


Review

# Artificial Intelligence in Orthopedic Radiography Analysis: A Narrative Review

Kenneth Chen <sup>1,2</sup>, Christoph Stotter <sup>1,2</sup>, Thomas Klestil <sup>1,2</sup> and Stefan Nehrer <sup>2,\*</sup> <sup>1</sup> Department for Orthopedics and Traumatology, Landeskrankenhaus Baden-Mödling, 2340 Mödling, Austria<sup>2</sup> Department for Health Sciences, Medicine and Research, Danube University Krems, 3500 Krems, Austria

\* Correspondence: stefan.nehrer@donau-uni.ac.at; Tel.: +43-2732-893-2600

**Abstract:** Artificial intelligence (AI) in medicine is a rapidly growing field. In orthopedics, the clinical implementations of AI have not yet reached their full potential. Deep learning algorithms have shown promising results in computed radiographs for fracture detection, classification of OA, bone age, as well as automated measurements of the lower extremities. Studies investigating the performance of AI compared to trained human readers often show equal or better results, although human validation is indispensable at the current standards. The objective of this narrative review is to give an overview of AI in medicine and summarize the current applications of AI in orthopedic radiography imaging. Due to the different AI software and study design, it is difficult to find a clear structure in this field. To produce more homogeneous studies, open-source access to AI software codes and a consensus on study design should be aimed for.

**Keywords:** musculoskeletal imaging; artificial intelligence; machine learning; deep learning; radiograph; X-ray



**Citation:** Chen, K.; Stotter, C.; Klestil, T.; Nehrer, S. Artificial Intelligence in Orthopedic Radiography Analysis: A Narrative Review. *Diagnostics* **2022**, *12*, 2235. <https://doi.org/10.3390/diagnostics12092235>

Academic Editors: Michail Klontzas and Apostolos Karantanas

Received: 9 August 2022

Accepted: 13 September 2022

Published: 16 September 2022

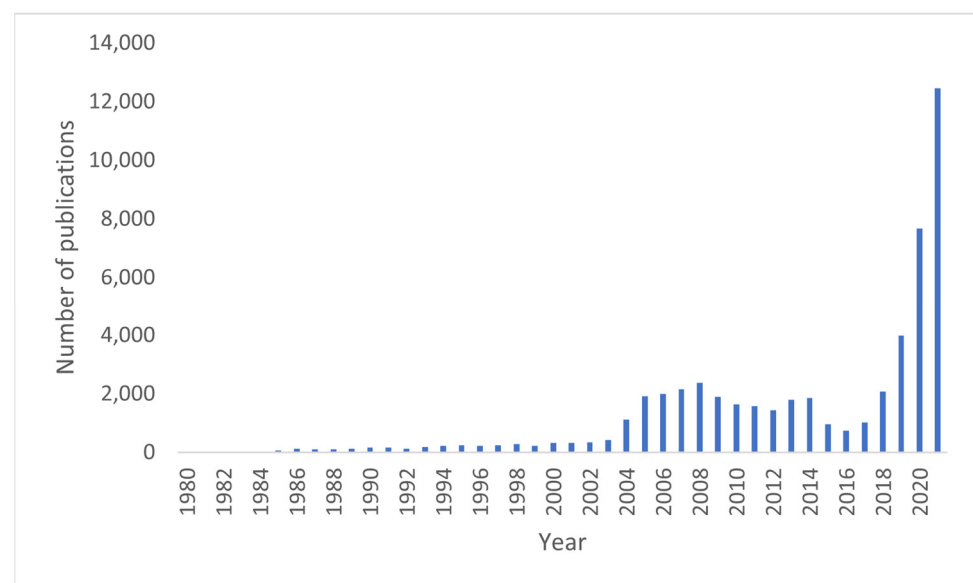
**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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

Artificial intelligence (AI) is a growing field that has experienced exponential growth, especially in the past five years (Figure 1). Reasons for that include the high-performance central- and graphics-processing units (CPU and GPU), the availability of huge amounts of data, and the development of learning algorithms [1].

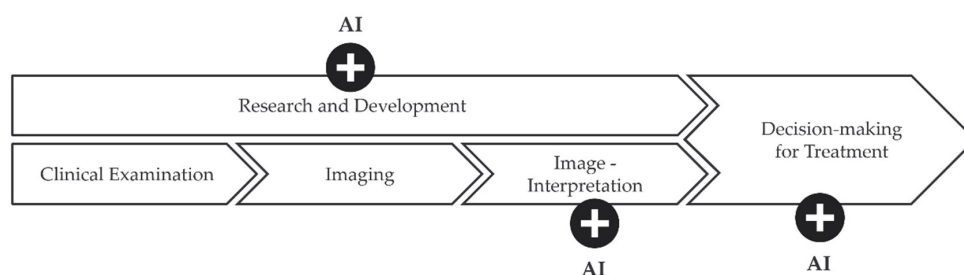


**Figure 1.** Number of PubMed search results for “artificial intelligence” with an exponential growth in numbers of published articles after 2016.

As early as 1955, John McCarthy et al. proposed a research project on AI to investigate their conjecture that every aspect of intelligence could be reproduced by a machine [2]; thus, the field of AI was founded, which became the foundation of subsequent computer research and development [3]. The early success of AI was due to the fact that tasks that were easy to formally program but difficult for humans to perform could be easily solved by AI. Ironically, tasks that seem easy for humans, like the recognition of objects or speech, are much harder to solve because they are done intuitively and are therefore difficult to put in a formal way such as code [4].

In the age of digitalization, where AI is applied in many fields of daily life, medicine is a field that has yet to make up the leeway. Although AI applications in medicine are increasing, most X-ray images are still assessed manually with narrative image descriptions [4–6]. AI in healthcare has the potential to create new methods to reduce the steadily increasing workload of clinicians, while also improving research and development, and recognizing underlying patterns that otherwise could not be recognized [7]. As there is always a risk of error when diagnosing radiographs, manually assessing them under severe and urgent conditions can elevate that risk to 40% or more [8]. As computers are not affected by their surroundings or stress, the performance of AI stays more consistent.

This narrative review gives an overview of AI in medicine, discusses the challenges, and summarizes the current applications of AI in orthopedic radiography analysis (Figure 2).



**Figure 2.** A schematic illustration of the role of artificial intelligence (AI) in orthopedic radiography analysis. AI can play a role in supporting and enhancing image interpretation, research and development, as well as decision-making for treatment plans.

### 1.1. Inter- and Intra-Observer Variability

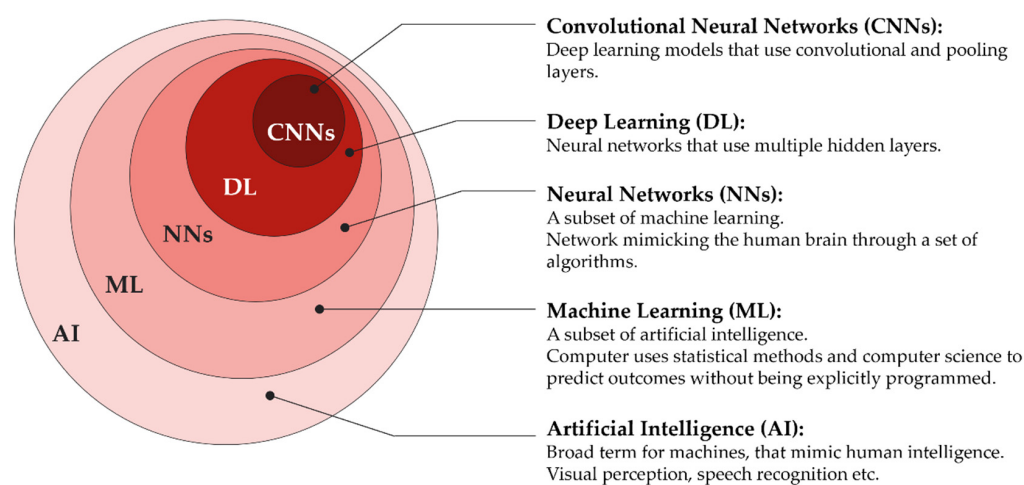
When manually measuring or classifying radiographs, aside from being prone to error under certain conditions, there is also the aspect of inter- and intraobserver variability. Inter- and intraobserver variability in orthopedic radiology can occur everywhere, for instance when measuring the acetabular component after hip arthroplasty [5] or assessing the sacroiliac joint for sacroiliitis [6]. While there have been studies comparing inter- and intraobserver variability, Rutgers and colleagues stated that although the quality of diagnosis by inexperienced clinicians can be improved through training, supervision by experienced observers is necessary [8]. It has been shown that it is possible to improve the variability and difference in quality through a computer-aided diagnosis, which is not exclusively found in orthopedic radiology, but also other fields of medicine [9].

While specialties such as oncology, cardiology, and neurosurgery have implemented AI in their daily practice, orthopedics and orthopedic surgery are only slowly embracing it, although its application is growing [10–12]. In 2017, Cabitza et al. could already see increasing demand for systematic reviews in the field of AI in orthopedics. Figure 1 demonstrates that the number of papers published has drastically increased over the last few years [13].

### 1.2. Artificial Intelligence and Machine Learning

Machine learning (ML) is part of the field of AI (Figure 3), which is based on algorithms that describe the relationships between variables. In conventional statistical analyses, the goal is to find inferences by understanding the relationships between variables to reach

conclusions or make predictions, based on the collected data. Data that is collected through the conversion of medical images, a process called radiomics, contains information on underlying pathophysiologies that can only be revealed by quantitative image analysis [14]. One of the advantages of ML is that it can formulate hypotheses based on the collected data, with only occasional input from humans, making it flexible. That flexibility, however, can also be a weakness [15]. While the primary goal of ML is to make predictions, in contrast to linear or logistic regression, the accuracy of the prediction or the “what” is more important than understanding the way or the “how” of reaching it, e.g., understanding the relationships of pixels of an X-ray is not important as long as the diagnosis of a fracture is correct [16]. Although this “black box” nature of ML models potentially offers more accurate performance in prediction, it comes with the cost of less understanding of the impact that variables might have on the outcome, making it more difficult to pinpoint the exact problems needed to treat [17].



**Figure 3.** Set diagram describing the relationship of Artificial intelligence (AI) being the umbrella term for everything inside the outer circle, Machine learning (ML) for everything inside the second-outer circle, etc. Neural networks (NN), deep learning (DL), and convolutional neural networks (CNNs).

The reason the above-mentioned flexibility of an ML model can be a weakness is overfitting. Overfitting is one major problem when it comes to ML. The term overfitting describes an error made by an ML model when it mistakenly extracts parameters out of a training dataset and assigns more importance to that parameter than it has in reality.

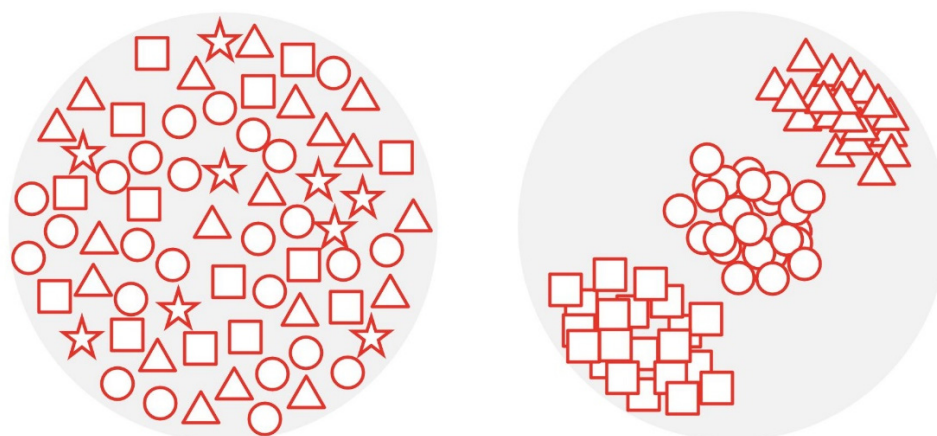
Typically, an ML algorithm runs the input data through its layers to get an output. The output is then compared with a “ground truth” (e.g., results of human readers as the ground truth for ML radiography analysis) and the parameters and their importance are adjusted repeatedly to reach an output with a minimal difference from the ground truth. This favours the parameters that have more “weight” in the decision-making for the output. If the ML model runs through the same dataset too many times, overfitting happens, and residual variation or “noise” are used for the predictions. One solution to overcome the overfitting problem is to collect bigger datasets for training. Although different amounts of data are required, depending on the study, a basic starting point of 1000 images per class will have a good chance for high accuracy. For medical data, it is often difficult to collect such large numbers due to confidentiality issues and high costs in acquiring the ground truth. To achieve better generalizability and reduce the risk of overfitting, external validation is necessary [18].

ML is broadly divided into three categories: supervised learning, unsupervised learning, and reinforcement learning [16,19,20].

Supervised learning is the most common form of learning used in medical practice. Typically, a supervised ML model is “trained” and validated on a dataset with given inputs

and outcomes. After successful training and validation, the model should be able to make predictions on new data. Those predictions can be either binary or continuous, which is then referred to as a classification algorithm or segmentation algorithm, respectively [16,20]. An example of a classification algorithm would be fracture detection, where there either is (true) or is no (false) fracture. A segmentation algorithm might be used to predict life expectancy for non-treated femur fractures.

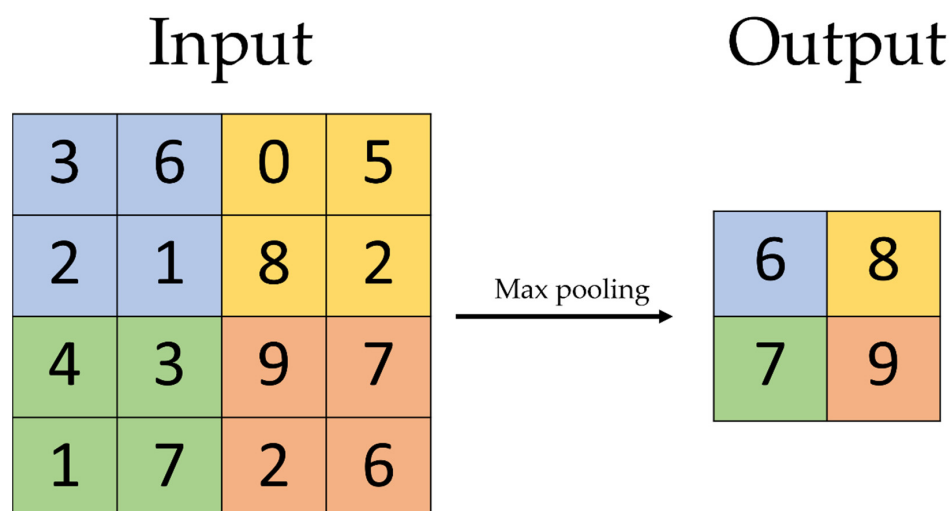
In contrast to supervised learning, in unsupervised learning the outcome is not put in manually. The goal here is to recognize patterns in a dataset, such as latent subspaces, by comparing the values of variables. If predefined variables show similar values, they are grouped together building a so-called cluster. By performing clusterization (Figure 4), patterns that would otherwise be difficult to recognize can be discovered [13,16,20,21].



**Figure 4.** Illustration describing the mechanism of clusterization: In the left image, no specific order can be seen. For clusterization, different shapes are recognized and organized in a group (cluster) as seen in the right image. Shapes that do not belong in any group (e.g., stars) are sorted out. This can help ML models to recognize patterns that would otherwise be difficult to recognize.

Reinforcement learning is the third category of ML, with the goal of optimizing the process for a certain reward. For example, in a game of chess or poker [22], the individual steps and sequences would be difficult to compute but the outcome of winning or losing can be predefined as a reward, and the algorithm learns the most optimal way of reaching it [20].

The most successful learning models recently developed are artificial neural networks (ANN) and deep learning (DL) models. Neural networks are learning models with an architecture similar to the mammalian cortex, which consists of two visible and a variable number of hidden layers. ANNs, with a high number of hidden layers, are considered DL models [16,20]. There is no clear cut-off of number of layers needed for a neural network to be defined as a DL model. Particularly suited for perceptual tasks, such as object recognition, are DL models that resemble the human visual cortex called convolutional neural networks (CNNs). The CNNs contain special types of layers called convolutional layers that apply filters on the input and pooling layers, which compute and filter the preceding layer, resulting in less computational load and a lower risk of overfitting (Figure 5). Today, CNNs are the majority of neural network architectures used in computer vision [20].



**Figure 5.** Max pooling layer: the input layer (left) is filtered by a  $2 \times 2$  max pooling layer computing the output with the highest activation (number) of the filtered area in the preceding layer into a smaller output layer. One can imagine a  $2 \times 2$  filter starting on the top left: The first filtered area will be the blue area. The number six, being the highest, will be computed as  $1 \times 1$  in the output, representing the  $2 \times 2$  input. After that, the filter will go to the next area that has not been filtered, e.g., two cells to the right, and repeat the process. With filtering methods like this, the required computational power can be reduced.

### 1.3. Applications of AI in Musculoskeletal Radiology

Musculoskeletal (MSK) diseases are the most common work-related health issues in the European Union (EU). According to the European Agency for Safety and Health at work, three out of five people have backache or pain in the neck or upper limbs [23]. In the “Global estimates for the need of rehabilitation . . . ” study, MSK disorders contributed most to the need for rehabilitation, with lower back pain being the most prevalent condition in more than half of the countries analyzed [24]. In addition, MSK trauma is among the most common reasons for unscheduled visits to the emergency department [25,26]. Missed or delayed diagnosis can lead to an increase in morbidity and mortality. With increasing numbers of images, physicians are pressured to work faster whilst maintaining quality standards. It has been estimated that the rate of radiologic error ranges from 4% to around 30% depending on the study sample with the two most common errors of underreading (a finding is present on the image but is missed) and satisfaction of search (findings missed because the search was not continued after one abnormality was found) [27]. When the 4% is extrapolated on one billion radiologic images a year, an estimated forty million radiologic errors are made [25]. Those errors might result in patients receiving inadequate treatment and the number of errors could be reduced if AI would be applied as a tool for clinicians to reduce workload [19].

## 2. Material and Methods

A PubMed search was performed on 5 May 2022 containing the following search string: (“Artificial Intelligence”[Title/Abstract] OR “deep learning”[Title/Abstract] OR “neural network”[Title/Abstract] OR “machine learning”[Title/Abstract]) AND (“musculoskeletal”[Title/Abstract] OR orthopedic\*[Title/Abstract] OR orthopaedic\*[Title/Abstract] OR muscle\*[Title/Abstract] OR skeletal\*[Title/Abstract] OR sceletal\*) AND (radiograph\*[Title/Abstract] OR “X-Ray”[Title/Abstract] OR “X-Rays”[Title/Abstract]). Additionally, reference lists of framework papers were searched.

## 2.1. AI in X-ray Imaging

### 2.1.1. Fracture Detection

There have been a number of studies with different products on fracture detection using AI on plain radiographs [28–34]. Fractures are the leading type of missed diagnosis [35]. DL models can classify images to detect whether there is a fracture or not and then use segmentation to assess localization and fracture type [36]. Yu and colleagues developed and tested a DL algorithm and evaluated its accuracy in the detection and localization of hip fractures. They found a sensitivity and specificity of 97.1% and 96.7%, respectively, for binary classification. In fracture localization, however, performance was lower with the sensitivity ranging from 95.8% in normal findings, 84.1% in subcapital/transcervical fractures, 76% in basicervical/intertrochanteric fractures, to 20% in subtrochanteric fractures. Although, when compared to subspecialty musculoskeletal radiologists the performance was lower, the DL algorithm showed high accuracy in fracture detection [33].

Liu and colleagues trained a DL algorithm for tibial plateau fractures (TPF) in radiographs and compared its performance with orthopedic physicians. The algorithm automatically labelled the suspicious TPF region with a rectangle. The accuracy of the algorithm was comparable to human performance with 0.91 and  $0.92 \pm 0.03$ , respectively. There was, however, a significant difference in the time spent assessing the X-rays, with the AI being 16-times faster than orthopedic physicians. The authors state that the working conditions for orthopedic physicians were not ideal to simulate real working conditions, since the study was conducted in nonemergency and under time-free constraints on the diagnosis process [29].

Missed diagnoses of vertebral fractures in plain radiographs can be as much as 30% [37]. Murata and colleagues developed a CNN to detect vertebral fractures and compared it to orthopedic residents, orthopedic surgeons, and spine surgeons. They found that there was no significant difference between the CNN, the orthopedic residents, and surgeons but spine surgeons had significantly higher accuracy, sensitivity, and specificity [30].

AI has been shown to be on par with or even outperform expert radiologists in fracture detection on radiographs and is a potential tool to aid clinicians to minimize errors and reduce burnout [25].

### 2.1.2. Classification

DL models can not only be trained to detect fractures but also for segmentation and classification. In knee osteoarthritis (OA), the measurement of joint space width (JSW)—which is the gap between the femur and tibia—and an indirect measure of cartilage width is the only recommended imaging biomarker as a structural endpoint in clinical trials, as recommended by the United States Food and Drug Administration (FDA) [38]. Automated knee segmentation has been investigated for accurate identification of the femur and tibia by several studies [39], and has been shown to deliver a more standardized and objective classification of joint-space narrowing [38]. The most widely used grading system for OA, however, is the Kellgren-Lawrence (KL) scale. The KL scale is recognized by the World Health Organization as the standard for clinical studies and includes radiological features: osteophytes, subchondral sclerosis, pseudocystic areas, and narrowing of the joint space [40]. Since several studies reported that the KL grading system suffers from high subjectivity and interobserver variability, studies investigating the benefit of AI-aided classification found that the consistency between physicians increased when automated knee OA software was used [41].

Since diagnosis and treatment in pediatric orthopedic disorders can depend on bone age, although a time-consuming activity, an accurate assessment is necessary. Bone age is closely related to skeletal maturity, whereas chronological age is based on an individual's date of birth. The two most common methods to assess bone age are the Greulich-Pyle and Tanner-Whitehouse methods. Both methods analyze the epi- and diaphyses of hand radiographs and compare them to reference images, leaving both methods prone to inter- and intraobserver variability [42]. Larson and colleagues tested a DL bone age assessment

model on 14,036 hand radiographs and found a comparable performance of the DL model to trained human reviewers, with a difference between bone age estimates of 0 years, root mean square of 0.63 years, and a mean absolute difference of 0.50 years [43]. It is important, however, that caution is taken when assessing bone age in patients with skeletal disorders like Hutchinson-Gilford progeria syndrome, where patients show a more advanced rate of skeletal maturation [44]. Large datasets that are needed to train ML models are much more difficult to acquire for rare diseases, which makes experience and supervision by humans a necessity.

The standard tool used for the diagnosis of osteoporosis is the assessment of bone mineral density (BMD). Dual-energy X-ray Absorptiometry (DXA) is the gold standard of methods used for the assessment of BMD and is the basis for the WHO guidelines for the diagnosis of osteoporosis [45]. Yamamoto and colleagues investigated a DL model to diagnose osteoporosis in hip radiographs. They tested five different CNNs and found high accuracy. When adding routinely available patient variables such as age, gender, and BMI, the diagnostic accuracy improved to an AUC of 0.906 with a ResNet50 CNN architecture [46,47].

Total joint replacement (TJR) is the most common elective surgical procedure for end-stage degenerative joint disease. The primary diagnosis leading to TJR is OA accounting for up to 81% of hip and 94% of knee replacements [48]. Implants fail within 15–20 years in around 10% of all TJRs and revision surgery is needed. The planning for revision surgery requires identification of the implant's model, which can be difficult when the primary surgery was performed more than 10 years ago and/or in a different country. Studies have shown that DL models can automatically identify implant models in the hip as well as shoulders on plain radiographs [49,50]. Borjali and colleagues compared the performance of a DL model to that of orthopedic surgeons and found similar or higher performance by the DL model. They also found that the DL model was significantly faster than surgeons, taking about 0.06 s and 8.4 min, respectively, given that the surgeon did not know the implant model by experience [49].

## 2.2. Measurements

Another application for AI in radiographs is in automated measurements, for example, leg alignment, joint orientation, leg length, or implant orientation [51–54]. Long-leg radiographs (LLR) are routinely acquired and manually measured to assess malalignment, which is considered a major contributing factor in OA [53]. The measurement of leg-length discrepancies (LLD), hip-knee-ankle angle (HKAA), or femoral anatomic-mechanical angle (AMA) is relatively simple but is a time-consuming task [55]. Several studies have tested DL models on automatic measurements of the lower limb measuring HKA, AMA, and LLD. The studies found that the AI was non-inferior when compared to human readers (accuracy ranging from 89.2 to 93%) but was significantly faster [52–54]. However, severe deformities of the lower limb and bad image quality are common exclusion criteria for the studies mentioned above, emphasizing again that experienced clinicians are indispensable as supervisors. This becomes even more important in undiagnosed bony abnormalities.

## 2.3. Computed Tomography and Magnetic Resonance Imaging

Apart from radiographs, studies on AI in MSK computed tomography (CT) have also been conducted. For example, for the detection and classification of fractures of the calcaneus, femur, and mandibula AI could be a useful tool for radiologists and orthopedic surgeons to reduce workload and improve accuracy [56–59]. When assessing rib fractures in CTs, radiologists achieved higher accuracy while reducing their reading time with the assistance of a DL model compared to without [60]. Aside from the analysis of bones, AI can also be used in the analysis of skeletal muscle and confirmation of atrophy in CT scans [61].

Magnetic resonance imaging (MRI) is unparalleled in detecting musculoskeletal diseases and is the preferred method for diagnosing knee injuries [62,63]. Several studies

have applied AI in the assessment of MRI, for example, in the detection of injuries of the anterior cruciate ligament (ACL) [57], meniscus tears [64], or cartilage lesions [65]. Machine learning in MRI can also be used for the assessment of vertebrae, discs, and muscles in patients with lower back pain [66]. In a systematic review, Siouras and colleagues stated that AI in MRI has the potential to be on par with human-level performance and has shown a prediction accuracy ranging from 72.5–100% [67].

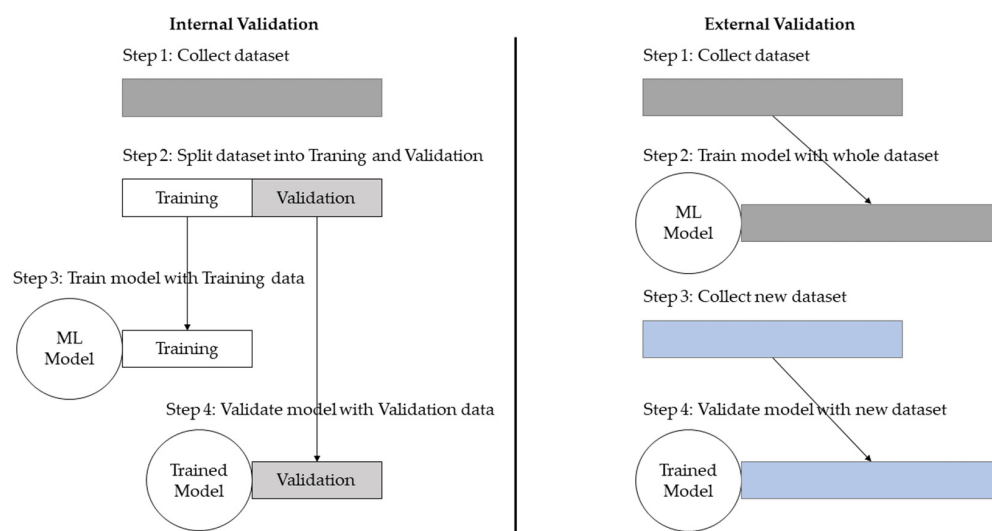
Research has shown that orthopedic problems can also be related to orthodontics. Neural networks have produced a good performance in treatment planning and can offer guidance for less experienced orthodontists [68]. For example, malocclusion has been shown to be related to scoliosis and weak body posture [69]. The implementation of MRI in orthodontics has shown great potential for radiation-free imaging. Compared to the currently used cone-beam CT, MRI has shown better results in the detection of periodontal structures, soft tissue debris, and unfilled spaces in radicular canals, which is necessary information for endodontic treatment [70–72]. Additionally, orthodontic treatment is mostly received during childhood, which makes MRI even more desirable [73]. As mentioned before, overfitting is a challenge for ML algorithms. The size and quality of data, especially with MRI images where the signal-noise ratio can be worse, determine and influence the performance quality. Although MRI images show a lot of information, they also have more noise [74].

#### 2.4. Internal and External Validation

The development of ML models requires training and validation, where validation is the process of evaluating the accuracy of the model. Generally, the validation processes can be divided into internal and external validation. Internal validation splits the same dataset to train and validate the ML model. The process where one part of the data is used for training and the rest is used for validation repeatedly, until all parts of the dataset have been used for training, is referred to as cross-validation. An advantage of using internal validation is that it is easier and more economical to perform than external validation since only one dataset is needed. The disadvantage, however, is that even if cross-validation shows high accuracy, it does not give information about the reproducibility or meaningfulness of findings. Ho and colleagues made a comparison between the learning of ML models and humans: When two humans study for an exam and one person critically evaluates and learns to understand the subject and the other person simply memorizes, both can reach a result of 99%. This result, however, does not give us information about the performance of those two humans on another exam of the same subject with different questions. The same principle can be applied to ML models: two models that are trained on the same dataset can reach the same result of 99% accuracy but the way the ML model learned to reach that accuracy may differ, which can lead to different results in performance on other datasets [75]. Additionally, the risk of bias in internally validated models is increased, because the training dataset may not represent the target population properly. Zech and colleagues noticed that the word “portable”, which was a label indicating that X-rays were taken with a portable scanner, was detected by the ML model and used as an indicator for a higher possibility of disease [76]. While that may have been the case in that specific hospital setting, the ML model suffered from overfitting because the reproducibility for different hospitals with different imaging procedures, that might not have “portable” written on the radiographs, can differ in performance.

External validation, on the other hand, uses one dataset for training and a different (external) dataset for validation. The validation dataset is supposed to be profoundly different from the training dataset, e.g., in a different geographical location. Although collecting a different dataset requires more effort, external validation can be used to counteract the above-mentioned disadvantages of internal validation and is considered important evidence for generalizability (Figure 6) [77].





**Figure 6.** Internal vs. external validation. Internal validation uses one dataset and splits it into training and validation, the ML model is then trained and validated on the same dataset resulting in higher risk of overfitting. External validation uses one dataset for training and a meaningfully different dataset for validation. When (geographically) different datasets are used, the risk of overfitting can be significantly reduced.

Since internally validated ML models tend to perform better on the data they were trained on [78,79], Oliveira and colleagues investigated ML models for fracture detection that were geographically (datasets from different locations) or temporally (datasets from different time points) externally validated. Out of 36 CNNs that were investigated, only four (11%) used a form of external validation, which limits the potential to use those CNNs in clinical practice [80].

### 3. Discussion

This narrative review gives an introduction to AI and summarizes the current applications of AI in orthopedic radiography imaging. The literature on AI in orthopedics is rapidly growing with different types of studies being conducted. Due to that heterogeneity, it is difficult to structure and compare data properly. AI in orthopedic radiographs is applied in the fields such as fracture detection, identification and classification of bone age, osteoporosis or implants, and measurement of alignment or implant position. Studies mostly show non-inferior performance when compared to physicians, orthopedic surgeons, and radiologists, whilst being faster at performing the tasks. For some, this now might raise the question of whether AI will replace clinicians, especially radiologists. When ML first had its hype in 2016, some authors hypothesized that the tasks of radiologists could be entirely replaced by AI, and some even stated that the training of radiologists should be stopped. This narrative, however, has been revised by many experts in the field that are working together with radiologists. Currently, AI is often trained on a single disease or a small set of diseases, which is far under the scope of what radiologists do. The performance of AI can reach that of a subspecialty thoracic radiologist when it comes to the diagnosis of common chest conditions; however, in contrast to the human radiologist, AI falls behind in detecting uncommon diseases. Hence, the very fitting comparison by Langlotz: “AI is impressive in identifying horses but is a long way from recognizing zebras” [81].

The performance of clinicians when aided by AI software has been shown to improve in comparison to AI or human performance alone. This could not only be applied in clinical practice to improve diagnosis quality but also as a training tool, especially for junior doctors with less experience [41].

Since medicine is in the process of digitalization, it is also important to familiarize junior doctors with the topic of AI for clinical practice. As Çalışkan and colleagues state,

medical students need to understand why AI can support doctors in clinical practice [82]. On the other hand, we cannot rule out the possibility that medical students and junior doctors could rely too much on AI, which can become a hindrance in developing analytical and observational skills as well as in learning the process of decision-making.

The importance of external validation has been explained above. As Oliveira and colleagues point out, external validation, which means validating a DL model on a different dataset than it was trained on, is hardly carried out and was only reported in 11% of studies investigating CNNs for fracture detection. To achieve a higher quality standard and reduce the risk of bias, geographically external validation should always be performed to assess performance and generalizability [80].

Another factor that is rarely investigated is ethnicity. In a systematic review, Dallora and colleagues stated that only two out of twenty-six studies considered ethnicity as a factor, and found that studies are predominantly conducted in the United States and Western Europe whilst it is known that, e.g., skeletal maturation, occurs earlier in Indonesians and later in African Americans when compared to Caucasians [42].

The application of AI models in orthopedics focused on predicting outcomes is still in its infancy. The studies investigating prediction models are mostly focused on medical management (predictions of delayed discharge or readmission) and the subspecialty of spine surgery [83]. We believe that there should be more attention given to AI prediction models in the future since providing predictions is a task that cannot be performed by humans. Currently, the predictions made by clinicians can only be made based on personal experience and statistical incidences; however, if the prediction of diseases could be made accurately, it could affect the following treatment.

The term big data was invented by computer scientists to describe the evolving technology based on digitally collected and rapidly expanding data. The idea of big data in healthcare is to share data between practitioners and patients to improve healthcare delivery [84]. Currently, most ML models are programmed on relatively small datasets by companies. An important step forward could be taken if data sharing would take place between all institutions.

Important aspects to discuss are the legal and ethical concerns. According to Naik et al., four major issues must be considered in this context: (1) informed consent to use the data, (2) safety and transparency, (3) algorithmic fairness and biases, and (4) data privacy. Transparency of the algorithm can be difficult to achieve because modern computing approaches can deliberately hide the thinking behind the output. To confront that problem, diverse and inclusive programming groups as well as frequent audits of the algorithm should be implemented. The data on which ML models are trained can be subject to underlying bias leading to biased ML models. For that, it is important to acquire good-quality unbiased data. Since technologists are not obligated by law to be held accountable for their actions to the same degree as doctors, the question of whether, and if yes how much, they should take responsibility for the AI system they created, remains. An effort on clear guidelines for the legal and ethical issues concerning AI should be made [85].

#### 4. Conclusions

Although AI in medicine is a rapidly growing field that shows promising results, clinical implementation is still lacking. Different studies develop and test their own ML algorithms without open-source access, leading to studies not being reproducible. Finding a clear structure for study design and the implementation of data sharing, as well as the regular use of external validation should be aimed for, in order to drive progress forward.

**Author Contributions:** Conceptualization, K.C. and C.S.; methodology, K.C. and C.S.; validation, T.K. and S.N.; investigation, K.C.; writing—original draft preparation, K.C.; writing—review and editing, K.C. and C.S.; visualization, K.C.; supervision, C.S., T.K. and S.N.; project administration, T.K. and S.N.; funding acquisition, C.S., T.K. and S.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Gesellschaft für Forschungsförderung Niederösterreich m.b.H. (GFF), grant number LS20-020.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Shen, D.; Wu, G.; Suk, H.-I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)] [[PubMed](#)]
2. McCarthy, J.; Minsky, M.L.; Rochester, N.; Shannon, C.E. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Mag.* **2006**, *27*, 12. [[CrossRef](#)]
3. Hamet, P.; Tremblay, J. Artificial intelligence in medicine. *Metabolism* **2017**, *69S*, S36–S40. [[CrossRef](#)] [[PubMed](#)]
4. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: Cambridge, MA, USA, 2016; ISBN 0262337371.
5. 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](#)]
6. Braun, J.; Sieper, J.; Bollow, M. Imaging of sacroiliitis. *Clin. Rheumatol.* **2000**, *19*, 51–57. [[CrossRef](#)]
7. Hameed, B.M.Z.; Dhavileswarapu, A.V.L.S.; Raza, S.Z.; Karimi, H.; Khanuja, H.S.; Shetty, D.K.; Ibrahim, S.; Shah, M.J.; Naik, N.; Paul, R.; et al. Artificial Intelligence and Its Impact on Urological Diseases and Management: A Comprehensive Review of the Literature. *J. Clin. Med.* **2021**, *10*, 1864. [[CrossRef](#)]
8. Rutgers, J.J.; Bánki, T.; van der Kamp, A.; Waterlander, T.J.; Scheijde-Vermeulen, M.A.; van den Heuvel-Eibrink, M.M.; van der Laak, J.A.W.M.; Fiocco, M.; Mavinkurve-Groothuis, A.M.C.; de Krijger, R.R. Interobserver variability between experienced and inexperienced observers in the histopathological analysis of Wilms tumors: A pilot study for future algorithmic approach. *Diagn. Pathol.* **2021**, *16*, 77. [[CrossRef](#)]
9. Ahmad, O.F.; Soares, A.S.; Mazomenos, E.; Brandao, P.; Vega, R.; Seward, E.; Stoyanov, D.; Chand, M.; Lovat, L.B. Artificial intelligence and computer-aided diagnosis in colonoscopy: Current evidence and future directions. *Lancet Gastroenterol. Hepatol.* **2019**, *4*, 71–80. [[CrossRef](#)]
10. Maffulli, N.; Rodriguez, H.C.; Stone, I.W.; Nam, A.; Song, A.; Gupta, M.; Alvarado, R.; Ramon, D.; Gupta, A. Artificial intelligence and machine learning in orthopedic surgery: A systematic review protocol. *J. Orthop. Surg. Res.* **2020**, *15*, 478. [[CrossRef](#)]
11. Senders, J.T.; Arnaout, O.; Karhade, A.V.; Dasenbrock, H.H.; Gormley, W.B.; Broekman, M.L.; Smith, T.R. Natural and Artificial Intelligence in Neurosurgery: A Systematic Review. *Neurosurgery* **2018**, *83*, 181–192. [[CrossRef](#)]
12. Federer, S.J.; Jones, G.G. Artificial intelligence in orthopaedics: A scoping review. *PLoS ONE* **2021**, *16*, e0260471. [[CrossRef](#)] [[PubMed](#)]
13. Cabitza, F.; Locoro, A.; Banfi, G. Machine Learning in Orthopedics: A Literature Review. *Front. Bioeng. Biotechnol.* **2018**, *6*, 75. [[CrossRef](#)] [[PubMed](#)]
14. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [[CrossRef](#)] [[PubMed](#)]
15. Demšar, J.; Zupan, B. Hands-on training about overfitting. *PLoS Comput. Biol.* **2021**, *17*, e1008671. [[CrossRef](#)]
16. Sidey-Gibbons, J.A.M.; Sidey-Gibbons, C.J. Machine learning in medicine: A practical introduction. *BMC Med. Res. Methodol.* **2019**, *19*, 64. [[CrossRef](#)]
17. Bayliss, L.; Jones, L.D. The role of artificial intelligence and machine learning in predicting orthopaedic outcomes. *Bone Jt. J.* **2019**, *101*, 1476–1478. [[CrossRef](#)]
18. Mutasa, S.; Sun, S.; Ha, R. Understanding artificial intelligence based radiology studies: What is overfitting? *Clin. Imaging* **2020**, *65*, 96–99. [[CrossRef](#)]
19. Lidströmer, N.; Ashrafian, H. (Eds.) *Artificial Intelligence in Medicine*; Springer International Publishing: Cham, Switzerland, 2020; ISBN 9783030580803.
20. Razavian, N.; Knoll, F.; Geras, K.J. Artificial Intelligence Explained for Nonexperts. *Semin. Musculoskelet. Radiol.* **2020**, *24*, 3–11. [[CrossRef](#)]
21. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)]
22. Brown, N.; Sandholm, T. Superhuman AI for multiplayer poker. *Science* **2019**, *365*, 885–890. [[CrossRef](#)]
23. De Kok, J.; Vroonhof, P.; Sniijders, J.; Roullis, G.; Clarke, M.; Peereboom, K.; Dorst, P.; Isusi, I. *Work-Related MSDs: Prevalence, Costs and Demographics in the EU*; European Agency for Safety and Health at Work (EU-OSHA): Bilbao, Spain, 2019.
24. Cieza, A.; Causey, K.; Kamenov, K.; Hanson, S.W.; Chatterji, S.; Vos, T. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **2020**, *396*, 2006–2017. [[CrossRef](#)]

25. Ajmera, P.; Kharat, A.; Botchu, R.; Gupta, H.; Kulkarni, V. Real-world analysis of artificial intelligence in musculoskeletal trauma. *J. Clin. Orthop. Trauma* **2021**, *22*, 101573. [[CrossRef](#)] [[PubMed](#)]
26. Laur, O.; Wang, B. Musculoskeletal trauma and artificial intelligence: Current trends and projections. *Skelet. Radiol.* **2022**, *51*, 257–269. [[CrossRef](#)] [[PubMed](#)]
27. Bruno, M.A.; Walker, E.A.; Abujudeh, H.H. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. *Radiographics* **2015**, *35*, 1668–1676. [[CrossRef](#)] [[PubMed](#)]
28. 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](#)]
29. Liu, P.-R.; Zhang, J.-Y.; Xue, M.; Duan, Y.-Y.; Hu, J.-L.; Liu, S.-X.; Xie, Y.; Wang, H.-L.; Wang, J.-W.; Huo, T.-T.; et al. Artificial Intelligence to Diagnose Tibial Plateau Fractures: An Intelligent Assistant for Orthopedic Physicians. *Curr. Med. Sci.* **2021**, *41*, 1158–1164. [[CrossRef](#)]
30. Murata, K.; Endo, K.; Aihara, T.; Suzuki, H.; Sawaji, Y.; Matsuoka, Y.; Nishimura, H.; Takamatsu, T.; Konishi, T.; Maekawa, A.; et al. Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. *Sci. Rep.* **2020**, *10*, 20031. [[CrossRef](#)]
31. Saun, T.J. Automated Classification of Radiographic Positioning of Hand X-rays Using a Deep Neural Network. *Plast. Surg.* **2021**, *29*, 75–80. [[CrossRef](#)]
32. Suzuki, T.; Maki, S.; Yamazaki, T.; Wakita, H.; Toguchi, Y.; Horii, M.; Yamauchi, T.; Kawamura, K.; Aramomi, M.; Sugiyama, H.; et al. Detecting Distal Radial Fractures from Wrist Radiographs Using a Deep Convolutional Neural Network with an Accuracy Comparable to Hand Orthopedic Surgeons. *J. Digit. Imaging* **2022**, *35*, 39–46. [[CrossRef](#)]
33. Yu, J.S.; Yu, S.M.; Erdal, B.S.; Demirel, M.; Gupta, V.; Bigelow, M.; Salvador, A.; Rink, T.; Lenobel, S.S.; Prevedello, L.M.; et al. Detection and localisation of hip fractures on anteroposterior radiographs with artificial intelligence: Proof of concept. *Clin. Radiol.* **2020**, *75*, 237.e1–237.e9. [[CrossRef](#)]
34. Zech, J.R.; Santomartino, S.M.; Yi, P.H. Artificial Intelligence (AI) for Fracture Diagnosis: An Overview of Current Products and Considerations for Clinical Adoption, from the AJR Special Series on AI Applications. *Am. J. Roentgenol.* **2022**. [[CrossRef](#)] [[PubMed](#)]
35. Kachalia, A.; Gandhi, T.K.; Puopolo, A.L.; Yoon, C.; Thomas, E.J.; Griffey, R.; Brennan, T.A.; Studdert, D.M. Missed and delayed diagnoses in the emergency department: A study of closed malpractice claims from 4 liability insurers. *Ann. Emerg. Med.* **2007**, *49*, 196–205. [[CrossRef](#)] [[PubMed](#)]
36. Kalmet, P.H.S.; Sanduleanu, S.; Primakov, S.; Wu, G.; Jochems, A.; Refaee, T.; Ibrahim, A.; Hulst, L.V.; Lambin, P.; Poeze, M. Deep learning in fracture detection: A narrative review. *Acta Orthop.* **2020**, *91*, 215–220. [[CrossRef](#)] [[PubMed](#)]
37. Berry, G.E.; Adams, S.; Harris, M.B.; Boles, C.A.; McKernan, M.G.; Collinson, F.; Hoth, J.J.; Meredith, J.W.; Chang, M.C.; Miller, P.R. Are plain radiographs of the spine necessary during evaluation after blunt trauma? Accuracy of screening torso computed tomography in thoracic/lumbar spine fracture diagnosis. *J. Trauma* **2005**, *59*, 1410–1413; discussion 1413. [[CrossRef](#)]
38. Paixao, T.; DiFranco, M.D.; Ljuhar, R.; Ljuhar, D.; Goetz, C.; Bertalan, Z.; Dimai, H.P.; Nehrer, S. A novel quantitative metric for joint space width: Data from the Osteoarthritis Initiative (OAI). *Osteoarthr. Cartil.* **2020**, *28*, 1055–1061. [[CrossRef](#)]
39. Kim, Y.J.; Lee, S.R.; Choi, J.-Y.; Kim, K.G. Using Convolutional Neural Network with Taguchi Parametric Optimization for Knee Segmentation from X-ray Images. *Biomed. Res. Int.* **2021**, *2021*, 5521009. [[CrossRef](#)]
40. Kohn, M.D.; Sassoon, A.A.; Fernando, N.D. Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. *Clin. Orthop. Relat. Res.* **2016**, *474*, 1886–1893. [[CrossRef](#)]
41. Nehrer, S.; Ljuhar, R.; Steindl, P.; Simon, R.; Maurer, D.; Ljuhar, D.; Bertalan, Z.; Dimai, H.P.; Goetz, C.; Paixao, T. Automated Knee Osteoarthritis Assessment Increases Physicians' Agreement Rate and Accuracy: Data from the Osteoarthritis Initiative. *Cartilage* **2021**, *13*, 957S–965S. [[CrossRef](#)]
42. Dallora, A.L.; Anderberg, P.; Kvist, O.; Mendes, E.; Diaz Ruiz, S.; Sanmartin Berglund, J. Bone age assessment with various machine learning techniques: A systematic literature review and meta-analysis. *PLoS ONE* **2019**, *14*, e0220242. [[CrossRef](#)]
43. Larson, D.B.; Chen, M.C.; Lungren, M.P.; Halabi, S.S.; Stence, N.V.; Langlotz, C.P. Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. *Radiology* **2018**, *287*, 313–322. [[CrossRef](#)]
44. Tsai, A.; Johnston, P.R.; Gordon, L.B.; Walters, M.; Kleinman, M.; Laor, T. Skeletal maturation and long-bone growth patterns of patients with progeria: A retrospective study. *Lancet Child Adolesc. Health* **2020**, *4*, 281–289. [[CrossRef](#)]
45. Lane, N.E. Epidemiology, etiology, and diagnosis of osteoporosis. *Am. J. Obstet. Gynecol.* **2006**, *194*, S3–S11. [[CrossRef](#)] [[PubMed](#)]
46. 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](#)]
47. Yamamoto, N.; Sukegawa, S.; Yamashita, K.; Manabe, M.; Nakano, K.; Takabatake, K.; Kawai, H.; Ozaki, T.; Kawasaki, K.; Nagatsuka, H.; et al. Effect of Patient Clinical Variables in Osteoporosis Classification Using Hip X-rays in Deep Learning Analysis. *Medicina* **2021**, *57*, 846. [[CrossRef](#)] [[PubMed](#)]
48. Snell, D.L.; Hipango, J.; Sinnott, K.A.; Dunn, J.A.; Rothwell, A.; Hsieh, C.J.; DeJong, G.; Hooper, G. Rehabilitation after total joint replacement: A scoping study. *Disabil. Rehabil.* **2018**, *40*, 1718–1731. [[CrossRef](#)]
49. Borjali, A.; Chen, A.F.; Bedair, H.S.; Melnic, C.M.; Muratoglu, O.K.; Morid, M.A.; Varadarajan, K.M. Comparing the performance of a deep convolutional neural network with orthopedic surgeons on the identification of total hip prosthesis design from plain radiographs. *Med. Phys.* **2021**, *48*, 2327–2336. [[CrossRef](#)]

50. 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](#)]
51. Lee, C.S.; Lee, M.S.; Byon, S.S.; Kim, S.H.; Lee, B.I.; Lee, B.-D. Computer-aided automatic measurement of leg length on full leg radiographs. *Skelet. Radiol.* **2021**, *51*, 1007–1016. [[CrossRef](#)]
52. 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](#)]
53. Schock, J.; Truhn, D.; Abrar, D.B.; Merhof, D.; Conrad, S.; Post, M.; Mittelstrass, F.; Kuhl, C.; Nebelung, S. Automated Analysis of Alignment in Long-Leg Radiographs by Using a Fully Automated Support System Based on Artificial Intelligence. *Radiol. Artif. Intell.* **2021**, *3*, e200198. [[CrossRef](#)]
54. Simon, S.; Schwarz, G.M.; Aichmair, A.; Frank, B.J.H.; Hummer, A.; DiFranco, M.D.; Dominkus, M.; Hofstaetter, J.G. Fully automated deep learning for knee alignment assessment in lower extremity radiographs: A cross-sectional diagnostic study. *Skelet. Radiol.* **2022**, *51*, 1249–1259. [[CrossRef](#)] [[PubMed](#)]
55. Tsai, A. Anatomical landmark localization via convolutional neural networks for limb-length discrepancy measurements. *Pediatr. Radiol.* **2021**, *51*, 1431–1447. [[CrossRef](#)] [[PubMed](#)]
56. Yoon, S.-J.; Hyong Kim, T.; Joo, S.-B.; Eel Oh, S. Automatic multi-class intertrochanteric femur fracture detection from CT images based on AO/OTA classification using faster R-CNN-BO method. *J. Appl. Biomed.* **2020**, *18*, 97–105. [[CrossRef](#)] [[PubMed](#)]
57. Wang, X.; Xu, Z.; Tong, Y.; Xia, L.; Jie, B.; Ding, P.; Bai, H.; Zhang, Y.; He, Y. Detection and classification of mandibular fracture on CT scan using deep convolutional neural network. *Clin. Oral Investig.* **2022**, *26*, 4593–4601. [[CrossRef](#)]
58. Pranata, Y.D.; Wang, K.-C.; Wang, J.-C.; Idram, I.; Lai, J.-Y.; Liu, J.-W.; Hsieh, I.-H. Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. *Comput. Methods Programs Biomed.* **2019**, *171*, 27–37. [[CrossRef](#)]
59. Aghnia Farda, N.; 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* **2021**, *52*, 616–624. [[CrossRef](#)]
60. Meng, X.H.; Di Wu, J.; Wang, Z.; Ma, X.L.; Dong, X.M.; Liu, A.E.; Chen, L. A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance. *Skelet. Radiol.* **2021**, *50*, 1821–1828. [[CrossRef](#)]
61. Kamiya, N. Deep Learning Technique for Musculoskeletal Analysis. *Adv. Exp. Med. Biol.* **2020**, *1213*, 165–176. [[CrossRef](#)]
62. Sage, J.E.; Gavin, P. Musculoskeletal MRI. *Vet. Clin. N. Am. Small Anim. Pract.* **2016**, *46*, 421–451. [[CrossRef](#)]
63. 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](#)]
64. Rizk, B.; Brat, H.; Zille, P.; Guillin, R.; Pouchy, C.; Adam, C.; Ardon, R.; d’Assignies, G. Meniscal lesion detection and characterization in adult knee MRI: A deep learning model approach with external validation. *Phys. Med.* **2021**, *83*, 64–71. [[CrossRef](#)] [[PubMed](#)]
65. 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](#)]
66. D’Antoni, F.; Russo, F.; Ambrosio, L.; Vollero, L.; Vadalà, G.; Merone, M.; Papalia, R.; Denaro, V. Artificial Intelligence and Computer Vision in Low Back Pain: A Systematic Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10909. [[CrossRef](#)]
67. Siouras, A.; Moustakidis, S.; Giannakidis, A.; Chalatsis, G.; Liampas, I.; Vlychou, M.; Hantes, M.; Tasoulis, S.; Tsaopoulos, D. Knee Injury Detection Using Deep Learning on MRI Studies: A Systematic Review. *Diagnostics* **2022**, *12*, 537. [[CrossRef](#)]
68. Li, P.; Kong, D.; Tang, T.; Su, D.; Yang, P.; Wang, H.; Zhao, Z.; Liu, Y. Orthodontic Treatment Planning based on Artificial Neural Networks. *Sci. Rep.* **2019**, *9*, 2037. [[CrossRef](#)] [[PubMed](#)]
69. Rohella, D.; Oswal, P.; Pizarla, M.; Maini, A.P.; Das, A.; Kandikatla, P.; Rajesh, D. Interrelation of Orthopedic and Orthodontic Findings from 6–12 Years Of Age: An Original Research. *J. Posit. Sch. Psychol.* **2022**, *6*, 4960–4968.
70. Capuani, S.; Gambarini, G.; Guarnieri, R.; Di Pietro, G.; Testarelli, L.; Di Nardo, D. Nuclear Magnetic Resonance Microimaging for the Qualitative Assessment of Root Canal Treatment: An Ex Vivo Preliminary Study. *Diagnostics* **2021**, *11*, 1012. [[CrossRef](#)]
71. Gaudino, C.; Cosgarea, R.; Heiland, S.; Csernus, R.; Beomonte Zobel, B.; Pham, M.; Kim, T.-S.; Bendszus, M.; Rohde, S. MR-Imaging of teeth and periodontal apparatus: An experimental study comparing high-resolution MRI with MDCT and CBCT. *Eur. Radiol.* **2011**, *21*, 2575–2583. [[CrossRef](#)]
72. Niraj, L.K.; Patthi, B.; Singla, A.; Gupta, R.; Ali, I.; Dhama, K.; Kumar, J.K.; Prasad, M. MRI in Dentistry—A Future Towards Radiation Free Imaging—Systematic Review. *J. Clin. Diagn. Res.* **2016**, *10*, ZE14–ZE19. [[CrossRef](#)]
73. Gili, T.; Di Carlo, G.; Capuani, S.; Auconi, P.; Caldarelli, G.; Polimeni, A. Complexity and data mining in dental research: A network medicine perspective on interceptive orthodontics. *Orthod. Craniofac. Res.* **2021**, *24* (Suppl. 2), 16–25. [[CrossRef](#)]
74. Mazurowski, M.A.; Buda, M.; Saha, A.; Bashir, M.R. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *J. Magn. Reson. Imaging* **2019**, *49*, 939–954. [[CrossRef](#)] [[PubMed](#)]
75. Ho, S.Y.; Wong, L.; Goh, W.W.B. Avoid Oversimplifications in Machine Learning: Going beyond the Class-Prediction Accuracy. *Patterns* **2020**, *1*, 100025. [[CrossRef](#)] [[PubMed](#)]
76. Zech, J.R.; Badgeley, M.A.; Liu, M.; Costa, A.B.; Titano, J.J.; Oermann, E.K. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med.* **2018**, *15*, e1002683. [[CrossRef](#)]

77. Ho, S.Y.; Phua, K.; Wong, L.; Bin Goh, W.W. Extensions of the External Validation for Checking Learned Model Interpretability and Generalizability. *Patterns* **2020**, *1*, 100129. [[CrossRef](#)] [[PubMed](#)]
78. Bongers, M.E.R.; Thio, Q.C.B.S.; Karhade, A.V.; Stor, M.L.; Raskin, K.A.; Lozano Calderon, S.A.; DeLaney, T.F.; Ferrone, M.L.; Schwab, J.H. Does the SORG Algorithm Predict 5-year Survival in Patients with Chondrosarcoma? An External Validation. *Clin. Orthop. Relat. Res.* **2019**, *477*, 2296–2303. [[CrossRef](#)] [[PubMed](#)]
79. Bleeker, S.; Moll, H.; Steyerberg, E.; Donders, A.; Derksen-Lubsen, G.; Grobbee, D.; Moons, K. External validation is necessary in prediction research. *J. Clin. Epidemiol.* **2003**, *56*, 826–832. [[CrossRef](#)]
80. Oliveira, E.; Carmo, L.; van den Merkhof, A.; Olczak, J.; Gordon, M.; Jutte, P.C.; Jaarsma, R.L.; IJpma, F.F.A.; Doornberg, J.N.; Prijs, J. An increasing number of convolutional neural networks for fracture recognition and classification in orthopaedics: Are these externally validated and ready for clinical application? *Bone Jt. Open* **2021**, *2*, 879–885. [[CrossRef](#)]
81. Langlotz, C.P. Will Artificial Intelligence Replace Radiologists? *Radiol. Artif. Intell.* **2019**, *1*, e190058. [[CrossRef](#)]
82. Çalışkan, S.A.; Demir, K.; Karaca, O. Artificial intelligence in medical education curriculum: An e-Delphi study for competencies. *PLoS ONE* **2022**, *17*, e0271872. [[CrossRef](#)]
83. Ogink, P.T.; Groot, O.Q.; Karhade, A.V.; Bongers, M.E.R.; Oner, F.C.; Verlaan, J.-J.; Schwab, J.H. Wide range of applications for machine-learning prediction models in orthopedic surgical outcome: A systematic review. *Acta Orthop.* **2021**, *92*, 526–531. [[CrossRef](#)]
84. Murdoch, T.B.; Detsky, A.S. The inevitable application of big data to health care. *JAMA* **2013**, *309*, 1351–1352. [[CrossRef](#)] [[PubMed](#)]
85. Naik, N.; Hameed, B.M.Z.; Shetty, D.K.; Swain, D.; Shah, M.; Paul, R.; Aggarwal, K.; Ibrahim, S.; Patil, V.; Smriti, K.; et al. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Front. Surg.* **2022**, *9*, 862322. [[CrossRef](#)] [[PubMed](#)]