模型 = 深度学习列线图；HR = 风险比。

对于原发性 RPS 患者，精准的术前病理分级对于制定合理的治疗方案至关重要[20]。近年来，越来越多研究证实，基于先进图像分析技术获得的影像生物标志物可用于肿瘤分级及预后预测[10] [11] [21] [22]。Arthur 等[8]近期提出了一种基于 CT 影像组学的预测方法，在 RPS 组织学亚型及肿瘤分级预测中表现出较好的效能。然而，该研究模型仅基于单一静脉期 CECT 图像构建，并依赖手工影像组学特征，其特征表达能力可能受到限制，难以充分捕获 RPS 复杂肿瘤异质性所包含的动态血流灌注信息及高维潜在影像模式。针对上述不足，本研究构建了一种基于深度学习的多模态融合预测模型，该模型充分利用深度学习自动提取多尺度影像特征的优势[23] [24]，并结合三期 CECT 提供的互补性动态强化信息，从而提高了模型在真实临床场景中的适用性。

此外，CAM 和 SHAP 可视化显示模型高权重激活区集中于肿瘤核心，对应于高级别肉瘤的高细胞密度、微血管新生及局灶坏死等核心病理特征[25]——这些均为 FNCLCC 分级系统中定义 2/3 级(高级别)肉瘤的核心病理学特征[26]。同时，该结果与影像科医师基于肿瘤内部异质性进行风险判断的诊断逻辑相一致，进一步验证了模型预测依据的合理性。

深度学习模型的误判模式对于理解其临床适用边界具有重要价值。本研究中，假阳性患者的误判提示，部分低级别肉瘤若伴有局灶性炎性充血或富血管纤维增生，可在 CECT 上形成类似于高级别肉瘤的强化异质性伪影，易引发模型的错误高估。而当高级别肉瘤处于早期或体积较小时，由于缺少大范围坏死及显著的纹理异质性，深度学习网络难以捕捉到足够的恶性征象。这表明在未来临床应用中，对于小体积病灶或伴有复杂继发改变的病例，仍需结合多序列 MRI 或穿刺病理进行综合研判。

既往研究显示，治疗前临床参数及影像组学模型可用于预测 RPS 患者生存结局。Gronchi 等[27]建立了公开的 Sarculator 预后预测工具，通过临床病理参数预测接受手术切除 RPS 患者的总体生存期和无病生存期。近期，Yaru 等[28]进一步证实，基于术前 CT 图像构建的深度学习模型能够预测原发性 RPS 患者局部复发无病生存，有助于患者风险分层。本研究结果进一步表明，DLN 评分不仅能够用于病理分级预测，同时可作为 RPS 患者 PFS 预测的独立相关因素。这表明 DLN 可以在术前作为一种强有力的补充

工具,辅助外科及肿瘤科医师识别高风险复发患者,从而调整随访频率或制定更为积极的综合治疗方案。

本研究尚存在局限性。第一,尽管包含外部验证,但样本量仍相对有限,未来需在大规模、多中心的前瞻性队列中进一步验证。第二,仅采用最大异质性单层(2D)特征以消除跨中心层厚伪影,损失了Z轴立体异质性。第三,受限于回顾性研究条件,本研究未将CT可解释性热图与离体大体病理切片进行空间对齐校验,未来前瞻性研究可结合空间转录组学或空间定位病理标本,进一步验证深度特征的生物学基础。

综上所述,基于三期增强CT与临床特征构建的DLN模型是一种无创且具备可解释性的辅助工具。该模型在预测RPS病理分级和评估PFS预后方面具有重要的临床应用价值,有助于推动腹膜后肉瘤个体化诊疗的发展。

致 谢

感谢所有参与本研究的放射医师、外科医师和工程师。

利益冲突

所有作者声明无利益冲突。

声 明

本研究获青岛大学附属医院(No. QYFY WZLL 50088)与宁夏医科大学总医院(No. KYLL-2026-0144)伦理委员会批准,并免除患者知情同意。

参考文献

- [1] Hayes, A.J., Nixon, I.F., Strauss, D.C., Seddon, B.M., Desai, A., Benson, C., *et al.* (2025) UK Guidelines for the Management of Soft Tissue Sarcomas. *British Journal of Cancer*, **132**, 11-31. <https://doi.org/10.1038/s41416-024-02674-y>
- [2] Stojadinovic, A., Yeh, A. and Brennan, M.F. (2002) Completely Resected Recurrent Soft Tissue Sarcoma: Primary Anatomic Site Governs Outcomes. *Journal of the American College of Surgeons*, **194**, 436-447. [https://doi.org/10.1016/s1072-7515\(02\)01120-1](https://doi.org/10.1016/s1072-7515(02)01120-1)
- [3] Thomas, D.M., O'Sullivan, B. and Gronchi, A. (2009) Current Concepts and Future Perspectives in Retroperitoneal Soft-Tissue Sarcoma Management. *Expert Review of Anticancer Therapy*, **9**, 1145-1157. <https://doi.org/10.1586/era.09.77>
- [4] Messiou, C., Moskovic, E., Vanel, D., Morosi, C., Benchimol, R., Strauss, D., *et al.* (2017) Primary Retroperitoneal Soft Tissue Sarcoma: Imaging Appearances, Pitfalls and Diagnostic Algorithm. *European Journal of Surgical Oncology (EJSO)*, **43**, 1191-1198. <https://doi.org/10.1016/j.ejso.2016.10.032>
- [5] Gronchi, A., Strauss, D.C., Miceli, R., Bonvalot, S., Swallow, C.J., Hohenberger, P., *et al.* (2016) Variability in Patterns of Recurrence after Resection of Primary Retroperitoneal Sarcoma (RPS). *Annals of Surgery*, **263**, 1002-1009. <https://doi.org/10.1097/sla.0000000000001447>
- [6] Pavlidis, E.T. and Pavlidis, T.E. (2023) New Trends in the Surgical Management of Soft Tissue Sarcoma: The Role of Preoperative Biopsy. *World Journal of Clinical Oncology*, **14**, 89-98. <https://doi.org/10.5306/wjco.v14.i2.89>
- [7] Schneider, N., Strauss, D.C., Smith, M.J., Miah, A.B., Zaidi, S., Benson, C., *et al.* (2017) The Adequacy of Core Biopsy in the Assessment of Smooth Muscle Neoplasms of Soft Tissues: Implications for Treatment and Prognosis. *American Journal of Surgical Pathology*, **41**, 923-931. <https://doi.org/10.1097/pas.0000000000000867>
- [8] Arthur, A., Orton, M.R., Emsley, R., Vit, S., Kelly-Morland, C., Strauss, D., *et al.* (2023) A CT-Based Radiomics Classification Model for the Prediction of Histological Type and Tumour Grade in Retroperitoneal Sarcoma (RADSARC-R): A Retrospective Multicohort Analysis. *The Lancet Oncology*, **24**, 1277-1286. [https://doi.org/10.1016/s1470-2045\(23\)00462-x](https://doi.org/10.1016/s1470-2045(23)00462-x)
- [9] Tirotta, F., Schut, A.W., Wemmers, D., Klein, S., Visser, J.J., Hanff, D.F., *et al.* (2025) Evaluation of Diagnostic Accuracy of Preoperative CT-Based Radiomics in Primary Retroperitoneal Sarcoma. *Annals of Surgical Oncology*, **32**, 7799-7807. <https://doi.org/10.1245/s10434-025-18040-y>
- [10] Yan, R., Hao, D., Li, J., Liu, J., Hou, F., Chen, H., *et al.* (2021) Magnetic Resonance Imaging-Based Radiomics Nomogram for Prediction of the Histopathological Grade of Soft Tissue Sarcomas: A Two-Center Study. *Journal of Magnetic*

- Resonance Imaging*, **53**, 1683-1696. <https://doi.org/10.1002/jmri.27532>
- [11] Guo, J., Li, Y., Guo, H., Hao, D., Xu, J., Huang, C., *et al.* (2025) Parallel CNN-Deep Learning Clinical-Imaging Signature for Assessing Pathologic Grade and Prognosis of Soft Tissue Sarcoma Patients. *Journal of Magnetic Resonance Imaging*, **61**, 807-819. <https://doi.org/10.1002/jmri.29474>
 - [12] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 6000-6010.
 - [13] Zhang, Z., Yu, L., Liang, X., Zhao, W. and Xing, L. (2021) TransCT: Dual-Path Transformer for Low Dose Computed Tomography. In: de Bruijne, M., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention*, Springer International Publishing, 55-64. https://doi.org/10.1007/978-3-030-87231-1_6
 - [14] Cai, L., Bai, R., Cao, Q., Sun, W., Wang, F., Liu, X., *et al.* (2025) A Non-Invasive MRI-Based Multimodal Fusion Deep Learning Model (MF-DLM) for Predicting Overall Survival in Bladder Cancer: A Multicentre Retrospective Study. *eClinicalMedicine*, **90**, Article 103640. <https://doi.org/10.1016/j.eclinm.2025.103640>
 - [15] Lundberg, S.M. and Lee, S.-I. (2017) A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4768-4777.
 - [16] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 618-626. <https://doi.org/10.1109/iccv.2017.74>
 - [17] Collins, G.S., Reitsma, J.B., Altman, D.G. and Moons, K.G.M. (2015) Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *BMJ*, **350**, g7594. <https://doi.org/10.1136/bmj.g7594>
 - [18] Trojani, M., Contesso, G., Coindre, J.M., Rouesse, J., Bui, N.B., De Mascarel, A., *et al.* (1984) Soft-Tissue Sarcomas of Adults; Study of Pathological Prognostic Variables and Definition of a Histopathological Grading System. *International Journal of Cancer*, **33**, 37-42. <https://doi.org/10.1002/ijc.2910330108>
 - [19] Li, C., He, Z., Lv, F., Liu, Y., Hu, Y., Zhang, J., *et al.* (2023) An Interpretable MRI-Based Radiomics Model Predicting the Prognosis of High-Intensity Focused Ultrasound Ablation of Uterine Fibroids. *Insights into Imaging*, **14**, Article No. 129. <https://doi.org/10.1186/s13244-023-01445-2>
 - [20] Swallow, C.J., Strauss, D.C., Bonvalot, S., Rutkowski, P., Desai, A., Gladdy, R.A., *et al.* (2021) Management of Primary Retroperitoneal Sarcoma (RPS) in the Adult: An Updated Consensus Approach from the Transatlantic Australasian RPS Working Group. *Annals of Surgical Oncology*, **28**, 7873-7888. <https://doi.org/10.1245/s10434-021-09654-z>
 - [21] Navarro, F., Dapper, H., Asadpour, R., Knebel, C., Spraker, M.B., Schwarze, V., *et al.* (2021) Development and External Validation of Deep-Learning-Based Tumor Grading Models in Soft-Tissue Sarcoma Patients Using MR Imaging. *Cancers*, **13**, Article 2866. <https://doi.org/10.3390/cancers13122866>
 - [22] Gitto, S., Cuocolo, R., Huisman, M., Messina, C., Albano, D., Omoumi, P., *et al.* (2024) CT and MRI Radiomics of Bone and Soft-Tissue Sarcomas: An Updated Systematic Review of Reproducibility and Validation Strategies. *Insights into Imaging*, **15**, Article No. 54. <https://doi.org/10.1186/s13244-024-01614-x>
 - [23] Byeon, Y., Park, Y.W., Lee, S., Park, D., Shin, H., Han, K., *et al.* (2025) Interpretable Multimodal Transformer for Prediction of Molecular Subtypes and Grades in Adult-Type Diffuse Gliomas. *npj Digital Medicine*, **8**, Article No. 140. <https://doi.org/10.1038/s41746-025-01530-4>
 - [24] Guo, R., Qu, X., Tian, S., Li, Z., Wang, X., Sun, Z., *et al.* (2026) Leveraging Vision Transformer for Histological Grade Prediction in Laryngeal and Hypopharyngeal Squamous Cell Carcinoma: A Large-Scale Multicenter Study. *Neuroradiology*, **68**, 115-125. <https://doi.org/10.1007/s00234-025-03876-8>
 - [25] Coindre, J. (2006) Grading of Soft Tissue Sarcomas: Review and Update. *Archives of Pathology & Laboratory Medicine*, **130**, 1448-1453. <https://doi.org/10.5858/2006-130-1448-gostsr>
 - [26] Arthur, A., Johnston, E.W., Winfield, J.M., Blackledge, M.D., Jones, R.L., Huang, P.H., *et al.* (2022) Virtual Biopsy in Soft Tissue Sarcoma. How Close Are We? *Frontiers in Oncology*, **12**, Article ID: 892620. <https://doi.org/10.3389/fonc.2022.892620>
 - [27] Gronchi, A., Miceli, R., Shurell, E., Eilber, F.C., Eilber, F.R., Anaya, D.A., *et al.* (2013) Outcome Prediction in Primary Resected Retroperitoneal Soft Tissue Sarcoma: Histology-Specific Overall Survival and Disease-Free Survival Nomograms Built on Major Sarcoma Center Data Sets. *Journal of Clinical Oncology*, **31**, 1649-1655. <https://doi.org/10.1200/jco.2012.44.3747>
 - [28] Ren, Y., Xue, Z., Liang, T., Sun, Y., Gao, J., Liu, K., *et al.* (2026) A CT-Based Deep Learning Model to Predict Local Recurrence-Free Survival in Primary Retroperitoneal Sarcoma. *Frontiers in Medicine*, **12**, Article ID: 1725377. <https://doi.org/10.3389/fmed.2025.1725377>