

Review

Not peer-reviewed version

Artificial Intelligence for Medical Image Analysis in Low-Resource Settings: A Systematic Review

[Tanya Akumu](#)^{*,†}, Victor M. Campello[†], Lidia Garrucho, Vien Ngoc Dang, Thomas Keller, Xue Zhao, Emmah Ng'ang'a, María J. Ledesma-Carbayo, Marius G. Linguraru, Karim Lekadir

Posted Date: 10 October 2026

doi: 10.20944/preprints202610.0843.v1

Keywords: artificial intelligence; low-resource settings; medical imaging; global health



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Artificial Intelligence for Medical Image Analysis in Low-Resource Settings: A Systematic Review

Tanya Akumu ^{1,*†}, Victor M. Campello ^{1,2,†}, Lidia Garrucho ¹, Vien Ngoc Dang ¹, Thomas Keller ¹, Xue Zhao ¹, Emmah Ng'ang'a ^{3,4,5}, María J. Ledesma-Carbayo ^{6,7}, Marius G. Linguraru ^{8,9} and Karim Lekadir ^{1,10}

- 1 Artificial Intelligence in Medicine Lab (BCN-AIM), Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Spain
- 2 Facultad de Educación, Economía y Tecnología de Ceuta, Universidad de Granada, Ceuta, Spain
- 3 ISGlobal, Barcelona, Spain
- 4 Facultat de Medicina i Ciències de la Salut, Universitat de Barcelona, Barcelona, Spain
- 5 Centre of Excellence in Women and Child Health, Aga Khan University, Nairobi, Kenya
- 6 Biomedical Image Technologies, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain
- 7 Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Instituto de Salud Carlos III, Madrid, Spain
- 8 Children's National Hospital, Washington, DC, USA
- 9 George Washington University, Washington, DC, USA
- 10 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
- * Correspondence: tanya.akumu@ub.edu
- † Tanya Akumu and Victor M. Campello contributed equally to this work.

Abstract

Artificial intelligence (AI) for medical image analysis is increasingly being developed for low-resource settings, where specialist expertise, imaging equipment, labelled data, and computing resources may be limited. We systematically reviewed 189 studies published between January 2021 and January 2026 to examine how these constraints relate to imaging modalities, clinical applications, and AI methods in low-resource settings. Each study was also assessed against 15 criteria adapted from the FUTURE-AI principles for trustworthy AI. Shortages of clinical expertise were the most common constraint, and portable or lower-cost imaging was widely used. Transfer learning was the most frequent technical strategy (59%), followed by data augmentation (54%) and efficient model design (36%). Medical-image pretraining and self-supervised learning were each used in 4% of studies, and model compression in 3%. Applications were concentrated in retinal disease, pulmonary disease, antenatal care, parasitic diseases, and cervical screening. Beyond these technical and clinical patterns, the trustworthy AI assessment showed a clear gap between documenting how systems were developed and demonstrating how they performed under relevant local conditions. Subgroup performance was evaluated in 24% of studies, intended users or realistic working conditions in 25%, and model explanations with intended users in 4%. Consistent with this gap, among 15 studies that reported the same model and metric during development and in the local setting, 12 performed worse locally. Future research should prioritise representative local evaluation, intended-user assessment, demonstrated resource suitability, and reproducible research practices.

Keywords: artificial intelligence; low-resource settings; medical imaging; global health

1. Introduction

Health inequities remain a pressing global problem. In low-resource settings (LRS), access to medical imaging is constrained by shortages of specialist clinicians, limited imaging and computational infrastructure, and access to high-quality annotated data [1,2]. Artificial intelligence (AI) could help address these constraints by providing diagnostic and screening capacity where specialist review is

limited [3,4]. Realising this potential, however, requires evidence that these systems remain effective in the populations, devices, and care settings for which they are intended.

Medical imaging is one area in which AI has been widely studied, with applications spanning image acquisition, segmentation, registration, annotation, and disease classification [5–7]. In global health research, however, it is important to distinguish between two related but different contexts. Low-resource settings describe constraints at the point of care, whereas low- and middle-income countries (LMICs) describe country-level economic status [8,9]. We use LRS to describe the care setting and LMIC to describe the country context.

Several characteristics of LRS can limit the development and translation of medical imaging AI. Training data often originate from high-resource institutions and may not represent the populations, disease profiles, or imaging devices encountered in LMICs [1,3]. Limited computing capacity, unreliable power, and restricted connectivity can constrain the deployment of imaging and AI systems [10,11]. Shortages of specialist clinicians can shift image acquisition and interpretation towards non-specialist health workers and portable modalities such as ultrasound and radiography [11,12]. Disease priorities also differ, with neglected tropical diseases, high-burden infectious diseases, and maternal and neonatal conditions carrying substantial burden in LMICs while receiving comparatively limited attention in AI research.

Existing reviews have examined individual diseases [13–15], regions [16], imaging modalities [17], and barriers to implementing AI in resource-constrained healthcare settings [18]. Recent work has expanded this evidence base to include AI implementation, deployment, and diagnostic imaging applications in LMICs [19,20]. Our review complements these studies by linking how medical imaging AI systems are developed and applied with the evidence reported for trustworthy AI in the settings they target. We examine this across imaging modalities and clinical applications using a structured assessment adapted from the FUTURE-AI consensus guideline [21], covering fairness, universality, traceability, usability, robustness, and explainability.

Specifically, we ask which constraints shape AI development and deployment in LRS (Q1); which methodologies are used to address these constraints (Q2); in which clinical applications and imaging modalities these approaches have been demonstrated (Q3); and to what extent studies report evidence relevant to trustworthy AI in the settings they target (Q4). Our aim is to characterise the reported evidence for trustworthy AI in LRS.

2. Identified Challenges in LRS

AI for medical imaging in low-resource settings must operate under constraints that affect data collection, model development, clinical use, and long-term sustainability. We organise these challenges into seven areas relevant to the studies in this review.

Infrastructure:

Limited access to imaging equipment remains a major constraint. LMICs have markedly fewer CT and MRI systems per capita than high-income countries [12], while unreliable power, limited computing capacity, and poor connectivity can restrict both imaging services and AI-supported workflows [11,19,22]. These conditions favour portable and lower-cost ultrasound and smartphone-based microscopy [13,23]. Health-system constraints beyond the clinic also affect whether AI-supported screening leads to clinical benefit, including referral, follow-up, and access to treatment [8,24].

Human resources:

Shortages of specialist clinicians limit both image acquisition and interpretation. Radiologist availability is much lower in low- and middle-income countries than in high-income countries [12]. These shortages have encouraged task shifting, in which selected imaging or clinical tasks are performed by trained non-specialist health workers. Portable ultrasound programmes, for example, rely on non-radiologist operators in resource-constrained settings [3,25]. AI systems intended for these settings must therefore account for user training, workload, clinical responsibility, and workflow inte-

gration [19,22]. Limited local technical expertise can further constrain model adaptation, monitoring, and maintenance.

Data and research representation:

Locally representative imaging data remain difficult to collect and share. Available datasets may be small, fragmented across institutions, or poorly representative of the populations and devices in which AI systems are ultimately evaluated [18,19]. Models trained predominantly on high-resource datasets may therefore perform differently when applied to local populations or acquisition conditions. Limited local data also restrict model adaptation and subgroup evaluation.

Geographic and demographic heterogeneity:

Patient populations, disease prevalence, imaging devices, operator experience, and care pathways can vary considerably between settings. Performance observed in one institution or population therefore cannot be assumed to generalise to another without appropriate evaluation. This is particularly relevant when models move from high-income to low-resource settings, between urban and rural services, or between acquisition devices [8,19].

Social and cultural context:

The use of AI in healthcare is shaped by the people and organisations that interact with it. Language, trust, clinical roles, user experience, and local workflows can influence whether AI-supported recommendations are accepted and used [18]. Evaluation should therefore extend beyond technical performance to consider intended users and the context in which decisions are made. Human-centred approaches similarly emphasise that successful adoption and safe use depend on how the technology interacts with its users and the wider care system [24,26].

Financial sustainability:

The long-term use of AI systems depends on more than initial development costs. Hardware, software, connectivity, integration, maintenance, training, and continuing technical support all contribute to sustainability [3]. Technologies that depend on short-term external funding or cannot be integrated into existing services may be difficult to sustain after research projects end. More broadly, long-term sustainability depends on whether a technology can be supported within local organisational and service capacity [26].

Regulation and governance:

Regulatory and governance frameworks for medical AI remain uneven across LMICs. Requirements for data governance, privacy, consent, accountability, and medical-device regulation differ between jurisdictions and may still be developing [27]. These differences affect how data can be shared, how AI systems can be evaluated and updated, and how responsibility is assigned when they are introduced into clinical practice.

3. Survey Methodology

We conducted this review following established recommendations for systematic and scoping searches [28,29]. We sought peer-reviewed studies reporting AI methods for medical image analysis in low-resource settings (LRS), including studies framed around low- and middle-income countries (LMICs), global health, and neglected tropical diseases.

Scope and definition of LRS:

We defined an LRS as a care setting in which at least one of four constraints applied: use of portable or lower-cost imaging in place of a reference modality; limited access to specialist interpretation; insufficient computing capacity, power, or connectivity for conventional AI workflows; or scarcity of locally representative annotated data. We distinguish LRS from LMIC, which is a country-level

economic classification based on gross national income [9]. We use LMIC for country-level statements and LRS for the operational conditions described within individual studies.

Search strategy:

We searched PubMed, Scopus, and IEEE Xplore using terms covering three concepts: low-resource settings, artificial intelligence, and medical imaging. The low-resource terms included LMIC, global-health, and neglected-tropical-disease framings. The specific search strings for each database are provided in Supplementary Table S1. The search ran on 20 January 2026, with Scopus and IEEE Xplore limited to conference proceedings, and returned 2,852 records from PubMed, 1,096 from Scopus, and 1,220 from IEEE Xplore resulting in a total of 5,168 records.

Eligibility criteria:

We included peer-reviewed original research published in English in first- or second-quartile journals according to the SCImago Journal Rank [30], or in conferences ranked C or higher in the CORE 2023 ranking [31]. Studies were required to report an AI methodology applied to medical imaging data from human subjects and to have an LRS or LMIC context as a central aim. We excluded reviews, opinion pieces, editorials, book chapters, and book sections, as well as studies without accessible full text or with insufficient methodological or results reporting for assessment.

Screening and study selection:

Study selection followed the PRISMA 2020 statement [32] and is summarised in Figure 1. Two reviewers (T.A., V.D.) independently assessed a stratified random sample of 250 records against the eligibility criteria. This reference set was used to evaluate Llama 3.1 8B-Instruct [33] and Gemma 2 9B-Instruct [34] using zero-shot and calibrated few-shot prompting for LLM-assisted screening. Calibrated Gemma 2 was selected because it retained all reviewer-included records in the reference set. The validation protocol and performance metrics are reported in Supplementary Section S6.

Of the 2,636 records screened at title and abstract level, 250 belonged to the reviewer-annotated reference set and 2,386 were screened by the selected model. A total of 1,181 articles proceeded to full-text review, which was divided across four reviewers. After full-text assessment, 189 studies met all eligibility criteria.

3.1. Data Analysis

The selected studies were analysed in two stages: a descriptive synthesis and an assessment of reported trustworthy AI evidence using the FUTURE-AI framework [21].

Descriptive synthesis:

For each study, we analysed the imaging modality, AI methodology, model architecture and reported parameter count, clinical application, dataset origin and size, study context, and resource constraints addressed. Studies were grouped by AI technique, imaging modality, clinical application, and resource constraint. Aggregate analyses included all 189 studies. Study-level characteristics are provided in Supplementary Section S8.

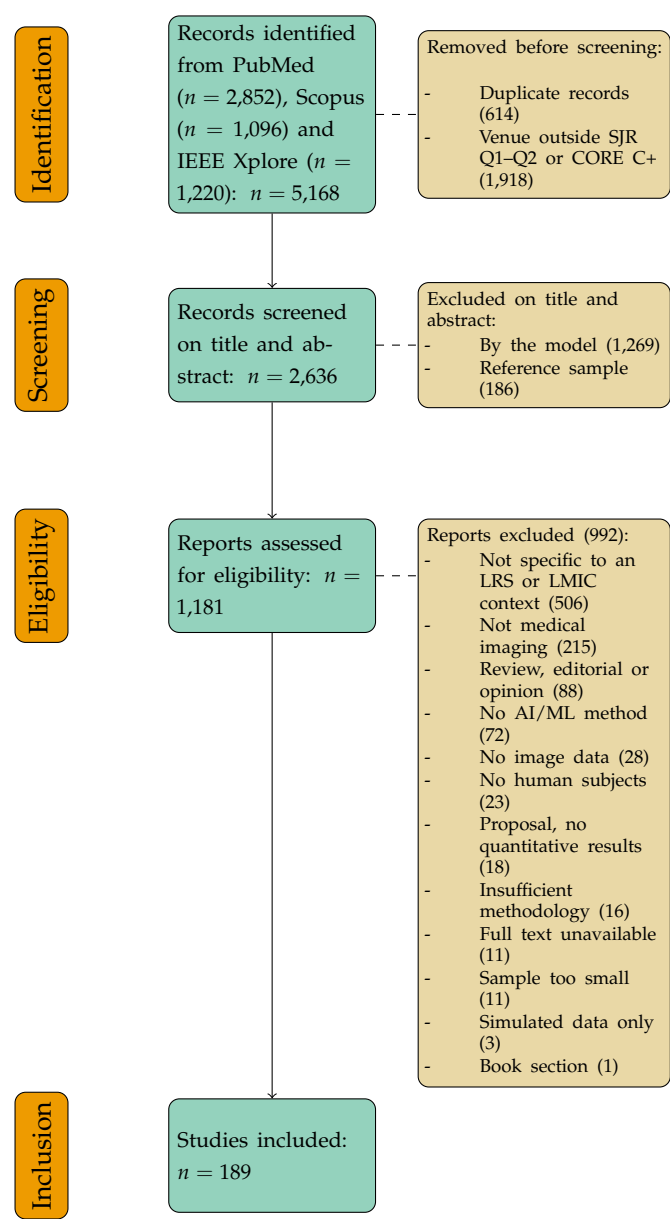


Figure 1. PRISMA flowchart for the study selection. Records identified are those dated within the review window, January 2021 to January 2026. Records at venues outside the venue criterion were removed before title and abstract screening.

Trustworthy AI assessment:

The second stage assessed the evidence each study reported for trustworthy AI in low-resource settings. We scored all 189 studies against 15 binary criteria adapted from the six FUTURE-AI principles: Fairness, Universality, Traceability, Usability, Robustness, and Explainability [21]. Table 1 defines each criterion. Each was scored as reported (1) or not reported (0), based on the published record. Two reviewers independently scored each study from the publication and supplementary material, with disagreements resolved by consensus. Further scoring details and reliability analyses are provided in Supplementary Section S3, and the released scoring sheet includes the final scores and supporting evidence.

Table 1. Trustworthy AI assessment criteria. Fifteen criteria adapted from the six FUTURE-AI principles for trustworthy AI assessment of published AI studies in low-resource settings. Each criterion was scored as reported (1) or not reported (0). Absence of evidence in the publication was scored as not reported.

#	Item	Criteria
<i>Fairness</i>		
1	Individual data characteristics	Characteristics of the individuals in the study population are reported, such as age, sex, ethnicity, or clinical characteristics.
2	Subgroup evaluation	Performance is reported separately for at least one relevant patient, population, or characteristic subgroup.
<i>Universality</i>		
3	Resource-appropriate design	The system is adapted to resource constraints of the intended setting, such as limited compute, hardware, connectivity, or cost.
4	Local involvement	People or institutions from the intended setting participate in the study design, data collection, system development, or evaluation. Affiliation alone does not qualify.
5	Local adaptation	The model is recalibrated, fine-tuned, retrained, or otherwise adapted using local data, with performance evaluated in the local setting. Fine-tuning a pretrained model on local data counts.
<i>Traceability</i>		
6	Method documentation	The model architecture, training procedure, and key training settings are reported in sufficient detail to understand how the model was developed.
7	Data provenance	The data source and acquisition process are reported, together with ethics approval and consent information or a documented waiver.
8	Open resources	A working public link is provided to the code, trained model, or dataset. Availability on request does not qualify.
<i>Usability</i>		
9	Intended user	The people expected to use the system are identified, such as community health workers, nurses, clinicians, sonographers, or radiologists.
10	Intended use	The clinical task and intended setting of use are described, including relevant information on the care setting, equipment, or user role.
11	Usability evaluation	The system is evaluated with intended users or under realistic conditions of use, beyond evaluation on a fixed retrospective dataset.
<i>Robustness</i>		
12	Representative training data	The training data include data from the target low-resource population or setting rather than exclusively from higher-resource settings.
13	Robustness evaluation	Performance is evaluated on independent data or under relevant variation in site, population, device, operator, acquisition, or image quality.
<i>Explainability</i>		
14	Explanation provided	An explainability or interpretability method is applied to the model and examples of its output are reported.
15	Explanation evaluation	The study evaluates whether intended users find the explanation useful or understandable.

For Universality, we additionally compared reported development and target-setting performance when the same model and metric were available on independent data from the low-resource setting addressed by the study. Comparisons across different models, configurations, or external reference values were excluded. Performance changes were calculated relative to each study’s development value.

4. Identified Technical Solutions

The 189 included studies used several technical strategies to address constraints in data availability, computation, and imaging hardware. Figure 2 summarises the main approaches, including efficient model design, transfer learning, data augmentation, self-supervised learning, and model compression. Study-level details are provided in Supplementary Section S8.

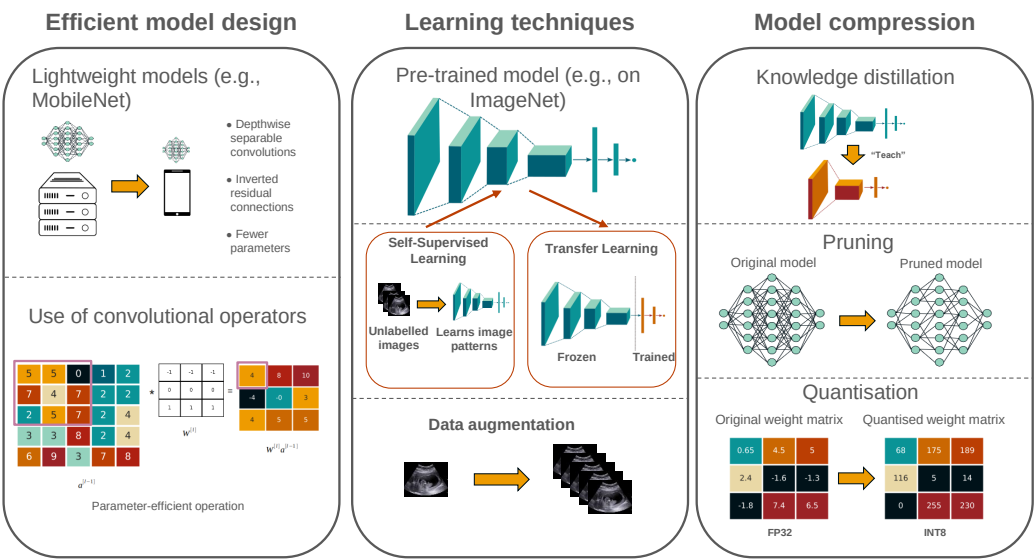


Figure 2. Technical approaches considered for addressing constraints in low-resource settings. The figure summarises efficient model design, learning techniques, and model compression approaches examined in the review. Knowledge distillation is included within model compression but was not reported in any of the 189 included studies.

4.1. Efficient Model Design

Efficient model design aims to reduce computational requirements by using architectures with fewer parameters or lower-cost operations. It was reported in 68 of 189 studies (36%), making it the third most common technical strategy after transfer learning and data augmentation. Common architecture families included MobileNet [35–37], ShuffleNet [38], EfficientNet [39,40], and lightweight attention-based models such as MobileViT [41]. MobileNet variants were used most often, particularly in smartphone-based optical imaging and digital microscopy. Song et al. [42] replaced standard U-Net convolutions with depthwise separable convolutions for oral cancer segmentation, reporting a model about six times smaller and twice as fast.

Some studies also measured performance on target hardware. Schilpzand et al. [43] reported placental segmentation in 19 ms per image on a smartphone paired with a low-cost ultrasound probe, and Gomes et al. [44] benchmarked a MobileNetV2 gestational-age model on four Android phones. Other studies used much larger models without reporting on-device performance, including an 80-million-parameter ViT-B/16 model [45] and a 76.65-million-parameter EfficientNetB7 model [46].

4.2. Learning Techniques

Learning techniques can reduce the need for labelled data or training from scratch. The main approaches reported in the included studies were data augmentation, transfer learning, and self-supervised learning.

Data augmentation:

Data augmentation increases training diversity without collecting new images, typically by modifying existing images or generating synthetic examples. It was reported in 103 of 189 studies (54%). Most studies used transformations such as rotation, flipping, cropping, colour changes, or intensity variation. Only three studies generated new images with generative models. Wang et al. [47] used a latent diffusion model to generate fetal ultrasound images and reported improved zero-shot classification on African cohorts, while Shuai et al. [48] combined real and diffusion-generated fundus photographs and reported improved F1 score on low-quality images. A smaller group used augmentation to reproduce imaging conditions expected in the target setting. Lin et al. [49] simulated low-field MRI from high-field scans, and Wu et al. [50] simulated noise associated with portable OCT.

Other studies modified existing images or transferred image regions containing cells or rare findings between images [51,52].

Transfer learning:

Transfer learning reuses a model or representation learned on another dataset and adapts it to a new task, reducing the need to train a model from scratch. It was the most frequently reported learning technique, used in 112 of 189 studies (59%). Most studies used ImageNet-pretrained general-purpose vision models [53], particularly ResNet-50 (47 studies), Inception-V3 (26), VGG-16 (24), and DenseNet-121 (16). For example, Sendra-Balcells et al. [54] used transfer learning for fetal-plane classification across five African countries with as few as two training patients from each target centre, while Capellán-Martín et al. [55] compared medical-image pretraining and training from scratch for paediatric tuberculosis detection. Medical-image pretraining was reported in 8 of 189 studies (4%). These included histopathology foundation models, public chest radiograph and fundus datasets, and fetal ultrasound data from another country [54–58].

Self-supervised learning:

Self-supervised learning uses unlabelled images to learn representations that can later be adapted to a labelled task. It was reported in 8 of 189 studies (4%). Applications included cervical cytology, histopathology, parasite detection, cardiovascular assessment, and chest radiography. Stegmüller et al. [51] used DINO pretraining on unlabelled Pap smear slides for cervical cytology, while Akhter et al. [59] used a self-supervised jigsaw task for silicosis detection on chest radiographs.

4.3. Model Compression

Model compression reduces model size or computational requirements through techniques such as knowledge distillation, pruning, and quantisation. Knowledge distillation trains a compact student model to reproduce the outputs of a larger teacher [60]. Six of the 189 studies (3%) used at least one compression technique. None used knowledge distillation. Pruning was used in three studies and quantisation in four, with one study applying both. Among the 68 studies using efficient model design, six also used pruning or quantisation.

Pruning:

Pruning removes model parameters judged to contribute least to performance [61]. It was reported in 3 of 189 studies, all of which also used efficient model design and transfer learning. Montalbo [62] reduced a pretrained network from 55 million to 441 thousand parameters, with accuracy decreasing from 98.67% to 97.41%. Bhoopalan et al. [63] combined pruning and quantisation in a vision transformer and reported 120 ms latency on a Raspberry Pi 4.

Weight quantisation:

Quantisation represents model weights at lower numerical precision to reduce memory and computational requirements [64]. It was reported in 4 of 189 studies (2%), including smartphone and edge-device applications [63,65–67]. Memon et al. [65] benchmarked its classifier across several platforms and reported 0.11 s inference on an Android phone, with a small loss of accuracy.

5. Clinical Applications

The 189 included studies covered a broad range of imaging modalities, clinical applications, and resource constraints. Most used data from low- or middle-income countries, with 177 of 189 studies (94%) including data from at least one LMIC. By design, 56 studies collected their data prospectively for the study, and one of them was a randomised trial [68]. We organise the findings by imaging modality, resource constraint, and clinical application. Figure 3 links these three dimensions, and study-level characteristics are provided in Supplementary Section S8.

5.1. By Modality

Five modalities accounted for 168 of 189 studies (89%): digital microscopy (53), optical imaging (34), ultrasound (33), X-ray (25), and fundus photography (23). The remaining studies used CT (5), MRI (7), optical coherence tomography (5), or more than one modality (4). Figure 4 shows publication trends across the leading modalities.

Several studies used portable or lower-cost modalities in place of more conventional imaging approaches. Examples included ultrasound for breast assessment [69], handheld cervical imaging [70], lung ultrasound [71], and low-field MRI [72].

5.2. By Resource Constraint

The most frequently reported constraint was limited clinical expertise, identified in 152 of 189 studies (80%). Several studies aimed to support non-specialist users, including fetal ultrasound systems designed for midwives or other frontline operators [44,73,74].

Low-cost or portable imaging was reported in 91 studies (48%). These studies addressed barriers such as limited access to fixed imaging equipment, specialist-dependent acquisition, or image quality from lower-cost devices. Some focused on improving the usability of constrained hardware, including super-resolution for ultrasound [75].

Connectivity or infrastructure limitations were reported in 79 studies (42%), limited sample size in 69 (37%), suboptimal modality or image quality in 61 (32%), and computational resource limitations in 52 (28%). Sixteen studies were framed around neglected tropical diseases, most commonly parasitology applications. Figure 3 links imaging modalities and clinical applications to these reported constraints. Because studies could address more than one constraint, the categories are not mutually exclusive.

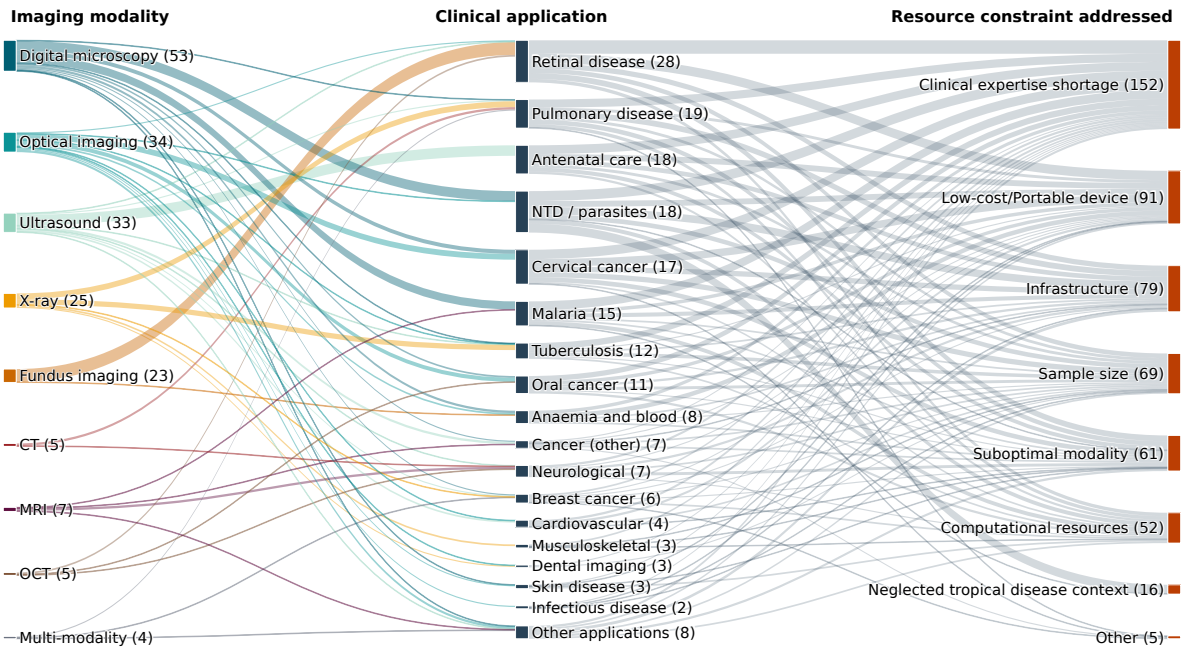


Figure 3. Imaging modality, clinical application, and reported resource constraint across the 189 included studies. Node labels show the number of studies in each category, and flow width represents the number of studies sharing each combination. Studies could report more than one resource constraint, so counts in the right-hand column exceed 189. Applications outside the named groups are pooled as “Other applications”.

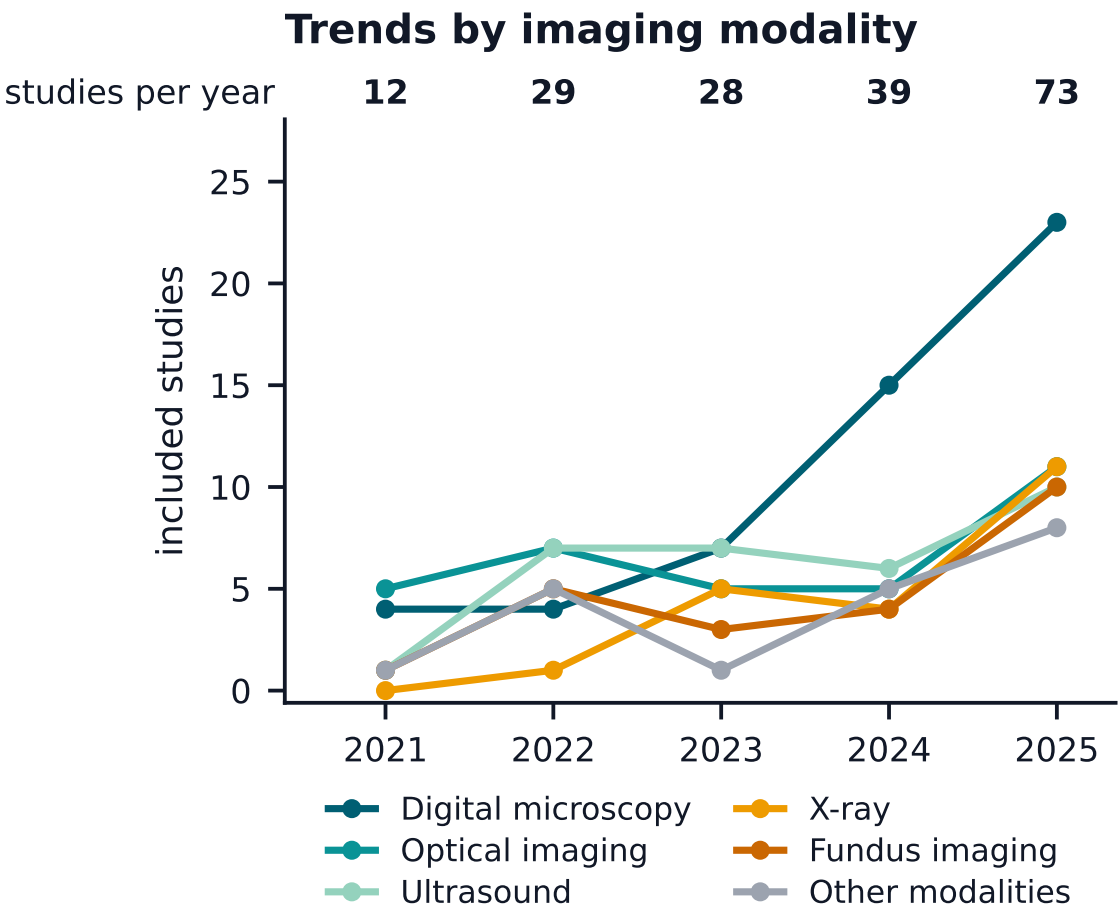


Figure 4. Trends by imaging modality. Annual numbers of included studies are shown for the five most frequently reported imaging modalities, with all remaining modalities pooled as “Other modalities”. Total studies per year are shown above the panel. The 2026 count covers only the three weeks up to the search date of 20 January 2026 and is therefore excluded from the trend plot.

5.3. By Medical Domain

Retinal disease was the largest application group (28 studies), followed by pulmonary disease other than tuberculosis (19), antenatal care (18), parasitic and other neglected tropical diseases (18), cervical cancer (17), malaria (15), and tuberculosis (12) (Figure 3). Children were the imaged population in 19 studies (10%), and a further 19 studies involved fetal imaging.

Antenatal care.

All 18 antenatal studies used ultrasound for tasks including gestational age estimation, fetal biometry, placenta localisation, fetal-plane assessment, and detection of congenital abnormalities. Examples included mobile gestational-age estimation from blind sweeps in the USA and Zambia [44], fetal-plane classification across five African countries [54], and prospective evaluation of an integrated portable ultrasound system in Zambia and the USA [76].

Cancer screening.

Cancer applications included cervical, oral, and breast screening. Of the 17 cervical cancer studies, 10 analysed cervical images and 7 microscopy slides. Smartphone or handheld imaging was used for automated visual assessment [70,77,78]. Smartphone-based oral cancer screening was evaluated in India [79], while portable ultrasound was used for breast assessment in Mexico [69].

Pulmonary disease.

Pulmonary applications focused mainly on tuberculosis and COVID-19. Nine of the 12 tuberculosis studies used chest radiographs, including prospective evaluations in Ethiopia and India [80,81]. Ultrasound was also evaluated for tuberculosis [71]. Nine studies addressed COVID-19, while smaller groups covered lung cancer, silicosis, and other pulmonary conditions.

Ophthalmology.

Retinal studies addressed diabetic eye disease, retinopathy of prematurity, glaucoma, and retinoblastoma, with 22 of 28 using fundus photographs. External validation of retinopathy-of-prematurity screening was reported in India, Nepal, and Mongolia [82,83]. Diabetic-retinopathy studies included feasibility and prospective evaluations in Brazil and India [84,85]. Fundus photographs were also used to estimate cardiovascular risk factors [86].

Infectious and neglected tropical diseases.

Fourteen of the 15 malaria studies used digital microscopy, including parasite detection in thick blood smears [87]. Low-field MRI was also used to assess cerebral malaria [72]. Eighteen studies addressed other parasitic and neglected tropical diseases, mainly schistosomiasis and soil-transmitted helminths. Most used microscopy, including edge-device detection of schistosomiasis eggs in Nigeria [67].

Other applications.

Smaller application groups included anaemia and other blood disorders, neurological disease, cardiovascular disease, and skin, dental, and musculoskeletal imaging. Examples included bone mineral density estimation from X-ray images in Vietnam [88] and anaemia detection from smartphone images of the conjunctiva and sclera [89].

6. Trustworthy AI Evidence

We assessed the 189 studies against 15 criteria adapted from the FUTURE-AI principles [21] for low-resource settings (Table 1). Across the eight criteria that describe a study's system, data, and intended use, evidence was reported in an average of 64% of studies. The corresponding average was 38% across the seven criteria requiring evidence beyond that description, such as evaluation, local participation or adaptation, and public release of research materials. Figure 5 summarises reporting across the six principles. Per-criterion results, reviewer agreement, and per-modality results are given in Supplementary Sections S2 to S4, and the per-study scores in the released scoring sheet.

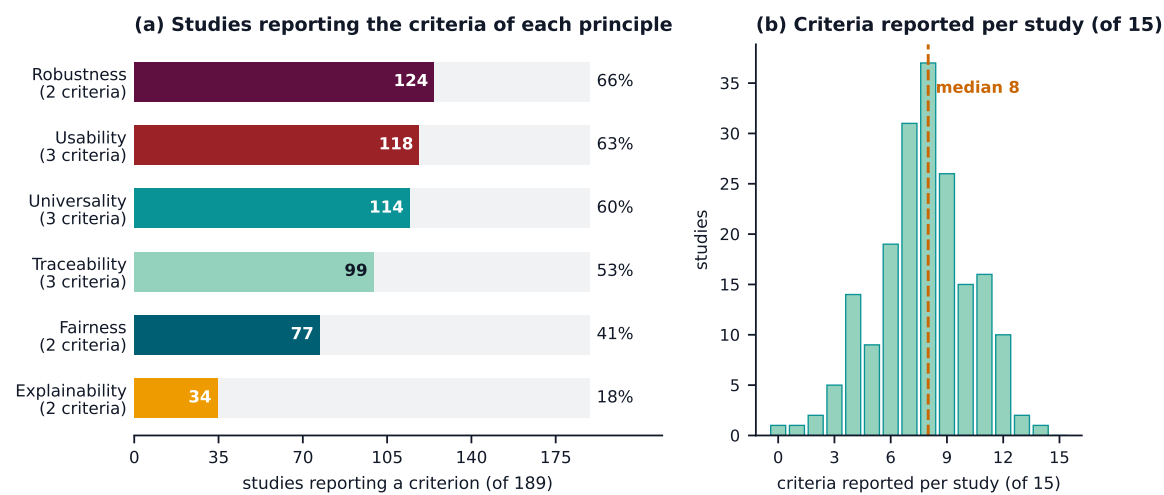


Figure 5. Reported trustworthy AI evidence across the 189 included studies. (a) Average reporting of the criteria within each FUTURE-AI principle. Each bar shows the mean number of studies reporting the criteria within that principle, with the corresponding percentage of the 189 studies. These percentages do not represent the proportion of studies satisfying every criterion within a principle. (b) Distribution of the number of criteria reported per study, out of 15. Per-criterion results are provided in Supplementary Section S2.

Fairness:

Fairness was assessed through reporting of individual characteristics and subgroup performance. Characteristics of the individuals whose images were used, such as age, sex, or clinical characteristics, were reported in 58% of studies, and performance for at least one subgroup in 24%.

Universality:

Universality was assessed through resource adaptation, local involvement, and local adaptation followed by evaluation. Local involvement was reported in 77% of studies, resource adaptation in 57%, and local adaptation followed by evaluation in 48%. The last criterion includes fine-tuning a model pretrained on general-purpose images on local data. Fifteen studies reported the same model and metric on both their development data and held-out data from the low-resource setting they name (Figure 6). Twelve reported lower performance in the local evaluation, and three reported performance that was unchanged or higher. Only one study reported directly comparable results before and after a local adjustment [90]. Every other study that adapted locally reported only the adapted result, or values that are not comparable (Supplementary Section S5).

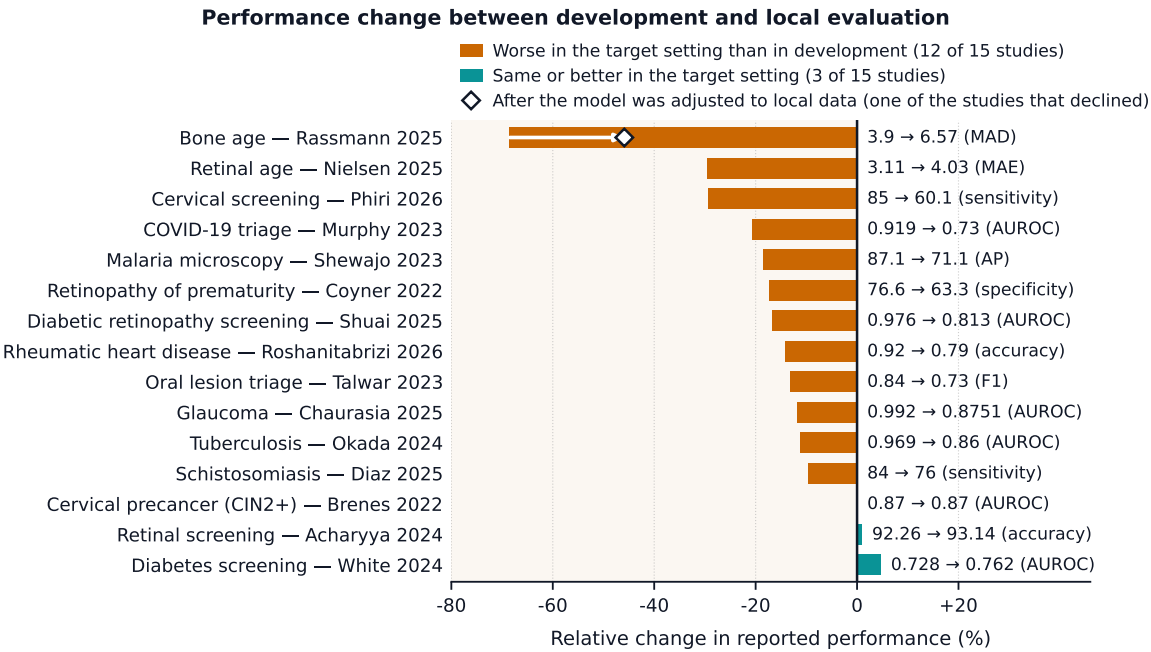


Figure 6. Change in reported performance between development and local evaluation for the 15 studies reporting the same model and metric in both [48,78,82,86,90–100]. Each bar shows the change relative to the study’s development value, with the two reported values shown beside it. Error metrics are signed so that a negative change always indicates worse performance. Because the studies use different performance metrics, these changes are not pooled and should not be interpreted as comparable clinical effects. On the top row, the arrow and diamond indicate performance after population-specific adjustment, the only study reporting directly comparable values before and after local adjustment [90]. AUROC, area under the receiver operating characteristic curve; AP, average precision; F1, macro-averaged F1 score; MAD, mean absolute difference in months; MAE, mean absolute error in years.

Traceability:

Traceability was assessed through method documentation, data provenance, and public availability of research materials. Method documentation was reported in 67% of studies and data provenance in 58%. A working public link to code, model weights, or a dataset was provided in 33%, a further 43 studies offered materials only on request, and 84 provided none.

Usability:

Usability was assessed through reporting of the intended user, the intended use, and evaluation with intended users or under realistic conditions. The intended user was identified in 80% of studies and the intended use in 83%. Evaluation with intended users or under realistic conditions was reported in 25%.

Robustness:

Robustness was assessed through the representativeness of training data and evaluation on independent data or under relevant variation. Representative training data were reported in 74% of studies and robustness evaluation in 58%. Of the 109 studies reporting robustness evaluation, 53 also reported local adaptation followed by evaluation.

Explainability:

Explainability was assessed through the provision of model explanations and their evaluation with intended users. An explanation, usually a saliency or attention map, was reported in 33% of studies, and an evaluation of the explanation with intended users in 4%. Explainability had the lowest average reporting of the six principles, at 18%.

7. Discussion

This review examined how medical imaging AI research addresses resource constraints and what evidence studies report for trustworthy AI in low-resource settings. The findings show that these constraints shape both imaging choices and model development. Portable and smartphone-based imaging address limitations in equipment and specialist access, while transfer learning, data augmentation, and efficient architectures address constraints in labelled data and computation. Clinical applications were also shaped by these constraints, with prominent areas combining high diagnostic need with portable or lower-cost imaging, including antenatal care, retinal disease, cervical screening, and parasitic diseases. The trustworthy AI assessment, adapted from the FUTURE-AI principles [21], shows a complementary pattern. Studies report how systems are developed more consistently than they report evidence of how those systems perform across the populations, devices, users, and conditions they target.

By examining these dimensions together, the review links resource constraints, methodological choices, clinical applications, and the evidence reported for trustworthy AI in the settings studied.

7.1. Constraints and Methods

Technical choices often followed the constraints identified by the studies. Transfer learning and data augmentation addressed limited labelled data, efficient architectures addressed computational constraints, and portable imaging addressed limited access to conventional equipment. AI-supported workflows also aimed to extend specialist tasks to non-specialist users. These approaches solve different parts of the problem, so no single method establishes suitability for a low-resource setting. A model can require little labelled data yet remain unsuitable for available hardware, while a lightweight architecture does not by itself demonstrate acceptable latency, memory use, or performance on the device where it is expected to run.

Imaging choices can also change the technical problem. Some studies used lower-cost or portable modalities where conventional equipment was less accessible, including handheld cervical imaging and low-field MRI. These substitutions can alter image quality and acquisition conditions. A small number of studies addressed this directly during model development by simulating lower-quality acquisition conditions [49,50].

Evidence of computational suitability was less consistent. Some studies proposed systems for low-resource use while reporting models of 77–80 million parameters without on-device measurements [45, 46], and model compression was reported in only 3% of studies. Efficient architecture design and demonstrated hardware suitability are therefore different claims. The latter requires evidence such as target-device latency, memory requirements, and connectivity needs, consistent with recommendations to evaluate systems under conditions representative of use [21].

Pretraining also relied heavily on general-purpose models. Most transfer-learning studies used architectures pretrained on ImageNet [53], while medical-image pretraining was reported in only eight studies. Self-supervised learning was also reported in eight studies, despite its potential to make use of unlabelled local data [101]. No included study used RadImageNet [102]. This does not imply that medical-domain pretraining would necessarily improve performance, since its value depends on the modality, task, and similarity between the pretraining and target data. The limited use of medical-domain pretraining, self-supervised learning, and model compression also identifies opportunities for comparative methodological research. Future studies should test whether these approaches reduce requirements for labelled data, computation, or memory under representative low-resource conditions, instead of assuming that their theoretical advantages translate to a particular application.

7.2. Clinical Evaluation Across Applications

Clinical applications differed in how far studies had progressed from model development to evaluation in realistic clinical settings. Cervical screening and tuberculosis radiography more often

included evaluation with intended users or under realistic conditions, whereas malaria microscopy studies more often remained at model-development evaluation. Diabetic retinopathy studies more frequently reported evaluation on independent data or under relevant variation. Similar technical approaches, including transfer learning and efficient architectures, appeared across applications at very different stages of evaluation. These comparisons are based on relatively small application groups, but they suggest that differences in reported trustworthy AI evidence may partly reflect differences in evaluation stage.

Programme context may also influence how far evaluation progresses. Tuberculosis and cervical screening include studies conducted within organised screening or diagnostic programmes, where threshold selection, workflow integration, and product evaluation become relevant [103,104]. Malaria microscopy studies in this review were more often evaluated at an earlier stage. Differences in development stage, programme context, and available evaluation infrastructure are plausible explanations, although this review did not test them.

Antenatal ultrasound shows that intended-user evaluation need not be deferred until late-stage implementation. In a prospective study, novice operators acquired blind-sweep ultrasound and AI-supported measurements were compared with standard biometry performed by expert sonographers [76]. When the intended user is central to the system design, user and workflow evaluation can therefore be incorporated earlier. This is consistent with DECIDE-AI, which distinguishes pre-clinical model evaluation from early live clinical evaluation, where human factors and safety become increasingly important [105].

7.3. Local Performance

Robustness evaluation and local adaptation address related but different questions. Evaluation on independent data or under relevant variation was reported in 58% of studies, while 48% reported local adaptation followed by evaluation. Independent evaluation can show whether performance holds beyond the development data, but it does not necessarily test the population, device, operator, or acquisition conditions of the setting the system targets. Local adaptation, in turn, shows that a model was adjusted to local data, but its value can only be judged if performance is then evaluated in that setting. The common requirement is therefore representative local evaluation.

Local adaptation should not be treated as necessary for every system. A model that performs adequately on representative local data may require no adjustment. Where performance does change, adaptation can range from recalibrating a decision threshold to fine-tuning model parameters. The World Health Organization, for example, recommends threshold selection for computer-aided tuberculosis detection according to intended use [104], while fine-tuning has improved performance in some external populations [106]. The relevant question is therefore not whether a model was adapted, but whether local evaluation showed that adaptation was needed and whether the adjustment improved performance.

The within-study comparison of development and local performance reinforces the importance of local evaluation (Figure 6). In 12 of the 15 studies, performance was lower locally, showing that performance can change across populations, devices, sites, and acquisition conditions. Comparable before-and-after evaluations of local adjustment were almost absent, so direct evidence of how much adaptation can recover remains scarce.

7.4. Research Priorities

The findings point to four priorities for future research in low-resource settings. These priorities concern representative evaluation, intended-user assessment, demonstrated computational suitability, and reproducible and reusable research practices. They should be interpreted in relation to the development stage and intended use of each system, since early model-development studies cannot be expected to provide the same evidence as prospective or implementation-stage evaluations.

First, future studies should evaluate models on data that represent the populations and conditions in which they are intended to be used. This includes subgroup analysis where relevant and comparison

before and after local adjustment when adaptation is performed. Multi-institutional, locally governed data collaborations could support larger and more representative datasets, independent evaluation, and analyses across relevant population groups [18].

Second, intended users should be incorporated more consistently into evaluation. Studies usually identified who was expected to use a system, but far fewer evaluated systems with those users or under realistic working conditions. Explainability showed the same imbalance. Explanations were often presented but rarely evaluated with the people expected to interpret them. This gap has also been reported across medical imaging AI more broadly [107]. Studies that incorporated intended users from the outset show how technical evaluation can be combined with evaluation by those users [76] and with assessment of usability [108].

Third, claims of suitability for resource-constrained settings should be supported by measurements on the hardware and infrastructure for which a system is proposed. Future work should therefore include target-device testing, memory and latency measurements, and connectivity requirements where these are relevant to the intended use.

Fourth, future research should make methods and research resources easier to reproduce, assess, and reuse. This includes clear documentation of methods and data provenance, release of code or model weights where appropriate, and explicit statements about access conditions when data cannot be shared openly. Openness is not equivalent to reproducibility, and unrestricted sharing is not always appropriate for sensitive clinical data. Clear access conditions can nevertheless support independent evaluation while respecting governance constraints [18,109].

These priorities sit within broader requirements for sustainable implementation. Financial sustainability depends on whether systems can be procured, maintained, supported, and integrated within local service capacity. Governance and regulation determine how data can be used, how systems are overseen, and how responsibility is assigned [26,27]. These dimensions were not systematically assessed in this review, but they affect whether systems with encouraging technical and clinical evaluation results can be sustained in low-resource health systems. Table 2 brings these research priorities together with the broader implementation considerations and identifies the stakeholders most directly involved.

7.5. Limitations

Two limitations are particularly important for interpreting these findings. They relate to the search and study selection, and to the operational definitions used in the trustworthy AI assessment.

The search and selection strategy may have affected the composition of the included literature. The review was limited to English-language records indexed in PubMed, Scopus, and IEEE Xplore and required low-resource terminology in the title or abstract. Studies framed only around a country, clinical programme, or specific technology may therefore have been missed. Restricting journals and conferences to ranked venues also excluded some workshops and biomedical-imaging proceedings that might otherwise have met the eligibility criteria. These choices may have affected both the geographic and methodological distribution of the included studies and the reporting patterns observed. LLM-assisted title and abstract screening was calibrated against reviewer decisions, but some eligible records may nevertheless have been excluded. The specific search terms for each database and screening evaluation are provided in Supplementary Sections S1 and S6.

Table 2. Priority gaps and recommended actions for medical imaging AI in low-resource settings. The first four domains follow directly from the findings of this review. Financial sustainability and governance and regulation are broader implementation considerations identified in the wider literature and were not systematically assessed in the trustworthy AI assessment adapted from FUTURE-AI [21]. Abbreviations: LMIC, low- and middle-income country; LRS, low-resource setting.

Domain	Priority gap	Recommended action	Primary stakeholders
Representative local evaluation	Local data are often available, but subgroup evaluation and comparison before and after local adjustment are rarely reported. Research activity is also concentrated in a limited number of countries and clinical areas.	Evaluate models on data that represent the target population, device, and care setting. Report subgroup performance where relevant and assess the effect of local adjustment when it is performed. Support multi-institutional, locally led data collaborations and research in under-represented settings.	Researchers, governments, data custodians, funders
Intended-user evaluation	Intended users are usually identified, but systems and model explanations are much less often evaluated with those users or under realistic working conditions.	Involve intended users in evaluation where appropriate, including assessment of usability, workflow fit, and whether model explanations are understandable and useful.	Researchers, end-users, communities, health-system managers
Resource suitability	Systems proposed for mobile, portable, or resource-constrained use often use methods intended to reduce labelled-data, memory, or computational requirements, but these gains are not always demonstrated under the data, hardware, and infrastructure constraints they are intended to address.	Develop and compare methods that reduce data, memory, and computational requirements, and evaluate these gains under representative hardware and infrastructure constraints. Where resource efficiency is claimed, report target-device latency, memory use, and connectivity requirements.	Researchers, industry, implementation partners, journals
Reproducible and reusable research	Methods are more consistently described than the availability of code, models, data, and their access conditions.	Document methods clearly, state where data came from and how they were collected, and release code, model weights, or datasets where appropriate. When materials cannot be shared openly, explain the access conditions and restrictions.	Researchers, journals, data custodians, ethics and governance bodies
Financial sustainability	Long-term use depends on affordability, procurement, maintenance, licensing, connectivity, and technical support, but these factors are often outside technical model evaluation.	Include realistic assumptions about costs, procurement, maintenance, and support when proposing implementation or sustained use in low-resource health systems.	Governments, funders, industry, procurement agencies
Governance and regulation	Long-term use also depends on clear rules for data use, accountability, independent evaluation, and ongoing oversight of deployed systems.	Strengthen regulatory and evaluation capacity, define responsibilities for data governance and system oversight, and establish mechanisms for independent evaluation and ongoing performance monitoring after deployment.	Regulators, governments, international bodies, funders, industry

The trustworthy AI assessment also depends on the operational definitions used in this review. Adapting the FUTURE-AI principles to low-resource settings required judgement, particularly for criteria such as robustness, resource-appropriate design, and local adaptation, and agreement between reviewers (Cohen’s κ) was lower for these criteria than for directly observable features such as a named intended user or publicly available resources. The reported percentages should therefore be interpreted as estimates under the definitions in Table 1, with detailed agreement and sensitivity analyses provided in Supplementary Section S3. The assessment was based on the published record. Evaluations that may have been performed but were not reported were scored as not reported. The scores therefore describe documented evidence, not all evaluation activities that may have occurred. No separate risk-of-bias assessment was conducted, so the review does not establish the validity of individual performance estimates.

8. Conclusion

This systematic review synthesised 189 studies of AI for medical image analysis in low-resource settings, covering resource constraints, technical approaches, clinical applications, and reported trustworthy AI evidence. The literature shows strong technical activity, particularly in portable imaging, transfer learning, data augmentation, and efficient model design. However, studies more consistently described how systems were developed than how they performed across the populations, devices, users, and conditions they target.

Future research should prioritise representative local evaluation, intended-user assessment, demonstrated resource suitability, and reproducible research practices. These priorities should be matched to the development stage and intended use of each system, alongside broader considerations of financial sustainability, governance, and regulation.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on [Preprints.org](https://preprints.org).

Author Contributions: Conceptualization, Methodology, Visualization: Tanya Akumu, Victor M. Campello. **Data curation:** Tanya Akumu, Victor M. Campello, Lidia Garrucho, Vien Ngoc Dang, Thomas Keller, Xue Zhao. **Formal analysis:** Tanya Akumu, Victor M. Campello, Lidia Garrucho, Vien Ngoc Dang, Thomas Keller, Xue Zhao. **Writing – original draft:** Tanya Akumu, Victor M. Campello, Lidia Garrucho, Emmah Ng'ang'a. **Writing – review & editing:** Tanya Akumu, Victor M. Campello, Lidia Garrucho, María J. Ledesma-Carbayo, Marius G. Linguraru, Karim Lekadir. **Funding acquisition, Supervision:** Karim Lekadir.

Data Availability Statement: The scoring sheet for the fifteen items with its scoring conventions, the screening prompts and few-shot calibration examples used for the language-model-assisted screening in this review are released at <https://doi.org/10.5281/zenodo.23091732>. The review analyses published studies, therefore no new patient data were collected.

Acknowledgments: The authors thank Bishesh Khanal and Richard Osuala for their helpful comments on an earlier version of the manuscript. This study was funded by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (AIMIX project, Grant Agreement No. 101044779). The funder had no role in the study design, data collection, analysis or interpretation, preparation of the manuscript, or decision to submit the article for publication.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Generative AI and AI-Assisted Technologies in the Manuscript Preparation Process

During the preparation of this work, the authors used Claude (Anthropic) to assist with coding for the review analyses and ChatGPT (OpenAI) to assist with language editing and refinement of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Wahl, B.; Cossy-Gantner, A.; Germann, S.; Schwalbe, N.R. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob. Health* **2018**, *3*, e000798. <https://doi.org/10.1136/bmjgh-2018-000798>.
2. Anazodo, U.C.; Adewole, M.; Dako, F. AI for population and global health in radiology. *Radiol. Artif. Intell.* **2022**, *4*, e220107. <https://doi.org/10.1148/ryai.220107>.
3. Mollura, D.J.; Culp, M.P.; Pollack, E.; Battino, G.; Scheel, J.R.; Mango, V.L.; Elahi, A.; Schweitzer, A.; Dako, F. Artificial Intelligence in Low- and Middle-Income Countries: Innovating Global Health Radiology. *Radiology* **2020**, *297*, 513–520. <https://doi.org/10.1148/radiol.2020201434>.
4. Schwalbe, N.; Wahl, B. Artificial intelligence and the future of global health. *Lancet* **2020**, *395*, 1579–1586. [https://doi.org/10.1016/S0140-6736\(20\)30226-9](https://doi.org/10.1016/S0140-6736(20)30226-9).
5. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>.
6. Haskins, G.; Kruger, U.; Yan, P. Deep learning in medical image registration: a survey. *Mach. Vis. Appl.* **2020**, *31*, 8. <https://doi.org/10.1007/s00138-020-01060-x>.
7. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Deep learning for cardiac image segmentation: a review. *Front. Cardiovasc. Med.* **2020**, *7*, 25. <https://doi.org/10.3389/fcvm.2020.00025>.
8. van Zyl, C.; Badenhurst, M.; Hanekom, S.; Heine, M. Unravelling 'low-resource settings': a systematic scoping review with qualitative content analysis. *BMJ Glob. Health* **2021**, *6*, e005190. <https://doi.org/10.1136/bmjgh-2021-005190>.
9. World Bank. World Bank country and lending groups: FY2025, 2024. Classifications updated annually on 1 July; accessed 2026-01-20.
10. Guo, J.; Li, B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. *Health Equity* **2018**, *2*, 174–181. <https://doi.org/10.1089/hec.2018.0037>.

11. Frija, G.; Blažić, I.; Frush, D.P.; Hierath, M.; Kawooya, M.; Donoso-Bach, L.; Brkljačić, B. How to Improve Access to Medical Imaging in Low- and Middle-Income Countries ? *eClinicalMedicine* **2021**, *38*. <https://doi.org/10.1016/j.eclinm.2021.101034>.
12. Hricak, H.; Abdel-Wahab, M.; Atun, R.; Lette, M.M.; Paez, D.; Brink, J.A.; Donoso-Bach, L.; Frija, G.; Hierath, M.; Holmberg, O.; et al. Medical Imaging and Nuclear Medicine: A Lancet Oncology Commission. *Lancet Oncol.* **2021**, *22*, e136–e172. [https://doi.org/10.1016/S1470-2045\(20\)30751-8](https://doi.org/10.1016/S1470-2045(20)30751-8).
13. Maturana, C.R.; de Oliveira, A.D.; Nadal, S.; Bilalli, B.; Serrat, F.Z.; Soley, M.E.; Igual, E.S.; Bosch, M.; Lluch, A.V.; Abelló, A.; et al. Advances and challenges in automated malaria diagnosis using digital microscopy imaging with artificial intelligence tools: A review. *Front. Microbiol.* **2022**, *13*. <https://doi.org/10.3389/fmicb.2022.1006659>.
14. Adeoye, J.; Akinshipo, A.; Thomson, P.; Su, Y.X. Artificial intelligence-based prediction for cancer-related outcomes in Africa: Status and potential refinements. *J. Glob. Health* **2022**, *12*, 03017. <https://doi.org/10.7189/jogh.12.03017>.
15. Umer, F.; Adnan, S.; Lal, A. Research and application of artificial intelligence in dentistry from lower-middle income countries – a scoping review. *BMC Oral Health* **2024**, *24*, 220. <https://doi.org/10.1186/s12903-024-03970-y>.
16. Etori, N.; Temesgen, E.; Gini, M. What We Know So Far: Artificial Intelligence in African Healthcare. arXiv preprint, 2023, [2305.18302]. <https://doi.org/10.48550/arXiv.2305.18302>.
17. Kim, S.; Fischetti, C.; Guy, M.; Hsu, E.; Fox, J.; Young, S.D. Artificial Intelligence (AI) Applications for Point of Care Ultrasound (POCUS) in Low-Resource Settings: A Scoping Review. *Diagnostics* **2024**, *14*, 1669. <https://doi.org/10.3390/diagnostics14151669>.
18. Ciecierski-Holmes, T.; Singh, R.; Axt, M.; Brenner, S.; Barteit, S. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *npj Digit. Med.* **2022**, *5*, 162. <https://doi.org/10.1038/s41746-022-00700-y>.
19. Marey, A.; Ambrozaite, O.; Afifi, A.; Agarwal, R.; Chellappa, R.; Adeleke, S.; Umair, M. A perspective on AI implementation in medical imaging in LMICs: challenges, priorities, and strategies. *Eur. Radiol.* **2025**, *36*, 2591–2602. <https://doi.org/10.1007/s00330-025-12031-z>.
20. Villamarín Marrugo, J.J.; Naranjo Piñeros, J.M.; Hernandez Rincon, E.H. Evidence on the utility of artificial intelligence in the interpretation of diagnostic radiological images in low and middle-income countries: A scoping review. *Acad. Radiol.* **2026**, *33*, 726–758. <https://doi.org/10.1016/j.acra.2025.11.012>.
21. Lekadir, K.; Frangi, A.F.; Porras, A.R.; Glocker, B.; Cintas, C.; Langlotz, C.P.; Weicken, E.; Asselbergs, F.W.; Prior, F.; Collins, G.S.; et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* **2025**, *388*, e081554. <https://doi.org/10.1136/bmj-2024-081554>.
22. Al-Ganad, A.; Al-Shahdhi, A.; Al-Dhaifi, O.; Hajeb, E.; Hajeb, H.; Al-Motarreb, A. Deploying medical AI in low-resource settings: a scoping review of challenges and strategies. *Front. Digit. Health* **2026**, *8*, 1743634. <https://doi.org/10.3389/fdgth.2026.1743634>.
23. Becker, D.M.; Tafoya, C.A.; Becker, S.L.; Kruger, G.H.; Tafoya, M.J.; Becker, T.K. The use of portable ultrasound devices in low- and middle-income countries: a systematic review of the literature. *Trop. Med. Int. Health* **2016**, *21*, 294–311. <https://doi.org/10.1111/tmi.12657>.
24. Carayon, P.; Wooldridge, A.; Hose, B.Z.; Salwei, M.; Benneyan, J. Challenges and opportunities for improving patient safety through human factors and systems engineering. *Health Aff.* **2018**, *37*, 1862–1869. <https://doi.org/10.1377/hlthaff.2018.0723>.
25. Luntsi, G.; Ugwu, A.C.; Nkubli, F.B.; Emmanuel, R.; Ochie, K.; Nwobi, C.I. Achieving universal access to obstetric ultrasound in resource constrained settings: a narrative review. *Radiography* **2021**, *27*, 709–715. <https://doi.org/10.1016/j.radi.2020.10.010>.
26. Greenhalgh, T.; Wherton, J.; Papoutsis, C.; Lynch, J.; Hughes, G.; A’Court, C.; Hinder, S.; Fahy, N.; Procter, R.; Shaw, S. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J. Med. Internet Res.* **2017**, *19*, e367. <https://doi.org/10.2196/jmir.8775>.
27. World Health Organization. Regulatory considerations on artificial intelligence for health. Technical report, World Health Organization, Geneva, 2023. Accessed: 2026-09-27.
28. Atkinson, L.Z.; Cipriani, A. How to Carry out a Literature Search for a Systematic Review: A Practical Guide. *BJPsych Adv.* **2018**, *24*, 74–82. <https://doi.org/10.1192/bja.2017.3>.

29. Bramer, W.M.; de Jonge, G.B.; Rethlefsen, M.L.; Mast, F.; Kleijnen, J. A Systematic Approach to Searching: An Efficient and Complete Method to Develop Literature Searches. *J. Med. Libr. Assoc.* **2018**, *106*, 531–541. <https://doi.org/10.5195/jmla.2018.283>.

30. SCImago Lab. SCImago Journal Rank (SJR) indicator, 2024 edition. <https://www.scimagojr.com/journalrank.php>, 2024. Accessed: 2026-01-20.

31. CORE. CORE Rankings Portal. <https://www.core.edu.au/conference-portal>, 2023. Accessed: 2024-06-30.

32. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. <https://doi.org/10.1136/bmj.n71>.

33. Grattafiori, A.; Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; et al. The Llama 3 Herd of Models. arXiv preprint, 2024, [2407.21783]. <https://doi.org/10.48550/arXiv.2407.21783>.

34. Gemma Team. Gemma 2: Improving Open Language Models at a Practical Size. arXiv preprint, 2024, [2408.00118]. <https://doi.org/10.48550/arXiv.2408.00118>.

35. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint, 2017, [1704.04861]. <https://doi.org/10.48550/arXiv.1704.04861>.

36. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>.

37. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>.

38. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6848–6856. <https://doi.org/10.1109/CVPR.2018.00716>.

39. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019, pp. 6105–6114.

40. Tan, M.; Le, Q. EfficientNetV2: Smaller Models and Faster Training. In Proceedings of the Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021, pp. 10096–10106.

41. Mehta, S.; Rastegari, M. MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer. International Conference on Learning Representations (ICLR 2022), 2022.

42. Song, B.; Li, S.; Sunny, S.; Gurushanth, K.; Mendonca, P.; Mukhia, N.; Patrick, S.; Peterson, T.; Gurudath, S.; Raghavan, S.; et al. Exploring uncertainty measures in convolutional neural network for semantic segmentation of oral cancer images. *J. Biomed. Opt.* **2022**, *27*. <https://doi.org/10.1117/1.JBO.27.11.115001>.

43. Schilpzand, M.; Neff, C.; Van Dillen, J.; Van Ginneken, B.; Heskes, T.; De Korte, C.; Van Den Heuvel, T. Automatic Placenta Localization From Ultrasound Imaging in a Resource-Limited Setting Using a Predefined Ultrasound Acquisition Protocol and Deep Learning. *Ultrasound Med. Biol.* **2022**, *48*, 663–674. <https://doi.org/10.1016/j.ultrasmedbio.2021.12.006>.

44. Gomes, R.G.; Vwalika, B.; Lee, C.; Willis, A.; Sieniek, M.; Price, J.T.; Chen, C.; Kasaro, M.P.; Taylor, J.A.; Stringer, E.M.; et al. A mobile-optimized artificial intelligence system for gestational age and fetal malpresentation assessment. *Commun. Med.* **2022**, *2*, 128. <https://doi.org/10.1038/s43856-022-00194-5>.

45. Park, S.y.; Ayana, G.; Wako, B.D.; Jeong, K.C.; Yoon, S.D.; Choe, S.w. Vision Transformers for Low-Quality Histopathological Images: A Case Study on Squamous Cell Carcinoma Margin Classification. *Diagnostics* **2025**, *15*, 260. <https://doi.org/10.3390/diagnostics15030260>.

46. Mousa, A.; Safwat, A.; Elgohr, A.T.; Elhadidy, M.S.; Abdelfatah, R.I.; Kasem, H.M. Attention-Enhanced CNNs and transformers for accurate monkeypox and skin disease detection. *Sci. Rep.* **2025**, *15*, 32924. <https://doi.org/10.1038/s41598-025-12216-y>.

47. Wang, F.; Whelan, K.; Silvestre, G.; Curran, K.M. Generative Diffusion Model Bootstraps Zero-Shot Classification of Fetal Ultrasound Images in Underrepresented African Populations. In *Perinatal, Preterm and Paediatric Image Analysis*; Link-Sourani, D.; Abaci Turk, E.; Macgowan, C.; Hutter, J.; Melbourne, A.; Licandro, R., Eds.; Springer Nature Switzerland: Cham, 2025; Vol. 14747, pp. 143–154. https://doi.org/10.1007/978-3-031-73260-7_13.

48. Shuai, Z.; Wu, C.; Tang, Z.; Restrepo, D.; Morley, M.; Nakayama, L.F. Enhancing AI-based diabetic retinopathy screening in low- and middle-income countries with synthetic data. *J. Biomed. Inform.* **2025**, *172*, 104938. <https://doi.org/10.1016/j.jbi.2025.104938>.

49. Lin, H.; Figini, M.; D'Arco, F.; Ogbole, G.; Tanno, R.; Blumberg, S.B.; Ronan, L.; Brown, B.J.; Carmichael, D.W.; Lagunju, I.; et al. Low-field magnetic resonance image enhancement via stochastic image quality transfer. *Med. Image Anal.* **2023**, *87*, 102807. <https://doi.org/10.1016/j.media.2023.102807>.
50. Wu, Z.; Wu, Q.; Fang, W.; Ou, W.; Wang, Q.; Zhang, L.; Chen, C.; Wang, Z.; Li, H. Harmonizing Unets: Attention Fusion module in cascaded-Unets for low-quality OCT image fluid segmentation. *Comput. Biol. Med.* **2024**, *183*, 109223. <https://doi.org/10.1016/j.compbiomed.2024.109223>.
51. Stegmüller, T.; Abbet, C.; Bozorgtabar, B.; Clarke, H.; Petignat, P.; Vassilakos, P.; Thiran, J.P. Self-supervised learning-based cervical cytology for the triage of HPV-positive women in resource-limited settings and low-data regime. *Comput. Biol. Med.* **2024**, *169*, 107809. <https://doi.org/10.1016/j.compbiomed.2023.107809>.
52. Giri, L.; Regmi, P.R.; Gurung, G.; Gurung, G.; Aryal, S.; Mandal, S.; Giri, S.; Chaulagain, S.; Acharya, S.; Umair, M. Diagnostic Accuracy of an Offline CNN Framework Utilizing Multi-View Chest X-Rays for Screening 14 Co-Occurring Communicable and Non-Communicable Diseases. *Diagnostics* **2025**, *16*, 66. <https://doi.org/10.3390/diagnostics16010066>.
53. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>.
54. Sendra-Balcells, C.; Campello, V.M.; Torrents-Barrena, J.; Ahmed, Y.A.; Elattar, M.; Ohene-Botwe, B.; Nyangulu, P.; Stones, W.; Ammar, M.; Benamer, L.N.; et al. Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five African countries. *Sci. Rep.* **2023**, *13*, 2728. <https://doi.org/10.1038/s41598-023-29490-3>.
55. Capellán-Martín, D.; Gómez-Valverde, J.J.; Sánchez-Jacob, R.; Hernanz-Lobo, A.; Schaaf, H.S.; García-Delgado, L.; Augusto, O.; Roshanitabrizi, P.; García-Basteiro, A.L.; Ribó, J.L.; et al. Multi-view deep learning framework for the detection of chest X-rays compatible with pediatric pulmonary tuberculosis. *Nat. Commun.* **2025**, *16*, 9170. <https://doi.org/10.1038/s41467-025-64391-1>.
56. Aldera, A.P.; Cifci, D.; Veldhuizen, G.P.; Tsai, W.J.; Pillay, K.; Boutall, A.; Brenner, H.; Hoffmeister, M.; Kather, J.N.; Ramesar, R. Deep learning predicts microsatellite instability status in colorectal carcinoma in an ethnically heterogeneous population in South Africa. *J. Clin. Pathol.* **2026**, *79*, 50–56. <https://doi.org/10.1136/jcp-2025-210053>.
57. Ellis, S.; Song, S.; Reiman, D.; Hui, X.; Zhang, R.; Shahriar, M.H.; Argos, M.; Kamal, M.; Shea, C.R.; Grossman, R.L.; et al. AI-assisted Diagnosis of Nonmelanoma Skin Cancer in Resource-Limited Settings. *Cancer Epidemiol. Biomarkers Prev.* **2025**, *34*, 1080–1088. <https://doi.org/10.1158/1055-9965.EPI-25-0132>.
58. Choopong, P.; Kusakunniran, W. Selection of pre-trained weights for transfer learning in automated cytomegalovirus retinitis classification. *Sci. Rep.* **2024**, *14*, 15899. <https://doi.org/10.1038/s41598-024-67121-7>.
59. Akhter, Y.; Ranjan, R.; Singh, R.; Vatsa, M. SHIELD: A Self-supervised, Silicosis-focused Hierarchical Imaging Framework for Occupational Lung Disease Diagnosis. In Proceedings of the Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, Montreal, Canada, 2025; pp. 9538–9546. <https://doi.org/10.24963/ijcai.2025/1060>.
60. Hinton, G.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network. arXiv preprint, 2015, [1503.02531]. <https://doi.org/10.48550/arXiv.1503.02531>.
61. Molchanov, P.; Tyree, S.; Karras, T.; Aila, T.; Kautz, J. Pruning Convolutional Neural Networks for Resource Efficient Inference. International Conference on Learning Representations (ICLR 2017), 2017.
62. Montalbo, F.J. Truncating fined-tuned vision-based models to lightweight deployable diagnostic tools for SARS-CoV-2 infected chest X-rays and CT-scans. *Multimed. Tools Appl.* **2022**, *81*, 16411–16439. <https://doi.org/10.1007/s11042-022-12484-0>.
63. Bhoopalan, R.; Sekar, P.; Nagaprasad, N.; Mamo, T.R.; Krishnaraj, R. Task optimized vision transformer for diabetic retinopathy detection and classification in resource constrained early diagnosis settings. *Sci. Rep.* **2025**, *15*, 39047. <https://doi.org/10.1038/s41598-025-25399-1>.
64. Yang, J.; Shen, X.; Xing, J.; Tian, X.; Li, H.; Deng, B.; Huang, J.; Hua, X.s. Quantization Networks. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7300–7308. <https://doi.org/10.1109/CVPR.2019.00748>.
65. Memon, K.; Yahya, N.; Yusoff, M.Z.; Remli, R.; Mustapha, A.W.M.M.; Hashim, H.; Ali, S.S.A.; Siddiqui, S. Edge Computing for AI-Based Brain MRI Applications: A Critical Evaluation of Real-Time Classification and Segmentation. *Sensors* **2024**, *24*, 7091. <https://doi.org/10.3390/s24217091>.

66. Lin, L.; Dacal, E.; Díez, N.; Carmona, C.; Martín Ramírez, A.; Barón Argos, L.; Bermejo-Peláez, D.; Caballero, C.; Cuadrado, D.; Darias-Plasencia, O.; et al. Edge artificial intelligence (AI) for real-time automatic quantification of filariasis in mobile microscopy. *PLoS Negl. Trop. Dis.* **2024**, *18*, e0012117. <https://doi.org/10.1371/journal.pntd.0012117>.
67. Oyibo, P.; Meulah, B.; Bengtson, M.; Van Lieshout, L.; Oyibo, W.; Diehl, J.C.; Vdovine, G.; Agbana, T. Two-stage automated diagnosis framework for urogenital schistosomiasis in microscopy images from low-resource settings. *J. Med. Imaging* **2023**, *10*. <https://doi.org/10.1117/1.JMI.10.4.044005>.
68. Tiyyarattanachai, T.; Apiparakoon, T.; Chaichuen, O.; Sukcharoen, S.; Yimsawad, S.; Jangsirikul, S.; Chaikajornwat, J.; Siri Wong, N.; Burana, C.; Siritaweechai, N.; et al. Artificial intelligence assists operators in real-time detection of focal liver lesions during ultrasound: A randomized controlled study. *Eur. J. Radiol.* **2023**, *165*, 110932. <https://doi.org/10.1016/j.ejrad.2023.110932>.
69. Berg, W.A.; López Aldrete, A.L.; Jairaj, A.; Ledesma Perea, J.C.; García, C.Y.; McClennan, R.C.; Cen, S.Y.; Larsen, L.H.; De Lara, M.T.S.; Love, S. Toward AI-supported US Triage of Women with Palpable Breast Lumps in a Low-Resource Setting. *Radiology* **2023**, *307*, e223351. <https://doi.org/10.1148/radiol.223351>.
70. Skerrett, E.; Miao, Z.; Asiedu, M.N.; Richards, M.; Crouch, B.; Sapiro, G.; Qiu, Q.; Ramanujam, N. Multicontrast Pocket Colposcopy Cervical Cancer Diagnostic Algorithm for Referral Populations. *BME Front.* **2022**, *2022*, 9823184. <https://doi.org/10.34133/2022/9823184>.
71. Weber, S.F.; Wolf, R.; Manten, K.; Thangakunam, B.; Isaac, B.; Shankar, D.; Mangal, D.; Dutta, A.K.; Vimala, L.R.; Irodi, A.; et al. Lung and abdominal ultrasound accuracy for tuberculosis: An Indian prospective cohort study. *PLoS One* **2025**, *20*, e0329670. <https://doi.org/10.1371/journal.pone.0329670>.
72. Tu, D.; Goyal, M.S.; Dworkin, J.D.; Kampondeni, S.; Vidal, L.; Biondo-Savin, E.; Juvvadi, S.; Raghavan, P.; Nicholas, J.; Chetcuti, K.; et al. Automated Analysis of Low-Field Brain MRI in Cerebral Malaria. *Biometrics* **2023**, *79*, 2417–2429. <https://doi.org/10.1111/biom.13708>.
73. Pokaparakarn, T.; Prieto, J.C.; Price, J.T.; Kasaro, M.P.; Sindano, N.; Shah, H.R.; Peterson, M.; Akapelwa, M.M.; Kapilya, F.M.; Sebastião, Y.V.; et al. AI Estimation of Gestational Age from Blind Ultrasound Sweeps in Low-Resource Settings. *NEJM Evid.* **2022**, *1*. <https://doi.org/10.1056/EVIDoA2100058>.
74. Akumu, T.; Elbatel, M.; Campello, V.M.; Osuala, R.; Martín-Isla, C.; Valenzuela, I.; Li, X.; Khanal, B.; Lekadir, K. Adaptive Frame Selection for Gestational Age Estimation from Blind Sweep Fetal Ultrasound Videos. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*; Gee, J.C.; Alexander, D.C.; Hong, J.; Iglesias, J.E.; Sudre, C.H.; Venkataraman, A.; Golland, P.; Kim, J.H.; Park, J., Eds.; Springer Nature Switzerland: Cham, 2026; Vol. 15973, pp. 3–12. https://doi.org/10.1007/978-3-032-05185-1_1.
75. Boumeridja, H.; Ammar, M.; Alzubaidi, M.; Mahmoudi, S.; Benamer, L.N.; Agus, M.; Househ, M.; Lekadir, K.; El Habib Daho, M. Enhancing fetal ultrasound image quality and anatomical plane recognition in low-resource settings using super-resolution models. *Sci. Rep.* **2025**, *15*, 8376. <https://doi.org/10.1038/s41598-025-91808-0>.
76. Stringer, J.S.A.; Pokaparakarn, T.; Prieto, J.C.; Vwalika, B.; Chari, S.V.; Sindano, N.; Freeman, B.L.; Sikapande, B.; Davis, N.M.; Sebastião, Y.V.; et al. Diagnostic Accuracy of an Integrated AI Tool to Estimate Gestational Age From Blind Ultrasound Sweeps. *JAMA* **2024**, *332*, 649. <https://doi.org/10.1001/jama.2024.10770>.
77. Hu, L.; Mwanahamuntu, M.H.; Sahasrabuddhe, V.V.; Barrett, C.; Horning, M.P.; Shah, I.; Laverriere, Z.; Banik, D.; Ji, Y.; Shibemba, A.L.; et al. Internal validation of Automated Visual Evaluation (AVE) on smartphone images for cervical cancer screening in a prospective study in Zambia. *Cancer Med.* **2024**, *13*, e7355. <https://doi.org/10.1002/cam4.7355>.
78. Phiri, H.; Uwinkindi, F.; Diop, M.; Mwanahamuntu, M.; Gondongwe, L.; Guzha, B.T.; Barrett, C.; Joseph, J.T.; Horning, M.P.; Anne, M.; et al. Performance of an artificial intelligence-based tool for cervical precancer screening in five countries in Africa: a prospective, observational, diagnostic accuracy study. *Lancet Glob. Health* **2026**, *14*, e49–e60. [https://doi.org/10.1016/S2214-109X\(25\)00352-3](https://doi.org/10.1016/S2214-109X(25)00352-3).
79. Birur N., P.; Song, B.; Sunny, S.P.; G., K.; Mendonca, P.; Mukhia, N.; Li, S.; Patrick, S.; G., S.; A.R., S.; et al. Field validation of deep learning based Point-of-Care device for early detection of oral malignant and potentially malignant disorders. *Sci. Rep.* **2022**, *12*, 14283. <https://doi.org/10.1038/s41598-022-18249-x>.
80. Guido, G.; Gulo, B.; Cotugno, S.; Nigussa, W.; Bobosha, K.; Segala, F.V.; Zauli, B.; Kenate Sori, B.; Putoto, G.; Cavallin, F.; et al. Improving Tuberculosis Diagnosis Through Artificial Intelligence (CAD4TB) and Stool Xpert MTB/RIF Testing: A Prospective Study From Oromia, Ethiopia. *Open Forum Infect. Dis.* **2025**, *13*, ofaf725. <https://doi.org/10.1093/ofid/ofaf725>.
81. Gupta, A.; Nair, A.M.; Ekka, S.; Gahwai, D.; Sharma, N.; Raza Khan, F.; Damani, M.; Kumar, S.; Pawar, S.; Antony Chiramal, J.; et al. Evaluating the Usefulness of Artificial Intelligence-based Chest X-Ray Screening

- in Improving Tuberculosis Detection Among the High-Risk Tribal Population of Chhattisgarh, India: A Prospective Multi-Centre Study. *Open Forum Infect. Dis.* **2025**, *13*, ofaf780. <https://doi.org/10.1093/ofid/ofaf780>.
82. Coyner, A.S.; Oh, M.A.; Shah, P.K.; Singh, P.; Ostmo, S.; Valikodath, N.G.; Cole, E.; Al-Khaled, T.; Bajimaya, S.; K.C., S.; et al. External Validation of a Retinopathy of Prematurity Screening Model Using Artificial Intelligence in 3 Low- and Middle-Income Populations. *JAMA Ophthalmol.* **2022**, *140*, 791. <https://doi.org/10.1001/jamaophthalmol.2022.2135>.
 83. Cole, E.; Valikodath, N.G.; Al-Khaled, T.; Bajimaya, S.; Kc, S.; Chuluunbat, T.; Munkhuu, B.; Jonas, K.E.; Chuluunkhuu, C.; MacKeen, L.D.; et al. Evaluation of an Artificial Intelligence System for Retinopathy of Prematurity Screening in Nepal and Mongolia. *Ophthalmol. Sci.* **2022**, *2*, 100165. <https://doi.org/10.1016/j.xops.2022.100165>.
 84. Korn Malerbi, F.; Barreto Melo, G. Feasibility of screening for diabetic retinopathy using artificial intelligence, Brazil. *Bull. World Health Organ.* **2022**, *100*, 643–647. <https://doi.org/10.2471/BLT.22.288580>.
 85. Duggal, M.; Chauhan, A.; Gupta, V.; Kankaria, A.; Budhija, D.; Verma, P.; Miglani, V.; Syal, P.; Kaur, G.; Kumar, L.; et al. Real-World Evaluation of AI-Driven Diabetic Retinopathy Screening in Public Health Settings: Validation and Implementation Study. *JMIR Med. Inform.* **2025**, *13*, e67529–e67529. <https://doi.org/10.2196/67529>.
 86. White, T.; Selvarajah, V.; Wolfhagen-Sand, F.; Svängård, N.; Mohankumar, G.; Fenici, P.; Rough, K.; Onyango, N.; Lyons, K.; Mack, C.; et al. Prediction of cardiovascular risk factors from retinal fundus photographs: Validation of a deep learning algorithm in a prospective non-interventional study in Kenya. *Diabetes Obes. Metab.* **2024**, *26*, 2722–2731. <https://doi.org/10.1111/dom.15587>.
 87. Abdurahman, F.; Fante, K.A.; Aliy, M. Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models. *BMC Bioinformatics* **2021**, *22*, 112. <https://doi.org/10.1186/s12859-021-04036-4>.
 88. Nguyen, H.G.; Nguyen, D.T.; Tran, T.S.; Ling, S.H.; Ho-Pham, L.T.; Van Nguyen, T. Artificial intelligence system for predicting areal bone mineral density from plain X-rays. *Osteoporos. Int.* **2025**, *36*, 2167–2176. <https://doi.org/10.1007/s00198-025-07634-7>.
 89. Ramos-Soto, O.; Ramos-Frutos, J.; Perez-Zarate, E.; Oliva, D.; Mousavirad, S.J.; Balderas-Mata, S.E. Non-invasive anemia detection from conjunctiva and sclera images using vision transformer with attention map explainability. *Sci. Rep.* **2025**, *15*, 44142. <https://doi.org/10.1038/s41598-025-32343-w>.
 90. Rassmann, S.; Abashishvili, L.; Melikidze, E.; Sukhiashvili, A.; Lartsuliani, M.; Chkhaidze, I.; Tskhvediani, N.; Gordeziani, T.; Kvaratskhelia, E.; Kheladze, N.; et al. Population-specific calibration and validation of an open-source bone age AI. *Sci. Rep.* **2025**, *15*, 32673. <https://doi.org/10.1038/s41598-025-20148-w>.
 91. Acharyya, M.; Moharana, B.; Jain, S.; Tandon, M. A double-blinded study for quantifiable assessment of the diagnostic accuracy of AI tool “ADVEN-i” in identifying diseased fundus images including diabetic retinopathy on a retrospective data. *Indian J. Ophthalmol.* **2024**, *72*, S46–S52. https://doi.org/10.4103/IJO.IJO_3342_22.
 92. Brenes, D.; Barberan, C.; Hunt, B.; Parra, S.G.; Salcedo, M.P.; Possati-Resende, J.C.; Cremer, M.L.; Castle, P.E.; Fregnani, J.H.; Maza, M.; et al. Multi-task network for automated analysis of high-resolution endomicroscopy images to detect cervical precancer and cancer. *Comput. Med. Imaging Graph.* **2022**, *97*, 102052. <https://doi.org/10.1016/j.compmedimag.2022.102052>.
 93. Chaurasia, A.K.; Liu, G.S.; Greatbatch, C.J.; Gharahkhani, P.; Craig, J.E.; Mackey, D.A.; MacGregor, S.; Hewitt, A.W. A generalised computer vision model for improved glaucoma screening using fundus images. *Eye* **2025**, *39*, 109–117. <https://doi.org/10.1038/s41433-024-03388-4>.
 94. Díaz De León Derby, M.; Delahunt, C.B.; Spencer, E.; Coulibaly, J.T.; Silué, K.D.; Bogoch, I.I.; Le Ny, A.L.; Fletcher, D.A. Multi-contrast machine learning improves schistosomiasis diagnostic performance. *PLoS Negl. Trop. Dis.* **2025**, *19*, e0012879. <https://doi.org/10.1371/journal.pntd.0012879>.
 95. Murphy, K.; Muhairwe, J.; Schalekamp, S.; Van Ginneken, B.; Ayakaka, I.; Mashate, K.; Katende, B.; Van Heerden, A.; Bosman, S.; Madonsela, T.; et al. COVID-19 screening in low resource settings using artificial intelligence for chest radiographs and point-of-care blood tests. *Sci. Rep.* **2023**, *13*, 19692. <https://doi.org/10.1038/s41598-023-46461-w>.
 96. Nielsen, C.; Wilms, M.; Forkert, N.D. The retinal age gap: an affordable and highly accessible biomarker for population-wide disease screening across the globe. *Proc. R. Soc. B* **2025**, *292*, 20242233. <https://doi.org/10.1098/rspb.2024.2233>.

97. Okada, K.; Yamada, N.; Takayanagi, K.; Hiasa, Y.; Kitamura, Y.; Hoshino, Y.; Hirao, S.; Yoshiyama, T.; Onozaki, I.; Kato, S. Applicability of artificial intelligence-based computer-aided detection (AI-CAD) for pulmonary tuberculosis to community-based active case finding. *Trop. Med. Health* **2024**, *52*, 2. <https://doi.org/10.1186/s41182-023-00560-6>.
98. Roshanitrabrizi, P.; Nath, V.; Brown, K.; Broudy, T.G.; Jiang, Z.; Parida, A.; Rwebembera, J.; Okello, E.; Beaton, A.; Roth, H.R.; et al. End-to-End Spatiotemporal Analysis of Color Doppler Echocardiograms: Application for Rheumatic Heart Disease Detection. *IEEE Trans. Med. Imaging* **2026**, *45*, 960–971. <https://doi.org/10.1109/TMI.2025.3615574>.
99. Shewajo, F.A.; Fante, K.A. Tile-based microscopic image processing for malaria screening using a deep learning approach. *BMC Med. Imaging* **2023**, *23*, 39. <https://doi.org/10.1186/s12880-023-00993-9>.
100. Talwar, V.; Singh, P.; Mukhia, N.; Shetty, A.; Birur, P.; Desai, K.M.; Sunkavalli, C.; Varma, K.S.; Sethuraman, R.; Jawahar, C.V.; et al. AI-Assisted Screening of Oral Potentially Malignant Disorders Using Smartphone-Based Photographic Images. *Cancers* **2023**, *15*, 4120. <https://doi.org/10.3390/cancers15164120>.
101. Ouyang, C.; Biffi, C.; Chen, C.; Kart, T.; Qiu, H.; Rueckert, D. Self-Supervised Learning for Few-Shot Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2022**, *41*, 1837–1848. <https://doi.org/10.1109/TMI.2022.3150682>.
102. Mei, X.; Liu, Z.; Robson, P.M.; Marinelli, B.; Huang, M.; Doshi, A.; Jacobi, A.; Cao, C.; Link, K.E.; Yang, T.; et al. RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning. *Radiol. Artif. Intell.* **2022**, *4*, e210315. <https://doi.org/10.1148/ryai.210315>.
103. Qin, Z.Z.; Naheyan, T.; Ruhwald, M.; Denking, C.M.; Gelaw, S.; Nash, M.; Creswell, J.; Kik, S.V. A new resource on artificial intelligence powered computer automated detection software products for tuberculosis programmes and implementers. *Tuberculosis* **2021**, *127*, 102049. <https://doi.org/10.1016/j.tube.2020.102049>.
104. World Health Organization. Determining the local calibration of computer-assisted detection (CAD) thresholds and other parameters: a toolkit to support the effective use of CAD for TB screening. Technical report, World Health Organization, Geneva, 2021. Accessed: 2026-09-24.
105. Vasey, B.; Nagendran, M.; Campbell, B.; Clifton, D.A.; Collins, G.S.; Denaxas, S.; Denniston, A.K.; Faes, L.; Geerts, B.; Ibrahim, M.; et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat. Med.* **2022**, *28*, 924–933. <https://doi.org/10.1038/s41591-022-01772-9>.
106. Palmer, M.; Seddon, J.A.; van der Zalm, M.M.; Hesseling, A.C.; Goussard, P.; Schaaf, H.S.; Morrison, J.; van Ginneken, B.; Melendez, J.; Walters, E.; et al. Optimising computer aided detection to identify intra-thoracic tuberculosis on chest x-ray in South African children. *PLOS Glob. Public Health* **2023**, *3*, e0001799. <https://doi.org/10.1371/journal.pgph.0001799>.
107. Chen, H.; Gomez, C.; Huang, C.M.; Unberath, M. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. *npj Digit. Med.* **2022**, *5*, 156. <https://doi.org/10.1038/s41746-022-00699-2>.
108. Nhat, P.T.H.; Van Hao, N.; Tho, P.V.; Kerdegari, H.; Pisani, L.; Thu, L.N.M.; Phuong, L.T.; Duong, H.T.H.; Thuy, D.B.; McBride, A.; et al. Clinical benefit of AI-assisted lung ultrasound in a resource-limited intensive care unit. *Crit. Care* **2023**, *27*, 257. <https://doi.org/10.1186/s13054-023-04548-w>.
109. Tejani, A.S.; Klontzas, M.E.; Gatti, A.A.; Mongan, J.T.; Moy, L.; Park, S.H.; Kahn, Jr., C.E.; CLAIM 2024 Update Panel. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update. *Radiol. Artif. Intell.* **2024**, *6*, e240300. <https://doi.org/10.1148/ryai.240300>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.