# Artificial intelligence in musculoskeletal oncology imaging: A critical review of current applications

Maxime Lacroix, Theodore Aouad, Jean Feydy, David Biau, Frédérique Larousserie, Laure Fournier, Antoine Feydy

Version of Record: https://www.sciencedirect.com/science/article/pii/S2211568422001929
Manuscript_d594cf29f19babc5ac984f270ee59bde

# Artificial intelligence in musculoskeletal oncology imaging: a critical review of current applications

## Authors

Maxime Lacroix [a,b,c*]
Theodore Aouad [d]
Jean Feydy [e]
David Biau [b,f]
Frédérique Larousserie [b,g]
Laure Fournier [a,b,c]
Antoine Feydy [b,h]

## Affiliations

[a] Department of Radiology, Hôpital Européen Georges Pompidou, Assistance Publique-Hôpitaux de Paris, Paris 75015, France.
[b] Université Paris Cité, Faculté de Médecine, Paris 75006, France.
[c] PARCC UMRS 970, INSERM, Paris 75015, France.
[d] Université Paris-Saclay, CentraleSupélec, Inria, Centre for Visual Computing, 91190 Gif-sur-Yvette, France.
[e] Université Paris Cité, HeKA team, Inria Paris, Inserm, 75006 Paris, France.
[f] Department of Orthopedic surgery, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris 75014, France.
[g] Department of Pathology, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris 75014, France.
[h] Department of Radiology, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris 75014, France.

*Corresponding author: maxime.lacroix@aphp.fr

**Artificial intelligence in musculoskeletal oncology imaging: a critical review of current applications**

**Abstract:**

Artificial intelligence (AI) methods are increasingly being studied in musculoskeletal oncology imaging. These tools have been applied to both primary and secondary bone tumors and assessed for various predictive tasks that include detection, segmentation, classification, and prognosis. Still, in the field of clinical research, further efforts are needed to improve AI studies reproducibility and reach an acceptable level of evidence in musculoskeletal oncology. This review describes the basic principles of the most common AI techniques including machine learning, deep learning and radiomics. Then, recent developments and current results of AI in the field of musculoskeletal oncology are presented. Finally, limitations and future perspectives of AI in this field are discussed.

**Keywords:**

Artificial intelligence; Bone tumor; Deep learning; Machine learning; Metastases

**Abbreviations:**

ADC: Apparent diffusion coefficient

AI: Artificial intelligence

AUC: Area under the operating characteristic curve

CNN: Convolutional neural network;

CT: Computed tomography;

DL: Deep learning;

DSC: Dice similarity coefficient;

FDG: Fluorodesoxyglucose;

PET/CT: Positron emission tomography computed tomography;

ML: Machine learning;

MRI: Magnetic resonance imaging;

PET-CT: Positron emission tomography with computed tomography;

RF: Random forest;

ROC: Receiver operating characteristic;

SPECT: Single photon emission computed tomography;

SVM: Support-vector machine;

## 1. Introduction

Artificial intelligence (AI) is increasingly used in clinical research for the study of both primary and secondary bone tumors (metastases) [1,2]. Accurate interpretation of bone tumors may be difficult for a general radiologist [3,4] and misdiagnosis can be detrimental to patient outcome [5]. Many patients with benign tumors are referred to bone biopsy leading to increased morbidity and cost [6,7]. Since imaging plays a key role in the assessment of bone tumors, it is expected that AI will facilitate and optimize its role in the future [8,9].

In this context, AI has been developed for a variety of applications, including predictive tasks such as detection, segmentation, classification and prognosis of malignant bone tumors [10,11]. However, while bone metastases are quite common, easier to identify or classify and thus better suited to AI training, primary tumors are rare and polymorphic and can thus be particularly challenging for AI models [11].

The purpose of this article was to sum-up and critically review recent literature about AI applied to musculoskeletal oncologic imaging, explain the main limitations, and discuss future developments.

## 2. Basic principles and process

AI refers to computer systems able to perform tasks that normally require human intelligence. This includes machine learning (ML) and deep learning (DL) models, while radiomics is another computer-assisted method commonly used to characterize tumors. These methods all rely on statistical tools and algorithmic structures that may be not familiar to clinicians. We thus start with a brief introduction to these topics.

### 2.1. Why do we do “learning”?

The main goal of AI in medical imaging is to automate tasks or subtasks that are currently performed by human operators [12]. In some favorable situations, clinical decision rules derived from expert medical knowledge can be directly translated into code, which enables "rule-based" models [13,14]. However, most tasks in image analysis are too complex to be solved using an explicit mathematical formula [15]. If "x" denotes a medical image

understood as a large array of numbers and "y" denotes a diagnosis understood as an integer label, designing from scratch a mathematical function "y = f(x)" that can reliably assign a diagnosis to an image is extremely difficult.

To create high-performance computer programs, engineers thus follow a two-steps workflow that distributes the complexity of the decision rule "f" between a relatively general mathematical structure and a task-specific dataset of example images. The first step is a "Design" step that specifies a general form for the model with a fixed mathematical structure "y = f(θ, x)" and a large number of free parameters θ. This program "architecture" is designed for a particular task (segmentation, classification, registration). The second step is a "Learning" or "training" step, which fits the parameters θ of the architecture according to the statistical analysis of a dataset of medical images. Engineers retrieve an optimized model "y = $f(\theta_{optimal}, x)$" whose numerical parameters $\theta_{optimal}$ have been fine-tuned to make the distinction between neighboring pathologies.

### 2.2 Overfitting and generalization

Most AI models are trained using "supervised learning", a generalization of least squares linear regression to complex models and data types [16,17]. Assuming that we have access to a set of input images "$x_i$" and corresponding diagnostics "$y_i$", we find the set of optimal parameters $\theta_{optimal}$ that induces the fewest errors "$y_i \neq f(\theta, x_i)$" on the dataset. This criterion is simple but may also lead to "overfitting": if the model architecture "y = f(θ, x)" is not constrained by expert knowledge, there is no guarantee that a trained model "y = $f(\theta_{optimal}, x)$" that performs well on the finite set of example images "$x_i$" will also extrapolate or "generalize" reliably to unseen images. Research is mainly concerned with the design of program structures that can "learn" important parameters from a training dataset while avoiding "overfitting" (Fig. 1). This needs to validate the performance and robustness of an AI model using an external dataset that has no overlap with the original training samples "$x_i$" is required.

### 2.3. Conventional statistics

Linear models provide a first baseline for learning tasks. Linear regression (for the prediction of numbers), logistic regression (for the prediction of class labels) and the Cox proportional hazards model (for survival analysis) all assume a simple relationship between a set of input

markers "x" and the output value "y": this makes them easy to implement and study mathematically [15]. However, linear models are also prone to overfitting when the number of markers available in every input "x" exceeds the number of patients. While having access to more patients is generally a good thing, using more markers per patient can be dangerous ("curse of dimensionality") [18].

### 2.4. Overcoming the "curse of dimensionality" with machine learning (ML)

ML methods are especially relevant in medical imaging, where we describe each patient using an image that contains millions of pixel values. To decrease the complexity of the learning task, ML methods first process the vast number of available markers (pixel values, physiological measurements) into a compact set of high-quality descriptors known as "features", before performing a robust statistical analysis.

### 2.5. Processing markers into hand-crafted features with radiomics

The design of a feature set often requires time-consuming interactions between domain experts and ML engineers. In order to streamline this process, researchers have developed a "standard toolbox" of quantitative features that provide a good baseline for image analysis. These mathematical formulas describe the shape, intensity distribution or texture characteristics of a region of interest and may be used in a wide range of settings [19,20]. Most of these "radiomic features" quantify the relationship between the value of a pixel and that of its close neighbors. In oncology, radiomic features may reflect tumor heterogeneity observed at the histological and genetic levels [21]. Radiomics is therefore largely investigated to assist cancer diagnosis, prognosis, and prediction of response to therapy. The main advantage of standard radiomic features is that they work "out of the box" and are easy to deploy. On the flip side, they are not optimized for any specific task and can only represent a limited set of decision rules.

### 2.6. Learning task-specific features with deep learning

To improve the performance of their models, engineers are thus increasingly working with expressive features that result from the iterative application of simple mathematical operations. For historical reasons, these models are known as "artificial neural networks" and

their parameters θ are called "neural weights" [22]. In radiology, we are especially interested in "convolutional" neural networks (CNN) that rely on weighted sums of neighboring pixel values, known as convolutions [23].

The weights of these convolutions are the free parameters that must be optimized to "train" a CNN on a specific task. General CNN feature extractors can be trained on large datasets of annotated images using supervised learning. This has led to the development of a large "zoo" of CNN architectures that are now commonly used in medical imaging, such as U-Nets for image segmentation [24] or ResNets [25] and EfficientNets [26] for image classification. Of note, that the link between the image and its resulting features is difficult to interpret for a human and this why DL can be seen as a "black box".

### 2.7. Statistical analysis and regularization

Once high-quality features have been computed on a set of images, engineers use robust statistical methods to obtain a decision rule. A first approach is to rely on decision trees and forests that naturally favor interpretability [27]. A second approach is to rely on linear transformations of the features. These are trained with error penalties such as the cross entropy or the max-margin loss, that respectively correspond to logistic regression or support-vector machines (SVM) [28] and are commonly found as a last step in CNN-based models.
For classification models, a common performance metric is the "area under the receiver operating characteristic" (AUC) [29]. This formula measures how well the model separates two populations: a value of 0.0 corresponds to a model that mis-classifies all images; 0.5 to a coin toss that makes no distinction between the two classes; and 1.0 to a perfect classifier.

### 2.8. Applying machine learning for segmentation

Segmentation consists in extracting a specific volume of interest from the entire image (typically to delineate a tumor). In other words, the purpose is to give a label to each pixel / voxel. Segmentation can be manual, semiautomatic or fully automatic. In semiautomatic segmentation, additional information concerning the output segmentation mask “y” is given beforehand. For example, some pixels can be labeled manually (by an experienced radiologist), or a bounding box can be placed on the region of interest, forcing all pixels outside of this region to be background. The effectiveness and accuracy of segmentation

methods are evaluated using the Dice similarity coefficient (DSC), which is calculated using manual or semi-automated segmentation as ground truth [30].

## 3. Current applications

### 3.1. Primary bone lesions

#### 3.1.1. Image segmentation

Even if most segmentation methods evaluated in musculoskeletal oncology are manual or semiautomatic, fully automatic methods were recently published [31–33]. Dionisío et al. compared manual and semiautomatic segmentation methods using MR images of 20 malignant bone lesions (osteosarcomas and Ewing sarcomas) [31]. There was high similarity when comparing manual and semiautomatic segmentations with a DSC reaching 96%, with a significant reduction of segmentation time using the semiautomatic method [31]. As limitations, the small sample size does not allow a generalizability of these methods on other tumors that may potentially be more difficult to segment. Zhang et al. assessed a semiautomatic segmentation method with a supervised residual network on 2,305 CT images from 23 patients with osteosarcoma [32]. The hierarchical features extracted could be learned directly from the images by the network [32]. Despite the use of many slices for each patient, the overall sample size was too small to validate the results convincingly: there is a risk that the variety of tumors is not sufficient to be representative. Besides, we can question the robustness of osteosarcomas segmentations made on CT since these lesions often have similar density to normal adjacent tissues. Qu et al. developed a DL-based automatic segmentation method for 105 pelvic bone tumors on MRI to extract three dimensional information before surgery [33]. The segmentation accuracy of this method (trained on 90 patients and tested on 15 patients) was superior to several competing methods and comparable to the expert annotation (DSC of up to 85%), while the average run time was significantly sped up (from 1820 to 19 seconds) [33].

#### 3.1.2. Lesion detection and classification

Several studies have evaluated DL models for the detection and classification of bone tumors either on radiographs, CT or MRI examinations [3,34,35].

He et al. developed a CNN to automatically classify primary bone tumors using a multi-institutional dataset of 2,899 plain radiographs from 1,356 patients into benign, intermediate or malignant tumors [3]. The model had a high performance with an AUC reaching up to 0.916 for malignant versus not malignant (benign or intermediate), and an accuracy of 72.1% for the three-way classification (benign versus intermediate versus malignant), with performances close to expert radiologists and better than junior radiologists [3]. Do et al. built a CNN to determine whether knee bone regions are normal, benign-tumor or malignant-tumor regions [34]. The model was applied on 1,576 plain radiographs (1,195 with tumors and 381 normal) and yielded 99% accuracy for the classification task [34]. As a limitation, the model was not tested on external sets for validation. Similarly, Liu et al. developed several DL models based on plain radiographs features to classify 982 bone tumors into benign, intermediate and malignant [35]. The model improved the performances of junior radiologists (AUC of 0.898 and 0.762, respectively; $P = 0.007$) and obtained performance similar to those of senior radiologists (AUC of 0.819; $P = 0.38$) [35].

Eweje et al. built a DL algorithm combining MR images and clinical characteristics to differentiate 1,060 bone tumors (582 benign and 478 malignant) [36]. The model showed similar accuracy (76% *vs.* 73%; $P = 0.7$), sensitivity (79% *vs.* 81%; $P = 1.0$) and improved specificity (75% *vs.* 66%; $P = 0.48$) by comparison with expert radiologists' performances [36]. One limitation was the need to perform manual lesion segmentation prior to analysis using the DL method [36].

Yin et al. developed a multiparametric MRI-based radiomic model from fat-saturated T2-weighted and contrast-enhanced T1-weighted images to differentiate 120 benign and malignant sacral tumors (54 chordomas, 30 metastases and 36 giant cell tumors) [37]. The best performance was found with the combination of the two sequences with an AUC of 0.77 and an accuracy of 71% [37]. An interesting point is that the combination of sequences generally improves the performances of radiomic models. Although the sample size is small, it remains suitable for radiomic studies dealing with rare tumors. Besides, there is a relatively balanced distribution of the studied histological subtypes in this study. Liu et al. evaluated a multi-model weighted fusion framework based on MRI data in 585 patients with spinal tumors that was designed to classify the tumors into benign or malignant [38]. The accuracy of the model was better than that of physicians for the classification task (82% and up to 74%, respectively) [38]. The main limitation is related to the recall rate of tumor regions improvement because the tumor detection model produced a certain number of false-positive regions, thus reducing the accuracy.

### 3.1.3. Pathologic tumor response

The heterogeneity of bone sarcomas may lead to inconsistent treatment outcomes among patients, in particular for those receiving neoadjuvant chemotherapy [7]. Thus, there are two potential applications in AI. One is the evaluation of the response to neoadjuvant chemotherapy directly on positron emission tomography (PET) or MRI (with the results of pathological analysis as the gold standard). This evaluation is performed on multiple pathological slices using a complex and time-consuming process [9]. This explains the interest for noninvasive methods to identify tumor necrosis caused by neoadjuvant chemotherapy and classify patients into responders and non-responders [39–41]. The second is the prediction of the response to chemotherapy based on the specific characteristics of each tumor, currently impossible to predict and which would be "seen" with AI, before any chemotherapy. MRI and metabolic imaging play a key role in this issue, motivating the development of dedicated AI models [41,42].

Zhong et al. implemented a pipeline on 144 patients with osteosarcoma (studying fat-saturated T2-weighted images from preoperative MRI examination) to predict good and bad responders to neoadjuvant chemotherapy [41]. The combination of clinical and radiomics nomogram demonstrated the best discriminative capabilities, with an AUC of 0.79, suggesting that this model could be applied to assist radiologists in predicting good responders to neoadjuvant chemotherapy before surgery. A limitation which must be discussed is that radiomics features were only extracted from fat-saturated T2-weighted images and at one time point. Multiple MRI sequences and images at different time points may improve model performance. Besides, this study involved a pediatric population, the results being not necessarily applicable to adults [41].

Kim et al. compared an ML approach using fluorine–$^{18}$fluorodeoxyglucose ($^{18}$F-FDG) uptake heterogeneity features and a CNN analysis to assess the accuracy of prediction of the response to neoadjuvant chemotherapy on a cohort of 105 patients with osteosarcoma [42]. The CNN network using $^{18}$F-FDG baseline PET images could predict the treatment response before prior chemotherapy with an AUC reaching 0.99 [42]. In view of this nearly perfect result based on a limited number of patients, overfitting cannot be excluded.

### 3.1.4. Prediction of tumor recurrence

A major issue for surgeons removing primary bone tumors is the possibility of local recurrence or secondary metastases. Therefore, predicting post-surgery recurrence of tumors based on pre-surgery medical images would be of significant interest. He et al. built a CNN model to predict the local recurrence of 56 giant cell bone tumors, considering clinical characteristics and pre-surgery MRI features [43]. The fusion model built by integrating all features available improved the accuracy and sensitivity for prediction (respectively 78.6% and 87.5%) [43]. Sheen et al. built and validated a radiomic imaging model for the prediction of future metastases development at the point of osteosarcoma diagnosis in 83 patients treated with surgery and chemotherapy, using $^{18}$F-FDG-PET data [44]. The final multivariable logistic model combining two radiomics features (SUVmax and Gray-Level Zone Length Matrix: Short-Zone Low Grey-Level Emphasis: GLZLM-SZLGE), achieved an AUC of 0.80 [44]. The most contributory features to the classification derive from GLZLM features, which is a radiomic feature likely correlated to the local heterogeneity in a tumor [19,45].

### 3.2. Secondary bone lesions

Both MRI and CT are commonly used for the detection of secondary lesions, but metabolic imaging also plays a central role. Thus, DL applied to the evaluation of bone metastases has predominantly involved bone scintigraphy / SPECT since it explores the whole-body with high sensitivity. Hong et al. built a CT radiomic–based ML model for differentiating benign bone islands from osteoblastic bone metastases [46]. A random forest (RF) model was built on 177 patients with 89 benign bone tumors and 88 metastasis. The AUC of the trained RF model was higher than that obtained by one of the inexperienced radiologists (0.96 and 0.8, respectively) [46]. Noguchi et al. developed a DL-based algorithm for automatic detection of bone metastases on CT scans (269 positive scans with 1,375 bone metastases and 463 without bone lesion) [47]. The model improved the overall performance of nine radiologists with AUC respectively of 0.75 and 0.9 ($P < 0.001$) without and with the use of the model. Furthermore, the mean interpretation time per case decreased from 168 to 85 seconds ($P = 0.004$) [47]. Xiong et al. built a ML method based on MRI sequences of 178 lesions from 107 patients to differentiate 60 multiple myeloma and 118 different tumor metastases of the lumbar vertebra [48]. Among 10 classifiers, the artificial neural networks classifier from the T2-weighted images achieved the best performance: accuracy, sensitivity, and specificity of 82%, 88%, and 79%, respectively, in the validation cohort [48].

Lin et al. constructed several deep classifiers to automatically diagnose metastases in 251 thoracic SPECT bone images, 85 with bone metastasis and 166 without tumor with good performances (AUC of 0.98) [49]. In view of such good performances, the risk of overfitting also seems likely here. Hsieh et al. evaluated DL techniques to improve the efficacy of bone metastasis detection on bone scintigraphy, with 37,427 image sets from 19,041 patients [50]. The overall performance was good with a negative predictive value reaching 96.5% [50]. This algorithm could help physicians safely exclude bone metastases, decreasing physician workload, and improving patient care.

## 4. Main limitations of AI

This critical review suggests that the quality of the articles related to musculoskeletal oncologic imaging is low on average with a lack of homogeneity and few studies for each objective and tumor types [51]. The main limitation of most studies is the sample size which is most often too small to apply DL algorithms and validate their results, even though we are aware that primary bone tumors are rare diseases. Unfortunately, CNN cannot be trained reliably on small datasets (less than a few hundred patients) because there is no way to control if the CNN has not simply memorized the characteristics of the few patients. Some authors divert this problem by considering each patient as a set of slices (for instance 100 two dimensional slices for 20 patients but representing only 20 three dimensional volumes) but the variety of tumors is not large enough to be representative. Also, acquisition parameters greatly influence CNN output. In our opinion, it seems essential to have multi-institution and multi-vendor datasets to be able to control all these parameters [37]. Besides, a specific issue of musculoskeletal oncology AI studies is that cohorts usually come from tertiary centers whose characteristics differ from non-expert centers [3].

Another limitation (not specific to musculoskeletal oncology) comes from the questionable quality of the ground truth labels used to train models. The type of label may vary (pathologic, clinical or imaging criteria), but usually requires musculoskeletal expertise [52,53]. Besides, all studies used data collected retrospectively with possible selection bias. The question of the comparability of the different studies performed with specific software and small cohorts must also be raised.

The full automation of bone tumor segmentation is more difficult to achieve than with other anatomical regions (lung / brain) [54,55] given the variability in bone lesion location and the non-uniform shape of bones based on anatomical location. Although fully automated

segmentation of bone tumors is not yet implemented in routine, it is an essential step to build entirely automated pipelines [10,56].

The "black-box" nature of CNN makes it challenging to identify the causes of false-positive or false-negative results. For instance, He et al. [3] observed that DL models were mistaken on radiographic criteria that were quite simple for radiologists (recognizing a permeative appearance or aggressive periostitis and associating it with malignancy [3]).

Concerning radiomics, a pretreatment procedure on raw images is recommended to improve the robustness of radiomics features [19]. Its goal is to reduce variability in voxel values, hence in radiomic features [52,53]. Although there is no universally accepted standard procedure, classical methods of data homogenization (such as N4 bias correction or gray-scale normalization) were performed on dataset before extracting radiomic features in some studies [38,48]. However, several studies did not apply these methods, raising the issue of the robustness of their results. Overall, the number of papers reporting the assessment of radiomic feature reproducibility and the use of independent or external clinical validation was relatively small [57]. Even if there are no clear arguments about the choice of MR sequences that should be used, most studies showed the interest of combining features extracted from several sequences to improve the performance of a model (for instance T2-weighted and contrast-enhanced T1-weighted images together), rather than treating them separately [37]. Moreover, few studies included clinico-biological features in their decision nomogram [43].

## 5. Future developments

The current AI tools developed in musculoskeletal oncology imaging are not yet used in clinical practice. In order to bring these AI tools from a preclinical research area to the clinical stage and daily use, some issues should be addressed in future studies [57]. They include: (*i*), Achieving reliable fully automated bone tumor segmentation; (*ii*), Ensuring the reproducibility of radiomic features (in particular applying standardization techniques) [58]; (*iii*), Controlling overfitting by selecting more reproducible features, screening and determining false discovery rates and determining a feature-selection algorithm suitable for small n-to-p data; (*iiii*), Performing bone lesions classifications by specific diagnosis, rather than binary (benign / malignant) [36,43] or ternary (benign/intermediate/malignant) classifications [35]; (*v*), Building models using different imaging modalities, clinico-biological features [59,60], radiomics and other features from the broad family of "-omics" (including genomics, proteomics or metabolomics) to get closer to precision medicine

[61,62]; (*vi*), Using independent datasets to validate the results; and (*vii*), Creating large public image databases to train and validate AI software, freely available to researchers [10].

## 6. Conclusion

In the field of clinical research, further efforts are still needed for AI imaging studies in musculoskeletal oncology to show a sufficient level of evidence to be used in daily practice and this mirrors the limitation found in other organs [51, 63]. Nevertheless, by applying rigorous methodological rules, some AI tools will certainly be integrated into practice in the future. They will free up medical time by automating tedious tasks of low added value and will probably improve radiologist performances. Such AI decision support tools may help to also scale expertise to communities outside of major academic centers, thereby increasing accessibility to specialist oncology care.

### Human rights

The authors declare that the work described has been performed in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans.

### Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patients.

### Disclosure of interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

### Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

### Funding

None of the authors was funded for this paper.

**Declaration of Competing Interest**

The authors declare that they have no competing interest in relation with this article.

**References**

[1] Akinci D'Antonoli T. Ethical considerations for artificial intelligence: an overview of the current radiology landscape. Diagn Interv Radiol 2020; 26: 504–511.

[2] Klontzas ME, Karantanas AH. Research in musculoskeletal radiology: setting goals and strategic directions. Semin Musculoskelet Radiol 2022; 26: 354–358.

[3] He Y, Pan I, Bao B, Halsey K, Chang M, Liu H, et al. Deep learning-based classification of primary bone tumors on radiographs: a preliminary study. EBioMedicine 2020; 62: 103121.

[4] Lacroix M, Burns R, Campagna R, Larousserie F, Drapé J-L. Acral fibromyxoma: findings on dynamic contrast-enhanced perfusion MRI. Diagn Interv Imaging 2022; 103: 59–61.

[5] Howe BM, Broski SM, Littrell LA, Pepin KM, Wenger DE. Quantitative musculoskeletal tumor imaging. Semin Musculoskelet Radiol 2020; 24: 428–440.

[6] Jones BC, Ahlawat S, Fayad LM. 3D MRI in musculoskeletal oncology. Semin Musculoskelet Radiol 2021; 25: 418–424.

[7] Vasilevska Nikodinovska V, Ivanoski S, Samardziski M, Janevska V. Percutaneous imaging-guided versus open musculoskeletal biopsy: concepts and controversies. Semin Musculoskelet Radiol 2020; 24: 667–675.

[8] Visser JJ, Goergen SK, Klein S, Noguerol TM, Pickhardt PJ, Fayad LM, et al. The value of quantitative musculoskeletal imaging. Semin Musculoskelet Radiol 2020; 24: 460–474.

[9] 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.

[10] Barat M, Chassagnon G, Dohan A, Gaujoux S, Coriat R, Hoeffel C, et al. Artificial intelligence: a critical review of current applications in pancreatic imaging. Jpn J Radiol 2021; 39: 514–23.

[11] Li MD, Ahmed SR, Choy E, Lozano-Calderon SA, Kalpathy-Cramer J, Chang CY. Artificial intelligence applied to musculoskeletal oncology: a systematic review. Skeletal Radiol 2022; 51: 245–256.

[12] Nakaura T, Higaki T, Awai K, Ikeda O, Yamashita Y. A primer for understanding radiology articles about machine learning and deep learning. Diagn Interv Imaging 2020; 101: 765–770.

[13] Sekar JAP, Tapia JJ, Faeder JR. Automated visualization of rule-based models. PLoS Comput Biol 2017; 13:e1005857.

[14] Hélie S, Shamloo F, Zhang H, Ell SW. The impact of training methodology and representation on rule-based categorization: an fMRI study. Cogn Affect Behav Neurosci 2021; 21: 717–735.

[15] Razavian N, Knoll F, Geras KJ. Artificial intelligence explained for nonexperts. Semin Musculoskelet Radiol 2020; 24: 3–11.

[16] Courot A, Cabrera DLF, Gogin N, Gaillandre L, Rico G, Zhang-Yin J, et al. Automatic cervical lymphadenopathy segmentation from CT data using deep learning. Diagn Interv Imaging 2021; 102: 675–681.

[17] Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D. Self-supervised learning for medical image analysis using image context restoration. Med Image Anal 2019; 58: 101539.

[18] Bach Cuadra M, Favre J, Omoumi P. Quantification in musculoskeletal imaging using computational analysis and machine learning: segmentation and radiomics. Semin Musculoskelet Radiol 2020; 24: 50–64.

[19] Lacroix M, Frouin F, Dirand AS, Nioche C, Orlhac F, Bernaudin JF, et al. Correction for magnetic field inhomogeneities and normalization of voxel values are needed to better reveal the potential of MR radiomic features in lung cancer. Front Oncol 2020; 10: 43.

[20] Rastegar S, Vaziri M, Qasempour Y, Akhash MR, Abdalvand N, Shiri I, et al. Radiomics for classification of bone mineral loss: a machine learning study. Diagn Interv Imaging 2020; 101: 599–610.

[21] Long L, Sun J, Jiang L, Hu Y, Li L, Tan Y, et al. MRI-based traditional radiomics and computer-vision nomogram for predicting lymphovascular space invasion in endometrial carcinoma. Diagn Interv Imaging 2021; 102: 455–462.

[22] Roca P, Attye A, Colas L, Tucholka A, Rubini P, Cackowski S, et al. Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. Diagn Interv Imaging 2020; 101: 795–802.

[23] Gao X, Wang X. Performance of deep learning for differentiating pancreatic diseases on contrast-enhanced magnetic resonance imaging: a preliminary study. Diagn Interv Imaging 2020; 101: 91–100.

[24] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 2020;39:1856–1867.

[25] He F, Liu T, Tao D. Why ResNet works? Residuals generalize. IEEE Trans Neural Netw Learn Syst 2020; 31: 5349–5362.

[26] Gessert N, Nielsen M, Shaikh M, Werner R, Schlaefer A. Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. MethodsX 2020; 7: 100864.

[27] Blanchet L, Vitale R, van Vorstenbosch R, Stavropoulos G, Pender J, Jonkers D, et al. Constructing bi-plots for random forest: tutorial. Analytica Chimica Acta 2020; 1131: 146–155.

[28] Shukla P, Verma A, Abhishek A, Verma S, Kumar M. Interpreting SVM for medical images using Quadtree. Multimed Tools Appl 2020; 79: 29353–29373.

[29] Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013; 4: 627–635.

[30] Lassau N, Bousaid I, Chouzenoux E, Lamarque JP, Charmettant B, Azoulay M, et al. Three artificial intelligence data challenges based on CT and MRI. Diagn Interv Imaging 2020; 101: 783–788.

[31] Dionísio FCF, Oliveira LS, Hernandes MA, Engel EE, Rangayyan RM, Azevedo-Marques PM, et al. Manual and semiautomatic segmentation of bone sarcomas on MRI have high similarity. Braz J Med Biol Res 2020; 53: e8962.

[32] Zhang R, Huang L, Xia W, Zhang B, Qiu B, Gao X. Multiple supervised residual network for osteosarcoma segmentation in CT images. Comput Med Imaging Graph 2018; 63: 1–8.

[33] Qu Y, Li X, Yan Z, Zhao L, Zhang L, Liu C, et al. Surgical planning of pelvic tumor using multi-view CNN with relation-context representation learning. Med Image Anal 2021; 69: 101954.

[34] Do NT, Jung ST, Yang HJ, Kim SH. Multi-level Seg-Unet model with global and patch-based X-ray images for knee bone tumor detection. Diagnostics 2021; 11: 691.

[35] Liu R, Pan D, Xu Y, Zeng H, He Z, Lin J, et al. A deep learning–machine learning fusion approach for the classification of benign, malignant, and intermediate bone tumors. Eur Radiol 2022; 32: 1371–1383.

[36] Eweje FR, Bao B, Wu J, Dalal D, Liao W, He Y, et al. Deep learning for classification of Bone lesions on routine MRI. EBioMedicine 2021; 68: 103402.

[37] Yin P, Mao N, Zhao C, Wu J, Chen L, Hong N. A triple-classification radiomics model for the differentiation of primary chordoma, giant cell tumor, and metastatic tumor of sacrum based on T2-weighted and contrast-enhanced T1-weighted MRI. J Magn Reson Imaging 2019; 49: 752–759.

[38] Liu H, Jiao M, Yuan Y, Ouyang H, Liu J, Li Y, et al. Benign and malignant diagnosis of spinal tumors based on deep learning and weighted fusion framework on MRI. Insights Imaging 2022; 13: 87.

[39] Saleh MM, Abdelrahman TM, Madney Y, Mohamed G, Shokry AM, Moustafa AF. Multiparametric MRI with diffusion-weighted imaging in predicting response to chemotherapy in cases of osteosarcoma and Ewing's sarcoma. Br J Radiol 2020; 93: 20200257.

[40] Huang L, Chen J, Hu W, Xu X, Liu D, Wen J, et al. Assessment of a radiomic signature developed in a general NSCLC cohort for predicting overall survival of ALK-positive patients with different treatment types. Clinical Lung Cancer 2019; 20: e638–e651.

[41] Zhong J, Zhang C, Hu Y, Zhang J, Liu Y, Si L, et al. Automated prediction of the neoadjuvant chemotherapy response in osteosarcoma with deep learning and an MRI-based radiomics nomogram. Eur Radiol 2022; 32: 6196–6206.

[42] Kim J, Jeong SY, Kim BC, Byun BH, Lim I, Kong CB, et al. Prediction of neoadjuvant chemotherapy response in osteosarcoma using convolutional neural network of tumor center 18F-FDG PET images. Diagnostics 2021; 11: 1976.

[43] He Y, Guo J, Ding X, van Ooijen PMA, Zhang Y, Chen A, et al. Convolutional neural network to predict the local recurrence of giant cell tumor of bone after curettage based on pre-surgery magnetic resonance images. Eur Radiol 2019; 29: 5441–5451.

[44] Sheen H, Kim W, Byun BH, Kong C-B, Song WS, Cho WH, et al. Metastasis risk prediction model in osteosarcoma using metabolic imaging phenotypes: a multivariable radiomics model. PLoS One 2019; 14: e0225242.

[45] Nardone V, Reginelli A, Scala F, Carbone SF, Mazzei MA, Sebaste L, et al. Magnetic-resonance-imaging texture analysis predicts early progression in rectal cancer patients undergoing neoadjuvant chemoradiation. Gastroenterol Res Pract 2019; 2019: 1–8.

[46] Hong JH, Jung JY, Jo A, Nam Y, Pak S, Lee SY, et al. Development and validation of a radiomics model for differentiating bone islands and osteoblastic bone metastases at abdominal CT. Radiology 2021; 299: 626–632.

[47] 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; doi.org/10.1007/s00330-022-08741-3.

[48] 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.

[49] Lin Q, Li T, Cao C, Cao Y, Man Z, Wang H. Deep learning based automated diagnosis of bone metastases with SPECT thoracic bone images. Sci Rep 2021; 11: 4223.

[50] Hsieh TC, Liao CW, Lai YC, Law KM, Chan PK, Kao CH. Detection of bone metastases on bone scans through image classification with contrastive learning. J Pers Med 2021; 11: 1248.

[51] Crombé A, Fadli D, Italiano A, Saut O, Buy X, Kind M. Systematic review of sarcomas radiomics studies: bridging the gap between concepts and clinical applications? Eur J Radiol 2020; 132: 109283.

[52] Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring computed tomography scanner variability of radiomics features. Invest Radiol 2015; 50: 757–765.

[53] Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. Radiology 2019; 291: 53–59.

[54] Autrusseau P-A, Labani A, De Marini P, Leyendecker P, Hintzpeter C, Ortlieb A-C, et al. Radiomics in the evaluation of lung nodules: intrapatient concordance between full-dose and ultra-low-dose chest computed tomography. Diagn Interv Imaging 2021; 102: 233–239.

[55] Li N, Mo Y, Huang C, Han K, He M, Wang X, et al. A clinical semantic and radiomics nomogram for predicting brain invasion in WHO grade II meningioma based on tumor and tumor-to-brain interface features. Front Oncol 2021; 11: 752158.

[56] Deniz CM, Xiang S, Hallyburton RS, Welbeck A, Babb JS, Honig S, et al. Segmentation of the proximal femur from MR images using deep convolutional neural networks. Sci Rep 2018; 8: 16485.

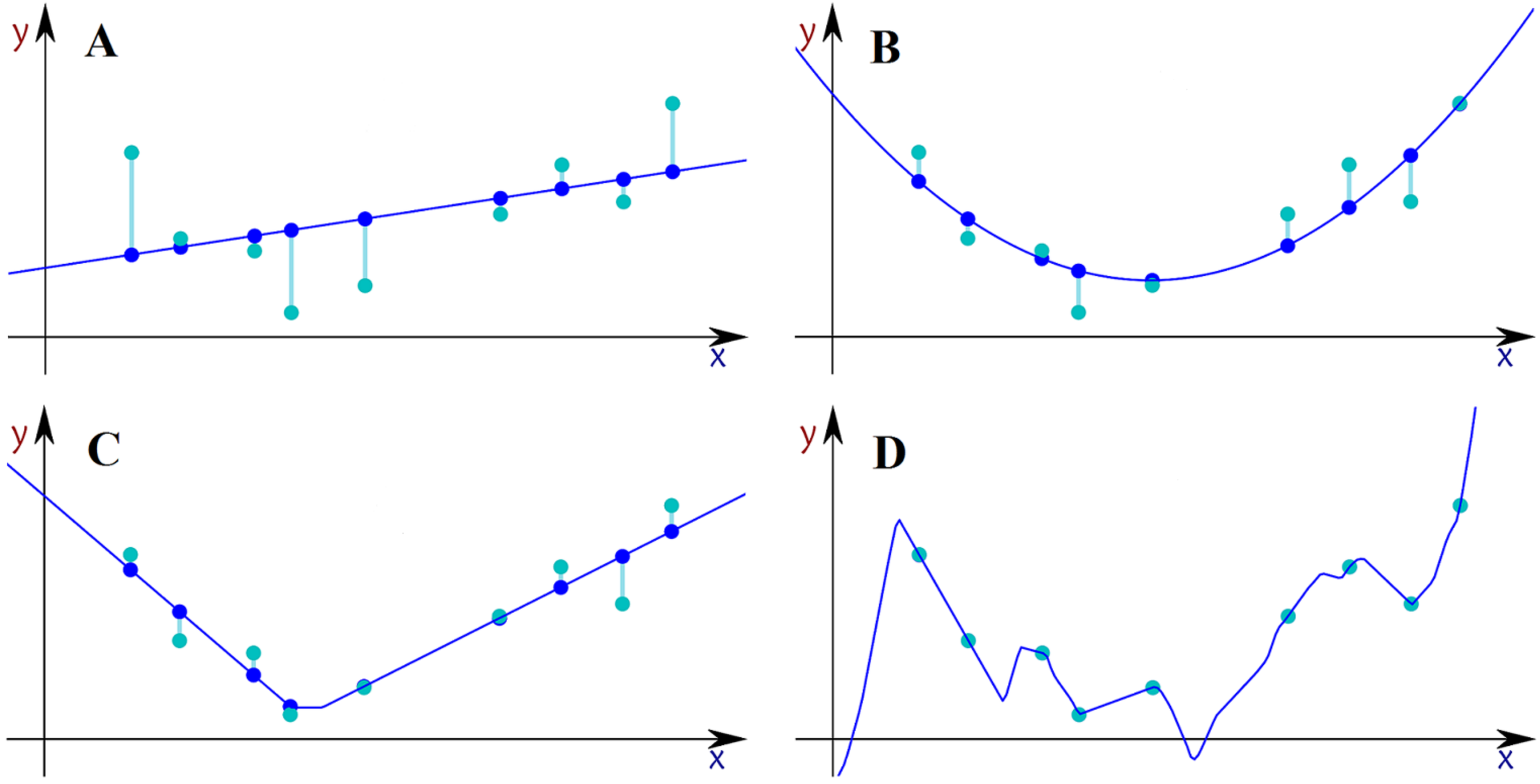
[57] Gitto S, Cuocolo R, Albano D, Morelli F, Pescatori LC, Messina C, et al. CT and MRI radiomics of bone and soft-tissue sarcomas: a systematic review of reproducibility and validation strategies. Insights Imaging 2021; 12: 68.

[58] Park H, Sholl LM, Hatabu H, Awad MM, Nishino M. Imaging of precision therapy for lung cancer: current state of the art. Radiology 2019: 293: 15–29.

[59] Shen G, Jia Z, Deng H. Apparent diffusion coefficient values of diffusion-weighted imaging for distinguishing focal pulmonary lesions and characterizing the subtype of lung cancer: a meta-analysis. Eur Radiol 2016; 26: 556–66.
[60] Gaume M, Chevret S, Campagna R, Larousserie F, Biau D. The appropriate and sequential value of standard radiograph, computed tomography and magnetic resonance imaging to characterize a bone tumor. Sci Rep 2022; 12: 6196.
[61] Acharya UR, Hagiwara Y, Sudarshan VK, Chan WY, Ng KH. Towards precision medicine: from quantitative imaging to radiomics. J Zhejiang Univ Sci B 2018; 19: 6–24.
[62] Vogrin M, Trojner T, Kelc R. Artificial intelligence in musculoskeletal oncological radiology. Radiol Oncol 2020; 55: 1–6.
[63] Chassagnon G, Dohan A. Artificial intelligence: from challenges to clinical implementation. Diagn Interv Imaging 2020; 101: 763–764.

**Figure legend**

**Fig. 1:** Graphs illustrate overfitting on a toy regression problem between a scalar input "x" and output "y". Nine training samples ($x_i$, $y_i$) are displayed using light blue dots. A), Linear regression estimates the linear trend (dark blue) that best minimizes the sum of least squared errors (vertical bars). B), Quadratic regression estimates the best fitting parabola and can model a slump followed by a growth. C), A simple neural network with two hidden neurons can represent a piecewise linear curve with two hinges. D), Complex neural networks have more degrees of freedom. Here, a fully connected network with 100 hidden neurons gets a "perfect fit" to the training data. Unfortunately, this is done at the cost of "overfitting" so that the interpolating curve has no robustness to noise and does not extract any meaningful trend.


y
A
x
y
B
x
y
C
x
y
D
x