

基于深度神经网络的骨质疏松自动检测方法

毛晓晖

衢州市中医医院医务部, 浙江 衢州

收稿日期: 2025年9月15日; 录用日期: 2025年10月17日; 发布日期: 2025年10月28日

摘要

本研究聚焦于自动检测老年人群中常见的骨质疏松性椎体骨折(OVF)。OVF在老年人中普遍存在, 常导致严重的个人痛苦和社会经济负担。由于早期OVF可能无症状, 导致其常被忽视, 未能及时诊断和报告, 从而影响患者的预后。本研究旨在利用深度卷积神经网络(CNN)从CT扫描片段中提取放射学特征, 并通过特征聚合模块进行最终诊断, 以提高OVF的早期检测率。在方法上, 本研究探索长短期记忆网络(LSTM)对于提升系统诊断性能的表现。本研究在1432个CT扫描(包含10,546个2D矢状视图图像)上训练和评估了系统性能。结果表明, 在129个CT扫描的保留测试集上, 系统达到了89.2%的准确率和90.8%的F1分数。这一表现与实际临床环境中的执业放射科医师相当。本研究开发的自动检测系统通过预先筛查常规CT检查, 并在放射科医生审阅前标记出可疑案例, 有望协助和改善临床环境中的OVF诊断。这项技术的重要性在于, 它不仅能提高诊断的准确率, 还能减少放射科医生的工作负担, 筛查出更多未被诊断的病例, 优化医疗资源的分配。

关键词

深度卷积神经网络, 骨质疏松性椎体骨折, 成像信息学, 体积CT分类, 递归神经网络

Automatic Detection Method for Osteoporosis Based on Deep Neural Networks

Xiaohui Mao

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

Received: September 15, 2025; accepted: October 17, 2025; published: October 28, 2025

Abstract

The study focuses on the automatic detection of Osteoporotic Vertebral Fractures (OVF), which are common in the elderly population. OVF is prevalent among the elderly and often leads to severe

personal suffering and significant social-economic burden. Since early OVF may be asymptomatic, it is often overlooked and not diagnosed or reported in time, which affects patient prognosis. This study aims to utilize Deep Convolutional Neural Network (CNN) to extract radiological features from CT scan slices, and to use a feature aggregation module for final diagnosis to improve the early detection rate of OVF. In terms of methodology, this study explores the performance of Long Short-Term Memory (LSTM) networks in enhancing the diagnostic performance of the system. The system was trained and evaluated on 1432 CT scans (containing 10,546 2D sagittal view images). Results indicate that on the 129 CT scans from the held-out test set, the system achieved an accuracy of 89.2% and an F1 score of 90.8%. This performance is comparable to that of practicing radiologists in a clinical environment. The automatic detection system developed in this study can assist in pre-screening routine CT scans and flagging suspicious cases before radiologists' review, potentially aiding and improving OVF diagnosis in clinical settings. The significance of this technology lies in its ability to not only improve diagnostic accuracy but also reduce radiologists' workload, screen more undiagnosed cases, and optimize the allocation of medical resources.

Keywords

Deep Convolutional Neural Networks, Osteoporotic Vertebral Fractures, Imaging Informatics, Volumetric CT Classification, Recurrent Neural Networks

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

骨质疏松症是一种在老年人群中常见且高发的疾病[1]。据统计,全球约有2亿人患有骨质疏松症,每年因此导致的骨折病例高达8900万例[2]。骨质疏松症主要表现为骨矿物质密度(Bone Mineral Density, BMD)和骨强度的下降,是一种严重的骨骼代谢疾病[3],可能引发腰痛、椎间盘退变及增加椎体骨折的风险[4]-[9]。因此,早期诊断骨质疏松症对于疾病的预防至关重要。

目前,评估骨质疏松症的方法主要包括双能X线吸收法(DXA)、定量计算机断层扫描(QCT)和定量超声(QUS),以及一些新兴的成像技术,如双层能量CT[10]、质子磁共振波谱(HMRS)[11]和正电子发射断层扫描(PET)[12]。早期诊断骨质疏松症的可靠方法之一是BMD的测量。DXA是一种常用的测量脊椎BMD的工具[13],其测量结果包括皮质骨和松质骨的总和,并考虑了二维结构。然而,DXA无法消除皮质骨、过度成骨和硬化对BMD测量的影响[14],可能导致实际骨量丢失的低估[15]。QCT是一种公认的三维骨密度评估方法[16],多项研究[17]-[20]表明,QCT在骨质疏松症检测中的灵敏度显著高于DXA。然而,QCT需要复杂的校准和标准化软件,并且其辐射剂量相对较高,这限制了其作为筛查技术的广泛应用。

每年,有数以千计的患者因尿路或消化系统疾病进行的CT扫描涵盖了脊柱的一部分或全部。这些CT扫描可以用于机会性骨质疏松症筛查,无需额外的放射暴露和高昂的成本[21]。多项研究表明,通过直接测量椎体松质骨的CT值来估算BMD,其常规诊断CT扫描的相关系数在0.399到0.891之间[22]。然而,CT值不仅取决于椎体内部因素,还受到设备、X射线管电压和CT设备等外部因素的影响[23]。因此,不同设备的CT值需要校准,这限制了常规CT扫描在骨质疏松症诊断中的应用。

深度学习在医学成像分析中的应用日益增多,并已进入快速发展阶段[24]-[28]。在骨质疏松症领域,已有几项关于深度学习技术应用的研究。Lee等人[29]结合机器学习和深度学习,通过整合脊柱X射线图

像预测 BMD 异常的患者。Pan 等人[30]开发了一个基于深度学习的系统,利用肺癌筛查的低剂量胸部 CT 扫描自动测量 BMD,用于机会性骨质疏松症筛查。然而,在这个系统中,3D 卷积神经网络模型仅用于将所有椎体分为三类,而每个椎体的隔离和标记则由传统的图像处理算法完成。Yasaka 等人[31]利用 DXA 获得的 BMD 值,开发了一种深度学习模型,从未增强的腹部 CT 图像预测腰椎的骨密度。然而,这项研究仅关注 BMD 预测,而不涉及椎体位置信息的提供。

先前的骨质疏松性椎体骨折(OVF)检测方法需要对每个椎体进行分割和计算高度损失比,这些步骤繁琐且效率低下。为了提高检测效率,本研究开发了一种基于端到端深度学习模型的自动化 OVF 检测系统,免去了对每个椎体进行多次分割和分析的必要性。该系统利用卷积神经网络(CNN)从胸部、腹部和骨盆的 CT 检查中提取放射学特征,并通过序列分类器将提取的特征序列进行聚合,以预测 CT 扫描中是否存在椎体骨折。研究的目的是找到一种有效的特征聚合方法来检测 OVFs,并将模型性能与机构内放射科医生的性能进行比较。本文接下来将详细描述所提出系统的设计和评估过程,旨在证明该自动化系统不仅能简化流程,减少人工参与,还能通过自动化减少误差,提高诊断准确性。端到端的学习模型从输入数据直接输出结果,减少了中间转换和处理步骤,使得整个检测流程更加高效和直接。通过与放射科医生的性能比较,可以验证该自动化系统在临床应用中的实用性和有效性。同时,本文提供设计系统的详细信息和评估结果,有助于同行评审和未来研究。

2. 方案设计

我们开发了一个用于骨折型椎体压缩性骨折(OVF)检测的系统,该系统采用了混合的神经网络架构,结合了卷积神经网络(CNN)和递归神经网络(RNN)。图 1 展示了基于 RNN 的 OVF 检测系统的整体流程。其可分为两个主要步骤:

首先是特征提取阶段,我们采用了深度残差网络(ResNet),从二维 CT 切片中提取高级特征。ResNet 在图像检测任务中表现优异[32],因此成为了我们特征提取的理想选择。

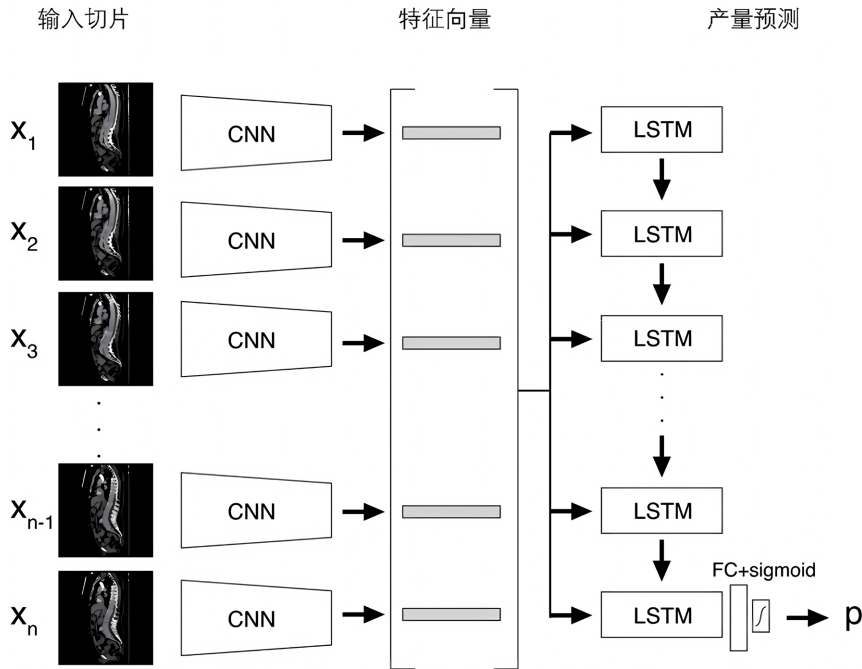


Figure 1. CNN/LSTM OVF detection system
图 1. CNN/LSTM OVF 检测系统

其次是特征聚合阶段，我们引入了长短期记忆网络(LSTM)，这是一种先进的 RNN 架构，用于处理从多个 CT 切片中提取的特征序列。LSTM 能有效处理顺序数据，特别适合整合多个切片的信息[33]，从而作出最终的诊断决策。

2.1. 特征提取

本文提出了一种基于卷积神经网络(CNN)的创新方法，用于自动提取计算机断层扫描(CT)图像中的特征，以辅助临床诊断。我们采用了深度残差网络 ResNet34 作为特征提取网络，并对其进行了适应性修改，以提升网络的诊断性能和训练效率。

传统的特征提取过程通常依赖于放射科医生的专业知识，这不仅耗时且容易产生人为误差。为了克服这些问题，我们选择使用 CNN 来自动进行特征提取，以提高效率和可重复性。具体而言，我们选择了 ResNet34 作为特征提取网络，因为其通过引入残差连接有效缓解了深度神经网络中的梯度消失问题，从而允许训练更深的网络。在预实验中，ResNet34 表现出与更深网络相似的性能，但训练消耗的时间更短，因此我们将其作为此研究的最佳选择。

在网络架构方面，ResNet34 的标准结构包含 33 个卷积层和 1 个全连接层。为了适应二分类任务并降低特征表示的维度，我们对其进行了改动，用两个全连接层(FC 层)替代了原始结构中的单个 FC 层。这种配置将特征从 512 维降至 32 维，并最终输出一个标量值。

此外，为了增强网络的性能，我们在两个全连接层之后加入了批量归一化和 ReLU 激活函数。批量归一化能够减少内部协变量偏移，提高训练稳定性并加速收敛；而 ReLU 函数作为非线性激活函数，可以增加网络的非线性能力，同时不影响残差结构的特点。

在输出层，我们采用了 Sigmoid 函数将全连接层的输出标量转换为介于 0 和 1 之间的概率值，适用于二分类任务。同时，使用二元交叉熵损失函数来衡量预测概率与目标分布之间的差异，从而指导网络在训练过程中改进其预测效果。

由于 CT 扫描通常包含多个切片，本文采用了标签传递策略，即将包含 CT 信息的标签(例如，是否存在某种疾病)传递给所有的切片。每个切片都被赋予与整个 CT 扫描相同的标签。这一策略简化了训练过程，并确保整个 CT 信息被用于训练，而不仅仅依赖单个切片。训练采用的损失函数可以表示为：

$$loss(\hat{y}, y) = -\frac{1}{n} \sum_i (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (1)$$

2.2. 特征聚合

本文通过特征聚合模块利用每个 CT 切片的特征序列来进行患者级别的诊断。由于每次 CT 检查具有不同数量的切片，因此我们的特征聚合器需能够处理任意数量的切片。

基于 RNN 的模型(CNN/RNN)是一种单层 LSTM 网络，该网络包含 256 个隐藏单元，随后是一个全连接层和一个 S 型激活函数层。每个 CT 切片中的 32 维特征向量依次输入网络，从第一个切片到最后一个切片。隐藏层的输出是一个 256 维的向量。最后一层的输出通过全连接层和 S 型激活函数层生成最终的预测结果(如图 1 所示)。最终的 S 型激活函数层产生一个在 0 到 1 之间的单一归一化值，代表 CT 扫描上 OVF 的概率(即置信评分)。我们将置信评分的阈值设定为 0.5，这便于区分阴阳性检测。

3. 实验

3.1. 标准数据集获取

本研究的数据集涵盖了 2008 年 4 月 16 日至 2017 年 1 月 9 日期间在衢州市某医院进行的所有 CT 检

查。这些数据通过服务器进行存储和检索。选择如此丰富的数据的目的是获取一个大型且历史数据较为完整的样本集，从而增加研究的可靠性和数据的代表性。

通过服务器的搜索功能，我们基于检查代码选取了胸部、腹部和盆腔的 CT 检查，并利用高级搜索功能根据特定标准选取 OVF 的阳性和阴性案例。

1) 阳性数据集

通过搜索包含“压缩性畸形或压缩性骨折”等术语的 CT 放射报告获取。

2) 阴性数据集

通过搜索不包含或否定阳性术语的 CT 放射报告，并从中随机选择与阳性相同数量的案例，同时匹配年龄和性别。

这种方法确保了样本中的阳性和阴性数量均衡，同时通过匹配年龄和性别，减少了这些变量可能带来的偏差，从而提高分析结果的准确性。我们将符合条件的案例的矢状位重建图像从 PACS 系统转移到本地加密服务器，同时排除了数据不完整和图像质量差(如有显著硬件或运动伪影)的检查。确保分析的图像数据具有足够的质量和完整性，是进行后续机器学习模型训练和测试的基础。

数据提取的详细过程如图 2 所示，筛选后的 1432 个 CT 扫描结果包含 713 例阳性以及 719 例阴性。总共被分为 80%的训练集、10%的验证集和 10%的测试集。这种划分方法旨在通过训练集和验证集来调整模型参数，并在开发完成后使用测试集评估模型在未见数据上的泛化能力。

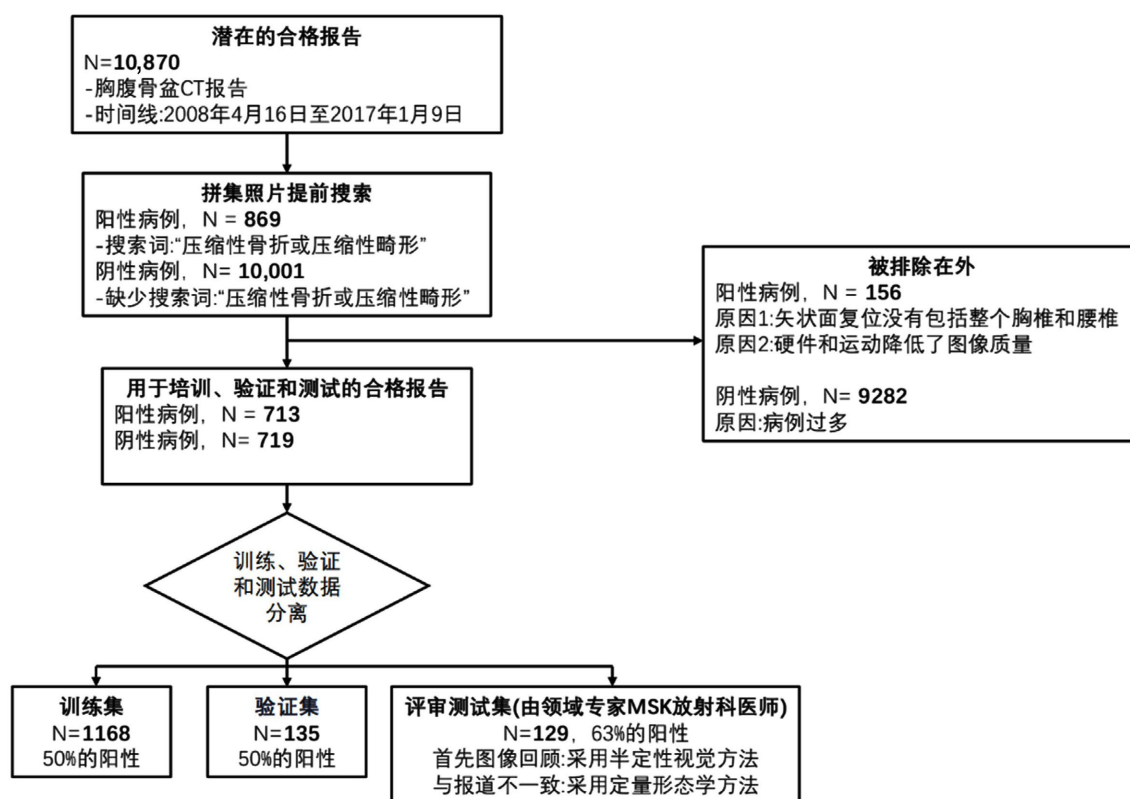


Figure 2. Data collection flow chart

图 2. 数据收集流程图

训练和验证集的标记是基于放射报告中的阳性/阴性标准，而测试集的评估则采取了额外的领域专家放射科医生的半定量和定量复评方法。

使用放射报告的原始标签最大限度地反映了现实世界中放射科医生的实践，从而增加了模型实际应用的相关性。尽管放射报告中可能存在漏诊或报告不规范的情况，但这种做法仍能较好地指导模型的训练。与此同时，为了更准确地评估模型在实际应用中的表现，我们通过专业放射科医生对测试集进行复评，确立了更加可靠的参考标准。

3.2. 数据预处理

为了降低存储空间的需求，我们将 CT 扫描从 DICOM 格式转换为 JPEG2000 格式，并以 10:1 的压缩比进行压缩。选择 JPEG2000 格式是因为它在高压比下仍能保持视觉上无损的图像质量，这有助于在减少数据量的同时，保留必要的图像细节[34]。

此外，为了消除来自不同扫描仪和患者间的变异，提高模型的泛化能力，我们对 CT 切片进行了逐像素平均值减法操作以归一化强度值。通过这种归一化处理，模型能够更专注于图像内容本身，而不受图像亮度等因素的干扰。

值得注意的是，在预处理过程中，我们仅提取了 CT 扫描中心 5% 的切片。因为这部分切片通常包含了判断椎体骨折的关键信息。此方式不仅可以减少无关信息对模型的干扰，还能大幅缩减训练和推断时的计算量。

为了提升模型对图像轻微变化的鲁棒性，我们仅采用了随机旋转和水平平移作为数据增强手段。经过实验验证，其他数据增强技术(如弹性变形、镜头畸变和随机噪声添加)[35][36]在验证集上并未带来显著的性能提升，因此并未采用。

为了适应 ResNet34 网络的输入格式，我们将 CT 切片从 512×512 下采样到 224×224 ，并复制到三个颜色通道。与其他的中心采样和缩放相比本文方法在验证集上展示了更好的泛化性能。这一步骤确保了输入数据与网络结构相适应，同时尽量保留了关键的图像信息。

3.3. 模型训练

3.3.1. 特征提取训练

ResNet34 模型包含两个全连接(FC)层，我们采用二元交叉熵损失函数和随机梯度下降算法进行训练。使用倒数第二个 FC 层输出的 32 维特征向量作为后续 LSTM 分类器的输入。

初始学习率设定为 0.01，批量大小为 192，并通过观察验证集上的损失不降低时，学习率减半的策略来帮助模型更好地收敛。Adam 优化器的参数设定为 $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ ， $\epsilon = 1e-8$ 。为减轻过拟合现象，我们应用了数据增强和权重衰减($1e-4$)来稳定训练过程。训练模型共进行 400 个周期，之后我们冻结了批量归一化层的所有参数，以确保在模型训练到一定程度后不再对特定层进行调整，从而保持特征提取的稳定性。这些步骤和选择确保了我们在特征提取阶段获得了最优的特征表示。

3.3.2. 特征聚合训练

为了对提取的特征进行聚合，我们使用了一个 LSTM 层加上一个全连接层和一个 Sigmoid 层构成的 RNN 聚合模型。LSTM 网络配置有 256 个隐藏单元，这一设置是通过预实验选择的最佳性能配置。需要注意的是，聚合网络的训练不是基于单个 CT 切片，而是基于从 CT 扫描中部提取的所有切片。

我们将最终诊断依赖于最后 Sigmoid 层的置信度得分。初始学习率设定为 0.1，训练过程中应用了数据增强，并使用了 dropout 技术($p = 0.2$)以提高模型的泛化能力，减少过拟合。

在聚合训练过程中，我们进一步微调了特征提取网络的 FC 层，微调学习率设为 $1e-4$ 。训练共进行了 600 个周期，随后在训练集和验证集的组合上进行微调，学习率调整为 $1e-6$ ，并额外进行了 50 个训练周期。这一系列的训练和微调步骤确保了模型能够在不同阶段逐步优化，最终达到最佳性能。

选择 LSTM 进行特征聚合的原因在于其能够处理序列数据，特别适合处理一系列的 CT 切片。隐藏单元数量的选择平衡了模型的复杂度和性能，使得模型在处理 CT 扫描的连续切片时表现出色。初始学习率和微调学习率的不同设置则帮助模型在训练初期快速收敛，在后期进行精细调整。

通过上述系统化的特征提取和聚合训练步骤，我们确保了模型能够高效地处理和分析 CT 扫描数据，并在椎体骨折检测任务中表现出色。

3.4. 测试集筛选

在测试集的筛选过程中，我们采用了半定量和定量两种方法来确立参考标准。具体而言，由一名具有超过 18 年肌肉骨骼放射学经验的认证放射科医生对测试集进行了盲评。这种评估方式确保了评估的客观性，避免了放射科医生受到原始诊断结果或系统结果的影响，从而减少了潜在的偏见。半定量评估基于视觉检查，这是一种快速且常用的方法，利用经验丰富的放射科医生的视觉检查来初步评估测试集。这种方法能够快速筛选出明显的阳性和阴性案例，提供一个初步的评估结果。当半定量评估和盲审下的 OVF 诊断或我们系统的结果之间存在分歧时，我们采用了定量形态测量法[37]进行了额外的调查，定量评估则基于对椎体尺寸的精确测量，计算高度损失比例。这种方法通过提供精确的测量数据，提供了一个更为客观的评估标准。

3.5. 模型评估

在系统性能评估阶段，我们将系统输出的每个 CT 扫描的概率值作为诊断结果。具体来说，我们使用 0.5 作为判断阳性或阴性的阈值。这一选择是基于 0.5 作为自然的分界点，表示阳性和阴性概率相等，是一个中立且常见的分类阈值。为了全面评估系统的性能，我们采用了标准的机器学习评价指标，包括准确率、精确率、敏感性、特异性和 F1 分数。这些指标能够从不同方面反映系统的性能，这些评估指标的定义如表 1 所示。

4. 结果

提出的模型在 129 个 CT 扫描的骨折检测测试集中表现优异，具体数据参见表 1。我们将模型的性能与 DHMC 放射科医生的诊断结果进行了对比。表 1 中的数据显示，CNN/RNN 组合在准确性和 F1 分数上均达到了最佳水平。此外，我们发现该模型的准确性和 F1 分数与 DHMC 放射科医生在临床实践中的骨折检测表现相当。

Table 1. Comparison of model performance
表 1. 模型性能对比

模型	精确率(%) (TP + TN)/ (TP + FP + FN + TN)	准确率(%) TP/(TP + FP)	召回率(%) TP/(TP + FN)	特异性(%) TN/(TN + FP)	F1 分数(%) 2TP/(2TP + FP + FN)
放射科医生报告	88.4 (81.5~93.3)	100.0 (97.2~100)	81.5 (73.6~87.7)	100.0 (97.2~100)	89.8 (83.4~94.5)
CNN/最大池化	81.4 (73.6~87.7)	83.5 (76.2~89.6)	87.7 (80.6~92.7)	70.8 (61.9~78.2)	85.5 (78.0~90.9)
CNN/平均池化	87.6 (80.6~92.7)	95.8 (91.2~98.7)	84.0 (76.2~89.6)	93.7 (88.1~97.3)	89.5 (82.5~93.9)
CNN/投票机制	88.4 (81.5~93.3)	97.1 (92.3~99.1)	84.0 (76.2~89.6)	95.8 (91.2~98.7)	90.1 (83.4~94.5)
CNN/LSTM	89.2 (82.5~93.9)	97.2 (92.3~99.1)	85.2 (78.0~90.9)	95.8 (91.2~98.7)	90.8 (84.3~95.1)

为了进一步研究我们方法的灵敏度和特异性，我们研究了系统的测试集工作特征(ROC)，体现了模型在不同阈值下的性能，区间为 0 到 1。我们还将放射科医生的表现纳入了考虑。尽管所有方法的 ROC 曲

线下面积(AUC)均高于 0.9, 但我们观察到 CNN/LSTM 方法的 ROC 曲线与放射科医生的表现相似。结果表明, 我们的 CNN/LSTM 方法在诊断骨质疏松性椎体骨折(OVFs)方面表现出很高的准确性, 其性能与专业放射科医生相当。

此外, 我们还采用了一种遮挡可视化技术[38]来验证我们系统识别 CT 扫描上骨质疏松性椎体骨折特征的相关性。图 3 展示了这种可视化技术的四个示例, 这些图像上的投影色彩图突出了每个 CT 切片上检测到的骨折。

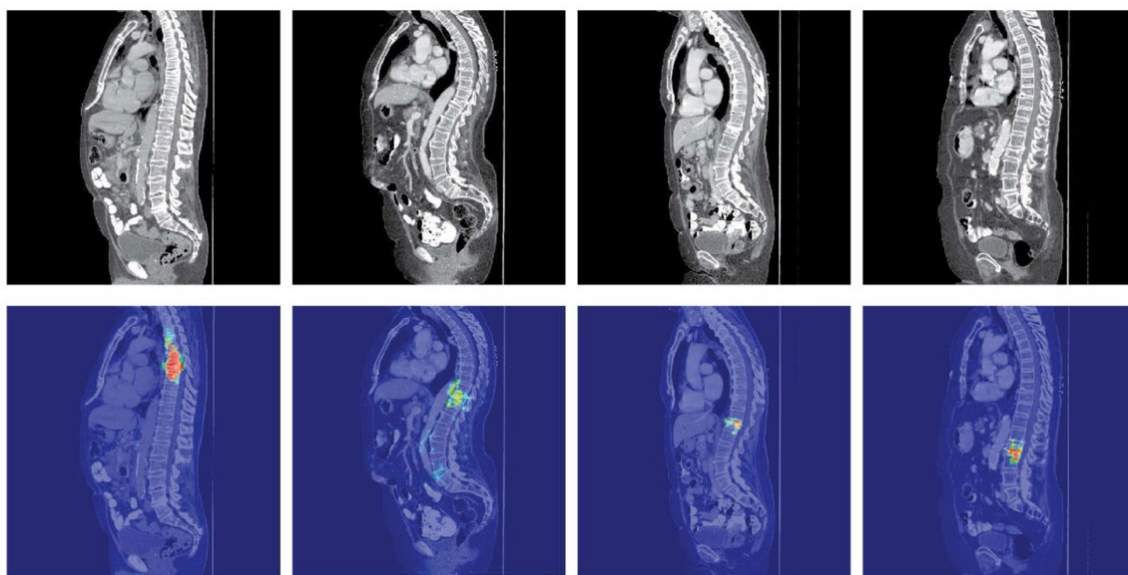


Figure 3. Performance of visualization technology in diagnostic atlases

图 3. 可视化技术在诊断图集上的表现

5. 研究回顾与分析

本研究构建了一个用于检测胸部、腹部和盆腔 CT 检查中骨质疏松性椎体骨折(OVF)的深度神经网络模型。该模型在测试集上的性能与实际放射科医生相当, 并且在准确性和 F1 分数等多个评估指标上表现出色。我们使用了一种基于卷积神经网络(CNN)的最佳 OVF 检测方法, 先从每个 CT 切片中提取特征, 再结合递归神经网络(RNN)序列分类器聚合这些特征, 从而基于整个 CT 扫描进行诊断。定性研究表明, 专家领域放射科医生的结果可视化及相关颜色图显示这些特征在诊断任务中是准确的。

本文采用的 RNN 聚合模块基于最新的长短期记忆网络(LSTM)架构, 能够处理一系列 CT 切片的高维特征, 并进行基于整个 CT 扫描的综合分析。与基于切片级决策的多种规则特征聚合方法相比, 此系统通过 RNN 聚合实现了最高的诊断准确性。通过模型的操作点而非 AUC 来比较模型与放射科医生的性能, 结果显示, 我们的 CNN/LSTM 模型在 OVF 检测任务中具有良好的准确性和 F1 分数平衡。

在评估过程中, 本文采取了严格措施以避免引入任何潜在的偏差。在测试集患者中, 我们提取了放射科报告中的诊断, 以便将本文方法与实际临床环境中的放射科医生表现进行比较。半定量和定量方法的结合, 建立了所有测试集病例的客观参考标准。与通常的评估策略相比, 此方法更真实地反映了放射科医生在临床实际中的表现。

在测试集中, 我们发现放射科报告诊断的差异主要集中在假阴性病例, 即原始放射科医生漏诊的病例。这些漏诊和欠报告可能由于“人员短缺和/或工作量过大”或“放射科医生对骨质疏松症信息缺乏”

所致。此外，OVF 通常是无症状的，不是进行 CT 的主要原因，忙碌的放射科医生为了节省时间，可能会在口述期间忽略这些偶然发现。尽管我们针对此情形作出了一些限制条件，但诊断偏差和专家间的变异性仍可能影响我们方法的性能和普适性。

最后使用卡方检验比较了系统在测试集上与放射科医生的性能。结果显示，尽管我们的系统优于放射科医生的表现没有统计学上的显著性，但至少达到了实践放射科医生的检测水平。

近年来，基于深度学习的 OVF 自动检测方法逐渐兴起，主要分为两类：基于分割的传统方法(需逐椎体定位与高度测量)与基于整体 CT 扫描的端到端分类方法。本研究属于后者，优势在于无需椎体分割，简化了流程，降低了人工干预。

6. 结论

本研究提出了一种不同于以往的方法：1) 整个系统在统一的深度神经网络框架中进行端到端训练和操作；2) 最终诊断基于 CT 总量，而非单个切片或小切片补丁。这对于脊柱变形(如脊柱侧弯)患者尤为重要。此外，我们的方法在高性能计算机上分析每个 CT 扫描的时间从 5 分钟缩短到不到 0.02 秒。因此，与以往的方法相比，我们的整体深度学习方法提供了一个快速、高效且准确的诊断工具。

7. 局限性及展望

本研究的局限性主要体现在以下几个方面。首先，我们的方法仅在一个学术机构的数据上进行了开发和评估，尚未在不同机构的数据上验证其普适性。因此，为了进一步验证模型的泛化能力，未来的工作需要其他机构的数据上进行研究和评估。其次，我们当前的分类架构仅考虑整个 CT 扫描的单一标签，这可能会导致混杂因素的学习，从而影响诊断的准确性。为了解决这个问题，我们计划开发基于深度学习的分割架构，以提高诊断的精确度和可靠性。此外，当前的架构不提供关于骨折位置和级别的重要诊断信息。我们预计，未来的分割框架将有助于提供这些关键信息，从而增强诊断的全面性和细致性。再者，目前的方法依赖于一种启发式的数据提取方法，选取中间 5% 的切片来加速训练和提高模型效率。未来，我们将探索使用视频动作识别任务中的时空方法，进一步优化这一数据提取步骤，并提高模型的整体性能。

未来，我们计划将这一方法扩展到胸部、腹部和盆腔以外的其他放射学检查模态。此外，我们计划在临床环境中部署所提出的系统，并进行前瞻性研究，以评估其对健康结果和成本的影响。这将包括与临床合作伙伴的密切合作，确保系统在实际应用中表现出色，并能够适应不同的临床需求和场景。

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] Devlin, H.B. and Goldman, M. (1966) Backache Due to Osteoporosis in an Industrial Population. A Survey of 481 Patients. *Irish Journal of Medical Science*, **41**, 141-148. <https://doi.org/10.1007/bf02943677>
- [2] Ott, S.M. (1991) Methods of Determining Bone Mass. *Journal of Bone and Mineral Research*, **6**, S71-S76. <https://doi.org/10.1002/jbmr.5650061416>
- [3] Ito, M., Hayashi, K., Yamada, M., Uetani, M. and Nakamura, T. (1993) Relationship of Osteophytes to Bone Mineral Density and Spinal Fracture in Men. *Radiology*, **189**, 497-502. <https://doi.org/10.1148/radiology.189.2.8210380>
- [4] Yang, J., Pham, S. and Crabbe, D. (2003) Effects of Oestrogen Deficiency on Rat Mandibular and Tibial Microarchitecture. *Dentomaxillofacial Radiology*, **32**, 247-251. <https://doi.org/10.1259/dmfr/12560890>

- [5] 尹梓名, 孙大运, 胡晓晖, 等. 人工智能在骨质疏松症中的应用研究综述[J]. 小型微型计算机系统, 2019, 40(9): 1839-1850.
- [6] Watanabe, M., Sakai, D., Yamamoto, Y., Sato, M. and Mochida, J. (2010) Upper Cervical Spine Injuries: Age-Specific Clinical Features. *Journal of Orthopaedic Science*, **15**, 485-492. <https://doi.org/10.1007/s00776-010-1493-x>
- [7] Engelke, K., Libanati, C., Liu, Y., Wang, H., Austin, M., Fuerst, T., *et al.* (2009) Quantitative Computed Tomography (QCT) of the Forearm Using General Purpose Spiral Whole-Body CT Scanners: Accuracy, Precision and Comparison with Dual-Energy X-Ray Absorptiometry (DXA). *Bone*, **45**, 110-118. <https://doi.org/10.1016/j.bone.2009.03.669>
- [8] 崔洋洋, 宫赫, 关夏莉, 等. 基于髋部骨骼属性预测骨折风险研究进展[J]. 医用生物力学, 2019, 34(5): 555-559.
- [9] Lee, J.J.Y., Aghdassi, E., Cheung, A.M., Morrison, S., Cymet, A., Peeva, V., *et al.* (2012) Ten-Year Absolute Fracture Risk and Hip Bone Strength in Canadian Women with Systemic Lupus Erythematosus. *The Journal of Rheumatology*, **39**, 1378-1384. <https://doi.org/10.3899/jrheum.111589>
- [10] Feng-tan, L., Dong, L. and Zhang, Y-T. (2013) Influence of Tube Voltage on CT Attenuation, Radiation Dose, and Image Quality: Phantom Study. *Chinese Journal of Radiology*, **47**, 458-461.
- [11] Li, N., Li, X., Xu, L., Sun, W., Cheng, X. and Tian, W. (2013) Comparison of QCT and DXA: Osteoporosis Detection Rates in Postmenopausal Women. *International Journal of Endocrinology*, **2013**, Article ID: 895474. <https://doi.org/10.1155/2013/895474>
- [12] Yang, Z., Griffith, J.F., Leung, P.C. and Lee, R. (2009) Effect of Osteoporosis on Morphology and Mobility of the Lumbar Spine. *Spine*, **34**, E115-E121. <https://doi.org/10.1097/brs.0b013e3181895aca>
- [13] Rand, T., Seidl, G., Kainberger, F., Resch, A., Hittmair, K., Schneider, B., *et al.* (1997) Impact of Spinal Degenerative Changes on the Evaluation of Bone Mineral Density with Dual Energy X-Ray Absorptiometry (DXA). *Calcified Tissue International*, **60**, 430-433. <https://doi.org/10.1007/s002239900258>
- [14] Ensrud, K.E., Blackwell, T.L., Cawthon, P.M., Bauer, D.C., Fink, H.A., Schousboe, J.T., *et al.* (2016) Degree of Trauma Differs for Major Osteoporotic Fracture Events in Older Men versus Older Women. *Journal of Bone and Mineral Research*, **31**, 204-207. <https://doi.org/10.1002/jbmr.2589>
- [15] Fechtenbaum, J., Etcheto, A., Kolta, S., Feydy, A., Roux, C. and Briot, K. (2016) Sagittal Balance of the Spine in Patients with Osteoporotic Vertebral Fractures. *Osteoporosis International*, **27**, 559-567. <https://doi.org/10.1007/s00198-015-3283-y>
- [16] 陈刘萍, 余卓, 潘亚玲, 等. 人工智能骨密度测量系统与 QCT 测量骨密度的一致性研究[J]. 中国医学计算机成像杂志, 2023, 29(2): 178-183.
- [17] Pisani, P., Renna, M.D., Conversano, F., Casciaro, E., Di Paola, M., Quarta, E., *et al.* (2016) Major Osteoporotic Fragility Fractures: Risk Factor Updates and Societal Impact. *World Journal of Orthopedics*, **7**, 171-181. <https://doi.org/10.5312/wjo.v7.i3.171>
- [18] Engelke, K. (2017) Quantitative Computed Tomography—Current Status and New Developments. *Journal of Clinical Densitometry*, **20**, 309-321. <https://doi.org/10.1016/j.jocd.2017.06.017>
- [19] Xu, X., Li, N., Li, K., Li, X., Zhang, P., Xuan, Y., *et al.* (2019) Discordance in Diagnosis of Osteoporosis by Quantitative Computed Tomography and Dual-Energy X-Ray Absorptiometry in Chinese Elderly Men. *Journal of Orthopaedic Translation*, **18**, 59-64. <https://doi.org/10.1016/j.jot.2018.11.003>
- [20] Löffler, M.T., Jacob, A., Valentinitisch, A., Rienmüller, A., Zimmer, C., Ryang, Y., *et al.* (2019) Improved Prediction of Incident Vertebral Fractures Using Opportunistic QCT Compared to DXA. *European Radiology*, **29**, 4980-4989. <https://doi.org/10.1007/s00330-019-06018-w>
- [21] 朱心雨, 郭立, 黄鹏, 等. CT 纹理特征联合机器学习对发生骨质疏松性压缩骨折的预测价值[J]. 中国临床医学影像杂志, 2023, 34(6): 428-432.
- [22] Gausden, E.B., Nwachukwu, B.U., Schreiber, J.J., Lorch, D.G. and Lane, J.M. (2017) Opportunistic Use of CT Imaging for Osteoporosis Screening and Bone Density Assessment. *Journal of Bone and Joint Surgery*, **99**, 1580-1590. <https://doi.org/10.2106/jbjs.16.00749>
- [23] Wolterink, J.M., Leiner, T., de Vos, B.D., van Hamersvelt, R.W., Viergever, M.A. and Išgum, I. (2017) Automatic Coronary Artery Calcium Scoring in Cardiac CT Angiography Using Paired Convolutional Neural Networks. *Medical Image Analysis*, **34**, 123-136. <https://doi.org/10.1016/j.media.2016.04.004>
- [24] 余科君. 深度学习模型在骨质疏松症诊断的初步研究[D]: [硕士学位论文]. 南充: 川北医学院, 2023.
- [25] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., *et al.* (2016) Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, **316**, 2402-2410. <https://doi.org/10.1001/jama.2016.17216>
- [26] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., *et al.* (2017) Dermatologist-Level Classification

- of Skin Cancer with Deep Neural Networks. *Nature*, **542**, 115-118. <https://doi.org/10.1038/nature21056>
- [27] González, G., Ash, S.Y., Vegas-Sánchez-Ferrero, G., Onieva Onieva, J., Rahaghi, F.N., Ross, J.C., *et al.* (2018) Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. *American Journal of Respiratory and Critical Care Medicine*, **197**, 193-203. <https://doi.org/10.1164/rccm.201705-0860oc>
- [28] 田峰, 谢雁鸣. 绝经后骨质疏松症危险因素、预测模型和筛检工具研究[J]. 中国骨质疏松杂志, 2011, 17(2): 166-171.
- [29] Lee, S., Choe, E.K., Kang, H.Y., Yoon, J.W. and Kim, H.S. (2019) The Exploration of Feature Extraction and Machine Learning for Predicting Bone Density from Simple Spine X-Ray Images in a Korean Population. *Skeletal Radiology*, **49**, 613-618. <https://doi.org/10.1007/s00256-019-03342-6>
- [30] Pan, Y., Shi, D., Wang, H., Chen, T., Cui, D., Cheng, X., *et al.* (2020) Automatic Opportunistic Osteoporosis Screening Using Low-Dose Chest Computed Tomography Scans Obtained for Lung Cancer Screening. *European Radiology*, **30**, 4107-4116. <https://doi.org/10.1007/s00330-020-06679-y>
- [31] Yasaka, K., Akai, H., Kunimatsu, A., Kiryu, S. and Abe, O. (2020) Prediction of Bone Mineral Density from Computed Tomography: Application of Deep Learning with a Convolutional Neural Network. *European Radiology*, **30**, 3549-3557. <https://doi.org/10.1007/s00330-020-06677-0>
- [32] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [33] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [34] Georgiev, V.T., Karahaliou, A.N., Skiadopoulos, S.G., Arikidis, N.S., Kazantzi, A.D., Panayiotakis, G.S., *et al.* (2012) Quantitative Visually Lossless Compression Ratio Determination of JPEG2000 in Digitized Mammograms. *Journal of Digital Imaging*, **26**, 427-439. <https://doi.org/10.1007/s10278-012-9538-7>
- [35] Simard, P.Y., Steinkraus, D. and Platt, J. (2003) Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. Institute of Electrical and Electronics Engineers, Inc.
- [36] Wu, R., Yan, S., Shan, Y., Dang, Q. and Sun, G. (2015) Deep Image: Scaling up Image Recognition. arXiv: 1501.02876.
- [37] Genant, H.K., Wu, C.Y., van Kuijk, C. and Nevitt, M.C. (1993) Vertebral Fracture Assessment Using a Semiquantitative Technique. *Journal of Bone and Mineral Research*, **8**, 1137-1148. <https://doi.org/10.1002/jbmr.5650080915>
- [38] Taylor, G.W., Fergus, R., LeCun, Y. and Bregler, C. (2010) Convolutional Learning of Spatio-Temporal Features. In: Daniilidis, K., Maragos, P. and Paragios, N., Eds., *Lecture Notes in Computer Science*, Springer, 140-153. https://doi.org/10.1007/978-3-642-15567-3_11