

Constructing an automated prediction model for osteoporosis based on chest radiographs: A multicenter study

LI Zhijian¹, HAO Yixiu², HU Didi³, ZENG Lingxiang⁴, ZHAO Yingnan⁵,
LONG Wansheng⁵, ZHANG Weicong^{5*}

(1. Department of Intervention, 4. Department of Orthopedics, Xinhui People's Hospital, Jiangmen 529100, China;
2. Department of Radiology, Guangzhou First People's Hospital, Guangzhou 510000, China; 3. Department of
Nephrology, 5. Jiangmen Key Laboratory of Artificial Intelligence in Medical Image Computation and
Application, Jiangmen Central Hospital, Jiangmen 529000, China)

[Abstract] **Objective** To construct an automated prediction model for osteoporosis (OP) based on chest radiographs. **Methods** A total of 4 542 patients from 3 hospitals (1 700 OP and 2 842 non-OP) were retrospectively enrolled. Deep learning (DL) model, clinical model and combined model for predicting OP were constructed. DL model adopted a hybrid architecture combining ResNet-18 and Transformer, and extracted local and global features based on a 6×6 grid image block partitioning strategy. The efficacy of these models for predicting OP were evaluated according to indicators such as the area under the curve (AUC) of receiver operating characteristic curve, accuracy and F1 score, while their clinical net benefits were observed through decision curve analysis, and Grad-CAM heatmap was used for visual interpretation. **Results** AUC of DL model for predicting OP in internal test set and external test set 1 and 2 was 0.917, 0.897 and 0.921, respectively, of clinical mode was 0.839, 0.759 and 0.532, of combined model was 0.919, 0.896 and 0.919, respectively. AUC of combined model was higher than that of clinical model in all sets, also higher than that of DL model in internal test set. The clinical net benefits of both DL model and combined model were higher than that of clinical model. Grad-CAM visualization showed that DL model mainly focused on the bony areas in chest X-rays. **Conclusion** Automated prediction model based on chest radiographs could be used to effectively predict OP, which had good generalization ability.

[Keywords] osteoporosis; deep learning; X-rays; multicenter study

DOI:10.13929/j.issn.1003-3289.2026.03.017

基于胸部 X 线片构建骨质疏松症自动化预测模型： 多中心研究

李志建¹, 郝以秀², 胡蒟蒻³, 曾凌翔⁴, 赵英男⁵, 龙晚生⁵, 张伟聪^{5*}

(1. 新会人民医院介入科, 4. 骨科一区, 广东 江门 529100; 2. 广州市第一人民医院放射科, 广东 广州 510000;
3. 江门市中心医院肾内科, 5. 医学影像智能计算及应用实验室, 广东 江门 529000)

[摘要] **目的** 基于胸部 X 线片构建骨质疏松症(OP)自动化预测模型。 **方法** 回顾性纳入 3 家医院共 4 542 例患者 (OP 1 700 例, 非 OP 2 842 例) 胸片数据, 分别构建深度学习(DL)模型、临床模型及联合模型用于预测 OP。 DL 模型采用 ResNet-18 与 Transformer 相结合混合架构, 基于 6×6 网格图像块划分策略提取局部与全局特征。 以受试者工作特征曲线的曲线下面积(AUC)、准确率和 F1 分数等指标评估模型效能, 以决策曲线分析评估临床净收益, 将 Grad-CAM 热力图用于可视化解释。 **结果** DL 模型预测内部测试集和外部测试集 1、2 OP 的 AUC 分别为 0.917、0.897 及 0.921, 临床模型

[第一作者] 李志建(1970—), 男, 湖南衡阳人, 本科, 副主任医师。 研究方向: 人工智能影像。 E-mail: 85390464@qq.com

[通信作者] 张伟聪, 江门市中心医院医学影像智能计算及应用实验室, 529000。 E-mail: zhangweichong@jmszxyy.com.cn

[收稿日期] 2025-08-29 **[修回日期]** 2025-11-07

分别为 0.839、0.759 及 0.532, 联合模型分别为 0.919、0.896 及 0.919。联合模型在各集的 AUC 均高于临床模型、在内部测试集高于 DL 模型; DL 模型和联合模型的临床净收益均高于临床模型。Grad-CAM 可视化显示, DL 模型主要聚焦胸片中的骨性区域。**结论** 基于胸部 X 线片构建的自动化预测模型能有效预测 OP 且泛化能力良好。

[关键词] 骨质疏松症; 深度学习; X 线; 多中心研究

[中图分类号] R589; R445 **[文献标识码]** A **[文章编号]** 1003-3289(2026)03-0403-05

全球老龄化背景下, 骨质疏松症 (osteoporosis, OP) 已成为严重威胁老年人健康的公共卫生问题, 我国 50 岁以上人群 OP 患病率为 19.2%、65 岁以上达 32.0%, OP 总数超过 9 000 万例^[1]。OP 引发的骨密度 (bone mineral density, BMD) 下降和骨微结构破坏显著增加骨折风险, 影响患者生活质量及死亡率^[2-4]; 但受限于双能 X 线吸收测定法 (dual-energy X-ray absorptiometry, DXA) 设备普及率不足和患者筛查依从性差^[5-7], 目前 OP 临床检出率偏低。利用深度学习 (deep learning, DL) 技术能在不增加额外成本的前提下利用常规影像学数据筛查病变^[8-10], 为机会性筛查 OP 提供了新的思路。OP 为全身性骨代谢疾病; 而胸部 X 线片可能隐含与 BMD 相关的结构特征信息, 具有开展人工智能辅助筛查的潜力。本研究观察基于多中心胸部 X 线片构建的自动化模型预测 OP 的效能。

1 资料与方法

1.1 研究对象 回顾性纳入江门市中心医院 2013 年 11 月—2024 年 7 月 4 278 例患者, 按 8:2 比例随机分为训练集 (3 422 例) 和内部测试集 (856 例); 并以广州市第一人民医院 2022 年 4 月—2024 年 3 月 188 例为外部测试集 1、新会人民医院 2019 年 4 月—2024 年 12 月 76 例作为外部测试集 2。纳入标准: ①年龄 >40 岁; ②于 6 个月内先后接受胸部正位 X 线检查和测量 DXA-BMD。排除标准: ①继发性 OP; ②胸片或 BMD 中的测量部位存在高密度异物; ③脊柱结构异常 (如骨折或侧弯); ④既往骨科手术史 (如内固定或椎体成形术); ⑤图像质量不佳或存在明显伪影。对接受多次检查者选取符合条件的首组影像和 BMD 数据进行分析。依据 WHO OP 诊断标准, 依据 DXA 所测中轴骨密度经换算后的最低 T 值进行分组, 以 T 值 ≤ -2.5 为 OP、> -2.5 为非 OP^[3-4]。本研究经 3 家医院伦理审查委员会批准 (江心医伦理审查 [2024]267 号 A、K-2025-016-01、医研伦审批 2024 第 34 号)。

1.2 图像预处理 采用标准化预处理方法统一各中心胸片处理流程。首先将从 PACS 系统中提取的 DICOM 图像进行匿名化处理, 之后转换为 16 位 PNG 格式; 随后采用 6×6 网格将图像分割成 36 个互不重叠

的矩形块, 并通过双线性插值将其统一缩放为 256×256 像素。对全部胸片进行曝光异常筛查, 剔除严重曝光不足或过度曝光者; 采用 Z-score 归一化降低成像设备及体型差异等引起的亮度与对比度偏差。在模型训练阶段对训练集实施数据增强, 包括随机水平翻转、旋转、裁剪缩放及色彩扰动, 以增强模型鲁棒性。

1.3 构建与训练 DL 模型 采用卷积神经网络与 Transformer 结合的混合架构捕捉图像局部纹理特征与区域间的全局依赖关系。以经 ImageNet 预训练的 ResNet-18 为特征提取器 f_{CNN} , 将其输入通道调整为 1 以处理灰度图像, 并移除全局池化层和分类层。将 36 个图像块分别输入共享权重的特征提取模块, 生成 512 维特征向量。引入可学习的 [BMD] 令牌 $Z_{bmd} \in \mathbb{R}^D$ 作为全局信息聚合器, 与特征序列拼接并加入位置编码后, 将其输入到由 6 层相同编码层堆叠的 Transformer 模块中, 每层均采用 Post-Norm 结构, 包含多头自注意力 (multi-headed self-attention, MHSA) 和前馈网络 (multi-layer perceptron, MLP): $E'_i = LayerNorm (MHSA (E_{i-1}) + E_{i-1})$, $E_i = LayerNorm [MLP (E'_i) + E'_i]$ 。模型通过双路预测头输出结果: 全局预测头以 [BMD] 令牌对应的输出特征 y_{bmd} 生成全局预测 $L_{global} = MLP_{global} (y_{bmd})$; 局部预测头则通过独立的 MLP 将每个图像块的特征 y_i 生成局部预测 $L_{local, i} = MLP_{local, i} (y_i)$; 全局预测头通过 Sigmoid 激活函数输出 0~1 之间的连续值, 该值代表 DL 模型预测概率, 将其定义为 OP 风险评分。见图 1。

在配备 NVIDIA GPU 的服务器上基于 PyTorch 框架对模型进行训练。参数优化器采用随机梯度下降, 初始学习率设为 0.000 1, 动量为 0.9, 批量大小为 8。损失函数包括全局损失与局部损失, 采用 5 折交叉验证, 并以 Grad-CAM 将 DL 模型可视化。

1.4 构建临床及联合模型 利用极限梯度提升 (extreme gradient boosting, XGBoost) 算法基于性别、年龄、身高、体质量和体质量指数 (body mass index, BMI) 构建临床模型。之后引入 DL 模型生成的 OP 风险评分, 以 XGBoost 算法构建联合模型。

1.5 统计学分析 采用 R 4.4.2 统计分析软件。以

$\bar{x} \pm s$ 表示连续变量,组间行独立样本 t 检验;以频数表示分类变量,组间行 Pearson χ^2 检验。绘制受试者工作特征(receiver operating characteristic, ROC)曲线,以曲线下面积(area under the curve, AUC)、敏感度、特异度、准确率、阳性预测值、阴性预测值及 F1 分数评估各模型预测效能。采用 DeLong 检验比较 AUC,以决策曲线分析(decision curve analysis, DCA)评估临床净收益。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 临床资料 全部 4 542 例中,OP 1 700 例、非 OP 2 842 例。内部数据集和外部测试集 1 内 OP 亚组年龄、女性占比均高于,而 BMI、BMD 均低于非 OP 亚组(P 均 < 0.05);外部测试集 2 内 OP 亚组 BMD 低于非 OP 亚组($P < 0.05$)。见表 1。

2.2 模型效能 DL 模型预测内部测试集和外部测试集 1、2 OP 的 AUC 分别为 0.917、0.897 及 0.921,

临床模型分别为 0.839、0.759 及 0.532,联合模型分别为 0.919、0.896 及 0.919。联合模型在各集的 AUC 均高于临床模型、在内部测试集的 AUC 高于 DL 模型(P 均 < 0.05)。见图 2 和表 2。

GRAD-CAM 热力图显示,多数情况下,DL 模型聚焦于骨骼区域,在部分病例突出肱骨近端、锁骨及胸腰椎等部位(图 3A),但对个别病例因关注非骨性区域而产生误判(图 3B)。DCA 显示,DL 模型与联合模型在各集的临床净收益均优于临床模型,见图 4。

3 讨论

胸部 X 线片为临床最常用的基础影像学检查之一,设备普及率高、操作便捷、成本低廉且辐射剂量低等,常用于大规模健康体检。近年来,基于胸片的大数据应用已从传统的诊断疾病拓展至筛查冠状动脉硬化、气胸等疾病及预测风险^[11-12]。胸片覆盖锁骨、肋骨及椎体等关键骨结构,其灰度、纹理及骨皮质厚度等特征与 BMD 密

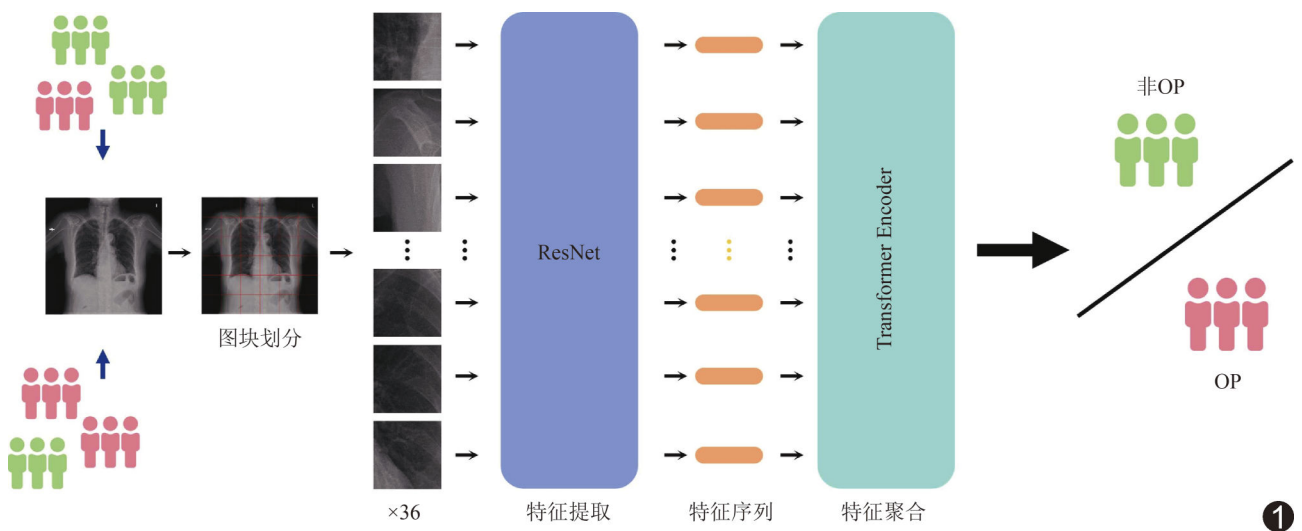


图 1 构建与训练 DL 模型过程

表 1 OP 与非 OP 患者临床基线资料比较

集别	亚组别	年龄(岁)	性别(例)		BMI(kg/m ²)	BMD(g/cm ²)
			男	女		
内部数据集($n=4\ 278$)	OP 亚组($n=1\ 518$)	71.1 \pm 8.9	206	1 312	22.2 \pm 3.5	0.6 \pm 0.1
	非 OP 亚组($n=2\ 760$)	58.8 \pm 11.3	840	1 920	24.3 \pm 3.4	0.9 \pm 0.1
	χ^2/t 值	39.062	150.784		-19.038	-71.804
	P 值	<0.001	<0.001		<0.001	<0.001
外部测试集 1($n=188$)	OP 亚组($n=119$)	74.2 \pm 9.8	20	99	22.9 \pm 5.2	0.6 \pm 0.1
	非 OP 亚组($n=69$)	67.3 \pm 11.8	31	38	24.5 \pm 3.8	0.8 \pm 0.1
	χ^2/t 值	4.098	-17.471		-2.427	-12.295
	P 值	<0.001	<0.001		<0.001	<0.001
外部测试集 2($n=76$)	OP 亚组($n=63$)	75.3 \pm 11.6	11	52	21.0 \pm 3.8	0.6 \pm 0.1
	非 OP 亚组($n=13$)	75.8 \pm 10.4	4	9	21.9 \pm 3.5	0.8 \pm 0.3
	χ^2/t 值	-0.179	1.205		-0.882	-2.676
	P 值	0.860	0.273		0.389	0.020

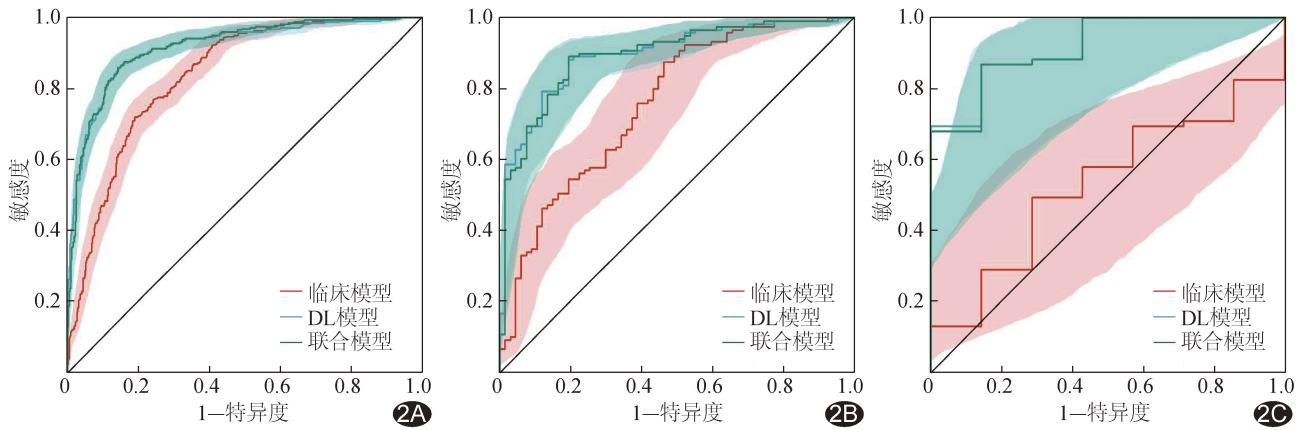


图 2 各模型区分各集 OP 的 ROC 曲线 A. 内部测试集; B. 外部测试集 1; C. 外部测试集 2

表 2 比较各模型预测 OP 效能

集别	模型	AUC	敏感度(%)	特异度(%)	准确率(%)	阳性预测值(%)	阴性预测值(%)	F1 分数
内部测试集	DL 模型	0.917*	84.01	85.85	85.17	79.36	89.45	0.815
	临床模型	0.839*	80.52	74.74	77.03	67.41	85.94	0.732
	联合模型	0.919	85.02	83.63	84.24	77.15	89.77	0.807
外部测试集 1	DL 模型	0.897	64.15	88.16	72.72	92.18	60.45	0.733
	临床模型	0.759*	76.84	60.03	70.80	77.74	59.58	0.772
	联合模型	0.896	79.55	77.05	78.63	86.52	68.18	0.826
外部测试集 2	DL 模型	0.921	52.93	99.74	62.95	99.52	24.71	0.683
	临床模型	0.532*	84.61	25.06	77.61	88.33	22.25	0.863
	联合模型	0.919	73.52	75.01	71.10	94.99	32.48	0.805

注: * :与集内联合模型比较 $P < 0.05$

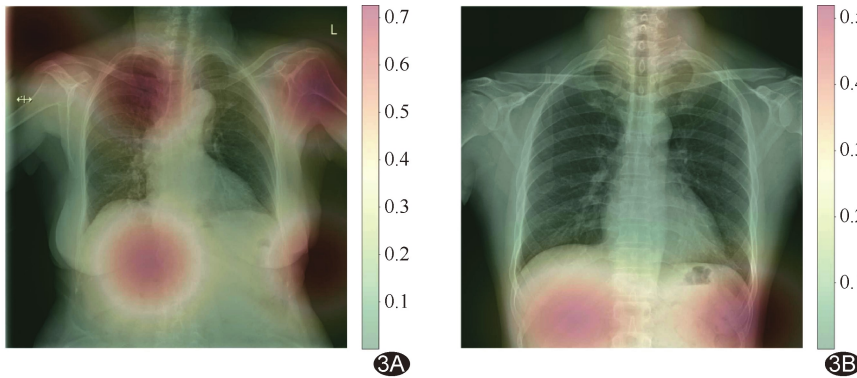


图 3 DL 模型的 Grad-CAM 热力图 A. 模型聚焦骨骼区域且分类正确; B. 模型聚焦非骨骼区域致分类错误 (红色区域代表模型赋予较高权重的部位,黄色与绿色区域权重较低)

块分割并结合 ResNet-18 与 Transformer 架构实现局部纹理捕捉与跨区域信息交互。本研究获 DL 模型在各集的 AUC 均超过 0.897,提示其具有较高预测效能及稳健的跨中心泛化能力,并验证了胸片作为机会性筛查 OP 的载体的潜力。Grad-CAM 显示模型主要关注骨性解剖区域,支持其表征学习具有生物学合理性,并提升了模型可解释性。

切相关。本研究基于胸片构建预测 OP 模型,旨在为 DXA 资源不足场景提供无额外辐射、低成本、易推广的机会性筛查手段,辅助识别高危人群并进行分层管理。

既往基于 X 线片预测 OP 多依赖手动标注 ROI 或检测骨性标志点^[8-9,13],耗时、费力、可重复性差,且易受定位误差影响。本研究所用全图网格化建模与自注意力机制能自动聚焦骨骼相关亚区域,可避免人工干预和检测误差累积;基于胸片以 DL 自动预测 OP,通过图像

结合临床可提升影像学预测多种疾病效能^[14-15]。本研究联合模型在各集的 AUC 均高于临床模型、且在内部测试集高于 DL 模型,表明胸片包含高信息密度骨质关联特征,而所纳入临床变量的贡献增益有限、甚至可能引入冗余噪声,提示未来需进一步优化多模态融合策略以提升协同效能。对本研究模型预测提示 OP、尤其 65 岁以上女性或存在骨折高风险的绝经后女性应进一步行 DXA 以确诊^[4]。

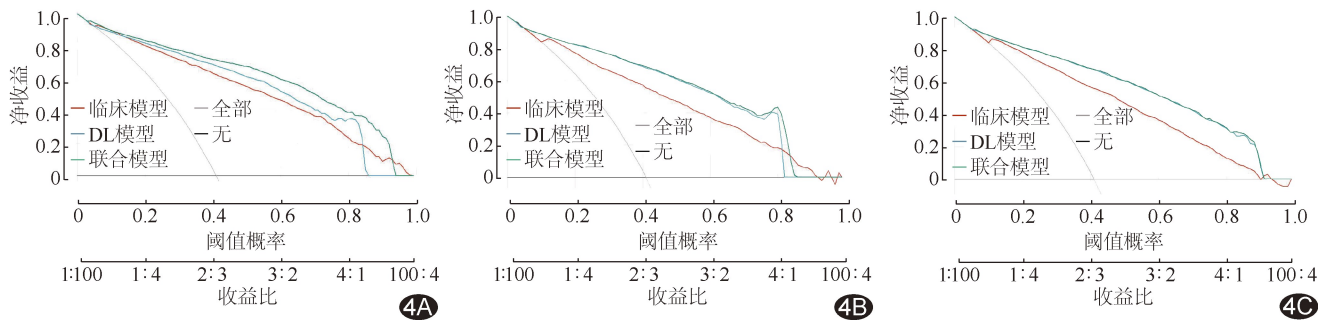


图 4 各模型在各集的 DCA 图 A. 内部测试集; B. 外部测试集 1; C. 外部测试集 2

综上所述,基于胸部 X 线片构建的自动化预测模型能有效预测 OP 且泛化能力良好,有望成为高效、可推广的机会性筛查 OP 工具。本研究的主要局限性:①样本量相对有限,需通过更大规模、更多样化人群进一步验证;②仅进行 OP 与非 OP 二分类,后续将拓展至三分类任务,以更精细地刻画骨量减少阶段表征;③仅基于正位胸片构建 DL 模型,未来可尝试构建融合胸部侧位片或 CT 等多模态影像信息的模型。

利益冲突:全体作者声明无利益冲突。

作者贡献:李志建研究设计、撰写和修改文章;郝以秀、胡蒟蒻和曾凌翔数据分析;赵英男图像处理、数据分析;龙晚生和张伟聪指导、审阅和修改文章。

[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)[J]. 中国全科医学, 2023, 26(14):1671-1691.
- [2] COMPSTON J E, McCLUNG M R, LESLIE W D. Osteoporosis[J]. Lancet, 2019, 393(10169):364-376.
- [3] ENSRUD K E, CRANDALL C J. Osteoporosis[J]. Ann Intern Med, 2024, 177(1):ITC1-ITC16.
- [4] US Preventive Services Task Force, NICHOLSON W K, SILVERSTEIN M, et al. Screening for osteoporosis to prevent fractures: US Preventive Services Task Force recommendation statement[J]. JAMA, 2025, 333(6):498-508.
- [5] LEBOFF M S, GREENSPAN S L, INSOGNA K L, et al. The clinician's guide to prevention and treatment of osteoporosis[J]. Osteoporos Int, 2022, 33(10):2049-2102.
- [6] ZENG Q, LI N, WANG Q, et al. The prevalence of osteoporosis in China, a nationwide, multicenter DXA survey [J]. J Bone Miner Res, 2019, 34(10):1789-1797.
- [7] CHENG X, ZHAO K, ZHA X, et al. Opportunistic screening using low-dose CT and the prevalence of osteoporosis in China: A nationwide, multicenter study [J]. J Bone Miner Res, 2021, 36(3):427-435.
- [8] SUKAGAWA S, FUJIMURA A, TAGUCHI A, et al. Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates [J]. Sci Rep, 2022, 12(1):6088.
- [9] HO C S, FAN T Y, KUO C F, et al. HarDNet-based deep learning model for osteoporosis screening and bone mineral density inference from hand radiographs [J]. Bone, 2025, 190:117317.
- [10] OH S, KANG W Y, PARK H, et al. Evaluation of deep learning-based quantitative computed tomography for opportunistic osteoporosis screening [J]. Sci Rep, 2024, 14(1):363.
- [11] HONG Y, JEONG H, JANG Y, et al. Predicting categories of coronary artery calcium scores from chest X-ray images using deep learning [J]. J Cardiovasc Comput Tomogr, 2025, 19(3):331-339.
- [12] HONG W, HWANG E J, LEE J H, et al. Deep learning for detecting pneumothorax on chest radiographs after needle biopsy: Clinical implementation [J]. Radiology, 2022, 303(2):433-441.
- [13] WANG F, ZHENG K, LU L, et al. Lumbar bone mineral density estimation from chest X-ray images: Anatomy-aware attentive multi-ROI modeling [J]. IEEE Trans Med Imaging, 2023, 42(1):257-267.
- [14] LIU Y, REN H, PEI Y, et al. Development of a CT-based comprehensive model combining clinical, radiomics with deep learning for differentiating pulmonary metastases from noncalcified pulmonary hamartomas: A retrospective cohort study [J]. Int J Surg, 2024, 110(8):4900-4910.
- [15] SAAD M B, HONG L, AMINU M, et al. Predicting benefit from immune checkpoint inhibitors in patients with non-small-cell lung cancer by CT-based ensemble deep learning: A retrospective study [J]. Lancet Digit Health, 2023, 5(7):e404-e420.