

Review

Deep Learning for Orthopedic Disease Based on Medical Image Analysis: Present and Future

JiHwan Lee  and Seok Won Chung *

Department of Orthopaedic Surgery, School of Medicine, Konkuk University,
120-1 Neungdong-ro (Hwayang-dong), Seoul 143-729, Korea; metaphoricusjh@gmail.com

* Correspondence: smilecsw@gmail.com

Abstract: Since its development, deep learning has been quickly incorporated into the field of medicine and has had a profound impact. Since 2017, many studies applying deep learning-based diagnostics in the field of orthopedics have demonstrated outstanding performance. However, most published papers have focused on disease detection or classification, leaving some unsatisfactory reports in areas such as segmentation and prediction. This review introduces research published in the field of orthopedics classified according to disease from the perspective of orthopedic surgeons, and areas of future research are discussed. This paper provides orthopedic surgeons with an overall understanding of artificial intelligence-based image analysis and the information that medical data should be treated with low prejudice, providing developers and researchers with insight into the real-world context in which clinicians are embracing medical artificial intelligence.

Keywords: artificial intelligence; orthopedics; neural network; deep learning



Citation: Lee, J.; Chung, S.W. Deep Learning for Orthopedic Disease Based on Medical Image Analysis: Present and Future. *Appl. Sci.* **2022**, *12*, 681. <https://doi.org/10.3390/app12020681>

Academic Editor: Fabio La Foresta

Received: 12 November 2021

Accepted: 4 January 2022

Published: 11 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/>).

1. Introduction

A convolutional neural network (CNN) is a deep learning algorithm architecture created based on a 1962 study investigating the visual process of feline brains, and it has been applied in a wide range of areas, from autonomous vehicles to medical diagnoses [1].

A traditional CNN consists of an input layer that transmits input information, a hidden layer that modifies information (filtering) received from the input layer and amplifies the features (pooling) and an output layer that finally synthesizes and outputs the information.

According to the universal approximation theorem, it has been confirmed that various linear classifications are possible even if the neural network has a shallow hidden layer, and some pioneering studies have shown that classification and detection are improved as the layers constituting the neural network become deeper (deep neural network) [2]. Since 2012, the performance of deep learning has rapidly increased in medical image analysis with the use of deep neural networks, and this has led to a decrease in the classification error rate from approximately 25% in 2011 to 3.6% in 2015.

The CNN model was developed using a pipeline in terms of classification and detection [3], and the improved CNN shows excellent judgment, essentially giving the computer a new visual organ. A CNN has thus been expected to be used for medical diagnoses. However, a CNN does not provide any information on the basis of the decision. Therefore, even if a CNN shows an excellent diagnostic ability, it can only be discussed within a limited scope in medicine, where the basis for a judgment is important [4].

This has been pointed out as a technical limitation that reduces the effectiveness of a CNN in various fields other than medicine [5]. Researchers have dubbed this limitation “black box issues” and worked to develop “explainable artificial intelligence (XAI)” to look inside the problem [6]. The term “explainable” can be expressed as “understandability”, “comprehensibility” or “interpretability” and has the same meaning. XAI should not degrade the classification or prediction performance of the model in any way and should

improve the explainability. Various strategies and suitable CNN architectures have been proposed to implement an appropriate XAI [7]. Unfortunately, the black box nature of deep learning has not been completely resolved, but there are some notable achievements [8]. As one of these achievements, in 2016, Zhou et al. introduced a method explaining how a CNN makes a decision through class activation mapping [9], and this method is widely used in the field of medical artificial intelligence (Figure 1) [10].

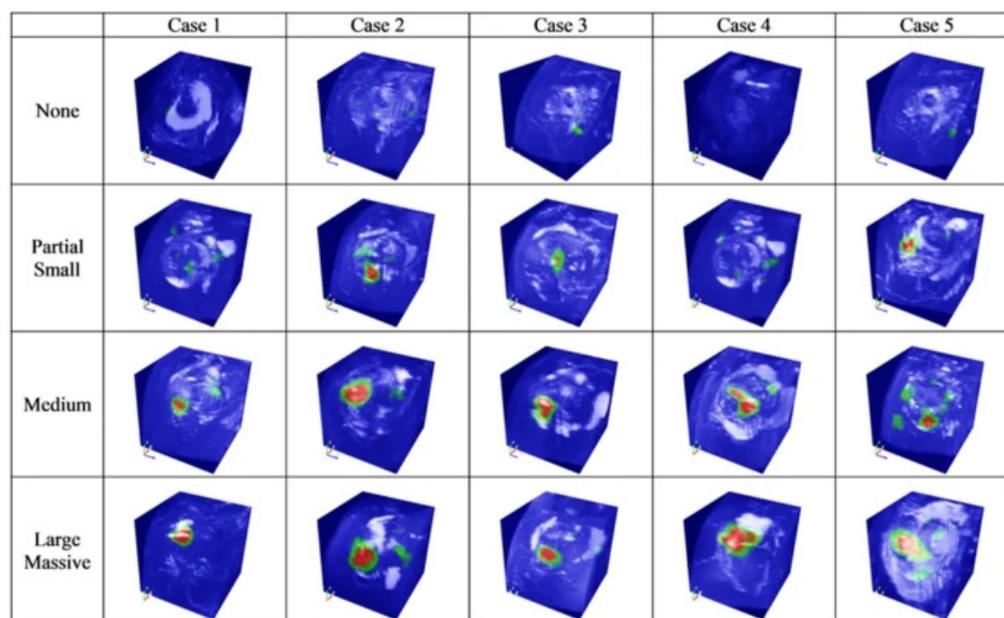


Figure 1. Image highlighting the location and size of a rotator cuff tear through a class activation map (CAM). Figure obtained from a study performed by Chung et al. [10].

In a similar context, there are attempts to improve the explainability by improving the existing CNN architecture [11]. Kim et al. modified U-Net, a CNN architecture that has strength in image segmentation, to appropriately increase the explainability. They presented an interpretable version of U-Net (SAU-Net) using an attention module for the decoder part [12].

Hence, studies introducing CNN models for diagnosing and classifying diseases using deep learning have been published in various fields of medicine, including ophthalmology and dermatology [13,14].

This trend is spreading rapidly in the field of orthopedics. Since 2017, when orthopedic disease research using deep learning was first introduced, the number of related papers has increased rapidly, and more than 300 papers in this area have been published. The search was conducted using Pubmed, MEDLINE and Embase, and papers were screened from 1 January 2017 to 2 November 2021. The search query was (orthopedic OR orthopedic) AND (deep learning). Among these studies, two orthopedic surgeons (S.W.C. and J.H.L.) independently reviewed the full text of the retrieved papers. Among these studies, 48 studies which both authors judged to be interesting and practical within the clinical context of orthopedic surgery are introduced and classified according to disease. This paper aims to provide insight into how medical artificial intelligence can help orthopedic surgeons treat patients vividly and in what context clinicians are accepting medical artificial intelligence from developers and researchers.

The authors introduce the selected papers by classifying them into the following sections: (1) Deep Learning for Fractures, (2) Deep Learning for Osteoarthritis and the Prediction of Arthroplasty Implants, (3) Deep Learning for Joint-Specific Soft Tissue Disease, (4) Miscellaneous and (5) Discussion.

2. Deep Learning for Fractures

Fractures are the most familiar ailments to orthopedists and the medical area in which deep learning methods were first applied. In 2018, Chung et al. published a CNN model for diagnosing and classifying proximal humerus fractures. Three specialists labeled 1891 anteroposterior shoulder radiographs as normal shoulders ($n = 515$) and 4 proximal humerus fracture types (greater tuberosity: 346; surgical neck: 514; 3-part: 269; and 4-part: 247) [15]. After labeling, a CNN model (ResNet-152) was trained with a training dataset created through augmentation of the labeled data. The CNN model recorded 96% accuracy for the normal shoulders and proximal humerus fractures, showing a higher accuracy than a general orthopedist (92.8% accuracy). This model showed a top-1 accuracy of 65–86% and an area under the curve (AUC) of 0.90–0.98 for classifying the fracture types. A recently published paper introduced a model with improved classification accuracy. In 2020, Demir et al. introduced a deep learning model to diagnose and classify humerus fractures using the exemplar pyramid method, a novel, stable feature extraction approach which showed a high classification accuracy of 99.12% [16].

Urakawa et al. trained the VGG-16 CNN model using hip plain radiographs (1773 intertrochanteric hip fracture images and 1573 normal hip images) and showed an accuracy of 95.5% [17]. Yamada et al. trained the CNN model (Xception architectural) based on 3123 hip plain and lateral radiography images, and the trained model classified fractures with 98% accuracy, which is better than orthopedists (92.2% accuracy) [18].

For the hip, as with the shoulder, there has been an attempt to classify fractures by training the CNN model. Lee et al. introduced a CNN model for training 786 anteroposterior pelvic plan radiographs using GoogLeNet-inception v3 [19]. The model classified a proximal femur fracture into type A (trochanteric region), type B (femur neck) and type C (femoral head) according to AO/OTA classification with an overall accuracy of 86.8%, showing a reasonable result. Lind et al. trained a ResNet-based CNN with anteroposterior and lateral knee radiographs, amounting to 6768 images [20]. The trained CNN model classified knee radiographic images according to the AO/OTA classification system and classified proximal tibia fractures, patellar fractures and distal femur fractures with AUCs of 0.87, 0.89 and 0.89, respectively.

The trained CNN diagnosed and classified fractures at a relatively high level in the large appendices of the shoulder, knee and hip. By contrast, a CNN model trained to diagnose and classify fractures in small joints or axial joints showed a relatively low AUC and accuracy. Farda et al. trained a PCANet-based CNN model that classified calcaneal fractures according to Sanders classification using computer tomography with 5534 datasets [21]. The trained CNN model showed 72% accuracy. In addition, Ozkaya et al. trained a CNN model based on ResNet50 with 390 anteroposterior wrist radiographic images [22]. The AUC of the learned CNN was 0.84, showing a relatively satisfactory result, but it was lower than that of experienced orthopedists.

Langerhuizen et al. compared the scaphoid fracture diagnostic accuracy between a deep learning algorithm and an orthopedist [23]. They trained the VGG16 CNN model with 150 radiographic images of scaphoid fractures and 150 images of normal wrist radiography without a fracture. Of the 150 images with scaphoid fractures, 23 could not be judged by the radiographic images and could only be confirmed through magnetic resonance imaging (MRI). The accuracy of the trained CNN model was 72%, which was lower than that of an orthopedic surgeon (84%). However, five of six occult scaphoid fractures were missed by all human observers.

An attempt was also made to diagnose the compression fractures in the spine using a trained CNN. The results showed a significant difference depending on the type of data used for learning. Chen et al. trained a ResNet-based CNN model using plain spine X-rays, and the trained CNN showed an accuracy of 73.59% [24]. By contrast, Yabu et al. presented a CNN model using MRI images as the training data. This model showed a higher accuracy (88%) than that of the surgeons [25].

In summary, fracture diagnosis using artificial intelligence showed a high level of accuracy. The trained CNN model conducted fracture diagnosis (binary classification) with a higher accuracy than fracture classification (multiclass classification), and this gap is expected to decrease as more advanced CNN models are developed.

In classifying fractures, small and axial joints showed a lower accuracy than large joints (Table 1). This may be a limitation of a CNN-based approach, which makes judgments by recognizing the contrast information (e.g., normal margin of the cortical bone and the fracture line or normal joint line) and spatial information of the images. The authors believe that this limitation can be overcome using more powerful CNN models.

Table 1. Summary of diagnostic performance for detecting/classifying orthopedic fracture.

Fracture Site	Image Used	Author. Year	CNN Used	Work	Dataset Size	Accuracy	AUC	Winner
Hip (femur neck)	X-ray	Matthew et al. 2019	GooLeNet	Binary classification	805	94%	0.98	
Hip	X-ray	Cheng et al. 2019	DenseNet	Binary classification	3605	91%	0.98	Orthopedist > CNN
Hip	X-ray	Takaaki et al. 2019	VGG-16	Binary classification	3346			CNN > Orthopedist
Hip	X-ray	Yamada et al. 2020	Xception, ImageNet	Binary classification	3123	98%		CNN > Orthopedist
Hip	X-ray	Lee et al. 2020	GoogLeNet-inception v3	Classification	686	86.8%		
Hip	X-ray	Tanzi et al. 2020	InceptionV3, VGG-16, ResNet50	Classification	2453	86% (3 class) 81% (5 class)		
Hip (Atypical fracture)	X-ray	Zdolsek et al. 2021	VGG19, InceptionV3, ResNet	Binary classification	982	91% (ResNet50) 83% (VGG19) 89% (InceptionV3)		
Shoulder (proximal humerus)	X-ray	Chung et al. 2018	ResNet	Binary classification	1891	95%	0.99	Orthopedist > CNN (specialized in the shoulder)
Knee	X-ray	Lind et al. 2021	ResNet-based CNN	Classification	6768		0.87 (Proximal tibia) 0.89 (Patella) 0.89 (Distal femur)	
Ankle	X-ray	Gene et al. 2019	Xception	Binary classification	596	75%		
Ankle (Malleolar)	X-ray	Olczak et al. 2021	ResNet	Classification	5495		0.90	
Ankle (Calcaneal)	CT	Farda et al. 2021	PCANet	Classification, Segmentation	5534	72%		
Wrist	X-ray	Kim et al. 2017	Inception	Binary classification	1389		0.95	
Wrist	X-ray	Thian et al. 2019	ResNet	Binary classification	7356	88.9%	0.90	

Table 1. Cont.

Fracture Site	Image Used	Author. Year	CNN Used	Work	Dataset Size	Accuracy	AUC	Winner
Wrist (Scaphoid)	X-ray	Langerhuizen et al. 2020	VGG-16	Binary classification	300	72%	0.77	Orthopedist > CNN
Wrist (Scaphoid)	X-ray	Ozkaya et al. 2020	ResNet50	Binary classification	390		0.84	Orthopedist > CNN
Vertebra	X-ray	Chen et al. 2021	ImageNet, ResNeXt	Binary classification	1306	73.6%	0.72	Orthopedist > CNN
Vertebra	MRI	Yabu et al. 2021	VGG-16,19, Inception V3, ResNet50	Binary classification	1624		0.95	CNN > Orthopedist

Most of the diagnosis and classification of fractures using deep learning have focused on osteoporotic fractures, and studies on osteoporotic fracture joints with low frequencies are relatively poor [26]. This may be because the dataset for training the CNN model is sufficient because osteoporotic fractures account for a high proportion of the total fracture frequency, and the fracture pattern is relatively standardized, making it suitable for use in fracture classification.

3. Deep Learning for Osteoarthritis and Prediction of Arthroplasty Implants

Osteoarthritis is as familiar to orthopedists as fractures. Therefore, several attempts have been made to diagnose and classify osteoarthritis using deep learning algorithms. Xue et al. trained a CNN model based on VGG-16 with 420 plain hip X-rays [27]. This is one of the earliest studies to apply deep learning methods to the orthopedic field, and the trained model diagnosed hip osteoarthritis with an accuracy of 92.8%. Ureten et al. also presented a model for diagnosing hip osteoarthritis using a similar research design, showing an accuracy of 90.2% [28].

Tiulpin et al. trained a CNN model to classify knee osteoarthritis according to the Kellgren–Lawrence grading scale using a Siamese classification CNN [29]. The model trained using plain knee X-rays showed a multiclass accuracy of 66.7%. In addition, Swiecicki et al. trained a Faster R-CNN using plain and lateral knee X-rays from the Multicenter Osteoarthritis Study dataset [30]. The multiclass accuracy of this model was 71.9%, which showed improved performance compared with the previous study conducted by Tiulpin et al.

Pedoia et al. trained a DenseNet-based CNN based on MRI-T2 images rather than X-ray data, as used in previous studies, and this model showed a high AUC of 0.83 [31]. Kim et al. trained an SE-ResNet-based CNN model using 4366 knee anteroposterior X-rays as a dataset. Furthermore, they trained the model by adding demographic information (age, sex and body mass index), alignment and metabolic data information that can affect knee osteoarthritis, in addition to image information [32]. The diagnostic performance of the image data with additional patient information showed a significantly higher AUC (Table 2).

Advanced osteoarthritis of the hip or knee often requires arthroplasty. Several studies have introduced a model for classifying arthroplasty implants used by patients with deep learning algorithms. Karnuta et al. trained the InceptionV3 network-based CNN model using anteroposterior knee X-rays with nine different implant models inserted [33]. The trained model showed an accuracy of 99% and an AUC of 0.99, classifying the implant models at an almost perfect level. A similar attempt was made at the hip joint. In addition, Borjali et al. created a CNN model trained on 252 plain hip X-rays containing 3 different implant designs, and this model classified implants with 100% accuracy (Figure 2) [34]. Kang et al. also developed a CNN model trained on 170 plain hip X-rays containing 29 different implant designs. This model also showed a high level of performance, with an AUC of 0.99 [35].

Table 2. Summary of diagnostic performance for classifying osteoarthritis.

Location	Image Used	Author. Year	CNN Used	Work	Dataset Size	Accuracy	AUC
Knee	X-ray	Tiulpin et al. 2018	Siamese CNN	Classification	5960	66.7%	
Knee	X-ray	Pedoaia et al. 2019	DenseNet	Classification	5042	75%	0.83
Knee	X-ray	Kim et al. 2020	SE-ResNet	Classification	4366	61.6% (with additional information)	0.75 (with additional information)
Knee	X-ray	Swiecicki et al. 2021	Faster R-CNN	Classification	2802	71.9%	
Hip	X-ray	Xue et al. 2017	VGG-16	Binary classification	420	92.8%	
Hip	X-ray	Ureten et al. 2020	VGG-16	Binary classification	434	90.2%	

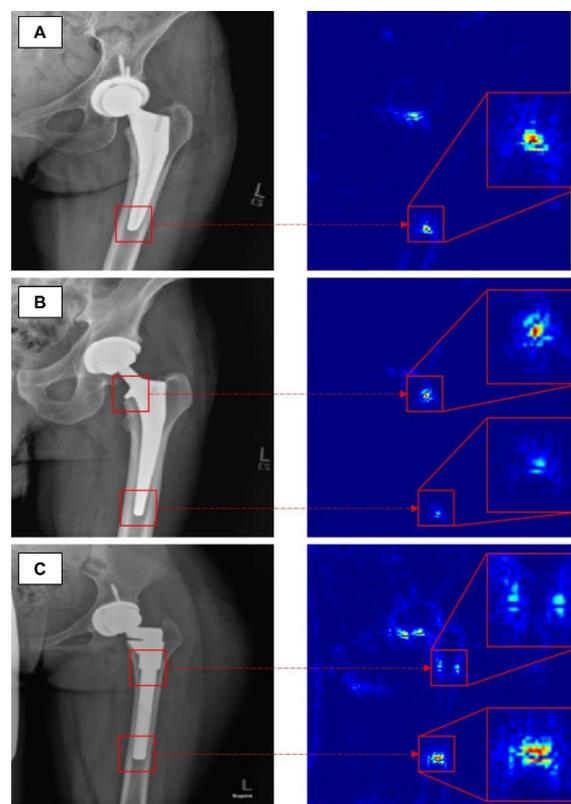


Figure 2. The figure shows how a trained convolutional neural network classifies total hip replacement implants of different designs in A, B and C. Figure obtained from a study performed by Borjali et al. [34].

By contrast, the model classifying shoulder arthroplasty implants showed a relatively low AUC. Urban et al. developed a CNN model trained on 597 plain shoulder X-rays with 16 different implant designs, showing an accuracy of 80% [36]. In addition, Sultan et al. proposed a model for classifying the different designs of four manufacturers using modified ResNet and DenseNet, showing an accuracy of 85.9% [37].

In summary, as in the case of using deep learning for fractures, binary classification of osteoarthritis has a higher accuracy than multiclass classification. In particular, the CNN-based model for specifying arthroplasty implants of the hip or knee shows a high accuracy. This may be because, unlike human bone, the implant design is highly standardized, demonstrating a clear margin on X-rays and providing clear contrast information to the CNN model. However, the classification of shoulder arthroplasty implants shows a low

level of accuracy. This may be due to the fact that a shoulder anteroposterior X-ray can show a wider range of positions than an anteroposterior radiograph of the knee or hip.

4. Deep Learning for Joint-Specific Soft Tissue Disease

As for deep learning approaches, an algorithm specialized for detection based on learned images and an algorithm for segmentation by analyzing features have structural differences and have developed into different areas of application [3]. In particular, segmentation has technical difficulties in that it is necessary to preserve spastic information that is easily lost in the outer-layer process of synthesizing the results of the CNN model being trained [38]. Recent studies have attempted to overcome these limitations through techniques such as FCN-based semantic segmentation.

These differences in deep learning algorithms also affect the use of deep learning in the orthopedic field. The deep learning-based studies introduced above are cases of diagnosing and classifying diseases based on X-ray images, and a CNN model specialized for segmentation is not always required [39]. By contrast, for diseases that are diagnosed and classified based on images such as ultrasound or MRI, a satisfactory level of accuracy can be obtained using only a CNN model specialized for segmentation. For example, a CNN model for diagnosing rotator cuff tears is more appropriate for inferring such tears based on the outline of the normal rotator cuff (segmentation) than a method of diagnosis applied by specifying the location where the tear occurred (regional detection).

Therefore, CNN models for diagnosing soft tissue disease in the orthopedic field have mainly been published after 2018, which was when the segmentation technology began to mature. Kim et al. trained a CNN model using a shoulder MRI dataset of 240 patients. The trained model identified the muscle region of the rotator cuff with an accuracy of 99.9% and graded fatty infiltration at a high level [40]. Taghizadeh et al. also conducted a similar study using a shoulder computed tomography of 103 patients as a dataset. The trained CNN model measured fatty infiltration with an accuracy of 91% [41].

Medina et al. introduced a model for segmenting the rotator cuff muscle with 98% accuracy by applying a CNN model trained using the shoulder MRIs of 258 patients [42]. Furthermore, Shim and Chung et al. introduced a model for evaluating the presence of tears and their sizes in the rotator cuff by training a Voxception-ResNet (VRN)-based CNN with 2124 shoulder MRIs. The trained CNN model diagnosed and classified rotator cuff tears with accuracies of 92.5% and 76.5%, respectively [10]. In addition, Lee et al. developed a new deep learning architecture using an integrated positive loss function and a pre-trained encoder. Using this, the location of the rotator cuff tear can be relatively accurately determined, even when imbalanced and noisy ultrasound images are provided [43].

Recent studies suggesting a CNN model for diagnosing meniscal tears, cartilage lesions and anterior cruciate ligament (ACL) ruptures in the knee joint have also been published. Couteaux et al. presented a model that trains a Mask-RCNN with 1828 T2-weighted 2D Fast Spin-Echo images to classify the torn part from the normal area of the meniscus and do so according to the location of the tear [44]. This model diagnosed and classified meniscal tears with an AUC of 0.91. Roblot et al. also proposed a model for diagnosing meniscal tears in a similar way, detecting meniscal tears with an AUC of 0.94 [45].

Chang et al. presented a model for diagnosing complete ACL tears by training a U-Net-based CNN using 320 coronal proton density-weighted 2D Fast Spin-Echo images, demonstrating an AUC of 0.97 [46]. In addition, Flannery et al. trained a modified U-Net-based CNN and evaluated the level of segmentation of the model. The segmentation level suggested by the trained model did not show a statistically significant difference from the ground truth (the value actually suggested by an expert) (Figure 3) [47].

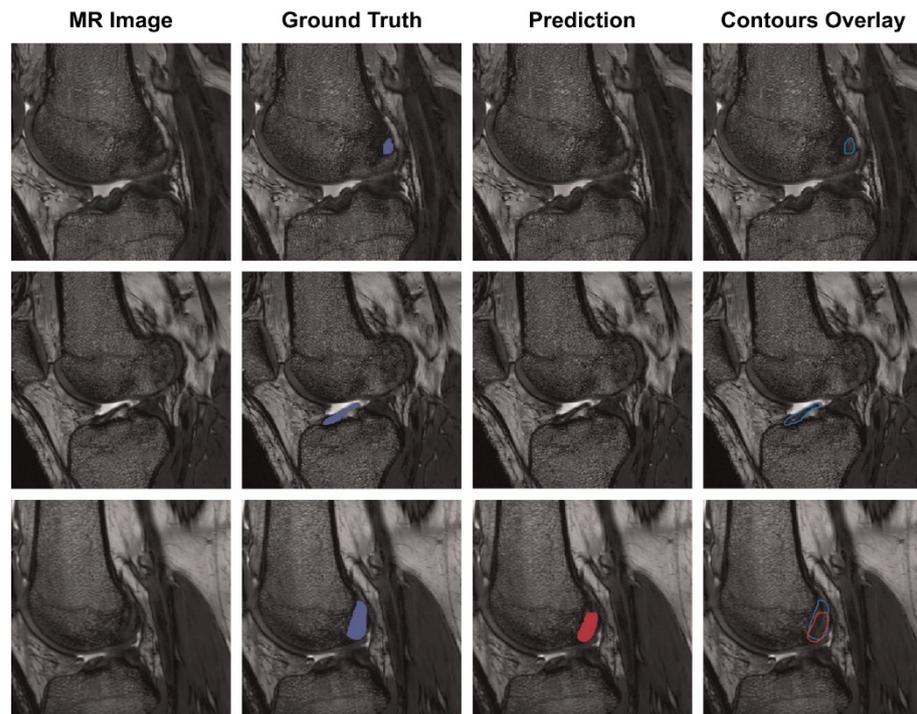


Figure 3. Each row is the same MR slice, and each column is an unsegmented slice (MR Image), an expert measured value (Ground Truth), a trained CNN model predicted value (Prediction) and an overlay of manual and predicted segmentations (Contours Overlay). Figure obtained from a study performed by Flannery et al. [47].

5. Miscellaneous

Concerning bone age, attempts to create a model that automatically predicts a bone's age through the learning of plain X-rays of carpal bones have been conducted since before the first deep learning algorithm was developed. Mahmoodi et al. presented a bone age prediction model with an accuracy of 82% in 2000, using a regression model and a Bayesian estimator [48]. A CNN model using a deep learning algorithm was developed, and it is now possible to predict the bone age with improved accuracy. In addition, Han et al. proposed a model with 97.6% accuracy by training the Inception ResNet v2 model with 5876 hand radiographs [39].

For pediatrics, developmental dysplasia of the hip is one of the most common hip joint disorders in infants and young children, and its diagnosis is difficult owing to the extensive variations in pediatric pelvic anatomy [49]. To create a deep learning algorithm that can diagnose developmental dysplasia of the hip, Zhang et al. trained a CNN model (based on ResNet-101) using 10,219 pelvic anteroposterior radiographs of children. The trained model showed a high AUC of 0.975 [50].

An acute pediatric elbow fracture is also difficult to diagnose, owing to the existence of multiple cartilaginous ossification centers and a highly variable appearance [51]. England et al. trained a CNN using 901 lateral elbow radiographs, and the trained model diagnoses elbow fractures with a high AUC of 0.985 [52].

Central dual-energy X-ray absorptiometry is the reference standard for diagnosing osteoporosis and osteopenia. A CNN model for diagnosing osteopenia and osteoporosis using plain radiography without dual-energy X-ray absorptiometry was recently introduced.

Zhang et al. trained a CNN model with 2564 lumbar X-ray images, and this model showed an AUC of 0.767 and 0.810 for osteoporosis and osteopenia, respectively [53]. Yamamoto et al. trained a CNN with 1131 hip X-rays, and this model diagnosed osteoporosis with an accuracy of 0.885 [54].

For alignment, Pei et al. published an interesting study using a deep learning algorithm to automatically measure the hip-knee-ankle angle. They trained a CNN model with

796 unilateral lower limb X-rays, showing a difference of 0.49° from the ground truth measured directly by orthopedic surgeons [55]. In addition, Rouzrokh and Pouria et al. trained a CNN model with 600 hip anteroposterior and 600 hip lateral X-rays taken after total hip arthroplasty and programmed this model to automatically derive the acetabular component inclination and version. Compared with the ground truth, this model showed a difference of 1.35° for the inclination and 1.39° for the anteversion [56].

Galbusera et al. presented a CNN model trained using biplanar radiographs of the spine. The model automatically calculated the T4-T12 kyphosis, L1-L5 lordosis, Cobb angle of scoliosis, pelvic incidence, sacral slope and pelvic tilt. Among them, the pelvic tilt showed a difference of 2.7° compared with the ground truth, whereas the L1-L5 lordosis showed a difference of 11.5° from the ground truth [56].

Concerning metastasis and infections in the spine, the spine is a joint that receives a high blood supply and is relatively easily exposed to metastasis compared with other joints [57]. Therefore, studies for diagnosing metastatic lesions using deep learning algorithms have mainly focused on the spine. Wang et al. reported that a CNN model trained with sagittal fat-suppressed T2 2D Fast Spin-Echo spine images localized metastatic lesions with a sensitivity of 90% [58]. In addition, Chmelik et al. trained a CNN with sagittal computed tomography images containing 1046 lytic lesions and 1135 sclerotic lesions, and the trained model detected lytic and sclerotic lesions with AUCs of 0.80 and 0.78, respectively [59].

Kim et al. published a CNN model to discriminate between tuberculous and pyogenic spondylitis. They trained the CNN using axial T2-weighted 2D Fast Spin-Echo images, and the trained CNN model divided the two conditions with an AUC of 0.80, with no significant difference from a human reader [60].

As for other applications, in addition to the previously introduced papers, studies using deep learning algorithms in the field of orthopedic surgery have been published. Won et al. introduced a model for grading spinal stenosis by training a Faster R-CNN [61]. Rouzrokh and Pouria et al. attempted to predict postoperative hip dislocation by training a CNN model with 92,584 hip X-rays taken after total hip arthroplasty. The trained model showed an AUC of 76.7% and an accuracy of 49.5% [62].

6. Discussion

Orthopedics, along with dermatology, ophthalmology and cardiology, is the medical field in which research into deep learning algorithms is most actively conducted. Related research has been explosively increasing since 2017, and this trend is expected to continue until the “new winter”, when the development of artificial intelligence will reach its limit.

To date, image analysis studies of orthopedic diseases using deep learning have shown excellent results overall. Several studies have reported that in fractures and osteoarthritis, a trained CNN model has a diagnostic accuracy comparable to that of an expert. The studies also presented satisfactory results for the classification of fractures and osteoarthritis. However, the accuracy of multiclass classification did not reach detection, and studies on small joints presented relatively poor results compared with studies on large joints.

Nevertheless, it is expected that this limitation can be overcome for two reasons. First, the CNN model for medical image analysis aims for accurate diagnosis and appropriate classification, and the types of classes required for this purpose are relatively small. When there are few class types, Basha et al. proved that the accuracy can be improved using a CNN model structured as a deeper layer [63]. Therefore, it is expected that the development of a CNN model with deep hyperparameters will increase the accuracy of multiclass classification through medical image analysis. Second, medical images are extremely refined data compared with images used to learn road traffic conditions or climate predictions; that is, researchers can relatively easily obtain appropriate image data without noise, such as different heights of traffic lights or flying birds. This means that even with simple data augmentation such as an affine transformation, an appropriate dataset for training the CNN model can be provided.

Therefore, the authors expect that the development of a CNN model and the accumulation of additional medical images will increase the classification accuracy of fractures and osteoarthritis, which are relatively weak compared with the accuracy of diagnosis. In the same context, it is also expected that the diagnosis and classification of joint-specific soft tissue will be improved, owing to the development of deep learning algorithms advantageous for segmentation. Indeed, there are several recent studies that have completed segmentation at a high level [64,65]. In particular, Hashimoto et al. and others segmented the psoas major muscle through a U-net-based CNN model, and the trained U-net-based CNN model showed an average of 86.6% intersection over union (IoU). U-net is one of the most important semantic segmentation frameworks of CNNs [66] and has the strength of having an architecture that can recognize structural edges. Therefore, U-net is expected to be widely used for segmentation of medical images [67]. Although not in the field of orthopedics, new CNN architectures based on U-Net are continuously being introduced and reporting notable results [68]. Rundo et al. performed prostate zonal segmentation with USE-Net, incorporating Squeeze-and-Excitation blocks (SE) into U-Net [69]. Yeung et al. showed that the model trained with a dual attention-gated CNN (Focus U-Net), which improved the U-Net, segmented the polyp of the colonoscopy image to a satisfactory level [70].

Studies published in the field of orthopedic surgery have thus far been unable to present a CNN model with a higher level of diagnosis and classification than experts. An in-depth discussion is needed as to whether these results are a problem that can be overcome through data accumulation or the development of a better CNN, or whether they are a natural limitation of a CNN model learned from image data.

The authors offer two approaches. First, experts do not solve problems with image data alone. Experts can utilize information other than images, such as the patient's demographic data, the degree of pain, the nature of the disease and a physical examination, which can affect the disease diagnosis and classification. Indeed, Kim et al. reported that a CNN model trained by adding demographic information (age, sex and body mass index), alignment and metabolic data that could affect knee osteoarthritis showed a statistically significantly higher AUC [32]. Therefore, even if an improved CNN model is developed and high-quality image data are accumulated, there is a possibility that the image analysis-based CNN model using a deep learning algorithm will not reach the level of experts.

Second, despite the opinions presented above, the possibility that CNN models will outperform experts in certain fields cannot be excluded, because the CNN model analyzes images from a different point of view than human beings. Among 150 images of scaphoid fractures, Langerhuizen et al. included 23 scaphoid fracture image data that could only be confirmed through an MRI. The trained CNN model showed a lower level of accuracy than orthopedic surgeons, but it detected five of six occult scaphoid fractures that were missed by all human observers [23]. It is therefore necessary to carefully discuss whether an image analysis model using deep learning can outperform experts.

It is clear that the present CNN models have room for improvement. However, this does not undermine the significance of the studies conducted to date. The currently developed CNN model can reduce the task intensity of the expert reader and can be used for the education of non-expert medical workers, such as medical students or specialists during training [71]. In addition, through a developed CNN model, a pediatrician can roughly estimate a patient's bone age using only X-rays without the help of an orthopedic surgeon.

A step away from the fate of clinical doctors and CNN's accuracy battle, there are interesting and more practical studies that give practical help to patients and doctors. Nie et al. converted native medical CT images to higher resolution images through generative adversarial networks (GANs) [72], and this study has the potential to be extended to MRI images [73]. Therefore, it can help a society that has no choice but to use low-quality MRI due to insufficient medical infrastructure or patients who have difficulty using high-quality MRI due to cost problems.

The authors reviewed deep learning approaches for orthopedic diseases applied through image analysis and found some limitations. First, there are no models approved by the Food and Drug Administration, other than a CNN model for predicting the bone age in children and a model for diagnosing wrist fractures [74]. In other medical departments, several models have been approved by the Food and Drug Administration, starting with a deep learning-based model for the automatic diagnosis of diabetic retinopathy in April 2018.

Second, no prospective studies have been conducted [75]. To improve the quality of research and continue applicable studies, a prospective and randomized trial according to the CONSORT-AI guidelines presented in 2020 will be necessary [76].

Third, recently described deep learning methods have mostly been designed to conduct a single task. To be useful in clinical practice, multiple deep learning algorithms will need to evaluate every possible abnormality. Some efforts have been made to overcome these limitations. For example, Grauhan et al. presented a CNN model for diagnosing fractures, joint dislocation and osteoarthritis through plain shoulder radiographs [77].

Finally, there is a need to reduce expert bias on a given dataset. Orthopedic surgeons have traditionally used ultrasound, computed tomography or MRIs to diagnose soft tissue diseases. However, deep learning algorithms often make appropriate judgments beyond human cognition. Kang et al. presented a model for diagnosing SSC tendon tears with a CNN model trained using axillary lateral radiographs, and the learned model showed an appropriate level of accuracy [78]. Thus, orthopedic surgeons may have the freedom to develop CNN models based on their imagination, free from prejudice.

In conclusion, image analysis using deep learning presents a clear milestone in the field of orthopedics and is experiencing explosive growth. The development of a CNN architecture and the accumulation of refined image data are expected to lead to the development of more sophisticated models. However, it is difficult to predict whether a deep learning model that exceeds the capability of experts can be created. Orthopedic surgeons who want to apply a deep learning algorithm to image analysis need to treat data with low prejudice, present research that meets the newly suggested guidelines and focus on developing models that can multitask.

Author Contributions: Conception and design of study, interpretation of data and approval of the version of the manuscript to be published, S.W.C.; interpretation of data, drafting the manuscript and approval of the version of the manuscript to be published, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This results was supported by “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE). (2021RIS-001(1345341783)).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank the authors for allowing the use of figures in this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J. Physiol.* **1962**, *160*, 106–154. [[CrossRef](#)] [[PubMed](#)]
2. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [[CrossRef](#)]
3. Zhao, Z.-Q.; Zheng, P.; Xu, S.-T.; Wu, X. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [[CrossRef](#)] [[PubMed](#)]
4. Wang, F.; Casalino, L.P.; Khullar, D. Deep learning in medicine—Promise, progress, and challenges. *JAMA Intern. Med.* **2019**, *179*, 293–294. [[CrossRef](#)] [[PubMed](#)]
5. Budd, S.; Robinson, E.C.; Kainz, B. A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Med. Image Anal.* **2021**, *71*, 102062. [[CrossRef](#)]

6. Fujita, H. AI-based computer-aided diagnosis (AI-CAD): The latest review to read first. *Radiol. Phys. Technol.* **2020**, *13*, 6–19. [[CrossRef](#)] [[PubMed](#)]
7. Castiglioni, I.; Rundo, L.; Codari, M.; Leo, G.D.; Salvatore, C.; Interlenghi, M.; Gallivanone, F.; Cozzi, A.; D'Amico, N.C.; Sardanelli, F. AI applications to medical images: From machine learning to deep learning. *Phys. Med.* **2021**, *83*, 9–24. [[CrossRef](#)]
8. Tjoa, E.; Guan, C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4793–4813. [[CrossRef](#)] [[PubMed](#)]
9. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 30 June 2016; pp. 2921–2929.
10. Shim, E.; Kim, J.Y.; Yoon, J.P.; Ki, S.-Y.; Lho, T.; Kim, Y.; Chung, S.W. Automated rotator cuff tear classification using 3D convolutional neural network. *Sci. Rep.* **2020**, *10*, 1–9. [[CrossRef](#)]
11. Singh, A.; Sengupta, S.; Lakshminarayanan, V. Explainable Deep Learning Models in Medical Image Analysis. *J. Imaging* **2020**, *6*, 52. [[CrossRef](#)]
12. Kim, B.; Wattenberg, M.; Gilmer, J.; Cai, C.; Wexler, J.; Viegas, F. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In Proceedings of the 35th International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; pp. 2668–2677.
13. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **2016**, *316*, 2402–2410. [[CrossRef](#)]
14. Bejnordi, B.E.; Veta, M.; Van Diest, P.J.; Van Ginneken, B.; Karssemeijer, N.; Litjens, G.; van der Laak, J.A.W.M.; The CAMELYON16 Consortium. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* **2017**, *318*, 2199–2210. [[CrossRef](#)]
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](#)]
16. Demir, S.; Key, S.; Tuncer, T.; Dogan, S. An exemplar pyramid feature extraction based humerus fracture classification method. *Med. Hypotheses* **2020**, *140*, 109663. [[CrossRef](#)]
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](#)]
18. Yamada, Y.; Maki, S.; Kishida, S.; Nagai, H.; Arima, J.; Yamakawa, N.; Iijima, Y.; Shiko, Y.; Kawasaki, Y.; Kotani, T.; et al. Automated classification of hip fractures using deep convolutional neural networks with orthopedic surgeon-level accuracy: Ensemble decision-making with antero-posterior and lateral radiographs. *Acta Orthop.* **2020**, *91*, 699–704. [[CrossRef](#)]
19. Lee, C.; Jang, J.; Lee, S.; Kim, Y.S.; Jo, H.J.; Kim, Y. Classification of femur fracture in pelvic X-ray images using meta-learned deep neural network. *Sci. Rep.* **2020**, *10*, 1–12. [[CrossRef](#)] [[PubMed](#)]
20. Lind, A.; Akbarian, E.; Olsson, S.; Näsell, H.; Sköldenberg, O.; Razavian, A.S.; Gordon, M. Artificial intelligence for the classification of fractures around the knee in adults according to the 2018 AO/OTA classification system. *PLoS ONE* **2021**, *16*, e0248809. [[CrossRef](#)] [[PubMed](#)]
21. Farda, N.A.; Lai, J.-Y.; Wang, J.-C.; Lee, P.-Y.; Liu, J.-W.; Hsieh, I.-H. Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques. *Injury* **2020**, *52*, 616–624. [[CrossRef](#)] [[PubMed](#)]
22. Ozkaya, E.; Topal, F.E.; Bulut, T.; Gursoy, M.; Ozuysal, M.; Karakaya, Z. Evaluation of an artificial intelligence system for diagnosing scaphoid fracture on direct radiography. *Eur. J. Trauma Emerg. Surg.* **2020**, 1–8. [[CrossRef](#)]
23. Langerhuizen, D.W.G.; Bulstra, A.E.J.; Janssen, S.J.; Ring, D.; Kerckhoffs, 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](#)]
24. Chen, H.-Y.; Hsu, B.W.-Y.; Yin, Y.-K.; Lin, F.-H.; Yang, T.-H.; Yang, R.-S.; Lee, C.-K.; Tseng, V.S. Application of deep learning algorithm to detect and visualize vertebral fractures on plain frontal radiographs. *PLoS ONE* **2021**, *16*, e0245992. [[CrossRef](#)] [[PubMed](#)]
25. Yabu, A.; Hoshino, M.; Tabuchi, H.; Takahashi, S.; Masumoto, H.; Akada, M.; Morita, S.; Maeno, T.; Iwamae, M.; Inose, H.; et al. Using artificial intelligence to diagnose fresh osteoporotic vertebral fractures on magnetic resonance images. *Spine J.* **2021**, *21*, 1652–1658. [[CrossRef](#)] [[PubMed](#)]
26. Moon, Y.L.; Jung, S.H.; Choi, G.Y. Evaluation of focal bone mineral density using three-dimensional of Hounsfield units in the proximal humerus. *CiSE.* **2015**, *18*, 86–90.
27. Xue, Y.; Zhang, R.; Deng, Y.; Chen, K.; Jiang, T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLoS ONE* **2017**, *12*, e0178992. [[CrossRef](#)]
28. Üreten, K.; Arslan, T.; Gültekin, K.E.; Demir, A.N.D.; Özer, H.F.; Bilgili, Y. Detection of hip osteoarthritis by using plain pelvic radiographs with deep learning methods. *Skelet. Radiol.* **2020**, *49*, 1369–1374. [[CrossRef](#)]
29. Tiulpin, A.; Thevenot, J.; Rahtu, E.; Lehenkari, P.; Saarakkala, S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. *Sci. Rep.* **2018**, *8*, 1–10. [[CrossRef](#)]

30. Swiecicki, A.; Li, N.; O'Donnell, J.; Said, N.; Yang, J.; Mather, R.C.; Jiranek, D.A.; Mazurowski, M.A. Deep learning-based algorithm for assessment of knee osteoarthritis severity in radiographs matches performance of radiologists. *Comput. Biol. Med.* **2021**, *133*, 104334. [[CrossRef](#)] [[PubMed](#)]
31. 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](#)]
32. Kim, D.H.; Lee, K.J.; Choi, D.; Lee, J.I.; Choi, H.G.; Lee, Y.S. Can Additional Patient Information Improve the Diagnostic Performance of Deep Learning for the Interpretation of Knee Osteoarthritis Severity. *J. Clin. Med.* **2020**, *9*, 3341. [[CrossRef](#)]
33. Karnuta, J.M.; Luu, B.C.; Roth, A.L.; Haeberle, H.S.; Chen, A.F.; Iorio, R.; Schaffer, J.L.; Mont, M.A.; Patterson, B.M.; Krebs, V.E.; et al. Artificial Intelligence to Identify Arthroplasty Implants From Radiographs of the Knee. *J. Arthroplast.* **2021**, *36*, 935–940. [[CrossRef](#)] [[PubMed](#)]
34. Borjali, A.; Chen, A.F.; Muratoglu, O.K.; Morid, M.A.; Varadarajan, K.M. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. *J. Orthop. Res.* **2020**, *38*, 1465–1471. [[CrossRef](#)] [[PubMed](#)]
35. Kang, Y.-J.; Yoo, J.-I.; Cha, Y.-H.; Park, C.H.; Kim, J.-T. Machine learning-based identification of hip arthroplasty designs. *J. Orthop. Transl.* **2020**, *21*, 13–17. [[CrossRef](#)] [[PubMed](#)]
36. Urban, G.; Porhemmat, S.; Stark, M.; Feeley, B.; Okada, K.; Baldi, P. Classifying shoulder implants in X-ray images using deep learning. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 967–972. [[CrossRef](#)] [[PubMed](#)]
37. Sultan, H.; Owais, M.; Park, C.; Mahmood, T.; Haider, A.; Park, K.R. Artificial Intelligence-Based Recognition of Different Types of Shoulder Implants in X-ray Scans Based on Dense Residual Ensemble-Network for Personalized Medicine. *J. Pers. Med.* **2021**, *11*, 482. [[CrossRef](#)]
38. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. A review of semantic segmentation using deep neural networks. *Int. J. Multimed. Inf. Retr.* **2018**, *7*, 87–93. [[CrossRef](#)]
39. Han, Y.; Wang, G. Skeletal bone age prediction based on a deep residual network with spatial transformer. *Comput. Methods Programs Biomed.* **2020**, *197*, 105754. [[CrossRef](#)]
40. Kim, J.Y.; Ro, K.; You, S.; Nam, B.R.; Yook, S.; Park, H.S.; Yoo, J.C.; Park, E.; Cho, K.; Cho, B.H.; et al. Development of an automatic muscle atrophy measuring algorithm to calculate the ratio of supraspinatus in supraspinous fossa using deep learning. *Comput. Methods Programs Biomed.* **2019**, *182*, 105063. [[CrossRef](#)] [[PubMed](#)]
41. Taghizadeh, E.; Truffer, O.; Becce, F.; Eminian, S.; Gidoin, S.; Terrier, A.; Farron, A.; Büchler, P. Deep learning for the rapid automatic quantification and characterization of rotator cuff muscle degeneration from shoulder CT datasets. *Eur. Radiol.* **2021**, *31*, 181–190. [[CrossRef](#)]
42. Medina, G.; Buckless, C.G.; Thomasson, E.; Oh, L.S.; Torriani, M. Deep learning method for segmentation of rotator cuff muscles on MR images. *Skelet. Radiol.* **2021**, *50*, 683–692. [[CrossRef](#)]
43. Lee, K.; Kim, J.Y.; Lee, M.H.; Choi, C.-H.; Hwang, J.Y. Imbalanced Loss-Integrated Deep-Learning-Based Ultrasound Image Analysis for Diagnosis of Rotator-Cuff Tear. *Sensors* **2021**, *21*, 2214. [[CrossRef](#)]
44. Couteaux, V.; Si-Mohamed, S.; Nempont, O.; Lefevre, T.; Popoff, A.; Pizaine, G.; Villain, N.; Bloch, I.; Cotten, A.; Bousset, L. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. *Diagn. Interv. Imaging* **2019**, *100*, 235–242. [[CrossRef](#)]
45. Roblot, V.; Giret, Y.; Antoun, M.B.; 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](#)]
46. Chang, P.D.; Wong, T.T.; Rasiej, M.J. Deep Learning for Detection of Complete Anterior Cruciate Ligament Tear. *J. Digit. Imaging* **2019**, *32*, 980–986. [[CrossRef](#)] [[PubMed](#)]
47. Flannery, S.W.; Kiapour, A.M.; Edgar, D.J.; Murray, M.M.; Fleming, B.C. Automated magnetic resonance image segmentation of the anterior cruciate ligament. *J. Orthop. Res.* **2021**, *39*, 831–840. [[CrossRef](#)]
48. Mahmoodi, S.; Sharif, B.S.; Chester, E.G.; Owen, J.P.; Lee, R. Skeletal growth estimation using radiographic image processing and analysis. *IEEE Trans. Inf. Technol. Biomed.* **2000**, *4*, 292–297. [[CrossRef](#)]
49. Kyung, B.S.; Lee, S.H.; Jeong, W.K.; Park, S.Y. Disparity between clinical and ultrasound examinations in neonatal hip screening. *CiOS* **2016**, *8*, 203–209. [[CrossRef](#)]
50. Zhang, S.-C.; Sun, J.; Liu, C.-B.; Fang, J.-H.; Xie, H.-T.; Ning, B. Clinical application of artificial intelligence-assisted diagnosis using anteroposterior pelvic radiographs in children with developmental dysplasia of the hip. *Bone Jt. J.* **2020**, *102*, 1574–1581. [[CrossRef](#)]
51. Rhyou, I.H.; Lee, J.H.; Park, K.J.; Kang, H.S.; Kim, K.W. The ulnar collateral ligament is always torn in the posterolateral elbow dislocation: A suggestion on the new mechanism of dislocation using MRI findings. *CiSE* **2011**, *14*, 193–198. [[CrossRef](#)]
52. England, J.R.; Gross, J.S.; White, E.A.; Patel, D.B.; England, J.T.; Cheng, P.M. Detection of Traumatic Pediatric Elbow Joint Effusion Using a Deep Convolutional Neural Network. *Am. J. Roentgenol.* **2018**, *211*, 1361–1368. [[CrossRef](#)]
53. Zhang, B.; Yu, K.; Ning, Z.; Wang, K.; Dong, Y.; Liu, X.; Liu, S.; Wang, J.; Zhu, C.; Yu, Q.; et al. Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study. *Bone* **2020**, *140*, 115561. [[CrossRef](#)] [[PubMed](#)]
54. Yamamoto, N.; Sukegawa, S.; Kitamura, A.; Goto, R.; Noda, T.; Nakano, K.; Takabatake, K.; Kawai, H.; Nagatsuka, H.; Kawasaki, K.; et al. Deep Learning for Osteoporosis Classification Using Hip Radiographs and Patient Clinical Covariates. *Biomolecules* **2020**, *10*, 1534. [[CrossRef](#)] [[PubMed](#)]

55. Pei, Y.; Yang, W.; Wei, S.; Cai, R.; Li, J.; Guo, S.; Li, Q.; Wang, J.; Li, X. Automated measurement of hip–knee–ankle angle on the unilateral lower limb X-rays using deep learning. *Phys. Eng. Sci. Med.* **2021**, *44*, 53–62. [[CrossRef](#)]
56. Rouzrokh, P.; Wyles, C.C.; Philbrick, K.A.; Ramazanian, T.; Weston, A.D.; Cai, J.C.; Taunton, M.J.; Lewallen, D.G.; Berry, D.J.; Erickson, B.J.; et al. A Deep Learning Tool for Automated Radiographic Measurement of Acetabular Component Inclination and Version After Total Hip Arthroplasty. *J. Arthroplast.* **2021**, *36*, 2510–2517.e6. [[CrossRef](#)] [[PubMed](#)]
57. Lee, B.J.; Kim, S.T.; Yoon, M.G.; Kim, S.S.; Moon, M.S. Chronic osteomyelitis of the lumbar transverse process. *CiOS* **2011**, *3*, 254–257. [[CrossRef](#)] [[PubMed](#)]
58. Wang, J.; Fang, Z.; Lang, N.; Yuan, H.; Su, M.-Y.; Baldi, P. A multi-resolution approach for spinal metastasis detection using deep Siamese neural networks. *Comput. Biol. Med.* **2017**, *84*, 137–146. [[CrossRef](#)]
59. Chmelik, J.; Jakubicek, R.; Walek, P.; Jan, J.; Ourednicek, P.; Lambert, L.; Amadori, E.; Gavelli, G. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. *Med. Image Anal.* **2018**, *49*, 76–88. [[CrossRef](#)]
60. Kim, K.; Kim, S.; Lee, Y.H.; Lee, S.H.; Lee, H.S.; Kim, S. Performance of the deep convolutional neural network based magnetic resonance image scoring algorithm for differentiating between tuberculous and pyogenic spondylitis. *Sci. Rep.* **2018**, *8*, 13124. [[CrossRef](#)]
61. Won, D.; Lee, H.-J.; Lee, S.-J.; Park, S.H. Spinal Stenosis Grading in Magnetic Resonance Imaging Using Deep Convolutional Neural Networks. *Spine* **2020**, *45*, 804–812. [[CrossRef](#)] [[PubMed](#)]
62. Rouzrokh, P.; Ramazanian, T.; Wyles, C.C.; Philbrick, K.A.; Cai, J.C.; Taunton, M.J.; Kremers, H.M.; Lewallen, D.G.; Erickson, B.J. Deep Learning Artificial Intelligence Model for Assessment of Hip Dislocation Risk Following Primary Total Hip Arthroplasty From Postoperative Radiographs. *J. Arthroplast.* **2021**, *36*, 2197–2203.e3. [[CrossRef](#)]
63. Huang, W.; Zhou, F. DA-CapsNet: Dual attention mechanism capsule network. *Sci. Rep.* **2020**, *10*, 11383. [[CrossRef](#)] [[PubMed](#)]
64. Hashimoto, F.; Kakimoto, A.; Ota, N.; Ito, S.; Nishizawa, S. Automated segmentation of 2D low-dose CT images of the psoas-major muscle using deep convolutional neural networks. *Radiol. Phys. Technol.* **2019**, *12*, 210–215. [[CrossRef](#)] [[PubMed](#)]
65. Kamiya, N.; Li, J.; Kume, M.; Fujita, H.; Shen, D.; Zheng, G. Fully automatic segmentation of paraspinal muscles from 3D torso CT images via multi-scale iterative random forest classifications. *Int. J. Comput. Assist. Radiol. Surg.* **2018**, *13*, 1697–1706. [[CrossRef](#)] [[PubMed](#)]
66. Du, G.; Cao, X.; Liang, J.; Chen, X.; Zhan, Y. Medical Image Segmentation based on U-Net: A Review. *J. Imaging Sci. Technol.* **2020**, *64*, 20508. [[CrossRef](#)]
67. Hiasa, Y.; Otake, Y.; Takao, M.; Ogawa, T.; Sugano, N.; Sato, Y. Automated Muscle Segmentation from Clinical CT Using Bayesian U-Net for Personalized Musculoskeletal Modeling. *IEEE Trans. Med. Imaging* **2020**, *39*, 1030–1040. [[CrossRef](#)]
68. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. [[CrossRef](#)] [[PubMed](#)]
69. Rundo, L.; Han, C.; Nagano, Y.; Zhang, J.; Hataya, R.; Militello, C.; Tangherloni, A.; Nobile, M.S.; Ferretti, C.; Besozzi, D.; et al. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing* **2019**, *365*, 31–43. [[CrossRef](#)]
70. Yeung, M.; Sala, E.; Schönlieb, C.B.; Rundo, L. Focus U-Net: A novel dual attention-gated CNN for polyp segmentation during colonoscopy. *Comput. Biol. Med.* **2021**, *137*, 104815. [[CrossRef](#)] [[PubMed](#)]
71. Kolachalama, V.B.; Garg, P.S. Machine learning and medical education. *NPJ Digit. Med.* **2018**, *1*, 54. [[CrossRef](#)]
72. Nie, D.; Trullo, R.; Lian, J.; Petitjean, C.; Ruan, S.; Wang, Q.; Shen, D. Medical image synthesis with context-aware generative adversarial networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 11–13 September 2017; Springer: Cham, Switzerland; pp. 417–425.
73. Ker, J.; Wang, L.; Rao, J.; Lim, C.T. Deep Learning Applications in Medical Image Analysis. *IEEE Access* **2017**, *6*, 9375–9389. [[CrossRef](#)]
74. Benjamins, S.; Dhunoo, P.; Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database. *NPJ Digit. Med.* **2020**, *3*, 118. [[CrossRef](#)] [[PubMed](#)]
75. Topol, E.J. Welcoming new guidelines for AI clinical research. *Nat. Med.* **2020**, *26*, 1318–1320. [[CrossRef](#)]
76. Liu, X.; Rivera, S.C.; Moher, D.; Calvert, M.J.; Denniston, A.K. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI Extension. *BMJ* **2020**, *370*, m3164. [[CrossRef](#)] [[PubMed](#)]
77. Grauhan, N.F.; Niehues, S.M.; Gaudin, R.A.; Keller, S.; Vahldiek, J.L.; Adams, L.C.; Bressemer, K.K. Deep learning for accurately recognizing common causes of shoulder pain on radiographs. *Skelet. Radiol.* **2021**, *51*, 355–362. [[CrossRef](#)]
78. Kang, Y.; Choi, D.; Lee, K.J.; Oh, J.H.; Kim, B.R.; Ahn, J.M. Evaluating subscapularis tendon tears on axillary lateral radiographs using deep learning. *Eur. Radiol.* **2021**, *31*, 9408–9417. [[CrossRef](#)] [[PubMed](#)]