



# **The Role of Artificial Intelligence in the Identification and Evaluation of Bone Fractures**

Andrew Tieu<sup>1</sup>, Ezriel Kroen<sup>2</sup>, Yonaton Kadish<sup>2</sup>, Zelong Liu<sup>1</sup>, Nikhil Patel<sup>1</sup>, Alexander Zhou<sup>1</sup>, Alara Yilmaz<sup>3</sup>, Stephanie Lee<sup>3</sup> and Timothy Deyer<sup>4,5,\*</sup>

- <sup>1</sup> BioMedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
- <sup>2</sup> New York Medical College, Valhalla, NY 10595, USA
- <sup>3</sup> Horace Mann School, Bronx, NY 10471, USA
- <sup>4</sup> East River Medical Imaging, New York, NY 10021, USA
- <sup>5</sup> Department of Radiology, Cornell Medicine, New York, NY 10021, USA
- \* Correspondence: tdeyer@eastriverimaging.com

**Abstract:** Artificial intelligence (AI), particularly deep learning, has made enormous strides in medical imaging analysis. In the field of musculoskeletal radiology, deep-learning models are actively being developed for the identification and evaluation of bone fractures. These methods provide numerous benefits to radiologists such as increased diagnostic accuracy and efficiency while also achieving standalone performances comparable or superior to clinician readers. Various algorithms are already commercially available for integration into clinical workflows, with the potential to improve healthcare delivery and shape the future practice of radiology. In this systematic review, we explore the performance of current AI methods in the identification and evaluation of fractures, particularly those in the ankle, wrist, hip, and ribs. We also discuss current commercially available products for fracture detection and provide an overview of the current limitations of this technology and future directions of the field.

Keywords: artificial intelligence; deep learning; bone fracture; fracture detection

# 1. Introduction

About 9.4 million fractures occur each year in the US [1], with an estimated annual expenditure of USD 22 billion from just osteoporotic fractures [2]. Due to an aging population, this is only expected to increase, with a projected cost of over USD 95 billion by 2040 [3]. Missed fractures are the most common diagnostic errors in the interpretation of musculoskeletal radiographs, and are especially prevalent in emergency department settings, where they account for approximately 80% of all misdiagnoses [4,5], leading to significant consequences such as delays in treatment and increased long-term morbidity [6].

Errors in fracture diagnosis are partially attributed to growing demands for radiological imaging, including radiography, computed tomography (CT), and magnetic resonance imaging (MRI) [7], with radiologist workloads estimated to have increased by 52% between 2012 and 2019 [8]. In the field of musculoskeletal radiology, increased usage of radiography is further compounded by the increasing incidence of fractures over the past 20 years, leading to radiologist fatigue and susceptibility to diagnostic errors [9]. Additionally, while the majority of skeletal radiographs are interpreted by radiologists in hospital settings, these reads may be performed by trainees or clinicians without specific musculoskeletal training [4], or even by nonradiologists [9], further contributing to missed fractures [5,10].

Artificial intelligence (AI), particularly machine learning, may provide a solution to several issues in the field of medical imaging analysis. Machine learning boasts a wide array of potential applications, including automated generation of diagnoses, image segmentation, and disease prognosis. One of the most common machine learning approaches to medical



Citation: Tieu, A.; Kroen, E.; Kadish, Y.; Liu, Z.; Patel, N.; Zhou, A.; Yilmaz, A.; Lee, S.; Deyer, T. The Role of Artificial Intelligence in the Identification and Evaluation of Bone Fractures. *Bioengineering* **2024**, *11*, 338. https://doi.org/10.3390/ bioengineering11040338

Academic Editor: Dongrong Xu

Received: 27 February 2024 Revised: 23 March 2024 Accepted: 26 March 2024 Published: 29 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). imaging analysis is deep learning, which uses deep neural network structures inspired by the human brain to interpret complex datasets [11]. In particular, convolutional neural networks (CNNs) exhibit strong performance with image-based tasks by using convolutional filters to automatically learn and extract features for image understanding. Trained using large imaging datasets, often with tens of thousands of images, these models are able to improve the accuracy and efficiency of clinician reads for fracture detection [12–14], with many standalone models achieving performances at the level of experienced clinicians or even outperforming them [13,15,16]. Applications of deep learning for fracture detection have been explored for a wide array of anatomical locations across various imaging modalities [12,17–19], and continue to grow as new capabilities of AI are being developed.

Given this rapidly evolving landscape, we provide a broad overview of the current state of machine learning in the identification and evaluation of bone fractures, particularly in the ankle, wrist, hip, and ribs. Specifically, "identification" refers to model tasks such as binary classification of fracture vs. no fracture, as well as fracture detection and segmentation, while "evaluation" describes additional tasks such as outcome prediction. We also highlight current commercial products available for fracture detection and discuss current challenges in the field, as well as future directions of this technology. As the applications of AI in medical imaging continue to broaden, an understanding of this technology will be invaluable as it begins to shape the practice of diagnostic radiology.

# 2. Methods

A systematic literature search was performed on three public journal databases (PubMed, IEEE, and Scopus) on 23 August 2023, as summarized in Figure 1. The PRISMA checklist [20] was used to facilitate transparent and complete reporting of the systematic review. Potential papers included those published since 1 January 2019, and were filtered using the following search terms: "ankle fracture" OR "wrist fracture" OR "hip fracture" OR "hip fracture" OR "rib fracture" AND "artificial intelligence" OR "deep learning".



Figure 1. Flowchart of PRISMA study selection.

All search records were first screened by title and abstract, and duplicate studies and data were excluded. Studies that did not use deep learning as their primary method of bone fracture detection were also excluded. From a total of 26 filtered papers during screening, the full texts were reviewed, and 14 studies were selected for discussion. Article selection was performed independently by two of the authors and collectively reviewed by the group before inclusion.

# 3. Performance Metrics

Given the large number of metrics used in evaluating the performance of imaging AI models, we provide here an overview of several commonly used metrics. Accuracy indicates the proportion of correctly classified cases relative to the total number of cases. Sensitivity, also known as recall, measures the proportion of true positives identified by the model among all ground truth positives, whereas specificity measures the ability of the model to identify true negatives among all ground truth negatives. Positive predictive value (PPV), also known as precision, complements sensitivity by indicating the proportion of true positives among all predicted positives by the model. Negative predictive value (NPV) similarly complements specificity, representing the proportion of true negatives among all predicted negatives by the model. The calculation of each of these metrics, bounded from 0 to 1, is shown below:

Sensitivity = 
$$\frac{TP}{TP + FN}$$
  
Specificity =  $\frac{TN}{TN + FP}$   
 $PPV = \frac{TP}{TP + FP}$   
 $NPV = \frac{TN}{TN + FN}$ 

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve provides a summary of the model's overall performance, quantifying its ability to discriminate between positive and negative instances at all thresholds. An AUC score of 1 indicates the perfect performance of a model in discriminating between positive and negative instances, while a score of 0.5 indicates performance equivalent to random chance. For segmentation tasks, Dice score and intersection over union (IoU) are both commonly used metrics that evaluate the overlap between AI-predicted segmentations and the ground truth. For both metrics, a score of 1 indicates perfect overlap, while a score of 0 indicates no overlap at all.

$$Dice \ score = \frac{2 \times Area \ of \ Intersection}{Total \ Area}$$
$$IoU = \frac{Area \ of \ Intersection}{Area \ of \ Union}$$

#### 4. Ankle Fractures

Ankle fractures are among the most common injuries treated by orthopedic surgeons, accounting for 9% of all bone fractures with an estimated cost of USD 10 billion per year in the US [21]. The incidence of ankle fractures continues to increase, having tripled in elderly women over the past 30 years [22]. Treatment of ankle fractures relies on careful identification of bone lesions and damage to soft tissue and ligaments on both clinical assessment and imaging [23]. However, nearly 23% of ankle fractures are missed on initial radiographic imaging due to factors such as anatomical variance, superposition of structures on radiographs, lack of experience, and high physician workload [24,25]. Left



untreated, these injuries can result in significant long-term morbidity [26,27]. An ankle fracture as seen on radiography is shown below in Figure 2.

Figure 2. Distal fibular fracture as seen on lateral ankle radiograph.

Ashkani-Esfahani et al. sought to investigate the performance of deep-learning algorithms in the detection of ankle fractures, especially occult fractures [28]. A dataset of 1050 normal ankle radiographs was collected and matched with 1050 radiographs with ankle fractures, 72 of which were labeled as occult fractures due to being missed initially and subsequently diagnosed using additional radiographs or CT images. However, the group also hypothesized that incorporating multiple radiographic views would increase the detection of previously occult fractures. Using transfer learning, InceptionV3 [29] and Resnet-50 [30] models pre-trained with ImageNet [31] were trained with the ankle radiographs, with inputs as either single-view (AP) radiographs or 3-input structures with multiple views (AP, mortise, and lateral). Overall, the InceptionV3 model outperformed Resnet-50 in all performance criteria when using 3-view radiographic image stacks, including sensitivity (99% vs. 98%), specificity (99% vs. 94%), PPV (99% vs. 95%), NPV (99% vs. 97%), accuracy (99% vs. 96%), F1 score (99% vs. 96%), and AUC (99% vs. 98%). The incorporation of multiple views was important in achieving this performance, improving InceptionV3's sensitivity (91% to 99%) and specificity (94% to 99%) when compared to the use of single views. Additionally, of the 72 occult fractures, the InceptionV3 model was able to detect 71/72 (98.6%) previously missed fractures, while Resnet-50 detected 69/72 (95.8%).

Given that previous work in ankle fracture detection was often bound by limitations such as large datasets [32], manual feature extraction [33], or pre-trained models [18], recent work by Kitamura et al. investigated whether comparable accuracy could be achieved by training CNNs de novo using a smaller dataset of 596 normal and abnormal ankle cases with multiple radiographic views [34]. The Inception V3, Resnet, and Xception [35] models were trained using single views, and ensembles were created from a combination of the trained models and then evaluated using three views for each ankle case. The ensembles achieved an accuracy of 81%, an impressive feat given the small dataset size and lack of reliance on manual feature extraction or pre-trained models.

The details and results of these studies are summarized in Table 1.

Lead Author	Year	Imaging Modality	Total Number of Images *	Models Used	Model Tasks	Performance Metrics
Ashkani- Esfahani	2022	X-ray	6300	InceptionV3 Resnet-50	Fracture detection	InceptionV3: sensitivity 99%, specificity 99%, PPV 99%, NPV 99%, accuracy 99%, F1 score 99%, AUC 99% Resnet-50: sensitivity 98%, specificity 94%, PPV 95%, NPV 97%, accuracy 96%, F1 score 96%, AUC 98%.
Kitamura	2019	X-ray	1681	InceptionV3 Resnet-101 Xception	Fracture detection	Accuracy 81%, sensitivity 80%, specificity 83%, PPV 82%, NPV 81%

Table 1. Summary of study characteristics for ankle fracture detection models.

\* Reported as the combined total of training, validation, and testing datasets.

# 5. Wrist Fractures

Fractures of the distal forearm and wrist are the most common sites of bone fracture in childhood, accounting for one-third of all cases [36]. Among the carpal bones, scaphoid fractures (as seen in Figure 3) occur most frequently, representing 2–7% of all skeletal fractures [37]. Left untreated, 12% of scaphoid fractures may progress to non-union [38], leading to further complications such as osteonecrosis, degenerative arthritis, and functional loss [39,40]. While radiography is generally the imaging modality of choice in evaluating such fractures, scaphoid fractures are often radiographically occult and difficult to diagnose, with estimations for occult fractures ranging from 7% to as high as 50% [41–43].



Figure 3. Scaphoid fracture as seen on ulnar-deviated PA scaphoid radiograph.

Given the challenges of scaphoid fracture detection, Langerhuizen et al. investigated the utility of deep-learning algorithms in the identification of such fractures, a field that has not yet been extensively explored [44]. A dataset comprised of 300 radiographic scaphoid series was assembled, consisting of 150 scaphoid fracture cases and 150 non-fracture cases, and used to train a convolutional neural network. On testing, the model achieved an AUC of 0.77 with 72% accuracy, 84% sensitivity, and 60% specificity. In comparison, a group of five orthopedic surgeons reading the images had similar accuracy (84%) and sensitivity (76%), but significantly increased specificity (93%; p < 0.01), owing to 13 false positive suggestions made by the model that were correctly identified by the clinicians. Interestingly, the model was able to detect 5 of 6 occult scaphoid fractures that were missed by all clinician readers.

A year later, Hendrix et al. expanded on these findings, developing a CNN capable of not only the detection of scaphoid fractures but also segmentation [14]. Given the small sample size of the previous study, a larger dataset was used, consisting of 1039 conventional radiographs of the hand, wrist, and scaphoid for training of the scaphoid segmentation CNN, as well as 3000 radiographs for scaphoid detection training. The segmentation model achieved a Dice score of 97.4%, while the fracture detection model achieved an AUC of 0.87 with 78% sensitivity, 84% specificity, and 83% PPV. This performance was similar to that of 11 clinician readers, who had an AUC of 0.83 (p = 0.09), thus achieving radiologist-level performance in the detection of scaphoid fractures.

Hendrix et al. continued this work the following year, using a further expanded dataset of 5004 conventional hand, wrist, and scaphoid radiographs for the training and testing of the scaphoid fracture detection model, this time also incorporating multi-view radiographs (PA, ulnar-deviated PA, oblique, lateral) [45]. The new model achieved a 72% sensitivity, 93% specificity, 81% PPV, and an AUC of 0.88. Although the AUCs for the model and that of five musculoskeletal radiologists (0.87) were again similar (p > 0.05), the reading time for four of the readers was significantly reduced with AI assistance (p < 0.001), with an average 51% reduction in reading time. Altogether, these three studies show that current state-of-the-art AI algorithms are able to match the performance of human clinicians in the detection of scaphoid fractures in conventional radiographs and improve diagnostic efficiency, but are yet to supersede the accuracy of human readers.

In addition to scaphoid fractures, AI methods have also been investigated in the detection of various other fractures of the wrist. In 2022, Hardalaç et al. evaluated the performances of several deep-learning models in detecting fractures of the radius and ulna using a collection of 542 pediatric wrist radiographs from 275 patients [46]. The models, which were pre-trained with the COCO dataset [47], included SABL [48], RegNet [49], RetinaNet [50], PAA [51], Libra R-CNN [52], FSAF [53], Faster R-CNN [54], Dynamic R-CNN [55], and DCN [56]. The best-performing individual model was the PAA model, achieving an AP50 of 0.754. Six ensemble models were then developed to further improve detection results, of which the WFD-C ensemble model displayed the strongest results with an AP50 of 0.864, an increase of 0.11 (14.59%) over the standalone PAA model.

Later work by Hržić et al. compared the performance of deep-learning models for wrist fracture detection to that of clinicians using a dataset of 19,700 pediatric wrist radiographs [13]. The exact types of wrist fractures included in the study and the distribution of these fractures were not specified. Several models based on YOLOv4 [57] and U-Net [58] architectures were trained, with the best-performing model (YOLO 512 Anchors) achieving an accuracy of 0.95 in binary fracture detection, 0.86 in appropriately counting the number of fractures, and 0.90 in fracture localization (defined as an IoU of greater than 0.5 between the predicted and true bounding boxes for each fracture). Compared to five radiologists, the YOLO 512 Anchors model performed better than four of the radiologists (p < 0.05) and at the same level as the fifth radiologist (p = 0.0654), with a model AUC of 0.965 compared to an average radiologist AUC of 0.831. When the radiologists utilized the model to assist their reads, their performances improved by an average F1-score of 0.067 (8.0%) with significant improvement in two of the readers (p < 0.05), leading the researchers to The details and results of these studies are summarized in Table 2.

Lead Author	Year	Imaging Modality	Region	Total Number of Images	Models Used	Model Tasks	Performance Metrics
Langerhuizen	2020	X-ray	Scaphoid	300	Pre-trained CNN (Visual Geometry Group) [59]	Fracture detection	AUC 0.77, accuracy 72%, sensitivity 84%, specificity 60%
Hendrix	2021	X-ray	Scaphoid	4229	DenseNet-121 [60]	Fracture detection, scaphoid segmentation Fracture	AUC 0.87, sensitivity 78%, specificity 84%, PPV 83%, Dice score 97.4%
Hendrix	2023	X-ray	Scaphoid	19,111	InceptionV3	detection and localization, scaphoid localization, laterality classification	AUC 0.88, sensitivity 72%, specificity 93%, PPV 81%
Hardalaç	2022	X-ray	Radius, Ulna (Pediatric)	542	SABL, RegNet, RetinaNet, PAA, Libra R-CNN, FSAF, Faster R-CNN, Dynamic R-CNN, DCN	Fracture detection and localization	AP50 0.864
Hržić	2022	X-ray	Wrist (Pediatric)	19,700	YOLOv4	Fracture detection, enumeration, and localization	AUC 0.965, accuracy 95%, sensitivity 95%, PPV 96%, F1 score 0.95 Fracture enumeration: Accuracy 86% Fracture localization: Accuracy 90%

Table 2. Summary of study characteristics for wrist fracture detection models.

# 6. Hip Fractures

Hip fractures are a common cause of hospitalization, morbidity, and mortality, accounting for the majority of fracture-related healthcare expenditure in men and women over 50 years old [61]. Among patients over 65, one in three will suffer a fall each year, with 10–15% of these falls resulting in a hip fracture [62]. While pelvic radiography is often used for the evaluation of these fractures (Figure 4), a minority of patients will have radiographically occult hip fractures, with reported rates between 3–10% among negative radiographs [63]. In such cases, additional imaging such as CT or MRI must be performed, resulting in increased diagnostic costs and delays in care [64]. AI may therefore provide a powerful tool in the diagnosis of previously occult hip fractures.

A systematic review and meta-analysis by Lex et al. evaluates the performance of AI in diagnosing hip fractures on pelvic radiographs as well as predicting postoperative outcomes such as mortality [65]. Of 39 included studies, 18 used AI for the diagnosis of hip fractures, while the other 21 focused on AI's ability to predict postoperative outcomes, with a combined total of 754,537 radiographs used for the training, validation, and testing of machine learning models. On pooled data analysis, the odds ratio for the diagnostic error of the models as compared to clinician readers was calculated to be 0.79 (p = 0.36). For mortality predictions, the AUC of the models was 0.84, while traditional statistical methods using multivariable linear or logistic regression had an AUC of 0.79 (p = 0.09). As such, this study demonstrates that AI methods provide a promising approach for hip fracture diagnosis and prognosis from pelvic radiographs, although current implementations do not yet provide a substantial improvement over traditional methodologies.

Recent work by Kitamura continues to broaden the applications of AI in pelvic radiograph analysis, investigating the utility of deep learning not only for the detection of hip fractures, but also for pelvic fractures, acetabular fractures, radiograph positioning, and the presence of hardware [66]. A total of 14,374 radiographic images from 7440 patients were used to train and test the deep-learning models, with labels created to denote the presence or absence of hardware as well as radiograph positioning, which included pelvic, hip, and chest images, with each position further including a number of different views. For proximal femoral fracture detection, the model achieved an AUC of 0.95, which was comparable to other recent studies with AUCs of 0.97–0.99 [32,67] that required manual isolation of the femur using bounding boxes. Performances for fracture detection in other anatomical locations varied, with AUCs as low as 0.70 for posterior pelvic fractures and as high as 0.85 for acetabular fractures. For the detection of radiograph position and presence of hardware, the models achieved an AUC of 0.99–1.00. Altogether, this work highlights novel applications of deep learning in multiple aspects of pelvic radiography.



Figure 4. Femoral neck fracture as seen on pelvic radiograph.

In order to more accurately omit false negatives during model development, Mawatari et al. developed a deep CNN for the detection of hip fractures in pelvic radiographs using CT and MRI as a gold standard [16]. The study used a dataset consisting of radiographs from a population of 316 patients with diagnosed proximal femoral fractures who had also received CT or MRI. The radiographs, which were manually annotated by radiologists with reference to CT and MRI for ROI selection, were then used to train the model, and the diagnostic performance of seven clinician readers with and without the CNN was then evaluated. The average AUC of the readers without the CNN was 0.832, which increased to 0.876 (p < 0.05) when guided by model output. Interestingly, the AUC of the CNN alone was 0.905, outperforming the combined readers even with CNN assistance. However, this was explained by variability in the experience levels of the readers, with the more experienced clinicians scoring higher than the CNN, achieving average AUCs of 0.934 and 0.920 with and without model interpretation, respectively.

The details and results of these studies are summarized in Table 3.

Lead Author	Year	Imaging Modality	Region	Total Number of Images	Models Used	Model Tasks	Performance Metrics
Lex	2023	X-ray	Femoral Neck Intertrochanteric Subtrocanteric	754,537 <sup>1</sup>	Various models, including: AlexNet [68], GoogLeNet [69], ResNet-50, DenseNet-121, ResNet-18, PelviXNet [70], Faster RCNN	Fracture detection, outcome prediction	Diagnosis: odds ratio 0.79, sensitivity 89.3%, specificity 87.5%, F1 score 0.90 Postop mortality: AUC 0.84 Provingel 6m orgali
Kitamura	2020	X-ray	Pelvic Acetabular Hip	14,374	DenseNet-121	Fracture detection, hardware detection, imaging position	AUC 0.95 Acetabular: AUC 0.85 Anterior pelvic: AUC 0.77 Posterior pelvic: AUC 0.70 Radiograph position: AUC 0.99 Hardware presence: AUC 1.00
Mawatari	2020	X-ray	Proximal Femoral	352	GoogLeNet	Fracture detection	AUC 0.905

## Table 3. Summary of study characteristics for hip fracture detection models.

<sup>1</sup> Pooled total, with 39,598 images used for fracture detection and 714,939 for outcome prediction.

# 7. Rib Fractures

Rib fractures are the most common injury in blunt chest trauma, with an estimated prevalence of 10–38% among all trauma patients [71]. The number and pattern of rib fractures are an important indicator of trauma severity, with an increased number of fractured ribs correlated with increased morbidity and mortality [72]. Unless identified and treated appropriately, rib fractures can present with life-threatening disease, particularly in elderly patients [73]. Among imaging modalities, plain radiography and CT are the most commonly used for rib fracture detection. Although radiography is fast and convenient, the detection rate is poor, missing over 50% of rib fractures [74]. On the other hand, while CT provides a more detailed assessment of rib fractures (Figure 5), it still presents with a misdiagnosis rate of 19.2–26.8% [75,76], and diagnosis can be tedious and difficult given the large amount of CT slices and the complex shape and course of the ribs across the numerous CT sections [77]. This presents a unique opportunity for machine learning to augment the accuracy and efficiency of rib fracture reads.



Figure 5. Left rib fracture (circled) as seen on axial CT.

In 2020, Jin et al. developed the deep-learning model FracNet for the automatic detection and segmentation of rib fractures on CT images [78]. The model utilizes a dataset of 900 chest-abdomen CT scans with a total of 7473 annotated traumatic rib fractures. On testing, FracNet achieved a detection sensitivity of 92.9%, outperforming deep neural network counterparts such as 3D FCN [79] (87.8%) and 3D DeepLab [80] (91.3%) as well as reads from two expert radiologists (83.1%). FracNet also boasted an 86% decrease in reading time as compared to human clinicians, although it was associated with a higher number of false positives per scan (5.27 vs. R1: 1.34 and R2: 0.92 from the two readers, respectively). In the segmentation of rib fractures, FracNet continued to exhibit impressive performance, with a Dice score of 71.5% as compared to 3D FCN (66.2%), 3D DeepLab (68.7%), and clinician readers (64.7%).

Later work carried out by Zhang et al. in 2021 similarly used deep learning for rib fracture detection using a collection of CT images from 198 patients [81]. On testing, the trained model was able to identify 687 of 865 true fractures (79.4%), including a large number of fractures not originally detected by the two radiologists included in the study (R1: 75, R2: 66). When the clinicians utilized the model to augment their reads, they exhibited increased sensitivity in rib fracture detection (R1: 6.1%, R2: 4.8%; *p* < 0.05) as well as decreased reading times (R1: 36%, R2: 34%). Again, however, the standalone model had a higher false positive rate (0.43) than the two human readers (R1: 0.16, R2: 0.19; *p* < 0.001). Together, these studies show that deep-learning models may be a valuable asset in improving the sensitivity and efficiency of clinician rib fracture reads, although there are still challenges to using them as a standalone tool.

Recent work by Yao et al. further improved the precision of deep-learning methods with their Rib Fracture Detection System, a model utilizing a three-step algorithm for the detection of rib fractures in CT imaging consisting of bone segmentation, rib location, and rib fracture classification [82]. Using a dataset of annotated chest CTs from 1707 patients, their model achieved a precision of 0.869 and a recall of 0.913, outperforming competing algorithms such as FracNet, Fast RCNN [83], Faster RCNN, and YOLOv3 [84], and exhibiting comparable precision to that of two clinician readers (R1: 0.935, R2: 0.928) but with higher recall (R1: 0.693, R2: 0.853).

Given the focus of previous models on CT images, Gao et al. proposed a deep-learning method for automated rib fracture detection in digital radiographs [85]. CCE-Net, a novel network architecture for rib fracture detection based on the Faster RCNN framework, integrates contralateral, contextual, and edge-enhanced modules to improve AI detection of fractures. Using a dataset of 1639 radiographs with 2703 rib fractures, CCE-Net attained an AP50 of 0.911 and a recall of 0.934, an improvement of 15.76% and 6.74%, respectively, over the original Faster RCNN (0.787, 0.875). The model similarly outperformed other methods such as Libra RCNN (0.825, 0.886), Dynamic RCNN (0.887, 0.903), Cascade RCNN [86] (0.910, 0.929), and YOLO v4 (0.813, 0.881). The details and results of these studies are summarized in Table 4.

Table 4. Summary of study characteristics for wrist fracture detection models.

Lead Author	Year	Imaging Modality	Total Number of Images <sup>1</sup>	Models Used	Model Tasks	Performance Metrics
Jin	2020	СТ	900	FracNet	Fracture detection and segmentation	Sensitivity 93%, Dice score 71.5%
Zhang	2021	СТ	198	Foveal Network [87] Faster R-CNN	Rib segmentation, fracture detection	Sensitivity 79.4%
Yao	2021	СТ	1707	U-Net 3D DenseNet	Bone segmentation, fracture detection	Sensitivity 91%, PPV 87%, NPV 97%, F1 score 0.890
Gao	2022	X-ray	1639	CCE-Net	Fracture detection and localization	Sensitivity 93%, AP50 0.911

<sup>1</sup> For CT, this is listed as total number of CT scans.

# 8. Commercial Availability

A number of AI products for bone fracture detection in these anatomical locations have already received clearance from the U.S. Food and Drug Administration (FDA). A comprehensive search for FDA-approved, commercially available products was conducted using the FDA webpage for AI and machine-learning-enabled medical devices [88], which is current as of 6 December 2023.

In May 2018, the FDA approved the marketing of the AI-based algorithm OsteoDetect (Imagen Technologies) for the detection of distal radial fractures in wrist radiographs [89]. The software localizes the fracture with a bounding box, achieving an AUC of 0.965, a sensitivity of 92.1%, and a specificity of 90.2%. Imagen Technologies also evaluated the performance of OsteoDetect in assisting 24 emergency medicine physicians with reading PA and lateral wrist radiographs, finding an improvement in overall AUC (0.84 to 0.89), sensitivity (75% to 80%), and specificity (89% to 91%).

Two years later, they also received clearance for their product FractureDetect, which similarly localizes bone fractures on radiographs but includes additional anatomical locations such as the shoulder, humerus, elbow, forearm, femur, knee, tibia, fibula, ankle, pelvis, hip, and clavicle [90]. The algorithm achieved an AUC of 0.98, a sensitivity of 95%, and a specificity of 89% on a testing dataset of 11,970 radiographs, and once again improved the performance of 24 clinician readers in overall AUC (0.91 to 0.95), sensitivity (82% to 90%), and specificity (89% to 92%).

uAI EasyTriage-Rib (Shanghai United Imaging Intelligence) received FDA clearance in 2021 as a workflow optimization algorithm that automatically detects and flags CT scans with three or more acute rib fractures, with an AUC of 0.94, a sensitivity of 93%, and a specificity of 85% [91]. Several months later, BriefCase for RibFx Triage (Aidoc Medical) received approval, an AI algorithm with similar functionality in notifying users of CT scans with three or more rib fractures [92]. FDA submission data showed an AUC of 0.976, a sensitivity of 96.7%, and a specificity of 90.4% in a multicenter study of 279 test cases. The company also showed that BriefCase had a time-to-notification of only 4.2 min, while the standard of care was associated with an 89.4 min delay between image acquisition and the time when a radiologist would first open the exam.

In 2022, BoneView (Gleamer) was approved, a software that provides bounding box localization of radiographic fractures in a variety of anatomical regions including the upper and lower extremities, pelvis, hip, shoulder, clavicle, ribs, and thoracic and lumbosacral spine, with a sensitivity of 93% and a specificity of 93% [93]. A study in 2021 funded by the company showed that BoneView enhanced the sensitivity (64.8% to 75.2%; *p* < 0.001) and specificity (90.6% to 95.6%; *p* = 0.001) of 24 clinician readers while reducing the overall reading time by 6.3 s per examination (*p* = 0.046) [94].

Later, that year saw the release of Rayvolve (AZmed), which offers similar functional capabilities as FractureDetect with additional fracture localization in the hands and feet [95]. The model's standalone performance showed a sensitivity of 98.8%, a specificity of 88.6%, and an AUC of 0.986. Research conducted by the company also showed that usage of Rayvolve improved the diagnostic performance of 24 clinician readers, with an average increase in AUC from 0.846 to 0.893 (p = 0.0041), alongside increases in sensitivity (0.866 to 0.955) and specificity (0.826 to 0.831).

The performance of these products has also been externally validated by various studies. Recent work by Oppenheimer et al. demonstrated that implementation of BoneView led to increased diagnostic performance among residents in the study, with 25 additional fractures being identified in a dataset of 367 ground truth fractures, along with improvements in sensitivity (84.7% to 91.3%) and specificity (97.1% to 97.35%) [96]. Another study by Bousson et al. investigated the performance of various commercial algorithms, including Rayvolve and BoneView, in detecting acute fractures for patients admitted to the emergency department. The products demonstrated strong performance in daily radiological practice, with sensitivities of 92.6% and 91.3% and specificities of 70.4% and 90.5% for the

Product (Company)	Approval Year	Imaging Modality	Region	Functionality	Performance Metrics
OsteoDetect (Imagen Technologies)	2018	X-ray	Distal radius	Fracture detection and localization	AUC 0.97, sensitivity 92%, specificity 90%
FractureDetect (Imagen Technologies)	2020	X-ray	Ankle, clavicle, elbow, femur, forearm, hip, humerus, knee, pelvis, shoulder, tibia, fibula, wrist	Fracture detection and localization	AUC 0.98, sensitivity 95%, specificity 89%
uAI EasyTriage-Rib (Shanghai United Imaging Alliance)	2021	СТ	Ribs	Notification if $\geq 3$ fractures	AUC 0.94, sensitivity 93%, specificity 85%
BriefCase (RibFx) (Aidoc Medical)	2021	СТ	Ribs	Notification if $\geq 3$ fractures	AUC 0.98, sensitivity 97%, specificity 90%
BoneView (Gleamer)	2022	X-ray	Ankle, foot, knee, tibia, fibula, wrist, hand, elbow, forearm, humerus, shoulder, clavicle, pelvis, hip, femur, ribs, thoracic spine, lumbocargal spine	Fracture detection and localization	AUC 0.93, sensitivity 93%, specificity 93%
Rayvolve (AZmed)	2022	X-ray	Ankle, clavicle, elbow, forearm, hip, humerus, knee, pelvis, shoulder, tibia, fibula, wrist, hand, foot	Fracture detection and localization	AUC 0.99, sensitivity 99%, specificity 89%

two algorithms, respectively [97]. A summary table detailing the characteristics of these products is provided in Table 5.

Table 5. Summary of FDA-approved AI products for bone fracture detection.

# 9. Discussion

In summary, we detail recent contributions of AI methodology to the detection of bone fractures of the ankle, wrist, hip, and ribs, as well as currently available FDA-approved products for these tasks. However, there are countless additional applications of AI in musculoskeletal imaging. Numerous studies have investigated the performance of deep-learning methods for fracture detection in other anatomical locations such as the vertebrae [98,99], humerus [15], femur [32,100,101], shoulder [102,103], elbow [104,105], and skull [106]. While not fully covered within the scope of this review, we briefly summarize current AI methodologies for these types of fractures:

In the vertebrae, Shen et al., in 2023, trained a multitask detection network on 11,397 radiographic images, achieving an overall internal accuracy, sensitivity, and specificity of 97.41%, 84.08%, and 97.25%, respectively, and an overall external accuracy, sensitivity, and specificity of 96.85%, 83.35%, and 94.70%, respectively [98]. Zhang et al. in 2023 trained a U-Net/GCN/ResNet-based CNN model on 1217 CT images. For fracture detection, they achieved a sensitivity of 95.23%, an accuracy of 97.93%, and a specificity of 98.35%. For fracture classification, they achieved AUCs of 0.904, 0.945, 0.878, and 0.942 for the four types of vertebral fractures, respectively [99].

In the humerus, Chung et al., in 2018, trained a ResNet CNN on 1891 radiograph images of four types of humerus fractures, achieving an overall AUC of 0.996, sensitivity of 0.99, and specificity of 0.97 [15]. In the femur, Beyaz et al., in 2020, trained a CNN on

234 radiographic images with a genetic algorithm approach to optimize hyperparameters, achieving an accuracy of 79.3%, sensitivity of 82.9%, and specificity of 72.9% [100]. Similarly, Gale et al., in 2017, trained a DenseNet CNN on 53,278 radiographic images, achieving an AUC of 0.994, an accuracy of 97%, a precision of 99%, a recall of 95%, and an F1 score of 0.97 [32]. This was further validated by Oakden-Rayner et al. in 2022, who trained the same model on 45,786 radiographic images and achieved an AUC of 0.994 on internal validation and an AUC of 0.969 on external validation [101].

In the shoulder, Uysal et al., in 2021, trained 26 deep-learning models on the MURA Xray dataset for two ensemble models, with the first achieving an accuracy of 0.846 and AUC of 0.886, and the second having an accuracy of 0.847 and AUC of 0.870 [102]. Magnéli et al. in 2023 trained a modified ResNet CNN on 7189 radiographic images, achieving an AUC of 0.96 for clavicle fractures and 0.87 for scapula fractures [103]. In the elbow, Rayan et al. in 2019 trained a multiview Xception CNN on 58,817 pediatric radiographic images, achieving an AUC of 0.95, accuracy of 88%, sensitivity of 91%, and specificity of 84% [104]. Luo et al., in 2021, took a knowledge-guided curriculum learning approach to train a multiview CNN model on 1964 radiographic images, achieving an AUC of 0.974 and an accuracy of 0.889 [105].

In the skull, Choi et al., in 2022, trained a YOLOv3 CNN on 413 radiographic images, achieving an AUC of 0.922, sensitivity of 81.1%, and specificity of 91.3% on an internal test set, and an AUC of 0.870, sensitivity of 78.9% and specificity of 88.2% on the external test set. Model-assisted AUC improvements of 0.094 and 0.069 were observed for radiology residents and emergency physicians, respectively, compared to diagnosis without AI assistance. However, no statistically significant improvement was observed in pediatric radiologists [106].

Expanding beyond bone fractures, models have also been developed for the diagnosis of ACL tears [107], meniscal tears [108,109], osteoarthritis [110], and cartilage lesions [111]. Other models are capable of automated grading of various musculoskeletal diseases [112–114] as well as augmenting several aspects of the image acquisition process, such as protocoling [115], reducing MRI acquisition times [116], and improving image quality [117]. These examples only touch the surface of what is possible with current technology, especially as the field continues to grow.

#### 9.1. Limitations

While recent advances in machine learning have made applications of AI in bone fracture detection more feasible than ever, there remain numerous challenges to be addressed for both the development and implementation of AI methods. Conventional deep-learning methods, as used by the majority of studies in this review, require large amounts of annotated data, which is a tedious, time-consuming, and often prohibitively expensive process. Given the complex regulatory and privacy concerns associated with the sharing of medical images, the field currently suffers from an overall lack of high-quality annotated images, as many datasets are not publicly available for research purposes [118,119]. This issue can be seen in many of the studies covered by this review, which were generally performed as retrospective, single-center studies on internal datasets unless otherwise specified. Because images are acquired using different protocols varying by institution, this lack of external validation presents a major limitation to the generalizability and accuracy of developed models, as they may suffer from decreased performance when used on external datasets. Lack of standardization across datasets also leads to intrinsic model biases depending on the geographic locations, pathologies, and imaging modalities represented in a given dataset [120,121].

Similarly, imaging quality may also affect model performance in fracture detection. Work by Lu et al. proposes a reinforcement learning and transformer-based solution to assess and account for image distortion (e.g., motion artifacts, noise, contrast dosing) to improve diagnostic accuracy in coronary CT angiography [122]. While this has not yet been explicitly studied in the context of fracture detection to the best of the authors' knowledge, techniques such as this may have utility in this space.

These issues generate questions surrounding the performance of proposed AI models in actual clinical settings. As shown in this review, AI methods are not devoid of making errors, with the inherent architecture of these models making it difficult to decipher the decision-making and rationale behind incorrect outputs. Due to the diversity of visual characteristics associated with any given pathology, even overall well-performing models may consistently misdiagnose specific subsets of cases, especially when imaging features are subtle or underrepresented [123]. Given that models generally do not report uncertainty behind decisions, this may lead to model outputs that would be clearly erroneous to a human reader. Issues surrounding the accountability of such decisions and their impact on patient care must be thoroughly and explicitly addressed before these models can be seamlessly integrated into clinical workflows.

## 9.2. Future Directions

The future of AI in radiology is promising. Current models are capable of clinical tasks such as diagnosis, prognosis, classification, and segmentation, with performances that match human clinicians and only continue to improve.

An important contribution to the field has been the recent improvement of selfsupervised and unsupervised learning methods, which are able to learn from large amounts of data without the need for manual annotations [124]. Models using these approaches have already achieved similar or even better performances than models trained with traditional supervised learning methods [125–127], giving rise to the idea of AI foundation models. While most current models are trained for specific applications with limited fluidity in adapting to new tasks, foundation models are able to be trained just once on large-scale datasets and subsequently fine-tuned to perform a wide array of downstream tasks, offering newfound flexibility and generalizability. Current medical imaging foundation models are capable of tasks such as the automatic segmentation of medical images [128,129], cardiac function assessment with text report generation from echocardiograms [130], and the diagnosis and prognosis of ocular diseases associated with retinal images [131].

The question of how these algorithms will be implemented into clinical workflows remains unanswered. Numerous studies have already demonstrated the feasibility of AI in improving the accuracy and efficiency of clinician readers in addition to boasting impressive standalone performance. Potential applications of AI models in radiology include augmentation of reads, validation of results, or perhaps the offloading of simple diagnoses to AI to allow radiologists to focus on more cognitively challenging tasks, as suggested by Jha and Topol [132]. However, there is no easy answer to this question, as it involves a complex interplay between patients, radiologists, and other healthcare providers in addition to associated reimbursement policies and legal ramifications. We continue to await further research on how these algorithms will impact patient outcomes, which may help uncover the optimal way to utilize AI in clinical practice, especially as the scope of AI continues to evolve.

## 10. Conclusions

Deep-learning models for bone fracture detection in the ankle, wrist, hip, and ribs have achieved performance that is comparable or superior to that of clinician readers. These algorithms are already becoming commercially available for integration into clinical workflows, providing numerous benefits to radiologists such as increased diagnostic accuracy and efficiency. The utility of AI in fracture detection and radiology on a broader scale is an area of active research, and the capabilities of generated models continue to rapidly evolve. While their exact applications in clinical practice remain to be determined, AI methods have the potential to optimize radiologist workflows and enhance healthcare delivery for both patients and providers, and clinicians should be informed on the current state of AI as it will likely impact the future practice of radiology. **Author Contributions:** Conceptualization, A.T., E.K., Y.K. and T.D.; methodology, Z.L.; data curation, A.T., A.Y., A.Z., E.K., N.P., S.L., Y.K. and Z.L.; writing—original draft preparation, A.T., A.Y., E.K., S.L., Y.K. and Z.L.; writing—review and editing, A.T., A.Z., N.P. and T.D.; supervision, T.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: T.D. is the Managing Partner of RadImageNet LLC.

# References

- Wu, A.-M.; Bisignano, C.; James, S.L.; Abady, G.G.; Abedi, A.; Abu-Gharbieh, E.; Alhassan, R.K.; Alipour, V.; Arabloo, J.; Asaad, M.; et al. Global, Regional, and National Burden of Bone Fractures in 204 Countries and Territories, 1990–2019: A Systematic Analysis from the Global Burden of Disease Study 2019. *Lancet Healthy Longev.* 2021, 2, e580–e592. [CrossRef] [PubMed]
- Blume, S.W.; Curtis, J.R. Medical Costs of Osteoporosis in the Elderly Medicare Population. Osteoporos. Int. 2011, 22, 1835–1844. [CrossRef] [PubMed]
- Lewiecki, E.M.; Ortendahl, J.D.; Vanderpuye-Orgle, J.; Grauer, A.; Arellano, J.; Lemay, J.; Harmon, A.L.; Broder, M.S.; Singer, A.J. Healthcare Policy Changes in Osteoporosis Can Improve Outcomes and Reduce Costs in the United States. *JBMR Plus* 2019, 3, e10192. [CrossRef]
- 4. Guly, H. Diagnostic Errors in an Accident and Emergency Department. Emerg. Med. J. 2001, 18, 263–269. [CrossRef] [PubMed]
- 5. Hussain, F.; Cooper, A.; Carson-Stevens, A.; Donaldson, L.; Hibbert, P.; Hughes, T.; Edwards, A. Diagnostic Error in the Emergency Department: Learning from National Patient Safety Incident Report Analysis. *BMC Emerg. Med.* **2019**, *19*, 77. [CrossRef]
- 6. Brennan, T.A.; Leape, L.L.; Laird, N.M.; Hebert, L.; Localio, A.R.; Lawthers, A.G.; Newhouse, J.P.; Weiler, P.C.; Hiatt, H.H. Incidence of Adverse Events and Negligence in Hospitalized Patients. *N. Engl. J. Med.* **1991**, 324, 370–376. [CrossRef] [PubMed]
- 7. Kim, D.H.; MacKinnon, T. Artificial Intelligence in Fracture Detection: Transfer Learning from Deep Convolutional Neural Networks. *Clin. Radiol.* **2018**, *73*, 439–445. [CrossRef] [PubMed]
- 8. Dan Lantsman, C.; Barash, Y.; Klang, E.; Guranda, L.; Konen, E.; Tau, N. Trend in Radiologist Workload Compared to Number of Admissions in the Emergency Department. *Eur. J. Radiol.* **2022**, *149*, 110195. [CrossRef] [PubMed]
- 9. Mizrahi, D.J.; Parker, L.; Zoga, A.M.; Levin, D.C. National Trends in the Utilization of Skeletal Radiography From 2003 to 2015. *J. Am. Coll. Radiol.* 2018, *15*, 1408–1414. [CrossRef] [PubMed]
- 10. Hallas, P.; Ellingsen, T. Errors in Fracture Diagnoses in the Emergency Department—Characteristics of Patients and Diurnal Variation. *BMC Emerg. Med.* **2006**, *6*, 4. [CrossRef] [PubMed]
- 11. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J.W.L. Artificial Intelligence in Radiology. *Nat. Rev. Cancer* 2018, 18, 500–510. [CrossRef] [PubMed]
- Lindsey, R.; Daluiski, A.; Chopra, S.; Lachapelle, A.; Mozer, M.; Sicular, S.; Hanel, D.; Gardner, M.; Gupta, A.; Hotchkiss, R.; et al. Deep Neural Network Improves Fracture Detection by Clinicians. *Proc. Natl. Acad. Sci. USA* 2018, 115, 11591–11596. [CrossRef] [PubMed]
- 13. Hržić, F.; Tschauner, S.; Sorantin, E.; Štajduhar, I. Fracture Recognition in Paediatric Wrist Radiographs: An Object Detection Approach. *Mathematics* **2022**, *10*, 2939. [CrossRef]
- Hendrix, N.; Scholten, E.; Vernhout, B.; Bruijnen, S.; Maresch, B.; de Jong, M.; Diepstraten, S.; Bollen, S.; Schalekamp, S.; de Rooij, M.; et al. Development and Validation of a Convolutional Neural Network for Automated Detection of Scaphoid Fractures on Conventional Radiographs. *Radiol. Artif. Intell.* 2021, *3*, e200260. [CrossRef] [PubMed]
- 15. Chung, S.W.; Han, S.S.; Lee, J.W.; Oh, K.-S.; Kim, N.R.; Yoon, J.P.; Kim, J.Y.; Moon, S.H.; Kwon, J.; Lee, H.-J.; et al. Automated Detection and Classification of the Proximal Humerus Fracture by Using Deep Learning Algorithm. *Acta Orthop.* **2018**, *89*, 468–473. [CrossRef] [PubMed]
- Mawatari, T.; Hayashida, Y.; Katsuragawa, S.; Yoshimatsu, Y.; Hamamura, T.; Anai, K.; Ueno, M.; Yamaga, S.; Ueda, I.; Terasawa, T.; et al. The Effect of Deep Convolutional Neural Networks on Radiologists' Performance in the Detection of Hip Fractures on Digital Pelvic Radiographs. *Eur. J. Radiol.* 2020, 130, 109188. [CrossRef] [PubMed]
- 17. Urakawa, T.; Tanaka, Y.; Goto, S.; Matsuzawa, H.; Watanabe, K.; Endo, N. Detecting Intertrochanteric Hip Fractures with Orthopedist-Level Accuracy Using a Deep Convolutional Neural Network. *Skelet. Radiol.* **2019**, *48*, 239–244. [CrossRef] [PubMed]
- Olczak, J.; Fahlberg, N.; Maki, A.; Razavian, A.S.; Jilert, A.; Stark, A.; Sköldenberg, O.; Gordon, M. Artificial Intelligence for Analyzing Orthopedic Trauma Radiographs. Acta Orthop. 2017, 88, 581–586. [CrossRef] [PubMed]
- 19. Tomita, N.; Cheung, Y.Y.; Hassanpour, S. Deep Neural Networks for Automatic Detection of Osteoporotic Vertebral Fractures on CT Scans. *Comput. Biol. Med.* **2018**, *98*, 8–15. [CrossRef] [PubMed]

- Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews. *BMJ* 2021, 372, n160. [CrossRef] [PubMed]
- Da Fonseca, L.L.; Nunes, I.G.; Nogueira, R.R.; Martins, G.E.V.; Mesencio, A.C.; Kobata, S.I. Reproducibility of the Lauge-Hansen, Danis-Weber, and AO Classifications for Ankle Fractures. *Rev. Bras. Ortop.* 2017, 53, 101–106. [CrossRef]
- Budny, A.M.; Young, B.A. Analysis of Radiographic Classifications for Rotational Ankle Fractures. *Clin. Podiatr. Med. Surg.* 2008, 25, 139–152. [CrossRef] [PubMed]
- 23. Yip, G.; Hay, D.; Stringfellow, T.; Ahluwalia, A.; Ahluwalia, R. Assessment and Management of Adult Ankle Fractures: Understanding the Evidence. *Br. J. Hosp. Med.* **2021**, *82*, 1–9. [CrossRef] [PubMed]
- 24. Moonen, P.-J.; Mercelina, L.; Boer, W.; Fret, T. Diagnostic Error in the Emergency Department: Follow up of Patients with Minor Trauma in the Outpatient Clinic. *Scand. J. Trauma. Resusc. Emerg. Med.* **2017**, *25*, 13. [CrossRef] [PubMed]
- Huysse, W.; Burssens, A.; Peiffer, M.; Cornelis, B.; Stufkens, S.A.S.; Kerkhoffs, G.M.M.J.; Buedts, K.; Audenaert, E.A. Morphometric Analysis of the Incisura Fibularis in Patients with Unstable High Ankle Sprains. *Skelet. Radiol.* 2021, 50, 1141–1150. [CrossRef] [PubMed]
- Valderrabano, V.; Horisberger, M.; Russell, I.; Dougall, H.; Hintermann, B. Etiology of Ankle Osteoarthritis. *Clin. Orthop. Relat. Res.* 2009, 467, 1800–1806. [CrossRef] [PubMed]
- 27. Delco, M.L.; Kennedy, J.G.; Bonassar, L.J.; Fortier, L.A. Post-Traumatic Osteoarthritis of the Ankle: A Distinct Clinical Entity Requiring New Research Approaches. *J. Orthop. Res.* 2017, *35*, 440–453. [CrossRef] [PubMed]
- Ashkani-Esfahani, S.; Mojahed Yazdi, R.; Bhimani, R.; Kerkhoffs, G.M.; Maas, M.; DiGiovanni, C.W.; Lubberts, B.; Guss, D. Detection of Ankle Fractures Using Deep Learning Algorithms. *Foot Ankle Surg.* 2022, 28, 1259–1265. [CrossRef] [PubMed]
- Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.
- Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
- 32. Gale, W.; Oakden-Rayner, L.; Carneiro, G.; Bradley, A.P.; Palmer, L.J. Detecting Hip Fractures with Radiologist-Level Performance Using Deep Neural Networks 2017. *arXiv* 2017, arXiv:1711.06504.
- Cao, Y.; Wang, H.; Moradi, M.; Prasanna, P.; Syeda-Mahmood, T.F. Fracture Detection in X-Ray Images through Stacked Random Forests Feature Fusion. In Proceedings of the 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), Brooklyn, NY, USA, 16–19 April 2015; pp. 801–805.
- Kitamura, G.; Chung, C.Y.; Moore, B.E. Ankle Fracture Detection Utilizing a Convolutional Neural Network Ensemble Implemented with a Small Sample, De Novo Training, and Multiview Incorporation. J. Digit. Imaging 2019, 32, 672–677. [CrossRef] [PubMed]
- Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
- Hedström, E.M.; Svensson, O.; Bergström, U.; Michno, P. Epidemiology of Fractures in Children and Adolescents. *Acta Orthop.* 2010, *81*, 148–153. [CrossRef] [PubMed]
- 37. Rhemrev, S.J.; Ootes, D.; Beeres, F.J.; Meylaerts, S.A.; Schipper, I.B. Current Methods of Diagnosis and Treatment of Scaphoid Fractures. *Int. J. Emerg. Med.* **2011**, *4*, 4. [CrossRef] [PubMed]
- Steinmann, S.P.; Adams, J.E. Scaphoid Fractures and Nonunions: Diagnosis and Treatment. J. Orthop. Sci. 2006, 11, 424–431. [CrossRef] [PubMed]
- 39. Roolker, W.; Maas, M.; Broekhuizen, A.H. Diagnosis and Treatment of Scaphoid Fractures, Can Non-Union Be Prevented? *Arch. Orth. Traum. Surg.* **1999**, *119*, 428–431. [CrossRef] [PubMed]
- 40. Prosser, G.H.; Isbister, E.S. The Presentation of Scaphoid Non-Union. Injury 2003, 34, 65–67. [CrossRef] [PubMed]
- De Zwart, A.D.; Beeres, F.J.P.; Rhemrev, S.J.; Bartlema, K.; Schipper, I.B. Comparison of MRI, CT and Bone Scintigraphy for Suspected Scaphoid Fractures. *Eur. J. Trauma. Emerg. Surg.* 2016, 42, 725–731. [CrossRef] [PubMed]
- 42. Buul, M.T.; van Beek, E.; Broekhuizen, A.H.; Bakker, A.J.; Bos, K.E.; van Royen, E. Radiography and Scintigraphy of Suspected Scaphoid Fracture. A Long-Term Study in 160 Patients. *J. Bone Jt. Surg. Br. Vol.* **1993**, 75-*B*, 61–65. [CrossRef]
- Gibney, B.; Smith, M.; Moughty, A.; Kavanagh, E.C.; Hynes, D.; MacMahon, P.J. Incorporating Cone-Beam CT Into the Diagnostic Algorithm for Suspected Radiocarpal Fractures: A New Standard of Care? *Am. J. Roentgenol.* 2019, 213, 1117–1123. [CrossRef] [PubMed]
- Langerhuizen, D.W.G.; Bulstra, A.E.J.; Janssen, S.J.; Ring, D.; Kerkhoffs, G.M.M.J.; Jaarsma, R.L.; Doornberg, J.N. Is Deep Learning on Par with Human Observers for Detection of Radiographically Visible and Occult Fractures of the Scaphoid? *Clin. Orthop. Relat. Res.* 2020, 478, 2653–2659. [CrossRef] [PubMed]

- Hendrix, N.; Hendrix, W.; van Dijke, K.; Maresch, B.; Maas, M.; Bollen, S.; Scholtens, A.; de Jonge, M.; Ong, L.-L.S.; van Ginneken, B.; et al. Musculoskeletal Radiologist-Level Performance by Using Deep Learning for Detection of Scaphoid Fractures on Conventional Multi-View Radiographs of Hand and Wrist. *Eur. Radiol.* 2023, *33*, 1575–1588. [CrossRef] [PubMed]
- 46. Hardalaç, F.; Uysal, F.; Peker, O.; Çiçeklidağ, M.; Tolunay, T.; Tokgöz, N.; Kutbay, U.; Demirciler, B.; Mert, F. Fracture Detection in Wrist X-Ray Images Using Deep Learning-Based Object Detection Models. *Sensors* **2022**, *22*, 1285. [CrossRef] [PubMed]
- Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; Volume 8693, pp. 740–755. ISBN 978-3-319-10601-4.
- Wang, J.; Zhang, W.; Cao, Y.; Chen, K.; Pang, J.; Gong, T.; Shi, J.; Loy, C.C.; Lin, D. Side-Aware Boundary Localization for More Precise Object Detection. In *Computer Vision—ECCV 20142020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12349, pp. 403–419. ISBN 978-3-030-58547-1.
- Xu, J.; Pan, Y.; Pan, X.; Hoi, S.; Yi, Z.; Xu, Z. RegNet: Self-Regulated Network for Image Classification. *IEEE Trans. Neural Netw.* Learn. Syst. 2023, 34, 9562–9567. [CrossRef] [PubMed]
- Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 42, 318–327. [CrossRef] [PubMed]
- Kim, K.; Lee, H.S. Probabilistic Anchor Assignment with IoU Prediction for Object Detection. In *Computer Vision—ECCV ECCV* 2020; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12370, pp. 355–371. ISBN 978-3-030-58594-5.
- Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; Lin, D. Libra R-CNN: Towards Balanced Learning for Object Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 821–830.
- 53. Zhu, C.; He, Y.; Savvides, M. Feature Selective Anchor-Free Module for Single-Shot Object Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 840–849.
- 54. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
- Zhang, H.; Chang, H.; Ma, B.; Wang, N.; Chen, X. Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training. In *Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12360, pp. 260–275. ISBN 978-3-030-58554-9.
- Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 764–773.
- 57. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection 2020. *arXiv* 2020, arXiv:2004.10934.
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
- Zhong, S.; Li, K.; Feng, R. Deep Convolutional Hamming Ranking Network for Large Scale Image Retrieval. In Proceedings of the 11th World Congress on Intelligent Control and Automation, Shenyang, China, 29 June–4 July 2014; pp. 1018–1023.
- Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2261–2269.
- Kanis, J.A.; Odén, A.; McCloskey, E.V.; Johansson, H.; Wahl, D.A.; Cooper, C. on behalf of the IOF Working Group on Epidemiology and Quality of Life. A Systematic Review of Hip Fracture Incidence and Probability of Fracture Worldwide. *Osteoporos. Int.* 2012, 23, 2239–2256. [CrossRef] [PubMed]
- 62. Berry, S.D.; Miller, R.R. Falls: Epidemiology, Pathophysiology, and Relationship to Fracture. *Curr. Osteoporos. Rep.* **2008**, *6*, 149–154. [CrossRef] [PubMed]
- 63. Rizzo, P.F.; Gould, E.S.; Lyden, J.P.; Asnis, S.E. Diagnosis of Occult Fractures about the Hip. Magnetic Resonance Imaging Compared with Bone-Scanning. *JBJS* **1993**, *75*, 395. [CrossRef] [PubMed]
- 64. Rehman, H.; Clement, R.G.E.; Perks, F.; White, T.O. Imaging of Occult Hip Fractures: CT or MRI? *Injury* **2016**, *47*, 1297–1301. [CrossRef] [PubMed]
- 65. Lex, J.R.; Di Michele, J.; Koucheki, R.; Pincus, D.; Whyne, C.; Ravi, B. Artificial Intelligence for Hip Fracture Detection and Outcome Prediction: A Systematic Review and Meta-Analysis. *JAMA Netw. Open* **2023**, *6*, e233391. [CrossRef] [PubMed]
- 66. Kitamura, G. Deep Learning Evaluation of Pelvic Radiographs for Position, Hardware Presence, and Fracture Detection. *Eur. J. Radiol.* 2020, *130*, 109139. [CrossRef] [PubMed]
- Krogue, J.D.; Cheng, K.V.; Hwang, K.M.; Toogood, P.; Meinberg, E.G.; Geiger, E.J.; Zaid, M.; McGill, K.C.; Patel, R.; Sohn, J.H.; et al. Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning 2019. *Adiology Artif. Intell.* 2020, 2, e190023.

- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 2017, 60, 84–90. [CrossRef]
- Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–9.
- Cheng, C.-T.; Wang, Y.; Chen, H.-W.; Hsiao, P.-M.; Yeh, C.-N.; Hsieh, C.-H.; Miao, S.; Xiao, J.; Liao, C.-H.; Lu, L. A Scalable Physician-Level Deep Learning Algorithm Detects Universal Trauma on Pelvic Radiographs. *Nat. Commun.* 2021, 12, 1066. [CrossRef] [PubMed]
- 71. Ziegler, D.W.; Agarwal, N.N. The morbidity and mortality of rib fractures. *J. Trauma. Acute Care Surg.* **1994**, 37, 975. [CrossRef] [PubMed]
- 72. Talbot, B.S.; Gange, C.P.; Chaturvedi, A.; Klionsky, N.; Hobbs, S.K.; Chaturvedi, A. Traumatic Rib Injury: Patterns, Imaging Pitfalls, Complications, and Treatment. *RadioGraphics* **2017**, *37*, 628–651. [CrossRef] [PubMed]
- Barnea, Y.; Kashtan, H.; Skornick, Y.; Werbin, N. Isolated Rib Fractures in Elderly Patients: Mortality and Morbidity. *Can. J. Surg.* 2002, 45, 43–46. [PubMed]
- 74. Traub, M.; Stevenson, M.; McEvoy, S.; Briggs, G.; Lo, S.K.; Leibman, S.; Joseph, T. The Use of Chest Computed Tomography versus Chest X-Ray in Patients with Major Blunt Trauma. *Injury* **2007**, *38*, 43–47. [CrossRef] [PubMed]
- 75. Banaste, N.; Caurier, B.; Bratan, F.; Bergerot, J.-F.; Thomson, V.; Millet, I. Whole-Body CT in Patients with Multiple Traumas: Factors Leading to Missed Injury. *Radiology* **2018**, *289*, 374–383. [CrossRef] [PubMed]
- Cho, S.H.; Sung, Y.M.; Kim, M.S. Missed Rib Fractures on Evaluation of Initial Chest CT for Trauma Patients: Pattern Analysis and Diagnostic Value of Coronal Multiplanar Reconstruction Images with Multidetector Row CT. *BJR* 2012, *85*, e845–e850. [CrossRef] [PubMed]
- 77. Ringl, H.; Lazar, M.; Töpker, M.; Woitek, R.; Prosch, H.; Asenbaum, U.; Balassy, C.; Toth, D.; Weber, M.; Hajdu, S.; et al. The Ribs Unfolded—A CT Visualization Algorithm for Fast Detection of Rib Fractures: Effect on Sensitivity and Specificity in Trauma Patients. *Eur. Radiol.* 2015, 25, 1865–1874. [CrossRef]
- 78. Jin, L.; Yang, J.; Kuang, K.; Ni, B.; Gao, Y.; Sun, Y.; Gao, P.; Ma, W.; Tan, M.; Kang, H.; et al. Deep-Learning-Assisted Detection and Segmentation of Rib Fractures from CT Scans: Development and Validation of FracNet. *EBioMedicine* 2020, 62, 103106. [CrossRef] [PubMed]
- Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 3431–3440.
- Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Computer Vision—ECCV 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 833–851.
- Zhang, B.; Jia, C.; Wu, R.; Lv, B.; Li, B.; Li, F.; Du, G.; Sun, Z.; Li, X. Improving Rib Fracture Detection Accuracy and Reading Efficiency with Deep Learning-Based Detection Software: A Clinical Evaluation. *Br. J. Radiol.* 2021, 94, 20200870. [CrossRef] [PubMed]
- 82. Yao, L.; Guan, X.; Song, X.; Tan, Y.; Wang, C.; Jin, C.; Chen, M.; Wang, H.; Zhang, M. Rib Fracture Detection System Based on Deep Learning. *Sci. Rep.* 2021, *11*, 23513. [CrossRef] [PubMed]
- 83. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
- 84. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement 2018. arXiv 2018, arXiv:1804.02767.
- 85. Gao, Y.; Liu, H.; Jiang, L.; Yang, C.; Yin, X.; Coatrieux, J.-L.; Chen, Y. CCE-Net: A Rib Fracture Diagnosis Network Based on Contralateral, Contextual, and Edge Enhanced Modules. *Biomed. Signal Process. Control* **2022**, 75, 103620. [CrossRef]
- Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving Into High Quality Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
- Brosch, T.; Saalbach, A. Foveal Fully Convolutional Nets for Multi-Organ Segmentation. In Proceedings of the Medical Imaging 2018: Image Processing; SPIE: Bellingham, WA, USA, 2018; Volume 10574, pp. 198–206.
- U.S. Food & Drug Administration. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. Available online: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learningaiml-enabled-medical-devices (accessed on 18 February 2024).
- U.S. FDA Center for Devices and Radiological Health. Evaluation of Automatic Class III Designation for OsteoDetect: Decision Summary. Available online: https://www.accessdata.fda.gov/cdrh\_docs/reviews/DEN180005.pdf (accessed on 20 January 2024).
- U.S. FDA Center for Devices and Radiological Health. K193417 FractureDetect (FX) Approval Letter. Available online: https://www.accessdata.fda.gov/cdrh\_docs/pdf19/K193417.pdf (accessed on 27 January 2024).
- 91. U.S. FDA Center for Devices and Radiological Health. K193271 uAI EasyTriage-Rib Approval Letter. Available online: https://www.accessdata.fda.gov/cdrh\_docs/pdf19/K193271.pdf (accessed on 27 January 2024).
- 92. U.S. FDA Center for Devices and Radiological Health. K202992 BriefCase for RibFx Triage Approval Letter. Available online: https://www.accessdata.fda.gov/cdrh\_docs/pdf20/K202992.pdf (accessed on 18 February 2024).

- 93. U.S. FDA Center for Devices and Radiological Health. K212365 BoneView Approval Letter. Available online: https://www.accessdata.fda.gov/cdrh\_docs/pdf21/K212365.pdf (accessed on 27 January 2024).
- Guermazi, A.; Tannoury, C.; Kompel, A.J.; Murakami, A.M.; Ducarouge, A.; Gillibert, A.; Li, X.; Tournier, A.; Lahoud, Y.; Jarraya, M.; et al. Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence. *Radiology* 2022, 302, 627–636. [CrossRef]
- 95. U.S. FDA Center for Devices and Radiological Health. K220164 Rayvolve Approval Letter. Available online: https://www.accessdata.fda.gov/cdrh\_docs/pdf22/K220164.pdf (accessed on 28 January 2024).
- Oppenheimer, J.; Lüken, S.; Hamm, B.; Niehues, S.M. A Prospective Approach to Integration of AI Fracture Detection Software in Radiographs into Clinical Workflow. *Life* 2023, 13, 223. [CrossRef] [PubMed]
- 97. Bousson, V.; Attané, G.; Benoist, N.; Perronne, L.; Diallo, A.; Hadid-Beurrier, L.; Martin, E.; Hamzi, L.; Depil Duval, A.; Revue, E.; et al. Artificial Intelligence for Detecting Acute Fractures in Patients Admitted to an Emergency Department: Real-Life Performance of Three Commercial Algorithms. *Acad. Radiol.* 2023, 30, 2118–2139. [CrossRef]
- Shen, L.; Gao, C.; Hu, S.; Kang, D.; Zhang, Z.; Xia, D.; Xu, Y.; Xiang, S.; Zhu, Q.; Xu, G.; et al. Using Artificial Intelligence to Diagnose Osteoporotic Vertebral Fractures on Plain Radiographs. J. Bone Miner. Res. 2023, 38, 1278–1287. [CrossRef] [PubMed]
- Zhang, J.; Liu, F.; Xu, J.; Zhao, Q.; Huang, C.; Yu, Y.; Yuan, H. Automated Detection and Classification of Acute Vertebral Body Fractures Using a Convolutional Neural Network on Computed Tomography. *Front. Endocrinol.* 2023, 14, 1132725. [CrossRef] [PubMed]
- 100. Beyaz, S.; Açıcı, K.; Sümer, E. Femoral Neck Fracture Detection in X-Ray Images Using Deep Learning and Genetic Algorithm Approaches. *Jt. Dis. Relat. Surg.* 2020, *31*, 175–183. [CrossRef] [PubMed]
- 101. Oakden-Rayner, L.; Gale, W.; Bonham, T.A.; Lungren, M.P.; Carneiro, G.; Bradley, A.P.; Palmer, L.J. Validation and Algorithmic Audit of a Deep Learning System for the Detection of Proximal Femoral Fractures in Patients in the Emergency Department: A Diagnostic Accuracy Study. *Lancet Digit. Health* 2022, 4, e351–e358. [CrossRef] [PubMed]
- 102. Uysal, F.; Hardalaç, F.; Peker, O.; Tolunay, T.; Tokgöz, N. Classification of Shoulder X-Ray Images with Deep Learning Ensemble Models. *Appl. Sci.* 2021, *11*, 2723. [CrossRef]
- 103. Magnéli, M.; Ling, P.; Gislén, J.; Fagrell, J.; Demir, Y.; Arverud, E.D.; Hallberg, K.; Salomonsson, B.; Gordon, M. Deep Learning Classification of Shoulder Fractures on Plain Radiographs of the Humerus, Scapula and Clavicle. *PLoS ONE* 2023, 18, e0289808. [CrossRef] [PubMed]
- 104. Rayan, J.C.; Reddy, N.; Kan, J.H.; Zhang, W.; Annapragada, A. Binomial Classification of Pediatric Elbow Fractures Using a Deep Learning Multiview Approach Emulating Radiologist Decision Making. *Radiol. Artif. Intell.* **2019**, *1*, e180015. [CrossRef]
- Luo, J.; Kitamura, G.; Arefan, D.; Doganay, E.; Panigrahy, A.; Wu, S. Knowledge-Guided Multiview Deep Curriculum Learning for Elbow Fracture Classification. *Mach. Learn. Med. Imaging* 2021, 12966, 555–564. [CrossRef] [PubMed]
- 106. Choi, J.W.; Cho, Y.J.; Ha, J.Y.; Lee, Y.Y.; Koh, S.Y.; Seo, J.Y.; Choi, Y.H.; Cheon, J.-E.; Phi, J.H.; Kim, I.; et al. Deep Learning-Assisted Diagnosis of Pediatric Skull Fractures on Plain Radiographs. *Korean J. Radiol.* **2022**, *23*, 343–354. [CrossRef] [PubMed]
- 107. Bien, N.; Rajpurkar, P.; Ball, R.L.; Irvin, J.; Park, A.; Jones, E.; Bereket, M.; Patel, B.N.; Yeom, K.W.; Shpanskaya, K.; et al. Deep-Learning-Assisted Diagnosis for Knee Magnetic Resonance Imaging: Development and Retrospective Validation of MRNet. *PLoS Med.* 2018, 15, e1002699. [CrossRef] [PubMed]
- 108. Couteaux, V.; Si-Mohamed, S.; Nempont, O.; Lefevre, T.; Popoff, A.; Pizaine, G.; Villain, N.; Bloch, I.; Cotten, A.; Boussel, L. Automatic Knee Meniscus Tear Detection and Orientation Classification with Mask-RCNN. *Diagn. Interv. Imaging* 2019, 100, 235–242. [CrossRef] [PubMed]
- Roblot, V.; Giret, Y.; Bou Antoun, M.; Morillot, C.; Chassin, X.; Cotten, A.; Zerbib, J.; Fournier, L. Artificial Intelligence to Diagnose Meniscus Tears on MRI. *Diagn. Interv. Imaging* 2019, 100, 243–249. [CrossRef] [PubMed]
- 110. Pedoia, V.; Lee, J.; Norman, B.; Link, T.M.; Majumdar, S. Diagnosing Osteoarthritis from T2 Maps Using Deep Learning: An Analysis of the Entire Osteoarthritis Initiative Baseline Cohort. *Osteoarthr. Cartil.* **2019**, *27*, 1002–1010. [CrossRef] [PubMed]
- 111. Liu, F.; Zhou, Z.; Samsonov, A.; Blankenbaker, D.; Larison, W.; Kanarek, A.; Lian, K.; Kambhampati, S.; Kijowski, R. Deep Learning Approach for Evaluating Knee MR Images: Achieving High Diagnostic Performance for Cartilage Lesion Detection. *Radiology* 2018, 289, 160–169. [CrossRef] [PubMed]
- 112. Antony, J.; McGuinness, K.; O'Connor, N.E.; Moran, K. Quantifying Radiographic Knee Osteoarthritis Severity Using Deep Convolutional Neural Networks. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1195–1200.
- 113. Lu, J.-T.; Pedemonte, S.; Bizzo, B.; Doyle, S.; Andriole, K.P.; Michalski, M.H.; Gonzalez, R.G.; Pomerantz, S.R. DeepSPINE: Automated Lumbar Vertebral Segmentation, Disc-Level Designation, and Spinal Stenosis Grading Using Deep Learning 2018. In Proceedings of the Machine Learning for Healthcare Conference 2018, Palo Alto, CA, USA, 17–18 August 2018.
- 114. Kapinski, N.; Zielinski, J.; Borucki, B.A.; Trzcinski, T.; Ciszkowska-Lyson, B.; Nowinski, K.S. Estimating Achilles Tendon Healing Progress with Convolutional Neural Networks. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, Granada, Spain, 16–20 September 2018; Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 949–957.
- 115. Trivedi, H.; Mesterhazy, J.; Laguna, B.; Vu, T.; Sohn, J.H. Automatic Determination of the Need for Intravenous Contrast in Musculoskeletal MRI Examinations Using IBM Watson's Natural Language Processing Algorithm. J. Digit. Imaging 2018, 31, 245–251. [CrossRef] [PubMed]

- 116. Hammernik, K.; Klatzer, T.; Kobler, E.; Recht, M.P.; Sodickson, D.K.; Pock, T.; Knoll, F. Learning a Variational Network for Reconstruction of Accelerated MRI Data. *Magn. Reson. Med.* **2018**, *79*, 3055–3071. [CrossRef] [PubMed]
- 117. Chaudhari, A.S.; Fang, Z.; Kogan, F.; Wood, J.; Stevens, K.J.; Gibbons, E.K.; Lee, J.H.; Gold, G.E.; Hargreaves, B.A. Super-Resolution Musculoskeletal MRI Using Deep Learning. *Magn. Reson. Med.* **2018**, *80*, 2139–2154. [CrossRef] [PubMed]
- 118. Li, J.; Zhu, G.; Hua, C.; Feng, M.; Bennamoun, B.; Li, P.; Lu, X.; Song, J.; Shen, P.; Xu, X.; et al. A Systematic Collection of Medical Image Datasets for Deep Learning. ACM Comput. Surv. 2023, 56, 116:1–116:51. [CrossRef]
- Langlotz, C.P.; Allen, B.; Erickson, B.J.; Kalpathy-Cramer, J.; Bigelow, K.; Cook, T.S.; Flanders, A.E.; Lungren, M.P.; Mendelson, D.S.; Rudie, J.D.; et al. A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology* 2019, 291, 781–791. [CrossRef] [PubMed]
- Willemink, M.J.; Koszek, W.A.; Hardell, C.; Wu, J.; Fleischmann, D.; Harvey, H.; Folio, L.R.; Summers, R.M.; Rubin, D.L.; Lungren, M.P. Preparing Medical Imaging Data for Machine Learning. *Radiology* 2020, 295, 4–15. [CrossRef] [PubMed]
- Park, S.H.; Han, K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology* 2018, 286, 800–809. [CrossRef]
- 122. Lu, Y.; Fu, J.; Li, X.; Zhou, W.; Liu, S.; Zhang, X.; Wu, W.; Jia, C.; Liu, Y.; Chen, Z. RTN: Reinforced Transformer Network for Coronary CT Angiography Vessel-Level Image Quality Assessment. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2022; Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S., Eds.; Springer: Cham, Switzerland, 2022; pp. 644–653.
- Oakden-Rayner, L.; Dunnmon, J.; Carneiro, G.; Ré, C. Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging. In Proceedings of the ACM Conference on Health, Inference, and Learning 2020, Toronto, ON, Canada, 2–4 April 2020; pp. 151–159. [CrossRef]
- 124. Liu, Z.; Kainth, K.; Zhou, A.; Deyer, T.W.; Fayad, Z.A.; Greenspan, H.; Mei, X. A Review of Self-Supervised, Generative, and Few-Shot Deep Learning Methods for Data-Limited Magnetic Resonance Imaging Segmentation. NMR Biomed. 2024, e5143. [CrossRef]
- 125. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria, 12–18 July 2020; Volume 119, pp. 1597–1607.
- 126. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollar, P.; Girshick, R. Masked Autoencoders Are Scalable Vision Learners. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 15979–15988.
- 127. Liu, Z.; Tieu, A.; Patel, N.; Zhou, A.; Soultanidis, G.; Fayad, Z.A.; Deyer, T.; Mei, X. VISION-MAE: A Foundation Model for Medical Image Segmentation and Classification 2024. *arXiv* 2024, arXiv:2402.01034.
- 128. Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment Anything in Medical Images. *Nat. Commun.* 2024, 15, 654. [CrossRef] [PubMed]
- Zhou, A.; Liu, Z.; Tieu, A.; Patel, N.; Sun, S.; Yang, A.; Choi, P.; Fauveau, V.; Soultanidis, G.; Huang, M.; et al. MRAnnotator: A Multi-Anatomy Deep Learning Model for MRI Segmentation 2024. arXiv 2024, arXiv:2402.01031.
- 130. Christensen, M.; Vukadinovic, M.; Yuan, N.; Ouyang, D. Multimodal Foundation Models for Echocardiogram Interpretation 2023. *arXiv* 2023, arXiv:2308.15670.
- 131. Zhou, Y.; Chia, M.A.; Wagner, S.K.; Ayhan, M.S.; Williamson, D.J.; Struyven, R.R.; Liu, T.; Xu, M.; Lozano, M.G.; Woodward-Court, P.; et al. A Foundation Model for Generalizable Disease Detection from Retinal Images. *Nature* 2023, 622, 156–163. [CrossRef] [PubMed]
- 132. Jha, S.; Topol, E.J. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *JAMA* **2016**, *316*, 2353–2354. [CrossRef] [PubMed]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.