

多模态 MRI 影像组学联合模型在骨质疏松症中的诊断价值

王文斌, 钟鉴, 曾威, 李春平, 刘英, 易雪冰

【摘要】 目的:探讨多模态 MRI 影像组学联合临床特征构建的机器学习模型对骨质疏松症(OP)的诊断效能。方法:回顾性纳入双中心 1095 例慢性腰痛患者,将来自中心 A 的 961 例患者按 8:2 的比例随机分为训练集($n=768$)与内部测试集($n=193$),来自中心 B 的患者($n=134$)作为外部验证集。采用 DeepLabv3+MobileNet 模型分割 L_1-L_3 椎体,计算 Dice 系数以评估模型分割的准确性。分别自 T_1WI 、 T_2WI 及 $FS-T_2WI$ 序列上提取 3 个椎体的影像组学特征,经 t 检验及 LASSO 回归进行特征筛选后,采用逻辑回归方法构建影像组学模型,并联合临床危险因素构建多模态联合模型。采用 ROC 曲线和决策曲线分析(DCA)评估各模型的效能,采用 Delong 检验比较模型间 AUC 的差异。采用沙普利加性解释(SHAP)图对联合模型的预测决策进行可解释性分析。结果:椎体分割的 Dice 系数达 97.5%。联合模型在内部测试集和外部验证集中的 AUC 分别为 0.807 和 0.741,显著高于影像组学模型(AUC 分别为 0.723 和 0.666)和临床模型(AUC 分别为 0.740 和 0.708),差异均有统计学意义($P<0.05$)。SHAP 分析显示在联合模型中影像组学特征的贡献度为 57%,其中源自 T_2WI 序列的影像组学特征的贡献度(33.6%)最高。结论:多模态联合模型有效整合 MRI 影像组学和临床特征,可为 OP 的影像标志物选择提供新依据;多序列中以 T_2WI 序列对 OP 识别的贡献度最显著。

【关键词】 腰椎;骨质疏松;磁共振成像;影像组学

【中图分类号】 R445.2;R681.5 **【文献标志码】** A **【文章编号】** 1000-0313(2026)04-0453-07

DOI:10.13609/j.cnki.1000-0313.2026.04.016

开放科学(资源服务)标识码(OSID):



Diagnostic value of multimodal MRI radiomics combined model in osteoporosis WANG Wen-bin, ZHONG Jian, ZENG Wei, et al. Department of Radiology, Sichuan Province Orthopedic Hospital, Chengdu 610041, China

【Abstract】 Objective: To investigate the diagnostic performance of a machine learning model constructed by combining multi-modal MRI radiomics with clinical features for osteoporosis (OP). **Methods:** A total of 1095 patients with chronic lower back pain from two medical centers were retrospectively included. Patients from Center A ($n=961$) were randomly divided into a training set ($n=768$) and an internal test set ($n=193$) at an 8:2 ratio, while patients from Center B ($n=134$) served as an external validation set. The DeepLabv3+MobileNet model was employed to segment L_1-L_3 vertebrae, and the Dice coefficient was calculated to evaluate segmentation accuracy. Radiomic features were extracted from T_1WI , T_2WI , and $FS-T_2WI$ sequences for the three vertebrae. Upon the completion of feature selection via t -test and LASSO regression, a radiomics model was constructed using logistic regression, and a multi-modal combined model was developed by integrating these radiomic signatures with clinical risk factors. The diagnostic performance of each model was evaluated using receiver operating characteristic (ROC) curves and decision curve analysis (DCA), while differences in AUC were compared using Delong test. Shapley Additive explanation plots (SHAP) were utilized to interpret the prediction decisions of the combined model. **Results:** The Dice coefficient for vertebral segmentation reached 97.5%. The vertebral segmentation achieved a Dice coefficient of 97.5%. The combined model achieved AUCs of 0.807 and 0.741 in the internal testing set and external validation set, respectively, which were significantly higher than those of the radiomics model (AUCs: 0.723 and 0.666) and the clinical model (AUCs: 0.740 and 0.708) (all $P<0.05$). SHAP analysis revealed that in the combined model, radiomics features contributed 57% to the prediction, with features derived from the T_2WI sequence showing the highest contribution (33.6%). **Conclusion:** The multi-modal combined model effectively integrates MRI radiomics and clinical features, offering a new basis for the selection of imaging biomarkers of OP. Among the multi-sequence features, those extracted from T_2WI sequence contribute most significantly to OP identification.

【Key words】 Lumbar vertebrae; Osteoporosis; Magnetic resonance imaging; Radiomics

骨质疏松症(osteoporosis, OP)是以骨密度(bone

mineral density, BMD)降低和骨微结构退变为特征的全身性骨骼疾病,可显著增加骨折风险,导致高致残率和死亡率^[1]。尽管双能 X 线吸收法(dual-energy X-ray absorptiometry, DXA)是临床诊断金标准^[2],但其为二维 BMD 测量方法,易受骨质增生、血管钙化等因素的干扰。定量 CT(quantitative computed tomo-

作者单位:610041 四川成都,四川省骨科医院医学影像科(王文斌、钟鉴、刘英、易雪冰);412012 湖南株洲,湖南中医药大学高等专科学校(曾威);637000 四川南充,川北医学院附属医院医学影像科(李春平)

作者简介:王文斌(1990-),男,湖南常德人,硕士,主治医师,主要从事骨骼肌肉影像学诊断工作。

通讯作者:易雪冰, E-mail:781630855@qq.com

graphy, QCT) 可测量 3D 容积 ROI 的 BMD 值, 因此能提供更精准的信息^[3]。但是 QCT 无法捕捉早期骨髓微结构变化的相关信息, 而骨髓微结构病变通常早于 BMD 降低^[4], 亟需新型影像标志物实现骨质疏松症临床前期的早期诊断。

影像组学可量化 MRI 图像上的纹理特征, 为骨微结构的评估提供了新视角^[5]。但现有的研究多局限于 CT 或 MRI 单模态影像组学分析, 且主要采用的是人工方法进行 ROI 的勾画, 存在效率低和可重复性差等问题^[5-8]。本研究主要有两方面的研究内容: ①开发基于深度学习的 L₁-L₃ 椎体自动分割模型, 实现影像组学特征标准化提取; ②构建融合多模态 MRI 影像组学和临床特征的可解释机器学习模型, 以提升 OP 的诊断准确性。

材料与方 法

1. 一般临床资料

回顾性纳入 2023 年 1 月—2025 年 1 月因慢性腰痛于四川省骨科医院天府院区 (中心 A) 或武侯院区 (中心 B) 就诊的患者。纳入标准: ①治疗前两周内完成腰椎 MRI 和 QCT 检查; ②电子病历中临床信息完整。排除标准: ①同时诊断为 OP 与非骨质疏松 (non-osteoporosis, NOP); ②MRI 及 QCT 检查前接受过抗 OP 药物治疗; ③MRI 图像质量差, 由具有 20 年诊断经验的资深骨肌影像医师评估确认; ④既往有腰椎骨折或内固定手术史, 或伴有恶性肿瘤、血液系统疾病、类风湿性关节炎和强直性脊柱炎等疾病。

本研究采用分层随机抽样, 将来自中心 A 的数据集 (n=961) 按 8:2 比例划分为训练集 (n=768) 与内部测试集 (n=193), 将中心 B 的数据作为外部验证集 (n=134)。搜集每例患者的临床资料, 主要包括年龄、性别、BMI 及吸烟、饮酒及糖尿病病史。

本研究获得四川省骨科医院伦理委员会批准 (KY2024-052-01), 遵循《赫尔辛基宣言》^[9], 回顾性研究故免除患者签订书面知情同意书。

2. QCT 扫描方法和 OP 诊断标准

QCT 扫描使用 Siemens Somatom CT 机, 并使用美国 Mindways Software 公司 QCT-Pro 软件进行

BMD 分析。扫描参数: 扫描范围自 Th₁₂ 椎体上缘至 L₃ 椎体下缘水平, 120 kV, 智能管电流调节, 扫描层厚 5 mm, 重建层厚 1.0 mm。BMD 诊断标准: BMD < 80 mg/cm³ 为骨质疏松; BMD = 80~120 mg/cm³ 为骨量减少; BMD > 120 mg/cm³ 为正常骨量^[2]。本研究将 BMD < 80 mg/cm³ 定义为 OP 组, BMD ≥ 80 mg/cm³ 为无 OP 组。

3. MRI 检查方法

使用 GE Signa Architect 3.0T (中心 A) 或 GE Signa Explorer 1.5T (中心 B) 磁共振仪行腰椎矢状面 MRI 扫描。患者取仰卧位, 扫描参数详见表 1。

4. 影像组学特征提取及特征筛选

图像预处理: 首先从医院 PACS 系统导出 DICOM 格式的原始 MRI 图像, 并转换为 NIFTI (.nii.gz) 格式。随后对 T₁WI、T₂WI 及 FS-T₂WI 图像依次进行以下标准化预处理: ①采用 N4 偏置场校正算法消除低频强度不均匀性; ②将图像强度线性归一化至 0~1 区间; ③使用双立方插值法重采样至各向同性体素 (1×1×1 mm³); ④对灰度值进行离散化处理, 设定 bin 宽度为 25。

掩膜生成: 本研究共纳入 1095 例受试者。为构建椎体自动分割模型, 从中随机抽取 150 例, 由两位放射科医师 (工作年限均 ≥ 5 年) 独立完成 L₁-L₃ 椎体感兴趣区域 (ROI) 的手动勾画。医师 A 使用 3D-Slicer 4.11.0 软件包对所有患者矢状面 T₁WI 图像进行逐层手动分割: 沿椎体骨皮质内轮廓勾画, 覆盖从上终板至下终板的全部层面, 以包含整个椎体骨髓区域; 同时尽可能避开椎静脉等非骨组织, 以减少干扰, 该手动勾画过程直接生成对应的二值化掩膜 (mask)。然后由医师 B 独立复审医师 A 勾画的 ROI。若两位医师对椎体边界 (尤其骨皮质内侧轮廓、椎静脉丛或退变区域) 的勾画范围存在分歧, 则提交至第三位资深医师 C (有 20 年以上工作经验) 进行仲裁, 最终掩膜以仲裁结果为准, 以确保标注的一致性。基于此训练集, 采用 DeepLabv3+MobileNet 架构的语义分割模型进行训练^[10], 并对剩余 945 例样本进行自动掩膜推理。

影像组学特征提取: 在预处理后的 T₁WI、T₂WI 及 FS-T₂WI 图像上, 分别结合对应椎体掩膜 (人工或

表 1 两个医疗中心 MRI 扫描序列和参数

扫描参数	中心 A			中心 B		
	T ₁ WI	T ₂ WI	FS-T ₂ WI	T ₁ WI	T ₂ WI	FS-T ₂ WI
TR/ms	726	3614	2682	583	4563	5184
TE/ms	10	123	71	160	123	43
层厚/mm	3.0	3.0	3.0	3.0	3.0	3.0
层间距/mm	0.3	0.3	0.3	0.3	0.3	0.3
视野/mm ²	280×280	280×280	280×280	300×300	300×300	300×300
矩阵	320×224	320×224	320×224	320×224	352×224	228×192

自动),使用 PyRadiomics 模块提取影像组学特征。每序列提取 107 个原始纹理特征及 279 个高斯拉普拉斯算子(Laplacian of Gaussian, LoG)滤波后特征($\sigma=1, 2, 3$),共计 $386 \times 3 = 1158$ 个初始特征。

特征筛选:评估观察者间一致性:随机抽取 20 例样本,采用组内相关系数(intraclass correlation coefficient, ICC)评估医师 A 初版勾画与医师 B 修订后掩膜的影像组学特征值的一致性,保留 $ICC > 0.75$ 的特征;再评估观察者内一致性:间隔 2 个月后,医师 A 对随机抽取的 20 例样本进行二次掩膜勾画,计算其 2 次勾画所提取组学特征的 ICC,保留 $ICC > 0.75$ 的影像组学特征。随后通过 t 检验,保留 $P < 0.05$ 的特征。最后采用最小绝对收缩与选择算子(least absolute shrinkage and selection operator, LASSO)回归分析(10 折交叉验证)进行特征降维,筛选出各序列(T_1 WI、 T_2 WI 及 FS- T_2 WI)的最优组学特征,并计算得到 3 个序列相应的影像组学评分(Radscore)。

5. 模型的构建

本研究依次构建临床模型、单模态影像组学模型及多模态联合模型,均采用 Logistic 回归作为分类器。所有连续变量在建模前进行 z -score 标准化;模型超参数通过贝叶斯优化确定,并采用 5 折交叉验证评估性能,以控制过拟合风险。①临床模型:纳入年龄、性别、BMI、糖尿病、吸烟及饮酒史共 6 个临床变量,采用单、多因素 logistic 回归分析,确定临床独立危险因素,并构建临床模型;②影像组学模型:基于 T_1 WI、 T_2 WI 及 FS- T_2 WI 序列的 Radscore 构建影像组学模型;③多模态联合模型:将上述筛选出的临床独立危险因素与三个 MRI 序列对应的 Radscore 共同作为输入特征,构建整合临床与多模态 MRI 序列的联合预测模型。

6. 联合模型解释

为阐明联合模型的决策机制,我们采用沙普利加性解释(Shapley additive explanations, SHAP)分析法对联合模型进行可视化解释^[11]。

7. 统计分析方法

使用 SPSS26.0、R4.4.1 和 Python 3.9.1 软件进行统计分析。连续变量以中位数(四分位距)表示,组间

比较采用曼-惠特尼 U 检验;分类变量以数量(百分比)表示,组间比较采用卡方检验。分别采用 ROC 曲线、校准曲线和决策曲线分析(decision curve analysis, DCA)评估模型的预测效能、拟合优度和临床净获益。采用点二列相关分析(point-biserial correlation)评估各变量(包括性别、年龄及 T_1 WI、 T_2 WI、FS- T_2 WI 的 Radscore)与 OP 之间的相关性。采用 DeLong 检验比较各模型间 AUC 的差异。以 $P < 0.05$ (双侧)为差异具有统计学意义。

结果

1. 语义分割模型

DeepLabv3+MobileNet 模型在矢状面 T_1 WI 图像中实现了对 L_1-L_3 椎体各扫描层面的分割推理;模型经过 100 个训练周期, Dice 系数值达 97.5%。

2. 临床资料分析和模型构建

本研究共纳入 1095 例患者,3 个数据集中 OP 组与无 OP 组基本临床资料的比较结果详见表 2。在 3 个数据集中,OP 组与无 OP 组之间年龄分布和性别构成的差异具有统计学意义($P < 0.05$);其它临床资料的组间差异无统计学意义($P > 0.05$)。

基于训练集数据,首先对 6 个临床变量进行单因素 Logistic 回归分析,结果见表 3。统计分析结果显示:年龄($OR = 1.09, 95\%CI: 1.07 \sim 1.12, P < 0.01$)和性别($OR = 0.49, 95\%CI: 0.36 \sim 0.65, P < 0.01$)与 OP 显著相关;BMI($P = 0.09$)、吸烟($P = 0.05$)、饮酒($P = 0.09$)处于边缘显著水平,糖尿病无统计学意义($P = 0.17$)。单因素分析结果显示,仅年龄和性别与 OP 显著相关(均 $P < 0.01$),因此将其纳入多因素 Logistic 回归分析。多因素结果显示,仅年龄($OR = 1.10, 95\%CI: 1.06 \sim 1.14, P < 0.001$)和性别($OR = 0.41, 95\%CI: 0.30 \sim 0.56, P < 0.001$)为独立危险因素;男性患 OP 的风险较女性低 59%;年龄每增加 1 岁,患 OP 的风险升高 10%。临床模型由年龄和性别 2 个变量构成,计算公式为: $\text{logit}(P) = -6.178 + (-0.885 \times \text{性别}) + 0.097 \times \text{年龄}$,其中,性别编码为:男性=1,女性=0; P 表示个体罹患 OP 的概率。

表 2 各数据集中 OP 组与无 OP 组基线临床资料的比较

指标	训练集(n=768)				内部测试集(n=193)				外部验证集(n=134)			
	NOP (n=346)	OP (n=422)	统计量	P 值	NOP (n=85)	OP (n=108)	统计量	P 值	NOP (n=57)	OP (n=77)	统计量	P 值
年龄/岁	67.0(10.0)	71.0(8.0)	49957.5 ^a	0.000	66.0(9.0)	70.0(8.0)	3150.0 ^a	0.000	69.0(5.0)	74.0(7.0)	1371.0 ^a	0.000
BMI/kg·m ⁻²	24.5(4.1)	23.9(3.9)	67119.0 ^a	0.054	24.0(4.6)	24.0(4.2)	4378.0 ^a	0.582	24.1(4.3)	23.6(4.8)	2055.0 ^a	0.530
性别/例			22.7 ^b	0.000			20.6 ^b	0.000			7.6 ^b	0.006
男	171(49.4%)	136(32.2%)			53(62.4%)	31(28.7%)			28(49.1%)	19(24.7%)		
女	175(50.6%)	286(67.8%)			32(37.6%)	77(71.3%)			29(50.9%)	58(75.3%)		
吸烟/例	74(21.4%)	67(15.9%)	3.5 ^b	0.062	18(21.2%)	15(13.9%)	1.3 ^b	0.253	9(15.8%)	6(7.8%)	1.4 ^b	0.240
饮酒/例	64(18.5%)	59(14.0%)	2.6 ^b	0.110	18(21.2%)	14(13.0%)	1.8 ^b	0.184	9(15.8%)	8(10.4%)	0.4 ^b	0.505
糖尿病/例	70(20.2%)	69(16.4%)	1.7 ^b	0.195	17(20.0%)	22(20.6%) [*]	0.0 ^b	1.000	12(21.1%)	20(26.0%)	0.2 ^b	0.649

注: * 内部测试集中 1 例患者的糖尿病状态缺失; ^a 为 Z 值; ^b 为 χ^2 值。

表 3 训练集中单因素和多因素 Logistic 回归分析

指标	单因素 Logistic 回归分析			多因素 Logistic 回归分析		
	OR(95%CI)	Z 值	P 值	OR(95%CI)	Z 值	P 值
年龄	1.09(1.07~1.12)	7.170	<0.01	1.10 (1.08~1.13)	7.605	<0.01
性别	0.49(0.36~0.65)	-4.811	<0.01	0.41 (0.30~0.56)	-5.555	<0.01
BMI	0.96(0.92~1.01)	-1.679	0.09	—	—	—
吸烟	0.69(0.48~1.00)	-1.957	0.05	—	—	—
饮酒	0.72(0.49~1.05)	-1.693	0.09	—	—	—
糖尿病	0.77(0.53~1.11)	-1.388	0.17	—	—	—

注:OR 为比值比。

表 4 内部测试集及外部验证集中各模型的诊断效能

模型类型	AUC(95%CI)	敏感度	特异度	符合率	精确度	F1 分数
临床模型						
内部测试集	0.740(0.720~0.761)	63.4%	70.3%	67.4%	66.3%	71.2%
外部验证集	0.708(0.678~0.738)	55.8%	65.9%	61.9%	65.9%	67.9%
影像组学模型						
内部测试集	0.723(0.696~0.750)	59.7%	66.4%	63.7%	73.3%	68.8%
外部验证集	0.666(0.661~0.670)	61.7%	67.8%	65.7%	67.8%	72.0%
联合模型						
内部测试集	0.806(0.741~0.864)	69.1%	69.6%	69.4%	69.6%	74.7%
外部验证集	0.741(0.655~0.823)	60.0%	68.3%	64.9%	68.3%	70.4%

3. 影像组学模型的构建

特征提取后通过 ICC 特征筛选后每个序列得到 287 个特征,包括 54 个原始影像特征及 233 个 LoG 特征;再经 t 检验及 LASSO 回归进行筛选后, T_1WI 、 T_2WI 及 FS- T_2WI 序列分别得到 11、18 及 15 个最佳影像组学特征,并根据 LASSO 回归公式生成相应的影像组学评分。基于 3 个序列的影像组学评分 (T_1WI_Score 、 T_2WI_Score 及 FS- T_2WI_Score),经逻辑回归分析构建影像组学模型。

4. 模型效能评估

在内部测试集和外部验证集中 3 个模型对 OP 的预测效能见表 4 及图 1。进一步比较发现,联合模型在内部测试集和外部验证集中的 AUC 显著高于其它 2 个模型(表 5)。绘制校准曲线来评估各模型预测结

果与观测结果的一致性,各模型在内部测试集及外部验证集中拟合良好(图 2),表明模型的预测结果与实际结果相符;DCA 显示各模型的临床净收益,联合模型在内部测试集及外部验证集中都表现出最大净收益(图 3)。

表 5 各模型间 AUC 的比较

模型比较	Z 值	P 值
内部测试集		
临床模型 vs. 影像组学模型	0.359	0.720
临床模型 vs. 联合模型	-2.053	0.040
影像组学模型 vs. 联合模型	-3.619	<0.001
外部验证集		
临床模型 vs. 影像组学模型	1.148	0.251
临床模型 vs. 联合模型	-2.005	0.045
影像组学模型 vs. 联合模型	-2.404	0.016

在内部测试集中分别按照性别和年龄对联合模型

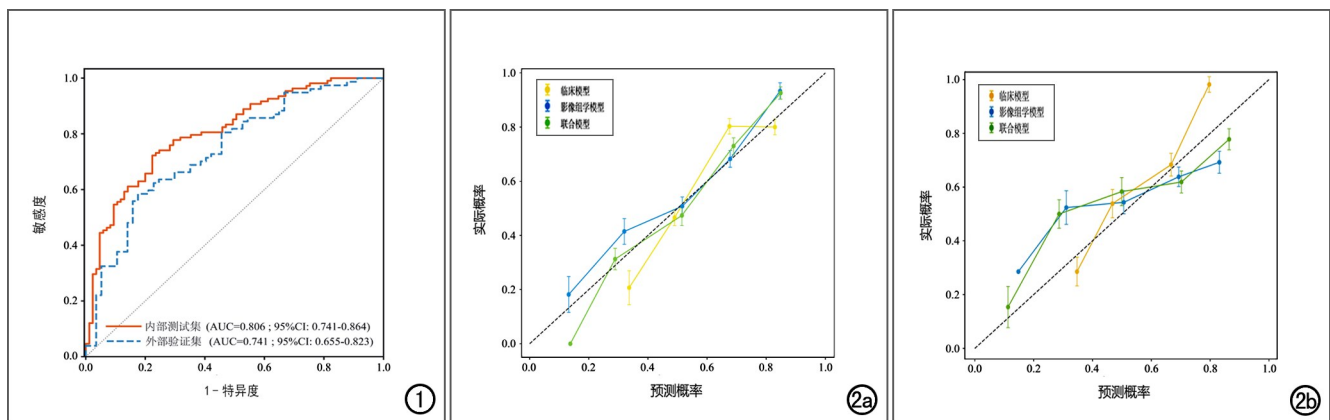


图 1 联合模型在内部测试集和外部验证集中的 ROC 曲线,显示联合模型在两个数据集中均具有良好的诊断效能,内部测试集的 ROC 曲线整体位于外部验证集上方,但 2 个曲线比较接近,表明联合模型在独立外部验证中表现出稳健的泛化能力。

图 2 各模型的校准曲线。a)内部测试集,显示联合模型的校准曲线(绿色线)最贴近理想虚线,表明其预测概率与真实发生率高度一致,拟合优度最佳;b)外部验证集,显示联合模型依然保持了良好的校准度,曲线紧密围绕理想线波动。

的 AUC 进行了亚组分析(图 4), DeLong 检验结果显示亚组间 AUC 的差异均无统计学意义($P>0.05$)。

联合模型在 2 个数据集中预测效能对比的雷达图显示在两个数据集中联合模型的各项效能指标分布均衡且数值较高, 尽管在外部验证集中的各项指标略低于内部测试集, 但整体形态高度重合, 表明联合模型在不同数据集间具有良好的一致性和稳健的泛化能力(图 5)。

5. 联合模型解释

对联合模型进行 SHAP 特征重要性和相关性分析, 结果显示 T_2 WI_Score 具有 33.6% 的贡献度, 是联合模型中最具影响力特征, 性别的贡献度为 22.7%, 年

龄的贡献度为 20.3%, T_1 WI_Score 的贡献度为 16.1%, FS- T_2 WI_Score 的贡献度仅占 7.3%; 性别与 OP 呈负相关, T_1 WI_Score、 T_2 WI_Score 及 FS- T_2 WI_Score 及年龄均与 OP 呈正相关(图 6)。

讨论

本研究通过分析 1095 例 OP 与无 OP 病例的临床和影像资料, 开发了基于 DeepLabv3+ MobileNet 的腰椎 MRI 语义分割模型, 确保了对 L_1-L_3 椎体的分割精准性; 构建了多模态 MRI 序列与临床独立危险因素(性别与年龄)的联合模型, 并在内部测试集、外部验证集及不同亚组(性别、年龄)中进行了验证; 并通过 SHAP 分析可视化各个特征对模型的贡献度。

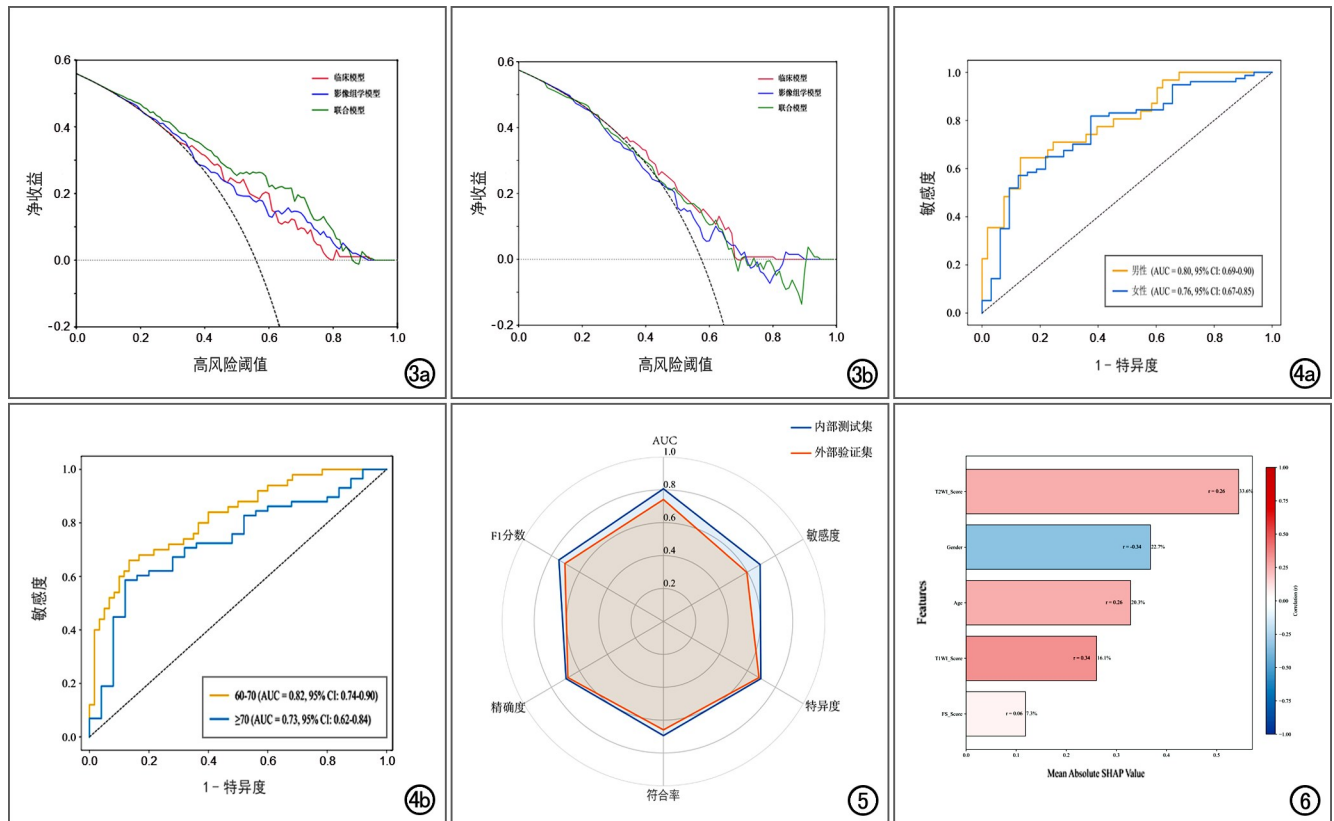


图 3 各模型的决策曲线。a) 内部测试集, 显示在大部分阈值概率范围内(0.1~0.8), 联合模型的决策曲线(绿色线)位于其它模型曲线的上方, 表明其净获益高于单一模型, 具有最佳的临床实用价值; b) 外部验证集, 尽管在高阈值区域(>0.7)受样本量限制联合模型的决策曲线(绿色线)出现一定波动, 但其整体有效阈值范围较单一模型分布更广泛, 表明在广泛的临床决策偏好下, 联合模型仍能提供相对更优或相当的净获益。图 4 不同性别和年龄分组中联合模型的 ROC 曲线。a) 不同性别间联合模型的曲线形态相似且重叠度高, 表明联合模型在男性和女性患者中均具有良好的诊断区分度, 且效能表现一致; b) 不同年龄组的 ROC 曲线, 显示高龄组(≥ 70 岁)的 AUC 略有下降, 但两条曲线均明显高于对角线, 且置信区间存在重叠, 提示联合模型在不同年龄段人群中均保持稳健的预测能力, 未受年龄因素的显著干扰。图 5 内部测试集与外部验证集中联合模型性能比较的雷达图, 显示在两个数据集中联合模型的各项效能指标分布均衡且数值较高, 尽管外部验证集的各项指标略低于内部测试集(橙色区域略小于蓝色区域), 但整体形态高度重合, 表明联合模型在不同数据集间具有良好的一致性和稳健的泛化能力。图 6 联合模型的 SHAP 特征重要性分析与相关性分析融合图, 显示联合模型中以 T_2 WI_Score 的平均绝对 SHAP 值最大, 贡献度占比(33.6%)最大, 是最重要的预测因子, 与 OP 风险呈正相关($r=0.26$), 表明 T_2 WI 序列提取的影像学特征对 OP 的判别能力最高; Gender(性别)是第二大重要特征(贡献度占比 22.7%), 且与 OP 呈负相关($r=-0.34$), 提示女性的患 OP 风险更高, 与临床常识一致; 其它特征的贡献度依次为 Age($r=0.26$)、 T_1 WI_Score($r=0.34$)和 FS- T_2 WI_Score($r=0.06$), 这些特征共同构成了模型的多维预测基础。

1. 语义分割模型

精确分割 L₁—L₃ 椎体是后续基于多模态 MRI 特征进行 OP 预测至关重要的预处理步骤,考虑到本研究涉及超过 1000 例的大样本量,手动分割在时间和人力成本上均不可行,且易引入主观误差。因此,我们开发并应用了基于 DeepLabv3+ 架构的自动化分割流程。此流程的核心是采用 MobileNet 作为主干网络的 DeepLabv3+ 模型。选择 MobileNet 的主要考量是其优异的计算效率和深度可分离卷积^[12],这对于处理本研究中庞大的数据集至关重要,确保了整个预测分析流程在可行的时间框架内完成,同时也符合未来潜在的临床部署对计算资源的要求。模型在验证集中的 Dice 系数非常高(>97.5%),这种高精度的自动化分割为后续从椎体区域稳定一致地提取影像组学特征提供了坚实保障。

2. 联合模型的性能及可解释性分析

本研究中,联合模型对 OP 的鉴别诊断效能最高,其 AUC 在内部测试集和外部验证集分别为 0.806 与 0.741,亚组分析结果表明模型在不同性别与年龄分层间性能稳定($P > 0.05$),表明联合模型具有较好的泛化性及稳健性;然而模型在外部验证集中的效能(AUC=0.741)较内部测试集(AUC=0.806)有所下降,提示影像组学特征在不同样本集间的稳定性可能存在挑战,跨中心设备差异(如场强不同)可能是潜在的影响因素。

SHAP 分析显示 T₂WI 和 T₁WI 为最主要的贡献序列,T₂WI 的高权重可能归因于其对骨基质内水分含量动态变化的高度敏感性,如骨髓水肿减少或脂肪浸润增加等病理生理过程;而 T₁WI 的优势则在于其能够有效量化骨髓内脂肪组织与红骨髓的相对比例,T₂WI 与 T₁WI 的协同作用,为评估骨质微结构提供了多维度的代谢信息,这与既往研究发现的 T₁WI 信号强度与 BMD 具有相关性以及 T₂WI 对骨代谢变化的敏感性高的结论相互印证^[13-14]。而 FS 序列在模型中的贡献度相对较低,其局限性可能源于两方面:其一,在 OP 早期阶段,骨髓水肿等病理特征往往不显著,削弱了 FS-T₂WI 序列的特异性诊断价值;其二,FS-T₂WI 序列对磁场均匀性较为敏感,椎体边缘易产生伪影,可能降低所提取特征的稳定性。尽管如此,FS-T₂WI 序列通过有效抑制黄骨髓信号,在识别椎体微骨折等细微结构改变方面仍具有辅助作用,为模型提供了补充诊断信息。

3. 临床-影像协同作用的生物学基础

性别和年龄的 SHAP 相关性方向与 Logistic 回归结果一致,其机制主要有以下 2 点。①增龄性骨丢失:成骨-破骨细胞活性失衡导致净骨量随年龄累积性

下降^[15];②性激素调控作用:雌激素缺乏加速绝经后女性骨吸收^[16],而雄激素通过促进骨形成赋予男性保护优势^[17]。联合模型通过融合临床变量的代谢调控信息与影像组学的骨微结构量化特征,实现了 OP 鉴别效能的跨维度提升。

本研究存在一定的局限:①仅采用了逻辑回归机器学习模型;②未区分骨量减少与正常骨量;③由于 OP 与基因、家族遗传史、运动状态和激素水平等多种因素相关,未来需通过引入更多的临床特征以提高模型的诊断效能,同时随着深度学习算法的发展,我们预期将进一步优化模型的诊断性能。

综上所述,本研究开发并验证了精准分割腰椎椎体的语义分割模型,为大样本影像组学量化评估 OP 提供了可靠的数据保障;同时通过构建一种标准化融合多模态 MRI 影像组学与临床特征的可解释机器学习模型,为 OP 评估提供了新思路,其中 T₂WI 序列对识别 OP 的贡献最为显著,为影像标志物的选择提供了新依据。

参考文献:

- [1] Gruenewald LD, Koch V, Martin SS, et al. Diagnostic accuracy of quantitative dual-energy CT-based volumetric bone mineral density assessment for the prediction of osteoporosis-associated fractures[J]. *Eur Radiol*, 2022, 32(5): 3076-3084.
- [2] 中华医学会放射学分会骨关节学组,中国医师协会放射医师分会肌骨学组,中华医学会骨科学分会骨质疏松学组,等.骨质疏松的影像学及骨密度诊断专家共识[J]. *中国骨质疏松杂志*, 2020, 26(9): 1249-1256.
- [3] Gao C, Xu Y, Li L, et al. Prevalence of osteoporotic vertebral fracture among community-dwelling elderly in Shanghai[J]. *Chin Med J*, 2019, 132(14): 1749-1751.
- [4] Sarfati M, Chapurlat R, Dufour A B, et al. Short-term risk of fracture is increased by deficits in cortical and trabecular bone microarchitecture independent of DXA BMD and FRAX: Bone Microarchitecture International Consortium (BoMIC) prospective cohorts[J]. *J Bone Miner Res*, 2024, 39(11): 1574-1583.
- [5] 康嗣如,田荣华.基于 T₂WI 序列腰椎磁共振影像组学诊断骨质疏松症[J]. *磁共振成像*, 2023, 14(11): 121-127.
- [6] 刘斋,高婷,许慧,等.基于腰椎 CT 影像组学列线图对骨质疏松的诊断价值[J]. *实用放射学杂志*, 2024, 40(7): 1133-1137.
- [7] 梁晓青,刘潺湲,王婧轶,等. RSNA2019 骨骼肌肉影像学[J]. *放射学实践*, 2020, 35(1): 13-18.
- [8] 程启焯,刘义军,王诗耕,等.基于腹部低 keV 图像的能谱 CT 影像组学模型评估骨质状态[J]. *放射学实践*, 2025, 40(4): 529-534.
- [9] World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects[J]. *JAMA*, 2013, 310(20): 2191-2194.
- [10] Chen LC, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [A/OL]. arXiv, 2018. (2018-08-22). <http://arxiv.org/abs/1802.02611>. DOI:10.48550/arXiv.1802.02611.
- [11] Lundberg SM, Lee SI. A unified approach to interpreting model

- predictions[C].Advances in Neural Information Processing Systems 30,2017;4765-4774.
- [12] Sandler M, Howard A, Zhu M, et al.MobileNetV2: inverted residuals and linear bottlenecks[C/OL].2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, IEEE, 2018; e4510-e4520[2024-04-29].https://ieeexplore.ieee.org/document/8578572/.DOI:10.1109/CVPR.2018.00474.
- [13] Bandirali M, Di Leo G, Papni GDE, et al.A new diagnostic score to detect osteoporosis in patients undergoing lumbar spine MRI[J].Eur Radiol,2015,25(10):2951-2959.
- [14] Chen L, Wu XY, Jin Q, et al.The correlation between osteoporotic vertebrae fracture risk and bone mineral density measured by quantitative computed tomography and dual energy X-ray absorptiometry: a systematic review and meta-analysis[J].Eur Spine J,2023,32(11):3875-3884.
- [15] 严伟,王中汉,刘贺.骨重建失衡导致骨质疏松发生的作用机制及其药物治疗策略[J].中国组织工程研究,2020,24(30):4866-4874.
- [16] 瞿霖,赵梦幻,王冬玉,等.雌激素受体 α 在绝经后骨质疏松中的研究进展及作用机制[J].中国骨质疏松杂志,2024,30(7):1021-1027.
- [17] 陈晓婷,宋迎香.选择性雄激素受体调节剂在男性骨质疏松中的作用及机制[J].中华骨质疏松和骨矿盐疾病杂志,2024,17(2):182-188.

(收稿日期:2025-05-25 修回日期:2025-09-07)

《放射学实践》杂志诚聘英才:青年编委招聘启事

一、基本条件

1.从事医学影像学临床相关工作,有明确的专业研究方向,副高级以上职称或具备博士学位,45周岁以下(优秀者年龄可适度放宽);

2.近5年在SCI期刊发表影响因子2.0分以上学术论文1篇或在《中华放射学杂志》发表2篇亦或在《放射学实践》杂志发表5篇以上论文。申请者本人以第一负责人主持过国家级或省部级科研项目者优先;

3.能熟练应用计算机和网络,通过网络投稿、审稿,熟练使用国内外重要期刊和数据库检索科研信息;具有较好的中英文专业论文写作能力,积极为《放射学实践》杂志撰写论文。熟练掌握统计学知识者优先;

4.治学严谨,有一定行业影响力,有较强的组织能力和社会活动能力,乐于奉献,热心杂志工作。

二、青年编委权益

1.编辑部为获聘者颁发青年编委聘书。聘期满后年龄超过45周岁者,若任期内胜任青年编委工作,即成为《放射学实践》编委候选人;

2.青年编委本人撰写或推荐的稿件,通过绿色通道优先审稿,经正常程序审稿为优质稿件可优先刊用。编辑部根据审稿数量、审稿质量,以及参加编辑部其他工作情况支付报酬。

三、青年编委义务

1.参与本刊审稿工作,审稿能秉持客观、公正态度,能在编辑部提交稿件后2周内认真、严谨完成任务;

2.任期内至少以第一作者或通讯作者身份在本刊发表2篇高水平学术论文;

3.及时向本刊传递行业动态、重要科研课题立项信息。积极反馈读者、作者意见,为杂志发展献计献策;

4.积极参加或者组织有关人员参加本刊学术交流活动。在参加国际、国内交流与合作的同时积极宣传、推介《放射学实践》,为杂志组织优秀稿件,为杂志发现和推荐优秀作者。

四、青年编委聘请办法及聘期

1.符合以上任职条件者,经编辑部审核、常务编委会讨论通过均可担任本刊青年编委;

2.青年编委聘期4年。任期内完成编辑部安排的工作后,可优先被聘请为下一届青年编委;

3.聘期内不能胜任青年编委工作者视为自动退出青年编委行列,编辑部保有解聘权利。

报名方式:将履历表随相关附件一并发送至编辑部邮箱 fsxsjzz@163.com

《放射学实践》编辑部