



OPEN ACCESS

EDITED BY

Zhidao Xia,
Swansea University, United Kingdom

REVIEWED BY

Fengxia Chen,
Zhongnan Hospital, Wuhan University,
China
Zhi Lv,
Second Hospital of Shanxi Medical
University, China
Songfeng Chen,
First Affiliated Hospital of Zhengzhou
University, China

*CORRESPONDENCE

Songxiang Liu
✉ 598628576@qq.com
Zhewei Ye
✉ yezhewei@hust.edu.cn

†These authors have contributed
equally to this work and share
first authorship

‡These authors have contributed
equally to this work and share
last authorship

SPECIALTY SECTION

This article was submitted to
Surgical Oncology,
a section of the journal
Frontiers in Oncology

RECEIVED 16 December 2022

ACCEPTED 13 January 2023

PUBLISHED 08 February 2023

CITATION

Huo T, Xie Y, Fang Y, Wang Z, Liu P,
Duan Y, Zhang J, Wang H, Xue M, Liu S and
Ye Z (2023) Deep learning-based algorithm
improves radiologists' performance in lung
cancer bone metastases detection on
computed tomography.
Front. Oncol. 13:1125637.
doi: 10.3389/fonc.2023.1125637

COPYRIGHT

© 2023 Huo, Xie, Fang, Wang, Liu, Duan,
Zhang, Wang, Xue, Liu and Ye. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Deep learning-based algorithm improves radiologists' performance in lung cancer bone metastases detection on computed tomography

Tongtong Huo^{1,2†}, Yi Xie^{1†}, Ying Fang^{1†}, Ziyi Wang², Pengran Liu¹,
Yuyu Duan³, Jiayao Zhang¹, Honglin Wang¹, Mingdi Xue¹,
Songxiang Liu^{1*‡} and Zhewei Ye^{1*‡}

¹Department of Orthopedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, ²Research Institute of Imaging, National Key Laboratory of Multi-Spectral Information Processing Technology, Huazhong University of Science and Technology, Wuhan, China, ³Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Purpose: To develop and assess a deep convolutional neural network (DCNN) model for the automatic detection of bone metastases from lung cancer on computed tomography (CT)

Methods: In this retrospective study, CT scans acquired from a single institution from June 2012 to May 2022 were included. In total, 126 patients were assigned to a training cohort (n = 76), a validation cohort (n = 12), and a testing cohort (n = 38). We trained and developed a DCNN model based on positive scans with bone metastases and negative scans without bone metastases to detect and segment the bone metastases of lung cancer on CT. We evaluated the clinical efficacy of the DCNN model in an observer study with five board-certified radiologists and three junior radiologists. The receiver operator characteristic curve was used to assess the sensitivity and false positives of the detection performance; the intersection-over-union and dice coefficient were used to evaluate the segmentation performance of predicted lung cancer bone metastases.

Results: The DCNN model achieved a detection sensitivity of 0.894, with 5.24 average false positives per case, and a segmentation dice coefficient of 0.856 in the testing cohort. Through the radiologists-DCNN model collaboration, the detection accuracy of the three junior radiologists improved from 0.617 to 0.879 and the sensitivity from 0.680 to 0.902. Furthermore, the mean interpretation time per case of the junior radiologists was reduced by 228 s (p = 0.045).

Conclusions: The proposed DCNN model for automatic lung cancer bone metastases detection can improve diagnostic efficiency and reduce the diagnosis time and workload of junior radiologists.

KEYWORDS

artificial intelligence, deep learning, deep convolutional neural network, lung cancer bone metastases, computer-aided diagnosis

1 Introduction

Lung cancer (LC) is the main cause of cancer-related deaths globally (1). Approximately 1.5 million people are diagnosed with LC every year, with 1.3 million deaths (2). Furthermore, bone is the most common and the initial site of metastases from LC (3). Approximately 30%–70% of bone metastases are associated with LC, and 20%–30% of patients with LC already have bone metastases upon initial diagnosis (4). LC is often asymptomatic at the initial stage. Therefore, patients possibly already have metastases at diagnosis (5, 6). Although bone metastases may progress to pathologic fracture and/or nerve and spinal cord compression; some patients have no painful symptoms at the time of detection (7, 8). Once a tumor metastasizes to the bone, it is practically incurable and has a high mortality rate (9). Therefore, early detection of bone metastasis is important for decreasing morbidity as well as for disease staging, outcome prediction, and treatment planning (10).

Imaging is an important part of the management of bone metastasis (11–13). Computed tomography (CT) has the advantages of good anatomical resolution, soft-tissue contrast, and detailed morphology (14, 15). It also facilitates simultaneous evaluation of the primary and metastatic lesions (12, 16). The most important advantage of CT is the relatively low cost, which is very patient-friendly (17). Thus, in the clinical setting, CT is the most commonly used imaging for the diagnosis of primary cancer and whole-body staging when bone metastases are suspected (17, 18). The measurements of all metastatic lesions are time-consuming, especially, if multiple metastases are present. The heavy workload of image evaluation can be tiresome for radiologists, thus increasing the risk of missing lesions and leading to decreased sensitivity (19). Therefore, automated analysis of CT images is ideal for assisting radiologists in the accurate diagnosis of bone metastasis from LC.

Deep learning has been identified as a key sector in which artificial intelligence could streamline pathways, acting as a triage or screening service, decision aid, or second-reader support for radiologists (20). By now, artificial intelligence with deep convolutional neural network (DCNN) has been exploited to develop automated diagnosis and classification of cancer, including prostate cancer (21, 22), pancreatic cancer (23), gastric cancer (24, 25), breast cancer (26, 27), and LC (28–30). Furthermore, there has been a line of research on DCNN-based automated classification of CT images for the detection of metastasis caused by various primary tumors including gastric cancer (30), breast cancer (31), LC (32), and thyroid cancer (33). There is emerging evidence suggesting that DCNN could also be used to extract information from bone scan images for the automatic detection of LCBM (34, 35).

In this study, we developed a DCNN model that automated detecting LC bone metastases (LCBM) on CT and validated the model internally and externally. We also compared the DCNN

Abbreviations: LCBM, Lung cancer bone metastases; AI, Artificial intelligence; DL, Deep learning; DCNN, Deep convolutional neural network; AUC, Area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, Negative predictive value; CT, Computed tomography; ROC, Receiver operating characteristic; ROI, Region of interest; FPs, False positive; IoU, Intersection-over-Union.

model with five radiologists and explored whether it could enhance the diagnostic accuracy of junior radiologists.

2 Materials and methods

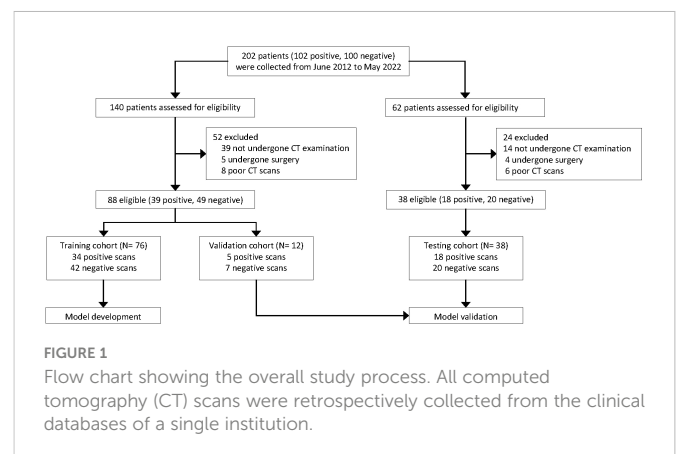
2.1 Patients

We collected 102 patients with pathologically confirmed primary LC who were confirmed to have synchronous or metachronous bone metastases and 100 patients who were confirmed to not have bone metastases by CT-guided biopsy pathology from June 2012 to May 2022. After reviewing the clinical and imaging data, we excluded patients who did not undergo CT examination ($n = 53$); who underwent surgery, chemotherapy, and radiation therapy for bone metastases before CT examination ($n = 9$); and who had poor CT image quality ($n = 14$). Finally, the CT images from 126 patients were included for the DCNN model development to detect bone metastases from LC, including a positive sample dataset of patients with biopsy-proven LC and bone metastases ($n = 57$), and a negative sample dataset of patients with biopsy-proven LC without bone metastases ($n = 69$). The process of patient enrollment is shown in Figure 1.

We randomly split the whole dataset ($n = 126$) into three cohorts: training (76 cases, to train the DCNN model), internal validation (12 cases, to fine-tune the hyper-parameters of the DCNN model), and external testing (38 cases, to evaluate the model and radiologists' performance). Furthermore, the patients included in the validation and testing datasets were excluded from the training dataset.

2.2 Imaging preprocessing and annotation

The clinical and imaging information of all patients was obtained through the medical record system and follow-up. To protect patients' privacy, all identifying information, such as name, sex, age, and ID, on CT was anonymized and omitted through image processing when data were first acquired. After image preprocessing in the Digital Imaging and Communications in Medicine (DICOM) format, complete thin-layer CT images were stored (see Supplementary Methods for CT protocols). The manual annotations of bone metastases were performed using LabelMe (36) with an image segmentation software (Mimics; Materialize, Belgium).



Two board-certified radiologists (8 and 9 years of experience in CT diagnosis) evaluated all CT examinations and, section by section, manually annotated five locations (spine, pelvis, limb, sternum, and clavicle) on the CT images of patients with LCBM. Two expert radiologists (14 and 20 years of experience in CT diagnosis, respectively) checked and manually delineated the volume of interest of the bone metastatic lesion in a voxel-wise manner on CT images using the diagnosis reports from two board-certified radiologists for establishing the reference standard of bone metastases. Furthermore, both of them repeated the annotations and modifications at least 3 weeks later and used them as ground truth (GT) labels for diagnosis and evaluation.

2.3 Model development

We developed a cascaded three-dimensional (3D) U-Net with 3D spatial SE modules and 3D GAU modules based on 3D U-Net (37), which is a robust state-of-the-art DCNN-based medical image segmentation method (see Supplementary Methods for training protocol). The Cascaded 3D U-Net (38) contains two 3D U-net architectures, wherein the first one is trained on down-sampled images and the second one is trained on full-resolution images. Training on down-sampled images first can enlarge the size of patches concerning the image and also enable the 3D U-Net network to learn more contextual information. Training on full-resolution images next refines the segmentation results predicted from the former 3D U-Net.

The 3D Spatial SE (39) module and 3D GAU (40) module are used more fully as the spatial attention module and the channel attention module to exploit the multiscale and multilevel features, respectively; they guide DCNN to efficiently focus on the targets rather than the background. The flowchart of our DCNN model for segmenting bone metastases is shown in Figure 2. The input of the end-to-end DCNN model is the 3D CT volume, and the output is the segmentation result of the bone metastases margin and the possibility of LCBM.

2.4 Observer study

To scale the proposed deep learning system with human readers, five radiologists (no overlap with the radiologists who labeled and checked the annotations) with 3–8 years of experience in CT diagnosis were required to participate in an independent human-only observer study. These radiologists were randomly shown the testing dataset to independently segment bone metastases and record the localization (corresponding CT layers); they were blinded to the bone metastases results and patient information.

To simulate the real clinical scenario, a radiologists-DCNN model collaboration study was conducted on three junior radiologists with 1–3 years of CT diagnostic experience, besides the independent observer study. These radiologists independently assessed the images to reach the first conclusion (whether bone metastases are present). Readouts of the DCNN model, including lesion labeling and probability of LCBM, were sent to the radiologists for reevaluating the images. The second assessment of radiologists served as the final output. In addition, we scheduled the 8 radiologists to perform the test in different locations to ensure their relative independence.

2.5 Statistical analysis

Our method followed a segmentation methodology to perform a detection task; therefore, both segmentation and detection metrics were important for evaluating the DCNN model performance. The metastases segmentation performance of the network was assessed using the metrics of dice coefficient and Intersection-over-Union (IoU) (41). The dice value indicates the overlapped voxels between the predicted results and GT. Its mathematical definition is as follows:

$$Dice(P, G) = \frac{2|P \cap G|}{|P| + |G|}$$

where denotes the number of labeled voxels, and P and G represent the predicted and GT values, respectively. The larger the

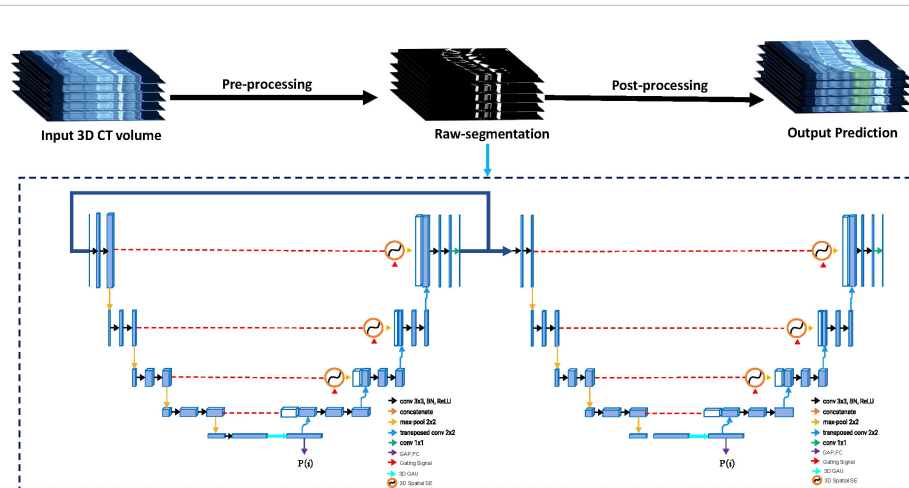


FIGURE 2 The flowchart of the proposed Cascaded 3D-Unet network. The network structure is divided into two parts: encoder and decoder. Between the encoder and decoder, we used the three-dimensional (3D) spatial squeeze and excitation modules (3D Spatial SE) and 3D global attention-up sample modules (3D GAU) to replace the original skip connections used in 3D U-Net.

value of the dice, the higher the degree of overlap between the segmentation prediction and GTs.

The evaluation of detection performance was based on case-based analysis; cases with at least one positive lesion were considered true positive (TP). We compared the model predictions with radiologists with TPs, false negatives (FNs), and false positives (FPs). According to the TP rate (sensitivity) versus the FP rate (1-specificity), we calculated the areas under the receiver operating characteristic (ROC) curve (AUC) and the 95% confidence interval (CI) for the five radiologists (averaged), DCNN model alone, three junior radiologists alone (averaged), and DCNN-assisted three junior radiologists. $P < 0.05$ was considered statistically significant. All statistical analyses were performed using SPSS 22.0 (IBM Corp., Armonk, NY, USA). The interpretation time for each scan was recorded automatically by the viewer and the detection time of the DCNN model was obtained from the computer terminal.

3 Results

No significant difference was noted in age or sex among the monocentric training, validation, and testing cohorts. In total, 34 positive and 42 negative scans were collected for the training dataset. The validation dataset comprised five positive and seven negative scans, and the testing dataset comprised 18 positive and 20 negative scans. The characteristics of the training, validation, and testing cohorts are listed in Table 1.

3.1 Performance of the DCNN model

A threshold of 0.5 (IoU > 0.5) was defined as the detection hit criterion; then, the segmentation metrics were computed with the GT labels generated in the image annotation procedure. At a threshold of

TABLE 1 Characteristics of patients in the three datasets.

Variables	Training Cohort (N=76)	Validation Cohort (N=12)	Testing Cohort (N=38)	Total (N=126)	P value
Age (years), mean \pm SD	61.55 \pm 13.22	61.48 \pm 13.14	61.41 \pm 13.45	61.52 \pm 13.25	<0.001
Gender					<0.001
Male	44 (57.8%)	7 (58.3%)	22 (57.9%)	73 (57.9%)	
Female	32 (42.2%)	5 (41.7%)	16 (42.1%)	53 (42.1%)	
Positive with bone metastases	33 (43.4%)	5 (41.7%)	19 (50.0%)	57 (45.2%)	
Bone metastasis site					0.417
Spine	14 (42.4%)	2 (40.0%)	8 (42.1%)	24 (42.1%)	
Pelvis	10 (30.3%)	1 (20.0%)	6 (31.6%)	17 (29.8%)	
Limb	5 (15.2%)	1 (20.0%)	3 (15.8%)	9 (15.8%)	
Sternum	3 (9.1%)	1 (20.0%)	2 (10.5%)	6 (10.5%)	
Clavicle	1 (3.0%)	0 (0.0%)	0 (0.0%)	1 (1.8%)	
Negative without bone metastases	43 (56.6%)	7 (58.3%)	19 (50.0%)	69 (54.8%)	
Brain metastases	9 (20.9%)	2 (28.6%)	4 (21.1%)	15 (21.7%)	
Liver metastases	26 (60.5%)	4 (57.1%)	12 (63.2%)	42 (60.9%)	
Lymph nodes metastases	8 (18.6%)	1 (14.3%)	3 (15.8%)	12 (17.4%)	
Primary Subtype					0.034
Small cell lung cancer (SCLC)	63 (82.9%)	10 (83.3%)	32 (84.2%)	105 (83.3%)	
Adenocarcinoma (LUAD)	13 (17.1%)	2 (16.7%)	6 (15.8%)	21 (16.7%)	
Number of tumors					0.002
Single	59 (77.6%)	9 (77.6%)	31 (81.6%)	99 (78.8%)	
Multiple (2)	17 (22.4%)	3 (22.4%)	8 (18.4%)	27 (21.4%)	
Tumor size					0.124
$0 \leq X \leq 3$	16 (20.5%)	2 (16.7%)	8 (21.1%)	26 (20.3%)	
$3 < X \leq 5$	25 (32.9%)	3 (25.0%)	13 (34.2%)	41 (32.5%)	
$5 < X \leq 7$	17 (22.4%)	4 (33.3%)	9 (23.7%)	30 (23.8%)	
> 7	18 (23.7%)	3 (25.0%)	8 (21.1%)	29 (23.0%)	
Grade					0.297

(Continued)

TABLE 1 Continued

Variables	Training Cohort (N=76)	Validation Cohort (N=12)	Testing Cohort (N=38)	Total (N=126)	P value
Grade I	4 (5.3%)	3 (25.0%)	9 (23.7%)	31 (24.6%)	
Grade II	19 (25.0%)	8 (66.7%)	23 (60.5%)	75 (59.5%)	
Grade III	44 (57.9%)	1 (8.3%)	4 (10.5%)	14 (11.1%)	
Grade IV	9 (11.8%)	3 (25.0%)	9 (23.7%)	31 (24.6%)	

For patient age, the mean age and standard deviation are presented, with a range of values in parentheses. For other data, the number of patients is presented, with percentages in parentheses.

0.5, the detection sensitivity of our DCNN model was 0.898 with 5.23 average FPs for the validation dataset and 0.894 with 5.24 average FPs for the testing dataset. Besides, our DCNN model achieved an acceptable segmentation performance (dice = 0.859 and 0.856 on the validation and testing datasets, respectively). The overall results of the DCNN model for the validation and testing datasets are shown in Table 2. An illustration of the predicted segmentation by DCNN is shown in Figure 3, where all cases were predicted with highly similar segmentation to GT.

3.2 Comparison with other networks and radiologists

We compared our model with several state-of-the-art deep neural networks in the validation and testing datasets to validate the

effectiveness of the proposed DCNN model (Figure 4 and Table 2). As shown in the results, the Cascaded 3D U-Net outperformed the 3D U-NET and 3D FCN (42) by large margins, which verified the effectiveness of network design in the proposed DCNN model. Moreover, the results of the ablation experiments demonstrated the best performance of our 3D U-Net with a 3D GAU module and 3D spatial SE module.

We performed observer studies and compared them with five radiologists using all images in the testing dataset to characterize the diagnostic value of the DCNN model. The DCNN model achieved high performance, outperforming any radiologists for LCBM detection with respect to both the primary metric AUROC (0.875 vs. 0.819 for the best radiologist) and sensitivity (0.894 vs. 0.892) tested in the observer-independent study. As for segmentation performance, our DCNN model outperformed all other networks and five radiologists (Table 2).

TABLE 2 Results of comparison with other networks and five radiologists on the testing cohort.

	AUROC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	IoU	Dice	Avg FP
Validation Cohort							
3D FCN	0.775 (0.746–0.803)	0.777 (0.724–0.823)	0.811 (0.792–0.840)	0.741 (0.720–0.766)	0.674 (0.643–0.750)	0.766 (0.736–0.796)	7.64
3D U-NET	0.781 (0.746–0.810)	0.790 (0.738–0.836)	0.810 (0.790–0.843)	0.752 (0.741–0.779)	0.681 (0.686–0.791)	0.785 (0.759–0.819)	7.20
3D GAU U-Net	0.805 (0.798–0.832)	0.807 (0.787–0.821)	0.784 (0.760–0.806)	0.834 (0.809–0.859)	0.693 (0.680–0.722)	0.794 (0.758–0.828)	5.97
3D SSE U-Net	0.853 (0.845–0.881)	0.853 (0.838–0.868)	0.860 (0.795–0.889)	0.850 (0.825–0.873)	0.750 (0.721–0.784)	0.827 (0.781–0.859)	5.65
Cascaded 3D U-Net	0.871 (0.849–0.880)	0.877 (0.821–0.907)	0.893 (0.881–0.905)	0.866 (0.835–0.875)	0.751 (0.701–0.790)	0.843 (0.812–0.888)	5.41
Our network	0.883 (0.878–0.901)	0.886 (0.842–0.922)	0.898 (0.881–0.905)	0.863 (0.835–0.875)	0.782 (0.741–0.810)	0.859 (0.826–0.884)	5.23
Testing Cohort							
3D FCN	0.771 (0.749–0.784)	0.773 (0.751–0.794)	0.806 (0.789–0.822)	0.736 (0.710–0.759)	0.677 (0.651–0.754)	0.762 (0.732–0.790)	7.74
3D U-NET	0.775 (0.756–0.796)	0.784 (0.787–0.801)	0.802 (0.785–0.818)	0.748 (0.723–0.770)	0.680 (0.664–0.782)	0.781 (0.750–0.818)	7.50
3D GAU U-Net	0.803 (0.776–0.834)	0.810 (0.799–0.853)	0.778 (0.761–0.818)	0.817 (0.801–0.844)	0.691 (0.674–0.718)	0.791 (0.764–0.823)	6.12
3D SSE U-Net	0.846 (0.828–0.867)	0.847 (0.783–0.897)	0.857 (0.786–0.891)	0.847 (0.819–0.868)	0.749 (0.720–0.798)	0.822 (0.779–0.855)	5.74
Cascaded 3D U-Net	0.868 (0.838–0.872)	0.875 (0.834–0.908)	0.887 (0.881–0.905)	0.854 (0.835–0.867)	0.750 (0.706–0.783)	0.841 (0.810–0.888)	5.58
Our network	0.875 (0.863–0.883)	0.878 (0.867–0.886)	0.894 (0.874–0.896)	0.857 (0.831–0.885)	0.789 (0.733–0.808)	0.856 (0.820–0.885)	5.24
Radiologist 1	0.785 (0.759–0.800)	0.771 (0.755–0.798)	0.846 (0.797–0.887)	0.766 (0.737–0.783)	0.642 (0.611–0.688)	0.726 (0.698–0.750)	3.25
Radiologist 2	0.792 (0.776–0.822)	0.793 (0.768–0.802)	0.861 (0.813–0.900)	0.769 (0.733–0.787)	0.668 (0.639–0.690)	0.741 (0.700–0.777)	3.10
Radiologist 3	0.795 (0.779–0.799)	0.799 (0.773–0.813)	0.876 (0.830–0.913)	0.776 (0.753–0.796)	0.679 (0.652–0.721)	0.750 (0.712–0.789)	2.68
Radiologist 4	0.804 (0.791–0.825)	0.810 (0.795–0.827)	0.880 (0.875–0.925)	0.792 (0.770–0.819)	0.691 (0.670–0.744)	0.769 (0.722–0.808)	2.14

(Continued)

TABLE 2 Continued

	AUROC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	IoU	Dice	Avg FP
Radiologist 5	0.819 (0.800–0.849)	0.822 (0.801–0.839)	0.892 (0.891–0.929)	0.807 (0.792–0.831)	0.717 (0.700–0.762)	0.798 (0.759–0.841)	1.80
Average of five radiologists	0.799 (0.772–0.821)	0.799 (0.782–0.834)	0.871 (0.841–0.899)	0.782 (0.758–0.800)	0.679 (0.654–0.702)	0.757 (0.730–0.783)	2.59

A comparison of detection and segmentation performance in the testing dataset of our proposed network (Cascaded 3D U-Net with 3D GAU modules and the 3D SSE modules), five other deep neural networks [Cascaded 3D U-Net, 3D FCN, 3D U-net, 3D GAU U-Net (3D U-Net with 3D global attention-up sample modules, and 3D SSE U-Net (3D U-Net with spatial squeeze and excitation modules)]. FP: false positives per scan. IoU: Intersection-over-Union. Dice: dice coefficient The bolded words “Validation Cohort” and “Testing Cohort” represent two equal cohorts containing different data.

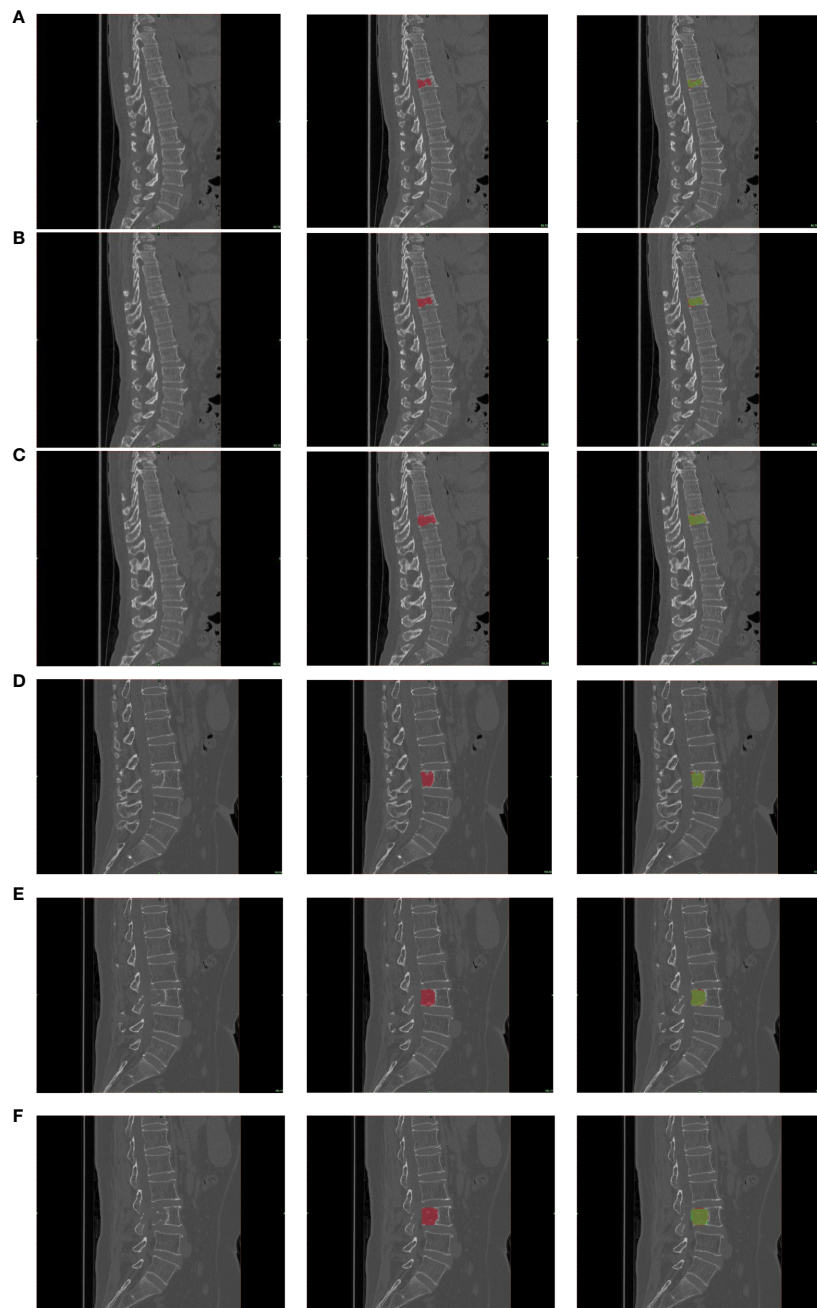


FIGURE 3 The segmentation results. Representative example of a patient with lung cancer bone metastasis with abnormal signals in the 4 lumbar spine (A–C) and a patient with lung cancer bone metastasis with abnormal signals in the 11 thoracic spine (D–F), and true-positive lesions with various appearances and locations. From left to right, the three images in a row are the original image, the corresponding GT label (red), and the candidate region output from the deep convolutional neural network (DCNN) (green). Note that the green region on the right image is a candidate region with a threshold of 0.5, and dice was calculated on this region.

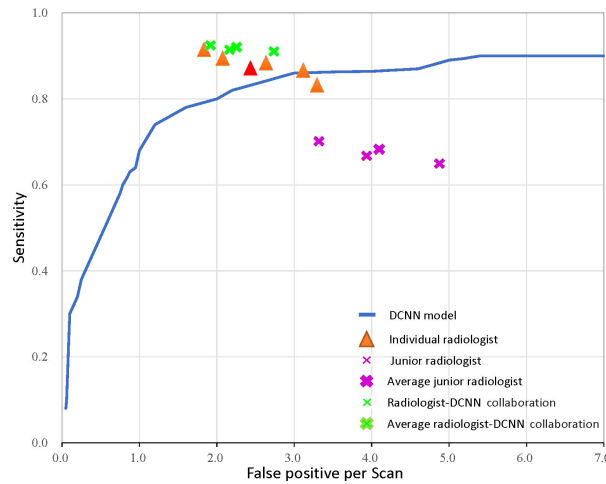


FIGURE 4 The receiver operator characteristic curve for the performance of the deep convolutional neural network (DCNN) model-only, radiologist-only, junior radiologist-only, and junior radiologists-DCNN model collaboration detection performance. Each orange triangle represents the performance of an individual radiologist; each pink fork represents the performance of a junior radiologist without the aid of DCNN; and each green fork represents the performance of a junior radiologist with the aid of DCNN. The red triangle indicates the average value of five radiologists, and the bolded forks indicate the average value of junior radiologists.

TABLE 3 A comparison of the DCNN model and three junior radiologists without and with the DCNN model.

		Age (year)	Experiences (year)	AUROC (95% CI)	Accuracy (95% CI)	Sensitivity (95% CI)	IoU	Dice	Avg Time (s)	Avg FP
DCNN Model		-	-	0.875 (0.863–0.883)	0.878 (0.867–0.886)	0.894 (0.874–0.896)	0.789 (0.733–0.808)	0.856 (0.820–0.885)	27	5.24
Junior radiologist only	R6	25	1.5	0.584 (0.567–0.612)	0.597 (0.575–0.614)	0.657 (0.531–0.768)	0.498 (0.480–0.515)	0.589 (0.568–0.610)	412	4.89
	R7	31	2	0.601 (0.588–0.620)	0.609 (0.590–0.632)	0.672 (0.546–0.782)	0.576 (0.534–0.604)	0.663 (0.638–0.686)	312	3.95
	R8	34	3	0.643 (0.611–0.668)	0.645 (0.622–0.671)	0.711 (0.695–0.741)	0.620 (0.586–0.645)	0.708 (0.684–0.733)	387	3.36
	Avg	30	2.17	0.609 (0.591–0.634)	0.617 (0.602–0.640)	0.680 (0.543–0.753)	0.565 (0.543–0.598)	0.653 (0.637–0.698)	370.3	4.07
Junior radiologist with DCNN	R6	25	1.5	0.871 (0.852–0.896)	0.875 (0.860–0.900)	0.898 (0.870–0.922)	0.771 (0.748–0.794)	0.833 (0.800–0.867)	204	2.72
	R7	31	2	0.873 (0.856–0.891)	0.879 (0.875–0.911)	0.901 (0.864–0.924)	0.790 (0.771–0.827)	0.861 (0.811–0.893)	104	2.23
	R8	34	3	0.878 (0.861–0.901)	0.883 (0.869–0.923)	0.907 (0.854–0.924)	0.802 (0.784–0.821)	0.886 (0.844–0.928)	119	1.98
	Avg	30	2.17	0.874 (0.857–0.914)	0.879 (0.863–0.910)	0.902 (0.844–0.921)	0.788 (0.764–0.823)	0.847 (0.812–0.869)	142.3	2.31

FP, false positives per scan. IoU, Intersection-over-Union. Dice, dice coefficient. The bolded words “DCNN Model” represent our proposed network (Cascaded 3D U-Net with 3D GAU modules and the 3D SSE modules). The bolded words “Avg” represent the average of the values in the three cells above.

3.3 Radiologists-DCNN model collaboration

We further validated the radiologists-DCNN model collaboration performance in the testing dataset. As demonstrated, the three junior radiologists (with the assistance of the DCNN model) showed substantial improvement in identifying LCBM (Figure 4 and Table 3).

The DCNN model assisted junior radiologists in diagnosing LCBM with a higher mean AUROC (0.874 [95% CI: 0.807–0.874] vs. 0.609 [95% CI: 0.591–0.634], $P < 0.001$), mean accuracy (0.879 vs. 0.617, $P < 0.001$), and mean sensitivity (0.902 vs. 0.680, $P = 0.009$) compared with those achieved alone. Moreover, the mean interpretation time per case of the junior radiologists was

significantly reduced from 370.3 s to 142.3 s (228 s decrease, $P = 0.045$) when assisted by the DCNN model.

4 Discussion

Herein, we proposed an improved Cascaded 3D U-Net based on the DCNN model to detect and segment LCBM on CT scans. Radiologists underperformed the DCNN model concerning detection sensitivities, although they achieved much lower average FPs. In the segmentation task, the proposed DCNN model achieved a mean dice of 0.856 and a mean FP of 5.24 in the testing dataset, showing that the proposed model achieved better results than other state-of-the-art networks.

In the radiologists-DCNN model collaboration study, the mean sensitivity of the junior radiologist for LCBM improved from 0.680 to 0.902 with acceptable FPs (2.59). The radiologists-DCNN model collaboration enhanced the detection sensitivity and FPs compared with radiologist-only or DCNN model-only diagnosis, demonstrating the existence of the DCNN model-detected bone metastases that were missed by junior radiologists and vice versa. Moreover, the DCNN model-assisted diagnosis significantly decreased approximately 62% clinical time (142 s vs. 370 s), which had never been evaluated in previous studies.

Prior to our study, two recent studies used deep learning to detect bone metastases from CT images (43, 18). However, both studies focused only on spinal lesions. The spine is the most common site of bone metastases; however, metastases can occur at any site in the entire skeleton (44). Furthermore, both studies formalized the task as two-dimensional detection, whereas our study formalized it as 3D segmentation. Besides, the data and annotation in our study were of a higher standard. In our study, high-quality thin-slice CT scans with a thickness of 1–1.25 mm were used to support the model development. Moreover, we followed a repetitive and retrospective labeling procedure by four radiologists to ensure the high quality of our annotations, thus reducing the risk of overvaluing model performance. Our model achieved significantly higher detection sensitivity and remained consistent across the training, validation, and testing datasets.

Nevertheless, there are still some limitations to this study. First, our study was retrospective and monocentric; therefore, future validations in prospective randomized settings can provide more powerful conclusions. In addition, the sample size was small. We believe that the main reason for such limitation is the somewhat low incidence and prevalence of LCBM (45). Although our results were encouraging, experiments in large multi-center datasets are needed to verify the results in further studies. Second, the training process of the DCNN model depends on the segmentation GT labels of LCBM by radiologists, who also have imperfect reliability. To address this issue, numerous studies have been conducted to train the DCNN model using weak labels (46). Third, the DCNN model was established to detect LCBM so that the patients included in the positive cohort had LC. However, bone metastases may also arise from other solid tumors, such as breast, prostate, colorectal, thyroid, and gynecologic cancers and melanoma (47). Therefore, more generalized DCNN models that can distinguish multiple origins of

bone metastases should be followed up. Finally, our DCNN models were designed to deal with a single task of LCBM detection on CT. However, in clinical practice, radiologists may not rely on a single medical file for a final diagnosis, instead, they need to combine other clinical or imaging reports to achieve the diagnosis of LCBM. If the models are built based on a single parameter, their clinical value may be significantly endangered. Therefore, more inclusive models combining various characteristics should be designed and emphasized in the future (48).

In conclusion, our DCNN model collaborated with junior radiologists helped to enhance the diagnostic effectiveness and efficiency in the diagnosis of LCBM on CT, indicating the great potential of DCNN-assisted diagnosis in clinical practice.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This retrospective study was approved by the hospital ethics committee (Union Hospital, Tongji Medical College, Huazhong University of Science and Technology). The ethics committee waived the requirement of written informed consent for participation.

Author contributions

TH and YX contributed equally to the manuscript. ZY and SZ present the conception of the work; YF and MX acquired the data; SL obtained funding to support the work and built up the research team for joint development; YD and LL labelled the images for further analysis; JZ and HL assisted in collecting data. TH, ZW and PL conducted the statistical analysis; TH drafted the manuscript with critical feedback from all authors. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (No. 82172524 and 81974355) and the Artificial Intelligence Major Special Project of Hubei (No. 2021BEA161).

Acknowledgments

We thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript. We especially thank for the support provided by the Research Institute of Imaging, National Key Laboratory of Multi-Spectral Information Processing Technology, Huazhong University of Science and Technology of China.

Conflict of interest

TH will apply for an International Invention Patent for the DCNN model used in this study pending Union Hospital, Tongji Medical College, Huazhong University of Science and Technology.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1125637/full#supplementary-material>

References

- Kim ST, Uhm JE, Lee J, Sun JM, Sohn I, Kim SW, et al. Randomized phase II study of gefitinib versus erlotinib in patients with advanced non-small cell lung cancer who failed previous chemotherapy. *Lung Cancer (Amsterdam Netherlands)* (2012) 75(1):82–8. doi: 10.1016/j.lungcan.2011.05.022
- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: Cancer J Clin* (2015) 65(2):87–108. doi: 10.3322/caac.21262
- Ebert W, Muley T, Herb KP, Schmidt-Gayk H. Comparison of bone scintigraphy with bone markers in the diagnosis of bone metastasis in lung carcinoma patients. *Anticancer Res* (2004) 24(5b):3193–201.
- Katakami N. [Lung cancer with bone metastasis]. *Gan to kagaku ryoho Cancer Chemother* (2006) 33(8):1049–53.
- Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, et al. The lung microenvironment: an important regulator of tumour growth and metastasis. *Nat Rev Cancer* (2019) 19(1):9–31. doi: 10.1038/s41568-018-0081-9
- Esposito M, Kang Y. Targeting tumor-stromal interactions in bone metastasis. *Pharmacol Ther* (2014) 141(2):222–33. doi: 10.1016/j.pharmthera.2013.10.006
- Costelloe CM, Rohren EM, Madewell JE, Hamaoka T, Theriault RL, Yu TK, et al. Imaging bone metastases in breast cancer: techniques and recommendations for diagnosis. *Lancet Oncol* (2009) 10(6):606–14. doi: 10.1016/S1470-2045(09)70088-9
- Hamaoka T, Madewell JE, Podoloff DA, Hortobagyi GN, Ueno NT. Bone imaging in metastatic breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol* (2004) 22(14):2942–53. doi: 10.1200/JCO.2004.08.181
- Roodman GD. Mechanisms of bone metastasis. *New Engl J Med* (2004) 350(16):1655–64. doi: 10.1056/NEJMra030831
- Chang CY, Gill CM, Joseph Simeone F, Taneja AK, Huang AJ, Torriani M, et al. Comparison of the diagnostic accuracy of 99m-Tc-MDP bone scintigraphy and 18F-FDG PET/CT for the detection of skeletal metastases. *Acta Radiol (Stockholm Sweden 1987)* (2016) 57(1):58–65. doi: 10.1177/0284185114564438
- Even-Sapir E. Imaging of malignant bone involvement by morphologic, scintigraphic, and hybrid modalities. *J Nucl Med Off Publication Soc Nucl Med* (2005) 46(8):1356–67.
- Heindel W, Gübitz R, Vieth V, Weckesser M, Schober O, Schäfers M. The diagnostic imaging of bone metastases. *Deutsches Arzteblatt Int* (2014) 111(44):741–7. doi: 10.3238/arztebl.2014.0741
- O'Sullivan GJ, Carty FL, Cronin CG. Imaging of bone metastasis: An update. *World J Radiol* (2015) 7(8):202–11. doi: 10.4329/wjr.v7.i8.202
- Kalogeropoulou C, Karachaliou A, Zampakis P. Radiologic evaluation of skeletal metastases: role of plain radiographs and computed tomography. *Bone Metastases: Springer* (2009) p:119–36. doi: 10.1007/978-1-4020-9819-2_6
- Yang H-L, Liu T, Wang X-M, Xu Y, Deng S-M. Diagnosis of bone metastases: a meta-analysis comparing 18FDG PET, CT, MRI and bone scintigraphy. *Eur Radiol* (2011) 21(12):2604–17. doi: 10.1007/s00330-011-2221-4
- Groves AM, Beadsmoore CJ, Cheow HK, Balan KK, Courtney HM, Kaptoge S, et al. Can 16-detector multislice CT exclude skeletal lesions during tumour staging? implications for the cancer patient. *Eur Radiol* (2006) 16(5):1066–73. doi: 10.1007/s00330-005-0042-z
- Hammon M, Dankerl P, Tsybal A, Wels M, Kelm M, May M, et al. Automatic detection of lytic and blastic thoracolumbar spine metastases on computed tomography. *Eur Radiol* (2013) 23(7):1862–70. doi: 10.1007/s00330-013-2774-5
- Chmelik J, Jakubicek R, Walek P, Jan J, Ourednicek P, Lambert L, et al. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. *Med Image Anal* (2018) 49:76–88. doi: 10.1016/j.media.2018.07.008
- Noguchi S, Nishio M, Sakamoto R, Yakami M, Fujimoto K, Emoto Y, et al. Deep learning-based algorithm improved radiologists' performance in bone metastases detection on CT. *Eur Radiol* (2022) 32(11):7976–87. doi: 10.1007/s00330-022-08741-3
- McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* (2020) 577(7788):89–94. doi: 10.1038/s41586-019-1799-6
- Acar E, Leblebici A, Ellidokuz BE, Başbınar Y, Kaya GÇJTBJOR. Machine learning for differentiating metastatic and completely responded sclerotic bone lesion in prostate cancer: A retrospective radiomics study. *Br Radiol* (2019) 92(1101):20190286. doi: 10.1259/bjr.20190286
- Kiljunen T, Akram S, Niemelä J, Löytyniemi E, Seppälä J, Heikkilä J, et al. A deep learning-based automated CT segmentation of prostate cancer anatomy for radiation therapy planning—a retrospective multicenter study. *Diagnost (Basel Switzerland)* (2020) 10(11):959. doi: 10.3390/diagnostics10110959
- Pelaez-Luna M, Takahashi N, Fletcher JG, Chari ST. Resectability of presymptomatic pancreatic cancer and its relationship to onset of diabetes: a retrospective review of CT scans and fasting glucose values prior to diagnosis. *Am J Gastroenterol* (2007) 102(10):2157–63. doi: 10.1111/j.1572-0241.2007.01480.x
- Jiang Y, Jin C, Yu H, Wu J, Chen C, Yuan Q, et al. Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: A multicenter, retrospective study. *Ann Surg* (2021) 274(6):e1153–e61. doi: 10.1097/SLA.0000000000003778
- Jiang Y, Zhang Z, Yuan Q, Wang W, Wang H, Li T, et al. Predicting peritoneal recurrence and disease-free survival from CT images in gastric cancer with multitask deep learning: a retrospective study. *Lancet Digital Health* (2022) 4(5):e340–e50. doi: 10.1016/S2589-7500(22)00040-1
- Rezaeijoo SM, Ghorvei M, Mofid B. Predicting breast cancer response to neoadjuvant chemotherapy using ensemble deep transfer learning based on CT images. *J X-ray Sci Technol* (2021) 29(5):835–50. doi: 10.3233/XST-210910
- Schreier J, Attanasio F, Laaksonen H. A full-image deep segmenter for CT images in breast cancer radiotherapy treatment. *frontiers in oncology. Front Oncol* (2019) 9:677. doi: 10.3389/fonc.2019.00677
- Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. Author correction: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* (2019) 25(8):1319. doi: 10.1038/s41591-019-0536-x
- Alahmari SS, Cherezov D, Goldgof D, Hall L, Gillies RJ, Schabath MB. Delta radiomics improves pulmonary nodule malignancy prediction in lung cancer screening. *IEEE Access Pract Innovations Open Solutions* (2018) 6:77796–806. doi: 10.1109/ACCESS.2018.2884126
- Cherezov D, Hawkins SH, Goldgof DB, Hall LO, Liu Y, Li Q, et al. Delta radiomic features improve prediction for lung cancer incidence: A nested case-control analysis of the national lung screening trial. *Cancer Med* (2018) 7(12):6340–56. doi: 10.1002/cam4.1852
- Yang X, Wu L, Ye W, Zhao K, Wang Y, Liu W, et al. Deep learning signature based on staging CT for preoperative prediction of sentinel lymph node metastasis in breast cancer. *Acad Radiol* (2020) 27(9):1226–33. doi: 10.1016/j.acra.2019.11.007
- Wang YW, Chen CJ, Huang HC, Wang TC, Chen HM, Shih JY, et al. Dual energy CT image prediction on primary tumor of lung cancer for nodal metastasis using deep learning. *Computerized Med Imaging Graphics Off J Computerized Med Imaging Society* (2021) 91:101935. doi: 10.1016/j.compmedimag.2021.101935
- Lee JH, Ha EJ, Kim JH. Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT. *Eur Radiol* (2019) 29(10):5452–7. doi: 10.1007/s00330-019-06098-8
- Zhou X, Wang H, Feng C, Xu R, He Y, Li L, et al. Emerging applications of deep learning in bone tumors: Current advances and challenges. *Front Oncol* (2022) 12:908873. doi: 10.3389/fonc.2022.908873

35. Li T, Lin Q, Guo Y, Zhao S, Zeng X, Man Z, et al. Automated detection of skeletal metastasis of lung cancer with bone scans using convolutional nuclear network. *Phys Med Biol* (2022) 67(1):015004. doi: 10.1088/1361-6560/ac4565
36. Torralba, Antonio, Russell BC, Yuen J. Labelme: Online image annotation and applications. *International Journal of Computer Vision* (2010) 98(8):1467–1484. doi: 10.1007/s11263-007-0090-8
37. Zhong Z, Kim Y, Zhou L, Plichta K, Allen B, Buatti J, et al. (2018). 3D fully convolutional networks for co-segmentation of tumors on PET-CT images, in: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* Washington, DC, USA, 4-7 April 2018.
38. Isensee F, Jäger PF, Kohl SA, Petersen J, Maier-Hein KH. Automated design of deep learning methods for biomedical image segmentation. *Nature Methods* (2020) 18(2):203–11. doi: 10.1038/s41592-020-01008-z
39. Roy AG, Navab N, Wachinger C. (2018). Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks, in: *International conference on medical image computing and computer-assisted intervention Lecture Notes in Computer Science* (2018), 11070. doi: 10.1007/978-3-030-00928-1_48
40. Li H, Xiong P, An J, Wang LJ. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180* (2018).
41. Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol* (2004) 11(2):178–89. doi: 10.1016/S1076-6332(03)00671-8
42. Li L, Zhao X, Lu W, Tan S. Deep learning for variational multimodality tumor segmentation in PET/CT. *Neurocomputing* (2020) 392:277–95. doi: 10.1016/j.neucom.2018.10.099
43. Fan X, Zhang X, Zhang Z, Jiang Y. Deep learning-based identification of spinal metastasis in lung cancer using spectral CT images. *Sci Program* (2021) 2021:2779390. doi: 10.1155/2021/2779390
44. Kakhki VRD, Anvari K, Sadeghi R, Mahmoudian A-S, Torabian-Kakhki MJNMR. Pattern and distribution of bone metastases in common malignant tumors. *Nucl Med Rev Cent East Eur* (2013) 16(2):66–9. doi: 10.5603/NMR.2013.0037
45. Łukaszewski B, Nazar J, Goch M, Łukaszewska M, Stepieński A, Jurczyk MU. Diagnostic methods for detection of bone metastases. *Contemp Oncol (Pozn)* (2017) 21(2):98–103. doi: 10.5114/wo.2017.68617
46. Wang Y, Lu L, Cheng C-T, Jin D, Harrison AP, Xiao J, et al. Weakly supervised universal fracture detection in pelvic x-rays, in: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science*, (2019) . Springer, Cham. doi: 10.1007/978-3-030-32226-7_51
47. Fornetti J, Welm AL, Stewart SA. Understanding the bone in cancer metastasis. *J Bone Mineral Res Off J Am Soc Bone Mineral Res* (2018) 33(12):2099–113. doi: 10.1002/jbmr.3618
48. Zhou X, Wang H, Feng C, Xu R, He Y, Li L, et al. Emerging applications of deep learning in bone tumors: Current advances and challenges. (2022) 12:908873. doi: 10.3389/fonc.2022.908873