

## 基于低剂量胸部 CT 深度学习模型自动测量骨密度研究

赵宇, 张晓岚, 郑超, 王敏红, 洪薇, 周运锋

**【摘要】** 目的:评价基于低剂量胸部 CT 深度学习模型全自动测量与定量 CT(QCT)手动测量骨密度的一致性。方法:回顾性分析 2018 年 6 月—2019 年 12 月在弋矶山医院行低剂量胸部 CT 筛查联合定量 CT(QCT)骨密度测量的 1406 例体检者的临床和影像资料。随机分为训练集(985 例)和测试集(421 例)。应用深度学习等方法构建骨分割模型等及内部组织校准模型,应用测试集检测其效能。以 QCT 结果作为参考标准,应用 Spearman 相关系数、组内相关系数、Bland-Altman 分析两种测量方法的相关性和一致性;以 QCT 测量结果的骨质疏松(OP)诊断作为参考标准,应用 ROC 曲线评价其对 OP 的诊断效能。结果:训练集及测试集中深度学习模型与 QCT 的骨密度测量结果均呈正相关(训练集: $r=0.957, P<0.001$ ;测试集: $r=0.955, P<0.001$ ),组内相关系数为 0.946(训练集)、0.945(测试集)。该模型在训练集中 ROC 曲线下面积(AUC)、灵敏度、特异度、准确度分别为 0.986、47.5%、100%、95.7%;在测试集中分别为 0.975、42.1%、100%、94.8%。结论:基于低剂量胸部 CT 深度学习模型和 QCT 的骨密度测量的一致性和相关性较好,初步实现了自动、快速的基于 LDCT 的骨质疏松筛查,但需扩大患者人群进一步优化和验证。

**【关键词】** 人工智能;骨密度;体层摄影术,X 线计算机;骨质疏松

**【中图分类号】** R-05; R336; R681; R814.4 **【文献标志码】** A

**【文章编号】** 1000-0313(2024)02-0262-05

DOI:10.13609/j.cnki.1000-0313.2024.02.019

开放科学(资源服务)标识码(OSID):



**Bone densitometry measurement based on low-dose chest CT with deep learning model** ZHAO Yu, ZHANG Xiao-lan, ZHENG Chao, et al. Department of Radiology, the First Affiliated Hospital of Wannan Medical College, Wuhu 241001, China

**【Abstract】 Objective:** To evaluate the consistency of bone mineral density (BMD) measurement between the low-dose chest CT based deep learning model and QCT. **Methods:** The present retrospective study was done and analyzed using the clinical and imaging data of 1406 patients who had low-dose chest CT screening and quantitative computed tomography (QCT) at Yijishan Hospital between June 28, 2018 and December 31, 2019. The patients were divided into a test set (985 cases) and a training set (421 cases) at random. Bone segmentation models and internal tissue calibration models were created using a deep neural network, and its efficiency was evaluated using the test set. The correlation and consistency of the two methods were examined using the QCT results as the reference standard and the Spearman correlation coefficient, interclass correlation coefficient, and Bland-Altman. **Results:** In both the training set and the test set, the deep learning model was positively correlated with QCT results in the training set ( $r=0.957, P<0.001$ ) and test set ( $r=0.955, P<0.001$ ). The intra-group correlation coefficient was 0.946 in the training set and 0.945 in the test set. In the training set, the AUC of this model was 0.986, with the sensitivity, specificity, and accuracy of 47.5%, 100%, and 95.7%, respectively. In the test set, the AUC of this model was 0.975, with the sensitivity, specificity, and accuracy of 42.1%, 100%, and 94.8%, respectively. **Conclusion:** The low-dose chest CT-based deep learning model's estimation of bone mineral density shows good consistency and correlation with the

作者单位:241001 安徽,芜湖皖南医学院第一附属医院/弋矶山医院放射科(赵宇、王敏红、洪薇、周运锋);100102 北京,数坤(北京)网络科技股份有限公司(张晓岚、郑超)

作者简介:赵宇(1985—),女,安徽巢湖人,学士,主治医师,主要从事骨肌肉影像工作。

通讯作者:周运锋, E-mail:zhouyunfeng808@163.com

基金项目:科研能力“高峰”骨干 KGF2019G13 XM\_LHJY2022\_05\_13

outcome determined by QCT. We have so far achieved automatic and quick osteoporosis screening using LDCT. For the further optimization and validation, we will increase the patient population.

**【Key words】** Artificial intelligence; Bone density; Osteoporosis; Tomography, X-ray computed

骨质疏松 (osteoporosis, OP) 是中老年人最常见的骨骼疾病, 表现为骨量减少, 易引发脆性骨折<sup>[1]</sup>。随着我国人口老龄化进程加快, 具有高致残率、高致死率的脆性骨折发病率升高, 带来了不可忽视的社会和医疗负担, 因此早期诊断 OP 至关重要<sup>[2]</sup>。双能 X 线吸收检测仪 (dual energy X-ray absorptiometry, DXA) 和定量 CT (quantitative computed tomography, QCT) 是目前最常用的骨矿物质密度 (bone mineral density, BMD) 测量方法, 但两者均对软硬件有一定要求, 且需额外的辐射和费用<sup>[3]</sup>。健康体检或针对其他适应证进行的低剂量胸部 CT (LDCT) 可用于评估体积骨矿物质密度 (volume bone mineral density, vBMD), 近期研究证明了 LDCT 联合 QCT 椎体骨密度测量在骨质疏松筛查中的作用<sup>[4-5]</sup>。但 LDCT 结合 QCT 测量骨密度需要医生在工作站上勾画兴趣区 (region of interest, ROI), 是一项费时、费力的工作。近年来, 人工智能在医学图像领域的发展为 BMD 的测量提供了新的机遇, 既往研究显示深度学习模型可从腹部平扫 CT 中预测腰椎 vBMD, 与 DXA 测量结果具有较好的一致性<sup>[6]</sup>。本研究笔者通过构建以 3D-ResUNet 和 Attention-UNet 为基础的深度学习模型, 从 LDCT 中预测椎体 vBMD, 并评估该模型诊断骨质疏松的性能。

## 材料与方法

### 1. 病例资料

回顾性分析 2018 年 6 月—2019 年 12 月弋矶山医院符合以下标准的体检人群的 LDCT 资料。纳入标准: ①年龄 ≥ 40 岁; ②同时行 LDCT 联合 QCT 椎体 vBMD 测量<sup>[7]</sup>。排除标准: ①图像质量不佳; ②近期注射静脉对比剂; ③脊柱手术史、测量区有严重的骨折、脊柱畸形或植入物; ④合并任何可能影响脊柱的疾病如脊柱肿瘤及其他代谢性骨疾病等。⑤QCT 报告中 vBMD L1/L2 椎体测量值的比值超过 3 个标准差 (SD)。根据身高、体重计算其 BMI 指数 (BMI = 体重/身高<sup>2</sup>)。最终 1406 例患者纳入研究, 男 889 例, 女 517 例, 年龄 40~94 (57 ± 10) 岁。1406 例患者采用随机数表法按 7:3 的比例分配至训练集 (train set) 及测试集 (test set), 分别为 985 例、421 例。

### 2. 数据采集

本研究采用 GE 公司 Optima CT540 机型, 扫描范围包括肺尖到 L2 椎体。扫描参数: 管电压 120 kV,

采用自动毫安技术, 螺距 1.75:1, 扫描视野 (scanning field of view, SFOV) 500 mm, 床高 135 mm; 重建方式: 矩阵 512 × 512, 算法: LUNG, 层厚 1.25 mm、层间距 1.25 mm。因本研究的目的是在 LDCT 中机会性筛查骨质疏松状态, 故未使用骨矿物质等效体模扫描。

### 3. QCT 测量骨密度

LDCT 数据传输至 QCT 工作站 (美国 Mindways QCT PRO 工作站) 进行异步分析处理。所有分析均由经过培训并具有使用 QCT 软件经验的放射科医师进行。操作者选取每个椎体中间层面 (采集 T12~L2 范围中的两个完整椎体, 一般是 L1~L2 椎体) 进行分析测量, 取两者平均值作为测量结果 vBMD<sub>QCT</sub><sup>[7]</sup>。

### 4. 数据处理与模型训练

图像分割: 使用 ITK-SNAP 软件 (版本 3.8.0, <http://www.itksnap.org/>) 对训练集数据进行图像分割。由 1 名具有 5 年经验的放射科医师在不知患者临床信息的前提下逐层勾画胸部的椎骨、锁骨、肩胛骨和肋骨的骨轮廓得到感兴趣区 (region of interest, ROI)。

模型训练: 本研究中构建的模型由数坤网络科技股份有限公司搭建平台并训练, 模型基于深度卷积神经网络 (deep convolutional neural networks, DCNN) 开发, 训练过程分为 4 步 (图 1), 最终得到训练好的骨密度检测模型。第一步, 训练骨分割模型, 以 UNet 作为基本骨架, 骨分割网络在每个卷积层之前加入了残差结构, 构建了 3D-ResUNet。模型的输入为 LDCT 检查的 3D 数据及人工标注的骨轮廓, 输出为胸部的椎骨、锁骨、肩胛骨和肋骨分割结果。第二步, 加入椎骨关键点 Attention-UNet 模型。在 UNet 的输出层之前加入了注意力机制, 使得网络能够对椎骨骨节的关键点进行定位, 输出 T12、L1、L2 椎体的位置。基于上述关键点选取腰大肌, 基于 CT 值选取前腹壁皮下脂肪 ROI。第三步, 去皮质骨处理。基于获取到的椎骨关键点, 分割出对应的椎体子区域并判定其扫描完整性, 通过一个半径为 4 个体素的球体对椎体掩模 (mask) 做侵蚀处理, 去除椎骨的皮质骨, 仅对椎体内的松质骨进行分析。第四步, 基于骨密度与 CT 值的线性关系, 使用脂肪、肌肉作为内部校正参数构建如下函数计算椎体骨密度预测结果 vBMD<sub>AI</sub>。

$$vBMD_{AI} = \frac{MUSCLEd + (SPINEct - MUSCLEct) \times (MUSCLEd - FATd)}{MUSCLEct - FATct}$$

vBMD<sub>AI</sub> 是预测的骨密度值, MUSCLEd、FATd

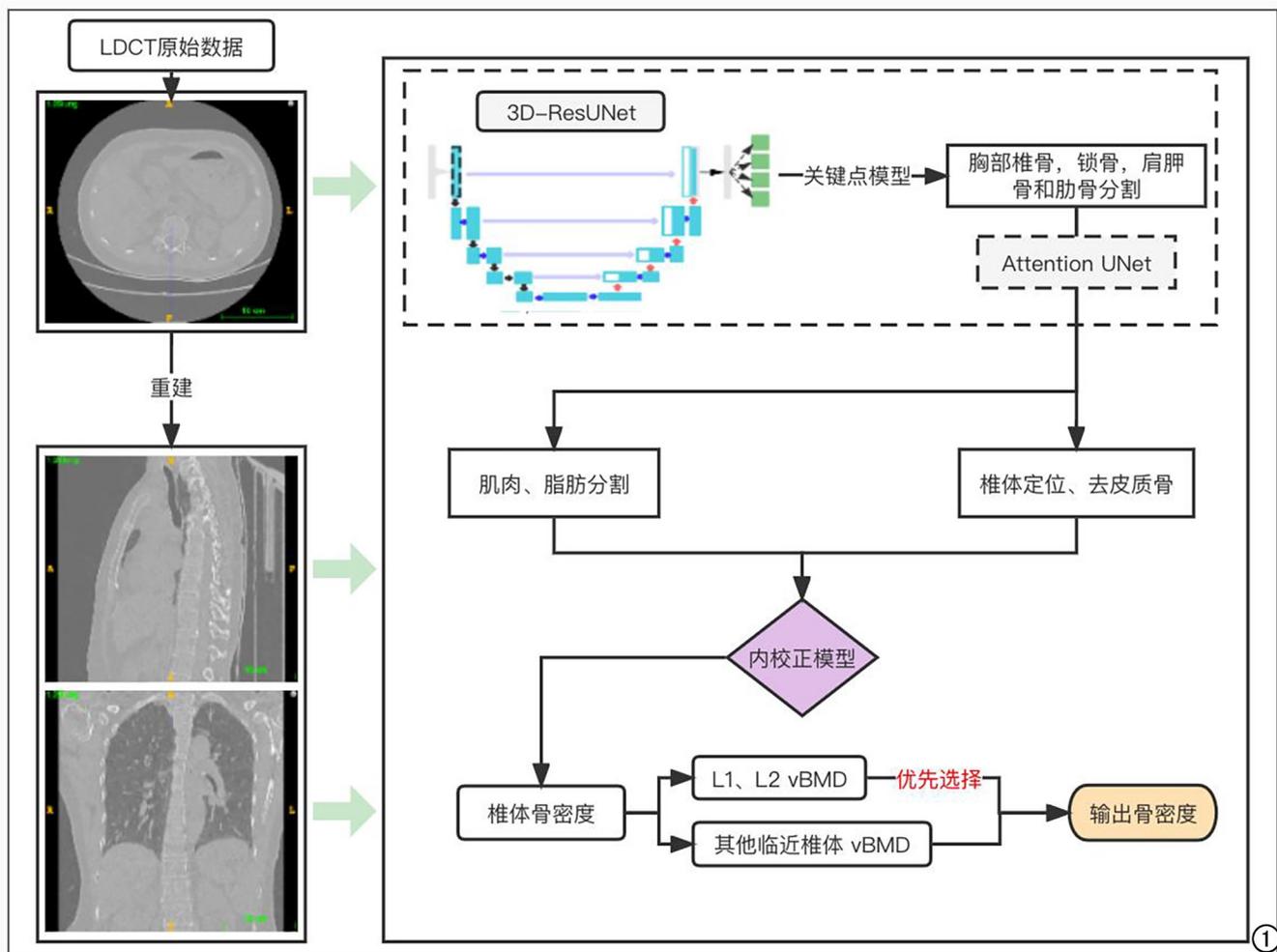


图 1 基于深度卷积神经网络构建骨密度自动测量模型。

分别是肌肉、脂肪密度值, SPINEct、MUSCLEct、FATct 分别是测量得到的椎体、肌肉和脂肪的 CT 值。

## 5. 统计学分析

使用 SPSS 26.0 版本软件进行数据统计。检验水准:  $P < 0.05$  有统计学意义。

基线数据处理: 对连续性变量使用 Shapiro-Wilk 检验进行正态性检验, 使用 Levene 检验进行方差齐性检验, 符合正态分布的资料用  $\bar{x} \pm s$  表示, 非正态分布资料用  $M(Q1, Q3)$  表示。采用独立样本  $t$  检验比较训练集、测试集的临床信息分布。对于定类变量资料采用频数表示, 两组间比较采用  $\chi^2$  检验。

深度学习模型与 QCT 测量的一致性验证: 应用 Spearman 相关系数分别在训练集和测试集中检验 AI 模型输出的骨密度值  $vBMD_{AI}$  和 QCT 测量的骨密度值  $vBMD_{QCT}$  的相关性。应用组内相关系数 (interclass correlation coefficient, ICC) 和 Bland-Altman 法验证两种方法的一致性。应用 ROC 曲线并计算曲线下面积 (area under the curve, AUC) 对 AI 模型诊断 OP 效能进行评估。

OP 的诊断参照《中国定量 CT(QCT)骨质疏松症

诊断指南 2018 版》<sup>[8]</sup>, 以  $vBMD < 80 \text{ mg/cm}^3$  作为骨质疏松症的阳性判定标准。

## 结果

### 1. 一般资料

训练集与测试集中患者间年龄、性别、身高、体重、BMI、QCT 测量的骨密度值  $vBMD_{QCT}$  差异均无统计学意义 ( $P$  值均  $> 0.05$ , 表 1)。

表 1 训练集和测试集中患者间临床资料比较

临床参数	训练集 (n=985)	测试集 (n=421)	统计值	P
年龄(岁)	56±10	57±10	-0.650	0.516
性别[例(%)]			0.021	0.885
男	624 (63.4)	265 (62.9)		
女	361 (36.6)	156 (37.1)		
身高(cm)	163.8±7.9	164.3±7.5	-1.098	0.273
体重(kg)	64.3±10.9	64.6±10.6	-0.492	0.623
BMI(kg/m <sup>2</sup> )	23.8±2.7	23.8±2.7	0.099	0.921
$vBMD_{QCT}$ (mg/cm <sup>3</sup> )	125.6±34.4	125.2±33.9	-0.195	0.845

注: 连续变量以  $\bar{x} \pm s$  表示, 分类变量“性别”的统计值为  $\chi^2$  值, 其他统计值为  $t$  值。

### 2. 骨密度测量

Spearman 相关系数显示 (图 2、3) 训练集中 AI 模

型输出的骨密度值与 QCT 测量的骨密度值呈正相关 ( $r=0.957, P<0.001$ ), 在测试集中 AI 模型输出的骨密度值与 QCT 测量的骨密度值呈正相关 ( $r=0.955, P<0.001$ )。在训练集中组内相关系数(95%置信区间)为 0.946 (0.940~0.952), 在测试集中为 0.945 (0.934~0.954)。Bland-Altman 法分析显示(图 4、5)训练集中 AI 模型与 QCT 测量的方法差值(95%一致性界限)为  $-1.425$  ( $-21.90\sim 19.05$ )  $\text{mg}/\text{cm}^3$ , 0.052% (51/421) 的点在 95% LoA 之外; 测试集中差值为  $-1.23$  ( $-21.51\sim 19.05$ )  $\text{mg}/\text{cm}^3$ , 0.057% (24/421) 的点在 95% LoA 之外。两种方法一致性较好。训练集中 ROC 曲线结果显示 AI 模型输出的骨密度值诊断骨质疏松的 AUC 为 0.986 (95%CI: 0.979~0.993), 测试集中 AUC 为 0.975 (95%CI: 0.953~0.997)。灵敏度为 42.1% (16/38), 特异度为 100% (383/383), 准确度为 94.8% (399/421), 阳性预测值为 100% (16/16), 阴性预测值为 94.6% (383/405) (图 6、7)。

## 讨论

早期筛查并及时采取防治措施是降低 OP 骨折发生率的重要策略, 但我国现有 BMD 测量仪的数量远

不能满足临床需求, 据统计我国 50 岁以上人群中接受过 BMD 检测的比例仅为 3.7%<sup>[9]</sup>。低剂量胸部 CT 体检和新冠肺炎筛查产生了大量 CT 数据<sup>[10]</sup>。

本研究通过构建了一套基于 DCNN 开发的全自动骨密度检测模型, 利用 LDCT 中预测椎体 vBMD 并与 QCT 测量结果进行了比较。研究结果显示该模型与 QCT 测量的骨密度值具有很高的 consistency, 对 OP 具有很高的诊断效能。

DXA 和 QCT 是测量骨密度的主要方式。QCT 采用专用体模和软件测量松质骨的体积骨密度, 不受骨骼体积、形态以及脊柱退行性变的影响, 准确性较 DXA 更高<sup>[11]</sup>。但 QCT 需人工勾画 ROI, 费时费力; 操作者间的一致性难以确定, 需要更高频的专业质控。通过深度学习模型进行 ROI 的自动勾画有助于克服人工操作的误差, 大大减少人工成本和节约时间成本, 测量结果具有可重复性。利用该模型可在患者进行 LDCT 检查时常规筛查 OP, 患者无需额外接受辐射和费用。

一些国外研究<sup>[6,12]</sup>探索了通过 CT 值估算 BMD 的可能性, 报告的相关系数介于 0.399 和 0.891 之间, 发现腰椎 CT 扫描的 CT 值与 DXA 或 QCT 测量结果

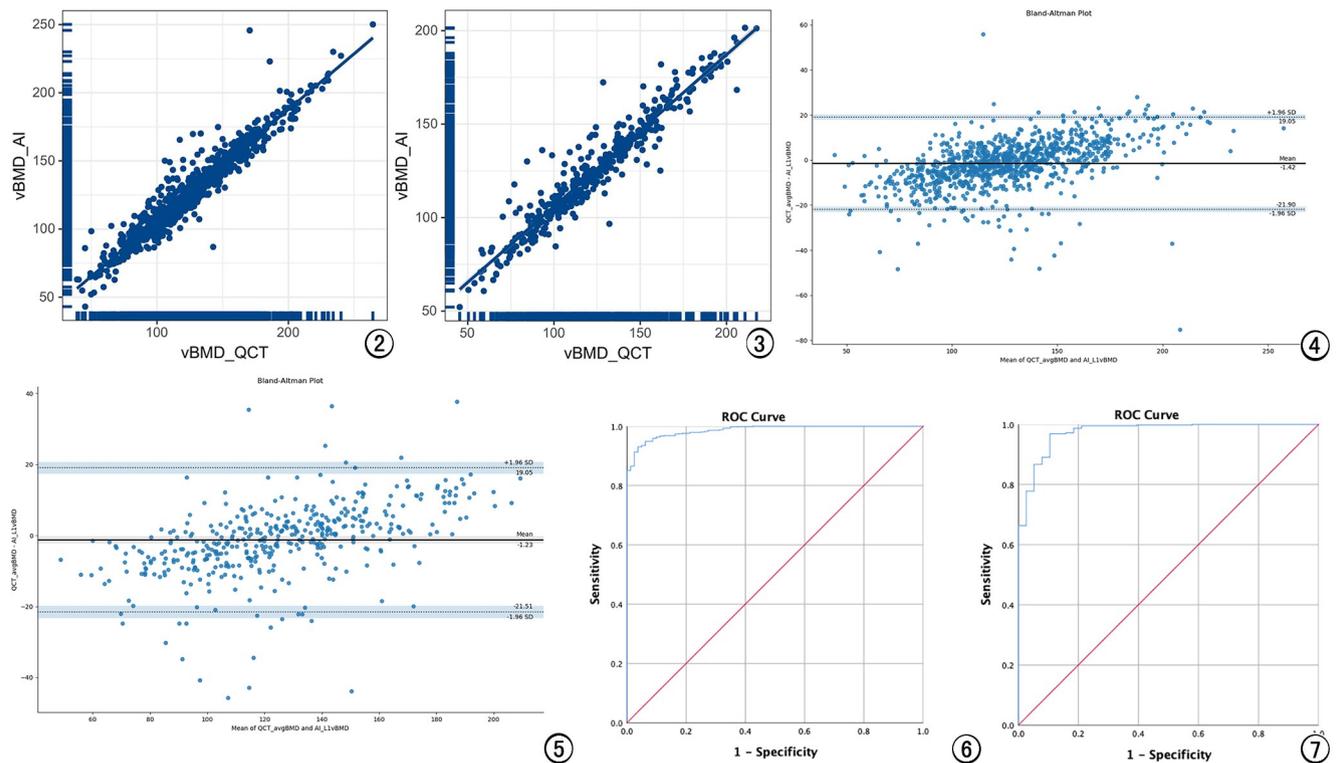


图 2 训练集 ( $n=985$ ), 模型与 QCT 骨密度测量值呈正相关 ( $r=0.957, P<0.001$ )。图 3 测试集 ( $n=421$ ), 模型与 QCT 骨密度测量值呈正相关 ( $r=0.955, P<0.001$ )。图 4 Bland-Altman 分析显示训练集中 AI 模型测量和 QCT 测量骨密度差值为  $-1.43$  ( $-21.90\sim 19.05$ )。图 5 Bland-Altman 分析显示测试集中 AI 模型测量和 QCT 测量骨密度差值为  $-1.23$  ( $-21.51\sim 19.05$ )。图 6 训练集 AI 模型测量的骨密度值诊断骨质疏松 ( $v\text{BMD}<80 \text{ mg}/\text{cm}^3$ ) 的 ROC 曲线, 曲线下面积 AUC 为 0.986。图 7 测试集 AI 模型测量的骨密度值诊断骨质疏松 ( $v\text{BMD}<80 \text{ mg}/\text{cm}^3$ ) 的 ROC 曲线, 曲线下面积 AUC 为 0.975。

有较好的相关性。大部分研究仍以腰椎 CT 为基础<sup>[13]</sup>, 基于低剂量胸部 CT 上进行的研究很少。Kaesmacher 等<sup>[14]</sup>探索了基于校正体模(包括同步和非同步方式)和无体模(基于内部组织校准)的可能, 提出了基于内部组织校准的公式。松质骨无明显界限, 提取过程无法直接进行标注分割, 本研究构建的模型从自动骨分割开始, 椎体定位、去皮质骨到内部组织校准, 最终实现全自动输出 BMD 值。在测试集中 AI 模型诊断 OP 的灵敏度仅为 42.1%(16/38), 这与数据组成有关。本研究中纳入的数据来自体检人群, 患病比例低, 仅有 9%(38/421), 模型后续仍需在患者中进行优化和调整。结果看 ICC 和 Bland-Altman 分析可知 AI 模型与 QCT 测量的骨密度结果具有很高的一致性。模型的阳性预测值、阴性预测值和准确性都很高, 可满足在健康体检中增加 OP 机会性筛查的需求。

笔者对测试集中表现较差的数据进行分析, 发现如下 3 类原因: ①脊柱存在轻度骨质增生的患者亦被纳入, 在松质骨分割时对增生部分过分割, 使测量结果偏大。②患者脊柱旋转或/和曲度较大, 影响椎体分割准确性。③受检者过瘦, 皮下脂肪偏薄, 影响 AI 对脂肪的提取。既往研究发现低剂量胸部 CT 联合 QCT 测量的腹内脂肪面积与血脂生化指标、骨密度均有相关性<sup>[15,16]</sup>, 不除外多或极少的腹内脂肪面积对骨密度测量产生影响。

本研究存在以下局限性: ①为回顾性单中心研究, 且数据来自健康体检人群, 无法代表全部人群, 尤其是存在骨质疏松风险的人群。下一步需多中心大样本数据进一步验证 AI 模型与 QCT 测量值的一致性和对 OP 的诊断效能。②AI 模型尚不能预测一些极端情况, 例如椎体形态变异、腰椎严重退行性病变、皮下脂肪薄等。③本研究以 QCT 测量结果作为参考标准, 可能存在误差, 需要进一步核查 QCT 勾画的 ROI 和体模校正情况。今后的研究中利用深度学习模型有望辅助定量分析腹内脂肪面积等更多指标, 挖掘更多影像、临床信息之间的相关性, 提高预测效能。

综上所述, 本研究通过基于 DCNN 的方法构建了 LDCT 的骨密度自动测量模型, 它不需要额外的设备、特殊的体模和复杂的质控流程, 无需依赖影像医生的经验, 此外, 可以回顾性地获取 vBMD 数据。因此, 该方法可用于扩大骨质疏松症的人群筛查, 特别是在双能 X 线吸收法(DXA)受限的国家或地区, 对增加普通人群的机会性筛查具有重要的意义。

利益冲突: 所有作者没有任何利益冲突, 未接受任何不当的职务或财务利益。

#### 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(2017)[J].中华骨质疏松和骨矿盐疾病杂志, 2017, 10(5): 413-443.
- [2] Si L, Winzenberg TM, Jiang Q, et al.Projection of osteoporosis-related fractures and costs in China:2010-2050[J].Osteoporos Int, 2015, 26(7):1929-1937.
- [3] 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].中华骨科杂志, 2020, 40(16): 1039-1046.
- [4] Wu Y, Jiang Y, Han X, et al.Application of low-tube current with iterative model reconstruction on Philips Brilliance iCT Elite FHD in the accuracy of spinal QCT using a European spine phantom [J].Quant Imaging Med Surg, 2018, 8(1):32-38.
- [5] 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.
- [6] Yasaka K, Akai H, Kunimatsu A, et al.Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network[J].Eur Radiol, 2020, 30(6): 3549-3557.
- [7] 王勇朋, 阳琰, 何生生, 等.低剂量胸部 CT 与 QCT 椎体骨密度测量一站式扫描可行性研究[J].放射学实践, 2018, 33(11):1194-1197.
- [8] 《中国定量 CT(QCT)骨质疏松症诊断指南》工作组, 程晓光, 王亮, 曾强, 等.中国定量 CT(QCT)骨质疏松症诊断指南(2018)[J].中国骨质疏松杂志, 2019, 25(6):733-737.
- [9] 中华医学会骨质疏松和骨矿盐疾病分会.中国骨质疏松症流行病学调查及“健康骨骼”专项行动结果发布[J].中华骨质疏松和骨矿盐疾病杂志, 2019, 12(4):317-318.
- [10] 魏璇, 赵磊, 庞少华, 等.新冠肺炎疫情防控常态化背景下胸部 CT 筛查的实践与思考[J].临床和实验医学杂志, 2021, 20(1): 73-78.
- [11] 包洁, 邹达, 李危石.椎体 CT 值评估腰椎退变患者骨密度的研究进展[J].中国脊柱脊髓杂志, 2020, 30(8):745-750.
- [12] Gausden EB, Nwachukwu BU, Schreiber JJ, et al.Opportunistic use of CT imaging for osteoporosis screening and bone density assessment:a qualitative systematic review[J].J Bone Joint Surg Am, 2017, 99(18):1580-1590.
- [13] Liu L, Si M, Ma H, et al.A hierarchical opportunistic screening model for osteoporosis using machine learning applied to clinical data and CT images[J].BMC Bioinformatics, 2022, 23(1):63-77.
- [14] Kaesmacher J, Liebl H, Baum T, et al.Bone mineral density estimations from routine multidetector computed tomography: a comparative study of contrast and calibration effects[J].J Comput Assist Tomogr, 2017, 41(2):217-223.
- [15] 王硕, 尤凤, 查晓娟, 等.不同性别健康体检者血生化指标、腹部脂肪分布与骨密度的关系[J].中华骨质疏松和骨矿盐疾病杂志, 2020, 13(3):210-215.
- [16] 玄淑兴, 张振国, 贾守强.基于定量 CT 分析不同年龄、生理时期女性腰椎骨密度与腹部脂肪的相关性[J].放射学实践, 2021, 36(5):642-647.