Current and emerging artificial intelligence applications for pediatric musculoskeletal radiology

Amaka C. Offiah [1,2]

Abstract

Artificial intelligence (AI) is playing an ever-increasing role in radiology (more so in the adult world than in pediatrics), to the extent that there are unfounded fears it will completely take over the role of the radiologist. In relation to musculoskeletal applications of AI in pediatric radiology, we are far from the time when AI will replace radiologists; even for the commonest application (bone age assessment), AI is more often employed in an AI-assist mode rather than an AI-replace or AI-extend mode. AI for bone age assessment has been in clinical use for more than a decade and is the area in which most research has been conducted. Most other potential indications in children (such as appendicular and vertebral fracture detection) remain largely in the research domain. This article reviews the areas in which AI is most prominent in relation to the pediatric musculoskeletal system, briefly summarizing the current literature and highlighting areas for future research. Pediatric radiologists are encouraged to participate as members of the research teams conducting pediatric radiology artificial intelligence research.

Keywords Artificial intelligence · Bone · Children · Musculoskeletal · Pediatric radiology

Introduction

Simply put, artificial intelligence (AI) can be defined as software that automates (or semi-automates) a cognitive task. AI applications in the musculoskeletal system can be fully automated (e.g., bone age assessment) or semi-automated (e.g., vertebral fracture assessment). Although AI tools in general can be categorized as AI-assist (helping the radiologist), AI-replace (replacing the radiologist) or AI-extend (exceeding the capability of the radiologist) [1], as far as the author is aware, no clinical tool functions in the AI-extend mode in pediatric musculoskeletal radiology practice.

Existing AI tools can help to improve image quality; aid in the measurement of lengths, angles and volumes; or aid in the detection of pathological processes, the last through recognition and classification of morphological or textural abnormalities.

Of the 144 AI products that are CE (Conformité Européenne) marked and commercially available, 74 also have United States Food and Drug Administration (FDA) approval and 18 are related to the musculoskeletal system [2]. Of these 18, one is for image enhancement and post-processing rather than being a diagnostic aid, per se. Considering the remaining 17 AI musculoskeletal products, the majority (14) have been designed for aiding diagnosis from radiographs, while use in pediatric radiology is only explicitly stated in the information available for 3 of the 17 tools. Table 1 summarizes these 17 available musculoskeletal AI tools; all 3 tools intended for use in pediatric radiology are for bone age assessment [2].

Although commercially available tools are currently only for bone age assessment, ongoing and published research pertains to tasks such as fracture diagnosis (appendicular and vertebral), scoliosis and leg-length discrepancy measurements. Other areas where pediatric research is being performed include determining bone health using the bone health index and diagnosing metopic craniosynostosis and developmental dysplasia of the hip. These emerging applications could achieve commercial release within the next decade.

✉ Amaka C. Offiah
a.offiah@sheffield.ac.uk

1 Department of Oncology and Metabolism, University of Sheffield, Damer Street Building, Sheffield S10 2TH, UK

2 Department of Radiology, Sheffield Children's NHS Foundation Trust, Sheffield, UK

Table 1 Currently available CE-marked musculoskeletal radiology artificial intelligence (AI) tools[a]

Date on market	Company	Product name	Disease targeted	Modality	Pediatrics[b]	CE class	FDA
Mar 2009	Visiana (Hørsholm, Denmark)	BoneXpert	Bone age	Radiography	Y	I	N
2012	Medimaps (Geneva, Switzerland)	TBS iNsight (Osteo)	Osteoporosis	Radiography	N	IIa	II
Nov 2017	Aidoc (Tel Aviv, Israel)	C-Spine	C-spine fracture	CT	N	I	II
May 2018	VUNO (Seoul, Korea)	VUNO Med-BoneAge	Bone age	Radiography	Y	IIa	N
Jan 2019	QUIBIM (Valencia, Spain)	2D Bone Microarchitecture QTS Score	Osteoporosis Oncology Osteopenia	Radiography	N	IIa	N
Jan 2019	QUIBIM	2D Bone Microarchitecture QTS Score	Osteoporosis Oncology Osteopenia	CT, MRI	N	IIa	N
Jan 2019	QUIBIM	Cartilage T2 Mapping	Osteoarthritis Degeneration Sports diseases	MRI	N	IIa	N
Jan 2019	QUIBIM	Texture Analysis	Tumors Osteoporosis Osteopenia Osteoarthritis	Radiography, CT, MRI, PET, US, SPECT	N	IIa	N
Jun 2019	AZMed (Paris, France)	Rayvolve	Fracture	Radiography	N	I	N
Aug 2019	ImageBiopsy Lab (Vienna, Austria)	IB Lab KOALA	Osteoarthritis (knee)	Radiography	N	I	II
Nov 2019	Radiobiotics (Copenhagen, Denmark)	RBknee	Osteoarthritis (knee)	Radiography	N	I	N
Feb 2020	Arterys (San Francisco, CA)	Chest/MSK AI	Fracture Dislocation	Radiography	N	IIa	N
Mar 2020	Gleamer (Paris, France)	BoneView	Fracture	Radiography	N	IIa	N
Oct 2020	ImageBiopsy Lab	IB Lab HIPPO	Hip measurements	Radiography	N	I	N
Oct 2020	ImageBiopsy Lab	IB Lab LAMA	Leg geometry	Radiography	N	I	N
Nov 2020	ImageBiopsy Lab	IB Lab PANDA	Bone age	Radiography	Y	I	N
U	Zebra (Kibbutz Shefayim, Israel)	Bone Health	Vertebral compression fractures	CT	N	U	II

CE Conformité Européene, *FDA* United States Food and Drug Administration approval, *MSK* musculoskeletal, *N* no, *PET* positron emission tomography, *SPECT* single-photon emission computed tomography, *U* unknown, *Y* yes

[a] Derived from [2]

[b] Only includes those tools for which pediatric use is explicitly stated

This review also identifies and briefly discusses areas in which very little AI research has been conducted but in which there is potential for AI to play a significant role; these areas include inflicted injury (child abuse) and skeletal dysplasias.

The main focus of this article is on diagnosis/detection of pathology. For AI applications related to image-quality improvement, image post-processing, quality control, etc., the reader is directed to other articles in this special issue and to the 2019 review by Koska [3].

Current applications: bone age assessment

Although three bone age assessment AI tools are on the market, the oldest and probably best known is BoneXpert (Visiana, Hørsholm, Denmark). Indeed, BoneXpert is the oldest musculoskeletal AI-replace software tool on the market (Table 1), and more than 150 departments are using BoneXpert in day-to-day clinical practice across Europe, each performing more than 100 analyses per year (personal communication with H.H. Thodberg and P. Bak, November 2020).

BoneXpert automatically calculates bone age according to the Greulich and Pyle and the Tanner and Whitehouse standards in a process that takes less than 15 s per hand and wrist radiograph. The method is based on traditional machine-learning methodology and involves prediction of bone age based on shape, intensity and texture scores as derived from principal component analysis. It is worth noting that there are no General Data Protection Regulation (GDPR)-related issues, because BoneXpert is configured as a Digital Imaging and Communications in Medicine (DICOM) node for local

picture archiving and communication systems (PACS) and is an image-analysis application only. In other words, BoneXpert does not store data, share data or transfer data outside the local PACS. The pathway and output are illustrated in Figs. 1 and 2, respectively, and the terminology used in the output is explained in Table 2.

Although launched in 2009 as an AI-replace tool, approximately 70% of departments that have the software installed use BoneXpert as an AI-assist tool (A. Offiah, unpublished work). The reason for this is simple: while BoneXpert rejects radiographs with significant abnormality (e.g., poor positioning, abnormal bone morphology, poor image quality), it does not reject radiographs with subtle abnormality of morphology (e.g., early rickets) or abnormality of texture (e.g., metaphyseal striations). If a radiologist does not review the radiographs, then these subtle changes will be missed. The detection of such abnormalities is outside the scope of the software as developed, and radiologists are advised to bear this in mind.

The percentage of radiographs rejected by BoneXpert because of abnormal anatomy depends on the types of patients seen, ranging from approximately 0.4% in general hospitals to up to 3% in hospitals specializing in skeletal dysplasias. The percentage of radiographs rejected by BoneXpert because of poor image quality is generally very low (reflecting radiographer competence). However, departments in which significant edge-enhancement is applied as part of the post-processing of images might see rejection rates of up to 10%, accompanied by the error message, "Too sharp" (personal communication with H.H. Thodberg and P. Bak, April 2021). The software has been tested in multiple populations and ethnicities, including Caucasian, African American, Hispanic, Asian Chinese and Saudi Arabian populations [4–9]; while generally applicable to all ethnicities, some caution is advised, but this is related to the applicability of the standards (Greulich and
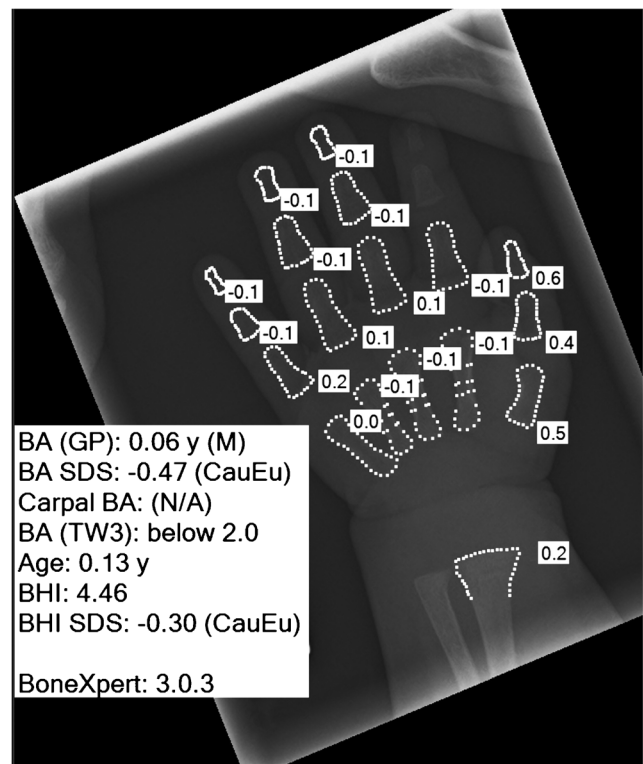
Fig. 2 Posteroanterior left hand and wrist radiograph in a 3-week-old boy, following interpretation of bone age by BoneXpert. The figures in the small white boxes represent the Greulich and Pyle bone ages of the individual bones. *BA (GP)* Greulich and Pyle bone age (gender), *BA SDS* Bone age standard deviation score (ethnicity), *BA (TW3)* Tanner and Whitehouse three-bone age, *BHI* bone health index (digital X-ray radiogram), *BHI SDS* bone health index standard deviation score (ethnicity), *CauEu* Caucasian European, *M* male, *N/A* not available (no ossified carpal bones in this child), *y* years. See also Table 2

Pyle, Tanner and Whitehouse 3) and not to the BoneXpert software itself. The latter claim can be made based on diagnostic accuracy studies that have compared manual to
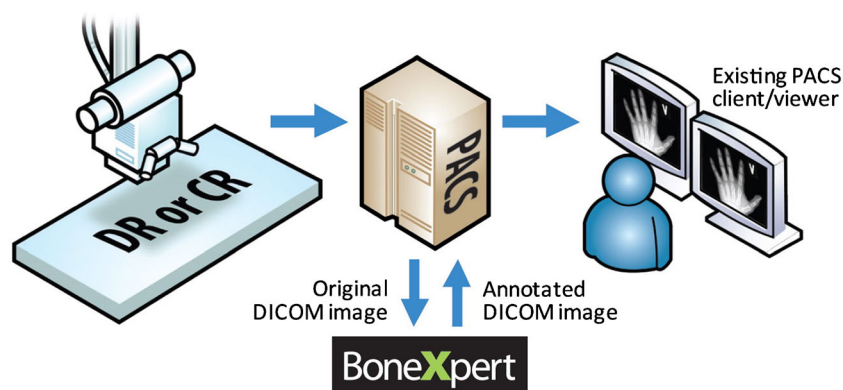
Fig. 1 BoneXpert workflow. The radiographer who has performed the left hand and wrist radiograph sends it to the picture archiving and communication system (PACS) in the usual way. The reporting radiologist (or radiographer) transfers the image to BoneXpert, where the automated bone age estimation is performed. BoneXpert then returns the annotated image to the PACS. The reporting radiologist now not only has access to the BoneXpert-derived Greulich and Pyle and the Tanner and Whitehouse 3 bone age assessments, but he or she can also review the original radiograph for disease-related abnormality (e.g., evidence of a skeletal dysplasia). Image courtesy of Peter Bak and Hans Henrik Thodberg (Visiana). *CR* computed radiography, *DICOM* digital imaging and communications in medicine, *DR* digital radiography

Table 2 BoneXpert outputs

Term	Interpretation
BA (GP)	Greulich and Pyle bone age (gender)
BA SDS	Bone age standard deviation score (ethnicity)
BA (TW3)	Tanner and Whitehouse three-bone age
Age	Chronological age
BHI	Bone health index (digital X-ray radiogram)
BHI SDS	Bone health index standard deviation score (ethnicity)

automated bone age assessments [10–17], with reported root mean square errors being as low as 0.63 [18].

In 2017, the Radiological Society of North America launched a Machine Learning Challenge, making freely available a set of more than 14,000 hand and wrist radiographs [19]. The best-performing entry achieved a concordance correlation coefficient of 0.99 and differed from ground truth by only 4.3 months, compared to 7.3 months for radiologists [20]. It is possible for readers to test the application for themselves but note that it is for demonstration purposes only (i.e., it is not for clinical use) [21].

Other bone age tools have also been tested and found to be reliable and accurate and to reduce reporting times [22–26] and it is worth pointing out the encouraging results obtained for bone age estimation of the index finger alone (as opposed to the entire left hand and wrist radiograph) when using a neural-network-based AI application, which paves the way for hand-held bone age estimation machines [26].

While some authors have focused their work on AI determination of bone age from other sites, such as the pelvis [27] or knee [28], and other modalities, such as MRI [29–32], a relatively recent systematic review highlighted the lack of such studies, in addition to the need for more research assessing potential socioeconomic and ethnic variations on the performance of such AI tools [33].

Emerging/future applications

Bone health index

There is no reliable method of predicting fracture risk in children. While dual-energy X-ray absorptiometry is the gold standard for bone mineral density assessment in children, it has limitations [34–36]. As such, other quantitative bone imaging techniques have been developed including AI applications (predominantly related to adults; Table 1). Of relevance to pediatrics is radiogrammetry. Originally performed manually [37], this technique lends itself to automation because it measures cortical thickness of the phalanges in relation to their lengths, thereby producing an index of bone strength.

In addition to determining bone age, the BoneXpert software discussed in the previous section also performs "digital X-ray radiogrammetry," providing an indication of bone health called the "bone health index" or "BHI" (Fig. 2 and Table 2). Bone health index is derived from a measurement of the cortical thickness, width and length of the three middle metacarpals. A standard deviation score is also provided, allowing comparison with the bone health index of healthy Caucasian children of the same age and gender.

While a few studies have been performed with favorable results [38–41], the clinical role of the bone health index in monitoring and assessing bone strength in children (of any ethnicity) has not been elucidated. In a recent systematic review, peripheral quantitative computed tomography (pQCT), bone health index and quantitative ultrasound (QUS) were compared with dual-energy X-ray absorptiometry. Meta-analysis showed BHI to have the strongest correlation with dual-energy X-ray absorptiometry, with a pooled estimate of correlation of 0.71 compared to 0.57 for both pQCT and QUS [42]. These results encourage further research into the potential clinical application of BHI.

Fracture assessment

Appendicular fractures

The few studies that have assessed the utility of AI for appendicular fracture detection in children have predominantly concentrated on the elbow joint, possibly because of the complexity of the elbow joint and multiple unossified epiphyseal centers that are found in children. England et al. [43] used a relatively small set of lateral radiographs to train (657 images), validate (115 images) and test (129 images) a convolutional neural network for the identification of elbow joint effusions. Compared to the reference standard of radiologists' reports, the network had sensitivity, specificity and accuracy of 0.91.

In a significantly larger study consisting of 21,456 anteroposterior and lateral elbow radiographs, Rayan et al. [44] determined the feasibility of deep learning to correctly classify elbow radiographs as normal or abnormal. The true positive rate (i.e., those radiographs correctly classified as abnormal) was highest for supracondylar fractures (0.996) and lowest for osteochondral lesions (0.000), although it should be noted that there were only two cases of osteochondral lesions in the entire dataset. Most recently, Choi et al. [45] assessed the ability of a convolutional neural network to correctly identify supracondylar fractures from 1,266 anteroposterior and lateral elbow radiographs.

The results of these three studies (summarized in Table 3; [43–45]) are encouraging. However, particularly given their

Table 3 Diagnostic accuracy of artificial intelligence (AI) applications for elbow trauma in children

Author [reference]	Sensitivity	Specificity	Accuracy	Positive predictive value	Negative predictive value
England et al. 2018 [43]	0.91	0.91	0.91	–	–
Rayan et al. 2019 [44]	0.91	0.84	0.88	0.87	0.89
Choi et al. 2020 [45] (temporal test set)	0.93	0.92	–	0.80	0.98
Choi et al. 2020 [45] (geographic test set)	0.10	0.86	–	0.70	0.10

relatively low positive predictive values, the current role of such AI tools would appear to be in the initial triage of elbow radiographs following trauma in children, in areas where a (pediatric) radiologist is not immediately available. Future research could also be directed toward determining the feasibility of AI tools for fracture detection in children at other appendicular sites, in a similar way to ongoing investigations in, e.g., the wrist [46] and proximal femur [47] of adults.

Axial fractures

A decision tree might be seen as a flowchart-like structure, with each branch representing a potential outcome. Optimal trees are predictive AI algorithms that limit the number of outcomes while encompassing as much of the available data as possible [48]. Bertsimas et al. [49] used an optimal trees artificial intelligence approach to predict cervical spine trauma in children. However, this model was based on history and clinical parameters (including Glasgow Coma Scale) and used imaging interpretation by radiologists as an outcome measure for presence or absence of fracture, rather than using the algorithm to classify radiographs. Given the difficulty associated with obtaining adequate views (particularly in younger children) and complexity of the cervical spine [50], this would be a worthy field for the development of an AI diagnostic tool.

Other studies assessing AI tools for detecting vertebral fractures in children relate to osteoporotic compression fractures rather than post-traumatic fractures and are briefly reviewed next.

The diagnosis of vertebral crush fractures from dual-energy X-ray absorptiometry scans is termed vertebral fracture assessment, and the Lunar iDXA machine (GE Healthcare Lunar, Buckinghamshire, UK) has been shown to be as reliable as radiographs for vertebral crush fracture diagnosis in children at a lower radiation dose penalty [51, 52]. The use of software tools to diagnose vertebral fractures from dual-energy X-ray morphometry scans is termed morphometric vertebral analysis [52]. Such software tools are widely available for clinical use in adults; however, they have not been licensed for use in children. Given both the wide variability in diagnosis of vertebral fractures in children [53] and that the recognition of vertebral

shape (morphometry) lends itself to AI applications, researchers have assessed the accuracy and reliability of existing adult software, specifically SpineAnalyzer (Optasia Medical, Cheadle, UK) and AVERT (Optasia Medical) in the diagnosis of vertebral fractures in children [54–56]. The AI tools SpineAnalyzer and AVERT are semi-automated; they require an individual to identify and label the centers of vertebral bodies T4 to L4 (any non- or poorly visible vertebrae can be omitted). The tools then automatically outline the vertebral bodies using 6 (SpineAnalyzer) or 33 (AVERT) points and provide an output indicating normal or fractured vertebrae and severity of fracture based on height loss ratios (Fig. 3; [56]). The reader can manually reposition any points that were erroneously identified by the software. The conclusion of these studies is that the diagnostic accuracy of existing adult (semi-automated) software tools for vertebral fracture assessment in children is insufficiently adequate for clinical use (Table 4). Reasons for this include unossified ring apophyses, variation in vertebral shape with age and normal variants in children, all issues that are less problematic (if at all) in adults. The adult tools were trained using the radiographs of post-menopausal women and (despite the misleading final column in Fig. 3, which suggests otherwise) are based on the Genant et al. [56] classification for vertebral fractures. If morphometric vertebral analysis is to be accurate in children, then any tool developed must be trained using the spine radiographs of a cohort of healthy children [57].

Inflicted fractures (child abuse)

Inflicted metaphyseal and rib fractures in infants and young children are often difficult to detect and yet are highly predictive of abuse [58, 59]. United Kingdom national guidelines (adopted by the European Society of Paediatric Radiology) advise that images be double-reported by at least one pediatric radiologist [60, 61]; therefore in centers where this is not possible because of staffing issues, it would be helpful to have an AI-assist tool, if only to highlight suspicious areas for closer review by the radiologist on either skeletal surveys performed for suspected abuse or (perhaps more important) on radiographs performed for other indications, e.g., a chest radiograph for cough. Work has been

Fig. 3 Lateral spine dual-energy X-ray absorptiometry scan in a 12-year-old boy, left, with deformity results right. Morphometric vertebral fracture assessment using SpineAnalyzer identifies four mild (T4, T10, T11, L1) and three moderate (T9, T12, L3) fractures. *Bicon.* biconcave, *SQ* semi-quantitative score (of Genant et al. [56])

Image

Not for diagnostic use.
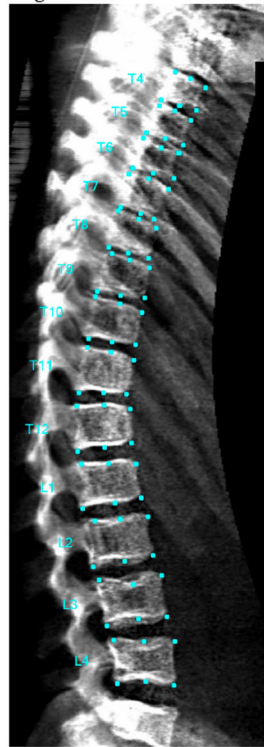
Deformity Results

Level	Wedge	Bicon.	Crush	Grade	SQ Score
	Morphometry (% Deformation)				
T4	24.3%	8.3%	-4.4%	Mild	-
T5	15.8%	2.9%	0.4%	Normal	-
T6	19.4%	16.9%	5.4%	Normal	-
T7	0.5%	-4.0%	-9.0%	Normal	-
T8	18.4%	6.6%	-2.1%	Normal	-
T9	25.1%	8.1%	6.7%	Moderate	-
T10	23.5%	9.2%	4.6%	Mild	-
T11	24.8%	10.1%	-2.0%	Mild	-
T12	25.4%	6.5%	1.2%	Moderate	-
L1	23.2%	10.7%	1.5%	Mild	-
L2	14.3%	8.2%	2.4%	Normal	-
L3	10.2%	25.0%	-7.4%	Moderate	-
L4	-2.5%	7.1%	-5.3%	Normal	-

0 - Normal < 20%	1 - Mild 20% < 25%	2 - Moderate 25% < 40%	3 - Severe ≥ 40%

Level	Differential Diagnosis
T4	-
T5	-
T6	-
T7	-
T8	-
T9	-
T10	-
T11	-
T12	-
L1	-
L2	-
L3	-
L4	-

done on the AI-assisted detection of rib fractures in adults, with encouraging results [62–65], including the development of fully automated convolutional neural networks to perform this task [66, 67]. However, to the author's knowledge, no such studies have been carried out for suspected abusive fractures in infants and young children. Research in this area should be encouraged.

To assist radiologists and others in the field, a web-based tool to unify the investigative protocol in suspected abuse and to support training and multicenter national and international research, a knowledge base to be populated with clinical information, radiographs and radiographic information has been described [68] and continues to be developed (ongoing work of author).

Other emerging/future pediatric musculoskeletal applications

Developmental hip dysplasia

Two studies have assessed the ability of neural networks to diagnose developmental hip dysplasia in children [69, 70]. Li et al. [69] used a training set of 11,473 anteroposterior pelvic

Table 4 Diagnostic accuracy of artificial intelligence (AI) applications for morphometric vertebral fracture assessment in children[a]

Author [reference]	Sensitivity	Specificity	False-positive rate	False-negative rate	Kappa
Crabtree et al. 2017 [52] (6-point technique[b])	0.79	0.71	0.13	0.27	0.24–0.60
Alqahtani 2019 [54] (SpineAnalyzer)	0.31	0.96	0.04	0.69	0.16–0.44
Alqahtani 2019 [54] (AVERT)	0.41	0.91	0.09	0.59	0.26–0.46
Alqahtani 2020 [55] (AVERT)					
Single radiographer	0.80	0.90	0.10	0.20	-
Additional observers	0.89	0.79	0.21	0.11	0.29–0.69

[a] From dual-energy X-ray absorptiometry scans

[b] The precise software tool used was not specified

radiographs and a test set of 101 images for the diagnosis of developmental dysplasia of the hip based on the Sharp angle (acetabular index). They found that accuracy was similar when compared to orthopedic surgeons and required less time, and they concluded that their AI tool could potentially replace orthopedic surgeons [69]. Zhang et al. [70] used 9,081 anteroposterior pelvic radiographs as their training set and a further 1,138 anteroposterior pelvic radiographs as their test set for the diagnosis of developmental hip dysplasia based on the acetabular index. They concluded that their deep-learning system improved consistency, convenience and effectiveness compared to clinician-led diagnosis and suggested that it might simplify current screening pathways [70].

As far as can be ascertained, neither of these tools was compared to a reference standard of pediatric radiologists. In the author's opinion, this would be an important next step before widespread use of such AI tools by pediatric radiology departments.

Spinal alignment

Several studies have been conducted to determine the degree of scoliosis and other measurements of the spine using conventional radiographic images [71], biplanar radiographic images [72] or moiré images [73]. All such studies have shown promising results and indeed automated spine and lower limb measurements are performed by the biplanar imaging system installed at the author's institution.

Other authors have assessed the ability of AI to predict scoliosis progression [74], assess the Risser stage [75], detect evidence of scoliosis treatment from radiographs [76] and to automate three-dimensional (3-D) spine reconstructions from biplanar images [77].

Miscellaneous

A few other studies conducted in children (or children and adults) and assessing AI applications in the musculoskeletal system are worthy of mention and include determining leg-length discrepancy from radiographs [78], quantifying the degree of metopic craniosynostosis from skull CT scans [79], predicting the presence of discoid lateral menisci from radiographs [80], determining muscle mass from dual-energy X-ray absorptiometry scans in cerebral palsy [81] and discerning sexual dimorphism from hand and wrist radiographs [82]. Further applications of some of these tools are obvious, e.g., diagnosis of premature fusion of sutures other than the metopic suture and determination of muscle mass in other conditions such as myopathies and juvenile dermatomyositis. The clinical utility of a tool that identifies gender from hand and wrist radiographs is limited to forensic imaging, perhaps helping with the identification of bodies destroyed by mass disasters, but it is significant because (to the author's

knowledge) it is the only example of a potential AI-extend tool in pediatric musculoskeletal imaging (i.e. a tool that performs a task over and above the capability of radiologists).

Computer-assisted diagnosis of skeletal dysplasias might be based both on AI-assisted morphological analysis and on the creation of "ontologies" in the skeletal dysplasia domain. An ontology organizes large datasets into sets of categories/concepts and forms relationships between them [83]. Ontologies related to skeletal dysplasias include the Human Phenotype Ontology [84], the Bone Dysplasia Ontology [85] and the dynamic Radiological Electronic Atlas of Malformation Syndromes (dREAMS) [86].

Pertaining to AI-assisted diagnosis of skeletal dysplasias based on skeletal morphometry, preliminary work using radiographs of infants from the dREAMS database has shown an accuracy of 78.0% to 87.5% for lateral spine, 68.0% to 75.0% for anteroposterior spine and 87.5% to 88.0% for anteroposterior chest radiographs in dichotomizing images to "achondroplasia" or "not achondroplasia" categories [87]. Accuracy and confidence intervals would be expected to improve using a dataset larger than that used by the authors (40 lateral spine, 16 anteroposterior spine and 26 anteroposterior chest radiographs in a ratio of 70% to 30% for training and testing, respectively). Nevertheless, the results provide proof of concept and suggest that the task is worth pursuing.

Conclusion

Bone age assessment tools are the only pediatric musculoskeletal AI tools available on the market. In recent years, increasing research has been conducted in areas such as elbow fractures, developmental hip dysplasia and scoliosis assessment. However, there is significant scope for more work, particularly in areas such as the diagnosis of vertebral fractures, inflicted injury, skeletal dysplasias and musculoskeletal oncology. Pediatric radiologists are encouraged to be members of the research teams conducting such studies, so that the reference standard used is the diagnostic accuracy of pediatric radiologists, rather than the diagnostic accuracy of clinicians from other specialties, which is the case in some publications. AI tools will not replace the pediatric musculoskeletal radiologist in the near future, if ever.

Declarations

References

1. van Ginneken B (2018) AI and radiologists — a painful divorce? ECR 2018 presentation. https://vimeo.com/258232453. Accessed 12 Mar 2021

2. No authors (n.d.) AI for radiology: an implementation guide. Products. Website. https://grand-challenge.org/aiforradiology/. Accessed 10 Mar 2021

3. Koska IO (2019) Applications of deep learning in radiology and pediatric radiology. Pediatr Radiol 49:247–317

4. Van Rijn RR, Lequin MH, Thodberg HH (2009) Automatic determination of Greulich and Pyle bone age in healthy Dutch children. Pediatr Radiol 39:591–597

5. Thodberg HH, Sävendahl L (2010) Validation and reference values of automated bone age determination for four ethnicities. Acad Radiol 17:1425–1432

6. Kim JR, Lee YS, Yu J (2015) Assessment of bone age in prepubertal healthy Korean children: comparison among the Korean standard bone age chart, Greulich-Pyle method, and Tanner-Whitehouse method. Korean J Radiol 16:201–205

7. Alshamrani K, Messina F, Offiah AC (2019) Is the Greulich & Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis. Eur Radiol 29:2910–2923

8. Alshamrani K, Offiah AC (2020) Applicability of two commonly used bone age assessment methods to 21st century UK children. Eur Radiol 30:504–513

9. Alshamrani K, Hewitt A, Offiah AC (2020) Applicability of two bone age assessment methods to children from Saudi Arabia. Clin Radiol 75:156.e1–156.e9

10. Pose Lepe G, Villacrés F, Fuente-Alba CS, Guiloff S (2018) Correlation in radiological bone age determination using the Greulich and Pyle method versus automated evaluation using BoneXpert software. Rev Chil Pediatr 89:606–611

11. Booz C, Yel I, Wichmann JL et al (2020) Artificial intelligence in bone age assessment: accuracy and efficiency of a novel fully automated algorithm compared to the Greulich-Pyle method. Eur Radiol Exp 4:6

12. Martin DD, Deusch D, Schweizer R et al (2009) Clinical application of automated Greulich-Pyle bone age determination in children with short stature. Pediatr Radiol 39:598–607

13. Unrath M, Thodberg HH, Schweizer R et al (2013) Automation of bone age reading and a new prediction model improve adult height prediction in children with short stature. Horm Res Paediatr 78:312–319

14. Martin DD, Meister K, Schweizer R et al (2011) Validation of automatic bone age rating in children with precocious and early puberty. J Pediatr Endocrinol Metab 24:1009–1014

15. Martin DD, Heil K, Heckmann C et al (2013) Validation of automatic bone age determination in children with congenital adrenal hyperplasia. Pediatr Radiol 43:1615–1621

16. Martin DD, Sato K, Sato M et al (2010) Validation of a new method for automated determination of bone age in Japanese children. Horm Res Paediatr 73:398–404

17. Van Rijn RR, Thodberg HH (2013) Bone age assessment: automated techniques coming of age? Acta Radiol 54:1024–1029

18. Martin DD, Thodberg HH (2019) Validation of a new version of BoneXpert bone age in children with congenital adrenal hyperplasia (CAH), precocious puberty (PP), growth hormone deficiency (GHD), turner syndrome (TS), and other short stature diagnoses. Horm Res Paediatr 91:26

19. Halabi SS, Prevedello LM, Kalpathy-Cramer J et al (2019) The RSNA pediatric bone age machine learning challenge. Radiology 290:498–503

20. Larson DB, Chen MC, Lungren MP et al (2018) Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology 287:313–322

21. Gaskin CM, Kähn SL, Bertozzi C, Bunch PM (2011) Predicting skeletal age. 16.Bit website. https://www.16bit.ai/bone-age. Accessed 13 Mar 2021

22. Mutasa C, Chang PD, Ruzal-Shapiro C, Ayyala R (2018) MABAL: a novel deep-learning architecture for machine-assisted bone age labeling. J Digit Imaging 31:513–519

23. Pan I, Thodberg HH, Halabi SS et al (2019) Improving automated pediatric bone age estimation using ensembles of models from the 2017 RSNA machine learning challenge. Radiol Artif Intell 1: e190053

24. Tajmir SH, Lee H, Shailam RS et al (2019) Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability. Skelet Radiol 48:275–283

25. Pan I, Baird GL, Mutasa S et al (2020) Rethinking Greulich and Pyle: a deep learning approach to pediatric bone age assessment using pediatric trauma hand radiographs. Radiol Artif Intell 2: e190198

26. Reddy NE, Rayan JC, Annapragada AV et al (2020) Bone age determination using only the index finger: a novel approach using a convolutional neural network compared with human radiologists. Pediatr Radiol 50:516–523

27. Li Y, Huang Z, Dong X et al (2019) Forensic age estimation for pelvic X-ray images using deep learning. Eur Radiol 29:2322–2329

28. Lei YY, Shen YS, Wang YH, Zhao H (2019) Regression algorithm of bone age estimation of knee-joint based on principal component analysis and support vector machine. Fa Yi Xue Za Zhi 35:194–199

29. Štern D, Payer C, Giuliani N, Urschler M (2019) Automatic age estimation and majority age classification from multi-factorial MRI data. IEEE J Biomed Health Inform 23:1392–1403

30. Štern D, Payer C, Urschler M (2019) Automated age estimation from MRI volumes of the hand. Med Image Anal 58:101538

31. Tang FH, Chan JLC, Chan BKL (2019) Accurate age determination for adolescents using magnetic resonance imaging of the hand and wrist with an artificial neural network-based approach. J Digit Imaging 32:283–289

32. Auf der Mauer M, Joop-van Well E, Herrmann J et al (2020) Automated age estimation of young individuals based on 3D knee MRI using deep learning. Int J Legal Med 135:649–663

33. Dallora L, Anderberg P, Kvist O et al (2019) Bone age assessment with various machine learning techniques: a systematic review and meta-analysis. PLoS One 14:e0220242

34. Adams JE (2013) Advances in bone imaging for osteoporosis. Nat Rev Endocrinol 9:28–42

35. Digby M, Bishop NJ, Paggiosi M, Offiah AC (2016) HR-pQCT: a non-invasive "biopsy" to assess bone structure and strength. Arch Dis Child Educ Pract Ed 101:268–270

36. Boyce A, Gordon C, Hogler W et al (2019) Skeletal health assessment in children from infancy to adolescence. Official pediatric positions of the International Society for Clinical Densitometry

https://iscd.org/learn/official-positions/pediatric-positions/. Accessed 13 Mar 2021

37. Barnett E, Nordin BE (1960) The radiological diagnosis of osteoporosis: a new approach. Clin Radiol 11:166–174

38. Nusman CM, Anink J, Otten MH et al (2015) Bone health of patients with juvenile idiopathic arthritis: a comparison between dual-energy X-ray absorptiometry and digital X-ray radiogrammetry. Eur J Radiol 84:1999–2003

39. Schündeln MM, Marschke L, Bauer JJ et al (2016) A piece of the puzzle: the bone health index of the BoneXpert software reflects cortical bone mineral density in pediatric and adolescent patients. PLoS One 11:e0151936

40. Neelis E, Rijnen N, Sluimer J et al (2018) Bone health of children with intestinal failure measured by dual energy X-ray absorptiometry and digital X-ray radiogrammetry. Clin Nutr 37:687–694

41. Alshamrani K, Messina F, Bishop NJ, Offiah AC (2019) Estimating bone mass in children: can bone health index replace dual energy X-ray absorptiometry? Pediatr Radiol 49:372–378

42. Shalof H, Dimitri P, Shuweihdi F, Offiah AC (2021) Which skeletal imaging modality is best for assessing bone health in children and young adults compared to DXA? A systematic review and meta-analysis. Bone. https://doi.org/10.1016/j.bone.2021.116013

43. England JR, Gross JS, White EA et al (2018) Detection of traumatic pediatric elbow joint effusion using a deep convolutional neural network. AJR Am J Roentgenol 211:1361–1368

44. Rayan JC, Reddy N, Kan JH et al (2019) Binomial classification of pediatric elbow fractures using a deep learning multiview approach emulating radiologist decision making. Radiol Artif Intell 1:e180015

45. Choi JW, Cho YJ, Lee S et al (2020) Using a dual-input convolutional neural network for automated detection of pediatric supracondylar fracture on conventional radiography. Investig Radiol 55:101–110

46. Kim DH, MacKinnon T (2018) Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 73:439–445

47. Yu JS, Yu SM, Erdal BS et al (2020) Detection and localisation of hip fractures on anteroposterior radiographs with artificial intelligence: proof of concept. Clin Radiol 75:237.e1–237.e9

48. Budnik M, Krawczyk B (2013) On optimal settings of classification tree ensembles for medical decision support. Health Informatics J 19:3–15

49. Bertsimas D, Masiakos PT, Mylonas KS, Wiberg H (2019) Prediction of cervical spine injury in young pediatric patients: an optimal trees artificial intelligence approach. J Pediatr Surg 54:2353–2357

50. McAllister AS, Nagaraj U, Radhakrishnan R (2019) Emergent imaging of pediatric cervical spine trauma. Radiographics 39:1126–1142

51. Adiotomre E, Summers L, Allison A et al (2016) Diagnostic accuracy of DXA compared to conventional spine radiographs for the detection of vertebral fractures in children. Eur Radiol 27:2188–2199

52. Crabtree N, Chapman S, Högler W et al (2017) Vertebral fractures assessment in children: evaluation of DXA imaging versus conventional spine radiography. Bone 97:168–174

53. Alqahtani FF, Offiah AC (2019) Diagnosis of vertebral fractures in children. Pediatr Radiol 49:283–296

54. Alqahtani FF, Messina F, Kruger E et al (2017) Evaluation of a semi-automated software program for the identification of vertebral fractures in children. Clin Radiol 72:904e20

55. Alqahtani FF, Messina F, Offiah AC (2019) Are semi-automated software program [sic] designed for adults accurate for the identification of vertebral fractures in children? Eur Radiol 29:6780–6789

56. Genant HK, Wu C, van Kuijk C et al (1993) Vertebal fracture assessment using a semiquantitative technique. J Bone Miner Res 8:1137–1148

57. Alqahtani FF, Crabtree NJ, Bromiley PA et al (2020) Diagnostic performance of morphometric vertebral fracture analysis (MXA) in children using a 33-point software program. Bone 133:115249

58. Paddock M, Sprigg A, Offiah AC (2017) Imaging and reporting considerations for suspected physical abuse (non-accidental injury) in infants and young children. Part 1: initial considerations and appendicular skeleton. Clin Radiol 72:179–188

59. Paddock M, Sprigg A, Offiah AC (2017) Imaging and reporting considerations for suspected physical abuse (non-accidental injury) in infants and young children. Part 2: axial skeleton and differential diagnoses. Clin Radiol 72:189–201

60. Royal College of Radiologists and Society and College of Radiographers (2017) The radiological investigation of suspected physical abuse in children. https://www.rcr.ac.uk/publication/radiological-investigation-suspected-physical-abuse-children. Accessed 15 Mar 2021

61. Offiah AC, Adamsbaum C, van Rijn RR (2014) ESPR adopts British guidelines for imaging in suspeted non-accidental injury as the European standard. Pediatr Radiol 44:1338

62. Kim J, Kim S, Kim YJ et al (2013) Quantitative measurement method for possible rib fractures in chest radiographs. Healthc Inform Res 19:196–204

63. Weikert T, Noordtzij LA, Bremerich J et al (2020) Assessment of a deep learning algorithm for detection of rib fractures on whole-body trauma computed tomography. Korean J Radiol 21:891–899

64. Jin L, Yang J, Kuang K et al (2020) Deep-learning-assisted detection and segmentation of rib fractures from CT scans: development and validation of FracNet. EBioMed 62:103106

65. Zhang B, Jia C, Wu R et al (2021) Improving rib fracture detection accuracy and reading efficiency with deep-learning-based detection software: a clinical evaluation. Br J Radiol 94:20200870

66. Zhou QQ, Tang W, Wang J et al (2020) Automatic detection and classification of rib fractures based on patients' CT images and clinical information via convolutional neural network. Eur Radiol 31:3815–3825

67. Meng XH, Wu DJ, Wan Z et al (2021) A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance. Skelet Radiol. https://doi.org/10.1007/s00256-021-03709-8

68. Offiah A, Hume J, Bamsey I et al (2011) ELECTRICA: electronic knowledge base for clinical care, teaching and research in child abuse. Pediatr Radiol 41:1433–1439

69. Li Q, Zhong L, Huang H et al (2019) Auxillary diagnosis of developmental dysplasia of the hip by automated detection of Sharp's angle on standardized anteroposterior pelvic radiographs. Medicine 98:e18500

70. Zhang SC, Sun J, Liu CB et al (2020) Clinical application of artificial intelligence-assisted diagnosis using anteroposterior pelvic radiographs in children with developmental dysplasia of the hip. Bone Joint J 11:1574–1581

71. Yang J, Zhang K, Fan H et al (2019) Development and validation of deep learning algorithms for scoliosis screening using back images. Comm Biol 2:390

72. Galbusera F, Niemeyer F, Wilke HJ et al (2019) Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. Eur Spine J 28:951–960

73. Watanabe K, Aoki Y, Matsumoto M (2019) An application of artificial intelligence to diagnostic imaging of spine disease: estimating spinal alignment from moiré images. Neurospine 16:697–702

74. Garcia-Cano E, Arámbula Cosio F, Duong L et al (2018) Prediction of spinal curve progression in adolescent idiopathic scoliosis using random forest regression. Comput Biol Med 103:34–43

75. Kaddioui H, Duong L, Joncas J et al (2020) Convolutional neural networks for automatic Risser stage assessment. Radiol Artif Intell 2:e180063

76. Vergari C, Skalli W, Gajny L (2020) A convolutional neural network to detect scoliosis treatment in radiographs. Int J Comput Assist Radiol Surg 15:1069–1074

77. Aubert B, Vazquez C, Cresson T et al (2019) Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting. IEEE Trans Med Imaging 38:2796–2806

78. Zheng Q, Shellikeri S, Huang H et al (2020) Deep learning measurement of leg length discrepancy in children based on radiographs. Radiology 296:152–158

79. Bhalodia R, Dvoracek LA, Ayyash AM et al (2020) Quantifying the severity of metopic craniosynostosis: a pilot study application of machine learning in craniofacial surgery. J Craniofac Surg 31:697–701

80. Ha CW, Kim SH, Lee DH et al (2020) Predictive validity of radiographic signs of complete discoid lateral menisci in children using machine learning techniques. J Orthop Res 38:1279–1288

81. Zhang C, Whitney DG, Singh H et al (2019) Statistical models to assess leg muscle mass in ambulatory children with spastic cerebral palsy using dual energy X-ray absorptiometry. J Clin Densitom 22:391–400

82. Yune S, Lee H, Kim M et al (2019) Beyond human perception: sexual dimorphism in hand and wrist radiographs is discernable by a deep learning model. J Digit Imaging 32:665–671

83. Offiah AC, Hall CM (2019) The radiologic diagnosis of skeletal dysplasias: past, present and future. Pediatr Radiol 50:1650–1657

84. Monarch Initiative (2021) The human phenotype ontology. Website. https://hpo.jax.org/app/. Accessed 15 Mar 2020

85. Groza T, Hunter J, Zankl A (2012) The bone dysplasia ontology: integrating genotype and phenotype information in the skeletal dysplasia domain. BMC Bioinforma 13:50

86. No authors (2021) Dynamic Radiological Electronic Atlas of Malformation Syndromes (dREAMS). https://d-reams.org/?page_id=82. Accessed 15 Mar 2020

87. Offiah A, Lanfranchi V, Davila Garcia M et al (2021) Image analysis and machine learning for skeletal dysplasias recognition. Poster presentation at the Child Health Technology Conference, 2–5 Mar 2021