Differentiating Multiple Myeloma and Osteolytic Bone Metastasis on Spinal CT Scan: A Comprehensive Study Using Convolutional Neural Network

Spinal BT Taramasında Multipl Miyelom ve Osteolitik Kemik Metastazının Ayırt Edilmesi: Konvolüsyonel Sinir Ağı Kullanan Kapsamlı Bir Çalışma

Muhammet Kursat SİMSEK [1] , Yusuf Kenan CETİNOGLU [2] , Resul BIRCAN [3] , Ali BALCI [4]

[1]Department of Radiology, Baskent University Hospital, Ankara, TÜRKİYE
[2]Department of Radiology, Behçet Uz Health Research Center For Pediatric Diseases And Surgery Hospital, İzmir, TÜRKİYE
[3]Department of Orthopedics and Traumatology, Dr. Abdurrahman Yurtaslan Ankara Oncology Training And Research Hospital, Ankara, TÜRKİYE
[4]Department of Radiology, Dokuz Eylul University Hospital, İzmir, TÜRKİYE

Abstract

Background: Accurate differentiation of spinal multiple myeloma (MM) and osteolytic metastatic bone tumor (OMBT) can be challenging. Usually, imaging methods, laboratory tests, and biopsy are performed for the correct diagnosis. In this study, we aimed to differentiate CT images from patients with MM and OMBT using CNN models.
Materials and Methods: 3707 CT images of 91 patients (1886 OMBT images and 1821 MM images; 46 males and 45 females; mean age: 61.2 years) obtained between January 2015 and January 2023 were reviewed. 2667 images were randomly selected for the training set, 740 for the validation set, and 300 for the test set. A transfer learning approach was used based on DenseNet121, DenseNet169, EfficientNetB0, MobileNet, MobileNetV2, VGG16, and Xception CNN architectures. The performance of the models was evaluated.
Results: When the sensitivity, specificity, positive predictive value, negative predictive value, accuracy, F1 score, and kappa measurements of the models in the MM and OMBT differentiation are evaluated, the most successful ones are MobileNetV2, MobileNet, and VGG16, with accuracy of 88%, 86.33%, and 86%, respectively.
Conclusions: Our study showed that CNN-based artificial intelligence models can differentiate MM and OMBT on CT images.

Keywords: Artificial Intelligence, Multiple Myeloma, Spinal Metastasis

Öz

Amaç: Spinal multipl miyelom (MM) ve osteolitik metastatik kemik tümörünün (OMKT) doğru bir şekilde ayırt edilmesi zorlayıcı olabilir, genellikle doğru tanı için görüntüleme yöntemleri, laboratuvar testleri ve biyopsinin kombinasyonu uygulanır. Bu çalışmada, MM ve OMKT hastalarından elde edilen BT görüntülerini CNN modelleri kullanarak ayırt etmeyi amaçladık.
Materyal ve Metod: Ocak 2015 ile Ocak 2023 arasında elde edilen 91 hastanın (1886 OMKT görüntüsü ve 1821 MM görüntüsü; 46 erkek ve 45 kadın; ortalama yaş: 61,2 yıl) 3707 BT görüntüsü incelendi. Eğitim seti için 2667, doğrulama seti için 740, test seti için 300 görüntü rastgele seçildi. DenseNet121, DenseNet169, EfficientNetB0, MobileNet, MobileNetV2, VGG16 ve Xception CNN mimarilerine dayanan bir transfer öğrenimi yaklaşımı kullanıldı. Modellerin performansı değerlendirildi.
Bulgular: Modellerin MM ve OMKT ayrımındaki duyarlılık, özgüllük, pozitif prediktif değer, negatif prediktif değer, doğruluk, F1 skoru ve kappa ölçümleri değerlendirildiğinde, en başarılı modeller sırasıyla %88, %86,33 ve %86 doğruluk oranları ile MobileNetV2, MobileNet ve VGG16 olmuştur.
Sonuç: Çalışmamızda CNN tabanlı yapay zekâ modellerinin BT görüntülerinde MM ve OMKT'yi ayırt edebileceğini gösterdik.

Anahtar Kelimeler: Yapay Zekâ, Multipl Miyelom, Spinal Metastaz

Introduction

Multiple myeloma (MM) is the most common primary malignant bone tumor while the most common malignant bone tumor is metastatic cancer (1). Both are generally seen in patients over 40 years old and present themselves with multiple osteolytic lesions (2). The most common site of osteolytic metastatic bone tumors (OMBT) is the spine where MM also generally presents itself (3). Distinguishing between these two entities is crucial, as their management strategies significantly differ, with implications for patient prognosis and overall survival.

Accurate identification of spinal lesions can be challenging, usually a combination of imaging methods, laboratory tests, and biopsy is performed for the correct diagnosis (4). Recent advancements in machine learning, specifically Convolutional Neural Networks (CNNs), have revolutionized the field of medical imaging analysis (5, 6). CNNs can automatically learn and extract relevant features from images, enabling them to discern intricate patterns and make accurate predictions. By leveraging this technology, we aim to develop a CNN-based approach that can effectively differentiate between spinal lesions in MM and OMBT using computed tomography (CT) scans which is the primary diagnostic imaging modality for MM, as proposed by the International Myeloma Working Group (IMWG) (4). Such an approach could streamline the diagnostic process, provide prompt and accurate identification of the underlying pathology, and guide appropriate treatment decisions.

In this study, we aimed to differentiate CT images from patients with MM and OMBT using different CNN models. To the best of our knowledge, this is the first study in the literature aimed at differentiating CT images of these two entities using CNN. We anticipate that our CNN-based approach will exhibit high sensitivity and specificity, offering a reliable and efficient tool for distinguishing between MM and OMBT in routine clinical practice. This study demonstrates the potential of the use of artificial intelligence technologies in cancer diagnosis and will form an important basis for future studies in the field of spinal lytic tumors.

Materials and Methods

This retrospective study was approved by the university institutional review board (Application no: 2023/18-19) and was conducted under the Declaration of Helsinki. Patient consent was waived.

Patient selection:

Non-contrast-enhanced spinal CT images of patients with MM and OMBT between January 2015 and January 2023 were evaluated.

Previous spinal surgery, unknown pathology diagnosis, and artifacts that impair diagnostic quality were exclusion criteria from the study. CT images were obtained from two scanners (Philips Brilliance 64; Philips Medical Systems, Best, The Netherlands, and Siemens SOMATOM Definition AS+, Siemens Healthcare, Germany).

Patient dataset:

All images on CT scans were evaluated by two neuroradiologists with seven years of neuroradiology experience (M.K.S. and Y.K.C.). DICOM images were analyzed with Sectra Workstation version 23.2.2.5087 and ProbelViewer programs. The spinal CT slices demonstrating pathology specific to OMBT and MM were manually selected through the interfaces of the programs. Images without pathology were excluded. Images were converted from DICOM (Digital Imaging and Communications in Medicine) format to JPEG (Joint Photographic Experts Group) format. The demographic data of the cases were anonymized. Selected images for both were divided into the relevant class. A dataset was created using a total of 3707 images, including 1886 OMBT images and 1821 MM images. 2667 images were randomly selected for the training set, 740 images for the validation set, and 300 images for the test set.

Training and validation:

The transfer learning method was used to differentiate MM and OMBT. DenseNet121, DenseNet169, EfficientNetB0, MobileNet, MobileNetV2, VGG16, and Xception, among the models that showed high performance in the ImageNet competition, were selected as the CNN models in which the transfer learning approach will be used. Xception accepts a 299x299 pixel image as its input shape, while in other models the input shape was 224x224 pixels. CT images were resized as accepted by the models and presented from three channels. The classifier layers of CNN models have been fine-tuned. The last classifier layers of the CNN models were replaced with two fully connected layers of 512 and 256 nodes, respectively. Rectified Linear Unit (ReLU) and SoftMax were used as the activation function. Pretrained weights from ImageNet were applied. The batch number was selected as 16. Epoch is set to 100 for all CNN models. The pipeline and architecture of the fine-tuned models are shown in figure 1.

Google Colaboratory (Colab), an open-source cloud-based tool owned by Google, was used for training, and testing all models.

Results

In our study, spinal CT images of 39 (43%) patients with MM and 52 (57%) patients with OMBT were reviewed. The causes of OMBT included in the study were: lung cancer (n=18), breast cancer (n=18), gastrointestinal cancer (n=7), hepatocellular cancer (n=2), neuroendocrine cancer (n=1), renal cell cancer (n=1) and prostate cancer (n=1). The primary cancer of 3 metastatic patients was unknown. In total, 45 (49%) of 91 cases were female and 46 (51%) were male. The mean age was 61.2 years (standard deviation ±5.6). All demographic data are shown in table 1.
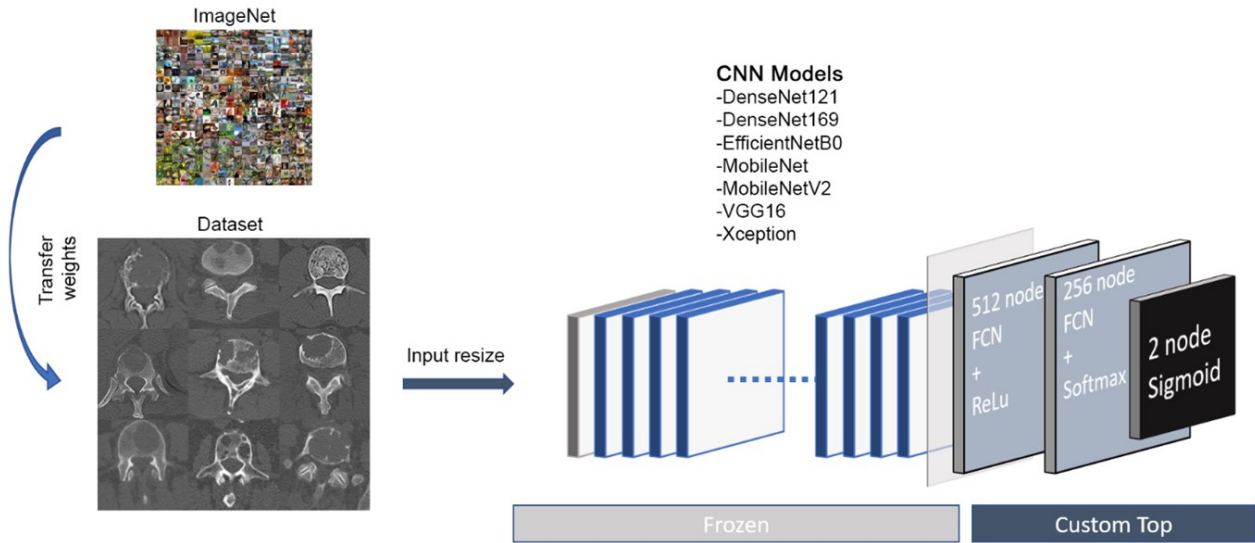
Figure 1. Artificial intelligence pipeline

Table 1. Demographics

	Myeloma	Metastasis	Total
Gender (n, %)			
Male	21 (53.8%)	25 (48.1%)	46 (50.5%)
Female	18 (46.2%)	27 (51.9%)	45 (49.5%)
Total	39 (100%)	52 (100%)	91 (100%)
Age (mean, Interval, Standard Deviation)	63.1 (23-83)	59.8 (34-85)	61.2 (23-85, ±5.6)

Performance of CNN models:

The highest accuracy rate in the CNN models we created in the differentiation of MM and OMBT was obtained in MobileNetV2 with 88%, and the lowest accuracy rate was obtained in EfficientNetB0 with 74.67%. Sensitivity, specificity, positive predictive value, negative predictive value, accuracy, F1-score, and kappa values obtained from the confusion matrix of the models are given in Table 2 and Figure 2 for each model. In addition, the ROC curve of the models is shown in figure 3.

Learning in the training graphs of pre-trained CNN models started after the first epochs and stabilized at 10-40 epochs. When the loss graphs of the CNN models are evaluated, it is seen that the loss decreases progressively in both training and validation sets, which shows that there is no overfitting problem. The accuracy and loss graphs for the fine-tuned models with the top three are presented in figure 4.

Table 2. Evaluation of models in MM and OMBT differentiation

|	DenseNet121	DenseNet169	EfficientNetB0	MobileNet	MobileNetV2	VGG16	Xception
Statistic	Value	Value	Value	Value	Value	Value	Value
Sensitivity	79.89%	80.86%	78.46%	84.28%	83.93%	81.03%	69.59%
Specificity	91.27%	86.23%	71.76%	88.65%	93.18%	92.86%	85.85%
Positive Predictive Value	92.67%	87.33%	68.00%	89.33%	94.00%	94.00%	90.00%
Negative Predictive Value	76.67%	79.33%	81.33%	83.33%	82.00%	78.00%	60.67%
Accuracy	84.67%	83.33%	74.67%	86.33%	88.00%	86.00%	75.33%
F1 Score	0.858	0.8397	0.7286	0.8673	0.8868	0.8704	0.7849
kappa	0.693	0.667	0.493	0.7227	0.760	0.720	0.507
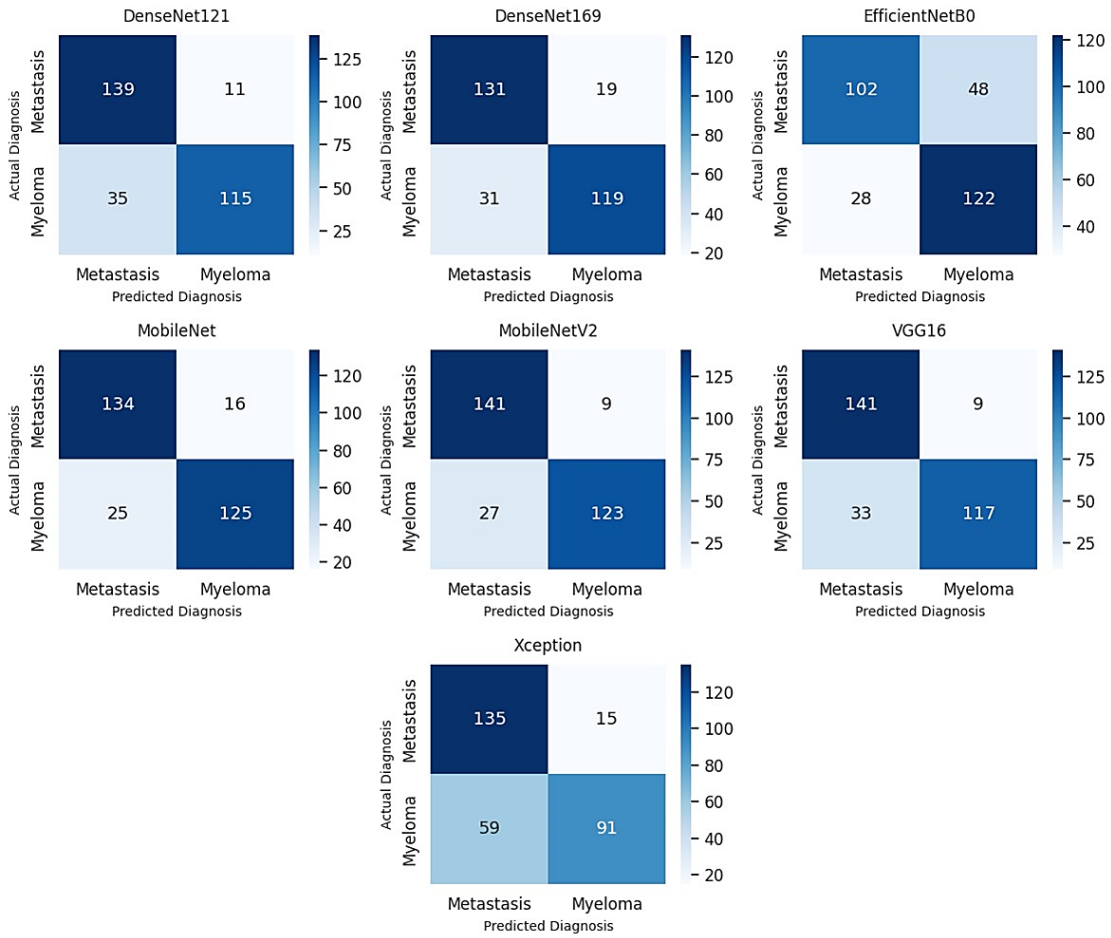
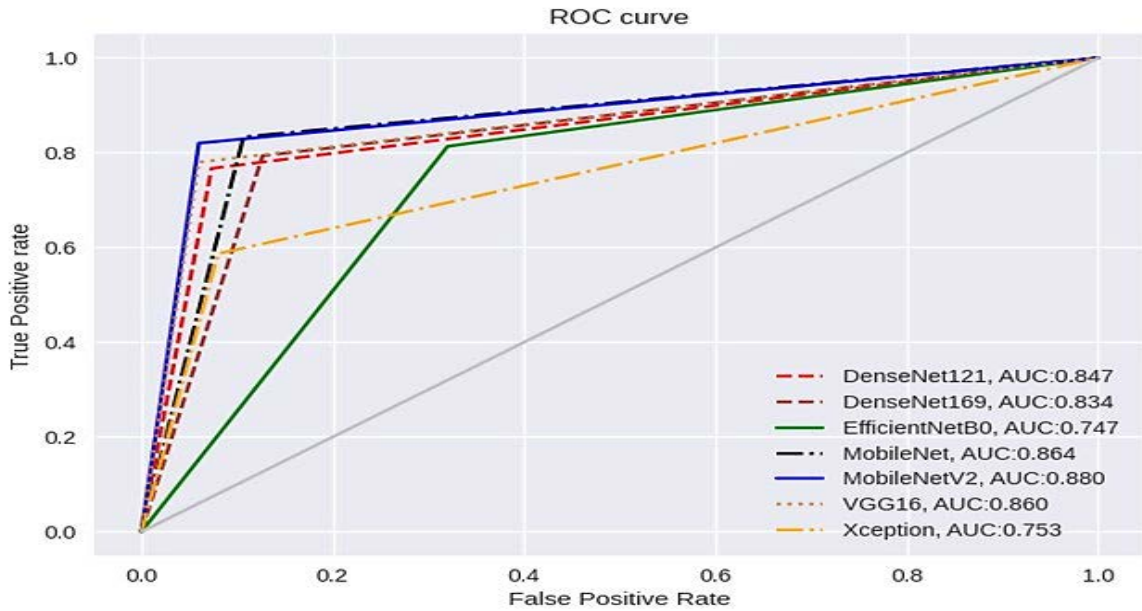**Figure 2.** Confusion matrices of the models

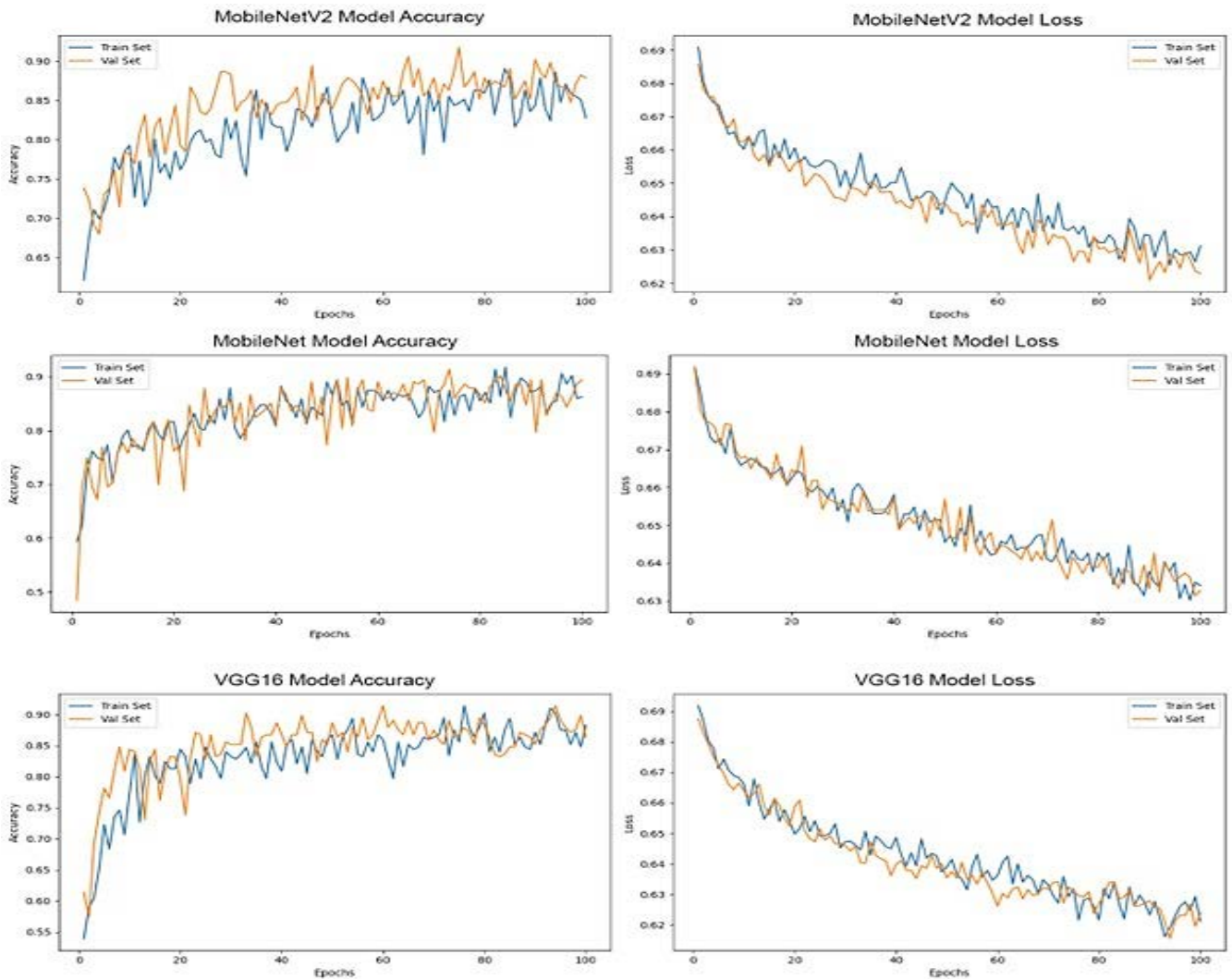**Figure 3.** Roc curves of the models

Figure 4. Accuracy and lose graphs

Discussion

In daily practice, CT scans for spinal bones are used a lot. Osteolytic bone lesions are the most commonly seen lesions in the spine. The spinal lesions detected on CT scans have both benign and malignant causes. The differential diagnosis of malignant causes includes MM and OMBT. Histopathological diagnosis of lesions with bone biopsy is the gold standard examination for benign and malignant tumors. Since the histopathological examination is an invasive procedure, clinical and radiological examinations play an important role in the diagnosis and follow-up of the disease in the differentiation of spinal lesions (7). Differentiation of MM and OMBT has an important role in the treatment planning of the patient. In patients with no prior known disease, radiological differentiation of MM and OMBT is very challenging. Moreover, publications are stating that pathologically diagnosed spinal bone lesions can be missed by the radiologist up to 24% (8). MM and OMBT are difficult to distinguish from each other in conventional radiology, therefore CNN-based deep learning models may have benefits such as increasing diagnostic accuracy and shortening the diagnosis time for these diseases.

In our study, CNN-based deep learning models were applied to differentiate the MM from OMBT. In our study, the accuracy of the VGG16 model was 86%, the accuracy of MobileNet was 86%, and the accuracy of the MobileNetV2 model was 88%. We used images from two different CT devices. The use of images from more than one CT device shows that the success of the models is generalizable. To the best of our knowledge, our study is the first study in the literature that uses CNN-based deep learning models to differentiate spinal MM and OMBT in CT. Among our CNN models, the most successful was the fine-tuned MobileNetV2 with 88% accuracy. The least successful model was EfficientNetB0. According to an article on the Keras API website, the difficulty of defining hyperparameters may cause EfficientNet variants to perform less successfully in classification tasks with fewer classes (9). In addition, there are studies in the literature showing that MobileNetV2 is more successful than EfficientNet in binary classification problems (10,11).

In a similar study, Lang et al. assessed lung cancer and other organ metastases in spinal bones using artificial intelligence

in DCE-MRI images. In their research, the accuracy achieved with conventional CNN ranged from 0.61 to 0.74, with a mean of 0.71 and a standard deviation of 0.043. In our study, all the CNN models employed demonstrated higher accuracy than the results reported by Lang et al. The improved performance of our models could be attributed to the use of the transfer learning approach. In this approach, the ImageNet dataset is utilized to pre-train the CNN, and the acquired functional weights are subsequently applied to the training of new custom datasets. As a result, this approach enhances the success of the models. Another possible explanation for the superior performance of our model may be the use of CT images instead of MRI. Because MRI imaging parameters may vary in different centers. MRI signals can be influenced by various parameters which can lead to magnetic field inhomogeneity. Moreover, CT images are more standardized than MRI images and are not affected by other signal differences, which may account for this outcome (12).

Xiong et al. used lumbar MRI images to distinguish between spinal metastasis and MM by texture analysis. In this study, they found that machine learning-based texture analysis can differentiate metastasis and MM in the lumbar spine. In their study, they achieved accuracy, sensitivity, and specificity performance of 0.815, 0.879, and 0.790, respectively, in the validation cohort with the Artificial Neural Networks classifier on T2WI images (13). Our study used CNN-based deep learning models, which are more advanced artificial intelligence techniques compared to the ones used in this study. Moreover, while the highest accuracy in this study was 81.5% in the validation set, in our study the accuracy was higher in the external validation set. Another advantage of our study is the ability to distinguish between OMBT and MM not only in lumbar vertebrae but also in thoracic vertebrae.

Baykara et al. aimed to differentiate MM and OMBT by histogram analysis using ADC maps in MRI images. They mentioned that the lower ADC values in myeloma patients compared to metastasis can be used in this differentiation (14). The deep learning approach is more feasible than histogram analysis. CNN models used in our study offer suggestions that will facilitate fast, effective, and clinical functioning.

There are deep learning studies performed to detect lesions in spinal bones. These studies can distinguish spinal lesions as lytic, sclerotic, and mixed types. It has also been shown that deep-learning models can detect spinal metastasis at a high rate (15,16). The aim of the studies in the literature is generally for lesion detection. Our study is the first to demonstrate successful differentiation of MM and OMBT using CT images in spinal bone lesions with CNN-based deep learning models.

Chen et al. evaluated MRI images in the differentiation of spinal MM and lung cancer metastasis by radiomics and deep learning model. Their study shows that the deep learning model has a stronger capability in differential diagnosis than that of the radiomics model and radiologist assessment (7). In our study, the differentiation of MM and OMBT is not limited to lung cancer metastasis. Another advantage of our

study is that this distinction can be made successfully in CT, which is cheaper than MRI.

There are publications in the literature that aim to differentiate spinal MM and OMBT. Among these publications, Mutlu et al., in their study of 207 patients with CT images, defined that features such as homogeneity of the lesion, high density, perilesional sclerosis, and border features may be useful in the differentiation of MM and OMBT. Despite the features that help in distinguishing these two conditions, they concluded that the distinction is not very clear in most patient groups (17). In addition, the subjective nature of the defined features is another diagnostic challenge. The use of deep learning models can yield useful findings in lesion differentiation. Subjective evaluation will be prevented thanks to the automation made with deep learning models. In addition to the conventional findings, our study shows artificial intelligence applications can make a significant contribution to this differentiation.

There are some limitations in our study. Although the loss graphs of the models in both the training and validation sets show a decreasing trend, indicating no over-fitting problem, the use of multiple images from the same patient may have resulted in higher model performance than expected. Furthermore, given the limited number of patients in our dataset despite the large number of images, further comprehensive studies are required such as including data from more diverse sources or a prospective study setup.

Consequently, it is very difficult to distinguish between MM and OMBT with CT images in daily practice. Our study showed that CNN-based artificial intelligence models can contribute to conventional radiological evaluations in this distinction with high success. Thanks to technological developments in models and artificial intelligence programs, the distinction between MM and OMBT will be easier.

References

1. van de Donk NWCJ, Pawlyn C, Yong KL. Multiple myeloma. Lancet. 2021;397(10272):410-427. doi: 10.1016/S0140-6736(21)00135-5.
2. Fornetti J, Welm AL, Stewart SA. Understanding the Bone in Cancer Metastasis. J Bone Miner Res. 2018;33(12):2099-2113. doi: 10.1002/jbmr.3618.

3. Sahgal A, Myrehaug SD, Siva S, Masucci GL, Maralani PJ, Brundage M, et all. Stereotactic body radiotherapy versus conventional external beam radiotherapy in patients with painful spinal metastases: an open-label, multicentre, randomised, controlled, phase 2/3 trial. Lancet Oncol. 2021;22(7):1023-1033. doi: 10.1016/S1470-2045(21)00196-0.

4. Hillengass J, Usmani S, Rajkumar SV, Durie BGM, Mateos MV, Lonial S, et all. International myeloma working group consensus recommendations on imaging in monoclonal plasma cell disorders. Lancet Oncol. 2019 Jun;20(6):e302-e312. doi: 10.1016/S1470-2045(19)30309-2. Erratum in: Lancet Oncol. 2019;20(7):e346. doi: 10.1016/S1470-2045(19)30423-1.

5. Gao R, Zhao S, Aishanjiang K, Cai H, Wei T, Zhang Y, et all. Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. J Hematol Oncol. 2021;14(1):154. doi: 10.1186/s13045-021-01167-2.

6. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. Skeletal Radiol. 2020;49(2):183-197. doi: 10.1007/s00256-019-03284-z.

7. Chen K, Cao J, Zhang X, Wang X, Zhao X, Li Q, et all. Differentiation between spinal multiple myeloma and metastases originated from lung using multi-view attention-guided network. Front Oncol. 2022;12:981769. doi: 10.3389/fonc.2022.981769.

8. D'Angelo T, Caudo D, Blandino A, Albrecht MH, Vogl TJ, Gruenewald LD, et all. Artificial intelligence, machine learning and deep learning in musculoskeletal imaging: Current applications. J Clin Ultrasound. 2022;50(9):1414-1431. doi: 10.1002/jcu.23321.

9. Y. Fu. Image classification via fine-tuning with EfficientNet, 2020. [accessed June 25, 2023, updated July 10, 2023]. 2023 [cited 2023 Jun 25]. Available from: https://keras.io/examples/vision/image_classification_efficientnet_fine_tuning/

10. Vollmer A, Saravi B, Vollmer M, Lang GM, Straub A, Brands RC, et all. Artificial Intelligence-Based Prediction of Oroantral Communication after Tooth Extraction Utilizing Preoperative Panoramic Radiography. Diagnostics (Basel). 2022;12(6):1406. doi: 10.3390/diagnostics12061406.

11. Cetinoglu YK, Koska IO, Uluc ME, Gelal MF. Detection and vascular territorial classification of stroke on diffusion-weighted MRI by deep learning. Eur J Radiol. 2021;145:110050. doi: 10.1016/j.ejrad.2021.110050.

12. Lang N, Zhang Y, Zhang E, Zhang J, Chow D, Chang P, et all. Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI. Magn Reson Imaging. 2019;64:4-12. doi: 10.1016/j.mri.2019.02.013.

13. Xiong X, Wang J, Hu S, Dai Y, Zhang Y, Hu C. Differentiating Between Multiple Myeloma and Metastasis Subtypes of Lumbar Vertebra Lesions Using Machine Learning-Based Radiomics. Front Oncol. 2021;11:601699. doi: 10.3389/fonc.2021.601699.

14. Baykara M, Yildirim M. Differentiation of multiple myeloma and metastases with apparent diffusion coefficient map histogram analysis. North Clin Istanb. 2022 Jul 5;9(3):256-260. doi: 10.14744/nci.2021.59376.

15. O'Connor SD, Yao J, Summers RM. Lytic metastases in thoracolumbar spine: computer-aided detection at CT--preliminary study. Radiology. 2007 Mar;242(3):811-6. doi: 10.1148/radiol.2423060260. Erratum in: Radiology. 2007;244(1):320.

16. Wels M, Kelm BM, Tsymbal A, Hammon M, Soza G, Sühling M, et al. Multi-stage osteolytic spinal bone lesion detection from CT data with internal sensitivity control. SPIE Proceedings. 2012;8315:831513. doi:10.1117/12.911169

17. Mutlu U, Balci A, Özsan GH, Özkal S, Şeyhanli A, Özgül HA. Computed tomography characteristics of multiple myeloma and other osteolytic metastatic bone lesions. Acta Radiol. 2021;62(12):1639-1647.