

R E V I E W

Diagnostic performance of artificial intelligence in fracture detection with ultrasound: A systematic review and meta-analysis

MICHAEL XIE^{1,*}, JUSTIN JY KIM^{1,*}, JAMES HB IM¹, LIZ DENNETT², ABHILASH HAREENDRANATHAN^{1,#}, JACOB L. JAREMKO^{1,#}

¹University of Alberta, Department of Radiology and Diagnostic Imaging, Edmonton, AB, Canada; ²University of Alberta, Sperber Health Sciences Library, Edmonton, AB, Canada; *These authors have contributed equally to this work; #These authors jointly supervised this work

ABSTRACT

Background: Bone fractures are common in acute care, and point-of-care ultrasound (POCUS) is an emerging diagnostic tool that can be complementary to or even in some cases an alternative to X-ray imaging. With the rise of artificial intelligence (AI) model incorporation in diagnostic interpretations, this systematic review and meta-analysis aimed to evaluate the diagnostic performance of deep learning models for detecting bone fractures on ultrasound images, using radiographic imaging with expert interpretation as the reference standard.

Methods: Comprehensive literature searches were conducted from inception to 15 August 2025, in databases including MEDLINE (Ovid interface), Embase (Ovid interface), CINAHL Plus with Full Text (EBSCOhost interface), Web of Science, ACM Digital Library, Scopus, and Google Scholar. We included studies evaluating deep learning models applied to ultrasound images or sweeps for fracture classification. The included studies predominantly evaluated pediatric cohorts presenting to emergency departments with suspected upper extremity fractures. Sensitivity and specificity were pooled using a bivariate random-effects model on the logit scale, and summary receiver operating characteristic (sROC) curve was constructed. Risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool.

Results: A total of 580 papers were identified in the preliminary literature search, and 333 studies were screened after duplicate removal. Screening 27 full-text articles resulted in eight studies that were included in the review. Across 21 reported model evaluations from five studies with extractable diagnostic data, pooled sensitivity was 0.77 (95% CI: 0.69-0.83) and pooled specificity was 0.90 (95% CI: 0.86-0.94), with an overall sROC AUC of 0.91.



Received: 11 March 2026 | Accepted: 23 April 2026

Correspondence: Michael Xie / 116 St. & 85 Ave, Edmonton, AB, T6G 2R3, Canada - Phone: +1 (403) 903-8178 / E-mail: kxie3@ualberta.ca

ORCID: 0009-0003-3580-1588

Justin Kim / 116 St. & 85 Ave, Edmonton, AB, T6G 2R3 / Canada - Phone: +1 (780) 271-4082 / E-mail: jjk2@ualberta.ca

ORCID: 0009-0009-7769-4147

Patient-level evaluations yielded higher pooled sensitivity (0.89; 95% CI: 0.81-0.94) and specificity (0.94; 95% CI: 0.81-0.98) compared to video-level analyses.

Conclusions: AI-assisted ultrasound demonstrates promising diagnostic performance for fracture detection with generally high specificity and variable sensitivity across architectures and anatomic targets. Future work should prioritize prospective, externally validated studies with standardized acquisition protocols and clinically meaningful reporting at the patient level.

Key words: Point-of-care ultrasound, artificial intelligence, bone fractures, deep learning models, musculoskeletal imaging

Introduction

Study background

Bone fractures are among the most common presentations to emergency and acute care settings where improper management may increase healthcare costs and long-term functional consequences (1,2). National U.S. data indicate that an estimated 7.9 million medical emergency department (ED) visits for fall-related fractures among children occurred between 2001 and 2015, which emphasizes the scale of acute fracture presentation (3). Therefore, accurate diagnosis is essential to guide appropriate management such as immobilization or surgical intervention for fracture reduction (4,5). Conventional radiography is the current gold standard for fracture detection, but it involves ionizing radiation and may be less accessible in remote healthcare settings (6,7).

Ultrasound is increasingly used in diagnosis of fractures; specifically, long bone, rib, metacarpal, and superficial bone fractures (7,8). It is a portable, relatively inexpensive, and radiation free tool which may be performed at bedside and can be integrated with existing imaging workflows. Diagnostic test accuracy supports ultrasound in detecting fractures in various extremities with moderately high sensitivities and specificities with some variability in the location of the fracture site (9–11). One randomized controlled trial displayed comparable diagnostic accuracy between ultrasonography to radiography in diagnosing distal forearm fractures (12). However, several studies

suggest that ultrasound can be limited by variability in skill level of the operator and quality of image acquisition (13–15). These factors contribute to inconsistent diagnostic performance. Therefore, there is a need in reducing user-dependence and improving reproducibility of ultrasound fracture assessments.

Artificial intelligence (AI), specifically deep learning-based image analysis, has shown strong performance for fracture detection on radiographs (2). Other studies which incorporated AI models in the diagnostic workflow of radiologists have shown to help increase accuracy in radiographic interpretation (16,17). Applying AI to augment ultrasound in fracture detection may reduce operator dependency and standardize interpretation by consistently identifying sonographic features of cortical disruption and associated soft-tissue changes (18). AI-assisted ultrasound could potentially expand access to remote areas where radiologists or trained sonographers are not available for image interpretation (19). Existing studies on AI-assisted ultrasound for fractures are few and heterogeneous, with variation in anatomical site scanned, patient population, and AI models used for detection. A systematic synthesis is needed to quantify pooled performance and guide future methods in studies to provide more consistent reporting methods and increase the likelihood of utilizing AI models in interpreting ultrasound scans clinically.

Study objectives

The objective of this systematic review and meta-analysis was to evaluate the diagnostic performance of

AI models for detecting bone fractures on ultrasound, using radiographic imaging as the reference standard.

Methods

The systematic review and meta-analysis was registered with the International Prospective Register of Systematic Reviews (PROSPERO: CRD420241114869). The review was registered before initiation and followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (20).

Search strategy

A comprehensive search was completed in the following databases from inception to 15 August 2025, to identify relevant publications: MEDLINE(Ovid interface), Embase (Ovid interface), CINAHL Plus with Full Text (EBSCOhost interface), Web of Science, ACM Digital Library, Scopus, and Google Scholar (on August 17, 2025). No language restrictions were applied. The search used a combination of subject headings and keywords related to ultrasound diagnosis, bone fractures, and artificial intelligence. Grey literature formats were included: conference abstracts, preprints, and clinical trial registry records from the various databases. The search was conducted by a health sciences librarian (LD) experienced in systematic review methodology. A sample OVID MEDLINE search strategy is detailed in the Supplementary Materials (Table S1). The full search strategy is also found in the Supplementary Materials.

Eligibility criteria

We define eligibility using the PICO framework. The population included patients of any age who have been evaluated with ultrasound for suspected musculoskeletal fractures. The index test was the use of a deep learning algorithm applied to ultrasound images or ultrasound sweeps for classifying and diagnosing fractures. Gold reference standard was radiographic imaging (X-ray, CT, MRI) with expert interpretation by the staff radiologist or orthopedic surgeon

confirming fracture status. The primary outcomes were the diagnostic performance metrics for the AI model used, including sensitivity, specificity, and accuracy at the image, video, or patient level. Secondary outcomes included sex, age, clinical setting, anatomical site scanned, AI model type and training dataset parameters, and ultrasonographer experience.

Observational studies, cohort studies, case-control studies, conference proceedings papers with sufficient data, and randomized controlled trials that evaluated deep learning models used for diagnosing fractures based on ultrasound imaging were included. Case reports and small case series that include less than 10 patients without quantitative diagnostic data, reviews, conference abstracts, and studies that applied AI to other types of imaging or used AI for tasks separate from classification of fractures were excluded.

Study selection

Any duplicates were removed after the first literature search, and the initial abstract screening were done independently by two reviewers (MX and JK). Full-text review of the included abstracts was also independently reviewed by two reviewers (MX and JK) against the eligibility criteria. Disagreements were resolved by discussion, and a third reviewer was consulted if a consensus was not reached. Reasons for full-text exclusion were summarized in the PRISMA flow diagram (Figure 1).

Data extraction and synthesis of results

Two reviewers independently extracted the data using a standardized data table. Study characteristics (author, year, country, study design, and setting), population details (sample and subgroup sample size, age), ultrasound parameters (probe type, operator proficiency, anatomical site scanned), and AI model characteristics (architecture, training dataset size, testing dataset size, task) were extracted from each included study.

We calculated true positives, true negatives, false positives, and false negatives using reported values of prevalence, sensitivity and specificity for each study in order to estimate standard errors. We produced pooled

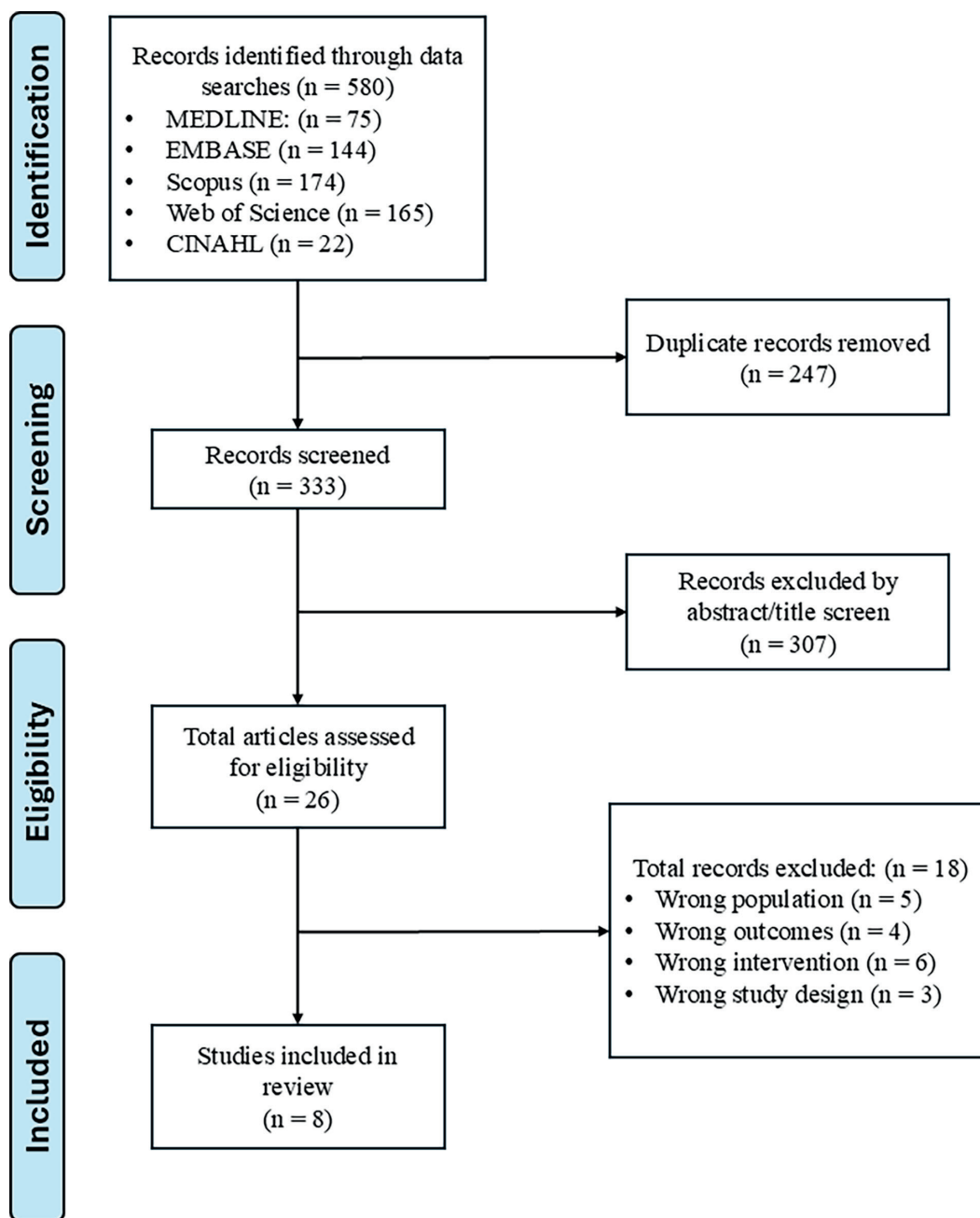


Figure 1. PRISMA flow diagram of included studies evaluating AI models in predicting MSK fractures using ultrasound as the diagnostic modality.

estimates of sensitivity and specificity and 95% confidence intervals (CI) for all studies. Additionally, we also performed subgroup analyses based on whether studies reported scanning at the patient level or the video level and reported pooled estimates of sensitivity and specificity. Because sensitivity and specificity are often related to each other through a pre-determined

cut-off value, we used a bivariate approach to take into account this correlation. Specifically, we used a linear mixed model to pool sensitivity and specificity. Additionally, we also constructed summary receiver operating characteristics (sROC) curves to further analyze overall accuracy of AI in detecting fractures using the area under the curve (AUC). The I^2 statistic was used

to assess the between-study heterogeneity. All statistical analyses were conducted using the *mada* packages in R Software version 4.4.1.

Risk of bias

Risk of bias was assessed with the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool across four domains (patient selection, index test, reference standard, and flow/timing) (21). Two reviewers performed assessments independently, with disagreements resolved by consensus. A traffic light and weighted bar plot of the distribution of risk-of-bias judgements were created.

Results

Study selection

We identified 580 studies, of which 247 duplicates were removed. A total of 26 articles were eligible for full-text screening whereby eight studies were included in the review (Figure 1). Of these eight studies, five provided extractable sensitivity and specificity data and were included in the quantitative meta-analysis (22-24,28,29). Three studies lacked sufficient diagnostic accuracy parameters for inclusion in the meta-analysis and were summarized qualitatively (25-27).

Study characteristics

Of these included studies, most of them were conducted in a pediatric emergency department (ED) setting and focused on radial fractures (22-29). Of the eight included studies, seven (87.5%) focused on pediatric populations (22-26,28,29), and one study (12.5%) included a mixed adult and pediatric cohort (27). Six studies were conducted in an ED setting (22,24-26,28,29). Ulnar, carpal, and supracondylar humerus fractures were also included in these studies (Table 1). One study also included lower limbs, but the exact anatomical fracture sites were not specified (27). Predominantly all studies were prospective with two retrospective observational studies were included (23,25). Ultrasound acquisition was most commonly performed using the linear high-frequency probes with some variation between 2D and 3D ultrasound sweeps. Diverse

model architectures such as the convolutional neural networks (CNN), recurrent models, and transformer-based approaches were used to perform a variety of AI tasks like classification, segmentation, and detection (Table 2). Ground-truth reference standards were based on radiography with expert interpretation with a couple studies providing additional clinical imaging follow-up. The sensitivities, specificities, and accuracies of each model were reported in Table 2.

Risk of bias results

Overall risk of bias and applicability concerns are summarized in Figure 2a and 2b. In the patient selection domain, four studies were rated as low risk of bias (22,25,26,29) and the remaining four did not fully describe inclusion and exclusion decisions (23,24,27,28). The index test domain showed two studies as unclear (23,27) and one study as high risk (26) as the thresholds were not prespecified or the results were not blinded, respectively. The reference standard domain was generally low risk due to the consistent use of X-rays as the gold standard. The flow and timing domain showed four studies at low risk (22,26,28,29), two at unclear risk (23,24), and two at high risk (25,27). Several studies used consecutive enrollment, but external validity may still be limited because recruitment occurred within similar clinical workflows. Overall risk of bias included 3 studies with low risk of bias (22,28,29), unclear risk of bias (23,24,27), and high risk of bias (25,26). Five studies either having the low risk of bias or unclear risk of bias were included in the quantitative meta-analysis as they also included measurable sensitivity and specificity values.

Quantitative synthesis

A meta-analysis of diagnostic accuracy was performed across 21 model evaluations drawn from five included articles (22-24,28,29). Three studies lacked extractable sensitivity, specificity, and accuracy data, therefore, were summarized qualitatively. Overall pooled performance demonstrated moderate-to-high diagnostic accuracy. The pooled sensitivity (95% CI) was 0.77 (0.69-0.83) with $I^2 = 56.72\%$, and the pooled specificity (95% CI) was 0.90 (0.86-0.94) with $I^2 = 58.55\%$ (Table 3, Figure 3). These I^2 values indicate moderate heterogeneity across model evaluations for both sensitivity and specificity, suggesting

Table 1. Characteristics of all included studies.

Author	Publication Year	Country	Study Design	Population Studied	Sample Size	Mean Age	Sex Distribution (M:F)	Patients with Fracture	Patients without Fracture	Clinical Setting	Type of Ultrasound	Anatomical Site of Fracture	Training Dataset Size	Technician Experience
Knight et al.	2023	Canada	Prospective	Pediatrics	127	NR	NR	NR	NR	ED	3D US - Philips IU22 machine using a 13 MHz VL13-5 probe 2D US - Philips Lumify L5-12 MHz probe	Distal radial	2D - 16865 images fractured (24.81%) 3D - 15882 images fractured (32.33%)	Novice, intermediate, expert
Großbröhmer et al.	2023	Germany	Conference Proceedings Paper: Retrospective	Pediatrics	43 (180 images)	NR	NR	79 images	101 images	NR	GE LOGIQ 57 Performed with WRIST-SAFE protocol	Distal radial and ulnar fractures	148 images	NR
Hareend-ranathan et al.	2023	Canada	Prospective	Pediatrics	30 (55 sweeps)	NR	NR	19 studies	31 studies	ED	3D US - Philips iU22 machine using a 13 MHz VL13-5 probe	Distal radius	28 sweeps	NR
†Zhou et al.	2022	Canada	Conference Proceedings Paper: Retrospective	Pediatrics	57 (12264 images, 113 sweeps)	NR	NR	22	35	ED	3D ultrasound (Philips iU22, 13VL5 probe)	Distal radius and carpals	10500 images and 93 sweeps	Expert sonographer
†Tripathi et al.	2024	India & Canada	Prospective	Pediatric	30	NR	NR	22	8	ED	3D ultrasound (Philips iU22, 13VL5 probe)	Distal radius	1024 images	NR

Author	Publication Year	Country	Study Design	Population Studied	Sample Size	Mean Age	Sex Distribution (M:F)	Patients with Fracture	Patients without Fracture	Clinical Setting	Type of Ultrasound	Anatomical Site of Fracture	Training Dataset Size	Technician Experience
†Teng, Pan, and Zhao	2023	China	Prospective	Adults + Pediatrics	164	45.8 (12-83)	87:77	164	0	In-patient post-fracture surgery	NR	Upper and lower limbs	104 images	NR
Zhang et al.	2021	Canada	Prospective	Pediatrics	36	9.9 (3.8-14.8)	21:9	13 scans	23 scans	ED	3D ultrasound (Philips iU22, 13VL5 probe)	Distal radius	6000 images	Medical students with 1 hour hands-on training
Zhou, Knight, et al.	2025	Canada	Prospective	Pediatric	Wrist: 111 Elbow: 53	NR	NR	Only Test Set reported: Wrist: 18 Elbow: 8	Only Test Set reported: Wrist: 89 Elbow: 16	ED	Philips Lumify L5-12MHz & Philips iU22, 13VL5 probe	Distal radius and supracondylar humerus	Wrist: 16865 Elbow: 6807	Expert sonographer

Legend: ED = Emergency Department; NR = Not reported; XR = X-ray; US = Ultrasound; M:F = Male:Female
 †Studies were not included in the quantitative meta-analysis

Table 2. Characteristics of models used for diagnostic interpretation of fractures by analyzing ultrasound imaging.

Author	AI Task (Detection, classification, segmentation)	Type of AI Model	Ground Truth Standard	Interpretation of Ground Truth	Original films or Follow-up imaging	Patient or Image/Sweep Level	Sensitivity	Specificity	Accuracy
Knight et al.	Classification	CNN ResNet34 DenseNet121	XR	Pediatric MSK radiologist read	Original clinical XR	Patient Level	ResNet34 2D: 0.82 DenseNet121 2D: 0.91 ResNet34 3D: 1.00 DenseNet121 3D: 1.00	ResNet34 2D: 1.00 DenseNet121 2D: 0.94 ResNet34 3D: 1.00 DenseNet121 3D: 0.94	ResNet34 2D: 0.89 DenseNet121 2D: 0.94 ResNet34 3D: 1.00 DenseNet121 3D: 0.94
Großbröhmer et al.	Classification, segmentation	ResNet18 (classification) nnUNet (segmentation) Mask R-CNN with ResNet50 backbone (detection/instance segmentation)	XR	NR	NR	Image Level	ResNet: 0.6969±0.1864 nnUNet: 0.6886 ± 0.1339 Mask R-CNN (τ = 0.5): 0.6807 ± 0.0781	ResNet: 0.7797 ± 0.0536 nnUNet: 0.8738 ± 0.1040 Mask R-CNN (τ = 0.5): Mask R-CNN (τ = 0.5): 0.9039 ± 0.0417	ResNet: 0.7625 ± 0.0508 nnUNet: 0.8000 ± 0.0545 Mask R-CNN (τ = 0.5): 0.8125 ± 0.0681
Hareendranathan et al.	Classification, segmentation	RNN (classification) LSTM (classification) GRU (classification) Alexnet (classification) Resnet 50 (classification) VGG 16 (classification) Inception V3 (classification)	Pediatric MSK radiologist US report	Pediatric MSK radiologist read	Original clinical XR	Image Level	RNN: 0.57 LSTM: 0.71 GRU: 0.86	RNN: 0.91 LSTM: 0.96 GRU: 0.87	RNN: 0.83 LSTM: 0.90 GRU: 0.87
†Zhou et al.	Segmentation	U-Net pix2pix GAN U-Net + GAN	XR	Pediatric MSK radiologist read	Original clinical XR	Image Level	NR	NR	U-Net: 0.783 pix2pix GAN: 0.644 U-Net + GAN: 0.731

†Tripathi et al.	Detection	Unsupervised Acoustic FF-CNN-KeyNet-RefineNet framework	XR	Pediatric MSK radiologist read	Original clinical XR	NR	NR	NR	NR	NR
†Teng, Pan, and Zhao	Classification	CNN VGG-11	XR	Pediatric MSK radiologist read	2, 4, 8, 12, week follow-up	NR	NR	NR	NR	Overall: 0.73
Zhang et al.	Classification	CNN	XR	Pediatric MSK radiologist read	Original clinical XR	Image, Sweep, and Patient Level	Per image: 0.98 Per sweep: 0.85 Per patient: 1.00	Per image: 0.89 Per sweep: 0.91 Per patient: 0.87	Per image: 0.94 Per sweep: 0.89 Per patient: 0.92	
Zhou, Knight, et al.	Classification, Segmentation	U-Net + ResNet-34; Segmentation ResNet-34; Classification VGG16; Classification Vision Transformer: Classification	XR	Central expert read by two radiologists	Original clinical XR	Patient and Image Level	Patient Classification: ResNet34 Wrist: 0.818 ResNet34 Elbow: 1.00 Image Classification: ResNet34 Wrist: 0.756 VGG16 Wrist: 0.459 ResNet34 Elbow: 0.014 VGG16 Elbow: 0.365	Patient Classification: ResNet34 Wrist: 1.000 ResNet34 Elbow: 0.714 Image Classification: ResNet34 Wrist: 0.964 VGG16 Wrist: 0.997 ResNet34 Elbow: 1.000 VGG16 Elbow: 0.975	Patient Classification: ResNet34 Wrist: 0.889 ResNet34 Elbow: 0.750 Image Classification: ResNet34 Wrist: 0.917 VGG16 Wrist: 0.876 ResNet34 Elbow: 0.951 VGG16 Elbow: 0.945	

Legend: ED = Emergency Department, NR = Not reported, XR = X-ray; US = Ultrasound; M:F = Male:Female
 †Studies were not included in the quantitative meta-analysis

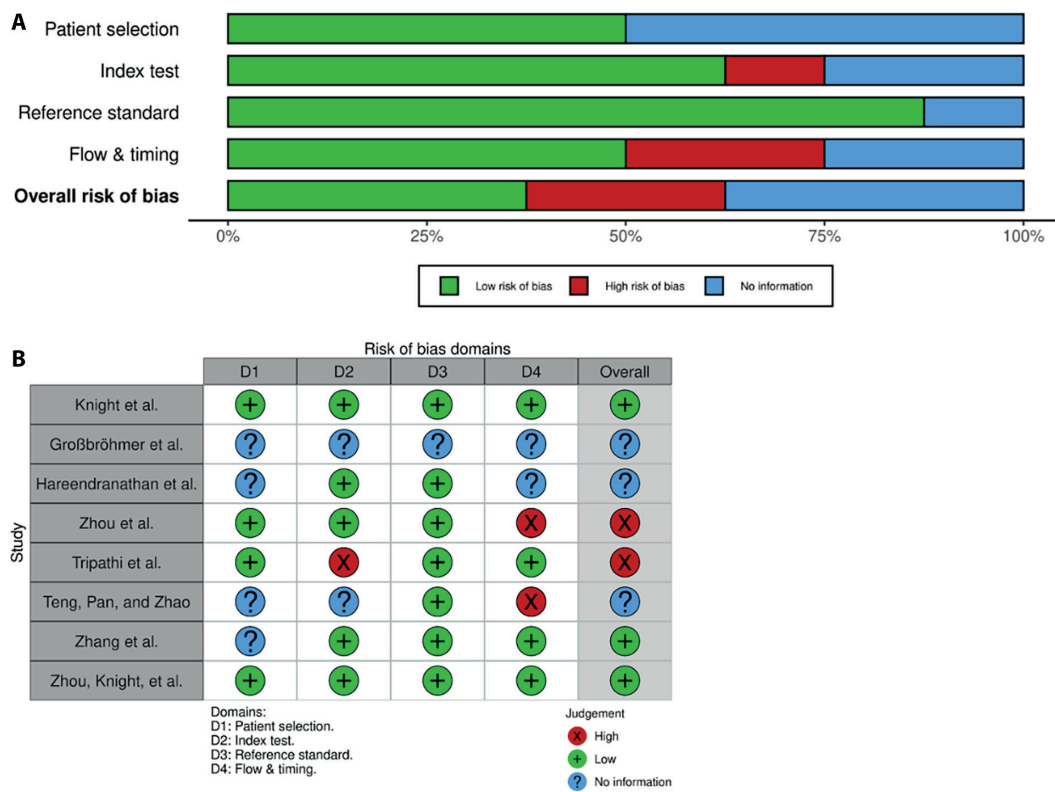


Figure 2. Risk of bias summary plot (a) depicting each domain using the QUADAS-2 assessment tool. Individual study-level traffic light plot (b) depicting QUADAS-2 risk of bias and applicability judgements for each included study.

Table 3. Characterizing overall, patient level, and video level sensitivity, specificity, and sROC AUC and the respective I² values.

	Overall	Patient Level	Video Level
Sensitivity	0.77 (0.69, 0.83), I ² = 0.5672	0.89 (0.81, 0.94), I ² = 0.4032	0.72 (0.62, 0.80), I ² = 0.7121
Specificity	0.90 (0.86, 0.94), I ² = 0.5855	0.94 (0.81, 0.98), I ² = 0.4262	0.91 (0.86, 0.94), I ² = 0.7200
sROC AUC	0.91	0.93	0.89

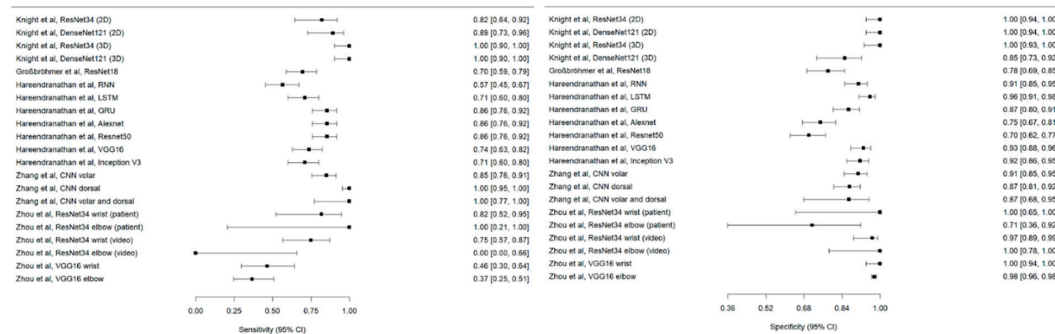


Figure 3. Forest plot of overall sensitivity and specificity of included AI models from 5 articles. Sensitivity estimates with 95% confidence intervals were included across all model evaluations.

that performance varied across studies and model configurations.

To explore whether diagnostic performance differed depending on how outcomes were aggregated, pooled estimates were also calculated at the patient level and video level (Table 3). At the patient level, pooled sensitivity increased to 0.89 (0.81-0.94) with $I^2 = 40.32\%$, while pooled specificity was 0.94 (0.81-0.98) with $I^2 = 42.62\%$ (Table 3, Figure 4). Compared with the overall analysis, patient-level pooling was associated with higher point estimates for both sensitivity and specificity with lower heterogeneity, indicating more consistent diagnostic performance when predictions are summarized per patient rather than per sweep.

At the video/sweep level, pooled sensitivity was lower at 0.72 (0.62-0.80) with $I^2 = 71.21\%$, whereas pooled specificity remained high at 0.91 (0.86-0.94) with $I^2 = 72.00\%$ (Table 3, Figures 5). The substantially higher I^2 values at the video level indicate greater between-evaluation variability. These findings suggest

that fracture detection performance was more reliable at the patient level than at the video level with patient level aggregation providing stronger sensitivity, specificity, and less heterogeneity.

Discriminative performance was further summarized using sROC curves. The overall sROC AUC was 0.91, consistent with moderately good discrimination (Table 3, Figure 6). When stratified by evaluation level, the sROC AUC was 0.93 at the patient level and 0.89 at the video level (Table 3, Figures 6). This pattern mirrors the pooled sensitivity and specificity results, with higher AUC at the patient level, indicating better overall diagnostic discrimination when models are evaluated between patients rather than between sweeps.

Discussion

This systematic review and meta-analysis synthesized evidence on AI-assisted fracture detection using ultrasound. Across 21 model evaluations, we found



Figure 4. Forest plot of patient sensitivity and specificity of included AI models from 5 articles. Sensitivity and specificity estimate for patient level with 95% confidence intervals were included across all model evaluations.

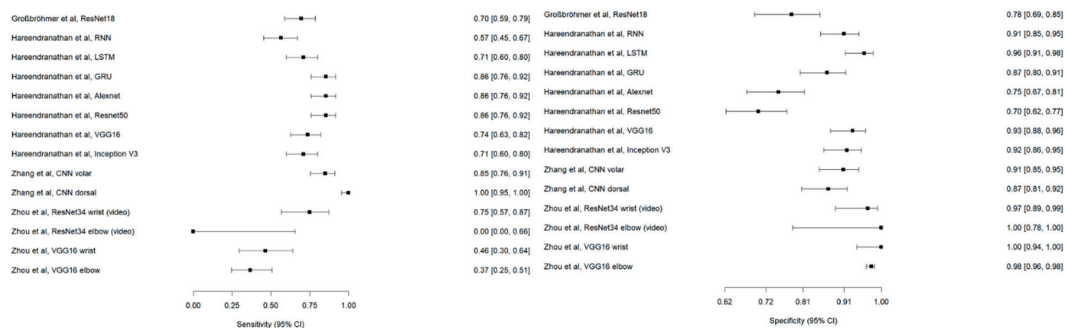


Figure 5. Forest plot of video level sensitivity and specificity included AI models from 5 articles. Sensitivity and specificity estimate for video level with 95% confidence intervals were included across all model evaluations.

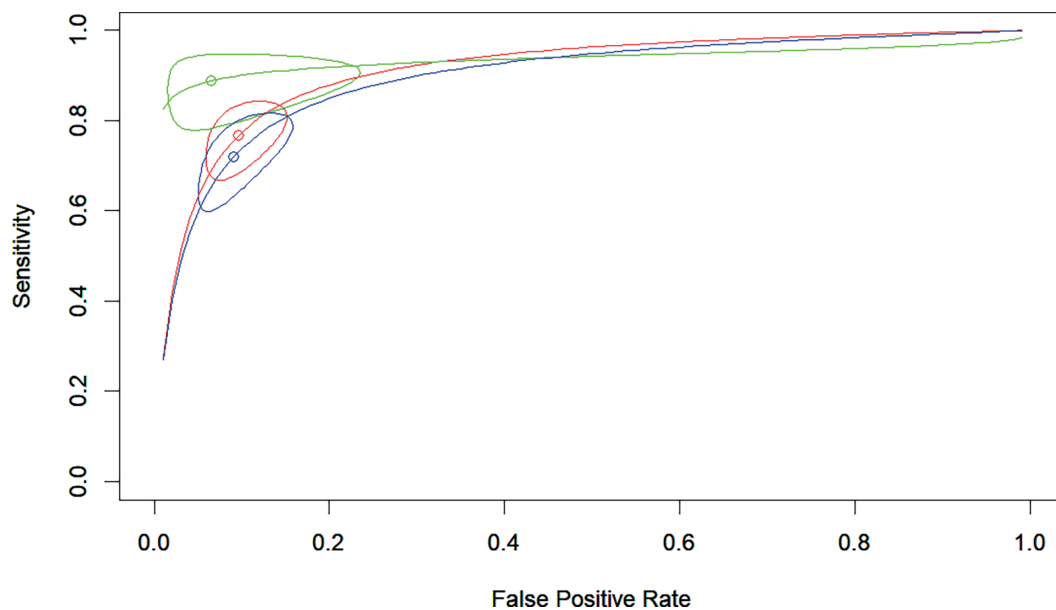


Figure 6. The sROC curve for AI-assisted fracture detection. The overall sROC curve (red), patient level sROC (green), and video level sROC curve (blue) are depicted.

that AI accuracy was higher on a per-patient level than on a per-video level and was high enough to be potentially clinically useful (pooled patient-level sensitivity 0.89, specificity 0.94). These performance patterns could help support decision making in settings where rapid triage of MSK injuries is desirable. AI-assisted ultrasound fracture detection consistently achieves high specificity across evaluation levels, while sensitivity is highest and most consistent at the patient level.

The studies that were included had similarities between fracture sites, where the ultrasound scans were conducted, population demographic, and the relatively consistent reference standard. This homogeneity supports cross-study comparisons with more emphasis on the diagnostic accuracy of the AI model. A practical strength of this review is that it synthesizes evidence between POCUS, AI, and fracture detection which can provide more direction on how these tools can be used to augment performance in a real-world setting like the ED. Prior studies using POCUS to assess distal forearm injuries in clinical practice showed strong diagnostic accuracy, supporting ultrasound-first pathways (30,31). The included studies focused primarily on distal forearm fractures; further studies in other frequently fractured sites would be useful, especially given the high diagnostic accuracy noted by others

using ultrasound by emergency physicians to find fractures in ribs, clavicles, and lateral malleolus (32–34).

AI fracture detection on radiographs has been found by others to be highly accurate in locations such as the wrist (35,36). Diagnostic accuracy of these models was even higher than ultrasound with AI, potentially due to the more standardized nature of radiographs, with less operator dependence than ultrasound (37,38). Most of the studies did not report the operator's level of expertise, which was a limitation since it would be impossible to tell if the model's accuracy was dependent on the user. Future studies may need inter-operator comparisons between model performance depending on which ultrasound sweep is being evaluated.

A key finding was the wide dispersion of sensitivity estimates, indicating performance instability across models and evaluation levels (at the image/video/patient level). This heterogeneity is clinically important as even with strong pooled specificity, sensitivity variability would result in low clinical utility without robust external validation and workflow safeguards. Additionally, I^2 values across all models indicated moderate levels of heterogeneity. Differences in model outputs and decision thresholds can vary substantially which may influence measured sensitivity and specificity which can complicate cross-study comparability.

A key limitation is the lack of independent replication across research groups. Although QUADAS-2 primarily evaluates individual study-level domains, multiple studies originated from the same research team. This indicates potential overlapping AI model development, patient selection patterns, and ultrasound acquisition protocols. As a result, the consistency in diagnostic performance may reflect the shared study designs, as the studies were likely conducted within similar clinical environments. This homogeneity limits the external validity of the pooled estimates, and it may not be readily applicable to other EDs and clinical workflow. Therefore, these outcomes should be interpreted as preliminary findings, underscoring the need for a variety of different ultrasound protocols in multicenter settings to evaluate the diagnostic utility of AI-assisted ultrasound.

Several AI-specific sources of bias should be considered when interpreting these results. Many of the included studies evaluated models on internal test sets drawn from the same and relatively small patient cohorts used for training, raising the possibility of overfitting where models learn dataset-specific patterns rather than generalizable features of bone fractures. None of the included studies performed external validation on fully independent datasets from separate institutions (22-29). Additionally, most studies did not prespecify decision thresholds where reported sensitivities and specificities may reflect optimized rather than prospectively determined cutoffs. These factors collectively may suggest that the pooled estimates may represent an upper bound of expected real-world performance.

The 21 model evaluations were drawn from five studies, with multiple models from the same dataset treated as separate observations. This approach attempted to capture the variability in diagnostic performance across different AI architectures (CNNs, recurrent models, transformer-based approaches). An important limitation that is drawn is that some of these models were trained on the same dataset sharing acquisition conditions; therefore, each model is not fully independent. An alternative approach of selecting a single representative model per study would reduce this correlation concern but would introduce selection bias as selecting the best-performing model would

inflate pooled estimates. All reported evaluations were retained in the primary analysis to provide a more comprehensive picture of current performance. Also, the findings of this review are derived primarily from pediatric populations where only one study included the adult population (27). Therefore, the findings may not be generalizable to adult fracture detection. Pediatric fractures, particularly buckle fractures of the distal radius, have distinct sonographic appearances compared with adult fracture patterns. Diagnostic performance of AI models may differ when applied to adult bone morphology and fracture types. A further consideration for generalizability is that most of the included studies (22,24-26,28) utilized 3D ultrasound transducers (Philips iU22 with 13VL5 probe), which may not be utilized in a typical ED setting. The volumetric data provided by 3D sweeps may facilitate AI classification in ways that are not replicated with conventional 2D imaging, and models trained on 3D data may not perform comparably when applied to 2D acquisitions.

Also, another limitation may be that the AI models evaluated in this review address the interpretation challenge by providing automated analysis of acquired sweeps, but they do not address the challenges associated with ultrasound acquisition. The latter is more operator-dependent and poor acquisition quality resulting from inadequate probe positioning, insufficient scanning coverage, or patient movement may degrade the AI model performance regardless of the model's intrinsic accuracy. Future studies should evaluate AI performance across varying acquisition quality levels and operator experience to understand the full clinical utility of these tools.

In terms of clinical utility, AI-assisted ultrasound for fracture detection is not yet at a level to substitute radiography or expert clinician ultrasound interpretation, but it may provide useful workflow-support. Ultrasound could, for example, be useful at the ER triage desk to speed up management when the model is highly confident in a positive finding, before having to wait long times for an X-ray (39-41). Further work should include prospective multicenter validation with patient-level outcomes comparing the work of ultrasound operators of varying experience to determine to what extent AI-augmented ultrasound assists in clinical workflow.

Conclusion

This systematic review found that AI-assisted fracture detection using ultrasound has meaningful diagnostic potential. However, the paucity of studies and heterogeneity in model sensitivity limit the confidence in broad clinical generalizability to date. Larger, prospective studies with standardized acquisition and patient-level evaluations are needed before routine clinical adoption.

Author's Contributions: Author contributions: MX, JK, AH, and JJ contributed to the study conception and design. Literature search was performed by LD. Material preparation and data collection were performed by MX and JK. Data analysis was performed by MX, JK, JI. The first draft of the manuscript was written by MX and JK, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. MX and JK have contributed equally as first authors. One or more authors are affiliated with the research group that produced some of the included primary studies. To mitigate potential bias, screening, data extraction, and risk-of-bias assessments were performed using prespecified criteria and the reviewers were not involved in the included primary studies.

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript

Ethics Approval and Consent to Participate: This study is a literature review based on publicly available data and, therefore, does not require ethical approval from an institutional review board.

Competing Interests: The authors do not have competing interests.

Declaration on the Use of AI: None

References

1. Miller AN. Evaluation of common musculoskeletal injuries in the urgent setting. *MedEdPORTAL* 2016 Dec 7. doi: 10.15766/mep_2374-8265.10514
2. Kuo RYL, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, et al. Artificial intelligence in fracture detection: a systematic review and meta-analysis. *Radiology* 2022;304:50-62. doi: 10.1148/radiol.211785
3. Orces CH, Orces J. Trends in the U.S. childhood emergency department visits for fall-related fractures, 2001-2015. *Cureus* 2020;12:e11629. doi: 10.7759/cureus.11629
4. Al-Naser Y, Alshadeedi F. Bringing imaging to the people: enhancing access and equity in healthcare through mobile imaging. *J Med Imaging Radiat Sci* 2024 Dec 1. doi: 10.1016/j.jmir.2024.101715
5. Pinto A, Berritto D, Russo A, Riccitiello F, Caruso M, Belfiore MP, et al. Traumatic fractures in adults: missed diagnosis on plain radiographs in the emergency department. *Acta Biomed* 2018;89 Suppl 1:111-23. doi: 10.23750/abm.v89i1-s.7015
6. Al-Baghdadi R, Yu TJ, Ferro A, Roberts J. Radiography versus sonography in the detection of acute fractures: a case series. *J Diagn Med Sonogr* 2021;37:400-6. doi: 10.1177/8756479321996314
7. Lin EC. Radiation risk from medical imaging. *Mayo Clin Proc* 2010;85:1142-6. doi: 10.4065/mcp.2010.0260
8. Du MJ, Lin YH, Chen WT, Zhao H. Advances in the application of ultrasound for fracture diagnosis and treatment. *Eur Rev Med Pharmacol Sci* 2022;26:7949-54. doi: 10.26355/eurrev_202211_30146
9. Tsou PY, Ma YK, Wang YH, Gillon JT, Rafael J, Deanehan JK. Diagnostic accuracy of ultrasound for upper extremity fractures in children: a systematic review and meta-analysis. *Am J Emerg Med* 2021;44:383-94. doi: 10.1016/j.ajem.2020.04.071
10. Atilla OD, Yesilaras M, Kilic TY, Tur FC, Reisoglu A, Sever M, et al. The accuracy of bedside ultrasonography as a diagnostic tool for fractures in the ankle and foot. *Acad Emerg Med* 2014;21:1058-61. doi: 10.1111/acem.12467
11. Bozorgi F, Shayesteh Azar M, Montazer SH, Chabra A, Heidari SF, Khalilian A. Ability of ultrasonography in detection of different extremity bone fractures; a case series study. *Emergency* 2017;5:e15. doi: 10.22037/emergency.v5i1.12821
12. Snelling PJ, Jones P, Bade D, Bindra R, Byrnes J, Davison M, et al. Ultrasonography or radiography for suspected pediatric distal forearm fractures. *N Engl J Med* 2023;388:2049-57. doi: 10.1056/nejmoa2213883
13. Ipsen S, Wulff D, Kuhlemann I, Schweikard A, Ernst F. Towards automated ultrasound imaging: robotic image acquisition in liver and prostate for long-term motion monitoring. *Phys Med Biol* 2021;66. doi: 10.1088/1361-6560/abf277
14. Li K, Xu Y, Meng MQH. An overview of systems and techniques for autonomous robotic ultrasound acquisitions. *IEEE Trans Med Robot Bionics* 2021;3:510-24. doi: 10.1109/tmrb.2021.3072190
15. Kojcev R, Khakzar A, Fuerst B, Zettinig O, Fahkry C, DeJong R, et al. On the reproducibility of expert-operated and robotic ultrasound acquisitions. *Int J Comput Assist Radiol Surg* 2017;12:1003-11. doi: 10.1007/s11548-017-1561-1
16. Gan K, Xu D, Lin Y, Shen Y, Zhang T, Hu K, 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. doi: 10.1080/17453674.2019.1600125
17. Nguyen T, Maarek R, Hermann AL, Kammoun A, Marchi A, Khelifi-Touhami MR, 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;52:2215-26. doi: 10.1007/s00247-022-05496-3
 18. Tenajas R, Miraut D, Illana CI, Alonso-Gonzalez R, Arias-Valcayo F, Herraiz JL. Recent advances in artificial intelligence-assisted ultrasound scanning. *Appl Sci* 2023;13:3693. doi: 10.3390/app13063693
 19. Baloescu C, Bailitz J, Cheema B, Agarwala R, Jankowski M, Eke O, et al. Artificial intelligence-guided lung ultrasound by nonexperts. *JAMA Cardiol* 2025;10:245-53. doi: 10.1001/jamacardio.2024.4991
 20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLOS Med* 2021;18:e1003583. doi: 10.1136/bmj.n71
 21. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36. doi: 10.7326/0003-4819-155-8-201110180-00009
 22. Knight J, Zhou Y, Keen C, Hareendranathan AR, Alves-Pereira F, Ghasseminia S, et al. 2D/3D ultrasound diagnosis of pediatric distal radius fractures by human readers vs artificial intelligence. *Sci Rep* 2023;13:14535. doi: 10.1038/s41598-023-41807-w
 23. Großbröhmer C, Bartram L, Rheinbay C, Heinrich MP, Tüshaus L. Leveraging semantic information for sonographic wrist fracture assessment within children. In: Deserno TM, Handels H, Maier A, Maier-Hein K, Palm C, Tolxdorff T, editors. *Bildverarbeitung für die Medizin 2023*. Wiesbaden: Springer Fachmedien; 2023. p. 102-7. doi: 10.1007/978-3-658-41657-7_23
 24. Hareendranathan AR, Tripathi A, Panicker MR, Zhang J, Boora N, Jaremko J. Deep learning approach for automatic wrist fracture detection using ultrasound bone probability maps. *SN Compr Clin Med* 2023;5:276. doi: 10.1007/s42399-023-01608-8
 25. Zhou Y, Hareendranathan AR, Keen C, Knight J, Jaremko J. Wrist ultrasound segmentation by deep learning. *Artif Intell Med* 2022 Jul 9. doi: 10.1007/978-3-031-09342-5_22
 26. Tripathi A, Panicker M, Zhang J, Boora N, Jaremko J, Hareendranathan AR. Domain specific transporter framework to detect fractures in ultrasound. *Annu Int Conf IEEE Eng Med Biol Soc* 2024;2024:1-4. doi: 10.1109/embc53108.2024.10782947
 27. Teng Y, Pan D, Zhao W. Application of deep learning ultrasound imaging in monitoring bone healing after fracture surgery. *J Radiat Res Appl Sci* 2023;16:100493. doi: 10.1016/j.jrras.2022.100493
 28. Zhang J, Boora N, Melendez S, Hareendranathan AR, Jaremko J. Diagnostic accuracy of 3D ultrasound and artificial intelligence for detection of pediatric wrist injuries. *Children* 2021;8:431. doi: 10.3390/children8060431
 29. Zhou Y, Knight J, Alves-Pereira F, Keen C, Hareendranathan AR, Jaremko JL. Wrist and elbow fracture detection and segmentation by artificial intelligence using point-of-care ultrasound. *J Ultrasound* 2025;28:465-73. doi: 10.1007/s40477-025-01019-6
 30. Sivrikaya S, Aksay E, Bayram B, Oray NC, Karakasli A, Altintas E. Emergency physicians performed point-of-care ultrasonography for detecting distal forearm fracture. *Turk J Emerg Med* 2016;16:98-101. doi: 10.1016/j.tjem.2016.04.002
 31. Snelling PJ, Jones P, Bade D, Bindra R, Davison M, Gillespie A, et al. Diagnostic accuracy of point-of-care ultrasound versus radiographic imaging for pediatric distal forearm fractures: a randomized controlled trial. *Ann Emerg Med* 2024;83:198-207. doi: 10.1016/j.annemergmed.2023.10.008
 32. Youseffard M, Baikpour M, Ghelichkhani P, Asady H, Darafarin A, Amini Esfahani MR, et al. Comparison of ultrasonography and radiography in detection of thoracic bone fractures; a systematic review and meta-analysis. *Emergency* 2016;4:55-64.
 33. Haak SL, Vos H, Borgstede MG, Boendermaker AE, Rietveld V, Kroon T, et al. Diagnostic accuracy of point-of-care ultrasound in detecting clavicle fractures. *Am J Emerg Med* 2025;88:156-61. doi: 10.1016/j.ajem.2024.11.008
 34. Ozturk P, Aksay E, Oray NC, Bayram B, Basci O, Tokgoz D. Emergency physician accuracy using ultrasonography to diagnose lateral malleolar fracture. *Am J Emerg Med* 2018;36:362-5. doi: 10.1016/j.ajem.2017.08.020
 35. Kuo RYL, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, et al. Artificial intelligence in fracture detection: a systematic review and meta-analysis. *Radiology* 2022;304:50-62. doi: 10.1148/radiol.211785
 36. Thian YL, Li Y, Jagmohan P, Sia D, Chan VEY, Tan RT. Convolutional neural networks for automated fracture detection and localization on wrist radiographs. *Radiol Artif Intell* 2019;1:e180001. doi: 10.1148/ryai.2019180001
 37. Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE. Machine learning for medical ultrasound: status, methods, and future opportunities. *Abdom Radiol* 2018;43:786-99. doi: 10.1007/s00261-018-1517-0
 38. Troxler D, Sanchez C, de Trey T, Mayr J, Walther M. Non-inferiority of point-of-care ultrasound compared to radiography to diagnose upper extremity fractures in children. *Children* 2022;9:1496. doi: 10.3390/children9101496
 39. Morello R, Mariani F, Snelling PJ, Buonsenso D. Point-of-care ultrasound for the diagnosis of distal forearm fractures in children and adolescents: a scoping review. *Eur J Pediatr* 2024;184:19. doi: 10.1007/s00431-024-05877-w
 40. Snelling PJ, Keijzers G, Byrnes J, Bade D, George S, Moore M, et al. Bedside ultrasound conducted in kids with distal upper limb fractures in the emergency department (BUCKLED):

a protocol for an open-label non-inferiority diagnostic randomised controlled trial. *Trials* 2021;22:282. doi: 10.1186/s13063-021-05239-z

41. Epema AC, Spanjer MJB, Ras L, Kelder JC, Sanders M. Point-of-care ultrasound compared with conventional

radiographic evaluation in children with suspected distal forearm fractures in the Netherlands: a diagnostic accuracy study. *Emerg Med J* 2019;36:613-6. doi: 10.1136/emmermed-2018-208380

Appendix – Supplementary Materials

Comprehensive search strategy:

Medline, Embase, CINAHL, Scopus, Web of Science, ACM Digital Library

Conference abstracts: yes

Preprints: yes

Clinical trial registry: yes

No date limit

No language limits

Ovid MEDLINE(R) ALL <1946 to August 15, 2025>

Date searched: Aug 17, 2025

Results: 75

Table S1. Ovid Medline search strategy for systematic review and meta-analysis

1.	exp artificial intelligence/ or machine learning/ or deep learning/ or supervised machine learning/ or support vector machine/ or unsupervised machine learning/ or latent class analysis/ or random forest/ or exp Cluster Analysis/
2.	("Artificial intelligence" or ((AI or "pattern recognition") adj8 (diagnos* or screen* or detect* or decision* or monitor* or insight* or algorithm* or tool or tools or analy* or program* or tech* or software or model* or "image enhancement")) or "deep learning" or "machine learning" or "predictive analytics" or "neural network*" or (classification adj2 (model or models or approach or approaches or algorithm*)) or "support vector machine" or "relevance vector machine" or "random forest" or "causal forest" or classifier* or "supervised learning" or "unsupervised learning" or "reinforcement learning" or "computer heuristic*" or "decision tree" or "feature detection" or (Markov adj3 model*) or "learning algorithm*" or ((multifactor* or multicriteria or multi-factor or multi-criteria) adj3 "decision analysis") or "nearest neighbo*" or kNN or k-NN or k-means or "neural net*" or "outlier detection" or "network learning" or "transfer learning" or "ensemble learning" or "Bayesian learning" or Bayesian method* or Bayesian network* or "gradient boost*" or (cluster* adj2 (analys* or algorithm*)) or clustering or ("computer vision" not "computer vision syndrome") or "sentiment analysis" or "probabilistic graphical model*" or "structured probabilistic model*" or "generative adversarial network*" or "dimension reduction" or "kernel method*" or "least squares kernel" or "multilayer perceptron" or "Multi-Layer Perceptron" or "principal component analysis" or "gradient boosting" or xgboost* or catboost* or "Pattern recognition" or "recurrence quantification analys*" or "functional data analys*" or "functional data boost*" or FDboost).mp.830664
3.	(classification or classifying).ti,ft. 105877
4.	1 or 2 or 3 956456
5.	Ultrasonography/ 210144
6.	(Ultrasound* or ultrasonograph*).mp.597367
7.	5 or 6 597367
8.	exp Fractures, Bone/ 222545
9.	(fracture* or ((broken or break or breaks) adj5 (femur* or femoral or ulnar* or ulna or humeral* or tibia* or fibula* or shoulder* or radius or elbow* or arm or arms or leg or legs or "upper extremit*" or "lower extremit*" or spine or spinal or back or neck or necks or skull or vertebra* or hip or hips or bone* or wrist or foot or feet or toe or toes or facial))).mp.
10.	8 or 9
11.	4 and 7 and 10 75

Embase <1974 to 2025 August 14> (OVID interface)

Date searched: Aug 17, 2025

Results: 144

Table S2. Embase search strategy for systematic review and meta-analysis

1.	exp artificial intelligence/ or exp machine learning/ or natural language processing/
2.	("Artificial intelligence" or ((AI or "pattern recognition") adj8 (diagnos* or screen* or detect* or decision* or monitor* or insight* or algorithm* or tool or tools or analy* or program* or tech* or software or model* or "image enhancement")) or "deep learning" or "machine learning" or "predictive analytics" or "neural network*" or (classification adj2 (model or models or approach or approaches or algorithm*)) or "support vector machine" or "relevance vector machine" or "random forest" or "causal forest" or classifier* or "supervised learning" or "unsupervised learning" or "reinforcement learning" or "computer heuristic*" or "decision tree" or "feature detection" or (Markov adj3 model*) or "learning algorithm*" or ((multifactor* or multicriteria or multi-factor or multi-criteria) adj3 "decision analysis") or "nearest neighbo*" or kNN or k-NN or k-means or "neural net*" or "outlier detection" or "network learning" or "transfer learning" or "ensemble learning" or "Bayesian learning" or Bayesian method* or Bayesian network* or "gradient boost*" or (cluster* adj2 (analys* or algorithm*)) or clustering or ("computer vision" not "computer vision syndrome") or "sentiment analysis" or "probabilistic graphical model*" or "structured probabilistic model*" or "generative adversarial network*" or "dimension reduction" or "kernal method*" or "least squares kernal" or "multilayer perceptron" or "Multi-Layer Perceptron" or "principal component analysis" or "gradient boosting" or xgboost* or catboost or "Pattern recognition" or "recurrence quantification analys*" or "functional data analys*" or "functional data boost*" or FDboost).mp.
3.	(classification or classifying).ti,bt. 120910
4.	1 or 2 or 3 1316167
5.	(Ultrasound* or ultrasonograph*).tw. 714012
6.	exp fracture/ 412476
7.	(fracture* or ((broken or break or breaks) adj5 (femur* or femoral or ulnar* or ulna or humeral* or tibia* or fibula* or shoulder* or radius or elbow* or arm or arms or leg or legs or "upper extremit*" or "lower extremit*" or spine or spinal or back or neck or necks or skull or vertebra* or hip or hips or bone* or wrist or foot or feet or toe or toes or facial))).mp.
8.	6 or 7 540644
9.	4 and 5 and 8 144

Scopus (Scopus Interface) (Advanced Search)

Date searched: Aug 17, 2025

Results: 174

Table S3. Scopus search strategy for systematic review and meta-analysis

1.	<p>((TITLE-ABS-KEY ("Artificial intelligence" OR ((AI OR "pattern recognition") W/8 (diagnos* OR screen* OR detect* OR decision* OR monitor* OR insight* OR algorithm* OR tool OR tools OR analy* OR program* OR tech* OR software OR model* OR "image enhancement")) OR "deep learning" OR "machine learning" OR "predictive analytics" OR "neural network*" OR (classification W/2 (model OR models OR approach OR approaches OR algorithm*)) OR "support vector machine" OR "relevance vector machine" OR "random forest" OR "causal forest" OR classifier* OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "computer heuristic*" OR "decision tree" OR "feature detection" OR (Markov W/3 model*) OR "learning algorithm*" OR ((multifactor* OR multicriteria OR multi-factor OR multi-criteria) W/3 "decision analysis") OR "nearest neighbo*" OR kNN OR k-NN OR k-means OR "neural net*" OR "outlier detection" OR "network learning" OR "transfer learning" OR "ensemble learning" OR "Bayesian learning" OR "Bayesian method*" OR "Bayesian network*" OR "gradient boost*" OR (cluster* W/2 (analys* OR algorithm*)) OR clustering OR "computer vision" OR "sentiment analysis" OR "probabilistic graphical model*" OR "structured probabilistic model*" OR "generative adversarial network*" OR "dimension reduction" OR "kernal method*" OR "least squares kernal" OR "multilayer perceptron" OR "Multi-Layer Perceptron" OR "principal component analysis" OR "gradient boosting" OR xgboost* OR catboost OR "Pattern recognition" OR "recurrence quantification analys*" OR "functional data analys*" OR "functional data boost*" OR FDboost) OR TITLE (classification OR classifying)) AND TITLE-ABS (Ultrasound* OR ultrasonograph*) AND TITLE-ABS-KEY (fracture* OR ((broken OR break OR breaks) W/5 (femur* OR femoral OR ulnar* OR ulna OR humeral* OR tibia* OR fibula* OR shoulder* OR radius OR elbow* OR arm OR arms OR leg OR legs OR "upper extremit*" OR "lower extremit*" OR spine OR spinal OR back OR neck OR necks OR skull OR vertebra* OR hip OR hips OR bone* OR wrist OR foot OR feet OR toe OR toes OR facial)))</p>
----	--

Web of Science Core Collection

Date searched: Aug 17, 2025

Results: 165

Table S4. Web of Science search strategy for systematic review and meta-analysis

1.	<p>(TS=("Artificial intelligence" OR ((AI OR "pattern recognition") NEAR/8 (diagnos* OR screen* OR detect* OR decision* OR monitor* OR insight* OR algorithm* OR tool OR tools OR analy* OR program* OR tech* OR software OR model* OR "image enhancement")) OR "deep learning" OR "machine learning" OR "predictive analytics" OR "neural network*" OR (classification NEAR/2 (model OR models OR approach OR approaches OR algorithm*)) OR "support vector machine" OR "relevance vector machine" OR "random forest" OR "causal forest" OR classifier* OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "computer heuristic*" OR "decision tree" OR "feature detection" OR (Markov NEAR/3 model*) OR "learning algorithm*" OR ((multifactor* OR multicriteria OR multi-factor OR multi-criteria) NEAR/3 "decision analysis") OR "nearest neighbo*" OR kNN OR k-NN OR k-means OR "neural net*" OR "outlier detection" OR "network learning" OR "transfer learning" OR "ensemble learning" OR "Bayesian learning" OR "Bayesian method*" OR "Bayesian network*" OR "gradient boost*" OR (cluster* NEAR/2 (analys* OR algorithm*)) OR clustering OR "computer vision" OR "sentiment analysis" OR "probabilistic graphical model*" OR "structured probabilistic model*" OR "generative adversarial network*" OR "dimension reduction" OR "kernel method*" OR "least squares kernel" OR "multilayer perceptron" OR "Multi-Layer Perceptron" OR "principal component analysis" OR "gradient boosting" OR xgboost* OR catboost OR "Pattern recognition" OR "recurrence quantification analysis*" OR "functional data analysis*" OR "functional data boost*" OR FDboost) OR TI=(classification OR classifying)) AND TS=(Ultrasound* OR ultrasonograph*) AND TS=(fracture* OR ((broken OR break OR breaks) NEAR/5 (femur* OR femoral OR ulnar* OR ulna OR humeral* OR tibia* OR fibula* OR shoulder* OR radius OR elbow* OR arm OR arms OR leg OR legs OR "upper extremit*" OR "lower extremit*" OR spine OR spinal OR back OR neck OR necks OR skull OR vertebra* OR hip OR hips OR bone* OR wrist OR foot OR feet OR toe OR toes OR facial)))</p>
----	---

CINAHL Plus with Full Text (EBSCOhost interface)

Date searched: Aug 17, 2025

Results: 22

Table S5. CINAHL search strategy for systematic review and meta-analysis

S1	(MH "Artificial Intelligence") OR (MH "Machine Learning+") OR (MH "Artificial Intelligence, Generative+") OR (MH "Intelligent Systems+") OR (MH "Natural Language Processing") OR (MH "Neural Networks (Computer)+") OR ("Artificial intelligence" OR ((AI OR "pattern recognition") N8 (diagnos* OR screen* OR detect* OR decision* OR monitor* OR insight* OR algorithm* OR tool OR tools OR analy* OR program* OR tech* OR software OR model* OR "image enhancement")) OR "deep learning" OR "machine learning" OR "predictive analytics" OR "neural network*" OR (classification N2 (model OR models OR approach OR approaches OR algorithm*)) OR "support vector machine" OR "relevance vector machine" OR "random forest" OR "causal forest" OR classifier* OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "computer heuristic*" OR "decision tree" OR "feature detection" OR (Markov N3 model*) OR "learning algorithm*" OR ((multifactor* OR multicriteria OR multi-factor OR multicriteria) N3 "decision analysis") OR "nearest neighbo*" OR kNN OR k-NN OR k-means OR "neural net*" OR "outlier detection" OR "network learning" OR "transfer learning" OR "ensemble learning" OR "Bayesian learning" OR "Bayesian method*" OR "Bayesian network*" OR "gradient boost*" OR (cluster* N2 (analys* OR algorithm*)) OR clustering OR "computer vision" OR "sentiment analysis" OR "probabilistic graphical model*" OR "structured probabilistic model*" OR "generative adversarial network*" OR "dimension reduction" OR "kernal method*" OR "least squares kernal" OR "multilayer perceptron" OR "Multi-Layer Perceptron" OR "principal component analysis" OR "gradient boosting" OR xgboost* OR catboost OR "Pattern recognition" OR "recurrence quantification analys*" OR "functional data analys*" OR "functional data boost*" OR FDboost) OR TI(classification OR classifying)
S2	(MH "Ultrasonography") OR (Ultrasound* OR ultrasonograph*)
S3	(MH "Fractures+") OR (fracture* OR ((broken OR break OR breaks) N5 (femur* OR femoral OR ulnar* OR ulna OR humeral* OR tibia* OR fibula* OR shoulder* OR radius OR elbow* OR arm OR arms OR leg OR legs OR "upper extremit*" OR "lower extremit*" OR spine OR spinal OR back OR neck OR necks OR skull OR vertebra* OR hip OR hips OR bone* OR wrist OR foot OR feet OR toe OR toes OR facial)))
S4	1 AND 2 AND 3

Copyright: The Author(s), 2026. Licensee Mattioli 1885, Fidenza, Italy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License (CC BY-NC-4.0).

Disclaimer/Publisher's Note: The statements, opinions, and data contained in this article are solely those of the author(s) and contributor(s) and do not necessarily reflect those of their affiliated organizations, the publisher, the editors, or the reviewers. The publisher and the editors disclaim any responsibility for injury to people or property resulting from any ideas, methods, instructions or products mentioned in the content. Any product that may be evaluated in this article, or claim made by its manufacturer, is not guaranteed or endorsed by the publisher.