diagnostics

MDPI

Review

AI-Assisted Fracture Detection in Orthopedic and Trauma Imaging: Where It Works, Where It Fails, and Principles for Safe Clinical Deployment

Wojciech Michał Glinkowski [1], Paweł Kamiński [2,3] and Rafał Obuchowicz [4,*]

[1] Center of Excellence "TeleOrto" for Telediagnostics and Treatment of Disorders and Injuries of the Locomotor System, Department of Medical Informatics and Telemedicine, Medical University of Warsaw, 02-091 Warsaw, Poland; w.glinkowski@gmail.com
[2] Clinic of Locomotor Disorders, Andrzej Frycz Modrzewski University, 30-705 Krakow, Poland; pawelkam@mp.pl
[3] Malopolan Orthopedic and Rehabilitation Hospital, Aleja Modrzewiowa 22, 30-224 Krakow, Poland
[4] Department of Diagnostic Imaging, Jagiellonian University Medical College, 31-008 Krakow, Poland
* Correspondence: r.obuchowicz@gmail.com

Abstract

Background: Missed fractures on initial imaging assessments remain a clinically significant source of diagnostic errors in orthopedic and trauma care. AI-assisted imaging tools are increasingly integrated into fracture detection workflows. However, their diagnostic benefits and safety vary substantially across anatomical regions, clinical contexts, and levels of reader experience. Purpose: To synthesize the current evidence on the diagnostic impact of AI-assisted fracture detection and to discuss evidence-informed principles for safe and selective clinical deployment. Methods: A structured narrative synthesis of meta-analyses, multi-reader, multi-case observer studies, and real-world implementation investigations was performed. Diagnostic performance patterns were examined across anatomical regions and levels of reader experience. No quantitative pooling or reanalysis of the primary data was performed. The findings were synthesized across anatomical regions, reader-experience groups, and implementation-relevant clinical contexts. Results: Across studies, AI-assisted interpretation was generally associated with moderate gains in sensitivity and lower missed-fracture rates compared with unaided human reading, while largely preserving specificity. The diagnostic benefit was greatest among less-experienced readers in high-volume emergency settings. Performance was strongly anatomy-dependent: consistent and clinically meaningful improvements were observed for hip and appendicular skeleton fractures; intermediate benefits with increased false-positive burden were reported for wrist and rib fractures; and inferior sensitivity relative to expert interpretation was documented for cervical and vertebral spine injuries. Conclusions: AI-assisted fracture detection improves diagnostic safety when implemented as a structured second-reader tool; however, its effectiveness depends heavily on anatomy. Available evidence supports selective, risk-stratified deployment, guided by anatomy-specific risk considerations and supervised clinical use, rather than indiscriminate or autonomous use, to maximize benefits and minimize patient safety risks in orthopedic and trauma imaging.

Keywords: artificial intelligence; fracture detection; musculoskeletal imaging; diagnostic accuracy; clinical governance; risk stratification; implementation science; second-reader systems; emergency radiology; orthopedic trauma

Check for updates

Academic Editor: Fahmi Khalifa

Received: 24 February 2026
Revised: 1 May 2026
Accepted: 4 May 2026
Published: 7 May 2026

Copyright: © 2026 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.

1. Introduction

Missed fractures during the initial imaging assessment remain a significant and clinically consequential challenge in orthopedics and trauma care [1,2]. Undetected injuries can lead to delayed or inappropriate treatment, prolonged morbidity, and increased medicolegal risk, particularly in high-throughput emergency settings, where time pressure, variable reader expertise, and subtle imaging findings increase the likelihood of diagnostic oversight [3–8].

Recent advances in artificial intelligence (AI), particularly deep learning-based image analysis, have accelerated the development of fracture detection tools, most commonly for plain radiographs [9–12]. Several commercial systems have been approved for clinical use, and early studies suggest that AI assistance can improve fracture detection when used to support, rather than replace, human readers, typically as a structured second reader or decision-support system [13–29].

Recent systematic reviews and meta-analyses suggest that stand-alone AI systems can achieve high diagnostic accuracy in radiographic fracture detection, in some settings, approaching expert-level performance and often exceeding that of non-specialist readers [11,21,30–32].

However, this benefit is inconsistent across clinical contexts. AI assistance appears to provide the most consistent and clinically meaningful gains when used as an assistive second reader, particularly for hip and appendicular skeletal fractures and among less-experienced readers in emergency settings [10,21,33–36].

In contrast, the performance is weaker and more variable in anatomically complex or subtle injury patterns, including cervical and vertebral spine fractures, multiple fractures, and selected avulsion injuries [10,36–40]. Recent reviews have summarized overall diagnostic accuracy well, whereas reader-experience effects, anatomy-specific variation, and implementation-relevant clinical considerations have been addressed less consistently across the literature [11,21,41].

The key clinical question is not merely whether AI can detect fractures, but rather where, in whom, and under what conditions it can reduce the rate of missed injuries without introducing unacceptable diagnostic risks. Accordingly, this review aims to: summarize reported diagnostic patterns of AI-assisted fracture detection, examine how these patterns vary by anatomical region and reader experience, and discuss selected considerations relevant to supervised clinical implementation. The primary focus of this review is fracture detection on plain radiographs, where AI is most commonly proposed as a second-reader or decision-support tool in initial trauma assessment. Additional non-radiographic evidence is considered only where directly relevant to reference-standard interpretation or clinically indicated escalation of diagnostic work-up.

2. Materials and Methods

2.1. Study Design

This study was designed as a structured narrative review of AI–assisted fracture detection in orthopedic and trauma imaging, with a primary focus on radiograph-based diagnostic support. A formal systematic review or meta-analysis was not performed because the available literature is highly heterogeneous with respect to anatomical targets, clinical settings, reader populations, AI tools, reference standards, and reported outcome measures. Accordingly, a structured narrative approach was used to identify clinically relevant patterns in diagnostic performance, reader-experience-dependent effects, anatomy-specific variation, and implementation-related risk. Additional non-radiographic studies were considered only when directly relevant to reference-standard interpretation in spinal fracture research or to clinically indicated escalation imaging for occult fractures. The core

review set was defined by the structured search strategy and eligibility criteria described below. Earlier or non-core references were cited selectively only for contextual framing, methodological background, or clinically relevant interpretation and were not treated as part of the core synthesis dataset.

2.2. Literature Search

A literature search was conducted in PubMed, Web of Science Core Collection, Scopus, ScienceDirect, and Google Scholar to identify studies that evaluated AI-assisted fracture detection on plain radiographs in clinically relevant musculoskeletal imaging settings. The searches were restricted to articles published from 1 January 2021, to 31 March 2026, and were limited to English-language articles and reviews. An initial exploratory PubMed search using a core combination of terms ("artificial intelligence" AND fracture AND x-ray/radiograph, Title/Abstract) yielded 178 records. A more clinically focused PubMed strategy incorporating diagnostic accuracy and implementation terms (e.g., sensitivity, specificity, observer, multi-reader, reader performance, implementation, workflow, decision support, and second reader) identified 325 records prior to screening and deduplication. Parallel searches with analogous clinically focused strategies were performed in the Web of Science Core Collection, Scopus, and ScienceDirect databases. Only articles reporting fracture-level diagnostic outcomes on plain radiographs were retained after manual screening. Google Scholar was used only as a supplementary tool for citation chasing and identifying recent or ahead-of-print articles, not as a primary search platform. The purpose of the search was to identify representative and methodologically informative evidence for structured narrative synthesis rather than to generate an exhaustive dataset for a quantitative pooling analysis. These searches were intended to identify a clinically informative core review set rather than an exhaustive corpus for formal quantitative synthesis. The full database-specific search strategies, including field restrictions and limits, are provided in Supplementary Table S1.

2.3. Eligibility Criteria and Study Selection

Studies were eligible if they evaluated AI systems for fracture detection in clinically relevant musculoskeletal imaging settings and reported diagnostic or workflow outcomes informative for clinical interpretation or implementation. Eligible designs included systematic reviews, meta-analyses, multi-reader multi-case observer studies, retrospective and prospective cohort studies, and real-world implementation studies. Priority was given to studies comparing AI-assisted human interpretation with unaided human reading, assessing stand-alone AI performance against human readers or reference standards, or examining deployment-related outcomes in emergency, trauma, or orthopedic imaging. Studies focused exclusively on technical model development without clinically interpretable diagnostic outcomes, non-fracture tasks, non-musculoskeletal applications, duplicate publications, and conference abstracts lacking sufficient methodological or outcome data were excluded from the review. Titles and abstracts were screened for relevance by W.M.G., with independent verification by R.O. and P.K. Full-text articles considered potentially eligible were assessed against the predefined inclusion criteria. Any uncertainties regarding study eligibility were resolved through discussion and consensus. Consistent with the structured narrative design, the study selection process prioritized clinical relevance, methodological informativeness, and applicability to real-world fracture diagnoses.

2.4. Data Extraction

Data extraction was performed using a predefined framework. The extracted variables included study design, clinical setting, anatomical region, imaging modality, AI system characteristics, mode of AI use (stand-alone versus assistive), reader characteristics and

level of experience, reference standard, diagnostic performance measures, and selected workflow-related outcomes. The diagnostic performance variables included sensitivity, specificity, area under the receiver operating characteristic curve (AUC), missed fracture rates, interpretation time, and inter-reader agreement. Initial data extraction was performed by W.M.G. and independently verified by R.O. and P.K. Any discrepancies were resolved through discussion and consensus among the authors. Reported values were extracted as published; no recalculation, statistical reweighting, or de novo meta-analytic synthesis of primary data was performed.

2.5. Qualitative Assessment of Bias and Heterogeneity

A formal study-level quality scoring tool was not applied because the included evidence comprised heterogeneous study designs, including meta-analyses, multi-reader observer studies, retrospective and prospective cohort studies, and real-world implementation studies. Instead, potential sources of bias and methodological limitations were qualitatively considered during the synthesis, informed by established frameworks for diagnostic accuracy studies, including QUADAS-2 and emerging AI-specific reporting guidance [42,43]. Particular attention was paid to selection bias, spectrum bias, verification bias, enriched fracture prevalence in observer studies, heterogeneity of reference standards, and differences between stand-alone algorithm evaluation and clinically embedded reader-assisted studies. Additional attention was given to the radiograph-versus-CT reference-standard mismatch in spinal fracture studies, as this may affect the interpretability of performance estimates derived from radiograph-based models trained or validated against CT-confirmed outcomes. These considerations informed the interpretation of the apparent diagnostic gains and the formulation of conclusions regarding anatomy-specific implementation.

2.6. Structured Narrative Synthesis and Table Construction

Because of substantial heterogeneity across anatomical domains, study designs, AI deployment modes, reader expertise, and reported outcomes, no pooled effect estimates, meta-regression analyses, or indirect comparative rankings of individual AI systems were performed. Instead, the findings were synthesized narratively across prespecified clinically relevant domains. The synthesis was organized along three primary axes:

(1) The overall diagnostic impact of AI assistance compared with unaided human readers and stand-alone AI systems; (2) reader-experience-dependent effects of AI support; and (3) anatomy-specific patterns of diagnostic benefit, uncertainty, and potential clinical risk.

The diagnostic performance findings were summarized using reported ranges and directional patterns rather than pooled summary estimates. The quantitative values presented in the text and tables were intended to reflect published findings from representative studies and reviews, not meta-analytic summary measures generated de novo by the authors. For anatomy-specific summary tables, the consistency of evidence was qualitatively assessed based on concordance across study designs, the direction of observed effects, and the breadth of supporting evidence. “High consistency” indicated broadly concordant findings across multiple studies or reviews; “moderate consistency” indicated generally similar directional findings with some heterogeneity; and “low consistency” indicates sparse, conflicting, or methodologically limited evidence.

Qualitative labels such as “high”, “moderate”, “variable”, or “moderate to strong” were used as interpretive synthesis descriptors based on the direction, consistency, and approximate magnitude of reported effects across the included studies; they should not be interpreted as formal evidence grades or standardized quantitative categories.

This synthesis approach has inherent methodological constraints. As a structured narrative review, it does not provide the same level of reproducibility as a formally reg-

istered systematic review with exhaustive database-specific search updating and formal study-level risk-of-bias scoring. In addition, the underlying literature is methodologically heterogeneous, which limits direct comparability across anatomical regions, clinical settings, and study designs. Some of the included evidence also derives from retrospective observer studies or enriched test sets that may not fully reflect real-world fracture prevalence or workflow conditions. Accordingly, the present synthesis was designed to prioritize clinical interpretability, implementation relevance, and diagnostic safety over formal quantitative aggregation. This approach is consistent with contemporary guidance on narrative evidence synthesis in complex and heterogeneous studies, which emphasizes the explicit justification of the review design and the transparent reporting of analytic procedures [44].

3. Results

3.1. Evidence Base and Study Landscape

The included literature comprised a heterogeneous yet clinically informative evidence base, including systematic reviews and meta-analyses of diagnostic test accuracy, multi-reader and multi-case (MRMC) observer studies, retrospective and prospective cohort studies, and real-world emergency department (ED) implementation studies. Across these investigations, AI systems were evaluated both as stand-alone diagnostic models and, more commonly, as assistive tools integrated into human image interpretation workflows. The evidence covered multiple anatomical regions, with the greatest number of studies involving hip fractures, wrist and hand fractures, and broader appendicular skeletal injuries. The study sizes ranged from small observer series to large retrospective datasets, including hundreds of thousands of radiographs, with reader panels typically ranging from a few to several dozen clinicians. The representative studies included in the structured narrative synthesis are summarized in Supplementary Table S1.

Across study designs, a consistent pattern emerged: the most clinically relevant and reproducible benefit of AI was observed when the technology was used to support human readers, rather than replacing them. Simultaneously, the magnitude and reliability of the diagnostic benefit varied by anatomical region, reader experience, reference standard, and study design, with substantial differences across these factors.

3.2. Overall Diagnostic Performance: Stand-Alone AI, Human Readers, and AI-Assisted Interpretation

Across recent meta-analyses and large observer studies, stand-alone AI systems generally demonstrated high diagnostic accuracy for fracture detection on radiographs, with reported sensitivities and specificities typically in the high-80% to low-90% range, and AUC values frequently exceeding 0.90 [11,21,31,32,45]. However, the stand-alone AI performance was not consistently superior to that of experienced radiologists or musculoskeletal imaging specialists, particularly in anatomically challenging regions or where fracture morphology was subtle, complex, or heterogeneous [11,33,37–39,46].

In contrast, AI-assisted human interpretation showed a more consistent and clinically relevant pattern of benefit. Across reviews and MRMC studies, AI support most commonly improved sensitivity, whereas specificity was usually preserved or changed only minimally [29,35,47–49]. In representative studies, AI assistance was commonly associated with moderate sensitivity gains and lower missed-fracture rates compared with unaided human interpretation [33,34,49–51]. These findings support the interpretation of AI as a diagnostic safety net that helps reduce perceptual oversight during the initial image assessment rather than as a substitute for expert clinical judgment [33,48,51,52]. The overall diagnostic performance patterns are summarized in Table 1. As outlined in Section 2.6,

these values are presented as directional summaries of the published findings rather than as de novo pooled estimates.

Table 1. Qualitative summary of overall diagnostic performance patterns in AI-assisted fracture detection.

Outcome Domain	Typical Reported Pattern	Main Clinical Interpretation	Evidence Base	Supporting References
Sensitivity	Usually improved with AI assistance; gains were commonly moderate to large in reader-assistance settings	AI most often functions as a diagnostic safety net by reducing perceptual oversight during initial image assessment	Meta-analyses; MRMC observer studies; selected implementation studies	[10,11,24,33–35,51,53]
Specificity	Usually preserved or only slightly affected	Sensitivity gains were often achieved without major loss of specificity, although this varied by anatomical region and task complexity	Meta-analyses; MRMC observer studies	[10,11,24,30,32,33,40,51]
Stand-alone AI discrimination	Often high, but not consistently superior to expert readers across all anatomical tasks	Stand-alone AI performance may approximate expert-level accuracy in selected tasks, but remains less reliable in anatomically complex or heterogeneous settings	Meta-analyses; stand-alone diagnostic accuracy studies	[11,21,24,31,32,45,54]
Missed fracture rate	Often reduced when AI is used as an assistive second-reader tool	The most clinically relevant benefit of AI assistance is reduction of overlooked fractures rather than replacement of expert judgment	MRMC studies; ED implementation studies	[10,26,33,35,40,47,49,51,53,55]
Reading time and workflow efficiency	Frequently shortened in observer and implementation studies, although effects vary by workflow and anatomical task	Potential efficiency benefits appear most relevant when AI is integrated into supervised human interpretation workflows	Implementation studies; observer studies; workflow analyses	[10,39,49,56–59]

Footnote: Reported patterns summarize the direction and approximate clinical magnitude of findings across recent meta-analyses, multi-reader multi-case (MRMC) observer studies, and real-world implementation reports. No quantitative pooling, reweighting, or reanalysis of primary data was performed. The table is intended to provide a qualitative overview of cross-study tendencies rather than precise comparative summary estimates.

3.3. Reader-Experience-Dependent Effects

The magnitude of the AI benefit was not uniform across the reader groups. Across multi-reader and subgroup analyses, less-experienced readers—including trainees, junior doctors, emergency physicians, and other nonspecialist clinicians—generally showed the greatest absolute gains in diagnostic performance with AI assistance [33,34,51,60,61]. In these settings, AI support most often improved sensitivity and reduced missed fractures, thereby narrowing, although not eliminating, the gap between inexperienced and expert readers [25,34,48,58,62].

In contrast, the incremental diagnostic benefit of AI was smaller among experienced radiologists and musculoskeletal imaging specialists. In expert readers, the contribution of AI was more often reflected in improved detection of subtle injuries, reduced interpretive variability, and, in some studies, shorter reading time rather than large gains in baseline diagnostic accuracy [33,53,63]. Across studies, the largest gains were generally reported in less-experienced reader groups [33,34,53]. The cross-study patterns by reader category are summarized in Table 2.

Table 2. Reader-experience-dependent effects of AI assistance in fracture detection.

Reader Category	Baseline Diagnostic Performance	Effect of AI Assistance	Typical Clinical Interpretation	Consistency Across Studies	Supporting References
Emergency physicians	Moderate sensitivity; high variability	Large sensitivity gain; substantial reduction in missed fractures	Strong safety-net effect in high-volume ED and acute-care settings	High	[26,33,49,55,58,64,65]
Junior radiology residents	Moderate sensitivity	Large sensitivity gain; improved detection of subtle errors	Compensation for limited experience; support for standardization	High	[10,26,40,47,53,55,66]
General radiologists	High sensitivity	Small to moderate improvement	Reduction in overlooked subtle findings; modest workflow support	Moderate	[10,11,33,35,49,67,68]
Musculoskeletal radiologists	Very high sensitivity	Minimal absolute gain; occasional subtle benefits	Main value is efficiency and detection of rare or easily missed lesions	Moderate	[10,35,47,67,69,70]

Footnote: Reader categories and effect magnitudes were synthesized qualitatively from subgroup analyses and comparative MRMC observer studies. The table reflects the direction and consistency of reported findings rather than pooled quantitative estimates.

3.4. Anatomical Region-Specific Diagnostic Performance

The diagnostic performance varied substantially across anatomical regions, and this variation was one of the findings of this review. The most consistent benefits of AI assistance were observed in appendicular skeletal imaging, including wrist, hand, ankle, elbow, knee, and long bone fractures. In these domains, AI-assisted readers generally achieve improved sensitivity with little or no major loss of specificity. In contrast, the performance for rib and spinal fractures was less favorable and more variable, with smaller or inconsistent gains in sensitivity, greater loss of specificity, and more interpretive noise, underscoring the need for greater caution in these regions [24,35,68,71–74].

Across studies, AI-assisted detection of hip and proximal femur fractures showed consistently strong diagnostic performance. Across studies, reported sensitivities and specificities for hip fracture detection were typically in the high-80% to low-90% range, with AUCs around 0.91–0.99, and several reports suggested that AI support could help less-experienced readers approach the performance of orthopedic or musculoskeletal imaging specialists [54,75–77]. However, occult or minimally displaced hip fractures remain an important limitation, and escalation of imaging may still be required in selected cases.

Wrist and hand fractures also showed strong performance, although with greater variability than hip fractures. AI assistance is particularly useful for detecting subtle or minimally displaced injuries. However, in some studies, this benefit was accompanied by an increased rate of false-positive prompts, especially in clinically ambiguous cases [68,78–80].

Rib fractures exhibited a more mixed pattern. Several studies have suggested that AI may improve sensitivity in emergency and after-hours settings; however, this benefit is sometimes offset by reduced specificity and increased interpretive noise [24,52,64,74].

In contrast, axial skeletal fracture detection was less reliable. The weakest performance was reported in radiograph-based detection of cervical spine trauma and vertebral fractures, where AI sensitivity was lower and less consistent than that of attending radiologists in representative studies [37,38,67]. Thoracic and lumbar vertebral fracture tasks also showed variable performance, indicating that spinal fracture detection should not be treated as a single homogeneous category. Across studies, spinal fracture detection showed lower and

less consistent performance than hip or appendicular radiography. The anatomy-specific benefit–risk profiles of AI-assisted fracture detection are summarized in Table 3.

Table 3. Qualitative synthesis of anatomy-specific patterns in AI-assisted fracture detection.

Anatomical Region	Stand-Alone AI Performance (Qualitative)	AI-Assisted Reader Performance (Qualitative)	Predominant Reported Benefit	Key Risks/Limitations	Evidence Consistency	Supporting References
Hip/proximal femur	High	Very high	Marked sensitivity gains and fewer missed fractures on radiographs	Occult or minimally displaced fractures may still require escalation of imaging in selected cases	High	[54,75–77]
Appendicular skeleton (general)	High	High	Consistent sensitivity gains across long bones and extremities	Performance varies across AI systems and specific fracture subtypes	High	[24,33,35,40,66]
Wrist/hand	Moderate to high	High	Improved detection of subtle and minimally displaced fractures	Increased false-positive prompts and greater interpretive noise in complex cases	Moderate	[68,78–80]
Rib fractures	Moderate	Moderate to high	Improved sensitivity in emergency and after-hours settings	Reduced specificity and possible downstream escalation to additional imaging, including CT in selected cases	Moderate	[24,52,64,74]
Cervical spine trauma on radiographs	Low	Low to moderate	Limited or selective benefit in some settings	Anatomical overlap, incomplete visualization, CT-reference mismatch, and high-risk false negatives	Low	[37,38,67,73]
Thoracic/lumbar vertebral fracture tasks	Low to moderate	Variable	Possible support in selected vertebral fracture tasks	Morphological heterogeneity, inconsistent performance, and non-equivalence to cervical trauma detection	Low to moderate	[24,71,73,81,82]

Footnote: The categories shown in the table are qualitative synthesis labels assigned by the authors based on concordance across reviews, observer studies, and implementation reports. They summarize the predominant patterns in the literature and do not constitute formal evidence grading or comparative ranking of individual AI systems.

3.5. Subtle and Occult Fractures: Diagnostic Gains and Trade-Offs

AI assistance was reported to be valuable in detecting subtle, non-obvious, and often overlooked fractures. Across studies, the relative diagnostic benefit of AI was often greatest for minimally displaced injuries, avulsion fractures, and other lesions prone to perceptual oversight during routine initial reading, with typical sensitivity gains of approximately 10–15 percentage points reported in these subgroups [65,83,84].

This pattern was also observed in the selected task-specific studies. For example, in a multicenter study of extremity radiographs, lesion-level detection of avulsion fractures was higher with the optimized AI model than that of radiologists (57.9% vs. 29.8%). In contrast, no significant difference was observed between non-avulsion fractures [85].

However, these gains were not without cost. Improved sensitivity to subtle findings is frequently accompanied by an increase in false-positive suggestions, particularly in the wrist, scaphoid, and other anatomically complex or low-signal contexts [24,70,86].

The benefit–risk balance across subtle and occult fracture patterns is presented in Table 4.

Table 4. Qualitative synthesis of benefit–risk patterns in AI assistance for subtle and commonly overlooked fractures.

Fracture Pattern	Typical AI-Related Gain	Main Diagnostic Benefit	Main Trade-Off	Suggested Clinical Safeguard	Supporting References
Minimally displaced fractures	Moderate	Improved detection of subtle cortical disruption	Increased interpretive noise and false positives	Human adjudication and local threshold calibration, where applicable	[65,83,84]
Avulsion fractures	Moderate to strong	Better lesion-level sensitivity for small osseous fragments	Overcalling anatomical variants or artifacts	Correlation with anatomy and clinical findings	[65,84,86]
Occult or initially overlooked fractures	Positive but variable	Increased vigilance during first-pass reading	More false-positive prompts and follow-up imaging	Structured second-reader workflow and escalation imaging when clinically indicated	[21,24,51,65,86]

Footnote: Patterns and effect magnitudes were qualitatively synthesized from subgroup analyses and task-specific studies focusing on subtle and occult fractures. The reported gains and trade-offs illustrate typical tendencies rather than pooled quantitative estimates. Qualitative gain descriptors reflect interpretive judgments of synthesis rather than standardized effect-size categories.

3.6. Workflow and Efficiency Outcomes

In addition to diagnostic accuracy, several studies have reported the workflow-related effects of AI assistance. Across observer and early implementation studies, AI support was associated with reductions in image interpretation time and improvements in inter-reader agreement. However, the magnitude of these effects varies substantially across anatomical regions and study designs [33,51,69,87].

Importantly, these efficiency gains were generally observed without major reductions in diagnostic accuracy in the reported studies. The available evidence remains heterogeneous, and the clinical relevance of faster interpretation depends on the balance between speed and diagnostic safety [10,51,71]. Several studies reported shorter interpretation times and, in some cases, improved inter-reader agreement when AI was used within supervised human workflows [59,66].

4. Discussion

4.1. Principal Findings

This review suggests that the clearest and most reliable clinical value of AI-assisted fracture detection lies in supporting human readers, rather than replacing them. Across a heterogeneous body of evidence, the most consistent and clinically relevant benefit was a reduction in missed fractures at the initial imaging assessment [10,51,57,71]. This pattern was especially evident in radiograph-based workflows and among less-experienced readers working in busy orthopedic, trauma, or emergency settings [35,61].

Simultaneously, the effect of AI assistance is context-dependent. Anatomical region and reader experience emerged as the main factors shaping both benefits and risks [13,24,34,39,51,60,88,89]. Therefore, the practical question is not whether AI works for fracture detection in general, but where it performs well enough to justify implementation, where the gains are modest, and where the limitations may introduce unacceptable diagnostic risk [51,57,71,90]. This anatomy- and context-specific variation is the central message of this review.

4.2. AI as a Diagnostic Safety Net Rather than an Autonomous Reader

A recurring finding across reviews, MRMC observer studies, and early implementation reports is that AI assistance tends to improve fracture detection primarily by increasing sensitivity while maintaining acceptable specificity [10,11]. This was most apparent in readers with limited experience in musculoskeletal imaging. In practical terms, AI seems to function best as a second reader or structured decision-support tool that helps reduce perceptual oversight and supports a more systematic review of radiographs under time pressure [11,35,51].

In contrast, stand-alone AI performance, although often strong on selected tasks, is generally comparable to, rather than consistently better than, expert radiologists or musculoskeletal imaging specialists [11,24,32,69]. Therefore, the real diagnostic value comes from human–AI collaboration rather than autonomous decision-making [51,71,90]. This has direct implications for deployment strategy, workflow design, and medico-legal responsibility and is broadly consistent with current regulatory and ethical expectations for clinical AI [71,91–93].

4.3. Reader Experience Modifies the Clinical Value of AI

One of the most clinically important findings of this review is that the benefit of AI is strongly influenced by reader experience. Less-experienced readers, including trainees, junior doctors, emergency physicians, and other non-specialists, generally derive the greatest benefit, most often through higher sensitivity and fewer missed fractures [26,35,55]. In practical terms, AI helps narrow the gap between inexperienced and expert readers, even if it does not completely close that gap [34,35,51,53,55].

Among experienced radiologists and musculoskeletal imaging specialists, the incremental benefit is usually smaller and more selective. In these settings, AI may still add value by reducing interpretive variability, improving vigilance for subtle findings, or improving efficiency; however, it is less likely to produce large absolute gains in baseline diagnostic accuracy [45,49,51,57]. This pattern supports a selective implementation strategy in which the rationale for AI use varies across institutions and reader groups, with the strongest case in high-volume environments staffed predominantly by less-experienced readers [24,94,95].

This experience-dependent pattern extends to pediatric practices. A recent systematic review and meta-analysis focusing on pediatric fracture detection on radiographs reported a pooled sensitivity of 93% (95% CI 92–94%), specificity of 91% (95% CI 88–93%), and an AUC of 0.96 (95% CI 0.92–0.97), with performance maintained in external validation cohorts [28]. At the same time, the authors highlighted substantial heterogeneity in anatomical coverage, case mix, and reader experience. They noted that many contributing datasets were single-center and retrospectively enriched, which may limit the direct generalizability of these pooled estimates across all emergency departments [28,96].

4.4. Anatomical Region as the Primary Factor Determining Benefits and Risks

The most important factor influencing the utility of AI for fracture diagnosis is the anatomical region under examination. Available evidence generally indicates a favorable benefit-to-risk ratio for skeletal imaging of the extremities and the detection of proximal femoral fractures [24,32,46,51,54,97]. In these anatomical regions, AI support most consistently improves sensitivity, reduces the number of missed fractures, and appears suitable for supervised routine use in trauma and emergency imaging [10,35,40,46,69,79]. Currently, these are the most mature and clinically justified applications of AI-assisted fracture detection [11,24,51,54].

However, anatomical regions differ not only in their inherent diagnostic difficulty but also in the clinical consequences of error [14,20,65]. In some situations, a moderate

increase in false-positive suggestions may be acceptable if the primary priority is reducing missed injuries. In other cases, especially where both the false reassurance of no injury and unnecessary escalation have serious consequences, the same AI behavior may be far less acceptable. Rib fractures occupy an intermediate position, where increased sensitivity is often offset by reduced specificity and greater interpretive noise [52,64,74]. Cervical and thoracolumbar spine injuries remain a high-risk area where current AI based on radiographs shows limited and inconsistent benefits compared to expert radiologists [38,71,98,99].

Similar difficulties are observed in cases of complex, multi-site, comminuted, displaced fractures. In a large-scale evaluation of commercial algorithms in real-world settings, performance for acute single fractures remained within the expected range, but dropped significantly in studies involving multiple fractures, highlighting the risk of overestimating AI reliability in polytrauma scenarios [24,39,51,59,65]. Similarly, although overall fracture detection accuracy in children is generally high, lesion-level sensitivity may still be lower for certain subtle injury subtypes, including specific avulsion fracture patterns [28,36,95,96]. In summary, the results reported in the literature suggest that AI is currently most reliable for typical tasks involving single fractures [11,24,33,51], whereas rare, multi-site, or difficult-to-diagnose injury patterns remain higher-risk applications for AI, requiring close expert supervision [36,39,59,98].

Accordingly, implementation decisions should be guided not only by overall accuracy metrics but also by anatomy-specific risk tolerance and the clinical consequences of missed or overcalled injuries.

4.5. Why Spinal Fracture AI Remains Problematic

The weakest and most clinically concerning performance pattern identified in this review involved spinal fracture detection, particularly radiograph-based assessments of traumatic cervical spine injuries. This should not be viewed solely as a governance or deployment issue. More fundamentally, it likely reflects structural imaging limitations, fracture heterogeneity, and a mismatch with reference standards, all of which make spinal fracture detection intrinsically more difficult for current AI systems than hip or appendicular fracture tasks [17,38,73,98,100–102].

Compared with appendicular radiography, cervical spine radiography is more challenging because of anatomical overlap, incomplete visualization, projection-dependent ambiguity, and the possibility that clinically important fractures may be occult or only subtly visible on plain films [73,100,103–105]. These features complicate both human interpretation and AI-based detection. In addition, spinal fractures are morphologically heterogeneous, spanning traumatic cervical injuries, thoracolumbar trauma, and osteoporotic vertebral compression fractures [71,81,99,103,106,107]. These tasks are not interchangeable and should not be treated as a single category.

Reference standard mismatch is another major issue. Many spinal fractures, especially in the cervical spine, are confirmed on CT despite being only subtly visible or not confidently visible on radiographs [3,103,105]. When radiograph-based AI systems are trained or validated against CT-confirmed endpoints, the target may be clinically relevant but methodologically problematic, as the model is effectively asked to infer findings that may not be fully resolvable from the input modality [108–110]. This likely contributes to the lower and less reliable performance of spinal fracture AI compared with hip and appendicular applications. For these reasons, traumatic cervical spine fracture detection on radiographs should still be regarded as a high-risk, lower-maturity use case that requires substantial caution. Vertebral fracture detection, including osteoporotic compression fractures, should be considered a related but distinct task with different imaging characteristics, clinical aims, and implementation implications [24,71,81,82].

4.6. Subtle Fractures: Benefit and Diagnostic Noise

AI assistance is particularly advantageous for detecting subtle, minimally displaced, and often overlooked fractures [35,51]. In these cases, its role is not to confirm obvious findings but to heighten awareness of abnormalities that might otherwise be missed during routine evaluations. AI-assisted fracture detection may be particularly helpful for subtle injuries that are at higher risk of being overlooked, but gains in sensitivity can also introduce interpretive noise and false-positive alerts. This tendency appears more pronounced in anatomically complex regions, in the presence of normal variants, and in workflows with limited opportunities for downstream adjudication. In such settings, AI support for subtle fractures is best positioned as a structured second reader, with routine human verification and clinical correlation, rather than as a signal to be accepted uncritically [21,24,26,29,47,51,111]. Table 4 provides a summary of the typical benefit–risk balance across major, subtle, and occult fracture patterns.

4.7. Clinical Governance and Implementation Considerations

The findings of this review support selective and supervised clinical implementation. In practical terms, safe implementation requires local validation, anatomy-specific deployment criteria, explicit clinician accountability for the final report, and post-deployment monitoring for performance drift or unintended workflow effects [112–114].

These safeguards are especially important because the clinical value of AI depends on more than model performance alone. It is also shaped by reader expertise, case mix, disease prevalence, and how the tool is integrated into reporting workflows [33,35,39,59,94]. Consequently, the same AI system may have very different practical value in a specialized trauma center, a general emergency department, or a lower-volume orthopedic service [33,35,39,59,94,112].

Patient-facing considerations, including transparency of data use, trust, and governance of image transfer or cloud-based infrastructure, are also relevant to implementation. However, because these issues were not systematically addressed across the included studies, they should be regarded as important contextual considerations rather than as conclusions directly established by the present evidence base [112–117].

Available evidence also suggests potential efficiency gains in selected high-volume settings. Still, current data remain insufficient to support strong economic conclusions [56,117], and diagnostic improvements should not be assumed to automatically translate into better patient outcomes [14,45,56,71].

4.8. Limitations of This Review

This review should be interpreted in light of several methodological limitations. First, as a structured narrative review, it does not provide the same degree of reproducibility as a formally documented systematic review, which includes exhaustive search updates, multiple verifications at all stages, and a formal assessment of the risk of bias at the individual study level. Therefore, the selection and synthesis of studies inevitably involved some degree of subjective judgment by the authors, which raises the possibility of selection bias and subjective interpretation.

Second, the evidence on which this review is based is heterogeneous in several important respects. The included literature comprises systematic reviews, multi-observer studies, retrospective and prospective cohort studies, and early implementation studies, which differ significantly in anatomical scope, case diversity, readers' expertise, AI implementation mode, and reference standards. This heterogeneity limits direct comparability across anatomical regions and clinical settings, diminishing the strength of any uniform cross-domain conclusions. The included evidence also spans different stages of techno-

logical and translational maturity, which further limits the stability of broad cross-study conclusions about current clinical usefulness.

Third, a significant portion of the available evidence remains dominated by retrospective studies and enriched datasets, in which fracture incidence and case composition may differ significantly from routine practice in emergency or trauma care. This may overestimate apparent diagnostic performance and limit the generalizability of reported benefits to real-world practice. Furthermore, differences in reference standards across studies introduce additional uncertainty in interpretation, particularly for anatomically challenging tasks such as detecting spinal fractures.

Fourth, the quantitative ranges presented in this review should be considered a descriptive and indicative summary of the included literature, rather than precise comparative estimates. Given the heterogeneity of study designs, anatomical tasks, reference standards, reader groups, and reported outcomes, direct quantitative comparisons between studies remain limited. These limitations do not negate the value of the review but require cautious interpretation. This synthesis aimed to identify clinically meaningful patterns regarding diagnostic benefits, risks, and implementation implications, rather than to create formal comparative rankings of AI systems or anatomy-specific quantitative summary measures.

Finally, the review was limited to English-language studies published between January 1, 2021, and March 31, 2026, focusing primarily on fracture detection based on X-ray images. Although this scope was deliberately chosen to address the main clinical question regarding AI as a second-reader tool during the initial assessment of trauma, it may have excluded earlier studies, studies in languages other than English, or studies not based on X-ray images, which could have provided additional contextual perspective.

5. Future Research

Future studies should not focus solely on reconfirming the high diagnostic accuracy of AI, but rather on its practical application in clinical practice, its limitations across anatomical areas, and its impact on patient outcomes.

First, prospective, realistic studies are needed to determine whether the use of AI actually improves diagnostic safety, workflow, and follow-up care for patients. Results obtained in retrospective studies do not always translate into real clinical benefits.

Second, detecting spinal fractures requires more precise studies. A clear distinction must be made between the detection of post-traumatic cervical spine fractures and the diagnosis of vertebral body fractures, such as osteoporotic fractures, as these are different diagnostic challenges. Better-suited datasets, clear task definitions, and robust reference standards are necessary.

Third, future studies should more closely examine how AI performance depends on the clinician's experience, the types of cases, and the work environment. The greatest benefits may be seen when working with clinicians in training, non-specialists, or in high-workload settings such as emergency departments—but this requires confirmation in prospective studies.

Fourth, greater attention should be paid to approaches that combine different data sources, particularly in diagnostically challenging cases, such as spinal injuries. Combining data from X-rays and computed tomography or integrating imaging data with clinical information can help overcome the limitations of individual methods.

Finally, the evaluation of AI implementation should consider not only the algorithm's accuracy but also implementation costs and challenges, oversight requirements, and the actual impact on clinical practice.

6. Conclusions

AI systems for fracture detection are of greatest clinical value when they support the physician rather than replace their judgment. Available studies show that the greatest benefit lies in reducing the number of missed fractures on X-rays, particularly in high-workload settings and among less-experienced physicians.

AI performance varies across anatomical regions. The best and most consistent results are seen in detecting proximal femur and limb fractures, where AI improves diagnostic sensitivity without compromising overall accuracy. For spinal fractures, the situation is more complex—current systems are less effective, mainly due to imaging challenges, the wide variety of fractures, and differences between X-ray and computed tomography images. These limitations stem from the very nature of the problem, not just from implementation issues, which define the current boundaries of AI applications in this field.

The physician's experience is also significant. AI better supports less-experienced users, whereas for experts, its role is more auxiliary—it aids in detecting subtle changes, increases vigilance, and helps maintain consistency in assessment, but does not lead to a significant improvement in basic diagnostic accuracy.

In summary, the available data indicate that AI should be implemented selectively, depending on the anatomical area, and always under clinical supervision. The key issue is not whether to use AI everywhere, but where it actually improves diagnostic safety in a repeatable and clinically acceptable manner. Currently, the most justified model is the use of AI as a "second reader" or decision-support tool, operating within the framework of responsible interpretation by a physician.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/diagnostics16101420/s1, Supplementary Table S1.

Author Contributions: Conceptualization, W.M.G., P.K., and R.O.; methodology, W.M.G., P.K., and R.O.; software, W.M.G., P.K., and R.O.; validation, W.M.G., P.K., and R.O.; formal analysis, W.M.G., P.K., and R.O.; investigation, W.M.G., P.K., and R.O.; resources, W.M.G.; data curation, W.M.G., P.K., and R.O.; writing—original draft preparation, W.M.G. and R.O.; writing—review and editing, W.M.G., P.K., and R.O.; visualization, W.M.G., P.K., and R.O.; supervision, W.M.G.; project administration, W.M.G.; funding acquisition, R.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing does not apply to this article.

Conflicts of Interest: The authors declare that they have no affiliations with or involvement in any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

References

1. Ha, A.S.; Porrino, J.A.; Chew, F.S. Radiographic Pitfalls in Lower Extremity Trauma. *Am. J. Roentgenol.* **2014**, *203*, 492–500. [CrossRef] [PubMed]
2. Kim, Y.W.; Mansfield, L.T. Fool Me Twice: Delayed Diagnoses in Radiology With Emphasis on Perpetuated Errors. *Am. J. Roentgenol.* **2014**, *202*, 465–470. [CrossRef]
3. Pinto, A.; Berritto, D.; Russo, A.; Riccitiello, F.; Caruso, M.; Belfiore, M.P.; Papapietro, V.R.; Carotti, M.; Pinto, F.; Giovagnoni, A.; et al. Traumatic fractures in adults: Missed diagnosis on plain radiographs in the Emergency Department. *Acta Biomed.* **2018**, *89*, 111–123. [CrossRef]

4. Pinto, A.; Reginelli, A.; Pinto, F.; Lo Re, G.; Midiri, F.; Muzj, C.; Romano, L.; Brunese, L. Errors in imaging patients in the emergency setting. *Br. J. Radiol.* **2016**, *89*, 20150914. [CrossRef]
5. Guly, H.R. Diagnostic errors in an accident and emergency department. *Emerg. Med. J.* **2001**, *18*, 263–269. [CrossRef]
6. Deakin, A.; Schultz, T.J.; Hansen, K.; Crock, C. Diagnostic error: Missed fractures in emergency medicine. *Emerg. Med. Australas.* **2015**, *27*, 177–178. [CrossRef] [PubMed]
7. Weber, M.A. Easily missed pathologies of the musculoskeletal system in the emergency radiology setting. *Rofo* **2025**, *197*, 277–287. [CrossRef]
8. Wei, C.-J.; Tsai, W.-C.; Tiu, C.-M.; Wu, H.-T.; Chiou, H.-J.; Chang, C.-Y. Systematic analysis of missed extremity fractures in emergency radiology. *Acta Radiol.* **2006**, *47*, 710–717. [CrossRef]
9. Olczak, J.; Fahlberg, N.; Maki, A.; Razavian, A.S.; Jilert, A.; Stark, A.; Sköldenberg, O.; Gordon, M. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop.* **2017**, *88*, 581–586. [CrossRef] [PubMed]
10. Guermazi, A.; Tannoury, C.; Kompel, A.J.; Murakami, A.M.; Ducarouge, A.; Gillibert, A.; Li, X.; Tournier, A.; Lahoud, Y.; Jarraya, M.; et al. Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence. *Radiology* **2022**, *302*, 627–636. [CrossRef]
11. Kuo, R.Y.L.; Harrison, C.; Curran, T.A.; Jones, B.; Freethy, A.; Cussons, D.; Stewart, M.; Collins, G.S.; Furniss, D. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology* **2022**, *304*, 50–62. [CrossRef]
12. Gan, K.; Xu, D.; Lin, Y.; Shen, Y.; Zhang, T.; Hu, K.; Zhou, K.; Bi, M.; Pan, L.; Wu, W.; et al. Artificial intelligence detection of distal radius fractures: A comparison between the convolutional neural network and professional assessments. *Acta Orthop.* **2019**, *90*, 394–400. [CrossRef]
13. Link, T.M.; Pedoia, V. Using AI to Improve Radiographic Fracture Detection. *Radiology* **2022**, *302*, 637–638. [CrossRef]
14. Luo, G.; Tan, S.; Luo, L.; Hu, K. Artificial intelligence and multimodal imaging in orthopaedics: From technological advances to clinical translation. *Front. Med.* **2026**, *12*, 1728248. [CrossRef]
15. Alwzwazy, H.A.; Alzubaidi, L.; Zhao, Z.; Gu, Y. FracNet: An end-to-end deep learning framework for bone fracture detection. *Pattern Recognit. Lett.* **2025**, *190*, 1–7. [CrossRef]
16. Abdellatif, N.A.M.; El-Rawy, A.S.; Abdellattif, A.R.; Al-Shatouri, M.A. Assessment of artificial intelligence-aided X-ray in diagnosis of bone fractures in emergency setting. *Egypt. J. Radiol. Nucl. Med.* **2025**, *56*, 153. [CrossRef]
17. AlSamhori, J.F.; Abdllah, A.N.; Fraihat, A.; Duncan, L.A.; Enayah, M.; Asha, S.Y.; Alhamwi, N.; Abujudeh, R.; Nashwan, A.J. The implication of artificial intelligence in radiographic evaluation of trauma and acute care. *Avicenna* **2025**, *2025*, 12. [CrossRef]
18. Bhatnagar, A.; Kekatpure, A.L.; Velagala, V.R.; Kekatpure, A. A Review on the Use of Artificial Intelligence in Fracture Detection. *Cureus* **2024**, *16*, e58364. [CrossRef]
19. Breitwieser, M.; Zirknitzer, S.; Poslusny, K.; Freude, T.; Scholsching, J.; Bodenschatz, K.; Wagner, A.; Hergan, K.; Schaffert, M.; Metzger, R.; et al. AI in Fracture Detection: A Cross-Disciplinary Analysis of Physician Acceptance Using the UTAUT Model. *Diagnostics* **2025**, *15*, 2117. [CrossRef]
20. Breu, R.; Avelar, C.; Bertalan, Z.; Grillari, J.; Redl, H.; Ljuhar, R.; Quadlbauer, S.; Hausner, T. Artificial intelligence in traumatology. *Bone Jt. Res.* **2024**, *13*, 588–595. [CrossRef]
21. Elkohail, A.; Soffar, A.; Paul, A.; Radu, L.; Ahamed, M.W.S.; Swealem, A.; Sha, A.A.M.A.; Veetil, H.H.M.; Millat, M.S.; Shah, R. Artificial Intelligence in Bone Fracture Detection: A Review of Evidence, Limitations, and Clinical Integration. *Cureus* **2025**, *17*, e97674. [CrossRef]
22. Endler, C.; Luetkens, J.; Nowak, S. Künstliche Intelligenz in der Frakturdiagnostik. *Die Unfallchirurgie* **2025**, *128*, 914–925. [CrossRef]
23. Hassan, A.; Munib, I.N.; Batool, A.; Noor, H. Fracture Detection In X-rays Using Custom Convolutional Neural Network (CNN) And Transfer Learning Models. *arXiv* **2025**, arXiv:2509.06228.
24. Husarek, J.; Hess, S.; Razaeian, S.; Ruder, T.D.; Sehmisch, S.; Müller, M.; Liodakis, E. Artificial intelligence in commercial fracture detection products: A systematic review and meta-analysis of diagnostic test accuracy. *Sci. Rep.* **2024**, *14*, 23053. [CrossRef] [PubMed]
25. Raj, S.; Sadegi, B.; Simon, J. Enhancing Pediatric Fracture Detection: Multicenter Evaluation of a Deep Learning AI Model and Its Impact on Radiologist Performance. *Acad. Radiol.* **2025**, *33*, 1121–1129. [CrossRef]
26. Sadat-Ali, M.; Omar, H.A.; Alneghaimshi, M.M.; AlHossan, A.; Baragabh, A. Role of Artificial Intelligence in Minimizing Missed and Undiagnosed Fractures Among Trainee Residents. *J. Multidiscip. Healthc.* **2025**, *18*, 3851–3858. [CrossRef] [PubMed]
27. Verma, K.; Johari, P.; Vermani, D. Revolutionizing Orthopaedic Diagnostics: An Innovative Deep Learning Framework for Wrist Fracture Detection. In Proceedings of the 2025 Seventh International Conference on Computational Intelligence andCommunication Technologies (CCICT), Sonepat, India, 11–12 April 2025; pp. 429–439.

28. Ximenes, G.F.; Costa, Á.L.; Leite, L.L.; Costa, L.L.; Ribeiro, M.O.; Baima Colares, P.G.; Cerqueira, G.S. Are Artificial Intelligence Models Reliable for Clinical Application in Pediatric Fracture Detection on Radiographs? A Systematic Review and Meta-analysis. *Clin. Orthop. Relat. Res.* **2026**, *484*, 371–385. [CrossRef]
29. 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**, *219*, 869–878. [CrossRef] [PubMed]
30. Zhang, B.; Jia, C.; Wu, R.; Lv, B.; Li, B.; Li, F.; Du, G.; Sun, Z.; Li, X. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: A clinical evaluation. *Br. J. Radiol.* **2021**, *94*, 20200870. [CrossRef]
31. Zhang, X.; Yang, Y.; Shen, Y.-W.; Zhang, K.-R.; Jiang, Z.-k.; Ma, L.-T.; Ding, C.; Wang, B.-Y.; Meng, Y.; Liu, H. Diagnostic accuracy and potential covariates of artificial intelligence for diagnosing orthopedic fractures: A systematic literature review and meta-analysis. *Eur. Radiol.* **2022**, *32*, 7196–7216. [CrossRef]
32. Nowroozi, A.; Salehi, M.A.; Shobeiri, P.; Agahi, S.; Momtazmanesh, S.; Kaviani, P.; Kalra, M.K. Artificial intelligence diagnostic accuracy in fracture detection from plain radiographs and comparing it with clinicians: A systematic review and meta-analysis. *Clin. Radiol.* **2024**, *79*, 579–588. [CrossRef]
33. Duron, L.; Ducarouge, A.; Gillibert, A.; Lainé, J.; Allouche, C.; Cherel, N.; Zhang, Z.; Nitche, N.; Lacave, E.; Pourchot, A.; et al. Assessment of an AI Aid in Detection of Adult Appendicular Skeletal Fractures by Emergency Physicians and Radiologists: A Multicenter Cross-sectional Diagnostic Study. *Radiology* **2021**, *300*, 120–129. [CrossRef] [PubMed]
34. Anderson, P.G.; Baum, G.L.; Keathley, N.; Sicular, S.; Venkatesh, S.; Sharma, A.; Daluiski, A.; Potter, H.; Hotchkiss, R.; Lindsey, R.V.; et al. Deep Learning Assistance Closes the Accuracy Gap in Fracture Detection Across Clinician Types. *Clin. Orthop. Relat. Res.* **2022**, *481*, 580–588. [CrossRef]
35. Bachmann, R.; Gunes, G.; Hangaard, S.; Nexmann, A.; Lisouski, P.; Boesen, M.; Lundemann, M.; Baginski, S.G. Improving traumatic fracture detection on radiographs with artificial intelligence support: A multi-reader study. *BJR Open* **2024**, *6*, tzae011. [CrossRef]
36. Hayashi, D.; Kompel, A.J.; Ventre, J.; Ducarouge, A.; Nguyen, T.; Regnard, N.-E.; Guermazi, A. Automated detection of acute appendicular skeletal fractures in pediatric patients using deep learning. *Skelet. Radiol.* **2022**, *51*, 2129–2139. [CrossRef]
37. van den Wittenboer, G.J.; Nijholt, I.M.; Maas, M.; Boomsma, M.F. How well does artificial intelligence detect fractures in the cervical spine on CT? *Ned. Tijdschr. Geneeskd.* **2024**, *168*, D8194. [PubMed]
38. van den Wittenboer, G.J.; van der Kolk, B.Y.M.; Nijholt, I.M.; Langius-Wiffen, E.; van Dijk, R.A.; van Hasselt, B.; Podlogar, M.; van den Brink, W.A.; Bouma, G.J.; Schep, N.W.L.; et al. Diagnostic accuracy of an artificial intelligence algorithm versus radiologists for fracture detection on cervical spine CT. *Eur. Radiol.* **2024**, *34*, 5041–5048. [CrossRef]
39. Luiken, I.; Lemke, T.; Komenda, A.; Marka, A.W.; Kim, S.H.; Graf, M.M.; Ziegelmayer, S.; Weller, D.; Mertens, C.J.; Bressem, K.K.; et al. Evaluation of commercial AI algorithms for the detection of fractures, effusions, and dislocations on real-world clinical data: A prospective registry study. *Radiography* **2025**, *31*, 103189. [CrossRef]
40. Nguyen, T.; Maarek, R.; Hermann, A.-L.; Kammoun, A.; Marchi, A.; Khelifi-Touhami, M.R.; Collin, M.; Jaillard, A.; Kompel, A.J.; Hayashi, D.; et al. Assessment of an artificial intelligence aid for the detection of appendicular skeletal fractures in children and young adults by senior and junior radiologists. *Pediatr. Radiol.* **2022**, *52*, 2215–2226. [CrossRef] [PubMed]
41. Debs, P.; Fayad, L.M. The promise and limitations of artificial intelligence in musculoskeletal imaging. *Front. Radiol.* **2023**, *3*, 1242902. [CrossRef]
42. Liu, X.; Cruz Rivera, S.; Moher, D.; Calvert, M.J.; Denniston, A.K.; Chan, A.-W.; Darzi, A.; Holmes, C.; Yau, C.; Ashrafian, H.; et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Nat. Med.* **2020**, *26*, 1364–1374. [CrossRef]
43. Whiting, P.F.; Rutjes, A.W.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.; Sterne, J.A.; Bossuyt, P.M. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **2011**, *155*, 529–536. [CrossRef]
44. Sukhera, J. Narrative Reviews: Flexible, Rigorous, and Practical. *J. Grad. Med. Educ.* **2022**, *14*, 414–417. [CrossRef]
45. Kutbi, M. Artificial Intelligence-Based Applications for Bone Fracture Detection Using Medical Images: A Systematic Review. *Diagnostics* **2024**, *14*, 1879. [CrossRef] [PubMed]
46. 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.* **2018**, *48*, 239–244. [CrossRef]
47. Zech, J.R.; Ezuma, C.O.; Patel, S.; Edwards, C.R.; Posner, R.; Hannon, E.; Williams, F.; Lala, S.V.; Ahmad, Z.Y.; Moy, M.P.; et al. Artificial intelligence improves resident detection of pediatric and young adult upper extremity fractures. *Skelet. Radiol.* **2024**, *53*, 2643–2651. [CrossRef]

48. Gasmi, I.; Calinghen, A.; Parienti, J.-J.; Belloy, F.; Fohlen, A.; Pelage, J.-P. Comparison of diagnostic performance of a deep learning algorithm, emergency physicians, junior radiologists and senior radiologists in the detection of appendicular fractures in children. *Pediatr. Radiol.* **2023**, *53*, 1675–1684. [CrossRef] [PubMed]
49. Canoni-Meynet, L.; Verdot, P.; Danner, A.; Calame, P.; Aubry, S. Added value of an artificial intelligence solution for fracture detection in the radiologist's daily trauma emergencies workflow. *Diagn. Interv. Imaging* **2022**, *103*, 594–600. [CrossRef]
50. Raj, M.; Ayub, A.; Pal, A.K.; Pradhan, J.; Varish, N.; Kumar, S.; Varikasuvu, S.R. Diagnostic Accuracy of Artificial Intelligence-Based Algorithms in Automated Detection of Neck of Femur Fracture on a Plain Radiograph: A Systematic Review and Meta-analysis. *Indian. J. Orthop.* **2024**, *58*, 457–469. [CrossRef]
51. Qin, H.; Ding, Y.; Ju, J.; Qu, Z.; Peng, L. Enhanced fracture detection on radiographs with AI assistance for clinicians: A systematic review and meta-analysis. *Ann. Med.* **2026**, *58*, 2610079. [CrossRef] [PubMed]
52. Huang, S.T.; Liu, L.R.; Tsai, M.F.; Huang, M.Y.; Chiu, H.W. Prospective Diagnostic Accuracy and Technical Feasibility of Artificial Intelligence-Assisted Rib Fracture Detection on Chest Radiographs: Observational Study. *JMIR Med. Inf.* **2026**, *14*, e77965. [CrossRef]
53. Lindsey, R.; Daluiski, A.; Chopra, S.; Lachapelle, A.; Mozer, M.; Sicular, S.; Hanel, D.; Gardner, M.; Gupta, A.; Hotchkiss, R.; et al. Deep neural network improves fracture detection by clinicians. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 11591–11596. [CrossRef] [PubMed]
54. Lex, J.R.; Di Michele, J.; Koucheki, R.; Pincus, D.; Whyne, C.; Ravi, B. Artificial Intelligence for Hip Fracture Detection and Outcome Prediction: A Systematic Review and Meta-analysis. *JAMA Netw. Open* **2023**, *6*, e233391. [CrossRef] [PubMed]
55. Meetschen, M.; Salhöfer, L.; Beck, N.; Kroll, L.; Ziegenfuß, C.D.; Schaarschmidt, B.M.; Forsting, M.; Mizan, S.; Umutlu, L.; Hosch, R.; et al. AI-Assisted X-ray Fracture Detection in Residency Training: Evaluation in Pediatric and Adult Trauma Patients. *Diagnostics* **2024**, *14*, 596. [CrossRef]
56. Gregory, L.; Boodhna, T.; Storey, M.; Shelmerdine, S.; Novak, A.; Lowe, D.; Harvey, H. Early Budget Impact Analysis of Artificial Intelligence to Support the Review of Radiographic Examinations for Suspected Fractures in National Health Service Emergency Departments. *Value Health* **2025**, *28*, 1161–1168. [CrossRef]
57. The National Institute for Health and Care Excellence (NICE). *Artificial Intelligence (AI) Technologies to Help Detect Fractures on X-Rays in Urgent Care: Early Value Assessment*; The National Institute for Health and Care Excellence (NICE): Manchester, UK, 2025.
58. Brurberg, K.G.; Kjelle, E.; Vardal, J.; Sivanandan, R. Artificial intelligence in emergency skeletal X-ray: Post-deployment monitoring and clinical impact of incorrect AI results. *Eur. J. Radiol.* **2026**, *199*, 112805. [CrossRef]
59. Prucker, P.; Lemke, T.; Mertens, C.J.; Ziegelmayer, S.; Graf, M.M.; Weller, D.; Kim, S.H.; Gassert, F.T.; Kader, A.; Dorfner, F.J.; et al. Real-world clinical impact of three commercial AI algorithms on musculoskeletal radiography interpretation: A prospective crossover reader study. *Int. J. Med. Inform.* **2026**, *205*, 106120. [CrossRef] [PubMed]
60. Shelmerdine, S.C.; Pauling, C.; Allan, E.; Langan, D.; Ashworth, E.; Yung, K.-W.; Barber, J.; Haque, S.; Rosewarne, D.; Woznitza, N.; et al. Artificial intelligence (AI) for paediatric fracture detection: A multireader multicase (MRMC) study protocol. *BMJ Open* **2024**, *14*, e084448. [CrossRef]
61. Tee, Y.S.; Liao, C.A.; Kuo, L.W.; Hsu, C.P.; Wang, C.C.; Fu, C.Y.; Liao, C.H.; Cheng, C.T. Artificial intelligence-assisted training for rib fracture interpretation: A prospective study in undergraduate medical students. *World J. Emerg. Surg.* **2026**, *21*, 14. [CrossRef]
62. Pervez, A.; Hasan, S.U.; Norrish, A.R. Convolutional neural networks in paediatric fracture detection: Pooled evidence from a systematic review and meta-analysis. *Eur. Radiol.* **2026**. [CrossRef]
63. Oettl, F.C.; Zsidai, B.; Oeding, J.F.; Hirschmann, M.; Feldt, R.; Fendrich, D.; Kraeutler, M.J.; Winkler, P.W.; Szaro, P.; Samuelsson, K. Artificial intelligence-assisted analysis of musculoskeletal imaging—A narrative review of the current state of machine learning models. *Knee Surg. Sports Traumatol. Arthrosc.* **2025**, *33*, 3032–3038. [CrossRef]
64. Sun, L.; Fan, Y.; Shi, S.; Sun, M.; Ma, Y.; Zhang, K.; Zhang, F.; Liu, H.; Yu, T.; Tong, H.; et al. AI-assisted radiologists vs. standard double reading for rib fracture detection on CT images: A real-world clinical study. *PLoS ONE* **2025**, *20*, e0316732. [CrossRef]
65. Tsang, B.; Gong, B.; Probyn, L.; Patlas, M.N. Artificial intelligence in emergency musculoskeletal imaging: A critical review of current applications. *Diagn. Interv. Imaging*, 2026; *Epub ahead of printing*. [CrossRef]
66. Kwee, R.M.; Kwee, T.C. Artificial intelligence-assisted detection of fractures on radiographs with BoneView: A systematic review. *Eur. J. Radiol.* **2025**, *190*, 112230. [CrossRef]
67. Pastor, M.; Dabli, D.; Lonjon, R.; Serrand, C.; Snene, F.; Trad, F.; de Oliveira, F.; Beregi, J.-P.; Greffier, J. Comparison between artificial intelligence solution and radiologist for the detection of pelvic, hip and extremity fractures on radiographs in adult using CT as standard of reference. *Diagn. Interv. Imaging* **2025**, *106*, 22–27. [CrossRef] [PubMed]
68. Huhtanen, J.T.; Nyman, M.; Blanco Sequeiros, R.; Koskinen, S.K.; Pudas, T.K.; Kajander, S.; Niemi, P.; Aronen, H.J.; Hirvonen, J. Comparative accuracy of two commercial AI algorithms for musculoskeletal trauma detection in emergency radiographs. *Emerg. Radiol.* **2025**, *32*, 569–580. [CrossRef] [PubMed]

69. Hendrix, N.; Hendrix, W.; van Dijke, K.; Maresch, B.; Maas, M.; Bollen, S.; Scholtens, A.; de Jonge, M.; Ong, L.S.; van Ginneken, B.; et al. Musculoskeletal radiologist-level performance by using deep learning for detection of scaphoid fractures on conventional multi-view radiographs of hand and wrist. *Eur. Radiol.* **2023**, *33*, 1575–1588. [CrossRef] [PubMed]
70. Cohen, M.; Puntonet, J.; Sanchez, J.; Kierszbaum, E.; Crema, M.; Soyer, P.; Dion, E. Artificial intelligence vs. radiologist: Accuracy of wrist fracture detection on radiographs. *Eur. Radiol.* **2023**, *33*, 3974–3983. [CrossRef]
71. Elbahi, M.K.; Muhammed, A.; Fadlelmola Abdalla Mohamednour, M.; Mukhtar, F.S. Artificial Intelligence in Fracture Diagnosis on Radiographs: Evidence, Pitfalls, and Pathways for Clinical Integration (2020–2025). *Cureus* **2025**, *17*, e93124. [CrossRef]
72. Sukhbaatar, T.; Davies, A.; Koye, A.; Hashem, M.; Sivaloganathan, S. Artificial intelligence in virtual fracture clinics: A systematic review of imaging and clinical-text tools. *J. Orthop. Surg. Res.* **2026**, *21*, 176. [CrossRef]
73. Oppenheimer, J.; Lüken, S.; Geveshausen, S.; Hamm, B.; Niehues, S.M. An overview of the performance of AI in fracture detection in lumbar and thoracic spine radiographs on a per vertebra basis. *Skelet. Radiol.* **2024**, *53*, 1563–1571. [CrossRef]
74. Collins, C.E.; Giammanco, P.A.; Trivedi, S.M.; Sarsour, R.O.; Kricfalusi, M.; Elsissy, J.G. Diagnostic Accuracy of Artificial Intelligence for Detection of Rib Fracture on X-ray and Computed Tomography Imaging: A Systematic Review. *J. Imaging Inf. Med.* **2025**, *38*, 2973–2982. [CrossRef] [PubMed]
75. Bonde, N.; Kjærgaard, K.; Aunaas, H.; Hangaard, S.; Daugaard, C.; Nybing, J.; Boesen, M.; Bachmann, R.; Lundemann, M.; Overgaard, S. Hip fracture detection on radiographs using an artificial intelligence-based support tool: A diagnostic accuracy study. *Acta Radiol.* **2026**, *67*, 314–323. [CrossRef]
76. Cha, Y.; Kim, J.T.; Park, C.H.; Kim, J.W.; Lee, S.Y.; Yoo, J.I. Artificial intelligence and machine learning on diagnosis and classification of hip fracture: Systematic review. *J. Orthop. Surg. Res.* **2022**, *17*, 520. [CrossRef]
77. Akbarian, E.; Mohammadi, M.; Tiala, E.; Ljungberg, O.; Razavian, A.; Magnéli, M.; Gordon, M. Development and validation of an artificial intelligence model for the classification of hip fractures using the AO-OTA framework. *Acta Orthop.* **2024**, *95*, 340–347. [CrossRef]
78. Wong, C.R.; Zhu, A.; Baltzer, H.L. The Accuracy of Artificial Intelligence Models in Hand/Wrist Fracture and Dislocation Diagnosis: A Systematic Review and Meta-Analysis. *JBJS Rev.* **2024**, *12*, e24. [CrossRef] [PubMed]
79. Jacques, T.; Cardot, N.; Ventre, J.; Demondion, X.; Cotten, A. Commercially-available AI algorithm improves radiologists' sensitivity for wrist and hand fracture detection on X-ray, compared to a CT-based ground truth. *Eur. Radiol.* **2024**, *34*, 2885–2894. [CrossRef] [PubMed]
80. Gharavi, A.; Navarro, S.M.; Tom, A.; Bankes, A.; Gish, M.; McGinnis, M.; Rich, M.D. Appraising the value of AI in wrist fracture detection: A clinical review. *Hand Surg. Rehabil.* **2026**, *45*, 102597. [CrossRef]
81. Namireddy, S.R.; Gill, S.S.; Peerbhai, A.; Kamath, A.G.; Ramsay, D.S.C.; Ponniah, H.S.; Salih, A.; Jankovic, D.; Kalasauskas, D.; Neuhoff, J.; et al. Artificial intelligence in risk prediction and diagnosis of vertebral fractures. *Sci. Rep.* **2024**, *14*, 30560. [CrossRef] [PubMed]
82. Silberstein, J.; Wee, C.; Gupta, A.; Seymour, H.; Ghotra, S.S.; Sá dos Reis, C.; Zhang, G.; Sun, Z. Artificial Intelligence-Assisted Detection of Osteoporotic Vertebral Fractures on Lateral Chest Radiographs in Post-Menopausal Women. *J. Clin. Med.* **2023**, *12*, 7730. [CrossRef]
83. Bhuria, R.; Gupta, S.; Ghoniem, R.M.; Singh, J.; Rani, S.; Taye, B.M.; Bharany, S. A transfer learning-based approach for automated bone fracture classification in X-ray imaging. *Ther. Adv. Musculoskelet. Dis.* **2026**, *18*, 1759720X251405099. [CrossRef]
84. Liu, J.; Sun, P.; Yuan, Y.; Chen, Z.; Tian, K.; Gao, Q.; Li, X.; Xia, L.; Zhang, J.; Xu, N. YOLOv12 Algorithm-Aided Detection and Classification of Lateral Malleolar Avulsion Fracture and Subfibular Ossicle Based on CT Images: Multicenter Study. *JMIR Med. Inf.* **2025**, *13*, e79064. [CrossRef]
85. Liu, Y.; Liu, W.; Chen, H.; Xie, S.; Wang, C.; Liang, T.; Yu, Y.; Liu, X. Artificial intelligence versus radiologist in the accuracy of fracture detection based on computed tomography images: A multi-dimensional, multi-region analysis. *Quant. Imaging Med. Surg.* **2023**, *13*, 6424–6433. [CrossRef]
86. Lee, K.C.; Choi, I.C.; Kang, C.H.; Ahn, K.S.; Yoon, H.; Lee, J.J.; Kim, B.H.; Shim, E. Clinical Validation of an Artificial Intelligence Model for Detecting Distal Radius, Ulnar Styloid, and Scaphoid Fractures on Conventional Wrist Radiographs. *Diagnostics* **2023**, *13*, 1657. [CrossRef] [PubMed]
87. Tee, Y.-S.; Huang, J.-F.; Huang, Y.-T.; Hsu, C.-P.; Chen, H.-W.; Hsieh, C.-H.; Fu, C.-Y.; Cheng, C.-T.; Liao, C.-H. Comparative analysis of AI support levels in clinical interpretation of traumatic pelvic radiographs. *npj Digit. Med.* **2025**, *8*, 518. [CrossRef]
88. Christoforides, E.; Rust, B.; Mulroy, D.; Oswald, S.; Muralidhar, R. Artificial Intelligence vs. Physician Expertise in Appendicular Skeleton Fracture Detection: A Scoping Review. *J. Am. Osteopath. Acad. Orthop.* **2024**, *5*, ZP987. [CrossRef]
89. Kim, C.-H.; Kim, J.W. Innovative applications of artificial intelligence in orthopedics focusing on fracture and trauma treatment: A narrative review. *J. Musculoskelet. Trauma.* **2025**, *38*, 178–185. [CrossRef]
90. Hussain, A.; Fareed, A.; Taseen, S. Bone fracture detection—Can artificial intelligence replace doctors in orthopedic radiography analysis? *Front. Artif. Intell.* **2023**, *6*, 1223909. [CrossRef]

91. McNair, D.; Price, W.N., II. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. In *Health Care Artificial Intelligence: Law, Regulation, and Policy*; National Academy of Medicine; The Learning Health System Series; Whicher, D., Ahmed, M., Israni, S.T., Matheny, M., Eds.; National Academies Press (US): Washington, DC, USA, 2023.
92. Potočnik, J.; Fujs, D. Navigating uncharted waters: Select practical considerations in radiology AI compliance with the EU AI Act. *npj Digit. Med.* **2025**, *8*, 630. [CrossRef]
93. Contaldo, M.T.; Pasceri, G.; Vignati, G.; Bracchi, L.; Triggiani, S.; Carrafiello, G. AI in Radiology: Navigating Medical Responsibility. *Diagnostics* **2024**, *14*, 1506. [CrossRef] [PubMed]
94. Lawrence, R.; Dodsworth, E.; Massou, E.; Sherlaw-Johnson, C.; Ramsay, A.I.G.; Walton, H.; O'Regan, T.; Gleeson, F.; Crellin, N.; Herbert, K.; et al. Artificial intelligence for diagnostics in radiology practice: A rapid systematic scoping review. *eClinicalMedicine* **2025**, *83*, 103228. [CrossRef]
95. Ashworth, E.; Allan, E.; Pauling, C.; Laidlow-Singh, H.; Arthurs, O.J.; Shelmerdine, S.C. Artificial intelligence (AI) in radiological paediatric fracture assessment: An updated systematic review. *Eur. Radiol.* **2025**, *35*, 5264–5286. [CrossRef]
96. Calleja, J.; Muscat, K.; Calleja, J.; Firth, G. Artificial Intelligence in Paediatric and Adolescent Fracture Detection: A Systematic Review and Meta-Analysis. *Cureus* **2025**, *17*, e92199. [CrossRef]
97. Zheng, Z.; Ryu, B.Y.; Kim, S.E.; Song, D.S.; Kim, S.H.; Park, J.W.; Ro, D.H. Deep learning for automated hip fracture detection and classification: Achieving superior accuracy. *Bone Jt. J.* **2025**, *107*, 213–220. [CrossRef] [PubMed]
98. Liawrungrueang, W.; Cholamjiak, W.; Promsri, A.; Jitpakdee, K.; Sunpaweravong, S.; Kotheeranurak, V.; Sarasombath, P. Artificial Intelligence for Cervical Spine Fracture Detection: A Systematic Review of Diagnostic Performance and Clinical Potential. *Glob. Spine J.* **2025**, *15*, 2547–2558. [CrossRef] [PubMed]
99. Lee, S.; Jung, J.-Y.; Mahatthanatrakul, A.; Kim, J.-S. Artificial Intelligence in Spinal Imaging and Patient Care: A Review of Recent Advances. *Neurospine* **2024**, *21*, 474–486. [CrossRef] [PubMed]
100. Ibrahim, M.T.; Milliron, E.; Yu, E. Artificial intelligence in spinal imaging—A narrative review. *Art. Int. Surg.* **2025**, *5*, 139–149. [CrossRef]
101. Li, Y.C.; Chen, H.H.; Horng-Shing Lu, H.; Hondar Wu, H.T.; Chang, M.C.; Chou, P.H. Can a Deep-learning Model for the Automated Detection of Vertebral Fractures Approach the Performance Level of Human Subspecialists? *Clin. Orthop. Relat. Res.* **2021**, *479*, 1598–1612. [CrossRef]
102. Voter, A.F.; Larson, M.E.; Garrett, J.W.; Yu, J.J. Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Cervical Spine Fractures. *AJNR Am. J. Neuroradiol.* **2021**, *42*, 1550–1556. [CrossRef]
103. Woodring, J.H.; Lee, C. Limitations of cervical radiography in the evaluation of acute cervical trauma. *J. Trauma.* **1993**, *34*, 32–39. [CrossRef]
104. Gale, S.C.; Gracias, V.H.; Reilly, P.M.; Schwab, C.W. The inefficiency of plain radiography to evaluate the cervical spine after blunt trauma. *J. Trauma.* **2005**, *59*, 1121–1125. [CrossRef] [PubMed]
105. Lin, J.T.; Lee, J.L.; Lee, S.T. Evaluation of occult cervical spine fractures on radiographs and CT. *Emerg. Radiol.* **2003**, *10*, 128–134. [CrossRef]
106. Shi, L.; Wang, H.; Shea, G.K. The Application of Artificial Intelligence in Spine Surgery: A Scoping Review. *J. Am. Acad. Orthop. Surg. Glob. Res. Rev.* **2025**, *9*, e24.00405. [CrossRef] [PubMed]
107. Wendt, K.; Nau, C.; Jug, M.; Pape, H.C.; Kdolsky, R.; Thomas, S.; Bloemers, F.; Komadina, R. ESTES recommendation on thoracolumbar spine fractures: January 2023. *Eur. J. Trauma Emerg. Surg.* **2024**, *50*, 1261–1275. [CrossRef]
108. Kocak, B.; Klontzas, M.E.; Stanzione, A.; Meddeb, A.; Demircioğlu, A.; Bluethgen, C.; Bressem, K.K.; Ugga, L.; Mercaldo, N.; Díaz, O.; et al. Evaluation metrics in medical imaging AI: Fundamentals, pitfalls, misapplications, and recommendations. *Eur. J. Radiol. Artif. Intell.* **2025**, *3*, 100030. [CrossRef]
109. Herrmann, J. Beyond Algorithms: Test Set Composition as a Determinant of AI Performance for Pediatric Fracture Detection. *Radiology* **2026**, *318*, e260207. [CrossRef]
110. Till, T.; Scherkl, M.; Stranger, N.; Singer, G.; Hankel, S.; Flucher, C.; Hržić, F.; Štajduhar, I.; Tschauner, S. Impact of test set composition on AI performance in pediatric wrist fracture detection in X-rays. *Eur. Radiol.* **2025**, *35*, 6853–6864. [CrossRef]
111. Najjar, R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics* **2023**, *13*, 2760. [CrossRef]
112. Magomedova, Z.; Pershina, E.S.; Keivan, D.; Nöldge, G.; Frank, M. The Human and Artificial Intelligence in Radiology: Current Status, Evidence, Regulation, and Future Perspectives. *Swiss J. Radiol. Nucl. Med.* **2026**, *27*, 1–28. [CrossRef]
113. Korfiatis, P.; Kline, T.L.; Meyer, H.M.; Khalid, S.; Leiner, T.; Loufek, B.T.; Blezek, D.; Vidal, D.E.; Hartman, R.P.; Joppa, L.J.; et al. Implementing Artificial Intelligence Algorithms in the Radiology Workflow: Challenges and Considerations. *Mayo Clin. Proc. Digit. Health* **2025**, *3*, 100188. [CrossRef] [PubMed]
114. Aldhafeeri, F.M. Governing Artificial Intelligence in Radiology: A Systematic Review of Ethical, Legal, and Regulatory Frameworks. *Diagnostics* **2025**, *15*, 2300. [CrossRef]

115. Mennella, C.; Maniscalco, U.; De Pietro, G.; Esposito, M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon* **2024**, *10*, e26297. [CrossRef] [PubMed]
116. Goisauf, M.; Cano Abadía, M. Ethics of AI in Radiology: A Review of Ethical and Societal Implications. *Front. Big Data* **2022**, *5*, 850383. [CrossRef] [PubMed]
117. Geis, J.R.; Brady, A.P.; Wu, C.C.; Spencer, J.; Ranschaert, E.; Jaremko, J.L.; Langer, S.G.; Kitts, A.B.; Birch, J.; Shields, W.F.; et al. Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. *J. Am. Coll. Radiol.* **2019**, *16*, 1516–1521. [CrossRef] [PubMed]

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