Korean Journal of Radiology

Automated Detection and Segmentation of Bone Metastases on Spine MRI Using U-Net: A Multicenter Study

Dong Hyun Kim[1,2], Jiwoon Seo[1,2], Ji Hyun Lee[1], Eun-Tae Jeon[1], DongYoung Jeong[3], Hee Dong Chae[2,4], Eugene Lee[2,5], Ji Hee Kang[6], Yoon-Hee Choi[7], Hyo Jin Kim[1,2], Jee Won Chai[1,2]

[1]Department of Radiology, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea
[2]College of Medicine, Seoul National University, Seoul, Republic of Korea
[3]DEEPNOID Inc., Seoul, Republic of Korea
[4]Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea
[5]Department of Radiology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
[6]Department of Radiology, Konkuk University Medical Center, Seoul, Republic of Korea
[7]Department of Physical Medicine and Rehabilitation, Soonchunhyang University Seoul Hospital, Seoul, Republic of Korea

Objective: To develop and evaluate a deep learning model for automated segmentation and detection of bone metastasis on spinal MRI.
Materials and Methods: We included whole spine MRI scans of adult patients with bone metastasis: 662 MRI series from 302 patients (63.5 ± 11.5 years; male:female, 151:151) from three study centers obtained between January 2015 and August 2021 for training and internal testing (random split into 536 and 126 series, respectively) and 49 MRI series from 20 patients (65.9 ± 11.5 years; male:female, 11:9) from another center obtained between January 2018 and August 2020 for external testing. Three sagittal MRI sequences, including non-contrast T1-weighted image (T1), contrast-enhanced T1-weighted Dixon fat-only image (FO), and contrast-enhanced fat-suppressed T1-weighted image (CE), were used. Seven models trained using the 2D and 3D U-Nets were developed with different combinations (T1, FO, CE, T1 + FO, T1 + CE, FO + CE, and T1 + FO + CE). The segmentation performance was evaluated using Dice coefficient, pixel-wise recall, and pixel-wise precision. The detection performance was analyzed using per-lesion sensitivity and a free-response receiver operating characteristic curve. The performance of the model was compared with that of five radiologists using the external test set.
Results: The 2D U-Net T1 + CE model exhibited superior segmentation performance in the external test compared to the other models, with a Dice coefficient of 0.699 and pixel-wise recall of 0.653. The T1 + CE model achieved per-lesion sensitivities of 0.828 (497/600) and 0.857 (150/175) for metastases in the internal and external tests, respectively. The radiologists demonstrated a mean per-lesion sensitivity of 0.746 and a mean per-lesion positive predictive value of 0.701 in the external test.
Conclusion: The deep learning models proposed for automated segmentation and detection of bone metastases on spinal MRI demonstrated high diagnostic performance.
Keywords: Bone neoplasms; Deep learning; Magnetic resonance imaging; Metastasis; Spine

INTRODUCTION

Bone is one of the most common metastatic sites for cancer, and bone metastases are a major cause of morbidity in patients with advanced-stage malignancies [1]. Comorbidities in spinal metastases, such as pathologic

fractures and spinal cord or nerve root compression, can impair patients' activities, deteriorate their quality of life, and affect their prognosis. This could also be linked to their general condition and volition for further treatment [2].

With advancements in cancer treatment, the survival of patients with advanced-stage cancer has improved [3]. Consequently, the life expectancy of patients with metastatic disease is increasing, despite the escalating incidence of skeletal metastasis [4]. Moreover, the management of bone metastasis is considered imperative, extending beyond palliative measures. Therefore, the early diagnosis of bone metastasis and detection of spinal complications are crucial for guiding treatment decisions.

MRI is the most sensitive imaging modality for detecting and evaluating metastases. However, the increasing number of patients undergoing spinal MRI poses challenges in interpretation for radiologists. Detecting metastases and analyzing the interval changes of these lesions on spinal MRI have become increasingly complex and laborious. Recently, increasing interest has been directed to deep learning and its ability to detect and segment medical imaging data [5-9]. Consequently, there is growing expectation for adopting deep learning algorithms in the diagnosis of bone metastasis.

The aim of this study was to develop and evaluate the performance of a deep learning model for the automated detection and segmentation of bone metastasis on multisequence whole-spine MRI.

MATERIALS AND METHODS

The Institutional Review Boards of the four participating centers approved the study design (IRB approval numbers: 30-2019-24, 2002-074-1101, B-2002/597-403, 2021-06-020) and waived the requirement for informed consent because of the retrospective nature of the study design.

Study Population

For the training and internal test sets, consecutive whole-spine MRI scans from three study centers (Centers 1–3), obtained between January 2015 and August 2021, were collected from the picture archiving communication systems of each center. The inclusion criteria were as follows: 1) confirmative diagnosis of malignancy and 2) first spinal MRI for evaluating metastasis, regardless of the previous treatment. Patients with 1) global bone marrow signal alteration, including diffuse bone metastases

or red marrow reconversion, which present an intricate background texture for input data, 2) innumerable bone metastases, 3) no bone metastasis, 4) suboptimal image quality, and 5) no reference imaging for the confirmative diagnosis of metastases were excluded from the study.

Another set of whole-spine MRIs were collected from another center (Center 4) for external testing. The images were obtained between January 2018 and August 2020. The inclusion and exclusion criteria were same as those used for the data for model development.

MRI Acquisition and Archiving

MRI acquisition followed the standard protocol at each study center, employing scanners from GE (GE Healthcare, Waukesha, WI, USA), Philips (Phillips Healthcare, Amsterdam, Netherlands), and Siemens (Siemens Healthineers, Erlangen, Germany), featuring both 1.5T and 3T platforms (Supplementary Tables 1, 2).

Non-contrast T1-weighted (T1), contrast-enhanced T1-weighted Dixon fat-only (FO), and contrast-enhanced fat-suppressed T1-weighted (CE) series were extracted from the Digital Imaging and Communications in Medicine files. The thickness of the image acquisition ranged from 3 mm (341 series) to 4 mm (322 series), following the standard protocol of each hospital. The MRI standard protocols varied among the study centers, leading to differences in the composition of the imaging sequences used for training and testing. CE images can be obtained using conventional fat suppression or the Dixon technique. Conversely, the FO images were exclusively obtained using the Dixon technique. In Center 4, where the external validation data were collected, the Dixon technique was not employed, leading to a lack of FO images. In contrast, the MRI protocols in other centers included the Dixon technique.

Patient movement during image acquisition may introduce variability in lesion localization, leading to spatial registration discrepancies and potentially yielding misleading information from the ground-truth labels used as training data. Discrepancies were identified when pixels of images from other sequences did not overlap with the lesion drawn on the T1-weighted image by 50% or more. To address this issue, series with significant discrepancies were excluded from the analysis.

Image Processing and Data Generation

The matrix size in the dataset varied from 320 x 320 to 1296 x 1296 in the dataset. The images were initially

scaled to 256 x 256 and the pixel values were then min-max normalized to integers within the range of (0, 255). Subsequently, the images were resized to 224 x 224 pixels, with 24 slices. Contrast-limited adaptive histogram equalization was applied for image preprocessing. The training datasets were augmented using rotation (randomly within ± 20°), translation (vertical and horizontal axes), and zooming (0.9–1.1 range). Conversion of 2D T1-weighted sagittal images to 3D data was achieved through slice-to-volume registration using a commercial image analysis platform (AVIEW Research, Coreline Soft, Seoul, Korea). SimpleITK version 2.1.1, Numpy version 1.19.5, and Opencv version 4.5.4 were used for image processing (Fig. 1).

Ground Truth

To establish the ground truth, two musculoskeletal radiologists (with 12 and 8 years of experience) collaboratively reviewed the processed 3D images. To delineate metastases, cuboid bounding volumes of interest (VOIs) were created around lesions. The subsequent 3D segmentation process was conducted based on the radiologists' consensus. Segmentation precision was achieved through the application of a semiautomatic 3D segmentation algorithm and the graph-cut technique. A refinement step involving 2D segmentation correction was incorporated into the process. All procedures were performed using an image analysis platform (AVIEW Research; Coreline Soft). The methodological workflow is illustrated in Figure 1.

All observable osseous lesions were assessed for the probability of metastasis through consensus. The primary diagnostic criteria for metastasis on MRI included nodular hypointense lesions with obliterated fatty marrow signals on T1 or FO along with avid enhancement of the corresponding lesion on CE. The secondary diagnostic criteria included extraosseous soft tissue extension of the lesion and presence
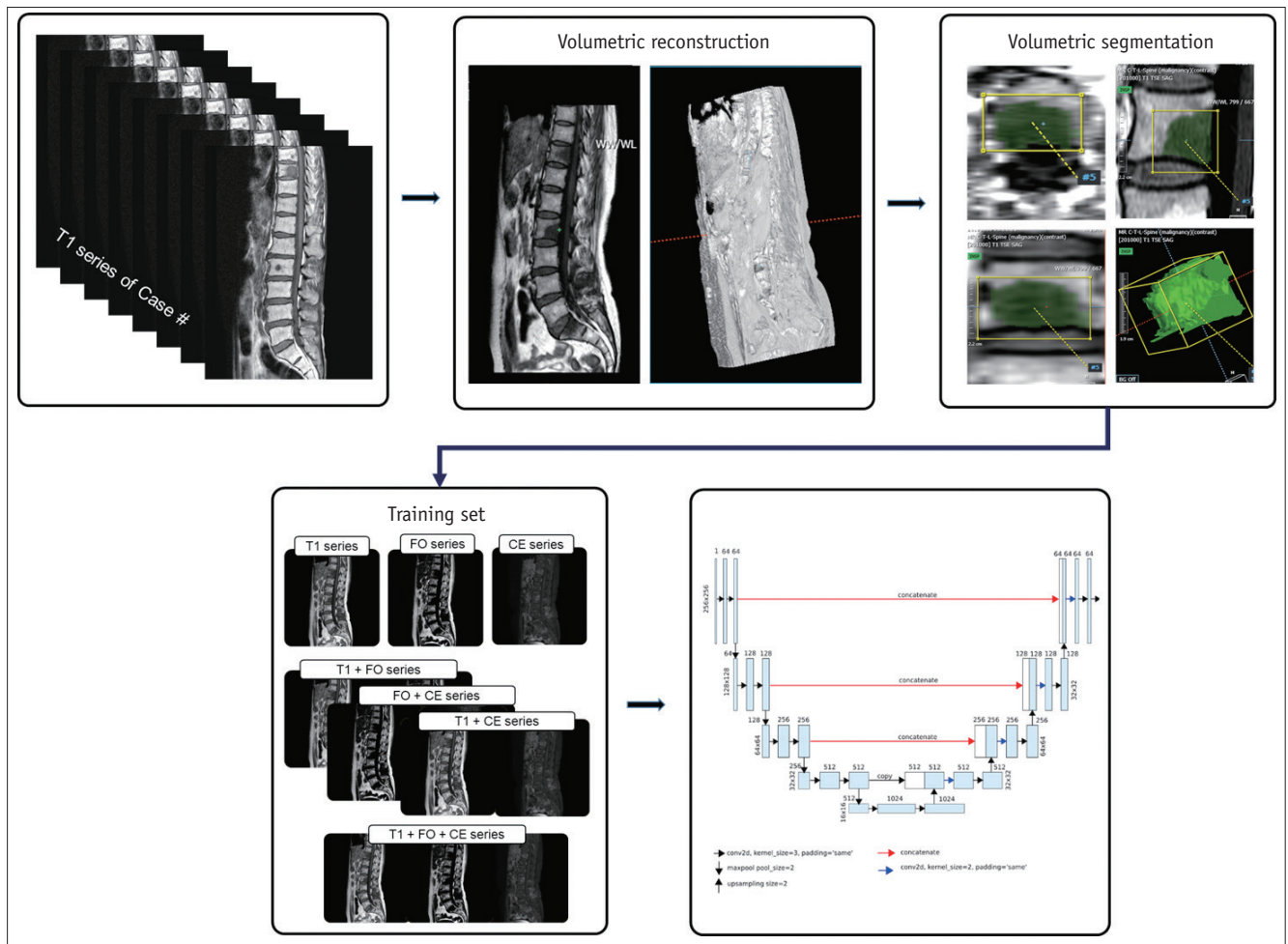

Fig. 1. Flowchart of the image data processing method and model development. T1 = non-contrast T1-weighted image, FO = contrast-enhanced T1-weighted Dixon fat-only image, CE = contrast-enhanced fat-suppressed T1-weighted image

of pathologic compression fractures. In patients in whom the diagnosis of metastasis remains uncertain, additional imaging studies such as PET, bone scan, chest or abdominal CT, performed within a month, were reviewed. If available, a follow-up whole-spine MRI was also examined to monitor the response or changes in the indeterminate lesion following cancer treatment.

The likelihood of metastasis was evaluated based on the following criteria. Indeterminate lesions partially met the primary diagnostic criteria, whereas probable metastatic lesions met the primary diagnostic criteria but lacked the observed features of the secondary diagnostic criteria. Lesions that fulfilled both primary and secondary diagnostic criteria were classified as definite metastasis. The volume of the lesion was calculated by multiplying the pixel size of each lesion with x-spacing, y-spacing, and slice thickness.

Pilot Study

During the preliminary phase, we conducted a pilot study to optimize imaging sequences for detecting bone metastases. Initially, T2-weighted images were included because of their good performance in brain and liver tumors [10-12]. However, in accordance with the observations of Hille et al. [11], T2-weighted images did not exhibit enhanced performance. Conversely, the FO sequence exhibited a superior contrast-to-noise ratio compared to the T1. Consequently, we excluded T2-weighted images and chose a combination of T1, FO, and CE images for the primary study modeling.

Data Characteristics

The dataset from Centers 1 to 3 was randomly split into training and test sets at a ratio of approximately 8:2. The area and number of metastases were also considered when assigning the training and test sets in an 8:2 ratio. Based on the MRI sequences, seven input data combinations (T1, FO, CE, T1 + FO, T1 + CE, FO + CE, T1 + FO + CE) were created. In combination models, different sequences were concatenated on the last axis of the NumPy array. For a 2D image, when data were represented as a 3D tensor, they included dimensions (height, width, and sequence). This implies that each sequence was considered a channel in the RGB images. The concatenation of sequences allowed the model to interpret the combined information across sequences, similar to the manner in which it processes different color channels in an RGB image.

Model Architecture and Training

Our models were trained using the 2D and 3D U-Net architectures, which are fully convolutional neural networks with customized hyperparameters. The model architectures were adopted from studies by Ronneberger et al. [13] and Çiçek et al. [14] (Supplementary Methods). Up-sampling layers with nearest-neighbor interpolation were used instead of up-convolution layers. A 3D U-Net extends a 2D tensor to a 3D tensor in the existing 2D U-Net. Training and evaluation of the models were performed in Python 3.8 using the Tensorflow library version 2.4.0, with four Nvidia GeForce GTX 1080 Ti GPUs.

Reader Study

Five radiologists, including three musculoskeletal radiologists (with 5–15 years of experience) and two musculoskeletal imaging (with 12-month training) from Center 1 participated in the study as independent and blinded test readers. None of them participated in the ground truth creation. Each radiologist received a uniform external dataset presented in randomized order. A radiologist reviewed the images and assessed the likelihood of metastasis for each detected lesion. They marked the VOI on the T1 image by drawing a cube encompassing the detected lesion using the image analysis platform (AVIEW Research, Coreline Soft).

Evaluation of Performance and Statistical Analysis

To assess segmentation performance, we employed the Sørensen–Dice coefficient, pixel-wise recall, and pixel-wise precision [15,16]. The detection performance of the selected models, chosen according to segmentation performance and data availability, and radiologists were assessed through the evaluation of per-lesion sensitivity, per-lesion positive predictive value, and free-response receiver operating characteristic (ROC) curve. Lesions detected by the model were considered true-positives if they had at least 10 pixels overlapping with a ground-truth area, and false-positives if they had less than 10 pixels overlapping with any of the ground-truth areas. Segmentation and detection performance were assessed with descriptive statistics and 95% confidence intervals [CIs], using the bootstrap technique. Continuous data were presented in mean ± standard deviation when applicable. Statistical analyses were conducted using the Scikit-learn library version 2.1.1 in Python.

RESULTS

Patient and Data Characteristics

A total of 662 MRI series from 302 patients (151 male and 151 female; age, 63.5 ± 11.5 years) obtained from three centers (Center 1, 220 series; Center 2, 210 series; Center 3, 232 series) were included as an internal dataset (Table 1, Fig. 2A). To train the models, 536 series were used. A total of 126 MRI series were used as the internal test sets. Notably, 4343 metastases were segmented, including 2519 lesions measuring ≤ 1.0 cm³. The number of lesions per patient was 12.5 ± 20.7.

For the external dataset, 49 MRI series of 20 patients (11 male and 9 female; 65.9 ± 11.5 years) were collected from Center 4. These were used for the external tests of the CE, T1, and T1 + CE models (Fig. 2B). Further details are provided in Table 1.

Segmentation Performance

Table 2 summarizes the segmentation performance of the selected models (2D CE, T1, and T1 + CE) in the external testing, with an illustrative example shown in Figure 3. Supplementary Table 3 presents the segmentation performance of all models in the internal test. The 2D models trained with T1 and CE sequences and T1, CE, and FO sequences showed consistently higher performance than that of other models in the internal testing. The

performance of the 3D models was consistently inferior to that of the 2D models. We chose the 2D T1 + CE model (and its single-image sequence models) for external testing, considering its performance in internal testing and the lack of an FO sequence in the external test set. The T1 + CE model showed a Dice coefficient of 0.699 (95% CI, 0.695–0.702) and pixel-wise recall of 0.653 (95% CI, 0.651–0.655) in external testing.

Detection Performance

Table 3 summarizes the per-lesion sensitivity for metastasis of the selected models (2D CE, T1, and T1 + CE), with an illustrative example shown in Figure 4. The T1 + CE model exhibited superior per-lesion sensitivity across all volume ranges in both internal and external testing. Notably, per-lesion sensitivity increased as lesion size increased. Despite the overall lower performance metrics, there were instances where the T1 or CE models detected lesions (true-positive detection by the T1 or CE models) that were overlooked by the T1 + CE model (Fig. 5). The per-lesion sensitivity for each center is presented in Supplementary Table 4.

Comparison of Detection Performance of Models with Human Readers

Table 4 presents an overview of the detection performance of the models and radiologists in external

Table 1. Characteristics of the patients and datasets

	Train set	Internal test set	External test set
Patients	242	60	20
Age, yr	63.5 ± 12.1	64.0 ± 9.7	65.9 ± 11.5
Sex			
Male	126	25	11
Female	116	35	9
Series*	536	126	49
Images	8615	2100	704
Labeled area, cm²	2478224	535952	223152
Metastasis lesions	3743	600	175
Volume of each metastasis, cm³			
≤ 0.5	769 (20.5)	81 (13.5)	24 (13.7)
> 0.5 to ≤ 1.0	1449 (38.7)	220 (36.7)	35 (20.0)
> 1.0 to ≤ 2.0	1016 (27.1)	200 (33.3)	34 (19.4)
> 2.0	509 (13.6)	99 (16.5)	82 (46.9)
Metastases per patient	13.1 ± 21.1	10.3± 13.6	8.75 ± 5.5

Data are number of patients or lesions with percentages in parentheses, or mean ± standard deviation.
*Series indicates a stack of T1-weighted sequences from each patient. For whole spine MRI is acquired in two or three subsequent scans for large coverage, from the skull base to the coccyx, MRI of each patient consists of two to three series of T1-weighted image, depending on the protocol of the study center and MRI machine
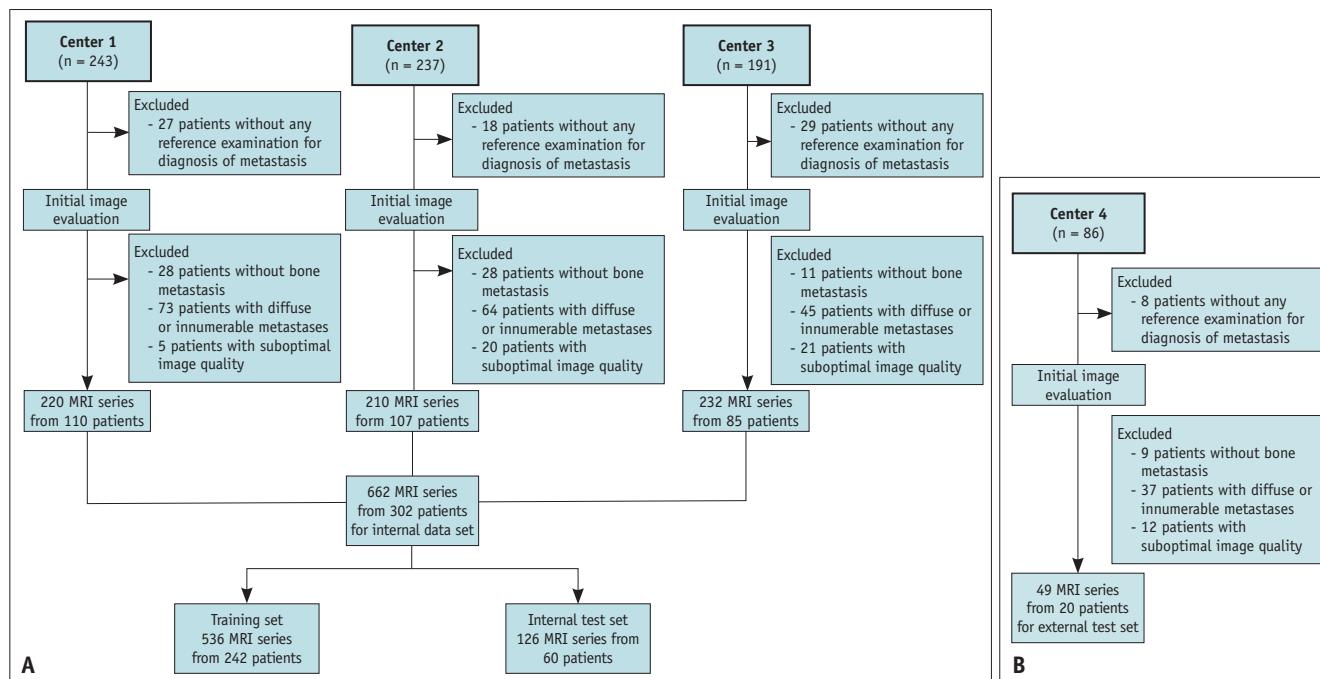
Fig. 2. Flowchart of data acquisition and selection. **A:** Internal data set. **B:** External test set.

Table 2. Segmentation performance at external test

Model	2D		
	Dice coefficient	Pixel-wise recall	Pixel-wise precision
T1 + CE	0.699 (0.695–0.702)	0.653 (0.651–0.655)	0.752 (0.750–0.754)
T1	0.570 (0.568–0.572)	0.465 (0.463–0.467)	0.736 (0.734–0.738)
CE	0.409 (0.404–0.413)	0.410 (0.408–0.412)	0.408 (0.406–0.410)

Data in parentheses are 95% confidence interval.
T1 = non-contrast T1-weighted image, CE = contrast-enhanced fat-suppressed T1-weighted image

testing. The T1 + CE model exhibited the highest per-lesion sensitivity of 0.857 (95% CI, 0.805–0.909) and per-lesion positive predictive value of 0.794 (95% CI, 0.766–0.821). The T1 model demonstrated a slightly higher per-lesion positive predictive value at 0.771 (95% CI, 0.709–0.834), whereas the CE model demonstrated a per-lesion sensitivity of 0.714 (95% CI, 0.647–0.781). The free-response ROC curves of the models are shown in Figure 6. In the reader study, the radiologists achieved a mean per-lesion sensitivity of 0.746 (95% CI, 0.543–0.960) and a mean per-lesion positive predictive value of 0.701 (95% CI, 0.511–0.882). Concerning false positives per series, the T1 + CE model demonstrated the lowest average at 0.384, followed by the radiologists at 1.338 (95% CI, -0.07–2.83), the T1 model at 0.797 (95% CI, 0.715–0.879), and the CE model at 2.263 (95% CI, 2.114–2.412).

False-Positive Model Results

The T1 + CE model, which exhibited the best performance,

presented false positive results involving 17 lesions mistakenly identified as metastases in the results of the internal test. These false-positive lesions consisted of compression fractures (n = 4), degenerative changes (n = 3), thyroid gland (n = 1), vertebroplasty (n = 1), piriformis muscle (n = 2), facet joint (n = 1), costovertebral joint (n = 1), intervertebral disc (n = 1), magnetic field inhomogeneity (n = 1), and contrast artifact (n = 2).

DISCUSSION

In the present study, we developed deep learning models to detect and segment bone metastases on spinal MRI and compared their performance with that of radiologists. The model trained with a combination of two sequences (T1 + CE) exhibited the best segmentation and detection performance in external testing. The per-lesion sensitivity of the T1 + CE model for metastases of > 1.0 cm^3 was 0.931 in the external test sets with a low false-positive rate. In the
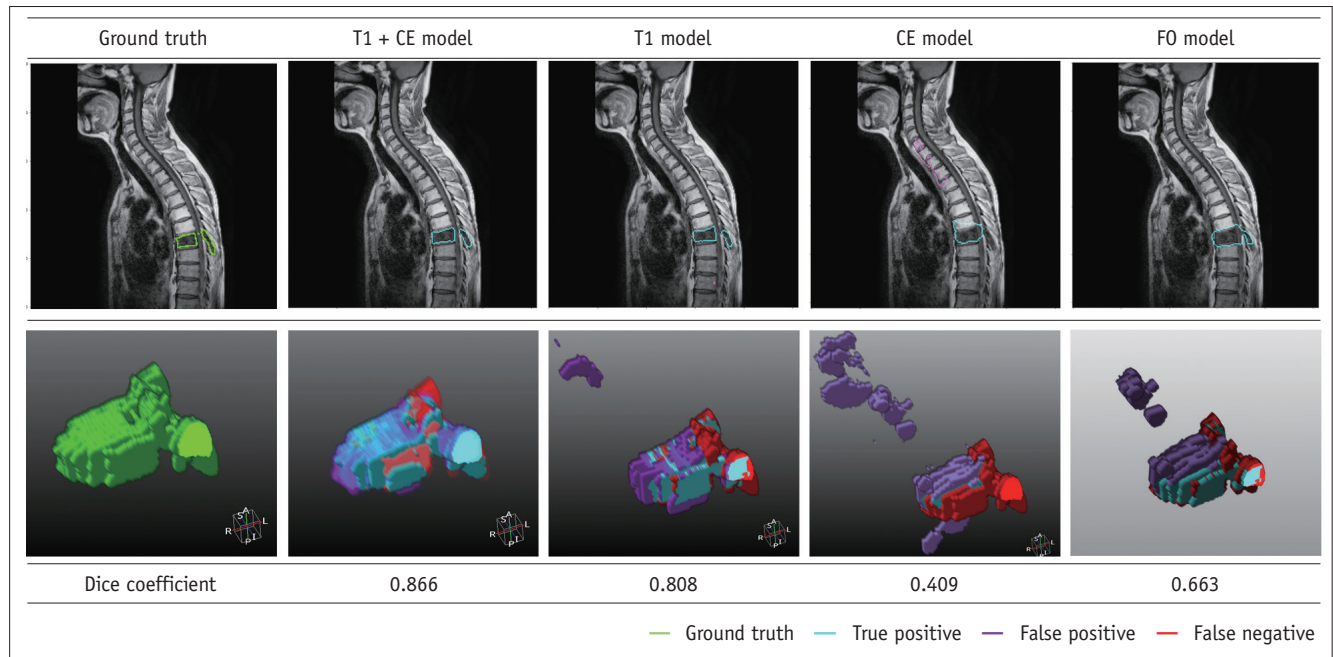
Fig. 3. Ground truth and prediction by models are labeled on the MRI of a 55-year-old female diagnosed with lung cancer. The images demonstrate the segmentation performance of the T1 + CE, T1, CE, and FO models. A prediction of the T7 vertebral body metastasis is outlined with light blue and shown as the volume of the blue region on volume rendering images. The prediction of the T1 + CE model shows more precise region segmentation than that of the rest of the models. T1 = non-contrast T1-weighted image, CE = contrast-enhanced fat-suppressed T1-weighted image, FO = contrast-enhanced T1-weighted Dixon fat-only image

Table 3. Per-lesion sensitivity of the models for metastasis detection

Lesion volume (cm³)	T1 + CE		T1		CE	
	Internal test	External test	Internal test	External test	Internal test	External test
≤ 0.5	0.457 (37/81)	0.667 (16/24)	0.136 (11/81)	0.542 (13/24)	0.296 (24/81)	0.333 (8/24)
> 0.5 to ≤ 1.0	0.814 (179/220)	0.743 (26/35)	0.450 (99/220)	0.686 (24/35)	0.459 (101/220)	0.600 (21/35)
> 1.0 to ≤ 2.0	0.915 (183/200)	0.794 (27/34)	0.705 (141/200)	0.735 (25/34)	0.725 (145/200)	0.588 (20/34)
> 2.0	0.990 (98/99)	0.988 (81/82)	0.980 (97/99)	0.890 (73/82)	0.889 (88/99)	0.927 (76/82)
Total	0.828 (497/600)	0.857 (150/175)	0.580 (348/600)	0.771 (135/175)	0.597 (358/600)	0.714 (125/175)

Data in parentheses are raw number (i.e., numerator/denominator).
T1 = non-contrast T1-weighted image, CE = contrast-enhanced fat-suppressed T1-weighted image

reader study, radiologists achieved a per-lesion sensitivity of 0.746 and a per-lesion positive predictive value of 0.701.

Despite progress in deep learning for other body parts, only a few approaches focus on diagnosing spinal metastases. Wang et al. [17] using the Siamese neural network architecture revealed the potential for the automated detection of spinal metastasis. They used only 26 cases of a single sequence (fat-suppressed T2-weighted MRI) as datasets. However, we used larger datasets of multisequence, multicenter, and multivendor MRI scans. Our findings provide further evidence of the model's application in clinical practice. Another study in which the researchers built a deep learning model based on the U-Net architecture reported a Dice coefficient per metastasis of 0.776, similar

to our results [15]. Although their results showed good segmentation accuracy, it should noted that the researchers used only 40 cropped images from the original MRI. Conversely, in our study, we used whole-image set of the spinal MRI as the training and testing sets, demonstrating similar or slightly superior performance compared with their results. Our models could predict metastases on original MRI, suggesting their applicability in actual clinical practice.

With advancements in the field of deep learning in medical imaging, there is an increasing interest in providing support to radiologists. Current technology cannot replace experienced radiologists in making diagnoses. Nevertheless, our reader study demonstrated comparable performance
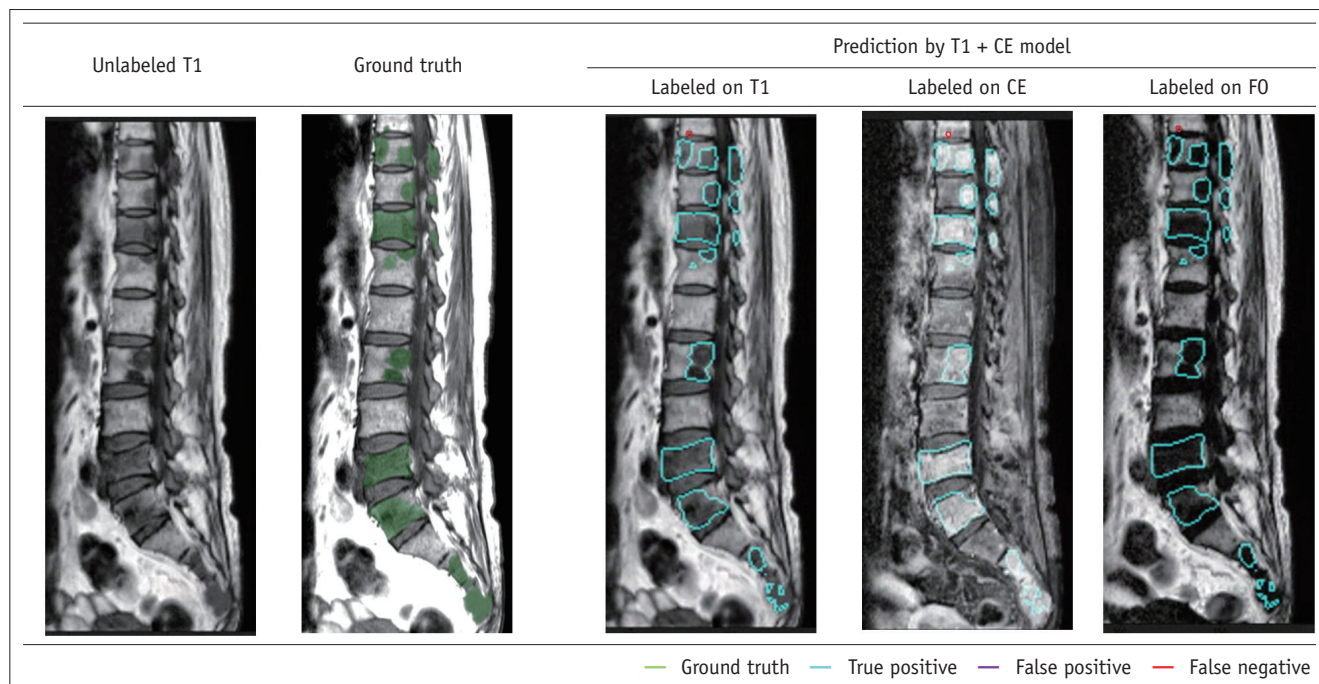
Fig. 4. Ground truth and prediction by the T1 + CE model are labeled on the MRI of a 62-year-old female diagnosed with breast cancer. The images illustrate the detection performance of the T1 + CE model. Each image sequence includes labeled indications of model-predicted true positive lesions, outlined in light blue, and false-negative lesions, outlined in red. The model predicted majority of the metastases, except for a diminutive lesion at T8 inferior endplate. T1 = non-contrast T1-weighted image, CE = contrast-enhanced fat-suppressed T1-weighted image, FO = contrast-enhanced T1-weighted Dixon fat-only image
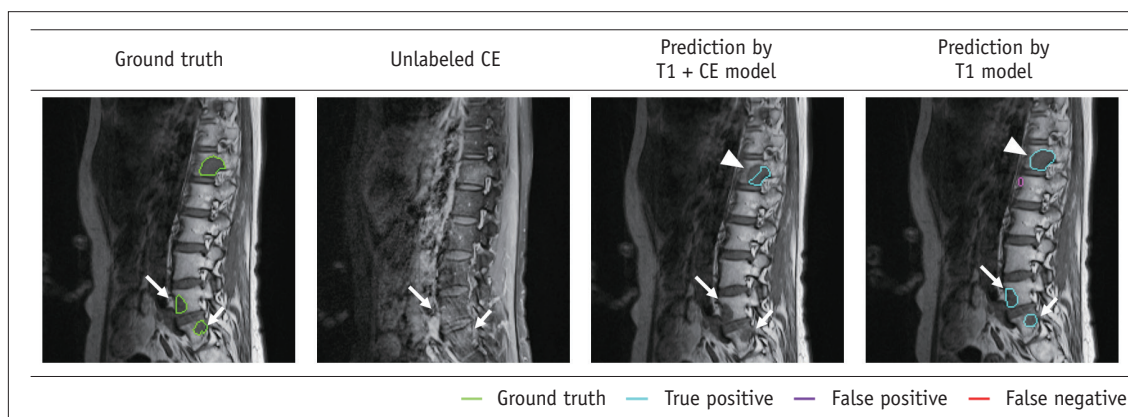
Fig. 5. Images of a 54-year-old female diagnosed with colon cancer. The metastases at L5 and the S1 vertebral bodies (arrows) were predicted by the T1 model, while the T1 + CE model missed both the lesions, showing faint enhancement on the CE. The T1 model also showed better segmentation prediction of another detected lesion (arrowheads) than that by the T1 + CE model. T1 = non-contrast T1-weighted image, CE = contrast-enhanced fat-suppressed T1-weighted image

Table 4. Detection performance of the models and radiologists in the external test

	T1 + CE	T1	CE	Radiologists*
Per-lesion sensitivity	0.857 (0.805–0.909)	0.771 (0.709–0.834)	0.714 (0.647–0.781)	0.746 (0.543–0.960)
Per-lesion positive predictive value	0.794 (0.766–0.821)	0.708 (0.678–0.739)	0.409 (0.388–0.429)	0.701 (0.511–0.882)
Average number of FP per series	0.384 (0.328–0.439)	0.797 (0.715–0.879)	2.263 (2.114–2.412)	1.338 (-0.071–2.833)

The data in parentheses are 95% confidence intervals.
*Data are the average values of five radiologists.
T1 = non-contrast T1-weighted image, CE = contrast-enhanced fat-suppressed T1-weighted image, FP = false positive
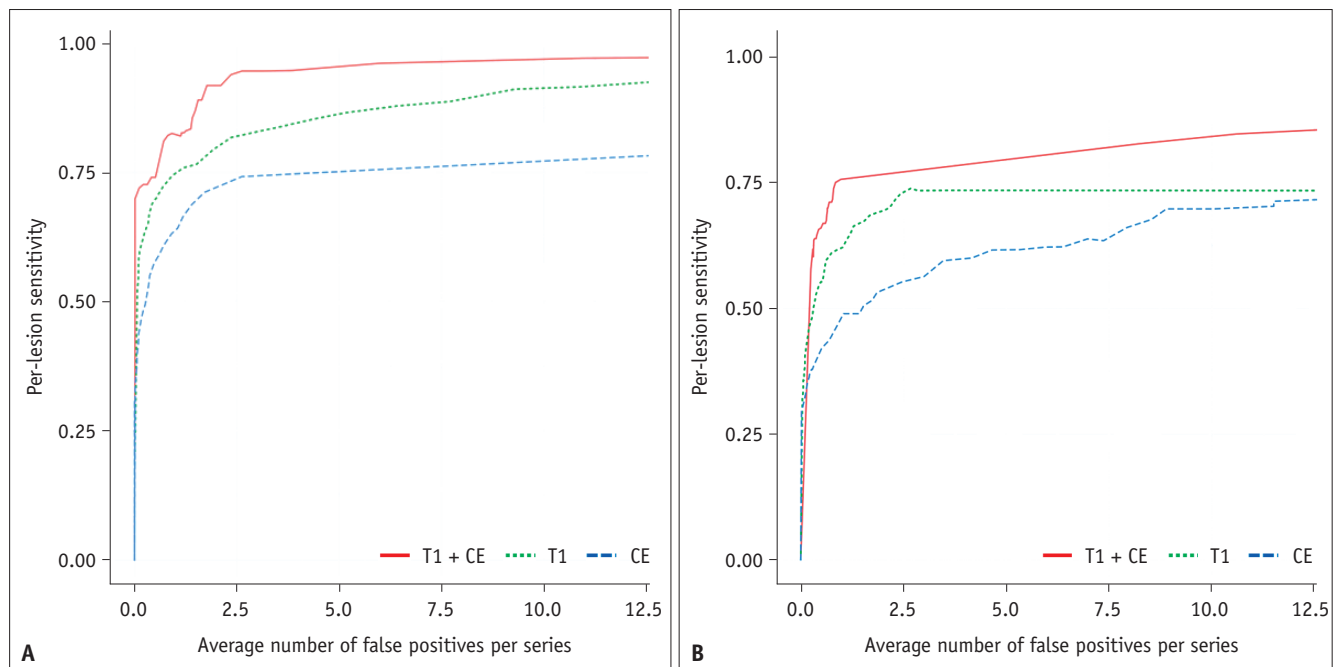
Fig. 6. Free-response receiver operating characteristics curve for the selected models in the internal **(A)** and external **(B)** test data sets. T1 = non-contrast T1-weighted image, CE = contrast-enhanced fat-suppressed T1-weighted image

of the combination (T1 + CE) model and radiologists. Its application in spinal metastasis detection can assist radiologists in identifying lesions more easily, especially those that are challenging to detect through routine radiological reviews. With the growing volume of medical images and radiologist workloads, further development of artificial intelligence is expected to play a role in improving diagnostic accuracy and efficiency in daily practice.

Our study has some limitations. First, the models were exclusively trained with metastatic cases, leading to false-positive predictions in benign lesions with similar appearances, such as compression fractures or Schmorl nodes. efinement is necessary to improve the distinction between lesions. Second, patients with global bone marrow signal alterations were excluded, as distinguishing metastases in the presence of diffusely altered bone marrow signal intensity is challenging for radiologists, complicating ground-truth drawing. This study serves as a preliminary step in metastasis detection and highlights the need for further development. Third, our dataset, although at baseline in each center, included images of pre- and post-treatment lesions. Post-treatment metastases may show different features, potentially affecting the models' performance. Finally, current models use sagittal images, limiting volumetric information. Future development with additional axial or 3D images is essential for comprehensive metastasis

diagnosis.

In conclusion, we developed an automated segmentation and detection model for spinal metastases using multisequence MRI obtained from three different centers. The combination model exhibited a high diagnostic performance in both segmentation and detection. This model may potentially support radiologists in clinical practice, leading to improvements in the diagnostic accuracy and efficiency for assessing spinal metastases.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2023.0671.

Availability of Data and Material
The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Conflicts of Interest
The authors have no potential conflicts of interest to disclose.

Author Contributions
Conceptualization: Dong Hyun Kim, Jiwoon Seo. Data

ORCID IDs
Dong Hyun Kim
   https://orcid.org/0000-0002-3871-7002
Jiwoon Seo
   https://orcid.org/0000-0002-1810-2342
Ji Hyun Lee
   https://orcid.org/0000-0001-5485-2776
Eun-Tae Jeon
   https://orcid.org/0000-0002-7000-9719
DongYoung Jeong
   https://orcid.org/0000-0001-8471-1595
Hee Dong Chae
   https://orcid.org/0000-0003-2624-1606
Eugene Lee
   https://orcid.org/0000-0003-4205-2362
Ji Hee Kang
   https://orcid.org/0000-0001-6498-9813
Yoon-Hee Choi
   https://orcid.org/0000-0001-6651-7908
Hyo Jin Kim
   https://orcid.org/0000-0002-0549-5722
Jee Won Chai
   https://orcid.org/0000-0003-1630-1863

REFERENCES

1. Coleman RE. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. *Cancer Treat Rev* 2001;27:165-176
2. Levack P, Graham J, Collie D, Grant R, Kidd J, Kunkler I, et al. Don't wait for a sensory level--listen to the symptoms: a prospective audit of the delays in diagnosis of malignant cord compression. *Clin Oncol (R Coll Radiol)* 2002;14:472-480
3. Haanen JB, Robert C. Immune checkpoint inhibitors. *Prog Tumor Res* 2015;42:55-66
4. Weber KL, Lewis VO, Randall RL, Lee AK, Springfield D. An approach to the management of the patient with metastatic bone disease. *Instr Course Lect* 2004;53:663-676
5. Ait Skourt B, El Hassani A, Majda A. Lung CT image segmentation using deep neural networks. *Procedia Comput Sci* 2018;127:109-113
6. Chen H, Dou Q, Yu L, Qin J, Heng PA. VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. *Neuroimage* 2018;170:446-455
7. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, et al. Fully automated deep learning system for bone age assessment. *J Digit Imaging* 2017;30:427-441
8. Zhao Z, Pi Y, Jiang L, Xiang Y, Wei J, Yang P, et al. Deep neural network based artificial intelligence assisted diagnosis of bone scintigraphy for cancer bone metastasis. *Sci Rep* 2020;10:17046
9. Sim Y, Chung MJ, Kotter E, Yune S, Kim M, Do S, et al. Deep convolutional neural network-based software improves radiologist detection of malignant lung nodules on chest radiographs. *Radiology* 2020;294:199-209
10. Kim J, Min JH, Kim SK, Shin SY, Lee MW. Detection of hepatocellular carcinoma in contrast-enhanced magnetic resonance imaging using deep learning classifier: a multi-center retrospective study. *Sci Rep* 2020;10:9458
11. Hille G, Dünnwald M, Becker M, Steffen J, Saalfeld S, Tönnies K. *Segmentation of vertebral metastases in MRI using an U-Net like convolutional neural network*. In: Handels H, Deserno T, Maier A, Maier-Hein K, Palm C, Tolxdorff T, eds. *Bildverarbeitung für die medizin 2019*. Wiesbaden: Springer Vieweg, 2019:31-36
12. Grøvik E, Yi D, Iv M, Tong E, Rubin D, Zaharchuk G. Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI. *J Magn Reson Imaging* 2020;51:175-182
13. Ronneberger O, Fischer P, Brox T. *U-net: convolutional networks for biomedical image segmentation*. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical image computing and computer-assisted intervention–MICCAI 2015*. Cham: Springer, 2015:234-241
14. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. *3D U-Net: learning dense volumetric segmentation from sparse*

annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, eds. *Medical image computing and computer-assisted intervention–MICCAI 2016*. Cham: Springer, 2016:424-432

15. Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945;26:297-302

16. Sørensen TJ. A method of establishing groups of equal amplitude in plant sociology based on similarity of species

content and its application to analyses of the vegetation on Danish commons [dissertation]. Copenhagen: I Kommission hos E. Munksgaard, 1948

17. Wang J, Fang Z, Lang N, Yuan H, Su MY, Baldi P. A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. *Comput Biol Med* 2017;84:137-146