The British Institute of Radiology

OXFORD

Commercially available artificial intelligence tools for fracture detection: the evidence

Cato Pauling, MEng[1,*], Baris Kanber [iD] , PhD[2,3], Owen J. Arthurs [iD] , PhD[1,4,5],
Susan C. Shelmerdine [iD] , PhD[1,4,5]

[1]UCL Great Ormond Street Institute of Child Health, University College London, London WC1E 6BT, United Kingdom
[2]Queen Square Multiple Sclerosis Centre, Department of Neuroinflammation, University College London (UCL) Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London WC1N 3BG, United Kingdom
[3]Department of Medical Physics and Biomedical Engineering, Centre for Medical Image Computing, University College London, London WC1E 6BT, United Kingdom
[4]Department of Clinical Radiology, Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, United Kingdom
[5]NIHR Great Ormond Street Hospital Biomedical Research Centre, Bloomsbury, London WC1N 1EH, United Kingdom
*Corresponding author: Cato Pauling, MEng, University College London, Gower Street, London WC1E 6BT, United Kingdom (cato.pauling.22@ucl.ac.uk)

Abstract

Missed fractures are a costly healthcare issue, not only negatively impacting patient lives, leading to potential long-term disability and time off work, but also responsible for high medicolegal disbursements that could otherwise be used to improve other healthcare services. When fractures are overlooked in children, they are particularly concerning as opportunities for safeguarding may be missed. Assistance from artificial intelligence (AI) in interpreting medical images may offer a possible solution for improving patient care, and several commercial AI tools are now available for radiology workflow implementation. However, information regarding their development, evidence for performance and validation as well as the intended target population is not always clear, but vital when evaluating a potential AI solution for implementation. In this article, we review the range of available products utilizing AI for fracture detection (in both adults and children) and summarize the evidence, or lack thereof, behind their performance. This will allow others to make better informed decisions when deciding which product to procure for their specific clinical requirements.

Keywords: machine learning; artificial intelligence; fracture; commercial; radiology; imaging.

Introduction

Missed fractures impact both patients and healthcare providers. Between 2015 and 2018, the total cost of missed fracture claims in the NHS was over £1.1 million, with an average cost per claim of approximately £14 000.[1] Although the number of claims were relatively few ($n = 78$) compared to over 1 million fracture attendances per year across the United Kingdom, they are seen as an avoidable cost, money which could be better spent improving other NHS services. Furthermore, missed fractures in young children pose an additional challenge, as these may reflect failed opportunities for safeguarding and referral to social services.

One means to reduce such errors may be to incorporate assistance from artificial intelligence (AI). Several systematic reviews and meta-analyses have investigated the use of AI for the detection of fractures. In adults, high sensitivity rates of 92% for plain radiographs and 89% for computerised tomography (CT) scans have been reported, with specificities of 91% for plain radiographs and 92% for CT scans[2-4]; with accuracy rates of between 89% and 98% in children.[5] Despite such promising results, over half of these studies had a high risk of bias,[2] were only used in a research setting, were not ready for clinical deployment, nor externally validated.[6] This is particularly concerning given that 24% of AI solutions in one study yielded a substantial decline in performance when evaluated on external data (compared to their internal data

set) and the majority (81% of algorithms) reported negative impact.[7]

Nonetheless, several AI vendors have now developed fracture detection models for routine practice (Figures 1–3), ready for commercial integration into radiological workflows. Although this brings cutting-edge technology a step closer to direct patient benefit, it is vital that such tools are independently evaluated prior to adoption. Of concern, van Leeuwen et al[8] found that in a review of 100 Conformité Européenne (CE) marked AI tools for different radiology use cases, only 36% had any associated peer reviewed verification of performance, of which fewer than half were independent of the vendor (ie, without flagrant conflict of interest). Without independent evidence, it can be very challenging to know how well an AI product might perform in a given clinical setting, and whether it is "worth" purchasing.[9]

In this market research review, we assess the range of available commercial products utilizing AI for fracture detection (in both adults and children) and summarize the evidence behind the performance of these tools. This will allow others to make better informed decisions when deciding which AI product, if any, to procure for their specific clinical requirements.

Methods

A search of commercially available AI solutions for fracture detection using medical images was performed by the lead

and senior author, using the methods detailed below to ensure as comprehensive a search as possible.

1) A search of both CE[8,10] certified and Food and Drug Administration (FDA) approved sites[11,12] were filtered for medical products that utilized machine learning (ML) algorithms, and then further refined to incorporate those that included the term "fractures" within the diseases targeted by the product.
2) A review of AI exhibitors and sponsors at several large annual radiology conferences in 2022 were identified relating to general, paediatric, and musculoskeletal imaging (ie, Radiological Society of North America—RSNA, European Congress of Radiology—ECR, European Society of Paediatric Radiology—ESPR, European Society for Skeletal Radiology—ESSR) to determine if they provided solutions for fracture detection.
3) A search of peer reviewed publications within the PubMed, Scopus, and EMBASE databases published between January 1, 2012 and December 31, 2022 using a Boolean search strategy to find those articles with keywords relating to "machine learning," "artificial intelligence," "imaging," and "fracture." This was then further manually reviewed for specific mention and evaluation of commercial AI solutions.

There were no restrictions placed on the type of imaging modality, body parts targeted, type of AI/ML methodology or the intended population for the AI tool. Products and applications returned through these search strategies were deemed eligible if they supported healthcare professionals at image diagnosis, triage, classification/detection, with respect to fracture detection. We excluded products that were advertised via software marketplace redistributions, and those that were categorized under medical image management and processing systems. Only software that was currently commercially available was included in the main analysis. Products which were still in development or at prototype stage were included separately in case of future relevance to readers. Applications that have been withdrawn from the market or no longer available were not included.

For each product, information was gathered through numerous sources. Company websites, FDA/CE certification documents, user manuals, and scientific articles were first collated to collect data about the developer, technical specifications, specific functionalities relating to clinical application, and any evidence that exists to support the product's performance. Levels of evidence for each AI solution were further classified into 6 levels according to an adapted hierarchical model of efficacy by Fryback and Thornbury,[13] and used in a prior article evaluating evidence for commercial AI products (Table 1).[8]

To ensure accuracy, timeliness and comprehensiveness of online information, all relevant AI vendors were contacted directly and further supplied a survey of questions (see Supplementary Material) to complete. A timeframe of 2 weeks was provided for return of the survey, with the option of having follow-up emails and online meetings to better discuss our survey queries, if preferred. We also contacted the MHRA directly with the vendor names and AI solutions to confirm whether United Kingdom Conformity Assessed (UKCA) certification had been awarded for the products in question.

Results

In total, 21 commercial AI products across 15 different AI vendors (Tables 2-5) were identified, with a further 3 products across 3 companies at prototype/pre-market stage. These prototypes comprised 2 products which could detect fractures on radiographs (Fraxpert, SeeAI,[54] Imera-MSK, Imera.ai[55]) and one for rib fractures on CT imaging (Dr. Wise@ChestFracture v1.0, no vendor name identified[56]).

The majority of the commercial AI products (14/21) were intended for fracture detection on plain radiographs (Figures 1–3), with the remainder (7/21) related to CT evaluation. Only 3 products specified they were intended for use in adults and children (all relating to radiographic interpretation), with the remainder intended solely for adult use. All products were intended to aid human interpretation or triage, not for autonomous usage (at this stage of their development or regulation).

Evidence levels

Predominantly, the AI products reviewed had evidence for their performance provided by the vendor for conformity certification, with 7 having independent, peer reviewed publications available (total publications = 18), with the greatest number originating from Gleamer (*n* = 8). The majority of the evidence levels for AI product performance were at Level 3 (ie, change in diagnosis with and without AI assistance) and with some products (eg, BoneView *Gleamer*,[36,38,39] Annalise Enterprise CXR v1.2 *Annalise.AI*,[46] Qure.AI *qXR*[48]) potentially demonstrating evidence at Level 4 (ie, demonstrating improvement in time for diagnosis, which could be argued may lead to swifter treatment or follow-up for the patient[38]). There was no evidence available to demonstrate a benefit in actual patient outcome (eg, reduced time for recovery, reduction in repeated hospital visits, etc.), nor any publications on the health economic cost savings. Only one external validation article specifically mentioned changes to patient management (for Annalise Enterprise CXR v1.2 *Annalise.AI*).[46] A summary of the available evidence associated with each product is reviewed in Table 4.

The largest independently published study to demonstrate improvement in human diagnostic accuracy with AI for fracture detection, included 480 radiographs (60 radiographs across 8 body parts; 50% abnormal) across 24 readers (comprising radiologists, emergency physicians, orthopaedic surgeons, and other healthcare professionals).[38] There was an overall improvement in sensitivity rates across all specialists with AI assistance of 10.4%, and shortened reading by 6.3 s per examination.

Only one of the AI products (HealthVCF, Nanox.Ai) was reviewed by the National Institute for Health and Care Excellence (NICE) in a Medtech innovation briefing document,[34] based on a published conference abstract[57] and one peer-reviewed article,[58] regarding the use of AI for the assessment of vertebral compression fractures on CT imaging. Whilst the NICE experts accepted that there would be clear patient benefit from the detection of vertebral compression fractures, and that the evidence was promising, it was nonetheless limited; the only published article was funded by the company.

Externally conducted studies that verify the performance of AI algorithms based on CT input are severely lacking.

Table 1. Hierarchical model of efficacy to assess the contribution of AI software to the diagnostic imaging process, reproduced from van Leeuwen et al,[8] originally adapted from Fryback and Thrornbury (1991)[13] under the Creative Commons Attribution 4.0 International Licence.[14]

Level	Explanation	Typical measures
Level 1t	Technical efficacy Article demonstrates the technical feasibility of the software	Reproducibility, inter-software agreement, error rate
Level 1c	Potential clinical efficacy Article demonstrates the feasibility of the software to be clinically applied	Correlation to alternative methods, potential predictive value, biomarker studies
Level 2	Diagnostic accuracy efficacy Article demonstrates the stand-alone performance of the software	Standalone sensitivity, specificity, area under the ROC curve, or Dice score
Level 3	Diagnostic thinking efficacy Article demonstrates the added value to the diagnosis	Radiologist performance with/without AI, change in radiological judgement
Level 4	Therapeutic efficacy Article demonstrates the impact of the software on the patient management decisions	Effect on treatment or follow-up examinations
Level 5	Patient outcome efficacy Article demonstrates the impact of the software on patient outcomes	Effect on quality of life, morbidity, or survival
Level 6	Societal efficacy Article demonstrates the impact of the software on society by performing an economic analysis	Effect on costs and quality-adjusted life years, incremental costs per quality-adjusted life year

Abbreviations: Level 1t = level 1, technical, Level 1c = level 1, clinical.

Only 3 such studies were identified, of which 2 are for the same product (Aidoc *C-Spine [CSF]*[51,52]). The study that included the largest number of cases involved 1904 CT scans and the performance of the AI algorithm was assessed against interpretation of the scans by a single attending neuroradiologist. The AI and neuroradiologist had agreeing reports in 91.5% of all cases. The AI was able to correctly identify 67 of 122 fracture cases (54.9%) and returned 106 cases that were false positives. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the AI algorithm were 54.9% (95% CI, 45.7-63.9), 94.1% (95% CI, 92.9-95.1), 38.7% (95% CI, 33.1-44.7), and 96.8% (95% CI, 96.2-97.4), respectively. The researchers also analysed the misdiagnosed fractures, finding that cases of chronic fractures were overrepresented, suggesting the AI algorithm is not well-adapted to handle this presentation.[52]

Evidence for usage in children

All AI products were intended for use in adults, with independent peer-reviewed evidence for accuracy of the product in children (and younger adults) available for 2 vendors (AZMed and Gleamer). In one study evaluating the performance of the Rayvolve product (AZMed),[44] a retrospective review of 2634 radiographs across 2549 children (<18 years age) from one single French centre was performed. This demonstrated an overall sensitivity of 95.7%, specificity of 91.2%, and accuracy of 92.6% for presence/absence of a fracture (regardless of number and whether the fracture was correctly localized or not). There was some reduction in the accuracy of the product for children aged <4 years old and those within a cast. While there was a similar sensitivity of fracture detection in patients with cast compared to without cast of 95.3% and 93.9% (1.4% difference), respectively, the difference in specificity was significant at 30.0% and 89.5% (50.9% difference), respectively. The accuracy of fracture detection also decreased to 83.0% in the case of patients with casts, compared to 90.7% in patients without cast (a

difference of 7.7%). Results in fracture detection from the 0-4 years and 5-18 years age subgroups also showed a difference with sensitivity of 90.5% (0-4 years), and 95.4% (5-18 years) (difference of 4.9%). The specificity did not demonstrate significant differences at 88.9% (0-4 years) and 88.8% (5-18 years), however, the accuracy did slightly decrease to 89.3% (0-4 years), compared to 90.7% (5-18 years) (difference of 1.4%).

Two publications evaluated the use of the BoneView product (Gleamer) in children and young adults[40,43] using the same data set of 300 musculoskeletal radiographs (half with fractures, in patients aged 2-21 years old) across 5 body parts acquired from a United States-based data provider. In the first study,[43] an external validation of the AI product alone was performed demonstrating sensitivity of 91.3%, specificity 90%, and patient-wise AUROC of 0.93. Avulsion fractures were noted to be challenging for the AI tool to detect (sensitivity per fracture of 72.7%). In the second publication,[40] differences in radiologist performance before and after AI assistance were evaluated across 8 radiologists (5 radiologists in training and 3 qualified paediatric radiologists). Across all 8 radiologists, the mean sensitivity was 73.3% without AI; and increased by almost 10% (*P* < .001) to 82.8% with AI. There was a statistically significant improvement in sensitivity for radiologists in training (by 10.3% [*P* < .001]) compared to specialist paediatric radiologists (8.2% [*P* = .08]) demonstrating greater benefit for less experienced radiologists.

Conformity certification

For medical devices to be commercialized in different countries, different types and levels of conformity certification are mandatory. These do not necessarily guarantee the safety or efficacy of a product, more that it has been assessed and found to meet a certain minimum requirement. A list of the certifications for various products in this review is listed in Table 3.

Table 2. Use cases and the intended population for commercial AI tools for fracture detection on medical imaging.

Company	Product, version	Modality	Disease(s) targeted	Fracture type	Notable inclusions (for fractures)	Notable exclusions (for fractures)	Target population
Gleamer	BoneView 1.1-US	Radiography	Bone fractures	Acute and healing	Ankle, foot, knee, tibia/fibula, wrist, hand, elbow, forearm, humerus, shoulder, clavicle, pelvis, hip, femur, ribs, thoracic spine, and lumbosacral spine	Cervical spine and skull radiographs	Adults (>21 years) for all body parts and children/adolescents (2-21 years) for all body parts except pelvis, hip, femur, ribs, thoracic spine, and lumbosacral spine
Radiobotics	RBfracture v.1	Radiography	Bone fractures	Acute and healing	Appendicular skeleton only	Spine, rib, and craniofacial fractures	Vendor states product is intended for adult and paediatric use (>2 years)
AZmed	Rayvolve v2.5.0	Radiography	Traumatic injuries (fractures, dislocations, joint effusions) and chest pathologies (pneumothoraces, cardiomegaly, pleural effusions, pulmonary oedema, consolidation, nodules)	Acute and healing	–	Dental, facial, skull, and spine radiographs.	Adult only according to FDA clearance, however, evidence for performance in children has been conducted
Rayscape	Chest X-ray	Radiography	17 classes of pathologies on chest radiographs, of which one is fractures (others include pulmonary- and cardiac-related findings, as well as scoliosis)	Acute	Only frontal AP or PA chest radiographs	Not lateral chest X-rays	Adults (>16 years old)
Milvue	SmartUrgences	Radiography	7 pathologies, including bone fractures, joint effusions, joint dislocations (for musculoskeletal radiographs) and pleural effusions, pneumothorax, pulmonary opacification, pulmonary nodules (on chest radiographs)	Acute	–	Excludes axial skeleton, dental, and abdominal radiographs	Vendor states product is intended for adult and paediatric use (lower age limit not defined)
Imagen Technologies	OsteoDetect	Radiography	Wrist fractures	Acute	Distal radial fractures only	–	Adults (>22 years)
Imagen Technologies	FractureDetect	Radiography	Fractures (upper/lower extremities)	Acute	Only includes ankle (frontal, lateral), clavicle (frontal, lateral, oblique), elbow (frontal, lateral), femur (frontal, lateral), forearm		Adults (>22 years)

(continued)

Table 2. (continued)

Company	Product, version	Modality	Disease(s) targeted	Fracture type	Notable inclusions (for fractures)	Notable exclusions (for fractures)	Target population
					(frontal, lateral), hip (frontal, frog leg lateral), humerus (frontal, lateral), knee (frontal, lateral), pelvis (frontal), shoulder (frontal, lateral, axillary), tibia/fibula (frontal, lateral), wrist (frontal, lateral, oblique)		
Annalise.AI	Annalise Enterprise CXR v1.2	Radiography	124 pathologies on chest radiographs (including fractures and bone lesions)	Acute and healing	Clavicle, spine, ribs, humerus, scapula	–	Adults >16 years
Deepnoid	DEEP: SPINE-CF-01	Radiography and MRI	Major spinal abnormalities such as compression fractures, scoliosis angles, intervertebral disc abnormalities	Acute	Spine only	–	Adult (>20 years)
Quibim	Chest X-Ray Classifier	Radiography	Atelectasis, cardiomegaly, consolidation, oedema, emphysema, enlarged cardiomediastinum, fibrosis, fracture, hernia, lung lesion, lung opacity, pleural effusion, pleural thickening, pneumothorax	Acute	Rib fractures	–	Age not specified, presumed only adults
SenseTime	SenseCare-Chest DR Pro	Radiography	Pneumonia, tuberculosis, pneumothorax, pleural effusion, cardiomegaly, rib fractures	Acute	Rib fractures	–	Age not specified, presumed only adults
Infervision	InferRead DR Chest v1.0.1.1	Radiography	Lung cancer, pneumothorax, fracture, tuberculosis, lung infection, aortic calcification, cord imaging, heart shadow enlargement, pleural effusion.	Acute	Rib fractures	–	Adults >16 years
Qure.AI	qXR	Radiography	30 findings including lung nodules, pneumothorax, pleural effusions, rib fractures, and pneumoperitoneum	Acute	Rib fractures only	–	Age not specified, presumed only adults

(continued)

Table 2. (continued)

Company	Product, version	Modality	Disease(s) targeted	Fracture type	Notable inclusions (for fractures)	Notable exclusions (for fractures)	Target population
Qure.AI	qMSK	Radiography	Bone fractures and joint dislocations	Acute	Wrist, hand, finger, fibula and tibia, ankle, foot, shoulder, ribs, forearm	–	Age not specified, presumed only adults
SenseTime	SenseCare Lung CT	CT	Pulmonary nodules, pneumonia (including COVID-19) lesions and fractures	Acute	Rib fractures	–	Age not specified, presumed only adults
Infervision	InferRead CT Rapid Triage v1	CT	Coronary artery stenosis, chest fracture, ICH	Acute	Rib fractures	–	Age not specified, presumed only adults
Qure.AI	qER	CT	11 findings, including cranial fractures	Acute	Skull only	–	Adults
Aidoc	Briefcase for C-Spine Fracture triage (CSF)	CT	Cervical spine fractures	Acute	–	–	Age not specified, presumed only adults
Aidoc	BriefCase for Rib Fracture (RibFx)	CT	Rib fractures	Acute	–	–	Age not specified, presumed only adults
Nanox.AI	Bone Health Solution/ HealthVCF	CT	Vertebral compression fractures	Acute and healing	–	–	Adults intended (>50 years)
Shanghai United Imaging Intelligence	uAI EasyTriage-Rib	CT	Rib fractures (detects when there are 3 or more fractures, not fewer)	Acute	–	–	Age not specified, presumed only adults

Abbreviations: MRI = magnetic resonance imaging, CT = computerized tomography, where available, the latest version of the AI product is provided.

Table 3. Licencing and user details of commercial AI tools for fracture detection on medical imaging.

Company	Product	CE certification	CE class	FDA certification	FDA class	FDA clearance date	Number of claimed users	Pricing strategy
Gleamer	BoneView 1.1-US	–	–	510(k)	2	January 31, 2023	Unknown	Annual or multi-year subscription (number of users, number of installations, number of analyses)
Gleamer	BoneView v2.0.2[a]	MDD (MDR pending)	2a	510(k) (Adult only)	2	March 1, 2022	>550	Annual or multi-year subscription (number of users, number of installations, number of analyses)
Radiobotics	RBfracture v.1	MDR	2a	–			>10	Pay-per-use, subscription, one-time license fee (number of analyses)
AZmed	Rayvolve v2.5.0	MDR	2a	510(k) (Adult only)	2	June 2, 2022	>700	Fixed-price annual subscription based on patient volume-try (trauma examinations per year). Free trial phase available
Rayscape	Chest X-ray	MDD	1	–			>100	Subscription (number of analyses)
Milvue	SmartUrgences	MDR	2a	–			>10	Subscription (number of users, number of installations, number of analyses)
Imagen Technologies	OsteoDetect	–		510(k) (Adult only)	2	May 24, 2018	Unknown	Unknown
Imagen Technologies	FractureDetect	–		510(k) (Adult only)	2	July 30, 2020	Unknown	Unknown
Annalise.AI	Annalise Enterprise CXR v1.2	MDR	2b	510(k) (Pneumothorax only)	2	February 24, 2022	>300	Subscription (number of analyses)
Quibim	Chest X-Ray Classifier	MDD	2a	–			Unknown	Licence (number of installations, number of analyses)
SenseTime	SenseCare-Chest DR Pro	MDR	2b	–			Unknown	Subscription, pay-per-use (number of users, number of installations, number of analyses)

Table 3. (continued)

Company	Product	CE certification	CE class	FDA certification	FDA class	FDA clear-ance date	Number of claimed users	Pricing strategy
Infervision	InferRead DR Chest v1.0.1.1	MDD	2a	–	–	–	Unknown	Subscription (number of installations)
Qure.ai	qXR	MDR	2b	510(k) (*Breathing tubes only*)	2	November 22, 2021	>1000 (*across all products*)	Pay-per-use
Qure.ai	qMSK	MDR	2b	–	–	–	>1000 (*across all products*)	Pay-per-use
SenseTime	SenseCare Lung CT	MDR	2b	–	–	–	Unknown	Subscription, pay-per-use (number of users, number of installations, number of analyses)
Qure.ai	qER	MDR	2b	510(k)	2	June 11, 2020	>1000 (*across all products*)	Pay-per-use
Aidoc	Briefcase for C-Spine Fracture triage (CSF)	MDD	1	510(k)	2	May 31, 2019	>500	Subscription (total imaging volume)
Aidoc	BriefCase for Rib Fracture (RibFx)	MDD	1	510(k)	2	April 14, 2021	>500	Subscription (total imaging volume)
Nanox.AI	Bone Health Solution/ HealthVCF	MDD	2a	510(k)	2	May 12, 2020	Unknown	Subscription (£38 000 to £90 000 p.a.—number of analyses)
Shanghai United Imaging Intelligence	uAI EasyTriage-Rib	MDD	2a	510(k)	2	January 15, 2021	Unknown	Unknown

Abbreviations: MDD = Medical Devices Directive (pre May 26, 2021), MDR = Medical Devices Regulation (post May 26, 2021).
Deepnoid's Deep: Spine-CF-01 only has approval from the Korean Ministry of Food and Drug Safety (No. 19-550).
[a]Previous version of the software from the company.

Table 4. Evidence for AI performance, based on FDA/CE conformity documentation or vendor endorsed studies.

Company	Product	Modality	Type of evidence	Predicate device (FDA)	Single/multicentre data	Readers/data set	Summary of evidence	Level of evidence	Ref.
Gleamer	BoneView 1.1-US	Radiography	FDA approval documentation	BoneView 1.0-US	Multicentre	2000 paediatric radiographs	High-sensitivity operating point: sensitivity 0.909 [95% CI, 0.889-0.926]. High-specificity operating point: specificity 0.965 [95% CI, 0.952-0.976].	Level 2	15
Gleamer	BoneView 1.1-US	Radiography	FDA approval documentation	BoneView 1.0-US	Multicentre	8918 adult radiographs	High-sensitivity operating point: sensitivity 0.928 [95% CI, 0.919-0.936] High-specificity operating point: specificity 0.932 [95% CI, 0.925-0.939]	Level 2	15
Gleamer	BoneView 1.1-US	Radiography	FDA approval documentation	BoneView 1.0-US	MCMR	480 cases, 14 clinical researchers	Reader specificity improved from 0.906 (95% CI, 0.898-0.913) to 0.956 (95% CI, 0.951-0.960). Reader sensitivity improved from 0.648 (95% CI, 0.640-0.656) to 0.752 (95% CI, 0.745-0.759).	Level 3	15,16
Gleamer	BoneView v2.0.2[a]	Radiography	FDA approval documentation	Imagen Technologies—FractureDetect	–	24 readers 480 examinations	Specificity improved from 0.906 to 0.956 Sensitivity improved from 0.648 to 0.752	Level 3	17
Radiobotics	RBfracture v.1	Radiography	Vendor conducted study	–	Multicentre (United States and Denmark)	8 readers 312 examinations	Sensitivity improved by 5.6% 47% reduction in false negative findings. AUROC was 0.97	Level 3	18
Azmed	Rayvolve v2.5.0	Radiography	FDA approval documentation	Imagen Technologies—FractureDetect	–	24 readers 186 examinations	Sensitivity improved from 0.86561 to 0.9554; Specificity improved from 0.82645 to 0.83116; AUROC improved from 0.84602 to 0.89327	Level 3	19
Rayscape	Chest X-ray	Radiography	Medical white paper by vendor	–	–	–	AUROC of 90.2.	Level 2	20
Milvue	SmartUrgences	Radiography	Medical white paper by vendor	–	Multicentre	8 readers 650 examinations	Sensitivity of junior radiologist with AI improved from 92% to 95%; Sensitivity of senior radiologists with AI improved from 93% to 95%	Level 3	21,22
Imagen Technologies	OsteoDetect	Radiography	FDA approval documentation	Not stated	Multicentre	24 readers 200 examinations	AUROC improved from 0.840 to 0.889; Sensitivity improved from 0.747 to 0.803; Specificity improved from 0.889 to 0.914.	Level 3	23

(continued)

Table 4. (continued)

Company	Product	Modality	Type of evidence	Predicate device (FDA)	Single/multicentre data	Readers/data set	Summary of evidence	Level of evidence	Ref.
Imagen Technologies	FractureDetect	Radiography	FDA approval documentation	Imagen Technologies—OsteoDetect		11 970 examinations for diagnostic accuracy 24 readers, 175 examinations for MRMC study	Sensitivity 0.951 (95% CI, 0.940-0.960), Specificity 0.893 (95% CI, 0.886-0.898), AUROC 0.982 (95% Bootstrap CI, 0.9790, 0.9850). AUROC improved from 0.912 to 0.952; Sensitivity improved from 0.819 to 0.900; Specificity improved from 0.890 to 0.918	Level 3	24
Imagen Technologies	FractureDetect	Radiography	Vendor conducted study	–	Multicentre (United States)	24 clinicians, 175 cases	Improved AUROC from 0.90 [95% CI, 0.89-0.92] to 0.94 [95% CI, 0.93-0.95]. Improved sensitivity from 82% [95% CI, 79-84] to 90% [95% CI, 88-92]. Improved specificity from 89% [95% CI, 88-90] to 92% [95% CI, 91-93]	Level 3	25
Annalise.AI	Annalise Enterprise CXR v1.2	Radiography	Vendor conducted study	–	–	20 readers 4568 examinations in 2568 patients	AUROC score of 0.713 improved to 0.808 with AI.	Level 3	26
Deepnoid	DEEP: SPINE-CF-01	Radiography and MRI	Vendor endorsed study	–	–	160 radiographs	Dice 91.60% Sensitivity 90.12% Specificity 99.59% Precision 84.57%	Level 2	27
Qure.AI	qER	CT	FDA approval documentation	Aidoc's Briefcase Software	Multicentre (United States)	1320 CT scans	Cranial fracture sensitivity of 96.77%, specificity of 92.72%, and AUROC of 0.9766	Level 2	28
Qure.AI	qER	CT	Vendor presentation	–	–	2971 scans	Per-scan image only AUROC was 0.72 and image with haemorrhage feature AUROC was 0.83.	Level 2	29
Qure.AI	qER	CT	Vendor poster	–	–	18 200 scans	Images only AUROC was 0.9599 and AP 0.7952 Images with scalp hematoma indicator AUROC was 0.9666 and AP 0.8190.	Level 2	30
Aidoc	Briefcase for C-Spine Fracture triage (CSF)	CT	FDA approval documentation	Aidoc Briefcase for ICH triage	Multicentre (3 sites)	186 examinations	Sensitivity of 91.7% Specificity of 88.6%	Level 2	16

(continued)

Table 4. (continued)

Company	Product	Modality	Type of evidence	Predicate device (FDA)	Single/ multicentre data	Readers/data set	Summary of evidence	Level of evidence	Ref.
Aidoc	BriefCase for Rib Fracture (RibFx)	CT	FDA approval documentation	Aidoc BriefCase for PE triage	Multicentre (3 sites)	279 examinations	Sensitivity of 96.7% Specificity of 90.4% AUROC of 0.976	Level 2	31,32
Nanox.AI	Bone Health Solution/ HealthVCF	CT	FDA approval documentation	cmTriage	Multicentre (United States and Israel)	611 examinations	AUROC of 0.9504 (95% CI, 0.9348-0.9660) Sensitivity of 90.2% (95% CI, 86.3-93.05) Specificity of 86.89% (95% CI, 82.63-90.22) Average performance time was 61.36 seconds.	Level 3	33
Nanox.AI	HealthVCF	CT	NICE review (includes vendor funded study)				Specificity of 94% Sensitivity of 59% and Major osteoporotic fractures: 66.5% sensitivity 64.7% specificity Hip fractures: 92.6% sensitivity 36.9% specificity	Level 2	34
Shanghai United Imaging Intelligence	uAI EasyTriage-Rib	CT	FDA approval documentation	NanoxAI, HealthVCF	Multicentre	200 examinations	Sensitivity of 92.7% Specificity of 84.7% AUROC of 0.939. Time to notification of all 76 positive findings in study were 69.56 seconds.	Level 3	35

Preferential evidence provided where MRMC (multireader, multicase) studies were conducted to evaluate clinical improvement (rather than standalone bench testing results). A hyphen denotes either unknown/not stated or not applicable information. Although the product for Annalise.ai does have FDA approval, this is only for detection of pneumothoraces rather than fractures, therefore, the FDA clearance evidence is not included below. Similarly, Qure.AI has FDA approval for their qXR product for "breathing tube placement" analysis only, therefore, the FDA clearance evidence is not included below.
aPrevious version of the software from the company.

Table 5. Evidence for AI performance, based on independent external peer reviewed publications.

Company	Product	Modality	Type of evidence	Readers/data set	Summary of evidence	Level of evidence	Ref.
Gleamer	BoneView v2.0.2[a]	Radiography	Retrospective study	500 patients, 3 radiologists	Sensitivity increased by 20% and specificity increased by 0.6%. PPV increased by 2.9% and the NPV by 10%. AUROC increased by 10.2%. Decreased mean reading time by 12.7 seconds.	Level 4	36
Gleamer	BoneView v2.0.2[a]	Radiography	Retrospective study	4774 radiographs	For fractures, dislocations, elbow effusions, and focal bone lesions, respectively: AI sensitivity higher by 24.4%, 26.6%, 6.8%, and 82%, specificity lower by 12%, 0.9%, 0.2%, and 4.4%.	Level 2	37
Gleamer	BoneView v2.0.2[a]	Radiography	Retrospective MRMC study	480 examinations, 24 readers	Per-patient sensitivity increased from 64.9% to 75.2%, specificity increased from 90.6% to 95.6%, decreased mean reading time by 6.3 seconds.	Level 4	38
Gleamer	BoneView v2.0.2[a]	Radiography	Retrospective MRMC study	600 patients, 6 radiologists and 6 emergency physicians	AI assistance improved sensitivity of physicians by 8.7%, specificity by 4.1%, reduced mean number of false positives in fracture diagnosis per-patient by 41.9% and reduced mean reading time by 15.0%. The stand-alone AI performance was better than all unaided readers with an AUROC of 0.94.	Level 4	39
Gleamer	BoneView v2.0.2[a]	Radiography	Retrospective study	300 radiographs, 3 senior paediatric radiologists and 5 resident radiologists	Using AI assistance, sensitivity for junior radiologists increased by 10.3%, senior radiologists by 8.2%. Junior radiologist specificity increased by 1.4% and senior radiologist specificity decrease by 0.2%. AI stand-alone sensitivity and specificity were 91% and 90%, respectively.	Level 3	40
Gleamer	BoneView v2.0.2[a]	Radiography	Retrospective study	1163 examinations, 2 resident radiologists	Radiologist unaided sensitivity was 84.74%, and AI algorithm stand-alone sensitivity was 86.92%. AI assistance increased sensitivity by 6.54% and specificity by 0.26%.	Level 3	41
Gleamer	BoneView v2.0.2[a]	Radiography	Retrospective study	1917 radiographs, 41 radiologists	Stand-alone AI sensitivity was 7% higher, specificity was the same. AI assistance increased sensitivity by 12% and decreased specificity by 4%.	Level 3	42
Gleamer	BoneView v2.0.2[a]	Radiography	Retrospective study	300 radiographs, 2 radiologists	Per-patient sensitivity across all fractures was 91.3% and specificity was 90.0%. The AUROC for all fractures was 0.93.	Level 2	43
AZmed	Rayvolve v2.5.0	Radiography	External validation study	5865 radiographs	95.7% sensitivity; 91.2% specificity; 92.6% accuracy.	Level 2	44
Milvue	SmartUrgences	Radiography	External validation study	300 radiographs, 26 radiologists	Accuracy was 79.5%, sensitivity as 83.6%, specificity was 75.2%.	Level 2	45
Annalise.AI	Annalise Enterprise CXR v1.2	Radiography	External MRMC study	2972 cases, 11 readers	Using AI assistance 92 cases (3.1%) had significant report changes, 43 cases (1.4%) had changed patient management and 29 cases (1.0%) resulted in further imaging recommendations.	Level 4	46

(continued)

Table 5. (continued)

Company	Product	Modality	Type of evidence	Readers/ data set	Summary of evidence	Level of evidence	Ref.
Annalise.AI	Annalise Enterprise CXR v1.2	Radiography	External validation study	1404 cases, 2 radiology residents	Radiologists performed better than AI for clavicle fracture ($P = .002$), humerus fracture ($P < .0015$) and scapula fracture ($P = .014$), no statistical difference for rib fractures.	Level 2	47
Qure.ai	qXR	Radiography	Prospective multicentre study	65 604 radiographs	AI rib fracture AUROC was 0.98 and NPV was 99.9%. Turnaround time decreased by 40.63% using AI.	Level 4	48
Qure.ai	qXR	Radiography	Retrospective multicentre study	279 cases	Rib fracture sensitivity was 87%, specificity was 100%, and accuracy was 94% in these cases that were previously initially missed or mislabelled in radiology reports.	Level 2	49
Qure.ai	qXR	Radiography	Retrospective study	127 cases, 5 radiologists	AI sensitivity was 26.4% higher. AI specificity was 15.9% lower.	Level 2	50
Aidoc	C-Spine (CSF)	CT	External validation study	665 examinations, (radiologists of different levels of expertise and training)	CNN accuracy lower than radiologist (92% [95% CI, 90-94] vs 95% [95% CI, 94-97]). CNN sensitivity lower than radiologist (76% [95% CI, 68-83] vs 93% [95% CI, 88-97]). CNN specificity higher than radiologist (97% [95% CI, 95-98] vs 96% [95% CI, 94-98]).	Level 2	51
Aidoc	C-Spine (CSF)	CT	External validation study	1904 cases, 1 attending neuroradiologist	AI and radiologist interpretation concordant in 91.5% of cases. AI correctly identified 54.9% of cases with 106 false positives. AI sensitivity was 54.9% [95% CI, 45.7-63.9], specificity was 94.1% [95% CI, 92.9-95.1], PPV was 38.7% [95% CI, 33.1-44.7], and NPV was 96.8% [95% CI, 96.2-97.4].	Level 2	52
Shanghai United Imaging Intelligence	uAI EasyTriage-Rib	CT	External validation study	393 cases	Per-patient level, AI set to detect all rib trauma—sensitivity was 90.91%, specificity was 76.21%, PPV was 77.63%, and NPV was 90.23%; AI set to detect displaced rib fractures—sensitivity was 95.56%, specificity was 74.59%, PPV was 52.76%, and NPV value was 98.26%.	Level 2	53

Abbreviation: MRMC = multireader, multi case study.
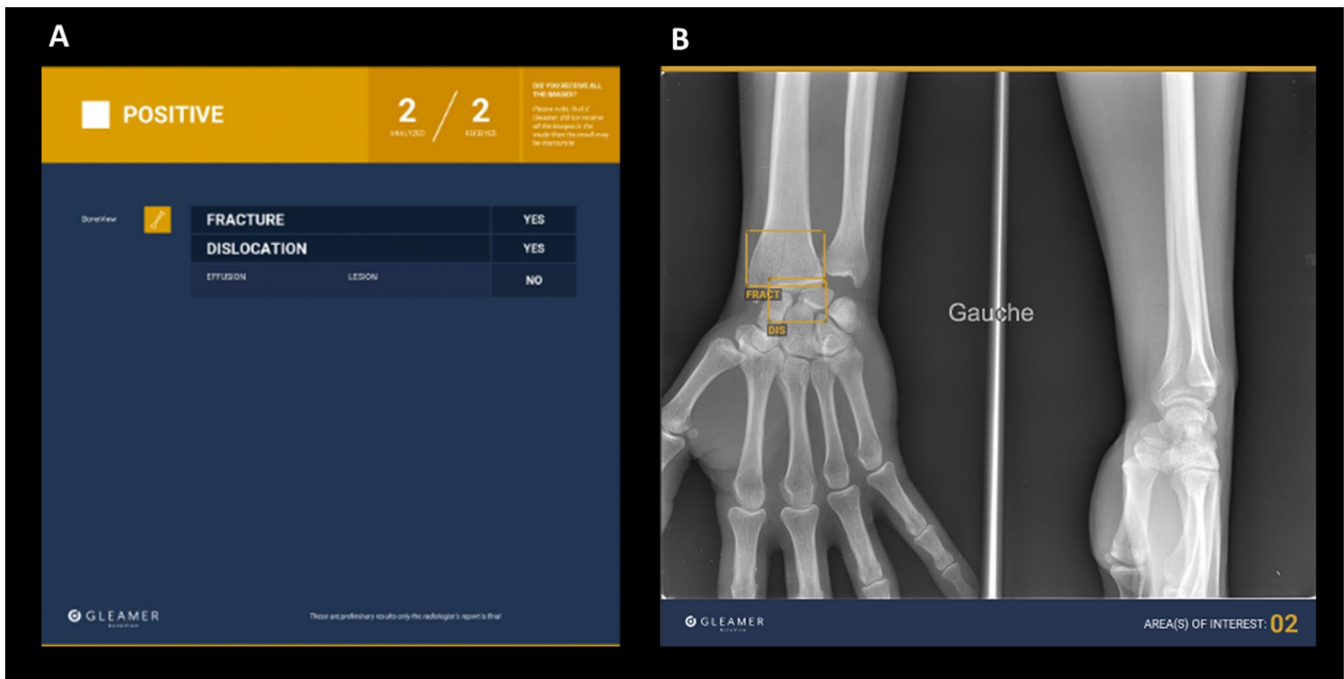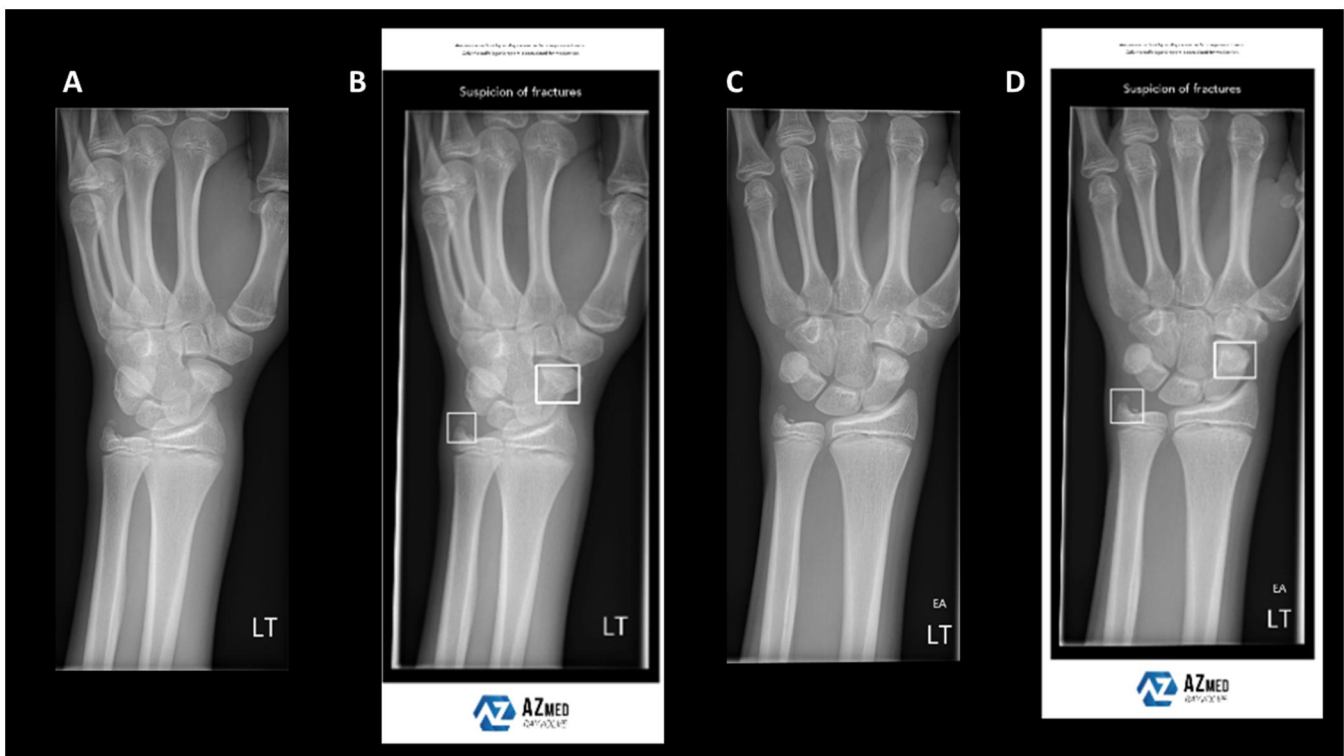aPrevious version of the software from the company.

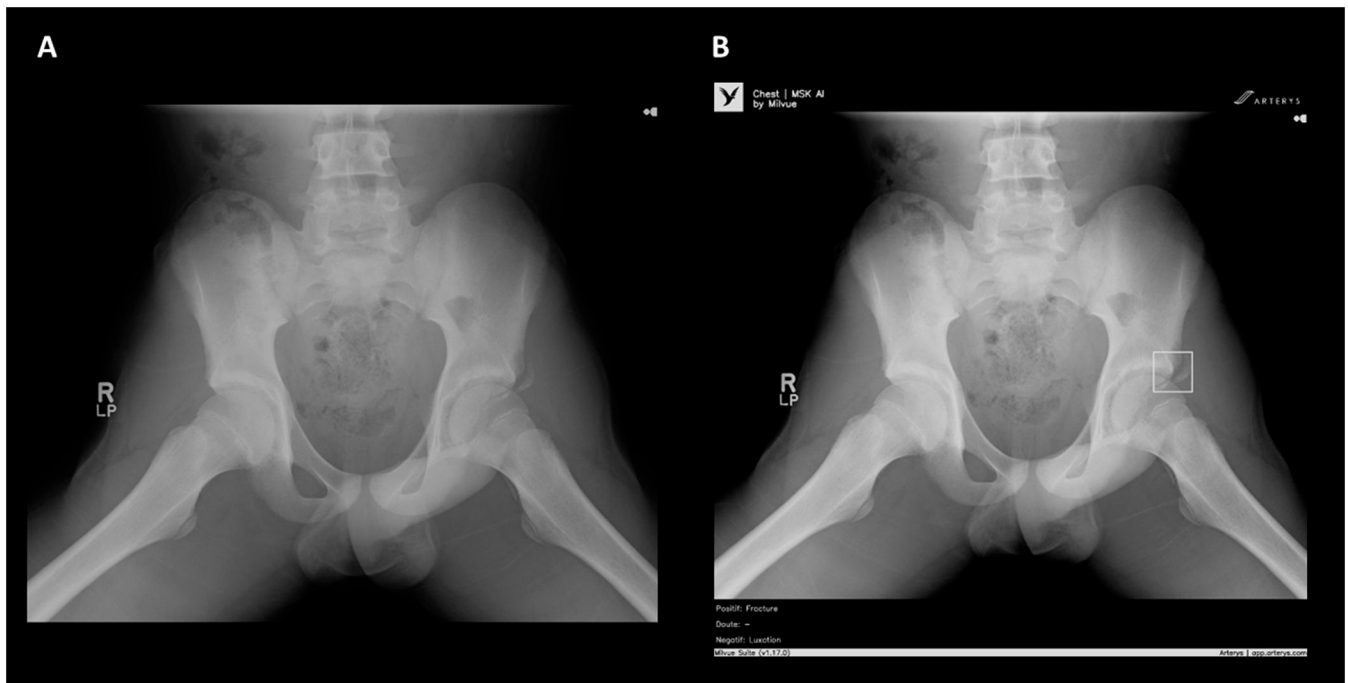**Figure 1.** This image demonstrates the results produced when using an artificial intelligence (AI) fracture detection tool by Gleamer, called BoneView. With this product, a summary image is sent to PACS, depicted in figure (A) which shows the number and type of pathologies detected on the radiograph. (B) A second image is also sent to PACS with bounding boxes and their associated labels (FRACT = fracture, DIS = dislocation) displayed as an "overlay" across the original radiographic image in question. In this example, the AI has flagged a fracture of the distal radius, with a scapholunate dislocation in a child. Image provided by Daniel Jones, Gleamer.

Figure 2. This image demonstrates the results produced when using an artificial intelligence (AI) fracture detection tool by AZMed, called Rayvolve. In this example, an oblique left wrist view (A) and DP wrist view (C) have been submitted for AI interpretation. The AI has flagged a fracture of the scaphoid and ulnar styloid (B, D) in a child by displaying bounding boxes as an "overlay" across the respective radiographic images. These are also sent to PACS for radiology reporter and clinician review. Image provided by Liza Alem, AZmed.

Figure 3. This image demonstrates an example of results produced when using an artificial intelligence (AI) fracture detection tool by Milvue, called Smarturgences. In this example, frog leg view of the pelvis in a child has been submitted for AI interpretation (A). The AI has correctly identified a fracture of the left anterior inferior iliac spine and placed a bounding box around the abnormality, as well as stating the pathology below the image (B).

In Europe and Northern Ireland, "CE certification" is required, however, recent changes to this regulation were introduced for medical devices. Prior to May 26, 2021, medical devices were CE certified under the "Medical Devices Directive" (MDD), however since then new standards known as the "Medical Devices Regulation" (MDR) have come into play. The MDR introduces more stringent requirements for clinical evidence, safety, post-market surveillance, and responsibilities of Notified Bodies (ie, the organizations designated to assess conformity with the regulations).[59] CE certified medical devices under MDD will therefore be required to re-certify under MDR before December 31, 2028 (for medium and lower risk medical devices)[60]; it is therefore important to know what CE certification a product has prior to purchase. In this review, 7/16 products have the more recent MDR CE certification, 6/16 have MDD CE certification, and 3 do not have current CE certification.

In Great Britain (ie, England, Wales, Scotland), the UKCA (UK Conformity Assessed) marking is a new regulatory marking that applied to medical devices following the end of the Brexit transition period (December 31, 2021),[61,62] although devices with CE marking will continue to be recognized until June 30, 2023 after which UKCA marking will be required for product use. At present, there is no central list of UKCA marked products, however, it is also possible to contact the relevant regulatory authority for further information (eg, the MHRA for UKCA). Although we contacted both bodies, we did not receive a timely response regarding which AI assisted fracture detection tools had this marking.

In the United States, FDA approval is required for medical devices and generally follows what is known as the "premarket notification (510(k)) process" for low to moderate risk devices (ie, Class 1 or 2 devices, which apply to the devices covered in this review). This pathway allows a vendor to demonstrate that their device is "substantially equivalent" to a legally marketed device (known as a "predicate device") already on the market and the vendor must include enough information (eg, intended use, comparison to predicate device, safety data) to prove that it can be marketed without requiring the more extensive "premarket approval" work-up (PMA).[63] In this review, 11/16 products reported FDA certification.

Discussion

Our market review highlighted a range of commercially available AI products for fracture detection across a variety of body parts and imaging modalities, with most for radiographic assessment and intended for an adult population. Relatively few products have published independent peer-reviewed evidence for their efficacy and diagnostic accuracy in children, although where tested, AI performance was found to be reduced for younger children.

For adults, there was a larger amount of peer-reviewed evidence across different body parts, with more studies evaluating the benefit of radiologists' imaging interpretation with and without AI and changes in speed of reporting for some products. It is hard to assess the best performing product purely through sensitivity and specificity due to varying levels and quality of evidence available; however, it is evident that Gleamer's BoneView is the most extensively externally validated product, whilst also reporting impressive sensitivities and specificities. Studies conducted on this product have also demonstrated reduced reading times of radiographs. This highlights possible future benefit for patients, especially if this leads to a faster referral for specialist care and treatment, although evidence demonstrating downstream improved patient outcomes and cost savings for a hospital department are yet to be evaluated (ie, Level 5 and 6 evidence).

It is important however to understand the type of conformity certification an AI product has prior to purchase, and we have tried to be as comprehensive yet concise as possible in our review of the market status. There have been some notable changes in the CE certification regulations and also those for sale in the UK market. Many products which have previously been awarded CE certification under MDD will require re-certification under MDR soon, and those wishing to use a product in the United Kingdom will need to check that their vendor has/will receive the UKCA certification in the near future.

Through conducting this investigation into commercially available tools that leverage AI to perform detection and diagnosis of fractures, we have been able derive some key conclusions about this market.

First is the divide in modality, between plain X-ray radiographs and CT scans. The latter constitutes a significantly smaller portion of the products available for fracture detection and all found in this analysis only target section(s) of the axial skeleton for fractures. All CT-based products also specifically target one type of fracture, either relating to the spine or ribs, whereas products based on radiographs often tend to include many different body parts and different pathologies. Whilst the results we provide in this review are for overall summary diagnostic accuracy rates, it is important for readers to review the listed references and FDA documentation, where available, if they wish to garner more detailed accuracy rates for specific fracture locations and types.

Second, the market for children is still significantly behind adults in terms of range of available products, with 4 of the 16 products (25%) stating their applicability to children. There are understandably significant technical challenges in adapting AI solutions to be effective in paediatrics due to the variability in bone structure, predominantly between the ages of 0-16 years.[5] Development in this more specialized field is also slowed and restricted by legal issues relating to the collection of data for training the algorithms and more complex and difficult procedures for obtaining certification or approval from respective legal bodies.[64]

Third, the amount of independent external validation is also significantly lacking for many products, with many only having vendor conducted validation for purposes of achieving FDA approval or CE certification. Future work in this domain that independently verifies the performance of specific commercially available products would provide a much clearer basis to evaluate which product is best for a clinical/health institution. Of all the evidence discovered throughout this review, the highest level of evidence was Level 4 (according to the levels of evidence suggested by Fryback and Thornberry[13]). This means study into the deeper impact of these tools is severely lacking, given that evidence Levels 5 and 6 assess the effect on patient outcomes through changes to quality of life and societal impact based on an economic analysis. As interest in this field continues to grow, such assessments will be fundamental in determining the greater value these products are able to provide.

We acknowledge that our review has limitations due to the ever-increasing number of commercial AI products coming to market and newer versions of existing tools being developed. It is likely that by the time of publication we may not have included some very recent tools, recent conformity accreditation and evidence to support usage in particular situations, which were unavailable at time of our search. We did contact as many AI companies as possible, including those that advertised only prototype versions of their software, to ensure we captured emerging products as well as those already established. We also offered the AI companies an opportunity to let us know of any updates in development. Some AI companies did not engage or respond to our request for information within the timeframe provided, including the MHRA regarding details on products with UKCA certification.

Conclusion

Overall, there is a scarcity of rigorous, independent evaluation of commercially available AI tools for fracture detection in adults and children, and some products will need to update their current conformity registration. The information in this article may help departmental and hospital leaders, as well as local AI champions, in understanding whether tools available are worth further investigation for their specific institution at this stage in their development.

Supplementary material

Supplementary material is available at *BJR|Open* online.

Funding

Conflicts of interest

None declared.

References

1. NHS Resolution. *Clinical Negligence Claims in Emergency Departments in England: Missed Fractures*. NHS Resolution; 2022. Accessed February 7, 2023. https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiQpMCIzYP9AhVOOMAKHfOSAgoQFnoECAsQAw&url=https%3A%2F%2Fresolution.nhs.uk%2Fwp-content%2Fuploads%2F2022%2F03%2F2-NHS-Resolution-ED-report-Missed-Fractures.pdf&usg=AOvVaw0vjwcy8RQezvCBltxqBRKP

2. Kuo RYL, Harrison C, Curran TA, et al. Artificial intelligence in fracture detection: a systematic review and meta-analysis. *Radiology*. 2022;304(1):50-62. https://doi.org/10.1148/radiol.211785

3. Zhang X, Yang Y, Shen YW, et al. Diagnostic accuracy and potential covariates of artificial intelligence for diagnosing orthopedic fractures: a systematic literature review and meta-analysis. *Eur Radiol*. 2022;32(10):7196-7216. https://doi.org/10.1007/s00330-022-08956-4

4. Langerhuizen DWG, Janssen SJ, Mallee WH, et al. What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? A

systematic review. *Clin Orthop Relat Res*. 2019;477(11): 2482-2491. https://doi.org/10.1097/corr.0000000000000848

5. Shelmerdine SC, White RD, Liu H, Arthurs OJ, Sebire NJ. Artificial intelligence for radiological paediatric fracture assessment: a systematic review. *Insights Imaging*. 2022;13(1):94. https://doi.org/10.1186/s13244-022-01234-3

6. Oliveira ECL, van den Merkhof A, Olczak J, et al.; Machine Learning Consortium. An increasing number of convolutional neural networks for fracture recognition and classification in orthopaedics: are these externally validated and ready for clinical application? *Bone Jt Open*. 2021;2(10):879-885. https://doi.org/10.1302/2633-1462.210.Bjo-2021-0133

7. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. *Radiol Artif Intell*. 2022;4(3):e210064. https://doi.org/10.1148/ryai.210064

8. van Leeuwen KG, Schalekamp S, Rutten M, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol*. 2021; 31(6):3797-3804. https://doi.org/10.1007/s00330-021-07892-z

9. Omoumi P, Ducarouge A, Tournier A, et al. To buy or not to buy-evaluating commercial AI solutions in radiology (the ECLAIR guidelines). *Eur Radiol*. 2021;31(6):3786-3796. https://doi.org/10.1007/s00330-020-07684-x

10. Van Leeuwen KG. *AI for Radiology: An Implementation Guide. Diagnostic Image Analysis Group*. Department of Medical Imaging, Radboud University Medical Center. Accessed February 7, 2023. https://grand-challenge.org/aiforradiology/

11. (FDA) USFaDA. *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*. U.S. Food and Drug Administration; 2023. Accessed February 7, 2023. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices#resources

12. The Medical Futurist. *FDA Approved AI Based Algorithms*. The Medical Futurist. Accessed February 7, 2023. https://medicalfuturist.com/fda-approved-ai-based-algorithms/

13. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making*. 1991;11(2):88-94.

14. Creative Commons Licences Attribution 4.0 International (CCBY4.0). Accessed February 10, 2023. https://creativecommons.org/licenses/by/4.0/

15. FDA. *Gleamer BoneView 1.1-US FDA Documentation Ref: K222176*. FDA; 2023. Accessed September 14, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf22/K222176.pdf

16. FDA. *Aidoc Medical Ltd., BriefCase FDA Documentation Ref: K190896*. FDA; 2019. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K190896.pdf

17. FDA. *Gleamer BoneView v2.5.0 FDA Documentation Ref: K212365*. FDA; 2022. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf21/K212365.pdf

18. Radiobotics. *RBFracture Improves the Diagnostic Accuracy of Emergency Care Professionals*. Radiobotics; 2021. Accessed February 24, 2023. https://static1.squarespace.com/static/5ea1660027ddc935cb0cfbfd/t/6295cbb133a872233ad4d461/1653984178738/RB05_RBfractureStudy.pdf

19. FDA. *AZMed Rayvolve FDA Documentation Ref: K220164*. FDA; 2022. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf22/K220164.pdf

20. Rayscape. *Rayscape Medical Whitepaper*. Rayscape; 2022. Accessed February 24, 2023. https://file.xvision.app/sources/Rayscape_Whitepaper.pdf

21. AI for Radiology An Implementation Guide. *Milvue Suite*. AI for Radiology An Implementation Guide; 2023. Accessed February 24, 2023. https://grand-challenge.org/aiforradiology/product/milvue-suite/

22. Milvue—MSK-AI Retrospective Study, Whitepaper. 2020. Accessed February 24, 2023. https://www.bing.com/ck/a?!&&p=726f270a092181e3JmltdHM9MTY3NzE5NjgwMCZpZ3VpZD0wYTM4ZGNkOC1hNDM1LTY0MmItMjM4NC1jZTkxYTU1NjY1MjgmaW5zaWQ9NTIxNA&ptn=3&hsh=3&fclid=0a38dcd8-a435-642b-2384-ce91a5566528&psq=milvue+parsy+et+al&u=a1aHR0cHM6Ly9wdWJsaWMuYXJjaW55cy5jb20vSW1hZ2luZ1dpcmmUvV2hpdGVwYXBlcl9NU0tfcmV0cm9zcGVjdGl2ZV9zdHVkeV9FTi5wZGY&ntb=1

23. FDA. *Imagen Technologies, OsteoDetect FDA Documentation De Novo No: DEN180005*. FDA; 2018. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180005.pdf

24. FDA. *Imagen Technologies, Inc. FractureDetect (FX) FDA Documentation Ref: K193417*. FDA; 2020. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K193417.pdf

25. Anderson PG, Baum GL, Keathley N, et al. Deep learning assistance closes the accuracy gap in fracture detection across clinician types. *Clin Orthop Relat Res*. 2023;481(3):580-588.

26. Seah JCY, Tang CHM, Buchlak QD, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *Lancet Digit Health*. 2021;3(8):e496-e506. https://doi.org/10.1016/s2589-7500(21)00106-0

27. Kim KC, Cho HC, Jang TJ, Choi JM, Seo JK. Automatic detection and segmentation of lumbar vertebrae from X-ray images for compression fracture evaluation. *Comput Methods Programs Biomed*. 2021;200:105833. https://doi.org/10.1016/j.cmpb.2020.105833

28. FDA. *Qure.ai Technologies, qER FDA Documentation Ref: K200921*. FDA; 2020. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf20/K200921.pdf

29. Ghosh R, Chilamkurthy S, Biviji M, Rao P. *Automated Detection and Localisation of Skull Fractures from CT Scans Using Deep Learning*. European Congress of Radiology (ECR); 2018. Accessed February 24, 2023. https://qure.ai/evidence/automated-detection-and-localisation-of-skull-fractures-from-ct-scans-using-deep-learning/

30. Tanamala S, Chilamkurthy S, Maniparambil M, Rao P, Biviji M. *Clinical Context Improves the Performance of AI models for Cranial Fracture Detection*. Abstract Archives of the RSNA; 2019. Accessed February 24, 2023. https://archive.rsna.org/2019/19013570.html

31. Weikert T, Noordtzij LA, Bremerich J, et al. Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *Korean J Radiol*. 2020;21(7):891-899. https://doi.org/10.3348/kjr.2019.0653

32. FDA. *Aidoc Medical Ltd. BriefCase for RibFx Triage FDA Documentation Ref: K202992*. FDA; 2021. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf20/K202992.pdf

33. FDA. *Zebra Medical Vision Ltd. HealthVCF FDA Documentation Ref: K192901*. FDA; 2020. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K192901.pdf

34. [MIB267] NICE MedTech Innovation Briefings (NMIB). *HealthVCF for Detecting Vertebral Compression Fractures on CT Scans*. 2021. Accessed February 21, 2023. https://www.nice.org.uk/advice/mib267/chapter/Clinical-and-technical-evidence

35. FDA. *Shanghai United Imaging Intelligence Co. Ltd. uAI EasyTriage-Rib. FDA Documentation Ref: K193271*. FDA; 2021. Accessed February 24, 2023. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K193271.pdf

36. 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(12):594-600. https://doi.org/10.1016/j.diii.2022.06.004

37. Regnard N-E, Lanseur B, Ventre J, et al. Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays. *Eur J Radiol*. 2022;154:110447. https://doi.org/101016/j.ejrad.2022.110447

38. Guermazi A, Tannoury C, Kompel AJ, et al. Improving radiographic fracture recognition performance and efficiency using artificial intelligence. *Radiology*. 2022;302(3):627-636. https://doi.org/10.1148/radiol.210937

39. Duron L, Ducarouge A, Gillibert 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(1):120-129. https://doi.org/10.1148/radiol.2021203886

40. Nguyen T, Maarek R, Hermann AL, 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(11):2215-2226. https://doi.org/10.1007/s00247-022-05496-3

41. Oppenheimer J, Lüken S, Hamm B, Niehues SM. A prospective approach to integration of AI fracture detection software in radiographs into clinical workflow. *Life (Basel)*. 2023;13(1):223.

42. Cohen M, Puntonet J, Sanchez J, et al. Artificial intelligence vs. radiologist: accuracy of wrist fracture detection on radiographs. *Eur Radiol*. 2022;33(6):3974-3983. https://doi.org/10.1007/s00330-022-09349-3

43. Hayashi D, Kompel AJ, Ventre J, et al. Automated detection of acute appendicular skeletal fractures in pediatric patients using deep learning. *Skeletal Radiol*. 2022;51(11):2129-2139. https://doi.org/10.1007/s00256-022-04070-0

44. Dupuis M, Delbos L, Veil R, Adamsbaum C. External validation of a commercially available deep learning algorithm for fracture detection in children: fracture detection with a deep learning algorithm. *Diagn Interv Imaging*. 2021;103(3):151-159. https://doi.org/10.1016/j.diii.2021.10.007

45. Shelmerdine SC, Martin H, Shirodkar K, Shamsuddin S, Weir-McCall JR; FRCR-AI Study Collaborators. Can artificial intelligence pass the Fellowship of the Royal College of Radiologists examination? Multi-reader diagnostic accuracy study. *BMJ*. 2022;379:e072826. https://doi.org/10.1136/bmj-2022-072826

46. Jones CM, Danaher L, Milne MR, et al. Assessment of the effect of a comprehensive chest radiograph deep learning model on radiologist reports and patient outcomes: a real-world observational study. *BMJ Open*. 2021;11(12):e052902. https://doi.org/10.1136/bmjopen-2021-052902

47. Gipson J, Tang V, Seah J, et al. Diagnostic accuracy of a commercially available deep-learning algorithm in supine chest radiographs following trauma. *Br J Radiol*. 2022;95(1134):20210979. https://doi.org/10.1259/bjr.20210979

48. Govindarajan A, Govindarajan A, Tanamala S, et al. Role of an automated deep learning algorithm for reliable screening of abnormality in chest radiographs: a prospective multicenter quality improvement study. *Diagnostics (Basel)*. 2022;12(11):2724.

49. Kaviani P, Digumarthy SR, Bizzo BC, et al. Performance of a chest radiography AI algorithm for detection of missed or mislabeled findings: a multicenter study. *Diagnostics (Basel)*. 2022;12 (9):2086.

50. Kvak D, Chromcová A, Biroš M, et al. Chest X-ray abnormality detection by using artificial intelligence: a single-site retrospective study of deep learning model performance. *BioMedInformatics*. 2023;3(1):82-101.

51. Small JE, Osler P, Paul AB, Kunst M. CT cervical spine fracture detection using a convolutional neural network. *AJNR Am J Neuroradiol*. 2021;42(7):1341-1347. https://doi.org/10.3174/ajnr.A7094

52. Voter AF, Larson ME, Garrett JW, Yu JPJ. Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of cervical spine fractures. *AJNR Am J Neuroradiol*. 2021; 42(8):1550-1556. https://doi.org/10.3174/ajnr.A7179

53. Liu T, Xie H, Xu YH, et al. A preliminary study on the application of artificial intelligence in CT diagnosis of rib fractures in thoracic trauma. 2021;41(7):920–25. https://doi.org/10.3969/j.issn.1674-8115.2021.07.012

54. See AI. Accessed February 21, 2023. https://www.seeai.co.uk/case-studies

55. Imera.ai. Accessed February 21, 2023. https://www.imera.ai/

56. Yang C, Wang J, Xu J, et al. Development and assessment of deep learning system for the location and classification of rib fractures via computed tomography. *Eur J Radiol*. 2022;154:110434. https://doi.org/10.1016/j.ejrad.2022.110434

57. Gunasingam C, Jaunalksnis A, Beaufement S, Chan V, Major G. *Opportunistic Identification of Vertebral Compression Fractures Using Artificial Intelligence Technology*. Internal Medicine Journal Supplement 2: Royal Australasian College of Physicians; 2020. Accessed February 21, 2023. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/imj.14932

58. Dagan N, Elnekave E, Barda N, et al. Automated opportunistic osteoporotic fracture risk assessment using computed tomography scans to aid in FRAX underutilization. *Nat Med*. 2020;26(1): 77-82. https://doi.org/10.1038/s41591-019-0720-z

59. *Factsheet for Manufacturers of Medical Devices*. European Commission; 2020. Accessed February 21, 2023. https://health.ec.europa.eu/system/files/2020-09/md_manufacturers_factsheet_en_0.pdf

60. European Commission. *Public Health: More Time to Certify Medical Devices to Mitigate Risks of Shortages*. European Commission; 2023. Accessed February 21, 2023. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_23

61. Gov.uk. *Medical Devices: Conformity Assessment and the UKCA Mark*. 2020. Accessed February 21, 2023. https://www.gov.uk/guidance/medical-devices-conformity-assessment-and-the-ukca-mark

62. Gov.uk. *Regulating Medical Devices in the UK*. 2022. Accessed February 21, 2023. https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk#full-publication-update-history

63. U.S. Food and Drug Administration. *Device Approvals, Denials and Clearances*. U.S. Food and Drug Administration; 2018. Accessed February 21, 2023. https://www.fda.gov/medical-devices/products-and-medical-procedures/device-approvals-denials-and-clearances#:~:text=A%20PMA%20is%20an%20application,its%20intended%20use%20or%20uses

64. Davendralingam N, Sebire NJ, Arthurs OJ, Shelmerdine SC. Artificial intelligence in paediatric radiology: future opportunities. *Br J Radiol*. 2020;94(1117):20200975. https://doi.org/10.1259/bjr.20200975