Does AI consider children as little adults?
Performance of adult radiology AI tools applied to paediatric populations – a scoping review

*Lene Laborie, Jennifer Lee, **Edward Antram**, Regina Küfner Lein, Susan Shelmerdine*

Aim: Evaluate how adult-trained radiology AI tools perform in a paediatric population without additional training.
Rationale: Underperforming tools in children risk inaccuracies, threaten patient safety, and highlight the need for paediatric-specific development and validation.

1. INTRODUCTION

- AI has become increasingly prominent in radiology, but most algorithms are developed and validated for adult populations.
- Of 196 radiology AI tools registered with the FDA in 2022, only a handful (6–11) were approved for paediatric use, leaving children underrepresented.
- Limited inclusion of children in AI development raises safety and efficacy concerns if adult-focused models are simply "repurposed" for paediatric care.
- This scoping review explores the extent to which adult-trained radiology AI algorithms have been validated in paediatric settings—and how well they perform.

2. METHODS

- **Design**: Scoping review following PRISMA-ScR guidelines; protocol registered in Open Science Framework on 22 February 2024.
- **Search Strategy**:
  - Searched **MEDLINE (Ovid), EMBASE (Ovid), Web of Science, Cochrane Library (Wiley)** from 1 Jan 2014 to 1 Oct 2024.
  - Used both subject headings and free text terms for "adult," "child," "artificial intelligence," "radiology/imaging," and "validation/performance."
  - Included retrospective and prospective studies that tested an **adult-intended AI tool** on a **paediatric population (<21 yrs)**.
- **Screening & Eligibility**: Excluded case reports, reviews, conference abstracts, and non-English papers. Final eligibility confirmed by two independent paediatric radiologists.
- **Data Extraction**: Collated study characteristics, AI modality, performance metrics, and key findings for descriptive analysis (no meta-analysis performed).

3. RESULTS

- **Search Yield**: 4459 abstracts → 18 eligible studies.
- **Imaging Modalities**: 11 used radiographs (mostly chest), 6 used CT/PET, and 1 used ultrasound.
- **Body Regions**: Chest (n=11), MSK (n=3), neuro (n=1), thyroid (n=1), whole body (n=2).
- **Study Designs**: All but one were retrospective.
- **Key Finding**: **All adult-intended AI tools had decreased performance in children**, with >10% lower sensitivity or >0.10 lower Dice scores for segmentation tasks.
  - Some tools remained marginally useful with mild performance drops.
  - Others showed severe accuracy deficits, raising safety concerns.

4. DISCUSSION

- **Performance Drop**: Algorithms trained on adult data often fail to account for paediatric anatomical and physiological differences, leading to systematic errors.
- **Underrepresentation**: Few AI tools are specifically built or validated for children, reflecting barriers in paediatric data access and the smaller volume of paediatric imaging.
- **Clinical Implications**: Relying on adult AI software in children without proper validation risks misdiagnosis, increased false positives/negatives, and compromised care.
- **Future Directions**:
  - Encourage **paediatric-inclusive datasets** and **transfer learning** to adapt adult models.
  - Focus on facilitating **data sharing** among children's hospitals to overcome limited sample sizes.
  - Standardized guidelines for **paediatric validation** are essential before deployment.

Study & Year	Modality & Task	Paediatric Cohort	Tool & Adult Baseline	Paediatric Performance	Key Conclusion
Candemir S (SPIE, 2015)	CXR: Lung boundary detection	1 day–18 yrs; retrospective	Adult-based lung segmentation (Dice ~95% in adults)	Improved overlap from ~26% to ~76% in <2-yr-olds after paeds adjustments	**Adult-trained model struggled in younger children, but age-tailored shapes helped.**
Alqahtani F, Messina, Kruger (Clin Radiol, 2017)	X-ray: Semiautomated vertebral fracture (6-pt SpineAnalyzer™)	5–15 yrs (n=137 scans), mostly low BMD	SpineAnalyzer™ (high sens/spec ~96–97% in adults)	Kappa up to 0.47; Sens ~18%, Spec ~97%	**Not reliable for paeds VF detection, especially mild fractures.**
Kermany DS, Goldbaum (Cell, 2018)	OCT + CXR: Transfer-learning for AMD + paediatric pneumonia	paediatric subset: 624 patients w/ pneumonia	Inception V3 (adults trained on retinal images, pneumonia CXR)	Accuracy ~92.8% for pneumonia vs normal; ~90.7% bacterial vs viral	**DL approach feasible; high overall accuracy for paediatric pneumonia classification.**
Rajaraman S, Candemir S (Appl Sci, 2018)	CXR: CNN classification of paediatric pneumonia	1–4 yrs; 624 paediatric CXRs	Customized VGG16 (trained on adult/paediatric mix)	Accuracy ~96.2% normal vs pneumonia; ~93.6% bacterial vs viral	**CNN very effective; visual interpretability methods demonstrated.**
Alqahtani F, Messina, Offiah (Eur Radiol, 2019)	DXA + X-ray: Compare 6-pt vs 33-pt vertebral fracture detection	5–15 yrs (n=50)	SpineAnalyzer™ vs AVERT™ (adult sens ~98–99%)	SpineAnalyzer sens 26–31%, AVERT 36–41%; both <0.5 kappa	**Both adult-based systems had poor accuracy in paediatric vertebral fractures.**
Alqahtani F, Crabtree (Bone, 2020)	DXA: Morphometric vertebral fracture analysis (33-pt AVERT™)	5–18 yrs (n=420 images)	AVERT™ (kappa>0.61 moderate/severe fx in adults)	Sens ~80%, Spec ~90%; good for severe fx	**Comparable to expert read for severe fx but weak for mild fractures.**
Chen Y (PLoS ONE, 2021)	CXR: DL classification of paediatric pneumonia (WHO standards)	1–59 months (n=431)	DenseNet121 pretrained on CheXpert (adult data)	AUC=0.928–0.977; Sens/Spec up to ~95% in high-agreement sets	**Strong performance on WHO-defined pneumonia; performs less well on ambiguous cases.**
Shin H (Sci Rep, 2022)	CXR: Adult-trained AI for paediatric chest findings (Lunit Insight)	<18 yrs (n=2273), mean 7 yrs	Lunit Insight (adult AUC ~97–99%)	Overall accuracy 87.5%, improved to 96.9% excluding <2 yrs & cardiomegaly	**Major drop in very young children; some lesion categories performed poorly in toddlers.**
Hardie RC (AJR, 2023)	Chest CT: Lung nodule detection (FlyerScan vs MONAI)	4–17 yrs (n=59)	Adult baseline: FlyerScan sens ~84%, MONAI ~95% at 2 FPs/scan	paediatric sens 68.4% (FlyerScan), 53.1% (MONAI) at 2 FPs/scan	**Both tools had notable drop from adult performance in paediatric scans.**
Morcos G (JACR, 2023)	CXR: Pneumonia classification in 1–5 yrs (TorchXRayVision)	5856 paediatric CXRs	TorchXRayVision (adult AUC up to 0.70–1.00)	Accuracy ~76.5%, Sens 79.8%, Spec 67.7%; viral pneumonia PPV ~70%	**Decent performance; lower specificity vs adult data, especially in viral pneumonia.**
Rajaraman S (Expert Syst Appl, 2023)	CXR: Adult lung segmentation models generalizing to paediatrics	Not specified (small test subsets)	UNet/TransUNet variants (adult Dice ~0.95)	paediatric Dice up to ~0.97 with domain adaptation	**With domain adaptation, adult-seg models can work well in older paediatric cohorts.**
Salman R (Eur J Pediatr, 2023)	Chest CT: Adult lung CAD for nodule detection at simulated low dose	12–18 yrs (n=30)	Syngo CT Lung CAD (adult sens 65–100%, spec 71–83%)	Overall low sens in paeds; no major drop at 50% dose simulation	**Adult-based CAD not optimal for smaller paediatric nodules; dose reduction feasible.**
Salman R (Clin Imaging, 2023)	Chest CT: Adult-based software for paediatric nodule detection	12–18 yrs (n=30)	Syngo CT Lung CAD (adult sensitivity up to 100% in some studies)	~68% overall sens at 1-mm slice; better for nodules >3 mm	**Misses smaller nodules in paeds; software not validated <12 yrs.**
Yang J (AJR, 2023)	Ultrasound: DL vs ACR TI-RADS in thyroid nodules <21 yrs	≤21 yrs (n=139 nodules)	In-house DL model (adult sens ~87%, spec ~52%)	Sens 87.5%, spec 36.1%	**Higher sensitivity than radiologists but lower specificity in paediatric nodules.**
Kumar K (IJROBP, 2024)	CT: Auto-segmentation (nnU-Net) trained on adults	5 days–16 yrs (n=50 scans)	Adult model DSC=0.80–0.95	Adding paeds data improved DSC drastically (e.g. 0.07→0.86 bladder)	**Adult-trained segmentations often fail in kids; combining paeds data boosts accuracy.**
Chen L (Phys Imaging Radiat Oncol, 2024)	CT: RayMachine auto-contouring for head & neck	Mixed adult + paeds (n=137 scans)	RayMachine (adult-based). Typically ~high DSC in adults	DSC/HD95 near interobserver range; minor re-edits needed in majority	**Clinically acceptable for most H&N OARs in paediatric & adult positioning variations.**
Chatterjee (J Imag Inform Med, 2024)	CT: "TotalSegmentator" (adult) generalizability to paediatric organs	0–16 yrs (n=64)	Adult DSC ~0.81 on AMOS dataset	DSC ~0.73 in paeds; improved to ~0.89 after paediatric fine-tuning	**Large performance gap without retraining; paediatric data critical for better results.**
Rollan-Martinez-Herrera (PLOS ONE, 2024)	CXR: Transfer from adult pneumonia model to paediatric classification	1–5 yrs (n=5856)	Xception-based CNN (adult AUC=0.95, Acc ~82%)	paediatric AUC=0.82, Acc=~78%; small drop from adult	**Adult model still feasible in paeds, but mild performance decline.**

CONCLUSION

Adult-intended radiology AI tools do not maintain the same performance standards in paediatric settings when used as-is.

Clinicians and stakeholders should exercise caution and demand robust paediatric validation or fine tuning of AI models prior to clinical implementation.

"Children are not small adults"—AI must be trained or adapted accordingly for safe, accurate paediatric care.

REFERENCES
1.Benjamens S, et al. *NPJ Digit Med*. 2020.
2.Tierradentro-Garcia LO, et al. *J Am Coll Radiol*. 2023.
3.American College of Radiology. ACR Data Science Institute AI Central Database. 2024.
4.Sammer MBK, et al. *J Am Coll Radiol*. 2023.
5.Coet P, et al. *Pediatr Radiol*. 2023.
6.Davendralingam N, et al. *Br J Radiol*. 2020.
7.Nelson BJ, et al. *J Am Coll Radiol*. 2023.
8.Farmakis SG, et al. *J Am Coll Radiol*. 2021.
9.Halliday K, et al. *Clin Radiol*. 2016.
10.PRISMA Extension for Scoping Reviews. *Ann Intern Med*. 2018.