

基于胸部CT平扫的机器学习影像组学模型筛查骨质疏松的研究

徐遵诚, 蒋天姿, 严梦涵, 许琦, 郝大鹏, 段峰*

青岛大学附属医院放射科, 山东 青岛

收稿日期: 2026年3月28日; 录用日期: 2026年4月22日; 发布日期: 2026年4月28日

摘要

目的: 利用常规胸部CT平扫的T12、L1椎体信息及对应DXA骨密度检查结果, 结合机器学习与影像组学方法建立辅助筛查模型, 实现胸部CT检查人群的骨质疏松机会性快速筛查。方法: 回顾性收集青岛大学附属医院2024年1月至2025年6月行胸部CT平扫及DXA检查(间隔 < 30日)的600例患者, 按7:3随机分为训练集(420例)和内部验证集(180例); 另纳入济宁市第一人民医院同期263例患者作为外部验证集。采用ITK SNAP 3.8软件勾画T12、L1椎体ROI, 通过Pyradiomics模块提取影像组学特征, 经Mann-Whitney U检验、Spearman相关性分析、mRMR、LASSO及10倍交叉验证筛选特征, 构建LR、KNN、SVM等6种机器学习模型; 结合临床独立影像因子构建临床模型及列线图模型, 以AUC、准确性等指标评估模型性能。结果: 每例患者提取1688个影像组学特征, 经筛选最终保留25个最佳特征; 6种模型中LR模型AUC最高, 确定为最终影像组学算法。影像组学模型在训练集、内部验证集、外部验证集的AUC分别为0.880、0.803、0.848, 准确率分别为0.838、0.767、0.791; 临床模型对应AUC分别为0.759、0.746、0.765, 准确率分别为0.757、0.744、0.787; 列线图模型对应AUC分别为0.897、0.849、0.875, 准确率分别为0.790、0.756、0.871。Delong检验显示, 列线图模型及影像组学模型预测效能均显著高于临床模型(均 $P < 0.05$)。结论: 基于胸部CT平扫影像组学特征构建的机器学习模型可有效区分骨质疏松与非骨质疏松患者; 结合临床变量与影像组学特征的列线图模型预测效能最佳, 能更全面精准评估骨质疏松风险, 为临床早期筛查及个性化干预提供可靠量化工具。

关键词

胸部CT, 骨质疏松, 机会性筛查, 机器学习, 影像组学

Study on Screening Osteoporosis Using Machine Learning Radiomics Model Based on Chest CT Plain Scan

Zuncheng Xu, Tianzi Jiang, Menghan Yan, Qi Xu, Dapeng Hao, Feng Duan*

*通讯作者。

文章引用: 徐遵诚, 蒋天姿, 严梦涵, 许琦, 郝大鹏, 段峰. 基于胸部 CT 平扫的机器学习影像组学模型筛查骨质疏松的研究[J]. 临床医学进展, 2026, 16(4): 4963-4979. DOI: 10.12677/acm.2026.1641769

Abstract

Objective: To establish an auxiliary screening model for opportunistic rapid screening of osteoporosis in patients undergoing chest CT examination by utilizing T12 and L1 vertebral information from routine non-contrast chest CT scans along with corresponding DXA bone mineral density results, combined with machine learning and radiomics methods. **Methods:** A retrospective collection of 600 patients who underwent non-contrast chest CT and DXA examination (interval < 30 days) at the Affiliated Hospital of Qingdao University from January 2024 to June 2025 was conducted. Patients were randomly divided into a training set (420 cases) and an internal validation set (180 cases) at a 7:3 ratio. Additionally, 263 patients from Jining First People's Hospital during the same period were included as an external validation set. Regions of interest (ROIs) of T12 and L1 vertebrae were delineated using ITK-SNAP version 3.8 software, and radiomic features were extracted using the Pyradiomics module. Feature selection was performed using Mann-Whitney U test, Spearman correlation analysis, mRMR, LASSO, and 10-fold cross-validation. Six machine learning models (LR, KNN, SVM, etc.) were constructed. Clinical and nomogram models were developed by incorporating independent clinical imaging factors. Model performance was evaluated using AUC, accuracy, and other metrics. **Results:** A total of 1688 radiomic features were extracted per patient, with 25 optimal features retained after selection. Among the six models, the LR model achieved the highest AUC and was selected as the final radiomic algorithm. The AUCs of the radiomic model in the training, internal validation, and external validation sets were 0.880, 0.803, and 0.848, with accuracies of 0.838, 0.767, and 0.791, respectively. The clinical model achieved corresponding AUCs of 0.759, 0.746, and 0.765, with accuracies of 0.757, 0.744, and 0.787. The nomogram model achieved corresponding AUCs of 0.897, 0.849, and 0.875, with accuracies of 0.790, 0.756, and 0.871. Delong tests showed that the predictive performance of both the nomogram model and the radiomic model was significantly superior to that of the clinical model (all $P < 0.05$). **Conclusion:** Machine learning models based on radiomic features from non-contrast chest CT scans can effectively distinguish between osteoporotic and non-osteoporotic patients. The nomogram model combining clinical variables and radiomic features demonstrates the best predictive performance, providing a more comprehensive and accurate assessment of osteoporosis risk and serving as a reliable quantitative tool for early clinical screening and personalized intervention.

Keywords

Chest CT, Osteoporosis, Opportunistic Screening, Machine Learning, Radiomics

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

骨质疏松症(Osteoporosis, OP)是以骨量低下、骨微结构破坏为核心特征的代谢性骨病,可导致脆性增加及脆性骨折风险升高,是威胁中老年人群健康的重要公共卫生问题[1]。我国 50 岁以上人群诊断率不足 20%, 全球漏诊率超 75%, 多数患者因脆性骨折才确诊, 错失最佳干预时机, 因此早期筛查与诊断至

关重要[2]。国际公认的诊断金标准是双能 X 射线吸收法(Dual-Energy X-ray Absorptiometry, DXA), 但由于 DXA 检测覆盖率仅 35%~40%, 临床应用中存在明显局限[3] [4]。这些问题导致骨质疏松症“高患病率、低诊断率”现状, 亟需便捷高效的替代筛查方案。

近年来, 机会性筛查在骨密度(Bone Mineral Density, BMD)评估中得到广泛重视, 由于受检者胸部 CT 平扫图像包含了胸腰椎骨密度的相关信息, 扫描范围通常覆盖上腰椎(尤其 L1 椎体), 为机会性筛查提供天然数据基础[5] [6]。国外多中心研究指出, L1 椎体在胸部 CT 中显示清晰, 伪影干扰率仅 3.2%, 松质骨占比约 75%, 对骨量变化更敏感, 与髌部骨折风险相关性($r = 0.78$)优于其他椎体[7] [8]。传统基于胸部 CT 的骨健康评估依赖亨斯菲尔德单位(Hounsfield Units, HU)测量, 但单一 HU 值仅反映平均密度, 无法捕捉骨微结构特征, 筛查准确性有限(AUC = 0.65~0.72) [9]。影像组学可高通量地从影像中提取大量信息, 将影像信息转化为量化指标(一阶统计特征、二阶纹理特征等)后, 可全面反映组织密度分布、异质性及微观结构, 为机会性筛查提供可靠支撑, 如果能利用影像组学挖掘 L1 椎体影像中的 BMD 信息构建模型, 实现对胸部 CT 平扫人群进行 BMD 机遇性筛查, 将大幅度提高筛查准确率和临床工作效率, 同时使患者受益[10]。

现有研究仍存在不足: 一是针对胸部 CT 的专项研究匮乏, 未明确 L1 为中心椎体的标准化分析方案; 二是多为单中心小样本研究, 泛化能力有待验证[11]-[13]。目前临床和科研中将 BMD 检查作为诊断 OP 的必要检查, 可用于评估骨质疏松严重程度、预测骨折风险和监测骨质疏松治疗效果等, 基于此, 本研究以 DXA 为诊断金标准, 以常规胸部 CT 第十二胸椎和第一腰椎椎体组合为分析靶点, 构建基于胸部 CT 平扫的机器学习影像组学模型并验证其在机会性筛查骨密度的可行性。

2. 材料与方法

2.1. 研究对象

本研究回顾性收集自 2024 年 1 月至 2025 年 6 月期间于青岛大学附属医院同时完成胸部 CT 检查和 DXA 检查的患者的图像信息和 DXA 报告信息, 另外收集外部医疗中心 2024 年 3 月至 2025 年 9 月期间同期完成胸部 CT 检查和 DXA 骨密度检查的患者图像信息和 DXA 报告信息。纳入和排除标准见图 1。

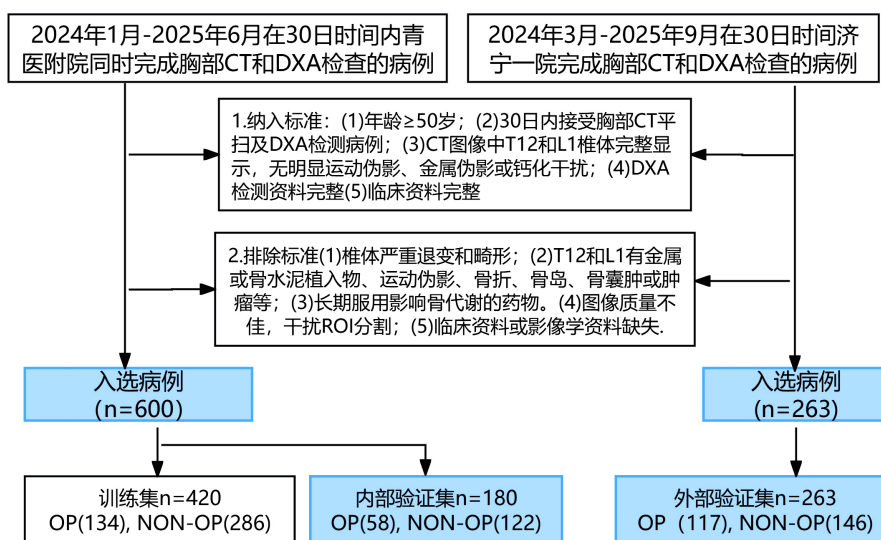


Figure 1. Flowchart of patient inclusion and exclusion criteria

图 1. 患者的纳入标准和排除标准路径流程图

根据纳入及排除标准,青岛大学附属医院最终纳入 600 例同时完成胸部 CT 检查和 DXA 检查的受检者,其中男性 140 例,女性 460 例,平均年龄 61.93 ± 8.9 (50~81)岁,按照 7:3 的比例随机分为训练集 420 例和内部验证集 180 例。按相同标准外院最终纳入 263 例患者作为外部验证集。其中男性 78 例,女性 185 例,平均年龄 61.15 ± 11.62 (50~85)岁。

2.2. CT 图像采集及 DXA 检查报告处理

本研究使用设备包括 GE Revolution、GE Optima、Siemens SOMATOM Force 及联影 uCT 760 CT 等多层螺旋 CT,胸部 CT 扫描参数为:120 kV,35 mA (自动调整),扫描层厚为 0.5 mm,扫描范围从肺炎至 L2 椎体。两家医院 DXA 设备均为 GE prodigy,检查前均严格按设备要求对设备进行每日质控校准工作。将最终纳入患者的胸部 CT 原始平扫图像以 DICOM 格式完整拷贝并导出,以 PACS 登记号命名并汇总。由两位医师记录患者所有病例 DXA 检查结果,存在分歧,则通过协商达成一致。骨质疏松阳性判断标准:以 DXA 检查技术作为准确诊断骨质疏松金标准,当 T 值 ≤ -2.5 时即提示存在骨质疏松症[14],反之则无骨质疏松症。

2.3. 图像预处理及图像分割

在图像分割前,使用 Simple-ITK 软件包(2.0.2 版)对 CT 图像进行预处理,以消除体素尺寸和设备差异的影响,将图像重采样至统一的 $1 \times 1 \times 1 \text{ mm}^3$ 体素大小,使用 25 HU 的固定二进制宽度对体素强度值进行灰度离散化处理。由 1 名具有 10 年以上工作经验的放射科医师在不知患者临床信息的前提下采用 ITK-SNAP 3.8 软件(<http://www.itksnap.org>)进行图像分割。每个椎体分上中下三层独立手动勾画 T12 及 L1 椎体感兴趣区域(ROI),勾画范围为椎体松质骨区域,避开皮质骨、椎弓根及钙化韧带。随机抽取中心 1 的 40 例患者,由另一名放射科医师按照同样方式进行图像分割,并计算两名医师的组内相关系数(ICC),以评估影像组学特征的可靠性和可重复性。ROI 勾画见图 2。

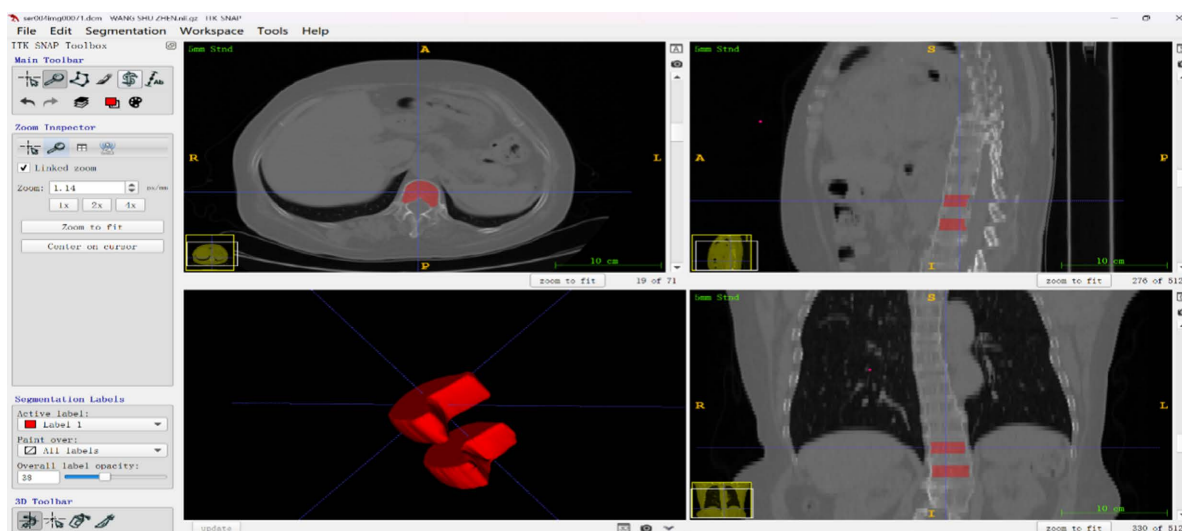


Figure 2. ROI delineation

图 2. ROI 勾画

2.4. 影像组学特征提取

根据图像生物标记标准化倡议(Image Biomarker Standardization Initiative, IBSI)的最新建议,使用

PyRadiomics 软件包(3.1.0 版, <https://pyradiomics.readthedocs.io/>)从 CT 图像 ROI 中提取影像组学特征,此外,还使用了 6 种图像滤波器(wavelet, log, square, gradient, exponential, logarithm)进一步提取 CT 图像的高维特征。最终,从每例患者的 CT 图像中共提取 1688 个特征。为确保提取特征集的可靠性和可重复性,对两位医师提取的特征进行了 ICC 分析,仅保留 ICC > 0.80 的特征用于后续分析。

2.5. 特征筛选及模型构建

使用 Z-score 法对所有影像组学特征归一化处理,剔除具有恒定值的特征。针对训练集中类别不平衡,使用合成少数类过采样技术(SMOTE)方法来增加训练队列中的样本数量并调整数据类别的平衡。对于内部验证队列及外部验证队列,我们使用从训练队列中得出的平均值和方差进行了归一化处理。将最终获得的最佳特征集与相应的系数进行加权,计算出每位患者的影像组学评分(Rad_score)。使用筛选出的特征分别建立相应的逻辑回归(LR)模型、支持向量机(SVM)模型、K 最近邻(KNN)模型、极限梯度提升(XGBoost)模型、随机森林(RF)模型、轻量级梯度提升(LightGBM)模型。使用受试者工作特征(ROC)曲线评估各种模型在训练集及内部、外部验证集中的预测效能。

通过单变量和多变量逻辑回归分析筛选出与骨质疏松状态相关的独立预测因子以构建临床模型。通过 LASSO 筛选出的最佳影像组学特征构建影像组学模型。基于影响骨质疏松状态的独立预测因子及最佳影像组学特征组合构建列线图模型。

2.6. 统计分析方法

采用 IBM SPSS 20.0、Python、R 软件进行统计分析。符合正态分布的连续变量以均值 \pm 标准差($\bar{x} \pm s$)表示,非正态分布的连续变量以中位数(四分位间距)表示,组间比较采用 Mann-Whitney U 检验。分类变量以频数和百分比(n, %)表示,组间比较采用 χ^2 检验或 Fisher 精确检验。通过 ROC 曲线、AUC、准确率、灵敏度、特异性评估模型性能。采用 DeLong 检验比较不同模型的 AUC 值间差异。采用决策曲线分析(Decision Curve Analysis, DCA)评估不同模型的临床实用性。 $P < 0.05$ 为差异具有统计学意义。

3. 结果

3.1. 研究对象基本特征

本研究共纳入 863 例患者,其中训练集 420 例,内部验证集 180 例,外部验证集 263 例。三组患者的年龄、性别等基本特征如表 1 所示。结果显示,三组患者中,骨质疏松患者的年龄均高于非骨质疏松患者(均 $P < 0.05$),女性患者发生骨质疏松的比例明显高于男性(均 $P < 0.05$)。在临床模型中,对训练集患者的临床变量进行单因素、多因素逻辑回归分析,结果如表 2 所示,结果显示,年龄、性别均是预测患者骨质疏松状态的独立预测因子(均 $P < 0.05$),因此选择年龄和性别作为构建临床模型的临床特征。

Table 1. Comparison of patient characteristics among the three groups

表 1. 三组患者基本特征比较

| 特征 | 训练集 (n = 420) | | | 内部验证集 (n = 180) | | | 外部验证集 (n = 263) | | |
|-------|-------------------|--------------------|--------|--------------------|--------------------|--------|--------------------|--------------------|--------|
| | 骨质疏松 (n = 134) | 非骨质疏松 (n = 286) | P 值 | 骨质疏松 (n = 58) | 非骨质疏松 (n = 122) | P 值 | 骨质疏松 (n = 117) | 非骨质疏松 (n = 146) | P 值 |
| 年龄(岁) | 66.09 \pm 8.15 | 59.24 \pm 8.34 | <0.001 | 67.72 \pm 9.20 | 60.92 \pm 8.17 | <0.001 | 68.26 \pm 9.33 | 56.55 \pm 10.64 | <0.001 |

续表

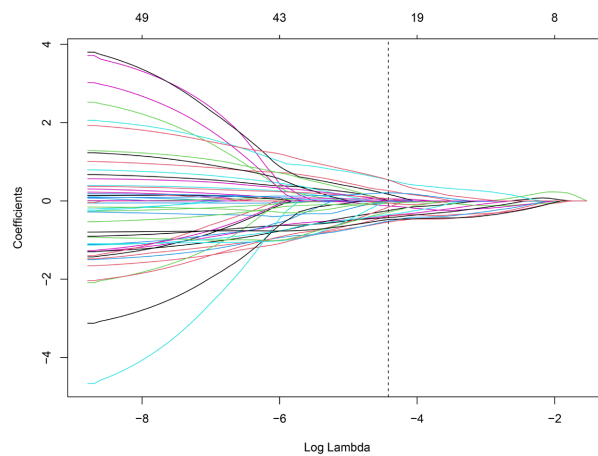
| 性别 | <0.001 | | 0.013 | | <0.001 | |
|----|--------------|--------------|-------------|-------------|--------------|-------------|
| 男 | 14 (10.45%) | 80 (27.97%) | 8 (13.79%) | 38 (31.15%) | 11 (9.40%) | 67 (45.89%) |
| 女 | 120 (89.55%) | 206 (72.03%) | 50 (86.21%) | 84 (68.85%) | 106 (90.60%) | 79 (54.11%) |

Table 2. Uni- and multi-variate logistic regression analysis of clinical variables in the training set**表 2.** 训练集中临床变量单因素及多因素逻辑回归分析结果

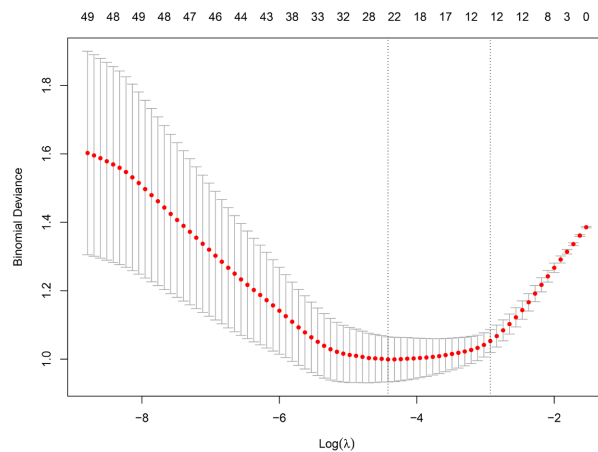
| 特征 | 单因素逻辑回归分析 | | 多因素逻辑回归分析 | |
|-------|---------------------|--------|---------------------|--------|
| | OR (95% CI) | P 值 | OR (95% CI) | P 值 |
| 年龄(岁) | 1.086 (1.054~1.121) | <0.001 | 1.104 (1.069~1.142) | <0.001 |
| 性别 | 0.173 (0.076~0.357) | <0.001 | 0.108 (0.043~0.246) | <0.001 |

3.2. 影像组学特征提取及筛选

每例患者共提取了 1688 个影像组学特征，通过 Mann-Whitney U 检验筛选出 877 个候选特征，经 Spearman 相关性分析去除高度相关特征后筛选出 175 个特征，经 mRMR 对剩余特征进行重要性排序，保留排序前 50 的特征，最后，使用 LASSO 算法和 10 倍交叉验证[λ 为 0.01206, $\ln(\lambda) = -4.4179$]进一步对特征进行降维处理，最终筛选出 25 个最佳特征，LASSO 影像组学特征筛选过程如图 3 所示。



(a)



(b)

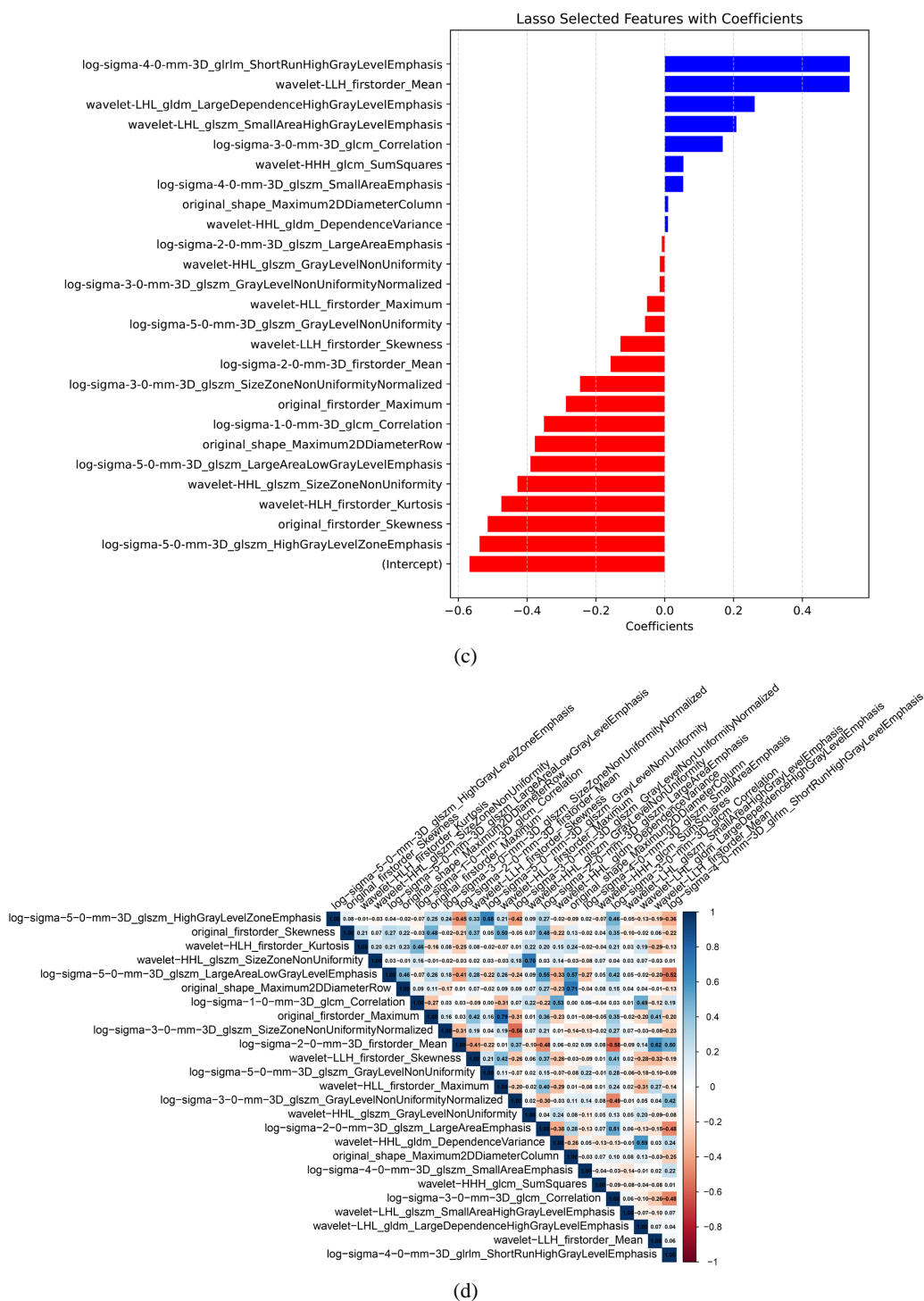


Figure 3. Selection of radiomic features by the least absolute shrinkage and selection operator (LASSO) regression model. (a) Selection of the tuning parameter (λ) via ten-fold cross-validation. The vertical dashed line indicates the optimal λ value. (b) LASSO coefficient curves of the selected radiomic features. The vertical line is drawn at the optimal (λ) value. (c) Distribution of the selected radiomic features and their coefficients from LASSO regression. (d) Visualization of the top 25 radiomic features with the highest selection frequency

图 3. 基于 LASSO 回归模型的影像组学特征筛选。(a) 十折交叉验证法确定调节参数(λ)，垂直虚线表示最优 λ 值；(b) 筛选得到的影像组学特征的 LASSO 系数曲线，垂直线绘制于最优 λ 值处；(c) LASSO 回归筛选得到的影像组学特征及其系数分布；(d) 选择频率最高的前 25 个影像组学特征的可视化结果

3.3. 不同机器学习模型区分骨质疏松患者与非骨质疏松患者的价值

利用筛选出的 25 个影像组学特征分别构建 LR、SVM、KNN、XGBoost、RF、LightGBM 模型，相应的 ROC 曲线见图 4。内部验证集与外部验证集中，不同机器学习模型两两间 AUC 值 Delong 检验结果见表 3 及表 4。结果显示，在内部验证集中，LR 模型的 AUC 值显著高于 KNN、SVM 模型(均 $P < 0.05$)，XGBoost 模型的 AUC 值显著高于 KNN 及 SVM 模型(均 $P < 0.05$)，其余各模型间差异均无统计学意义。在外部验证集中，LR 模型及 XGBoost 的 AUC 值显著高于 LightGBM 模型(均 $P < 0.05$)，其余各模型间差异均无统计学意义。由于内部及外部验证集中，LR 模型的 AUC 最高，因此选定其为最终影像组学采用的机器学习模型。

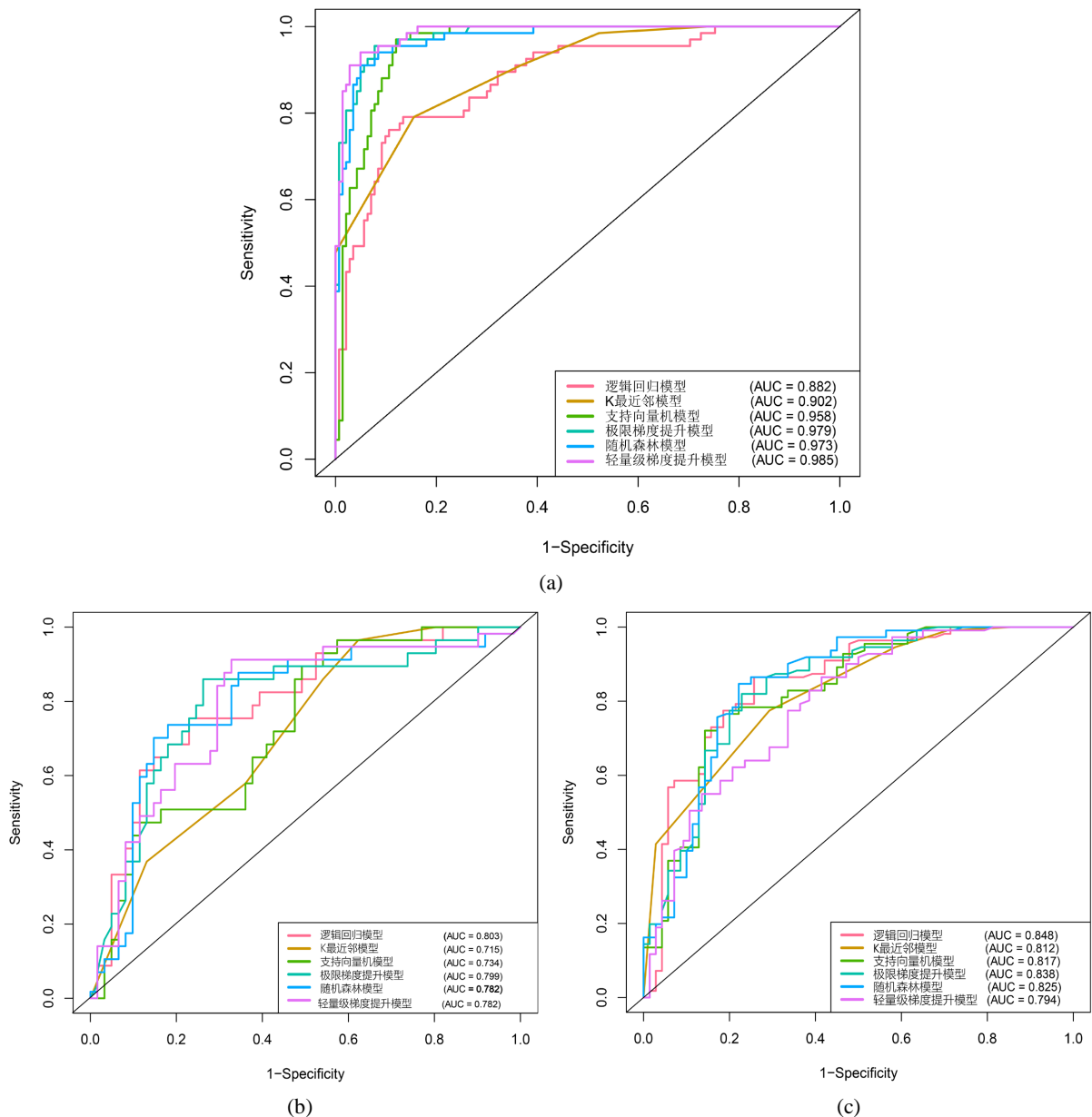


Figure 4. (a) Training set ROC curve, (b) internal validation set ROC curve, (c) external validation set ROC curve
图 4. (a) 训练集 ROC 曲线, (b) 内部验证集 ROC 曲线, (c) 外部验证集 ROC 曲线

Table 3. Value of different machine learning models in distinguishing between osteoporotic and non-osteoporotic patients
表 3. 不同机器学习模型预测骨质疏松患者与非骨质疏松患者的价值

| 模型 | AUC (95% CI) | 敏感性 | 特异性 | 准确率 | PPV | NPV | |
|----------|--------------|---------------------|-------|-------|-------|-------|-------|
| LR | 训练集 | 0.880 (0.845~0.915) | 0.791 | 0.86 | 0.838 | 0.726 | 0.898 |
| | 内部验证集 | 0.803 (0.735~0.871) | 0.759 | 0.77 | 0.767 | 0.611 | 0.87 |
| | 外部验证集 | 0.848 (0.801~0.895) | 0.855 | 0.74 | 0.791 | 0.725 | 0.864 |
| KNN | 训练集 | 0.902 (0.873~0.931) | 0.791 | 0.846 | 0.829 | 0.707 | 0.896 |
| | 内部验证集 | 0.715 (0.642~0.788) | 0.966 | 0.377 | 0.567 | 0.424 | 0.958 |
| | 外部验证集 | 0.812 (0.763~0.860) | 0.752 | 0.705 | 0.726 | 0.672 | 0.78 |
| SVM | 训练集 | 0.956 (0.939~0.974) | 0.97 | 0.874 | 0.905 | 0.783 | 0.984 |
| | 内部验证集 | 0.734 (0.659~0.808) | 0.897 | 0.508 | 0.633 | 0.464 | 0.912 |
| | 外部验证集 | 0.817 (0.766~0.867) | 0.718 | 0.836 | 0.783 | 0.778 | 0.787 |
| XGBoost | 训练集 | 0.979 (0.968~0.990) | 0.955 | 0.916 | 0.929 | 0.842 | 0.978 |
| | 内部验证集 | 0.799 (0.727~0.872) | 0.862 | 0.738 | 0.778 | 0.61 | 0.918 |
| | 外部验证集 | 0.838 (0.790~0.885) | 0.812 | 0.781 | 0.795 | 0.748 | 0.838 |
| RF | 训练集 | 0.973 (0.960~0.986) | 0.91 | 0.944 | 0.933 | 0.884 | 0.957 |
| | 内部验证集 | 0.782 (0.706~0.859) | 0.724 | 0.82 | 0.789 | 0.656 | 0.862 |
| | 外部验证集 | 0.825 (0.776~0.875) | 0.803 | 0.76 | 0.779 | 0.729 | 0.828 |
| LightGBM | 训练集 | 0.985 (0.976~0.994) | 0.94 | 0.951 | 0.948 | 0.9 | 0.971 |
| | 内部验证集 | 0.782 (0.708~0.855) | 0.897 | 0.672 | 0.744 | 0.565 | 0.932 |
| | 外部验证集 | 0.794 (0.741~0.847) | 0.872 | 0.589 | 0.715 | 0.63 | 0.851 |

LR: 逻辑回归模型; KNN: K 最近邻模型; SVM: 支持向量机模型; XGBoost: 极限梯度提升模型; RF: 随机森林模型; LightGBM: 轻量级梯度提升模型。

Table 4. Delong test results for pairwise comparisons of AUC values of different machine learning models in the external validation set
表 4. 外部验证集中不同机器学习模型两两间 AUC 值 Delong 检验结果

| | LR | KNN | SVM | XGBoost | RF | LightGBM |
|----------|--------|-------|-------|---------|-------|----------|
| LR | 1.000 | 0.160 | 0.178 | 0.631 | 0.279 | 0.016* |
| KNN | 0.160 | 1.000 | 0.847 | 0.373 | 0.618 | 0.561 |
| SVM | 0.178 | 0.847 | 1.000 | 0.433 | 0.747 | 0.414 |
| XGBoost | 0.631 | 0.373 | 0.433 | 1.000 | 0.438 | 0.009* |
| RF | 0.279 | 0.618 | 0.747 | 0.438 | 1.000 | 0.120 |
| LightGBM | 0.016* | 0.561 | 0.414 | 0.009* | 0.120 | 1.000 |

LR: 逻辑回归模型; KNN: K 最近邻模型; SVM: 支持向量机模型; XGBoost: 极限梯度提升模型; RF: 随机森林模型; LightGBM: 轻量级梯度提升模型; * $P < 0.05$ 。

3.4. 临床模型、影像组学模型、列线图模型预测性能评估

结合临床变量及放射组学特征的列线图模型的可视化结果如图 5 所示, 在 3 个数据集中, 临床模型、影像组学模型及列线图模型的预测性能详见表 5。图 6(a)~(c)展示了不同预测模型预测性能的 ROC 曲线。训练集中, 列线图模型、影像组学模型、临床模型的 AUC 分别为 0.897 (95% CI: 0.867~0.927)、0.880 (95% CI: 0.845~0.915)、0.759 (95% CI: 0.710~0.808)。列线图模型、影像组学模型的预测效能均显著高于临床模型, 差异具有统计学意义(Delong 检验: Z 值 = 5.994, $P < 0.001$; Z 值 = 4.162, $P < 0.001$)。列线图模型的预测效能高于影像组学模型, 差异有统计学意义(Delong 检验: Z 值 = 2.071, $P = 0.038 < 0.05$)。内部验证集中, 列线图模型、影像组学模型、临床模型的 AUC 分别为 0.849 (95% CI: 0.794~0.903)、0.803 (95% CI: 0.735~0.871)、0.746 (95% CI: 0.667~0.824)。列线图模型的预测效能显著高于临床模型及影像组学模型, 差异具有统计学意义(Delong 检验: Z 值 = 2.971, $P = 0.003 < 0.05$; Z 值 = 2.038, $P = 0.042 < 0.05$)。影像组学模型的预测性能高于临床模型, 但差异无统计学意义(Delong 检验: Z 值 = 1.115, $P = 0.265 > 0.05$)。外部验证集中, 列线图模型、影像组学模型、临床模型的 AUC 分别为 0.875 (95% CI: 0.794~0.903)、0.803 (95% CI: 0.735~0.871)、0.746 (95% CI: 0.667~0.824)。列线图模型的预测效能显著高于临床模型及影像组学模型, 差异具有统计学意义(Delong 检验: Z 值 = 3.939, $P < 0.001$; Z 值 = 1.057, $P = 0.290 > 0.05$)。影像组学模型的预测效能显著高于临床模型, 差异具有统计学意义(Delong 检验: Z 值 = 2.197, $P = 0.028$)。

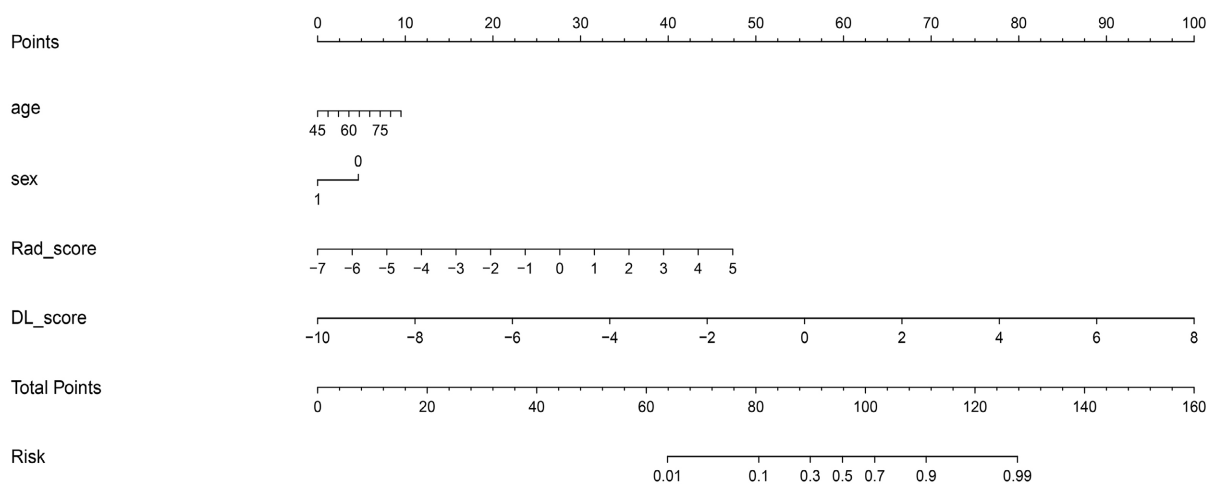


Figure 5. A nomogram for predicting osteoporosis status in patients based on clinical variables and radiomics

图 5. 基于临床变量及影像组学预测患者骨质疏松状态的列线图

Table 5. Value of different prediction models in distinguishing between osteoporotic and non-osteoporotic patients

表 5. 不同预测模型区分骨质疏松患者与非骨质疏松患者的价值

| 模型 | AUC (95% CI) | 敏感性 | 特异性 | 准确率 | PPV | NPV | |
|--------|--------------|---------------------|-------|-------|-------|-------|-------|
| 临床模型 | 训练集 | 0.759 (0.710~0.808) | 0.582 | 0.839 | 0.757 | 0.629 | 0.811 |
| | 内部验证集 | 0.746 (0.667~0.824) | 0.586 | 0.820 | 0.744 | 0.607 | 0.806 |
| | 外部验证集 | 0.765 (0.701~0.829) | 0.726 | 0.836 | 0.787 | 0.78 | 0.792 |
| 影像组学模型 | 训练集 | 0.880 (0.845~0.915) | 0.791 | 0.860 | 0.838 | 0.726 | 0.898 |
| | 内部验证集 | 0.803 (0.735~0.871) | 0.759 | 0.770 | 0.767 | 0.611 | 0.870 |
| | 外部验证集 | 0.848 (0.801~0.895) | 0.855 | 0.740 | 0.791 | 0.725 | 0.864 |

续表

| | | | | | | | |
|-------|-------|---------------------|-------|-------|-------|-------|-------|
| | 训练集 | 0.897 (0.867~0.927) | 0.910 | 0.734 | 0.790 | 0.616 | 0.946 |
| 列线图模型 | 内部验证集 | 0.849 (0.794~0.903) | 0.897 | 0.689 | 0.756 | 0.578 | 0.933 |
| | 外部验证集 | 0.875 (0.828~0.922) | 0.880 | 0.863 | 0.871 | 0.837 | 0.900 |

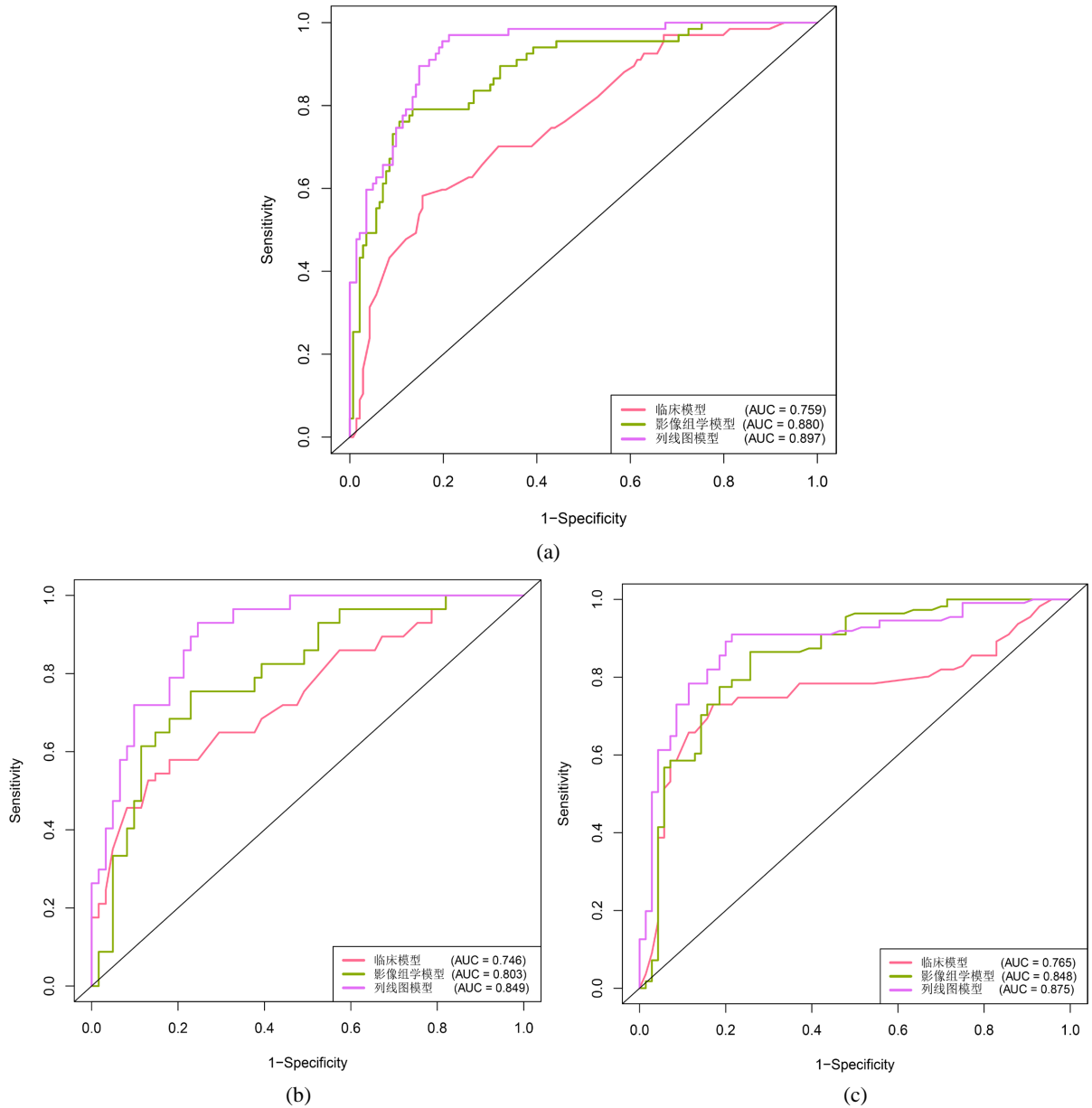
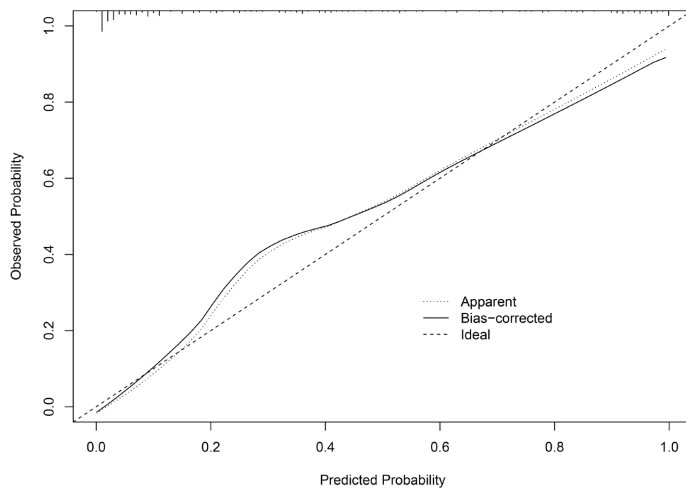


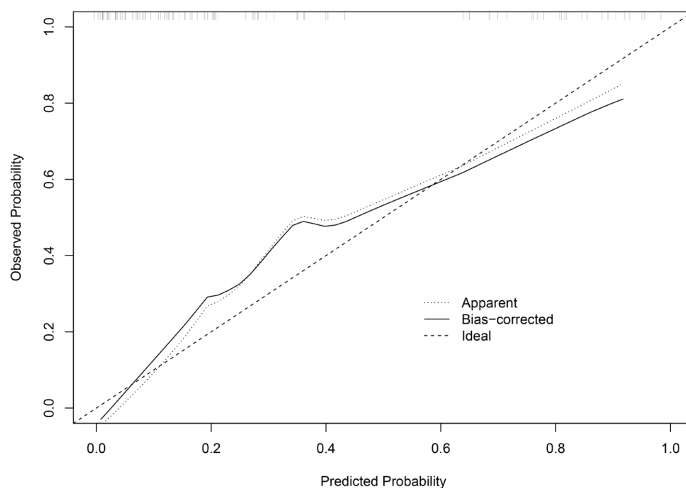
Figure 6. (a) training set ROC curve, (b) Internal validation set ROC curve, (c) External validation set ROC curve

图 6. (a) 训练集 ROC 曲线, (b) 内部验证集 ROC 曲线, (c) 外部验证集 ROC 曲线

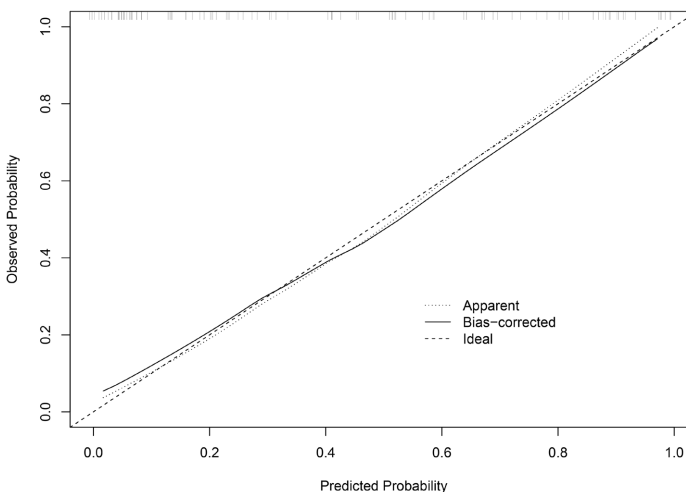
列线图模型的校准曲线详见图 7(a)~(c), 结果显示, 列线图模型的预测概率与实际结果具有较好的一致性。DCA 结果(图 8(a)~(c))显示, 内部验证集中风险阈值范围在 0.18~0.46, 外部验证集中风险阈值范围在 0.22~0.87 时, 列线图模型具有最高的净获益。



(a)



(b)



(c)

Figure 7. (a) Training set calibration curve, (b) internal validation set calibration curve, (c) external validation set calibration curve

图 7. (a) 训练集校准曲线, (b) 内部验证集校准曲线, (c) 外部验证集校准曲线

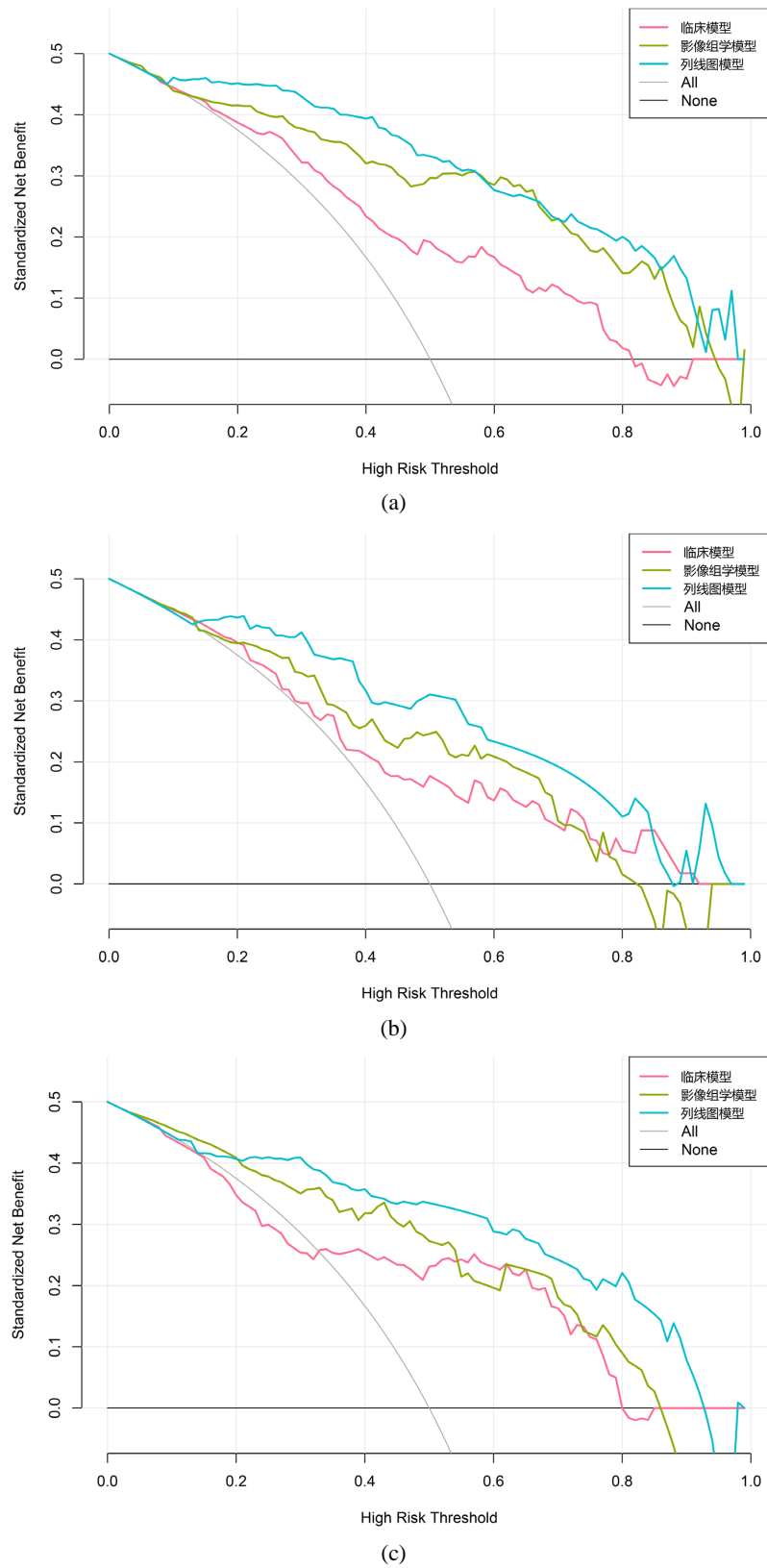


Figure 8. (a) Training set decision curve, (b) Internal validation set decision curve, (c) External validation set decision curve
图 8. (a) 训练集决策曲线, (b) 内部验证集决策曲线, (c) 外部验证集决策曲线

4. 讨论

本研究构建了融合胸部 CT 平扫图像的机器学习影像组学与临床变量特征的列线图模型，用于骨质疏松的机会性筛查。研究结果显示，该列线图模型在各数据集中均表现出最优的筛查性能，训练集、内部验证集和外部验证集中 AUC 分别达 0.897、0.849 和 0.875，准确率分别为 0.790、0.756 和 0.871。本研究结果与既往采用胸部 CT 进行骨质疏松症机会性筛查的研究结果一致，如汪洋等[15]基于胸部 CT 图像放射组学特征构建的骨质疏松症筛查模型 AUC 为 0.85，与本研究影像组学模型的 AUC (0.796~0.896)相近；Yi M 等[16]构建的结合临床因素（年龄、性别、糖皮质激素使用史等）和影像学参数的筛查模型 AUC 为 0.87，与本研究列线图模型的 AUC 结果相符。

本研究从六种机器学习模型，最终选定 LR 模型作为核心影像组学模型，主要因为 LR 模型综合特性更具优势：虽然 LR 模型外部验证集 AUC 为 0.848，与 XGBoost 模型(0.838)预测效能相近，差异无统计学意义($P = 0.631$)，但 LR 模型抗过拟合与稳定性最优，在训练集、内部验证集、外部验证集 AUC 波动幅度最小(0.880, 0.803, 0.848)；反观 XGBoost、LightGBM 等集成模型训练集 AUC 虽高达 0.979、0.985，内部验证集 AUC 骤降至 0.80 以下，SVM 模型也从 0.956 降至 0.734，过拟合特征显著。另外 LR 模型为线性模型，能直接量化影像组学特征对骨质疏松症的预测权重，便于临床理解预测逻辑；而集成模型属于“黑箱模型”，无法直观解析特征贡献度，难以适配临床诊断场景。

本研究经 SMOTE 法处理数据不平衡后，各模型仍存在训练 - 验证集性能落差，其核心原因主要是模型复杂度与数据特征不匹配，集成模型拟合能力过强，易过度捕捉训练集局部噪声，而 LR 线性结构对噪声耐受性更强；二是数据集潜在异质性以及影像组学特征提取受图像分割、重建等主观因素干扰，进一步加剧过拟合。需明确的是，SMOTE 法仅能校正样本类别不平衡问题，无法消除数据集特征异质性、模型 - 数据量不匹配等问题，因此无法彻底规避过拟合风险。

值得注意的是，本模型在三个数据集中的表现存在一定差异：影像组学模型在训练集中 AUC 达 0.880，但在内部验证集降至 0.803，提示模型可能存在一定程度的过拟合；而列线图模型在外部验证集中的准确率(0.871)高于训练集(0.790)和内部验证集(0.756)，表明临床变量引入在一定程度上增强了模型的泛化能力，这与既往研究结论一致。例如，Pickhardt 等[17]在 CT 机会性筛查研究中指出，单纯影像指标虽具有较好的诊断效能，但联合临床因素有望进一步提升模型的稳健性。Song 等[18]最新发表的机器学习研究证实，在容积特征基础上加入身体成分信息后，模型 AUC 从 0.87 提升至 0.95，且在外部验证集中表现出更佳的泛化能力。这一结果为后续研究指明了优化方向，包括开展多中心联合研究、探索更稳健的特征筛选算法，以进一步提升模型的稳定性和外推性。

许多研究表明，CT 上单个或多椎体的骨密度信息不仅与骨密度测量金标准 DXA 的 T 值评分有很好的相关性，李考等[19]发现 T6-L1 椎体 CT 值与 DXA T 值呈显著正相关，通过胸部 CT 测量多椎体 CT 值可有效预测骨质疏松。Weber 等[20]研究显示，QCT (Quantitative Computed Tomography)生成的 BMD T 值与 DXA 高度相关($R^2 = 0.84$)。刘义等[21]开发的 AI 骨密度测量系统与 DXA T 值高度一致($r = 0.767$)，诊断骨质疏松的 AUC 达 0.941，证实以 DXA 为金标准的基于胸部 CT 的筛查模型具有与 QCT 等效的临床应用价值，进一步支持了这一结论。

本研究共纳入 863 例多中心患者，分为训练组、内部验证组和外部验证组。三组患者在性别、年龄、骨质疏松患病率等基本特征方面没有显著差异，具有可比性，这与既往研究一致。Mattia 等[22]在 POSE 研究中证实，分组后两组患者的年龄、性别、BMI、骨密度及骨折主要危险因素均无显著差异，确保了后续疗效比较的可靠性。Ebina 等[23]和 Cosman 等[24]的随机对照试验同样报告了组间基线特征无显著差异的研究结果。

本研究中, 研究对象的平均年龄约为 56 岁, 女性占比约为 74%, 骨质疏松症的患病率约为 63%, 远高于全国 50 岁以上人群的平均水平(19.2%) [14] [25] [26], 可能与所选研究对象的样本中女性占绝大多数的原因。女性骨质疏松患病率高于男性, 主要是与女性绝经后的雌激素水平下降所导致骨吸收增加有关 [27]-[29]。患者年龄是骨质疏松症的重要危险因素, 随着年龄不断增长, 成骨细胞的功能减退, 导致骨量逐渐减少, 因此本研究中, 性别和年龄被确定为骨质疏松症的独立危险因素, 这就为临床模型的构建提供了相当可靠的依据。此外, 有研究证明, 胸腰段交界区(T12, L1)的骨密度变化相对最为敏感, 可能是早期骨量流失的“预警窗口” [30]-[32]。因此本研究选择以胸腰段(T12, L1)椎体为中心勾画 ROI 进行模型构建对胸腰椎体骨密度进行精准评估具有非常重要的临床意义。该区域骨密度细微变化可比较早地反映全身骨骼健康状况, 为干预提供关键时间窗口[33]。

本研究构建的列线图模型基于常规胸部 CT 检查数据, 无需额外增加检查成本和时间, 适用于多种临床场景, 对于 50 岁以上长期吸烟、有肺癌家族史等肺癌高危人群和健康体检人群, 在进行胸部 CT 肺癌筛查的同时, 可采用本模型进行骨质疏松症机会性筛查, 实现两种疾病的同步风险评估, 这与当前影像学研究的前沿方向高度一致[34]; 另外对于骨科术前评估患者, 在进行胸椎相关手术前, 可利用既往胸部 CT 评估骨质疏松风险, 指导术中操作和术后管理; 该模型的应用有望推动骨质疏松症的早诊早治, 成为老年人群的常规体检骨质疏松症筛查的有效工具, 降低骨质疏松性骨折的发生风险并降低医疗负担, 尤其在资源有限的地区更具推广价值。

本研究为回顾性研究, 不可避免地存在选择和信息偏倚。本研究中椎体分割用的是人工手动的方式, 可能带有一定的偏差, 未来的研究可以开发出使用自动分割算法来提高 ROI 的勾画速度、勾画的准确性和一致性。另外, 本文使用的放射组学特征提取方法、机器学习算法上也做了改进, 未来研究中可以尝试使用更多放射组学特征(深度学习特征)来提高模型的准确率。另外外部验证组的患者只来自于一家医院, 存在地域上的局限性, 未来的研究可以包含更多的地区、更多的医疗机构患者的样本, 进而检验模型的普适性。

5. 结论

综上所述, 本研究构建的基于胸部 CT 平扫的机器学习影像组学列线图模型, 在骨质疏松机会性筛查中表现出良好的效能和稳定性, 优于单纯临床模型, 与金标准 DXA 结果具有高度一致性。该模型能够更全面、更精准地评估骨质疏松风险, 为临床早期筛查与个性化干预提供可靠的量化工具。

声明

本研究为回顾性研究, 遵循《赫尔辛基宣言》的伦理原则, 并获得青岛大学附属医院伦理委员会的批准(批件号: QYFY-WZLL-42131)。因研究仅使用既往临床诊疗中常规采集的匿名化数据, 不涉及前瞻性干预及患者隐私泄露风险, 伦理委员会同意豁免患者知情同意。

参考文献

- [1] 中华医学会骨质疏松和骨矿盐疾病分会. 原发性骨质疏松症诊疗指南(2022 年版) [J]. 中华骨质疏松和骨矿盐疾病杂志, 2022, 15(6): 573-611.
- [2] Khan, A.A., *et al.* (2024) Osteoporotic Fractures: Diagnosis, Evaluation, and Significance from the International Working Group on DXA Best Practices. *Mayo Clinic Proceedings*, **99**, 1127-1141. <https://doi.org/10.1016/j.mayocp.2024.01.011>
- [3] Gómez, M.P.A., Wáng, Y.J., Yu, J.S., Johnson, R. and Chang, C.Y. (2025) Dual-Energy X-Ray Absorptiometry for Osteoporosis Screening: AJR Expert Panel Narrative Review. *American Journal of Roentgenology*, **225**, e2532802. <https://doi.org/10.2214/ajr.25.32802>

- [4] Slart, R.H.J.A., Punda, M., Ali, D.S., Bazzocchi, A., Bock, O., Camacho, P., *et al.* (2024) Updated Practice Guideline for Dual-Energy X-Ray Absorptiometry (DXA). *European Journal of Nuclear Medicine and Molecular Imaging*, **52**, 539-563. <https://doi.org/10.1007/s00259-024-06912-6>
- [5] Pan, Y., Wu, Y., Wang, H., Yu, T., He, D., Lu, X., *et al.* (2024) Opportunistic Use of Chest Low-Dose Computed Tomography (LDCT) Imaging for Low Bone Mineral Density and Osteoporosis Screening: Cutoff Thresholds for the Attenuation Values of the Lower Thoracic and Upper Lumbar Vertebrae. *Quantitative Imaging in Medicine and Surgery*, **14**, 4792-4803. <https://doi.org/10.21037/qims-24-59>
- [6] 王旭, 刘磊, 刘义军, 等. 胸腹部平扫 CT 值用于机会性筛查骨质疏松的可行性[J]. 放射学实践, 2024, 39(3): 393-398.
- [7] Buckens, C.F., Dijkhuis, G., de Keizer, B., Verhaar, H.J. and de Jong, P.A. (2015) Opportunistic Screening for Osteoporosis on Routine Computed Tomography? An External Validation Study. *European Radiology*, **25**, 2074-2079. <https://doi.org/10.1007/s00330-014-3584-0>
- [8] Kim, Y., Kim, H.Y., Lee, S., Hong, S. and Lee, J.W. (2024) Age-Dependent Changes in CT Vertebral Attenuation Values in Opportunistic Screening for Osteoporosis: A Nationwide Multi-Center Study. *European Radiology*, **35**, 3519-3527. <https://doi.org/10.1007/s00330-024-11263-9>
- [9] Lee, H., Park, S., Kwack, K. and Yun, J.S. (2023) CT and MR for Bone Mineral Density and Trabecular Bone Score Assessment in Osteoporosis Evaluation. *Scientific Reports*, **13**, Article No. 16574. <https://doi.org/10.1038/s41598-023-43850-z>
- [10] Pan, J., Lin, P., Gong, S., Wang, Z., Cao, R., Lv, Y., *et al.* (2024) Feasibility Study of Opportunistic Osteoporosis Screening on Chest CT Using a Multi-Feature Fusion DCNN Model. *Archives of Osteoporosis*, **19**, Article No. 98. <https://doi.org/10.1007/s11657-024-01455-7>
- [11] 姜文蓁, 张宇威, 崔效楠, 等. 胸部低剂量 CT 结合定量 CT 测量下段胸椎骨密度诊断骨质疏松[J]. 中国医学影像技术, 2022, 38(5): 734-738.
- [12] Welsner, M., Navel, H., Hosch, R., Rathsmann, P., Stehling, F., Mathew, A., *et al.* (2024) Opportunistic Screening for Low Bone Mineral Density in Adults with Cystic Fibrosis Using Low-Dose Computed Tomography of the Chest with Artificial Intelligence. *Journal of Clinical Medicine*, **13**, Article 5961. <https://doi.org/10.3390/jcm13195961>
- [13] Lin, X., Shen, R., Zheng, X., Shi, S., Dai, Z. and Fang, K. (2024) Utilizing Radiomics Techniques to Isolate a Single Vertebral Body from Chest CT for Opportunistic Osteoporosis Screening. *BMC Musculoskeletal Disorders*, **25**, Article No. 785. <https://doi.org/10.1186/s12891-024-07903-2>
- [14] 中华医学会骨质疏松和骨矿盐疾病分会. 原发性骨质疏松症诊疗指南(2022) [J]. 中国全科医学, 2023, 26(14): 1671-1691.
- [15] 汪洋, 张联合. 腹部影像组学对乳腺癌相关骨质疏松机会性筛查的研究[J]. 中国医学计算机成像杂志, 2023, 29(5): 560-566.
- [16] Yi, M., Lin, W., Zhang, Y., Fang, X., Zhang, G., Song, J., *et al.* (2025) Development and Internal Validation of a Clinical-Radiological Nomogram for Osteoporosis Screening: A Cohort Retrospective Study. *European Spine Journal*, **34**, 5053-5064. <https://doi.org/10.1007/s00586-025-09330-w>
- [17] Pickhardt, P.J. (2022) Value-added Opportunistic CT Screening: State of the Art. *Radiology*, **303**, 241-254. <https://doi.org/10.1148/radiol.211561>
- [18] Song, J., Cho, S.W., Yoo, H.J., Cho, S.J., Hong, N. and Yoon, S.H. (2026) Enhanced Opportunistic CT Screening for Osteoporosis Using Machine Learning Derived Volumetric Vertebral and Complementary Body Composition Information. *European Journal of Radiology*, **194**, Article ID: 112555. <https://doi.org/10.1016/j.ejrad.2025.112555>
- [19] 李考, 王萍, 闫如意, 等. 慢性阻塞性肺疾病患者椎骨的 CT 值与骨密度的关系[J]. 中华结核和呼吸杂志, 2018, 41(5): 340-344.
- [20] Weber, N.K., Fidler, J.L., Keaveny, T.M., Clarke, B.L., Khosla, S., Fletcher, J.G., *et al.* (2014) Validation of a CT-Derived Method for Osteoporosis Screening in IBD Patients Undergoing Contrast-Enhanced CT Enterography. *American Journal of Gastroenterology*, **109**, 401-408. <https://doi.org/10.1038/ajg.2013.478>
- [21] 崔璨, 温庆祥, 刘妮, 等. 基于胸部 CT 的人工智能骨密度测量系统机会性筛查骨质疏松症的可行性研究[J]. CT 理论与应用研究(中英文), 2025, 34(6): 1182-1187.
- [22] Mattia, L., Davis, S., Mark-Wagstaff, C., Abrahamsen, B., Peel, N., Eastell, R., *et al.* (2022) Utility of PINP to Monitor Osteoporosis Treatment in Primary Care, the POSE Study (PINP and Osteoporosis in Sheffield Evaluation). *Bone*, **158**, Article ID: 116347. <https://doi.org/10.1016/j.bone.2022.116347>
- [23] Ebina, K., Hashimoto, J., Kashii, M., Hirao, M., Kaneshiro, S., Noguchi, T., *et al.* (2016) The Effects of Switching Daily Teriparatide to Oral Bisphosphonates or Denosumab in Patients with Primary Osteoporosis. *Journal of Bone and Mineral Metabolism*, **35**, 91-98. <https://doi.org/10.1007/s00774-015-0731-x>

-
- [24] Cosman, F., McMahon, D., Dempster, D. and Nieves, J.W. (2019) Standard versus Cyclic Teriparatide and Denosumab Treatment for Osteoporosis: A Randomized Trial. *Journal of Bone and Mineral Research*, **35**, 219-225. <https://doi.org/10.1002/jbmr.3850>
- [25] 汤淑女, 尹香君, 余卫, 等. 中国 40 岁及以上绝经后女性骨质疏松症患病率及其影响因素研究[J]. 中华流行病学杂志, 2022, 43(4): 509-516.
- [26] Cheng, X., Zhao, K., Zha, X., Du, X., Li, Y., Chen, S., *et al.* (2020) Opportunistic Screening Using Low-Dose CT and the Prevalence of Osteoporosis in China: A Nationwide, Multicenter Study. *Journal of Bone and Mineral Research*, **36**, 427-435. <https://doi.org/10.1002/jbmr.4187>
- [27] Cui, Z., *et al.* (2019) Estimation and Prediction of Standardized Prevalence of Osteoporosis in China. *Archives of Osteoporosis*, **14**, 110.
- [28] Yao, Y., Cai, X., Chen, Y., Zhang, M. and Zheng, C. (2024) Estrogen Deficiency-Mediated Osteoimmunity in Postmenopausal Osteoporosis. *Medicinal Research Reviews*, **45**, 561-575. <https://doi.org/10.1002/med.22081>
- [29] Zhao, H., Yu, F. and Wu, W. (2025) New Perspectives on Postmenopausal Osteoporosis: Mechanisms and Potential Therapeutic Strategies of Sirtuins and Oxidative Stress. *Antioxidants*, **14**, Article 605. <https://doi.org/10.3390/antiox14050605>
- [30] Melton, L.J., Kan, S.H., Frye, M.A., Wahner, H.W., O'Fallon, W.M. and Riggs, B.L. (1989) Epidemiology of Vertebral Fractures in Women. *American Journal of Epidemiology*, **129**, 1000-1011. <https://doi.org/10.1093/oxfordjournals.aje.a115204>
- [31] Briggs, A.M., van Dieën, J.H., Wrigley, T.V., Greig, A.M., Phillips, B., Lo, S.K., *et al.* (2007) Thoracic Kyphosis Affects Spinal Loads and Trunk Muscle Force. *Physical Therapy*, **87**, 595-607. <https://doi.org/10.2522/ptj.20060119>
- [32] Guglielmi, G., Floriani, I., Torri, V., Li, J., van Kuijk, C., Genant, H.K., *et al.* (2005) Effect of Spinal Degenerative Changes on Volumetric Bone Mineral Density of the Central Skeleton as Measured by Quantitative Computed Tomography. *Acta Radiologica*, **46**, 269-275. <https://doi.org/10.1080/02841850510012661>
- [33] Erdem, I., Truumees, E. and van der Meulen, M.C.H. (2013) Simulation of the Behaviour of the L1 Vertebra for Different Material Properties and Loading Conditions. *Computer Methods in Biomechanics and Biomedical Engineering*, **16**, 736-746. <https://doi.org/10.1080/10255842.2011.636741>
- [34] Pickhardt, P.J., Graffy, P.M., Zea, R., Lee, S.J., Liu, J., Sandfort, V., *et al.* (2020) Automated CT Biomarkers for Opportunistic Prediction of Future Cardiovascular Events and Mortality in an Asymptomatic Screening Population: A Retrospective Cohort Study. *The Lancet Digital Health*, **2**, e192-e200. [https://doi.org/10.1016/s2589-7500\(20\)30025-x](https://doi.org/10.1016/s2589-7500(20)30025-x)