



# Artificial intelligence in paediatric chest imaging: applications, challenges, and future directions

John Joseph Muringathuparambil<sup>1</sup> · Shamiek Maharaj<sup>1</sup> · Bradley Max Segal<sup>1</sup> · Nasreen Mahomed<sup>1</sup>

Received: 10 August 2025 / Revised: 14 May 2026 / Accepted: 16 May 2026  
© The Author(s) 2026

## Abstract

**Background** Paediatric chest imaging is central to diagnosing respiratory and cardiopulmonary disease, particularly in low- and middle-income countries (LMICs) where pneumonia remains a leading cause of childhood mortality and radiology expertise is scarce. Artificial intelligence (AI) could expand access, standardise quality and support task-shifting in these “diagnostic deserts,” yet most systems are trained and validated on adult datasets from high-income settings, and paediatric radiographs form only a small minority of major public training cohorts — raising concerns about safety, generalisability and equity when such models are deployed in children.

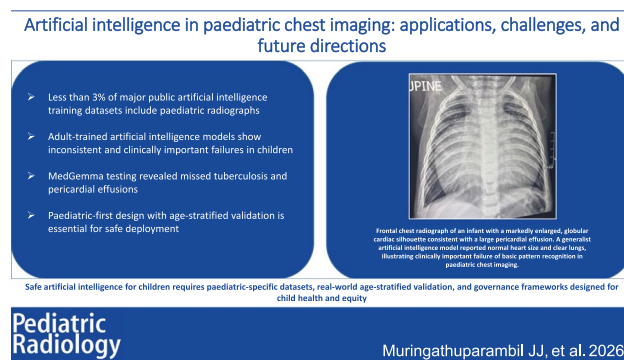
**Objective** To synthesise current applications of AI across paediatric chest radiography, lung ultrasound, computed tomography and MRI, with emphasis on LMIC workflows, and to define what is required for safe, paediatric-specific deployment.

**Materials and methods** Narrative/pictorial review of the published literature, complemented by the authors’ real-world evaluation of a generalist vision–language model (MedGemma) on adult and paediatric chest radiographs, illustrated with representative clinical cases.

**Results** Beyond diagnosis, AI shows potential in quality assurance, lung-ultrasound guidance and multilingual reporting. Real-world experience from CAD4Kids and from MedGemma’s evaluation — including critical failures in detecting pericardial effusion and tuberculosis without explicit clinical context — illustrates common failure modes and the ethical implications of domain shift. Key challenges include infrastructure constraints, dataset scarcity and the need for age-aware, explainable models.

**Conclusion** Rather than adapting adult systems, future tools must be designed from inception for children and the environments in which they live, prioritising federated learning, multimodal integration, robust validation across age strata and multilingual communication with caregivers.

## Graphical abstract



**Keywords** Artificial intelligence · Child · Diagnostic imaging · Low- and middle-income countries · Lung diseases · Radiography, thoracic · Tomography, X-ray computed · Ultrasonography

Extended author information available on the last page of the article

## Introduction and clinical context

### Global burden and the paediatric imaging gap

Paediatric chest imaging is a core diagnostic tool for respiratory infections, congenital abnormalities, and cardiopulmonary disease across diverse health systems. Pneumonia remains a leading infectious cause of death in children under five, with the burden disproportionately concentrated in low- and middle-income countries (LMICs) [1]. This clinical burden sits alongside structural constraints: radiologist shortages in many regions and limited paediatric subspecialty coverage, even where imaging volume is high.

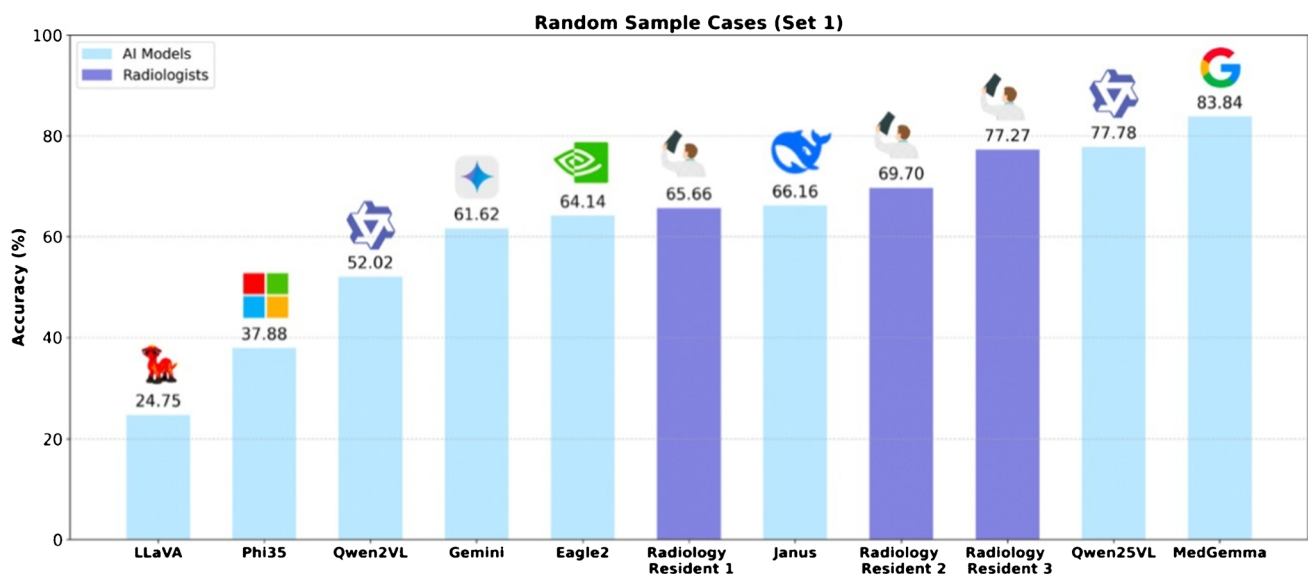
Paediatric interpretation is not simply “adult radiology in miniature”. Age-dependent anatomy, evolving physiology and normal developmental variants introduce recurring pitfalls for non-paediatric readers [2, 3]. Image acquisition is often challenging—motion, rotation and suboptimal inspiration are common—whilst radiation sensitivity narrows the margin for repeat imaging. These constraints make diagnostic support attractive, but they also heighten the consequences of error.

Artificial intelligence (AI) therefore presents both opportunity and risk. In principle, AI could support task-shifting, improve consistency and strengthen workflows in resource-constrained settings. In practice, many systems have been trained and evaluated predominantly on adult data acquired in high-resource environments, with

limited representation of paediatric imaging features or LMIC acquisition variability. This mismatch motivates the central question of this review: what is currently feasible, what fails in real-world paediatric settings and what would be required for safe integration.

### The validation gap: when adult artificial intelligence meets paediatric reality

A central barrier to paediatric AI adoption is the validation gap created by adult-dominant training and evaluation. Major public datasets include only a small proportion of paediatric images, limiting the reliability of performance claims in children [3]. Benchmarks may nevertheless appear compelling. MedGemma (4B multimodal variant; developmental at time of writing) achieved the highest overall accuracy among the eight models evaluated on the ReXVQA chest-radiograph interpretation benchmark (83.24%); on the accompanying 200-case reader study it scored 83.84%, exceeding the best-performing radiology resident (77.27%) (Fig. 1) [4]. In ReXVQA, the comparison readers were radiology residents rather than subspecialist thoracic radiologists, so “expert” performance cannot be inferred from that benchmark alone. In report-generation evaluations, board-certified radiologist review has also been used, highlighting that “good” performance depends strongly on task definition and evaluation design. Where adult-derived evidence is discussed in this review, it should be read as methodological precedent rather than validated paediatric performance, and



**Fig. 1** ReXVQA reader-study performance of MedGemma (4B) compared with the three radiology residents in that benchmark [4]. On the 200 randomly sampled cases shown, MedGemma achieved 83.84%

accuracy versus 77.27% for the best-performing resident. This result should not be interpreted as superiority over subspecialist thoracic radiologists or as evidence of paediatric clinical readiness

extrapolation to children requires age-stratified data to support it.

Such results are a signal of technical progress, but they do not establish safety or effectiveness in paediatric populations, nor do they guarantee robustness under acquisition variability and epidemiologic patterns typical of LMIC practice—including settings where tuberculosis is endemic and image quality constraints are common.

Ethical and governance considerations are intensified in paediatrics. Errors can alter life-course outcomes, and children cannot meaningfully consent or advocate for themselves. However, the generalisation problem is not reducible to geography alone: models can fail when transferred across contexts that differ in disease prevalence, comorbidity patterns, equipment and protocols, workforce skill mix and temporal shifts in practice and population health [5]. For these reasons, paediatric AI should be assessed using context-specific validation and monitored performance in the environments where children are imaged, rather than relying on adult-dominant benchmarks as a proxy for clinical readiness.

### Low- and middle-income countries realities and innovation imperatives

LMIC deployment barriers extend beyond workforce limitations to infrastructure constraints that shape both image acquisition and AI implementation. Analogue equipment persists in some settings; PACS integration and stable connectivity may be limited; and imaging may be acquired by non-specialist staff and interpreted by general clinicians, if interpretation occurs at all. These “diagnostic deserts” are precisely where AI could provide disproportionate benefit—if tools are engineered for portability, low-bandwidth or offline operation and resilience to interruptions in power and workflow.

The CAD4Kids initiative illustrates both promise and the need for local calibration. Developed through collaboration between South African clinicians and Dutch engineers, CAD4Kids achieved 76% sensitivity and 80% specificity for WHO-defined pneumonia detection at a large tertiary hospital serving children in an LMIC context [6]. Importantly, performance varied by subgroup (including HIV exposure status and demographic factors), emphasising that paediatric AI tools require population-aware validation and ongoing monitoring rather than assuming uniform performance across settings.

### Beyond diagnosis: expanding artificial intelligence applications

Although most clinical AI discussions centre on pathology detection, there are adjacent applications that may

meaningfully improve paediatric care in LMIC contexts. Natural language processing (NLP) systems could support translation of radiology outputs into local languages or generate caregiver-facing explanations, including voice-based communication, where health literacy and language barriers affect comprehension and adherence.

AI-enabled quality assurance tools may also assist at the point of acquisition. Real-time detection of motion or positioning issues could prompt corrective action before poor-quality studies enter the reporting pathway. Standardised acquisition guidance may improve consistency in environments where paediatric-specific training is limited.

Taken together, these considerations argue for paediatric-first development: globally inclusive datasets, explicit context-specific validation and deployment designs that reflect real LMIC workflow constraints and the ethical stakes of child health. The goal should not be to transplant adult systems into paediatrics, but to build approaches that begin with paediatric anatomy, disease patterns and care pathways as the design baseline.

### Current applications and real-world performance

#### Chest radiography: promise meets practical limitations

Chest radiography remains the backbone of paediatric thoracic imaging and, in many low- and middle-income countries (LMICs), the only modality consistently available. This reality explains why thoracic AI development has been heavily centred on chest radiographs. Many landmark models trained on large adult datasets report strong performance for pneumonia, pleural effusion and cardiomegaly detection [7–11]; however, these results derive primarily from adult cohorts and should be interpreted as methodological precedents rather than validated paediatric solutions. Paediatric images constitute only a small minority of major public datasets (Table 1), and external validation studies repeatedly show that adult-trained systems may generalise poorly to children [2].

Our evaluation of a contemporary vision–language model during an African health-AI exercise illustrates both capability and fragility in real clinical use. The model performed reliably on clear adult-pattern pneumonias and reproduced its previously reported benchmark performance on the ReX-VQA dataset (Supplementary Material 1; see Section 1 for benchmark context). When applied to paediatric clinical radiographs from an LMIC setting, outputs became more variable and sensitive to acquisition conditions and local disease context. In a confirmed paediatric pulmonary tuberculosis case, the model initially described a non-specific

**Table 1** Overview of selected publicly available datasets used in paediatric chest imaging artificial intelligence (AI) research

Dataset	Total images	Paediatric %	Primary use case
NIH Chest X-ray14	~112,000	<1%	Adult pathology detection
CheXpert	~224,000	<2%	Multi-label classification
MIMIC-CXR	~377,000	<1%	Report generation
PadChest	~160,000	~3%	Mixed pathology
Guangzhou Pediatric	5,856	100%	Paediatric pneumonia
CAD4Kids	1,000+	100%	LMIC paediatric TB/pneumonia

Dataset characteristics compiled from systematic review findings [3] and original dataset publications [8–10]. CAD4Kids performance and cohort characteristics are reported in [6]. The Guangzhou paediatric pneumonia dataset is described in [11]. PadChest cohort statistics are summarised in [3]. Analysis based on publicly reported statistics and research literature

LMIC low- and middle-income country, TB tuberculosis

“left upper lobe opacity” without identifying cavitation or suggesting tuberculosis (Supplementary Material 2). In settings where tuberculosis is endemic, this is a clinically important limitation because TB recognition is one of the few high-yield reasons chest radiographs remain routinely used in many paediatric respiratory pathways. In a second case, the model identified an oesophageal foreign body on an upright frontal radiograph but failed to detect the identical object on a rotated, supine projection (Supplementary Material 3). More concerning, the generated output in this case also contained implausible additional findings and diagnoses not supported by the image - an example of confabulation risk when generative systems are used outside tightly constrained tasks. This behaviour is not a cosmetic limitation: in paediatrics, confidently stated but incorrect “report-like” outputs can misdirect escalation decisions, antibiotic use, referral pathways and caregiver counselling unless the model is tightly constrained to validate tasks with explicit uncertainty signalling. In a separate evaluation on an adult case from a high-burden setting, the model similarly returned broad differentials without explicitly naming tuberculosis despite classic radiographic features (Supplementary Material 4). In a third case, the model missed a pericardial effusion despite a globular cardiac silhouette (Fig. 2).

These cases highlight recurring limitations of current paediatric CXR AI and multi-modal generative models: (1) limited paediatric representation in training data, (2) vulnerability to projectional and technical variation typical of paediatric imaging and (3) incomplete modelling of locally prevalent conditions. They also underscore a safety-relevant point: fluent “explanations” can obscure uncertainty and may introduce fabricated diagnoses unless systems are bounded by validated outputs, calibrated confidence and appropriate human oversight.

Near-term clinical value is therefore most plausible in narrow, constrained applications—line and tube localisation, automated quality assessment, prioritisation of grossly abnormal studies and decision support for non-specialist clinicians—provided these tools undergo explicit

paediatric-specific and LMIC-relevant validation and are evaluated for failure modes, including confabulation. Ultimately, progress in this domain depends on child-specific dataset development, real-world benchmarking and evaluation within the health systems where these tools are intended to function.

### Lung ultrasound at the point of care: reducing operator dependence

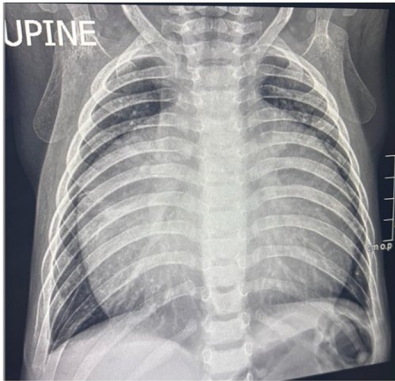
AI development in paediatric chest imaging has skewed toward chest radiography largely because adult CXR datasets are abundant, not because CXR is always the optimal modality in children. For several common paediatric indications, lung ultrasound (LUS) performs well and avoids ionising radiation [12, 13]. The practical limitation of LUS is not its diagnostic potential but its operator dependence—an area where AI may offer meaningful support [14].

Two directions are particularly relevant. First, AI-assisted acquisition guidance can provide real-time feedback to help less-experienced users obtain diagnostically adequate images. A multi-centre study demonstrated that healthcare professionals without prior ultrasound experience, when assisted by an AI guidance tool, could acquire scans of diagnostic quality comparable to expert-acquired studies [14]. Second, automated interpretation models have been developed to detect and quantify key sonographic features such as consolidation, B-lines and pleural effusions, with the goal of improving consistency at the bedside [15, 16]. Whilst early results are encouraging, paediatric evidence remains limited relative to adult literature, and performance is likely to be sensitive to device type, operator experience, scanning protocol and case mix - therefore multi-centre paediatric validation remains a prerequisite for routine deployment.

If these tools remain robust outside controlled development settings, they could support task-shifting in contexts where expert sonographers and radiologists are scarce, enabling earlier diagnosis and triage for high-burden diseases

[ User ]

Shortness and breath. Describe this X-ray.



[ MedGemma ]

Based on the provided chest X-ray, here's a description:

**Overall Impression:**

The image shows a normal heart size and shape. The lungs appear clear, with no obvious consolidation, effusions, or masses. The mediastinum is unremarkable.

**Specific Findings:**

- **Heart Size:** The heart size appears to be within normal limits.
- **Lung Fields:** The lung fields are clear, with no evidence of consolidation, effusions, or masses.
- **Mediastinum:** The mediastinum is unremarkable.
- **Ribs:** The ribs appear intact.

**Differential Diagnosis:**

Given the normal appearance of the lungs and heart, the most likely cause of shortness of breath is not apparent on this single chest X-ray. However, other conditions that could cause shortness of breath, such as pulmonary embolism, pneumothorax, or heart failure, may not be visible on this image.

**Additional Information Needed:**

To further evaluate the cause of shortness of breath, additional imaging studies, such as a CT scan of the chest, may be necessary.

**Disclaimer:** This interpretation is based solely on the provided image and is not a definitive diagnosis. A complete evaluation requires a thorough clinical history and potentially additional imaging studies.

**Fig. 2** A young male infant with a large pericardial effusion. Frontal chest radiograph demonstrates a markedly enlarged, globular cardiac silhouette consistent with significant pericardial effusion;

MedGemma 4B-it reported normal heart size and clear lungs, illustrating clinically important failure of basic pattern recognition under domain shift and paediatric acquisition variability

such as pneumonia. However, real-world paediatric deployment depends on careful validation across ages and clinical environments, and on ensuring that AI outputs integrate safely into clinical decision-making rather than replacing training and supervision [17, 18].

### Computed tomography: evolving capability and persistent limitations

Computed tomography (CT) provides detailed structural assessment of the paediatric thorax, including congenital anomalies, complications of pneumonia (e.g. necrotising infection/empyema), airway disease and oncological staging. Access remains uneven globally, and radiation exposure requires careful paediatric dose optimisation. Although AI research in chest CT is expanding, paediatric-specific evidence remains limited for several common paediatric indications—including congenital lung malformations and other routine paediatric CT referral questions—reflecting the broader scarcity of paediatric CT datasets.

Drawing primarily on adult data, AI advances in CT have been particularly visible in reconstruction and image optimisation—supporting low-dose imaging and potentially improving quality where motion and noise degrade interpretability. Emerging techniques such as denoising and motion correction may further improve image quality at lower dose, whilst synthetic augmentation approaches may help address data scarcity [19–22]. Nevertheless, paediatric-specific CT datasets remain relatively limited, and heterogeneity in acquisition parameters and age-related anatomy can constrain generalisability. For this reason, adult-derived performance claims should not be assumed to translate directly to paediatric practice. Multi-centre paediatric validation remains essential before broad clinical adoption.

### Magnetic resonance imaging: precision, motion, complexity and the promise of acceleration

Magnetic resonance imaging (MRI) is increasingly used in paediatric thoracic assessment for mediastinal masses,

congenital structural abnormalities, cardiac evaluation and chest wall pathology. Its advantages are offset by longer acquisition times, motion artefact and the potential need for sedation in younger children. Developed largely in adult populations, AI-enabled acceleration, reconstruction and motion-correction techniques have shown promise in reducing scan times and improving signal-to-noise ratios [23–26]. Although many of these approaches were developed in adult neuro-oncology and cardiac imaging, their technical value is directly relevant to paediatric thoracic MRI.

Federated learning may help address the chronic shortage of paediatric MRI datasets by enabling multi-institutional training without centralising raw patient data [27–29]. In parallel, radiomics and longitudinal modelling are being explored for risk stratification and outcome prediction in paediatric oncology [30–32]. As with CT and radiography, protocol heterogeneity, age variation and limited ground-truth standards remain barriers, reinforcing the need for deliberate paediatric-specific validation prior to routine clinical deployment.

## Technical and deployment challenges

### The infrastructure reality

Real-world deployment of AI in paediatric imaging faces constraints that laboratory benchmarks rarely capture. Experience within a large public healthcare system in a low- and middle-income country (LMIC) illustrates how basic infrastructure limitations—intermittent connectivity, ageing PACS hardware, limited storage and inconsistent integration between systems—can negate the practical value of otherwise capable algorithms. This is the central implementation paradox: the diagnostic gap AI could help address is often greatest in precisely the settings where deployment is hardest.

Children add a second layer of complexity at the point of acquisition. Motion artefact, rotation, variable inspiration and age-dependent anatomy are common in paediatric chest radiographs and can destabilise adult-trained systems. In our limited evaluation of MedGemma 4B-it during the AfriDSAI Health AI Datathon (Fig. 2), even modest projectional variation was associated with confident but incorrect outputs, including clinically important misses—an operational risk in busy services where downstream teams may over-trust fluent model responses [33].

### Paediatric generalisability: data scarcity, domain shift and failure modes

The gap between benchmark performance and clinical utility becomes most visible when models encounter paediatric

normal variants and paediatric-specific context. In one infant case (Fig. 2), the model did not identify a large pericardial effusion despite a markedly enlarged, globular cardiac silhouette. This should not be framed as a “paediatric-only” pathology problem; rather, it illustrates a more general failure mode of brittle pattern recognition under domain shift, where model confidence exceeds competence.

This is not limited to isolated examples. The thymus, age-dependent cardiothoracic ratios and developmental variation in lung appearance create interpretive contexts that adult-dominant training distributions may mishandle [2, 3]. In children, these errors have outsized consequences: false reassurance can delay escalation, whilst over-calling normal variants can trigger unnecessary antibiotics, repeat imaging, radiation exposure or referral. For paediatric deployment, the core question is therefore not whether a system can score well on a benchmark, but whether it can remain reliable across the real-world variability that defines paediatric imaging services—and whether its uncertainty is communicated in a way that discourages over-trust.

### Workflow integration and governance are safety problems, not software features

Even accurate algorithms can fail clinically if they sit outside routine workflows. Safe implementation requires integration into worklists and PACS viewers, transparent display of uncertainty, clear logging of model outputs and mechanisms for clinician override and feedback. “Invisible co-pilot” metaphors are appealing, but in paediatrics they can obscure accountability: who is responsible when an algorithm influences a missed diagnosis, delayed treatment or unnecessary intervention? Deployment should instead prioritise explicit audit trails, calibrated uncertainty display and clear responsibility pathways when outputs influence decisions.

Accordingly, implementation should be treated as a governance problem. Health systems require auditable outputs, version control, post-deployment monitoring for drift and subgroup underperformance and defined escalation pathways when model behaviour deviates from expectations. These requirements are amplified in LMIC settings due to older vendor software, limited interoperability and constrained IT support. Hybrid deployment strategies are therefore often more realistic than cloud-only designs. Edge-capable systems that can function offline—with periodic synchronisation when connectivity permits—better match many public-sector environments. The technical target in these settings is not maximal performance under ideal conditions, but predictable performance, graceful degradation when systems fail and a deployment pathway that can be supported long-term.

Accountability and liability must be defined before deployment. If an AI output contributes to harm (through missed diagnosis, biased outputs or fabricated “findings”), responsibility cannot be ambiguous. Implementation policies should specify: (i) whether AI outputs enter the medical record and how they are labelled, (ii) who is accountable for AI-influenced decisions (clinician, institution, developer or shared), (iii) how version changes are governed and audited and (iv) how privacy, cybersecurity and access controls are enforced—particularly where systems operate offline or on edge devices in LMIC settings.

### **Toward paediatric-centric design: robustness, calibration and communication**

Paediatric imaging AI should be designed around three priorities: robustness, calibration and communication. Robustness means resilience to paediatric acquisition issues (rotation, motion, low inspiration) and device/protocol heterogeneity. Calibration means that model confidence tracks true reliability, enabling clinicians to recognise when outputs should be discounted. Communication means outputs that can be interpreted across users (radiologists, general clinicians and—in selected applications—care teams), including explicit uncertainty indicators and, where appropriate, multilingual or caregiver-facing explanations.

Crucially, “acceptable performance” cannot be defended using a single headline accuracy figure. Thresholds must be task-specific and harm-aware: triage and quality assurance tools carry different risk than systems that influence treatment decisions. For this reason, paediatric deployment should be tied to prospective evaluation stratified by paediatric age bands, with monitoring frameworks that detect drift, failure clusters and subgroup underperformance over time.

The trajectory of models such as MedGemma—showing strong performance compared with radiology resident readers on standardised benchmarks—does not eliminate the paediatric validation problem; it sharpens it [4]. The urgent work is to define safe tasks, validate them in paediatric cohorts under real acquisition conditions and embed them into workflows with auditable outputs and clear accountability.

### **The technological roadmap to paediatric-centric artificial intelligence**

Having established why adult-centric performance does not automatically translate to paediatrics—particularly in resource-constrained settings—the next step is a pragmatic

roadmap. The goal is not simply “more paediatric data”, but deliberate choices about what tasks to automate, how models learn from scarce data and how outputs are governed in clinical care.

### **Quantitative biomarkers and opportunistic screening**

The future of paediatric AI extends beyond detecting overt pneumonia or effusion. One opportunity is extracting quantitative, often sub-visual biomarkers from routine chest radiographs and converting them into reproducible longitudinal measures [34–37]. For chronic disease such as cystic fibrosis, models may generate severity scores over time that support monitoring and stratification, potentially reducing reliance on repeated CT when appropriate and validated. Adult work suggests that subtle opacity and interstitial change may be detectable beyond human sensitivity; paediatric translation would require age-specific normative standards and careful validation before clinical use [36, 37].

Opportunistic screening is a related direction. Radiographs obtained for respiratory symptoms could be automatically assessed for skeletal demineralisation patterns or body-composition proxies relevant to nutrition and cardiometabolic risk, analogous to adult work using chest imaging for osteoporosis risk stratification and body-composition assessment [38–41]. In paediatrics, any such approach would require explicit governance: age-appropriate reference ranges, defined care pathways for incidental findings and safeguards against inequitable downstream consequences.

### **Multimodal integration and predictive analytics**

AI performance can improve when imaging is interpreted with clinical context rather than in isolation. Transformer-based architectures can integrate radiographs with structured clinical variables and unstructured notes [42]. This directly relates to observed failure modes in our own testing, where provision of clinical context altered outputs; mature systems should ingest such context automatically rather than relying on ad hoc prompting [42].

Longitudinal modelling extends the aim from diagnosis to prediction. Adult studies show that combining radiographs with early electronic health record data can forecast deterioration or ICU transfer with clinically useful accuracy [43]. In paediatric emergency and critical-care settings, analogous models could support earlier escalation triggers—provided they are validated prospectively and assessed for subgroup performance and unintended consequences.

## Learning from scarce, distributed paediatric data

Paediatric imaging is data-sparse by nature: rarer diseases, smaller volumes, fragmented systems. Few-shot and meta-learning techniques aim to learn meaningful patterns from limited labelled examples [44]. Self-supervised learning using large volumes of unlabelled paediatric radiographs can provide robust representations of anatomy and normal variants before task-specific fine-tuning [45], reducing dependence on heavily curated labels.

Federated learning offers a complementary path where privacy and data governance prevent centralisation. Hospitals can collaboratively train shared models whilst keeping images behind local firewalls, exchanging only model updates [28, 29]. This is particularly relevant for building globally representative paediatric tools without excluding LMIC institutions that cannot export data.

## Explainable and age-aware artificial intelligence

Safe adoption depends on systems that are interpretable and explicitly age-aware. Explainability tools (e.g. saliency methods) can reveal where a model is “looking”, sometimes helping radiologists detect obvious failure modes [46]. However, plausible-looking explanations can also worsen automation bias if clinicians over-trust convincing but incorrect rationales. Explanations therefore require validation, and radiologists need training to interpret them critically rather than as proof [47].

Age-aware evaluation is non-negotiable. Lung size, thymic contour, cardiothoracic ratio and ossification patterns evolve dramatically from neonate to adolescent. Adult-trained models may misclassify normal paediatric variants as pathology [2, 3]. Safer design strategies include using age or developmental stage as explicit inputs, training age-stratified models and reporting performance separately across predefined paediatric age bands with prospective validation thresholds.

## From assistant to more agentic workflows

In the near term, AI is most plausibly deployed as constrained workflow support: prioritising worklists, flagging critical or discrepant studies, performing routine measurements and suggesting structured report elements whilst surfacing relevant clinical context [50]. This can reduce cognitive load and variability whilst keeping radiologists responsible for final interpretation.

Longer-term research is exploring more agentic systems capable of executing multi-step workflows (e.g. detecting a critical finding, cross-checking prior imaging and vitals, drafting a report and issuing a time-stamped alert). Whether such systems are acceptable in paediatric care remains uncertain, but progress in this direction increases the importance of audit trails, data standards, and safety governance—requirements that will benefit any clinical AI deployed in children [48, 49].

## Future directions: building for children, not datasets

### Breaking free from adult-centric development paradigms

Paediatric AI development remains constrained by a common pattern: adult-trained models that are subsequently adapted with relatively small paediatric datasets. This is not only a data limitation; it is a design misalignment. Children differ from adults in anatomy, disease prevalence and the frequency of acquisition artefacts, and these differences shape both model failure modes and clinical risk. An approach designed from inception for children therefore requires deliberate problem selection (which tasks are safe and valuable), age-aware evaluation, and deployment pathways designed for the clinical environments in which children are actually imaged.

The ReXVQA benchmark exemplifies the evolution of visual question answering (VQA) evaluation, but its paediatric representation remains limited, reinforcing an adult-centric bias despite technical sophistication [3]. Addressing this is not simply a call for “more data”. Paediatric imaging data are inherently fragmented across institutions and constrained by governance, consent, and small case volumes. Future progress will depend on multi-institutional collaboration, including privacy-preserving approaches such as federated learning, and on clearly defined paediatric evaluation sets that reflect endemic disease patterns, acquisition variability and age-stratified normal appearances.

Our datathon evaluation of MedGemma was undertaken with explicit awareness of its limitations: an adult-trained generalist model applied to paediatric chest radiographs without domain-specific adaptation and evaluated only in a 4-billion parameter configuration. The failures we observed were therefore not surprising; they illustrate known constraints when general-purpose models are applied outside their training distribution. However, the same evaluation also clarified where near-term value may plausibly exist: narrow, safety-oriented tasks such

as foreign body detection, line and tube localisation, and acquisition quality assessment. In under-resourced settings where radiologists are scarce, targeted tools that reduce avoidable harm (missed tubes/lines, delayed detection of gross abnormalities, poor-quality images entering the pathway) may be more realistic and defensible than attempts at broad, autonomous diagnosis.

The key question is therefore not whether AI will be used, but how it will be built, who it will serve, and whether its deployment will be aligned with the needs and constraints of under-resourced health systems. In paediatrics, “technical performance” must be interpreted through a safety lens: error modes, subgroup performance, confidence calibration and the downstream clinical consequences of both misses and false positives. Importantly, current generalist models should be treated as research tools or narrowly scoped assistants, not as near-equivalents of trained paediatric radiologists requiring only “local induction” into a local environment. Their clinical use should therefore be limited to clearly defined tasks with prospective, age-stratified validation, explicit uncertainty display and auditable governance.

### **Explainability as communication, not just transparency**

Explainability in paediatric AI extends beyond showing “where the model looked”. Clinicians need outputs that support accountable decision-making (structured findings, uncertainty indicators and auditability), whilst caregivers require explanations that are comprehensible, culturally appropriate and available in local languages. Evidence from an international survey of health-care professionals suggests that acceptance of AI in paediatric imaging is strongly influenced by clarity of explanation and visible human oversight [51].

Future systems should therefore support layered communication: technical annotations for radiologists, concise summaries for referring clinicians and simplified or voice-based explanations for families where health literacy and language barriers limit comprehension. Importantly, caregiver-facing tools should be framed as communication support rather than as diagnostic authority, and they require governance to prevent overconfidence, misunderstanding and unintended coercion.

### **Human-centred co-design and participatory deployment**

A recurring limitation of AI deployment is the assumption that technical success will translate automatically into clinical adoption. In LMIC paediatric imaging, this is rarely

true. Co-design—where clinicians, radiographers, caregivers and local healthcare workers participate as collaborators—should be treated as a core development requirement rather than an implementation afterthought. Participatory engagement can shape what is built (high-value tasks), how it is presented (interfaces, language, uncertainty) and how it is governed (escalation pathways, accountability, documentation). This approach also improves the realism of evaluation: models can be tested against the actual failure conditions that matter (poor positioning, intermittent connectivity, non-specialist users) rather than idealised datasets.

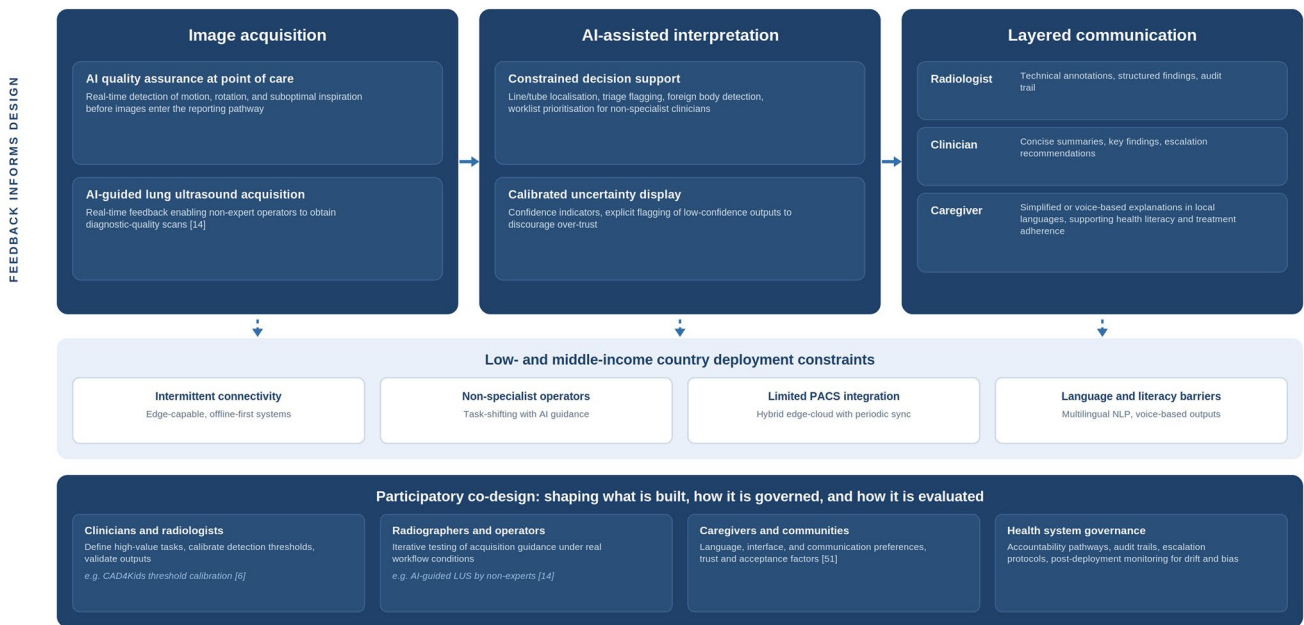
Participatory design has already been demonstrated in several paediatric imaging AI initiatives. The CAD4Kids programme, for example, involved active collaboration between South African clinicians, radiographers and Dutch engineers to define clinically meaningful diagnostic targets and calibrate model detection thresholds for local disease patterns, including in a population with high HIV exposure rates [6]. Similarly, AI-assisted lung ultrasound guidance projects have incorporated frontline healthcare workers—including those with no prior ultrasound experience—into iterative development and testing cycles, with the aim of ensuring that acquisition guidance tools reflect real bedside workflows rather than idealised laboratory conditions [14].

These examples illustrate that participatory co-design is not merely aspirational: it is practically achievable and demonstrably improves both the relevance and the real-world performance of AI systems deployed in resource-constrained paediatric settings. Figure 3 provides a conceptual overview of how these elements connect across the clinical workflow, from image acquisition and AI-assisted interpretation through layered communication with clinicians and caregivers, within the deployment constraints and participatory processes that define safe implementation in resource-limited settings.

### **Training AI-literate practitioners and strengthening accountability**

As AI becomes more embedded in paediatric imaging workflows, radiologists will need skills beyond operating software. Training should emphasise domain shift awareness, uncertainty interpretation, automation bias and the practicalities of audit trails and monitoring, so that clinicians can act as responsible stewards of AI outputs rather than passive recipients [50]. This is particularly important in LMIC settings, where radiologists may be the primary interface between AI tools and broader clinical teams.

In parallel, ethical and legal accountability must be explicit. Paediatric patients are uniquely vulnerable to harm from missed diagnoses and biased outputs. Health systems therefore need clear policies on responsibility for AI-influenced decisions, documentation of AI outputs in the medical



**Fig. 3** Proposed framework for practical integration of artificial intelligence in paediatric chest imaging in low- and middle-income country settings, illustrating the workflow from image acquisition through layered communication, the deployment constraints that shape imple-

mentation and the participatory co-design process involving clinicians, operators, caregivers and health system governance. *LUS*, lung ultrasound; *NLP*, natural language processing; *PACS*, picture archiving and communication system

record, privacy and security safeguards, and defined escalation pathways when model behaviour is inconsistent with clinical expectations. Without these safeguards, “explainability” risks becoming cosmetic rather than protective.

### Conclusion

Paediatric chest imaging sits at the intersection of high clinical burden and high uncertainty in medical AI. Adult benchmark performance has not yet translated into reliable, equitable tools for children—particularly in LMIC contexts where tuberculosis, pneumonia and congenital cardiopulmonary disease remain major causes of morbidity and mortality.

Our experience with MedGemma and CAD4Kids illustrates two divergent trajectories: powerful generalist models that can fail on endemic paediatric disease and acquisition variability, and locally calibrated systems that can perform well within narrowly defined, clinically relevant tasks. Moving forward, the most defensible route is paediatric-first development: validation across developmental age bands and context-specific evaluation in the settings where children are imaged. AI must be explainable not only to clinicians but also communicable to caregivers, aligned to local languages and health literacy, and embedded within governance structures that preserve accountability.

The central question is no longer whether AI can interpret chest radiographs, but whether we design and deploy these systems in ways that genuinely benefit children. At present, most AI systems applied to paediatric chest imaging should be regarded as research tools or narrowly scoped clinical assistants rather than autonomous diagnostic systems—a distinction that must be preserved in both clinical practice and policy. The technology may be advancing quickly; paediatric safety, equity and real-world utility must advance faster.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1007/s00247-026-06671-6>.

**Acknowledgements** Not applicable.

**Author contributions** All authors contributed to the conception and design of this review. JJ Muringathuparambil and S Maharaj performed the literature search and drafted the manuscript. BM Segal contributed to the literature review and critical analysis. N Mahomed conceived and supervised the review, contributed the de-identified clinical images, and critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

**Funding** Open access funding provided by University of the Witwatersrand.

**Data availability** No datasets were generated or analysed during the current study.

## Declarations

**Use of artificial intelligence** Artificial intelligence–assisted tools were used to support language editing, structural refinement and drafting of phrasing. All scientific statements, interpretations and conclusions were developed and verified by the authors using the cited literature and original clinical expertise. All AI-assisted suggestions were critically reviewed and edited, and the authors take full responsibility for the accuracy and integrity of the final manuscript.

**Ethics approval and consent to participate** this article does not contain any studies with human participants or animals performed by any of the authors.

**Conflicts of interest** None

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Liu L, Oza S, Hogan D et al (2015) Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. *Lancet* 385:430–440
- Shin HJ, Son NH, Kim MJ, Kim EK (2022) Diagnostic performance of artificial intelligence approved for adults for the interpretation of pediatric chest radiographs. *Sci Rep*. <https://doi.org/10.1038/s41598-022-14519-w>
- Padash S, Mohebbian MR, Adams SJ et al (2022) Pediatric chest radiograph interpretation: how far has artificial intelligence come? A systematic literature review. *Pediatr Radiol* 52:1568–1580
- Pal A, Lee JO, Zhang X, Sankarasubbu M, Roh S, Kim WJ, et al (2026) ReXVQA: a large-scale visual question answering benchmark for generalist chest X-ray understanding. *Pac Symp Biocomput* 31:251–264
- Ciet P, Eade C, Ho ML et al (2024) The unintended consequences of artificial intelligence in paediatric radiology. *Pediatr Radiol* 54:585–593
- Mahomed N, van Ginneken B, Philipsen R et al (2020) Computer-aided diagnosis for World Health Organization-defined chest radiograph primary-endpoint pneumonia in children. *Pediatr Radiol* 50:482–491
- Kitamura FC, Prevedello LM, Colak E et al (2024) Lessons learned in building expertly annotated multi-institution datasets and hosting the RSNA AI challenges. *Radiol Artif Intell* 6:e230227
- Wang X, Peng Y, Lu L et al (2017) ChestX-ray8: hospital-scale chest X-ray database and benchmarks for weakly-supervised classification and localization of common thorax diseases. *Proc IEEE Conf Comput Vis Pattern Recognit (CVPR)* pp 3462–3471. <https://doi.org/10.1109/cvpr.2017.369>
- Irvin J, Rajpurkar P, Ko M et al (2019) CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *Proc AAAI Conf Artif Intell* 33(1):590–597. <https://doi.org/10.1609/aaai.v33i01.3301590>
- Johnson AEW, Pollard TJ, Berkowitz SJ et al (2019) MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci Data* 6:317
- Kermayn DS, Goldbaum M, Cai W et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172:1122–1131.e9
- Shi C, Xu X, Xu Y (2024) Systematic review and meta-analysis of the accuracy of lung ultrasound and chest radiography in diagnosing community acquired pneumonia in children. *Pediatr Pulmonol* 59:3130–3147
- Yang Y, Wu Y, Zhao W (2024) Comparison of lung ultrasound and chest radiography for detecting pneumonia in children: a systematic review and meta-analysis. *Ital J Pediatr* 50:12
- Baloescu C, Bailitz J, Cheema B et al (2025) Artificial intelligence-guided lung ultrasound by nonexperts. *JAMA Cardiol* 10:245–253
- Kessler D, Zhu M, Gregory CR et al (2024) Development and testing of a deep learning algorithm to detect lung consolidation among children with pneumonia using hand-held ultrasound. *PLoS One* 19:e0309109
- Shea DE, Kulhare S, Millin R et al (2023) Deep learning video classification of lung ultrasound features associated with pneumonia. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); Vancouver, BC, Canada: IEEE; pp 3103–12
- Correa M, Zimic M, Barrientos F et al (2018) Automatic classification of pediatric pneumonia based on lung ultrasound pattern recognition. *PLoS One* 13:e0206410
- Riaz A, Cambaco O, Ellington LE et al (2021) Feasibility, usability and acceptability of paediatric lung ultrasound among healthcare providers and caregivers for the diagnosis of childhood pneumonia in resource-constrained settings: a qualitative study. *BMJ Open* 11:e042547
- Zhang R, Szczykutowicz TP, Toia GV (2025) Artificial intelligence in computed tomography image reconstruction: a review of recent advances. *J Comput Assist Tomogr* 49(4):521–530. <https://doi.org/10.1097/RCT.0000000000001734>
- Goodfellow I, Pouget-Abadie J, Mirza M et al (2020) Generative adversarial networks. *Commun ACM* 63:139–144
- Kazerouni A, Aghdam EK, Heidari M et al (2023) Diffusion models in medical imaging: a comprehensive survey. *Med Image Anal* 88:102846
- Motamed S, Rogalla P, Khalvati F (2021) Data augmentation using generative adversarial networks (GANs) for GAN-based detection of pneumonia and COVID-19 in chest X-ray images. *Inform Med Unlocked* 27:100779
- Serai SD, Rapp JB, States LJ et al (2021) Pediatric lung MRI: currently available and emerging techniques. *AJR Am J Roentgenol* 216:781–790
- Chen Y, Schonlieb CB, Lio P et al (2022) AI-based reconstruction for fast MRI—a systematic review and meta-analysis. *Proc IEEE* 110:224–245
- Klemenz AC, Reichardt L, Gorodezky M et al (2024) Accelerated cardiac MRI with deep learning-based image reconstruction for cine imaging. *Radiol Cardiothorac Imaging* 6:e230419
- Gearhart A, Anjewierden S, Buddha S, Tandon A (2025) Review of the current state of artificial intelligence in pediatric cardiovascular magnetic resonance imaging. *Children* 12:416
- Lee EH, Han M, Wright J et al (2024) An international study presenting a federated learning AI platform for pediatric brain tumors. *Nat Commun* 15:7615

28. Sheller MJ, Edwards B, Reina GA et al (2020) Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 10:12598
29. Tzortzis IN, Gutierrez-Torre A, Sykiotis S et al (2025) Towards generalizable federated learning in medical imaging: a real-world case study on mammography data. *Comput Struct Biotechnol J* 28:106–117
30. Lv L, Zhang Z, Zhang D et al (2023) Machine-learning radiomics to predict bone marrow metastasis of neuroblastoma using magnetic resonance imaging. *Cancer Innovation* 2:405–415
31. Madhogarhia R, Haldar D, Bagheri S et al (2022) Radiomics and radiogenomics in pediatric neuro-oncology: a review. *Neuro-Oncology Advances* 4:vdac083
32. Tak D, Garomsa BA, Zapaishchikova A et al (2024) Longitudinal risk prediction for pediatric glioma with temporal deep learning. medRxiv. <https://doi.org/10.1101/2024.06.04.24308434>
33. AfriDSAI, Emory University HiTi Lab, DSFSI, Pretoria Uo, editors. AfriDSAI Health AI Datathon 2025 – Programme Brochure. AfriDSAI Health AI Datathon 2025; 2025; Pretoria: University of Pretoria
34. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278:563–577
35. Viswanathan VS, Gupta A, Madabhushi A (2022) Novel imaging biomarkers to assess oncologic treatment-related changes. *Am Soc Clin Oncol Educ Book* 42:1–13
36. Rea G, Sverzellati N, Bocchino M et al (2023) Beyond visual interpretation: quantitative analysis and artificial intelligence in interstitial lung disease diagnosis “Expanding Horizons in Radiology.” *Diagnostics (Basel)* 13:2333
37. Kim W, Lee SM, Kim JI et al (2022) Utility of a deep learning algorithm for detection of reticular opacity on chest radiography in patients with interstitial lung disease. *AJR Am J Roentgenol* 218:642–650
38. Farina JM, Pereyra M, Mahmoud AK et al (2023) Artificial intelligence-based prediction of cardiovascular diseases from chest radiography. *J Imaging* 9:236
39. Ueda D, Matsumoto T, Ehara S et al (2023) Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study. *Lancet Digit Health* 5:e525–e533
40. Chinapuvvula NR, Hatley M, Khan R et al (2025) Opportunistic use of chest X-ray for identifying older adults at risk of osteoporosis and not meeting criteria for screening. *J Prim Care Community Health*. <https://doi.org/10.1177/21501319251353372>
41. Xu K, Khan MS, Li TZ, Gao R, Terry JG, Huo Y et al (2023) AI body composition in lung cancer screening: added value beyond lung cancer detection. *Radiology* 308:e222937
42. Jandoubi B, Akhloufi MA (2025) Multimodal artificial intelligence in medical diagnostics. *Information (Basel)* 16:591
43. Rahman M, Gao J, Carey KA et al (2025) Comparison of deep learning approaches using chest radiographs for predicting clinical deterioration: retrospective observational study. *JMIR AI* 4:e67144
44. Pachetti E, Colantonio S (2024) A systematic review of few-shot learning in medical imaging. *Artif Intell Med* 156:102949
45. Cheng S, Starosolski ZA, Subramanian D (2024) Self-supervised learning for building robust pediatric chest X-ray classification models. arXiv:2409.00231
46. Ihongbe I, Fouad S, Mahmoud TF, et al (2024) Evaluating explainable artificial intelligence (XAI) techniques in chest radiology imaging through a human-centered lens. *PLoS One* 19:e0308758
47. Prinster D, Mahmood A, Saria S et al (2024) Care to explain? AI explanation types differentially impact chest radiograph diagnostic performance and physician trust in AI. *Radiology* 313:e233261
48. Zou J, Topol EJ (2025) The rise of agentic AI teammates in medicine. *Lancet* 405:457
49. Jamil Abusamra HN, Ali SHM, Khidir Elhussien WA et al (2025) Ethical and practical considerations of artificial intelligence in pediatric medicine: a systematic review. *Cureus* 17:e79024
50. Najjar R (2023) Redefining radiology: a review of artificial intelligence integration in medical imaging. *Diagnostics (Basel)* 13:2760
51. Shelmerdine SC, Rosendahl K, Arthurs OJ (2022) Artificial intelligence in paediatric radiology: international survey of health care professional opinions. *Pediatr Radiol* 52:30–41

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

John Joseph Muringathuparambil<sup>1</sup> · Shamiek Maharaj<sup>1</sup> · Bradley Max Segal<sup>1</sup> · Nasreen Mahomed<sup>1</sup>

✉ Nasreen Mahomed  
nasreen.mahomed@wits.ac.za

John Joseph Muringathuparambil  
muringaradiology@gmail.com

Shamiek Maharaj  
shamiekmaharaj25@gmail.com

Bradley Max Segal  
bradleymaxsegal@gmail.com

<sup>1</sup> Faculty of Health Sciences, University of the Witwatersrand, 7 York Road, Parktown, Johannesburg 2193, South Africa