© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718207

Artificial Intelligence for Chest Radiograph Analysis in Pediatric Tuberculosis: A Comprehensive Review

AJIBOYE FEHINTI PRISCA[1], BUKOLA AJIDE[2]
[1, 2]Department of Biological Sciences, Faculty of Science and Technology,
Bingham University, Karu, Nasarawa State, Nigeria

Abstract-Tuberculosis (TB) is a leading infectious cause of childhood morbidity and mortality worldwide, with 1.1 million new paediatric cases estimated in 2020. Early detection is paramount, as untreated TB carries a 70% risk of death within ten years. Paediatric diagnosis is uniquely challenging owing to non-specific clinical manifestations, paucibacillary disease, and the difficulty of obtaining adequate respiratory specimens. Chest radiography (CXR) is a highly sensitive, accessible, and affordable screening tool, but accurate interpretation requires considerable expertise that is frequently unavailable in high-burden, resource-limited settings. The World Health Organization (WHO) now recommends the use of computer-aided detection (CAD) software to automate CXR interpretation for TB screening in individuals aged 15 years and older, yet the application of artificial intelligence (AI) to paediatric populations remains nascent. This comprehensive review critically examines the evolution, current evidence, and future potential of AI-driven CXR analysis for paediatric TB. We trace the progression from early conventional CAD systems limited by handcrafted features and modest generalizability to modern deep learning models that automatically learn hierarchical representations from large datasets. Performance metrics are reviewed in depth: while selected algorithms have achieved area under the receiver operating characteristic curve (AUC) values of up to 0.99 in curated datasets, real-world clinical evaluations demonstrate more modest AUCs ranging from 0.71 to 0.94, with sensitivity and specificity varying widely across settings. Only one commercial product, CAD4TB version 6, is currently licensed for use in children over four years. We explore the multifactorial benefits of AI, including enhanced diagnostic accuracy, reduction of interobserver variability, high-throughput screening, augmentation of limited expert capacity, and early detection of subclinical disease. Against these benefits, we highlight persistent challenges: a critical scarcity of large, microbiologically confirmed, and demographically diverse paediatric CXR datasets; the risk of overfitting and dataset-specific bias; variable reference standards; ethical, medicolegal, and regulatory hurdles; and the near-total absence of prospective clinical implementation studies. Future directions including federated learning, lightweight smartphone-deployable models, multimodal diagnostic algorithms integrating clinical and radiological data, and rigorous randomised controlled trials are discussed. With sustained investment, interdisciplinary collaboration, and an emphasis on equity, AI promises to revolutionize paediatric TB diagnosis and contribute meaningfully to the WHO End TB Strategy.

Keywords: Artificial Intelligence, Computer-Aided Detection, Chest Radiography, Paediatric Tuberculosis, Diagnosis, Screening

I. INTRODUCTION

Tuberculosis (TB), an airborne infection caused by Mycobacterium tuberculosis, remains one of the most devastating infectious diseases affecting children globally. In 2020, despite being both curable and preventable, an estimated 1.1 million children developed active TB, with the highest case burdens concentrated in sub-Saharan Africa and Southeast Asia (Chaparro & Suchdev, 2019). The COVID-19 pandemic severely disrupted TB care cascades: global TB notifications fell by 18% between 2019 and 2020, dropping from 7.1 million to 5.8 million diagnosed cases, a decline attributed to lockdowns, restricted mobility, reduced health-seeking behavior, and diversion of resources (Naidoo et al., 2023). This setback imperils the ambitious targets of the WHO End TB Strategy, which aims for a 90% reduction in TB mortality and an 80% reduction in TB incidence by 2030.

Paediatric TB presents a distinct set of diagnostic obstacles. Children often exhibit non-specific symptoms such as failure to thrive, prolonged cough, or fever, and due to the paucibacillary nature of their disease, conventional sputum smear microscopy has low sensitivity. Even advanced molecular assays such as the Xpert MTB/RIF Ultra, while highly sensitive in adults, face reduced yield because obtaining adequate respiratory specimens from young children is technically demanding and often requires invasive procedures (Naidoo et al., 2023; Nischal et al., 2023). Consequently, imaging particularly chest radiography has become an indispensable tool for the diagnosis of pulmonary TB in children, especially when bacteriological confirmation is elusive.

Chest radiography boasts a pooled sensitivity of 98% for pulmonary TB, making it an excellent screening modality (Ortiz et al., 2020). It is inexpensive, widely available, uses minimal ionizing radiation, and is recommended by the WHO for systematic screening of paediatric contacts and symptomatic children (Van't Hoog et al., 2022). However, the interpretation of paediatric CXRs is notoriously difficult. Radiographic manifestations of TB in children differ from those in adults, with hilar and mediastinal lymphadenopathy often representing the sole abnormality, while cavitary lesions are less common. Furthermore, the superimposition of thymic tissue, vascular structures, and evolving age-related anatomy introduces substantial interpretative challenges (Alshammeri et al., 2024). Unsurprisingly, interobserver agreement in reading paediatric CXRs for TB is suboptimal, with kappa values reported as low as −0.03 for mediastinal lymphadenopathy (Acanfora et al., 2024). In many high-burden countries, the shortage of radiologists, and particularly paediatric radiologists, is acute; clinicians with limited imaging training are often forced to make diagnostic and therapeutic decisions based on uncertain radiographic readings (Andom et al., 2022).

Artificial intelligence (AI) has emerged as a transformative technology across medical imaging. In thoracic radiology, deep learning-based computer-aided detection (CAD) algorithms have demonstrated the ability to detect and classify a range of abnormalities, including pulmonary nodules, consolidation, and pleural effusions, with an accuracy that rivals or occasionally surpasses that of human experts (Chassagnon et al., 2020). For TB specifically, the WHO's March 2021 updated guidelines on TB screening recommended the use of CAD software to automate the interpretation of digital CXRs in individuals aged 15 years and older, generating a numerical score that reflects the probability of active TB (Naidoo et al., 2023). This landmark endorsement was informed by a growing body of evidence that CAD systems can augment or even replace human readers in certain screening contexts, while simultaneously reducing costs and improving efficiency.

Despite these advances, the literature on AI for paediatric TB diagnosis remains sparse. The vast majority of CAD products have been developed and validated exclusively in adult populations, and only one CAD4TB version 6 is licensed for use in children older than four years (Ciet et al., 2023). Paediatric CXR datasets are small, often poorly annotated, and rarely incorporate microbiological confirmation as a reference standard. Moreover, the biological and radiographic heterogeneity of childhood TB threatens the generalizability of models trained predominantly on adult images. As interest and investment in AI for global health surge, there is an urgent need for a comprehensive, critical appraisal of the field specific to children.

The present review provides a thorough synthesis of the role of AI in chest radiograph analysis for paediatric TB. We describe the fundamental concepts of AI, machine learning, and deep learning; trace the evolution of CAD from early rule-based systems to modern neural networks; summarize diagnostic performance data from peer-reviewed studies and meta-analyses; explore the multidimensional benefits of AI-assisted interpretation; analyze the persisting challenges related to data, bias, regulation, and implementation; and outline future directions that could bridge the gap between promising research prototypes and scalable clinical tools.

II. CHEST RADIOGRAPHY IN PAEDIATRIC TB: STRENGTHS AND LIMITATIONS

2.1. Radiographic Features and Diagnostic Yield

Pulmonary TB is the most common form of TB in children, with up to 80% of paediatric cases manifesting thoracic disease (Allwood *et al*., 2021). The chest radiograph can reveal a spectrum of abnormalities, including parenchymal infiltrates, airspace consolidation, nodules, cavities, pleural effusion, and most characteristically in children hilar and mediastinal lymphadenopathy (Boddu *et al*., 2017; Hantous-Zannad *et al*., 2022). Lymphadenopathy, often the sole radiological indicator of primary TB, may cause compression of adjacent airways, leading to segmental or lobar atelectasis and air trapping, yet it is also one of the most difficult signs to detect due to overlap with normal mediastinal structures (Taka *et al*., 2023). Cavitary lesions, while more common in adolescents and adults with reactivation disease, can occur in children and are associated with higher bacillary loads and transmission risk.

The diagnostic accuracy of CXR, however, is heavily operator-dependent. Studies that have evaluated the detection of specific radiographic findings report widely variable interobserver agreement. For example, kappa values for identifying lymphadenopathy range from −0.03 to 0.52, while agreement for consolidation and cavitation is somewhat better but still far from perfect (Acanfora *et al*., 2024). A landmark study by Drozdov *et al*. (2021) demonstrated that even in emergency department settings, classification of CXRs as normal or abnormal showed only low-to-moderate agreement among physicians. This variability can lead to both over diagnosis exposing children to unnecessary treatment and underdiagnoses, with the attendant risks of disease progression, transmission, and mortality.

The WHO, in its latest consolidated guidelines, has attempted to mitigate some of these limitations by recommending a shift in focus toward radiographic signs that demonstrate better inter-reader agreement and that correlate with clinical severity. This allows classification of children into severity categories, some of whom may be eligible for shorter treatment regimens (Van't Hoog *et al*., 2022). However, the sensitivity of this approach still depends on the availability of adequately trained readers, which remains a critical bottleneck.

2.2. CXR as a Screening Tool for Paediatric TB

Beyond its diagnostic role, CXR is a potent screening tool, particularly among asymptomatic children with known TB exposure. In a prospective cohort of 4,468 children exposed to TB in Korea, Hwang *et al*. (2019, cited in Naidoo *et al*., 2023) found that asymptomatic children with abnormal baseline CXRs had a 25.1-fold increased risk of coexistent TB and a 26.7-fold increased risk of developing active TB during follow-up compared to children with normal radiographs. This study elegantly illustrates that a substantial proportion of children harbour subclinical TB disease that would be missed by symptom-based screening alone, and that CXR can identify these high-risk individuals for early intervention.

The WHO therefore recommends that CXR be integrated into systematic TB screening protocols for children across all age groups, in conjunction with symptom assessment and, where available, molecular testing. However, the large volumes of CXRs generated by such screening programmes can overwhelm the limited radiological workforce. In high-burden settings such as Lesotho, qualitative research has identified the lack of trained readers and delays in CXR reporting as major barriers to effective implementation (Andom *et al*., 2022). It is precisely this gap that AI-based CAD systems are designed to fill.

III. ARTIFICIAL INTELLIGENCE IN CHEST RADIOGRAPH ANALYSIS FOR TB

3.1. Definitions: AI, Machine Learning, and Deep Learning

Artificial intelligence refers to the broad capability of machines to perform tasks that would normally require human intelligence, including pattern recognition, decision-making, and learning from experience (Alkatheiri, 2022). Machine learning (ML), a subset of AI, involves the use of statistical algorithms that improve their performance on a specific task as they are exposed to more data, without being explicitly programmed for every scenario (Nassehi *et al*., 2022). Deep learning (DL), a more recent and powerful subfield of ML, employs artificial neural networks with many layers (hence “deep”) to automatically extract hierarchical features from raw data, such as the pixel intensities in a chest radiograph (Ma *et al*., 2021). Convolutional neural networks (CNNs), the most common DL architecture in imaging, use learnable filters to detect edges, textures, shapes, and finally complex objects, making them exquisitely suited for medical image analysis.

3.2. Conventional Computer-Aided Detection (CAD) Systems

Before the rise of deep learning, CAD systems for TB detection relied on handcrafted feature engineering. Radiographic characteristics such as lung texture, shape-based features, and optic flow were manually defined by human experts and fed into classifiers such as support vector machines or random forests. These early systems demonstrated the feasibility of automated CXR analysis but suffered from limited robustness, as the explicitly programmed features could not fully capture the protean manifestations of TB.

For instance, Xu *et al*. (2013) developed a CAD algorithm specifically designed to detect TB cavities using a coarse-to-fine dual-scale technique. Based on a modest dataset of 35 CXRs containing 50 cavities from a Canadian hospital, the system achieved a sensitivity of 78.8%, specificity of 86.8%, and overall accuracy of 82.8%. While impressive for its time, its exclusive focus on cavities rendered it unsuitable for detecting the full range of paediatric TB presentations. Using a different methodology, Singh *et al*. (2022) computationally extracted the rib cage from CXR images, making any residual opacity highly conspicuous; their system detected focal TB lesions with 85% accuracy on a dataset of 200 images.

A more comprehensive approach was taken by Jaeger *et al*., who developed a multi-stage TB detection pipeline incorporating lung field segmentation, texture analysis, and feature classification (Kulkarni & Jha, 2020). Trained on 138 CXRs representing the varied features of primary TB, the system achieved an accuracy of 83%, approaching the performance of human readers for that specific task. Nevertheless, each of these systems was constrained by the limited size and diversity of training datasets, the fixed nature of the feature sets, and the inability to learn from new data.

The most prominent conventional CAD system, CAD4TB (Delft Imaging Systems, The Netherlands), has undergone multiple iterations. Early versions used predefined image features and were evaluated in a systematic review by Pande *et al*. (2016), which included five studies from Zambia, Tanzania, South Africa, and the United Kingdom. The pooled AUC ranged from 0.71 to 0.84, and sensitivity and specificity varied considerably. Attempts to augment the CAD4TB score with clinical features such as haemoptysis, night sweats, or fever produced only marginal improvements: Melendez *et al*. (2016) reported an AUC of 0.84, with a sensitivity of 49% and a specificity of 95% when clinical data were combined with the imaging score. These findings underscored the limitations of feature-engineering approaches and prompted a shift toward deep learning.

3.3. Deep Learning Models for TB Detection

The application of deep learning, particularly transfer learning, catalyzed a quantum leap in the accuracy and flexibility of TB detection from CXRs. Transfer learning involves taking a CNN that has been pre-trained on a large generic image dataset most commonly ImageNet, which contains over 14 million non-medical images across thousands of categories and fine-tuning it on a smaller medical imaging dataset. This strategy dramatically reduces the quantity of labelled medical images needed and leverages features learned from natural images that are surprisingly useful for radiographic analysis.

Hwang *et al*. (2018) were among the first to apply a pre-trained AlexNet architecture to TB detection, using 10,848 CXRs from the Korean Institute of Tuberculosis (KIT). The model achieved an AUC of 0.964 on the KIT test set, but its performance declined to an AUC of 0.88 on the U.S. National Institutes of Health (NIH) dataset and 0.93 on the Shenzhen dataset from China, illustrating the challenge of domain shift performance degradation when a model is applied to data from a different source with different acquisition parameters, patient demographics, or disease prevalence. This phenomenon, known as overfitting to the training domain, remains one of the most critical barriers to generalizable AI.

Lakhani and Sundaram (2017) conducted a rigorous comparison of pre-trained and untrained AlexNet and GoogLeNet models on a dataset of 1,007 CXRs sourced from four different institutions. The untrained versions achieved AUCs of 0.90 and 0.88, respectively, while the pre-trained versions reached AUCs of 0.98 and 0.97. Most remarkably, an ensemble of both models combined with a radiologist to adjudicate discrepant cases yielded a sensitivity of 97.3%, specificity of 100%, and AUC of 0.99, demonstrating the synergistic potential of combining AI with human expertise. This study also provided strong evidence for cross-dataset validity, a necessary attribute for global deployment.

The computational burden of large networks can be prohibitive in low-resource settings. Pasa *et al*. (2023) addressed this by training a lightweight neural network from scratch on only 1,104 CXRs from multiple public datasets, including the NIH, Shenzhen, and Belarus Tuberculosis Portal databases. Their model contained merely 230,000 parameters compared to 7 million in GoogLeNet and 60 million in AlexNet and required only 350 mega-FLOPs of computing power, enabling it to run on a central processing unit costing under $200. The AUCs achieved were 0.811 (NIH), 0.900 (Shenzhen), and 0.925 (combined), rivalling some of the larger, more computationally expensive networks. This work demonstrated that diagnostic accuracy need not be sacrificed for deployability.

In addition to diagnosis, researchers have explored the potential of AI to extract “hidden” information from CXRs. Jaeger *et al*. (2018) trained a DL model to classify drug-sensitive versus multi-drug-resistant TB directly from radiographic appearance. Using a small dataset of 135 images (45% drug-sensitive, 54% multi-drug-resistant), the model attained an AUC of 0.66 discerning above chance but well below clinical utility. While this specific application remains experimental, it points toward the tantalizing possibility that AI might detect imaging signatures of drug resistance before they are clinically apparent.

3.4. Commercial CAD Products and Meta-analyses

A rapid proliferation of commercial CAD products has occurred in parallel with academic development. The

Stop TB Partnership and the Foundation for Innovative New Diagnostics (FIND) have established an online resource that catalogues and compares these products. Schalekamp *et al*. (2022) reviewed 40 Conformité Européenne (CE)-marked commercial AI software packages for chest imaging and found that none were specifically designed or validated for paediatric populations. Most products were intended for adult screening, and their performance characteristics in children were extrapolated rather than empirically demonstrated.

Several systematic reviews and meta-analyses have assessed the global performance of AI-based CAD for TB. Harris *et al*. (2019) systematically reviewed the diagnostic accuracy of AI programs for pulmonary TB on CXRs, finding a high risk of bias particularly in development studies that used curated databases rather than clinical cohorts. The authors cautioned that higher AUCs observed in development studies may not translate to real-world settings. Nabulsi *et al*. (2021) reported AUCs ranging from 0.78 to 0.99, sensitivities from 0.56 to 0.97, and specificities from 0.36 to 0.95 across included studies, highlighting considerable heterogeneity. A more recent meta-analysis by Hua *et al*. (2023) confirmed that certified AI products achieved high pooled sensitivity and specificity for detecting active TB, even among individuals with previous TB, supporting their role as triage or screening tools.

Suchitra *et al*. (2023) reviewed 53 articles 40 on development and 13 on clinical evaluation and concluded that while DL-based CAD systems are promising, the evidence base is dominated by small, single-centre studies lacking external validation. The authors called for prospective, multicentre clinical studies that minimize spectrum bias and use microbiologically confirmed reference standards.

3.5. AI Performance in Paediatric Populations

The direct evidence for AI performance in paediatric TB is extremely limited. Mouton and Vaezipour (2022) appear to have conducted one of the first studies evaluating AI detection of abnormalities in paediatric CXRs from a high-TB-incidence setting, achieving an AUC of 0.78 for correctly identifying abnormal regions. While this is a promising start, the modest AUC underscores the difficulty of the task and the need for dedicated paediatric algorithms. The CAD4TB system, now re-engineered as a deep learning model, has been used in children in some field settings, but only version 6 has received regulatory approval for those older than four years. Its diagnostic accuracy in children younger than four years, who have the highest risk of progression from infection to disease and the most non-specific radiographic findings, remains unknown.

It is also noteworthy that many AI models have been trained predominantly on images from adult pulmonary TB, which features apical infiltrates, cavitation, and fibrosis, whereas paediatric TB is characterized by lymphadenopathy, lower lobe consolidation, and airway compression. A model trained on adult patterns may systematically misinterpret paediatric pathology or, conversely, overcall normal paediatric anatomy as abnormal. The few studies that have explicitly tested adult-trained algorithms on paediatric images have generally observed a degradation in performance, reinforcing the need for dedicated paediatric training data.

IV. BENEFITS OF AI-ASSISTED PAEDIATRIC TB DIAGNOSIS

4.1. Improved Diagnostic Accuracy and Consistency

AI algorithms, once trained, are immune to the fatigue, cognitive biases, and experiential variability that affect human readers. They can consistently identify subtle radiographic abnormalities such as low-contrast lymphadenopathy or faint infiltrates that might be overlooked, particularly by non-radiologist clinicians working in under-resourced settings (Kotei & Thirunavukarasu, 2024). By providing an objective, standardized second read, AI has the potential to reduce both false-positive and false-negative rates, leading to more judicious use of anti-TB therapy and fewer missed cases. In a multicentre study by Choi *et al.* (2021), both radiology residents and board-certified radiologists demonstrated improved sensitivity for detecting active TB when assisted by a DL-based CAD algorithm, with the greatest gains among less experienced readers. This "levelling" effect could be particularly impactful in settings where paediatric radiologists are unavailable.

4.2. Increased Efficiency and High-Throughput Screening

AI systems can analyze a CXR in seconds, generating a probability score that can automatically triage cases for urgent review or confirmatory testing. This speed is transformative for mass screening campaigns, such as those conducted in schools, refugee camps, or household contact investigations. Simulation studies using CAD4TB in adults demonstrated that setting a threshold score of 85 (on a scale of 0–100 for abnormality) could reduce the need for costly molecular testing by 60%, cutting costs per screened individual and per notified TB case by more than 50% (Naidoo *et al.*, 2023). Similar cost-effectiveness analyses in paediatric populations, though lacking, are likely to show even greater relative benefits given the higher proportion of normal CXRs and the greater cost of invasive specimen collection in children.

4.3. Augmentation of Expertise and Task-Shifting

In many endemic regions, TB screening is performed by community health workers or nurses who have limited training in radiograph interpretation. AI-enabled smartphone applications can bypass the need for PACS, local IT infrastructure, or even a radiologist on site, empowering frontline providers to make informed decisions about referral or treatment initiation (Hwang *et al.*, 2024). AI can also serve as a continuous training tool, offering annotated heatmaps that highlight suspicious regions and providing feedback to novice readers. This model of task-shifting, where AI augments the capabilities of lower-cadre health workers, is particularly appealing for achieving universal health coverage and the End TB goals.

4.4. Early Detection and Reduced Transmission

Children with subclinical TB radiographic abnormalities in the absence of overt symptoms represent a hidden reservoir of infection that sustains community transmission. AI-based screening of large paediatric CXR repositories, including those stored from prior contact investigations or hospital admissions, could identify these cases and prompt early treatment. By shortening the duration of infectiousness and preventing progression to cavitary or disseminated disease, such an approach could reduce both morbidity and transmission. Furthermore, integration with electronic health records could enable automated surveillance, alerting clinicians to children with CXR abnormalities consistent with TB even when TB was not initially suspected.

V. CHALLENGES AND LIMITATIONS

5.1. Scarcity and Quality of Paediatric Training Data

The single greatest barrier to the development of accurate paediatric AI models is the shortage of large, well-annotated datasets. Most publicly available CXR datasets (e.g., NIH ChestX-ray14, CheXpert, MIMIC-CXR) contain predominantly adult images and only a small fraction of paediatric cases. Even among available paediatric datasets, TB-specific labelling is rare, and the reference standard often relies on radiological interpretation rather than microbiological confirmation. The misclassification inherent in such labels propagates error, limiting the ceiling of achievable model performance. Crowdsourcing initiatives, data sharing agreements, and the creation of federated data networks are essential to overcome this bottleneck.

5.2. Domain Shift, Overfitting, and Bias

AI models are highly sensitive to the characteristics of their training data. Differences in X-ray equipment, kilovoltage peak (kVp) and milliampere-seconds (mAs) settings, post-processing algorithms, patient positioning, and digital versus film-based capture can all degrade performance when a model is deployed in a new setting (Chassagnon *et al*., 2020). Models trained on high-resolution digital radiographs from a tertiary centre in South Korea may perform poorly on smartphone photographs of film-based CXRs taken in a rural clinic in sub-Saharan Africa. This domain shift is compounded by biological differences: nutritional status, HIV co-infection, and age-specific chest anatomy further widen the gap between training and deployment populations. Bias can also arise from the over-representation of certain demographic groups in training data, leading to systematically lower accuracy in under-represented populations—a critical ethical and fairness concern (Harris *et al*., 2019).

5.3. Reference Standard and Labelling Dilemmas

Accurate labels are the foundation of supervised learning. In paediatric TB, the paucibacillary nature of the disease means that even the best molecular tests may be negative in true cases, while culture the gold standard is slow and insensitive. As a result, many studies use a composite reference standard that includes radiographic and clinical criteria, introducing circularity when evaluating AI that is trained on those same radiological features. Independent, microbiologically confirmed cohorts are urgently needed to establish robust ground truth. The same issue applies to "normal" labels: a CXR read as normal by a general practitioner may harbour early lymphadenopathy that a paediatric radiologist would detect, causing inconsistent labelling and limiting the effective signal for AI training.

5.4. Regulatory, Ethical, and Medicolegal Challenges

Regulatory frameworks for AI-based medical devices are evolving but remain immature, especially for paediatric indications. In most jurisdictions, AI CAD products are classified as Class II medical devices, requiring demonstration of safety and efficacy through clinical validation studies. To date, only CAD4TB v6 has obtained a paediatric licence, and no product has undergone the rigorous pre-market approval process that would be required for widespread implementation in children. Ethical questions abound: who is liable if an AI algorithm misses a case of TB? How is patient privacy protected when images are processed in the cloud? How do we obtain informed consent for AI-based screening in communities with low digital literacy? Addressing these questions is essential to building trust among providers, patients, and policymakers.

5.5. Absence of Prospective Implementation Studies

The vast majority of AI studies in TB are retrospective, using pre-existing datasets and simulated workflows. While these provide proof-of-concept, they do not capture the complexities of real-world clinical environments workflow integration, user acceptability, impact on clinical decision-making, or patient outcomes. No randomized controlled trial has yet compared AI-assisted CXR interpretation to standard-of-care reading in paediatric populations. Without such evidence, the true clinical utility and cost-effectiveness of these tools remain speculative. Implementation science frameworks should be applied to study the adoption, sustainability, and scalability of AI in diverse health systems.

5.6. Paediatric-Specific Anatomical and Pathological Considerations

The radiographic appearance of the paediatric chest varies markedly with age: the thymus is prominent in infants, the cardiothoracic ratio is larger, and the lungs are less hyper inflated than in adults. Normal variants such as thymic sail signs or physiological peribronchial thickening can mimic pathology. AI models trained on adults may mistake these normal paediatric features for disease or, conversely, normalize true pathology that does not conform to adult patterns. TB itself manifests differently; primary disease with lymphadenopathy and segmental consolidation may be misinterpreted by a model trained to detect apical cavities. Dedicated paediatric datasets that span the full age spectrum from infancy to adolescence, annotated by expert paediatric radiologists, are needed to develop age-appropriate AI systems.

VI. FUTURE DIRECTIONS

6.1. Building Large, Diverse, and Microbiologically Confirmed Paediatric Datasets

The foremost priority is the creation of multicentre, multi-national paediatric CXR repositories with linked clinical and microbiological data. Initiatives such as the Global Paediatric TB Consortium and collaborations between academic centres in high- and low-burden countries could pool de-identified images and standardize annotation protocols. Microbiological confirmation using Xpert MTB/RIF Ultra or culture should serve as the reference standard wherever possible. Data should represent the full diversity of age, nutritional status, HIV co-infection, and disease severity to ensure model robustness.

6.2. Federated Learning and Privacy-Preserving AI

Federated learning enables AI models to be trained collaboratively across multiple institutions without the need to share raw patient data, thus preserving privacy and navigating complex data governance regulations (Guan & Liu, 2023). This approach is particularly well-suited to paediatric TB, where any single centre may have too few cases to train a deep model but a network of centres could collectively provide thousands. Algorithms such as secure aggregation and differential privacy can further enhance data protection.

6.3. Lightweight and Smartphone-Deployable Models

Research into model compression, quantization, and efficient neural network architectures (e.g., MobileNet, EfficientNet) can yield AI systems that run directly on smartphones or low-cost hardware without internet connectivity. This would obviate the need for PACS and high-speed internet, enabling point-of-care AI in the most remote settings. Paediatric versions of such models could be loaded onto community health workers' devices, integrated with symptom screening apps, and coupled with portable X-ray units to create a self-contained diagnostic platform.

6.4. Multimodal and Longitudinal AI

Future AI systems should not rely on CXR data in isolation. Integrating clinical variables (age, weight, symptoms, TB contact history), laboratory results (C-reactive protein, GeneXpert results), and even point-of-care ultrasound findings into a multimodal AI model could substantially improve diagnostic accuracy. Additionally, longitudinal models that compare serial CXRs over time could detect subtle progression, assess treatment response, and predict the risk of paradoxical reactions such as immune reconstitution inflammatory syndrome (IRIS) in HIV-coinfected children (Mahomed *et al*., 2023). Such comprehensive approaches, while technically demanding, would mirror the integrative diagnostic reasoning of expert clinicians.

6.5. Rigorous Clinical Validation and Implementation Trials

The roadmap to clinical adoption must include well-designed, prospective studies that measure patient-centered outcomes. Cluster-randomized trials comparing AI-assisted versus standard CXR interpretation in community or primary care settings could evaluate time to diagnosis, treatment initiation rates, and TB-related mortality. Mixed-methods research should explore user acceptability, workflow integration, and barriers to adoption among healthcare workers, patients, and families. Economic evaluations should determine cost-effectiveness across different epidemiological and health system contexts, informing investment decisions by governments and donors.

6.6. Regulatory Harmonization and Ethical Frameworks

International regulatory bodies, including the WHO, the U.S. Food and Drug Administration, and the European Medicines Agency, should collaborate to develop harmonized standards for the validation and approval of paediatric AI devices. Fast-track pathways for AI tools that address pressing global health needs, combined with rigorous post-market surveillance, could accelerate deployment without compromising safety. Ethical frameworks must address data ownership, algorithmic fairness, transparency, and accountability, ensuring that AI benefits all children equitably, regardless of geography or socioeconomic status.

VII. CONCLUSION

Artificial intelligence stands at the threshold of transforming paediatric TB diagnosis. The trajectory from simple CAD algorithms that detected cavities with 82% accuracy to sophisticated deep learning ensembles that rival or exceed human performance in some settings is remarkable. AI offers the promise of overcoming the chronic shortage of paediatric radiology expertise, standardizing CXR interpretation, enabling high-throughput screening, and detecting disease at its earliest, most treatable stages. The WHO's endorsement of CAD for adult TB screening validates this vision, and the next frontier is unequivocally paediatric.

Realizing this potential, however, demands recognition of the unique biological, epidemiological, and technical challenges of childhood TB. The development of large, diverse paediatric datasets with robust microbiological reference standards is the most critical unmet need. Algorithms must be purpose-built for children, not simply adapted from adults. Prospective clinical and implementation trials are needed to move from proof-of-concept to proof-of-impact. Regulatory and ethical frameworks must evolve in tandem with technological advances to ensure safety, fairness, and trust.

The synergy of human expertise and AI, as demonstrated by the ensemble models where radiologists and algorithms together achieved near-perfect accuracy, points toward a future where AI is not a replacement for clinicians but a powerful extension of their capabilities. In the global fight to end TB, ensuring that children are not left behind is both a moral imperative and an epidemiological necessity. With sustained investment, interdisciplinary collaboration, and a commitment to equity, AI-driven chest radiograph analysis can become a cornerstone of paediatric TB care, saving countless young lives.

REFERENCES

[1] Acanfora, G., Carillo, A. M., Dello Iacovo, F., Salatiello, M., Pisapia, P., Bellevicine, C., ... & Vigliar, E. (2024). Interobserver variability in cytopathology: How much do we agree? *Cytopathology*.

[2] Ahmad, H. K., Milne, M. R., Buchlak, Q. D., Ektas, N., Sanderson, G., Chamtie, H., ... & Jones, C. (2023). Machine learning augmented interpretation of chest X-rays: A systematic review. *Diagnostics*, 13(4), 743.

[3] Alkatheiri, M. S. (2022). Artificial intelligence assisted improved human-computer interactions for computer systems. *Computers and Electrical Engineering*, 101, 107950.

[4] Allwood, B. W., Byrne, A., Meghji, J., Rachow, A., van der Zalm, M. M., & Schoch, O. D. (2021). Post-tuberculosis lung disease: Clinical review of an under-recognised global challenge. *Respiration*, 100(8), 751–763.

[5] Alshammeri, A. F., Alhamaid, Y. A., Alshakhs, A. M., Bohulaigah, Z. H., Eissa, G. A., Almutairi, M. S., ... & Algafly, H. A. (2024). X-ray interpretation in emergency department in the Kingdom of Saudi Arabia: Do we need the radiologist? *Emergency Radiology*, 1–10.

[6] Andom, A. T., Gilbert, H. N., Ndayizigiye, M., Mukherjee, J. S., Nthunya, J., Marole, T. A., ... & Yuen, C. M. (2022). Understanding reasons for suboptimal tuberculosis screening in a low-resource setting: A mixed-methods study in the Kingdom of Lesotho. *PLOS Global Public Health*, 2(3), e0000249.

[7] Boddu, P., Parimi, V., Taddonio, M., Kane, J. R., & Yeldandi, A. (2017). Pathologic and radiologic correlation of adult cystic lung disease: A comprehensive review. *Pathology Research International*, 2017.

[8] Chaparro, C. M., & Suchdev, P. S. (2019). Anemia epidemiology, pathophysiology, and

etiology in low-and middle-income countries. *Annals of the New York Academy of Sciences*, 1450(1), 15–31.

[9] Chassagnon, G., Vakalopoulou, M., Paragios, N., & Revel, M. P. (2020). Artificial intelligence applications for thoracic imaging. *European Journal of Radiology*, 123, 108774.

[10] Choi, S. Y., Park, S., Kim, M., Park, J., Choi, Y. R., & Jin, K. N. (2021). Evaluation of a deep learning-based computer-aided detection algorithm on chest radiographs: Case control study. *Medicine*, 100(16), e25663.

[11] Ciet, P., Eade, C., Ho, M. L., Laborie, L. B., Mahomed, N., Naidoo, J., ... & Shelmerdine, S. C. (2023). The unintended consequences of artificial intelligence in paediatric radiology. *Pediatric Radiology*, 1–9.

[12] Drozdov, I., Szubert, B., Reda, E., Makary, P., Forbes, D., Chang, S. L., ... & Lowe, D. J. (2021). Development and prospective validation of COVID-19 chest X-ray screening model for patients attending emergency departments. *Scientific Reports*, 11(1), 20384.

[13] Guan, H., & Liu, M. (2023). DomainATM: Domain adaptation toolbox for medical data analysis. *NeuroImage*, 268, 119863.

[14] Hantous-Zannad, S., Néji, H., Affes, M., Attia, M., Baccouche, I., Kechaou, S., ... & Miled-M'rad, K. B. (2022). *Imaging of thoracic tuberculosis. In Imaging of Tuberculosis*. Springer (pp. 185–223).

[15] Harris, M., Qi, A., Jeagal, L., Torabi, N., Menzies, D., Korobitsyn, A., ... & Ahmad Khan, F. (2019). A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. *PLoS ONE*, 14(9), e0221339.

[16] Hua, D., Nguyen, K., Young, N., Cho, J. G., Yap, A., & others. (2023). Benchmarking the diagnostic test accuracy of certified AI products for screening pulmonary tuberculosis in digital chest radiographs: Preliminary evidence from a rapid review and meta-analysis. *International Journal of Medical Informatics*, 105159.

[17] Hwang, E. J., Jeong, W. G., David, P. M., Arentz, M., Ruhwald, M., & Yoon, S. H. (2024). AI for detection of tuberculosis: Implications for global health. *Radiology: Artificial Intelligence*, 6(2), e230327.

[18] Hwang, E. J., Nam, J. G., Lim, W. H., Park, S. J., Jeong, Y. S., Kang, J. H., ... & Park, C. M. (2019). Deep learning for chest radiograph diagnosis in the emergency department. *Radiology*, 293(3), 573–580.

[19] Hwang, S., Kim, K. Y., Kang, S. K., Seo, S., Paeng, J. C., Lee, D. S., & Lee, J. S. (2018). Improving the accuracy of simultaneously reconstructed activity and attenuation maps using deep learning. *Journal of Nuclear Medicine*, 59(10), 1624–1629.

[20] Jaeger, S., Juarez-Espinosa, O. H., Candemir, S., Poostchi, M., Yang, F., Kim, L., ... & Thoma, G. (2018). Detecting drug-resistant tuberculosis in chest radiographs. *International Journal of Computer Assisted Radiology and Surgery*, 13, 1915–1925.

[21] Kotei, E., & Thirunavukarasu, R. (2024). A comprehensive review on advancement in deep learning techniques for automatic detection of tuberculosis from chest X-ray images. *Archives of Computational Methods in Engineering*, 31(1), 455–474.

[22] Kulkarni, S., & Jha, S. (2020). Artificial intelligence, radiology, and tuberculosis: A review. *Academic Radiology*, 27(1), 71–75.

[23] Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2), 574–582.

[24] Ma, W., Liu, Z., Kudyshev, Z. A., Boltasseva, A., Cai, W., & Liu, Y. (2021). Deep learning for the design of photonic structures. *Nature Photonics*, 15(2), 77–90.

[25] Mahomed, N., Kilborn, T., Smit, E. J., Chu, W. C. W., Young, C. Y. M., Koranteng, N., ... & Sodhi, K. S. (2023). Tuberculosis revisted:

classic imaging findings in childhood. *Pediatric Radiology*, 53(9), 1799–1828.

[26] Melendez, J., Sánchez, C. I., Philipsen, R. H., Maduskar, P., Dawson, R., Theron, G., ... & Van Ginneken, B. (2016). An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Scientific Reports*, 6(1), 25265.

[27] Mouton, J., & Vaezipour, N. (2022). Towards accurate point-of-care tests for tuberculosis in children. *Pathogens*, 11, 327.

[28] Nabulsi, Z., Sellergren, A., Jamshy, S., Lau, C., Santos, E., Kiraly, A. P., ... & Shetty, S. (2021). Deep learning for distinguishing normal versus abnormal chest radiographs and generalization to two unseen diseases tuberculosis and COVID-19. *Scientific Reports*, 11(1), 15523.

[29] Naidoo, J., Shelmerdine, S. C., Charcape, C. F. U., & Sodhi, A. S. (2023). Artificial intelligence in paediatric tuberculosis. *Pediatric Radiology*, 53(9), 1733–1745.

[30] Nassehi, A., Zhong, R. Y., Li, X., & Epureanu, B. I. (2022). *Review of machine learning technologies and artificial intelligence in modern manufacturing systems*. In Design and Operation of Production Networks for Mass Personalization in the Era of Cloud Technology (pp. 317–348). Elsevier.

[31] Nischal, N., Nath, R., Rathi, V., & Ish, P. (2023). Diagnosing and treating extrapulmonary tuberculosis in India: Challenges and solutions. *Preventive Medicine: Research and Reviews*, 10–4103.

[32] Ortiz, M. I. G., de Melo Alencar, C., De Paula, B. L. F., Magno, M. B., Maia, L. C., & Silva, C. M. (2020). Accuracy of near-infrared light transillumination (NILT) compared to bitewing radiograph for detection of interproximal caries in the permanent dentition: A systematic review and meta-analysis. *Journal of Dentistry*, 98, 103351.

[33] Pande, T., Cohen, C., Pai, M., & Ahmad Khan, F. (2016). Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. *The International Journal of Tuberculosis and Lung Disease*, 20(9), 1226–1230.

[34] Pasa, L., Navarin, N., Erb, W., & Sperduti, A. (2023). Empowering simple graph convolutional networks. IEEE Transactions on Neural Networks and Learning Systems.

[35] Schalekamp, S., Klein, W. M., & van Leeuwen, K. G. (2022). Current and emerging artificial intelligence applications in chest imaging: a pediatric perspective. *Pediatric Radiology*, 52(11), 2120–2130.

[36] Singh, M., Pujar, G. V., Kumar, S. A., Bhagyalalitha, M., Akshatha, H. S., Abuhaija, B., ... & Gandomi, A. H. (2022). Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications. *Electronics*, 11(17), 2634.

[37] Suchitra, S., Ibrahim, S. J. A., Sathya, M., Sahini, V., Chakravarthy, K., Surya, N., & Kumar, V. (2023). Recent advances in analysis and detection of tuberculosis system in chest X-ray using artificial intelligence (AI) techniques: A review. *Current Materials Science*, 16(1), 43–51.

[38] Taka, M., Kobayashi, S., Mizutomi, K., Inoue, D., Takamatsu, S., Gabata, T., ... & Abo, H. (2023). Diagnostic approach for mediastinal masses by imaging and histological analysis. *European Journal of Radiology*, 110767.

[39] Van't Hoog, A., Viney, K., Biermann, O., Yang, B., Leeflang, M. M., & Langendam, M. W. (2022). Symptom- and chest-radiography screening for active pulmonary tuberculosis in HIV-negative adults and adults with unknown HIV status. *Cochrane Database of Systematic Reviews*, (3).

[40] Xu, T., Cheng, I., Long, R., & Mandal, M. (2013). Novel coarse-to-fine dual scale technique for tuberculosis cavity detection in chest radiographs. *EURASIP Journal on Image and Video Processing*, 2013, 1–18.