

基于DCNN的椎体分割与骨密度计算方法研究

毛晓晖

衢州市中医医院医务部, 浙江 衢州

收稿日期: 2025年8月14日; 录用日期: 2025年9月17日; 发布日期: 2025年9月25日

摘要

本研究旨在探讨深度学习在原发性骨质疏松症患者中的应用。通过采用基于深度卷积神经网络(DCNN)的全自动方法, 实现椎体分割和CT图像中的骨密度(BMD)计算。采用588例患者作为训练数据, 以及863例患者作为测试数据。首先使用U-Net全卷积神经网络对椎体进行自动分割, 并以手动勾画的椎体区域作为对照。其次采用DenseNet-121卷积神经网络对BMD进行计算, 并与定量CT (QCT)的标准值进行比较。测试集分为三组: 测试集1 (463例)、测试集2 (200例)和测试集3 (200例)。实验结果显示自动分割结果与手动分割结果高度相关, 三个测试集的Dice系数分别可达0.823、0.786和0.782。不同供应商的测试集显示, 自动计算的平均BMD值与QCT结果高度相关(相关系数 $r > 0.98$)。最终研究结论表明, 基于深度学习的方法能够实现CT图像中骨质疏松症、骨量减少和正常骨密度的全自动识别, 具有良好的精确度和一致性。

关键词

深度学习, 脊椎, 骨质疏松症, 骨密度, X射线计算机断层扫描

Research on Vertebral Segmentation and Bone Mineral Density Calculation Method Based on DCNN

Xiaohui Mao

Medical Department of Quzhou Hospital of Traditional Chinese Medicine, Quzhou Zhejiang

Received: August 14, 2025; accepted: September 17, 2025; published: September 25, 2025

Abstract

The study aims to explore the application of deep learning in patients with primary osteoporosis. By adopting a fully automated method based on Deep Convolutional Neural Networks (DCNN),

vertebral segmentation and Bone Mineral Density (BMD) calculation from CT images are achieved. A total of 588 patients were used for training data, and 863 patients were used for testing data. First, a U-Net fully convolutional neural network was employed to automatically segment the vertebrae, with manually drawn vertebral regions as a reference. Next, the DenseNet-121 convolutional neural network was used to calculate BMD and compare it with the standard values of quantitative CT (QCT). The test set was divided into three groups: Test Set 1 (463 cases), Test Set 2 (200 cases), and Test Set 3 (200 cases). Experimental results showed that the automatic segmentation results were highly correlated with manual segmentation, with Dice coefficients of 0.823, 0.786, and 0.782 for the three test sets, respectively. The test sets from different vendors showed that the automatically calculated average BMD values were highly correlated with the QCT results (correlation coefficient $r > 0.98$). The final conclusion of the study indicates that the deep learning-based method can achieve fully automated identification of osteoporosis, osteopenia, and normal bone density in CT images, with good accuracy and consistency.

Keywords

Deep Learning, Spine, Osteoporosis, Bone Mineral Density, X-Ray Computed Tomography

Copyright © 2025 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

骨质疏松症是一种常见的慢性骨病，影响着全球超过 2 亿人口，主要集中在老年群体[1]。骨质疏松症的特征是骨强度下降和骨矿物质密度低。骨脆性的增加导致由于骨组织微结构退化而引发的骨折风险上升[2]。患者及其亲友可能会因为长期失去行动能力而经历身体、情感和经济上的困扰。每三位 50 岁以上的女性中就有一位，每五位男性中就有一位可能因骨质疏松症而骨折[3]。骨质疏松症在风险人群中经常被低估和治疗不足，但通过早期诊断和治疗可以预防这种疾病[4]。经调查超过一半的严重骨折发生在从未进行过骨质疏松症筛查的人群中[5][6]。双能 X 线吸收法(DXA)测试的有限可用性，是导致这种疾病诊断不足和治疗不足的主要原因之一。

此外，使用 X 射线机和其他显微镜捕获的数字图像可通过多种深度学习技术进行探索，以开发可用于大规模筛查的生物医学应用[7][8]。在临床中诊断骨质疏松症的传统方法是测量骨矿物质密度，并对检测结果进行 T 评分，然后将其与健康成年人的密度进行比较。T 评分为 1 表示骨密度正常。T 评分在 1 到 2.5 之间表示低骨密度(骨质减少)。T 评分为 2.5 表示存在骨质疏松症。然而，体内矿物质密度在预测骨折风险方面显示出一定的局限性[9]。由于骨折的不同症状和复杂原因，在没有达到骨质疏松症严重程度的情况下也出现了更多的骨折事件，这使得通过 DXA 评估骨骼质量的准确性较低[10][11]。

医学影像在预测骨质疏松症方面展现了巨大的潜力。多项研究已经探讨了如何从无对照剂头部 CT 扫描中提取纹理特征，并比较了这些特征在骨质疏松症患者和非患者之间的差异，以评估其在机会性筛查中的潜在应用[11][12]。在骨质疏松症筛查中，研究者们还使用了从 CT 扫描中获得的多种三维纹理特征以及区域体积骨矿物质密度[13]。除了 CT 扫描，X 射线和 MRI 扫描也被用于评估因骨质疏松症引起的骨折风险[14][15]。为了提高诊断的准确性和自动化水平，研究者们采用了多种纹理分析方法、局部二值模式以及不同的图像特征变换方法[16]-[18]。

近年来，深度神经网络在计算机视觉领域表现出色。深度学习方法尝试从输入图像数据及其相关类

标签中找出对应的映射关系。由于架构的不断进步和硬件计算能力的快速提升, 神经网络在包括医学影像在内的多种应用中表现出强大的性能[19]-[25]。在 X 射线图像的感兴趣区域, 研究者们使用自编码器和支撑向量机的半监督技术来诊断骨质疏松症[26]。一种基于机器学习的骨质疏松症机会性筛查模型被提出, 该模型结合了人口统计变量、常规实验室测试数据和 CT 图像来检测骨质疏松症[27]。

尽管 CT 扫描更为普及, 但其辐射剂量远高于 X 射线成像。此外, X 射线扫描成本较低, 且在临床环境中由于创伤、疼痛等各种原因拍摄了大量 X 射线片。因此, 开发基于深度学习的方法, 利用 X 射线图像来诊断骨质疏松症具有重要意义。据我们所知, 目前尚无研究利用深度学习全自动定位腰椎椎体并计算类似 QCT 值的骨密度。

本研究提出了一种深度学习算法, 用于在 CT 扫描中自动定位腰椎椎体, 并计算骨密度(BMD), 使其精准度与 QCT 值相似。研究的目的是评估自动腰椎椎体定位方法的准确性及基于深度学习的 BMD 计算性能, 以探讨深度学习在骨质疏松症诊断中的潜在应用。

2. 数据来源与预处理

本研究在 2018 年 3 月至 2020 年 5 月期间, 按照 QCT 检查的登记顺序, 从医院收集了 CT 扫描图像和数据。入选条件为: (a) 包含腰椎的 CT 检查图像, 如腰椎检查和腹部检查; (b) 同意参与本研究。为了消除其他因素的干扰, 使 QCT 测量结果作为 BMD 的参考标准, 我们也设置了一些排除条件: (a) 不符合 CT 检查后处理要求; (b) 患有继发性骨质疏松症, 如因肾功能衰竭、糖尿病和甲状旁腺功能亢进引起的骨质疏松症; (c) 存在 L1-L4 椎体压缩性骨折; (d) 有术后金属或骨水泥植入物。最终, 共有 1449 名患者符合所有标准并被纳入本次研究。其中, 588 名患者用于模型开发的训练队列, 另外三个独立的测试队列用于评估和分析训练后的模型, 每个测试队列分别包含 463 名、200 名和 200 名患者。

所有用于训练的 CT 扫描和 463 个用于测试的 CT 扫描均是通过 128 通道多层探测器 CT 扫描仪(uCT760, 联影医疗)获得的。剩余的 400 个用于测试的 CT 扫描来自两家不同厂商的 CT 扫描仪。CT 参数设置遵循“中国健康定量 CT 大数据项目研究计划”[28], 具体设置如下: 切片厚度为 0.625 毫米; 管电压为 120 kVp; 管电流为自动模式。重建间隔为 1.0 毫米。所有 CT 图像均使用供应商 CT 扫描仪提供的迭代重建算法重新构建为 512×512 矩阵。

本研究中, 我们以 QCT 测量值作为诊断和预测骨密度的标准。所有 CT 图像均经 QCTProModel4 (Mindways Software, Inc.) 进行后处理, 质量控制分析采用统一的欧洲脊柱幻影(ESP, NO.145)。通过选取椎体的中央层计算平均骨密度值。根据国际临床密度学学会(ISCN)和美国放射学会(ACR)的指南, 骨密度值低于 80 mg/cm^3 属于骨质疏松症; 80 到 120 mg/cm^3 为骨量减少症; 超过 120 mg/cm^3 为正常[29][30]。

3. 方案设计

该方法主要分为两个步骤: (1) 进行腰椎体分割, (2) 进行骨密度计算。本研究中, 我们采用深度学习技术来自动完成这一过程。自动计算的准确性基于准确的分割。具体来说, 我们采用 U-Net 来执行分割任务。U-Net 是一种常用的卷积神经网络架构, 特别适用于生物医学图像的分割任务。其独特的 U 形结构能够有效地捕捉和融合多尺度特征信息, 从而提高分割精度。

在完成分割后, 我们需要对腰椎体进行精确定位。这一步骤至关重要, 因为只有精确定位了腰椎体, 才能确保从中提取出的信息是相关且准确的。腰椎体的定位使用分割后的图像, 通过识别腰椎体的边界和形状特征来实现。

接下来, 我们从定位后的腰椎体中提取用于骨密度计算的内部信息。骨密度是衡量骨骼强度的重要指标, 通常通过分析腰椎体的 CT 图像来计算。为了实现这一目的, 我们采用回归模型进行骨密度计算。

回归模型能够根据提取的特征与已知的骨密度值之间的关系，预测出新的骨密度值。

图 1 展示了我们提出的框架概述。首先，我们将 CT 图像输入 U-Net 模型，进行腰椎体的自动分割。然后，对分割结果进行腰椎体的定位处理，提取出腰椎体的内部信息。最后，通过回归模型对提取的信息进行分析和计算，得到骨密度值。整个过程利用深度学习技术，实现了骨密度的自动计算，确保了计算的准确性和效率。

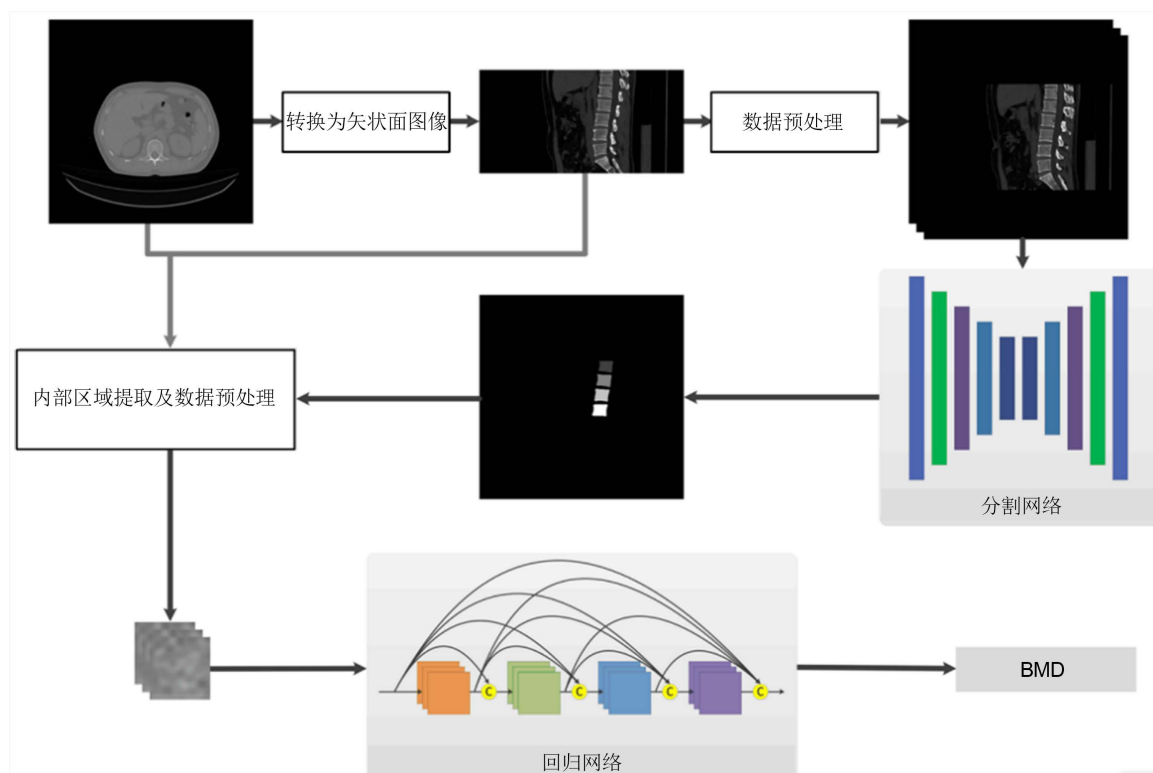


Figure 1. Overall framework

图 1. 整体框架

3.1. 图像预处理

在本研究中，使用分割模型来确定前四个腰椎椎体(L1~L4)的位置。每位患者的 CT 横断面图像被转换为矢状面图像。为了简化这一过程，我们在横断面图像中进行了椎孔检测。具体操作是，给定一张横断面图像，首先通过阈值方法获得二值图像，该阈值依据训练队列中 HU 值的分布进行选择，本研究设定 HU 为 150。接下来，通过一系列操作，包括形态学闭合操作、NOT 操作、背景去除和 3D 去噪处理，最终在二值图像中仅保留椎孔区域。之后，给定一系列患者的 CT 切片，经过以上检测获取每个横断面切片的椎孔中心。所有切片的椎孔中心的平均值被视为脊柱中心，并用于选择矢状面图像。我们最终连续选择包含该平均中心的 10 张矢状面图像作为分割任务的候选图像。

为了减少人工注释的时间消耗，同时确保图像中所有相关椎体(L1~L4)的可见性，我们使用了 20 张连续矢状面图像的平均图像，并在研究平台 InferScholar3.0 (Infervision)上进行椎体轮廓描绘，由多名放射科医生提供数据注释。所有矢状面图像首先由两位放射科医生分别描绘(分别为一年经验的住院医师和三年经验的住院医师)。然后，所有注释由具有七年经验的认证放射科医生审查和修改。这些描绘区域作为分割的真实值。为了使网络能够直接识别四个不同的椎体，在每个分割的真实值中，L1 的标记区域填

充“1”，L2 填充“2”，L3 填充“3”，L4 填充“4”，其他部分填充“0”。

3.2. 分割模型

首先，对所有矢状面图像进行窗口处理，窗口级别(WL)和窗口宽度(WW)分别设定为[350,1000]。接着，采用最小 - 最大归一化方法将图像中所有像素值缩放到[0,1]。最后，将所有图像尺寸调整为 512×512 像素。如果图像尺寸超过这个范围，则围绕中心裁剪；如果尺寸不足，则进行零填充。

由于后续的回归任务依赖于对椎体关键部位的准确提取，因此需要高精度的分割。U-net 在医学图像分析中表现出色，即使在训练数据不足的情况下，也能精确地分割图像。因此，本文采用了带有残差块的残差 U-net (ResU-net)，并最终构建了带有循环过程的循环残差 U-net (R2U-Net)。

与其他卷积神经网络(CNN)相比，R2U-Net 在医学图像分割方面表现突出。该模型由残差单元和循环卷积神经网络组成，其中多个卷积操作共享一个核权重，并进行多次迭代操作。通过逐元操作将输入值添加到相应层的输出值，可以增强特征值的表达能力，从而实现深度结构学习并积累特征值。

该模型包含多个编码器和解码器，每个编码器和解码器都由时间步长设为二的循环卷积 2D 滤波器 (Recurrent Conv2D)、批量归一化(BN)和激活函数修正线性单元(ReLU)组成。编码器通过核大小为 3×3 、步长为 1×1 的 Recurrent Conv2D、BN 和 ReLU 进行处理，并通过最大池化(核大小为 2×2 ，步长为 2×2)捕获上下文信息并减小特征图的大小。解码器通过循环上采样卷积 2D (Recurrent Up-Conv2D)层、BN 和 ReLU 进行处理，通过 2×2 大小的上采样防止空间信息丢失，并与编码器使用的神经网络进行连接。最终，通过使用 Recurrent Conv2D、 1×1 卷积和 Sigmoid 激活函数提取分割图。

在独立测试队列中评估分割模型性能时，我们将分割结果区域与人工真实值注释进行比较，分割性能通过以下方法进行评估：

(a) Dice 相似系数(DSC)，用于测量分割和真实值空间重叠的方法，定义为

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}$$

(b) 阳性预测值(PPV)，用于评估所有阳性预测中 TP 分割数量的方法，定义为

$$PPV = \frac{TP}{TP + FP}$$

(c) 灵敏度，用于评估所有真实阳性中 TP 分割数量的方法，定义为

$$\text{灵敏度} = \frac{TP}{TP + FN}$$

其中 TP、FP 和 FN 分别是真阳性、假阳性和假阴性分割的数量。

3.3. 回归模型

为了进行回归和图像标注，我们从横断面切片中提取三通道图像。首先，在矢状图像上覆盖掩模以确定椎体位置。矢状图像位于 Y-Z 平面，而横断面切片位于 X-Y 平面。通过连通组件操作定位椎体的中心位置，获取中心位置(y,z)，其中 z 坐标对应所需横断面切片的位置，y 坐标对应该横断面切片的相应 y 坐标。接着，从横断面切片中裁剪一条高度为“h”的窄条图像，以 y 坐标为中点，裁剪出的图像大小为 $h \times 512$ 像素。然后使用阈值法将窄条图像转换为二值图像，并保留最大连通区域，该区域位于横断面切片的椎体中。之后，通过处理后的二值图像作为窄条的掩模，从中提取大小为 $h \times w$ 像素的图像，确保提取的图像不包含皮质。在本研究中，“h”和“w”均设为 32。最后对于第一通道和第三通道图像，则分

别从上述横断面切片的前一个和后一个切片的相应位置提取图像。

最终,提取的三通道图像用于回归任务,其中每个通道的图像大小为 32×32 像素。所有提取的图像在输入自动回归模型之前都进行了预处理。处理流程包括: (1) 使用上述窗口将灰度值转换为[0,255]; (2) 使用最小-最大归一化方法将其归一化为[0,1]。并以 QCT 得出的相应 BMD (骨密度)作为参考标准。

在训练集上开发回归模型使用 DenseNet-121 网络结构计算 BMD 值。DenseNet-121 主要包含 4 个密集块和 3 个过渡层。每个密集块由一系列 1×1 卷积层、 3×3 卷积层和跳跃连接组成。每个过渡层包括一个 1×1 卷积层用于减少网络参数和一个 2×2 的平均池化操作用于采样。

按照前面的描述完成图像分割和图像预处理后,样本被送入 DenseNet-121 网络进行训练。使用均方误差(MSE)评估训练过程中的损失。在迭代训练过程中,网络的损失逐渐减小,并在每次训练的第 20 个迭代周期自动保存模型。然后使用测试集样本对保存的模型进行测试,以获取每个测试样本的预测信息。当损失降低到足够低的水平时,网络训练结束。

为了评估模型的拟合优度,我们使用决定系数(r^2)来评价自动获取的 BMD 与参考标准之间的回归性能。此外,为了进一步验证回归模型的有效性,我们计算了平均 BMD,并将其根据 QCT 研究的评估指南分为三类。在测试集上,我们通过平均 BMD 分类的真阳性、假阳性、真阴性和假阴性数量构建了一个 3×3 的列联表,从而形成混淆矩阵。

3.4. 统计分析和软件

在进行统计分析和软件比较时,我们通常使用多种方法来评估数据之间的差异、关联和一致性。对于连续变量,使用均值和标准差来表示数据的中心趋势和离散程度;对于分类变量,则使用频率来表示不同类别在样本中出现的次数。在比较分类变量时,卡方检验用于检验两个或多个分类变量之间的关联性或差异性。对于配对 BMD 值的比较,使用 Wilcoxon 符号秩检验来比较两组相关配对数据的差异,这种方法不依赖于数据的潜在分布。统计显著性的判定标准为 $p < 0.05$,即当计算得到的 p 值小于 0.05 时,表示存在统计显著性差异。

为了评估 BMD 计算的可靠性,使用 Pearson 相关系数来评估两个连续变量之间的线性相关性。同时,Bland-Altman 分析用于比较两种测量方法之间的一致性和差异,通常用于评估同一样本的两次测量结果。为了评估自动回归与参考标准的一致性,使用 Cohen's kappa 系数,该系数用于评估两个分类者之间的一致性,并常用于比较自动分类和参考标准之间的一致性。根据 Cohen's kappa 系数的取值范围,不同的一致性程度包括: 0 到 0.20 为轻微一致性,0.21 到 0.40 为一般一致性,0.41 到 0.60 为中等一致性,0.61 到 0.80 为显著一致性,0.81 到 1 为几乎完美的一致性。

本实验用于构建基于 DCNN 模型的软件运行在 Ubuntu 16.04 操作系统上,使用的是 MXNet 深度学习工具包。训练过程是在 Intel® Core™ i7-5930K CPU 3.50 GHz 和 GeForce GTX TITAN X GPU 上进行的。整个神经网络的实现、评估和统计分析都在 Python 3.7 环境中完成。

4. 结果

共有 1449 名患者符合最终分析条件,年龄从 15 岁到 98 岁不等(平均 53.8 岁)。表 1 列出了训练和测试数据集的临床和人口学特征。根据指南,所有患者分为三类:骨质疏松症($n = 244$,占 16.8%)、骨量减少($n = 605$,占 41.8%)和正常($n = 600$,占 41.4%)。

结果显示训练和测试数据集在年龄特征上没有显著差异($p > 0.05$),但在 BMD 分布上有显著差异($p < 0.01$)。性别和 CT 检查部位在训练数据集和测试集 1 之间没有显著差异($p > 0.05$),但在测试集 2 和测试集 3 之间显示显著差异($p < 0.001$)。

Table 1. Features of training and test sets
表 1. 训练和测试集特征

变量	训练队列	测试队列		
		测试集 1	测试集 2	测试集 3
n	586	463	200	200
年龄	51.7±14.3	56.5±15.1	55.5±13.2	52.3±15.3
性别				
男性	231	197	120	96
女性	355	266	80	104
CT 检查部位				
脊椎	528	401	0	27
其它部位	58	62	200	173
BMD				
骨质疏松症	85	89	31	39
骨质减少症	348	124	77	56
正常的	153	250	92	105

4.1. 在独立测试队列中分割模型的表现

表 2 显示了分割模型的评估结果。结果表明，基于深度学习的自动分割与手动分割在四个腰椎体上具有良好的相关性。

Table 2. Model segmentation results on different test sets
表 2. 不同测试集下的模型分割效果

测试队列	椎体	分割模型			回归模型
		DSC	Sensitivity	PPV	r ²
测试集 1	L1	0.823 ± 0.267	0.836 ± 0.270	0.823 ± 0.270	0.948 (0.930, 0.960)
	L2	0.825 ± 0.266	0.822 ± 0.271	0.843 ± 0.266	0.958 (0.954, 0.962)
	L3	0.862 ± 0.202	0.861 ± 0.216	0.891 ± 0.191	0.949 (0.942, 0.956)
	L4	0.899 ± 0.153	0.900 ± 0.162	0.909 ± 0.153	0.962 (0.959, 0.965)
测试集 2	L1	0.786 ± 0.290	0.787 ± 0.286	0.798 ± 0.301	-
	L2	0.793 ± 0.291	0.783 ± 0.291	0.814 ± 0.305	0.946 (0.937, 0.955)
	L3	0.813 ± 0.273	0.792 ± 0.282	0.870 ± 0.256	0.946 (0.938, 0.954)
	L4	0.883 ± 0.189	0.870 ± 0.197	0.908 ± 0.192	0.963 (0.959, 0.966)
测试集 3	L1	0.789 ± 0.237	0.764 ± 0.240	0.844 ± 0.237	0.780 (0.724, 0.811)
	L2	0.786 ± 0.242	0.759 ± 0.240	0.836 ± 0.250	0.877 (0.862, 0.889)
	L3	0.801 ± 0.226	0.771 ± 0.230	0.871 ± 0.222	0.828 (0.780, 0.905)
	L4	0.782 ± 0.293	0.758 ± 0.293	0.843 ± 0.294	0.886 (0.870, 0.899)

在测试集 1 中，69%的案例表现出分割模型的良好性能，而在测试集 2 和测试集 3 中，这一比例分别为 63.5%和 44%。在四个腰椎体中，有至少三个的 DSC 值超过 0.90。特别是，测试集 1 中有 58%的患

者、测试集 2 中有 50% 的患者以及测试集 3 中有 30% 的患者，其所有四个腰椎体的 DSC 值都达到或超过 0.90。图 2 展示了横断面 CT 图像、放射科医师的手动标注与我们方法的自动分割结果示例。

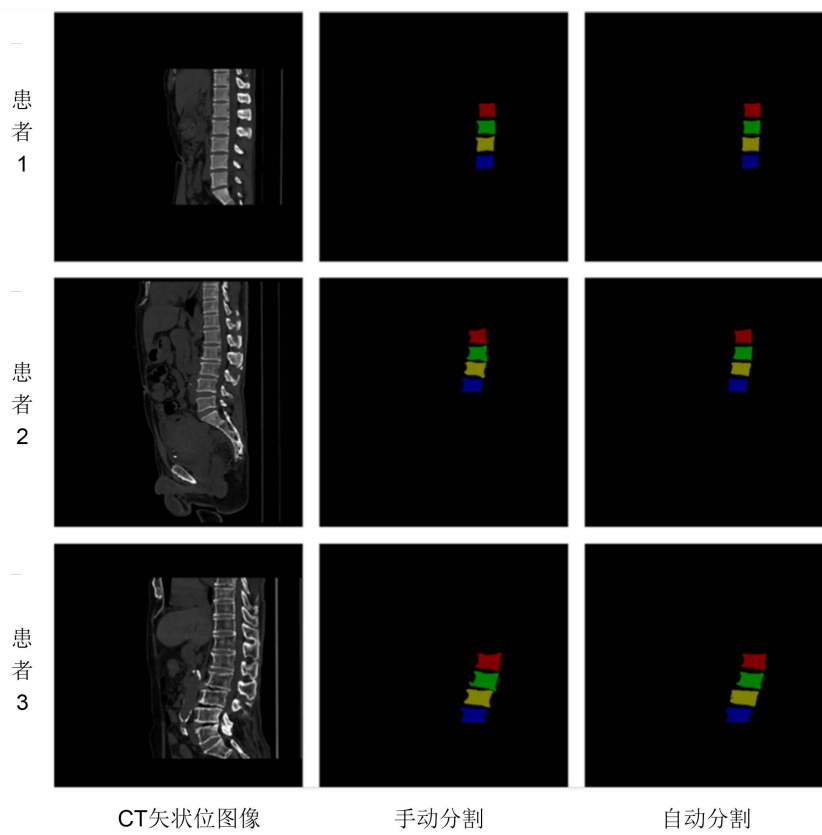


Figure 2. Automatic segmentation and manual segmentation

图 2. 自动分割与手动分割

由于本研究使用的分割模型用于定位每个图像中的腰椎体，当四个 DSC 值均大于 0.5 时，分割结果被认为是有效的。按照这一标准，测试集 1 中约 86% 的案例、测试集 2 中 81.5% 的案例以及测试集 3 中 65.5% 的案例获得了良好的分割结果。

在独立测试队列中，回归模型的性能分析基于部分有效的分割结果进行评估。具体来看，测试集 1 中的 86% (398/463)、测试集 2 中的 81.5% (163/200) 以及测试集 3 中的 65.5% (131/200) 案例均展示了回归模型的良好表现。

我们对回归模型中得到的骨密度(BMD)和通过 QCT 获得的 BMD 进行了比较，结果见表 2。本研究中，每位患者只进行了三个椎体的 QCT 处理，因此平均骨密度(BMD)是指这三个椎体预测 BMD 值的均值。三个测试队列中的平均 BMD 与参考值高度相关(相关系数分别为 $r=0.992$ 、 0.986 和 0.980 ，所有 $p < 0.001$)。通过自动回归得到的平均 BMD 与参考值的一致性为：测试集 1 为 -11.84 至 13.64 mg/cm^3 ，测试集 2 为 -16.8 至 12.6 mg/cm^3 ，测试集 3 为 -8.3 至 29.3 mg/cm^3 。Wilcoxon 符号秩检验显示，测试集 1 中的平均 BMD 与参考标准无显著差异($p > 0.3$)，而在测试集 2 中被低估，在测试集 3 中则被高估($p < 0.001$)。此外，以 QCT 获得的平均 BMD 值作为参考标准，三个测试队列中的预测值与参考标准高度一致(Cohen's kappa 分别为 0.888 、 0.868 和 0.879)。表 3 展示了平均 BMD 的真阳性、假阳性、真阴性和假阴性结果的分类混淆矩阵。

Table 3. Confusion matrix of test set and label
表 3. 测试集与标签的混淆矩阵

		测试集 1			测试集 2			测试集 3		
		骨质疏松症	骨质减少症	正常	骨质疏松症	骨质减少症	正常	骨质疏松症	骨质减少症	正常
真实	骨质疏松症	77	5	1	26	0	0	18	7	0
	骨质减少症	7	97	14	6	51	6	0	25	4
	正常	0	9	188	0	5	69	0	1	76

4.2. 分析

本文介绍了一种基于深度学习的框架，该框架通过分割脊柱或腹部的 CT 扫描图像来自动计算腰椎椎体的骨密度(BMD)值。该框架包括两个主要部分：一个用于分割的 U-Net 模型和一个用于计算 BMD 值的回归网络。首先，分割任务采用 U-Net 模型，从 CT 扫描图像中分割出椎体部分；然后，通过回归网络利用矢状面图像训练来计算椎体的 BMD 值。实验结果表明，深度卷积神经网络(DCNN)在 L1 至 L4 椎体的定位和分割方面表现出色，且 DCNN 计算的 BMD 值与 QCT 方法获得的 BMD 值高度相关。腰椎椎体在矢状面图像中的准确定位是 BMD 测量的基础，测试结果显示，U-Net 模型在四个椎体的平均 Dice 相似系数(DSC)接近或超过 0.8，表明自动分割结果与手动标注高度一致。约 80.2%的患者中，L1-L4 的 DSC 均超过 0.5，但由于结构相似的椎体区分困难、语义分割而非实例分割、标签噪声和训练数据多样性不足等原因，部分分割效果未达预期。

在 BMD 测量方法的比较中，BMD 测量被认为是早期诊断骨质疏松的方法。尽管双能 X 线吸收法(DXA)广泛使用，但易受腹主动脉钙化和脊柱退变的影响。QCT 作为补充，通过几何和小梁骨评估骨质疏松症，但其复杂的后处理设备和工作使得大多数医院未广泛使用。基于 DCNN 的方法则自动计算 BMD 值，解决了 QCT 复杂后处理的问题。测试结果显示，在测试集(测试集 1、测试集 2 和测试集 3)中，自动回归计算的平均 BMD 的 Pearson 相关系数均超过 0.98，表明自动方法获得的平均 BMD 与 QCT 获得的值之间存在强相关性。

尽管自动回归与参考值在不同供应商的 CT 扫描中存在差异，但临床上可接受此差异。差异的主要原因是训练数据的多样性不足。然而，基于简单卷积神经网络的 BMD 估算方法比 QCT 更高效、更便捷，且在机会性筛查中具有显著优势。利用腹部或脊柱 CT 扫描进行“机会性筛查”，可以发现患有其他指征的患者可能伴有骨质疏松或骨量减少。尽管图像重建算法和电压辐射管的多样性限制了 CT 值评估骨质疏松症的能力，但实验结果表明，该方法可以使用常规 CT 扫描检测骨质疏松或骨量减少，有助于早期骨质疏松症的筛查和预防。

5. 研究优势与局限性

5.1. 研究优势

本研究基于深度卷积神经网络(DCNN)方法，实现了腰椎体的自动分割与骨密度(BMD)的精准计算，显示出与定量 CT(QCT)方法高度一致的结果。该方法显著降低了对专用后处理设备的依赖，提升了骨质疏松症筛查的自动化水平与临床可行性，尤其适用于机会性筛查场景，具有一定的临床转化潜力。优势部分保留原意，但表述更简洁、客观。

5.2. 局限性

尽管本研究在椎体分割与 BMD 计算方面取得了良好效果,但仍存在若干重要局限性,需在后续研究与临床应用中予以重视:

1) 本研究训练数据主要来源于单一中心的同一型号 CT 设备,导致模型在面对不同厂商、不同扫描参数的 CT 图像时出现系统性 BMD 偏差。这种偏差在测试集 2 和测试集 3 中表现为 BMD 值的系统性低估或高估,可能影响骨质疏松症的诊断一致性,尤其在临界值附近易导致误诊。

2) 当前模型尚未涵盖严重脊柱畸形、术后植入物或明显钙化等复杂情况,限制了其在真实临床环境中的泛化能力。未来需纳入多中心、多设备、多病种的异构数据,进行大规模外部验证,以提升模型的稳健性与普适性。

3) 本研究采用的语义分割方法虽在整体椎体定位中表现良好,但在区分相邻椎体、避免部分容积效应等方面仍存在不足。语义分割无法区分实例,可能导致椎体间混淆,进而影响 BMD 计算的区域准确性,尤其在单个椎体进行独立评估时可能引入误差。

综上所述,尽管本研究展示了基于深度学习的自动 BMD 计算方法在骨质疏松症筛查中的潜力,但其在实际临床应用前仍需进一步优化数据多样性、模型鲁棒性与跨设备一致性,并开展多中心前瞻性研究以验证其临床效用。

基金项目

本项目得到衢州市科技计划项目“基于卷积神经网络的磁共振图像建构骨质疏松症的分类方法”(项目编号:2023K159)资助。

参考文献

- [1] Kim, W.J., *et al.* (2004) Clinical Outcome of Conservative Treatment for Osteoporotic Compression Fractures in Thoracolumbar Junction. *Journal of Korean Society of Spine Surgery*, **14**, 240-246.
- [2] Alexandru, D. and So, W. (2012) Evaluation and Management of Vertebral Compression Fractures. *The Permanente Journal*, **16**, 46-51. <https://doi.org/10.7812/tpp/12-037>
- [3] Choi, S.H., Kim, D., Koo, J.W., Lee, S.G., Jeong, S. and Kang, C. (2020) Incidence and Management Trends of Osteoporotic Vertebral Compression Fractures in South Korea: A Nationwide Population-Based Study. *Asian Spine Journal*, **14**, 220-228. <https://doi.org/10.31616/asj.2019.0051>
- [4] Kiel, D. (1995) Report Assessing Vertebral Fractures. *Journal of Bone and Mineral Research*, **10**, 518-523. <https://doi.org/10.1002/jbmr.5650100403>
- [5] 高琬俞, 朱新进. 双能量 CT 在评估椎体骨密度中的应用进展[J]. 中国医学创新, 2023, 20(14): 179-183.
- [6] Yüksel, M.O., *et al.* (2016) The Association between Sagittal Index, Canal Compromise, Loss of Vertebral Body Height, and Severity of Spinal Cord Injury in Thoracolumbar Burst Fractures. *Journal of Neurosciences in Rural Practice*, **7**, 57-61.
- [7] Lee, S.H., Lee, S.G., Son, S. and Kim, W.K. (2014) Influence of Compression Ratio Differences between Magnetic Resonance Images and Simple Radiographs on Osteoporotic Vertebral Compression Fracture Prognosis after Vertebroplasty. *Korean Journal of Spine*, **11**, 62-67. <https://doi.org/10.14245/kjs.2014.11.2.62>
- [8] Sadiqi, S., Verlaan, J., Lehr, A.M., Chapman, J.R., Dvorak, M.F., Kandziora, F., *et al.* (2016) Measurement of Kyphosis and Vertebral Body Height Loss in Traumatic Spine Fractures: An International Study. *European Spine Journal*, **26**, 1483-1491. <https://doi.org/10.1007/s00586-016-4716-9>
- [9] Son, K.H., Chung, N.S. and Jeon, C.H. (2010) Measurement of Vertebral Compression and Kyphosis in the Thoracolumbar and Lumbar Fractures. *Journal of Korean Society of Spine Surgery*, **17**, 120-126. <https://doi.org/10.4184/jkss.2010.17.3.120>
- [10] 尹梓名, 张震宇, 胡晓晖, 等. 基于多层次集成学习的骨质疏松辅助诊断研究[J]. 中国骨质疏松杂志, 2022, 28(5): 663-669.
- [11] Leslie, W.D., Schousboe, J.T., Morin, S.N., Martineau, P., Lix, L.M., Johansson, H., *et al.* (2020) Measured Height Loss

- Predicts Incident Clinical Fractures Independently from FRAX: A Registry-Based Cohort Study. *Osteoporosis International*, **31**, 1079-1087. <https://doi.org/10.1007/s00198-020-05313-3>
- [12] Platzer, P., Hauswirth, N., Jandl, M., Chatwani, S., Vecsei, V. and Gaebler, C. (2006) Delayed or Missed Diagnosis of Cervical Spine Injuries. *The Journal of Trauma: Injury, Infection, and Critical Care*, **61**, 150-155. <https://doi.org/10.1097/01.ta.0000196673.58429.2a>
- [13] 宋浩然, 张玉强, 谷娜, 等. 基于 Citespace 对人工智能在骨创伤研究的可视化分析[J]. 中国组织工程研究, 2025, 29(3): 493-502.
- [14] Arpitha, A. and Rangarajan, L. (2020) Computational Techniques to Segment and Classify Lumbar Compression Fractures. *La Radiologia Medica*, **125**, 551-560. <https://doi.org/10.1007/s11547-020-01145-7>
- [15] Ribeiro, E.A., Nogueira-Barbosa, M.H., Rangayyan, R.M. and Azevedo-Marques, P.M. (2010). Detection of Vertebral Plateaus in Lateral Lumbar Spinal X-Ray Images with Gabor Filters. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Vol. 32, 4052-4055. <https://doi.org/10.1109/iembs.2010.5627625>
- [16] Ker, J., Wang, L., Rao, J. and Lim, T. (2018) Deep Learning Applications in Medical Image Analysis. *IEEE Access*, **6**, 9375-9389. <https://doi.org/10.1109/access.2017.2788044>
- [17] Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D. and Chen, M. (2014) Medical Image Classification with Convolutional Neural Network. *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, Singapore, 10-12 December 2014, 844-848. <https://doi.org/10.1109/icarcv.2014.7064414>
- [18] Yadav, S.S. and Jadhav, S.M. (2019) Deep Convolutional Neural Network Based Medical Image Classification for Disease Diagnosis. *Journal of Big Data*, **6**, 113-131. <https://doi.org/10.1186/s40537-019-0276-2>
- [19] 林冰洁, 王梅云. 深度学习在医学影像学中的研究现状及发展前景[J]. 山东大学学报(医学版), 2023, 61(12): 21-29.
- [20] 何伟焘, 陈德春, 王新涛. 人工神经网络在脊柱外科中的应用进展[J]. 临床骨科杂志, 2024, 27(2): 289-293.
- [21] 姜良, 张程, 魏德健, 等. 深度学习在骨质疏松辅助诊断中的应用[J]. 计算机工程与应用, 2024, 60(7): 26-40.
- [22] Park, C., Took, C.C. and Seong, J. (2018) Machine Learning in Biomedical Engineering. *Biomedical Engineering Letters*, **8**, 1-3. <https://doi.org/10.1007/s13534-018-0058-3>
- [23] 段维, 杨国庆, 杨连. 人工智能辅助诊断骨质疏松及预测骨折的研究进展[J]. 中国医药导报, 2024, 21(4): 58-61.
- [24] Al Arif, S.M.M.R., Knapp, K. and Slabaugh, G. (2018) Fully Automatic Cervical Vertebrae Segmentation Framework for X-Ray Images. *Computer Methods and Programs in Biomedicine*, **157**, 95-111. <https://doi.org/10.1016/j.cmpb.2018.01.006>
- [25] Lu, J.T., et al. (2018) DeepSPINE: Automated Lumbar Vertebral Segmentation, Disc-Level Designation, and Spinal Stenosis Grading Using Deep Learning.
- [26] Kim, Y.J., Ganbold, B. and Kim, K.G. (2020) Web-Based Spine Segmentation Using Deep Learning in Computed Tomography Images. *Healthcare Informatics Research*, **26**, 61-67. <https://doi.org/10.4258/hir.2020.26.1.61>
- [27] Lessmann, N., van Ginneken, B., de Jong, P.A. and Išgum, I. (2019) Iterative Fully Convolutional Neural Networks for Automatic Vertebra Segmentation and Identification. *Medical Image Analysis*, **53**, 142-155. <https://doi.org/10.1016/j.media.2019.02.005>
- [28] Pang, S., Leung, S., Ben Nachum, I., Feng, Q. and Li, S. (2018) Direct Automated Quantitative Measurement of Spine via Cascade Amplifier Regression Network. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2018*, Granada, 16-20 September 2018, 940-948. https://doi.org/10.1007/978-3-030-00934-2_104
- [29] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [30] Du, G., Cao, X., Liang, J., Chen, X. and Zhan, Y. (2020) Medical Image Segmentation Based on U-Net: A Review. *Journal of Imaging Science and Technology*, **64**, Article ID: 020508. <https://doi.org/10.2352/j.imagingsci.technol.2020.64.2.020508>