

# 基于深度学习的核素骨显像中骨转移瘤的诊断及评估

刘思敏<sup>1</sup> 冯明<sup>2</sup> 蔡海东<sup>1</sup> 孙明<sup>1</sup> 王胤<sup>2</sup> 吕中伟<sup>1</sup> 李丹<sup>1</sup>

<sup>1</sup>同济大学附属第十人民医院核医学科,上海 200072;<sup>2</sup>同济大学电子与信息工程学院计算机科学与技术系,上海 201804

通信作者:李丹, Email: plumredlinda@163.com

**【摘要】目的** 基于深度学习对<sup>99</sup>Tc<sup>m</sup>-亚甲基二膦酸盐(MDP)全身骨显像图像中骨转移瘤进行智能诊断,并设计肿瘤负荷的定量评估指标。**方法** 回顾性纳入同济大学附属第十人民医院核医学科 2018 年 3 月至 2019 年 7 月间 621 例患者(男 389 例、女 232 例;年龄 12~93 岁)的骨显像图像,分为骨转移瘤组和非骨转移瘤组。从 2 组分别抽取 80%作为训练集,余 20%作为测试集。利用深度残差卷积神经网络 ResNet34 构建骨转移瘤诊断分类及分割模型。计算灵敏度、特异性、准确性以评估分类模型性能,分析分类模型在<50 岁(15 例)、≥50 且<70 岁(75 例)及≥70 岁(33 例)组的性能差异。利用模型分割骨转移瘤区域,以骰子系数评估分割模型结果与人工标注结果的比对。计算骨显像肿瘤负荷系数(BSTBI)以定量评估骨转移瘤肿瘤负荷。**结果** 骨转移瘤图像 280 例,非骨转移瘤图像 341 例;其中,训练集 498 例,测试集 123 例。诊断分类模型识别骨转移瘤的灵敏度、特异性及准确性分别为 92.59%(50/54)、85.51%(59/69)和 88.62%(109/123)。分类模型在<50 岁组表现最佳(灵敏度 2/2,特异性 12/13,准确性 14/15),其特异性在≥70 岁组中最低(8/12)。分割模型中,骨转移瘤区域骰子系数为 0.739,膀胱区域骰子系数为 0.925,模型在 3 个年龄组表现相当。初步结果显示,BSTBI 随病灶数目的增多、<sup>99</sup>Tc<sup>m</sup>-MDP 摄取程度的增高而增大。构建的骨转移瘤智能诊断模型从输入原始数据到最终完成 BSTBI 计算所需时间为(0.48±0.07) s。**结论** 基于深度学习的骨转移瘤智能诊断模型能较准确地识别骨转移瘤、进行自动区域分割及计算肿瘤负荷,为骨显像图像的解读提供了新方法。研究提出的 BSTBI 有望成为骨转移瘤肿瘤负荷的定量评估指标。

**【关键词】** 肿瘤转移;骨骼;放射性核素显像;<sup>99m</sup>Tc 锝美罗酸盐;神经网络(计算机)

DOI:10.3760/cma.j.cn321828-20200906-00337

## Automatic diagnosis and assessment of bone metastases on bone scans based on deep learning

Liu Simin<sup>1</sup>, Feng Ming<sup>2</sup>, Cai Haidong<sup>1</sup>, Sun Ming<sup>1</sup>, Wang Yun<sup>2</sup>, Lyu Zhongwei<sup>1</sup>, Li Dan<sup>1</sup>

<sup>1</sup>Department of Nuclear Medicine, Tenth People's Hospital of Tongji University, Shanghai 200072, China;

<sup>2</sup>Computer Science and Technology, School of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

Corresponding author: Li Dan, Email: plumredlinda@163.com

**【Abstract】Objective** To develop an approach for the automatic diagnosis of bone metastasis and to design a parameter of quantitative evaluation for tumor burden on bone scans based on deep learning technology. **Methods** A total of 621 cases (389 males, 232 females, age: 12–93 years) of bone scan images from the Department of Nuclear Medicine in Tenth People's Hospital of Tongji University from March 2018 to July 2019 were retrospectively analyzed. Images were divided into bone metastasis group and non-bone metastasis group. Eighty percent of the cases were randomly extracted from both groups as the training set, and the rest of cases were used as the test set. A deep residual convolutional neural network ResNet34 was used to construct the classification model and the segmentation model. The sensitivity, specificity and accuracy were calculated and the performance differences of the classification model in different age groups (15 cases of <50 years, 75 cases of ≥50 and <70 years, 33 cases of ≥70 years) were analyzed. The regions of metastatic bone lesions were automatically segmented by the segmentation model. The Dice coefficient was used to evaluate the effect of the segmentation model and the manual labeled results. Finally, the bone scans tumor burden index (BSTBI) was calculated to assess the tumor burden of bone metastases. **Results** There were 280 cases with bone metastases and 341 cases with non-bone metastases, including 498 in training set and 123 in test set. The classification model could accurately identify bone metastases, with the sensitivity, specificity and accuracy of 92.59% (50/54), 85.51% (59/69) and 88.62% (109/123), respectively, and it performed best in the <50 years group (sensitivity, 2/2; specificity, 12/13; accuracy, 14/15). The spe-

cificity in the  $\geq 70$  years group (8/12) was the lowest. The Dice coefficient of bone metastatic area and bladder area were 0.739 and 0.925 in the segmentation model, which performed similarly in the three age groups. Preliminary results showed that the value of BSTBI increased with the increase of the number of bone metastatic lesions and the degree of  $^{99}\text{Tc}^{\text{m}}$ -MDP uptake. The machine learning model in this study took  $(0.48 \pm 0.07)$  s for the entire analysis process from input to the final BSTBI calculation. **Conclusions** The deep learning based on automatic diagnosis framework for bone metastases can automatically and accurately identify segment bone metastases and calculate tumor burden. It provides a new way for the interpretation of bone scans. The proposed BSTBI may be used as a quantitative evaluation indicator in the future to assess the tumor burden of bone metastases based on bone scans.

**【Key words】** Neoplasm metastasis; Skeleton; Radionuclide imaging; Technetium Tc 99m medronate; Neural networks (computer)

DOI:10.3760/cma.j.cn321828-20200906-00337

$^{99}\text{Tc}^{\text{m}}$ -亚甲基二膦酸盐 (methylene diphosphate, MDP) 骨显像已广泛用于检测恶性肿瘤骨转移及评估骨转移瘤治疗效果。但人工解读图像任务重而繁琐,且易受个人经验的影响。骨转移瘤在临床上也尚缺乏客观准确的定量评价方法。目前,用于检测骨显像中骨病变的计算机辅助诊断系统多通过阈值法提取图像中的放射性浓聚区<sup>[1-6]</sup>,人为设计特征,但这会造成部分信息遗失,降低结论的可靠性。近年来,深度学习已在 CT、超声等主流医学影像的识别方面取得了不错的进展,但关于全身骨显像图像深度学习的研究尚少。本研究采用深度学习技术,提取骨显像图像特征,构建图像分类、分割模型等,期望实现快速智能判别骨显像图像,以提升阅片效率,帮助临床医师定量评价骨转移瘤。

## 资料与方法

1. 研究对象。回顾性纳入同济大学附属第十人民医院核医学科 2018 年 3 月至 2019 年 7 月 621 例患者(男 389 例、女 232 例,年龄 12~93 岁,中位 66 岁)的骨显像图像。由 2 位 10 年以上工作经验的核医学科医师将数据分为骨转移瘤组和非骨转移瘤组。

纳入标准:(1)骨转移瘤组。恶性肿瘤病史,有明确原发灶,图像上  $^{99}\text{Tc}^{\text{m}}$ -MDP 异常浓聚,且 CT、MRI 示骨质破坏,或经穿刺活组织检查(简称活

检),或随访半年骨病灶进展等证实。(2)非骨转移瘤组。无异常浓聚灶,但含四肢关节、脊椎等退行性病变,外伤或术后引起的  $^{99}\text{Tc}^{\text{m}}$ -MDP 浓聚等,随访半年骨病灶好转、未进展及无新的骨病灶。排除标准:(1)原发性骨肿瘤;(2)骨异常浓聚灶,未经其他相关检查证实;(3)骨转移瘤仅表现为放射性分布缺损;(4)图像对比度差,不清晰;(5)2 位医师对浓聚病灶意见不一致,不能通过其他相关检查证实。

本研究经同济大学附属第十人民医院伦理委员会通过(批件号 SHSY-IEC-KY-3.0/18-147/01)。研究符合《赫尔辛基宣言》的原则。

2. 显像方法。静脉注射  $^{99}\text{Tc}^{\text{m}}$ -MDP 370~740 MBq (上海原子科兴药业有限公司), 2~3 h 后行全身骨显像。SPECT 仪为美国 GE Discovery NM/CT 670 型,配低能高分辨平行孔准直器,窗宽 10%,能峰 140 keV,矩阵 256×1 024,扫描速率 10~20 cm/min。常规采集前后位图像。采用美国 GE Medical Systems 处理图像。

3. 数据处理(图 1)。原始数据包含 2 种不同灰度值下的前、后位 4 幅图像(图 1A)。由 2 位医师分别使用 Labelme 软件勾画骨转移瘤及膀胱区域(图 1B)<sup>[7]</sup>。使用双线性差值的方式统一原始图像尺寸。按图像像素数量将图像分为 7 个层级(零像素级、个像素级、十像素级、百像素级、千像素级、万像素

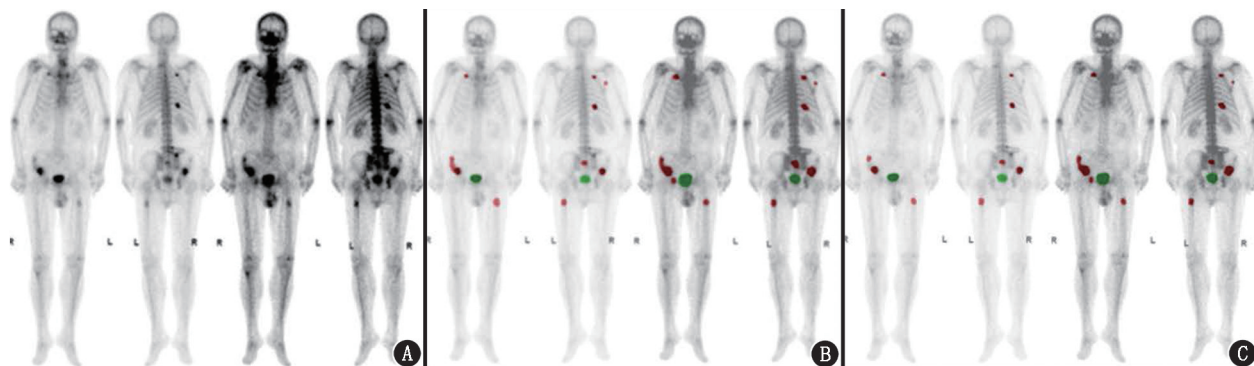


图 1 肺癌多发骨转移患者骨显像数据处理示意图。A. 原始图像; B. 人工标注; C. 模型预测结果; 可见模型预测能从像素级精度勾勒出更为精确的边界; 红色区域为骨转移瘤, 绿色区域为膀胱; L5 及右侧髂后上嵴浓聚影系污染所致

素级和十万像素级);以患者为单位,从每层级数据中随机挑选 80%作为训练集用于训练模型,余 20%作为测试集用来评估模型性能。

4.骨转移瘤智能诊断模型的建立。骨转移瘤智能诊断模型包含 2 个部分:骨转移瘤与膀胱分类模型、区域分割模型。前者判断输入的图像是否含骨转移瘤或膀胱,后者以像素级精确度分割骨转移瘤和膀胱区域。

诊断分类模型以 256×256 裁剪后的子图像作为输入。模型为多标签分类,结构采用深度残差卷积神经网络(ResNet34)。阈值根据分类“金标准”设置为 0.5<sup>[6]</sup>。在区域分割模型中输入含骨转移瘤或膀胱的数据,进行像素级分割,采用经典 unet 编码与解码结构。采用骰子系数(一种评估相似度的函数)计算 2 个样本(人工标注与区域分割模型)的相似度或重叠度,范围为[0, 1],系数越大,对应的分割效果越好。使用单个 1 080Ti 图形处理器(graphics processing unit, GPU)训练,采用 Pytorch 框架,训练时采用 Adam 优化器,初始学习率 0.01,批大小设置为 16。

将测试集患者按年龄分为<50 岁、≥50 且<70 岁及 ≥70 岁 3 组,分析骨转移瘤智能诊断模型在不同年龄组的性能差异。

5.骨显像肿瘤负荷系数(bone scan tumor burden index, BSTBI)的计算。在肿瘤疗效评价标准中,骨病灶因多发、形态不规则而无法定量测量<sup>[8]</sup>,缺少客观的骨转移瘤疗效评价方法。本研究计算 BSTBI 来近似评估骨转移瘤总体肿瘤负荷状态。公式如下:

BSTBI = K \* (sum(pixel in L (255 - pixel)) / ((K - 1) \* sum(pixel in L (255 - pixel)) + sum(pixel in R (255 - pixel)) - sum(pixel in B (255 - pixel)))) ..... 公式 1

其中,K 表示缩放因子,本研究选为 5;L 表示分割的骨转移瘤像素总和;pixel 表示像素值,MDP 越浓聚,

数值越小,对应到公式肿瘤负荷越大;R 表示图像中所有的像素;B 表示分割的膀胱像素。BSTBI ∈ [0,1]。

将前、后位 4 幅子图的平均 BSTBI 作为最终 BSTBI。在测试集中,对比不同患者间及同一患者治疗前后 BSTBI 变化,简单验证 BSTBI 是否能反映骨显像图像中骨转移瘤的负荷状态。

6.统计学处理。采用 Python 3.7.1 软件分析数据。符合正态分布的定量资料用 x̄±s 表示,定性资料用例数(或百分比)表示。采用 ROC AUC 分析骨转移瘤分类模型识别及区分骨转移瘤与膀胱的性能。

### 结 果

1.一般资料。纳入 621 例患者的骨显像图像中,骨转移瘤组 280 例、非骨转移瘤组 341 例。2 组患者原发肿瘤类型见表 1。训练集 498 例,含骨转移瘤组 226 例、非骨转移瘤组 272 例;测试集 123 例,含骨转移瘤组 54 例、非骨转移瘤组 69 例。

2.骨转移瘤智能诊断模型结果。123 例测试集中,骨转移瘤分类模型的灵敏度为 92.59%(50/54)、特异性为 85.51%(59/69)、阳性预测值为 83.33%(50/60)、阴性预测值为 93.65%(59/63)、准确性为 88.62%(109/123)。模型判断骨转移瘤的 ROC AUC 为 0.926;判断膀胱 AUC 为 0.997。

区域分割模型预测结果:骨转移瘤区域骰子系数为 0.739,膀胱区域骰子系数为 0.925。分割结果见图 1C,相比人工标注,区域分割模型能够从像素级的精度勾画出更为精确的边界。

3.骨转移瘤智能诊断模型对不同年龄组的性能差异(表 2)。诊断分类模型在 3 个年龄组的灵敏度均达 90%以上,以<50 岁组最高,为 2/2;特异性在 ≥70 岁组最低,为 8/12;准确性 3 组相当,<50 岁组略高,为 14/15。区域分割模型在 3 组的性能差异不

表 1 骨转移瘤组及非骨转移瘤组患者的原发肿瘤类型

组别	例数	肺癌	乳腺癌	前列腺癌	消化道肿瘤	肝癌	泌尿系肿瘤	鼻咽癌	骨肿瘤	卵巢癌
骨转移组	280	112	84	48	23	9	4	0	0	0
非骨转移瘤组	341	153	65	31	48	11	6	14	10	3
合计	621	265	149	79	71	20	10	14	10	3

表 2 骨转移瘤智能诊断模型对不同年龄组患者骨转移瘤识别及分割效果的对比

组别	例数	灵敏度	特异性	准确性	阳性预测值	阴性预测值	骰子系数
<50 岁	15	2/2	12/13	14/15	2/3	12/12	0.717
≥50 且<70 岁	75	90.32%(28/31)	88.64%(39/44)	89.33%(67/75)	84.85%(28/33)	92.86%(39/42)	0.742
≥70 岁	33	95.24%(20/21)	8/12	84.85%(28/33)	83.33%(20/24)	8/9	0.740
合计	123	92.59%(50/54)	85.51%(59/69)	88.62%(109/123)	83.33%(50/60)	93.65%(59/63)	0.739

注:括号内为对应例数比

大,骰子系数为 0.717~0.742。

4. BSTBI 的横向、纵向评估结果。由图 2A~2C 可知,BSTBI 随病灶数目的增多、 $^{99}\text{Tc}^{\text{m}}$ -MDP 摄取程度的增高而增大。应用模型对 1 例乳腺癌多发骨转移患者唑来膦酸治疗前后的骨显像(图 2D~2E)进行对比,治疗前后 BSTBI 分别为 0.367 和 0.249,提示 BSTBI 或可用于评估同一患者骨转移的治疗效果,以及评估不同患者间肿瘤骨转移负荷差异。在单 1 080Ti GPU 环境下,机器学习模型从原始图像输入到最终完成 BSTBI 计算所需时间约为  $(0.48 \pm 0.07)$  s,相比人工阅片具有明显的时间优势。

## 讨 论

Koizumi 等<sup>[6]</sup>应用计算机辅助诊断系统 BONENAVI v2 评估核素骨显像图像,采用了人工神经网络(artificial neural network, ANN)的学习方法,其诊断骨转移瘤的灵敏度为 85%,特异性为 82%,对前列腺癌骨转移瘤的诊断准确性最高达 72%。Brown 等<sup>[3]</sup>应用类似的辅助系统诊断前列腺癌骨转移瘤,灵敏度、特异性及准确性分别为 94.1%、89.2% 和 89.4%,高于前者。既往研究多利用阈值法提取热区、人工提取特征,如以每个热区的像素数、面积及位置作为输入的数值特征<sup>[1-6]</sup>。本研究采用经典的 ResNet34 网络构建诊断分类模型,输入原始图像,利用模型抽象图像特征,诊断模型灵敏度、特异性、准确性均高于 Koizumi 等<sup>[6]</sup>的研究结果,准确性与 Brown 等<sup>[3]</sup>的结果相似,表明构建的模型能较好地识别并分割骨转移瘤。本研究的诊断分类模型在 <50 岁的患者中性能最佳,这可能与该组样本量小、骨转移概率低有关。对于  $\geq 70$  岁组患者,分类模型的特异性最低,这与该组骨转移率高、正常骨图像相对较少相符。区域分割模型在 3 个年龄组的性能差异小,表明分割提取病灶与年龄无明显关联。相比既往,本研究群体更广,模型更具有普及性。ResNet34 模型能够

将病灶精确到像素级进行分割提取,具有更精确的边界,是构建骨转移瘤定量分析模型的基础。

评估肿瘤负荷大多依赖于影像学,核素骨显像是临床评估骨转移瘤疗效的常用方法。目前,骨显像图像中骨转移瘤定量评估指标主要包括骨显像系数(bone scan index, BSI)和骨显像病灶面积(bone scan lesion area, BSLA)等<sup>[3,8-9]</sup>。BSLA 包括自动图像标准化、病灶分割以及计算骨显像前后位视图上分割的病灶面积总和作为肿瘤负荷。BSI 则是在视觉判定或病灶分割的基础上,将估算质量与每个骨骼累及范围的乘积相加得出。两者均基于 ANN,需要大量人工标注、人工设计特征,耗时长,BSLA 还易受骨骼本底的影响。

膀胱内 $^{99}\text{Tc}^{\text{m}}$ -MDP 滞留会影响平面图像上骨盆转移瘤的显示。本研究在设计 ResNet34 模型时对膀胱的特点进行了学习,构建的模型能够很好地识别并分割膀胱,从而将膀胱与骨盆转移瘤鉴别开来。在分类、分割并提取骨转移瘤及膀胱的基础上,本研究预设计算 BSTBI 来评估骨转移瘤的总体肿瘤负荷。BSTBI 是以骨显像图像中除去膀胱以外的所有像素为背景,所有骨转移瘤病灶像素总和占图像背景的百分比。相较于 BSI 和 BSLA, BSTBI 能够减少膀胱对骨盆转移瘤的影响,减少由于药物注射、个体差异导致的药物代谢及分布的不同、扫描技术和图像后处理等多方面的影响,能更加准确地反映骨转移瘤负荷状态。从图像上可看出, BSTBI 随病灶数目增多、 $^{99}\text{Tc}^{\text{m}}$ -MDP 摄取程度增高而增大。从模型开发角度看(即据公式推论), BSTBI 能够反映骨转移瘤肿瘤负荷,以及治疗前后肿瘤负荷变化,尤其是从视觉分析角度不易发现的轻度变化。但受测试集数据量的限制,尚未对 BSTBI 进行大数据的临床验证,这将是下一步的工作重点。

本研究揭示了深度学习与骨显像相结合的临床前景,但数据与方法仍属概念验证,存在不足。首

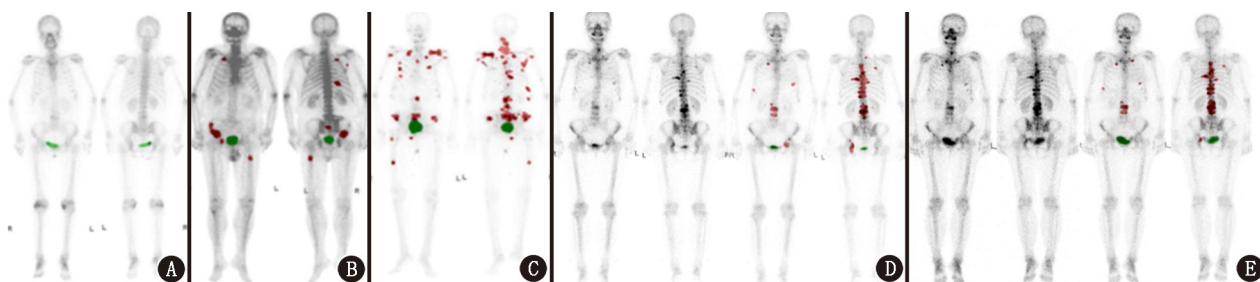


图 2 不同程度的骨转移瘤患者及同一患者治疗前后骨显像肿瘤负荷系数(BSTBI)的对比。A.乳腺癌患者(女,67岁),未见骨转移,BSTBI 为 0;B.肺癌多发骨转移患者(男,84岁),BSTBI 为 0.105;C.前列腺癌广泛骨转移患者(90岁),BSTBI 为 0.603;D~E.乳腺癌多发骨转移患者(女,61岁)唑来膦酸治疗前后 BSTBI 的变化,治疗前为 0.367(D),治疗后为 0.249(E)

先,数据来源于同一个中心,相比实际数据分布可能存在一定偏差,导致模型不能在所有数据上都有比较好的表现。这有待选择不同医学中心患者进一步验证模型,以减少差异。其次,骨转移瘤边界模糊,标注的工具不能勾画出区域最接近实际的边界,对模型有一定影响。第三,由于骨显像存在灵敏度高、特异性低的局限性,骨折、脊椎退行性病变等可能会被模型识别为骨转移瘤,需要医师进一步判断。本研究初步设定将误诊病灶删除后再行 BSTBI 计算。最后,骨显像的闪烁现象是 BSTBI 作为定量评价骨转移瘤肿瘤负荷指标的限制因素,理论上其会引起 BSTBI 增高,目前尚无有效的解决方法。

综上所述,本研究通过深度学习构建的骨显像骨转移瘤智能诊断模型取得了较好效果。初步结果显示,BSTBI 可作为骨转移瘤肿瘤负荷定量分析的指标之一,用于骨转移瘤疗效评估,但仍存在一些不足,需收集更多不同类别数据来提高其泛化能力并进行 BSTBI 的进一步临床验证。

**利益冲突** 所有作者声明无利益冲突

**作者贡献声明** 刘思敏、冯明:研究实施、论文撰写;蔡海东、孙明:图像采集、编辑及说明;王胤、吕中伟、李丹:研究指导、论文修改、经费支持

**参 考 文 献**

[1] Chiu JS, Wang YF, Su YC, et al. Artificial neural network to predict skeletal metastasis in patients with prostate cancer[J]. J Med

Syst, 2009, 33(2): 91-100. DOI:10.1007/s10916-008-9168-2.  
 [2] Sadik M, Hamadeh I, Nordblom P, et al. Computer-assisted interpretation of planar whole-body bone scans[J]. J Nucl Med, 2008, 49(12): 1958-1965. DOI:10.2967/jnumed.108.055061.  
 [3] Brown MS, Chu GH, Kim HJ, et al. Computer-aided quantitative bone scan assessment of prostate cancer treatment response[J]. Nucl Med Commun, 2012, 33(4): 384-394. DOI:10.1097/MNM.0b013e3283503ebf.  
 [4] Tokuda O, Harada Y, Ohishi Y, et al. Investigation of computer-aided diagnosis system for bone scans: a retrospective analysis in 406 patients[J]. Ann Nucl Med, 2014, 28(4): 329-339. DOI:10.1007/s12149-014-0819-8.  
 [5] Horikoshi H, Kikuchi A, Onoguchi M, et al. Computer-aided diagnosis system for bone scintigrams from Japanese patients: importance of training database[J]. Ann Nucl Med, 2012, 26(8): 622-626. DOI:10.1007/s12149-012-0620-5.  
 [6] Koizumi M, Wagatsuma K, Miyaji N, et al. Evaluation of a computer-assisted diagnosis system, BONENAVI version 2, for bone scintigraphy in cancer patients in a routine clinical setting[J]. Ann Nucl Med, 2015, 29(2): 138-148. DOI:10.1007/s12149-014-0921-y.  
 [7] Russell BC, Torralba A, Murphy KP, et al. LabelMe: a database and web-based tool for image annotation[J]. Int J Comput Vis, 2008, 77(1-3): 157-173. DOI:10.1007/s11263-007-0090-8.  
 [8] Imbriaco M, Larson SM, Yeung HW, et al. A new parameter for measuring metastatic bone involvement by prostate cancer: the Bone Scan Index[J]. Clin Cancer Res, 1998, 4(7): 1765-1772.  
 [9] Brown MS, Kim G, Chu GH, et al. Quantitative bone scan lesion area as an early surrogate outcome measure indicative of overall survival in metastatic prostate cancer[J]. J Med Imaging (Bellingham), 2018, 5(1): 011017. DOI:10.1117/1.JML.5.1.011017.

(收稿日期:2020-09-06)

· 读者 · 作者 · 编者 ·

**关于投稿提供伦理委员会批准文件及受试对象知情同意书的通告**

根据中华医学会杂志社的相关规定,当论文的主体是以人为研究对象时,作者应该说明其遵循的程序是否符合伦理审核委员会(单位性的、地区性的或国家性的)所制订的伦理学标准,并提供该委员会的批准文件(批准文号著录于论文中)及受试对象或其亲属的知情同意书。

本刊编辑部